



HAL
open science

Covariance matrices : diffusions, free probability, and deep learning

Ezéchiél Kahn

► **To cite this version:**

Ezéchiél Kahn. Covariance matrices : diffusions, free probability, and deep learning. Probability [math.PR]. Ecole des Ponts ParisTech, 2021. English. NNT : . tel-03531875

HAL Id: tel-03531875

<https://hal.science/tel-03531875v1>

Submitted on 18 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Matrices de covariance : diffusions, probabilités libres, et apprentissage profond

École doctorale N° 532, Mathématiques et sciences et technologies de l'information et de la communication (MSTIC)

Mathématiques

Thèse préparée au sein du CERMICS

Thèse soutenue le 05/07/21, par
Ezechiel KAHN

Composition du jury:

Mylène, MAÏDA
Professeure, Université de Lille

Présidente

Catherine, DONATI-MARTIN
Professeure, Université de Versailles-Saint-Quentin-en-Yvelines

Rapporteuse

Florent, BENAYCH-GEORGES
Directeur adjoint de la recherche, Capital Fund Management

Rapporteur

Nizar, DEMNI
Maître de Conférences, Aix-Marseille Université

Examineur

Reda, CHHAIBI
Maître de Conférences, Université Paul Sabatier

Examineur

Benjamin, JOURDAIN
Professeur, ENPC

Co-directeur de thèse

Djalil, CHAFAÏ
Professeur, Université Paris-Dauphine

Directeur de thèse

Remerciements (Acknowledgements)

Bien que mon manuscrit de thèse ne soit signé que par une seule personne, et bien que mes précédents diplômes, réussites personnelles ou professionnelles ne portent parfois qu'un nom, je suis heureux d'avoir l'occasion aujourd'hui de remercier toutes celles et ceux qui y ont participé -parfois sans s'en rendre compte, qui m'ont accompagné, qui m'ont soutenu pendant tout ou une partie de mon parcours, en particulier jusqu'au bout de ces trois années de recherche.

Mes premiers remerciements s'adressent évidemment à mes directeurs de thèse Benjamin Jourdain et Djalil Chafaï qui, avant même le début ce travail, m'ont fait confiance, et m'ont proposé un sujet à la frontière de leurs domaines de recherche. J'ai voyagé avec eux dans de magnifiques paysages mathématiques d'une façon qui m'était encore jusque-là étrangère, le voyage se transformant peu à peu en aventures, en expéditions, en explorations. Je suis très reconnaissant pour leur patience, pour leur rigueur, pour leur disponibilité et leur soutien. Merci pour toute la richesse de leur encadrement qui a été essentiel au bon déroulement de ma thèse.

Je tiens aussi à remercier sincèrement Catherine Donati-Martin et Florent Benaych-Georges pour l'intérêt qu'ils ont porté à mon travail en acceptant d'être les rapporteurs de ma thèse, et pour leur relecture minutieuse de mon mémoire. Je remercie Mylène Maïda d'avoir accepté de présider le jury de ma thèse. Je remercie également Nizar Demni de m'avoir fait l'honneur de participer à mon jury, et dont j'ai croisé les travaux tout au long de ces trois années.

Je remercie Reda Chhaïbi que j'ai eu la chance de rencontrer pendant ma deuxième année de thèse. Il a eu la gentillesse de partager avec moi ses géniales intuitions et son enthousiasme pour plusieurs de ses projets scientifiques. Nous avons partagé bien au-delà de notre sujet de travail. Ce fut et c'est encore aujourd'hui un immense plaisir de travailler et d'échanger avec lui. En outre, merci à lui d'avoir participé à mon jury de thèse.

Je suis très heureux de remercier ma famille qui m'a soutenu tout au long de ma vie. Merci à mes parents qui m'ont tant apporté qu'il serait impossible pour moi de le concevoir entièrement, et encore moins de le retranscrire ici. Merci en particulier à ma mère qui était présente à chaque moment de mon parcours. C'est grâce à eux que je suis là aujourd'hui et j'ai beaucoup de chance de les avoir à mes côtés. Merci à mes grands-mères Fernande et Christiane. Merci à mes tantes Babou et Lise. Merci à Robert, avec qui j'ai tant partagé. Enfin, merci Clara pour ton soutien indéfectible et pour la lumière que tu m'apportes au quotidien.

Je remercie également tous mes anciens professeurs, qui m'ont transmis le goût de leur matière et l'amour des sciences, je pense pour n'en citer que quelques-uns à Françoise Tavenard, Agnès Carbonell, Albert Djaoui, Elizabeth Altschull, Roger Muffat, Bernard Randé, et Louis Capéran. Je mesure la chance que j'ai eue de côtoyer des personnes si passionnées toutes ces années, cette thèse et le reste de mon parcours découle de leurs enseignements.

Je tiens à remercier également l'ensemble des chercheurs et employés du CERMICS, qui offrent un cadre privilégié à leurs thésards. Mention spéciale à Isabelle Simunic, secrétaire générale du CERMICS pour son travail exceptionnel.

Je remercie Michel pour tous les précieux moments que nous avons partagé. Il était un grand ami et je pense souvent à lui, et à ses conseils.

Enfin, merci à tous mes amis et à l'ensemble des doctorants du CERMICS : Adel, Adrien, Alex, Angelica, Antoine, Armand, Baptiste, Benjamin, Benoît, Bérengère, la team CFOA, Charlotte, Claire, Clément, Cyrille, Diane, Elise, François-Xavier, Gabriel, Gaspard, Grégoire, Guillaume, Guillaume, Ingrid, Jean-Claude, Jean-Maxime, Jérémie, Joachim, Joséphine, Jules, Julia, Juliette, Laurent, Léa, Louise, Lucie, Maël, Marc, Margaux, Margot, Marie, Marion, Martin, Martin, Max, Maximilien, Mélissa, Michel, Mouad, Oumaïma, Patrice, Paul, Paul, Pauline, Pierre, Prunelle, Quentin, Quentin, Mathis, Rachel, Rafaël, Raphaëla, Rémi, Rétox (oui c'est son nom), Samuel, Sarah, Sébastien, Sofiane, Sopheavy, Théo, Thibaud, Thomas, Thomas, Thomas, Toma, Tiphaine, Véronique, Viviane, Walid, William (huh huh huh), Yana, Yixin. Merci pour tous les bons moments passés ensemble, tous les rires, les délires, les discussions plus ou moins profondes, votre soutien.

Titre : Matrices de covariance : diffusions, probabilités libres, et apprentissage profond

Résumé : Cette thèse est motivée par l'étude des matrices de covariance, et s'articule naturellement en trois parties. Dans la première partie, nous étudions des modèles dynamiques liés aux matrices de covariance. Nous étudions plus précisément les systèmes d'équations différentielles stochastiques hérités de la dynamique des valeurs propres de processus matriciels appelés processus de Wishart et processus de Jacobi. Les solutions de ces systèmes sont respectivement les processus de β -Wishart et de β -Jacobi. Nous étendons les résultats connus d'existence et d'unicité de solutions à ces équations et caractérisons leur comportement en temps long. Dans la seconde partie, nous étudions à la lumière de résultats modernes de probabilités libres, en particulier sur la convolution libre rectangulaire additive, le comportement limite en grande dimension de la mesure empirique des particules constituant le processus de β -Wishart, et établissons la commutativité entre la limite en temps long et la limite en grande dimension pour cette suite de processus à valeurs mesures. La troisième partie porte sur l'étude de la stabilité de la procédure de rétro-propagation dans la phase d'apprentissage d'un réseau de neurones à propagation avant dont les couches sont de largeur variable. Cette étude se concentre sur la matrice Jacobiennne du réseau, qui peut se décrire comme un long de produit de matrices, et dont les propriétés spectrales sont déterminantes pour la stabilité de la descente du gradient. Cela demande la définition d'une opération de convolution libre rectangulaire multiplicative. Nous proposons un algorithme efficace de calcul de la mesure du carré des valeurs singulières de cette matrice. Les conclusions de ce travail donnent des indicateurs permettant d'évaluer la stabilité d'un réseau de neurones, à vocation d'aide au design du réseau pour les praticiens.

Mots-clefs : Matrices aléatoires, Calcul stochastique, Diffusions, Systèmes de particules, Probabilités libres, Apprentissage profond, Réseaux de neurones

Title : Covariance matrices : diffusions, free probability and deep learning

Abstract : This thesis is motivated by the study of covariance matrices, and is naturally structured in three parts. In the first part, we study dynamic models related to covariance matrices. More precisely, we study the systems of stochastic differential equations inherited from the dynamics of the eigenvalues of matrix valued processes named the Wishart process and the Jacobi process. The solutions to these systems are respectively the β -Wishart process and the β -Jacobi process. We extend the known results on the existence and uniqueness of solutions to these equations and we characterize their long term behaviour. In the second part, in the light of modern results from the free probability theory, especially about the rectangular additive free convolution, we study the behaviour of the empirical measure of the particles of the β -Wishart process in the large dimension limit, and establish the commutativity between the long time and the high dimensional limits for this sequence of measure valued process. The third part is related to the study of stability of the backpropagation procedure in the learning phase of a feed-forward neural network with a variable width profile. This work focuses on the Jacobian matrix of the network, which can be seen as a long product of matrices, and whose spectral properties are crucial for the stability of the gradient descent. This appeals for the definition of a free rectangular multiplicative convolution. We suggest an efficient algorithm to compute the empirical measure of the square of the singular values of this matrix. The conclusion of this work give insights allowing to assess the stability of a feed-forward neural network, aiming to assist practitioners in the choice of the features in the design of neural networks.

Keywords : Random matrices, Stochastic calculus, Diffusions, System of particles, Free probability, Deep learning, Neural networks

Contents

1	Introduction (in French)	9
1.1	Théorie des matrices aléatoires	9
1.1.1	Histoire	9
1.1.2	Applications	11
1.1.3	Résultats principaux et outils d'analyse	12
1.2	Matrices de Wishart et de Jacobi	18
1.2.1	Matrices de covariance empiriques et matrices de Wishart	18
1.2.2	Matrices de Wishart, ensembles de Jacobi et statistiques	21
1.3	Processus matriciels et gaz de particules	24
1.3.1	Systèmes de particules en interaction	26
1.3.2	Systèmes de particules en grande dimension	33
1.4	Matrices aléatoires et réseaux de neurones	35
1.4.1	Description du modèle	37
1.4.2	Contributions	39
1.5	Publications	42
2	Strong solutions to a beta-Wishart particle system	43
2.1	Introduction	43
2.2	Main results	48
2.3	Properties of the solutions	51
2.4	Proof of Theorem 2.2.2	56
2.5	Proof of the other results	71
2.6	Appendix	76
3	Strong solutions to a beta-Jacobi particle system	79
3.1	Introduction	79
3.2	Results	84
3.3	Properties of the solutions	86
3.4	Proof of Theorem 3.2.3	90
3.5	Appendix	112

4	Wishart processes : mean-field limit, long time behavior, and free probability	115
4.1	Introduction	115
4.2	The complex framework	120
4.3	Application to more general Wishart processes	121
4.4	Useful tools	124
4.4.1	The Cauchy-Stieltjes transform	124
4.4.2	Free convolution and rectangular free convolution from a random matrix point of view	125
4.5	Proofs	127
5	Newton lily pads and Jacobians of neural networks	137
5.1	Introduction	137
5.1.1	Contributions	140
5.1.2	Structure of the paper	141
5.2	Free Probability	141
5.2.1	Definitions and notations	141
5.2.2	Product of rectangular free matrices	143
5.3	Theoretical resolution of the model	143
5.3.1	The model	144
5.3.2	Resolution and formulas	145
5.3.3	Master equation	146
5.3.4	Moments of J	146
5.4	Numerical resolution	147
5.4.1	On the classical Newton-Raphson scheme	147
5.4.2	Kantorovich's criterion	148
5.4.3	Doubling strategies and chaining	149
5.5	Benchmarks	150
5.5.1	Computational speed	150
5.5.2	Plots	151
5.5.3	Description of the experiments	153
5.6	Conclusion	153
5.7	Acknowledgements	153
5.8	Supplementary material	154
5.8.1	Computations of moments	154
5.8.2	Specializing to classical weight distributions	155
5.8.3	Proofs	155

Chapter 1

Introduction (in French)

1.1 Théorie des matrices aléatoires

1.1.1 Histoire

Si la théorie des matrices aléatoires apparaît aujourd’hui dans de nombreux problèmes théoriques, ce sont bien les applications qui ont motivé sa genèse, et, comme on le montrera dans cette thèse, nombre de ses développements actuels. Il est d’usage de considérer son émergence en 1928 avec le travail fondateur [Wis28] du statisticien *John Wishart*, dont la problématique était d’analyser une population de données multivariées, et en particulier les causes communes à l’origine des variations de ces données. Cela le conduisit naturellement à l’étude du spectre de la matrice de variance covariance empirique des données (voir Chapitre 3).

En 1951, le physicien lauréat du prix Nobel *Eugène Wigner*, dans son travail de modélisation de noyaux lourds en mécanique quantique [Wig51, Wig55, Wig58], fit un pas décisif dans le développement de la théorie des matrices aléatoires. À l’époque, les modèles développés à partir de ceux fonctionnant sur des systèmes physiques plus simples échouaient à prédire correctement les niveaux d’énergie des noyaux lourds observés expérimentalement. Ces niveaux d’énergie étaient donnés par les valeurs propres de l’hamiltonien du système. Le raisonnement de Wigner fut le suivant : le système physique des noyaux lourds est trop compliqué pour être approximé par les modèles décrivant les systèmes plus simples. En revanche, on peut essayer de décrire ce système complexe en introduisant de l’aléa dans le modèle. Son idée fut donc de modéliser le système de noyaux lourds par le modèle le plus aléatoire possible tout en tenant compte des contraintes connues du système, telles que le caractère auto-adjoint de l’hamiltonien le décrivant. L’aléatoire fut donc utilisé comme souvent pour décrire ce que l’on ne connaît pas. Ainsi, l’opérateur hamiltonien du système, généralement défini sur un espace continu de dimension infinie, fut approximé par un opérateur discret en dimension finie : une grande matrice auto-ajointe. N’ayant pas plus d’informations sur les propriétés de l’hamiltonien du système, Wigner choisit de considérer un modèle simple où les entrées de la matrice sont des Gaussiennes indépendantes. La distribution des valeurs propres de ces matrices, en particulier la distance typique entre les valeurs propres, donna une descrip-

tion concordante avec les mesures expérimentales des niveaux d'énergie de certains systèmes physiques. Le physicien *Freeman Dyson* proposa alors la réflexion suivante : si un tel modèle aléatoire ne décrit pas correctement un système physique quantique, alors il y a probablement une contrainte du système physique se répercutant sur la structure de son hamiltonien, par exemple une relation de dépendance entre les entrées de la matrice, qui fut ignorée dans le modèle et qui doit être retrouvée et intégrée afin que le modèle fonctionne.

Ces deux premières motivations pour l'étude des grandes matrices aléatoires conduirent les chercheurs à se poser une question emblématique et structurante de la théorie : comment étudier le spectre des grandes matrices aléatoires quand leur taille tend vers l'infini ?

À la suite de la découverte de Wigner, Dyson, *Michel Gaudin* et *Madan Lal Mehta* étudièrent rigoureusement plusieurs ensembles aujourd'hui classiques de théorie des matrices aléatoires : les *Gaussian orthogonal ensembles (GOE)*, les *Gaussian unitary ensembles (GUE)*, auxquels vinrent s'ajouter les *Gaussian symplectic ensembles (GSE)* dont on donnera une définition Chapitre 1.1.3. Dyson et Mehta formulèrent la conjecture connue aujourd'hui sous le nom de conjecture d'universalité de Wigner-Dyson-Mehta (WDM), que l'on peut retrouver dans le livre [Lal67]. Elle affirme que la densité de probabilité jointe de k valeurs propres arbitraires de toute matrice symétrique (respectivement hermitienne et quaternionique auto-adjointe) à entrées i.i.d. converge en probabilité quand la taille de la matrice tend vers l'infini vers la fonction de corrélation à k points du GOE (respectivement du GUE et du GSE), et ne dépend ainsi pas de la distribution des éléments de la matrice.

Un des résultats les plus marquants de la théorie concernant la limite en grande dimension du spectre du symétrisé (de l'hermitisé) d'une matrice aléatoire rectangulaire à entrées i.i.d. fut prouvée par *Vladimir Marcenko* et *Leonid Pastur* en 1967 [MP67] et ouvrit la voie à de nouvelles applications.

Dans les années 1990, *Dan-Virgil Voiculescu* développa dans un contexte de théorie des algèbres d'opérateurs et d'analyse des algèbres de von Neumann la théorie des probabilités libres, une théorie des probabilités pour des variables aléatoires non commutatives munies d'une nouvelle notion d'indépendance nommée la liberté. Son observation fondamentale résidait dans le constat qu'équiper les algèbres de von Neumann d'une application de trace avait une forte ressemblance avec la notion d'espérance en probabilités, et que la nouvelle notion de liberté construite à partir de ces traces définissait une notion similaire à l'indépendance en théorie des probabilités classique. Cette théorie s'applique naturellement aux matrices aléatoires, dont le produit est non-commutatif, munies de l'espérance de la trace normalisée. On peut se référer à [Voi85, Voi86, Voi91, BV93, Voi87, BV92] et bien d'autres pour les articles historiques et aux livres [NS06, HP00, MS17] pour une introduction à cette théorie.

Le monde des matrices aléatoires est d'une grande richesse et de nombreux autres développements majeurs sont apparus, par exemple dans l'étude du comportement des vecteurs propres des grandes matrices aléatoires. Le lecteur de cette thèse pourra les retrouver dans de nombreux ouvrages tels que [TVV04, ABDF11], mais nous avons choisi de n'évoquer dans cette introduction que les aspects historiques proches de nos préoccupations lors de cette thèse.

Avant de présenter plus en détails les résultats fondateurs et les outils classiques de la théorie des matrices aléatoires, nous donnons quelques exemples des nombreuses applications de ce champ de recherche qui le rendent si séduisant.

1.1.2 Applications

Comme nous l'avons vu, les premiers pas du développement de la théorie des matrices aléatoires furent motivés par des applications en statistiques ainsi qu'en physique quantique. Les matrices aléatoires sont par nature la combinaison de l'objet linéaire simple par excellence que sont les matrices, utiles par exemple pour approcher des phénomènes non linéaires plus complexes (par exemple la dérivée ou la matrice jacobienne), et de la théorie des probabilités dont les domaines d'applications sont immenses, que l'on retrouve même pour mieux comprendre des problèmes déterministes (voir par exemple la méthode probabiliste de *Paul Erdős* [AS04]). On s'attend donc à que les domaines d'applications de la théorie des matrices aléatoires soit riches et variés, et c'est bien le cas, comme le suggèrent les quelques exemples suivants.

Statistique en grande dimension et machine learning : Sur les traces des travaux de Wishart en statistiques multivariées, le comportement du spectre des matrices aléatoires en grande dimension montre que les estimateurs statistiques classiques des matrices de covariance échouent à décrire précisément les relations de corrélations entre les caractéristiques des données. Ce régime de la grande dimension est courant aujourd'hui où règnent les *Big Data*, toujours plus nombreuses et surtout de grande dimension. Voir par exemple [BBP17, CL21].

Théorie des nombres : Un des problèmes encore non résolu des plus connus en mathématiques aujourd'hui est la fameuse hypothèse de Riemann, qui conjecture les zéros non-triviaux de la fonction zêta de Riemann sont tous de partie réelle égale à $\frac{1}{2}$. Hilbert et Pólya ont formulé l'idée que ces zéros non triviaux pourraient être les valeurs propres d'un opérateur auto-adjoint. *Hugh Montgomery* apporta une importante contribution à cette observation en 1972 en calculant la fonction de corrélation entre deux zéros non-triviaux et en montrant qu'ils se repoussaient de la même façon que se repoussent deux valeurs propres de grandes matrices aléatoires unitaires. De nombreux développements suivirent sur cette connexion. Voir par exemple [Mon73, BK13].

Théorie des graphes et applications : Un des outils centraux de la théorie des graphes est la matrice d'adjacence, qui encode les arrêtes entre les sommets du graphe. L'étude de graphes complexes et de systèmes dynamiques sur ces graphes, provenant par exemple de *neurosciences* pour modéliser les connexions synaptiques, d'*épidémiologie*, de réseaux sociaux et d'apparition de communautés, utilisent les matrices aléatoires dans leurs modélisations pour décrire les propriétés statistiques des objets étudiés.

Télécommunications : Les matrices aléatoires sont très utilisées dans les modèles de communication sans fil où elles caractérisent les canaux de communication. Par exemple, dans les modèles multi-antennes comme le CDMA, la matrice canal modélise la propagation de l'information entre chaque paire d'antennes émettrice et réceptrice. La performance du canal dépend alors du spectre de la matrice aléatoire en question. Voir par exemple [TVV04, CD11].

Théorie de l'information quantique : En théorie de l'information quantique, on s'intéresse à des états et canaux quantiques modélisés par des matrices complexes. Afin de comprendre leur comportement typique, on en étudie une version aléatoire. De plus, certains problèmes tels que la minimisation de l'entropie de sortie additive d'un système n'ont pas de réponse déterministe évidente aujourd'hui, alors que leur version aléatoire s'étudie plus facilement. Voir [Nec19, AS].

1.1.3 Résultats principaux et outils d'analyse

Nous introduisons ici des résultats sur les modèles les plus classiques de matrices aléatoires liés aux préoccupations de cette thèse, à savoir les matrices de Wigner ainsi que les matrices de (variance) covariance. Nous utiliserons tout au long de cette introduction les abréviations i.i.d. pour "indépendantes identiquement distribuées", p.s. pour "presque sûrement", p.p. pour "presque partout", et la notation $*$ pour opérateur adjoint.

Définition 1.1.1. Soit $n \geq 1$. Une *matrice de Wigner* $X = (X_{i,j})_{1 \leq i,j \leq n}$ est une matrice dont les familles d'entrées $\{X_{i,j}\}_{1 \leq i < j \leq n}$ ainsi que $\{X_{i,i}\}_{1 \leq i \leq n}$ sont des variables aléatoires réelles (respectivement complexes, quaternioniques) centrées, de seconds moments finis, i.i.d., et respectant la contrainte d'hermiticité $X_{i,j} = \bar{X}_{j,i}$ pour tout $1 \leq i \leq j \leq n$ (de matrice auto-adjointe dans le cas quaternionique).

Les matrices de Wigner à coefficients gaussiens furent les premières étudiées. Elles sont d'un intérêt capital comme le suggère la conjecture WDM qui prédit que la densité jointe de k valeurs propres arbitraires d'une matrice de Wigner converge en probabilité quand la taille de la matrice tend vers l'infini vers la fonction de corrélation à k points de l'ensemble gaussien orthogonal (GOE) quand les coefficients de la matrice sont des variables aléatoires à valeurs réelles, de l'ensemble gaussien unitaire (GUE) quand les coefficients de la matrice sont des variables aléatoires à valeurs complexes, et de l'ensemble gaussien symplectique (GSE) quand les coefficients de la matrice sont des variables aléatoires à valeurs quaternioniques. Ces ensembles sont invariants par l'action des groupes respectivement orthogonal, unitaire et symplectique. Les nombreux résultats et conjectures d'universalité dans la théorie des matrices aléatoires motivent ainsi à s'intéresser à des modèles pour lesquels on sait faire les calculs, afin de se construire une intuition sur les cas plus généraux. On associe à ces trois ensembles le coefficient β valant respectivement 1, 2 et 4 pour le GOE, le GUE et le GSE. Ces ensembles se définissent par les matrices de Wigner telles que $X_{1,1}$ suit la loi $\mathcal{N}(0, 1)$ et $X_{1,2}$ suit la loi $\mathcal{N}_\beta\left(0, \frac{1}{2}I_\beta\right)$ où I_β est la matrice identité de taille $\beta \times \beta$. L'importance de ces modèles réside donc en partie dans le fait que l'on peut calculer explicitement la loi jointe des valeurs propres, qui a une densité par rapport à la mesure de Lebesgue sur \mathbb{R}^n proportionnelle à

$$(\lambda^1, \dots, \lambda^n) \mapsto \prod_{1 \leq i < j \leq n} |\lambda^i - \lambda^j|^\beta \exp\left(-\frac{\beta}{2} \sum_{j=1}^n (\lambda^j)^2\right).$$

Nous avons jusqu'à présent introduit le concept de matrices aléatoires comme des matrices dont les entrées sont des variables aléatoires. Évidemment, l'objet matrices aléatoires est plus

général, et se conçoit comme une variable aléatoire à valeur matricielle, se définissant donc à l'aide d'une mesure de probabilité sur un ensemble matriciel donné. Des classes de matrices aléatoires très étudiées sont celles des *ensembles invariants*, qui généralisent les ensembles GOE, GUE et GSE. Elles se définissent sur l'ensemble des matrices carrées de taille $n \times n$ réelles symétriques, complexes Hermitiennes ou quaternioniques auto-adjointes (auxquelles on associe à nouveau le coefficient β valant respectivement 1, 2 et 4) par une densité définie sur cet ensemble de matrices de la forme d'une distribution de Boltzmann :

$$P(M)dM = \mathcal{Z}^{-1} \exp\left(-\frac{\beta n}{2} \text{Tr}V(M)\right) dM,$$

où $dM = \prod_{i \leq j} dM_{i,j}$ est la mesure de Lebesgue plate ($dM_{i,i}$ est la mesure de Lebesgue sur \mathbb{R} , et $dM_{i,j}$ pour $i < j$ est la mesure de Lebesgue sur \mathbb{R} , \mathbb{C} ou \mathbb{H} en fonction des ensembles considérés), la fonction $V : \mathbb{R} \mapsto \mathbb{R}$ est appelée *potentiel* de l'ensemble invariant et s'applique aux valeurs propres de la matrice dans l'expression ci-dessus, et \mathcal{Z} est une constante de normalisation. Ces distributions sont invariantes par conjugaison orthogonale, unitaire ou symplectique, mais les entrées de ces matrices ne sont pas indépendantes, sauf dans le cas gaussien (GOE, GUE et GSE). Dans ce contexte, une propriété est dite universelle si elle ne dépend pas de V . Contrairement aux matrices de Wigner, la densité jointe des valeurs propres des ensembles invariants est explicite, ce qui rend leur étude plus accessible. Cette densité par rapport à la mesure de Lebesgue sur \mathbb{R}^n est proportionnelle à

$$(\lambda^1, \dots, \lambda^n) \mapsto \exp\left(-\frac{\beta n}{2} \sum_{j=1}^n V(\lambda^j)\right) \prod_{1 \leq i < j \leq n} |\lambda^i - \lambda^j|^\beta. \quad (1.1.1)$$

Pour $V(x) = \frac{x^2}{2n}$, on retrouve en particulier la densité des valeurs propres des GOE, GUE et GSE.

Cette densité peut se réécrire sous la forme d'un facteur de Boltzmann

$$e^{-\frac{\beta n}{2} \sum_{j=1}^n V(\lambda^j) + \beta \sum_{1 \leq i < j \leq n} \ln|\lambda^i - \lambda^j|}$$

, où la différence entre deux particules apparaît comme un potentiel d'interaction, qui donne un point de vue gaz de Coulomb ou log-gaz à cette distribution. Sous le point de vue des gaz de particules, on pourrait même envisager de généraliser le type de gaz considéré en remplaçant $\ln|\lambda^i - \lambda^j|$ par un potentiel d'interaction générique $W(\lambda^i, \lambda^j)$. Ces gaz de particules sont des objets d'étude à part entière, en particulier l'analyse "statique" de leur position d'équilibre. Ils sont souvent appelés β -ensembles, et sont notamment documentés dans [For10]. Dans ce cadre, la densité (1.1.1) s'étudie pour tout paramètre $\beta > 0$, les cas $\beta = 1, 2, 4$ étant donc rattachés à un modèle de matrices aléatoires. Le point de vue des gaz de Coulomb fournit d'ailleurs des méthodes d'analyse du spectre des modèles matriciels sous-jacents. Deux β -ensembles nous ont intéressé particulièrement dans cette thèse, les β -Laguerre et les β -Jacobi, qui s'obtiennent à partir de (1.1.1) en prenant respectivement $V(x) = \frac{x}{n} - \left(\frac{m}{n} - 1 + \frac{1}{n} - \frac{2}{\beta n}\right) \ln(x)$ avec $m \geq n$ et $V(x) = -\left(\frac{p}{n} - 1 + \frac{1}{n} - \frac{2}{\beta n}\right) \ln(x) - \left(\frac{q}{n} - 1 + \frac{1}{n} - \frac{2}{\beta n}\right) \ln(1-x)$ avec $p \geq n$ et $q \geq n$, voir [For10, p. 111]. Les ensembles

de Laguerre orthogonal, unitaire et symplectique (respectivement LOE, LUE et LSE pour $\beta = 1, 2, 4$) se retrouvent dans la loi jointe des valeurs propres des matrices de covariance empirique XX^* où X est de taille $n \times m$ avec $n \leq m$, les $X_{i,j}$ étant respectivement des gaussiennes réelles, complexes et quaternioniques i.i.d. De même, si X et Y sont deux matrices de tailles respectives $n \times p$ et $n \times q$, les $X_{i,j}$ et $Y_{i,j}$ étant des gaussiennes réelles, complexes ou quaternioniques i.i.d., alors la loi jointe des valeurs propres de la matrice $(XX^* + YY^*)^{-1/2}XX^*(XX^* + YY^*)^{-1/2}$ correspond respectivement à la distribution des ensembles de Jacobi orthogonal (JOE, $\beta=1$), unitaire (JUE, $\beta = 2$) et symplectique (JSE, $\beta = 4$).

La théorie des matrices aléatoires contient de nombreux résultats macroscopiques comme microscopiques sur le comportement des valeurs propres. On s'intéressera ici plutôt aux comportements macroscopiques, et on donne ainsi la définition suivante.

Définition 1.1.2. La mesure spectrale empirique d'une matrice (aléatoire) X de taille $n \times n$ est une mesure de probabilités (aléatoire) définie par

$$\mu_X^n = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda^i(X)},$$

où $\lambda^1(X), \dots, \lambda^n(X)$ sont les valeurs propres de X .

Le théorème de Wigner, démontré pour la première fois dans [Wig58] dans le cas du GOE et considéré par certains comme le point de départ de la théorie des matrices aléatoires, donne le comportement asymptotique universel de la mesure spectrale empirique d'une matrice de Wigner.

Théorème 1.1.3 (Wigner [AGZ10], Théorème 2.1.21). *Soit $X = (X_{i,j})_{1 \leq i,j \leq n}$ une matrice de Wigner telle que $\mathbb{E}[X_{1,2}^2] = 1$. Alors en convergence presque sûre :*

$$\mu_{X/\sqrt{n}}^n \xrightarrow[n \rightarrow \infty]{} \mu^{sc},$$

où μ^{sc} est la distribution semi-circulaire définie par sa densité par rapport à la mesure de Lebesgue sur \mathbb{R} :

$$\sigma(x) = \frac{1}{2\pi} \sqrt{4 - x^2} \mathbf{1}_{[-2,2]}(x).$$

Dans la même optique de propriétés macroscopiques du spectre, d'autres résultats de "loi des grands nombres" sur les valeurs propres de matrices aléatoires furent démontrés, dont le célèbre théorème de Marcenko-Pastur [MP67] que nous détaillerons plus loin. De nombreuses techniques furent développées pour démontrer ce type de résultats, comme la combinatoire avec la méthode des moments, la transformée de Cauchy-Stieltjes, les probabilités libres, les polynômes orthogonaux, la théorie du potentiel, et bien d'autres. Nombre de ces techniques furent héritées ou inspirées de la physique, comme l'analogie des gaz de Coulomb ou les diagrammes de Feynman, et le lecteur pourra les retrouver par exemple dans [ABDF11, AGZ10, Tao12, BS10, CD11, BIPZ93, BGJJ04].

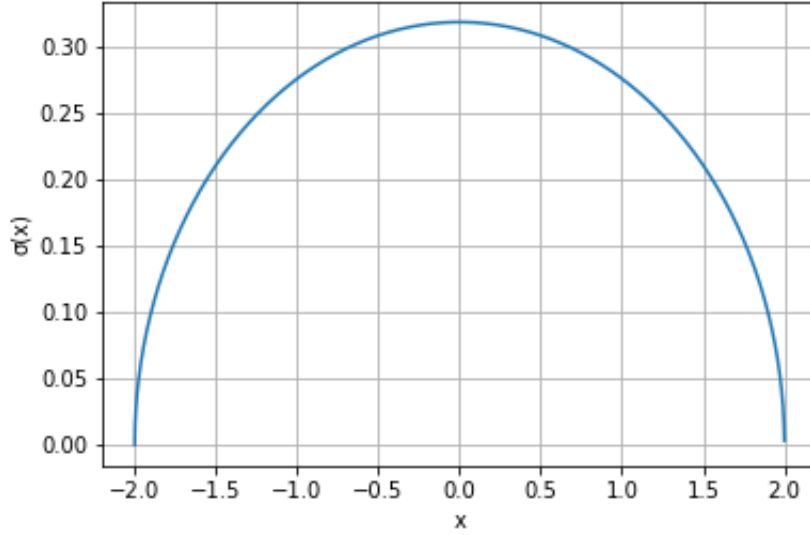


Figure 1.1 – Densité de la loi semi-circulaire.

La version convergence en norme L^1 du Théorème 1.1.3 se montre communément par la méthode dite *méthode des moments*, méthode qui fut utilisée par Wigner dans [Wig58]. Elle permet de faire le lien entre les moments de $\mu_{X/\sqrt{n}}^n$, que l'on peut aussi voir comme les traces normalisées des puissances de la matrice X . En montrant que ces moments s'expriment asymptotiquement en fonction des nombres de Catalan, on démontre la convergence des moments de $\mu_{X/\sqrt{n}}^n$ vers les moments de la mesure semi-circulaire.

Une méthode alternative est d'étudier la mesure spectrale empirique de X par sa transformée de Cauchy-Stieltjes, qui se définit pour toute mesure de probabilité sur l'axe réel μ par

$$G_\mu : z \in \mathbb{C}_+ \mapsto \int_{\mathbb{R}} \frac{d\mu(x)}{z - x} \in \mathbb{C}_-,$$

où $\mathbb{C}_\pm = \{z \in \mathbb{C} : \pm \text{Im}(z) > 0\}$. Pour une matrice X de taille $n \times n$, on écrira $G_X = G_{\mu_X^n}$. Pour une mesure μ à support compact, on réalise rapidement par un développement en série entière pour z suffisamment loin de l'axe réel que G_μ contient l'information de tous les moments de μ . On a en fait le résultat suivant qui montre que la transformée de Cauchy-Stieltjes d'une mesure de probabilité la caractérise entièrement, et qui montre également comment la retrouver à partir de G_μ .

Théorème 1.1.4 ([MS17]). (1) Soit μ une mesure de probabilité sur \mathbb{R} . Alors,

1. G_μ est analytique sur \mathbb{C}_+ ,

2. on a

$$\lim_{y \rightarrow \infty} iyG_\mu(iy) = 1.$$

(2) Toute mesure de probabilité sur \mathbb{R} peut être retrouvée à partir de sa transformée de Cauchy-Stieltjes G_μ via la formule d'inversion de Stieltjes : pour tout $a, b \in \mathbb{R}$, $a < b$,

$$-\lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \int_a^b \operatorname{Im}(G_\mu(x + i\varepsilon)) dx = \mu((a, b)) + \frac{1}{2}\mu(\{a, b\}).$$

La transformée de Cauchy-Stieltjes G_μ caractérise donc la mesure μ .

(3) Soit $G : \mathbb{C}_+ \mapsto \mathbb{C}_-$ une fonction analytique telle que

$$\limsup_{y \rightarrow \infty} |G(iy)| = 1.$$

Alors il existe une unique mesure de probabilité μ sur \mathbb{R} telle que $G = G_\mu$.

Dans le cadre du Théorème 1.1.3, en utilisant le fait que $G_X(z) = \frac{1}{n} \mathbb{E} \operatorname{Tr} \frac{1}{zI_n - X}$, on peut montrer que quand n tend vers l'infini, les valeurs d'adhérence de G_X vérifient une équation dont la seule solution correspondant à la transformée de Cauchy-Stieltjes d'une mesure de probabilité est celle de la mesure semi-circulaire.

Le dernier outil que nous allons introduire ici, et qui constitue le fil directeur de cette thèse est la théorie des probabilités libres développée par Voiculescu dans les années 1990. Cette théorie donne des règles de calcul pour les opérateurs non commutatifs fondées sur la notion de liberté.

Définition 1.1.5 ([Voi85]).

- Un *espace de probabilité non commutatif* est un couple (\mathcal{A}, φ) où \mathcal{A} est une \mathbb{C} -algèbre associative et $\varphi : \mathcal{A} \mapsto \mathbb{C}$ est une forme linéaire telle que $\varphi(1_{\mathcal{A}}) = 1$.
- Soit (\mathcal{A}, φ) un espace de probabilité non commutatif. Une famille $\{\mathcal{A}_i\}_{i \in I}$ de sous-algèbres de \mathcal{A} est dite libre si pour tout $n \in \mathbb{N}^*$, pour tous $i_1, \dots, i_n \in I$ tels que $i_k \neq i_{k+1}$ pour tout $k \in \{1, \dots, n-1\}$ et pour tout $(a_1, \dots, a_n) \in \mathcal{A}_{i_1} \times \dots \times \mathcal{A}_{i_n}$, on a

$$\varphi(a_1) = \dots = \varphi(a_n) = 0 \implies \varphi(a_1 \dots a_n) = 0.$$

Des familles de variables non commutatives sont dites *libres* si les sous-algèbres qu'elles engendrent sont libres.

Par exemple, l'algèbre $\mathcal{A} = \bigcap_{p \geq 1} L^p(\Omega, \mathcal{M}_n(\mathbb{C}))$ des matrices aléatoires de taille $n \times n$ dont les coefficients ont des moments de tout ordre, munie de $\varphi(a) = \frac{1}{n} \mathbb{E}[\operatorname{Tr}(a)]$ est un espace de probabilité non commutatif. Quelques années plus tard, Voiculescu [Voi91] comprit que les matrices aléatoires invariantes par rotation satisfont un critère de liberté asymptotique (dans la limite en dimension infinie), ce qui eut de nombreuses conséquences dans l'étude des matrices aléatoires. Intuitivement, deux matrices sont asymptotiquement libres lorsque la base de vecteurs propres de l'une est une rotation aléatoire de celle de l'autre.

Théorème 1.1.6. [Voi98] Soient A_n et B_n deux matrices de taille n dont les moments des mesures spectrales empiriques convergent quand $n \rightarrow +\infty$. Soit $(U_n)_n$ une suite de matrices unitaires de taille n . Alors $U_n A_n U_n^*$ et B_n sont asymptotiquement libres pour $n \rightarrow +\infty$.

On peut retrouver ce résultat sous différentes formes dont [NS06, Théorème 23.14].

Deux notions fondamentales de la théorie des probabilités libres sont la convolution libre additive et la convolution libre multiplicative. Avant de les définir, donnons tout d'abord la définition de trois fonctionnelles de mesures de probabilités, la fonction génératrice des moments, la *R-transformée* et la *S-transformée*.

Définition 1.1.7. Soit μ une mesure de probabilités définie sur \mathbb{R} .

1. La fonction génératrice des moments de μ se définit comme

$$M_\mu(z) = zG_\mu(z) - 1,$$

2. la R-transformée de μ se définit dans un voisinage de 0 comme

$$R_\mu(z) = G_\mu^{-1}(z) - \frac{1}{z},$$

3. pour μ mesure de probabilités définie sur \mathbb{R}_+ différente de δ_0 , la S-transformée de μ se définit sur un voisinage de 0 comme

$$S_\mu(z) = \frac{1+z}{zM_\mu^{-1}(z)}.$$

Comme la transformée de Cauchy-Stieltjes, ces fonctionnelles caractérisent la mesure μ . On a alors le résultat/définition suivant :

Théorème 1.1.8 ([Voi86, BV92]). Soient $(A_n)_n$ et $(B_n)_n$ deux suites de matrices de taille $n \times n$ asymptotiquement libres dont les suites de mesures spectrales empiriques $(\mu_{A_n}^n)_n$ et $(\mu_{B_n}^n)_n$ convergent presque sûrement quand n tend vers l'infini vers les mesures déterministes respectives μ_A et μ_B . Alors

$$\mu_{A_n+B_n}^n \xrightarrow{n \rightarrow \infty} \mu_A \boxplus \mu_B,$$

où $\mu_A \boxplus \mu_B$ se caractérise par la relation

$$R_{\mu_A \boxplus \mu_B} = R_{\mu_A} + R_{\mu_B}.$$

On appelle $\mu_A \boxplus \mu_B$ la convolution libre additive entre μ_A et μ_B . Si de plus μ_A et μ_B sont des mesures de probabilité sur \mathbb{R}_+ différentes de δ_0 , on a

$$\mu_{A_n B_n}^n \xrightarrow{n \rightarrow \infty} \mu_A \boxtimes \mu_B,$$

où $\mu_A \boxtimes \mu_B$ se caractérise par la relation

$$S_{\mu_A \boxtimes \mu_B} = S_{\mu_A} S_{\mu_B}.$$

On appelle $\mu_A \boxtimes \mu_B$ la convolution libre multiplicative entre μ_A et μ_B .

Ce résultat donne un aperçu de la puissance des probabilités libres dans l'analyse du spectre asymptotique de polynômes de matrices aléatoires quand une relation de liberté asymptotique existe entre les différentes matrices en jeu.

Les précédents résultats et outils décrivent le spectre de grandes matrices carrées. Dans cette thèse nous nous sommes intéressés au cas des matrices de variance covariance et aux matrices de Wishart, ce qui nous mène naturellement à nous interroger sur l'importance de ces matrices dans la théorie ainsi que dans des applications plus spécifiques, et sur l'existence de résultats analogues pour les matrices rectangulaires.

1.2 Matrices de Wishart et de Jacobi

À l'ère des "Big Data", de grandes quantités de données sont enregistrées quotidiennement dans de nombreux domaines de recherche et d'application tels que la physique, l'analyse d'images, la génomique, l'épidémiologie, l'ingénierie, l'économie et la finance. De nouvelles méthodes d'analyse de données furent nécessaires pour les analyser, en particulier quand ces données vivent dans des espaces de grande dimension, cas qui nous intéressera ici. Il est très naturel de rechercher les causes communes à l'origine des dynamiques de ces données. L'objet le plus simple quantifiant la variation d'une donnée par rapport à une autre est la matrice de covariance. Ses valeurs propres et vecteurs propres peuvent être utilisés pour caractériser les combinaisons de données les plus importantes et déterminantes dans l'explication des variations de leurs variables aléatoires sous-jacentes. C'est ce que l'on appelle l'analyse en composantes principales (ACP, ou PCA en anglais), et des livres entiers furent dédiés à ce sujet (voir par exemple [Jol86]).

Une autre méthode statistique très répandue est l'analyse canonique des corrélations (ACC ou CCA en anglais). Contrairement à l'ACP dont le but est de trouver une projection des données qui contient le maximum de variance, l'ACC opère sur deux jeux de données et a pour but de trouver des projections de ces deux jeux de données qui sont le plus corrélées entre elles.

Il est bien compris aujourd'hui dans le domaine de l'analyse statistique multivariée que la majorité des techniques standards telles que l'ACP, l'ACC, l'analyse de variance multivariée (MANOVA en anglais) ou l'analyse discriminante linéaire reposent sur l'analyse du spectre de matrices de covariance (voir par exemple [MMS76, Joh01, Joh01]). Là où l'ACP est liée à l'analyse des matrices de Wishart, voir [Joh01], l'ACC est intimement liée à l'analyse des matrices de Jacobi, voir [Joh08].

1.2.1 Matrices de covariance empiriques et matrices de Wishart

Alors que les matrices de Wigner ont un grand nombre d'applications notamment en physique théorique, c'est dans le contexte des statistiques puis des télécommunications qu'un intérêt se développa pour l'étude des matrices de covariance empiriques.

Comme dit précédemment, le travail fondateur en théorie des matrices aléatoires de Wishart s'attela à comprendre la corrélation entre les différentes caractéristiques d'un jeu

de données multivariées : donnons-nous un jeu de m vecteurs de données de \mathbb{C}^n i.i.d. :

$$\begin{cases} Y^1 = (Y_1^1, \dots, Y_n^1) \\ Y^2 = (Y_1^2, \dots, Y_n^2) \\ \vdots \\ Y^m = (Y_1^m, \dots, Y_n^m), \end{cases}$$

et définissons la matrice de taille $n \times m$: $Y = (Y^1, \dots, Y^m)$. La *matrice de covariance* C de Y^1 se définit par $C_{i,j} = \mathbb{E}[Y_i^1 Y_j^1]$.

Estimer la matrice C à partir de données trouve son importance dans de nombreux contextes. Par exemple, dans le cas de la régression linéaire multivariée, où l'on cherche à expliquer le vecteur Y^1 à partir d'une entrée X de taille $n \times p$ en utilisant le modèle linéaire

$$Y^1 = X\beta + \varepsilon,$$

où β représente le coefficient de régression à p facteurs et ε représente le résidu. En supposant $\mathbb{E}[\varepsilon|X] = 0$ et $\mathbb{E}[\varepsilon\varepsilon^*|X] = C$, l'estimateur des moindres carrés de β s'écrit

$$\hat{\beta} = (X^*CX)^{-1}X^*C^{-1}Y^1.$$

Obtenir un bon estimateur de C est donc un préalable à l'utilisation de l'estimateur des moindres carrés.

De même, trouver un bon estimateur de C est primordial dans l'utilisation de la méthode des moindres carrés généralisée, ou encore en classification, pour classifier des données entre plusieurs classes définies par des Gaussiennes, en utilisant la méthode d'analyse discriminante linéaire (ADL ou LDA en anglais, voir par exemple [HTF09]).

Classiquement, l'estimateur de la matrice de covariance utilisée est la *matrice de (variance) covariance empirique*, connue aussi sous le nom d'estimateur de *Pearson*, définie par

$$\hat{C} = \frac{1}{m}YY^*.$$

En effet on a alors

$$\hat{C}_{i,j} = \frac{1}{m} \sum_{k=1}^m Y_i^k Y_j^k,$$

qui, à n fixé, converge presque sûrement par la loi forte des grands nombres vers $C_{i,j}$ quand m tend vers l'infini. Cependant, quand n est grand et n'est plus négligeable devant la taille m de l'échantillon de données, l'estimation des éléments ou des valeurs propres de C devient problématique. C'est par exemple le cas en génomique, où l'on peut avoir la donnée de milliers de gènes pour un patient, avec une population de quelques centaines de patients. Dans ce régime où n et m deviennent très grand avec $\frac{n}{m} \xrightarrow{m \rightarrow +\infty} \rho > 0$, la matrice de variance covariance empirique définie ci-dessus a un comportement très différent de la matrice de covariance théorique C que l'on cherche à estimer. Ce régime est appelé le "régime de Kolmogorov".

Comme pour les matrices de Wigner, des résultats d'universalité furent prouvés pour les matrices de covariance empiriques, et une importance particulière est donnée au cas où les Y^i sont des vecteurs gaussiens, dans lequel des calculs explicites sont possibles, et qui donne une intuition du comportement des matrices de covariance empiriques dans des configurations plus générales. On appelle ainsi *matrice de Wishart* la matrice de covariance empirique définie positive \hat{C} dans le cas où les Y^i sont des vecteurs gaussiens. Dire que \hat{C} est une matrice de Wishart est équivalent à dire que $\hat{C} = \sqrt{C}ZZ^*\sqrt{C}$ où Z est une matrice de taille $n \times m$ dont les entrées sont i.i.d. et suivent une loi normale centrée réduite et C est de taille $n \times n$ (voir par exemple [BBP17]). Dans le papier considéré par beaucoup comme le point de départ de la théorie des matrices aléatoires [Wis28], John Wishart donne la densité explicite de \hat{C} sachant C , qui s'écrit :

$$\frac{m^{nm/2}}{2^{nm/2}\Gamma_n(m/2)} \frac{\det(\hat{C})^{\frac{m-n-1}{2}}}{\det(C)^{m/2}} e^{\frac{m}{2}\text{Tr}(C^{-1}\hat{C})},$$

où Γ_n est la fonction Gamma multivariée de paramètre n . On dit que \hat{C} suit la *loi de Wishart* $W_n(C/m, m)$. La loi des valeurs propres est également connue pour $C = I_n$ et a une densité par rapport à la mesure de Lebesgue sur \mathbb{R}^n proportionnelle à :

$$(\lambda^1, \dots, \lambda^n) \mapsto \prod_{i=1}^n (\lambda^i)^{\frac{n-m-1}{2}} e^{-\lambda^i} \prod_{i < j} |\lambda^i - \lambda^j|.$$

Le premier résultat de description de la limite en grande dimension de la mesure spectrale empirique de \hat{C} fut prouvé par Marcenko et Pastur en 1967 dans l'article [MP67]. La formule de la mesure limite montre que les valeurs propres de \hat{C} deviennent bruitées en grande dimension, quand $\rho > 0$, ce que l'on appelle en machine learning la *malédiction de la dimension*. Quand $C = I_n$, alors que l'on s'attend à trouver les valeurs propres de \hat{C} proches de 1, la mesure empirique des valeurs propres suit la *loi de Marcenko-Pastur* de paramètre ρ , et ces valeurs propres sont de plus en plus bruitées quand ρ grandit, comme on peut le voir Figure 1.2.

Théorème 1.2.1 (Marcenko-Pastur [MP67]). *Soit M une matrice aléatoire de taille $n \times m$ avec $(M_{i,j})_{1 \leq i \leq n, 1 \leq j \leq m}$ des variables aléatoires i.i.d. à valeurs réelles, complexes ou quaternioniques. Supposons $m = m_n$ avec*

$$\lim_{n \rightarrow \infty} \frac{n}{m_n} = \rho \geq 0.$$

Alors, presque sûrement,

$$\mu_{\frac{1}{m}MM^*}^n \xrightarrow{n \rightarrow \infty} \left(1 - \frac{1}{\max(\rho, 1)}\right) \delta_0 + \mu^{MP_{\rho,1}},$$

où $\mu^{MP_{\rho,\sigma}}$ est la loi de Marcenko-Pastur de paramètre de forme ρ et de paramètre d'échelle σ , et de densité par rapport à la mesure de Lebesgue sur \mathbb{R}

$$MP : x \mapsto \frac{\sqrt{(a_+ - x)(x - a_-)}}{2\pi\rho x\sigma^2} \mathbb{1}_{[a_-, a_+]} \text{ où } a_{\pm} = \sigma^2(1 \pm \sqrt{\rho})^2.$$

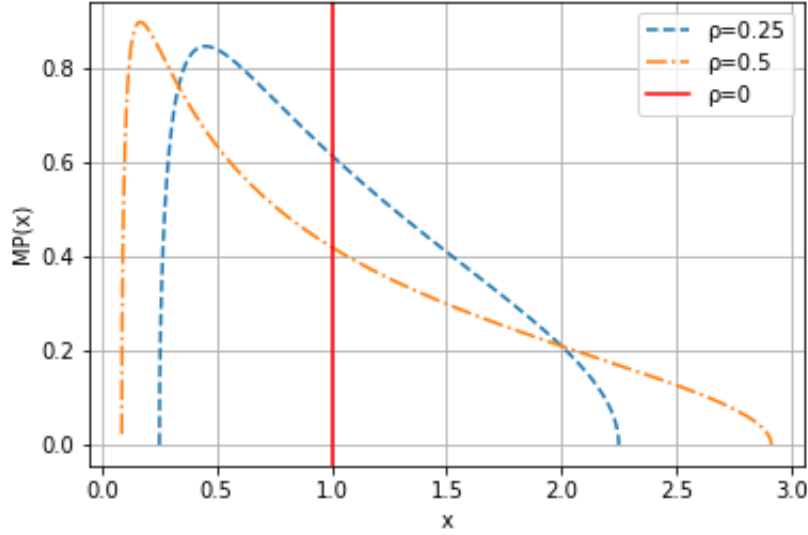


Figure 1.2 – Densité de la loi de Marcenko-Pastur.

Dans leur résultat, Marcenko et Pastur obtinrent en fait une équation pour la transformée de Cauchy-Stieltjes de \hat{C} dépendant de celle de C . Par exemple, dans un modèle où les données Y peuvent s'écrire $Y = \sqrt{C}X$ avec X une matrice de taille $n \times m$ à entrées non corrélées vérifiant $\mathbb{E}[X_{i,j}] = 0$, $\mathbb{E}[X_{i,j}^2] = \frac{1}{m}$ et $\mathbb{E}[X_{i,j}^4] < +\infty$, on a $\hat{C} = \sqrt{C}XX^*\sqrt{C}$ et l'équation quand n tend vers l'infini avec $n/m \xrightarrow[n \rightarrow \infty]{} \rho$:

$$zG_{\hat{C}}(z) = \frac{z}{1 - \rho + \rho zG_{\hat{C}}(z)} G_C \left(\frac{z}{1 - \rho + \rho zG_{\hat{C}}(z)} \right).$$

Ce résultat permet de nombreuses avancées en statistiques en grande dimension, et peut se démontrer de plusieurs façons, par exemple à l'aide des probabilités libres et de la convolution libre multiplicative introduite précédemment.

1.2.2 Matrices de Wishart, ensembles de Jacobi et statistiques

Plusieurs méthodes de statistiques multivariées sont liées, comme cas particulier ou cas limite, à l'étude du problème *Wishart double*, c'est à dire à l'étude de la matrice $A(A+B)^{-1}$ où A et B sont des matrices de Wishart, et ce lien est détaillé dans [Joh08]. La loi jointe des valeurs propres de $A(A+B)^{-1}$ est donnée par la distribution des ensembles de Jacobi orthogonaux.

Dans le cas de l'ACC, supposons que nous avons m observations de $p+q$ variables avec $p \leq q$. Les p premières variables sont regroupées dans la matrice de données de taille $m \times p$: $X = (x_1, \dots, x_p)$ et les q dernières dans la matrice de taille $m \times q$: $Y = (y_1, \dots, y_q)$. La version centrée de l'ACC cherche les combinaisons linéaires a^*x et b^*y qui sont le plus

corrélées, i.e. qui maximisent

$$r = \text{Corr}(a^*x, b^*x) = \frac{a^*X^*Yb}{\sqrt{a^*X^*Xa}\sqrt{b^*Y^*Yb}}.$$

On obtient de cette démarche une corrélation r_1 et deux vecteurs a_1 et b_1 . On peut réitérer cette opération en se restreignant à chaque pas aux vecteurs orthogonaux à ceux déjà trouvés :

$$r_k = \max\{a^*X^*Yb : a^*X^*Xa = b^*Y^*Yb = 1, a^*X^*Xa_j = b^*Y^*Yb_j = 0 \forall 1 \leq j < k\}.$$

Les corrélations successives sont les racines de l'équation

$$\det(X^*Y(Y^*Y)^{-1}Y^*X - r^2X^*X),$$

voir [KBM79, p. 284]. En notant $A = X^*(I - Y(Y^*Y)^{-1}Y^*)X$ et $B = X^*Y(Y^*Y)^{-1}Y^*X$, cette équation se réécrit $\det(B - r^2(A + B)) = 0$. En supposant que les données sont des gaussiennes centrées de matrice de covariance

$$\Sigma = \begin{pmatrix} I_p & \Sigma_{XY} \\ \Sigma_{YX} & I_q \end{pmatrix},$$

on a sous l'hypothèse nulle $H_0 : \Sigma_{XY} = 0$ que les matrices A et B sont indépendantes et suivent les distributions de Wishart $A \sim W_p(I, m - q)$ et $B \sim W_q(I, m - p)$.

Ces matrices apparaissent également dans le problème de l'égalité entre des matrices de covariance. Supposons que nous avons deux jeux de données indépendants provenant de deux distributions normales de matrices de covariance respectives C_1 et C_2 de taille $p \times p$. On a vu précédemment que leurs estimateurs \hat{C}_1 et \hat{C}_2 suivent des distribution de Wishart. Alors le test de l'hypothèse nulle $H_0 : C_1 = C_2$ repose sur la plus grande valeur propre de $(C_1 + C_2)^{-1}C_2$ (voir [Mui09]).

Les matrices de cette forme apparaissent également dans l'analyse discriminante linéaire multiple, le modèle de régression linéaire multivarié, et dans l'étude des angles et distances entre des sous-espaces vectoriels aléatoires, ces liens étant également documentés dans [Joh08]. Les ensembles de Jacobi sont donc d'un grand intérêt dans l'analyse statistique multivariée.

Les probabilités libres rectangulaires

Cette section montre l'importance, en statistiques notamment, de l'analyse spectrale des matrices rectangulaires. Soit X une matrice à coefficients réels ou complexes de taille $n \times m$. On appelle *valeurs singulières* de X les valeurs propres de la matrice $\sqrt{XX^*}$. De même, on appelle mesure singulière empirique de X la mesure spectrale empirique de $\sqrt{XX^*}$. Une théorie des probabilités libres rectangulaire, développée par Florent Benaych-Georges [BG09] permet de décrire, comme en probabilités libres "classiques", le comportement asymptotique de la mesure singulière empirique de la somme de deux matrices rectangulaires libres quand le rapport des dimensions n/m tend, quand n tend vers l'infini, vers un coefficient $\rho \geq 0$. C'est l'équivalent de la convolution libre additive dans ce contexte, appelée convolution libre rectangulaire de ratio ρ . Elle est définie dans les résultats suivants issus de [BG09].

Théorème 1.2.2 (Convolution libre rectangulaire additive de ratio ρ pour les matrices aléatoires). *Pour tous $n, m \in \mathbb{N}^*$, soient $M_{n,m}$ et $N_{n,m}$ deux matrices aléatoires à coefficients réels (complexes) indépendantes de taille $n \times m$ telles que*

- *la distribution de $M_{n,m}$ est invariante sous l'action du groupe orthogonal (unitaire) de chaque côté,*
- *les suites de mesures empiriques $(\mu^{\sqrt{M_{n,m}M_{n,m}^*}})_{n,m}$ et $(\mu^{\sqrt{N_{n,m}N_{n,m}^*}})_{n,m}$ convergent respectivement en probabilité, quand n et m tendent vers l'infini avec n/m tendant vers $\rho \in (0, 1]$, vers les mesures de probabilité déterministes μ_∞^M et μ_∞^N .*

Alors, au sens de la convergence faible en probabilité,

$$\text{sym} \left(\mu^{\sqrt{(M_{n,m}+N_{n,m})(M_{n,m}+N_{n,m})^*}} \right) \xrightarrow[n/m \rightarrow \rho]{n, m \rightarrow \infty} \text{sym}(\mu_\infty^M) \boxplus_\rho \text{sym}(\mu_\infty^N),$$

où l'opération sym est définie Chapitre 4.

Cette opération peut de manière équivalente être définie pour des variables libres d'un espace de probabilité non commutatif rectangulaire.

Soit μ une mesure de probabilité symétrique sur \mathbb{R} . Sa transformée de Cauchy rectangulaire de ratio ρ se définit comme

$$H_\mu : z \in \mathbb{C} \setminus [0, +\infty) \mapsto z(\rho M_{\mu^2}(z) + 1)(M_{\mu^2}(z) + 1),$$

où

$$M_{\mu^2} : z \in \mathbb{C} \setminus [0, +\infty) \mapsto \int_{\mathbb{R}} \frac{zv^2}{1-zv^2} d\mu(v) = \frac{1}{\sqrt{z}} G_\mu \left(\frac{1}{\sqrt{z}} \right) - 1.$$

La R-transformée rectangulaire de ratio ρ de μ est définie sur un voisinage de zéro par

$$C_\mu(z) = U \left(\frac{z}{H_\mu^{-1}(z)} - 1 \right),$$

où sur un voisinage de zéro

$$U(z) = \begin{cases} \frac{-\rho - 1 + [(\rho + 1)^2 + 4\rho z]^{1/2}}{2\rho} & \text{if } \rho \neq 0 \\ z & \text{if } \rho = 0. \end{cases}$$

On a alors comme en Théorie des Probabilités Libres classique le résultat suivant.

Théorème 1.2.3 ([BG09]). *La R-transformée rectangulaire de ratio ρ linéarise la convolution libre rectangulaire de ratio ρ : pour μ et ν des mesures de probabilités symétriques sur l'axe réel, et pour z dans un voisinage de zéro,*

$$C_{\mu \boxplus_\rho \nu}(z) = C_\mu(z) + C_\nu(z),$$

et $\mu \boxplus_\rho \nu$ est l'unique mesure de probabilité symétrique vérifiant cette relation.

Théorème 1.2.4 (Injectivité de la R-transformée rectangulaire, [BG09]). *Si les R-transformées rectangulaires de ratio ρ de deux mesures de probabilité symétriques coïncident sur un voisinage de 0 dans $(-\infty, 0]$, alors les deux mesures sont identiques.*

1.3 Processus matriciels et gaz de particules

Nous cherchons ici, avant de présenter les résultats développés pendant cette thèse, à montrer différentes motivations de l'introduction de processus matriciels.

Dans [Dys62], Dyson a pour objectif de généraliser le principe de matrices aléatoires de façon à donner aux modèles de gaz de Coulomb un "sens", par delà celui de l'étude d'un modèle statique d'équilibre thermodynamique : un modèle qui pourrait ne pas être à l'équilibre, et qui évoluerait avec le temps comme un système dynamique. Il propose donc d'étudier des processus stochastiques à valeur matricielle. En effet, le calcul d'Itô permet, pour certains processus matriciels, de montrer que les processus des valeurs propres du processus matriciel suivent un système équations différentielles stochastiques (EDS), et il en est parfois de même pour les vecteurs propres. Plus précisément, on a le résultat suivant.

Théorème 1.3.1 ([AGZ10]). *Soient $((B_t^{j,k})_t, (\tilde{B}_t^{j,k})_t, 1 \leq j \leq k \leq n)$ une collection de mouvements Browniens réels indépendants. Définissons le processus matriciel $(H_t)_t$ symétrique pour $\beta = 1$, hermitien pour $\beta = 2$ dont les coefficients sont pour tout $t \geq 0, j \leq k$:*

$$H_t^{j,k} = \begin{cases} \frac{1}{\sqrt{\beta n}}(B_t^{j,k} + i(\beta - 1)\tilde{B}_t^{j,k}) & \text{if } j < k, \\ \sqrt{\frac{2}{\beta n}}B_t^{j,j} & \text{if } j = k. \end{cases}$$

Supposons que les valeurs propres de la matrice symétrique (hermitienne) M_0 appartiennent à la fermeture $\bar{\Delta}$ du simplexe $\Delta = \{(x^i)_{1 \leq i \leq n} \in \mathbb{R}^n : x^1 < x^2 < \dots < x^n\}$. Supposons de plus que pour tout $t \geq 0$ les valeurs propres $(\lambda_t^1, \dots, \lambda_t^n)$ de $M_t = M_0 + H_t$ appartiennent presque sûrement à $\bar{\Delta}$.

Alors les processus $(\lambda_t^1, \dots, \lambda_t^n)_t$ sont des semi-martingales satisfaisant le système d'équations différentielles stochastiques

$$d\lambda_t^i = \frac{\sqrt{2}}{\sqrt{\beta n}}dB_t^i + \frac{1}{n} \sum_{j \neq i} \frac{1}{\lambda_t^i - \lambda_t^j} dt \text{ pour tout } i \in \{1, \dots, n\},$$

où les $(B_t^i)_t$ sont des mouvements Browniens réels indépendants.

Le système de particules défini dans le Théorème 1.3.1 est appelé le *mouvement Brownien de Dyson*. La démarche de rechercher une diffusion de particules soumises à un potentiel donné peut être généralisée par le fait bien connu que pour toute fonction $V : \mathbb{R}^n \mapsto \mathbb{R}$ satisfaisant les bonnes propriétés de régularité et de croissance, l'équation différentielle stochastique gradient

$$dx_t^i = dB_t^i - \nabla_i V(x_t^1, \dots, x_t^n) dt \text{ pour tout } i \in \{1, \dots, n\}, \quad (1.3.1)$$

où $(x_t^1, \dots, x_t^n)_t$ est un vecteur de procesus stochastiques de taille n et les $(B_t^i)_t$ sont des mouvements Browniens indépendants sur \mathbb{R} , a une mesure stationnaire de densité par rapport

à la mesure de Lebesgue sur \mathbb{R}^n proportionnelle à $(x^1, \dots, x^n) \in \mathbb{R}^n \mapsto e^{-2V(x^1, \dots, x^n)}$. On peut donc fabriquer des systèmes de particules dynamiques sous certaines hypothèses sur V (voir par exemple le mouvement Brownien de Dyson généralisé [LLX14, Théorème 1.1], pour que l'EDS (1.3.1) admette une unique solution forte) à partir de leur mesure stationnaire. Sous certaines hypothèses sur V , la loi jointe des n particules convergera vers cette mesure stationnaire en temps long. Les dynamiques des systèmes de particules de mesure stationnaire proportionnelle à (1.1.1) peuvent ainsi s'étudier dans le cas général $\beta > 0$ dans le cadre des systèmes d'équations différentielles stochastiques, et il existe plusieurs résultats rattachant ces dynamiques pour $\beta > 0$ à des modèles matriciels, notamment tridiagonaux (voir par exemple [AGZ10, Section 4.5]).

En physique statistique comme en matrices aléatoires, on s'intéresse aux limites en grandes dimension. La question de la commutativité des limites pour les systèmes de particules dynamiques en grande dimension et en temps long se pose naturellement, et est adressée pour les ensembles β -Laguerre dans le Chapitre 4.

Outre l'étude dynamique de gaz de Coulomb, le calcul stochastique est également utilisé comme outil d'analyse des grandes matrices aléatoires. Par exemple, dans le cas du mouvement Brownien de Dyson (Théorème 1.3.1) avec $\beta = 1$, si l'on note pour tout $t \geq 0$ la mesure empirique $\mu_t = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i}$ et que l'on suppose que la mesure spectrale empirique $\mu_{M_0}^n$ de M_0 converge presque sûrement vers une mesure de probabilité $\mu_{M_0}^\infty$ quand n tend vers l'infini, le calcul d'Itô montre que la transformée de Cauchy-Stieltjes $(z, t) \mapsto G(z, t)$ de μ_t vérifie asymptotiquement (pour $n \rightarrow +\infty$) une équation aux dérivées partielles complexe dite de Burgers s'écrivant

$$\frac{\partial G}{\partial t}(z, t) = -G(z, t) \frac{\partial G}{\partial z}(z, t), \quad G(z, 0) = G_{\mu_{M_0}^\infty}(z),$$

qui se résout par la méthode des caractéristiques et permet de décrire le comportement de μ_t en grande dimension pour tout $t \geq 0$.

Récemment, le calcul stochastique matriciel fut utilisé par le groupe de chercheurs *László Erdős, Benjamin Schlein* et *Hornng-Tzer Yau* (voir par exemple [ESY09, EPR⁺10, EY17]) pour démontrer que la convergence de la mesure spectrale empirique des matrices de Wigner vers la loi du semi-cercle a aussi lieu localement, plus précisément à peu de choses près à n'importe quelle échelle supérieure à la distance typique entre les valeurs propres normalisées, qui est de l'ordre de grandeur n^{-1} . Ils parlent donc d'universalité locale. La stratégie de la preuve est notamment introduite dans le livre [EY17].

Historiquement, une autre motivation pour l'introduction du calcul stochastique dans le monde des matrices aléatoires est l'étude de la stabilité de la distributions des valeurs propres de modèles matriciels par l'introduction d'un bruit, par exemple gaussien. Les distributions qui résultent d'une telle perturbation sont compliquées et c'est notamment ce qui conduit *Marie-France Bru* dans son article [Bru89] à étudier un modèle où le bruit ajouté est dynamique : une matrice dont les entrées sont des mouvement Browniens indépendants. La démarche mathématique est en fait la même qu'au Théorème 1.3.1. Ces matrices ont été le point de départ de cette thèse.

Les systèmes d'équations différentielles stochastiques étant déjà un sujet d'investigation du domaine de recherche du calcul stochastique, les préoccupations liées aux gaz de Coulomb et aux matrices aléatoires sont apparues en plusieurs occasions dans cette littérature, par exemple avec l'étude des processus de Dunkl (voir par exemple [Chy06, Dem09b, GY+06]). Des outils développés dans le contexte des processus stochastiques tels que les équations différentielles multivoques introduites et étudiées par Emmanuel Cépa ([Cép94, Cép95]) sont notamment utiles pour étendre les résultats d'existence forte et d'unicité de solutions à des systèmes de particules hérités des β -ensembles, en particulier dans ce travail de thèse.

1.3.1 Systèmes de particules en interaction

Une des contributions de cette thèse fut d'étendre les résultats d'existence et d'unicité de solutions fortes à des systèmes d'équations différentielles stochastiques issus de problèmes de matrices aléatoires et de β -ensembles : le processus de β -Wishart et le processus de β -Jacobi, liés respectivement aux ensembles de β -Laguerre et aux ensembles de β -Jacobi.

Processus de β -Wishart

Ces résultats sont issus d'un travail en collaboration avec mon directeur de thèse *Benjamin Jourdain*.

Soient $\alpha \geq 0$, $\gamma \in \mathbb{R}$, $\beta > 0$, $n \geq 1$, et $\mathbf{B} = (B_t^1, \dots, B_t^n)_t$ un mouvement Brownien en dimension n . Le processus de β -Wishart (que l'on trouve parfois sous le nom de β -Laguerre dans la littérature) est solution du système d'équations différentielles stochastiques suivant, décrivant des particules positives, évoluant comme des processus de Cox-Ingersoll-Ross [CIR85], et se repoussant selon une interaction coulombienne.

$$d\lambda_t^i = 2\sqrt{\lambda_t^i}dB_t^i + \left(\alpha - 2\gamma\lambda_t^i + \beta \sum_{j \neq i} \frac{\lambda_t^i + \lambda_t^j}{\lambda_t^i - \lambda_t^j} \right) dt \text{ pour tout } i \in \{1, \dots, n\} \quad (1.3.2)$$

$$0 \leq \lambda_t^1 < \dots < \lambda_t^n, \text{ p.s. } dt - \text{ presque partout.}$$

Ce processus est apparu dans les travaux de *Marie-France Bru* [Bru89] dont la motivation première, comme expliqué plus tôt, était l'analyse de la stabilité de l'algorithme d'analyse en composantes principales par une perturbation gaussienne en ajoutant une matrice de mouvements Browniens indépendants à une matrice de données. Dans [Bru91], elle constate que cette diffusion matricielle constitue une extension matricielle naturelle des carrés de Bessel.

En effet, un carré de Bessel de dimension $m \geq 1$ satisfaisant l'équation différentielle stochastique $dx_t = 2\sqrt{x_t}dB_t + mdt$ où $(B_t)_t$ est un mouvement Brownien a pour générateur $m\frac{\partial}{\partial x} + 2x\frac{\partial^2}{\partial x^2}$. Bru constate qu'en considérant une matrice $(M_t)_t$ de taille $m \times n$ avec $n \geq m$ dont les entrées mouvements Browniens indépendants, alors le processus $(M_t^*M_t)_t$ est solution de l'EDS suivante sur l'ensemble des matrices réelles symétriques :

$$dS_t = \sqrt{S_t}dN_t + dN_t^*\sqrt{S_t} + mI_n dt, \quad (1.3.3)$$

où $(N_t)_t$ est une matrice $n \times n$ de mouvements Browniens indépendants. Le générateur de S est alors $Tr \left(m \frac{\partial}{\partial x} + 2x \frac{\partial^2}{\partial x^2} \right)$. En plus de la généralisation du paramètre $m \geq n$ entier à un paramètre $\alpha \geq n$ réel, Bru introduit le processus suivant : $(M_t)_t$ est un processus stochastique prenant ses valeurs dans l'espace des matrices réelles de taille $m \times n$ vérifiant l'EDS

$$dM_t = dW_t - \gamma M_t dt, \quad M_0 = m_0,$$

où $\gamma \geq 0$, $(W_t)_t$ est une matrice de taille $m \times n$ dont les entrées sont des mouvements Browniens indépendants et m_0 est une matrice réelle déterministe de taille $m \times n$. Les entrées de M sont donc des processus d'Ornstein-Uhlenbeck indépendants. Un analogue complexe de ce modèle a été étudié dans [KO01]. Bru démontre dans [Bru91] que le processus des valeurs propres $(\lambda_t^1, \dots, \lambda_t^n)_t$ de $(M_t^* M_t)_t$ est solution du système d'EDS

$$d\lambda_t^i = 2\sqrt{\lambda_t^i} dB_t^i + m dt - 2\gamma m \lambda_t^i dt + \sum_{j \neq i} \frac{\lambda_t^i + \lambda_t^j}{\lambda_t^i - \lambda_t^j} dt \text{ pour tout } i \in \{1, \dots, n\}$$

$$0 \leq \lambda_t^1 < \dots < \lambda_t^n \text{ p.s. dt-p.p.}$$

Le système (1.3.2) généralise ce dernier par l'introduction d'une intensité $\beta > 0$ donnée aux termes d'interaction, et par le remplacement du paramètre entier $m \geq n$ par un paramètre réel $\alpha \geq 0$. Ce remplacement de m par α est l'analogie de la généralisation de la dynamique de la norme d'un mouvement Brownien réel m -dimensionnel aux carrés de Bessel de paramètre α (voir [RY99, chapitre XI]).

On montre le résultat d'existence et d'unicité forte suivant (le cas $\beta \geq 1$ est déjà traité dans la littérature dans [YCG08, Dem09b, GM13, GM14], voir Chapitre 2 pour plus de détails).

Théorème 1.3.2. *Supposons $\beta < 1, \alpha - (n - 1)\beta > 0$. Considérons la condition initiale $\Lambda_0 = (\lambda_0^1, \dots, \lambda_0^n)$ indépendante du mouvement Brownien \mathbf{B} , telle que $0 \leq \lambda_0^1 \leq \dots \leq \lambda_0^n$ p.s. et $\lambda_0^2 > 0$ p.s.*

Alors le système d'EDS (1.3.2) a une unique solution forte sur l'intervalle

$$[0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon),$$

où, pour $\varepsilon > 0$,

$$\zeta_\varepsilon = \inf\{t \geq 0 : \lambda_t^1 \leq \varepsilon \text{ and } \lambda_t^2 - \lambda_t^1 \leq \varepsilon\}.$$

De plus,

- (i) pour $\gamma \in \mathbb{R}$, si $\alpha - (n - 1)\beta \geq 1 - \beta$ alors $\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon = \infty$ p.s.
- (ii) pour $\gamma \geq 0$, si $\alpha - (n - 1)\beta < 1 - \beta$ alors $\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon < \infty$ p.s.
- (iii) Il n'y a pas de collision multiple entre particules, i.e. pour tout $1 \leq i < j \leq n - 1$,

$$\mathbb{P}\left\{\exists t \in (0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon) : \lambda_t^i = \lambda_t^{i+1} \text{ et } \lambda_t^j = \lambda_t^{j+1}\right\} = 0.$$

On montre également que pour $\beta < 1$, il y a collision entre particules voisines en temps fini presque sûrement.

Proposition 1.3.3. *Soient $\beta \in (0, 1)$, $\alpha > 0$ et considérons la condition initiale $\Lambda_0 = (\lambda_0^1, \dots, \lambda_0^n)$ indépendante du mouvement Brownien \mathbf{B} et telle que $0 \leq \lambda_0^1 \leq \dots \leq \lambda_0^n$. Supposons que $\gamma > 0$ et que le système (1.3.2) admette une solution définie sur \mathbb{R}_+ , ou $\gamma \geq 0$, $\alpha - (n-1)\beta \geq 1 - \beta$ et $\lambda_0^2 > 0$ p.s. (de façon que, par le Theorem 1.3.2, il existe également une solution définie sur \mathbb{R}_+ au système (1.3.2)). Alors pour tout $i \in \{2, \dots, n\}$, le temps d'arrêt*

$$T^{(i)} = \inf\{t > 0 : \lambda_t^i = \lambda_t^{i-1}\}$$

vérifie

$$\mathbb{P}\{T^{(i)} < \infty\} = 1.$$

Le lien avec les ensembles de β -Laguerre se comprend par le résultat suivant.

Proposition 1.3.4. *Supposons $\gamma > 0$ et $\alpha - (n-1)\beta > 0$. L'unique mesure de probabilité stationnaire du système d'EDS (1.3.2) est ρ_{inv} admettant pour densité par rapport à la mesure de Lebesgue sur \mathbb{R}^n :*

$$d\rho_{inv}(\lambda^1, \dots, \lambda^n) = \frac{1}{\mathcal{Z}} \times \prod_{i=1}^n \left((\lambda^i)^{\frac{\alpha - (n-1)\beta}{2} - 1} e^{-\gamma \lambda^i} \prod_{j \neq i} |\lambda^j - \lambda^i|^{\beta/2} \right) \mathbb{1}_{\{0 \leq \lambda^1 \leq \dots \leq \lambda^n\}} d\lambda^1 \dots d\lambda^n,$$

où \mathcal{Z} est une constante de normalisation. Plus précisément, si une solution $\Lambda = (\lambda_t^1, \dots, \lambda_t^n)_t$ à (1.3.2) est telle que la distribution de Λ_t ne dépend pas de t , alors cette distribution est ρ_{inv} . Réciproquement, (1.3.2) admet une unique solution Λ de condition initiale Λ_0 distribuée selon ρ_{inv} et indépendante du mouvement Brownien \mathbf{B} . De plus, pour tout $t \in \mathbb{R}_+$, Λ_t est distribuée selon ρ_{inv} .

On montre de plus que si $\gamma > 0$, si les coordonnées de la condition initiale sont intégrables, i.e.

$$\mathbb{E} \sum_{i=1}^n \lambda_0^i < +\infty,$$

et si la solution Λ est globale, alors la distribution de $(\lambda_t^1, \dots, \lambda_t^n)$ converge étroitement vers ρ_{inv} quand $t \rightarrow +\infty$. La distribution ρ_{inv} donne une généralisation de la distribution de l'ensemble de β -Laguerre qui s'écrit

$$\frac{1}{\mathcal{Z}} \prod_{i=1}^n \left((\lambda^i)^{\frac{m\beta - (n-1)\beta}{2} - 1} e^{-\frac{\beta}{2} \lambda^i} \prod_{j \neq i} |\lambda^j - \lambda^i|^{\beta/2} \right) \mathbb{1}_{\{0 \leq \lambda^1 \leq \dots \leq \lambda^n\}}, \quad m \in \{n, n+1, \dots\}, \quad \beta > 0,$$

en remplaçant $m\beta$ par α dans l'exposant de λ^i et $\frac{\beta}{2}$ par γ dans le terme exponentiel.

Une étape clé dans la preuve de ces résultats est d'étudier le système d'EDS obtenu en calculant formellement $d\sqrt{\lambda_t^i}$ en appliquant la formule d'Itô :

$$dx_t^i = dB_t^i + \left(\frac{\alpha - 1}{2} \frac{1}{x_t^i} - \gamma x_t^i + \frac{\beta}{2x_t^i} \sum_{j \neq i} \frac{(x_t^i)^2 + (x_t^j)^2}{(x_t^i)^2 - (x_t^j)^2} \right) dt \quad \text{pour tout } i \in \{1, \dots, n\} \quad (1.3.4)$$

$$0 \leq x_t^1 < \dots < x_t^n, \text{ p.s., } dt - p.p.,$$

qui peut s'écrire comme une diffusion gradient

$$dx_t^i = dB_t^i - \partial_i V(x_t^1, \dots, x_t^n) dt \text{ pour tout } i \in \{1, \dots, n\}, \quad (1.3.5)$$

de potentiel

$$V(x_1, \dots, x_n) = - \sum_{i=1}^n \left\{ \frac{\alpha - (n-1)\beta - 1}{2} \ln |x^i| - \frac{1}{2} \gamma (x^i)^2 + \frac{\beta}{4} \sum_{j \neq i} (\ln |x^i - x^j| + \ln |x^i + x^j|) \right\}. \quad (1.3.6)$$

Processus de β -Jacobi

Soient $p, q \geq 0, \beta > 0, n \geq 1$, et $\mathbf{B} = (B_t^1, \dots, B_t^n)_t$ un mouvement Brownien n -dimensionnel. Le processus de β -Jacobi décrit l'évolution de particules sur le segment $[0, 1]$ se repoussant par une force coulombienne, et est solution du système d'EDS suivant :

$$d\lambda_t^i = 2\sqrt{\lambda_t^i(1-\lambda_t^i)} dB_t^i + \beta \left[p - (p+q)\lambda_t^i + \sum_{j \neq i} \frac{\lambda_t^i(1-\lambda_t^j) + \lambda_t^j(1-\lambda_t^i)}{\lambda_t^i - \lambda_t^j} \right] dt \text{ pour tout } i \in \{1, \dots, n\} \quad (\text{J}(p,q))$$

$$0 \leq \lambda_t^1 < \dots < \lambda_t^n \leq 1, \text{ p.s. } dt - \text{presque partout.}$$

Pour $n = 1$, le processus de Jacobi réel fut déjà étudié pour des modèles de génétique (voir par exemple [EK09]) ou financiers (voir par exemple [DS02]). Plus généralement, les processus β -Jacobi apparaissent dans l'étude du processus de Jacobi matriciel introduit et analysé par Yan Doumerc dans [Dou05] : soient n, m, p trois entiers naturels non nuls, Θ un mouvement Brownien orthogonal de dimension $m \times m$ (voir [Lév17] pour une construction du mouvement Brownien orthogonal), et soit X la matrice constituée des n premières lignes et des p premières colonnes de Θ . Alors, si $p \geq n + 1, q \geq n + 1$ et $p + q = m$, le processus $J = XX^*$, est une diffusion et est solution de l'EDS

$$dJ = \sqrt{J} dN \sqrt{I_n - J} + \sqrt{I_n - J} dN^* \sqrt{J} + (pI - (p+q)J) dt, \quad (1.3.7)$$

où N est une matrice de taille $n \times n$ dont les entrées sont des mouvements Browniens indépendants (voir [Dou05, Theorem 9.2.3]). Les valeurs propres de J vérifient alors le système d'EDS (J(p,q)) avec $\beta = 1$ (voir [Dou05, Theorem 9.3.1]). En prenant Θ un mouvement Brownien unitaire plutôt qu'orthogonal, les valeurs propres de J vérifient le système d'EDS (J(p,q)) avec $\beta = 2$ (voir [Dem09a]). Le système (J(p,q)) est donc une généralisation à des coefficients généraux $p, q \geq 0$ et $\beta > 0$ de ces systèmes issus de processus matriciels.

On montre le résultat d'existence et d'unicité suivant (le cas $\beta \geq 1$ est déjà traité dans la littérature dans [Dem09a, GM13, GM14], et on étend ici ces résultats, voir Chapitre 2 pour plus de détails).

Théorème 1.3.5. *Supposons $\beta < 1$ et $p \wedge q - n + 1 > 0$. Considérons la condition initiale $\Lambda_0 = (\lambda_0^1, \dots, \lambda_0^n)$ indépendante du mouvement Brownien \mathbf{B} et telle que $0 \leq \lambda_0^1 \leq \dots \leq \lambda_0^n \leq 1$ p.s. et $\lambda_0^2(1 - \lambda_0^{n-1}) > 0$ p.s.*

Alors le système d'EDS $(J(p, q))$ admet une unique solution forte définie sur l'intervalle de temps

$$[0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon),$$

où, pour tout $\varepsilon > 0$,

$$\zeta_\varepsilon = \inf\{t \geq 0 : \lambda_t^1 \leq \varepsilon \text{ et } \lambda_t^2 - \lambda_t^1 \leq \varepsilon\} \wedge \inf\{t \geq 0 : \lambda_t^n \geq 1 - \varepsilon \text{ et } \lambda_t^n - \lambda_t^{n-1} \leq \varepsilon\}.$$

De plus,

(i) *si $q - n + 1 \geq \frac{1}{\beta} - 1$ alors $\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon = \infty$ p.s.*

(ii) *Il n'y a pas de collision multiple entre particules, i.e. pour tout $1 \leq i < j \leq n - 1$,*

$$\mathbb{P}\left\{\exists t \in (0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon) : \lambda_t^i = \lambda_t^{i+1} \text{ et } \lambda_t^j = \lambda_t^{j+1}\right\} = 0.$$

Le résultat équivalent de la Proposition 1.3.3 pour les processus de β -Jacobi s'énonce de la façon suivante, et est démontrée dans [Dem09a].

Proposition 1.3.6. *Supposons $\beta(p \wedge q - n + 1) > 0$. L'unique mesure de probabilité stationnaire du système d'EDS $(J(p, q))$ est ρ_{inv} admettant pour densité par rapport à la mesure de Lebesgue sur \mathbb{R}^n :*

$$d\rho_{inv}(\lambda^1, \dots, \lambda^n) = \frac{1}{\mathcal{Z}} \times \prod_{i=1}^n \left((\lambda^i)^{\beta \frac{p-n+1}{2} - 1} (1 - \lambda^i)^{\beta \frac{q-n+1}{2} - 1} \prod_{j \neq i} |\lambda^j - \lambda^i|^{\beta/2} \right) \mathbb{1}_{\{0 \leq \lambda^1 \leq \dots \leq \lambda^n \leq 1\}} d\lambda^1 \dots d\lambda^n,$$

où \mathcal{Z} est une constante de normalisation.

Plus précisément, si une solution $\Lambda = (\lambda_t^1, \dots, \lambda_t^n)_t$ de $(J(p, q))$ est telle que la distribution de Λ_t ne dépend pas de t , alors cette distribution est ρ_{inv} . Réciproquement, $(J(p, q))$ admet une unique solution Λ de condition initiale Λ_0 distribuée selon ρ_{inv} et indépendante du mouvement Brownien \mathbf{B} . De plus, pour tout $t \in \mathbb{R}_+$, Λ_t est distribuée selon ρ_{inv} .

On reconnaît ici en ρ_{inv} la densité de l'ensemble de β -Jacobi. On montre de plus que si la solution Λ est globale, alors la distribution de $(\lambda_t^1, \dots, \lambda_t^n)$ converge étroitement vers ρ_{inv} quand $t \rightarrow +\infty$.

Comme pour les β -Wishart, une étape clé dans la preuve de ces résultats est d'étudier le système d'EDS obtenu en calculant formellement $d \arcsin(\sqrt{\lambda_t^i})$ en appliquant la formule d'Itô :

$$d\phi_t^i = dB_t^i + \left\{ \frac{\beta(p-q)}{2} \cot \phi_t^i + (\beta(q-n+1) - 1) \cot(2\phi_t^i) + \frac{\beta}{2} \sum_{j \neq i} [\cot(\phi_t^i + \phi_t^j) + \cot(\phi_t^i - \phi_t^j)] \right\} dt \quad (1.3.8)$$

pour tout $i \in \{1, \dots, n\}$,
 $0 \leq \phi_t^1 < \dots < \phi_t^n \leq \frac{\pi}{2}$, p.s. dt - p.p.,

qui peut s'écrire

$$d\phi_t^i = dB_t^i - \partial_i V(\phi_t^1, \dots, \phi_t^n) dt \text{ pour tout } i \in \{1, \dots, n\}, \quad (1.3.9)$$

avec

$$V(\phi^1, \dots, \phi^n) = - \sum_{i=1}^n \left\{ \frac{\beta(p-q)}{2} \ln |\sin \phi^i| + \frac{\beta(q-n+1)-1}{2} \ln |\sin(2\phi_t^i)| \right. \\ \left. + \frac{\beta}{4} \sum_{j \neq i} (\ln |\sin(\phi^i + \phi^j)| + \ln |\sin(\phi^i - \phi^j)|) \right\}. \quad (1.3.10)$$

Liens avec d'autres points de vue sur les systèmes de particules

Plus généralement que le système (1.3.1), les systèmes de particules en interaction du type

$$dx_t^i = b_i(x_t^i) dt + \sigma_i(x_t^i) dB_t^i - \partial_i V(x_t^1, \dots, x_t^n) dt, \quad (1.3.11)$$

où $V : \mathbb{R}^n \mapsto (-\infty, +\infty]$ ont été étudiés par de nombreux auteurs. En ajoutant une moyennisation $\frac{1}{n}$ devant V , elles apparaissent dans le livre de *Henry P. McKean* [McK69], et chez *Alain-Sol Sznitman* dans le contexte de la propagation du chaos, qui relève de nombreuses autres références dans son cours [Szn91]. Dans le cas particulier $\sigma_i(x^i) = \sigma > 0$, $b_i(x^i) = 0$ et $V(x^1, \dots, x^n) = -\frac{\beta}{2} \sum_{i=1}^n \sum_{j \neq i} \ln |x^i - x^j| + \frac{\theta}{2} \sum_{i=1}^n (x^i)^2$ avec $\beta \geq \frac{\sigma^2}{2}$ et $\theta > 0$, l'existence et l'unicité d'une solution forte du système (1.3.11) ont été prouvé dans [RS93].

Les systèmes du type (1.3.11) où b_i et σ_i sont des fonctions Lipschitz, où V est une fonction convexe semi-continue inférieurement telle que $D = \{x \in \mathbb{R}^n : V(x) < +\infty\}$ est un ensemble convexe non vide et V est continuellement différentiable sur D , et où la condition initiale $(x_0^1, \dots, x_0^n) \in \bar{D}$ ont été étudié en profondeur par *Cépa* et *Lépingle* par exemple dans [CL97] et [CL01] où ils appliquent la *théorie des équations multivoques* de Cépa développée dans [Cép94] et [Cép95]. Cette théorie traite les problèmes d'existence et d'unicité de solution d'EDS multivoques associées à une fonction convexe définie sur un domaine D de \mathbb{R}^n . Dans [Lé10], Lépingle applique cette théorie à un mouvement Brownien contraint à rester dans une chambre de Weyl en rebondissant ou en étant repoussé par ses frontières. Le comportement du potentiel convexe aux frontières de D impose le comportement du processus sur ces mêmes frontières (s'il atteint ou non la frontière en temps fini, s'il est réfléchi sur la frontière, etc.). Nos systèmes d'EDS de β -Wishart et de β -Jacobi réécrites sous la forme de diffusion gradient après changement de variables peuvent être vues sous cet angle, et cette connexion est exploitée dans la démonstration des Théorèmes 1.3.2 et 1.3.5.

Les processus de β -Wishart et de β -Jacobi furent également abordés par une autre veine de la littérature s'intéressant aux processus de Dunkl radiaux (voir [Chy06, Dem09b, GY+06] pour une introduction détaillée). On définit un système de racines R par un ensemble fini $\mathbb{R}^n \setminus \{0\}$ générant l'espace vectoriel \mathbb{R}^n tel que

- pour tout $\alpha \in R$, $R \cap \mathbb{R}\alpha = \{\alpha, -\alpha\}$,
- pour tout $\alpha \in R$, $\sigma_\alpha(R) = R$,

où σ_α est la réflexion par rapport à l'hyperplan orthogonal à α . Un système simple Δ est une base de \mathbb{R}^n imposant un ordre total sur R de la manière suivante : une racine $\alpha \in R$ est positive si elle s'écrit comme une combinaison linéaire à coefficients positifs d'éléments de Δ . Un système simple Δ étant fixé, on peut alors définir l'ensemble des racines positives de R que l'on note R_+ . Quand $\sigma_i = 1$ et V prend la forme

$$V : x \mapsto - \sum_{\alpha \in R_+} k(\alpha) \ln(\langle \alpha, x \rangle), \quad x \in D,$$

où D est la chambre de Weyl définie par

$$D = \{x \in \mathbb{R}^n, \langle \alpha, x \rangle > 0 \forall \alpha \in R_+\},$$

Demni démontre dans [Dem09b, Theorem 1] l'existence et l'unicité d'une solution forte à (1.3.11) sur le domaine \bar{D} quand $k(\alpha) > 0$ pour tout $\alpha \in R_+$. Pour ce faire, il applique la théorie des équations multivoques développée par Cépa évoquée plus tôt. Ce système correspond à notre système β -Wishart (1.3.2) pour un certain choix de R_+ et des k . En remplaçant la fonction convexe $\phi \mapsto -\ln(\langle \alpha, \phi \rangle)$ par $\phi \mapsto -\ln(\sin(\langle \alpha, \phi \rangle))$ dans V , et pour un certain choix de R_+ et des k , ce système correspond également à notre système β -Jacobi ($J(p,q)$).

On retrouve aussi les processus de β -Wishart et β -Jacobi dans les travaux de *Piotr Graczyk* et de *Jacek Malecki* ([GM13], [GM14]) qui s'intéressent à des systèmes d'EDS de la forme

$$d\lambda_t^i = \sigma_i(\lambda_t^i) dB_t^i + \left(b_i(\lambda_t^i) + \sum_{j \neq i} \frac{H_{i,j}(\lambda_t^i, \lambda_t^j)}{\lambda_t^i - \lambda_t^j} \right) dt, \quad \text{for all } i \in \{1, \dots, n\}$$

$$\lambda_t^1 \leq \dots \leq \lambda_t^n, \quad t \geq 0.$$

où les fonctions σ_i, b_i et $H_{i,j}$ sont continues, et où $H_{i,j}$ est positive et symétrique au sens où $H_{i,j}(x, y) = H_{j,i}(y, x)$ pour tout $x, y \in \mathbb{R}$. Les processus de β -Wishart et β -Jacobi sont des cas particuliers de ces systèmes de particules pour les bons choix de σ_i, b_i et $H_{i,j}$. La motivation première des auteurs est liée à des préoccupations matricielles évoquées par Bru dans [Bru91] : démontrer un théorème de Yamada-Watanabe multidimensionnel. C'est ainsi qu'ils considèrent dans [GM13] ([GM14]) l'EDS à valeur matricielle sur $\mathcal{S}_n(\mathcal{H}_n)$, l'espace des matrices symétriques (hermitiennes) de taille $n \times n$, de la forme

$$dX_t = g(X_t) dW_t h(X_t) + h(X_t) d(W_t)^* g(X_t) + b(X_t) dt, \quad X_0 \in \mathcal{H}_n, \quad (1.3.12)$$

où les fonctions $g, h, b : \mathbb{R} \rightarrow \mathbb{R}$ sont continues et s'appliquent sur le spectre de X_t , i.e. si $X \in \mathcal{H}_n$, U unitaire et D diagonale sont telles que $X = U^\dagger D U$, alors g agit comme l'application $\mathcal{H}_n \ni X \mapsto U^\dagger \text{Diag}(g(D_{1,1}), \dots, g(D_{n,n})) U$. Ici, $W = (W_t)_t$ est une matrice de taille $n \times n$ dont les entrées sont des mouvements Browniens complexes (voir Chapitre 4

pour plus de précisions). Cette formulation générale englobe les équations pour les modèles de β -Wishart et de β -Jacobi (en particulier (1.3.3) et (1.3.7)), comme c'est indiqué dans [GM13]. Si $(X_t)_t$ est solution de (1.3.12), il est prouvé dans [GM14] que les processus valeurs propres $(\lambda_t^1, \dots, \lambda_t^n)_t$ de la solution $(X_t)_t$ satisfont le système d'EDS suivant :

$$d\lambda_t^i = 2g(\lambda_t^i)h(\lambda_t^i)dB_t^i + 2 \left(b(\lambda_t^i) + \sum_{j \neq i} \frac{G(\lambda_t^i, \lambda_t^j)}{\lambda_t^i - \lambda_t^j} \right) dt, \text{ for all } i \in \{1, \dots, n\} \quad (1.3.13)$$

$$\lambda_t^1 \leq \dots \leq \lambda_t^n, \quad t \geq 0.$$

où $G(x, y) = g^2(x)h^2(y) + g^2(y)h^2(x)$.

Après l'étude de résultats d'existence et d'unicité, de mesures stationnaires ainsi que de comportement en temps long de ces processus de β -Wishart et β -Jacobi, une question naturelle est de s'interroger sur leur comportement en grande dimension.

1.3.2 Systèmes de particules en grande dimension

De nombreux travaux ont pour objectif de décrire la mesure empirique asymptotique de systèmes de particules issus de matrices aléatoire, comme [DG01, PPA16, JPP⁺19]. Certains des travaux cités précédemment dans le contexte des systèmes d'EDS ([RS93, CL97, CL01]) se sont de même intéressés au comportement asymptotique en grande dimension de la mesure empirique des particules. Les méthodes pour prouver ce type de convergence reposent généralement dans la littérature sur l'étude de l'équation aux dérivées partielles (EDP) de Burgers complexe que l'on trouve pour la transformée de Cauchy-Stieltjes de la mesure empirique asymptotique ([RS93], [CL01]). L'EDP en question se résoud quand c'est possible par la méthode des caractéristiques ou par une transformation de Hopf-Cole. *José Luis Pérez* et *Malecki* ont étudié les solutions de (1.3.13) sous cet angle en étudiant les moments de la mesure empirique des particules dans [MP19] (voir aussi [SYY19, TT20b, ABMV13]). Dans ces travaux, on trouve des résultats d'existence de la limite en grande dimension de la mesure empirique des particules, ainsi qu'une forme explicite de cette limite quand la condition initiale est un dirac en zéro. Nos résultats sur les β -Wishart généralisent les précédents en partant de conditions initiales dont la seule contrainte est que le huitième moment est fini. On couple pour cela l'étude d'une EDP de Burgers complexe à des outils issus des probabilités libres rectangulaires de [BG09].

L'objet limite peut être interprété comme la loi d'un processus de Wishart libre, défini dans l'article [CDM05]. On notera que le même type de travail fut réalisé pour construire le processus de Jacobi libre ([Dem08]), dont la recherche de la loi, de même que dans le cas Wishart, est envisagée comme une piste de recherche future. Le lecteur trouvera dans [Bia97, BS98, BS01] une introduction à la théorie des processus stochastiques libres.

Mesure empirique des beta-Wishart en grande dimension

Nous introduisons ici une forme généralisée des processus de β -Wishart englobant les différentes définitions que l'on peut trouver dans la littérature. Soient $\beta_1, \beta_2 > 0$, $\kappa \geq 0$, $m \geq n$

et considérons le *processus de Wishart généralisé* défini par le système d'EDS :

$$d\lambda_t^{i,n,W} = \frac{2\kappa}{\sqrt{m}} \sqrt{\lambda_t^{i,n,W}} dB_t^i - 2\gamma \lambda_t^{i,n,W} dt + \beta_1 \kappa^2 \left(1 + \frac{\beta_2}{m} \sum_{j \neq i} \frac{\lambda_t^{i,n,W} + \lambda_t^{j,n,W}}{\lambda_t^{i,n,W} - \lambda_t^{j,n,W}} \right) dt \text{ pour tout } i \in \{1, \dots, n\}$$

$$0 \leq \lambda_t^{1,n,W} < \dots < \lambda_t^{n,n,W} \text{ p.s. dt-p.p.} \quad (1.3.14)$$

où B^1, \dots, B^n sont des mouvements Browniens indépendants.

Définissons pour tout $t \geq 0$:

$$\nu_t^{n,W} = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_t^{i,n,W}}, \quad \mu_t^{n,W} = \sqrt{\nu_t^{n,W}} = \frac{1}{2n} \sum_{i=1}^n \left(\delta_{\sqrt{\lambda_t^{i,n,W}}} + \delta_{-\sqrt{\lambda_t^{i,n,W}}} \right).$$

On a le résultat suivant, que l'on trouve déjà dans [MP19] mais que l'on rappelle dans le cas du processus de Wishart généralisé, et qui exhibe les étapes classiques de l'étude en grande dimension de la mesure spectrale empirique d'un système de particules. On exhibe notamment l'équation intégrale-différentielle vérifiée par les valeurs d'adhérence de la suite $((\nu_t^{n,W})_t)_n$ en (1.3.15), ainsi que l'EDP de Burgers complexe (1.3.16) vérifiée par les transformées de Cauchy-Stieltjes des valeurs d'adhérences de $((\nu_t^{n,W})_t)_n$.

Théorème 1.3.7 (Burgers complexe et limite de champ moyen : formulation faible). *Supposons que la suite de mesures $(\nu_0^{n,W})_n$ converge faiblement en probabilité quand $n \rightarrow +\infty$ avec $\frac{n}{m} \rightarrow \alpha \in (0, 1]$ vers une mesure de probabilité limite notée ν_0^W et que*

$$\sup_n \int x^8 \nu_0^{n,W}(dx) < \infty.$$

Alors la famille $\{(\nu_t^{n,W})_{t \geq 0}, n \in \mathbb{N}^\}$ est tendue et tout flot de mesures limite quand n tend vers l'infini avec $\frac{n}{m} \rightarrow \alpha \in (0, 1]$ satisfait pour toute fonction test f réelle doublement continuellement différentiable l'équation :*

$$\langle \nu_t, f \rangle = \langle \nu_0, f \rangle + \int_0^t \langle \nu_s, (\beta_1 \kappa^2 - 2\gamma \Phi) f' \rangle ds + \frac{\alpha \beta_1 \beta_2 \kappa^2}{2} \int_0^t \left(\iint (x+y) \frac{f'(x) - f'(y)}{x-y} \nu_s(dx) \nu_s(dy) \right) ds \quad (1.3.15)$$

où $\Phi : x \mapsto x$ et avec la convention $\frac{f'(x) - f'(y)}{x-y} = f''(x)$ quand $x = y$. Si ν_0 admet une fonction caractéristique, et si cette fonction est analytique dans un voisinage de l'origine, alors cette équation admet une unique solution.

De plus, si $(\nu_t)_{t \geq 0}$ est une solution de l'équation (1.3.15) et si pour tout $t \geq 0$, G_t est la transformée de Cauchy-Stieltjes de ν_t , alors G est solution de l'EDP de Burgers complexe

$$\frac{\partial}{\partial t} G_t(z) = (\alpha \beta_1 \beta_2 \kappa^2 - \beta_1 \kappa^2 + 2\gamma z) \frac{\partial}{\partial z} G_t(z) - 2\alpha \beta_1 \beta_2 \kappa^2 z G_t(z) \frac{\partial}{\partial z} G_t(z) - \alpha \beta_1 \beta_2 \kappa^2 G_t^2(z) + 2\gamma G_t(z),$$

$$G_0(z) = \int \frac{\nu_0^W(dv)}{z-v} = \phi(z). \quad (1.3.16)$$

Le résultat suivant donne une réponse complète à la limite de la suite $((\mu_t^{n,W})_t)_n$ et donc $((\nu_t^{n,W})_t)_n$ pour toute condition initiale $\nu_0^W = \nu_0^{n,W}$ de huitième moment fini.

Théorème 1.3.8 (Limite de champ moyen pour le processus de Wishart généralisé). *Sous les hypothèses du Théorème 1.3.7, supposons de plus $\beta_2\alpha \leq 1$, que $(\lambda_t^{1,n,W}, \dots, \lambda_t^{n,n,W})_t$ est définie sur \mathbb{R}_+ et que $\mu_0^W = \sqrt{\nu_0^W}$. Supposons de plus que ν_0^W admet une fonction caractéristique analytique dans un voisinage de zéro. Alors pour tout $t \geq 0$, $(\mu_t^{n,W})_{n \geq 0}$ converge faiblement en probabilité et*

$$\mu_t^W := \lim_{n \rightarrow \infty} \mu_t^{n,W} = (e^{-\gamma t} \mu_0^W) \boxplus_{\beta_2\alpha} \sqrt{\mu^{\text{MP}_{\beta_2\alpha, \sigma_t}}}$$

où $\boxplus_{\beta_2\alpha}$ est la convolution libre rectangulaire de paramètre $\beta_2\alpha$, où $e^{-\gamma t} \mu_0^W$ est la mesure image de μ_0^W par l'application $v \mapsto e^{-\gamma t} v$, et où pour tout $t \geq 0$:

$$\sigma_t^2 = \begin{cases} \frac{\beta_1 \kappa^2}{2\gamma} (1 - e^{-2\gamma t}) & \text{if } \gamma \neq 0 \\ \beta_1 \kappa^2 t & \text{if } \gamma = 0. \end{cases}$$

De plus, si $\gamma \neq 0$ et avec $\sigma_\infty = \sqrt{\frac{\beta_1 \kappa^2}{2\gamma}}$,

$$\lim_{t \rightarrow +\infty} \mu_t^W = \sqrt{\mu^{\text{MP}_{\beta_2\alpha, \sigma_\infty}}}.$$

On conclut de plus sur la commutativité des limites en temps long et en grande dimension.

Théorème 1.3.9 (Commutativité des limites). *Sous les hypothèses du Théorème 1.3.8, supposons de plus $\gamma \neq 0, \beta_2 = 1$ et que pour tout $n \in \mathbb{N}^*$: $\sum_{i=1}^n \mathbb{E}[\lambda_0^{i,n,W}] < +\infty$. Notons $\sigma_\infty = \sqrt{\frac{\beta_1 \kappa^2}{2\gamma}}$. Alors on a au sens de la convergence faible*

$$\lim_{\substack{n \rightarrow \infty \\ n/m \rightarrow \alpha}} \lim_{t \rightarrow \infty} \nu_t^{n,W} = \lim_{t \rightarrow \infty} \lim_{\substack{n \rightarrow \infty \\ n/m \rightarrow \alpha}} \nu_t^{n,W} = \mu^{\text{MP}_{\alpha, \sigma_\infty}}.$$

1.4 Matrices aléatoires et réseaux de neurones

Les algorithmes de réseaux de neurones profonds, ou d'apprentissage profond, sont à l'origine d'avancées majeures en intelligence artificielle. Leur popularisation récente est due à deux principaux facteurs matériels : l'augmentation de la capacité de stockage des données, dont le grand nombre est nécessaire au bon fonctionnement de ces algorithmes, et l'augmentation de la capacité de calcul des ordinateurs. À la suite des travaux fondateurs [RHW85], [LBD+89] et [LBBH98], l'introduction des réseaux convolutifs dans [KSH12] eut un impact considérable pour ce domaine de recherche en dominant la compétition de reconnaissance d'images ImageNet (voir [DDS+09]). Depuis, les réseaux de neurones profonds furent d'une efficacité redoutable en surpassant les méthodes pré-existantes dans de nombreux domaines, tels que le traitement automatique du langage (voir par exemple [WCA14, Rud18]), la reconnaissance

vocale (voir par exemple [HDY⁺12]), dans l'amélioration de l'Imagerie par Résonance Magnétique (voir [ZKS⁺19]) ou plus récemment dans la résolution du problème du repliement des protéines avec le travail [SEJ⁺20]. L'apprentissage profond a en parallèle ouvert de nouvelles perspectives dans des domaines plus théoriques, comme en mécanique quantique (voir par exemple [SAC⁺17]), ou même dans la résolution de certains problèmes de mathématiques symboliques (voir par exemple [LC19]).

Si l'efficacité des réseaux de neurones profonds dans la résolution de nombreuses tâches est considérable, la compréhension de leur fonctionnement est encore aujourd'hui une question largement ouverte. En effet, non seulement ces réseaux ont de très nombreux paramètres, ce qui pour des modèles statistiques classiques mènerait à du surapprentissage, mais la phase d'apprentissage de ces paramètres se fait en apprentissage supervisé par une procédure de minimisation d'une fonction de coût hautement non convexe, dont, dans la pratique, on ne retrouve jamais le minimum global. Cette procédure de recherche de minimum se fait en utilisant des algorithmes de descente de gradient stochastique (voir [Bot98]), où le gradient est évalué par une méthode de rétro-propagation (voir [LBOM12]). Comme souligné dans [CHM⁺15], la forme générale de la fonction de coût est encore mal comprise. Cette fonction possède de nombreux minima locaux, et des expérimentations multiples ont montré que la performance du réseau de neurones semble peu dépendre du minimum local vers lequel converge la descente de gradient. En revanche, la convergence vers le minimum global, rare dans des réseaux de neurones profonds, mène souvent à une situation de surapprentissage.

Pour mieux comprendre le paysage des points critiques de la fonction de coût, le travail [CHM⁺15] tire parti de la similarité de cette fonction, sous certaines hypothèses sur le réseau, avec l'hamiltonien des verres de spin étudié dans [AA⁺13, AAČ13]. Dans ces travaux, les auteurs décrivent la répartition des points critiques de l'hamiltonien à l'aide de la formule de Kac-Rice (voir [AT09]) qui lie le décompte des points critiques à l'intégrale du déterminant de la hessienne de l'hamiltonien étudié selon les lignes de niveau de ce dernier. Cette intégrale est reliée au comportement des valeurs propres du GOE.

De nombreux travaux ont pour objectif de comprendre la forme de la sortie d'un réseau de neurones en fonction de son entrée, de ses non-linéarités et de ses matrices de poids qui seront définies dans le prochain paragraphe, et les phénomènes de concentration au sein du réseau. Ils étudient pour cela à l'aide de la théorie des matrices aléatoires des réseaux de neurones à entrées aléatoires et matrices de poids aléatoires (voir [PW17, LLC⁺18, BP19, FW20, CL20]).

L'étude des réseaux de neurones aléatoires trouve de plus son intérêt dans l'étude de l'initialisation de la phase d'apprentissage. Des travaux comme [GB10] montrent l'intérêt d'une initialisation aléatoire des poids et des biais du réseau. Cette initialisation est très utilisée en pratique et implémentée dans diverses bibliothèques de référence pour les praticiens telle que *PyTorch*. En partant du constat que la concentration du spectre du Jacobien d'un réseau de neurones autour de 1 à l'initialisation du réseau est déterminante dans la rapidité de la phase d'apprentissage des paramètres, *Jeffrey Pennington, Samuel S. Schoenholz et Surya Ganguli* étudient dans [PSG18] l'impact des choix de profondeur du réseau, de non-linéarité, et de variance à l'initialisation sur les phénomènes d'explosion ou de disparition du gradient lors de la procédure rétropropagation en utilisant des outils de probabilité libre. Les couches du modèle de réseau qu'ils étudient sont de largeur constante. En collaboration avec *Reda*

Chhaibi et Tariq Daouda, et à la lumière du papier [PS20], nous cherchons à généraliser cette approche à des réseaux dont les couches sont de largeur variable, et dont le choix des non-linéarités est libre pour chaque couche. Cette nouvelle approche fait intervenir des matrices rectangulaires plutôt que carrées, et demande la définition d’une convolution libre multiplicative rectangulaire, plus simple à définir que la convolution libre additive rectangulaire définie dans [BG09].

D’autres travaux exploitent cette connexion entre réseaux de neurones et matrices aléatoires afin de mieux comprendre le comportement des algorithmes d’apprentissage profond tels que [Yan19, Yan20a, Yan20b].

1.4.1 Description du modèle

Considérons un réseau de neurones à propagation avant de profondeur $L \in \mathbb{N}$, c’est à dire à L couches. À la profondeur $\ell = 1, 2, \dots, L$, chaque couche a pour activation $x_\ell \in \mathbb{R}^{N_\ell}$, où N_ℓ est la largeur ou la taille de la couche ℓ . Le vecteur $x_0 \in \mathbb{R}^{N_0}$ est l’entrée du réseau de neurones qui renvoie la sortie $x_L \in \mathbb{R}^{N_L}$. Le vecteur de la taille des couches s’écrit :

$$\mathbf{N} := (N_0, N_1, \dots, N_L) .$$

On a la relation de récurrence suivante entre les couches :

$$x_\ell = \phi_\ell \left(W_\ell^{(\mathbf{N})} x_{\ell-1} + b_\ell^{(\mathbf{N})} \right) ,$$

où ϕ_ℓ est une non-linéarité qui s’applique par entrées, $W_\ell^{(\mathbf{N})} \in M_{N_\ell, N_{\ell-1}}(\mathbb{R})$ est une matrice de poids et $b_\ell^{(\mathbf{N})} \in \mathbb{R}^{N_\ell}$ est le vecteur des biais. On écrit $h_\ell := W_\ell^{(\mathbf{N})} x_{\ell-1} + b_\ell^{(\mathbf{N})}$ pour les pré-activations.

La matrice Jacobienne calculée pendant la rétro-propagation peut s’écrire explicitement en utilisant la règle de dérivation des fonctions composées. On a alors :

$$\begin{aligned} J^{(\mathbf{N})} &:= \frac{\partial x_L}{\partial x_0} & (1.4.1) \\ &= \frac{\partial x_L}{\partial x_{L-1}} \frac{\partial x_{L-1}}{\partial x_{L-2}} \cdots \frac{\partial x_1}{\partial x_0} \\ &= D_L^{(\mathbf{N})} W_L^{(\mathbf{N})} D_{L-1}^{(\mathbf{N})} W_{L-1}^{(\mathbf{N})} \cdots D_1^{(\mathbf{N})} W_1^{(\mathbf{N})} , \end{aligned}$$

où les matrices D_ℓ sont les matrices diagonales telles que :

$$\left[D_\ell^{(\mathbf{N})} \right]_{i,i} = \phi'_\ell ([h_\ell]_i) .$$

L’importance de la matrice Jacobienne pendant la phase d’apprentissage peut se comprendre de la façon suivante. Une étape de descente de gradient modifie les poids et les biais selon :

$$\left(W_\ell^{(\mathbf{N})}, b_\ell^{(\mathbf{N})} \right) \leftarrow \left(W_\ell^{(\mathbf{N})}, b_\ell^{(\mathbf{N})} \right) - \alpha \frac{\partial \mathcal{L}}{\partial (W_\ell^{(\mathbf{N})}, b_\ell^{(\mathbf{N})})} , \quad (1.4.2)$$

pour tout $\ell = 1, \dots, L$. Le coefficient $\alpha > 0$ est le taux d'apprentissage et \mathcal{L} la fonction de coût sur un mini-lot (minibatch en anglais). Si le mini-lot est de taille $B \in \mathbb{N}$, et correspond à un petit échantillon $((X_i, Y_i) ; i = 1, \dots, B)$ des données d'entraînement, on a :

$$\mathcal{L} = \frac{1}{B} \sum_{i=1}^B d(x_L(X_i), Y_i) .$$

Ici d est une distance, les X_i sont pris en entrée du réseau de neurones et les Y_i sont les sorties, c'est à dire les labels dans le cas d'un classifieur, $Y_i = X_i$ dans le cas d'un auto-encodeur etc. En dérivant les fonctions composées on obtient :

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial (W_\ell^{(\mathbf{N})}, b_\ell^{(\mathbf{N})})} &= \frac{\partial \mathcal{L}}{\partial x_L} \frac{\partial x_L}{\partial x_{L-1}} \frac{\partial x_{L-1}}{\partial x_{L-2}} \cdots \frac{\partial x_{\ell+1}}{\partial x_\ell} \frac{\partial x_\ell}{\partial (W_\ell^{(\mathbf{N})}, b_\ell^{(\mathbf{N})})} \\ &= \frac{\partial \mathcal{L}}{\partial x_L} J_\ell^{(\mathbf{N})} \frac{\partial h_\ell}{\partial (W_\ell^{(\mathbf{N})}, b_\ell^{(\mathbf{N})})} , \end{aligned} \quad (1.4.3)$$

où

$$\frac{\partial \mathcal{L}}{\partial x_L} = \frac{1}{B} \sum_{i=1}^B \partial_1 d(x_L(X_i), Y_i) \in M_{1, N_L}(\mathbb{R}) , \quad (1.4.4)$$

$$J_\ell^{(\mathbf{N})} = D_L^{(\mathbf{N})} W_L^{(\mathbf{N})} \dots D_{\ell+1}^{(\mathbf{N})} W_{\ell+1}^{(\mathbf{N})} D_\ell^{(\mathbf{N})} \in M_{N_L, N_\ell}(\mathbb{R}) . \quad (1.4.5)$$

Le problème auquel nous nous sommes intéressés est le suivant : un grand produit de grandes matrices peut facilement devenir instable au sens où

- si un grand nombre de valeurs singulières de $J_\ell^{(\mathbf{N})}$ sont $\ll 1$, on est confronté à un problème de disparition du gradient.
- Si un grand nombre de valeurs singulières de $J_\ell^{(\mathbf{N})}$ sont $\gg 1$, on est confronté à un problème d'explosion du gradient.

Intuition. Cette instabilité peut se comprendre aisément en considérant le cas plus simple de la dimension un, pour lequel une suite géométrique q^n quand $n \rightarrow \infty$ est l'archétype d'un long produit. Cette suite converge extrêmement vite, vers 0 si $|q| < 1$ ou vers l'infini si $|q| > 1$.

Une façon moins naïve de se forger une intuition du problème est d'observer que l'échantillonnage des mini-lots (1.4.4) est très bruité. Sans davantage d'informations sur les données et sur la méthode d'échantillonnage, on peut faire l'hypothèse que $\frac{\partial \mathcal{L}}{\partial x_L}$ se comporte comme un vecteur gaussien de matrice de covariance proportionnelle à I_{N_L} . Ainsi, chaque pas de gradient $\alpha \frac{\partial \mathcal{L}}{\partial (W_\ell^{(\mathbf{N})}, b_\ell^{(\mathbf{N})})}$ dans (1.4.2) est approximativement un vecteur gaussien de matrice de covariance proportionnelle à :

$$\alpha^2 \left(\frac{\partial h_\ell}{\partial (W_\ell^{(\mathbf{N})}, b_\ell^{(\mathbf{N})})} \right)^T (J_\ell^{(\mathbf{N})})^T J_\ell^{(\mathbf{N})} \frac{\partial h_\ell}{\partial (W_\ell^{(\mathbf{N})}, b_\ell^{(\mathbf{N})})} .$$

Le comportement du spectre de $(J_\ell^{(\mathbf{N})})^T J_\ell^{(\mathbf{N})}$ est donc déterminant pour la stabilité de la descente de gradient, et les vecteurs propres de $(J_\ell^{(\mathbf{N})})^T J_\ell^{(\mathbf{N})}$ sont les directions selon lesquelles l'intuition en dimension un s'applique.

Dans un but de simplification, nous nous concentrons sur la matrice Jacobienne $J^{(\mathbf{N})}$ donnée par (1.4.1) qui est proche de la matrice $J_0^{(\mathbf{N})}$ donnée par (1.4.5) à multiplication près à droite par $D_0^{(\mathbf{N})}$.

Aléa. Dans le travail pionnier de *Xavier Glorot* et *Yoshua Bengio* [GB10] sur les initialisations aléatoires, il fut suggéré que les propriétés spectrales de $J^{(\mathbf{N})}$ sont un très bon indicateur de stabilité et de performance d'apprentissage. Depuis, différentes initialisations aléatoires furent proposées et sont aujourd'hui implémentées dans les bibliothèques modernes d'apprentissage automatique [PGC⁺17, AAB⁺15].

Nous avons donc pris dans ce travail des choix classiques d'initialisation aléatoire. Les biais $b_\ell^{(\mathbf{N})}$ sont des vecteurs dont les entrées sont des variables aléatoires normales centrées i.i.d. de variance $\sigma_{b_\ell}^2$. Concernant les matrices de poids, nous considérons les entrées $[W_\ell^{(\mathbf{N})}]_{i,j}$ comme des variables aléatoires centrées i.i.d. de variance $\sigma_{W_\ell}^2/N_\ell$ et de quatrième moment fini comme dans [PS20], dont le travail théorique permet de justifier la liberté requise dans notre travail. En effet, sur les traces de [PSG18], nous utilisons des outils issus des probabilités libres pour analyser les valeurs singulières de la matrice $J^{(\mathbf{N})}$ quand la largeur du réseau tend vers l'infini. Cette limite en grande dimension, i.e. en largeur infinie des couches est particulièrement intéressante dans l'étude de grands réseaux de neurones profonds, notamment parce que le régime décrit par les probabilités libres apparaît déjà pour des réseaux de largeur relativement petite par des propriétés de concentration, voir [LLC⁺18].

1.4.2 Contributions

L'analyse des propriétés spectrales de J nécessite de donner le résultat suivant sur les produits de matrices rectangulaires.

Théorème 1.4.1. *Considérons trois suites d'entiers $(p_N)_{N \geq 1}, (q_N)_{N \geq 1}, (r_N)_{N \geq 1}$, telles que*

$$p_N, q_N, r_N \xrightarrow{N \rightarrow \infty} \infty, \quad \frac{r_N}{q_N} \xrightarrow{N \rightarrow \infty} c > 0 .$$

Considérons pour tout $N \geq 1$ les matrices aléatoires $A^{(N)}, B^{(N)}$ de tailles respectives $p_N \times q_N$ et $q_N \times r_N$ dont les mesures empiriques des carrés des valeurs singulières convergent faiblement. Supposons de plus que les matrices $B^{(N)} (B^{(N)})^T$ et $(A^{(N)})^T A^{(N)}$ sont asymptotiquement libres. Alors on a quand $N \rightarrow \infty$:

$$S_{(AB)^T AB}(m) = S_{A^T A}(cm) S_{B^T B}(m) ,$$

où ici et dans la suite, on omet l'exposant (N) pour indiquer l'opérateur asymptotique associé à la limite de la mesure spectrale empirique de la suite de matrices en question.

Dans ce résultat, la fonctionnelle S est celle donnée dans la définition 1.1.7.

Ce résultat définit implicitement une convolution libre rectangulaire multiplicative, dans le même esprit que la convolution libre rectangulaire développée dans [BG09]. Cependant, associer un coefficient c à cette convolution en la notant \boxtimes_c semble maladroit puisqu'elle ne serait alors pas associative. Une solution plus raisonnable est d'associer un ratio c à chaque mesure, comme chaque matrice à laquelle on associe un rapport de dimensions. On a alors la définition suivante :

Définition 1.4.2. On définit l'opération \boxtimes sur l'ensemble des paires (μ, c) des mesures de probabilités sur \mathbb{R}_+ différentes δ_0 et des réels strictement positifs par :

$$(\mu_1, c_1) \boxtimes (\mu_2, c_2) := (\nu, c_1 c_2) ,$$

où ν est l'unique mesure de probabilité sur \mathbb{R}_+ telle que $S_\nu = S_{\mu_1}(c_2 \cdot) S_{\mu_2}$. Cette définition étend la définition de la convolution libre multiplicative usuelle qui se retrouve de la façon suivante :

$$(\mu_1, 1) \boxtimes (\mu_2, 1) := (\mu_1 \boxtimes \mu_2, 1) .$$

Cette opération est bien associative et permet une formulation commode de la mesure qui nous intéresse dans ce travail, la mesure empirique du carré des valeurs singulières de $J^{(\mathbf{N})}$.

On choisira ici les notation suivante : soit une suite de matrices $A^{(\mathbf{N})}$ de taille $N_\ell \times N_{\ell-1}$ telle que quand $N_{\ell-1}$ et N_ℓ tendent vers l'infini, le ratio $\frac{N_{\ell-1}}{N_\ell}$ converge vers λ_ℓ . Alors si la mesure empirique des carrés des valeurs singulières de $A^{(\mathbf{N})}$ converge faiblement, on notera ν_A cette limite. On notera de plus $M_{A^T A}$ et $S_{A^T A}$ respectivement la fonction génératrice des moments et la S-transformée de ν_A .

Sous les bonnes hypothèses motivées par les travaux [PLR⁺16, SGGSD16, PSG18] et justifiées plus rigoureusement dans [Pas20, PS20], pour tout $\ell = 1, \dots, L$, les mesures des carrés des valeurs singulières de $W_\ell^{(\mathbf{N})}$ et $D_\ell^{(\mathbf{N})}$ convergent respectivement vers ν_{W_ℓ} et ν_{D_ℓ} , et on a les propriétés de liberté asymptotique nécessaires pour démontrer le résultat suivant.

Théorème 1.4.3. *Supposons qu'en régime de largeur infinie du réseau, c'est à dire $N_\ell \rightarrow \infty$, on a $\ell = 0, 1, 2, \dots, L$:*

$$\frac{N_{\ell-1}}{N_\ell} \xrightarrow{N \rightarrow \infty} \lambda_\ell > 0 ,$$

et notons

$$\Lambda_\ell := \lim_{N \rightarrow \infty} \frac{N_0}{N_\ell} = \prod_{k=1}^{\ell} \lambda_k .$$

Alors en termes de convolution libre rectangulaire multiplicative, la mesure empirique du carré des valeurs singulières de $J^{(\mathbf{N})}$ converge faiblement vers

$$\nu_J = (\nu_{D_L}, 1) \boxtimes (\nu_{W_L}, \lambda_L) \boxtimes \dots \boxtimes (\nu_{D_1}, 1) \boxtimes (\nu_{W_1}, \lambda_1) . \quad (1.4.6)$$

De plus, la S-transformée de $J^T J$ dans le régime de largeur infinie vérifie

$$S_{J^T J}(m) = \prod_{\ell=1}^L \left[S_{D_\ell^2}(\Lambda_\ell m) S_{W_\ell^T W_\ell}(\Lambda_{\ell-1} m) \right] . \quad (1.4.7)$$

En particulier, sous l'hypothèse que les entrées de W_ℓ sont *i.i.d.* :

$$S_{J^T J}(m) = \prod_{\ell=1}^L \left(S_{D_\ell^2}(\Lambda_\ell m) \frac{1}{\sigma_{W_\ell}^2} \frac{1}{1 + \Lambda_\ell m} \right),$$

$$M_{J^T J}^{-1}(m) = \frac{m+1}{m} \prod_{\ell=1}^L \frac{\sigma_{W_\ell}^2 (1 + \Lambda_\ell m)}{S_{D_\ell^2}(\Lambda_\ell m)}.$$

Le but de ce travail est de donner un algorithme efficace de calcul de ν_J . Cela permet de dégager des indications sur la stabilité de la rétro-propagation, et participe à l'objectif d'une plus grande automatisation du design des réseaux de neurones. Les contributions de ce travail sont à la fois théoriques, numériques et expérimentales, et peuvent se résumer de la façon suivante :

- La contribution théorique est incrémentale : dans [PSG18], la largeur du réseau est supposée constante. Nous généralisons le modèle avec un profil de largeur des couches variable, ce qui reflète la façon dont les réseaux de neurones sont utilisés en pratique. Cela nécessite de développer une convolution libre rectangulaire multiplicative, dont les règles de calcul, si nous ne les avons pas trouvées dans la littérature des probabilités libres, sont probablement déjà connues des experts de la communauté. Nous ne supposons pas non plus que la non linéarité ϕ_ℓ est la même dans toutes les couches du réseau.
- Notre principale contribution est numérique. Nous proposons et implémentons un algorithme efficace de calcul des densités spectrales à partir de leurs transformées complexes. Cette méthode repose sur l'inversion de S -transformées en utilisant l'algorithme de Newton-Raphson. Alors que cet algorithme de Newton-Raphson fonctionne localement, nous parvenons à une résolution globale en chaînant les bassins d'attraction de l'algorithme.

Cette contribution est d'autant plus intéressante que, même dans la communauté des probabilités libres, l'inversion de telles fonctions est considérée comme généralement impossible à réaliser en pratique (voir [BGD08]). Cette communauté s'est plutôt intéressée jusqu'à aujourd'hui à des méthodes combinatoires utilisant les moments, et des algorithmes de point fixe (voir [Tar20]). Notre algorithme est très performant, ce qui laisse place à des travaux en cours et futurs sur des problèmes plus généraux de probabilités libres comme la déconvolution.

Notre code est disponible sur Github à l'adresse <https://github.com/redachhaibi/FreeNN>.¹

- Nous comparons les courbes d'apprentissage pour différents choix de design des réseaux de neurones, éclairés ou non par les conclusions de nos résultats théoriques. Nous nous concentrons notamment sur l'utilisation des non-linéarités ReLU, Hard Tanh et

¹Au moment de l'écriture de ces lignes, le répertoire "FreeNN" est privé. Il sera rendu public lors de la soumission ou de la publication de ce travail de recherche, sous une license standard Apache 2.0.

Hard Sin. Les indications données par nos résultats théoriques pourront permettre la définition d'un indicateur de stabilité des réseaux se calculant rapidement, pouvant servir de support à des meilleurs choix de design.

Nous soulignons de plus deux méthodes utilisées communément pour palier les phénomènes de disparition ou d'explosion du gradient : l'écrêtage de gradient (gradient clipping en anglais) et la connection saute-couche (skip-connexion en anglais). La première pourrait aisément être ajoutée au modèle proposé, et la seconde est une direction de recherche future très motivante ! Nous précisons que le Chapitre 5 constitue une version encore préliminaire d'un article de recherche.

1.5 Publications

Les Chapitres 2, 3, 4 font l'objet d'une pré-publication que l'on peut retrouver par les liens suivants.

Strong solutions to a beta-Wishart particle system avec *Benjamin Jourdain*, soumis à *Journal of Theoretical Probability*

Strong solutions to a beta-Jacobi particle system, soumis à *Séminaire de Probabilités*

Wishart processes : mean-field limit, long time behavior, and free probability, soumis à *Random Matrices : Theory and Applications*

Le Chapitre 5 est un travail préliminaire ayant vocation à être publié dans les prochains mois.

Chapter 2

Strong solutions to a beta-Wishart particle system

This Section is a joint work with *Benjamin Jourdain* submitted to *Journal of Theoretical Probability* that one can find at the address <https://arxiv.org/abs/2003.08699>.

Abstract

The purpose of this paper is to study the existence and uniqueness of solutions to a Stochastic Differential Equation (SDE) coming from the eigenvalues of Wishart processes. The coordinates are non-negative, evolve as Cox-Ingersoll-Ross (CIR) processes and repulse each other according to a Coulombian like interaction force. We show the existence of strong and pathwise unique solutions to the system until the first multiple collision, and give a necessary and sufficient condition on the parameters of the SDEs for this multiple collision not to occur in finite time.

Keywords: stochastic differential equations, diffusions with gradient drift, singular interaction, random matrices.

AMS Subject Classification (2020): 60H10, 60J60, 60B20, 60G17, 60J70.

2.1 Introduction

Let $\alpha \geq 0$, $\gamma \in \mathbb{R}$, $\beta > 0$, $n \geq 1$, and $\mathbf{B} = (B_t^1, \dots, B_t^n)_t$ be a n -dimensional Brownian motion. Our SDEs system of interest is the following : for all $i \in \{1, \dots, n\}$,

$$d\lambda_t^i = 2\sqrt{\lambda_t^i}dB_t^i + \left(\alpha - 2\gamma\lambda_t^i + \beta \sum_{j \neq i} \frac{\lambda_t^i + \lambda_t^j}{\lambda_t^i - \lambda_t^j} \right) dt \quad (2.1.1)$$

$$= 2\sqrt{\lambda_t^i}dB_t^i + \left(\alpha - (n-1)\beta - 2\gamma\lambda_t^i + 2\beta\lambda_t^i \sum_{j \neq i} \frac{1}{\lambda_t^i - \lambda_t^j} \right) dt \quad (2.1.2)$$

$$0 \leq \lambda_t^1 < \dots < \lambda_t^n, \text{ a.s. } dt - \text{almost everywhere.} \quad (2.1.3)$$

The strict inequalities (2.1.3) allow the interaction terms in the system (2.1.1) to make sense. We will look for continuous solutions to the SDEs (2.1.1). Thus, by continuity, we have for all $t \geq 0$, $0 \leq \lambda_t^1 \leq \dots \leq \lambda_t^n$ a.s. While, for $n = 1$, the system reduces to the Cox-Ingersoll-Ross SDE $d\lambda_t^1 = 2\sqrt{\lambda_t^1}dB_t^1 + \alpha dt - 2\gamma\lambda_t^1 dt$, the coordinates are repulsed by a Coulombian interaction when $n \geq 2$.

Our goal is to study the existence and uniqueness of strong solutions to this system of SDEs in a general setting for the parameters α , β and γ , especially in the case $\beta < 1$ and $\alpha - (n-1)\beta \in (0, 1)$ which is not covered to our knowledge by the literature. The reader will find a summary of the conditions on the parameters for which the system (2.1.1) admits a unique strong solution in Table 1 below.

The difficulty in proving the existence of solutions to (2.1.1) comes from the fact that there are singularities both when a particle touches zero, where the derivative of the square root diffusion coefficient explodes, and when two particles touch each other. If we define $D = \{0 < \lambda^1 < \lambda^2 < \dots < \lambda^n\}$, a collision occurs when the process $\Lambda = (\lambda_t^1, \dots, \lambda_t^n)_t$ hits the boundary ∂D made of the union of $\{\lambda^i = \lambda^{i+1}\}$ for $i \in \{1, \dots, n-1\}$ and $\{\lambda^1 = 0\}$. A multiple collision occurs when two of these sets are reached at the same time and we will speak about "collision between particles" when two particles touch each other. Because of the singularity at the origin, it is not enough to show that there is no collision between the particles to prove the existence of a solution, as it is for instance the case for the Dyson Brownian motions which satisfy up to a change of time

$$dx_t^i = \sqrt{2}dB_t^i + \beta \sum_{j \neq i} \frac{dt}{x_t^i - x_t^j}, \text{ for all } i \in \{1, \dots, n\}. \quad (2.1.4)$$

When $\beta \geq 1$, there is no collision between these particles in finite time and (2.1.4) admits a unique strong solution (see [RS93]).

Our results about the SDEs (2.1.1) are the following. In Proposition 2.2.4, we state that $k(\alpha - (n-k)\beta) < 2$ is a necessary and sufficient condition for multiple collisions between k particles to occur at position zero in finite time. Our main result Theorem 2.2.2 gives the existence and uniqueness of solutions to the SDEs (2.1.1) when $\beta \in (0, 1)$ and $\alpha - (n-2)\beta \geq 1$ (so that there is no multiple collision at the origin). In Proposition 2.2.6 we state that when $\beta < 1$, then every pair of neighbour particles collides in finite time, and we explicit in Proposition 2.2.7 the unique stationary probability measure of the SDEs (2.1.1).

The paper is organized as follows. The remaining of the introduction is devoted to the bibliographical background of this work. In Section 2, we state our main results. We prove in Section 3 some useful properties of the solutions to the system (2.1.1), before checking Proposition 2.2.4 and Theorem 2.2.2 in Section 4. We prove the rest of the results in Section 5. Section 6 is an Appendix stating some well-known results that we use in our proofs.

The system (2.1.1) originally comes from the following matrix valued SDE : let $m \geq n$ and $(M_t)_t$ be a stochastic process taking its values in the space of $m \times n$ matrices with real

entries verifying the following stochastic differential equation

$$dM_t = dW_t - \gamma M_t dt, \quad M_0 = m_0,$$

where W is a $m \times n$ matrix filled with independent Brownian motions, m_0 is a $m \times n$ deterministic matrix. The entries of the matrix M are independent Ornstein-Uhlenbeck processes just as the one considered in [Bru91]. The process $M^\dagger M$ (where \dagger is the transpose operator) is called a *Wishart process*, and it was shown in [Bru89] and [Bru91] that the eigenvalues of $M^\dagger M$ satisfy the system of SDEs

$$d\lambda_t^i = 2\sqrt{\lambda_t^i} dB_t^i + m dt - 2\gamma \lambda_t^i dt + \sum_{j \neq i} \frac{\lambda_t^i + \lambda_t^j}{\lambda_t^i - \lambda_t^j} dt \text{ for all } i \in \{1, \dots, n\},$$

where B^1, \dots, B^n are independent Brownian motions. The reader will find in [KO01] an analysis of the complex analog of Bru's model. The system (2.1.1) generalizes the matrix inherited system by the introduction of an intensity of the interaction factor $\beta > 0$ before the sum and by the replacement of the integer parameter $m \in \mathbb{N}^* = \{1, 2, \dots\}$ by $\alpha > 0$. This replacement is analogous to the generalisation of the dynamics of the square of the norm of a m -dimensional real Brownian motion to the square α -dimensional Bessel dynamics (see for instance [RY99, chapter XI]).

Like in the correspondence between square Bessel processes and Bessel processes, let us consider the square root change of variables $x_t^i = \sqrt{\lambda_t^i}$ and set $X = (x_t^1, \dots, x_t^n)_t$. We apply Itô's formula, formally after the stopping time $\inf \{s \geq 0 : \lambda_s^1 = 0\}$ since the square root is not twice continuously differentiable at 0, and obtain for all $i \in \{1, \dots, n\}$:

$$dx_t^i = dB_t^i + \left(\frac{\alpha - 1}{2} \frac{1}{x_t^i} - \gamma x_t^i + \frac{\beta}{2x_t^i} \sum_{j \neq i} \frac{(x_t^i)^2 + (x_t^j)^2}{(x_t^i)^2 - (x_t^j)^2} \right) dt \quad (2.1.5)$$

$$= dB_t^i + \left(\frac{\alpha - (n-1)\beta - 1}{2} \frac{1}{x_t^i} - \gamma x_t^i + \beta x_t^i \sum_{j \neq i} \frac{1}{(x_t^i)^2 - (x_t^j)^2} \right) dt \quad (2.1.6)$$

$$0 \leq x_t^1 < \dots < x_t^n, \text{ a.s., } dt - a.e.$$

When X is a solution to (2.1.5), then $((x_t^1)^2, \dots, (x_t^n)^2)_t$ is a solution to (2.1.1), but it is not necessarily true the other way around.

The system (2.1.6) can be rewritten as a gradient diffusion

$$dx_t^i = dB_t^i - \partial_i V(x_t^1, \dots, x_t^n) dt \text{ for all } i \in \{1, \dots, n\}, \quad (2.1.7)$$

with potential

$$V(x_1, \dots, x_n) = - \sum_{i=1}^n \left\{ \frac{\alpha - (n-1)\beta - 1}{2} \ln |x^i| - \frac{1}{2} \gamma (x^i)^2 + \frac{\beta}{4} \sum_{j \neq i} (\ln |x^i - x^j| + \ln |x^i + x^j|) \right\}. \quad (2.1.8)$$

Systems of interacting particles following equations of the type

$$dx_t^i = b_i(x_t^i)dt + \sigma_i(x_t^i)dB_t^i - \partial_i V(x_t^1, \dots, x_t^n)dt, \quad (2.1.9)$$

where $V : \mathbb{R}^n \mapsto (-\infty, +\infty]$ is a lower semi continuous convex function such that $D = \{x \in \mathbb{R}^n : V(x) < +\infty\}$ is a non-empty convex set and V is continuously differentiable in D have been studied by many authors. For instance, in the peculiar case $\sigma_i(x^i) = \sigma > 0$, $b_i(x^i) = 0$ and $V(x^1, \dots, x^n) = -\frac{\beta}{2} \sum_{i=1}^n \sum_{j \neq i} \ln |x^i - x^j| + \frac{\theta}{2} \sum_{i=1}^n (x^i)^2$ with $\beta \geq \frac{\sigma^2}{2}$ and $\theta > 0$, the existence and uniqueness of a strong solution to (2.1.9) were derived in [RS93].

Link with the multivalued stochastic differential equations theory. The systems of type (2.1.9) with b_i and σ_i Lipschitz and $(x_0^1, \dots, x_0^n) \in \bar{D}$ were deeply studied by Cépa and Lépingle for instance in [CL97] and [CL01] where they apply Cépa's multivalued stochastic differential equations theory developed in [Cép95]. This theory treats the existence and uniqueness of solutions to multivalued SDEs associated with a convex function defined on a domain of \mathbb{R}^n . In [Lé10], Lépingle applied this theory to a constrained Brownian motion between reflecting or repellent walls of Weyl chambers. The boundary behavior of the convex function dictates the behavior of the process on these same boundaries (hitting or not the boundary in finite time, reflection on the boundary, etc.). Our SDEs of interest rewritten in the form (2.1.7) thanks to the square root change of variables can be seen this way, and we will exploit this connection in the paper.

Link with radial Dunkl processes (see [Dem09b] for a more complete description of the theory)

Let us define a reduced root system R by a finite set in $\mathbb{R}^n \setminus \{0\}$ spanning \mathbb{R}^n such that

- for all $\alpha \in R$, $R \cap \mathbb{R}\alpha = \{\alpha, -\alpha\}$,
- for all $\alpha \in R$, $\sigma_\alpha(R) = R$,

where σ_α is the reflection with respect to the hyperplane orthogonal to α . A simple system Δ is a basis of \mathbb{R}^n which induces a total ordering in R the following way : a root $\alpha \in R$ is positive if it is a positive linear combination of elements of Δ . A simple system Δ being fixed, we can thus define R_+ as the set of positive roots of R . When $\sigma_i = 1$ and V takes the form

$$V : x \mapsto - \sum_{\alpha \in R_+} k(\alpha) \ln(\langle \alpha, x \rangle), \quad x \in D,$$

where D is the positive Weyl chamber defined by

$$D = \{x \in \mathbb{R}^n, \langle \alpha, x \rangle > 0 \forall \alpha \in R_+\},$$

Demni proved in [Dem09b, Theorem 1] the existence and uniqueness of a solution to (2.1.9) on the domain \bar{D} when $k(\alpha) > 0$ for all $\alpha \in R_+$. To do so, he applied Cépa's multivalued

stochastic differential equations theory. This system corresponds to (2.1.1) for some choice of R_+ and k . Indeed, when the root system is of so-called B_n -type, it is defined by

$$\begin{aligned} R &= \{\pm e_i, \pm e_i \pm e_j, 1 \leq i < j \leq n\}, \\ \Delta &= \{e_{i+1} - e_i, 1 \leq i \leq n-1, e_1\}, \\ R_+ &= \{e_i, 1 \leq i \leq n, e_j \pm e_i, 1 \leq i < j \leq n\}, \\ D &= \{x \in \mathbb{R}^n, 0 < x^1 < \dots < x^n\}, \end{aligned}$$

which, with the right choice of k gives equation (2.1.5) with $\gamma = 0$ (see (2.1.8)). The condition $k(\alpha) > 0$ for all $\alpha \in R_+$ implies $\alpha - (n-1)\beta > 1$. We seek here to obtain the existence of a solution to (2.1.1) while relaxing the last inequality.

Link with other works. The reader will find in Graczyk and Malecki [GM13] and [GM14] a treatment of equations of the form

$$\begin{aligned} d\lambda_t^i &= \sigma_i(\lambda_t^i) dB_t^i + \left(b_i(\lambda_t^i) + \sum_{j \neq i} \frac{H_{i,j}(\lambda_t^i, \lambda_t^j)}{\lambda_t^i - \lambda_t^j} \right) dt, \text{ for all } i \in \{1, \dots, n\} \\ \lambda_t^1 &\leq \dots \leq \lambda_t^n, \quad t \geq 0, \end{aligned}$$

where the functions σ_i, b_i and $H_{i,j}$ are assumed continuous, with $H_{i,j}$ non-negative and symmetric in the sense that $H_{i,j}(x, y) = H_{j,i}(y, x)$ for all $x, y \in \mathbb{R}$. The system (2.1.1) is recovered in the particular case when $H_{i,j}(\lambda^i, \lambda^j) = \beta(\lambda^i + \lambda^j)$, $\sigma_i(\lambda^i) = 2\sqrt{\lambda^i}$ and $b_i(\lambda^i) = \alpha - 2\gamma\lambda^i$. According to [GM14, Section 6.4], (2.1.1) admits a strong solution on the time interval $[0, +\infty)$ when $\beta \geq 1$. In this regime, the authors proved that there is no collision between the particles. They also demonstrated the pathwise uniqueness of the solutions for every $\beta > 0$, as recalled in Lemma 2.3.1 below.

The same authors studied in [GM19] the system of SDEs defined for all $i \in \{1, \dots, n\}$ by

$$\begin{aligned} d\lambda_t^i &= 2\sqrt{|\lambda_t^i|} dB_t^i + \bar{\beta} \left(\bar{\alpha} + \sum_{j \neq i} \frac{|\lambda_t^i| + |\lambda_t^j|}{\lambda_t^i - \lambda_t^j} \mathbf{1}_{\{\lambda_t^i \neq \lambda_t^j\}} \right) dt \\ \lambda_t^1 &\leq \dots \leq \lambda_t^n, \quad \forall t \geq 0. \end{aligned} \tag{2.1.10}$$

They prove existence of a unique strong solution for all $\bar{\alpha} \in \mathbb{R}$ and $\bar{\beta} \geq 1$. The main difference here with our problem (2.1.1) is the relaxation of the hard edge at zero by the replacement of $\sqrt{\lambda_t^i}$ by $\sqrt{|\lambda_t^i|}$ in the diffusion coefficient. This enables the authors to establish the existence and uniqueness of a strong solution for all $\bar{\alpha} \in \mathbb{R}$ whereas in our framework, as we will see in Remark 2.2.1, we necessarily need $\alpha \geq (n-1)\beta$ for a solution to exist on the whole interval \mathbb{R}_+ . The introduction of the absolute value in the diffusion coefficient was presented earlier in [GJY03] in the case of squared Bessel processes. Moreover, our work differs from [GM19] by the fact that our main objective is to tackle the $0 < \beta < 1$ regime.

In all these references, β is identified as a fundamental parameter, its position relative to 1 governing the possibility of collisions between the particles.

Acknowledgement : We thank Djalil Chafai for numerous fruitful discussions. We also thank the referees for their remarks which helped us to improve the first version of the manuscript.

2.2 Main results

The form (2.1.2) of the SDEs hints that $\alpha - (n-1)\beta$ is a fundamental parameter impacting the existence of solutions. Consequently, we will study the range of values of this coefficient for which the system has a solution. For instance, if we assume $\alpha - (n-1)\beta < 0$, we have:

$$\begin{aligned} d\lambda_t^1 &= 2\sqrt{\lambda_t^1}dB_t^1 + \left(\alpha - (n-1)\beta - 2\gamma\lambda_t^1 + 2\beta\lambda_t^1 \sum_{j \neq 1} \frac{1}{\lambda_t^1 - \lambda_t^j} \right) dt \\ &\leq 2\sqrt{\lambda_t^1}dB_t^1 - 2\gamma\lambda_t^1 dt. \end{aligned} \quad (2.2.1)$$

Then, according to the pathwise comparison theorem of Ikeda and Watanabe (that we recall in Theorem 2.6.2 below) :

$$\lambda_t^1 \leq r_t \text{ a.s. for all } t \geq 0,$$

where

$$r_t = \lambda_0^1 + 2 \int_0^t \sqrt{r_s} dB_s^1 - 2 \int_0^t \gamma r_s ds \text{ for all } t \geq 0,$$

which is a CIR process (see for instance [LL08, Theorem 6.2.2]). By standard results on CIR processes recalled in Lemma 2.6.1 we can conclude that the stopping time $T = \inf\{t \geq 0 : r_t = 0\}$ verifies $\mathbb{P}(T < \infty) = 1$ for $\gamma \geq 0$, and $0 < \mathbb{P}(T < \infty) < 1$ for $\gamma < 0$. On $\{T < \infty\}$, after T , r stays at zero indefinitely. As the drift in (2.2.1) is strictly negative when $\lambda_t^1 = 0$ and therefore when $r_t = 0$, it will stay strictly negative on a time interval of positive measure. Consequently, the system of SDEs has no global solution.

We thus proved the following result:

Remark 2.2.1. A necessary condition for the existence of a global solution to (2.1.1) is $\alpha - (n-1)\beta \geq 0$.

We will thus assume this condition in the remaining of the paper. This condition is of course not necessary in the framework of (2.1.10) (see [GM19]).

It is proved in [GM13, Corollary 8] with condition $0 \leq \lambda_0^1 < \dots < \lambda_0^n$ that for $\beta \geq 1$, the SDEs (2.1.1) has a unique global strong solution and that there actually is no collision between particles.

Demni proved in [Dem09b, subsection 5.1] by applying Cépa's multivalued equations theory [Cép95] that under the conditions $\alpha - (n-1)\beta > 1$, $\beta > 0$, $\gamma = 0$ and $0 < \lambda_0^1 < \dots < \lambda_0^n$, the system of SDEs (2.1.5) admits a unique strong solution and that for $0 < \beta < 1$ and $\alpha - (n-1)\beta > 1$, there is collision between any neighbour particles λ^i and λ^{i+1} for $i \in \{1, \dots, n-1\}$ in finite time almost surely. The condition $\alpha - (n-1)\beta > 1$ ensures the

convexity of the potential V defined in (2.1.8) which is needed in Cépa's multivalued equations theory. In Theorem 2.2.2 and Proposition 2.2.6, we tackle respectively the existence problem and the collision problem on wider ranges for the parameters $\alpha - (n - 1)\beta$ and γ .

Our following result applies in the case $\alpha - (n - 1)\beta \leq 1$:

Theorem 2.2.2. *Let us assume $\beta < 1, \alpha - (n - 1)\beta > 0$. Let the initial condition $\Lambda_0 = (\lambda_0^1, \dots, \lambda_0^n)$ be independent from the Brownian motion \mathbf{B} and such that $0 \leq \lambda_0^1 \leq \dots \leq \lambda_0^n$ a.s. and $\lambda_0^2 > 0$ a.s.*

Then, the system of SDEs (2.1.1) has a unique strong solution defined on the time interval

$$[0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon),$$

where, for $\varepsilon > 0$,

$$\zeta_\varepsilon = \inf\{t \geq 0 : \lambda_t^1 \leq \varepsilon \text{ and } \lambda_t^2 - \lambda_t^1 \leq \varepsilon\}.$$

Moreover,

- (i) for $\gamma \in \mathbb{R}$, if $\alpha - (n - 1)\beta \geq 1 - \beta$ then $\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon = \infty$ a.s.
- (ii) for $\gamma \geq 0$, if $\alpha - (n - 1)\beta < 1 - \beta$ then $\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon < \infty$ a.s.
- (iii) There is no double collision between particles, i.e.

$$\mathbb{P}\left\{\exists t \in (0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon) : \lambda_t^i = \lambda_t^{i+1} \text{ and } \lambda_t^j = \lambda_t^{j+1} \text{ for some } 1 \leq i < j \leq n - 1\right\} = 0.$$

Theorem 2.2.2 is proved in Section 4.

Remark 2.2.3. This result states the existence of a unique strong solution $(\lambda_t^1, \dots, \lambda_t^n)_t$ to (2.1.1) defined on the time-interval \mathbb{R}_+ for $\alpha - (n - 1)\beta \geq 1 - \beta$ and on $[0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon)$ if $0 < \alpha - (n - 1)\beta < 1 - \beta$. When $\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon < +\infty$, then, according to the first step in the proof of assertion (ii) of Proposition 2.2.4 below, the solution can be continuously extended to the closed time-interval $[0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon]$. From the definition of ζ_ε , then $\lambda_{\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon}^1 = \lambda_{\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon}^2 = 0$. The next step would be to find how to start back from $(\lambda_{\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon}^1, \dots, \lambda_{\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon}^n)$ to define a solution on the whole interval \mathbb{R}_+ .

The disjunction (i) – (ii) comes from the application with $k = 2$ of the following proposition, which gives a condition for k particles to collide at the position zero.

Proposition 2.2.4 (Multiple collision at zero). *Let $k \in \{1, \dots, n\}$. Let the initial condition $\Lambda_0 = (\lambda_0^1, \dots, \lambda_0^n)$ be independent from the Brownian motion \mathbf{B} and such that $0 \leq \lambda_0^1 \leq \dots \leq \lambda_0^n$ a.s. Then,*

(i) if $\gamma \geq 0$, if (2.1.1) has a global solution $\Lambda = (\lambda_t^1, \dots, \lambda_t^n)_t$ and if $k(\alpha - (n - k)\beta) < 2$, then

$$\mathbb{P}(\exists t \geq 0 : \lambda_t^1 + \lambda_t^2 + \dots + \lambda_t^k = 0) = 1,$$

(ii) if $\gamma \in \mathbb{R}$, $\lambda_0^k > 0$ a.s., if (2.1.1) has a local solution $\Lambda = (\lambda_t^1, \dots, \lambda_t^n)_t$ defined up to a stopping time \mathcal{T} and if $k(\alpha - (n - k)\beta) \geq 2$, then there is no collision of k particles at zero almost surely. More precisely,

$$\mathbb{P}(\mathcal{T} = +\infty, \exists t \geq 0 : \lambda_t^1 + \lambda_t^2 + \dots + \lambda_t^k = 0) + \mathbb{P}(\mathcal{T} < +\infty, \inf_{t \in [0, \mathcal{T})} \lambda_t^1 + \lambda_t^2 + \dots + \lambda_t^k = 0) = 0.$$

Remark 2.2.5. Under the assumptions made in the first assertion of Proposition 2.2.4 but with $\gamma < 0$, we can in fact prove that

$$\mathbb{P}(\exists t \geq 0 : \lambda_t^1 + \lambda_t^2 + \dots + \lambda_t^k = 0) \in (0, 1),$$

using Lemma 2.6.1 point 3. As a consequence, under the assumptions made in (ii) of Theorem 2.2.2 but with $\gamma < 0$,

$$\mathbb{P}(\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon < \infty) \in (0, 1).$$

This proposition is proved in the beginning of Section 4.

Demni proved in [Dem09b] that for $\beta < 1$ and $\alpha - (n - 1)\beta > 1$, a collision between the particles λ^i and λ^{i+1} occurs in finite time almost surely for all $i \in \{1, \dots, n - 1\}$. We strengthen here this result by showing that for $\beta < 1$ and $\alpha - (n - 1)\beta > 0$, every particle touches its neighbour particles in finite time almost surely.

Proposition 2.2.6. Let $\beta \in (0, 1)$, $\alpha > 0$ and the initial condition $\Lambda_0 = (\lambda_0^1, \dots, \lambda_0^n)$ be independent from the Brownian motion \mathbf{B} and such that $0 \leq \lambda_0^1 \leq \dots \leq \lambda_0^n$. Let us assume that either $\gamma > 0$ and there is a global solution to (2.1.1) or $\gamma \geq 0$, $\alpha - (n - 1)\beta \geq 1 - \beta$ and $\lambda_0^2 > 0$ a.s. (so that, by Theorem 2.2.2, there also exists a global solution to (2.1.1)). Then for all $i \in \{2, \dots, n\}$, the stopping time

$$T^{(i)} = \inf\{t > 0 : \lambda_t^i = \lambda_t^{i-1}\}$$

is such that

$$\mathbb{P}\{T^{(i)} < \infty\} = 1.$$

To check this result, we prove that

$$\int_0^{+\infty} \lambda_s^1 ds = +\infty \text{ a.s.} \quad (2.2.2)$$

using the next result.

Proposition 2.2.7. Let us assume $\gamma > 0$ and $\alpha - (n - 1)\beta > 0$. The unique stationary probability measure of the system of SDEs (2.1.1) is ρ_{inv} with density with respect to the Lebesgue measure

$$d\rho_{inv}(\lambda^1, \dots, \lambda^n) = \frac{1}{\mathcal{Z}} \times \prod_{i=1}^n \left((\lambda^i)^{\frac{\alpha - (n-1)\beta}{2} - 1} e^{-\gamma \lambda^i} \prod_{j \neq i} |\lambda^j - \lambda^i|^{\beta/2} \right) \mathbf{1}_{\{0 \leq \lambda^1 \leq \dots \leq \lambda^n\}} d\lambda^1 \dots d\lambda^n,$$

where \mathcal{Z} is a normalizing constant. More precisely, if a solution to (2.1.1) is such that the distribution of Λ_t does not depend on t , then this distribution is ρ_{inv} . Conversely, (2.1.1) admits a unique solution Λ starting from Λ_0 distributed according to ρ_{inv} and independent from the Brownian motion \mathbf{B} . Moreover, for all $t \in \mathbb{R}_+$, Λ_t is distributed according to ρ_{inv} .

Remark 2.2.8. To prove (2.2.2), one could consider applying the ergodic theorem, but we were not able to prove that the process defined by the system of SDEs (2.1.1) is a Markov process. The difficulty came from the choice of the state space: when $\lambda_0^2 = 0$, we do not know how to prove the existence of a solution to (2.1.1). If the state space is $\{0 \leq \lambda^1 \leq \dots \leq \lambda^n, \lambda^2 > 0\}$ we do not know how to prove that for all $t > 0$, $\mathbb{P}(\lambda_t^2 > 0) = 1$ when $\alpha - (n-1)\beta < 1 - \beta$.

Remark 2.2.9. The density of ρ_{inv} generalizes the one of the beta-Laguerre ensemble which writes

$$\frac{1}{Z} \prod_{i=1}^n \left((\lambda^i)^{\frac{m\beta - (n-1)\beta}{2} - 1} e^{-\frac{\beta}{2}\lambda^i} \prod_{j \neq i} |\lambda^j - \lambda^i|^{\beta/2} \right) \mathbf{1}_{\{0 \leq \lambda^1 \leq \dots \leq \lambda^n\}}, \quad m \in \{n, n+1, \dots\}, \quad \beta > 0,$$

by replacement of $m\beta$ by α in the exponent of λ^i and of $\frac{\beta}{2}$ by γ in the exponential factor. The reader will find in [For10] a documentation on beta-Laguerre ensembles. This ensemble is related to a random matrix model. The parameters m and n are then respectively the number of lines and columns of the underneath random matrix model, and $\beta = 1, 2$ or 4 depending on the dimension of the underlying algebra (\mathbb{R}, \mathbb{C} or \mathbb{H}).

Table 2.1 shows, for $\gamma \geq 0$, the conditions on the coefficients of the SDEs (2.1.1) for the existence of strong solutions.

2.3 Properties of the solutions

One can first remark, and it will be useful in several proofs, that the sum of the n coordinates of a solution $(\lambda_t^1, \dots, \lambda_t^n)_t$ to SDEs (2.1.1) follows a CIR dynamics. Indeed,

$$\begin{aligned} d\left(\sum_{i=1}^n \lambda_t^i\right) &= 2 \sum_{i=1}^n \sqrt{\lambda_t^i} dB_t^i - 2\gamma \sum_{i=1}^n \lambda_t^i dt + n\alpha dt \\ &= 2 \sqrt{\sum_{i=1}^n \lambda_t^i} dW_t - 2\gamma \sum_{i=1}^n \lambda_t^i dt + n\alpha dt, \end{aligned} \quad (2.3.1)$$

where W defined by $W_0 = 0$ and $dW_t = \sum_{i=1}^n \left(\mathbf{1}_{\{\sum_{j=1}^n \lambda_t^j \neq 0\}} \frac{\sqrt{\lambda_t^i}}{\sqrt{\sum_{j=1}^n \lambda_t^j}} + \mathbf{1}_{\{\sum_{j=1}^n \lambda_t^j = 0\}} \frac{1}{\sqrt{n}} \right) dB_t^i$ is a Brownian motion according to Lévy's characterization.

The pathwise uniqueness part of the next Lemma is proved in [GM14, Theorem 5.3], but we reproduce the proof for the sake of completeness.

Lemma 2.3.1. *Let $\gamma \in \mathbb{R}$. The solutions to (2.1.1) are pathwise unique.*

Moreover, if $Z = (z_t^1, \dots, z_t^n)_t$ and $\tilde{Z} = (\tilde{z}_t^1, \dots, \tilde{z}_t^n)_t$ are two global solutions to (2.1.1) with the same driving Brownian motion and verifying

$$\mathbb{E} \sum_{i=1}^n [z_0^i + \tilde{z}_0^i] < +\infty,$$

$\alpha - (n-1)\beta$	$\beta \geq 1$	$\beta < 1$
< 0	Defined on $[0, \liminf_{\varepsilon \rightarrow 0} \{t \geq 0 : \lambda_t^1 \leq \varepsilon\})$, no collision, Proposition 2.5.3	Defined on $[0, \liminf_{\varepsilon \rightarrow 0} \{t \geq 0 : \lambda_t^1 \leq \varepsilon\})$ which is finite a.s., Proposition 2.5.3
$\in (0, 1 - \beta)$	empty interval	Defined on $[0, \liminf_{\varepsilon \rightarrow 0} \{t \geq 0 : \lambda_t^1 \leq \varepsilon \text{ and } \lambda_t^2 - \lambda_t^1 < \varepsilon\})$, which is finite a.s., Theorem 2.2.2
$\geq 1 - \beta$	Defined on \mathbb{R}_+ , no collision between particles, proved in [GM13] with separate initial condition	Defined on \mathbb{R}_+ , collision in finite time but no multiple collision at zero, Theorem 2.2.2
> 1	Defined on \mathbb{R}_+ , no collision between particles, proved in [GM13] with separate initial condition or in [Dem09b] with $\gamma = 0$	Defined on \mathbb{R}_+ , collision between particles in finite time, [Dem09b], Proposition 2.2.6
≥ 2	Defined on \mathbb{R}_+ , no collision, proved in [GM13] with separate initial condition or in [Dem09b] with $\gamma = 0$	Defined on \mathbb{R}_+ , collision between particles in finite time, $\inf\{t > 0 : \lambda^1 = 0\} = +\infty$ a.s., Proposition 2.2.4

Table 2.1 – Conditions on the coefficients of SDEs (2.1.1) for the existence of strong solutions when $\gamma \geq 0$.

then for all $t \geq 0$,

$$\sum_{i=1}^n \mathbb{E}|z_t^i - \tilde{z}_t^i| \leq \left(\sum_{i=1}^n \mathbb{E}|z_0^i - \tilde{z}_0^i| \right) \exp(-2\gamma t). \quad (2.3.2)$$

Proof. Let Z and \tilde{Z} be two solutions to (2.1.1). Because of the square root diffusion coefficient, the local time of $z^i - \tilde{z}^i$ at 0 is zero ([RY99, Lemma 3.3 p.389]). Applying the Tanaka formula then the integration by parts formula to compute $de^{2\gamma s}|z_s^i - \tilde{z}_s^i|$ we get

$$\begin{aligned} de^{2\gamma s} \sum_{i=1}^n |z_s^i - \tilde{z}_s^i| &= 2 \sum_{i=1}^n e^{2\gamma s} |\sqrt{z_s^i} - \sqrt{\tilde{z}_s^i}| dB_s^i + \beta e^{2\gamma s} \sum_{i=1}^n \text{sgn}(z_s^i - \tilde{z}_s^i) \sum_{j \neq i} \left(\frac{z_s^i + z_s^j}{z_s^i - z_s^j} - \frac{\tilde{z}_s^i + \tilde{z}_s^j}{\tilde{z}_s^i - \tilde{z}_s^j} \right) ds \\ &\leq 2 \sum_{i=1}^n e^{2\gamma s} |\sqrt{z_s^i} - \sqrt{\tilde{z}_s^i}| dB_s^i, \end{aligned} \quad (2.3.3)$$

where $\text{sgn}(x) = 1$ if $x > 0$ and $\text{sgn}(x) = -1$ if $x \leq 0$. The inequality (2.3.3) comes from the fact that we have for all $i < j$:

$$\begin{aligned} &\left[\frac{z_s^i + z_s^j}{z_s^i - z_s^j} - \frac{\tilde{z}_s^i + \tilde{z}_s^j}{\tilde{z}_s^i - \tilde{z}_s^j} \right] (\text{sgn}(z_s^i - \tilde{z}_s^i) - \text{sgn}(z_s^j - \tilde{z}_s^j)) \\ &= 2 \frac{z_s^j \tilde{z}_s^i - z_s^i \tilde{z}_s^j}{(z_s^i - z_s^j)(\tilde{z}_s^i - \tilde{z}_s^j)} (\text{sgn}(z_s^i - \tilde{z}_s^i) - \text{sgn}(z_s^j - \tilde{z}_s^j)) \end{aligned} \quad (2.3.4)$$

$$\begin{aligned}
&= 2 \frac{z_s^j(\tilde{z}_s^i - z_s^i) + z_s^i(z_s^j - \tilde{z}_s^j)}{(z_s^i - z_s^j)(\tilde{z}_s^i - \tilde{z}_s^j)} (\operatorname{sgn}(z_s^i - \tilde{z}_s^i) - \operatorname{sgn}(z_s^j - \tilde{z}_s^j)) \\
&= -2 \frac{z_s^j|\tilde{z}_s^i - z_s^i| + z_s^i|z_s^j - \tilde{z}_s^j|}{(z_s^i - z_s^j)(\tilde{z}_s^i - \tilde{z}_s^j)} |\operatorname{sgn}(z_s^i - \tilde{z}_s^i) - \operatorname{sgn}(z_s^j - \tilde{z}_s^j)| \leq 0,
\end{aligned} \tag{2.3.5}$$

as the denominator is non-negative. Let the two solutions Z and \tilde{Z} be respectively defined on $[0, T]$ and $[0, \tilde{T}]$, where T and \tilde{T} are stopping times for a filtration $(\mathcal{F}_t)_{t \geq 0}$ with respect to which \mathbf{B} is a Brownian motion and Z_0 and \tilde{Z}_0 are \mathcal{F}_0 -measurable.

Let $M > 0$ and

$$\tau_M = \inf\{t \in [0, T \wedge \tilde{T}] : \tilde{z}_t^n + z_t^n \geq M\},$$

with the convention $\inf \emptyset = T \wedge \tilde{T}$. As Z and \tilde{Z} are continuous and assumed well defined on respectively $[0, T]$ and $[0, \tilde{T}]$, $\tau_M \uparrow T \wedge \tilde{T}$ when $M \uparrow \infty$.

As for all $i \in \{1, \dots, n\}$:

$$\mathbb{E} \int_0^{t \wedge \tau_M} e^{4\gamma s} |\sqrt{z_s^i} - \sqrt{\tilde{z}_s^i}|^2 ds < \infty,$$

the stochastic integrals $\int_0^{t \wedge \tau_M} e^{2\gamma s} |\sqrt{z_s^i} - \sqrt{\tilde{z}_s^i}| dB_s^i$ have zero expectation so that integrating (2.3.3) on $[0, t \wedge \tau_M]$ and taking expectations, we obtain that

$$\mathbb{E} e^{2\gamma(t \wedge \tau_M)} \sum_{i=1}^n |z_{t \wedge \tau_M}^i - \tilde{z}_{t \wedge \tau_M}^i| \leq \mathbb{E} \sum_{i=1}^n |z_0^i - \tilde{z}_0^i|.$$

When $Z_0 = \tilde{Z}_0$ we deduce that for all $t \geq 0$ and $M > 0$

$$\sum_{i=1}^n \mathbb{E} |z_{t \wedge \tau_M}^i - \tilde{z}_{t \wedge \tau_M}^i| = 0,$$

and conclude using Fatou's Lemma to take the limit $M \rightarrow \infty$, that for all $t \geq 0$

$$\sum_{i=1}^n \mathbb{E} |z_{t \wedge T \wedge \tilde{T}}^i - \tilde{z}_{t \wedge T \wedge \tilde{T}}^i| = 0,$$

which concludes the proof of pathwise uniqueness.

Let us now prove (2.3.2) by assuming that Z and \tilde{Z} are defined globally, with integrable initial conditions, i.e.

$$\mathbb{E} \sum_{i=1}^n [z_0^i + \tilde{z}_0^i] < +\infty.$$

As the equation (2.3.1) shows that the sum of the coordinates of Z and \tilde{Z} are both CIR processes, we deduce from Lemma 2.6.1 point 4. that

$$\mathbb{E} \left[\int_0^t e^{4\gamma s} \left(\sum_{i=1}^n z_s^i + \sum_{i=1}^n \tilde{z}_s^i \right) ds \right] < \infty,$$

so that $\mathbb{E} \sum_{i=1}^n \int_0^t e^{2\gamma s} |\sqrt{z_s^i} - \sqrt{\tilde{z}_s^i}| dB_s^i = 0$. Integrating (2.3.3) between 0 and t and taking expectations, we conclude that

$$\sum_{i=1}^n \mathbb{E} |z_t^i - \tilde{z}_t^i| \leq \left(\sum_{i=1}^n \mathbb{E} |z_0^i - \tilde{z}_0^i| \right) \exp(-2\gamma t).$$

□

Proof of Proposition 2.2.7. We start by proving the uniqueness of the invariant distribution. Let us first show that an invariant distribution ρ_{inv} has a finite first order moment. To do so, one can remark (see (2.3.1)) that the image of ρ_{inv} by the sum of the n coordinates is invariant for the CIR process

$$dr_t = 2\sqrt{r_t} dW_t + (n\alpha - 2\gamma r_t) dt.$$

It is known (see for instance [CIR85]) that the invariant distribution of such a process is a gamma law of positive parameters, whose density is

$$r \mapsto \frac{\gamma^{\frac{n\alpha}{2}}}{\Gamma(\frac{n\alpha}{2})} r^{\frac{n\alpha}{2}-1} e^{-\gamma r},$$

which has a finite first order moment. We can thus first apply the second part of Lemma 2.3.1 for two solutions to (2.1.1) starting respectively according to two invariant distributions to deduce that these two invariant distributions are equal.

The candidate density, obtained by the square of each coordinate change of variables from the density proportional to $e^{-2V(x^1, \dots, x^n)}$ with potential V defined in (2.1.8) candidate to be stationary for the gradient diffusion (2.1.5), writes

$$f_{inv}(\lambda^1, \dots, \lambda^n) = \frac{e^{-2V(\sqrt{\lambda^1}, \dots, \sqrt{\lambda^n})}}{\mathcal{Z} \sqrt{\lambda^1} \dots \lambda^n} = \frac{1}{\mathcal{Z}} \times \prod_{i=1}^n \left((\lambda^i)^{\frac{\alpha-2-(n-1)\beta}{2}} e^{-\gamma \lambda^i} \prod_{j \neq i} |\lambda^j - \lambda^i|^{\beta/2} \right) \mathbf{1}_{\{0 \leq \lambda^1 \leq \dots \leq \lambda^n\}}, \quad (2.3.6)$$

with \mathcal{Z} a normalizing constant. The second factor is indeed integrable since for $0 \leq \lambda^1 \leq \dots \leq \lambda^n$,

$$\prod_{i=1}^n \left((\lambda^i)^{\frac{\alpha-2-(n-1)\beta}{2}} e^{-\gamma \lambda^i} \prod_{j \neq i} |\lambda^j - \lambda^i|^{\beta/2} \right) \leq (\lambda^n)^{\frac{(n-1)^2 \beta + \alpha - 2}{2}} e^{-\gamma \lambda^n} \prod_{i=1}^{n-1} \left((\lambda^i)^{\frac{\alpha-2-(n-1)\beta}{2}} e^{-\gamma \lambda^i} \right),$$

and $\alpha - (n-1)\beta > 0 \implies \frac{\alpha-2-(n-1)\beta}{2} > -1$. Let us check that the probability measure ρ_{inv} with density f_{inv} with respect to the Lebesgue measure solves the Fokker-Planck equation in the sense of distributions

$$\mathcal{A}^* \rho_{inv} = 0,$$

where \mathcal{A} is the infinitesimal generator associated with the dynamics (2.1.1) :

$$\mathcal{A} = \sum_{i=1}^n \left(\alpha - 2\gamma \lambda^i + \beta \sum_{j \neq i} \frac{\lambda^i + \lambda^j}{\lambda^i - \lambda^j} \right) \frac{\partial}{\partial \lambda^i} + 2 \sum_{i=1}^n \lambda^i \frac{\partial^2}{\partial (\lambda^i)^2}.$$

For a test function $\phi : \mathbb{R}^n \mapsto \mathbb{R}$, compactly supported and twice continuously differentiable, since f_{inv} vanishes for $\lambda^1 = 0$, $\lambda^i = \lambda^{i+1}$ when $i \in \{1, \dots, n-1\}$, and for $\lambda^n \rightarrow +\infty$, we obtain by integration by parts that for $i \in \{1, \dots, n\}$

$$\begin{aligned} & \int_{0 \leq \lambda^1 \leq \dots \leq \lambda^n} \lambda^i \frac{\partial^2 \phi}{\partial (\lambda^i)^2}(\lambda^1, \dots, \lambda^n) f_{inv}(\lambda^1, \dots, \lambda^n) d\lambda^1 \dots d\lambda^n \\ &= - \int_{0 \leq \lambda^1 \leq \dots \leq \lambda^n} \frac{\partial \phi}{\partial \lambda^i}(\lambda^1, \dots, \lambda^n) \left(f_{inv}(\lambda^1, \dots, \lambda^n) + \lambda^i \frac{\partial f_{inv}}{\partial \lambda^i}(\lambda^1, \dots, \lambda^n) \right) d\lambda^1 \dots d\lambda^n. \end{aligned}$$

Since

$$\begin{aligned} & \left(\alpha - 2\gamma\lambda^i + \beta \sum_{j \neq i} \frac{\lambda^i + \lambda^j}{\lambda^i - \lambda^j} - 2 \right) f_{inv}(\lambda^1, \dots, \lambda^n) - 2\lambda^i \frac{\partial f_{inv}}{\partial \lambda^i}(\lambda^1, \dots, \lambda^n) \\ &= f_{inv}(\lambda^1, \dots, \lambda^n) \left[\alpha - 2\gamma\lambda^i + \beta \sum_{j \neq i} \frac{\lambda^i + \lambda^j}{\lambda^i - \lambda^j} - 2 + 2\lambda^i \left(\frac{1}{2\lambda^i} + \frac{2}{2\sqrt{\lambda^i}} \frac{\partial V}{\partial x^i}(\sqrt{\lambda^1}, \dots, \sqrt{\lambda^n}) \right) \right] \\ &= f_{inv}(\lambda^1, \dots, \lambda^n) \left[\alpha - 2\gamma\lambda^i + \beta \sum_{j \neq i} \frac{\lambda^i + \lambda^j}{\lambda^i - \lambda^j} - 2 + 1 \right. \\ & \quad \left. + 2\sqrt{\lambda^i} \left(-\frac{\alpha-1}{2} \frac{1}{\sqrt{\lambda^i}} + \gamma\sqrt{\lambda^i} - \frac{\beta}{2\sqrt{\lambda^i}} \sum_{j \neq i} \frac{\lambda^i + \lambda^j}{\lambda^i - \lambda^j} \right) \right], \end{aligned}$$

where the last factor vanishes, we conclude that $\int_{0 \leq \lambda^1 \leq \dots \leq \lambda^n} \mathcal{A}\phi(\lambda^1, \dots, \lambda^n) f_{inv}(\lambda^1, \dots, \lambda^n) d\lambda^1 \dots d\lambda^n = 0$. To deduce the existence of a weak solution to (2.1.1) whose marginals follow the law ρ_{inv} , we may apply [Tre16, Theorem 2.5], as soon as

$$\int_{0 \leq \lambda^1 \leq \dots \leq \lambda^n} \sum_{i=1}^n \left\{ 2\lambda^i + \left| \alpha - 2\gamma\lambda^i + \beta \sum_{j \neq i} \frac{\lambda^i + \lambda^j}{\lambda^i - \lambda^j} \right| \right\} d\rho_{inv}(\lambda^1, \dots, \lambda^n) < +\infty.$$

This property can be proved by remarking that in the definition (2.3.6) of f_{inv} , the exponential factors ensure integrability at infinity while the factor $|\lambda^j - \lambda^i|^{\beta/2}$ makes the singularity of the denominator of the interaction term $\frac{\lambda^i + \lambda^j}{\lambda^i - \lambda^j}$ integrable. The exponential factors in f_{inv} crush every other term at $+\infty$. If $\beta \geq 1$, the boundary $\lambda^i = \lambda^j$ is not singular. Let us discuss further this boundary when $\beta < 1$ by looking at the term below for $1 \leq j < i \leq n$:

$$\begin{aligned} & \int_{0 \leq \lambda^1 \leq \dots \leq \lambda^n} \frac{\lambda^i + \lambda^j}{\lambda^i - \lambda^j} f_{inv}(\lambda^1, \dots, \lambda^n) d\lambda^1 \dots d\lambda^n \\ & \leq \frac{2}{Z} \int_{0 \leq \lambda^1 \leq \dots \leq \lambda^n} (\lambda^i - \lambda^j)^{-1} \lambda^n \prod_{k=1}^n \left((\lambda^k)^{\frac{\alpha-2-(n-1)\beta}{2}} e^{-\gamma\lambda^k} \prod_{l \neq k} |\lambda^l - \lambda^k|^{\beta/2} \right) d\lambda^1 \dots d\lambda^n \end{aligned} \tag{2.3.7}$$

$$\leq \frac{2}{Z} \int_{0 \leq \lambda^1 \leq \dots \leq \lambda^n} (\lambda^i - \lambda^j)^{-1+\beta} (\lambda^n)^{1+\frac{\alpha-2-(n-1)\beta}{2}n+\frac{\beta}{2}(n(n-1)-2)} e^{-\gamma\lambda^n} d\lambda^1 \dots d\lambda^n \tag{2.3.8}$$

$$\begin{aligned}
&\leq \frac{2}{\mathcal{Z}} \int_{0 \leq \lambda^1 \leq \dots \leq \lambda^n} (\lambda^i - \lambda^j)^{-1+\beta} (\lambda^n)^{1-\beta+\frac{n(\alpha-2)}{2}} e^{-\gamma \lambda^n} d\lambda^1 \dots d\lambda^n & (2.3.9) \\
&\leq \frac{2}{\mathcal{Z}} \int_{(\mathbb{R}_+)^{n-1}} \left(\int_{\mathbb{R}_+} \mathbb{1}_{\{0 \leq \lambda^1 \leq \lambda^2\}} d\lambda^1 \right) (\lambda^i - \lambda^j)^{-1+\beta} (\lambda^n)^{1-\beta+\frac{n(\alpha-2)}{2}} e^{-\gamma \lambda^n} \mathbb{1}_{\{0 \leq \lambda^2 \leq \dots \leq \lambda^n\}} d\lambda^2 \dots d\lambda^n \\
&\leq \frac{2}{\mathcal{Z}} \int_{(\mathbb{R}_+)^{n-1}} (\lambda^i - \lambda^j)^{-1+\beta} (\lambda^n)^{2-\beta+\frac{n(\alpha-2)}{2}} e^{-\gamma \lambda^n} \mathbb{1}_{\{0 \leq \lambda^2 \leq \dots \leq \lambda^n\}} d\lambda^2 \dots d\lambda^n \\
&\vdots \\
&\leq \frac{2}{\mathcal{Z}} \int_{(\mathbb{R}_+)^{n-j+1}} (\lambda^i - \lambda^j)^{-1+\beta} (\lambda^n)^{j-\beta+\frac{n(\alpha-2)}{2}} e^{-\gamma \lambda^n} \mathbb{1}_{\{0 \leq \lambda^j \leq \dots \leq \lambda^n\}} d\lambda^j \dots d\lambda^n \\
&\leq \frac{2}{\mathcal{Z}} \int_{(\mathbb{R}_+)^{n-j}} \left(\int_{\mathbb{R}_+} (\lambda^i - \lambda^j)^{-1+\beta} \mathbb{1}_{\{0 \leq \lambda^j \leq \lambda^{j+1}\}} d\lambda^j \right) (\lambda^n)^{j-\beta+\frac{n(\alpha-2)}{2}} e^{-\gamma \lambda^n} \mathbb{1}_{\{0 \leq \lambda^{j+1} \leq \dots \leq \lambda^n\}} \\
&\hspace{20em} d\lambda^{j+1} \dots d\lambda^n \\
&\leq \frac{1}{2^{n-1}(1-\beta)} \times \frac{1}{\mathcal{Z}} \int_{(\mathbb{R}_+)^{n-j}} (\lambda^i - \lambda^j)^\beta (\lambda^n)^{j-\beta+\frac{n(\alpha-2)}{2}} e^{-\gamma \lambda^n} \mathbb{1}_{\{0 \leq \lambda^{j+1} \leq \dots \leq \lambda^n\}} d\lambda^{j+1} \dots d\lambda^n & (2.3.10) \\
&\leq \frac{1}{1-\beta} \times \frac{2}{\mathcal{Z}} \int_{(\mathbb{R}_+)^{n-j}} (\lambda^n)^{j+\frac{n(\alpha-2)}{2}} e^{-\gamma \lambda^n} \mathbb{1}_{\{0 \leq \lambda^{j+1} \leq \dots \leq \lambda^n\}} d\lambda^{j+1} \dots d\lambda^n < +\infty \\
&\vdots \\
&\leq \frac{1}{1-\beta} \times \frac{2}{\mathcal{Z}} \int_{\mathbb{R}_+} (\lambda^n)^{-1+\frac{n\alpha}{2}} e^{-\gamma \lambda^n} d\lambda^n < +\infty,
\end{aligned}$$

as $\alpha > (n-1)\beta > 0$. To go from (2.3.7) to (2.3.9), we bound from above $(\lambda^k)^{\frac{\alpha-2-(n-1)\beta}{2}}$ by $(\lambda^n)^{\frac{\alpha-2-(n-1)\beta}{2}}$, $|\lambda^l - \lambda^k|^{\beta/2}$ for $\{l, k\} \neq \{i, j\}$ by $(\lambda^n)^{\frac{\beta}{2}}$ and $e^{-\gamma \lambda^k}$ for $k \neq n$ by 1. We can thus apply [Tre16, Theorem 2.5] to deduce that we can define a process solving (2.1.1) whose marginals follow the law ρ_{inv} so that ρ_{inv} is invariant. By pathwise uniqueness proved in Lemma 2.3.1, this weak solution to (2.1.1) is a strong solution. \square

2.4 Proof of Theorem 2.2.2

We start this section by the proof of Proposition 2.2.4 since this result is crucial in the proof of Theorem 2.2.2.

Proof of Proposition 2.2.4

(i) To prove this assertion, we study for all $k \in \{1, \dots, n\}$ the process $\lambda^1 + \dots + \lambda^k$ and show that, as its interaction terms with the particles $\lambda^{k+1}, \dots, \lambda^n$ are non-positive, it is smaller than a CIR process hitting zero in finite time.

Let us define W^k by $W_0^k = 0$ and

$$dW_t^k = \sum_{i=1}^k \left(\mathbf{1}_{\{\sum_{j=1}^k \lambda_t^j \neq 0\}} \frac{\sqrt{\lambda_t^i}}{\sqrt{\sum_{j=1}^k \lambda_t^j}} + \mathbf{1}_{\{\sum_{j=1}^k \lambda_t^j = 0\}} \frac{1}{\sqrt{k}} \right) dB_t^i.$$

According to Lévy's characterization, W^k is a Brownian motion. For $t \geq 0$,

$$\begin{aligned} d(\lambda_t^1 + \dots + \lambda_t^k) &= 2\sqrt{\lambda_t^1 + \dots + \lambda_t^k} dW_t^k - 2\gamma(\lambda_t^1 + \dots + \lambda_t^k)dt + k(\alpha - (n-k)\beta)dt \\ &\quad + 2\beta \sum_{i=1}^k \lambda_t^i \sum_{j=k+1}^n \frac{1}{\lambda_t^i - \lambda_t^j} dt \\ &\leq 2\sqrt{\lambda_t^1 + \dots + \lambda_t^k} dW_t^k - 2\gamma(\lambda_t^1 + \dots + \lambda_t^k)dt + k(\alpha - (n-k)\beta)dt. \end{aligned} \tag{2.4.1}$$

By the pathwise comparison theorem of Ikeda and Watanabe (that we recall in Theorem 2.6.2 below),

$$\lambda_t^1 + \dots + \lambda_t^k \leq r_t \text{ for all } t \geq 0 \text{ a.s.},$$

where

$$r_t = \lambda_0^1 + \dots + \lambda_0^k + 2 \int_0^t \sqrt{r_s} dW_s^k - 2\gamma \int_0^t r_s ds + k(\alpha - (n-k)\beta)t \tag{2.4.2}$$

is a CIR process. Applying Lemma 2.6.1 with $a = k(\alpha - (n-k)\beta)$, $b = 2\gamma$ and $\sigma = 2$, which satisfy $a < \frac{\sigma^2}{2}$ and $b \geq 0$, we can conclude.

(ii) Before proving the assertion by backward induction on k , let us first check that, whatever $\alpha \geq 0$, $\beta > 0$ and $\gamma \in \mathbb{R}$, a local solution to (2.1.1) defined up to a stopping time \mathcal{T} is actually continuous and solves (2.1.1) on the closed time interval $[0, \mathcal{T}]$ on $\{\mathcal{T} < \infty\}$. For all $k \in \{1, \dots, n\}$, we define the Brownian motion W^k the following way :

$$dW_t^k = \mathbf{1}_{\{t \in [0, \mathcal{T})\}} \sum_{i=1}^k \left(\mathbf{1}_{\{\sum_{j=1}^k \lambda_t^j \neq 0\}} \frac{\sqrt{\lambda_t^i}}{\sqrt{\sum_{j=1}^k \lambda_t^j}} + \mathbf{1}_{\{\sum_{j=1}^k \lambda_t^j = 0\}} \frac{1}{\sqrt{k}} \right) dB_t^i + \mathbf{1}_{\{t \geq \mathcal{T}\}} \frac{1}{\sqrt{k}} \sum_{i=1}^k dB_t^i.$$

For $k = n$ the inequality (2.4.1) is an equality for $t < \mathcal{T}$, and according to Lemma 2.6.1, the CIR process r satisfying (2.4.2) is defined globally and for $t \in [0, \mathcal{T})$ we have $\lambda_t^1 + \dots + \lambda_t^n = r_t$. Moreover, for $k \in \{1, \dots, n\}$ and $t \in [0, \mathcal{T})$,

$$\sum_{i=1}^k \lambda_t^i = \sum_{i=1}^k \lambda_0^i + 2 \sum_{i=1}^k \int_0^t \sqrt{\lambda_s^i} dB_s^i - 2\gamma \int_0^t \sum_{i=1}^k \lambda_s^i ds + k(\alpha - (n-k)\beta)t - 2\beta \int_0^t \sum_{i=1}^k \lambda_s^i \sum_{j=k+1}^n \frac{1}{\lambda_s^j - \lambda_s^i} ds.$$

On $\{\mathcal{T} < \infty\}$, $\int_0^{\mathcal{T}} \sum_{i=1}^n \lambda_s^i ds = \int_0^{\mathcal{T}} r_s ds < +\infty$ so that the second and third terms in the right-hand side are continuous functions of t on the time-interval $[0, \mathcal{T}]$. Since the fourth term in the right-hand side is also obviously continuous, the integrand in the last term in the right-hand side is non-negative and the left-hand side is also non-negative, we deduce

that $\int_0^{\mathcal{T}} \sum_{i=1}^k \lambda_s^i \sum_{j=k+1}^n \frac{1}{\lambda_s^j - \lambda_s^i} ds < +\infty$ so that the last term in the right-hand side is also continuous on the time interval $[0, \mathcal{T}]$. Therefore, still on $\{\mathcal{T} < \infty\}$, $(\lambda_t^1, \lambda_t^1 + \lambda_t^2, \dots, \lambda_t^1 + \dots + \lambda_t^n)$ admits a limit as $t \rightarrow \mathcal{T}^-$ and so does $(\lambda_t^1, \lambda_t^2, \dots, \lambda_t^n)$. Defining $(\lambda_{\mathcal{T}}^1, \lambda_{\mathcal{T}}^2, \dots, \lambda_{\mathcal{T}}^n)$ as this limit, we conclude that (2.1.1) is satisfied on the closed time interval $[0, \mathcal{T}]$ on $\{\mathcal{T} < \infty\}$.

Let us now prove the assertion by backward induction on k noticing that $\min_{k \leq \ell \leq n} \ell(\alpha - (n - \ell)\beta) = k(\alpha - (n - k)\beta)$. The idea is to show that the process $\lambda^1 + \dots + \lambda^k$ is not smaller than a CIR process which never hits zero. For $k = n$, if $n\alpha \geq 2$, then Lemma 2.6.1 applied with $a = n\alpha$, $b = 2\gamma$ and $\sigma = 2$ which satisfy $a \geq \frac{\sigma^2}{2}$, implies that the CIR process r defined by (2.4.2) and which coincides with $\lambda^1 + \dots + \lambda^n$ on $[0, \mathcal{T}]$ does not hit 0.

Let us now assume that for some $k \in \{1, \dots, n - 1\}$, $k(\alpha - (n - k)\beta) \geq 2$ and that the desired property holds for the sum of the $k+1$ first coordinates. Since $(k+1)(\alpha - (n - k - 1)\beta) > k(\alpha - (n - k)\beta) \geq 2$, the probability for the $k+1$ smallest particles to collide at zero is null. Consequently, the probability for the k smallest particles to collide at zero is the probability for exactly the k smallest particles to collide at zero and not the $(k+1)$ th which stays away from zero. We thus have :

$$\begin{aligned} \mathbb{P}\left(\mathcal{T} = \infty, \exists t \geq 0 : \lambda_t^1 + \lambda_t^2 + \dots + \lambda_t^k = 0\right) \\ &= \lim_{\varepsilon \downarrow 0} \mathbb{P}\left(\mathcal{T} = \infty, \exists t \geq 0 : \lambda_t^1 + \lambda_t^2 + \dots + \lambda_t^k = 0 \text{ and } \lambda_t^{k+1} - \lambda_t^k \geq \varepsilon\right) \\ &= \lim_{\varepsilon \downarrow 0} \mathbb{P}\left(\mathcal{T} = \infty, \exists t \geq 0 : \lambda_t^1 + \lambda_t^2 + \dots + \lambda_t^k + \mathbf{1}_{\{\lambda_t^{k+1} - \lambda_t^k < \varepsilon\}} = 0\right), \end{aligned}$$

and

$$\mathbb{P}\left(\mathcal{T} < \infty, \inf_{t \in [0, \mathcal{T})} (\lambda_t^1 + \dots + \lambda_t^k) = 0\right) = \lim_{\varepsilon \downarrow 0} \mathbb{P}\left(\mathcal{T} < \infty, \inf_{t \in [0, \mathcal{T})} (\lambda_t^1 + \dots + \lambda_t^k + \mathbf{1}_{\{\lambda_t^{k+1} - \lambda_t^k < \varepsilon\}}) = 0\right).$$

For $\varepsilon > 0$, starting with

$$\tau_\varepsilon^0 = \inf\{t \in [0, \mathcal{T}) : \lambda_t^{k+1} - \lambda_t^k \geq \varepsilon\},$$

let us define inductively for $j \in \mathbb{N}$

$$\begin{aligned} \sigma_\varepsilon^j &= \inf\{t \in (\tau_\varepsilon^j, \mathcal{T}) : \lambda_t^{k+1} - \lambda_t^k \leq \frac{\varepsilon}{2}\}, \\ \tau_\varepsilon^{j+1} &= \inf\{t \in (\sigma_\varepsilon^j, \mathcal{T}) : \lambda_t^{k+1} - \lambda_t^k \geq \varepsilon\}, \end{aligned}$$

with the convention $\inf \emptyset = \mathcal{T}$. As the function $t \rightarrow \lambda_t^{k+1} - \lambda_t^k$ is continuous on $[0, \mathcal{T})$ and even on $[0, \mathcal{T}]$ when $\mathcal{T} < +\infty$,

$$\sigma_\varepsilon^j, \tau_\varepsilon^j \xrightarrow{j \rightarrow +\infty} \mathcal{T} \text{ and } \max\{j \in \mathbb{N} : \tau_\varepsilon^j < \mathcal{T}\} < +\infty \text{ a.s. on } \{\mathcal{T} < \infty\}.$$

As $\lambda_t^{k+1} - \lambda_t^k < \varepsilon$ on $[0, \tau_\varepsilon^0)$ and $(\sigma_\varepsilon^j, \tau_\varepsilon^{j+1})$ for $j \in \mathbb{N}$, we deduce that

$$\left\{ \mathcal{T} < \infty, \inf_{t \in [0, \mathcal{T})} (\lambda_t^1 + \dots + \lambda_t^k + \mathbf{1}_{\{\lambda_t^{k+1} - \lambda_t^k < \varepsilon\}}) = 0 \right\}$$

$$\begin{aligned}
&= \left\{ \mathcal{T} < \infty, \exists j \in \mathbb{N}^*, \inf_{t \in [\tau_\varepsilon^j, \sigma_\varepsilon^j]} (\lambda_t^1 + \dots + \lambda_t^k) = 0 \right\}, \\
\text{and } &\left\{ \mathcal{T} = \infty, \exists t \geq 0 : \lambda_t^1 + \lambda_t^2 + \dots + \lambda_t^k + \mathbb{1}_{\{\lambda_t^{k+1} - \lambda_t^k < \varepsilon\}} = 0 \right\} \\
&= \left\{ \mathcal{T} = \infty, \exists j \in \mathbb{N}^*, \exists t \in [\tau_\varepsilon^j, \sigma_\varepsilon^j] : \lambda_t^1 + \lambda_t^2 + \dots + \lambda_t^k = 0 \right\}.
\end{aligned}$$

Therefore it is enough to check that

$$\begin{aligned}
&\mathbb{P} \left(\left\{ \mathcal{T} < +\infty, \exists j \in \mathbb{N}^*, \inf_{t \in [\tau_\varepsilon^j, \sigma_\varepsilon^j]} (\lambda_t^1 + \dots + \lambda_t^k) = 0 \right\} \right) \\
&+ \mathbb{P} \left(\left\{ \mathcal{T} = +\infty, \exists j \in \mathbb{N}^*, \exists t \in [\tau_\varepsilon^j, \sigma_\varepsilon^j] : \lambda_t^1 + \lambda_t^2 + \dots + \lambda_t^k = 0 \right\} \right) = 0.
\end{aligned}$$

We have for all $t \in [\tau_\varepsilon^j, \sigma_\varepsilon^j]$, $\lambda_t^{k+1} - \lambda_t^k > \frac{\varepsilon}{2}$ and

$$\begin{aligned}
d(\lambda_t^1 + \dots + \lambda_t^k) &= 2\sqrt{\lambda_t^1 + \dots + \lambda_t^k} dW_t^k - \gamma(\lambda_t^1 + \dots + \lambda_t^k) dt + k(\alpha - (n-k)\beta) dt \\
&\quad + 2\beta \sum_{i=1}^k \lambda_t^i \sum_{j=k+1}^n \frac{1}{\lambda_t^i - \lambda_t^j} dt \\
&\geq 2\sqrt{\lambda_t^1 + \dots + \lambda_t^k} dW_t^k - \gamma(\lambda_t^1 + \dots + \lambda_t^k) dt + k(\alpha - (n-k)\beta) dt - \frac{4}{\varepsilon} \beta(n-k) \sum_{i=1}^k \lambda_t^i dt \\
&\geq 2\sqrt{\lambda_t^1 + \dots + \lambda_t^k} dW_t^k - \left(\gamma + \frac{4}{\varepsilon} \beta(n-k) \right) (\lambda_t^1 + \dots + \lambda_t^k) dt + k(\alpha - (n-k)\beta) dt.
\end{aligned} \tag{2.4.3}$$

We can then define on $\{\tau_\varepsilon^j < \infty\}$ the process r^j by $r_0^j = (\lambda_{\tau_\varepsilon^j}^1 + \dots + \lambda_{\tau_\varepsilon^j}^k) \mathbb{1}_{\{\tau_\varepsilon^j < \mathcal{T}\}} + \mathbb{1}_{\{\tau_\varepsilon^j = \mathcal{T}\}}$ and for all $t \geq 0$:

$$\begin{aligned}
dr_t^j &= 2\sqrt{r_t^j} dW_{t+\tau_\varepsilon^j}^k + \left[- \left(\gamma + \frac{4}{\varepsilon} \beta(n-k) \right) r_t^j + k(\alpha - (n-k)\beta) \right] dt \\
&= 2\sqrt{r_t^j} d(W_{t+\tau_\varepsilon^j}^k - W_{\tau_\varepsilon^j}^k + W_{\tau_\varepsilon^j}^k) + \left[- \left(\gamma + \frac{4}{\varepsilon} \beta(n-k) \right) r_t^j + k(\alpha - (n-k)\beta) \right] dt \\
&= 2\sqrt{r_t^j} d\tilde{W}_t^k + \left[- \left(\gamma + \frac{4}{\varepsilon} \beta(n-k) \right) r_t^j + k(\alpha - (n-k)\beta) \right] dt,
\end{aligned}$$

where conditionally on $\{\tau_\varepsilon^j < \infty\}$, by the strong Markov property, $(\tilde{W}_t^k = W_{t+\tau_\varepsilon^j}^k - W_{\tau_\varepsilon^j}^k)_{t \geq 0}$ is a Brownian motion independent from $\mathcal{F}_{\tau_\varepsilon^j}^j$. Conditionally on $\{\tau_\varepsilon^j < \infty\}$, the process r^j is a CIR process defined globally according to Lemma 2.6.1 with $a = k(\alpha - (n-k)\beta)$ and $\sigma = 2$ which satisfy $a \geq \frac{\sigma^2}{2}$, and it stays positive on \mathbb{R}_+ .

We can conclude since by (2.4.3) and Theorem 2.6.2, for all $t \in [\tau_\varepsilon^j, \sigma_\varepsilon^j]$,

$$\lambda_t^1 + \dots + \lambda_t^k \geq r_{t-\tau_\varepsilon^j}^j.$$

Proof of Theorem 2.2.2 As explained in the introduction, the main difficulty in proving this result comes from the fact that we have to deal with both singularities when a particle hits zero and when two particles collide at the same time. For $\varepsilon > 0$, our method precisely consists in separating these difficulties by defining two new SDEs (\hat{A}_ε) and (B_ε) which each remove one type of singularity and coincide with (2.1.1) on domains that cover $\{t \geq 0 : 0 \leq \lambda_t^1 \leq \dots \leq \lambda_t^n \text{ and } \lambda_t^1 \vee (\lambda_t^2 - \lambda_t^1) \geq \varepsilon\}$. This allows us to build a solution to (2.1.1) by piecing together solutions to (\hat{A}_ε) and (B_ε).

Let us consider in this proof the Brownian motion $\mathbf{B} = (B_t^1, \dots, B_t^n)_t$, $\mathcal{F}_t = \sigma((\lambda_0^1, \dots, \lambda_0^n), (\mathbf{B}_s)_{s \leq t})$ for all $t \geq 0$, and the SDEs defined by (2.1.1). Let us define for all $\varepsilon > 0$ the following SDEs :

$$d\hat{\lambda}_t^{i,\varepsilon} = 2\sqrt{\hat{\lambda}_t^{i,\varepsilon}} dB_t^i + \left[\alpha - (n-1)\beta + 1 - \left(0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sqrt{\hat{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 \right) - 2\gamma \hat{\lambda}_t^{i,\varepsilon} + 2\beta \hat{\lambda}_t^{i,\varepsilon} \sum_{j \neq i} \frac{1}{\hat{\lambda}_t^{i,\varepsilon} - \hat{\lambda}_t^{j,\varepsilon}} \right] dt \quad (\hat{A}_\varepsilon)$$

$$0 \leq \hat{\lambda}_t^{1,\varepsilon} < \dots < \hat{\lambda}_t^{n,\varepsilon}, \text{ a.s., } dt - a.e.$$

$$\left\{ \begin{array}{l} d\tilde{\lambda}_t^{1,\varepsilon} = 2\sqrt{\tilde{\lambda}_t^{1,\varepsilon}} dB_t^1 + (\alpha - (n-1)\beta)dt - 2\gamma \tilde{\lambda}_t^{1,\varepsilon} dt - 2\beta \sum_{j \neq 1} \frac{\tilde{\lambda}_t^{1,\varepsilon} \wedge \varepsilon}{(\tilde{\lambda}_t^{j,\varepsilon} - \tilde{\lambda}_t^{1,\varepsilon} \wedge \varepsilon) \vee \varepsilon} dt \\ \forall i \in \{2, \dots, n\}, d\tilde{\lambda}_t^{i,\varepsilon} = 2\sqrt{\tilde{\lambda}_t^{i,\varepsilon}} dB_t^i + (\alpha - (n-1)\beta + 1)dt - 2\gamma \tilde{\lambda}_t^{i,\varepsilon} dt + 2\beta \sum_{j \geq 2, j \neq i} \frac{\tilde{\lambda}_t^{i,\varepsilon}}{\tilde{\lambda}_t^{i,\varepsilon} - \tilde{\lambda}_t^{j,\varepsilon}} dt \\ - \left[0 \vee \frac{2}{\sqrt{\varepsilon}} \left(\sqrt{\tilde{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2} \right) \wedge 1 \right] dt + 2\beta \frac{\tilde{\lambda}_t^{i,\varepsilon}}{(\tilde{\lambda}_t^{i,\varepsilon} - \tilde{\lambda}_t^{1,\varepsilon} \wedge \varepsilon) \vee \varepsilon} dt \end{array} \right. \quad (B_\varepsilon)$$

$$0 \leq \tilde{\lambda}_t^{1,\varepsilon} \text{ and } 0 \leq \tilde{\lambda}_t^{2,\varepsilon} < \dots < \tilde{\lambda}_t^{n,\varepsilon}, \text{ a.s., } dt - a.e.$$

These systems are built such as :

$$(\hat{A}_\varepsilon) \quad \text{coincides with (2.1.1) on } \{t, \hat{\lambda}_t^{1,\varepsilon} \geq \frac{\varepsilon}{2}\},$$

$$(B_\varepsilon) \quad \text{coincides with (2.1.1) on } \{t, \tilde{\lambda}_t^{1,\varepsilon} \leq \varepsilon \text{ and } \tilde{\lambda}_t^{2,\varepsilon} - \tilde{\lambda}_t^{1,\varepsilon} \geq \varepsilon\}.$$

Lemmas 2.4.1 and 2.4.2 give existence of global pathwise unique strong solutions to (\hat{A}_ε) and (B_ε) for any random initial condition with ordered non-negative coordinates and independent from the driving Brownian motion.

For $\xi \in \mathbb{R}_+^n$ deterministic with ordered coordinates, let $\hat{\Lambda}^{\varepsilon,T,\xi}$ denotes the process solution to (\hat{A}_ε) on $[T, +\infty)$ starting from ξ at time T and equal to 0 on $(-\infty, T)$. Likewise, $\tilde{\Lambda}^{\varepsilon,T,\xi}$ denotes the process solution to (B_ε) on $[T, +\infty)$ starting from ξ at time T and equal to 0 on $(-\infty, T)$. We define by induction :

$$\begin{aligned}
\tau_0^\varepsilon &= 0; \\
\text{let us note } \Lambda^{(0)} &= (\lambda_t^{(0),1}, \dots, \lambda_t^{(0),n})_t = \hat{\Lambda}^{\varepsilon,0,\Lambda_0}; \\
\tau_1^\varepsilon &= \mathbf{1}_{\{\lambda_0^1 \geq \varepsilon\}} \inf \left\{ t \geq 0 : \lambda_t^{(0),1} \leq \frac{\varepsilon}{2} \right\}; \\
\text{let us note } \Lambda^{(1)} &= (\lambda_t^{(1),1}, \dots, \lambda_t^{(1),n})_t = \tilde{\Lambda}^{\varepsilon, \tau_1^\varepsilon, \Lambda_{\tau_1^\varepsilon}^{(0)}}; \\
\tau_2^\varepsilon &= \inf \left\{ t \geq \tau_1^\varepsilon : \lambda_t^{(1),1} \geq \varepsilon \right\}; \\
&\vdots \\
\text{let us note } \Lambda^{(2i)} &= (\lambda_t^{(2i),1}, \dots, \lambda_t^{(2i),n})_t = \hat{\Lambda}^{\varepsilon, \tau_{2i}^\varepsilon, \Lambda_{\tau_{2i}^\varepsilon}^{(2i-1)}}; \\
\tau_{2i+1}^\varepsilon &= \inf \left\{ t \geq \tau_{2i}^\varepsilon : \lambda_t^{(2i),1} \leq \frac{\varepsilon}{2} \right\}; \\
\text{let us note } \Lambda^{(2i+1)} &= (\lambda_t^{(2i+1),1}, \dots, \lambda_t^{(2i+1),n})_t = \tilde{\Lambda}^{\varepsilon, \tau_{2i+1}^\varepsilon, \Lambda_{\tau_{2i+1}^\varepsilon}^{(2i)}}; \\
\tau_{2i+2}^\varepsilon &= \inf \left\{ t \geq \tau_{2i+1}^\varepsilon : \lambda_t^{(2i+1),1} \geq \varepsilon \right\}; \\
&\vdots
\end{aligned}$$

and as for all $i \in \mathbb{N}$, the τ_i^ε defined before are stopping times for the filtration $(\mathcal{F}_t)_{t \geq 0}$, the random vectors $\mathbf{1}_{\{\tau_i^\varepsilon < +\infty\}} \Lambda_{\tau_i^\varepsilon}^{(i)}$ are $\mathcal{F}_{\tau_i^\varepsilon}$ -measurable, the construction makes sense.

We finally define for all $\varepsilon > 0$ and $t \geq 0$:

$$\mathcal{Z}_t^\varepsilon = \mathbf{1}_{\{0 \leq t \leq \tau_1^\varepsilon\}} \Lambda_t^{(0)} + \sum_{i=1}^{+\infty} \mathbf{1}_{\{\tau_i^\varepsilon < t \leq \tau_{i+1}^\varepsilon\}} \Lambda_t^{(i)}.$$

For all $i \in \mathbb{N}$ and for $t \in [\tau_{2i+1}^\varepsilon, \tau_{2i+2}^\varepsilon)$, the equation for the smallest coordinate in (B_ε) and the non-negativity of $\lambda_t^{(2i+1),1}$ gives

$$d\lambda_t^{(2i+1),1} \leq 2\sqrt{\lambda_t^{(2i+1),1}} dB_t^1 + (\alpha - (n-1)\beta)dt. \quad (2.4.4)$$

Then, according to the pathwise comparison theorem of Ikeda and Watanabe (that we recall in Theorem 2.6.2 below), for all $i \in \mathbb{N}^*$ and for all $t \in [\tau_{2i+1}^\varepsilon, \tau_{2i+2}^\varepsilon)$

$$\lambda_t^{(2i+1),1} \leq r_{t-\tau_{2i+1}^\varepsilon}^{2i+2},$$

where for all $t \geq 0$

$$r_t^{2i+2} = \frac{\varepsilon}{2} + 2 \int_{\tau_{2i+1}^\varepsilon}^{\tau_{2i+1}^\varepsilon + t} \sqrt{r_s} dB_s^1 + (\alpha - (n-1)\beta)t,$$

which is a CIR process.

For each $i \in \mathbb{N}^*$, the delay $\tau_{2i+2}^\varepsilon - \tau_{2i+1}^\varepsilon$ is larger than the time needed by the CIR process r^{2i+2} to go from $\frac{\varepsilon}{2}$ to ε . Moreover, the times for the processes r^{2i+2} to go from $\frac{\varepsilon}{2}$ to ε are iid

positive random variables. Consequently, there is no accumulation of the stopping times τ_ε^j which go to infinity as $j \rightarrow \infty$.

The stochastic process \mathcal{Z}^ε is thus defined globally.

We recall that (\hat{A}_ε) and (B_ε) respectively coincide with (2.1.1) when $\hat{\lambda}_t^{1,\varepsilon} \geq \frac{\varepsilon}{2}$, and when $\tilde{\lambda}_t^{1,\varepsilon} \leq \varepsilon$ and $\tilde{\lambda}_t^{2,\varepsilon} - \tilde{\lambda}_t^{1,\varepsilon} \geq \varepsilon$. On the other hand, on $[\tau_{2i}^\varepsilon, \tau_{2i+1}^\varepsilon]$, \mathcal{Z}^ε evolves according to (\hat{A}_ε) and $\mathcal{Z}^{1,\varepsilon} \geq \frac{\varepsilon}{2}$ while on $[\tau_{2i+1}^\varepsilon, \tau_{2i+2}^\varepsilon]$, \mathcal{Z}^ε evolves according to (B_ε) and $\mathcal{Z}^{1,\varepsilon} \leq \varepsilon$. By induction on i we deduce that \mathcal{Z}^ε is a solution to (2.1.1) until

$$\inf \left\{ t \in \bigcup_{i \in \mathbb{N}} [\tau_{2i+1}^\varepsilon, \tau_{2i+2}^\varepsilon], \mathcal{Z}_t^{2,\varepsilon} - \mathcal{Z}_t^{1,\varepsilon} \leq \varepsilon \right\} \geq \inf \{ t \geq 0 : \mathcal{Z}_t^{1,\varepsilon} \leq \varepsilon \text{ and } \mathcal{Z}_t^{2,\varepsilon} - \mathcal{Z}_t^{1,\varepsilon} \leq \varepsilon \} =: \zeta_\varepsilon.$$

From Lemmas 2.4.1 and 2.4.2, we have :

$$\mathbb{P}\{\exists i \in \mathbb{N}, \exists t \in (\tau_{2i}^\varepsilon, \tau_{2i+1}^\varepsilon] : \mathcal{Z}_t^{j,\varepsilon} = \mathcal{Z}_t^{j+1,\varepsilon} \text{ and } \mathcal{Z}_t^{k,\varepsilon} = \mathcal{Z}_t^{k+1,\varepsilon} \text{ for some } 0 \leq j < k \leq n-1\} = 0,$$

where by convention $\mathcal{Z}^{0,\varepsilon} \equiv 0$, and

$$\mathbb{P}\{\exists i \in \mathbb{N}, \exists t \in (\tau_{2i+1}^\varepsilon, \tau_{2i+2}^\varepsilon] : \mathcal{Z}_t^{j,\varepsilon} = \mathcal{Z}_t^{j+1,\varepsilon} \text{ and } \mathcal{Z}_t^{k,\varepsilon} = \mathcal{Z}_t^{k+1,\varepsilon} \text{ for some } 2 \leq j < k \leq n-1\} = 0.$$

On the time intervals $[\tau_{2i+1}^\varepsilon \wedge \zeta_\varepsilon, \tau_{2i+2}^\varepsilon \wedge \zeta_\varepsilon]$, we have $\mathcal{Z}_t^{2,\varepsilon} - \mathcal{Z}_t^{1,\varepsilon} \geq \varepsilon$. Consequently,

$$\mathbb{P}\left\{ \exists t \in (0, \zeta_\varepsilon] : \mathcal{Z}_t^{i,\varepsilon} = \mathcal{Z}_t^{i+1,\varepsilon} \text{ and } \mathcal{Z}_t^{j,\varepsilon} = \mathcal{Z}_t^{j+1,\varepsilon} \text{ for some } 1 \leq i < j \leq n-1 \right\} = 0. \quad (2.4.5)$$

As the solutions to equation (2.1.1) are pathwise unique (see Lemma 2.4.2), for $n \in \mathbb{N}^*$, the processes $\mathcal{Z}^{\frac{1}{n}}$ and $\mathcal{Z}^{\frac{1}{n+1}}$ coincide on $[0, \zeta_{\frac{1}{n}} \wedge \zeta_{\frac{1}{n+1}}]$. Thus, $\zeta_{\frac{1}{n}} \wedge \zeta_{\frac{1}{n+1}} = \zeta_{\frac{1}{n}}$ and the sequence $(\zeta_{\frac{1}{n}})_{n \in \mathbb{N}^*}$ is non-decreasing. Moreover, for all $n \in \mathbb{N}^*$, $\mathcal{Z}^{\frac{1}{n}}$ verifies (2.4.5). Consequently, we can define for all $t \in [0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon)$

$$\Lambda_t = \mathcal{Z}_t^1 \mathbf{1}_{\{0 \leq t \leq \zeta_1\}} + \sum_{n \geq 1} \mathcal{Z}_t^{\frac{1}{n+1}} \mathbf{1}_{\{\zeta_{\frac{1}{n}} < t \leq \zeta_{\frac{1}{n+1}}\}}. \quad (2.4.6)$$

which is a solution to the SDEs (2.1.1) on $[0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon)$ verifying (iii) of Theorem 2.2.2.

Finally, as the solutions to (2.1.1) are pathwise unique (Lemma 2.3.1), we can apply the Yamada-Watanabe theorem (see for instance [RY99, Theorem 1.7 p.368]) to deduce the existence of strong solutions to the equation.

Since on $\{\zeta_\varepsilon < +\infty\}$, $\lambda_{\zeta_\varepsilon}^1 + \lambda_{\zeta_\varepsilon}^2 = 2\lambda_{\zeta_\varepsilon}^1 + \lambda_{\zeta_\varepsilon}^2 - \lambda_{\zeta_\varepsilon}^1 \leq 3\varepsilon$, on $\{\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon < +\infty\}$ we have $\inf_{t \in [0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon)} \lambda_t^1 + \lambda_t^2 = 0$. We then use Proposition 2.2.4 (ii) with $k = 2$ to conclude that $\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon = +\infty$ when $\alpha - (n-1)\beta \geq 1 - \beta$ which is (i) from Theorem 2.2.2.

For $\alpha - (n - 1)\beta < 1 - \beta$ and $\gamma \geq 0$, let us prove assertion (ii) of Theorem 2.2.2. Following the steps of the proof of Proposition 2.2.4 with $k = 2$ until (2.4.1), we have for all $0 \leq t < \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon$:

$$d(\lambda_t^1 + \lambda_t^2) \leq 2\sqrt{\lambda_t^1 + \lambda_t^2} dW_t^2 - 2\gamma(\lambda_t^1 + \lambda_t^2)dt + 2(\alpha - (n - 2)\beta)dt,$$

where by Lévy's characterization, W^2 defined by $W_0^2 = 0$ and

$$dW_t^2 = \mathbb{1}_{\left\{0 \leq t < \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon\right\}} \frac{\sqrt{\lambda_t^1} dB_t^1 + \sqrt{\lambda_t^2} dB_t^2}{\sqrt{\lambda_t^1 + \lambda_t^2}} + \mathbb{1}_{\left\{t \geq \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon\right\}} \frac{dB_t^1 + dB_t^2}{\sqrt{2}}$$
 is a Brownian motion.

By the pathwise comparison theorem of Ikeda and Watanabe (that we recall in Theorem 2.6.2 below),

$$\lambda_t^1 + \lambda_t^2 \leq r_t \text{ for all } 0 \leq t < \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon \text{ a.s.}$$

where for all $t \geq 0$

$$r_t = \lambda_0^1 + \lambda_0^2 + 2 \int_0^t \sqrt{r_s} dW_s^2 - 2\gamma \int_0^t r_s ds + 2(\alpha - (n - 2)\beta)t$$

is a CIR process. Applying Lemma 2.6.1 with $a = 2(\alpha - (n - 2)\beta)$, $b = 2\gamma$ and $\sigma = 2$ which satisfy $a < \frac{\sigma^2}{2}$ and $b \geq 0$, we conclude that the hitting time of 0 by r is finite almost surely. Consequently, $\mathbb{P}\left(\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon = +\infty\right) = 0$ which concludes the proof of Theorem 2.2.2 (ii).

Lemma 2.4.1. *Let us assume $\alpha - (n - 1)\beta > 0$. The system*

$$d\hat{\lambda}_t^{i,\varepsilon} = 2\sqrt{\hat{\lambda}_t^{i,\varepsilon}} dB_t^i + \left[\alpha - (n - 1)\beta + 1 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sqrt{\hat{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 - 2\gamma \hat{\lambda}_t^{i,\varepsilon} + 2\beta \hat{\lambda}_t^{i,\varepsilon} \sum_{j \neq i} \frac{1}{\hat{\lambda}_t^{i,\varepsilon} - \hat{\lambda}_t^{j,\varepsilon}} \right] dt$$

$$0 \leq \hat{\lambda}_t^{1,\varepsilon} < \dots < \hat{\lambda}_t^{n,\varepsilon}, \text{ a.s., } dt - a.e.$$

(\hat{A}_ε)

has a global pathwise unique strong solution $(\hat{\lambda}_t^{1,\varepsilon}, \dots, \hat{\lambda}_t^{n,\varepsilon})_t$ starting from any random initial condition $\Lambda_0 = (\lambda_0^1, \dots, \lambda_0^n)$ independent from \mathbf{B} such that $0 \leq \lambda_0^1 \leq \dots \leq \lambda_0^n$ a.s.

Moreover,

$$\mathbb{P}\{\exists t > 0 : \hat{\lambda}_t^{i,\varepsilon} = \hat{\lambda}_t^{i+1,\varepsilon} \text{ and } \hat{\lambda}_t^{j,\varepsilon} = \hat{\lambda}_t^{j+1,\varepsilon} \text{ for some } 0 \leq i < j \leq n - 1\} = 0, \quad (2.4.7)$$

where by convention $\hat{\lambda}^{0,\varepsilon} \equiv 0$.

Proof. Let us consider $\mathcal{F}_t = \sigma\left(\left(\sqrt{\lambda_0^1}, \dots, \sqrt{\lambda_0^n}\right), (\mathbf{B}_s)_{s \leq t}\right)$ and the system of SDEs defined by

$$dx_t^{i,\varepsilon} = dB_t^i + \frac{\alpha - (n - 1)\beta - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(x_t^{i,\varepsilon} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}}\right) \wedge 1}{2x_t^{i,\varepsilon}} dt - \gamma x_t^{i,\varepsilon} dt + \beta x_t^{i,\varepsilon} \sum_{j \neq i} \frac{dt}{(x_t^{i,\varepsilon})^2 - (x_t^{j,\varepsilon})^2} \text{ for } i \in \{1, \dots, n\}$$

$$0 \leq x_t^{1,\varepsilon} < \dots < x_t^{n,\varepsilon}, \text{ a.s., } dt - a.e.$$

(A_ε)

with random initial condition $(\sqrt{\lambda_0^1}, \dots, \sqrt{\lambda_0^n})$ such that $0 \leq \sqrt{\lambda_0^1} \leq \dots \leq \sqrt{\lambda_0^n}$.

We are going to apply Cépa's multivalued equations theory ([Cép95]) to conclude that there exists a unique strong solution to (A_ε) .

To do so, we define

$$\begin{aligned}
D &= \{0 < x^1 < x^2 < \dots < x^n\}, \\
\Phi_\gamma &: (x^1, \dots, x^n) \in \mathbb{R}^n \rightarrow \begin{cases} -\sum_{i=1}^n \left[\frac{\alpha - (n-1)\beta}{2} \ln |x^i| - \frac{\gamma}{2} (x^i)^2 + \frac{\beta}{4} \sum_{j \neq i} (\ln |x^i - x^j| + \ln |x^i + x^j|) \right] & \text{if } x \in D \\ +\infty & \text{if } x \notin D, \end{cases} \\
g &: (x^1, \dots, x^n) \in \mathbb{R}^n \rightarrow \left(-\frac{0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(x_t^{1,\varepsilon} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1}{2x_t^{1,\varepsilon}}, \dots, -\frac{0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(x_t^{n,\varepsilon} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1}{2x_t^{n,\varepsilon}} \right),
\end{aligned}$$

to rewrite the system of SDEs on D with $X = (x_t^{1,\varepsilon}, \dots, x_t^{n,\varepsilon})_t$ the following way

$$dX_t = d\mathbf{B}_t + g(X_t)dt - \nabla\Phi_\gamma(X_t)dt. \quad (A_\varepsilon)$$

Since g is globally Lipschitz and Φ_γ is convex, Cépa's multivalued equations theory and in particular [CL01, Theorem 2.2] combined with the remark following [Cép95, Theorem 3.1] show the existence of a unique strong solution to equation

$$\begin{aligned}
d\tilde{X}_t &= d\mathbf{B}_t + g(\tilde{X}_t)dt - \nabla\Phi_\gamma(\tilde{X}_t)dt - \nu(\tilde{X}_t)dL_t \text{ for all } t \geq 0 & (\tilde{A}_\varepsilon) \\
\forall t \geq 0, \tilde{X}_t &\in \bar{D} \text{ a.s.} \\
\tilde{X}_0 &= (\sqrt{\lambda_0^1}, \dots, \sqrt{\lambda_0^n}),
\end{aligned}$$

where \tilde{X} is a continuous adapted to $(\mathcal{F}_t)_{t \geq 0}$ process, L is a continuous non-decreasing adapted to $(\mathcal{F}_t)_{t \geq 0}$ process with $L_0 = 0$ verifying

$$L_t = \int_0^t \mathbf{1}_{\{\tilde{X}_s \in \partial D\}} dL_s,$$

and $\nu(x) \in \pi(x)$ ($\pi(x)$ is the set of unitary outward normals to ∂D at $x \in \partial D$). The solution to equation (\tilde{A}_ε) follows the conditions : for all $t > 0$

$$\begin{aligned}
\mathbb{E} \left[\int_0^t \mathbf{1}_{\{\tilde{X}_s \in \partial D\}} ds \right] &= 0, \\
\mathbb{E} \left[\int_0^t |\nabla\Phi_\gamma(\tilde{X}_s)| ds \right] &< \infty.
\end{aligned}$$

Let us now prove that the boundary process L is equal to zero.

For all $m \in \{1, \dots, n\}$, for all $t \geq 0$, we have with $C = \frac{\alpha - (n-1)\beta}{2}$:

$$\mathbb{E} \int_0^t \left| \frac{C}{x_s^{m,\varepsilon}} + \beta x_s^{m,\varepsilon} \sum_{j \neq m} \frac{1}{(x_s^{m,\varepsilon})^2 - (x_s^{j,\varepsilon})^2} \right| ds < \infty. \quad (2.4.8)$$

Let us prove by backward induction on m that

$$\text{for all } 1 < m \leq n \text{ and for all } t \geq 0, \mathbb{E} \int_0^t \left| \frac{1}{x_s^{m,\varepsilon}} \right| + \sum_{l < m} \left| \frac{x_s^{m,\varepsilon}}{(x_s^{m,\varepsilon})^2 - (x_s^{l,\varepsilon})^2} \right| ds < \infty. \quad (H_m)$$

- $m = n$

As all the terms in the absolute value of (2.4.8) have the same sign for $m = n$, we deduce the individual integrability.

- Let $1 < m \leq n$ and let us assume (H_j) for all $j \in \{m+1, \dots, n\}$.

We have :

$$\frac{C}{x_s^{m,\varepsilon}} + \beta x_s^{m,\varepsilon} \sum_{j \neq m} \frac{1}{(x_s^{m,\varepsilon})^2 - (x_s^{j,\varepsilon})^2} = \frac{C}{x_s^{m,\varepsilon}} + \beta x_s^{m,\varepsilon} \sum_{j < m} \frac{1}{(x_s^{m,\varepsilon})^2 - (x_s^{j,\varepsilon})^2} - \beta \sum_{j > m} \frac{x_s^{m,\varepsilon}}{(x_s^{j,\varepsilon})^2 - (x_s^{m,\varepsilon})^2}.$$

Let us remark that for all $s \geq 0$

$$\frac{C}{x_s^{m,\varepsilon}} + \beta x_s^{m,\varepsilon} \sum_{j < m} \frac{1}{(x_s^{m,\varepsilon})^2 - (x_s^{j,\varepsilon})^2} \leq \left| \frac{C}{x_s^{m,\varepsilon}} + \beta x_s^{m,\varepsilon} \sum_{j \neq m} \frac{1}{(x_s^{m,\varepsilon})^2 - (x_s^{j,\varepsilon})^2} \right| + \beta \sum_{j > m} \frac{x_s^{j,\varepsilon}}{(x_s^{j,\varepsilon})^2 - (x_s^{m,\varepsilon})^2}, \quad (2.4.9)$$

by the triangle inequality and since $x_s^{j,\varepsilon} \geq x_s^{m,\varepsilon}$ for $j > m$.

By (2.4.8) and the induction hypothesis for $j \in \{m+1, \dots, n\}$, each term in the right-hand side of (2.4.9) is integrable, which ends the induction argument.

Consequently, for all $1 \leq l < m \leq n$ and for all $t \geq 0$ we have

$$\mathbb{E} \int_0^t \frac{1}{x_s^{m,\varepsilon} - x_s^{l,\varepsilon}} ds = \mathbb{E} \int_0^t \frac{x_s^{m,\varepsilon} + x_s^{l,\varepsilon}}{(x_s^{m,\varepsilon})^2 - (x_s^{l,\varepsilon})^2} ds \leq 2 \mathbb{E} \int_0^t \frac{x_s^{m,\varepsilon}}{(x_s^{m,\varepsilon})^2 - (x_s^{l,\varepsilon})^2} ds < \infty. \quad (2.4.10)$$

As in the second part of the proof of [CL01, Theorem 2.2] (equation (2.40)), using the occupation times formula and (2.4.10), we have for $1 \leq l < m \leq n, t \geq 0$

$$\int_0^{+\infty} \frac{L_t^a(x^{m,\varepsilon} - x^{l,\varepsilon})}{a} da = \int_0^t \frac{d\langle x^{m,\varepsilon} - x^{l,\varepsilon} \rangle_s}{x_s^{m,\varepsilon} - x_s^{l,\varepsilon}} = 2 \int_0^t \frac{1}{x_s^{m,\varepsilon} - x_s^{l,\varepsilon}} ds < +\infty, \quad (2.4.11)$$

and

$$\int_0^{+\infty} \frac{L_t^a(x^{1,\varepsilon})}{a} da = \int_0^t \frac{d\langle x^{1,\varepsilon} \rangle_s}{x_s^{1,\varepsilon}} = \int_0^t \frac{1}{x_s^{1,\varepsilon}} ds < +\infty,$$

where $L_t^a(\mathcal{X})$ is the local time at time t and on level a for a real continuous semimartingale \mathcal{X} . Since the function $a \mapsto \frac{1}{a}$ is not integrable at 0 and $(L_t^a(X))$ is cadlag at a by [RY99, Theorem 1.7 p.225], one deduces that $L_t^0(x^{m,\varepsilon} - x^{l,\varepsilon}) = L_t^0(x^{1,\varepsilon}) = 0$.

From there, the reasoning made in the proof of [CL01, Theorem 2] allows to conclude that the boundary process L is equal to zero.

Then, setting $\hat{\lambda}^{i,\varepsilon} = (x^{i,\varepsilon})^2$ for all $i \in \{1, \dots, n\}$ we obtain a global solution to (\hat{A}_ε) . Following the approach used in the proof of Lemma 2.4.2 below to demonstrate pathwise uniqueness for the slightly more complicated system (B_ε) , we obtain that the solutions to (\hat{A}_ε) are pathwise unique. The Yamada-Watanabe Theorem (see for instance [RY99, Theorem 1.7 p.368]), allows to conclude that (\hat{A}_ε) has a pathwise unique global strong solution.

Let us now prove (2.4.7).

Let us consider the system defined by (A_ε) with initial condition $0 \leq \sqrt{\lambda_0^1} \leq \dots \leq \sqrt{\lambda_0^n}$.

Let us define for all $\varepsilon > 0, M > 0$

$$\tau_M = \inf\{t \geq 0 : \exists i \in \{1, \dots, n\}, x_t^{i,\varepsilon} \geq M\},$$

and for $t \in [0; \tau_M) : \Theta(t) = (\theta_1(t), \dots, \theta_n(t))$ with

$$\forall i \in \{1, \dots, n\}, \theta_i(t) = -\frac{0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(x_t^{i,\varepsilon} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}}\right) \wedge 1}{2x_t^{i,\varepsilon}} + \gamma x_t^{i,\varepsilon},$$

and for all $t \geq 0$

$$Z(t) = \exp\left\{-\int_0^{t \wedge \tau_M} \Theta(u) \cdot d\mathbf{B}_u - \frac{1}{2} \int_0^{t \wedge \tau_M} \|\Theta(u)\|^2 du\right\}.$$

We have for all $i \in \{1, \dots, n\}$,

$$\begin{aligned} \theta_i^2(t) &= \left(-\frac{0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(x_t^{i,\varepsilon} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}}\right) \wedge 1}{2x_t^{i,\varepsilon}} + \gamma x_t^{i,\varepsilon}\right)^2 \\ &\leq 2 \left(\frac{0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(x_t^{i,\varepsilon} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}}\right) \wedge 1}{2x_t^{i,\varepsilon}}\right)^2 + 2(\gamma x_t^{i,\varepsilon})^2 \\ &\leq \frac{1}{\varepsilon} + 2\gamma^2 M^2. \end{aligned}$$

We thus have

$$\mathbb{E}\left[\exp\left\{\frac{1}{2} \int_0^{t \wedge \tau_M} \|\Theta(u)\|^2 du\right\}\right] < \infty \text{ for all } t \geq 0.$$

Then, according to Novikov's criterion (see for instance [KS91, Proposition 5.12 p.198]), Z is a \mathbb{P} -martingale, and $\mathbb{E}[Z(t)] = 1$. Consequently, recalling that $\mathcal{F}_t = \sigma\left((\sqrt{\lambda_0^1}, \dots, \sqrt{\lambda_0^n}), (\mathbf{B}_s)_{s \leq t}\right)$ and defining \mathbb{Q} such that

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_t} = Z(t),$$

and

$$\text{for all } i \in \{1, \dots, n\}, \check{B}_t^{i,M} = B_t^i + \int_0^{t \wedge \tau_M} \theta_i(s) ds$$

$$= B_t^i - \int_0^{t \wedge \tau_M} \frac{0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(x_s^{i,\varepsilon} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1}{2x_s^{i,\varepsilon}} + \gamma x_s^{i,\varepsilon} ds, \quad t \geq 0,$$

$\check{\mathbf{B}}^M = (\check{B}_t^{1,M}, \dots, \check{B}_t^{n,M})_t$ is a \mathbb{Q} -Brownian motion according to the Girsanov theorem (see for instance [KS91, Proposition 5.4 p.194]).

Thus, (A_ε) can be rewritten in terms of $\check{\mathbf{B}}^M$ as

$$dx_t^{i,\varepsilon} = d\check{B}_t^{i,M} + \frac{\alpha - (n-1)\beta}{2x_t^{i,\varepsilon}} dt - \gamma x_t^{i,\varepsilon} \mathbb{1}_{\{t \geq \tau_M\}} dt + \beta x_t^{i,\varepsilon} \sum_{j \neq i} \frac{dt}{(x_t^{i,\varepsilon})^2 - (x_t^{j,\varepsilon})^2} \quad \text{for } i \in \{1, \dots, n\}. \quad (\check{A}_{\varepsilon, M})$$

By the same arguments as in the beginning of the proof, the normally reflected SDE

$$\begin{aligned} d\check{X}_t &= d\check{\mathbf{B}}_t^M - \nabla \Phi_0(\check{X}_t) dt - \nu(\check{X}_t) d\check{L}_t \quad \text{for all } t \geq 0 \\ \forall t \geq 0, \check{X}_t &\in \bar{D} \quad \text{a.s.} \\ \check{X}_0 &= (\sqrt{\lambda_0^1}, \dots, \sqrt{\lambda_0^n}) \end{aligned} \quad (2.4.12)$$

admits a global solution and the term $\nu(\check{X}_t) d\check{L}_t$ is zero. We can apply [Lé10, Theorem 3.1] to the SDE (2.4.12) to conclude that its solutions cannot have multiple collisions. This last SDE can be rewritten

$$\begin{aligned} d\check{x}_t^{i,\varepsilon} &= d\check{B}_t^{i,M} + \frac{\alpha - (n-1)\beta}{2\check{x}_t^{i,\varepsilon}} dt + \beta \check{x}_t^{i,\varepsilon} \sum_{j \neq i} \frac{dt}{(\check{x}_t^{i,\varepsilon})^2 - (\check{x}_t^{j,\varepsilon})^2} \quad \text{for } i \in \{1, \dots, n\} \\ 0 &\leq \check{x}_t^{1,\varepsilon} < \dots < \check{x}_t^{n,\varepsilon}, \quad \text{a.s., } dt - a.e. \end{aligned}$$

By pathwise uniqueness deduced from the application of Lemma 2.3.1 after the square of each coordinate change of variable, the solutions to this last system of SDEs coincide with the solutions to $(\check{A}_{\varepsilon, M})$ on $[0, \tau_M)$, which implies that there is no collision of $(x_t^{1,\varepsilon}, \dots, x_t^{n,\varepsilon})_t$ on $[0, \tau_M)$ under the probability \mathbb{Q} . There is thus no multiple collision of $(x_t^{1,\varepsilon}, \dots, x_t^{n,\varepsilon})_t$ under the probability \mathbb{P} on $[0, \tau_M)$:

$$\mathbb{P}\{\exists t \in [0, \tau_M) : x_t^{i,\varepsilon} = x_t^{i+1,\varepsilon} \text{ and } x_t^{j,\varepsilon} = x_t^{j+1,\varepsilon} \text{ for some } 0 \leq i < j \leq n-1\} = 0.$$

Since it is true for all $M > 0$, and since as the system of SDEs (A_ε) admits a continuous global solution, $\tau_M \xrightarrow{M \rightarrow +\infty} +\infty$ $\mathbb{P} - a.s.$, we have the result for (A_ε) , and thus for (\hat{A}_ε) , which concludes the proof. \square

Lemma 2.4.2. *Let us assume $\alpha - (n-1)\beta > 0$. The system of SDEs (B_ε) with random initial condition $(\tilde{\lambda}_0^{1,\varepsilon}, \dots, \tilde{\lambda}_0^{n,\varepsilon})$ such that $0 \leq \tilde{\lambda}_0^{1,\varepsilon} \leq \dots \leq \tilde{\lambda}_0^{n,\varepsilon}$ a.s. and independent from the Brownian motion $\mathbf{B} = (B_t^1, \dots, B_t^n)_t$ has a global pathwise unique strong solution $(\tilde{\lambda}_t^{1,\varepsilon}, \dots, \tilde{\lambda}_t^{n,\varepsilon})_t$.*

Moreover,

$$\mathbb{P}\{\exists t > 0 : \tilde{\lambda}_t^{i,\varepsilon} = \tilde{\lambda}_t^{i+1,\varepsilon} \text{ and } \tilde{\lambda}_t^{j,\varepsilon} = \tilde{\lambda}_t^{j+1,\varepsilon} \text{ for } 2 \leq i < j \leq n-1\} = 0. \quad (2.4.13)$$

Proof. Let us consider a Brownian motion $\tilde{\mathbf{B}} = (\tilde{B}_t^1, \dots, \tilde{B}_t^n)_t$ and the system

$$\begin{aligned} d\tilde{\lambda}_t^{1,\varepsilon} &= 2\sqrt{\tilde{\lambda}_t^{1,\varepsilon}} d\tilde{B}_t^1 + (\alpha - (n-1)\beta)dt - 2\gamma\tilde{\lambda}_t^{1,\varepsilon}dt & (2.4.14) \\ \forall i \in \{2, \dots, n\}, d\tilde{\lambda}_t^{i,\varepsilon} &= 2\sqrt{\tilde{\lambda}_t^{i,\varepsilon}} d\tilde{B}_t^i + (\alpha - (n-1)\beta + 1)dt - \left[0 \vee \frac{2}{\sqrt{\varepsilon}} \left(\sqrt{\tilde{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2} \right) \wedge 1 \right] dt \\ &\quad - 2\gamma\tilde{\lambda}_t^{i,\varepsilon}dt + 2\beta \sum_{j \geq 2, j \neq i} \frac{\tilde{\lambda}_t^{i,\varepsilon}}{\tilde{\lambda}_t^{i,\varepsilon} - \tilde{\lambda}_t^{j,\varepsilon}} dt. \end{aligned}$$

Let us remark that $\tilde{\lambda}^1$ is a CIR process, and that the coordinates $i \in \{2, \dots, n\}$ satisfy an autonomous system of SDEs for $n-1$ particles similar to equation (\hat{A}_ε) . The only differences are the coefficient $(n-1)\beta$ in (\hat{A}_ε) which remains $(n-1)\beta$ here and is thus unaffected by the change in the number of particles, and the terms $\left[0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sqrt{\tilde{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 \right]$ in (\hat{A}_ε) which becomes $\left[0 \vee \frac{2}{\sqrt{\varepsilon}} \left(\sqrt{\tilde{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2} \right) \wedge 1 \right]$ here. We can still consider the root process $(\sqrt{\tilde{\lambda}_t^{2,\varepsilon}}, \dots, \sqrt{\tilde{\lambda}_t^{n,\varepsilon}})_t$, and apply the same method as in the proof of Lemma 2.4.1 to prove the existence of a strong solution to this subsystem. Equation (2.4.14) also has a strong solution (Lemma 2.6.1), and consequently the whole n -particles system considered here admits a global strong solution.

Let us define for all $\varepsilon > 0$ and for $t \geq 0$: $\Theta(t) = (\theta_1(t), \dots, \theta_n(t))$ with

$$\begin{aligned} \theta_1(t) &= -\frac{\beta}{\sqrt{\tilde{\lambda}_t^{1,\varepsilon}}} \sum_{j \neq 1} \frac{\tilde{\lambda}_t^{1,\varepsilon} \wedge \varepsilon}{(\tilde{\lambda}_t^{j,\varepsilon} - \tilde{\lambda}_t^{1,\varepsilon} \wedge \varepsilon) \vee \varepsilon}, \\ \forall i \in \{2, \dots, n\}, \theta_i(t) &= \frac{\beta}{\sqrt{\tilde{\lambda}_t^{i,\varepsilon}}} \frac{\tilde{\lambda}_t^{i,\varepsilon}}{(\tilde{\lambda}_t^{i,\varepsilon} - \tilde{\lambda}_t^{1,\varepsilon} \wedge \varepsilon) \vee \varepsilon}, \end{aligned}$$

and for all $t \geq 0$

$$Z(t) = \exp \left\{ \int_0^t \Theta(u) \cdot d\tilde{\mathbf{B}}_u - \frac{1}{2} \int_0^t \|\Theta(u)\|^2 du \right\}.$$

We have

$$\begin{aligned} \theta_1^2(t) &= \frac{\beta^2}{\tilde{\lambda}_t^{1,\varepsilon}} \left(\sum_{j > 1} \frac{\tilde{\lambda}_t^{1,\varepsilon} \wedge \varepsilon}{(\tilde{\lambda}_t^{j,\varepsilon} - \tilde{\lambda}_t^{1,\varepsilon} \wedge \varepsilon) \vee \varepsilon} \right)^2 \\ &\leq \frac{\beta^2 (\tilde{\lambda}_t^{1,\varepsilon} \wedge \varepsilon)^2}{\tilde{\lambda}_t^{1,\varepsilon}} \left(\sum_{j > 1} \frac{1}{\varepsilon} \right)^2 \\ &\leq \frac{(n-1)^2 \beta^2}{\varepsilon}. \end{aligned} \tag{2.4.15}$$

For all $1 < i \leq n$,

$$\theta_i^2(t) = \frac{\beta^2}{\tilde{\lambda}_t^{i,\varepsilon}} \frac{(\tilde{\lambda}_t^{i,\varepsilon})^2}{((\tilde{\lambda}_t^{i,\varepsilon} - \tilde{\lambda}_t^{1,\varepsilon} \wedge \varepsilon) \vee \varepsilon)^2}$$

$$\begin{aligned}
&\leq \beta^2 \frac{(\tilde{\lambda}_t^{i,\varepsilon} - \tilde{\lambda}_t^{1,\varepsilon} \wedge \varepsilon)^+ + \tilde{\lambda}_t^{1,\varepsilon} \wedge \varepsilon}{((\tilde{\lambda}_t^{i,\varepsilon} - \tilde{\lambda}_t^{1,\varepsilon} \wedge \varepsilon) \vee \varepsilon)^2} \\
&\leq \frac{2\beta^2}{\varepsilon}.
\end{aligned} \tag{2.4.16}$$

We thus have

$$\mathbb{E} \left[\exp \left\{ \frac{1}{2} \int_0^t \|\Theta(u)\|^2 du \right\} \right] < \infty \text{ for all } t \geq 0.$$

Then, according to Novikov's criterion (see for instance [KS91, Proposition 5.12 p.198]), Z is a \mathbb{P} -martingale, and $\mathbb{E}[Z(t)] = 1$ for all $t \geq 0$. Consequently, defining for all $t \geq 0$, $\tilde{\mathcal{F}}_t = \sigma \left((\tilde{\mathbf{B}}_s)_{s \leq t}, (\tilde{\lambda}_0^{1,\varepsilon}, \dots, \tilde{\lambda}_0^{n,\varepsilon}) \right)$ and \mathbb{Q} such that

$$\frac{d\mathbb{Q}}{d\mathbb{P} |_{\tilde{\mathcal{F}}_t}} = Z(t),$$

and

$$\begin{aligned}
\tilde{B}_t^1 &= \tilde{B}_t^1 - \int_0^t \theta_1(s) ds \\
&= \tilde{B}_t^1 + \int_0^t \frac{\beta}{\sqrt{\tilde{\lambda}_s^{1,\varepsilon}}} \sum_{j \neq 1} \frac{\tilde{\lambda}_s^{1,\varepsilon} \wedge \varepsilon}{(\tilde{\lambda}_s^{j,\varepsilon} - \tilde{\lambda}_s^{1,\varepsilon} \wedge \varepsilon) \vee \varepsilon} ds, \\
\text{for all } i \in \{2, \dots, n\}, \tilde{B}_t^i &= \tilde{B}_t^i - \int_0^t \theta_i(s) ds \\
&= \tilde{B}_t^i - \int_0^t \frac{\beta}{\sqrt{\tilde{\lambda}_s^{i,\varepsilon}}} \frac{\tilde{\lambda}_s^{i,\varepsilon}}{(\tilde{\lambda}_s^{i,\varepsilon} - \tilde{\lambda}_s^{1,\varepsilon} \wedge \varepsilon) \vee \varepsilon} ds, \quad 0 \leq t,
\end{aligned}$$

$\tilde{\mathbf{B}} = (\tilde{B}_t^1, \dots, \tilde{B}_t^n)_t$ is a \mathbb{Q} -Brownian motion according to the Girsanov theorem (see for instance [KS91, Proposition 5.4 p.194]).

Consequently, (B_ε) has a global weak solution.

We now have to prove the pathwise uniqueness of the solutions to (B_ε) . The differences with Lemma 2.3.1 are the term

$$\left[0 \vee \frac{2}{\sqrt{\varepsilon}} \left(\sqrt{\tilde{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2} \right) \wedge 1 \right],$$

and the interaction terms between the first particle and the others.

Let $Z = (z_t^1, \dots, z_t^n)_t$ and $\tilde{Z} = (\tilde{z}_t^1, \dots, \tilde{z}_t^n)_t$ be two global solutions to (B_ε) with $Z_0 = \tilde{Z}_0$ independent from the same driving Brownian motion $\mathbf{B} = (B_t^1, \dots, B_t^n)_t$.

Let $M > 0$ and

$$\tau_M = \inf \{ t \geq 0 : \tilde{z}_t^1 + z_t^1 + \tilde{z}_t^n + z_t^n \geq M \}.$$

As Z and \tilde{Z} are continuous and assumed well defined on \mathbb{R}_+ , $\tau_M \uparrow \infty$ when $M \uparrow \infty$.

The local time of $z^i - \tilde{z}^i$ at 0 is zero ([RY99, Lemma 3.3 p.389]). Applying the Tanaka formula to the process $z^i - \tilde{z}^i$ stopped at τ_M and summing over i ,

$$\begin{aligned} \sum_{i=1}^n |z_{t \wedge \tau_M}^i - \tilde{z}_{t \wedge \tau_M}^i| &= \sum_{i=1}^n |z_0^i - \tilde{z}_0^i| + \sum_{i=1}^n \int_0^{t \wedge \tau_M} \operatorname{sgn}(z_s^i - \tilde{z}_s^i) d(z_s^i - \tilde{z}_s^i) \\ &= 2 \sum_{i=1}^n \int_0^{t \wedge \tau_M} |\sqrt{z_s^i} - \sqrt{\tilde{z}_s^i}| dB_s^i - 2\gamma \int_0^{t \wedge \tau_M} \sum_{i=1}^n |z_s^i - \tilde{z}_s^i| ds \\ &\quad + 2\beta \int_0^{t \wedge \tau_M} \sum_{i=2}^n \operatorname{sgn}(z_s^i - \tilde{z}_s^i) \sum_{j \geq 2, j \neq i} \left(\frac{z_s^i}{z_s^i - z_s^j} - \frac{\tilde{z}_s^i}{\tilde{z}_s^i - \tilde{z}_s^j} \right) ds \end{aligned} \quad (2.4.17)$$

$$\begin{aligned} &\quad + 2\beta \int_0^{t \wedge \tau_M} \sum_{i=2}^n \left\{ \operatorname{sgn}(z_s^i - \tilde{z}_s^i) \left(\frac{z_s^i}{(z_s^i - z_s^1 \wedge \varepsilon) \vee \varepsilon} - \frac{\tilde{z}_s^i}{(\tilde{z}_s^i - \tilde{z}_s^1 \wedge \varepsilon) \vee \varepsilon} \right) \right. \\ &\quad \left. - \operatorname{sgn}(z_s^1 - \tilde{z}_s^1) \left(\frac{z_s^1 \wedge \varepsilon}{(z_s^i - z_s^1 \wedge \varepsilon) \vee \varepsilon} - \frac{\tilde{z}_s^1 \wedge \varepsilon}{(\tilde{z}_s^i - \tilde{z}_s^1 \wedge \varepsilon) \vee \varepsilon} \right) \right\} ds \end{aligned} \quad (2.4.18)$$

$$\begin{aligned} &\quad - \int_0^{t \wedge \tau_M} \sum_{i=2}^n \operatorname{sgn}(z_s^i - \tilde{z}_s^i) \left\{ \left[0 \vee \frac{2}{\sqrt{\varepsilon}} \left(\sqrt{z_t^i} - \frac{\sqrt{\varepsilon}}{2} \right) \wedge 1 \right] - \left[0 \vee \frac{2}{\sqrt{\varepsilon}} \left(\sqrt{\tilde{z}_t^i} - \frac{\sqrt{\varepsilon}}{2} \right) \wedge 1 \right] \right\} ds. \end{aligned} \quad (2.4.19)$$

As the processes are stopped at τ_M , the expectation of the stochastic integrals is zero. As in the proof of Lemma 2.3.1, the terms (2.4.17) are not positive. To deal with the expectation of (2.4.18), one remarks that the function

$$(x, y) \mapsto \frac{x}{(x - y \wedge \varepsilon) \vee \varepsilon}$$

is Lipschitz on $[0, M]^2$.

As for the term (2.4.19), the function $f : z \mapsto \left[0 \vee \frac{2}{\sqrt{\varepsilon}} \left(\sqrt{z} - \frac{\sqrt{\varepsilon}}{2} \right) \wedge 1 \right]$ defined on \mathbb{R}_+ is constant on $\left[0, \frac{\varepsilon}{4} \right)$ and on $[\varepsilon, +\infty)$ and is differentiable on $\left(\frac{\varepsilon}{4}, \varepsilon \right)$ with

$$f'(z) = \frac{1}{\sqrt{\varepsilon z}} \leq \frac{2}{\varepsilon} \quad \text{for all } z \in \left(\frac{\varepsilon}{4}, \varepsilon \right).$$

Consequently, f is Lipschitz with constant $\frac{2}{\varepsilon}$. Then for all $M > 0$ there exists a constant $K_M \geq 0$ depending on M such that for all $t \geq 0$

$$\begin{aligned} \sum_{i=1}^n \mathbb{E} |z_{t \wedge \tau_M}^i - \tilde{z}_{t \wedge \tau_M}^i| &\leq K_M \mathbb{E} \left[\int_0^{t \wedge \tau_M} \sum_{i=1}^n |z_s^i - \tilde{z}_s^i| ds \right] \\ &\leq K_M \int_0^t \sum_{i=1}^n \mathbb{E} |z_{s \wedge \tau_M}^i - \tilde{z}_{s \wedge \tau_M}^i| ds. \end{aligned}$$

The Grönwall Lemma allows to conclude that for all $M > 0$ and $t \geq 0$

$$\sum_{i=1}^n \mathbb{E} |z_{t \wedge \tau_M}^i - \tilde{z}_{t \wedge \tau_M}^i| = 0.$$

Using Fatou's Lemma to take the limit M going to infinity we deduce that for all $t \geq 0$

$$\sum_{i=1}^n \mathbb{E}|z_t^i - \tilde{z}_t^i| = 0,$$

which concludes the proof on the existence and pathwise uniqueness.

Let us now prove (2.4.13). To do so, we can use the same method used to prove (2.4.7) on the square root of the coordinates $(\tilde{\lambda}_t^{2,\varepsilon}, \dots, \tilde{\lambda}_t^{n,\varepsilon})_t$, which solve, as explained in the beginning of the proof, a system of SDEs similar to (A_ε) for $n - 1$ particles. □

2.5 Proof of the other results

Let us now prove the result on the collision time between particles. To do so, we study the difference between two neighbour coordinates and bound it from above by a time changed Bessel process hitting zero in finite time.

Proof of Proposition 2.2.6. For $i \in \{2, \dots, n\}$,

$$\begin{aligned} d(\lambda_t^i - \lambda_t^{i-1}) &= 2\sqrt{\lambda_t^i}dB_t^i - 2\sqrt{\lambda_t^{i-1}}dB_t^{i-1} + 2\gamma(\lambda_t^{i-1} - \lambda_t^i)dt + 2\beta\frac{\lambda_t^i + \lambda_t^{i-1}}{\lambda_t^i - \lambda_t^{i-1}}dt \\ &\quad + 2\beta \sum_{j \notin \{i, i-1\}} \lambda_t^j \frac{\lambda_t^{i-1} - \lambda_t^i}{(\lambda_t^i - \lambda_t^j)(\lambda_t^{i-1} - \lambda_t^j)} dt \\ &\leq 2\sqrt{\lambda_t^i}dB_t^i - 2\sqrt{\lambda_t^{i-1}}dB_t^{i-1} + 2\beta\frac{\lambda_t^i + \lambda_t^{i-1}}{\lambda_t^i - \lambda_t^{i-1}}dt, \end{aligned} \tag{2.5.1}$$

because

$$2\gamma(\lambda_t^{i-1} - \lambda_t^i) \leq 0,$$

and the contribution of the greater and smaller coordinates is non-positive (in the sum, the numerator of the ratio factor is always non-positive and the denominator always non-negative). Let us fix $i \in \{2, \dots, n\}$ and let us define the continuous local martingale

$$M_t^i = 2 \int_0^t \sqrt{\lambda_s^i} dB_s^i - \sqrt{\lambda_s^{i-1}} dB_s^{i-1} \text{ for } t \geq 0,$$

and

$$\langle M^i \rangle_t = 4 \int_0^t (\lambda_s^i + \lambda_s^{i-1}) ds \text{ for } t \geq 0,$$

with generalized inverse

$$C_t^i = \inf\{s \geq 0 : \langle M^i \rangle_s > t\}.$$

The process $(\langle M^i \rangle_t)_t$ is continuous, and according to Lemma 2.5.2 below, $\lim_{t \rightarrow \infty} \langle M^i \rangle_t = +\infty$ a.s., which implies that for all $t \geq 0$, $C_t^i < \infty$. We can apply the Dambins Dubins-Schwarz

Theorem (see for instance [RY99, Theorem 1.6 p.181]) to conclude that the process defined by

$$(B_t^{(i)})_t = \left(2 \int_0^{C_t^i} \sqrt{\lambda_s^i} dB_s^i - \sqrt{\lambda_s^{i-1}} dB_s^{i-1} \right)_t$$

is a $(\mathcal{F}_{C_t^i})_t$ -Brownian motion.

We define for all $t \in \mathbb{R}_+$: $D_t^i = \lambda_{C_t^i}^i - \lambda_{C_t^i}^{i-1}$. By (2.5.1) and the definition of $B^{(i)}$, we have

:

$$\begin{aligned} dD_t^i &= d(\lambda_{C_t^i}^i - \lambda_{C_t^i}^{i-1}) \\ &\leq dB_t^{(i)} + \frac{\beta}{2(\lambda_{C_t^i}^i - \lambda_{C_t^i}^{i-1})} dt = dB_t^{(i)} + \frac{\beta}{2D_t^i} dt. \end{aligned}$$

Let us define the Bessel process

$$r_t^i = \lambda_0^i - \lambda_0^{i-1} + B_t^{(i)} + \frac{\beta}{2} \int_0^t \frac{1}{r_s^i} ds.$$

Then,

$$d(r_t^i - D_t^i) \geq \frac{\beta}{2} \frac{D_t^i - r_t^i}{r_t^i D_t^i} dt,$$

and as long as neither r^i nor D^i touches 0

$$d \left(e^{\frac{\beta}{2} \int_0^t \frac{ds}{r_s^i D_s^i}} (r_t^i - D_t^i) \right) \geq 0,$$

and $r_t^i \geq D_t^i$ thanks to the equality $r_0^i = D_0^i$. As the trajectories are continuous, as soon as r^i reaches 0, which is the case in finite time almost surely when $\beta \leq 1$ (see for instance [RY99, Chapter XI, (ii) p.442]), so does D^i , which concludes the proof. \square

Remark 2.5.1. It turns out that whatever $\gamma \in \mathbb{R}$ and $\alpha, \beta > 0$, we can conclude that $\int_0^{+\infty} (\lambda_s^n + \lambda_s^{n-1}) ds = +\infty$ a.s. as soon as (2.1.1) admits a global solution. Since the coordinates are ordered, for all $t \geq 0$

$$\int_0^t (\lambda_s^n + \lambda_s^{n-1}) ds \geq \frac{2}{n} \int_0^t \left(\sum_{i=1}^n \lambda_s^i \right) ds.$$

Let us define W by $W_0 = 0$ and $dW_t = \sum_{i=1}^n \left(\mathbf{1}_{\{\sum_{j=1}^n \lambda_t^j \neq 0\}} \frac{\sqrt{\lambda_t^i}}{\sqrt{\sum_{j=1}^n \lambda_t^j}} + \mathbf{1}_{\{\sum_{j=1}^n \lambda_t^j = 0\}} \frac{1}{\sqrt{n}} \right) dB_t^i$. According to Lévy's characterization, W is a Brownian motion. Then, using the equality

(2.3.1) in the introduction, $(\sum_{i=1}^n \lambda_s^i)_{s \geq 0}$ is a CIR process. Proposition 6.2.4 in [LL08] gives an expression of the Laplace transform of integrated CIR processes : for any $\mu > 0$,

$$\mathbb{E} \left[e^{-\mu \int_0^t \sum_{i=1}^n \lambda_s^i ds} \right] = \exp(-2\alpha\phi_\mu(t)) \exp \left(- \left(\sum_{i=1}^n \lambda_0^i \right) \psi_\mu(t) \right),$$

where

$$\phi_\mu(t) = -\frac{1}{2} \ln \left(\frac{2\sqrt{\gamma^2 + 2\mu} e^{(\gamma - \sqrt{\gamma^2 + 2\mu})t}}{(\sqrt{\gamma^2 + 2\mu} - \gamma) e^{-2t\sqrt{\gamma^2 + 2\mu}} + \sqrt{\gamma^2 + 2\mu} + \gamma} \right) \xrightarrow{t \rightarrow +\infty} +\infty,$$

$$\text{and } \psi_\mu(t) = \frac{\mu \left(e^{2t\sqrt{\gamma^2 + 2\mu}} - 1 \right)}{\sqrt{\gamma^2 + 2\mu} - \gamma + e^{2t\sqrt{\gamma^2 + 2\mu}}(\sqrt{\gamma^2 + 2\mu} + \gamma)} \xrightarrow{t \rightarrow +\infty} \frac{\mu}{\sqrt{\gamma^2 + 2\mu} + \gamma}.$$

Thus,

$$\mathbb{E} \left[e^{-\mu \int_0^{+\infty} \sum_{i=1}^n \lambda_s^i ds} \right] = \lim_{t \rightarrow +\infty} \mathbb{E} \left[e^{-\mu \int_0^t \sum_{i=1}^n \lambda_s^i ds} \right] = 0,$$

so that

$$\int_0^{+\infty} \sum_{i=1}^n \lambda_s^i ds = +\infty \text{ a.s.}$$

and we can conclude.

Lemma 2.5.2. *Let us assume that $\alpha, \beta > 0$ with $\alpha - (n-1)\beta \geq 0$ and that $\Lambda = (\lambda_t^1, \dots, \lambda_t^n)_t$ is a global solution to (2.1.1). If either $\gamma > 0$ or $\gamma \leq 0$, $\alpha - (n-1)\beta \geq 1 - \beta$ and $\lambda_0^2 > 0$ a.s., then,*

$$\int_0^{+\infty} \lambda_s^1 ds = +\infty \text{ a.s.}$$

Proof. The pathwise uniqueness result in Lemma 2.3.1 ensures that the solution is strong. Therefore it is enough to deal with the case of a deterministic initial condition $\Lambda_0 = (\lambda_0^1, \dots, \lambda_0^n)$. We thus suppose without loss of generality that $\mathbb{E} \sum_{i=1}^n \lambda_0^i < \infty$ still with $\lambda_0^2 > 0$ a.s. in the second setting.

Let us first deal with the case $\gamma > 0$. To do so, let $\tilde{\Lambda}_0$ be distributed according to ρ_{inv} and independent from the Brownian motion \mathbf{B} and let $\tilde{\Lambda} = (\tilde{\lambda}_t^1, \dots, \tilde{\lambda}_t^n)_t$ be the solution to (2.1.1) starting from $\tilde{\Lambda}_0$ given by Proposition 2.2.7. For all $t \geq 0$, $\tilde{\Lambda}_t$ is distributed according to ρ_{inv} . Let us show that

$$\mathbb{P} \left(\int_0^{+\infty} \tilde{\lambda}_s^1 ds = +\infty \right) = 1.$$

Since $(\tilde{\Lambda}_{t+1})_{t \geq 0}$ is a solution to (2.1.1) starting from $\tilde{\Lambda}_1$ distributed according to ρ_{inv} for the Brownian motion $(B_{t+1} - B_1)_{t \geq 0}$ and pathwise uniqueness implies weak uniqueness, $(\tilde{\Lambda}_{t+1})_{t \geq 0}$ has the same distribution as $(\tilde{\Lambda}_t)_{t \geq 0}$. Thus, $\int_0^{+\infty} \tilde{\lambda}_s^1 ds$ has the same distribution as $\int_1^{+\infty} \tilde{\lambda}_s^1 ds$. Consequently, since

$$e^{-\int_0^{+\infty} \tilde{\lambda}_s^1 ds} \leq e^{-\int_1^{+\infty} \tilde{\lambda}_s^1 ds} \text{ a.s.,}$$

$$e^{-\int_0^{+\infty} \tilde{\lambda}_s^1 ds} = e^{-\int_1^{+\infty} \tilde{\lambda}_s^1 ds} \text{ a.s.},$$

which also writes

$$\left(1 - e^{\int_0^1 \tilde{\lambda}_s^1 ds}\right) e^{-\int_0^{+\infty} \tilde{\lambda}_s^1 ds} = 0 \text{ a.s.}$$

As

$$\rho_{inv}(\{x \in \mathbb{R}^n, x^1 > 0\}) = 1,$$

we have

$$\tilde{\lambda}_0^1 > 0 \text{ a.s.}$$

and

$$e^{\int_0^1 \tilde{\lambda}_s^1 ds} > 1 \text{ a.s.},$$

and one can deduce that

$$e^{-\int_0^{+\infty} \tilde{\lambda}_s^1 ds} = 0 \text{ a.s.}$$

so that

$$\int_0^{+\infty} \tilde{\lambda}_s^1 ds = +\infty \text{ a.s.}$$

By (2.3.2), for all $s \geq 0$

$$\mathbb{E}|\lambda_s^1 - \tilde{\lambda}_s^1| \leq \left(\sum_{i=1}^n \mathbb{E}|\lambda_0^i - \tilde{\lambda}_0^i|\right) \exp(-2\gamma s),$$

and then

$$\begin{aligned} \mathbb{E}\left|\int_0^{+\infty} \lambda_s^1 ds - \int_0^{+\infty} \tilde{\lambda}_s^1 ds\right| &\leq \int_0^{+\infty} \mathbb{E}|\lambda_s^1 - \tilde{\lambda}_s^1| ds \\ &\leq \int_0^{+\infty} \sum_{i=1}^n \mathbb{E}|\lambda_0^i - \tilde{\lambda}_0^i| \exp(-2\gamma s) ds. \end{aligned}$$

As $\gamma > 0$, the right-hand side is finite. Since

$$\int_0^{+\infty} \tilde{\lambda}_s^1 ds = +\infty \text{ a.s.},$$

we can deduce that

$$\int_0^{+\infty} \lambda_s^1 ds = +\infty \text{ a.s.}$$

Let us now suppose that $\gamma \leq 0$ and $\alpha - (n-1)\beta \geq 1 - \beta$. Let $\hat{\gamma} > 0$ and $\hat{\Lambda}$ denote the global solution to (2.1.1) with γ replaced by $\hat{\gamma}$ starting from $\hat{\Lambda}_0 = \Lambda_0$ given by Theorem 2.2.2. By the previous reasoning,

$$\int_0^{+\infty} \hat{\lambda}_s^1 ds = +\infty \text{ a.s.}$$

For $M > 0$ let

$$\tau_M = \inf\{t \geq 0 : \lambda_t^n + \hat{\lambda}_t^n \geq M\},$$

with the convention $\inf \emptyset = +\infty$. As Λ and $\hat{\Lambda}$ are continuous global solutions, $\tau_M \uparrow +\infty$ when $M \uparrow \infty$. Reasoning like in the proof of Lemma 2.3.1, we obtain

$$\begin{aligned} \mathbb{E} \left[\sum_{i=1}^n (\hat{\lambda}_{t \wedge \tau_M}^i - \lambda_{t \wedge \tau_M}^i)^+ \right] &= \mathbb{E} \left[\int_0^{t \wedge \tau_M} \sum_{i=1}^n \mathbf{1}_{\{\hat{\lambda}_s^i > \lambda_s^i\}} \left(-2\hat{\gamma} \hat{\lambda}_s^i + 2\gamma \lambda_s^i + \beta \sum_{j \neq i} \left(\frac{\hat{\lambda}_s^i + \hat{\lambda}_s^j}{\hat{\lambda}_s^i - \hat{\lambda}_s^j} - \frac{\lambda_s^i + \lambda_s^j}{\lambda_s^i - \lambda_s^j} \right) \right) ds \right] \\ &\leq \beta \mathbb{E} \left[\int_0^{t \wedge \tau_M} \sum_{i=1}^n \mathbf{1}_{\{\hat{\lambda}_s^i > \lambda_s^i\}} \sum_{j \neq i} \left(\frac{\hat{\lambda}_s^i + \hat{\lambda}_s^j}{\hat{\lambda}_s^i - \hat{\lambda}_s^j} - \frac{\lambda_s^i + \lambda_s^j}{\lambda_s^i - \lambda_s^j} \right) ds \right] \quad (2.5.2) \\ &\leq 0. \end{aligned}$$

The last inequality comes from the fact that, as in (2.3.5), we have for all $i < j$:

$$\left[\frac{\hat{\lambda}_s^i + \hat{\lambda}_s^j}{\hat{\lambda}_s^i - \hat{\lambda}_s^j} - \frac{\lambda_s^i + \lambda_s^j}{\lambda_s^i - \lambda_s^j} \right] (\mathbf{1}_{\{\hat{\lambda}_s^i > \lambda_s^i\}} - \mathbf{1}_{\{\hat{\lambda}_s^j > \lambda_s^j\}}) = -2 \frac{\lambda_s^j |\hat{\lambda}_s^i - \lambda_s^i| + \lambda_s^i |\lambda_s^j - \hat{\lambda}_s^j|}{(\hat{\lambda}_s^i - \hat{\lambda}_s^j)(\lambda_s^i - \lambda_s^j)} |\mathbf{1}_{\{\hat{\lambda}_s^i > \lambda_s^i\}} - \mathbf{1}_{\{\hat{\lambda}_s^j > \lambda_s^j\}}| \leq 0,$$

as the denominator is non-negative.

Using Fatou's Lemma to take the limit M going to infinity, we deduce that for all $t \geq 0$

$$\mathbb{E} \left[\sum_{i=1}^n (\hat{\lambda}_t^i - \lambda_t^i)^+ \right] = 0,$$

and thus, for all $i \in \{1, \dots, n\}$,

$$\hat{\lambda}_t^i < \lambda_t^i \text{ a.s.}$$

Since $\int_0^{+\infty} \hat{\lambda}_s^1 ds = +\infty$ a.s., we conclude that

$$\int_0^{+\infty} \lambda_s^1 ds = +\infty \text{ a.s.}$$

□

Proposition 2.5.3. *Let us assume $\alpha - (n-1)\beta < 0$, $\beta > 0$ and $0 \leq \lambda_0^1 \leq \dots \leq \lambda_0^n$.*

The system (2.1.1) has a unique strong solution defined on the time interval $[0, \lim_{\varepsilon \rightarrow 0} S_\varepsilon)$ where for all $\varepsilon > 0$,

$$S_\varepsilon = \inf\{t \geq 0 : \lambda_t^1 \leq \varepsilon\}.$$

Proof. Let $\varepsilon > 0$ and let us consider the system

$$dx_t^{i,\varepsilon} = dB_t^i + \frac{\alpha - (n-1)\beta - 1}{2} \frac{1}{x_t^{i,\varepsilon} \vee \sqrt{\varepsilon}} dt - \gamma x_t^{i,\varepsilon} dt + \beta x_t^{i,\varepsilon} \sum_{j \neq i} \frac{dt}{(x_t^{i,\varepsilon})^2 - (x_t^{j,\varepsilon})^2} \text{ for } i \in \{1, \dots, n\} \quad (C_\varepsilon)$$

$$0 \leq x_t^{1,\varepsilon} < \dots < x_t^{n,\varepsilon} \text{ a.s. } dt - a.e.$$

with initial condition $0 \leq \sqrt{\lambda_0^1} \leq \dots \leq \sqrt{\lambda_0^n}$.

As the function $x \mapsto \frac{1}{x\sqrt{\varepsilon}}$ is globally Lipschitz, we can apply Ceba's multivalued equations theory ([Cép95]) to conclude that there exists a unique strong solution to (C_ε) defined

globally. Moreover, one can remark that the solution to (C_ε) is a solution to (2.1.5) on $[0, \inf\{t \geq 0 : x_t^1 \leq \sqrt{\varepsilon}\}]$ for all $\varepsilon > 0$. By taking the square of each coordinate, we obtain a solution to (2.1.1) up to S_ε for all $\varepsilon > 0$. By pathwise uniqueness for (2.1.1) (see Lemma 2.3.1), these solutions are consistent and we can conclude. \square

2.6 Appendix

The next lemma deals with the existence and uniqueness to the CIR SDE and with the probability for the solution to hit zero. It is proved for instance in [LL08, Theorem 6.2.2 and Proposition 6.2.3]. The point 4. comes directly from [CIR85].

Lemma 2.6.1. *Let $a \geq 0, b, \sigma \in \mathbb{R}$. Suppose that W is a standard Brownian motion defined on \mathbb{R}_+ . For any real number $x \geq 0$, there is a unique continuous, adapted process X , taking values in \mathbb{R}_+ , satisfying $X_0 = x$ and*

$$dX_t = (a - bX_t)dt + \sigma\sqrt{X_t}dW_t \text{ on } [0, \infty).$$

Moreover, if we denote by X^x the solution to this SDE starting at x and by $\tau_0^x = \inf\{t \geq 0 : X_t^x = 0\}$,

1. If $a \geq \sigma^2/2$, we have $\mathbb{P}(\tau_0^x = \infty) = 1$, for all $x > 0$.
2. If $0 \leq a < \sigma^2/2$ and $b \geq 0$, we have $\mathbb{P}(\tau_0^x < \infty) = 1$, for all $x > 0$.
3. If $0 \leq a < \sigma^2/2$ and $b < 0$, we have $0 < \mathbb{P}(\tau_0^x < \infty) < 1$, for all $x > 0$.
4. For all $s > t$,

$$\mathbb{E}[r_s | r_t] = r_t e^{-b(s-t)} + \frac{a}{b}(1 - e^{-b(s-t)}).$$

The following result is the Ikeda-Watanabe Theorem, which allows to compare two Itô processes if their starting points and their drift coefficients are comparable, and if their diffusion coefficients are regular enough. It is proved for instance in [RW00, Theorem V.43.1 p.269].

Theorem 2.6.2. *(Ikeda-Watanabe) Suppose that, for $i = 1, 2$,*

$$X_t^i = X_0^i + \int_0^t \sigma(X_s^i)dB_s + \int_0^t \beta_s^i ds, \tag{2.6.1}$$

and that there exist $b : \mathbb{R} \mapsto \mathbb{R}$, such that

$$\beta_s^1 \geq b(X_s^1), \quad b(X_s^2) \geq \beta_s^2.$$

Suppose also that

1. σ is measurable and there exists an increasing function $\rho : \mathbb{R}_+ \mapsto \mathbb{R}_+$ such that

$$\int_{0^+} \rho(u)^{-1} du = \infty,$$

and for all $x, y \in \mathbb{R}$,

$$(\sigma(x) - \sigma(y))^2 \leq \rho(|x - y|);$$

2. $X_0^1 \geq X_0^2$ a.s.;

3. b is Lipschitz.

Then $X_t^1 \geq X_t^2$ for all t a.s.

Chapter 3

Strong solutions to a beta-Jacobi particle system

This Section an article submitted to *Séminaire de Probabilités* that one can find at the address <https://arxiv.org/abs/2104.10142>.

Abstract

The purpose of this paper is to study the existence and uniqueness of solutions to a system of Stochastic Differential Equations (SDEs). The coordinates are bounded by zero and one, and repulse each other according to a Coulombian like interaction force. We show the existence of strong and pathwise unique solutions to the system until the first multiple collision at zero or one, and give a sufficient condition on the parameters of the SDEs for this multiple collision not to occur in finite time.

Keywords: stochastic differential equations, diffusions with gradient drift, singular interaction, random matrices.

AMS Subject Classification (2020): 60H10, 60J60, 60B20, 60G17, 60J70.

3.1 Introduction

Let $p, q \geq 0, \beta > 0, n \geq 1$, and $\mathbf{B} = (B_t^1, \dots, B_t^n)_t$ be a n -dimensional Brownian motion. Our SDEs system of interest is the following :

$$\begin{aligned} d\lambda_t^i &= 2\sqrt{\lambda_t^i(1-\lambda_t^i)}dB_t^i + \beta \left[p - (p+q)\lambda_t^i + \sum_{j \neq i} \frac{\lambda_t^i(1-\lambda_t^j) + \lambda_t^j(1-\lambda_t^i)}{\lambda_t^i - \lambda_t^j} \right] dt & (\text{J}(p,q)) \\ &= 2\sqrt{\lambda_t^i(1-\lambda_t^i)}dB_t^i + \beta \left[p - n + 1 - (p+q)\lambda_t^i + 2\lambda_t^i \sum_{j \neq i} \frac{1-\lambda_t^j}{\lambda_t^i - \lambda_t^j} \right] dt \text{ for all } i \in \{1, \dots, n\} \end{aligned} \quad (3.1.1)$$

$$0 \leq \lambda_t^1 < \dots < \lambda_t^n \leq 1, \text{ a.s. } dt - \text{almost everywhere.} \quad (3.1.2)$$

The form (3.1.1) can be found thanks to the manipulation

$$\begin{aligned} \sum_{j \neq i} \frac{\lambda_t^i(1 - \lambda_t^j) + \lambda_t^j(1 - \lambda_t^i)}{\lambda_t^i - \lambda_t^j} &= \sum_{j \neq i} \frac{2\lambda_t^i(1 - \lambda_t^j) + \lambda_t^j - \lambda_t^i}{\lambda_t^i - \lambda_t^j} \\ &= -(n-1) + 2\lambda_t^i \sum_{j \neq i} \frac{1 - \lambda_t^j}{\lambda_t^i - \lambda_t^j}. \end{aligned} \quad (3.1.3)$$

The strict inequalities (3.1.2) allow the interaction terms in the system (J(p,q)) to make sense. We will look for continuous solutions to the SDEs (J(p,q)). Thus, by continuity, we have for all $t \geq 0$, $0 \leq \lambda_t^1 \leq \dots \leq \lambda_t^n \leq 1$, a.s. While, for $n = 1$, the system reduces to the real Jacobi SDE $d\lambda_t^1 = 2\sqrt{\lambda_t^1(1 - \lambda_t^1)}dB_t^1 + \beta[p - (p+q)\lambda_t^1]dt$ studied in [Dou05] (see Lemma 3.5.1 in the Appendix), the coordinates are repulsed by a Coulombian like interaction when $n \geq 2$.

Our goal is to study the existence and uniqueness of strong solutions to this system of SDEs in a general setting for the parameters p, q , and β , especially in the case $\beta < 1$ and $p \wedge q - n + 1 < 1/\beta$ which is not covered to our knowledge by the literature. This system is related to the famous Dyson Brownian motions which satisfy

$$d\phi_t^i = \sqrt{2}dB_t^i + \beta \sum_{j \neq i} \frac{dt}{\phi_t^i - \phi_t^j}, \text{ for all } i \in \{1, \dots, n\},$$

where the only type of singularity is when two particles collide. This system admits a unique strong solution (see [RS93]) when $\beta \geq 1$.

The difficulty in proving the existence of solutions to (J(p,q)) comes from the fact that there are singularities both when a particle touches zero or one, where the derivative of the square root diffusion coefficient explodes, and when two particles touch each other. If we define $D = \{0 < \lambda^1 < \lambda^2 < \dots < \lambda^n < 1\}$, a collision occurs when the process $\Lambda = (\lambda_t^1, \dots, \lambda_t^n)_t$ hits the boundary ∂D made of the union of $\{\lambda^i = \lambda^{i+1}\}$ for $i \in \{1, \dots, n-1\}$ and $\{\lambda^1 = 0\}$ and $\{\lambda^n = 1\}$. A multiple collision occurs when two of these sets are reached at the same time and we will speak about "collision between particles" when two particles touch each other.

The same kind of difficulties emerge in the study of the beta-Wishart system of SDEs (see [JK20]) which writes

$$\begin{aligned} d\lambda_t^i &= 2\sqrt{\lambda_t^i}dB_t^i + \left(\alpha - 2\gamma\lambda_t^i + \beta \sum_{j \neq i} \frac{\lambda_t^i + \lambda_t^j}{\lambda_t^i - \lambda_t^j} \right) dt \text{ for all } i \in \{1, \dots, n\}, \\ 0 &\leq \lambda_t^1 < \dots < \lambda_t^n, \text{ a.s. } dt - \text{a.e.}, \end{aligned}$$

where $\alpha \geq 0, \gamma \in \mathbb{R}$ and $\beta > 0$. Indeed, we can find in this system the same kind of singularities both when a particle touches zero and when two particles collide. The methods developed in [JK20] to overcome the two types of singularities and to prove the existence and uniqueness of a strong solution to the system of SDEs are adapted to the beta-Jacobi system here.

Our results about the SDEs $(J(p,q))$ are the following. In Proposition 3.2.5, we state that $k\beta(p \wedge q - n + k) \geq 2$ is a sufficient condition for multiple collisions between k particles not to occur at position zero or one in finite time. Our main result Theorem 3.2.3 gives the existence and uniqueness of solutions to the SDEs $(J(p,q))$ when $\beta \in (0, 1)$ and $p \wedge q - n + 1 \geq \frac{1}{\beta} - 1$ (so that there is no multiple collision at the origin or at position one). We explicit in Proposition 3.2.6 the unique stationary probability measure of the SDEs $(J(p,q))$.

The paper is organized as follows. The remaining of the introduction is devoted to the bibliographical background of this work. In Section 2, we state our main results. We prove in Section 3 some useful properties of the solutions to the system $(J(p,q))$, before checking Proposition 3.2.5 and Theorem 3.2.3 in Section 4. Section 5 is an Appendix stating some well-known results that we use in our proofs.

The beta-Jacobi process emerges from the matrix Jacobi process introduced and studied in [Dou05] : let $n, m, p > 0$ be three integers, Θ a $m \times m$ orthogonal Brownian motion (see [Lév17] for a construction of the orthogonal Brownian motion), and let X be the matrix made of the n first lines and the p first columns of Θ . Then, if $p \geq n + 1$, $q \geq n + 1$ and $p + q = m$, the process $J = XX^*$, where $*$ is the adjoint operator, is a diffusion and a solution to the SDE

$$dJ = \sqrt{J}dB\sqrt{I_n - J} + \sqrt{I_n - J}dB^*\sqrt{J} + (pI - (p + q)J)dt,$$

where B is a $n \times n$ matrix filled with Brownian motions (see [Dou05, Theorem 9.2.3]). The eigenvalues of J then verify the system of SDEs $(J(p,q))$ with $\beta = 1$ (see [Dou05, Theorem 9.3.1]). When the orthogonal Brownian motion Θ is replaced by a unitary Brownian motion, the eigenvalues of J then verify the system of SDEs $(J(p,q))$ with $\beta = 2$ (see [Dem09a]). The system $(J(p,q))$ is thereby a generalization of this matrix inherited system of SDEs for general $p, q \geq 0$ and $\beta > 0$.

Let us consider the change of variable $\phi_t^i = \arcsin(\sqrt{\lambda_t^i})$ and set $\Phi = (\phi_t^1, \dots, \phi_t^n)_t$. We apply Itô's formula, formally after the stopping time $\inf\{s \geq 0 : \lambda_s^1(1 - \lambda_s^n) = 0\}$ since the square root function is not twice differentiable at zero, and obtain, after trigonometric computations that the reader can find in [Dem09a] :

$$d\phi_t^i = dB_t^i + \left\{ \frac{\beta(p - q)}{2} \cot \phi_t^i + (\beta(q - n + 1) - 1) \cot(2\phi_t^i) + \frac{\beta}{2} \sum_{j \neq i} [\cot(\phi_t^i + \phi_t^j) + \cot(\phi_t^i - \phi_t^j)] \right\} dt \quad (3.1.4)$$

$$0 \leq \phi_t^1 < \dots < \phi_t^n \leq \frac{\pi}{2}, \text{ a.s. } dt - \text{a.e.},$$

which can be rewritten

$$d\phi_t^i = dB_t^i - \partial_i V(\phi_t^1, \dots, \phi_t^n) dt \text{ for all } i \in \{1, \dots, n\}, \quad (3.1.5)$$

with

$$V(\phi^1, \dots, \phi^n) = - \sum_{i=1}^n \left\{ \frac{\beta(p-q)}{2} \ln |\sin \phi^i| + \frac{\beta(q-n+1)-1}{2} \ln |\sin(2\phi_t^i)| \right. \\ \left. + \frac{\beta}{4} \sum_{j \neq i} (\ln |\sin(\phi^i + \phi^j)| + \ln |\sin(\phi^i - \phi^j)|) \right\}. \quad (3.1.6)$$

Systems of interacting particles following equations of the type

$$d\phi_t^i = b_i(\phi_t^i)dt + \sigma_i(\phi_t^i)dB_t^i - \partial_i V(\phi_t^1, \dots, \phi_t^n)dt \text{ for all } i \in \{1, \dots, n\}, \quad (3.1.7)$$

where $V : \mathbb{R}^n \mapsto (-\infty, +\infty]$ is a lower semi continuous convex function such that $D = \{x \in \mathbb{R}^n : V(x) < +\infty\}$ is a non-empty convex set and V is continuously differentiable in D have been studied by many authors. For instance, in the peculiar case $\sigma_i(\phi^i) = \sigma > 0$, $b_i(\phi^i) = 0$ and $V(\phi^1, \dots, \phi^n) = -\frac{\beta}{2} \sum_{i=1}^n \sum_{j \neq i} \ln |\phi^i - \phi^j| + \frac{\theta}{2} \sum_{i=1}^n (\phi^i)^2$ with $\beta \geq \frac{\sigma^2}{2}$ and $\theta > 0$, the existence and uniqueness of a strong solution to (3.1.7) were derived in [RS93].

Link with the multivalued stochastic differential equations theory. The systems of type (3.1.7) with b_i and σ_i Lipschitz and $(\phi_0^1, \dots, \phi_0^n) \in \bar{D}$ were deeply studied by Cépa and Lépingle for instance in [CL97] and [CL01] where they apply Cépa's multivalued stochastic differential equations theory developed in [Cép95]. This theory treats the existence and uniqueness of solutions to multivalued SDEs associated with a convex function defined on a domain of \mathbb{R}^n . In [Lé10], Lépingle applied this theory to a constrained Brownian motion between reflecting or repellent walls of Weyl chambers. The boundary behavior of the convex function dictates the behavior of the process on these same boundaries (hitting or not the boundary in finite time, reflection on the boundary, etc.). Our SDEs of interest rewritten in the form (3.1.5) thanks to the square root change of variables can be seen this way, and we will exploit this connection in the paper.

Link with radial Dunkl processes (see [Dem09b] for a more complete description of the theory)

Let us define a reduced root system R by a finite set in $\mathbb{R}^n \setminus \{0\}$ spanning \mathbb{R}^n such that

- for all $\alpha \in R$, $R \cap \mathbb{R}\alpha = \{\alpha, -\alpha\}$,
- for all $\alpha \in R$, $\sigma_\alpha(R) = R$,

where σ_α is the reflection with respect to the hyperplane orthogonal to α . A simple system Δ is a basis of \mathbb{R}^n which induces a total ordering in R the following way : a root $\alpha \in R$ is positive if it is a positive linear combination of elements of Δ . A simple system Δ being fixed, we can thus define R_+ as the set of positive roots of R . When $\sigma_i = 1$ and V takes the form

$$V : \phi \mapsto - \sum_{\alpha \in R_+} k(\alpha) \ln(\langle \alpha, \phi \rangle), \quad x \in D,$$

where D is the positive Weyl chamber defined by

$$D = \{\phi \in \mathbb{R}^n, \langle \alpha, \phi \rangle > 0 \forall \alpha \in R_+\},$$

Demni proved in [Dem09b, Theorem 1] and [Dem09a] the existence and uniqueness of a solution to (3.1.7) on the domain \bar{D} when $k(\alpha) > 0$ for all $\alpha \in R_+$. To do so, he applied Cépa's multivalued stochastic differential equations theory. This system corresponds to $(J(p,q))$ when we make the following modification in V : the convex function $\phi \mapsto -\ln(\langle \alpha, \phi \rangle)$ should be substituted by $\phi \mapsto -\ln(\sin(\langle \alpha, \phi \rangle))$, and for some choice of R_+ and k . Indeed, when the root system is of so-called BC_n -type, it is defined by

$$\begin{aligned} R &= \{\pm e_i, \pm 2e_i, \pm e_i \pm e_j, 1 \leq i < j \leq n\}, \\ \Delta &= \{e_{i+1} - e_i, 1 \leq i \leq n-1, e_1\}, \\ R_+ &= \{e_i, 2e_i, 1 \leq i \leq n, e_j \pm e_i, 1 \leq i < j \leq n\}, \\ D &= \left\{ \phi \in \left[0, \frac{\pi}{2}\right]^n, 0 < \phi^1 < \dots < \phi^n < \frac{\pi}{2} \right\}, \end{aligned}$$

which, with the right choice of k gives equation (3.1.5) (see (3.1.6)). The condition $k(\alpha) > 0$ for all $\alpha \in R_+$ implies $p > q$ and $q - n + 1 > \frac{1}{\beta}$. As a corollary, Demni deduces the existence and uniqueness of a strong solution to $(J(p,q))$ under the condition $p \wedge q - n + 1 > \frac{1}{\beta}$. We seek here to obtain the existence of a solution to $(J(p,q))$ while relaxing this inequality. Demni moreover proved in [Dem09a] that for $\beta < 1$, there is collision between any two neighbour particles in finite time almost surely.

Link with other works. The reader will find in Graczyk and Malecki [GM13] and [GM14] a treatment of equations of the form

$$\begin{aligned} d\lambda_t^i &= \sigma_i(\lambda_t^i)dB_t^i + \left(b_i(\lambda_t^i) + \sum_{j \neq i} \frac{H_{i,j}(\lambda_t^i, \lambda_t^j)}{\lambda_t^i - \lambda_t^j} \right) dt, \text{ for all } i \in \{1, \dots, n\} \\ \lambda_t^1 &\leq \dots \leq \lambda_t^n, \quad t \geq 0, \end{aligned}$$

where the functions σ_i, b_i and $H_{i,j}$ are assumed continuous, with $H_{i,j}$ non-negative and symmetric in the sense that $H_{i,j}(x, y) = H_{j,i}(y, x)$ for all $x, y \in \mathbb{R}$. The system $(J(p,q))$ is recovered in the particular case when $H_{i,j}(\lambda^i, \lambda^j) = \beta(\lambda^i(1+\lambda^j) + \lambda^j(1+\lambda^i))$, $\sigma_i(\lambda^i) = 2\sqrt{\lambda^i(1-\lambda^i)}$ and $b_i(\lambda^i) = \beta(p - (p+q)\lambda^i)$. According to [GM14, Section 6.5], $(J(p,q))$ admits a strong solution on the time interval $[0, +\infty)$ when $\beta \geq 1$. In this regime, the authors proved that there is no collision between the particles. They also demonstrated the pathwise uniqueness of the solutions for every $\beta > 0$, as recalled in Lemma 3.3.1 below.

In all these references, β is identified as a fundamental parameter, its position relative to 1 governing the possibility of collisions between the particles.

Acknowledgement : I thank Benjamin Jourdain and Djalil Chafaï for numerous fruitful discussions.

3.2 Results

Let us begin by the fundamental following remark.

Remark 3.2.1. If $(\lambda_t^1, \dots, \lambda_t^n)_t$ is solution to $(J(p, q))$, then $(1 - \lambda_t^n, \dots, 1 - \lambda_t^1)_t$ is solution to $J(q, p)$ for the Brownian motion $-\mathbf{B}$.

The form (3.1.1) of the SDEs combined with Remark 3.2.1 hint that $\beta(p \wedge q - n + 1)$ is a fundamental parameter impacting the existence of solutions. Consequently, we will study the range of values of this coefficient for which the system has a solution. For instance, if we assume $p - n + 1 < 0$, we have:

$$\begin{aligned} d\lambda_t^1 &= 2\sqrt{\lambda_t^1(1 - \lambda_t^1)}dB_t^1 + \beta \left[p - n + 1 - (p + q)\lambda_t^1 + 2\lambda_t^1 \sum_{j>1} \frac{1 - \lambda_t^j}{\lambda_t^1 - \lambda_t^j} \right] dt \\ &\leq 2\sqrt{\lambda_t^1(1 - \lambda_t^1)}dB_t^1. \end{aligned} \quad (3.2.1)$$

Then, according to the pathwise comparison theorem of Ikeda and Watanabe (that we recall in Theorem 3.5.3 below) :

$$\lambda_t^1 \leq r_t \text{ a.s. for all } t \geq 0,$$

where

$$r_t = \lambda_0^1 + 2 \int_0^t \sqrt{r_s(1 - r_s)}dB_s^1 \text{ for all } t \geq 0,$$

which is a real Jacobi process. By Lemma 3.5.1, the stopping time $T = \inf\{t \geq 0 : r_t = 0\}$ verifies $\mathbb{P}(T < \infty) = 1$. On $\{T < \infty\}$, after T , r stays at zero indefinitely. As the drift in (3.2.1) is negative when $\lambda_t^1 = 0$ and therefore when $r_t = 0$, it will stay strictly negative on a time interval of positive measure. Consequently, the system of SDEs has no global solution. If we assume $q - n + 1 < 0$ instead of $p - n + 1 < 0$, the same reasoning can be applied replacing $(\lambda_t^1, \dots, \lambda_t^n)$ by $(1 - \lambda_t^n, \dots, 1 - \lambda_t^1)$ and interverting p and q in (3.2.1).

We thus proved the following result :

Remark 3.2.2. A necessary condition for the existence of a global solution to $(J(p, q))$ is $p \wedge q - n + 1 > 0$.

We will thus assume this condition in the remaining of the paper.

It is proved in [GM13, Corollary 9] with condition $0 \leq \lambda_0^1 < \dots < \lambda_0^n \leq 1$ a.s. that for $\beta \geq 1$, the SDEs $(J(p, q))$ has a unique global strong solution and that there is no collision between the particles.

Before relaxing the condition $p > q$, Demni proved in [Dem09a] by applying Cépa's multivalued equations theory (see [Cép95]) that under the conditions $p > q$, $p \wedge q - n + 1 > 1/\beta$, $\beta > 0$ and $0 \leq \lambda_0^1 \leq \dots \leq \lambda_0^n \leq 1$, the system of SDEs $(J(p, q))$ admits a unique strong solution. The conditions $p > q$ and $p \wedge q - n + 1 > 1/\beta$ ensure the convexity of the potential V defined in (3.1.6) which is needed in Cépa's multivalued equations theory. In Theorem 3.2.3, we tackle the existence and uniqueness problem on wider ranges for the parameter $p \wedge q - n + 1$.

Our following result applies in the case $p \wedge q - n + 1 \leq 1/\beta$.

Theorem 3.2.3. *Let us assume $\beta < 1$, $p \wedge q - n + 1 > 0$. Let the initial condition $\Lambda_0 = (\lambda_0^1, \dots, \lambda_0^n)$ be independent from the Brownian motion \mathbf{B} and such that $0 \leq \lambda_0^1 \leq \dots \leq \lambda_0^n \leq 1$ a.s. and $\lambda_0^2(1 - \lambda_0^{n-1}) > 0$ a.s.*

Then, the system of SDEs $(J(p, q))$ has a unique strong solution defined on the time interval

$$[0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon),$$

where, for $\varepsilon > 0$,

$$\zeta_\varepsilon = \inf\{t \geq 0 : \lambda_t^1 \leq \varepsilon \text{ and } \lambda_t^2 - \lambda_t^1 \leq \varepsilon\} \wedge \inf\{t \geq 0 : \lambda_t^n \geq 1 - \varepsilon \text{ and } \lambda_t^n - \lambda_t^{n-1} \leq \varepsilon\}.$$

Moreover,

(i) if $p \wedge q - n + 1 \geq \frac{1}{\beta} - 1$ then $\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon = \infty$ a.s.

(ii) There is no double collision between the particles, i.e. (3.2.2)

$$\mathbb{P}\left\{\exists t \in (0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon) : \lambda_t^i = \lambda_t^{i+1} \text{ and } \lambda_t^j = \lambda_t^{j+1} \text{ for some } 1 \leq i < j \leq n-1\right\} = 0.$$

Theorem 3.2.3 is proved in Section 3.4.

Remark 3.2.4. This result states the existence of a unique strong solution $(\lambda_t^1, \dots, \lambda_t^n)_t$ to $(J(p, q))$ defined on the time-interval \mathbb{R}_+ for $p \wedge q - n + 1 \geq \frac{1}{\beta} - 1$ and until $\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon$ which might be finite if $p \wedge q - n + 1 < \frac{1}{\beta} - 1$. When $\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon < +\infty$, then, according to the first step of the proof of assertion (i) of Proposition 3.2.5 below, the solution can be continuously extended to the closed-time-interval $[0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon]$. The next step would be to find how to start back from $(\lambda_{\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon}^1, \dots, \lambda_{\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon}^n)$ to define a solution on the whole interval \mathbb{R}_+ .

The assertion (i) of Theorem 3.2.3 comes from the application of the following proposition for $k = 2$.

Proposition 3.2.5 (Multiple collision at zero). *Let $k \in \{1, \dots, n\}$. Let the initial condition $\Lambda_0 = (\lambda_0^1, \dots, \lambda_0^n)$ be independent from the Brownian motion \mathbf{B} and such that $0 \leq \lambda_0^1 \leq \dots \leq \lambda_0^n \leq 1$ a.s. Then,*

(i) *if $\lambda_0^k > 0$ a.s., if $(J(p, q))$ has a local solution $\Lambda = (\lambda_t^1, \dots, \lambda_t^n)_t$ defined up to a stopping time \mathcal{T} and if $k\beta(p - n + k) \geq 2$, then there is no collision of k particles at zero. More precisely,*

$$\mathbb{P}(\mathcal{T} = +\infty, \exists t \geq 0 : \lambda_t^1 + \lambda_t^2 + \dots + \lambda_t^k = 0) + \mathbb{P}(\mathcal{T} < +\infty, \inf_{t \in [0, \mathcal{T})} \lambda_t^1 + \lambda_t^2 + \dots + \lambda_t^k = 0) = 0.$$

(ii) *if $\lambda_0^{n-k+1} < 1$ a.s., if $(J(p, q))$ has a local solution $\Lambda = (\lambda_t^1, \dots, \lambda_t^n)_t$ defined up to a stopping time \mathcal{T} and if $k\beta(q - n + k) \geq 2$, then there is no collision of k particles at 1. More precisely,*

$$\mathbb{P}(\mathcal{T} = +\infty, \exists t \geq 0 : \lambda_t^{n-k+1} + \lambda_t^{n-k+2} + \dots + \lambda_t^n = k) + \mathbb{P}(\mathcal{T} < +\infty, \sup_{t \in [0, \mathcal{T})} \lambda_t^{n-k+1} + \lambda_t^{n-k+2} + \dots + \lambda_t^n = k) = 0.$$

This proposition is proved in the beginning of Section 3.4. On top of the existence and uniqueness result, we exhibit in the next proposition the stationary probability measure of the system of SDEs $(J(p,q))$.

Proposition 3.2.6. *Let us assume $\beta(p \wedge q - n + 1) > 0$. The unique stationary probability measure of the system of SDEs $(J(p,q))$ is ρ_{inv} with density with respect to the Lebesgue measure*

$$d\rho_{inv}(\lambda^1, \dots, \lambda^n) = \frac{1}{\mathcal{Z}} \times \prod_{i=1}^n \left((\lambda^i)^{\beta \frac{p-n+1}{2}-1} (1-\lambda^i)^{\beta \frac{q-n+1}{2}-1} \prod_{j \neq i} |\lambda^j - \lambda^i|^{\beta/2} \right) \mathbf{1}_{\{0 \leq \lambda^1 \leq \dots \leq \lambda^n \leq 1\}} d\lambda^1 \dots d\lambda^n,$$

where \mathcal{Z} is a normalizing constant.

More precisely, if a solution to $(J(p,q))$ is such that the distribution of Λ_t does not depend on t , then this distribution is ρ_{inv} . When Λ_0 is distributed according to ρ_{inv} and is independent from the Brownian motion \mathbf{B} , the unique solution to $(J(p,q))$ is such that for all $t \in \mathbb{R}_+$, Λ_t is distributed according to ρ_{inv} , and it is a strong solution.

Remark 3.2.7. The density of ρ_{inv} is the density of the beta-Jacobi ensembles, about which the reader will find in [For10] a documentation. These ensembles are related to the distribution of the eigenvalues of a random matrix model involving a $p \times n$ matrix and a $q \times n$ matrix with the constraints $n \geq p + q$, and $\beta = 1, 2$ or 4 depending on the dimension of the underlying algebra (\mathbb{R}, \mathbb{C} or \mathbb{H}).

3.3 Properties of the solutions

The pathwise uniqueness part of the next Lemma is proved in [GM14, Theorem 5.3] but we reproduce the proof for the sake of completeness.

Lemma 3.3.1. *The solutions to $(J(p,q))$ are pathwise unique.*

Moreover, if $Z = (z_t^1, \dots, z_t^n)_t$ and $\tilde{Z} = (\tilde{z}_t^1, \dots, \tilde{z}_t^n)_t$ are two global in time solutions to $(J(p,q))$ with the same driving Brownian motion, then for all $t \geq 0$,

$$\sum_{i=1}^n \mathbb{E} |z_t^i - \tilde{z}_t^i| \leq \left(\sum_{i=1}^n \mathbb{E} |z_0^i - \tilde{z}_0^i| \right) \exp(-\beta(p+q)t). \quad (3.3.1)$$

Proof of Lemma 3.3.1.

Let $Z = (z_t^1, \dots, z_t^n)_t$ and $\tilde{Z} = (\tilde{z}_t^1, \dots, \tilde{z}_t^n)_t$ be two solutions to $(J(p,q))$ with the same driving Brownian motion. As for all $i \in \{1, \dots, n\}$, $0 < \varepsilon < 1$ and $t \geq 0$:

$$\begin{aligned} \int_0^t \mathbf{1}_{\{0 < z_s^i - \tilde{z}_s^i \leq \varepsilon\}} \frac{d\langle z^i - \tilde{z}^i \rangle_s}{z_s^i - \tilde{z}_s^i} ds &= 4 \int_0^t \mathbf{1}_{\{0 < z_s^i - \tilde{z}_s^i \leq \varepsilon\}} \frac{\left(\sqrt{z_s^i(1-z_s^i)} - \sqrt{\tilde{z}_s^i(1-\tilde{z}_s^i)} \right)^2}{z_s^i - \tilde{z}_s^i} ds \\ &= 4 \int_0^t \mathbf{1}_{\{0 < z_s^i - \tilde{z}_s^i \leq \varepsilon\}} \frac{\left(\sqrt{z_s^i(1-z_s^i)} - \sqrt{\tilde{z}_s^i(1-\tilde{z}_s^i)} \right)^2 \left(\sqrt{z_s^i(1-z_s^i)} + \sqrt{\tilde{z}_s^i(1-\tilde{z}_s^i)} \right)}{(z_s^i - \tilde{z}_s^i) \left(\sqrt{z_s^i(1-z_s^i)} + \sqrt{\tilde{z}_s^i(1-\tilde{z}_s^i)} \right)} ds \end{aligned}$$

$$\begin{aligned}
&= 4 \int_0^t \mathbf{1}_{\{0 < z_s^i - \tilde{z}_s^i \leq \varepsilon\}} \left| \frac{(z_s^i(1 - z_s^i) - \tilde{z}_s^i(1 - \tilde{z}_s^i)) (\sqrt{z_s^i(1 - z_s^i)} - \sqrt{\tilde{z}_s^i(1 - \tilde{z}_s^i)})}{(z_s^i - \tilde{z}_s^i) (\sqrt{z_s^i(1 - z_s^i)} + \sqrt{\tilde{z}_s^i(1 - \tilde{z}_s^i)})} \right| ds \\
&= 4 \int_0^t \mathbf{1}_{\{0 < z_s^i - \tilde{z}_s^i \leq \varepsilon\}} \left| \frac{(1 - z_s^i - \tilde{z}_s^i) (\sqrt{z_s^i(1 - z_s^i)} - \sqrt{\tilde{z}_s^i(1 - \tilde{z}_s^i)})}{(\sqrt{z_s^i(1 - z_s^i)} + \sqrt{\tilde{z}_s^i(1 - \tilde{z}_s^i)})} \right| ds \\
&\leq 4 \int_0^t |1 - z_s^i - \tilde{z}_s^i| ds < \infty,
\end{aligned}$$

the local time of $z^i - \tilde{z}^i$ at 0 is zero ([RY99, Lemma 3.3 p.389]).

Applying the Tanaka formula then the integration by parts formula to compute $de^{\beta(p+q)t} |z_t^i - \tilde{z}_t^i|$ we get

$$\begin{aligned}
de^{\beta(p+q)t} \sum_{i=1}^n |z_t^i - \tilde{z}_t^i| &= 2 \sum_{i=1}^n e^{\beta(p+q)t} \operatorname{sgn}(z_t^i - \tilde{z}_t^i) \left(\sqrt{z_t^i(1 - z_t^i)} - \sqrt{\tilde{z}_t^i(1 - \tilde{z}_t^i)} \right) dB_t^i \\
&\quad + \beta e^{\beta(p+q)t} \sum_{i=1}^n \operatorname{sgn}(z_t^i - \tilde{z}_t^i) \sum_{j \neq i} \left(\frac{z_t^i(1 - z_t^j) + z_t^j(1 - z_t^i)}{z_t^i - z_t^j} - \frac{\tilde{z}_t^i(1 - \tilde{z}_t^j) + \tilde{z}_t^j(1 - \tilde{z}_t^i)}{\tilde{z}_t^i - \tilde{z}_t^j} \right) ds \\
&\leq 2 \sum_{i=1}^n e^{\beta(p+q)t} \operatorname{sgn}(z_t^i - \tilde{z}_t^i) \left(\sqrt{z_t^i(1 - z_t^i)} - \sqrt{\tilde{z}_t^i(1 - \tilde{z}_t^i)} \right) dB_t^i, \quad (3.3.2)
\end{aligned}$$

where $\operatorname{sgn}(x) = 1$ if $x > 0$ and $\operatorname{sgn}(x) = -1$ if $x \leq 0$. The inequality (3.3.2) comes from the fact that we have for all $i < j$:

$$\begin{aligned}
&\left[\frac{z_s^i(1 - z_s^j) + z_s^j(1 - z_s^i)}{z_s^i - z_s^j} - \frac{\tilde{z}_s^i(1 - \tilde{z}_s^j) + \tilde{z}_s^j(1 - \tilde{z}_s^i)}{\tilde{z}_s^i - \tilde{z}_s^j} \right] (\operatorname{sgn}(z_s^i - \tilde{z}_s^i) - \operatorname{sgn}(z_s^j - \tilde{z}_s^j)) \\
&= 2 \left[\frac{z_s^j \tilde{z}_s^i - z_s^i \tilde{z}_s^j}{(z_s^i - z_s^j)(\tilde{z}_s^i - \tilde{z}_s^j)} + \frac{-z_s^i z_s^j \tilde{z}_s^i + z_s^i z_s^j \tilde{z}_s^j + z_s^i \tilde{z}_s^j \tilde{z}_s^i - z_s^j \tilde{z}_s^j \tilde{z}_s^i}{(z_s^i - z_s^j)(\tilde{z}_s^i - \tilde{z}_s^j)} \right] (\operatorname{sgn}(z_s^i - \tilde{z}_s^i) - \operatorname{sgn}(z_s^j - \tilde{z}_s^j)) \\
&= 2 \left[\frac{z_s^j(\tilde{z}_s^i - z_s^i) + z_s^i(z_s^j - \tilde{z}_s^j)}{(z_s^i - z_s^j)(\tilde{z}_s^i - \tilde{z}_s^j)} + \frac{z_s^i \tilde{z}_s^i(\tilde{z}_s^j - z_s^j) + z_s^j \tilde{z}_s^j(z_s^i - \tilde{z}_s^i)}{(z_s^i - z_s^j)(\tilde{z}_s^i - \tilde{z}_s^j)} \right] (\operatorname{sgn}(z_s^i - \tilde{z}_s^i) - \operatorname{sgn}(z_s^j - \tilde{z}_s^j)) \\
&= 2 \frac{z_s^j(1 - \tilde{z}_s^j)(\tilde{z}_s^i - z_s^i) + z_s^i(1 - \tilde{z}_s^i)(z_s^j - \tilde{z}_s^j)}{(z_s^i - z_s^j)(\tilde{z}_s^i - \tilde{z}_s^j)} (\operatorname{sgn}(z_s^i - \tilde{z}_s^i) - \operatorname{sgn}(z_s^j - \tilde{z}_s^j)) \\
&= -2 \frac{z_s^j(1 - \tilde{z}_s^j)|\tilde{z}_s^i - z_s^i| + z_s^i(1 - \tilde{z}_s^i)|z_s^j - \tilde{z}_s^j|}{(z_s^i - z_s^j)(\tilde{z}_s^i - \tilde{z}_s^j)} |\operatorname{sgn}(z_s^i - \tilde{z}_s^i) - \operatorname{sgn}(z_s^j - \tilde{z}_s^j)| \leq 0, \quad (3.3.3)
\end{aligned}$$

as the denominator is non-negative. Let the two solutions Z and \tilde{Z} be respectively defined on $[0, T]$ if $T < +\infty$ and $[0, T)$ if $T = +\infty$, and $[0, \tilde{T}]$ if $\tilde{T} < +\infty$ and $[0, \tilde{T})$ if $\tilde{T} = +\infty$, where T and \tilde{T} are stopping times for a filtration $(\mathcal{F}_t)_{t \geq 0}$ with respect to which \mathbf{B} is a Brownian motion and Z_0 is \mathcal{F}_0 -measurable.

As for all $i \in \{1, \dots, n\}$:

$$\mathbb{E} \int_0^t e^{2\beta(p+q)s} \left| \sqrt{z_s^i(1 - z_s^i)} - \sqrt{\tilde{z}_s^i(1 - \tilde{z}_s^i)} \right|^2 ds < \infty,$$

the stochastic integrals $\int_0^t e^{\beta(p+q)s} \text{sgn}(z_s^i - \tilde{z}_s^i) \left(\sqrt{z_s^i(1-z_s^i)} - \sqrt{\tilde{z}_s^i(1-\tilde{z}_s^i)} \right) dB_s^i$ have zero expectation so that integrating (3.3.2) on $[0, t \wedge T \wedge \tilde{T}]$ and taking expectations, we obtain that

$$\mathbb{E} e^{\beta(p+q)t \wedge T \wedge \tilde{T}} \sum_{i=1}^n |z_{t \wedge T \wedge \tilde{T}}^i - \tilde{z}_{t \wedge T \wedge \tilde{T}}^i| \leq \mathbb{E} \sum_{i=1}^n |z_0^i - \tilde{z}_0^i|. \quad (3.3.4)$$

When $Z_0 = \tilde{Z}_0$ we deduce that for all $t \geq 0$

$$\sum_{i=1}^n \mathbb{E} |z_{t \wedge T \wedge \tilde{T}}^i - \tilde{z}_{t \wedge T \wedge \tilde{T}}^i| = 0,$$

which concludes the proof of pathwise uniqueness.

Taking $T = \tilde{T} = +\infty$ in (3.3.4), we have (3.3.1). □

Proof of Proposition 3.2.6. We start by proving the uniqueness of the invariant distribution. As the process $(\Lambda_t)_t$ lives in $[0, 1]^n$, ρ_{inv} has a finite first order moment. We can thus apply the second part of Lemma 3.3.1 for two solutions to (J(p,q)) starting respectively according to two invariant distributions to deduce that these two invariant distributions are equal.

The candidate density, obtained by a $t \mapsto \sin^2(t)$ change of variables from the density proportional to $e^{-2V(\phi^1, \dots, \phi^n)}$ with potential V defined in (3.1.6) which is candidate to be stationary for the gradient diffusion (3.1.5), writes

$$f_{inv}(\lambda^1, \dots, \lambda^n) = \frac{2^{n(\beta(q-n+1)-2)}}{\mathcal{Z}'} \times \prod_{i=1}^n \left((\lambda^i)^{\beta \frac{p-n+1}{2}-1} (1-\lambda^i)^{\beta \frac{q-n+1}{2}-1} \prod_{j \neq i} |\lambda^j - \lambda^i|^{\beta/2} \right) \mathbb{1}_{\{0 \leq \lambda^1 \leq \dots \leq \lambda^n \leq 1\}}, \quad (3.3.5)$$

where $\mathcal{Z}' = \int e^{-2V(\phi^1, \dots, \phi^n)} d\phi^1 \dots d\phi^n$. The second factor is indeed integrable since the exponents of λ^i and $1 - \lambda^i$ are bigger than -1 .

Let us check that the probability measure ρ_{inv} with density f_{inv} with respect to the Lebesgue measure solves the Fokker-Plack equation in the sense of distributions

$$\mathcal{A}^* \rho_{inv} = 0,$$

where \mathcal{A} is the infinitesimal generator associated with the dynamics (J(p,q)):

$$\mathcal{A} = \sum_{i=1}^n b_i(\lambda^1, \dots, \lambda^n) \frac{\partial}{\partial \lambda^i} + 2 \sum_{i=1}^n \lambda^i (1 - \lambda^i) \frac{\partial^2}{\partial (\lambda^i)^2},$$

and where $b_i(\lambda^1, \dots, \lambda^n) = \beta \left[p - (p+q)\lambda_t^i + \sum_{j \neq i} \frac{\lambda_t^i(1-\lambda_t^j) + \lambda_t^j(1-\lambda_t^i)}{\lambda_t^i - \lambda_t^j} \right]$.

For a test function ϕ twice continuously differentiable, since f_{inv} vanishes for $\lambda^i = \lambda^{i+1}$ for $i \in \{1, \dots, n-1\}$ and since the exponents of λ^i and $1 - \lambda^i$ for $i \in \{1, \dots, n\}$ in f_{inv} are bigger than -1 , we obtain by integration by parts that for $i \in \{1, \dots, n\}$

$$\begin{aligned} & \int_{0 \leq \lambda^1 \leq \dots \leq \lambda^n \leq 1} \lambda^i (1 - \lambda^i) \frac{\partial^2 \phi}{\partial (\lambda^i)^2} (\lambda^1, \dots, \lambda^n) f_{inv}(\lambda^1, \dots, \lambda^n) d\lambda^1 \dots d\lambda^n \\ &= - \int_{0 \leq \lambda^1 \leq \dots \leq \lambda^n \leq 1} \frac{\partial \phi}{\partial \lambda^i} (\lambda^1, \dots, \lambda^n) \left((1 - 2\lambda^i) f_{inv}(\lambda^1, \dots, \lambda^n) + \lambda^i (1 - \lambda^i) \frac{\partial f_{inv}}{\partial \lambda^i} (\lambda^1, \dots, \lambda^n) \right) d\lambda^1 \dots d\lambda^n. \end{aligned}$$

Then we have

$$\begin{aligned} & \int_{0 \leq \lambda^1 \leq \dots \leq \lambda^n \leq 1} \mathcal{A}\phi(\lambda^1, \dots, \lambda^n) d\rho_{inv}(\lambda^1, \dots, \lambda^n) \\ &= \sum_{i=1}^n \int_{0 \leq \lambda^1 \leq \dots \leq \lambda^n \leq 1} \left[(b_i(\lambda^1, \dots, \lambda^n) - 2(1 - 2\lambda^i)) f_{inv}(\lambda^1, \dots, \lambda^n) - 2\lambda^i (1 - \lambda^i) \frac{\partial f_{inv}}{\partial \lambda^i} (\lambda^1, \dots, \lambda^n) \right] \\ & \quad \times \frac{\partial \phi}{\partial \lambda^i} (\lambda^1, \dots, \lambda^n) d\lambda^1 \dots d\lambda^n. \end{aligned}$$

As for $0 < \lambda^1 < \dots < \lambda^n < 1$:

$$\frac{\partial f_{inv}}{\partial \lambda^i} (\lambda^1, \dots, \lambda^n) = \left[\left(\beta \frac{p-n+1}{2} - 1 \right) \frac{1}{\lambda^i} - \left(\beta \frac{q-n+1}{2} - 1 \right) \frac{1}{1-\lambda^i} + \beta \sum_{j \neq i} \frac{1}{\lambda^i - \lambda^j} \right] f_{inv}(\lambda^1, \dots, \lambda^n),$$

a calculus gives us

$$\begin{aligned} & (b_i(\lambda^1, \dots, \lambda^n) - 2(1 - 2\lambda^i)) f_{inv}(\lambda^1, \dots, \lambda^n) - 2\lambda^i (1 - \lambda^i) \frac{\partial f_{inv}}{\partial \lambda^i} (\lambda^1, \dots, \lambda^n) \\ &= f_{inv}(\lambda^1, \dots, \lambda^n) \left[\beta \left[p - (p+q)\lambda^i + \sum_{j \neq i} \frac{\lambda^i(1-\lambda^j) + \lambda^j(1-\lambda^i)}{\lambda^i - \lambda^j} \right] - 2(1 - 2\lambda^i) \right. \\ & \quad \left. - (\beta(p-n+1) - 2)(1 - \lambda^i) + (\beta(q-n+1) - 2)\lambda^i - 2\beta \sum_{j \neq i} \frac{\lambda^i(1-\lambda^i)}{\lambda^i - \lambda^j} \right] \\ &= 0, \end{aligned}$$

where we used the manipulation (3.1.3) and $\sum_{j \neq i} \frac{\lambda^i(1-\lambda^j)}{\lambda^i - \lambda^j} = (n-1)\lambda^i + \sum_{j \neq i} \frac{\lambda^i(1-\lambda^i)}{\lambda^i - \lambda^j}$.

We conclude that $\int_{0 \leq \lambda^1 \leq \dots \leq \lambda^n \leq 1} \mathcal{A}\phi(\lambda^1, \dots, \lambda^n) f_{inv}(\lambda^1, \dots, \lambda^n) d\lambda^1 \dots d\lambda^n = 0$. To deduce the existence of a weak solution to $(J(p,q))$ whose marginals follow the law ρ_{inv} , we may apply [Tre16, Theorem 2.5], as soon as

$$\int_{0 \leq \lambda^1 \leq \dots \leq \lambda^n \leq 1} \sum_{i=1}^n \left\{ \lambda^i (1 - \lambda^i) + |b_i(\lambda^1, \dots, \lambda^n)| \right\} d\rho_{inv}(\lambda^1, \dots, \lambda^n) < +\infty.$$

This property can be proved by remarking that in the definition of (3.3.5) of f_{inv} , the factor $|\lambda^j - \lambda^i|^{\beta/2}$ makes the singularity of the denominator of the interaction term $\frac{\lambda^i(1-\lambda^j) + \lambda^j(1-\lambda^i)}{\lambda^i - \lambda^j}$ integrable. By pathwise uniqueness proved in Lemma 3.3.1, this weak solution to $(J(p,q))$ is a strong solution. □

3.4 Proof of Theorem 3.2.3

We start this section by the proof of Proposition 3.2.5 since this result is crucial in the proof of Theorem 3.2.3.

Proof of Proposition 3.2.5. Before proving the assertion, let us first check by backwards induction that, whatever $p, q \geq 0$ and $\beta > 0$, a solution to $(J(p, q))$ defined up to a stopping time \mathcal{T} is actually continuous and solves $(J(p, q))$ on the closed time interval $[0, \mathcal{T}]$ on $\{\mathcal{T} < \infty\}$. For $k = n$ and for $0 \leq t < \mathcal{T}$, integrating the SDE verified by the largest particle on $[0, t]$, we have :

$$\lambda_t^n = \lambda_0^n + \int_0^t 2\sqrt{\lambda_s^n(1-\lambda_s^n)}dB_s^n + \beta pt - \beta(p+q) \int_0^t \lambda_s^n ds + \beta \sum_{j \neq n} \int_0^t \frac{\lambda_s^n(1-\lambda_s^j) + \lambda_s^j(1-\lambda_s^n)}{\lambda_s^n - \lambda_s^j} ds.$$

As $\lambda_t^n \in [0, 1]$ for all $0 \leq t \leq \mathcal{T}$, the right-hand side of this last equality is bounded too, and the second and the fourth terms of the right-hand side are continuous in t and bounded, and thus admit a limit in $\mathcal{T}-$. Consequently, the term $\beta \sum_{j \neq n} \int_0^t \frac{\lambda_s^n(1-\lambda_s^j) + \lambda_s^j(1-\lambda_s^n)}{\lambda_s^n - \lambda_s^j} ds$ is bounded, and as all its integrands are non-negative, all the integrals in this sum are bounded and non-decreasing, and thus admits a finite limit in $\mathcal{T}-$.

Let $k \in \{1, \dots, n-1\}$ and let us assume that for all $l \in \{k+1, \dots, n\}$,

$$\sum_{j \neq l} \int_0^{\mathcal{T}} \left| \frac{\lambda_s^l(1-\lambda_s^j) + \lambda_s^j(1-\lambda_s^l)}{\lambda_s^l - \lambda_s^j} \right| ds < \infty.$$

For $0 \leq t < \mathcal{T}$, we have :

$$\begin{aligned} \lambda_t^k &= \lambda_0^k + \int_0^t 2\sqrt{\lambda_s^k(1-\lambda_s^k)}dB_s^k + \beta pt - \beta(p+q) \int_0^t \lambda_s^k ds \\ &\quad + \beta \sum_{j < k} \int_0^t \frac{\lambda_s^k(1-\lambda_s^j) + \lambda_s^j(1-\lambda_s^k)}{\lambda_s^k - \lambda_s^j} ds - \beta \sum_{j > k} \int_0^t \frac{\lambda_s^k(1-\lambda_s^j) - \lambda_s^j(1-\lambda_s^k)}{\lambda_s^j - \lambda_s^k} ds. \end{aligned}$$

By induction hypothesis, each integral in the last sum of the right-hand side of this equality is bounded, and consequently as each integrand is non-negative, each integral of this term admits a limit in $\mathcal{T}-$. Taking this into account and by the same arguments as for $(\lambda_t^n)_t$, the second and the fourth terms of the right-hand side are continuous in t and bounded, and thus admit a limit in $\mathcal{T}-$. Consequently the term $\beta \sum_{j < k} \int_0^t \frac{\lambda_s^k(1-\lambda_s^j) + \lambda_s^j(1-\lambda_s^k)}{\lambda_s^k - \lambda_s^j} ds$ is bounded. As all its integrands are non-negative, all the integrals in this sum are bounded and non-decreasing, and thus admits a finite limit in $\mathcal{T}-$.

We moreover proved that $\sum_{j \neq k} \int_0^{\mathcal{T}} \left| \frac{\lambda_s^k(1-\lambda_s^j) + \lambda_s^j(1-\lambda_s^k)}{\lambda_s^k - \lambda_s^j} \right| ds < \infty$, which ends the induction argument. Therefore, still on $\{\mathcal{T} < \infty\}$, $(\lambda_t^1, \dots, \lambda_t^n)$ admits a limit as $t \rightarrow \mathcal{T}-$. Defining $(\lambda_{\mathcal{T}}^1, \lambda_{\mathcal{T}}^2, \dots, \lambda_{\mathcal{T}}^n)$ as this limit, we conclude that $(J(p, q))$ is satisfied on the closed time interval $[0, \mathcal{T}]$ on $\{\mathcal{T} < \infty\}$.

Now, the idea of this proof is to show that the process $\lambda^1 + \dots + \lambda^k$ hits zero only when another process defined below hits zero. We then show that this last process is not smaller than a Cox-Ingersoll-Ross (CIR) process which never hits zero.

Let $s = \frac{1}{2}$, $\varepsilon_s = \frac{1}{16}$, and let us define the following functions on \mathbb{R} :

$$\begin{aligned} E(x) &:= e^{-\frac{1}{x}} \mathbf{1}_{(0, +\infty)}(x) \\ F(x) &:= E(x-s)E(s+\varepsilon_s-x) \\ I(x) &:= 1 - \frac{\int_{-\infty}^x F(z)dz}{\int_{-\infty}^{+\infty} F(z)dz}. \end{aligned}$$

By construction, E , F and I are non-negative smooth functions, F has as support $[s, s + \varepsilon_s]$ and I is non-increasing, equal to 1 on $(-\infty, s]$ and to 0 on $[s + \varepsilon_s, +\infty)$. Let us moreover remark that

$$\begin{aligned} |I'(x)| &= \frac{F(x)}{\int_{-\infty}^{+\infty} F(z)dz} \leq \frac{e^{-\frac{2}{s}}}{\int_{-\infty}^{+\infty} F(z)dz} \\ |I''(x)| &= \left| \frac{1}{(s-x)^2} - \frac{1}{(x-s-\varepsilon_s)^2} \right| \frac{|I'(x)|}{\int_{-\infty}^{+\infty} F(z)dz} < +\infty, \end{aligned}$$

as the last function is continuous and compactly supported on \mathbb{R} . We then define

$$f := x \in [0, 1] \mapsto I(x) \arcsin^2(\sqrt{x}) + (1 - I(x)) \left(\frac{1}{10}x + \frac{9}{10} \right) \in [0, 1].$$

The function f thus coincides with $\arcsin^2(\sqrt{x})$ on $[0, s]$ and with $\frac{1}{10}x + \frac{9}{10}$ on $[s + \varepsilon_s, 1]$, is increasing and is twice continuously differentiable on the whole interval $[0, 1]$.

Let us define for all $i \in \{1, \dots, n\}$: $\psi^i = f(\lambda^i)$. Applying Itô's formula (formally after the stopping time $\inf\{t \geq 0 : \lambda_t^1(1 - \lambda_t^n) = 0\}$), we get

$$\begin{aligned} d\psi_t^i &= f'(\lambda_t^i) d\lambda_t^i + 2f''(\lambda_t^i) \lambda_t^i (1 - \lambda_t^i) dt \\ &= \left\{ 2\sqrt{\psi_t^i} dB_t^i + \left\{ 1 + \beta(p - q) \sqrt{\psi_t^i} \cot \sqrt{\psi_t^i} + 2(\beta(q - n + 1) - 1) \sqrt{\psi_t^i} \cot \left(2\sqrt{\psi_t^i} \right) \right. \right. \\ &\quad \left. \left. + \beta \sqrt{\psi_t^i} \sum_{j \neq i} \left[\cot \left(\sqrt{\psi_t^i} + \sqrt{\tilde{\psi}_t^j} \right) + \cot \left(\sqrt{\psi_t^i} - \sqrt{\tilde{\psi}_t^j} \right) \right] \right\} dt \right\} \mathbf{1}_{\{\psi_t^i \leq f(s)\}} \\ &\quad + \left\{ f'(\lambda_t^i) d\lambda_t^i + 2f''(\lambda_t^i) \lambda_t^i (1 - \lambda_t^i) dt \right\} \mathbf{1}_{\{\psi_t^i > f(s)\}}, \end{aligned}$$

where we denote $\tilde{\psi}_t^j = \arcsin^2(\sqrt{f^{-1}(\psi_t^j)})$ which is always defined as $f : [0, 1] \mapsto [0, 1]$ is increasing and thus injective. Let us note that $\tilde{\psi}_t^j$ coincides with ψ_t^j when $\psi_t^j \leq f(s)$.

For all $k \in \{1, \dots, n\}$, starting with

$$\sigma^{k,0} = 0,$$

$$\sigma^{k,1} = \inf \left\{ t \in [0, \mathcal{T}) : \psi_t^k \leq f\left(\frac{s}{2}\right) \right\} \mathbf{1}_{\{\psi_0^k \geq f(s)\}},$$

let us define inductively for $j \in \mathbb{N}$

$$\begin{aligned} \sigma^{k,2j+1} &= \inf \left\{ t \in (\sigma^{k,2j}, \mathcal{T}) : \psi_t^k \leq f\left(\frac{s}{2}\right) \right\}, \\ \sigma^{k,2j+2} &= \inf \left\{ t \in (\sigma^{k,2j+1}, \mathcal{T}) : \psi_t^k \geq f(s) \right\}, \end{aligned}$$

with the convention $\inf \emptyset = \mathcal{T}$. We can then rewrite :

$$\begin{aligned} d\psi_t^i &= \left\{ 2\sqrt{\psi_t^i} dB_t^i + \left\{ 1 + \beta(p-q)\sqrt{\psi_t^i} \cot \sqrt{\psi_t^i} + 2(\beta(q-n+1) - 1)\sqrt{\psi_t^i} \cot \left(2\sqrt{\psi_t^i} \right) \right. \right. \\ &\quad \left. \left. + \beta\sqrt{\psi_t^i} \sum_{j \neq i} \left[\cot \left(\sqrt{\psi_t^i} + \sqrt{\tilde{\psi}_t^j} \right) + \cot \left(\sqrt{\psi_t^i} - \sqrt{\tilde{\psi}_t^j} \right) \right] \right\} dt \right\} \mathbf{1}_{\{t \in \cup_{l \in \mathbb{N}} [\sigma^{i,2l+1}, \sigma^{i,2l+2})\}} \\ &\quad + \left\{ f'(\lambda_t^i) d\lambda_t^i + 2f''(\lambda_t^i) \lambda_t^i (1 - \lambda_t^i) dt \right\} \mathbf{1}_{\{t \in \cup_{l \in \mathbb{N}} [\sigma^{i,2l}, \sigma^{i,2l+1})\}}. \end{aligned} \quad (3.4.1)$$

We moreover define for all $i \in \{1, \dots, n\}$ and $k \geq i$,

$$\begin{aligned} \theta_i^k(t) &= - \left\{ \frac{1}{2} \beta(p-q) \left(\cot \sqrt{\psi_t^i} - \frac{1}{\sqrt{\psi_t^i}} \right) + (\beta(q-n+1) - 1) \left(\cot \left(2\sqrt{\psi_t^i} \right) - \frac{1}{2\sqrt{\psi_t^i}} \right) \right. \\ &\quad \left. + \frac{\beta}{2} \sum_{j \neq i} \left[\cot \left(\sqrt{\psi_t^i} + \sqrt{\tilde{\psi}_t^j} \right) - \frac{1}{\sqrt{\psi_t^i} + \sqrt{\tilde{\psi}_t^j}} + \cot \left(\sqrt{\psi_t^i} - \sqrt{\tilde{\psi}_t^j} \right) - \frac{1}{\sqrt{\psi_t^i} - \sqrt{\tilde{\psi}_t^j}} \right] \right\} \\ &\quad \times \mathbf{1}_{\{t \in \cup_{l \in \mathbb{N}} [\sigma^{k,2l+1}, \sigma^{k,2l+2})\}}, \end{aligned}$$

and $\Theta^k(t) = (\theta_1^k(t), \dots, \theta_n^k(t), 0, \dots, 0)$ for $k \in \{1, \dots, n\}$.

The function \cot is continuous on $(0, \pi)$, $\cot(x) - \frac{1}{x} \underset{x \rightarrow 0}{\sim} -\frac{x}{3}$, and the function $x \mapsto \cot(x) - \frac{1}{x}$ is decreasing on $[0, \pi)$. Thus, as $0 \leq \psi_t^i \leq \psi_t^k \leq f(s) < \frac{\pi^2}{4}$ on $\cup_{l \in \mathbb{N}} [\sigma^{k,2l+1}, \sigma^{k,2l+2}) \subset \cup_{l \in \mathbb{N}} [\sigma^{i,2l+1}, \sigma^{i,2l+2})$, we have for all $t \in \cup_{l \in \mathbb{N}} [\sigma^{k,2l+1}, \sigma^{k,2l+2})$:

$$\begin{aligned} |\theta_i^k(t)| &\leq \frac{\beta^2(p-q)^2}{4} \cot^2 \left(\sqrt{f(s)} \right) + (\beta(q-n+1) - 1)^2 \cot^2 \left(2\sqrt{f(s)} \right) \\ &\quad + \frac{\beta^2}{4} \sum_{j \neq i} \left[\cot^2 \left(2\sqrt{f(s)} \right) + \cot^2 \left(\sqrt{f(s)} \right) \right] \\ &< +\infty. \end{aligned} \quad (3.4.2)$$

Let us now prove the assertion by first remarking that by construction

$$\mathbb{P}(\exists t \geq 0 : \lambda_t^1 + \lambda_t^2 + \dots + \lambda_t^k = 0) = \mathbb{P}(\exists t \geq 0 : \psi_t^1 + \psi_t^2 + \dots + \psi_t^k = 0).$$

We proceed by backward induction on k noticing that $\min_{k \leq l \leq n} l\beta(p-n+l) = k\beta(p-n+k)$. The idea is to show that the process $\psi^1 + \dots + \psi^k$ is on $t \in \bigcup_{l \in \mathbb{N}} [\sigma^{k,2l+1}, \sigma^{k,2l+2})$ not smaller than a CIR process which never hits zero.

For $k = n$ let us consider for all $t \geq 0$:

$$Z^n(t) = \exp \left\{ \int_0^t \Theta^n(u) d\mathbf{B}_u - \frac{1}{2} \int_0^t \|\Theta^n(u)\|^2 du \right\}.$$

We have thanks to the inequality (3.4.2) :

$$\mathbb{E} \left[\exp \left\{ \frac{1}{2} \int_0^t \|\Theta^n(u)\|^2 du \right\} \right] < \infty \text{ for all } t \geq 0.$$

Then, according to Novikov's criterion (see for instance [KS91, Proposition 5.12 p.198]), Z^n is a \mathbb{P} -martingale, and $\mathbb{E}[Z^n(t)] = 1$. Consequently, recalling that $\mathcal{F}_t = \sigma((\lambda_0^1, \dots, \lambda_0^n), (\mathbf{B}_s)_{s \leq t})$ and defining \mathbb{Q}^n such that

$$\frac{d\mathbb{Q}^n}{d\mathbb{P}} \Big|_{\mathcal{F}_t} = Z(t)$$

and for all $i \in \{1, \dots, n\}$,

$$\tilde{B}_t^i = B_t^i - \int_0^t \theta_i^n(s) ds,$$

$\tilde{\mathbf{B}} = (\tilde{B}_t^1, \dots, \tilde{B}_t^n)_t$ is a \mathbb{Q}^n -Brownian motion according to the Girsanov Theorem (see for instance [KS91, Proposition 5.4 p.194]).

Thus, (3.4.1) can be rewritten in terms of $\tilde{\mathbf{B}}$ as

$$\begin{aligned} d\psi_t^i = & \left\{ 2\sqrt{\psi_t^i} d\tilde{B}_t^i + \left[\beta(p-n+1) + 2\beta \sum_{j \neq i} \frac{\psi_t^i}{\psi_t^i - \psi_t^j} \right] dt \right\} \mathbf{1}_{\{t \in \bigcup_{l \in \mathbb{N}} [\sigma^{n,2l+1}, \sigma^{n,2l+2})\}} \\ & + \left\{ f'(\lambda_t^i) d\lambda_t^i + 2f''(\lambda_t^i) \lambda_t^i (1 - \lambda_t^i) dt \right\} \mathbf{1}_{\{t \in \bigcup_{l \in \mathbb{N}} [\sigma^{n,2l}, \sigma^{n,2l+1})\}} \text{ for all } i \in \{1, \dots, n\}. \end{aligned}$$

We then have for all $t \in \bigcup_{l \in \mathbb{N}} [\sigma^{n,2l+1}, \sigma^{n,2l+2}) \cap [0, \mathcal{T})$:

$$\begin{aligned} d(\psi_t^1 + \dots + \psi_t^n) &= 2 \sum_{i=1}^n \sqrt{\psi_t^i} d\tilde{B}_t^i + n\beta p dt \\ &= 2\sqrt{\psi_t^1 + \dots + \psi_t^n} dW_t^n + n\beta p dt, \end{aligned}$$

where W^n is the Brownian motion defined by :

$$\begin{aligned} dW_t^n = & \mathbf{1}_{\{t \in [0, \mathcal{T}) \cap (\bigcup_{l \in \mathbb{N}} [\sigma^{n,2l+1}, \sigma^{n,2l+2}))\}} \sum_{i=1}^n \left(\mathbf{1}_{\{\sum_{j=1}^n \psi_t^j \neq 0\}} \frac{\sqrt{\psi_t^i}}{\sqrt{\sum_{j=1}^n \psi_t^j}} + \mathbf{1}_{\{\sum_{j=1}^n \psi_t^j = 0\}} \frac{1}{\sqrt{n}} \right) d\tilde{B}_t^i \\ & + \left(\mathbf{1}_{\{t \geq \mathcal{T}\}} + \mathbf{1}_{\{t \in [0, \mathcal{T}) \cap (\bigcup_{l \in \mathbb{N}} [\sigma^{n,2l}, \sigma^{n,2l+1}))\}} \right) \frac{1}{\sqrt{n}} \sum_{i=1}^n d\tilde{B}_t^i. \end{aligned}$$

According to Lemma 3.5.2, the process $(\psi_t^1 + \dots + \psi_t^n)_t$ coincides with CIR processes on each interval of $\bigcup_{l \in \mathbb{N}} [\sigma^{n, 2l+1}, \sigma^{n, 2l+2})$ and each of those CIRs are defined globally, are positive on \mathbb{R}_+ and especially at \mathcal{T} a.s. when $\mathcal{T} < \infty$. If $n\beta p \geq 2$, we can thus conclude for $k = n$.

Let us now assume that for some $k \in \{1, \dots, n-1\}$, $k\beta(p-n+k) \geq 2$ and that the desired property holds for the sum of the $k+1$ smallest particles. Since $(k+1)\beta(p-n+k+1) > k\beta(p-n+k) \geq 2$, the probability of the $k+1$ smallest particles to collide at zero is null. Consequently, the probability for k smallest particles to collide at zero is the probability for exactly the k smallest particles to collide at zero and not the $k+1$ th which stays away from zero. We thus have :

$$\begin{aligned} & \mathbb{P} \left(\mathcal{T} < \infty, \inf_{t \in [0, \mathcal{T})} (\lambda_t^1 + \lambda_t^2 + \dots + \lambda_t^{k+1}) = 0 \right) + \mathbb{P} \left(\mathcal{T} = \infty, \exists t \geq 0 : \lambda_t^1 + \lambda_t^2 + \dots + \lambda_t^{k+1} = 0 \right) \\ &= \mathbb{P} \left(\mathcal{T} < \infty, \inf_{t \in [0, \mathcal{T})} (\psi_t^1 + \psi_t^2 + \dots + \psi_t^{k+1}) = 0 \right) + \mathbb{P} \left(\mathcal{T} = \infty, \exists t \geq 0 : \psi_t^1 + \psi_t^2 + \dots + \psi_t^{k+1} = 0 \right) = 0. \end{aligned}$$

Then,

$$\begin{aligned} & \mathbb{P} \left(\mathcal{T} = \infty, \exists t \geq 0 : \psi_t^1 + \psi_t^2 + \dots + \psi_t^k = 0 \right) \\ &= \lim_{\varepsilon \downarrow 0} \mathbb{P} \left(\mathcal{T} = \infty, \exists t \geq 0 : \psi_t^1 + \psi_t^2 + \dots + \psi_t^k = 0 \text{ and } \psi_t^{k+1} - \psi_t^k \geq \varepsilon \right) \\ &= \lim_{\varepsilon \downarrow 0} \mathbb{P} \left(\mathcal{T} = \infty, \exists t \geq 0 : \psi_t^1 + \psi_t^2 + \dots + \psi_t^k + \mathbf{1}_{\{\psi_t^{k+1} - \psi_t^k < \varepsilon\}} = 0 \right), \end{aligned}$$

and

$$\mathbb{P} \left(\mathcal{T} < \infty, \inf_{t \in [0, \mathcal{T})} (\psi_t^1 + \dots + \psi_t^k) = 0 \right) = \lim_{\varepsilon \downarrow 0} \mathbb{P} \left(\mathcal{T} < \infty, \inf_{t \in [0, \mathcal{T})} (\psi_t^1 + \dots + \psi_t^k + \mathbf{1}_{\{\psi_t^{k+1} - \psi_t^k < \varepsilon\}}) = 0 \right).$$

For $\varepsilon > 0$, starting with

$$\tau_\varepsilon^0 = \inf\{t \in [0, \mathcal{T}) : \psi_t^{k+1} - \psi_t^k \geq \varepsilon\},$$

let us define inductively for $j \in \mathbb{N}$

$$\begin{aligned} \tau_\varepsilon^{2j+1} &= \inf\{t \in (\tau_\varepsilon^{2j}, \mathcal{T}) : \psi_t^{k+1} - \psi_t^k \leq \frac{\varepsilon}{2}\}, \\ \tau_\varepsilon^{2j+2} &= \inf\{t \in (\tau_\varepsilon^{2j+1}, \mathcal{T}) : \psi_t^{k+1} - \psi_t^k \geq \varepsilon\}, \end{aligned}$$

with the convention $\inf \emptyset = \mathcal{T}$. As the function $t \rightarrow \psi_t^{k+1} - \psi_t^k$ is continuous on $[0, \mathcal{T}]$ when $\mathcal{T} < \infty$ and on $[0, \mathcal{T})$ when $\mathcal{T} = \infty$,

$$\tau_\varepsilon^j \xrightarrow{j \rightarrow +\infty} \mathcal{T} \text{ and } \max\{j \in \mathbb{N} : \tau_\varepsilon^j < \mathcal{T}\} < +\infty \text{ a.s. on } \{\mathcal{T} < \infty\}.$$

As $\psi_t^{k+1} - \psi_t^k < \varepsilon$ on $[0, \tau_\varepsilon^0)$ and $[\tau_\varepsilon^{2j+1}, \tau_\varepsilon^{2j+2})$ for $j \in \mathbb{N}$,

$$\begin{aligned} & \left\{ \mathcal{T} < \infty, \inf_{t \in [0, \mathcal{T})} (\psi_t^1 + \dots + \psi_t^k + \mathbf{1}_{\{\psi_t^{k+1} - \psi_t^k < \varepsilon\}}) = 0 \right\} \\ &= \left\{ \mathcal{T} < \infty, \exists j \in \mathbb{N}, \inf_{t \in [\tau_\varepsilon^{2j}, \tau_\varepsilon^{2j+1})} (\psi_t^1 + \dots + \psi_t^k) = 0 \right\} \\ \text{and } & \left\{ \mathcal{T} = \infty, \exists t \geq 0 : \psi_t^1 + \psi_t^2 + \dots + \psi_t^k + \mathbf{1}_{\{\psi_t^{k+1} - \psi_t^k < \varepsilon\}} = 0 \right\} \\ &= \left\{ \mathcal{T} = \infty, \exists j \in \mathbb{N}, \exists t \in [\tau_\varepsilon^{2j}, \tau_\varepsilon^{2j+1}) : \psi_t^1 + \psi_t^2 + \dots + \psi_t^k = 0 \right\}, \end{aligned}$$

it is enough to check that

$$\begin{aligned} & \mathbb{P} \left(\left\{ \mathcal{T} < +\infty, \exists j \in \mathbb{N}, \inf_{t \in [\tau_\varepsilon^{2j}, \tau_\varepsilon^{2j+1})} (\psi_t^1 + \dots + \psi_t^k) = 0 \right\} \right) \\ &+ \mathbb{P} \left(\left\{ \mathcal{T} = +\infty, \exists j \in \mathbb{N}, \exists t \in [\tau_\varepsilon^{2j}, \tau_\varepsilon^{2j+1}) : \psi_t^1 + \psi_t^2 + \dots + \psi_t^k = 0 \right\} \right) = 0. \end{aligned}$$

Moreover, these events can only happen on $t \in \bigcup_{l \in \mathbb{N}} [\sigma^{k, 2l+1}, \sigma^{k, 2l+2})$ so it is in fact enough to check that

$$\begin{aligned} & \mathbb{P} \left(\left\{ \mathcal{T} < +\infty, \exists j, l \in \mathbb{N}, \inf_{t \in [\tau_\varepsilon^{2j}, \tau_\varepsilon^{2j+1}) \cap [\sigma^{k, 2l+1}, \sigma^{k, 2l+2})} (\psi_t^1 + \dots + \psi_t^k) = 0 \right\} \right) \\ &+ \mathbb{P} \left(\left\{ \mathcal{T} = +\infty, \exists j, l \in \mathbb{N}, \exists t \in [\tau_\varepsilon^{2j}, \tau_\varepsilon^{2j+1}) \cap [\sigma^{k, 2l+1}, \sigma^{k, 2l+2}) : \psi_t^1 + \psi_t^2 + \dots + \psi_t^k = 0 \right\} \right) = 0. \end{aligned} \tag{3.4.3}$$

Let us consider for all $t \geq 0$

$$Z^k(t) = \exp \left\{ \int_0^t \Theta^k(u) d\mathbf{B}_u - \frac{1}{2} \int_0^t \|\Theta^k(u)\|^2 du \right\}.$$

We have thanks to the inequality (3.4.2) :

$$\mathbb{E} \left[\exp \left\{ \frac{1}{2} \int_0^t \|\Theta^k(u)\|^2 du \right\} \right] < \infty \text{ for all } t \geq 0.$$

Then, according to Novikov's criterion (see for instance [KS91, Proposition 5.12 p.198]), Z^k is a \mathbb{P} -martingale, and $\mathbb{E}[Z^k(t)] = 1$. Consequently, recalling that $\mathcal{F}_t = \sigma((\lambda_0^1, \dots, \lambda_0^n), (\mathbf{B}_s)_{s \leq t})$ and defining \mathbb{Q}^k such that

$$\frac{d\mathbb{Q}^k}{d\mathbb{P}} \Big|_{\mathcal{F}_t} = Z(t)$$

and for all $i \in \{1, \dots, k\}$,

$$\tilde{B}_t^i = B_t^i - \int_0^t \theta_i^k(s) ds,$$

$\tilde{\mathbf{B}}^{(k)} = (\tilde{B}_t^1, \dots, \tilde{B}_t^k, B_t^{k+1}, \dots, B_t^n)_t$ is a \mathbb{Q}^k -Brownian motion according to the Girsanov Theorem (see for instance [KS91, Proposition 5.4 p.194]). The equality (3.4.3) is still true under the probability \mathbb{Q}^k .

We define the Brownian motion W^k the following way :

$$\begin{aligned} dW_t^k &= \mathbb{1}_{\{t \in [0, \mathcal{T}) \cap (\cup_{l \in \mathbb{N}} [\sigma^{k, 2l+1}, \sigma^{k, 2l+2}])\}} \sum_{i=1}^k \left(\mathbb{1}_{\{\sum_{j=1}^k \psi_t^j \neq 0\}} \frac{\sqrt{\psi_t^i}}{\sqrt{\sum_{j=1}^k \psi_t^j}} + \mathbb{1}_{\{\sum_{j=1}^k \psi_t^j = 0\}} \frac{1}{\sqrt{k}} \right) d\tilde{B}_t^i \\ &+ \left(\mathbb{1}_{\{t \geq \mathcal{T}\}} + \mathbb{1}_{\{t \in [0, \mathcal{T}) \cap (\cup_{l \in \mathbb{N}} [\sigma_s^{k, 2l}, \sigma^{k, 2l+1}])\}} \right) \frac{1}{\sqrt{k}} \sum_{i=1}^k d\tilde{B}_t^i. \end{aligned}$$

We have for all $t \in [\tau_\varepsilon^{2j}, \tau_\varepsilon^{2j+1}) \cap (\cup_{l \in \mathbb{N}} [\sigma^{k, 2l+1}, \sigma^{k, 2l+2}])$, $\psi_t^{k+1} - \psi_t^k > \frac{\varepsilon}{2}$ and

$$\begin{aligned} d(\psi_t^1 + \dots + \psi_t^k) &= 2\sqrt{\psi_t^1 + \dots + \psi_t^k} dW_t^k + k\beta(p - n + k)dt + 2\beta \sum_{i=1}^k \sum_{j=k+1}^n \frac{\psi_t^i}{\psi_t^i - \psi_t^j} dt \\ &\geq 2\sqrt{\psi_t^1 + \dots + \psi_t^k} dW_t^k + k\beta(p - n + k)dt - \frac{4}{\varepsilon}(n - k)\beta \left(\sum_{i=1}^k \psi_t^i \right) dt. \end{aligned} \tag{3.4.4}$$

We can then define on $\{\tau_\varepsilon^{2j} \vee \sigma^{k, 2l+1} < \infty\}$, for each interval $[\tau_\varepsilon^{2j}, \tau_\varepsilon^{2j+1}) \cap [\sigma^{k, 2l+1}, \sigma^{k, 2l+2})$ the process $r^{j, l}$ by $r_0^{j, l} = \left(\lambda_{\tau_\varepsilon^{2j} \vee \sigma^{k, 2l+1}}^1 + \dots + \lambda_{\tau_\varepsilon^{2j} \vee \sigma^{k, 2l+1}}^k \right) \mathbb{1}_{\{\tau_\varepsilon^{2j} \vee \sigma^{k, 2l+1} < \mathcal{T}\}} + \mathbb{1}_{\{\tau_\varepsilon^{2j} \vee \sigma^{k, 2l+1} = \mathcal{T}\}}$ and for all $t \geq 0$:

$$\begin{aligned} dr_t^{j, l} &= 2\sqrt{r_t^{j, l}} dW_{t+\tau_\varepsilon^{2j} \vee \sigma^{k, 2l+1}}^k + \left[-\frac{4}{\varepsilon}\beta(n - k)r_t^{j, l} + k\beta(p - n + k) \right] dt \\ &= 2\sqrt{r_t^{j, l}} d(W_{t+\tau_\varepsilon^{2j} \vee \sigma^{k, 2l+1}}^k - W_{\tau_\varepsilon^{2j} \vee \sigma^{k, 2l+1}}^k + W_{\tau_\varepsilon^{2j} \vee \sigma^{k, 2l+1}}^k) + \left[-\frac{4}{\varepsilon}\beta(n - k)r_t^{j, l} + k\beta(p - n + k) \right] dt \\ &= 2\sqrt{r_t^{j, l}} d\tilde{W}_t^k + \left[-\frac{4}{\varepsilon}\beta(n - k)r_t^{j, l} + k\beta(p - n + k) \right] dt \end{aligned}$$

where conditionally on $\{\tau_\varepsilon^{2j} \vee \sigma^{k, 2l+1} < \infty\}$, by strong Markov property, $(\tilde{W}_t^k = W_{t+\tau_\varepsilon^{2j} \vee \sigma^{k, 2l+1}}^k - W_{\tau_\varepsilon^{2j} \vee \sigma^{k, 2l+1}}^k)_{t \geq 0}$ is a Brownian motion independent from $\mathcal{F}_{\tau_\varepsilon^{2j} \vee \sigma^{k, 2l+1}}$. Conditionally on $\{\tau_\varepsilon^{2j} \vee \sigma^{k, 2l+1} < \infty\}$, the process $r^{j, l}$ is a CIR process defined globally in time according to Lemma 3.5.2 with $a = k\beta(p - n + k)$ and $\sigma = 2$ which satisfy $a \geq \frac{\sigma^2}{2}$, and it stays positive on \mathbb{R}_+ .

This together with (3.4.4) and Theorem 3.5.3 give that for all $t \in [\tau_\varepsilon^{2j}, \tau_\varepsilon^{2j+1}) \cap [\sigma^{k, 2l+1}, \sigma^{k, 2l+2})$,

$$\psi_t^1 + \dots + \psi_t^k \geq r_{t-\tau_\varepsilon^{2j} \vee \sigma^{k, 2l+1}}^{j, l}.$$

We can thus conclude the proof of (i).

To prove (ii), we just need to apply (i) to the process $(1 - \lambda_t^n, \dots, 1 - \lambda_t^1)_t$ which satisfies $(J(q, p))$ according to Remark 3.2.1. \square

Proof of Theorem 3.2.3. By Remark 3.2.1, it is enough to prove Theorem 3.2.3 with the additional assumption $p \geq q$. Indeed, if the theorem is true under this additional assumption, then, if $q > p$, we apply the theorem to the system of SDEs $(J(q,p))$ instead of $(J(p,q))$, which allows to build a solution to $(J(p,q))$ using Remark 3.2.1, and this solution still verifies the assertion (ii) of the theorem. By Lemma 3.3.1, we can then conclude of the existence and uniqueness of a strong solution to $(J(p,q))$, and this solution verifies (ii). We thus make the assumption $p \geq q$ in the rest of the proof.

As explained in the introduction, the main difficulty in proving this result comes from the fact that we have to deal with both the singularity related to the interaction between particles and the singularity related to the edges (zero and one) when two particles hit zero (or one) at the same time. For $\varepsilon > 0$, our method precisely consists in separating these difficulties by defining four new systems of SDEs (\hat{J}_ε) , (\tilde{J}_ε) , (\check{J}_ε) and (\bar{J}_ε) which each remove three types of singularity and coincide with $(J(p,q))$ on domains that cover $\{t \geq 0 : 0 \leq \lambda_t^1 \leq \dots \leq \lambda_t^n \leq 1 \text{ and } (\lambda_t^1 \geq \varepsilon \text{ or } \lambda_t^2 - \lambda_t^1 \geq \varepsilon) \text{ and } (\lambda_t^n \leq 1 - \varepsilon \text{ or } \lambda_t^n - \lambda_t^{n-1} \geq \varepsilon)\}$. This allows us to build a solution to $(J(p,q))$ by piecing together solutions to (\hat{J}_ε) , (\tilde{J}_ε) , (\check{J}_ε) and (\bar{J}_ε) .

Let us consider in this proof the Brownian motion $\mathbf{B} = (B_t^1, \dots, B_t^n)_t$, $\mathcal{F}_t = \sigma((\lambda_0^1, \dots, \lambda_0^n), (\mathbf{B}_s)_{s \leq t})$ for all $t \geq 0$, and the system of SDEs defined by $(J(p,q))$. Let us define for all $\varepsilon > 0$ the following SDEs :

$$\begin{aligned} \forall i \in \{1, \dots, n\}, d\hat{\lambda}_t^{i,\varepsilon} &= 2\sqrt{\hat{\lambda}_t^{i,\varepsilon}(1 - \hat{\lambda}_t^{i,\varepsilon})}dB_t^i \\ &+ \left\{ (1 - 2\hat{\lambda}_t^{i,\varepsilon}) \left[2 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sqrt{1 - \hat{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sqrt{\hat{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 \right] \right. \\ &+ \beta \left[p - (p + q)\hat{\lambda}_t^{i,\varepsilon} + \sum_{j \neq i} \frac{\hat{\lambda}_t^{i,\varepsilon}(1 - \hat{\lambda}_t^{j,\varepsilon}) + \hat{\lambda}_t^{j,\varepsilon}(1 - \hat{\lambda}_t^{i,\varepsilon})}{\hat{\lambda}_t^{i,\varepsilon} - \hat{\lambda}_t^{j,\varepsilon}} \right] \left. \right\} dt \\ &0 \leq \hat{\lambda}_t^{1,\varepsilon} < \dots < \hat{\lambda}_t^{n,\varepsilon} \leq 1, \text{ a.s., } dt - a.e. \end{aligned} \tag{\hat{J}_\varepsilon}$$

$$\left\{ \begin{aligned} d\tilde{\lambda}_t^{1,\varepsilon} &= 2\sqrt{\tilde{\lambda}_t^{1,\varepsilon}(1 - \tilde{\lambda}_t^{1,\varepsilon})}dB_t^1 \\ &+ \beta \left[p - n + 1 - (p + q)\tilde{\lambda}_t^{1,\varepsilon} - 2 \frac{(1 - \tilde{\lambda}_t^{1,\varepsilon}) \wedge \frac{\varepsilon}{2}}{\frac{\varepsilon}{2}} \sum_{j \neq 1} \frac{\tilde{\lambda}_t^{1,\varepsilon}(1 - \tilde{\lambda}_t^{j,\varepsilon})}{(\tilde{\lambda}_t^{j,\varepsilon} - \tilde{\lambda}_t^{1,\varepsilon} \wedge \varepsilon) \vee \varepsilon} \right] dt \\ \forall i \in \{2, \dots, n\}, d\tilde{\lambda}_t^{i,\varepsilon} &= 2\sqrt{\tilde{\lambda}_t^{i,\varepsilon}(1 - \tilde{\lambda}_t^{i,\varepsilon})}dB_t^i \\ &+ \left\{ (1 - 2\tilde{\lambda}_t^{i,\varepsilon}) \left[2 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sqrt{1 - \tilde{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 - 0 \vee \frac{2}{\sqrt{\varepsilon}} \left(\sqrt{\tilde{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2} \right) \wedge 1 \right] \right. \\ &+ \beta \left[p - n + 1 - (p + q)\tilde{\lambda}_t^{i,\varepsilon} + 2 \frac{(1 - \tilde{\lambda}_t^{i,\varepsilon}) \wedge \frac{\varepsilon}{2}}{\frac{\varepsilon}{2}} \frac{\tilde{\lambda}_t^{i,\varepsilon}(1 - \tilde{\lambda}_t^{1,\varepsilon})}{(\tilde{\lambda}_t^{i,\varepsilon} - \tilde{\lambda}_t^{1,\varepsilon}) \vee \varepsilon} + 2 \sum_{j \geq 2, j \neq i} \frac{\tilde{\lambda}_t^{i,\varepsilon}(1 - \tilde{\lambda}_t^{j,\varepsilon})}{\tilde{\lambda}_t^{i,\varepsilon} - \tilde{\lambda}_t^{j,\varepsilon}} \right] \left. \right\} dt \\ &0 \leq \tilde{\lambda}_t^{1,\varepsilon} \leq 1 \text{ and } 0 \leq \tilde{\lambda}_t^{2,\varepsilon} < \dots < \tilde{\lambda}_t^{n,\varepsilon} \leq 1, \text{ a.s., } dt - a.e. \end{aligned} \right. \tag{\tilde{J}_\varepsilon}$$

$$\left\{ \begin{array}{l}
d\check{\lambda}_t^{n,\varepsilon} = 2\sqrt{\check{\lambda}_t^{n,\varepsilon}(1-\check{\lambda}_t^{n,\varepsilon})}dB_t^n \\
+ \beta \left[p-n+1 - (p+q)\check{\lambda}_t^{n,\varepsilon} + 2\frac{\check{\lambda}_t^{n,\varepsilon} \wedge \frac{\varepsilon}{2}}{\frac{\varepsilon}{2}} \sum_{j \neq n} \frac{\check{\lambda}_t^{n,\varepsilon}(1-\check{\lambda}_t^{j,\varepsilon})}{(\check{\lambda}_t^{n,\varepsilon} - \check{\lambda}_t^{j,\varepsilon} \wedge \varepsilon) \vee \varepsilon} \right] dt \\
\forall i \in \{1, \dots, n-1\}, d\check{\lambda}_t^{i,\varepsilon} = 2\sqrt{\check{\lambda}_t^{i,\varepsilon}(1-\check{\lambda}_t^{i,\varepsilon})}dB_t^i \\
+ \left\{ (1-2\check{\lambda}_t^{i,\varepsilon}) \left[2-0 \vee \frac{2}{\sqrt{\varepsilon}} \left(\sqrt{1-\check{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2} \right) \wedge 1 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sqrt{\check{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 \right] \right. \\
\left. + \beta \left[p-n+1 - (p+q)\check{\lambda}_t^{i,\varepsilon} - 2\frac{(1-\check{\lambda}_t^{i,\varepsilon}) \wedge \frac{\varepsilon}{2}}{\frac{\varepsilon}{2}} \frac{\check{\lambda}_t^{i,\varepsilon}(1-\check{\lambda}_t^{i,\varepsilon})}{(\check{\lambda}_t^{n,\varepsilon} - \check{\lambda}_t^{i,\varepsilon}) \vee \varepsilon} + 2 \sum_{j \leq n-1, j \neq i} \frac{\check{\lambda}_t^{i,\varepsilon}(1-\check{\lambda}_t^{j,\varepsilon})}{\check{\lambda}_t^{i,\varepsilon} - \check{\lambda}_t^{j,\varepsilon}} \right] \right\} dt \\
0 \leq \check{\lambda}_t^{n,\varepsilon} \leq 1 \text{ and } 0 \leq \check{\lambda}_t^{1,\varepsilon} < \dots < \check{\lambda}_t^{n-1,\varepsilon} \leq 1, \text{ a.s., } dt - a.e.
\end{array} \right. \quad (\check{J}_\varepsilon)$$

$$\left\{ \begin{array}{l}
d\bar{\lambda}_t^{1,\varepsilon} = 2\sqrt{\bar{\lambda}_t^{1,\varepsilon}(1-\bar{\lambda}_t^{1,\varepsilon})}dB_t^1 \\
+ \beta \left[p-n+1 - (p+q)\bar{\lambda}_t^{1,\varepsilon} - 2\frac{(1-\bar{\lambda}_t^{1,\varepsilon}) \wedge \frac{\varepsilon}{2}}{\frac{\varepsilon}{2}} \sum_{j \neq 1} \frac{\bar{\lambda}_t^{1,\varepsilon}(1-\bar{\lambda}_t^{j,\varepsilon})}{(\bar{\lambda}_t^{j,\varepsilon} - \bar{\lambda}_t^{1,\varepsilon} \wedge \varepsilon) \vee \varepsilon} \right] dt \\
d\bar{\lambda}_t^{n,\varepsilon} = 2\sqrt{\bar{\lambda}_t^{n,\varepsilon}(1-\bar{\lambda}_t^{n,\varepsilon})}dB_t^n \\
+ \beta \left[p-n+1 - (p+q)\bar{\lambda}_t^{n,\varepsilon} + 2\frac{\bar{\lambda}_t^{n,\varepsilon} \wedge \frac{\varepsilon}{2}}{\frac{\varepsilon}{2}} \sum_{j \neq n} \frac{\bar{\lambda}_t^{n,\varepsilon}(1-\bar{\lambda}_t^{j,\varepsilon})}{(\bar{\lambda}_t^{n,\varepsilon} - \bar{\lambda}_t^{j,\varepsilon} \wedge \varepsilon) \vee \varepsilon} \right] dt \\
\forall i \in \{2, \dots, n-1\}, d\bar{\lambda}_t^{i,\varepsilon} = 2\sqrt{\bar{\lambda}_t^{i,\varepsilon}(1-\bar{\lambda}_t^{i,\varepsilon})}dB_t^i \\
+ \left\{ (1-2\bar{\lambda}_t^{i,\varepsilon}) \left[2-0 \vee \frac{2}{\sqrt{\varepsilon}} \left(\sqrt{1-\bar{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2} \right) \wedge 1 - 0 \vee \frac{2}{\sqrt{\varepsilon}} \left(\sqrt{\bar{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2} \right) \wedge 1 \right] \right. \\
\left. + \beta \left[p-n+1 - (p+q)\bar{\lambda}_t^{i,\varepsilon} + 2\frac{(1-\bar{\lambda}_t^{i,\varepsilon}) \wedge \frac{\varepsilon}{2}}{\frac{\varepsilon}{2}} \left(\frac{\bar{\lambda}_t^{i,\varepsilon}(1-\bar{\lambda}_t^{1,\varepsilon})}{(\bar{\lambda}_t^{i,\varepsilon} - \bar{\lambda}_t^{1,\varepsilon}) \vee \varepsilon} - \frac{\bar{\lambda}_t^{i,\varepsilon}(1-\bar{\lambda}_t^{n,\varepsilon})}{(\bar{\lambda}_t^{n,\varepsilon} - \bar{\lambda}_t^{i,\varepsilon}) \vee \varepsilon} \right) \right. \right. \\
\left. \left. + 2 \sum_{2 \leq j \leq n-1, j \neq i} \frac{\bar{\lambda}_t^{i,\varepsilon}(1-\bar{\lambda}_t^{j,\varepsilon})}{\bar{\lambda}_t^{i,\varepsilon} - \bar{\lambda}_t^{j,\varepsilon}} \right] \right\} dt \\
0 \leq \bar{\lambda}_t^{1,\varepsilon} \leq 1, 0 \leq \bar{\lambda}_t^{n,\varepsilon} \leq 1 \text{ and } 0 \leq \bar{\lambda}_t^{2,\varepsilon} < \dots < \bar{\lambda}_t^{n-1,\varepsilon} \leq 1, \text{ a.s., } dt - a.e.
\end{array} \right. \quad (\bar{J}_\varepsilon)$$

These systems are built such that :

- (\hat{J}_ε) coincides with $(J(p,q))$ on $\{t, \hat{\lambda}_t^{1,\varepsilon} \geq \frac{\varepsilon}{2} \text{ and } \hat{\lambda}_t^{n,\varepsilon} \leq 1 - \frac{\varepsilon}{2}\}$,
- (\tilde{J}_ε) coincides with $(J(p,q))$ on $\{t, \tilde{\lambda}_t^{1,\varepsilon} \leq \varepsilon \text{ and } \tilde{\lambda}_t^{2,\varepsilon} - \tilde{\lambda}_t^1 \geq \varepsilon \text{ and } \tilde{\lambda}_t^{n,\varepsilon} \leq 1 - \frac{\varepsilon}{2}\}$,
- (\check{J}_ε) coincides with $(J(p,q))$ on $\{t, \check{\lambda}_t^{1,\varepsilon} \geq \frac{\varepsilon}{2} \text{ and } \check{\lambda}_t^{n,\varepsilon} \geq 1 - \varepsilon \text{ and } \check{\lambda}_t^{n,\varepsilon} - \check{\lambda}_t^{n-1,\varepsilon} \geq \varepsilon\}$,

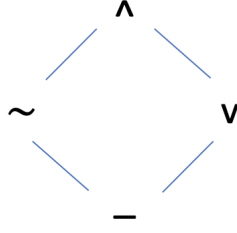


Figure 3.1 – Graphical representation of (Δ, d) .

(\bar{J}_ε) coincides with $(J(p,q))$ on $\{t, \bar{\lambda}_t^{1,\varepsilon} \leq \varepsilon \text{ and } \bar{\lambda}_t^{2,\varepsilon} - \bar{\lambda}_t^1 \geq \varepsilon \text{ and } \bar{\lambda}_t^{n,\varepsilon} \geq 1 - \varepsilon \text{ and } \bar{\lambda}_t^{n,\varepsilon} - \bar{\lambda}_t^{n-1,\varepsilon} \geq \varepsilon\}$.

Lemmas 3.4.1, 3.4.2, 3.4.3 and 3.4.4 give the existence of global pathwise unique strong solutions to (\hat{J}_ε) , (\check{J}_ε) , (\check{J}_ε) and (\bar{J}_ε) with any random initial condition with ordered non-negative coordinates and independent from the driving Brownian motion.

For $\xi \in [0, 1]^n$ deterministic with ordered coordinates, let $\hat{\Lambda}^{\varepsilon,T,\xi} = (\hat{\lambda}_t^{1,\varepsilon,T,\xi}, \dots, \hat{\lambda}_t^{1,\varepsilon,T,\xi})_t$, $\tilde{\Lambda}^{\varepsilon,T,\xi} = (\tilde{\lambda}_t^{1,\varepsilon,T,\xi}, \dots, \tilde{\lambda}_t^{1,\varepsilon,T,\xi})_t$, $\check{\Lambda}^{\varepsilon,T,\xi} = (\check{\lambda}_t^{1,\varepsilon,T,\xi}, \dots, \check{\lambda}_t^{1,\varepsilon,T,\xi})_t$ and $\bar{\Lambda}^{\varepsilon,T,\xi} = (\bar{\lambda}_t^{1,\varepsilon,T,\xi}, \dots, \bar{\lambda}_t^{1,\varepsilon,T,\xi})_t$ denotes the process solution to respectively (\hat{J}_ε) , (\check{J}_ε) , (\check{J}_ε) and (\bar{J}_ε) on $[T, +\infty)$ starting from ξ at time T and equal to 0 on $(-\infty, T)$.

Let us note d the distance on the set $\Delta = \{\wedge, \sim, \vee, -\}$ defined by

$$\begin{aligned} d(\sim, \vee) &= d(\wedge, -) = 2, \\ d(\wedge, \sim) &= d(\wedge, \vee) = d(-, \sim) = d(-, \vee) = 1, \end{aligned}$$

which kind be more easily visualised in Figure 3.1.

Let us moreover define the application defined on a subset of $\Delta \times \Delta \times [0, +\infty) \times [0, 1]^n$ by

$$T(\text{ymb}, \text{ymb}', t, x) = \begin{cases} \inf \left\{ t' \geq t : \binom{\text{ymb}}{\lambda}^{1,\varepsilon,t,x} \leq \frac{\varepsilon}{2} \right\} & \text{if } (\text{ymb}, \text{ymb}') \in \{(\wedge, \sim), (\vee, -)\}, \\ \inf \left\{ t' \geq t : \binom{\text{ymb}}{\lambda}^{n,\varepsilon,t,x} \geq 1 - \frac{\varepsilon}{2} \right\} & \text{if } (\text{ymb}, \text{ymb}') \in \{(\wedge, \vee), (\sim, -)\}, \\ \inf \left\{ t' \geq t : \binom{\text{ymb}}{\lambda}^{n,\varepsilon,t,x} \leq 1 - \varepsilon \right\} & \text{if } (\text{ymb}, \text{ymb}') \in \{(\vee, \wedge), (-, \sim)\}, \\ \inf \left\{ t' \geq t : \binom{\text{ymb}}{\lambda}^{1,\varepsilon,t,x} \geq \varepsilon \right\} & \text{if } (\text{ymb}, \text{ymb}') \in \{(\sim, \wedge), (-, \vee)\}. \end{cases}$$

We define by induction

$$\tau_0^\varepsilon = 0,$$

$$\begin{aligned} \text{symp}_0^\varepsilon &= \begin{cases} \wedge & \text{on } \{\lambda_0^1 \geq \varepsilon \text{ and } \lambda_0^n \leq 1 - \varepsilon\}, \\ \sim & \text{on } \{\lambda_0^1 < \varepsilon \text{ and } \lambda_0^n \leq 1 - \varepsilon\}, \\ \vee & \text{on } \{\lambda_0^1 \geq \varepsilon \text{ and } \lambda_0^n > 1 - \varepsilon\}, \\ - & \text{on } \{\lambda_0^1 < \varepsilon \text{ and } \lambda_0^n > 1 - \varepsilon\}, \end{cases} \\ \tau_1^\varepsilon &= \bigwedge_{\text{symp} \in \Delta, d(\text{symp}, \text{symp}_0^\varepsilon)=1} \mathbb{T}(\text{symp}_0^\varepsilon, \text{symp}, 0, \Lambda_0), \end{aligned}$$

if $\#\{\text{symp} \in \Delta \setminus \{\text{symp}_0^\varepsilon\} \text{ such that } \tau_1^\varepsilon = \mathbb{T}(\text{symp}_0^\varepsilon, \text{symp}, 0, \Lambda_0)\} = 1$ then $\text{symp}_1^\varepsilon$ is the only element of that set. Otherwise, $\text{symp}_1^\varepsilon$ is the only element of Δ such that $d(\text{symp}_0^\varepsilon, \text{symp}_1^\varepsilon) = 2$,

$$\begin{aligned} \mathcal{X}_t^{(1)} &= \left(\begin{array}{c} \text{symp}_0^\varepsilon \\ \Lambda \end{array} \right)^{\varepsilon, 0, \Lambda_0} \mathbb{1}_{\{0 \leq t \leq \tau_1^\varepsilon\}} \text{ for all } t \in \mathbb{R}, \\ \tau_2^\varepsilon &= \bigwedge_{\text{symp} \in \Delta, d(\text{symp}, \text{symp}_1^\varepsilon)=1} \mathbb{T}(\text{symp}_1^\varepsilon, \text{symp}, \tau_1^\varepsilon, \mathcal{X}_{\tau_1^\varepsilon}^{(1)}), \\ \mathcal{X}_t^{(2)} &= \left(\begin{array}{c} \text{symp}_1^\varepsilon \\ \Lambda \end{array} \right)^{\varepsilon, \tau_1^\varepsilon, \mathcal{X}_{\tau_1^\varepsilon}^{(1)}} \mathbb{1}_{\{\tau_1^\varepsilon < t \leq \tau_2^\varepsilon\}} \text{ for all } t \in \mathbb{R}, \\ &\vdots \\ \tau_{i+1}^\varepsilon &= \bigwedge_{\text{symp} \in \Delta, d(\text{symp}, \text{symp}_i^\varepsilon)=1} \mathbb{T}(\text{symp}_i^\varepsilon, \text{symp}, \tau_i^\varepsilon, \mathcal{X}_{\tau_i^\varepsilon}^{(i)}), \end{aligned}$$

if $\#\{\text{symp} \in \Delta \setminus \{\text{symp}_i^\varepsilon\} \text{ such that } \tau_{i+1}^\varepsilon = \mathbb{T}(\text{symp}_i^\varepsilon, \text{symp}, \tau_i^\varepsilon, \mathcal{X}_{\tau_i^\varepsilon}^{(i)})\} = 1$ then $\text{symp}_{i+1}^\varepsilon$ is the only element of that set. Otherwise, $\text{symp}_{i+1}^\varepsilon$ is the only element of Δ such that $d(\text{symp}_i^\varepsilon, \text{symp}_{i+1}^\varepsilon) = 2$,

$$\begin{aligned} \mathcal{X}_t^{(i+1)} &= \left(\begin{array}{c} \text{symp}_i^\varepsilon \\ \Lambda \end{array} \right)^{\varepsilon, \tau_i^\varepsilon, \mathcal{X}_{\tau_i^\varepsilon}^{(i)}} \mathbb{1}_{\{\tau_i^\varepsilon < t \leq \tau_{i+1}^\varepsilon\}} \text{ for all } t \in \mathbb{R}, \\ &\vdots \end{aligned}$$

and as for all $i \in \mathbb{N}$, the τ_i^ε defined before are stopping times for the filtration $(\mathcal{F}_t)_{t \geq 0}$, the random vectors $\mathbb{1}_{\{\tau_i^\varepsilon < +\infty\}} \mathcal{X}_{\tau_i^\varepsilon}^{(i)}$ are $\mathcal{F}_{\tau_i^\varepsilon}$ -measurable, the construction makes sense.

We finally define for all $\varepsilon > 0$ and $t \geq 0$:

$$\mathcal{Z}_t^\varepsilon = \sum_{i=1}^{+\infty} \mathcal{X}_t^{(i)}.$$

Let us verify that the process $(\mathcal{Z}_t^\varepsilon)_t$ is defined globally. To do so, we show that there is no accumulation of the stopping times τ_i^ε by first separating them between the ones coping with the singularity at 0 and the ones coping with the singularity at 1.

On $\text{symp}_0^\varepsilon \in \{\vee, \wedge\}$, let us define $(\sigma_i^{0, \varepsilon})_{i \geq 0}$ as the subsequence of $(\tau_i^\varepsilon)_{i \geq 0}$ corresponding to the stopping times involving the position of the smallest particle of the considered system regarding ε and $\frac{\varepsilon}{2}$ and the stopping times that can be attributed both to a change of position of the smallest particle and of the biggest particle. This means that on $\text{symp}_0^\varepsilon \in \{\vee, \wedge\}$, we can write $(\sigma_i^{0, \varepsilon})_{i \geq 0} = (\tau_{\varphi(i)}^\varepsilon)_{i \geq 0}$ where φ is the only bijective increasing map from \mathbb{N}

to $\{j \in \mathbb{N}^*, (symb_{j-1}^\varepsilon, symb_j^\varepsilon) \in \{(\wedge, \sim), (\vee, -), (\sim, \wedge), (-, \vee), (\wedge, -), (-, \wedge), (\sim, \vee), (\vee, \sim)\}$, where $\mathbb{N} = \{0, 1, 2, \dots\}$ and $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$.

On $symb_0^\varepsilon \in \{\sim, -\}$, let us define $\sigma_0^{0,\varepsilon} = 0$ and $(\sigma_i^{0,\varepsilon})_{i \geq 1}$ as the subsequence of $(\tau_i^\varepsilon)_{i \geq 0}$ corresponding to the stopping times involving the position of the smallest particle of the considered system regarding ε and $\frac{\varepsilon}{2}$ and the stopping times that be attributed both to a change of position of the smallest particle and of the biggest particle. This means that on $symb_0^\varepsilon \in \{\sim, -\}$, we can write $(\sigma_i^{0,\varepsilon})_{i \geq 1} = (\tau_{\varphi(i)}^\varepsilon)_{i \geq 0}$ where φ is the only bijective increasing map from \mathbb{N}^* to $\{j \in \mathbb{N}^*, (symb_{j-1}^\varepsilon, symb_j^\varepsilon) \in \{(\wedge, \sim), (\vee, -), (\sim, \wedge), (-, \vee), (\wedge, -), (-, \wedge), (\sim, \vee), (\vee, \sim)\}$.

On $symb_0^\varepsilon \in \{\sim, \wedge\}$, let us define the sequence $(\sigma_i^{1,\varepsilon})_{i \geq 0}$ as the subsequence of $(\tau_i^\varepsilon)_{i \geq 0}$ corresponding to the stopping times involving the position of the biggest particle of the considered system regarding $1 - \varepsilon$ and $1 - \frac{\varepsilon}{2}$. This means that on $symb_0^\varepsilon \in \{\sim, \wedge\}$, we can write $(\sigma_i^{1,\varepsilon})_{i \geq 0} = (\tau_{\varphi(i)}^\varepsilon)_{i \geq 0}$ where φ is the only bijective increasing map from \mathbb{N} to $\{j \in \mathbb{N}^*, (symb_{j-1}^\varepsilon, symb_j^\varepsilon) \in \{(\wedge, \vee), (\sim, -), (\vee, \wedge), (-, \sim)\}$.

On $symb_0^\varepsilon \in \{\vee, -\}$, let us define the sequence $(\sigma_i^{1,\varepsilon})_{i \geq 0}$ as $\sigma_0^{1,\varepsilon} = 0$ and $(\sigma_i^{1,\varepsilon})_{i \geq 1}$ as the subsequence of $(\tau_i^\varepsilon)_{i \geq 0}$ corresponding to the stopping times involving the position of the biggest particle of the considered system regarding $1 - \varepsilon$ and $1 - \frac{\varepsilon}{2}$. This means that on $symb_0^\varepsilon \in \{\vee, -\}$, we can write $(\sigma_i^{1,\varepsilon})_{i \geq 1} = (\tau_{\varphi(i)}^\varepsilon)_{i \geq 0}$ where φ is the only bijective increasing map from \mathbb{N}^* to $\{j \in \mathbb{N}^*, (symb_{j-1}^\varepsilon, symb_j^\varepsilon) \in \{(\wedge, \vee), (\sim, -), (\vee, \wedge), (-, \sim)\}$.

We have for $i \geq 1$:

- for all $t \in (\sigma_{2i+1}^{0,\varepsilon}, \sigma_{2i+2}^{0,\varepsilon}]$,

$$d\mathcal{Z}_t^{\varepsilon,1} \leq 2\sqrt{\mathcal{Z}_t^{\varepsilon,1}(1 - \mathcal{Z}_t^{\varepsilon,1})}dB_t^1 + \beta [p - n + 1 - (p + q)\mathcal{Z}_t^{\varepsilon,1}] dt,$$

and thus, according to the pathwise comparison theorem of Ikeda and Watanabe (that we recall in Theorem 3.5.3 below)

$$\mathcal{Z}_t^{\varepsilon,1} \leq r_{t - \sigma_{2i+1}^{0,\varepsilon}}^{0,i},$$

where for all $t \geq 0$

$$r_t^{0,i} = \frac{\varepsilon}{2} + 2 \int_0^t \sqrt{r_s^{0,i}(1 - r_s^{0,i})} dB_{s + \sigma_{2i+1}^{0,\varepsilon}}^1 + \beta(p - n + 1)t - \beta(p + q) \int_0^t r_s^{0,i} ds,$$

which is a real Jacobi process. The times $\sigma_{2i+2}^{0,\varepsilon} - \sigma_{2i+1}^{0,\varepsilon}$ for all $i \geq 1$ are larger than the time interval for the Jacobi processes $r^{0,i}$ to go from $\frac{\varepsilon}{2}$ to ε . Moreover, the times for the $r^{0,i}$ processes to go from $\frac{\varepsilon}{2}$ to ε are iid positive random variables. Consequently, there is no accumulation of the stopping times $\sigma_i^{0,\varepsilon}$ which go to infinity as $i \rightarrow \infty$.

- for all $t \in (\sigma_{2i+1}^{1,\varepsilon}, \sigma_{2i+2}^{1,\varepsilon}]$,

$$d\mathcal{Z}_t^{\varepsilon,n} \geq 2\sqrt{\mathcal{Z}_t^{\varepsilon,n}(1 - \mathcal{Z}_t^{\varepsilon,n})}dB_t^n + \beta [p - n + 1 - (p + q)\mathcal{Z}_t^{\varepsilon,n}] dt,$$

and thus

$$\mathcal{Z}_t^{\varepsilon,n} \geq r_{t-\sigma_{2i+1}^{1,\varepsilon}}^{1,i},$$

where for all $t \geq 0$

$$r_t^{1,i} = 1 - \frac{\varepsilon}{2} + 2 \int_0^t \sqrt{r_s^{1,i}(1-r_s^{1,i})} dB_{s+\sigma_{2i+1}^{1,\varepsilon}}^n + \beta(p-n+1)t - \beta(p+q) \int_0^t r_s^{1,i} ds,$$

which is a real Jacobi process. The times $\sigma_{2i+2}^{1,\varepsilon} - \sigma_{2i+1}^{1,\varepsilon}$ for all $i \geq 1$ are larger than the time interval for the Jacobi processes $r^{1,i}$ to go from $1 - \frac{\varepsilon}{2}$ to $1 - \varepsilon$. Moreover, the times for the $r^{1,i}$ processes to go from $1 - \frac{\varepsilon}{2}$ to $1 - \varepsilon$ are iid positive random variables. Consequently, there is no accumulation of the stopping times $\sigma_i^{1,\varepsilon}$ which go to infinity as $i \rightarrow \infty$.

The stochastic process \mathcal{Z}^ε is thus defined globally.

We recall that the SDEs (\hat{J}_ε) , (\check{J}_ε) , (\bar{J}_ε) , (\tilde{J}_ε) respectively coincide with the system of SDEs (J(p,q)) on $\{t, \hat{\lambda}_t^{1,\varepsilon} \geq \frac{\varepsilon}{2} \text{ and } \hat{\lambda}_t^{n,\varepsilon} \leq 1 - \frac{\varepsilon}{2}\}$, $\{t, \tilde{\lambda}_t^{1,\varepsilon} \leq \varepsilon \text{ and } \tilde{\lambda}_t^{2,\varepsilon} - \tilde{\lambda}_t^1 \geq \varepsilon \text{ and } \tilde{\lambda}_t^{n,\varepsilon} \leq 1 - \frac{\varepsilon}{2}\}$, $\{t, \check{\lambda}_t^{1,\varepsilon} \geq \frac{\varepsilon}{2} \text{ and } \check{\lambda}_t^{n,\varepsilon} \geq 1 - \varepsilon \text{ and } \check{\lambda}_t^{n,\varepsilon} - \check{\lambda}_t^{n-1,\varepsilon} \geq \varepsilon\}$ and $\{t, \bar{\lambda}_t^{1,\varepsilon} \leq \varepsilon \text{ and } \bar{\lambda}_t^{2,\varepsilon} - \bar{\lambda}_t^1 \geq \varepsilon \text{ and } \bar{\lambda}_t^{n,\varepsilon} \geq 1 - \varepsilon \text{ and } \bar{\lambda}_t^{n,\varepsilon} - \bar{\lambda}_t^{n-1,\varepsilon} \geq \varepsilon\}$. On the other hand, $(\mathcal{Z}_t^\varepsilon)_t$ evolves according to those four SDEs, and verify $\mathcal{Z}_t^{\varepsilon,1} \geq \frac{\varepsilon}{2}$ on $[\tau_i^\varepsilon, \tau_{i+1}^\varepsilon]$ for i such that $\text{symp}_i^\varepsilon \in \{\wedge, \vee\}$ and $\mathcal{Z}_t^{\varepsilon,n} \leq 1 - \frac{\varepsilon}{2}$ on $[\tau_i^\varepsilon, \tau_{i+1}^\varepsilon]$ for i such that $\text{symp}_i^\varepsilon \in \{\wedge, \sim\}$. By induction on i , $(\mathcal{Z}_t^\varepsilon)_t$ is a solution to (J(p,q)) until

$$\begin{aligned} & \inf \left\{ t \in \bigcup_{i \in \mathbb{N}, \text{symp}_i^\varepsilon \in \{\sim, -\}} [\tau_i^\varepsilon, \tau_{i+1}^\varepsilon], \mathcal{Z}_t^{2,\varepsilon} - \mathcal{Z}_t^{1,\varepsilon} \leq \varepsilon \right\} \wedge \inf \left\{ t \in \bigcup_{i \in \mathbb{N}, \text{symp}_i^\varepsilon \in \{\vee, -\}} [\tau_i^\varepsilon, \tau_{i+1}^\varepsilon], \mathcal{Z}_t^{n,\varepsilon} - \mathcal{Z}_t^{n-1,\varepsilon} \leq \varepsilon \right\} \\ & \geq \inf \left\{ t \geq 0 : \left(\mathcal{Z}_t^{1,\varepsilon} \leq \varepsilon \text{ and } \mathcal{Z}_t^{2,\varepsilon} - \mathcal{Z}_t^{1,\varepsilon} \leq \varepsilon \right) \text{ or } \left(\mathcal{Z}_t^{n,\varepsilon} \geq 1 - \varepsilon \text{ and } \mathcal{Z}_t^{n,\varepsilon} - \mathcal{Z}_t^{n-1,\varepsilon} \leq \varepsilon \right) \right\} =: \zeta_\varepsilon. \end{aligned}$$

From Lemmas 3.4.1, 3.4.2, 3.4.3 and 3.4.4 we have :

$$\mathbb{P}\{\exists i \in \{j \in \mathbb{N}, \text{symp}_j^\varepsilon = \wedge\}, \exists t \in (\tau_i^\varepsilon, \tau_{i+1}^\varepsilon] : \mathcal{Z}_t^{k,\varepsilon} = \mathcal{Z}_t^{k+1,\varepsilon} \text{ and } \mathcal{Z}_t^{l,\varepsilon} = \mathcal{Z}_t^{l+1,\varepsilon} \text{ for some } 0 \leq k < l \leq n\} = 0 \quad (3.4.5)$$

$$\mathbb{P}\{\exists i \in \{j \in \mathbb{N}, \text{symp}_j^\varepsilon = \sim\}, \exists t \in (\tau_i^\varepsilon, \tau_{i+1}^\varepsilon] : \mathcal{Z}_t^{k,\varepsilon} = \mathcal{Z}_t^{k+1,\varepsilon} \text{ and } \mathcal{Z}_t^{l,\varepsilon} = \mathcal{Z}_t^{l+1,\varepsilon} \text{ for some } 2 \leq k < l \leq n\} = 0 \quad (3.4.6)$$

$$\begin{aligned} & \mathbb{P}\{\exists i \in \{j \in \mathbb{N}, \text{symp}_j^\varepsilon = \vee\}, \exists t \in (\tau_i^\varepsilon, \tau_{i+1}^\varepsilon] : \mathcal{Z}_t^{k,\varepsilon} = \mathcal{Z}_t^{k+1,\varepsilon} \text{ and } \mathcal{Z}_t^{l,\varepsilon} = \mathcal{Z}_t^{l+1,\varepsilon} \\ & \quad \text{for some } 0 \leq k < l \leq n-2\} = 0 \quad (3.4.7) \end{aligned}$$

$$\begin{aligned} & \mathbb{P}\{\exists i \in \{j \in \mathbb{N}, \text{symp}_j^\varepsilon = -\}, \exists t \in (\tau_i^\varepsilon, \tau_{i+1}^\varepsilon] : \mathcal{Z}_t^{k,\varepsilon} = \mathcal{Z}_t^{k+1,\varepsilon} \text{ and } \mathcal{Z}_t^{l,\varepsilon} = \mathcal{Z}_t^{l+1,\varepsilon} \\ & \quad \text{for some } 2 \leq k < l \leq n-2\} = 0 \quad (3.4.8) \end{aligned}$$

where by convention $\mathcal{Z}^{0,\varepsilon} \equiv 0$ and $\mathcal{Z}^{n+1,\varepsilon} \equiv 1$.

On the time intervals $[\tau_i^\varepsilon \wedge \zeta_\varepsilon, \tau_{i+1}^\varepsilon \wedge \zeta_\varepsilon]$ for $i \in \{j \in \mathbb{N}, \text{symp}_j^\varepsilon \in \{\sim, -\}\}$, we have $\mathcal{Z}_t^{2,\varepsilon} - \mathcal{Z}_t^{1,\varepsilon} \geq \varepsilon$ and on the time intervals $[\tau_i^\varepsilon \wedge \zeta_\varepsilon, \tau_{i+1}^\varepsilon \wedge \zeta_\varepsilon]$ for $i \in \{j \in \mathbb{N}, \text{symp}_j^\varepsilon \in \{\vee, -\}\}$

, we have $\mathcal{Z}_t^{n,\varepsilon} - \mathcal{Z}_t^{n-1,\varepsilon} \geq \varepsilon$. This together with (3.4.5,3.4.6,3.4.7,3.4.8) allows to conclude :

$$\mathbb{P}\left\{\exists t \in (0, \zeta_\varepsilon] : \mathcal{Z}_t^{i,\varepsilon} = \mathcal{Z}_t^{i+1,\varepsilon} \text{ and } \mathcal{Z}_t^{j,\varepsilon} = \mathcal{Z}_t^{j+1,\varepsilon} \text{ for some } 1 \leq i < j \leq n-1\right\} = 0. \quad (3.4.9)$$

As the solutions to equation (J(p,q)) are pathwise unique (see Lemma 3.3.1), for $n \in \mathbb{N}^*$, the processes $\mathcal{Z}_t^{\frac{1}{n}}$ and $\mathcal{Z}_t^{\frac{1}{n+1}}$ coincide on $\left[0, \zeta_{\frac{1}{n}} \wedge \zeta_{\frac{1}{n+1}}\right]$. Thus, $\zeta_{\frac{1}{n}} \wedge \zeta_{\frac{1}{n+1}} = \zeta_{\frac{1}{n}}$ and the sequence $\left(\zeta_{\frac{1}{n}}\right)_{n \in \mathbb{N}^*}$ is non-decreasing. Moreover, for all $n \in \mathbb{N}^*$, $\mathcal{Z}_t^{\frac{1}{n}}$ verifies (3.4.9). Consequently, we can define for all $t \in [0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon)$

$$\Lambda_t = \mathcal{Z}_t^1 \mathbf{1}_{\{0 \leq t \leq \zeta_1\}} + \sum_{n \geq 1} \mathcal{Z}_t^{\frac{1}{n+1}} \mathbf{1}_{\left\{\zeta_{\frac{1}{n}} < t \leq \zeta_{\frac{1}{n+1}}\right\}}, \quad (3.4.10)$$

which is a solution to the system of SDEs (J(p,q)) on $[0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon)$ verifying (iii) of Theorem 3.2.3.

Finally, as the solutions to (J(p,q)) are pathwise unique (Lemma 3.3.1), we can apply the Yamada-Watanabe theorem (see for instance [RY99, Theorem 1.7 p.368]) to deduce the existence of strong solutions to the equation.

Since on $\{\zeta_\varepsilon < +\infty\}$ we have $\lambda_{\zeta_\varepsilon}^1 + \lambda_{\zeta_\varepsilon}^2 = 2\lambda_{\zeta_\varepsilon}^1 + \lambda_{\zeta_\varepsilon}^2 - \lambda_{\zeta_\varepsilon}^1 \leq 3\varepsilon$, or $(1 - \lambda_{\zeta_\varepsilon}^n) + (1 - \lambda_{\zeta_\varepsilon}^{n-1}) = 2(1 - \lambda_{\zeta_\varepsilon}^n) + (1 - \lambda_{\zeta_\varepsilon}^{n-1}) - (1 - \lambda_{\zeta_\varepsilon}^n) \leq 3\varepsilon$, we have on $\left\{\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon < +\infty\right\} : \inf_{t \in [0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon)} \lambda_t^1 + \lambda_t^2 = 0$

or $\sup_{t \in [0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon)} \lambda_t^n + \lambda_t^{n-1} = 2$. On $\left\{\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon < +\infty\right\} \cap \left\{\inf_{t \in [0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon)} \lambda_t^1 + \lambda_t^2 = 0\right\}$ we use Proposition

3.2.5 (i) with $k = 2$ to conclude that

$$\liminf_{\varepsilon \rightarrow 0} \left\{t \geq 0 : \mathcal{Z}_t^{1,\varepsilon} \leq \varepsilon \text{ and } \mathcal{Z}_t^{2,\varepsilon} - \mathcal{Z}_t^{1,\varepsilon} \leq \varepsilon\right\} = +\infty \text{ when } p - n + 1 \geq \frac{1}{\beta} - 1.$$

Likewise, on $\left\{\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon < +\infty\right\} \cap \left\{\sup_{t \in [0, \lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon)} \lambda_t^n + \lambda_t^{n-1} = 2\right\}$ we use Proposition 3.2.5 (ii)

with $k = 2$ to conclude that

$$\liminf_{\varepsilon \rightarrow 0} \left\{t \geq 0 : \mathcal{Z}_t^{n,\varepsilon} \geq 1 - \varepsilon \text{ and } \mathcal{Z}_t^{n,\varepsilon} - \mathcal{Z}_t^{n-1,\varepsilon} \leq \varepsilon\right\} = +\infty \text{ when } q - n + 1 \geq \frac{1}{\beta} - 1.$$

Consequently, when $p \wedge q - n + 1 \geq \frac{1}{\beta} - 1$,

$$\lim_{\varepsilon \rightarrow 0} \zeta_\varepsilon = +\infty,$$

which is (i) from Theorem 3.2.3. □

Lemma 3.4.1. *Let us assume $p \geq q$ and $q - n + 1 > 0$. The system of SDEs*

$$\begin{aligned}
d\hat{\lambda}_t^{i,\varepsilon} &= 2\sqrt{\hat{\lambda}_t^{i,\varepsilon}(1-\hat{\lambda}_t^{i,\varepsilon})}dB_t^i \\
&+ \left\{ (1-2\hat{\lambda}_t^{i,\varepsilon}) \left[2-0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sqrt{1-\hat{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sqrt{\hat{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 \right] \right. \\
&+ \left. \beta \left[p - (p+q)\hat{\lambda}_t^{i,\varepsilon} + \sum_{j \neq i} \frac{\hat{\lambda}_t^{i,\varepsilon}(1-\hat{\lambda}_t^{j,\varepsilon}) + \hat{\lambda}_t^{j,\varepsilon}(1-\hat{\lambda}_t^{i,\varepsilon})}{\hat{\lambda}_t^{i,\varepsilon} - \hat{\lambda}_t^{j,\varepsilon}} \right] \right\} dt \\
0 \leq \hat{\lambda}_t^{1,\varepsilon} &< \dots < \hat{\lambda}_t^{n,\varepsilon} \leq 1, \quad \text{a.s., } dt - \text{a.e.}
\end{aligned} \tag{J_\varepsilon}$$

has a global pathwise unique strong solution $(\hat{\lambda}_t^{1,\varepsilon}, \dots, \hat{\lambda}_t^{n,\varepsilon})_t$ starting from any random initial condition $\Lambda_0 = (\lambda_0^1, \dots, \lambda_0^n)$ independent from \mathbf{B} such that $0 \leq \lambda_0^1 \leq \dots \leq \lambda_0^n \leq 1$ a.s. Moreover,

$$\mathbb{P}\{\exists t > 0 : \hat{\lambda}_t^{i,\varepsilon} = \hat{\lambda}_t^{i+1,\varepsilon} \text{ and } \hat{\lambda}_t^{j,\varepsilon} = \hat{\lambda}_t^{j+1,\varepsilon} \text{ for some } 0 \leq i < j \leq n\} = 0, \tag{3.4.11}$$

where by convention $\hat{\lambda}^{0,\varepsilon} \equiv 0$ and $\hat{\lambda}^{n+1,\varepsilon} \equiv 1$.

Proof. Let us consider $\mathcal{F}_t = \sigma\left(\left(\arcsin\left(\sqrt{\lambda_0^1}\right), \dots, \arcsin\left(\sqrt{\lambda_0^n}\right)\right), (\mathbf{B}_s)_{s \leq t}\right)$ and the system of SDEs defined by

$$\begin{aligned}
d\hat{\phi}_t^{\varepsilon,i} &= dB_t^i + \left\{ \frac{\beta(p-q)}{2} \cot \hat{\phi}_t^{\varepsilon,i} \right. \\
&+ \left(\beta(q-n+1) + \left[1 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\cos \hat{\phi}_t^{\varepsilon,i} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sin \hat{\phi}_t^{\varepsilon,i} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 \right] \right) \cot(2\hat{\phi}_t^{\varepsilon,i}) \\
&+ \left. \frac{\beta}{2} \sum_{j \neq i} \left[\cot(\hat{\phi}_t^{\varepsilon,i} + \hat{\phi}_t^{\varepsilon,j}) + \cot(\hat{\phi}_t^{\varepsilon,i} - \hat{\phi}_t^{\varepsilon,j}) \right] \right\} dt \\
0 \leq \hat{\phi}_t^{\varepsilon,1} &< \dots < \hat{\phi}_t^{\varepsilon,n} \leq \frac{\pi}{2}, \quad \text{a.s. } dt - \text{a.e.},
\end{aligned} \tag{J_\varepsilon^\phi}$$

with random initial condition $\left(\arcsin\left(\sqrt{\lambda_0^1}\right), \dots, \arcsin\left(\sqrt{\lambda_0^n}\right)\right)$ such that $0 \leq \arcsin\left(\sqrt{\lambda_0^1}\right) \leq \dots \leq \arcsin\left(\sqrt{\lambda_0^n}\right) \leq \frac{\pi}{2}$.

We are going to apply Cepa's multivoque equations theory ([Cép95]) to conclude that there exists a unique strong solution to $(\hat{J}_\varepsilon^\phi)$.

To do so, we define

$$D = \{0 < \phi^1 < \phi^2 < \dots < \phi^n < \frac{\pi}{2}\}$$

$$\mathcal{V} : (\phi^1, \dots, \phi^n) \in \mathbb{R}^n \rightarrow \begin{cases} -\sum_{i=1}^n \left\{ \frac{\beta(p-q)}{2} \ln |\sin \phi^i| + \beta \frac{q-n+1}{2} \ln |\sin(2\phi^i)| \right. \\ \left. + \frac{\beta}{4} \sum_{j \neq i} (\ln |\sin(\phi^i + \phi^j)| + \ln |\sin(\phi^i - \phi^j)|) \right\} & \text{if } x \in D \\ +\infty & \text{if } x \notin D \end{cases}$$

$$g : (\phi^1, \dots, \phi^n) \in \mathbb{R}^n \rightarrow \left(\left[1 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\cos \phi^1 - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sin \phi^1 - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 \right] \cot(2\phi^1), \right. \\ \left. \dots, \left[1 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\cos \phi^n - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sin \phi^n - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 \right] \cot(2\phi^n) \right),$$

to rewrite the system of SDEs on D with $\hat{\Phi} = (\hat{\phi}_t^{1,\varepsilon}, \dots, \hat{\phi}_t^{n,\varepsilon})_t$ the following way

$$d\hat{\Phi}_t = d\mathbf{B}_t + g(\hat{\Phi}_t)dt - \nabla \mathcal{V}(\hat{\Phi}_t)dt. \quad (\hat{\mathcal{J}}_\varepsilon^\phi)$$

Since g is globally Lipschitz and \mathcal{V} is convex, Cépa's multivoque equations theory shows the existence and uniqueness of a strong solution to equation

$$d\hat{\varphi}_t = d\mathbf{B}_t + g(\hat{\varphi}_t)dt - \nabla \mathcal{V}(\hat{\varphi}_t)dt - \nu(\hat{\varphi}_t)dL_t \text{ for all } t \geq 0 \quad (\hat{\mathcal{J}}_\varepsilon^\varphi)$$

$$\forall t \geq 0, \hat{\varphi}_t \in \bar{D} \text{ a.s.}$$

$$\hat{\varphi}_0 = \left(\arcsin \left(\sqrt{\lambda_0^1} \right), \dots, \arcsin \left(\sqrt{\lambda_0^n} \right) \right),$$

where $\hat{\varphi}$ is a continuous adapted to $(\mathcal{F}_t)_{t \geq 0}$ process, L is a continuous non-decreasing adapted to $(\mathcal{F}_t)_{t \geq 0}$ process with $L_0 = 0$ verifying

$$L_t = \int_0^t \mathbf{1}_{\{\hat{\varphi}_s \in \partial D\}} dL_s,$$

and $\nu(x) \in \pi(x)$ ($\pi(x)$ is the set of unitary outward normals to ∂D at $x \in \partial D$). The solution to equation $(\hat{\mathcal{J}}_\varepsilon^\varphi)$ follows the conditions : for all $t > 0$

$$\mathbb{E} \left[\int_0^t \mathbf{1}_{\{\hat{\varphi}_s \in \partial D\}} ds \right] = 0,$$

$$\mathbb{E} \left[\int_0^t |\nabla \mathcal{V}(\hat{\varphi}_s)| ds \right] < \infty.$$

We apply [CL01, Theorem 2.2] which is an application of Cépa's multivoque equations theory to this kind of SDE and the remark following [Cép95, Theorem 3.1] to deduce that $(\hat{\mathcal{J}}_\varepsilon^\varphi)$ has a unique strong solution. To prove that the boundary process L is equal to zero, we just follow the steps of the proof of [Dem09a, Lemma 2.2], itself coming from the proof of [Dem09b, Lemma 1] which is an adaptation of the proof [CL01, Theorem 2.2, Step 1] and [CL01, Lemma 3.8].

Then, setting with $\hat{\lambda}^{i,\varepsilon} = \sin^2(\varphi^{i,\varepsilon})$ for all $i \in \{1, \dots, n\}$ we obtain a global solution to (\hat{J}_ε) . Following the approach used in the proof of Lemma 3.4.3 below to demonstrate the pathwise uniqueness of a slightly more complicated system (\tilde{J}_ε) , we obtain that the solutions to (\hat{J}_ε) are pathwise unique. The Yamada-Watanabe Theorem (see for instance [RY99, Theorem 1.7 p.368]) allows to conclude that (\hat{J}_ε) has a pathwise unique global strong solution.

Let us now prove (3.4.11).

Let us consider the system of SDEs defined by $(\hat{J}_\varepsilon^\phi)$ with initial condition $0 \leq \arcsin\left(\sqrt{\lambda_0^1}\right) \leq \dots \leq \arcsin\left(\sqrt{\lambda_0^n}\right) \leq \frac{\pi}{2}$.

Let us define for all $\varepsilon > 0$ and for $t \geq 0$: $\Theta(t) = (\theta_1(t), \dots, \theta_n(t))$ with

$$\forall i \in \{1, \dots, n\}, \theta_i(t) = \left[1 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\cos \hat{\phi}_t^{\varepsilon,i} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sin \hat{\phi}_t^{\varepsilon,i} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 \right] \cot(2\hat{\phi}_t^{\varepsilon,i}),$$

and for all $t \geq 0$

$$Z(t) = \exp \left\{ - \int_0^t \Theta(u) \cdot d\mathbf{B}_u - \frac{1}{2} \int_0^t \|\Theta(u)\|^2 du \right\}.$$

We have for all $i \in \{1, \dots, n\}$,

$$\begin{aligned} \theta_i^2(t) &= \left(\left[1 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\cos \hat{\phi}_t^{\varepsilon,i} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sin \hat{\phi}_t^{\varepsilon,i} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 \right] \cot(2\hat{\phi}_t^{\varepsilon,i}) \right)^2 \\ &\leq \left(\frac{1 - \frac{\varepsilon}{4}}{2\sqrt{\frac{\varepsilon}{8}} \left(1 - \frac{\varepsilon}{8} \right)} \right)^2, \end{aligned}$$

where we used the fact that the support of the map

$$x \in [0, 1] \mapsto \left[1 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sqrt{1-x} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sqrt{x} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 \right] \in [0, 1],$$

is $\left[\frac{\varepsilon}{8}, 1 - \frac{\varepsilon}{8} \right]$ and that the map $\phi \mapsto \cot(2\phi) = \frac{1 - \sin^2(\phi)}{2\sqrt{\sin^2(\phi)(1 - \sin^2(\phi))}}$ is monotonous on the interval $\left(0, \frac{\pi}{2}\right)$.

We thus have

$$\mathbb{E} \left[\exp \left\{ \frac{1}{2} \int_0^t \|\Theta(u)\|^2 du \right\} \right] < \infty \text{ for all } t \geq 0.$$

Then, according to Novikov's criterion (see for instance [KS91, Proposition 5.12 p.198]), Z is a \mathbb{P} -martingale, and $\mathbb{E}[Z(t)] = 1$.

Consequently, recalling that $\mathcal{F}_t = \sigma \left(\left(\arcsin \left(\sqrt{\lambda_0^1} \right), \dots, \arcsin \left(\sqrt{\lambda_0^n} \right) \right), (\mathbf{B}_s)_{s \leq t} \right)$ and defining \mathbb{Q} such that

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_t} = Z(t),$$

and for all $i \in \{1, \dots, n\}$,

$$\begin{aligned}\hat{B}_t^i &= B_t^i + \int_0^t \theta_i(s) ds \\ &= B_t^i + \int_0^t \left[1 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\cos \hat{\phi}_s^{\varepsilon,i} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sin \hat{\phi}_s^{\varepsilon,i} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 \right] \cot(2\hat{\phi}_s^{\varepsilon,i}) ds, \quad t \geq 0,\end{aligned}$$

$\hat{\mathbf{B}} = (\hat{B}_t^1, \dots, \hat{B}_t^n)_t$ is a \mathbb{Q} -Brownian motion according to the Girsanov theorem (see for instance [KS91, Proposition 5.4 p.194]).

Thus, $(\hat{J}_\varepsilon^\phi)$ can be rewritten in terms of $\hat{\mathbf{B}}$ as

$$\begin{aligned}d\hat{\phi}_t^{\varepsilon,i} &= d\hat{B}_t^i + \left\{ \frac{\beta(p-q)}{2} \cot \hat{\phi}_t^{\varepsilon,i} + \beta(q-n+1) \cot(2\hat{\phi}_t^{\varepsilon,i}) + \frac{\beta}{2} \sum_{j \neq i} \left[\cot(\hat{\phi}_t^{\varepsilon,i} + \hat{\phi}_t^{\varepsilon,j}) + \cot(\hat{\phi}_t^{\varepsilon,i} - \hat{\phi}_t^{\varepsilon,j}) \right] \right\} dt \\ 0 &\leq \hat{\phi}_t^{\varepsilon,1} < \dots < \hat{\phi}_t^{\varepsilon,n} \leq \frac{\pi}{2}, \quad \text{a.s. } dt - \text{a.e.}\end{aligned}\tag{3.4.12}$$

By the same arguments as in the beginning of the proof, the normally reflected SDE

$$\begin{aligned}d\hat{\varphi}_t &= d\hat{\mathbf{B}}_t - \nabla \mathcal{V}(\hat{\varphi}_t) dt - \nu(\hat{\varphi}_t) d\hat{L}_t \quad \text{for all } t \geq 0 \\ \forall t \geq 0, \hat{\varphi}_t &\in \bar{D} \quad \text{a.s.} \\ \hat{\varphi}_0 &= \left(\arcsin \left(\sqrt{\lambda_0^1} \right), \dots, \arcsin \left(\sqrt{\lambda_0^n} \right) \right),\end{aligned}\tag{3.4.13}$$

admits a global solution and the term $\nu(\hat{\varphi}_t) d\hat{L}_t$ is zero, so that $\hat{\varphi}$ solves (3.4.12). We can apply [Lé10, Theorem 3.1], the proof of which can be applied to our case of SDE (3.4.13) to conclude that its solutions cannot have multiple collisions, which implies that there is no multiple collision of $(\hat{\phi}_t^{1,\varepsilon}, \dots, \hat{\phi}_t^{n,\varepsilon})_t$ under the probability \mathbb{Q} . There is thus no multiple collision of $(\hat{\phi}_t^{1,\varepsilon}, \dots, \hat{\phi}_t^{n,\varepsilon})_{t \geq 0}$ under the probability \mathbb{P} :

$$\mathbb{P}\{\exists t \geq 0 : \hat{\phi}_t^{i,\varepsilon} = \hat{\phi}_t^{i+1,\varepsilon} \text{ and } \hat{\phi}_t^{j,\varepsilon} = \hat{\phi}_t^{j+1,\varepsilon} \text{ for some } 0 \leq i < j \leq n\} = 0.$$

□

Lemma 3.4.2. *Let us assume $p \geq q$ and $q-n+1 > 0$. The system of SDEs (\tilde{J}_ε) with random initial condition $(\tilde{\lambda}_0^{1,\varepsilon}, \dots, \tilde{\lambda}_0^{n,\varepsilon})$ such that $0 \leq \tilde{\lambda}_0^{1,\varepsilon} \leq \dots \leq \tilde{\lambda}_0^{n,\varepsilon} \leq 1$ a.s. and independent from the Brownian motion $\mathbf{B} = (B_t^1, \dots, B_t^n)_t$ has a global pathwise unique strong solution $(\tilde{\lambda}_t^{1,\varepsilon}, \dots, \tilde{\lambda}_t^{n,\varepsilon})_{t \geq 0}$.*

Moreover,

$$\mathbb{P}\{\exists t > 0 : \tilde{\lambda}_t^{i,\varepsilon} = \tilde{\lambda}_t^{i+1,\varepsilon} \text{ and } \tilde{\lambda}_t^{j,\varepsilon} = \tilde{\lambda}_t^{j+1,\varepsilon} \text{ for some } 2 \leq i < j \leq n\} = 0.\tag{3.4.14}$$

Proof. Let us consider a Brownian motion $\tilde{\mathbf{B}} = (\tilde{B}_t^1, \dots, \tilde{B}_t^n)_t$, and the system of SDEs

$$\begin{aligned}
d\tilde{\lambda}_t^{1,\varepsilon} &= 2\sqrt{\tilde{\lambda}_t^{1,\varepsilon}(1-\tilde{\lambda}_t^{1,\varepsilon})}d\tilde{B}_t^1 + \beta \left[p - n + 1 - (p+q)\tilde{\lambda}_t^{1,\varepsilon} \right] dt \\
\forall i \in \{2, \dots, n\}, d\tilde{\lambda}_t^{i,\varepsilon} &= 2\sqrt{\tilde{\lambda}_t^{i,\varepsilon}(1-\tilde{\lambda}_t^{i,\varepsilon})}d\tilde{B}_t^i \\
&+ \left\{ (1-2\tilde{\lambda}_t^{i,\varepsilon}) \left[2 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sqrt{1-\tilde{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 - 0 \vee \frac{2}{\sqrt{\varepsilon}} \left(\sqrt{\tilde{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2} \right) \wedge 1 \right] \right. \\
&+ \left. \beta \left[p - 1 - (p+q)\tilde{\lambda}_t^{i,\varepsilon} + \sum_{j \geq 2, j \neq i} \frac{\tilde{\lambda}_t^{i,\varepsilon}(1-\tilde{\lambda}_t^{j,\varepsilon}) + \tilde{\lambda}_t^{j,\varepsilon}(1-\tilde{\lambda}_t^{i,\varepsilon})}{\tilde{\lambda}_t^{i,\varepsilon} - \tilde{\lambda}_t^{j,\varepsilon}} \right] \right\} dt \\
0 \leq \tilde{\lambda}_t^{1,\varepsilon} \leq 1 \text{ and } 0 \leq \tilde{\lambda}_t^{2,\varepsilon} < \dots < \tilde{\lambda}_t^{n,\varepsilon} \leq 1, \text{ a.s., } dt - a.e.
\end{aligned}$$

Let us remark that $\tilde{\lambda}^1$ is a one dimensional Jacobi process and admits a strong solution defined on \mathbb{R}_+ (see Lemma 3.5.1), and that the coordinates $i \in \{2, \dots, n\}$ satisfy an autonomous system of SDEs for $n-1$ particles similar to equation (\hat{J}_ε) . The only differences are the terms $\left[0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sqrt{\hat{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 \right]$ and βp in (\hat{J}_ε) which become respectively $\left[0 \vee \frac{2}{\sqrt{\varepsilon}} \left(\sqrt{\tilde{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2} \right) \wedge 1 \right]$ and $\beta(p-1)$ here.

We can still consider the process $\left(\arcsin \left(\sqrt{\tilde{\lambda}_t^{2,\varepsilon}} \right), \dots, \arcsin \left(\sqrt{\tilde{\lambda}_t^{n,\varepsilon}} \right) \right)_t$, and apply the same method as in the proof of Lemma 3.4.1 (indeed, if $q-n+1 > 0$, then $q-(n-1)+1 > 0$) to prove the existence of a pathwise unique strong solution to this subsystem. As the equation for $\tilde{\lambda}^{1,\varepsilon}$ also has a strong solution, the whole n-particles system considered here admits a global strong solution.

Let us define for all $\varepsilon > 0$ and for $t \geq 0$: $\Theta(t) = (\theta_1(t), \dots, \theta_n(t))$ with

$$\begin{aligned}
\theta_1(t) &= -\frac{\beta}{\sqrt{\tilde{\lambda}_t^{1,\varepsilon}(1-\tilde{\lambda}_t^{1,\varepsilon})}} \times \frac{(1-\tilde{\lambda}_t^{1,\varepsilon}) \wedge \frac{\varepsilon}{2}}{\frac{\varepsilon}{2}} \sum_{j \neq 1} \frac{\tilde{\lambda}_t^{1,\varepsilon}(1-\tilde{\lambda}_t^{j,\varepsilon})}{(\tilde{\lambda}_t^{j,\varepsilon} - \tilde{\lambda}_t^{1,\varepsilon} \wedge \varepsilon) \vee \varepsilon}, \\
\forall i \in \{2, \dots, n\}, \theta_i(t) &= \frac{\beta}{\sqrt{\tilde{\lambda}_t^{i,\varepsilon}(1-\tilde{\lambda}_t^{i,\varepsilon})}} \times \frac{(1-\tilde{\lambda}_t^{i,\varepsilon}) \wedge \frac{\varepsilon}{2}}{\frac{\varepsilon}{2}} \times \frac{\tilde{\lambda}_t^{i,\varepsilon}(1-\tilde{\lambda}_t^{1,\varepsilon})}{(\tilde{\lambda}_t^{i,\varepsilon} - \tilde{\lambda}_t^{1,\varepsilon}) \vee \varepsilon},
\end{aligned}$$

and for all $t \geq 0$

$$Z(t) = \exp \left\{ \int_0^t \Theta(u) \cdot d\tilde{\mathbf{B}}_u - \frac{1}{2} \int_0^t \|\Theta(u)\|^2 du \right\}.$$

We have

$$\begin{aligned}
\theta_1^2(t) &= \frac{\beta^2}{\tilde{\lambda}_t^{1,\varepsilon}(1-\tilde{\lambda}_t^{1,\varepsilon})} \left(\frac{(1-\tilde{\lambda}_t^{1,\varepsilon}) \wedge \frac{\varepsilon}{2}}{\frac{\varepsilon}{2}} \sum_{j \neq 1} \frac{\tilde{\lambda}_t^{1,\varepsilon}(1-\tilde{\lambda}_t^{j,\varepsilon})}{(\tilde{\lambda}_t^{j,\varepsilon} - \tilde{\lambda}_t^{1,\varepsilon} \wedge \varepsilon) \vee \varepsilon} \right)^2 \\
&\leq \beta^2 \tilde{\lambda}_t^{1,\varepsilon} \frac{(1-\tilde{\lambda}_t^{1,\varepsilon}) \wedge \frac{\varepsilon}{2}}{\frac{\varepsilon^2}{4}} \left(\sum_{j \neq 1} \frac{1-\tilde{\lambda}_t^{j,\varepsilon}}{(\tilde{\lambda}_t^{j,\varepsilon} - \tilde{\lambda}_t^{1,\varepsilon} \wedge \varepsilon) \vee \varepsilon} \right)^2
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{2\beta^2}{\varepsilon} \left(\sum_{j>1} \frac{1}{\varepsilon} \right)^2 \\
&\leq \frac{2(n-1)^2\beta^2}{\varepsilon^3}.
\end{aligned}$$

For all $2 < i \leq n$,

$$\begin{aligned}
\theta_i^2(t) &= \frac{\beta^2}{\tilde{\lambda}_t^{i,\varepsilon}(1-\tilde{\lambda}_t^{i,\varepsilon})} \times \frac{\left((1-\tilde{\lambda}_t^{i,\varepsilon}) \wedge \frac{\varepsilon}{2} \right)^2}{\frac{\varepsilon^2}{4}} \times \frac{\left(\tilde{\lambda}_t^{i,\varepsilon}(1-\tilde{\lambda}_t^{1,\varepsilon}) \right)^2}{\left((\tilde{\lambda}_t^{i,\varepsilon} - \tilde{\lambda}_t^{1,\varepsilon}) \vee \varepsilon \right)^2} \\
&\leq \frac{2\beta^2}{\varepsilon} \times \frac{\tilde{\lambda}_t^{i,\varepsilon} (1-\tilde{\lambda}_t^{1,\varepsilon})^2}{\varepsilon^2} \\
&\leq \frac{2\beta^2}{\varepsilon^3}.
\end{aligned}$$

We thus have

$$\mathbb{E} \left[\exp \left\{ \frac{1}{2} \int_0^t \|\Theta(u)\|^2 du \right\} \right] < \infty \text{ for all } t \geq 0.$$

Then, according to Novikov's criterion (see for instance [KS91, Proposition 5.12 p.198]), Z is a \mathbb{P} -martingale, and $\mathbb{E}[Z(t)] = 1$ for all $t \geq 0$. Consequently, defining for all $t \geq 0$, $\tilde{\mathcal{F}}_t = \sigma \left((\tilde{\mathbf{B}}_t)_{s \leq t}, (\tilde{\lambda}_0^{1,\varepsilon}, \dots, \tilde{\lambda}_0^{n,\varepsilon}) \right)$ and \mathbb{Q} such that

$$\frac{d\mathbb{Q}}{d\mathbb{P} |_{\tilde{\mathcal{F}}_t}} = Z(t),$$

and

$$\begin{aligned}
\tilde{B}_t^1 &= \tilde{B}_t^1 - \int_0^t \theta_1(s) ds \\
&= \tilde{B}_t^1 + \int_0^t \frac{\beta}{\sqrt{\tilde{\lambda}_s^{1,\varepsilon}(1-\tilde{\lambda}_s^{1,\varepsilon})}} \times \frac{(1-\tilde{\lambda}_s^{1,\varepsilon}) \wedge \frac{\varepsilon}{2}}{\frac{\varepsilon}{2}} \sum_{j \neq 1} \frac{\tilde{\lambda}_s^{1,\varepsilon}(1-\tilde{\lambda}_s^{j,\varepsilon})}{(\tilde{\lambda}_s^{j,\varepsilon} - \tilde{\lambda}_s^{1,\varepsilon}) \vee \varepsilon} ds
\end{aligned}$$

$$\begin{aligned}
\text{for all } i \in \{2, \dots, n\}, \tilde{B}_t^i &= \tilde{B}_t^i - \int_0^t \theta_i(s) ds \\
&= \tilde{B}_t^i - \int_0^t \frac{\beta}{\sqrt{\tilde{\lambda}_s^{i,\varepsilon}(1-\tilde{\lambda}_s^{i,\varepsilon})}} \times \frac{(1-\tilde{\lambda}_s^{i,\varepsilon}) \wedge \frac{\varepsilon}{2}}{\frac{\varepsilon}{2}} \times \frac{\tilde{\lambda}_s^{i,\varepsilon}(1-\tilde{\lambda}_s^{1,\varepsilon})}{(\tilde{\lambda}_s^{i,\varepsilon} - \tilde{\lambda}_s^{1,\varepsilon}) \vee \varepsilon} ds, \quad 0 \leq t,
\end{aligned}$$

$\tilde{\mathbf{B}} = (\tilde{B}_t^1, \dots, \tilde{B}_t^n)_t$ is a \mathbb{Q} -Brownian motion according to the Girsanov theorem (see for instance [KS91, Proposition 5.4 p.194]).

Consequently, (\tilde{J}_ε) has a global weak solution.

We now have to prove the pathwise uniqueness of the solutions to (\tilde{J}_ε) . The differences with Lemma 3.3.1 are the terms

$$(1 - 2\tilde{\lambda}_t^{i,\varepsilon}) \left[2 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sqrt{1 - \tilde{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 - 0 \vee \frac{2}{\sqrt{\varepsilon}} \left(\sqrt{\tilde{\lambda}_t^{i,\varepsilon}} - \frac{\sqrt{\varepsilon}}{2} \right) \wedge 1 \right],$$

and the interaction terms between the first particle and the others.

Let $Z = (z_t^1, \dots, z_t^n)_t$ and $\tilde{Z} = (\tilde{z}_t^1, \dots, \tilde{z}_t^n)_t$ be two global solutions to (\tilde{J}_ε) with $Z_0 = \tilde{Z}_0$ independent from the same driving Brownian motion $\mathbf{B} = (B_t^1, \dots, B_t^n)_t$.

The local time of $z^i - \tilde{z}^i$ at 0 is zero ([RY99, Lemma 3.3 p.389]). Applying the Tanaka formula to the process $z^i - \tilde{z}^i$ and summing over i ,

$$\begin{aligned} \sum_{i=1}^n |z_t^i - \tilde{z}_t^i| &= 2 \sum_{i=1}^n \int_0^t \operatorname{sgn}(z_s^i - \tilde{z}_s^i) \left(\sqrt{z_s^i(1-z_s^i)} - \sqrt{\tilde{z}_s^i(1-\tilde{z}_s^i)} \right) dB_s^i \\ &+ 2\beta \int_0^t \sum_{i=2}^n \operatorname{sgn}(z_s^i - \tilde{z}_s^i) \sum_{j \geq 2, j \neq i} \left(\frac{z_s^i(1-z_s^j)}{z_s^i - z_s^j} - \frac{\tilde{z}_s^i(1-\tilde{z}_s^j)}{\tilde{z}_s^i - \tilde{z}_s^j} \right) ds \end{aligned} \quad (3.4.15)$$

$$- \beta(p+q) \int_0^t \sum_{i=1}^n |z_s^i - \tilde{z}_s^i| ds \quad (3.4.16)$$

$$\begin{aligned} &+ 2\beta \int_0^t \sum_{i=2}^n \left\{ \operatorname{sgn}(z_s^i - \tilde{z}_s^i) \left(\frac{(1-z_s^i) \wedge \frac{\varepsilon}{2}}{\frac{\varepsilon}{2}} \cdot \frac{z_s^i(1-z_s^1)}{(z_s^i - z_s^1) \vee \varepsilon} - \frac{(1-\tilde{z}_s^i) \wedge \frac{\varepsilon}{2}}{\frac{\varepsilon}{2}} \cdot \frac{\tilde{z}_s^i(1-\tilde{z}_s^1)}{(\tilde{z}_s^i - \tilde{z}_s^1) \vee \varepsilon} \right) \right. \\ &\left. - \operatorname{sgn}(z_s^1 - \tilde{z}_s^1) \left(\frac{(1-z_s^1) \wedge \frac{\varepsilon}{2}}{\frac{\varepsilon}{2}} \cdot \frac{z_s^1(1-z_s^i)}{(z_s^1 - z_s^i) \wedge \varepsilon} \vee \varepsilon - \frac{(1-\tilde{z}_s^1) \wedge \frac{\varepsilon}{2}}{\frac{\varepsilon}{2}} \cdot \frac{\tilde{z}_s^1(1-\tilde{z}_s^i)}{(\tilde{z}_s^1 - \tilde{z}_s^i) \wedge \varepsilon} \vee \varepsilon \right) \right\} ds \end{aligned} \quad (3.4.17)$$

$$\begin{aligned} &- \int_0^t \sum_{i=2}^n \operatorname{sgn}(z_s^i - \tilde{z}_s^i) \left\{ (1-2z_s^i) \left[2 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sqrt{1-z_s^i} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 - 0 \vee \frac{2}{\sqrt{\varepsilon}} \left(\sqrt{z_s^i} - \frac{\sqrt{\varepsilon}}{2} \right) \wedge 1 \right] \right. \\ &\left. - (1-2\tilde{z}_s^i) \left[2 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sqrt{1-\tilde{z}_s^i} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 - 0 \vee \frac{2}{\sqrt{\varepsilon}} \left(\sqrt{\tilde{z}_s^i} - \frac{\sqrt{\varepsilon}}{2} \right) \wedge 1 \right] \right\} ds. \end{aligned} \quad (3.4.18)$$

As in the proof of Lemma 3.3.1, the expectation of the stochastic integrals is zero and the terms (3.4.15) are not positive (see inequality (3.3.3)). To deal with the expectation of (3.4.17), one remarks that the function

$$(x, y) \mapsto \frac{(1-x) \wedge \frac{\varepsilon}{2}}{\frac{\varepsilon}{2}} \cdot \frac{x(1-y)}{(x-y) \vee \varepsilon}$$

is Lipschitz on $[0, 1]^2$.

As for the term (3.4.18), the function

$$f : z \mapsto (1-2z) \left[2 - 0 \vee \frac{2\sqrt{2}}{\sqrt{\varepsilon}} \left(\sqrt{1-z} - \frac{\sqrt{\varepsilon}}{2\sqrt{2}} \right) \wedge 1 - 0 \vee \frac{2}{\sqrt{\varepsilon}} \left(\sqrt{z} - \frac{\sqrt{\varepsilon}}{2} \right) \wedge 1 \right] \text{ defined on}$$

$[0, 1]$ is continuous and can be rewritten

$$f(z) = \begin{cases} 1 - 2z & \text{if } z \in \left[0, \frac{\varepsilon}{4}\right) \cup \left(1 - \frac{\varepsilon}{8}, 1\right], \\ (1 - 2z) \left(\frac{2\sqrt{z}}{\sqrt{\varepsilon}} - 1\right) & \text{if } z \in \left[\frac{\varepsilon}{4}, \varepsilon\right], \\ (1 - 2z) \left(\frac{2\sqrt{2(1-z)}}{\sqrt{\varepsilon}} - 1\right) & \text{if } z \in \left[1 - \frac{\varepsilon}{2}, 1 - \frac{\varepsilon}{8}\right], \\ 0 & \text{if } z \in \left[\varepsilon, 1 - \frac{\varepsilon}{2}\right], \end{cases}$$

and is thus Lipschitz on $[0, 1]$.

Then there exists a constant $K \geq 0$ such that for all $t \geq 0$

$$\begin{aligned} \sum_{i=1}^n \mathbb{E}|z_t^i - \tilde{z}_t^i| &\leq K \mathbb{E} \left[\int_0^t \sum_{i=1}^n |z_s^i - \tilde{z}_s^i| ds \right] \\ &\leq K \int_0^t \sum_{i=1}^n \mathbb{E}|z_s^i - \tilde{z}_s^i| ds. \end{aligned}$$

The Grönwall Lemma allows to conclude that for all $t \geq 0$

$$\sum_{i=1}^n \mathbb{E}|z_t^i - \tilde{z}_t^i| = 0,$$

which concludes the proof on the existence and pathwise uniqueness.

Let us now prove (3.4.14). To do so, we apply to the process $\left(\arcsin\left(\sqrt{\tilde{\lambda}_t^{2,\varepsilon}}\right), \dots, \arcsin\left(\sqrt{\tilde{\lambda}_t^{n,\varepsilon}}\right)\right)_t$ the method used to prove (3.4.11). Indeed, this process solves, as explained in the beginning of the proof, a system of SDEs similar to (\hat{J}_ε) for $n - 1$ particles. \square

Lemma 3.4.3. *Let us assume $p \geq q$ and $q - n + 1 > 0$. The system of SDEs (\check{J}_ε) with random initial condition $(\check{\lambda}_0^{1,\varepsilon}, \dots, \check{\lambda}_0^{n,\varepsilon})$ such that $0 \leq \check{\lambda}_0^{1,\varepsilon} \leq \dots \leq \check{\lambda}_0^{n,\varepsilon} \leq 1$ a.s. and independent from the Brownian motion $\mathbf{B} = (B_t^1, \dots, B_t^n)_t$ has a global pathwise unique strong solution $(\check{\lambda}_t^{1,\varepsilon}, \dots, \check{\lambda}_t^{n,\varepsilon})_{t \geq 0}$.*

Moreover,

$$\mathbb{P}\{\exists t > 0 : \check{\lambda}_t^{i,\varepsilon} = \check{\lambda}_t^{i+1,\varepsilon} \text{ and } \check{\lambda}_t^{j,\varepsilon} = \check{\lambda}_t^{j+1,\varepsilon} \text{ for some } 0 \leq i < j \leq n - 2\} = 0.$$

Proof. The proof of this Lemma follows the same steps as the proof of Lemma 3.4.2 interchanging the role of the particles indexed by 1 and n . \square

Lemma 3.4.4. *Let us assume $p \geq q$ and $q - n + 1 > 0$. The system of SDEs (\bar{J}_ε) with random initial condition $(\bar{\lambda}_0^{1,\varepsilon}, \dots, \bar{\lambda}_0^{n,\varepsilon})$ such that $0 \leq \bar{\lambda}_0^{1,\varepsilon} \leq \dots \leq \bar{\lambda}_0^{n,\varepsilon} \leq 1$ a.s. and independent from the Brownian motion $\mathbf{B} = (B_t^1, \dots, B_t^n)_t$ has a global pathwise unique strong solution $(\bar{\lambda}_t^{1,\varepsilon}, \dots, \bar{\lambda}_t^{n,\varepsilon})_{t \geq 0}$.*

Moreover,

$$\mathbb{P}\{\exists t > 0 : \bar{\lambda}_t^{i,\varepsilon} = \bar{\lambda}_t^{i+1,\varepsilon} \text{ and } \bar{\lambda}_t^{j,\varepsilon} = \bar{\lambda}_t^{j+1,\varepsilon} \text{ for some } 2 \leq i < j \leq n - 2\} = 0.$$

Proof. The proof of this Lemma follows the same steps as in Lemma 3.4.2, but with both $\bar{\lambda}^{1,\varepsilon}$ and $\bar{\lambda}^{n,\varepsilon}$ playing a peculiar role like $\tilde{\lambda}^{1,\varepsilon}$ in (\tilde{J}_ε) and $\check{\lambda}^{n,\varepsilon}$ in (\check{J}_ε) . \square

3.5 Appendix

This Lemma deals with the existence and uniqueness to the real Jacobi process and was proved in [Dou05].

Lemma 3.5.1 (Real Jacobi Processes). *The real Jacobi process of parameters $d, d' \geq 0$ is the unique strong solution to*

$$dJ_t = 2\sqrt{J_t(1 - J_t)}dB_t + (d - (d + d')J_t)dt$$

where B is a real Brownian motion. The process J remains in $(0, 1)$ a.s. when $d \wedge d' \geq 2$, hits 0 a.s. if $0 \leq d < 2$ and hits 1 a.s. if $0 \leq d' < 2$.

The next lemma deals with the existence and uniqueness to the CIR SDE and with the probability for the solution to hit zero. It is proved for instance in [LL08, Theorem 6.2.2 and Proposition 6.2.3].

Lemma 3.5.2. *Let $a \geq 0, b, \sigma \in \mathbb{R}$. Suppose that W is a standard Brownian motion defined on \mathbb{R}_+ . For any real number $x \geq 0$, there is a unique continuous, adapted process X , taking values in \mathbb{R}_+ , satisfying $X_0 = x$ and*

$$dX_t = (a - bX_t)dt + \sigma\sqrt{X_t}dW_t \text{ on } [0, \infty).$$

Moreover, if we denote by X^x the solution to this SDE starting at x and by $\tau_0^x = \inf\{t \geq 0 : X_t^x = 0\}$,

1. If $a \geq \sigma^2/2$, we have $\mathbb{P}(\tau_0^x = \infty) = 1$, for all $x > 0$.
2. If $0 \leq a < \sigma^2/2$ and $b \geq 0$, we have $\mathbb{P}(\tau_0^x < \infty) = 1$, for all $x > 0$.
3. If $0 \leq a < \sigma^2/2$ and $b < 0$, we have $0 < \mathbb{P}(\tau_0^x < \infty) < 1$, for all $x > 0$.

The following result is the Ikeda-Watanabe Theorem, which allows to compare two Itô processes if their starting points and their drift coefficients are comparable, and if their diffusion coefficients are regular enough. It is proved for instance in [RW00, Theorem V.43.1 p.269].

Theorem 3.5.3. (Ikeda-Watanabe) Suppose that, for $i = 1, 2$,

$$X_t^i = X_0^i + \int_0^t \sigma(X_s^i) dB_s + \int_0^t \beta_s^i ds, \quad (3.5.1)$$

and that there exist $b : \mathbb{R} \mapsto \mathbb{R}$, such that

$$\beta_s^1 \geq b(X_s^1), \quad b(X_s^2) \geq \beta_s^2.$$

Suppose also that

1. σ is measurable and there exists an increasing function $\rho : \mathbb{R}_+ \mapsto \mathbb{R}_+$ such that

$$\int_{0^+} \rho(u)^{-1} du = \infty,$$

and for all $x, y \in \mathbb{R}$,

$$(\sigma(x) - \sigma(y))^2 \leq \rho(|x - y|);$$

2. $X_0^1 \geq X_0^2$ a.s.;
3. b is Lipschitz.

Then $X_t^1 \geq X_t^2$ for all t a.s.

Chapter 4

Wishart processes : mean-field limit, long time behavior, and free probability

This Section an article submitted to *Random Matrices : Theory and Applications* that one can find at the address <https://arxiv.org/abs/2009.09874>.

Abstract

This paper is devoted to the study of the eigenvalues of the Wishart process which are the analog of the Dyson Brownian Motion for covariance matrices. Such processes were in particular studied by Bru. The mean field convergence of the empirical measure of these eigenvalues was proved Malecki and Perez. In this paper, we provide a new approach to the mean field convergence problem using tools from the free rectangular convolution theory developed by Benaych-Georges, which in particular allows to compute explicitly the limit measure valued flow. We highlight the link with the integro-differential equation related to the mean field limit and its translation into a complex Burgers partial differential equation.

Keywords: stochastic differential equations, random matrices, free probability theory.

AMS Subject Classification (2020): 60B20, 60G07, 46L54, 60F05, 60H15.

4.1 Introduction

We will denote by $\mathbb{N} = \{0, 1, 2, \dots\}$ the set of non-negative integers, and by $\mathbb{N}^* = \{1, 2, 3, \dots\}$ the set of positive integers. Let $n, m \in \mathbb{N}^*$ such that $n \leq m$ and let $(M_t)_t$ be a stochastic process taking its values in the space of $n \times m$ matrices with real entries verifying the following stochastic differential equation (SDE)

$$dM_t = \kappa dW_t - \gamma M_t dt, \tag{4.1.1}$$

where W is a $n \times m$ matrix filled with independent Brownian motions, $\gamma \geq 0$ and $\kappa \geq 0$. M is thus a matrix whose entries are independent Ornstein-Uhlenbeck processes just as the one considered in [Bru91] by Bru. The reader will find in [KO01] an analysis of the complex analog of Bru's model.

For all $t \geq 0$ let $(x_t^{1,n}, \dots, x_t^{n,n})$ and $(\lambda_t^{1,n}, \dots, \lambda_t^{n,n})$ be respectively the eigenvalues of the $n \times n$ square matrices $\sqrt{\frac{1}{m}M_tM_t^*}$ and $\frac{1}{m}M_tM_t^*$ where $*$ is the transposition operator.

It is proved in [Bru89] and [Bru91] that the eigenvalues $(\lambda_t^{1,n}, \dots, \lambda_t^{n,n})$ satisfy the system of SDEs

$$d\lambda_t^{i,n} = \frac{2\kappa}{\sqrt{m}}\sqrt{\lambda_t^{i,n}}dB_t^i + \kappa^2 dt - 2\gamma\lambda_t^{i,n}dt + \frac{\kappa^2}{m}\sum_{j \neq i} \frac{\lambda_t^{i,n} + \lambda_t^{j,n}}{\lambda_t^{i,n} - \lambda_t^{j,n}}dt \text{ for all } i \in \{1, \dots, n\} \quad (4.1.2)$$

$$0 \leq \lambda_t^{1,n} < \dots < \lambda_t^{n,n} \text{ a.s. dt-a.e.,}$$

where B^1, \dots, B^n are independent Brownian motions. It is proved for instance in [GM14] that the SDE (4.1.2) admits a strong pathwise unique solution.

If $(\lambda_t^{1,n}, \dots, \lambda_t^{n,n})_t$ is solution to the SDE (4.1.2), then a formal calculus gives for $x^{i,n} = \sqrt{\lambda^{i,n}}$:

$$dx_t^{i,n} = \frac{\kappa}{\sqrt{m}}dB_t^i + \left(\left(1 - \frac{1}{m}\right) \frac{\kappa^2}{2x_t^{i,n}} - \gamma x_t^{i,n} + \frac{\kappa^2}{2mx_t^{i,n}} \sum_{j \neq i} \frac{(x_t^{i,n})^2 + (x_t^{j,n})^2}{(x_t^{i,n})^2 - (x_t^{j,n})^2} \right) dt$$

$$0 \leq x_t^{1,n} < \dots < x_t^{n,n} \text{ a.s. dt-a.e.}$$

We respectively define the empirical measure of the eigenvalues of $\frac{1}{m}M_tM_t^*$ and the symmetrized empirical measure of the eigenvalues of $\sqrt{\frac{1}{m}M_tM_t^*}$ by

$$\nu_t^n = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_t^{i,n}} \quad \text{and} \quad \mu_t^n = \sqrt{\nu_t^n} = \frac{1}{2n} \sum_{i=1}^n (\delta_{x_t^{i,n}} + \delta_{-x_t^{i,n}}),$$

where here and in the rest of the paper, for any probability measure μ on $[0, +\infty)$, $\sqrt{\mu} = \text{sym}(\sqrt{\cdot} \# \mu)$ denotes the symmetrization of the push-forward of μ by the map $v \mapsto \sqrt{v}$. The application sym is fully defined in Section 4.4. These measure sequences are random variables sequences in the space of probability measures on respectively \mathbb{R}_+ and \mathbb{R} . We will thus speak about weak convergence in probability or almost surely throughout the paper in the sense that the sequence converges in probability or almost surely in the space of probability measures on \mathbb{R}_+ (or \mathbb{R}) with the distance of Prokhorov (see for instance [Bil13]). Let us note that by non negativness of the $(\lambda_t^{1,n}, \dots, \lambda_t^{n,n})_{t \geq 0}$, symmetry of μ_t^n and continuity of the functions $v \mapsto v^2$ and $v \mapsto \sqrt{v}$, there is equivalence between the convergence of $(\nu^n)_n$ and the convergence of $(\mu^n)_n$.

In this note, we are interested in the convergence and limit of these measure processes when n and m go to infinity at a rate $n/m \rightarrow \alpha \in (0, 1]$.

Our main result gives a complete answer to the limit of $(\mu^n)_n$ problem, using free probability tools (in particular the rectangular free convolution of index α denoted by \boxplus_α defined

in Section 4.4). Let us first denote by $\mu^{\text{MP}_{\rho,\sigma}}$ the Marcenko-Pastur distribution with shape parameter ρ and scale parameter σ which admits the density with respect to the Lebesgue measure :

$$x \mapsto \frac{\sqrt{(a_+ - x)(x - a_-)}}{2\pi\rho x\sigma^2} \mathbb{1}_{[a_-, a_+]} \text{ where } a_{\pm} = \sigma^2(1 \pm \sqrt{\rho})^2.$$

Theorem 4.1.1 (Mean-field limit). *Let us assume $n \leq m$, and that when n grows to infinity, m grows to infinity too, with $\frac{n}{m} \rightarrow \alpha \in (0, 1]$. Let us also assume that $(\mu_0^n)_n$ converges weakly in probability towards a non random probability measure μ_0 . Then for all $t \geq 0$, $(\mu_t^n)_{n \geq 0}$ converges weakly in probability and*

$$\mu_t := \lim_{n \rightarrow \infty} \mu_t^n = (e^{-\gamma t} \mu_0) \boxplus_{\alpha} \sqrt{\mu^{\text{MP}_{\alpha, \sigma_t}}},$$

where \boxplus_{α} is the rectangular free convolution of parameter α , $e^{-\gamma t} \mu_0$ denotes the push-forward of μ_0 by the map $v \mapsto e^{-\gamma t} v$, and where for all $t \geq 0$:

$$\sigma_t^2 = \begin{cases} \frac{\kappa^2}{2\gamma}(1 - e^{-2\gamma t}) & \text{if } \gamma \neq 0 \\ \kappa^2 t & \text{if } \gamma = 0. \end{cases}$$

Moreover, if $\gamma \neq 0$ and with $\sigma_{\infty} = \sqrt{\frac{\kappa^2}{2\gamma}}$, we have :

$$\lim_{t \rightarrow +\infty} \mu_t = \sqrt{\mu^{\text{MP}_{\alpha, \sigma_{\infty}}}}.$$

For $\gamma = 0$ and $\kappa = 1$, this measure valued flow $(\mu_t)_t$ can be seen as the law of the free Wishart process introduced in [CDM05]. The reader will find in [Bia97, BS98, BS01] an introduction to the free stochastic calculus theory.

This last result is in fact true in a complex framework (with complex Brownian motions of the form $(B_t^1 + iB_t^2)_{t \geq 0}$ where $(B_t^1)_{t \geq 0}$ and $(B_t^2)_{t \geq 0}$ are independent real Brownian motions), as the reader will see in Section 4.2. This result might also be extended to a quaternionic framework but the rectangular free convolution theory is not developed yet in this case. We moreover proved the following result.

Theorem 4.1.2 (Commutativity of the limits). *Under the assumption of Theorem 4.1.1, let us moreover suppose $\gamma \neq 0$ and that for all $n \in \mathbb{N}^*$: $\sum_{i=1}^n \mathbb{E}[\lambda_0^{i,n}] < +\infty$. Let us note $\sigma_{\infty} = \sqrt{\frac{\kappa^2}{2\gamma}}$. Then we have in the sense of weak convergence*

$$\lim_{\substack{n \rightarrow \infty \\ n/m \rightarrow \alpha}} \lim_{t \rightarrow \infty} \nu_t^n = \lim_{t \rightarrow \infty} \lim_{\substack{n \rightarrow \infty \\ n/m \rightarrow \alpha}} \nu_t^n = \mu^{\text{MP}_{\alpha, \sigma_{\infty}}}.$$

The SDE (4.1.2) is related to the β -Wishart (or β -Laguerre) process, the system of particles defined by the SDE

$$d\lambda_t^{i,n} = 2\sqrt{\lambda_t^{i,n}} dB_t^i - 2\eta\lambda_t^{i,n} dt + \delta dt + \beta \sum_{j \neq i} \frac{\lambda_t^{i,n} + \lambda_t^{j,n}}{\lambda_t^{i,n} - \lambda_t^{j,n}} dt \text{ for all } i \in \{1, \dots, n\}$$

$$0 \leq \lambda_t^{1,n} < \dots < \lambda_t^{n,n} \text{ a.s. dt-a.e.}$$

with $\beta, \delta, \eta > 0$ and which, for $\beta = 1$, $\eta = m\gamma/\kappa^2$, $\delta = m$, and after the change of time $t \rightarrow \frac{\kappa^2}{m}t$, corresponds to the dynamics of the process followed by the eigenvalues of $\frac{1}{m}M_t M_t^*$. Sometimes in the literature, the β -Wishart process is considered with $\eta = 0$ (see for instance [Dem09b]) or with $\eta = 1/2$ (see for instance [TT20b]). Sometimes, the β -Wishart process is also found in the literature as the system of particles given by the SDE

$$d\lambda_t^{i,n} = 2\sqrt{\lambda_t^{i,n}} dB_t^i - 2\eta\lambda_t^{i,n} dt + \beta \left(\delta + \sum_{j \neq i} \frac{\lambda_t^{i,n} + \lambda_t^{j,n}}{\lambda_t^{i,n} - \lambda_t^{j,n}} \right) dt \text{ for all } i \in \{1, \dots, n\}$$

$$0 \leq \lambda_t^{1,n} < \dots < \lambda_t^{n,n} \text{ a.s. dt-a.e.,}$$

as for instance in [ABMV13].

The reader will find in Section 4.3 an extension of Theorems 4.1.1 and 4.1.2 to the β -Wishart case.

Similar problems were tackled before, often by the direct study of the limit of empirical measure processes of a system of interacting diffusive particles. In [RS93], the authors studied the empirical measure of diffusing particles with electrostatic repulsion and a negative linear elastic force. They proved that the accumulation points of the sequence of empirical measures of the particles sequence (the sequence being indexed by the number of particles) satisfy an integro-differential equation. The method of their proof was used in other papers and is reproduced here in our context. They proved the uniqueness of the limit by translating the integro-differential equation problem into a complex partial differential equation (PDE) problem, a Burgers equation, which they could solve using the method of characteristics.

In [CL97] was studied a more general system of diffusing particles with electrostatic repulsion, a linear elastic force, and an additional constant drift. In this case, an integro-differential equation for the limit of the empirical measure of the particles was found and translated into a Burgers-like complex PDE using the methods developed in [RS93]. However, the existence and uniqueness of the solution to the PDE were proved using Fourier transforms methods. A second article from the same authors [CL01] followed, studying Brownian particles with electrostatic repulsion on the circle. There, the same procedure allowed the authors to get another Burgers-like PDE of order two, the existence and uniqueness of a solution to which were this time proved using the Hopf-Cole transformation. This allowed the authors to find an explicit expression of the limit measure flow when each particle starts from the position 0, i.e. when the initial condition of the empirical measure is δ_0 .

More recently, these questions regained in interest and two articles, [SY19] and [MP19] tackled more general problems. They got interested into the empirical (spectral) measure of the first introduced in [GM13] general matrix valued stochastic differential equations on \mathcal{H}_n , the space of Hermitian $n \times n$ matrices, of the form

$$dX_t = g(X_t)dW_t h(X_t) + h(X_t)d(W_t)^\dagger g(X_t) + \frac{1}{n}b_n(X_t)dt, \quad X_0 \in \mathcal{H}_n,$$

where \dagger is the conjugate transpose operator, and where the continuous functions $g, h, b_n : \mathbb{R} \rightarrow \mathbb{R}$ act spectrally on X_t , i.e. if $X \in \mathcal{H}_n$, U unitary and D diagonal are such that $X = U^\dagger D U$, then g is identified with the map $\mathcal{H}_n \ni X \mapsto U^\dagger \text{Diag}(g(D_{1,1}), \dots, g(D_{n,n})) U$. Here $W = (W_t)_t$ stands for a $n \times n$ complex-valued Brownian motion, i.e. the matrix valued process with entries being independent one-dimensional complex-valued Brownian motions. This general formulation encompasses the β -Wishart (or β -Laguerre) model and the β -Jacobi model, as it is stated in [GM13]. After proving the tightness of the family of measure valued processes, these two papers use a [RS93]-like method to get the integro-differential equation that is satisfied by the accumulation points of the measure valued processes family. In [MP19], the more general of the two articles, the authors derive existence and uniqueness of the solution to the integro-differential equation, and thus the convergence of the empirical measure valued process when $n \rightarrow \infty$ using the fact that the moments of the limit measure valued flow are uniquely determined by the integro-differential equation. Moreover, in some peculiar cases, they give an explicit expression of the limit, but only for the initial condition $X_0 = 0$. In [SY19], rather than using a moment approach, they translate as in [RS93] the integro-differential equation problem into a Burgers equation, which they do not solve in general, but only for the initial condition $X_0 = 0$, using scaling properties of the initial SDE problem. For this initial condition, the Marcenko-Pastur distribution appears naturally in the expression of the measure valued flow limit.

The β -Wishart (or β -Laguerre) problem was also tackled in [ABMV13]. They compute there in the high temperature regime (when $\beta n \xrightarrow{n \rightarrow \infty} 2c \in (0, \infty)$) the limit integro-differential equation, parametrized by c , the stationary probability measure of this equation, and remark that this limit distribution interpolates the Marcenko-Pastur distribution on the $c \rightarrow +\infty$ limit and the Gamma distribution on the $c \rightarrow 0$ limit.

In [TT20b] and [TT20a] were studied the same questions in the peculiar cases of respectively β -Wishart (or β -Laguerre) and β -Jacobi processes using a moment based method as in [MP19] to prove the uniqueness the limit measure process. The authors moreover compute, in the high temperature regime (when $\beta n \xrightarrow{n \rightarrow \infty} 2c \in (0, \infty)$), the long-time behaviour of the limit measure flow for both process families.

The interest of this note is to understand the limit behaviour of the empirical spectral measure of $\frac{1}{m} M_t M_t^*$ and $\sqrt{\frac{1}{m} M_t M_t^*}$ both in a real and a complex framework, using rectangular free probability tools. This point of view, more random matrix theory and free probability oriented, sheds a new light on the problem of the limit behaviour of the empirical measure process of β -Wishart particles for all $\beta > 0$, and allows to compute the limit measure valued flow for any initial condition, and to recover the long-time behaviour of the limit measure valued flow. For the sake of completeness, we also derive the integro-differential equation and the complex PDE approach of the problem.

The note is organized the following way. After the Introduction where the main results are given, the complex version of these results are given in Section 2. The reader will find in Section 3 an application of Theorem 4.1.1 to the general β -Wishart processes. Preliminary

results about the Cauchy-Stieltjes transform and about the free and the rectangular free convolution are given in Section 4, and the results are proven in Section 5.

Acknowledgement : I thank Djalil Chafai and Benjamin Jourdain for numerous fruitful discussions.

4.2 The complex framework

Theorem 4.1.1 is in fact true in a complex framework, i.e. for $(M_t)_{t \geq 0}$ being a complex valued matrix process, following the SDE

$$dM_t = \kappa dW_t - \gamma M_t dt,$$

with W a $n \times m$ matrix filled with independent complex Brownian motions (of the form $(B_t^1 + iB_t^2)_{t \geq 0}$ where $(B_t^1)_{t \geq 0}$ and $(B_t^2)_{t \geq 0}$ are independent real Brownian motions), M_0 a complex valued random matrix. In this Section, we keep the definitions of the $(\lambda_t^{1,n}, \dots, \lambda_t^{n,n})_t$ and $(x_t^{1,n}, \dots, x_t^{n,n})_t$ as the eigenvalues of respectively $\frac{1}{m} M_t M_t^*$ and $\sqrt{\frac{1}{m}} M_t M_t^*$ but with the symbol $*$ denoting the conjugate transpose rather than the transpose operator.

Theorem 4.2.1 (Complex case). *Theorem 4.1.1 applies in this framework with for all $t \geq 0$:*

$$\sigma_t^2 = \begin{cases} \frac{\kappa^2}{\gamma} (1 - e^{-2\gamma t}) & \text{if } \gamma \neq 0 \\ 2\kappa^2 t & \text{if } \gamma = 0. \end{cases}$$

and with $\sigma_\infty = \sqrt{\frac{\kappa^2}{\gamma}}$.

Let us note that in this context, the eigenvalues $(\lambda_t^{1,n}, \dots, \lambda_t^{n,n})_t$ follow the SDE (see [KO01]) :

$$d\lambda_t^{i,n} = \frac{2\kappa}{\sqrt{m}} \sqrt{\lambda_t^{i,n}} dB_t^i + 2\kappa^2 dt - 2\gamma \lambda_t^{i,n} dt + \frac{2\kappa^2}{m} \sum_{j \neq i} \frac{\lambda_t^{i,n} + \lambda_t^{j,n}}{\lambda_t^{i,n} - \lambda_t^{j,n}} dt \text{ for all } i \in \{1, \dots, n\}$$

$$0 \leq \lambda_t^{1,n} < \dots < \lambda_t^{n,n} \text{ a.s. dt-a.e.,}$$

where B^1, \dots, B^n are independent Brownian motions. It corresponds to a β -Wishart process with $\beta = 2$, or to a $W(2, 1)$ process, which is defined in the next Section.

4.3 Application to more general Wishart processes

For $\beta_1, \beta_2 > 0$, let us consider the changed of time *general Wishart process* defined by the SDE :

$$d\lambda_t^{i,n,W} = \frac{2\kappa}{\sqrt{m}} \sqrt{\lambda_t^{i,n,W}} dB_t^i - 2\gamma \lambda_t^{i,n,W} dt + \beta_1 \kappa^2 \left(1 + \frac{\beta_2}{m} \sum_{j \neq i} \frac{\lambda_t^{i,n,W} + \lambda_t^{j,n,W}}{\lambda_t^{i,n,W} - \lambda_t^{j,n,W}} \right) dt \text{ for all } i \in \{1, \dots, n\}$$

$$0 \leq \lambda_t^{1,n,W} < \dots < \lambda_t^{n,n,W} \text{ a.s. dt-a.e.,} \quad (4.3.1)$$

where B^1, \dots, B^n are independent real Brownian motions. We will refer to this SDE by $W(\beta_1, \beta_2)$. As stated in the introduction, the classical β -Wishart processes found in the literature are of the form $W(1, \beta)$ or $W(\beta, 1)$ (after the change of variables $t \mapsto \frac{m}{\kappa^2} t$). We also remark that the SDE (4.1.2) corresponds to $W(1, 1)$. Results from [GM14] and [JK20] together show (after applying the change of variables $t \mapsto \frac{m}{\kappa^2} t$ to the SDE (4.3.1)) that this SDE admits a strong pathwise unique solution defined on \mathbb{R}_+ as soon as $m - (n - 1)\beta_2 > 0$ and

- $\beta_1 \beta_2 \geq 1$;
- or if $0 < \beta_1 \beta_2 < 1$ when $m\beta_1 + (2 - n)\beta_1 \beta_2 \geq 1$.

Let us note that the long time behaviour of such a process was studied before : the change of variables $t \mapsto \frac{m}{\kappa^2} t$ in the SDE (4.3.1) coupled with the results [JK20, Lemma 3.1 and Proposition 2.8] prove that, given integrable initial conditions, i.e. $\sum_{i=1}^n \mathbb{E}[\lambda_0^{i,n,W}] < +\infty$, the distribution of $(\lambda_t^{1,n,W}, \dots, \lambda_t^{n,n,W})$ converges weakly when $t \rightarrow +\infty$ to a unique stationary probability measure with density with respect to the Lebesgue measure

$$(\lambda^1, \dots, \lambda^n) \mapsto \frac{1}{\mathcal{Z}} \prod_{i=1}^n \left((\lambda^i)^{\frac{\beta_1 m - (n-1)\beta_1 \beta_2}{2} - 1} e^{-\frac{m\gamma}{\kappa^2} \lambda^i} \prod_{j \neq i} |\lambda^j - \lambda^i|^{\beta_1 \beta_2 / 2} \right) \mathbf{1}_{0 \leq \lambda^1 \leq \dots \leq \lambda^n}, \quad (4.3.2)$$

where \mathcal{Z} is a normalizing constant. For $\beta_2 = 1$, this Gibbs measure is the β -Laguerre ensemble, on real symmetric matrices for $\beta_1 = 1$, on complex hermitian matrices for $\beta_1 = 2$ and on quaternion self-dual matrices for $\beta_1 = 4$, see for instance [For10]. This distribution is moreover related to the distribution of the singular values of $n \times m$ random matrix with independent identically distributed real (for $\beta_1 = 1$) or complex (for $\beta_2 = 2$) Gaussian entries.

It is a well known fact that, for $\beta_2 = 1$, if $(\lambda^1, \dots, \lambda^n)$ is a random vector distributed according to (4.3.2), then

$$\nu^{n,W} = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda^i} \xrightarrow[n/m \rightarrow \alpha]{n \rightarrow \infty} \mu^{MP_{\alpha, \sigma_\infty}}, \quad (4.3.3)$$

with $\sigma_\infty = \sqrt{\frac{\beta_1 \kappa^2}{2\gamma}}$, see for instance [HP00, Theorem 5.5.7].

Let us define for all $t \geq 0$:

$$\nu_t^{n,W} = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_t^{i,n,W}}, \quad \mu_t^{n,W} = \sqrt{\nu_t^{n,W}} = \frac{1}{2n} \sum_{i=1}^n \left(\delta_{\sqrt{\lambda_t^{i,n,W}}} + \delta_{-\sqrt{\lambda_t^{i,n,W}}} \right).$$

The next result tackles the limit of the sequence $(\nu_t^{n,W})_n$. As written in the introduction, it is a peculiar case of the results of [MP19] reproduced here partly for the sake of completeness. The limit of the empirical measure problem is transformed into a complex PDE problem, a complex Burgers equation, to which the Cauchy-Stieltjes transform of the limit measure process, if it exists, must be solution.

Definitions and properties about the Cauchy-Stieltjes transform are given in Section 3.

Theorem 4.3.1 (Complex Burgers and mean field limit : weak formulation). *Let us assume that the measure valued sequence $(\nu_0^{n,W})_n$ converges weakly in probability when $n \rightarrow +\infty$ with $\frac{n}{m} \rightarrow \alpha \in (0, 1]$ to a limit probability measure denoted by ν_0^W and that*

$$\sup_n \int x^8 \nu_0^{n,W}(dx) < \infty.$$

Then the family $\{(\nu_t^{n,W})_{t \geq 0}, n \in \mathbb{N}^\}$ is tight and any limiting measure valued flow when n goes to infinity with $\frac{n}{m} \rightarrow \alpha \in (0, 1]$ satisfies for all twice continuously differentiable real test function f the equation :*

$$\langle \nu_t, f \rangle = \langle \nu_0, f \rangle + \int_0^t \langle \nu_s, (\beta_1 \kappa^2 - 2\gamma \Phi) f' \rangle ds + \frac{\alpha \beta_1 \beta_2 \kappa^2}{2} \int_0^t \left(\iint (x+y) \frac{f'(x) - f'(y)}{x-y} \nu_s(dx) \nu_s(dy) \right) ds \quad (4.3.4)$$

where $\Phi : x \mapsto x$ and with the convention $\frac{f'(x) - f'(y)}{x-y} = f''(x)$ when $x = y$. If ν_0 admits a characteristic function, and if this function is analytic on a neighborhood of the origin, then this equation admits a unique solution.

Moreover, if $(\nu_t)_{t \geq 0}$ is a solution to equation (4.3.4) and if for all $t \geq 0$, G_t is the Cauchy-Stieltjes transform of ν_t , then G satisfies the complex Burgers PDE

$$\begin{aligned} \frac{\partial}{\partial t} G_t(z) &= (\alpha \beta_1 \beta_2 \kappa^2 - \beta_1 \kappa^2 + 2\gamma z) \frac{\partial}{\partial z} G_t(z) - 2\alpha \beta_1 \beta_2 \kappa^2 z G_t(z) \frac{\partial}{\partial z} G_t(z) - \alpha \beta_1 \beta_2 \kappa^2 G_t^2(z) + 2\gamma G_t(z), \\ G_0(z) &= \int \frac{\nu_0^W(dv)}{z-v} = \phi(z). \end{aligned} \quad (4.3.5)$$

Remark 4.3.2 (Moments dynamics). Let $\nu = (\nu_t)_t$ be a real measure valued flow verifying equation (4.3.4), and let us denote for all $k \in \mathbb{N}^*, t \geq 0$: $m_t^k = \mathbb{E}[\nu_t^k]$. Then for all $t \geq 0$:

$$(i) \quad m_t^1 = \frac{\beta_1 \kappa^2}{2\gamma} + \left(m_0^1 - \frac{\beta_1 \kappa^2}{2\gamma} \right) e^{-2\gamma t};$$

$$(ii) \quad m_t^2 = \frac{1}{4\gamma^2} \left[(\beta_1^2 \kappa^4 + (\alpha \beta_1 \beta_2 \kappa^2 - 4\gamma m_0^1) \beta_1 \kappa^2 + 4\gamma^2 m_0^2 - 4\alpha \beta_1 \beta_2 \kappa^2 \gamma m_0^1) e^{-4\gamma t} + 4\beta_1 \kappa^2 (\beta_2 \alpha + 1) \left(\left(\gamma m_0^1 - \frac{\beta_1 \kappa^2}{2} \right) e^{-2\gamma t} + \frac{\beta_1 \kappa^2}{4} \right) \right];$$

$$(iii) \quad \forall k \geq 3, \quad \frac{d}{dt} m_t^k = k \beta_1 \kappa^2 m_t^{k-1} - 2k\gamma m_t^k + \frac{k\alpha \beta_1 \beta_2 \kappa^2}{2} \sum_{j=0}^{k-2} \left(m_t^{j+1} m_t^{k-2-j} + m_t^j m_t^{k-j-1} \right).$$

These equations are used in [MP19, TT20b] to prove existence and uniqueness of the solution to the integro-differential equation (4.3.4).

Remark 4.3.3. The PDE (4.3.5) satisfies the assumptions of the Cauchy-Kowalevski Theorem, see for instance [Fol95, Theorem 1.25]. However, this theorem only gives local existence and uniqueness of an analytic solution (both in space and time) to the PDE. We did not manage to expand this approach to conclude to global existence and uniqueness of a solution.

The next Proposition gives results on the PDE (4.3.5) using a different method than the ones used in [MP19].

Proposition 4.3.4 (Marcenko Pastur stability along the Burgers PDE and stationarity).

(i) Let $\rho_0, \sigma_0 \in \mathbb{R}_+$. Let us assume $\kappa \neq 0$ and that $\nu_0^W = \lim_{n \rightarrow \infty} \nu_0^{n,W}$ follows a Marcenko-Pastur distribution of shape parameter ρ_0 and of scale parameter σ_0 .

Then, the PDE (4.3.5) admits a solution which is the Cauchy-Stieltjes transform of a Marcenko-Pastur distribution for all $t > 0$ if and only if $\rho_0 = \beta_2 \alpha$. In this case, this solution can be written as the Cauchy-Stieltjes transform of the Marcenko-Pastur distribution $\mu^{MP_{\beta_2 \alpha, \sigma(t)}}$ with

$$\sigma : t \in \mathbb{R}_+ \mapsto \begin{cases} \sqrt{\left(\sigma_0^2 - \frac{\beta_1 \kappa^2}{2\gamma}\right) e^{-2\gamma t} + \frac{\beta_1 \kappa^2}{2\gamma}} & \text{if } \gamma \neq 0 \\ \sqrt{\sigma_0^2 + \beta_1 \kappa^2 t} & \text{if } \gamma = 0 \end{cases}$$

(ii) Let us assume $\gamma \neq 0$ and let $\sigma_0 = \frac{\sqrt{\beta_1 \kappa}}{\sqrt{2\gamma}}$. Then $G_{\mu^{MP_{\beta_2 \alpha, \sigma_0}}}$ is the unique stationary solution to the equation (4.3.5) corresponding to the Cauchy-Stieltjes transform of a probability measure on \mathbb{R} . Moreover, under the assumptions of Theorem 4.3.1 and if $\nu_0 = \mu^{MP_{\beta_2 \alpha, \sigma_0}}$, then for all $t \geq 0$, $\nu_t^W = \lim_{n \rightarrow \infty} \nu_t^{n,W} = \mu^{MP_{\beta_2 \alpha, \sigma_0}}$.

To our knowledge, the assertion (i) was not remarked yet in the past literature, and (ii) can be seen as in the continuity of the results of [ABMV13]. Indeed, their computations are in the high temperature regime (when $\beta n \xrightarrow{n \rightarrow \infty} 2c \in (0, \infty)$) so that the limit of their integro-differential equation and its stationary probability measure are parametrized by c , and they remark that this limit stationary probability measure converges to the Marcenko-Pastur distribution in the limit $c \rightarrow +\infty$.

The next result gives a complete answer to the limit of $(\mu^{n,W})_n$ problem, using free probability tools (in particular the rectangular free convolution defined in Section 3). It is the analog of Theorem 4.1.1 in this context.

Theorem 4.3.5 (Mean-field limit for the general Wishart process). *Under the assumptions of Theorem 4.3.1, let us moreover assume $\beta_2 \alpha \leq 1$, that $(\lambda_t^{1,n,W}, \dots, \lambda_t^{n,n,W})_t$ is defined on \mathbb{R}_+ and that $\mu_0^W = \sqrt{\nu_0^W}$. Let us moreover assume that ν_0^W has a characteristic function*

which is analytic on a neighborhood of the origin. Then for all $t \geq 0$, $(\mu_t^{n,W})_{n \geq 0}$ converges weakly in probability and

$$\mu_t^W := \lim_{n \rightarrow \infty} \mu_t^{n,W} = (e^{-\gamma t} \mu_0^W) \boxplus_{\beta_2 \alpha} \sqrt{\mu^{\text{MP}_{\beta_2 \alpha, \sigma_t}}}$$

where $\boxplus_{\beta_2 \alpha}$ is the rectangular free convolution of parameter $\beta_2 \alpha$, where $e^{-\gamma t} \mu_0^W$ denotes the push-forward of μ_0^W by the map $v \mapsto e^{-\gamma t} v$, and where for all $t \geq 0$:

$$\sigma_t^2 = \begin{cases} \frac{\beta_1 \kappa^2}{2\gamma} (1 - e^{-2\gamma t}) & \text{if } \gamma \neq 0 \\ \beta_1 \kappa^2 t & \text{if } \gamma = 0. \end{cases}$$

Moreover, if $\gamma \neq 0$ and with $\sigma_\infty = \sqrt{\frac{\beta_1 \kappa^2}{2\gamma}}$,

$$\lim_{t \rightarrow +\infty} \mu_t^W = \sqrt{\mu^{\text{MP}_{\beta_2 \alpha, \sigma_\infty}}}.$$

Theorem 4.3.5 coupled with the limit (4.3.3) allows to show the following result.

Theorem 4.3.6 (Commutativity of the limits). *Under the assumptions of Theorem 4.3.5, let us moreover suppose $\gamma \neq 0, \beta_2 = 1$ and that for all $n \in \mathbb{N}^*$: $\sum_{i=1}^n \mathbb{E}[\lambda_0^{i,n,W}] < +\infty$. Let us note $\sigma_\infty = \sqrt{\frac{\beta_1 \kappa^2}{2\gamma}}$. Then we have in the sense of weak convergence*

$$\lim_{\substack{n \rightarrow \infty \\ n/m \rightarrow \alpha}} \lim_{t \rightarrow \infty} \nu_t^{n,W} = \lim_{t \rightarrow \infty} \lim_{\substack{n \rightarrow \infty \\ n/m \rightarrow \alpha}} \nu_t^{n,W} = \mu^{\text{MP}_{\alpha, \sigma_\infty}}.$$

4.4 Useful tools

4.4.1 The Cauchy-Stieltjes transform

Let us denote

$$\mathbb{C}_\pm = \{z \in \mathbb{C}, \pm \text{Im}(z) > 0\}.$$

Let μ be a probability measure on \mathbb{R} . Its Cauchy-Stieltjes transform is defined by

$$z \in \mathbb{C}_+ \mapsto G_\mu(z) = \int_{\mathbb{R}} \frac{\mu(dv)}{z - v} \in \mathbb{C}_-.$$

The next result shows that this transformation characterizes the probability measure on \mathbb{R} .

Theorem 4.4.1 ([SM17]). *(1) Let μ be a probability measure on \mathbb{R} . Then,*

(i) G_μ is analytic on \mathbb{C}_+ ,

(ii) we have

$$\lim_{y \rightarrow \infty} iy G_\mu(iy) = 1.$$

(2) Any probability measure on \mathbb{R} can be recovered from its Cauchy-Stieltjes transform G_μ via the Stieltjes inversion formula : for all $a, b \in \mathbb{R}$ with $a < b$,

$$-\lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \int_a^b \operatorname{Im}(G_\mu(x + i\varepsilon)) dx = \mu((a, b)) + \frac{1}{2} \mu(\{a, b\}).$$

It follows that the Cauchy-Stieltjes transform G_μ characterises the measure μ .

(3) Let $G : \mathbb{C}_+ \mapsto \mathbb{C}_-$ be an analytic function which satisfies

$$\limsup_{y \rightarrow \infty} |G(iy)| = 1.$$

Then, there exists a unique probability measure μ on \mathbb{R} such that $G = G_\mu$.

4.4.2 Free convolution and rectangular free convolution from a random matrix point of view

Let M be a $n \times n$ real symmetric matrix. We will denote its eigenvalues by

$$\lambda^1(M) \leq \dots \leq \lambda^n(M),$$

and its empirical (spectral) measure by

$$\mu^M = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda^i(M)}.$$

Let N be a $n \times m$ real valued matrix. Its empirical singular measure is the empirical measure of the positive semi-definite symmetric matrix NN^* .

Free convolutions are operations on probability measures on the real line which allow to compute the empirical spectral or singular measures of large random matrices (i.e. matrices whose size goes to infinity) which are expressed as sums or products of independent large random matrices whose spectral measures are already known. The free additive convolution is denoted by \boxplus .

Theorem 4.4.2 (Additive Free Convolution for random matrices [VDN92]). *For all $n \in \mathbb{N}^*$ let us define M_n and N_n two independent $n \times n$ random symmetric (hermitian) matrices, such that*

- *the distribution of M_n is invariant under the action of the orthogonal (unitary) group by conjugation,*
- *the empirical measures sequences $(\mu^{M_n})_n$ and $(\mu^{N_n})_n$ converge weakly almost surely when n goes to infinity to non-random probability measures, respectively μ_∞^M and μ_∞^N .*

Then, in the sense of weakly almost sure convergence,

$$\mu^{M_n + N_n} \xrightarrow{n \rightarrow \infty} \mu_\infty^M \boxplus \mu_\infty^N.$$

This operation can be equivalently defined in reference to free elements of a non commutative probability space. Let μ be a probability measure on \mathbb{R} . Its R-transform is defined on a neighbourhood of zero by

$$R_\mu(z) = G_\mu^{-1}(z) - \frac{1}{z}.$$

Theorem 4.4.3. *The R-transform linearizes the free convolution : for μ and ν probability measures on the real line, and for z in a neighbourhood of zero,*

$$R_{\mu \boxplus \nu}(z) = R_\mu(z) + R_\nu(z),$$

and $\mu \boxplus \nu$ is the unique probability measure verifying this relation.

Let us denote by $\mathcal{B}(\mathbb{R})$ the Borel sets of \mathbb{R} . Let us define the application from $\mathcal{M}_1(\mathbb{R}_+)$, the set of probability measures on \mathbb{R}_+ , to $\mathcal{M}_1^S(\mathbb{R}) = \{\mu \in \mathcal{M}_1(\mathbb{R}), \forall A \in \mathcal{B}(\mathbb{R}), \mu(A) = \mu(-A)\}$, the set of symmetric probability measures on \mathbb{R} :

$$\text{sym} : \mu \in \mathcal{M}_1(\mathbb{R}_+) \mapsto \left(\text{sym}(\mu) : A \in \mathcal{B}(\mathbb{R}) \mapsto \frac{\mu(A \cap \mathbb{R}_+) + \mu(-A \cap \mathbb{R}_+)}{2} \right) \in \mathcal{M}_1^S(\mathbb{R}).$$

This application is bijective from $\mathcal{M}_1(\mathbb{R}_+)$ to $\mathcal{M}_1^S(\mathbb{R})$ and admits the inverse:

$$\text{sym}^{-1}(\nu) : A \in \mathcal{B}(\mathbb{R}_+) \mapsto \mathbf{1}_A(0)\nu(\{0\}) + 2\nu(A \setminus \{0\}). \quad (4.4.1)$$

For any $\alpha \in [0, 1]$, the rectangular free convolution denoted \boxplus_α can be defined the following way.

Theorem 4.4.4 (Additive free rectangular convolution of ratio α for random matrices [BG09]). *For all $n, m \in \mathbb{N}^*$ let us define $M_{n,m}$ and $N_{n,m}$ two independent $n \times m$ random matrices with real (complex)-valued entries, such that*

- *the distribution of $M_{n,m}$ is invariant under the action of the orthogonal (unitary) group by conjugation on any side,*
- *the empirical measures sequences $(\mu^{\sqrt{M_{n,m} M_{n,m}^*}})_{n,m}$ and $(\mu^{\sqrt{N_{n,m} N_{n,m}^*}})_{n,m}$ respectively converge in probability, when n and m go to infinity with n/m tending to $\alpha \in (0, 1]$, to non-random probability measures μ_∞^M and μ_∞^N .*

Then in the sense of weak convergence in probability,

$$\text{sym} \left(\mu^{\sqrt{(M_{n,m} + N_{n,m})(M_{n,m} + N_{n,m})^*}} \right) \xrightarrow[n/m \rightarrow \alpha]{n, m \rightarrow \infty} \text{sym}(\mu_\infty^M) \boxplus_\alpha \text{sym}(\mu_\infty^N).$$

This operation can also be equivalently defined in reference to free elements of a rectangular non commutative probability space.

Let μ be a symmetric probability measure on \mathbb{R} . Its rectangular Cauchy transform with ratio α is defined by

$$H_\mu : z \in \mathbb{C} \setminus [0, +\infty) \mapsto z(\alpha M_{\mu^2}(z) + 1)(M_{\mu^2}(z) + 1),$$

where

$$M_{\mu^2} : z \in \mathbb{C} \setminus [0, +\infty) \mapsto \int_{\mathbb{R}} \frac{zv^2}{1-zv^2} d\mu(v) = \frac{1}{\sqrt{z}} G_{\mu} \left(\frac{1}{\sqrt{z}} \right) - 1.$$

This equality can be derived the following way in a neighbourhood of zero :

$$\begin{aligned} G_{\mu} \left(\frac{1}{\sqrt{z}} \right) &= z^{\frac{1}{2}} \int_{\mathbb{R}} \frac{1}{1-z^{\frac{1}{2}}v} d\mu(v) \\ &= z^{\frac{1}{2}} \sum_{k \geq 0} \int_{\mathbb{R}} z^{\frac{k}{2}} v^k d\mu(v) \\ &= z^{\frac{1}{2}} \left(1 + zv^2 \sum_{k \geq 0} \int_{\mathbb{R}} z^k v^{2k} d\mu(v) \right), \end{aligned}$$

where we use in the last equality the fact that μ is symmetric.

The rectangular R-transform with ratio α of μ is defined on a neighbourhood of zero by

$$C_{\mu}(z) = U \left(\frac{z}{H_{\mu}^{-1}(z)} - 1 \right),$$

where on a neighbourhood of zero

$$U(z) = \begin{cases} \frac{-\alpha - 1 + [(\alpha + 1)^2 + 4\alpha z]^{1/2}}{2\alpha} & \text{if } \alpha \neq 0 \\ z & \text{if } \alpha = 0. \end{cases}$$

Theorem 4.4.5 ([BG09]). *The rectangular R-transform with ratio α linearizes the rectangular free convolution with ratio α : for μ and ν symmetric probability measures on the real line, and for z in a neighbourhood of zero,*

$$C_{\mu \boxplus_{\alpha} \nu}(z) = C_{\mu}(z) + C_{\nu}(z),$$

and $\mu \boxplus_{\alpha} \nu$ is the unique symmetric probability measure verifying this relation.

Theorem 4.4.6 (Injectivity of the rectangular R-transform, [BG09]). *If the rectangular R-transforms with ratio α of two symmetric probability measures coincide on a neighbourhood of 0 in $(-\infty, 0]$, then the measures are equal.*

4.5 Proofs

Proof of Theorem 4.1.1. The entries of the matrix M are independent Ornstein-Uhlenbeck processes just as the one considered in [Bru91]. We thus can write for all $t \geq 0$

$$M_t = M_0 e^{-\gamma t} + \kappa \int_0^t e^{\gamma(s-t)} dW_s.$$

The idea is to use a suitable rectangular free convolution, see [BG09]. More precisely, the matrices $A_{n,t} = M_0 e^{-\gamma t}$ and $B_{n,t} = \kappa \int_0^t e^{\gamma(s-t)} dW_s$ are independent and we may use a version of Voiculescu asymptotic freeness theorem, see [HP00]. The precise result to use is Theorem 4.4.4 (see [BG09, Theorem 3.13]). Indeed, $B_{n,t}$ is a matrix filled with i.i.d Gaussian random variables of variance

$$\sigma_t^2 = \mathbb{E} \left[\left| \kappa \int_0^t e^{-\gamma(s-t)} dW_s \right|^2 \right] = \begin{cases} \kappa^2 \int_0^t e^{2\gamma(s-t)} ds = \frac{\kappa^2}{2\gamma} (1 - e^{-2\gamma t}) & \text{if } \gamma \neq 0 \\ \kappa^2 t & \text{if } \gamma = 0 \end{cases}$$

by Ito's isometry, and is thus bi-unitary invariant under the action of the orthogonal group. The Marcenko-Pastur theorem (see for instance [BS10, Theorem 3.10]) tells us that in the sense of convergence in probability,

$$\mu^{\frac{1}{m} B_{n,t} B_{n,t}^*} = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda^i(\frac{1}{m} B_{n,t} B_{n,t}^*)} \xrightarrow[n \rightarrow \infty]{} \mu^{MP_{\alpha, \sigma t}},$$

weakly, and thus we have :

$$\mu^{\sqrt{\frac{1}{m} B_{n,t} B_{n,t}^*}} = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda^i(\sqrt{\frac{1}{m} B_{n,t} B_{n,t}^*})} \xrightarrow[n \rightarrow \infty]{} \sqrt{\cdot} \# \mu^{MP_{\alpha, \sigma t}},$$

which is a non random measure.

An application of Theorem 4.4.4 to $A_{n,t}$ and $B_{n,t}$ for all $t > 0$ ends the first part of the proof.

Let us assume $\gamma \neq 0$. According to [BG09, Theorem 2.12], the binary operation \boxplus_{α} is continuous (with respect to the weak convergence) on the set of symmetric probability measures on the real line, and so does the the rectangular R-transform C . Thus, as

$$\lim_{t \rightarrow +\infty} \text{sym}(e^{-\gamma t} \mu_0) = \delta_0 \quad \text{and} \quad \lim_{t \rightarrow +\infty} \sqrt{\mu^{MP_{\alpha, \sigma t}}} = \sqrt{\mu^{MP_{\alpha, \sigma \infty}}},$$

we have

$$\lim_{t \rightarrow +\infty} \text{sym}(\mu_t) = \delta_0 \boxplus_{\alpha} \sqrt{\mu^{MP_{\alpha, \sigma \infty}}}.$$

Moreover, the formula in Subsection 4.4.2 allows to compute for all $z \in \mathbb{C}$:

$$\begin{aligned} M_{(\delta_0)^2}(z) &= 0, \\ H_{\delta_0}(z) &= z, \\ C_{\delta_0}(z) &= U(0) = 0. \end{aligned}$$

Consequently, applying Theorem 4.4.5,

$$C_{\delta_0 \boxplus_{\alpha} \sqrt{\mu^{MP_{\alpha, \sigma \infty}}}} = C_{\delta_0} + C_{\sqrt{\mu^{MP_{\alpha, \sigma \infty}}}} = C_{\sqrt{\mu^{MP_{\alpha, \sigma \infty}}}},$$

which allows to conclude, applying Theorem 4.4.6, that

$$\delta_0 \boxplus_{\alpha} \sqrt{\mu^{MP_{\alpha, \sigma \infty}}} = \sqrt{\mu^{MP_{\alpha, \sigma \infty}}},$$

which ends the proof. □

Proof of Theorem 4.1.2. Let us first show that if $(\lambda^{1,n}, \dots, \lambda^{n,n})$ is a random vector distributed according to the distribution with density with respect to the Lebesgue measure:

$$(\lambda^1, \dots, \lambda^n) \mapsto \frac{1}{\mathcal{Z}} \prod_{i=1}^n \left((\lambda^i)^{\frac{m-n+1}{2}-1} e^{-\frac{m\gamma}{\kappa^2} \lambda^i} \prod_{j \neq i} |\lambda^j - \lambda^i|^{1/2} \right) \mathbb{1}_{0 \leq \lambda^1 \leq \dots \leq \lambda^n}, \quad (4.5.1)$$

where \mathcal{Z} is a normalizing constant, and if we define the empirical measure

$$\nu^n = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda^{i,n}},$$

and note $\sigma_\infty = \sqrt{\frac{\kappa^2}{2\gamma}}$, then, in the sense of weak convergence,

$$\nu^n \xrightarrow[n/m \rightarrow \alpha]{n \rightarrow +\infty} \mu^{MP_{\alpha, \sigma_\infty}},$$

which is a particular case of (4.3.3).

Let us consider the $n \times m$ random matrix M_0 whose coordinates are independent identically distributed centered real Gaussian random variables of variance $\sigma_\infty^2 = \frac{\kappa^2}{2\gamma}$. An application of [PS11, Proposition 7.4.1] shows that of the matrix $\frac{1}{m} M_0 M_0^*$ follow the density (4.5.1). Thus, we have

$$\mu^{\frac{1}{m} M_0 M_0^*} = \nu^n,$$

in the sense of equality in law.

Moreover, an application of the Marcenko-Pastur theorem (see for instance [BS10, Theorem 3.10]) shows that in the sense of convergence in probability

$$\mu^{\frac{1}{m} M_0 M_0^*} \xrightarrow[n/m \rightarrow \alpha]{n \rightarrow +\infty} \mu^{MP_{\alpha, \sigma_\infty}},$$

weakly which allows to conclude.

For all $n \in \mathbb{N}^*$, an application of [JK20, Lemma 3.1 and Proposition 2.8] shows that in the sense of weak convergence,

$$\nu_t^n \xrightarrow[t \rightarrow +\infty]{} \nu^n.$$

We proved earlier that we have in the sense of weak convergence

$$\nu^n \xrightarrow[n/m \rightarrow \alpha]{n \rightarrow +\infty} \mu^{MP_{\alpha, \sigma_\infty}}.$$

Theorem 4.1.1 gives the two other limits and concludes the proof. \square

Proof of Theorem 4.2.1. The proof of this Theorem mimics the proof of Theorem 4.1.1. \square

Proof of Theorem 4.3.1. For f a twice continuously differentiable real test function, we have thanks to the SDE (4.3.1) :

$$\begin{aligned} d\langle \nu_t^{n,W}, f \rangle &= \frac{2\kappa}{n\sqrt{m}} \sum_{j=1}^n f'(\lambda_t^{j,n,W}) \sqrt{\lambda_t^{j,n,W}} dB_t^j \\ &+ \left(\frac{2\kappa^2}{nm} \sum_{j=1}^n \lambda_t^{j,n,W} f''(\lambda_t^{j,n,W}) + \frac{1}{n} \sum_{j=1}^n (\beta_1 \kappa^2 - 2\gamma \lambda_t^{j,n,W}) f'(\lambda_t^{j,n,W}) \right) dt \\ &+ \left(\frac{\beta_1 \beta_2 \kappa^2}{nm} \sum_{j=1}^n f'(\lambda_t^{j,n,W}) \sum_{k \neq j} \frac{\lambda_t^{j,n,W} + \lambda_t^{k,n,W}}{\lambda_t^{j,n,W} - \lambda_t^{k,n,W}} \right) dt. \end{aligned}$$

We have :

$$\begin{aligned} \frac{1}{nm} \sum_{j=1}^n f'(\lambda_t^{j,n,W}) \sum_{k \neq j} \frac{\lambda_t^{j,n,W} + \lambda_t^{k,n,W}}{\lambda_t^{j,n,W} - \lambda_t^{k,n,W}} &= \frac{1}{2nm} \sum_{j=1}^n \sum_{k \neq j} (f'(\lambda_t^{j,n,W}) - f'(\lambda_t^{k,n,W})) \frac{\lambda_t^{j,n,W} + \lambda_t^{k,n,W}}{\lambda_t^{j,n,W} - \lambda_t^{k,n,W}} \\ &= \frac{n}{2m} \iint_{\{x \neq y\}} (f'(x) - f'(y)) \frac{x+y}{x-y} \nu_t^{n,W}(dx) \nu_t^{n,W}(dy) \\ &= \frac{n}{2m} \iint (x+y) \frac{f'(x) - f'(y)}{x-y} \nu_t^{n,W}(dx) \nu_t^{n,W}(dy) \\ &\quad - \frac{1}{m} \int x f''(x) \nu_t^{n,W}(dx). \end{aligned}$$

Finally we have

$$\begin{aligned} d\langle \nu_t^{n,W}, f \rangle &= \frac{2\kappa}{n\sqrt{m}} \sum_{j=1}^n f'(\lambda_t^{j,n,W}) \sqrt{\lambda_t^{j,n,W}} dB_t^j \\ &+ \left(\frac{2\kappa^2}{nm} \sum_{j=1}^n \lambda_t^{j,n,W} f''(\lambda_t^{j,n,W}) + \frac{1}{n} \sum_{j=1}^n (\beta_1 \kappa^2 - 2\gamma \lambda_t^{j,n,W}) f'(\lambda_t^{j,n,W}) \right) dt \\ &+ \beta_1 \beta_2 \kappa^2 \left(\frac{n}{2m} \iint (x+y) \frac{f'(x) - f'(y)}{x-y} \nu_t^{n,W}(dx) \nu_t^{n,W}(dy) - \frac{1}{m} \int x f''(x) \nu_t^{n,W}(dx) \right) dt \\ &= \frac{2\kappa}{n\sqrt{m}} \sum_{j=1}^n f'(\lambda_t^{j,n,W}) \sqrt{\lambda_t^{j,n,W}} dB_t^j + \langle \nu_t^{n,W}, \frac{\kappa^2(2 - \beta_1 \beta_2) \Phi}{m} f'' + (\beta_1 \kappa^2 - 2\gamma \Phi) f' \rangle dt \\ &+ \beta_1 \beta_2 \kappa^2 \left(\frac{n}{2m} \iint (x+y) \frac{f'(x) - f'(y)}{x-y} \nu_t^{n,W}(dx) \nu_t^{n,W}(dy) \right) dt \\ &= dM_t^{(n,f)} + \langle \nu_t^{n,W}, \frac{\kappa^2(2 - \beta_1 \beta_2) \Phi}{m} f'' + (\beta_1 \kappa^2 - 2\gamma \Phi) f' \rangle dt \\ &+ \beta_1 \beta_2 \kappa^2 \left(\frac{n}{2m} \iint (x+y) \frac{f'(x) - f'(y)}{x-y} \nu_t^{n,W}(dx) \nu_t^{n,W}(dy) \right) dt, \end{aligned} \tag{4.5.2}$$

with $\Phi : x \rightarrow x$ defined on \mathbb{R} and $M^{(n,f)}$ a continuous martingale satisfying

$$d\langle M^{(n,f)} \rangle_t = \frac{4\kappa^2}{n^2 m} \sum_{i=1}^n |\lambda_t^{i,n,W} f'(\lambda_t^{i,n,W})|^2 dt.$$

The reader will find in [MP19] a proof of the tightness of the family $\{(\nu_t^{n,W})_{t \geq 0}; n \geq 1\}$, which allows to conclude by the virtue of the previous computations that any accumulation point of this family satisfies the evolution equation (4.3.4). The reader will also find in [MP19] the proof of the uniqueness of the solution to the equation (4.3.4) in the case where ν_0 admits a characteristic function and this function is analytic on a neighbourhood of the origin, this proof being based on the considerations made in Remark 4.3.2.

Let us now prove the second part of the Theorem. Applying (4.3.4) with

$$f(v) = \frac{1}{z - v},$$

we get that $G_t(z)$ obeys

$$G_t(z) = G_0(z) + \int_0^t \beta_1 \kappa^2 \frac{1}{(z-x)^2} - 2\gamma \frac{x}{(z-x)^2} \nu_s(dx) ds + \frac{\alpha \beta_1 \beta_2 \kappa^2}{2} \iint (x+y) \frac{f'(x) - f'(y)}{x-y} \nu_t(dx) \nu_t(dy).$$

We have

$$\begin{aligned} & \frac{1}{2} \iint (x+y) \frac{f'(x) - f'(y)}{x-y} \nu_t(dx) \nu_t(dy) \\ &= \frac{1}{2} \iint \left(\frac{1}{(z-x)^2} - \frac{1}{(z-y)^2} \right) \frac{x+y}{x-y} \nu_t(dx) \nu_t(dy) \\ &= \frac{1}{2} \iint \frac{(2z-x-y)(x+y)}{(z-x)^2(z-y)^2} \nu_t(dx) \nu_t(dy) \\ &= \iint \frac{2zx - xy - x^2 + z^2 - z^2}{(z-x)^2(z-y)^2} \nu_t(dx) \nu_t(dy) \\ &= \iint -\frac{1}{(z-y)^2} - \frac{xy}{(z-x)^2(z-y)^2} + \frac{z^2}{(z-x)^2(z-y)^2} \nu_t(dx) \nu_t(dy) \\ &= \iint -\frac{1}{(z-y)^2} + \frac{z^2}{(z-x)^2(z-y)^2} \nu_t(dx) \nu_t(dy) - \left(\int \frac{x}{(z-x)^2} \nu_t(dx) \right)^2 \\ &= \iint -\frac{1}{(z-y)^2} + \frac{z^2}{(z-x)^2(z-y)^2} \nu_t(dx) \nu_t(dy) - \left(\int \frac{1}{x-z} + \frac{z}{(z-x)^2} \nu_t(dx) \right)^2 \\ &= -\int \frac{1}{(z-x)^2} \nu_t(dx) + z^2 \left(\int \frac{1}{(z-x)^2} \nu_t(dx) \right)^2 - \left(\int \frac{1}{z-x} \nu_t(dx) \right)^2 \\ &+ 2z \left(\int \frac{1}{z-x} \nu_t(dx) \right) \left(\int \frac{1}{(z-x)^2} \nu_t(dx) \right) - z^2 \left(\int \frac{1}{(z-x)^2} \nu_t(dx) \right)^2, \end{aligned}$$

so that $G_t(z)$ obeys

$$G_t(z) = G_0(z) + \int_0^t \beta_1 \kappa^2 \frac{1}{(z-x)^2} - 2\gamma \left(\frac{1}{x-z} + \frac{z}{(z-x)^2} \right) \nu_s(dx) ds \\ - \alpha \beta_1 \beta_2 \kappa^2 \left[\int \frac{1}{(z-x)^2} \nu_t(dx) + \left(\int \frac{1}{z-x} \nu_t(dx) \right)^2 - 2z \left(\int \frac{1}{z-x} \nu_t(dx) \right) \left(\int \frac{1}{(z-x)^2} \nu_t(dx) \right) \right].$$

The conclusion is given replacing the previous terms by the corresponding derivatives of G . \square

Proof of Proposition 4.3.4. (i) We first recall the Cauchy-Stieltjes transform of a Marcenko-Pastur law (see for instance [BS10, Lemma 3.11]) :

$$G_{\mu^{MP_{\rho,\sigma}}}(z) = \int \frac{\mu^{MP_{\rho,\sigma}}(dv)}{z-v} = \frac{-\sigma^2(1-\rho) + z - \sqrt{(z-\sigma^2 - \rho\sigma^2)^2 - 4\rho\sigma^4}}{2\rho z\sigma^2},$$

for all $z \in \mathbb{C}_+$.

We now want to find conditions on the functions $t \rightarrow \sigma(t) \in \mathbb{R}_+$ and $t \rightarrow \rho(t) \in \mathbb{R}_+$ with $\sigma(0) = \sigma_0$ and $\rho(0) = \rho_0$ such that $(t, z) \rightarrow G_{\mu^{MP_{\rho(t),\sigma(t)}}}(z)$ is solution to the PDE (4.3.5):

$$\frac{\partial}{\partial t} G_t(z) = (\alpha \beta_1 \beta_2 \kappa^2 - \beta_1 \kappa^2 + 2\gamma z) \frac{\partial}{\partial z} G_t(z) - 2\alpha \beta_1 \beta_2 \kappa^2 z G_t(z) \frac{\partial}{\partial z} G_t(z) - \alpha \beta_1 \beta_2 \kappa^2 G_t^2(z) + 2\gamma G_t(z), \\ G_0(z) = \int \frac{\nu_0^W(dv)}{z-v} = \phi(z).$$

We have

$$\frac{\partial}{\partial t} G_{\mu^{MP_{\rho(t),\sigma(t)}}}(z) = -\frac{\sigma(z-\sigma^2)\dot{\rho} + 2\dot{\sigma}\rho z}{2\rho^2 z\sigma^3} \\ - \frac{\sigma((1-\rho)\sigma^4 - z(\rho+2)\sigma^2 + z^2)\dot{\rho} + 2(z-\sigma^2(1+\rho))z\dot{\sigma}\rho}{2\rho^2 z\sigma^3 \sqrt{(z-\sigma^2 - \rho\sigma^2)^2 - 4\rho\sigma^4}} \\ \frac{\partial}{\partial z} G_{\mu^{MP_{\rho(t),\sigma(t)}}}(z) = \frac{1-\rho}{2\rho z^2} - \frac{z(1+\rho) - \sigma^2(1-\rho)^2}{2\rho z^2 \sqrt{(z-\sigma^2 - \rho\sigma^2)^2 - 4\rho\sigma^4}} \\ (G_{\mu^{MP_{\rho(t),\sigma(t)}}}(z))^2 = \frac{\sigma^4(1-\rho)^2 - 2z\sigma^2 + z^2 + [z - \sigma^2(1-\rho)]\sqrt{(z-\sigma^2 - \rho\sigma^2)^2 - 4\rho\sigma^4}}{2\rho^2 z^2 \sigma^4} \\ G_{\mu^{MP_{\rho(t),\sigma(t)}}}(z) \frac{\partial}{\partial z} G_{\mu^{MP_{\rho(t),\sigma(t)}}}(z) = \frac{-(1-\rho)^2 \sigma^2 + z}{2\rho^2 z^3 \sigma^2} + \frac{z^2 + z\sigma^2[\rho^2 \rho - 2] + \sigma^4(1-\rho)^3}{2\rho^2 z^3 \sigma^2 \sqrt{(z-\sigma^2 - \rho\sigma^2)^2 - 4\rho\sigma^4}}.$$

The equation solved by G can be written for $z \in \mathbb{C}_+, t \geq 0$:

$$H(z, t) = 0 \tag{4.5.3}$$

with

$$H(z, t) = \frac{z\sigma^2(z - \sigma^2)\dot{\rho} + 2\sigma\rho\dot{\sigma}z^2 + \beta_1\kappa^2(1 - \rho)(\beta_2\alpha - \rho)\sigma^4 + 2\sigma^2\rho\gamma z^2 - \beta_1\beta_2\alpha\kappa^2 z^2}{2\rho^2\sigma^4 z^2}$$

$$- \frac{z((1 - \rho)\sigma^4 - z(\rho + 2)\sigma^2 + z^2)\sigma^2\dot{\rho} + 2z^2(z - (1 + \rho)\sigma^2)\rho\sigma\dot{\sigma} + \beta_1\kappa^2(\rho - 1)^2(\beta_2\alpha - \rho)\sigma^6}{2\rho^2\sigma^4 z^2 \sqrt{(z - \sigma^2 - \rho\sigma^2)^2 - 4\rho\sigma^4}}$$

$$+ \frac{-z((2\gamma z + \beta_1\kappa^2)\rho - \beta_1\beta_2\alpha\kappa^2)(\rho + 1)\sigma^4 + z^2((2\gamma z + \beta_1\beta_2\alpha\kappa^2)\rho + \beta_1\beta_2\alpha\kappa^2)\sigma^2 - \beta_1\beta_2\alpha\kappa^2 z^3}{2\rho^2\sigma^4 z^2 \sqrt{(z - \sigma^2 - \rho\sigma^2)^2 - 4\rho\sigma^4}}.$$

Using the fact that

$$((z - \sigma^2 - \rho\sigma^2)^2 - 4\rho\sigma^4)^{-\frac{1}{2}} = \frac{1}{z} + \frac{\sigma^2(1 + \rho)}{z^2} + \sigma^4 \frac{\rho^2 + 4\rho + 1}{z^3} + o_{|z| \rightarrow +\infty} \left(\frac{1}{|z|^3} \right),$$

we have for $z \in \mathbb{C}_+$, $t \geq 0$:

$$H(z, t) = \frac{\sigma^2\dot{\rho} + 2\sigma\rho\dot{\sigma} - \beta_1\beta_2\alpha\kappa^2 + 2\gamma\rho\sigma^2}{\sigma^4\rho^2} - \frac{\dot{\rho}}{\rho^2} \frac{1}{z}$$

$$- \frac{\beta_1\kappa^2(\rho - 1)(\beta_2\alpha - \rho) - 4\sigma\rho^2\dot{\sigma} + (\beta_1\kappa^2 - 4\gamma\sigma^2)\rho^2 + \beta_1\kappa^2(1 + \beta_2\alpha)\rho - \beta_1\beta_2\alpha\kappa^2}{2\rho^2} \frac{1}{z^2} + o_{|z| \rightarrow +\infty} \left(\frac{1}{|z|^2} \right)$$

which gives in particular by identity (4.5.3) :

$$\begin{aligned} \sigma^2\dot{\rho} + 2\sigma\rho\dot{\sigma} - \beta_1\beta_2\alpha\kappa^2 + 2\gamma\rho\sigma^2 &= 0, \\ \dot{\rho} &= 0, \\ \beta_1\kappa^2(\rho - 1)(\beta_2\alpha - \rho) - 4\sigma\rho^2\dot{\sigma} + (\beta_1\kappa^2 - 4\gamma\sigma^2)\rho^2 + \beta_1\kappa^2(1 + \beta_2\alpha)\rho - \beta_1\beta_2\alpha\kappa^2 &= 0. \end{aligned}$$

We thus have

$$\begin{aligned} \rho(t) &= \beta_2\alpha, \\ \dot{\sigma}^2 &= \beta_1\kappa^2 - 2\gamma\sigma^2, \end{aligned}$$

and

$$\sigma^2(t) = \begin{cases} \left(\sigma_0^2 - \frac{\beta_1\kappa^2}{2\gamma} \right) e^{-2\gamma t} + \frac{\beta_1\kappa^2}{2\gamma} & \text{if } \gamma \neq 0 \\ \sigma_0^2 + \beta_1\kappa^2 t & \text{if } \gamma = 0, \end{cases}$$

for all $t \in \mathbb{R}_+$.

Reciprocally, we verify with this definition of σ that $G_\mu^{MP_{\beta_2\alpha, \sigma(t)}}$ is a solution to the PDE (4.3.5).

(ii) Let us consider the stationary version of the PDE (4.3.5) :

$$(\beta_1\beta_2\alpha\kappa^2 - \beta_1\kappa^2 + 2\gamma z) \frac{d}{dz} G(z) - 2\beta_1\beta_2\alpha\kappa^2 z G(z) - \frac{d}{dz} G(z) - \beta_1\beta_2\alpha\kappa^2 G^2(z) + 2\gamma G(z) = 0.$$

We can integrate it

$$-\beta_1\beta_2\alpha\kappa^2zG^2(z) + (2\gamma z + (\beta_2\alpha - 1)\beta_1\kappa^2)G(z) = C,$$

where $C \in \mathbb{C}$ is an integration constant. It gives

$$G(z) = \frac{2\gamma z + (\beta_2\alpha - 1)\beta_1\kappa^2 \pm \sqrt{(2\gamma z + (\beta_2\alpha - 1)\beta_1\kappa^2)^2 - 4C\beta_1\beta_2\alpha\kappa^2z}}{2\beta_1\beta_2\alpha\kappa^2z}.$$

For this function to be the Cauchy-Stieltjes transform of a probability measure on \mathbb{R} , we need it to verify the condition given in assertion (3) of Theorem 4.4.1 :

$$\limsup_{y \rightarrow \infty} y|G(iy)| = 1,$$

necessarily, the sign \pm must be replaced by a minus sign and $C = 2\gamma$. Thus, the only stationary solution to the PDE (4.3.5) corresponding to the Cauchy-Stieltjes transform of a probability measure on \mathbb{R} is

$$\begin{aligned} G(z) &= \frac{2\gamma z + (\beta_2\alpha - 1)\beta_1\kappa^2 - \sqrt{(2\gamma z + (\beta_2\alpha - 1)\beta_1\kappa^2)^2 - 8\gamma\beta_1\beta_2\alpha\kappa^2z}}{2\beta_1\beta_2\alpha\kappa^2z} \\ &= \frac{-\sigma^2(1 - \rho) + z - \sqrt{(z - \sigma^2 - \rho\sigma^2)^2 - 4\rho\sigma^4}}{2\rho z\sigma^2}, \end{aligned}$$

with $\rho = \beta_2\alpha$ and $\sigma^2 = \frac{\beta_1\kappa^2}{2\gamma}$. We recognize the Cauchy-Stieltjes transform of the Marcenko-Pastur distribution with parameters ρ and σ , and verify that G is solution to the stationary version of the PDE (4.3.5). Under the assumptions of Theorem 4.3.1, and applying Theorem 4.3.1 with $\nu_0 = \mu^{MP_{\beta_2\alpha, \sigma_0}}$, then $\mu^{MP_{\beta_2\alpha, \sigma_0}}$ is the unique solution to the equation (4.3.4). □

Proof of Theorem 4.3.5. By Theorem 4.3.1, the integro-differential equation (4.3.4) is verified by any accumulation point of the family $\{(\nu_t^{n,W})_{t \geq 0}, n \geq 1\}$.

Let $n \leq m$ with $\frac{n}{m} \xrightarrow{n \rightarrow +\infty} \beta_2\alpha$. Let us define the stochastic process $(M_t)_t$ taking its values in the space of $n \times m$ matrices with real entries verifying the following SDE

$$dM_t = \sqrt{\beta_1}\kappa dW_t - \gamma M_t dt.$$

The eigenvalues of $\frac{1}{m}M_t M_t^*$ verify the SDE (4.1.2) with κ replaced by $\sqrt{\beta_1}\kappa$. By Theorem 4.3.1 (see more precisely the computation (4.5.2) in the proof), the integro-differential equation (4.3.4) is also verified by any accumulation point of the family of the empirical spectral measures of $\left(\frac{1}{m}M_t M_t^*\right)_{1 \leq n \leq m}$ with $\frac{n}{m} \xrightarrow{n \rightarrow +\infty} \beta_2\alpha$, which ends the proof.

By Theorem 4.3.1, we have uniqueness of the solution to the integro-differential equation (4.3.4), and by Theorem 4.1.1 we have the expression of the limit of $\mu_t^{n,W} = \sqrt{\nu_t^{n,W}}$ for all $t \geq 0$. □

Proof of Theorem 4.3.6. For all $n \in \mathbb{N}^*$, an application of [JK20, Lemma 3.1 and Proposition 2.8] shows that in convergence in law

$$\nu_t^{n,W} \xrightarrow[t \rightarrow +\infty]{} \nu^{n,W},$$

where $\nu^{n,W}$ is defined such as in equation (4.3.3). By equation (4.3.3), we have

$$\nu^{n,W} \xrightarrow[n \rightarrow +\infty, n/m \rightarrow \alpha]{} \mu^{MP_{\alpha, \sigma_{\infty}}}.$$

Theorem 4.3.5 gives the two other limits and concludes the proof. □

Acknowledgement : I thank Benjamin Jourdain and Djalil Chafaï for numerous fruitful discussions.

Chapter 5

Newton lilypads and Jacobians of neural networks

This Section is the preliminary version of a joint work with *Reda Chhaibi* and *Tariq Daouda*.

Abstract

Gradient descent during the learning process of a neural network can be subject to many instabilities. The spectral density of the Jacobian is a key component for analyzing robustness. Following the works of Pennington et al., such Jacobians are modelled using free multiplicative convolutions from Free Probability Theory.

We present a reliable and very fast method for computing the associated spectral densities, which is certainly of independent interest. Our technique is based on an adaptative Newton-Raphson scheme, by finding and chaining basins of attraction: the Newton algorithm finds contiguous lilypad-like basins and steps from one to the next, heading towards the objective.

This allows to investigate how the learning process is influenced by design choices such as width ratios, depth, non-linearities and the variances at initialization.

Keywords : Machine learning, Robustness and design of neural networks, Back-propagation, Free probability, Free convolution

AMS Subject Classification (2010) : Primary 60F99; Secondary 60G60, 81P15

5.1 Introduction

Consider a feed-forward network of depth $L \in \mathbb{N}$, that is to say that there are L layers. At depth $\ell = 1, 2, \dots, L$, each layer has activation $x_\ell \in \mathbb{R}^{N_\ell}$, where N_ℓ is the current width. The vector $x_0 \in \mathbb{R}^{N_0}$ takes in the neural network's input, while $x_L \in \mathbb{R}^{N_L}$ gives the output. The vector of widths is written:

$$\mathbf{N} := (N_0, N_1, \dots, N_L) ,$$

and will often appear in superscript to indicate the dependence in any of the N_ℓ 's. The following recurrence relation holds between layers:

$$x_\ell = \phi_\ell \left(W_\ell^{(\mathbf{N})} x_{\ell-1} + b_\ell^{(\mathbf{N})} \right) ,$$

where ϕ_ℓ is a choice of non-linearity applied entry-wise, $W_\ell^{(\mathbf{N})} \in M_{N_\ell, N_{\ell-1}}(\mathbb{R})$ is a weight matrix and $b_\ell^{(\mathbf{N})} \in \mathbb{R}^{N_\ell}$ is the vector of biases. We write $h_\ell := W_\ell^{(\mathbf{N})}x_{\ell-1} + b_\ell^{(\mathbf{N})}$ for the pre-activations.

The Jacobian computed during back-propagation can be written explicitly by using the chain rule. Indeed, we have:

$$\begin{aligned} J^{(\mathbf{N})} &:= \frac{\partial x_L}{\partial x_0} \\ &= \frac{\partial x_L}{\partial x_{L-1}} \frac{\partial x_{L-1}}{\partial x_{L-2}} \cdots \frac{\partial x_1}{\partial x_0} \\ &= D_L^{(\mathbf{N})} W_L^{(\mathbf{N})} D_{L-1}^{(\mathbf{N})} W_{L-1}^{(\mathbf{N})} \cdots D_1^{(\mathbf{N})} W_1^{(\mathbf{N})} , \end{aligned} \tag{5.1.1}$$

where D_ℓ 's are the diagonal matrices given by

$$\left[D_\ell^{(\mathbf{N})} \right]_{i,i} = \phi'_\ell([h_\ell]_i) . \tag{5.1.2}$$

Technically, a step of gradient descent updates weights and biases following

$$\left(W_\ell^{(\mathbf{N})}, b_\ell^{(\mathbf{N})} \right) \leftarrow \left(W_\ell^{(\mathbf{N})}, b_\ell^{(\mathbf{N})} \right) - \alpha \frac{\partial \mathcal{L}}{\partial (W_\ell^{(\mathbf{N})}, b_\ell^{(\mathbf{N})})} , \tag{5.1.3}$$

for each $\ell = 1, \dots, L$. Here $\alpha > 0$ is the learning rate and \mathcal{L} is the loss on a minibatch. If the minibatch has size $B \in \mathbb{N}$, and corresponds a small sample $((X_i, Y_i) ; i = 1, \dots, B)$ of the dataset, we have:

$$\mathcal{L} = \frac{1}{B} \sum_{i=1}^B d(x_L(X_i), Y_i) .$$

Here d is a real-valued distance or similarity function, the X_i 's are the neural net's inputs while the Y_i 's are the outputs: labels in the case of classifier, $Y_i = X_i$ in the case of an autoencoder etc...

The chain rule dictates:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial (W_\ell^{(\mathbf{N})}, b_\ell^{(\mathbf{N})})} &= \frac{\partial \mathcal{L}}{\partial x_L} \frac{\partial x_L}{\partial x_{L-1}} \frac{\partial x_{L-1}}{\partial x_{L-2}} \cdots \frac{\partial x_{\ell+1}}{\partial x_\ell} \frac{\partial x_\ell}{\partial (W_\ell^{(\mathbf{N})}, b_\ell^{(\mathbf{N})})} \\ &= \frac{\partial \mathcal{L}}{\partial x_L} J_\ell^{(\mathbf{N})} \frac{\partial h_\ell}{\partial (W_\ell^{(\mathbf{N})}, b_\ell^{(\mathbf{N})})} , \end{aligned} \tag{5.1.4}$$

where

$$\frac{\partial \mathcal{L}}{\partial x_L} = \frac{1}{B} \sum_{i=1}^B \partial_1 d(x_L(X_i), Y_i) \in M_{1, N_L}(\mathbb{R}) , \tag{5.1.5}$$

$$J_\ell^{(\mathbf{N})} = D_L^{(\mathbf{N})} W_L^{(\mathbf{N})} \cdots D_{\ell+1}^{(\mathbf{N})} W_{\ell+1}^{(\mathbf{N})} D_\ell^{(\mathbf{N})} \in M_{N_L, N_\ell}(\mathbb{R}) . \tag{5.1.6}$$

Therefore, for the sake of simplicity, we shall focus on the Jacobian $J^{(\mathbf{N})}$ given in Eq. (5.1.1) since it has exactly the same form as the $J_\ell^{(\mathbf{N})}$ given in Eq. (5.1.6). The issue is that a large product of (even larger) matrices can easily become unstable:

- If many singular values are $\ll 1$, we have gradient vanishing.
- If many singular values are $\gg 1$, we have gradient explosion.

Intuition. This instability is easily understood thanks to the naive analogy with the one-dimensional case. Indeed, the geometric progression q^n with $n \rightarrow \infty$ is the archetype of a long product and it converges extremely fast, to either 0 if $|q| < 1$ or to ∞ if $|q| > 1$.

A less naive intuition consists in observing that mini-batch sampling in Eq. (5.1.5) is very noisy. Without further information on the dataset and sampling procedure, it is fair to assume that $\frac{\partial \mathcal{L}}{\partial x_L}$ has a Gaussian behavior with covariance proportional to I_{N_L} . Therefore, each gradient step $\alpha \frac{\partial \mathcal{L}}{\partial (W_\ell^{(\mathbf{N})}, b_\ell^{(\mathbf{N})})}$ in Eq. (5.1.3) is approximately a Gaussian vector with covariance proportional to:

$$\alpha^2 \left(\frac{\partial h_\ell}{\partial (W_\ell^{(\mathbf{N})}, b_\ell^{(\mathbf{N})})} \right)^T \left(J_\ell^{(\mathbf{N})} \right)^T J_\ell^{(\mathbf{N})} \frac{\partial h_\ell}{\partial (W_\ell^{(\mathbf{N})}, b_\ell^{(\mathbf{N})})}.$$

Simplifying further, we see the importance of the spectrum of $\left(J_\ell^{(\mathbf{N})} \right)^T J_\ell^{(\mathbf{N})}$ for stability. Basically, eigenvectors of $\left(J_\ell^{(\mathbf{N})} \right)^T J_\ell^{(\mathbf{N})}$ are the directions along which the one-dimensional intuition applies.

Randomness. Starting from the pioneering works of Glorot and Bengio [GB10] on random initializations, it was suggested that the spectral properties of $J^{(\mathbf{N})}$ are an excellent indicator for stability and learning performance. In particular, an appropriate random initialization was suggested and since implemented in all modern ML frameworks [PGC+17, AAB+15].

Here are classical choices of random initializations. The biases $b_\ell^{(\mathbf{N})}$ are taken as random vectors which entries are centered i.i.d. Gaussian random variables with standard deviation σ_{b_ℓ} . For the weights, we will consider the following matrix ensembles: the $[W_\ell^{(\mathbf{N})}]_{i,j}$ are drawn from i.i.d. centered random variables with variance $\sigma_{W_\ell}^2/N_\ell$ and finite fourth moment as in [PS20].

Modeling spectrum thanks to Free Probability Theory. Now, following the works of Pennington et al. [PSG18], the tools of Free Probability Theory (FPT) can be used to quantitatively analyze the singular values of $J^{(\mathbf{N})}$ in the large width limit. The large width limit is particularly attractive when studying large deep networks, especially because free probability appears at relatively small sizes because of strong concentration properties [LLC+18].

For the purposes of this paragraph, we restrict ourselves to square matrices and assume $N_\ell = N$ for all $\ell = 1, \dots, L$. In fact, FPT is concerned with the behavior of spectral measures as $N \rightarrow \infty$. For any diagonalizable $A_N \in M_N(\mathbb{R})$, the associated spectral measure on the real line is:

$$\mu_{A^{(N)}}(dx) := \frac{1}{N} \sum_{i=1}^N \delta_{a_i^{(N)}}(dx)$$

with the $a_i^{(N)}$'s being the eigenvalues of A_N . For ease of notation, the spectrum of (squared) singular values is written $\nu_{A^{(N)}} := \mu_{(A^{(N)})^T A^{(N)}}$. A fundamental assumption for invoking tools

from Free Probability Theory, is the assumption of *asymptotic freeness*. Without defining the notion, which can be found in [MS17], let us describe the important computation it allows, discovered in the seminal work of Voiculescu [Voi87]. Given two sequences of square matrices $A^{(N)}, B^{(N)}$ in $M_N(\mathbb{R})$, with converging spectral measures:

$$\lim_{N \rightarrow \infty} \nu_{A^{(N)}} = \nu_A, \quad \lim_{N \rightarrow \infty} \nu_{B^{(N)}} = \nu_B,$$

we have that, under the assumption of asymptotic freeness:

$$\lim_{N \rightarrow \infty} \nu_{A^{(N)}B^{(N)}} = \nu_A \boxtimes \nu_B,$$

where \boxtimes is a deterministic operation between measures called multiplicative free convolution. The letter A (as well as B) does not correspond to a limiting matrix but to an abstract operator, with associated spectral measure μ_A and measure of squared singular values ν_A . For such limiting operators, we drop the superscript (N) .

Under suitable assumptions which are motivated and detailed later following the works of [PSG18, HN19, Pas20, PS20, CH21], for all $\ell = 1, \dots, L$, the measures $\nu_{W_\ell^{(N)}}$ and $\nu_{D_\ell^{(N)}}$ will respectively converge to ν_{W_ℓ} and ν_{D_ℓ} . In the limit, asymptotic freeness will also hold. Therefore, we will see that the measure of interest is:

$$\lim_{N \rightarrow \infty} \nu_{J^{(N)}} = \nu_J := \nu_{D_L} \boxtimes \nu_{W_L} \boxtimes \dots \boxtimes \nu_{D_1} \boxtimes \nu_{W_1}. \quad (5.1.7)$$

The goal of this paper is to give a very fast and stable computation of this measure, in the more general setup of rectangular matrices. This allows for a number of insights on the stability of back-propagation and paves the road towards more automatic tuning of neural networks.

5.1.1 Contributions

More precisely, we aim at streamlining the approach of Pennington et al. by providing the tools for a systematic use of FPT. The contributions of this paper can be categorized as:

- Theoretical: At this level, the contribution is incremental. In Pennington et al., a constant width is assumed. We generalize the model to allow for a width profile, which reflects how neural networks are designed in practice. This requires us to develop a rectangular multiplicative free convolution.

We also do not assume that the non-linearity ϕ_ℓ is the same throughout the neural network.

- Numerical: This is our main contribution. We propose and implement an efficient computational scheme for computing spectral densities. The method relies on adaptive inversions of S -transforms using the Newton-Raphson algorithm. If the Newton-Raphson is only local, we achieve a global resolution by chaining basins of attractions, thanks to a doubling strategy.

Interestingly, even in the FPT community, inverting S -transforms has been considered impossible to realize in practice [BGD08]. Instead, the focus was on combinatorial methods that use moments, or fixed-point algorithms [ATV⁺20, MNN⁺20, Tar20]. Our algorithm has a high performance, which allows for a large body of experiments.

A complete implementation is provided in the Github repository

<https://github.com/redachhaibi/FreeNN>¹.

- Experiments: We compare the learning curves for various design choices in neural networks. We focus specifically on the comparison of Relu vs Hard tanh vs Hard Sine.

5.1.2 Structure of the paper

We start in Section 5.2 by stating facts from Free Probability Theory. Most of it is available in the literature, except the definition of the product of rectangular free matrices. There, we establish in the rectangular setting an analogue of Eq. (5.1.7) in Theorem 5.2.3. Although it is probably known to specialists, the specifics were never made explicit.

In Section 5.3, we explain in detail the FPT model for random neural networks. Then thanks to the results of [Pas20, PS20] and our rectangular setting, we show that the spectral measure of the Jacobian $J^{(\mathbf{N})}$ converges to ν_J and we encode the limit in explicit generating series in Theorem 5.3.1. This gives how ν_J can *theoretically* be recovered.

Section 5.4 presents the numerical resolution which inverts the generating series. There, we start by a quick primer on the classical Newton-Raphson scheme, give Kantorovich's criterion for detecting (local) basins of attraction and then give our algorithm. By chaining different basins of attractions, we obtain a global resolution method.

Finally Section 5.5 gives benchmarks. We benchmark the numerical method, then present experiments using real neural networks.

5.2 Free Probability

5.2.1 Definitions and notations

Free Probability Theory provides a framework to analyze eigenvalues and singular values of large random matrices. Let us now introduce various complex-analytic generating series which encode the measures and the basic operations on them. First, the Cauchy-Stieltjes transform of μ , a probability measure on \mathbb{R}_+ is:

$$G_\mu : \mathbb{C}_+ \rightarrow \mathbb{C}_- \\ z \mapsto \int_{\mathbb{R}_+} \frac{\mu(dv)}{z-v} ,$$

where

$$\mathbb{C}_\pm := \{z \in \mathbb{C} \mid \pm \operatorname{Im} z > 0\} .$$

¹At the moment of writing these lines, the github repository "FreeNN" is private. It will be made public upon arxivng or submitting the paper, with a standard Apache 2.0 license.

The transform G_μ encodes the measure μ and reciprocally, the measure can be recovered thanks to:

Lemma 5.2.1 (Cauchy-Stieltjes inversion formula – Theorem 6 in [MS17]). *We have the weak convergence of probability measures:*

$$\lim_{y \rightarrow 0} -\frac{1}{\pi} \operatorname{Im} G_\mu(x + iy) dx = \mu(dx) .$$

The moment generating function is

$$M_\mu(z) = zG_\mu(z) - 1 = \sum_{k=1}^{+\infty} \frac{m_k(\mu)}{z^k},$$

where for all $k \in \mathbb{N}$, $m_k(\mu) := \int_{\mathbb{R}} x^k \mu(dx)$ is the k -th moment of μ . For $\mu \neq \delta_0$, M_μ is invertible in the neighborhood of ∞ and the inverse is denoted by $M_\mu^{(-1)}$. The S-transform of μ is defined as

$$S_\mu(m) = \frac{1 + m}{m M_\mu^{(-1)}(m)},$$

and is analytic in a neighborhood of $m = 0$. Furthermore, the variable z will always denote an element of \mathbb{C}_+ , while the variables g and m will denote elements in the image of G_μ and M_μ .

For a diagonalizable matrix $A^{(N)} \in M_N(\mathbb{R})$, we write $S_{A^{(N)}} := S_{\mu_{A^{(N)}}}$, $G_{A^{(N)}} := G_{\mu_{A^{(N)}}}$, $M_{A^{(N)}} := M_{\mu_{A^{(N)}}}$.

A landmark result in the field introduces free multiplicative convolution in a natural way, and shows that this operation is linearized by the S -transform:

Theorem 5.2.2 (Voiculescu, [Voi87]). *Consider two sequences of positive matrices, each element in $M_N(\mathbb{R})$*

$$\left(A^{(N)} ; N \geq 1 \right) , \quad \left(B^{(N)} ; N \geq 1 \right) ,$$

such that:

$$\lim_{N \rightarrow \infty} \mu_{A^{(N)}} = \mu_A , \quad \lim_{N \rightarrow \infty} \mu_{B^{(N)}} = \mu_B .$$

Under the assumption of asymptotic freeness for $A^{(N)}$ and $B^{(N)}$, there exists a deterministic probability measure $\mu_A \boxtimes \mu_B$ such that:

$$\lim_{N \rightarrow \infty} \mu_{A^{(N)} B^{(N)}} = \mu_A \boxtimes \mu_B .$$

The operation \boxtimes is the multiplicative free convolution. Moreover

$$S_{AB}(m) = S_A(m) S_B(m) . \tag{5.2.1}$$

This convergence akin to a law of large numbers is the key ingredient which allows to build the deterministic model for the back-propagation of gradients in Eq. (5.1.7).

5.2.2 Product of rectangular free matrices

As a generalization of Eq. (5.2.1) to rectangular matrices, we state:

Theorem 5.2.3. *Let $(p_N)_{N \geq 1}, (q_N)_{N \geq 1}, (r_N)_{N \geq 1}$, be three sequences of integers satisfying*

$$p_N, q_N, r_N \xrightarrow{N \rightarrow \infty} \infty, \quad \frac{r_N}{q_N} \xrightarrow{N \rightarrow \infty} c > 0 .$$

Consider for all $N \geq 1$ let $A^{(N)}, B^{(N)}$ be random matrices of respective sizes $p_N \times q_N$ and $q_N \times r_N$ such that the (squared) singular laws of $A^{(N)}, B^{(N)}$ converge weakly. Assuming that $B^{(N)} (B^{(N)})^T$ and $(A^{(N)})^T A^{(N)}$ are asymptotically free, we have that in the limit $N \rightarrow \infty$:

$$S_{(AB)^T AB}(m) = S_{A^T A}(cm) S_{B^T B}(m) .$$

Proof. See supplementary material, Subsection 5.8.3. □

Implicitly this defines a rectangular multiplicative free convolution, which could be denoted \boxtimes_c in the spirit of the rectangular free additive convolution [BG09]. But, in the current setting, this is not a good idea. Indeed, if one defines $\mu_1 \boxtimes_c \mu_2$ as the measure whose S -transform is $S_{\mu_1}(c \cdot) S_{\mu_2}$, then a quick computation shows that \boxtimes_c is not associative i.e. for a triplet (μ_1, μ_2, μ_3) of probability measures and a pair $(c_1, c_2) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$, we generically have:

$$\mu_1 \boxtimes_{c_1} (\mu_2 \boxtimes_{c_2} \mu_3) \neq (\mu_1 \boxtimes_{c_1} \mu_2) \boxtimes_{c_2} \mu_3 .$$

A better idea is to treat the dimension ratio c as part the data:

Definition 5.2.4. On the set of pairs (μ, c) such that μ is a probability measure on \mathbb{R}_+ and $c \in \mathbb{R}_+^*$, define the operation \boxtimes as:

$$(\mu_1, c_1) \boxtimes (\mu_2, c_2) := (\nu, c_1 c_2) ,$$

where ν is the unique probability measure such that $S_\nu = S_{\mu_1}(c_2 \cdot) S_{\mu_2}$. This extends the classical definition as the usual free convolution is recovered with:

$$(\mu_1, 1) \boxtimes (\mu_2, 1) := (\mu_1 \boxtimes \mu_2, 1) .$$

Such an operation is associative and will allow a neat formulation of the measure of interest in the upcoming Theorem 5.3.1, entirely analogous to Eq. (5.1.7).

5.3 Theoretical resolution of the model

Let us start by describing the model and the underlying assumptions.

5.3.1 The model

Width profile: Pennington et al. [PSG18] consider $N_\ell = N$ for $\ell = 1, 2, \dots, N$. Here, we consider that the width of layers is not uniform in the depth ℓ , which is mostly the case in practice. Indeed, many classical architectures have a specified topology. For example, classifiers start very wide with a narrow end. Auto-encoders start wide and get narrower and narrower in the encoding phase, then reach a bottle-neck which is interpreted as a latent space, and then start widening again until the initial dimension is reached.

Let us assume that we are in the infinite width regime in the sense that $N_\ell \rightarrow \infty$, for all $\ell = 0, 1, 2, \dots, L$:

$$\frac{N_{\ell-1}}{N_\ell} \xrightarrow{N \rightarrow \infty} \lambda_\ell \geq 0 ,$$

and let us denote

$$\Lambda_\ell := \lim_{N \rightarrow \infty} \frac{N_0}{N_\ell} = \prod_{k=1}^{\ell} \lambda_k ,$$

with the convention $\Lambda_0 = 1$.

FPT limits: Throughout the paper, we write \mathcal{N} for a standard Gaussian random variable on the real line \mathbb{R} :

$$\mathbb{P}(\mathcal{N} \in dx) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx .$$

An important subtlety is that $D_\ell^{(\mathbf{N})}$ is diagonal with entries $\phi'_\ell([h_\ell]_i)$ (see Eq. (5.1.2)), and the pre-activations

$$h_\ell = W_\ell^{(\mathbf{N})} x_{\ell-1} + b_\ell^{(\mathbf{N})} = W_\ell^{(\mathbf{N})} \phi_\ell(h_{\ell-1}) + b_\ell^{(\mathbf{N})}$$

clearly depend on the previous layers. Because of this lack of independence, the standard results of FPT cannot be applied directly i.e. asymptotic freeness does not obviously hold. Based on an information propagation argument, the papers [PLR⁺16, SGGSD16] argue that the entries of h_ℓ behave as the i.i.d. samples of a Gaussian distribution with zero mean and variance q^ℓ . A basic law of large numbers applied to Eq. (5.1.2) gives a limit for the empirical measure:

$$\mu_{D_\ell} = \lim_{N_\ell \rightarrow \infty} \mu_{D_\ell^{(\mathbf{N})}} = \phi'_\ell \left(\sqrt{q^\ell} \mathcal{N} \right) .$$

Also the recurrence for the variance is:

$$q^\ell = f_\ell(q^{\ell-1}) = \sigma_{W_\ell}^2 \mathbb{E} \left[\phi_\ell \left(\sqrt{q^{\ell-1}} \mathcal{N} \right)^2 \right] + \sigma_{b_\ell}^2, \quad (5.3.1)$$

with initial condition $q^1 = \frac{\sigma_{W_1}^2}{N_1} \sum_{i=1}^{N_1} (x_0^i)^2 + \sigma_{b_1}^2$.

Recently in Pastur et al. completed this heuristic thanks to a swapping trick – see [Pas20, Lemma 3.3] and [PS20, Remark 3.4]. They basically proved that, regarding the asymptotical

spectral properties of $J^{(\mathbf{N})}$, one can replace each $D_\ell^{(\mathbf{N})}$ by a diagonal matrix with independent Gaussian entries $\sqrt{q^\ell \mathcal{N}}$ independent from the rest. In that setting, one can apply the results on products of asymptotically free matrices which were given in Section 5.2.

5.3.2 Resolution and formulas

We then have the following result in the infinite width limit.

Theorem 5.3.1. *In terms of the rectangular multiplicative free convolution, the measure of (squared) singular values of $J^{(\mathbf{N})}$ converges to*

$$\nu_J = (\nu_{D_L}, 1) \boxtimes (\nu_{W_L}, \lambda_L) \boxtimes \cdots \boxtimes (\nu_{D_1}, 1) \boxtimes (\nu_{W_1}, \lambda_1) . \quad (5.3.2)$$

Moreover, the S -transform of $J^T J$ in the infinite width regime verifies

$$S_{J^T J}(m) = \prod_{\ell=1}^L \left[S_{D_\ell^2}(\Lambda_\ell m) S_{W_\ell^T W_\ell}(\Lambda_{\ell-1} m) \right] . \quad (5.3.3)$$

In particular, under the assumption that the entries of W_ℓ are i.i.d. :

$$S_{J^T J}(m) = \prod_{\ell=1}^L \left(S_{D_\ell^2}(\Lambda_\ell m) \frac{1}{\sigma_{W_\ell}^2} \frac{1}{1 + \Lambda_\ell m} \right) ,$$

$$M_{J^T J}^{(-1)}(m) = \frac{m+1}{m} \prod_{\ell=1}^L \frac{\sigma_{W_\ell}^2 (1 + \Lambda_\ell m)}{S_{D_\ell^2}(\Lambda_\ell m)} .$$

Proof. See supplementary material, Subsection 5.8.3. □

Let us now collect the various required formulas, and specialize them to a selection of nonlinearities: ReLu, Hard Tanh and Hard Sine. As discussed in the introduction, the empirical distribution of $D_\ell^{(\mathbf{N})}$ converges to the law of $\phi'_\ell(\sqrt{q^\ell \mathcal{N}})$. From this observation was deduced in [PSG18] the following formula :

$$M_{D_\ell^2}(z) = \sum_{k \geq 1} \frac{m_k(D_\ell^2)}{z^k} = \mathbb{E} \left[\frac{\phi'_\ell(\sqrt{q^\ell \mathcal{N}})^2}{z - \phi'_\ell(\sqrt{q^\ell \mathcal{N}})^2} \right] , \quad (5.3.4)$$

$$S_{D_\ell^2}(m) = \frac{1+m}{m M_{D_\ell^2}^{(-1)}(m)} , \quad (5.3.5)$$

The reader will find in Table 5.1 the computations of $M_{D_\ell^2}$, $S_{D_\ell^2}$ and the recurrence relation given by f_ℓ where

$$f_\ell(q^{\ell-1}) = \sigma_{W_\ell}^2 g_\ell(q^{\ell-1}) + \sigma_{b_\ell}^2 , \text{ with } g_\ell(q) := \mathbb{E} \left[\phi_\ell(\sqrt{q} \mathcal{N})^2 \right] ,$$

and

$$C_\ell = \mathbb{P} \left(0 \leq \mathcal{N} \leq \frac{1}{\sqrt{q^\ell}} \right) .$$

	Linear	ReLU	Hard Tanh	Hard sine / Triangle
$\phi_\ell(h)$	h	$[h]_+$	$[h + 1]_+ - [h - 1]_+ - 1$	$\frac{2}{\pi} \arcsin \circ \sin(\frac{\pi}{2}h)$
$M_{D_\ell^2}(z)$	$\frac{1}{z-1}$	$\frac{1}{2} \frac{1}{z-1}$	$C_\ell \frac{1}{z-1}$	$\frac{1}{z-1}$
$m_k(D_\ell^2)$	1	$\frac{1}{2}$	C_ℓ	1
$S_{D_\ell^2}(m)$	1	$\frac{1+m}{\frac{1}{2}+m}$	$\frac{1+m}{C_\ell+m}$	1
$g_\ell(q)$	q	$\frac{q}{2}$	$2qC_\ell - \sqrt{\frac{2q}{\pi}} e^{-\frac{1}{2q}}$	$\frac{1}{3} + \frac{4}{\pi^2} \sum_{n \geq 1} \frac{(-1)^n}{n^2} e^{-q \frac{\pi^2 n^2}{2}}$

Table 5.1 – Table of formulas for moment generating functions and S -transforms.

5.3.3 Master equation

In the end, we only need to fix widths and non-linearities in order to form $M_J^{(-1)}(m)$. This is the master equation which we solve numerically thanks to an adaptive Newton-Raphson scheme.

The non-linearities ReLu, Hard Tanh and Hard Sine yield similar formulas. If W_ℓ has i.i.d. entries,

$$M_{J^T J}^{(-1)}(m) = \frac{m+1}{m} \prod_{\ell=1}^L \sigma_{W_\ell}^2 (c_\ell + \Lambda_\ell m) , \quad (5.3.6)$$

where $c_\ell = \frac{1}{2}$ when ϕ_ℓ is ReLu, $c_\ell = C_\ell = \mathbb{P}\left(0 \leq \mathcal{N} \leq \frac{1}{\sqrt{q^\ell}}\right)$ if ϕ_ℓ is Hard Tanh and $c_\ell = 1$ if ϕ_ℓ is Hard Sine.

5.3.4 Moments of J

We can reach an early understanding of the behavior of J 's singular values by computing mean and variance. For ease of notation, we write:

$$m_k^{(s)}(A) = m_k(A^T A)$$

for any operator A , which admits a measure of singular values. We have:

$$m_1^{(s)}(J) = \prod_{\ell=1}^L \left(m_1^{(s)}(D_\ell) m_1^{(s)}(W_\ell) \right) = \prod_{\ell=1}^L (c_\ell \sigma_{W_\ell}^2) , \quad (5.3.7)$$

$$\begin{aligned} & m_2^{(s)}(J) - m_1^{(s)}(J) \\ &= m_1^{(s)}(J)^2 \left(\sum_{\ell=1}^L \Lambda_\ell \left(\frac{m_2^{(s)}(D_\ell) - m_1^{(s)}(D_\ell)^2}{m_1^{(s)}(D_\ell)^2} + \frac{m_2^{(s)}(W_\ell) - m_1^{(s)}(W_\ell)^2}{m_1^{(s)}(W_\ell)^2} \right) \right) \end{aligned} \quad (5.3.8)$$

$$= \left(\prod_{\ell=1}^L c_\ell^2 \sigma_{W_\ell}^4 \right) \left(\sum_{\ell=1}^L \Lambda_\ell \left(\frac{1 - c_\ell}{c_\ell} + \lambda_\ell \right) \right) .$$

The proofs of these formulas are given in Supplementary material, Subsection 5.8.1. These formulas need to be interpreted:

- Variance grows with L , showing increased instability with depth.
- Larger layers, relative to N_0 , give larger Λ_ℓ 's and thus the same effect.

5.4 Numerical resolution

In this section, we completely describe the numerical scheme aimed at computing the spectral density of $J^T J$ in Eq. (5.1.1).

Here are the steps:

- Because of the Cauchy-Stieltjes inversion formula given in Lemma 5.2.1, pick a small $y > 0$ in order to compute:

$$-\frac{1}{\pi} \operatorname{Im} G_{J^T J}(z = x + iy) .$$

The smaller the better, and in practice our method works for up to $y = 10^{-10}$.

- Because

$$M(z) = zG(z) - 1 ,$$

we need to compute $M(z)$.

- $M^{(-1)}(m)$ is available thanks to the master equation in Theorem (5.3.1). Therefore, we need to invert $m \mapsto M^{(-1)}(m)$.

Clearly, this last step is the crucial one. And this strategy has been deemed hopeless by many mathematicians working in Free Probability Theory, see [BGD08, p.218].

5.4.1 On the classical Newton-Raphson scheme

First, let us start by recalling the classical Newton-Raphson scheme in order to invert the equation $z = f(m)$ where $z \in \mathbb{C}_+$ is fixed and f is rational. A neat trick which leverages the fact that f is rational and that $z \in \mathbb{C}_+$ is to define:

$$\varphi_z(m) := P(m)/z - Q(z) . \tag{5.4.1}$$

As such, we have:

$$z = f(m) = \frac{P(m)}{Q(m)} \iff \varphi_z(m) = 0 .$$

There are several advantages of doing that:

- Inversion is recast into finding the zero of a polynomial function.
- Since we have $\lim_{z \rightarrow \infty} M(z) = 0$, if z is large in modulus, $m = 0$ is a natural starting point for the algorithm when z is large.

In the end, the pseudo-code is given in Algorithm 5.4.1.

Algorithm 5.4.1 Newton-Raphson scheme for a rational function f

Name: NEWTON_RAPHSON

Input:

Numerical precision: $\varepsilon > 0$ (Default: 10^{-12}),

Image value: $z \in \mathbb{C}_+$,

Polynomials: P, Q such that $f = \frac{P}{Q}$,

(Optional) Guess: $m_0 \in \mathbb{C}$, (Default: $m_0 = 0$).

Code:

$m \leftarrow m_0$

while True **do**

 value $\leftarrow \varphi_z(m)$ # See Eq. (5.4.1)

if |value| $< \varepsilon$ **then**

return m

end if

 grad $\leftarrow \varphi'_z(m)$

$m \leftarrow m - \text{value}/\text{grad}$

end while

It is well-known that the Newton-Raphson scheme fails unless the initial guess $m_0 \in \mathbb{C}$ belongs to a basin of attraction for the method. And, provided such a guarantee, the Newton-Raphson scheme is exceptionnally fast with a quadratic convergence speed (the convergence of a sequence $(u_n)_n$ towards a limit l is said to be quadratic if the sequence $\left(\frac{\|u_{n+1}-l\|}{\|u_n-l\|^2}\right)_n$ is bounded from above). Kantorovich's seminal work in 1948 provides such a guarantee locally.

5.4.2 Kantorovich's criterion

Here we give the optimal criterion from [GT74] adapted to this paper. Fix $z \in \mathbb{C}_+$ and recall that φ_z in Eq. (5.4.1) is the map whose zero we want to find.

Theorem 5.4.1 (Kantorovich's criterion, [Kan48]). *Consider a starting point $m_0 \in \mathbb{C}$, and define:*

$$\delta := \left| \frac{\varphi_z(m_0)}{\varphi'_z(m_0)} \right|, \quad \kappa := \left| \frac{1}{\varphi'_z(m_0)} \right|.$$

If the starting point satisfies $h := \delta\kappa\lambda < \frac{1}{2}$, where

$$\lambda := \sup_{|m-m_0| \leq t^*} |\varphi''_z(m)|, \quad t^* := \frac{2\delta}{1 + \sqrt{1-h}} < 2\delta.$$

Then, the Newton-Raphson scheme, starting from m_0 converges to m^ such that $\varphi_z(m^*) = 0$. Furthermore, the convergence at each step is at least quadratic.*

Therefore, we assume that we have at our disposal a function

$$\text{IS_IN_BASIN}(z, m)$$

which indicates if the Kantorovich criterion is satisfied for φ_z at any $m \in \mathbb{C}$. It is particularly easy to program with φ_z polynomial.

5.4.3 Doubling strategies and chaining

Now we have all the (local) ingredients in order to describe a global strategy which solves in $m \in \mathbb{C}$ the equation:

$$\varphi_z(m) = 0 .$$

First, one has to notice that this problem is part of a family parametrized by $z \in \mathbb{C}_+$. And the solution is $m \approx 0$ for z large. Therefore, one can find a proxy solution for $z \in \mathbb{C}_+$ high enough. This is done thanks to a doubling strategy until a basin of attraction is reached.

Second, if a proxy (z, m) is available, we can use the Newton-Raphson algorithm to find a solution $(z + \Delta z, m + \Delta m)$ starting from m . To do so, we need Δz small enough. This on the other hand is done by dichotomy.

Tying the pieces together allows to chain the different basins of attraction and leads to Algorithm 5.4.2. Notice that in the description of the algorithm, we chose to make implicit the dependence in the function f , since it is only passed along as a parameter. Technically, f is a parameter for all three functions

- NEWTON__RAPHSON,
- IS__IN__BASIN,
- NEWTON__LILYPADS.

Algorithm 5.4.2 Newton lilypads, chaining basins of attraction

Name: NEWTON_LILYPADS

Input:

Image value: $z_{objective} \in \mathbb{C}_+$,

(Optional) Proxy: $(z_0, m_0) \in \mathbb{C}_+ \times \mathbb{C}$.

Code:

Find a proxy $(z_0, m_0 = 0)$ using a doubling strategy, if None given

if (z_0, m_0) is None **then**

$m \leftarrow 0$

$z \leftarrow z_{objective}$

while not IS_IN_BASIN(z, m) **do**

$z \leftarrow z + i \operatorname{Im}(z)$

end while

$m \leftarrow \text{NEWTON_RAPHSON}(z, \text{Guess} = m)$

else

$(z, m) \leftarrow (z_0, m_0)$

end if

Starts heading towards $z_{objective}$ using dichotomy

while $|z_{objective} - z| > 0$ **do**

$\Delta z \leftarrow z_{objective} - z$

while not IS_IN_BASIN($z + \Delta z, m$) **do**

$\Delta z \leftarrow 0.5 * \Delta z$

end while

$z \leftarrow z + \Delta z$

$m \leftarrow \text{NEWTON_RAPHSON}(z, \text{Guess} = m)$

end while

return m

5.5 Benchmarks

5.5.1 Computational speed

A table of the computational time required for computing N density points is given in Table 5.2. Notice that the recorded time scales sublinearly with the number of required points. This is easily understood by the fact that smaller N requires the computation of more basins of attraction per point.

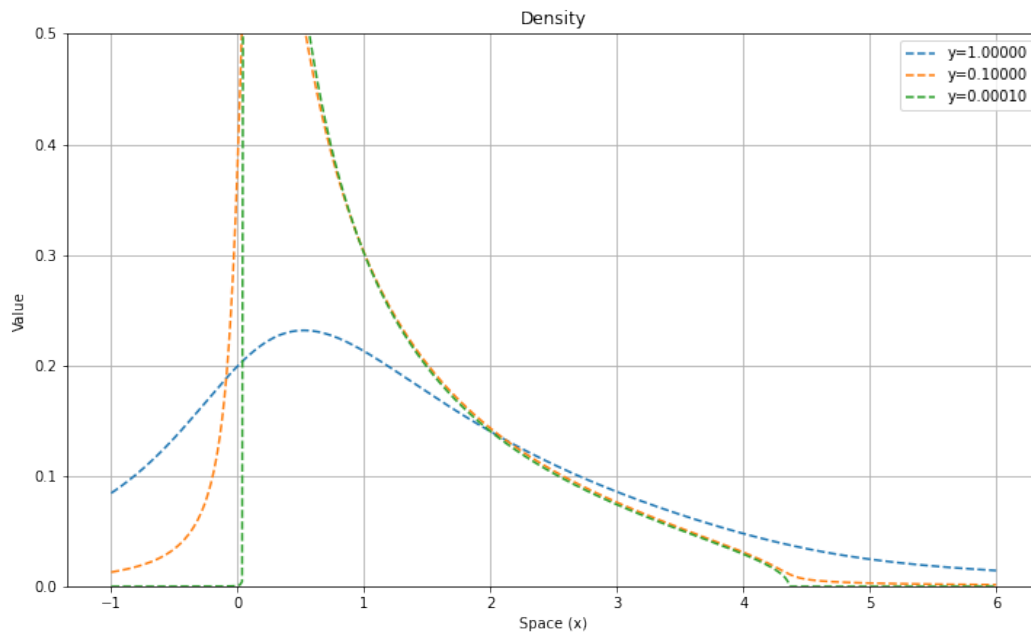
N	100	200	400	1000
Time (ms)	38.1	57.4	95.4	220.7

Table 5.2 – Table of computational time, on a laptop with an Intel Core i7-9850H CPU, for computing N density points. Nearest points are used as proxy for the Newton-Raphson scheme.

5.5.2 Plots

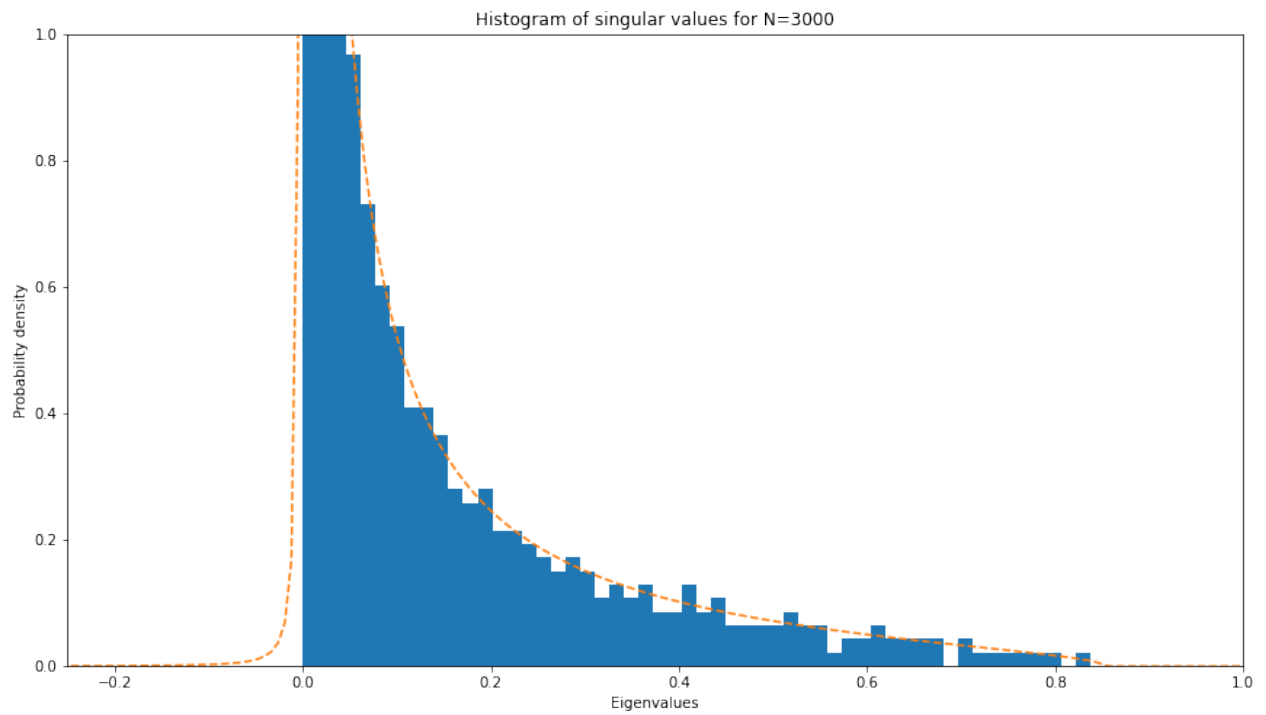
Now we give multiple plots of the technique’s results.

A first plot in Figure 5.1 shows the same target distribution but convolved with various Cauchy distributions $y\mathcal{C}$ where $y \in \{1, 10^{-1}, 10^{-4}\}$. This corresponds to computing the density $\frac{-1}{\pi} \text{Im} G_\mu(\cdot + iy)$ for different y ’s.



Plot of $\frac{-1}{\pi} \text{Im} G_\mu(\cdot + iy)$ for $y \in \{1, 10^{-1}, 10^{-4}\}$. Here μ is the multiplicative free convolution of three Marchenko-Pastur distributions, with different parameters.

A second plot in Figure 5.2 which compares the output of our method versus a Monte-Carlo sampling with matrices of size $N = 3000$. Not only Monte-Carlo is imprecise because of the noise, but its performance scales very poorly with N since one needs to diagonalize ever larger matrices. Of course, Monte-Carlo remains the easiest method to implement.



Plot of density of singular values for the Jacobian matrix J . The network has constant width and Relu non-linearities. Monte-Carlo sampling uses matrices of size $N_\ell = N = 3000$. Computation time is 22467.3 ms with the same laptop as the one used for Table 5.2.

5.5.3 Description of the experiments

At this stage, we are still in the process of designing experiments with real neural networks, together with the deterministic model given by FPT. The idea is to consider various network architectures and compare the stability of the learning process thanks to learning curves. In particular, it is desirable to have a comparable format to [GB10, PSG18].

To do so, we also need to stabilize our current code base.

5.6 Conclusion

In this paper, we gave an efficient numerical method in FPT geared towards the application in machine learning. The computation method itself is fast and reliable, which can be used to automatically tune some design choices in neural networks such as the width profile, and initialization variances.

We finish by pointing out two methods commonly used for mitigating gradient explosion and gradient vanishing. The former is very easily included in the current setup. The latter is an interesting direction for future investigations.

Gradient clipping: This is the most natural heuristic method to alleviate gradient explosion. At the theoretical level, this only amounts to modify the non-linearity "program" φ . The automatically computed gradient φ' , say via autograd if one uses Pytorch [PGC⁺17], is replaced by a surrogate gradient:

$$\tilde{\varphi}' = \varphi' \mathbf{1}_{\{|\varphi'| \leq C\}} ,$$

where C is a clipping threshold.

Skip-connections: When putting in place skip-connections, the goal is to alleviate the problem of gradient vanishing. In terms of equations, the recurrence changes via:

$$x_\ell = \varphi \left(\sum_{i=1}^{K-1} W_{\ell,i}^{(\mathbf{N})} x_{\ell-i} + b_\ell^{(\mathbf{N})} \right) ,$$

where K is the maximum lag and $W_{\ell,i}^{(\mathbf{N})} \in M_{N_\ell, N_{\ell-i}}$ is a weight matrix.

Compared to the framework of this paper, the chain rule used for back-propagation changes in a very fundamental way. The Jacobian cannot be approximated by a simple product of free matrices: the same free variable will appear at multiple locations.

5.7 Acknowledgements

R.C. recognizes the support of French ANR STARS held by Guillaume Cébron.

The authors would like to thank Mireille Capitaine, Guillaume Cébron and Fabrice Gamboa for fruitful conversations.

5.8 Supplementary material

5.8.1 Computations of moments

The following remark is useful in the computation of moments.

Remark 5.8.1 (Moments). At the neighborhood of $z \sim \infty$:

$$M_\mu(z) = \frac{m_1(\mu)}{z} + \frac{m_2(\mu)}{z^2} + \mathcal{O}(z^{-3}) .$$

By inversion, at the neighborhood of $m \sim 0$:

$$\begin{aligned} M_\mu^{(-1)}(m) &= \frac{m_1(\mu)}{m} + \mathcal{O}(1) . \\ M_\mu^{(-1)}(m) &= \frac{m_1(\mu)}{m} + \frac{m_2(\mu)}{m_1(\mu)} + \mathcal{O}(m) . \end{aligned}$$

Hence:

$$\begin{aligned} S_\mu(m) &= \frac{1+m}{m_1(\mu) + m \frac{m_2(\mu)}{m_1(\mu)} + \mathcal{O}(m^2)} \\ &= \frac{1}{m_1(\mu)} (1+m) \left(1 - m \frac{m_2(\mu)}{m_1(\mu)^2} + \mathcal{O}(m^2) \right) \\ &= \frac{1}{m_1(\mu)} + \frac{m}{m_1(\mu)} \left(1 - \frac{m_2(\mu)}{m_1(\mu)^2} \right) + \mathcal{O}(m^2) . \end{aligned}$$

Thanks to this, we can prove Eq. (5.3.7) and (5.3.8). By Remark 5.8.1 and Theorem 5.3.1, we have as $m \rightarrow 0$:

$$\begin{aligned} S_{J^T J}(m) &= \prod_{\ell=1}^L [S_{D_\ell^2}(\Lambda_\ell m) S_{W_\ell^T W_\ell}(\Lambda_\ell m)] \\ &= \prod_{\ell=1}^L \left[\left(\frac{1}{m_1^{(s)}(D_\ell)} + m \frac{\Lambda_\ell}{m_1^{(s)}(D_\ell)} \left(1 - \frac{m_2^{(s)}(D_\ell)}{m_1^{(s)}(D_\ell)^2} \right) + \mathcal{O}(m^2) \right) \right. \\ &\quad \left. \left(\frac{1}{m_1^{(s)}(W_\ell)} + m \frac{\Lambda_\ell}{m_1^{(s)}(W_\ell)} \left(1 - \frac{m_2^{(s)}(W_\ell)}{m_1^{(s)}(W_\ell)^2} \right) + \mathcal{O}(m^2) \right) \right] \\ &= \left[\prod_{\ell=1}^L \frac{1}{m_1^{(s)}(D_\ell) m_1^{(s)}(W_\ell)} \right] \prod_{\ell=1}^L \left[1 + m \Lambda_\ell \left(2 - \frac{m_2^{(s)}(D_\ell)}{m_1^{(s)}(D_\ell)^2} - \frac{m_2^{(s)}(W_\ell)}{m_1^{(s)}(W_\ell)^2} \right) + \mathcal{O}(m^2) \right] . \end{aligned}$$

Identifying the first order term, one finds indeed Eq. (5.3.7). Continuing the previous computation:

$$S_{J^T J}(m) = \frac{1}{m_1^{(s)}(J)} + \frac{m}{m_1^{(s)}(J)} \left(\sum_{\ell=1}^L \Lambda_\ell \left(2 - \frac{m_2^{(s)}(D_\ell)}{m_1^{(s)}(D_\ell)^2} - \frac{m_2^{(s)}(W_\ell)}{m_1^{(s)}(W_\ell)^2} \right) \right) + \mathcal{O}(m^2) .$$

Applying Remark 5.8.1 again for $S_{J^T J}$, we get :

$$\sum_{\ell=1}^L \Lambda_\ell \left(2 - \frac{m_2^{(s)}(D_\ell)}{m_1^{(s)}(D_\ell)^2} - \frac{m_2^{(s)}(W_\ell)}{m_1^{(s)}(W_\ell)^2} \right) = 1 - \frac{m_2^{(s)}(J)}{m_1^{(s)}(J)^2}$$

which is equivalent to Eq. (5.3.8).

5.8.2 Specializing to classical weight distributions

Under the assumption that the entries of W_ℓ are i.i.d. we have $m_1^{(s)}(W_\ell) = \sigma_{W_\ell}^2$ and $m_2^{(s)}(W_\ell) = \sigma_{W_\ell}^4(1 + \lambda_\ell)$ which gives

$$\begin{aligned} m_2^{(s)}(J) &= \left(1 - \sum_{\ell=1}^L \left(\Lambda_\ell \left(1 - \frac{m_2^{(s)}(D_\ell)}{m_1^{(s)}(D_\ell)^2} - \lambda_\ell \right) \right) \right) \prod_{\ell=1}^L \left(m_1^{(s)}(D_\ell)^2 \sigma_{W_\ell}^4 \right) \\ m_2^{(s)}(J) - (m_1^{(s)}(J))^2 &= \sum_{\ell=1}^L \left(\Lambda_\ell \left(\frac{m_2^{(s)}(D_\ell)}{m_1^{(s)}(D_\ell)^2} + \lambda_\ell - 1 \right) \right) \prod_{\ell=1}^L \left(m_1^{(s)}(D_\ell)^2 \sigma_{W_\ell}^4 \right). \end{aligned}$$

5.8.3 Proofs

Proof of Theorem 5.2.3

For $k \in \mathbb{N}$, we have :

$$\begin{aligned} \text{Tr} \left[\left((A^{(N)} B^{(N)})^T A^{(N)} B^{(N)} \right)^k \right] &= \text{Tr} \left[\left((B^{(N)})^T (A^{(N)})^T A^{(N)} B^{(N)} \right)^k \right] \\ &= \text{Tr} \left[(B^{(N)} (B^{(N)})^T (A^{(N)})^T A^{(N)})^k \right]. \end{aligned}$$

As $(A^{(N)} B^{(N)})^T A^{(N)} B^{(N)} \in M_{r_N}(\mathbb{C})$ and $B^{(N)} (B^{(N)})^T (A^{(N)})^T A^{(N)} \in M_{q_N}(\mathbb{C})$, this shows that

$$M_{(A^{(N)} B^{(N)})^T A^{(N)} B^{(N)}}(z) = \frac{q_N}{r_N} M_{B^{(N)} (B^{(N)})^T (A^{(N)})^T A^{(N)}}(z),$$

and then

$$M_{(A^{(N)} B^{(N)})^T A^{(N)} B^{(N)}}^{(-1)}(m) = M_{B^{(N)} (B^{(N)})^T (A^{(N)})^T A^{(N)}}^{(-1)} \left(\frac{r_N}{q_N} m \right).$$

Consequently,

$$\begin{aligned} S_{(A^{(N)} B^{(N)})^T A^{(N)} B^{(N)}}(m) &= \frac{1 + m}{m M_{(A^{(N)} B^{(N)})^T A^{(N)} B^{(N)}}^{(-1)}(m)} \\ &= \frac{1 + m}{m M_{B^{(N)} (B^{(N)})^T (A^{(N)})^T A^{(N)}}^{(-1)} \left(\frac{r_N}{q_N} m \right)} \times \frac{1 + \frac{r_N}{q_N} m}{\frac{r_N}{q_N} m} \times \frac{\frac{r_N}{q_N} m}{1 + \frac{r_N}{q_N} m} \end{aligned}$$

$$= \frac{r_N}{q_N} \frac{1+m}{1 + \frac{r_N}{q_N} m} S_{B^{(N)}(B^{(N)})^T (A^{(N)})^T A^{(N)}} \left(\frac{r_N}{q_N} m \right). \quad (5.8.1)$$

As $(A^{(N)})^T A^{(N)}$ and $B^{(N)} (B^{(N)})^T$ are asymptotically free, taking the limit $N \rightarrow +\infty$ and applying Voiculescu's Theorem 5.2.2, we get

$$S_{(AB)^* AB}(m) = \alpha \frac{1+m}{1+\alpha m} S_{BB^*}(\alpha m) S_{A^* A}(\alpha m).$$

Moreover, the above equality is true replacing A with the identity I . $S_I(m) = 1$ yields:

$$S_{B^* B}(m) = \alpha \frac{1+m}{1+\alpha m} S_{BB^*}(\alpha m).$$

Finally we have

$$S_{(AB)^* AB}(m) = S_{A^* A}(\alpha m) S_{B^* B}(m).$$

This concludes the proof.

Proof of Theorem 5.3.1

Thanks to a swapping trick justified in [PS20], we can assume that the matrices D_ℓ have i.i.d. entries independent from the rest of the network, distributed as $\phi'_\ell(\sqrt{q_\ell} \mathcal{N})$. Notice that we can also replace the D_ℓ 's by deterministic matrices that use the quantiles of the same distribution. This together with standard results from FPT such as [MS17] gives asymptotic freeness – see [BC17] for a more general result reflecting the current state of the art. Therefore, we can apply Theorem 5.2.3. Starting from Eq. (5.1.1), we get in the infinite width regime and by induction:

$$\begin{aligned} S_{(J^{(N)})^T J^{(N)}}(m) &= S_{\left(D_L^{(N)} W_L^{(N)} D_{L-1}^{(N)} W_{L-1}^{(N)} \dots D_1^{(N)} W_1^{(N)} \right)^T W_1^{(N)} D_1^{(N)} \dots W_{L-1}^{(N)} D_{L-1}^{(N)} W_L^{(N)} D_L^{(N)}}(m) \\ &= S_{\left(D_L^{(N)} W_L^{(N)} D_{L-1}^{(N)} W_{L-1}^{(N)} \dots D_2^{(N)} W_2^{(N)} D_1^{(N)} \right)^T D_1^{(N)} W_2^{(N)} D_2^{(N)} \dots W_{L-1}^{(N)} D_{L-1}^{(N)} W_L^{(N)} D_L^{(N)}}(\lambda_1 m) \\ &\quad \times S_{\left(W_1^{(N)} \right)^T W_1^{(N)}}(m) \\ &= S_{\left(D_L^{(N)} W_L^{(N)} D_{L-1}^{(N)} W_{L-1}^{(N)} \dots D_2^{(N)} W_2^{(N)} \right)^T W_2^{(N)} D_2^{(N)} \dots W_{L-1}^{(N)} D_{L-1}^{(N)} W_L^{(N)} D_L^{(N)}}(\lambda_1 m) \\ &\quad \times S_{\left(D_1^{(N)} \right)^2}(\lambda_1 m) S_{\left(W_1^{(N)} \right)^T W_1^{(N)}}(m) \\ &\quad \vdots \\ &= \prod_{\ell=1}^L \left[S_{\left(D_\ell^{(N)} \right)^2} \left(\prod_{k=1}^{\ell} \lambda_k m \right) S_{\left(W_\ell^{(N)} \right)^T W_\ell^{(N)}} \left(\prod_{k=1}^{\ell-1} \lambda_k m \right) \right] \end{aligned}$$

$$= \prod_{\ell=1}^L \left[S_{(D_\ell^{(\mathbf{N})})^2}(\Lambda_\ell m) S_{(W_\ell^{(\mathbf{N})})^T W_\ell^{(\mathbf{N})}}(\Lambda_{\ell-1} m) \right],$$

with the convention $\Lambda_0 = \prod_{k=1}^0 \lambda_k = 1$.

Under the assumption that the entries of W_ℓ are i.i.d., the Marcenko-Pastur Theorem gives

$$S_{(W_\ell^{(\mathbf{N})})^T W_\ell^{(\mathbf{N})}}(m) = \frac{1}{\sigma_{W_\ell}^2} \frac{1}{1 + \lambda_\ell m},$$

which leads to

$$S_{(J^{(\mathbf{N})})^T J^{(\mathbf{N})}}(m) = \prod_{\ell=1}^L \left(S_{(D_\ell^{(\mathbf{N})})^2}(\Lambda_\ell m) \frac{1}{\sigma_{W_\ell}^2} \frac{1}{1 + \Lambda_\ell m} \right).$$

We thus have

$$\begin{aligned} M_{(J^{(\mathbf{N})})^T J^{(\mathbf{N})}}^{(-1)}(m) &= \frac{m+1}{m S_{(J^{(\mathbf{N})})^T J^{(\mathbf{N})}}(m)} \\ &= \frac{m+1}{m \prod_{\ell=1}^L \left(S_{(D_\ell^{(\mathbf{N})})^2}(\Lambda_\ell m) \frac{1}{\sigma_{W_\ell}^2} \frac{1}{1 + \Lambda_\ell m} \right)} \\ &= \frac{(m+1) \prod_{\ell=1}^L \sigma_{W_\ell}^2 (1 + \Lambda_\ell m)}{m \prod_{\ell=1}^L S_{(D_\ell^{(\mathbf{N})})^2}(\Lambda_\ell m)} \\ &= \frac{(m+1) \prod_{\ell=1}^L \sigma_{W_\ell}^2 (1 + \Lambda_\ell m)}{m \prod_{\ell=1}^L S_{(D_\ell^{(\mathbf{N})})^2}(\Lambda_\ell m)}. \end{aligned}$$

Proof of formulas in Table 5.1

Hard Sine: This is the most tricky formula to establish. If $\phi(x) = \frac{2}{\pi} \arcsin \circ \sin\left(\frac{\pi}{2}x\right)$ and $\widehat{f}(\xi) = \int dx f(x) e^{i\xi x}$ is the Fourier transform on f , then the application of the Plancherel formula yields:

$$\begin{aligned} &\mathbb{E} \left[\phi(\sqrt{q}\mathcal{N})^2 \right] \\ &= \int_{-\infty}^{\infty} dx \frac{e^{-\frac{x^2}{2q}}}{\sqrt{2\pi q}} \phi^2(x) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\xi \widehat{\phi^2}(\xi) e^{-q\frac{\xi^2}{2}}. \end{aligned}$$

But in term of Fourier series:

$$\phi^2(x) = \sum_{n \in \mathbb{Z}} \left(\widehat{\phi^2} \right)_n e^{i\pi n x}$$

as $\phi^2(x) = x^2$ on $[-1, 1]$ and extended in order to become 2-periodic. In terms of Schwartz distributions:

$$\widehat{\phi^2}(\xi) = \sum_{n \in \mathbb{Z}} \left(\widehat{\phi^2}\right)_n (x \mapsto e^{i\pi n x}) = \sum_{n \in \mathbb{Z}} \left(\widehat{\phi^2}\right)_n 2\pi \delta_{-\pi n}(d\xi) .$$

Hence:

$$\mathbb{E} \left[\phi(\sqrt{q}\mathcal{N})^2 \right] = \sum_{n \in \mathbb{Z}} \left(\widehat{\phi^2}\right)_n e^{-q \frac{\pi^2 n^2}{2}} .$$

We conclude by computing the Fourier coefficients of ϕ^2 .

$$\left(\widehat{\phi^2}\right)_0 = \frac{1}{2} \int_{-1}^1 dx x^2 = \frac{1}{3} .$$

$$\begin{aligned} \left(\widehat{\phi^2}\right)_n &= \frac{1}{2} \int_{-1}^1 dx x^2 e^{-i\pi n x} \\ &= \int_0^1 dx x^2 \cos(\pi n x) \\ &= -\frac{1}{\pi n} \int_0^1 dx 2x \sin(\pi n x) \\ &= \frac{2}{(\pi n)^2} [x \cos(n\pi x)]_0^1 - \frac{2}{(\pi n)^2} \int_0^1 dx \cos(\pi n x) \\ &= \frac{2}{(\pi n)^2} (-1)^n . \end{aligned}$$

In the end:

$$\mathbb{E} \left[\phi(\sqrt{q}\mathcal{N})^2 \right] = \frac{1}{3} + \frac{4}{\pi^2} \sum_{n \geq 1} \frac{(-1)^n}{n^2} e^{-q \frac{\pi^2 n^2}{2}} .$$

Bibliography

- [AA⁺13] Antonio Auffinger, Gerard Ben Arous, et al., *Complexity of random smooth functions on the high-dimensional sphere*, Annals of Probability **41** (2013), no. 6, 4214–4247.
- [AAB⁺15] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, *TensorFlow: Large-scale machine learning on heterogeneous systems*, 2015, Software available from tensorflow.org.
- [AAČ13] Antonio Auffinger, Gérard Ben Arous, and Jiří Černý, *Random matrices and complexity of spin glasses*, Communications on Pure and Applied Mathematics **66** (2013), no. 2, 165–201.
- [ABDF11] Gernot Akemann, Jinho Baik, and Philippe Di Francesco, *The Oxford Handbook of Random Matrix Theory*, Oxford Handbooks in Mathematics, Oxford University Press, 9 2011.
- [ABMV13] Romain Allez, Jean-Philippe Bouchaud, Satya N. Majumdar, and Pierpaolo Vivo, *Invariant β -Wishart ensembles, crossover densities and asymptotic corrections to the Marčenko-Pastur law*, J. Phys. A **46** (2013), no. 1, 015001, 22. MR 3001575
- [AGZ10] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni, *An introduction to random matrices*, no. 118, Cambridge university press, 2010.
- [AS] Guillaume Aubrun and Stanisław Szarek, *Alice and bob meet banach: The interface of asymptotic geometric analysis and quantum information theory*, Bull. Amer. Math. Soc.

- [AS04] Noga Alon and Joel H Spencer, *The probabilistic method*, John Wiley & Sons, 2004.
- [AT09] Robert J Adler and Jonathan E Taylor, *Random fields and geometry*, Springer Science & Business Media, 2009.
- [ATV⁺20] Octavio Arizmendi, Pierre Tarrago, Carlos Vargas, et al., *Subordination methods for free deconvolution*, Annales de l’Institut Henri Poincaré, Probabilités et Statistiques, vol. 56, Institut Henri Poincaré, 2020, pp. 2565–2594.
- [BBP17] Joël Bun, Jean-Philippe Bouchaud, and Marc Potters, *Cleaning large correlation matrices: tools from random matrix theory*, Physics Reports **666** (2017), 1–109.
- [BC17] Serban T Belinschi and Mireille Capitaine, *Spectral properties of polynomials in independent wigner and deterministic matrices*, Journal of Functional Analysis **273** (2017), no. 12, 3901–3963.
- [BG09] Florent Benaych-Georges, *Rectangular random matrices, related convolution*, Probability Theory and Related Fields **144** (2009), no. 3-4, 471–515.
- [BGD08] Florent Benaych-Georges and Mérouane Debbah, *Free deconvolution: from theory to practice*, Paradigms for Biologically-Inspired Autonomic Networks and Services (2008), 201–224.
- [BGJJ04] Zdzislaw Burda, A Görlich, Andrzej Jarosz, and Jerzy Jurkiewicz, *Signal and noise in correlation matrix*, Physica A: Statistical Mechanics and its Applications **343** (2004), 295–310.
- [Bia97] Philippe Biane, *Free brownian motion, free stochastic calculus and random matrice, in free probability theory*, Fields Inst. Commun. **12** (1997), 1–19.
- [Bil13] Patrick Billingsley, *Convergence of probability measures*, Wiley Series in Probability and Statistics, Wiley, 2013.
- [BIPZ93] Edouard Brézin, Claude Itzykson, Giorgio Parisi, and Jean-Bernard Zuber, *Planar diagrams*, The Large N Expansion In Quantum Field Theory And Statistical Physics: From Spin Systems to 2-Dimensional Gravity, World Scientific, 1993, pp. 567–583.
- [BK13] Paul Bourgade and Jonathan P Keating, *Quantum chaos, random matrix theory, and the riemann ζ -function*, Chaos, Springer, 2013, pp. 125–168.
- [Bot98] Léon Bottou, *Online learning and stochastic approximations*, On-line learning in neural networks **17** (1998), no. 9, 142.
- [BP19] Lucas Benigni and Sandrine Péché, *Eigenvalue distribution of nonlinear models of random matrices*, arXiv preprint arXiv:1904.03090 (2019).

- [Bru89] Marie-France Bru, *Diffusions of perturbed principal component analysis*, Journal of multivariate analysis **29** (1989), no. 1, 127–136.
- [Bru91] ———, *Wishart processes*, Journal of Theoretical Probability **4** (1991), no. 4, 725–751.
- [BS98] Philippe Biane and Roland Speicher, *Stochastic calculus with respect to free brownian motion and analysis on wigner space*, Probability theory and related fields **112** (1998), no. 3, 373–409.
- [BS01] ———, *Free diffusions, free entropy and free fisher information*, Annales de l’Institut Henri Poincaré (B) Probability and Statistics, vol. 37, Elsevier, 2001, pp. 581–606.
- [BS10] Zhidong Bai and Jack W Silverstein, *Spectral analysis of large dimensional random matrices*, vol. 20, Springer, 2010.
- [BV92] Hari Bercovici and Dan-Virgil Voiculescu, *Lévy-hinčin type theorems for multiplicative and additive free convolution*, Pacific journal of mathematics **153** (1992), no. 2, 217–248.
- [BV93] Hari Bercovici and Dan Voiculescu, *Free convolution of measures with unbounded support*, Indiana University Mathematics Journal **42** (1993), no. 3, 733–773.
- [CD11] Romain Couillet and Merouane Debbah, *Random matrix methods for wireless communications*, Cambridge University Press, 2011.
- [CDM05] Mireille Capitaine and Catherine Donati-Martin, *Free wishart processes*, Journal of Theoretical Probability **18** (2005), no. 2, 413–438.
- [Cép94] Emmanuel Cépa, *Équations différentielles stochastiques multivoques*, Ph.D. thesis, Orléans, 1994.
- [Cép95] ———, *Equations différentielles stochastiques multivoques*, Séminaire de Probabilités XXIX, Springer, 1995, pp. 86–107.
- [CH21] Benoit Collins and Tomohiro Hayase, *Asymptotic freeness of layerwise jacobians caused by invariance of multilayer perceptron: The haar orthogonal case*, arXiv preprint arXiv:2103.13466 (2021).
- [CHM⁺15] Anna Choromanska, Mikael Henaff, Michael Mathieu, Gérard Ben Arous, and Yann LeCun, *The loss surfaces of multilayer networks*, Artificial intelligence and statistics, PMLR, 2015, pp. 192–204.
- [Chy06] Oleksandr Chybiryakov, *Processus de dunkl et relation de lamperti*.
- [CIR85] John C. Cox, Jonathan E. Ingersoll, and Stephen Ross, *A theory of the term structure of interest rates*, Econometrica **53** (1985), no. 2, 385–407.

- [CL97] Emmanuel Cépa and Dominique Lépingle, *Diffusing particles with electrostatic repulsion*, Probability theory and related fields **107** (1997), no. 4, 429–449.
- [CL01] ———, *Brownian particles with electrostatic repulsion on the circle: Dyson’s model for unitary random matrices revisited*, ESAIM: Probability and Statistics **5** (2001), 203–224.
- [CL20] Romain Couillet and Cosme Louart, *Concentration of solutions to random equations with concentration of measure hypotheses*.
- [CL21] Romain Couillet and Zhenyu Liao, *Random matrix theory for machine learning*, 2021.
- [DDS⁺09] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, *Imagenet: A large-scale hierarchical image database*, 2009 IEEE conference on computer vision and pattern recognition, Ieee, 2009, pp. 248–255.
- [Dem08] Nizar Demni, *Free jacobi process*, Journal of Theoretical Probability **21** (2008), no. 1, 118–143.
- [Dem09a] Nizar Demni, *Beta jacobi processes*, arXiv: Probability (2009).
- [Dem09b] Nizar Demni, *Radial Dunkl processes: existence, uniqueness and hitting time*, C. R. Math. Acad. Sci. Paris **347** (2009), no. 19-20, 1125–1128. MR 2566989
- [DG01] T Cabanal Duvillard and A Guionnet, *Large deviations upper bounds for the laws of matrix-valued processes and non-communicative entropies*, Annals of probability (2001), 1205–1261.
- [Dou05] Yan Doumerc, *Matrices aléatoires, processus stochastiques et groupes de réflexions*, Ph.D. thesis, Toulouse 3, 2005.
- [DS02] Freddy Delbaen and Hiroshi Shirakawa, *An interest rate model with upper and lower bounds*, Asia-Pacific Financial Markets **9** (2002), no. 3, 191–209.
- [Dys62] Freeman J Dyson, *A brownian-motion model for the eigenvalues of a random matrix*, Journal of Mathematical Physics **3** (1962), no. 6, 1191–1198.
- [EK09] Stewart N Ethier and Thomas G Kurtz, *Markov processes: characterization and convergence*, vol. 282, John Wiley & Sons, 2009.
- [EPR⁺10] László Erdős, Sandrine Péché, José A Ramírez, Benjamin Schlein, and Horng-Tzer Yau, *Bulk universality for wigner matrices*, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences **63** (2010), no. 7, 895–925.

- [ESY09] László Erdős, Benjamin Schlein, and Horng-Tzer Yau, *Local semicircle law and complete delocalization for wigner random matrices*, Communications in Mathematical Physics **287** (2009), no. 2, 641–655.
- [EY17] László Erdős and Horng-Tzer Yau, *A dynamical approach to random matrix theory*, vol. 28, American Mathematical Soc., 2017.
- [Fol95] Gerald B. Folland, *Introduction to partial differential equations*, Princeton university press, 1995.
- [For10] Peter J. Forrester, *Log-gases and random matrices*, London Mathematical Society Monographs Series, vol. 34, Princeton University Press, Princeton, NJ, 2010. MR 2641363
- [FW20] Zhou Fan and Zhichao Wang, *Spectra of the conjugate kernel and neural tangent kernel for linear-width neural networks*, arXiv preprint arXiv:2005.11879 (2020).
- [GB10] Xavier Glorot and Yoshua Bengio, *Understanding the difficulty of training deep feedforward neural networks*, Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings, 2010, pp. 249–256.
- [GJY03] Anja Göing-Jaeschke and Marc Yor, *A survey and some generalizations of bessel processes*, Bernoulli **9** (2003), no. 2, 313–349.
- [GM13] Piotr Graczyk and Jacek Malecki, *Multidimensional Yamada-Watanabe theorem and its applications to particle systems*, J. Math. Phys. **54** (2013), no. 2, 021503, 15. MR 3076363
- [GM14] ———, *Strong solutions of non-colliding particle systems*, Electron. J. Probab. **19** (2014), no. 119, 21. MR 3296535
- [GM19] ———, *On squared bessel particle systems*, Bernoulli **25** (2019), no. 2, 828–847.
- [GT74] WB Gragg and RA Tapia, *Optimal error bounds for the newton–kantorovich theorem*, SIAM Journal on Numerical Analysis **11** (1974), no. 1, 10–13.
- [GY+06] Léonard Gallardo, Marc Yor, et al., *A chaotic representation property of the multidimensional dunkl processes*, Annals of probability **34** (2006), no. 4, 1530–1549.
- [HDY+12] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al., *Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups*, IEEE Signal processing magazine **29** (2012), no. 6, 82–97.

- [HN19] Boris Hanin and Mihai Nica, *Products of many large random matrices and gradients in deep neural networks*, Communications in Mathematical Physics (2019), 1–36.
- [HP00] Fumio Hiai and Dénes Petz, *The semicircle law, free random variables and entropy*, Math. Surveys and Monographs, vol. 77, Amer. Math. Soc., Providence, 2000. MR 1746976 (2001j:46099)
- [HTF09] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The elements of statistical learning: data mining, inference, and prediction*, Springer Science & Business Media, 2009.
- [JK20] Benjamin Jourdain and Ezéchiél Kahn, *Strong solutions to a beta-Wishart particle system*, preprint 2003.08699 (2020).
- [Joh01] Iain M Johnstone, *On the distribution of the largest eigenvalue in principal components analysis*, Annals of statistics (2001), 295–327.
- [Joh08] ———, *Multivariate analysis and jacobi ensembles: Largest eigenvalue, tracy-widom limits and rates of convergence*, Annals of statistics **36** (2008), no. 6, 2638.
- [Jol86] Ian T Jolliffe, *Principal components in regression analysis*, Principal component analysis, Springer, 1986, pp. 129–155.
- [JPP⁺19] Arturo Jaramillo, Juan Carlos Pardo, José Luis Pérez, et al., *Convergence of the empirical spectral distribution of gaussian matrix-valued processes*, Electronic Journal of Probability **24** (2019).
- [Kan48] LV Kantorovich, *On newton’s method for functional equations*, Dokl. Akad. Nauk SSSR, vol. 59, 1948, pp. 1237–1240.
- [KBM79] JT Kent, John Bibby, and KV Mardia, *Multivariate analysis*, Academic Press Amsterdam, 1979.
- [KO01] Wolfgang König and Neil O’Connell, *Eigenvalues of the Laguerre process as non-colliding squared Bessel processes*, Electron. Comm. Probab. **6** (2001), 107–114. MR 1871699
- [KS91] Ioannis Karatzas and Steven E. Shreve, *Brownian motion and stochastic calculus*, second ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991. MR 1121940
- [KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, *Imagenet classification with deep convolutional neural networks*, Advances in neural information processing systems **25** (2012), 1097–1105.

- [Lal67] Mehta Madan Lal, *Random matrices and the statistical theory of energy levels / m. l. mehta*, Academic press, New York London, 1967 (eng).
- [LBBH98] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, *Gradient-based learning applied to document recognition*, Proceedings of the IEEE **86** (1998), no. 11, 2278–2324.
- [LBD⁺89] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel, *Backpropagation applied to handwritten zip code recognition*, Neural computation **1** (1989), no. 4, 541–551.
- [LBOM12] Yann A LeCun, Léon Bottou, Genevieve B Orr, and Klaus-Robert Müller, *Efficient backprop*, Neural networks: Tricks of the trade, Springer, 2012, pp. 9–48.
- [LC19] Guillaume Lample and François Charton, *Deep learning for symbolic mathematics*, arXiv preprint arXiv:1912.01412 (2019).
- [Lév17] Thierry Lévy, *The master field on the plane.*, Asterisque **388** (2017).
- [LL08] Damien Lambertson and Bernard Lapeyre, *Introduction to stochastic calculus applied to finance*, second ed., Chapman & Hall/CRC Financial Mathematics Series, Chapman & Hall/CRC, Boca Raton, FL, 2008. MR 2362458
- [LLC⁺18] Cosme Louart, Zhenyu Liao, Romain Couillet, et al., *A random matrix approach to neural networks*, The Annals of Applied Probability **28** (2018), no. 2, 1190–1248.
- [LLX14] Songzi Li, Xiang-Dong Li, and Yong-Xiao Xie, *On the law of large numbers for the empirical measure process of generalized dyson brownian motion*, arXiv preprint arXiv:1407.7234 (2014).
- [Lé10] Dominique Lépingle, *Boundary behavior of a constrained Brownian motion between reflecting-repellent walls*, Probab. Math. Statist. **30** (2010), no. 2, 273–287. MR 2792586
- [McK69] Henry P McKean, *Stochastic integrals*, vol. 353, American Mathematical Soc., 1969.
- [MMS76] Donald F Morrison, Lauriston C Marshall, and Harry L Sahlin, *Multivariate statistical methods*.
- [MNN⁺20] Mylène Maïda, Tien Dat Nguyen, Thanh Mai Pham Ngoc, Vincent Rivoirard, and Viet Chi Tran, *Statistical deconvolution of the free fokker-planck equation at fixed time*, arXiv preprint arXiv:2006.11899 (2020).
- [Mon73] Hugh L Montgomery, *The pair correlation of zeros of the zeta function*, Proc. Symp. Pure Math, vol. 24, 1973, pp. 181–193.

- [MP67] V.A. Marčenko and Leonid Pastur, *Distribution of eigenvalues for some sets of random matrices*, Math USSR Sb **1** (1967), 457–483.
- [MP19] Jacek Malecki and José Luis Pérez, *Universality classes for general random matrix flows*, 2019.
- [MS17] James A Mingo and Roland Speicher, *Free probability and random matrices*, vol. 35, Springer, 2017.
- [Mui09] Robb J Muirhead, *Aspects of multivariate statistical theory*, vol. 197, John Wiley & Sons, 2009.
- [Nec19] Ion Nechita, *Applications of random matrices in quantum information theory*.
- [NS06] Alexandru Nica and Roland Speicher, *Lectures on the combinatorics of free probability*, vol. 13, Cambridge University Press, 2006.
- [Pas20] Leonid Pastur, *On random matrices arising in deep neural networks. gaussian case*, arXiv preprint arXiv:2001.06188 (2020).
- [PGC⁺17] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer, *Automatic differentiation in pytorch*.
- [PLR⁺16] Ben Poole, Subhaneil Lahiri, Maithra Raghu, Jascha Sohl-Dickstein, and Surya Ganguli, *Exponential expressivity in deep neural networks through transient chaos*, arXiv preprint arXiv:1606.05340 (2016).
- [PPPA16] Juan Carlos Pardo, José-Luis Pérez, and Victor Pérez-Abreu, *A random matrix approximation for the non-commutative fractional brownian motion*, Journal of Theoretical Probability **29** (2016), no. 4, 1581–1598.
- [PS11] Leonid Pastur and Mariya Shcherbina, *Eigenvalue distribution of large random matrices*, Mathematical surveys and monographs, American Mathematical Society, Providence, RI, 2011.
- [PS20] L Pastur and V Slavin, *On random matrices arising in deep neural networks: General iid case*, arXiv preprint arXiv:2011.11439 (2020).
- [PSG18] Jeffrey Pennington, Samuel Schoenholz, and Surya Ganguli, *The emergence of spectral universality in deep networks*, International Conference on Artificial Intelligence and Statistics, PMLR, 2018, pp. 1924–1932.
- [PW17] Jeffrey Pennington and Pratik Worah, *Nonlinear random matrix theory for deep learning*.

- [RHW85] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams, *Learning internal representations by error propagation*, Tech. report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.
- [RS93] L. Chris G. Rogers and Zhan Shi, *Interacting Brownian particles and the Wigner law*, *Probab. Theory Related Fields* **95** (1993), no. 4, 555–570. MR 1217451 (94h:60146)
- [Rud18] Sebastian Ruder, *NLP’s ImageNet moment has arrived*, <https://ruder.io/nlp-imagenet/>, 2018.
- [RW00] L. Chris G. Rogers and David Williams, *Diffusions, Markov processes, and martingales. Vol. 2*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000, Itô calculus, Reprint of the second (1994) edition. MR 1780932
- [RY99] Daniel Revuz and Marc Yor, *Continuous martingales and Brownian motion*, third ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 293, Springer-Verlag, Berlin, 1999. MR 1725357
- [SAC⁺17] Kristof T Schütt, Farhad Arbabzadah, Stefan Chmiela, Klaus R Müller, and Alexandre Tkatchenko, *Quantum-chemical insights from deep tensor neural networks*, *Nature communications* **8** (2017), no. 1, 1–8.
- [SEJ⁺20] Andrew W. Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Židek, Alexander W. R. Nelson, Alex Bridgland, Hugo Penedones, Stig Petersen, Karen Simonyan, Steve Crossan, Pushmeet Kohli, David T. Jones, David Silver, Koray Kavukcuoglu, and Demis Hassabis, *Improved protein structure prediction using potentials from deep learning*, *Nature* **577** (2020), no. 7792, 706–710.
- [SGGSD16] Samuel S Schoenholz, Justin Gilmer, Surya Ganguli, and Jascha Sohl-Dickstein, *Deep information propagation*, arXiv preprint arXiv:1611.01232 (2016).
- [SM17] Roland Speicher and James Mingo, *Free probability and random matrices*, 01 2017.
- [SYY19] Jian Song, Jianfeng Yao, and Wangjun Yuan, *High-dimensional limits of eigenvalue distributions for general wishart process*, arXiv preprint arXiv:1901.02190 (2019).
- [Szn91] Alain-Sol Sznitman, *Topics in propagation of chaos*, Ecole d’été de probabilités de Saint-Flour XIX—1989, Springer, 1991, pp. 165–251.
- [Tao12] Terence Tao, *Topics in random matrix theory*, vol. 132, American Mathematical Soc., 2012.

- [Tar20] Pierre Tarrago, *Spectral deconvolution of unitarily invariant matrix models*, arXiv preprint arXiv:2006.09356 (2020).
- [Tre16] Dario Trevisan, *Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients*, Electron. J. Probab. **21** (2016), Paper No. 22, 41. MR 3485364
- [TT20a] Hoang Dung Trinh and Khanh Duy Trinh, *Beta jacobi ensembles and associated jacobi polynomials*, 2020.
- [TT20b] ———, *Beta laguerre processes in a high temperature regime*, 2020.
- [TVV04] Antonia M Tulino, Sergio Verdú, and Sergio Verdu, *Random matrix theory and wireless communications*, Now Publishers Inc, 2004.
- [VDN92] Dan V. Voiculescu, Kenneth J. Dykema, and Alexandru Nica, *Free random variables*, CRM Monograph Series, American Mathematical Society, 1992.
- [Voi85] Dan Voiculescu, *Symmetries of some reduced free product c^* -algebras*, Operator algebras and their connections with topology and ergodic theory, Springer, 1985, pp. 556–588.
- [Voi86] Dan Voiculescu, *Addition of certain non-commuting random variables*, Journal of Functional Analysis **66** (1986), no. 3, 323–346.
- [Voi87] Dan Voiculescu, *Multiplication of certain non-commuting random variables*, Journal of Operator Theory (1987), 223–235.
- [Voi91] ———, *Limit laws for random matrices and free products*, Inventiones mathematicae **104** (1991), no. 1, 201–220.
- [Voi98] ———, *A strengthened asymptotic freeness result for random matrices with applications to free entropy*, International Mathematics Research Notices **1998** (1998), no. 1, 41–63.
- [WCA14] Jason Weston, Sumit Chopra, and Keith Adams, *# tagspace: Semantic embeddings from hashtags*, Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1822–1827.
- [Wig51] Eugene P. Wigner, *On the statistical distribution of the widths and spacings of nuclear resonance levels*, Mathematical Proceedings of the Cambridge Philosophical Society **47** (1951), no. 4, 790–798.
- [Wig55] Eugene P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Annals of Mathematics **62** (1955), no. 3, 548–564.
- [Wig58] Eugene P Wigner, *On the distribution of the roots of certain symmetric matrices*, Annals of Mathematics (1958), 325–327.

- [Wis28] John Wishart, *The generalised product moment distribution in samples from a normal multivariate population*, *Biometrika* (1928), 32–52.
- [Yan19] Greg Yang, *Tensor programs i: Wide feedforward or recurrent neural networks of any architecture are gaussian processes*, arXiv preprint arXiv:1910.12478 (2019).
- [Yan20a] ———, *Tensor programs ii: Neural tangent kernel for any architecture*, arXiv preprint arXiv:2006.14548 (2020).
- [Yan20b] ———, *Tensor programs iii: Neural matrix laws*, arXiv preprint arXiv:2009.10685 (2020).
- [YCG08] Marc Yor, Oleksandr. Chybiryakov, and Léonard Gallardo, *Dunkl processes and their radial parts relative to a root system*, *Harmonic and stochastic analysis of Dunkl processes* (P. Graczyk, M. Rösler, and M. Yor, eds.), *Travaux en cours*, vol. 71, Hermann, 2008, pp. 113–197.
- [ZKS⁺19] Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J. Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdal, Adriana Romero, Michael Rabbat, Pascal Vincent, Nafissa Yakubova, James Pinkerton, Duo Wang, Erich Owens, C. Lawrence Zitnick, Michael P. Recht, Daniel K. Sodickson, and Yvonne W. Lui, *fastmri: An open dataset and benchmarks for accelerated mri*, 2019.