Optimal ecological transition path of a credit portfolio distribution, based on Multidate Monge-Kantorovich formulation

Emmanuel Gobet, Clara Lage

Optimal ecological transition path of a credit portfolio distribution, based on Multidate Monge-Kantorovich formulation*

E. Gobet† and C. Lage‡

November 9, 2021

Abstract: Accounting for climate transition risks is one of the most important challenges in the transition to a low-carbon economy. Banks are encouraged to align their investment portfolios to CO2 trajectories fixed by international agreements, showing the necessity of a quantitative methodology to implement it. We propose a mathematical formulation for this problem and a multistage optimization criterion for a transition between the current bank portfolio and a target one. The optimization Problem combines the Monge-Kantorovich formulation of optimal transport, for which the cost is defined according to the financial context, and a credit risk measure. We show that the problem is well-posed, and can be embedded into a saddle-point problem for which Primal-Dual algorithms can be used. We design a numerical scheme that is able to solve the problem in available time, with nice scalability properties according to the number of decision times; its numerical convergence is analysed. Last we test the model using real financial data, illustrating that the optimal portfolio alignment may differ from the naive interpolation between the initial portfolio and the target.

Keywords: Optimal transport, Environmental economics, Credit risk, Primal-Dual algorithm

MSC2020: 49Q22, 91B76, 91G40, 49M29

1 Introduction

Climate change context. Since the advent of the industrial age, anthropogenic greenhouse gas (GHG) emissions have continued to increase, causing ongoing climate change. As the greenhouse effect intensifies, climate change is increasingly materializing in the form of physical hazards that threaten the stability and sustainability of today's societies as well as human lives. Faced with

*This research is supported by the *Chair Stress Test, RISK Management and Financial Steering of the Foundation Ecole Polytechnique*. The authors thank the STFS and CSR teams from BNP Paribas for their feedbacks regarding an intermediate version of this work.

†CMAP, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, Route de Saclay, 91128 Palaiseau Cedex, France. email:emmanuel.gobet@polytechnique.edu

‡CMAP, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, Route de Saclay, 91128 Palaiseau Cedex, France. email:clara.lage@polytechnique.edu

the increasing frequency and severity of physical damages, the pressure on legislators is increasing, encouraging them to take measures to respect the environment and, in particular, to significantly reduce greenhouse gas emissions from human activities. These new climate risks, particularly physical and transition risks, have been described in the banking domain by [9] in a resounding speech. To make echo his call, the banking regulators have tackled the issue, in particular by gathering under the egide of the Network for the Greening of the Financial System (NGFS) [28]. The Basel Committee on Banking Supervision published in April 2021 a report [4] exploring how climate-related risk drivers, including physical risks and transition risks, can arise and affect both banks and the banking system via micro- and macroeconomic transmission channels.
In parallel with these descriptive works on climate risks and their transmission channels, various decisions to green the economy and finance have been taken. Five international banks commit themselves in the COP24 agreement (Katovice 2018) to align their portfolio with the CO2 trajectories of the Paris agreements [11, p.22]. The European Banking Authority (EBA) published in December 2019 its Action Plan for Sustainable Finance laying the groundwork for, on the one hand, better controlling the risks related to the climate change; on the other hand, shifting progressively capital towards responsible investments so as to achieve the objectives of sustainable growth [12]. In June 2021, the World Bank Group announced its new Climate Change Action Plan that aims to deliver record levels of climate finance to developing countries, reduce emissions, strengthen adaptation, and align financial flows with the goals of the Paris Agreement [43].
Our work is in this vein and aims to provide methodological and quantitative tools to progressively align an investment portfolio with a green and sustainable objective over a certain long-term horizon. The purpose is to find, given a current portfolio (typically that of a corporate and investment bank), an investment path that achieves a green score target at the time horizon of 2050 say, while controlling credit risk.

Our contributions. Although the goal of greening investment portfolios is becoming a concern shared by most (if not all) financial institutions, to the best of our knowledge, there is no quantitative methodology to implement it. Our main contributions are, in a few words, to provide a suitable framework for risk modeling, to determine the optimal investment path as the solution of an optimization problem in the space of probability distributions and to provide a numerical algorithm for computing effectively its solution. Numerical experiments confirm the convergence of the algorithm and illustrate the outputs of such optimisation program on realistic data.

To better grasp our original contributions and make some connections with related works, let us quickly summarize how we model the problem, the details being given in Section 2. Changing an investment portfolio of a bank cannot be done abruptly, neither without cost nor without risk. Given an invested capital, an investment portfolio is represented by its distribution of obligors, represented by their creditworthiness (propensity to reimburse their debts) and an ecological rating (linked to their Environmental Social and Governance (ESG for short) scores, their carbon emissions, etc). The objective is to transport the initial investment distribution of the bank to a target (greener) distribution: the costs are related to the modification of the portfolio over time (we will take discrete dates) and to the credit risk (capital cost given by a risk metric, like the Value at Risk). There is a non trivial tradeoff between the objective and the costs: for instance, some "green" obligors may

correspond to innovative business, and as such, they may be riskier than well-established "brown" obligors; therefore, going fully green may be far from financially optimal regarding the credit risk incurred by the bank.
In Section 2, we model this problem as a new multidate optimal transport problem of Monge-Kantorovich type (see [42] for an overview of optimal transport), the solution of which being a path of investment distributions. We show that finding the solution is equivalent to that of a point-saddle problem (Theorem 2.5) for which we use a Primal-Dual algorithm (see Algorithms 2 and 5). The output of this work is two fold: it allows risk managers of banks to get projections of the alignment cost; while getting investment path for greening their portfolio, it supports them in a better forward-looking financial steering to achieve a carbon-neutral economy.

Background results.
On the financial modelling side. In the last years the terms as: green portfolio, sustainable finance and price of climate change have gained some relevance in the financial debate and science fields. In [6, 13], the authors analyse equity portfolios within the prism of an asset manager (using mean-variance optimization criterion) under climate change schemes. Instead, we focus on the credit portfolio of a bank and account for credit risk.

Regarding the impact of climate change on banks, we refer to the work [5] on climate stress-test, providing a description of transmission channels of the climate risks to a bank, using a network approach to financial dependencies. The authors discuss the impact of climate policy timing on the stability and resilience of the financial system. The topic is different from the current work.

Defining alignment with climate impact target (like those of the Paris agreement) is a topic with several possible answers (depending on whether the criteria taken into account give more or less importance to carbon emissions, environmental impact, temperature rise, ...). See the detailed discussion in [33] and references therein. Our work will not go into these aspects and we consider directly that each obligor has a certain ecological score (whether it is linked to its CO2 emissions, its impact on the environment, its contribution to temperature rise ....) which is an input of our decision modelling. The alignment target is considered as given, as one of the few existing solutions mentioned in [33], and in practice it should account also for the corporate strategy of the bank; rather than defining this target, our approach is rather to determine how to reach it dynamically accounting for credit risk.

On the mathematical side. The multidate model proposed in the present work can be compared with the two stage and multi-stage risk problems that already have large applications and theoretical aspects present in the literature (see [19, 38, 31]). Usually, some liquidity constraints are introduced to be realistic with the portfolio management application, which translated into linear inequalities on the state space. Some transaction costs can be included too. A significant difference with our framework is that these costs and constraints are acting on the state space and not at the level of portfolio distributions as we do. In our case, the decision variable at each decision time is a distribution of obligors and the cost of changing the decision (for progressively targeting the green objective) shall be written as a transport cost between two probability measures μ, ν: naturally, we choose the Monge-Kantorovich optimal transport cost $\mathbf{MK}(\mu, \nu)$ (see [42]). Besides, since the

cost c defining $\mathbf{MK}(.,.)$ has no reason to be related to a distance, $\mathbf{MK}(.,.)$ may be different from a Wasserstein distance from which many numerical tools are available, see for instance [32]. To the best of our knowledge, multidate Monge-Kantorovich optimal transport problems (involving a sequence of pairs of measures) like those we introduce in Section 2 have not been considered so far in the literature. As discussed in Subsection 2.4, this is different from a traditional geodesic problem. Last, for the sake of completeness, we refer to [14, Chapter 9] for various applications of optimal transport methods in economics.

Plan of the paper. This paper contributes to existing literature in both methodological and applied aspects. Section 2 is devoted to modeling the problem of optimal credit portfolio alignment accounting for ecological scores and credit risk, while achieving a given green investment portfolio. Section 3 gathers numerical algorithms to compute the solution, and numerical tests to assess the convergence algorithm. Experiments on realistic financial data are performed in Section 4. Some technical results are postponed to Appendix A.

2 Modeling the portfolio alignment and credit risk

2.1 Path of obligors portfolio

We are to consider the time evolution of a credit portfolio of a bank which grants loans to institutions, companies, individuals, commonly called obligors. The date $t = 0$ refers to today and the final date $T \in \mathbb{N}^*$ corresponds for instance to 2050 (a given horizon to achieve a prescribed portfolio alignment). For a bank, modifying the credit portfolio is in practice a continuous-time process but in our study, we take the simplified assumption that reinvestment can be adjusted only at $t \in \{0, 1, ..., T\}$ and the distance between two dates is constant (e.g. $H = 5$ years for instance). Furthermore, we consider the problem of portfolio alignment under the angle of investment reallocation, i.e. ignoring the initiation of new investments that could occur owing to new available capital: in other words, for the current analysis, the total invested capital is K_0 and constant from $t = 0, \dots, T$. In doing so, we ignore the possible expansion of the bank and the interest rates (with rates close to 0, the assumption is realistic at present).

At time t, the adjusted portfolio is described by a probability distribution $\mathbf{P}_t(\mathrm{d}\mathfrak{o})$ of obligor attributes $\mathfrak{o}$. The obligor attributes are made of two components $\mathfrak{o} = (\mathfrak{e}, \mathfrak{g})$ representing green scores on the one hand, and economic activities on the other hand. The types of $\mathfrak{e}$ and $\mathfrak{g}$ can be arbitrary (number, vectors, categories,...), they are just split into two components to facilitate the interpretation.

1. The state space of the green score $\mathfrak{e}$ is denoted by $\mathcal{E}$: $\mathfrak{e}$ could be a CO2 emission rate (typically kg of CO2 by € invested and the amount of CO2 is evaluated over a period of H years). It could be described by ESG scores like in the experiments of Section 4.2. In any case, it reflects the greenyness of the obligors.

2. The state space of the group $\mathfrak{g}$ is denoted by $\mathcal{G}$. Different groups represent the variety of obligors in the portfolio according to their sector of activity (Consumer Services, Energy Min-

erals, Productor Manufacturing, Utilities, etc) and their geographical area (France, Europe, Asia, North America, etc.): for the nomenclature of these groups, see for instance SP500 sectors on https://www.spglobal.com/spdji/en/landing/investment-themes/sectors/, the "North American Industry Classification System" (NAICS) or the "Nomenclature statistique des Activités économiques dans la Communauté Européenne" (NACE) according to geographic zones.

There are many other ways to classify obligors according to their groups. What is important is that this classification is meaningful for determining the credit risk of obligors. By highlighting the sectors of activity and location, it makes it possible to set up credit risk models with sectorial factors, as is often done among practioners. For such a multi-factor copula model in credit risk, see [36, p.222].

This modeling point of view is placed at a mesoscopic scale, which is a compromise between the fineness of the portfolio description, the tractability of the mathematical/numerical analysis of the optimal transport problem, the interpretability of the results. If we adopted a microscopic description of the portfolio, it would require collecting the carbon scores of each obligor, its credit risk dependencies with other obligors and economic factors: this raises data collection difficulties, and then numerical complexity problems (related to the number of points in the optimal transport problem). Our further experiments show that the current approach remains tractable in practice.

In addition, we denote by $\mathbf{P}_{\text{now}}$ the initial credit portfolio of the bank, before changing to $\mathbf{P}_0$, and by $\mathbf{P}_{\text{target}}$ the target portfolio (desired distribution) to be reached with $\mathbf{P}_T$ (as close as possible). Changing the portfolio distribution represents a cost, encoded by a function $c_t : (\mathfrak{o}, \mathfrak{o}') \in \mathcal{O} \times \mathcal{O} \mapsto c_t(\mathfrak{o}, \mathfrak{o}') \in [0, +\infty)$ describing both the cost of terminating the credit contract (decontracting) with the obligor $\mathfrak{o}$ and the cost of recontracting with a new obligor $\mathfrak{o}'$. To account for normalization, we assume c is set for a nominal of 1€ . In addition, the cost function may depend on t: brown obligors could be harder and harder to decontract (effect of stranded assets), International Climate Policy and Intended Nationally Determined Contributions (INDCs) could increase the incentives to financially support greener obligors. Besides, the dependence in time of c could be inspired by a chosen Shared Socio-economic Pathway (SSP), typically the scenario of an orderly transition to a low-carbon economy.

Changing the portfolio distribution from $\mathbf{P}_{t-1}$ to $\mathbf{P}_t$ will be done according to Monge-Kantorovich (MK in short) paradigm (see Section A.2 for standard properties), i.e. by minimizing the average cost $\mathbf{MK}_t(\mathbf{P}_{t-1}, \mathbf{P}_t)$ where

$$\mathbf{MK}_t(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{O}\times\mathcal{O}} c_t(\mathfrak{o}, \mathfrak{o}')\pi(\mathrm{d}\mathfrak{o}, \mathrm{d}\mathfrak{o}') \tag{2.1}$$

for arbitrary probability measures μ and ν. The financial cost for a total capital K_0 is

$$K_0 \ \mathbf{MK}_t(\mathbf{P}_{t-1}, \mathbf{P}_t) \in .$$

2.2 Modelling the credit risk

There are many possible modelling approaches to credit risk, see [24, 36] for some references. For the current work, it could be any of them, but for the sake of presentation and further experiments,

we shall focus on one of them, namely the Gaussian copula model: this model is supported by its simplicity of use, as reported recently in [29] by the Bank of International Settlements about Risk-Weighted Assets. More precisely, we model the default probability over a period of H years using a firm model: the credit quality of the obligor depends on the distance between the obligor asset value and a default threshold, the time-evolution depends on d Gaussian factors $F = (F^{(1)}, \dots, F^{(d)})$. The ratings between obligors may differ because their respective thresholds and asset values differ; the ratings change because of the evolution of activity/geographic factors F to which the asset value is correlated. All in all, the individual default probability of an obligor [17, Section 2][36, p.222] writes typically as

$$\mathbb{P}\left(X \geq \gamma \mid F\right), \qquad X = \sum_{j=1}^{d} \beta_j F^{(j)} + \beta_0 \varepsilon, \quad |\beta_{1:d}| := \sqrt{\sum_{j=1}^{d} \beta_j^2} < 1, \quad \beta_0 := \sqrt{1 - |\beta_{1:d}|^2},$$

with parameters $\theta = (\beta_0, \dots, \beta_d, \gamma)$ that depend on the obligor.

The components are independent, ε is an idiosyncratic risk factor independent from F; the components and ε are distributed as standard normal random variable. We have

$$\mathbb{P}\left(X \geq \gamma \mid F\right) = \Phi\left(\left(\sum_{j=1}^{d} \beta_j F^{(j)} - \gamma\right) / \beta_0\right), \tag{2.2}$$

where Φ is the c.d.f. of the standard normal random variable. We will simply write the above $\varphi(F, \theta)$ for a certain function $\varphi(.)$ depending on the type of credit modelling (here Gaussian copula) and on some parameters θ entering in the model (here $(\beta_j)_j, \gamma$). All obligors with the attribute $\mathfrak{o}$ have a default probability of the form

$$\varphi(Z, \theta)$$

for some parameter θ depending stochastically on the attribute $\mathfrak{o}$. Conditionally to $\mathfrak{o}$, the distribution of θ at time t is denoted by

$$q_{t,\mathfrak{o}}(\mathrm{d}\theta).$$

See the discussion in Subsection 4.3.3 about how to calibrate such distribution. Once again, the dependence in time reflects the possible change of the risk structure of the economy in the future, inspired by SSP.

Finally, the credit risk loss per € at point $\mathfrak{o}$ (at time t) is given by

$$\int_{\Theta} \varphi(F_t, \theta) q_{t,\mathfrak{o}}(\mathrm{d}\theta)$$

For a given portfolio allocation $\mathbf{P}$, the aggregated credit risk loss (at time t) per € invested is then

$$L_t(\mathbf{P}) = \int_{\Theta \times \mathcal{O}} \varphi(F_t, \theta) q_{t,\mathfrak{o}}(\mathrm{d}\theta) \mathbf{P}(\mathrm{d}\mathfrak{o}). \tag{2.3}$$

As usually done [16], we assume that the time-evolution of F follows a mean-reverting process representing the fluctuating cycles in the economy, typically a AR(1) process: this specification

of model for F is in no way a restriction, but simply a possible choice for the experiments that follow. In the current framework of portfolio alignment (typically in a given transition scenario to a low-carbon economy), it would make sense to use the distribution of F in that scenario: however, to the best of our knowledge, there is no consensus on how to calibrate such risk factors F in a given future scenario, we leave this modelling question for further research. Alternatively, one could calibrate the random evolution of F according to its usual (historically observed) distribution, and apply expert-judgement shocks to account for the transition scenario.

Given a potential aggregated credit risk loss $L_t(\mathbf{P})$, we assume that the cost of capital is computed through a risk metrics $\varrho(.)$, which should satisfy in general the axioms of coherent risk measures [3]. For our work, we specifically require some convexity, continuity and positive homogeneity properties for ϱ (detailed later in Properties 2.1).

All in all, the cost of capital due to credit risk (renormalized by the invested capital) is

$$\varrho\left(K_0\ L_t(\mathbf{P})\right) \text{€}\ .$$

2.3 Optimization problem and properties of solutions

If $\mathbf{P}_{\text{now}}$ is the current portfolio distribution and $\mathbf{P}_{\text{target}}$ is the desired distribution at horizon T, we seek the distribution $\mathbf{P}_t$ at the intermediate dates $t \in \{0, 1, ..., T\}$ that minimizes the cost of the changing trajectory in the portfolio and the cost of credit risk. The optimization criterion

$$\min_{\mathbf{P}_0,\dots,\mathbf{P}_T} \left[K_0 \sum_{t=0}^{T+1} \lambda_t \mathbf{MK}_t(\mathbf{P}_{t-1}, \mathbf{P}_t) + \sum_{t=0}^{T} \varrho\left(K_0 L_{t+1}(\mathbf{P}_t)\right) \right] \tag{2.4}$$

with the convention

$$\mathbf{P}_{-1} = \mathbf{P}_{\text{now}}, \qquad \mathbf{P}_{T+1} = \mathbf{P}_{\text{target}}.$$

The positive coefficients $(\lambda_t)_t$ are weighting more or less the transport cost along the timeline and according the credit risk. Since the risk metric is taken as positive homogenous (Properties 2.1), $\varrho\left(K_0 L_{t+1}(\mathbf{P}_t)\right) = K_0\ \varrho\left(L_{t+1}(\mathbf{P}_t)\right)$, therefore the solution is independent on K_0; in the following, we set

$$K_0 = 1.$$

Generally speaking, we refer to (2.4) as *Multidate Monge Kantorovich (MMK)* problem. Changing the portfolio will help to achieve an ideally aligned portfolio $\mathbf{P}_{\text{target}}$ in terms of green sector and economic group to support, but it will change the risk pattern of investment because of the risk metrics ϱ. There is no reason for which changing abruptly be optimal, and we seek for optimal portfolio distribution path, accounting for reallocation cost and credit risk.

2.4 Similarities and differences with the geodesic problem in Wasserstein spaces

We observe that the MMK problem (2.4), when excluding the risk part, have some similarities with the problem of finding a geodesic in metric spaces. To alleviate the next statements and discussions, we assume in this section that the underlying state space (previously denoted $\mathcal{O}$) is denoted X, it

is an Euclidean space equipped with a distance d. Particularly, when the MK cost $c := c_p = d^p$ is the p-th power of the distance d for $p \geq 1$, and if we set $\mathbf{W_p}(\mu, \nu) = (\mathbf{MK}^{(c_p)}(\mu, \nu))^{1/p}$, the latter defines a distance on the set of probability measures with *finite p-moment* denoted by

$$\mathcal{P}_p(X) := \{\mu \in \mathcal{P}(X) : \int_X d^p(x, x_0)\mu(\mathrm{d}x) < +\infty\}$$

(see [42, Definition 6.1][1]). $\mathbf{W_p}$ (resp. $\mathcal{P}_p(X)$ is the so-called p-Wasserstein distance (resp. space), see [42, Definition 6.4]. We now recall the definition of a geodesic in the metric space $(\mathcal{P}_p(X), \mathbf{W_p})$.

Definition 2.1. *A geodesic is a curve $\gamma : I = [0, 1] \to \mathcal{P}_p(X)$ that minimizes the length $\mathcal{L}(\gamma)$, where:*

$$\mathcal{L}(\gamma) = \sup_{N \in \mathbb{N}} \sup_{0=s_0 \leq s_1 \ldots \leq s_N = 1} \sum_{i=0}^{N-1} \mathbf{W_p}(\gamma(s_i), \gamma(s_{i+1})).$$

Not all metric spaces admit geodesics. But in our setting, the p-Wasserstein space $(\mathcal{P}_p(X), \mathbf{W_p})$ does, the minimal length curve exists.

Theorem 2.1. *The p-Wassertein space $\big(\mathcal{P}_p(X), \mathbf{W_p}\big)$ is a geodesic space, in the sense that*

$$\mathbf{W_p}(\mu, \nu) = \min_{\gamma \in C([0,1], \mathcal{P}_p(X))} \{\mathcal{L}(\gamma) : \gamma(0) = \mu, \gamma(1) = \nu\},$$

where $C([0, 1], \mathcal{P}_p(X))$ stands for the set of continuous functions γ from $[0, 1]$ to $\mathcal{P}_p(X)$.

Proof. See [37, Theorem 5.27]. □

We are now in a position to connect a simplified MMK problem to a geodesic problem in $\mathcal{P}_p(X)$.

Theorem 2.2. *The optimal solutions $(\mathbf{P}_0, ..., \mathbf{P}_T)$ of the problem*

$$\min_{\mathbf{P}_0, \ldots, \mathbf{P}_T} \left[\sum_{t=0}^{T+1} \lambda_t \mathbf{W_p}(\mathbf{P}_{t-1}, \mathbf{P}_t)\right] \quad \textit{such that} \quad \mathbf{P}_{-1} = \mathbf{P}_{\text{now}},\ \mathbf{P}_{T+1} = \mathbf{P}_{\text{target}} \textit{ are given}, \tag{2.5}$$

lie in a geodesic curve $\gamma : [0, 1] \to \mathcal{P}_p(X)$ such that $\gamma(0) = \mathbf{P}_{\text{now}}$ and $\gamma(1) = \mathbf{P}_{\text{target}}$.

Proof. ▷ First, the triangular inequality for the distance function $\mathbf{W_p}$ implies that for any $\mathbf{P}_0, \ldots, \mathbf{P}_T$, we have

$$\mathbf{W_p}(\mathbf{P}_{\text{now}}, \mathbf{P}_{\text{target}}) \leq \sum_{t=0}^{T+1} \mathbf{W_p}(\mathbf{P}_{t-1}, \mathbf{P}_t). \tag{2.6}$$

Let $s \in \{0, 1, ..., T\}$. Note that, if inequality (2.6) is an equality, $\mathbf{P}_s$ lies in a geodesic curve between $\mathbf{P}_{\text{now}}$ and $\mathbf{P}_{\text{target}}$. To see that, consider γ_1 a geodesic between $\mathbf{P}_{\text{now}}$ and $\mathbf{P}_s$ and γ_2 a geodesic between $\mathbf{P}_s$ and $\mathbf{P}_{\text{target}}$, the concatenation $\gamma = \gamma_1 \cdot \gamma_2$ is also a continuous curve and $\mathcal{L}(\gamma) = \mathcal{L}(\gamma_1) + \mathcal{L}(\gamma_2)$ (see Proposition 2.3.4 in [8]). If (2.6) is an equality it implies in particular that:

[1] in this definition the reference point x_0 does not play any role

$\mathbf{W_p}(\mathbf{P}_{\text{now}}, \mathbf{P}_s) + \mathbf{W_p}(\mathbf{P}_s, \mathbf{P}_{\text{target}}) = \mathbf{W_p}(\mathbf{P}_{\text{now}}, \mathbf{P}_{\text{target}})$. In this case, $\mathcal{L}(\gamma) = \mathbf{W_p}(\mathbf{P}_{\text{now}}, \mathbf{P}_{\text{target}})$, that is, γ is a geodesic joining $\mathbf{P}_{\text{now}}$ and $\mathbf{P}_{\text{target}}$, and $\mathbf{P}_s$ lies in the image of γ.

$\triangleright$ Let $(\mathbf{P}_0, ..., \mathbf{P}_T)$ be a solution to (2.5). Consider the set of weights $\{\lambda_t : t \in \{0, ..., T\}\}$, $\tilde{\lambda} = \min_t\{\lambda_t\}$, and the index $\tilde{t}$ such that $\lambda_{\tilde{t}} = \tilde{\lambda}$. Consider a geodesic $\gamma^\star$ between $\mathbf{P}_{\text{now}}$ and $\mathbf{P}_{\text{target}}$, and choose a partition of the interval $[0, 1]$ such that $s_0 = s_1 = ... = s_{\tilde{t}} = 0$ and $s_{\tilde{t}+1} = ... = s_{T+1} = 1$. Then:

$$\begin{aligned}\sum_{i=0}^{T+1} \lambda_i \mathbf{W_p}(\gamma^\star(s_i), \gamma^\star(s_{i+1})) &= \tilde{\lambda}\mathbf{W_p}(\mathbf{P}_{\text{now}}, \mathbf{P}_{\text{target}}) \\ &\leq \sum_{t=0}^{T+1} \tilde{\lambda}\mathbf{W_p}(\mathbf{P}_{t-1}, \mathbf{P}_t) \leq \sum_{t=0}^{T+1} \lambda_t \mathbf{W_p}(\mathbf{P}_{t-1}, \mathbf{P}_t).\end{aligned} \tag{2.7}$$

Since $(\mathbf{P}_0, ..., \mathbf{P}_T)$ is a solution, all inequalities in (2.7) need to be equalities, which implies in particular that (2.6) is an equality. This means that $\mathbf{P}_t$ lies in a geodesic between $\mathbf{P}_{\text{now}}$ and $\mathbf{P}_{\text{target}}$, as it is announced in the Theorem statement.

Moreover, note that the last inequality in (2.7) is strict when $\lambda_t > \tilde{\lambda}$ and $\mathbf{W_p}(\mathbf{P}_{t-1}, \mathbf{P}_t) \neq 0$. In other words, for any $t \notin \mathcal{T} := \{s : \lambda_s = \tilde{\lambda}\}$, an optimal solution must be such that $\mathbf{P}_{t-1} = \mathbf{P}_t$, which reduces the problem (2.5) to seek the solutions over $\#\mathcal{T}$ dates only and with a constant λ. $\square$

When $p = 2$ and $T = 0$, problem (2.4) (without the credit risk term) can also be compared with the barycenter problem between measures $\mathbf{P}_{\text{now}}$ and $\mathbf{P}_{\text{target}}$ with weights λ_0, and λ_1, that is:

$$\min_{\mathbf{P}_0} \left[\lambda_0 \mathbf{W_2}(\mathbf{P}_{\text{now}}, \mathbf{P}_0) + \lambda_1 \mathbf{W_2}(\mathbf{P}_0, \mathbf{P}_{\text{target}})\right]. \tag{2.8}$$

When we have just one decision $\mathbf{P}_0$, and the cost $c = c_2$ is the squared distance, the MMK problem (2.4) reads as the Wasserstein barycenter between measures $\mathbf{P}_{\text{now}}$ and $\mathbf{P}_{\text{target}}$ (2.8). The barycenter between two measures is a well-known problem with explicit solution that relies in the geodesic curve of the Wasserstein space $\gamma : [0, 1] \to \mathcal{P}_2(X)$ connecting $\mathbf{P}_{\text{now}}$ and $\mathbf{P}_{\text{target}}$ (See [1, Example 6.2]). The geodesic may not be unique and depends itself on the optimal transport plan $\bar{\pi}$, solution of the Kantorovich problem (2.1) transporting the measure $\mathbf{P}_{\text{now}}$ to $\mathbf{P}_{\text{target}}$. The geodesic is unique when there is a unique optimal transport plan $\bar{\pi}$, which is always the case when considering $\mathbf{P}_{\text{now}}$ and $\mathbf{P}_{\text{target}}$ continuous distributions; in such an uniqueness case, parameters λ_0 and λ_1 completely determine the unique barycenter.

The previous discussion about the similarities of MMK problem (2.4) with the geodesic problem gives a possible intuition about the optimal MMK trajectory $(\mathbf{P}_0, ..., \mathbf{P}_T)$: it can be viewed as an equilibrium between a geodesic curve (in which lie solutions of problem (2.5)), and the risk along the path $(\sum_{t=0}^{T} \rho(L_{t+1}(\mathbf{P}_t)))$. Nevertheless, although tasting similarly, there are some differences between the MMK problem and the geodesic problem. Indeed, it is also important to highlight the ambient assumptions of Theorem 2.2, in particular the cost $c_p = d^p$. Our problem **MK** considers more general possibilities of cost, that are well-adapted to the financial context. Besides, the credit risk may alter significantly the optimisation problem.

2.4.1 The discrete case

To elaborate on the solution to (2.4) and its properties, we assume that the universe of attributes $\mathcal{O}$ is finite:

$$\mathcal{O} = \{\mathfrak{o}_1, \dots, \mathfrak{o}_N\}, \qquad p_t^{(i)} := \mathbf{P}_t(\mathfrak{o}_i)$$

and write $\mathbf{p}_t := (p_t^{(i)} : i \in \{1, \dots, N\})$ which takes value in the probability simplex:

$$\Sigma_N := \left\{ p \in \mathbb{R}^N : p^{(i)} \geq 0, \sum_{i=1}^N p^{(i)} = 1 \right\}.$$

We denote:

$$\begin{cases} c_t^{i,j} := c_t(\mathfrak{o}_i, \mathfrak{o}_j), \\ \mathbf{p}_{\text{now}} := (\mathbf{P}_{\text{now}}(O = \mathfrak{o}_i) : i \in \{1, \dots, N\}), \\ \mathbf{p}_{\text{targ}} := (\mathbf{P}_{\text{target}}(O = \mathfrak{o}_i) : i \in \{1, \dots, N\}). \end{cases}$$

We are still looking for solutions to (2.4) but with discrete-time notations:

$$L_{t+1}(\mathbf{P}_t) \underset{(2.3)}{=} \sum_{i=1}^N p_t^{(i)} \int_\Theta \varphi(F_{t+1}, \theta) q_{t+1,\mathfrak{o}_i}(\mathrm{d}\theta) =: L_{t+1}(\mathbf{p}_t). \tag{2.9}$$

The risk measure term $\varrho\left(L_{t+1}(\mathbf{P}_t)\right)$ is now a function of the vector $\mathbf{p}_t$. The function $\mathbf{p}_t \in \Sigma_N \mapsto \varrho(L_{t+1}(\mathbf{p}_t)) \in \mathbb{R}$ should fulfill some properties to ensure the existence of solution of (2.4) in the discrete case:

Properties 2.1 (Risk Measure Properties)**.**

a) Positive Homogeneity: For any $\delta \geq 0$ and any vector $\mathbf{p}_t \in \Sigma_N$, then

$$\varrho(L_{t+1}(\delta\ \mathbf{p}_t)) = \delta\ \varrho(L_{t+1}(\mathbf{p}_t)).$$

b) Continuity: $\mathbf{p}_t \in \Sigma_N \mapsto \varrho(L_{t+1}(\mathbf{p}_t)) \in \mathbb{R}$ is continuous.

c) Convexity with respect to $\mathbf{p}_t$: For any $\delta \in (0,1)$ and any vectors $\mathbf{p}_t^1, \mathbf{p}_t^2$ in Σ_N, then

$$\varrho(L_{t+1}(\delta \mathbf{p}_t^1 + (1-\delta)\mathbf{p}_t^2)) \leq \delta\ \varrho(L_{t+1}(\mathbf{p}_t^1)) + (1-\delta)\ \varrho(L_{t+1}(\mathbf{p}_t^2)).$$

These properties are automatically satisfied by the α-Expected Shortfall (`ES`) defined by [35]

$$\texttt{ES}_\alpha(L) = \inf_{\ell \in \mathbb{R}} \left(\frac{\mathbb{E}\left((L-\ell)_+\right)}{1-\alpha} + \ell \right)$$

for some $\alpha \in (0,1)$. They also hold when ρ is the α-`VaR` in a 1-factor Gaussian copula model with non-negative correlations, since in that case, the credit risk term is linear w.r.t. $\mathbf{p}_t$ (see Lemma A.1).

The $\mathbf{MK}_t$ problem is then a linear optimization problem with primal and dual forms:

$$\mathbf{MK}_t(\mathbf{p}_{t-1}, \mathbf{p}_t) = \begin{cases} \min\limits_{\pi \geq 0} & \sum\limits_{i,j} c_t^{i,j} \pi_{i,j} \\ \text{s.t} & \sum_j \pi_{i,j} = p_{t-1}^{(i)},\ \sum_i \pi_{i,j} = p_t^{(j)}, \end{cases} \tag{2.10}$$

$$= \begin{cases} \max\limits_{f,g} & \sum\limits_i f_i p_{t-1}^{(i)} + \sum\limits_j g_j p_t^{(j)} \\ \text{s.t} & f_i + g_j \leq c_t^{i,j}, \end{cases} \tag{2.11}$$

using the primal and dual formulations of MK problems in the discrete case (Section A.2). The problem (2.4) (with $K_0 = 1$) writes now

$$\min_{\mathbf{p}_0 \in \Sigma_N, \dots, \mathbf{p}_T \in \Sigma_N} \left[\sum_{t=0}^{T+1} \lambda_t \mathbf{MK}_t(\mathbf{p}_{t-1}, \mathbf{p}_t) + \sum_{t=0}^{T} \varrho \left(L_{t+1}(\mathbf{p}_t) \right) \right], \tag{2.12}$$

and is still refereed to as *Multidate Monge Kantorovich (MMK)* problem.

Theorem 2.3 (Existence of Solution)**.** *If the risk measure ρ satisfies the risk measure properties 2.1, the set of solutions for the Multidate Monge Kantorovich problem (MMK) (2.12) is a non-empty convex and closed.*

Proof. Note that $\mathbf{MK}_t(\mathbf{p}_{t-1}, \mathbf{p}_t)$ is finite valued for all t since the primal formulation is clearly bounded from below (non-negative cost c_t). Regarding the dual formulation of the MK problem (2.11), we observe $\mathbf{MK}_t(\mathbf{p}_{t-1}, \mathbf{p}_t)$ as a maximum of linear functions with coefficients valued in a non-empty convex polyhedral set. Applying [30, Theorem 8.7.2], this polyhedral set can be decomposed as the sum of a convex hull (of a finite number of extremal points) and a conical hull: observe that these extreme points can achieve all possible optimal values of the dual problem. Therefore, $\mathbf{MK}_t(\mathbf{p}_{t-1}, \mathbf{p}_t)$ can be viewed as a maximum of linear functions over this *finite* set of extreme points: consequently, on the one hand $(\mathbf{p}_{t-1}, \mathbf{p}_t) \in \Sigma_N^2 \mapsto \mathbf{MK}_t(\mathbf{p}_{t-1}, \mathbf{p}_t)$ is a piecewise linear function, hence (Lipschitz) continuous, and on the other hand, it is convex. In addition, $\mathbf{p}_t \in \Sigma_N \mapsto \varrho \left(L_{t+1}(\mathbf{p}_t) \right)$ is convex and continuous, thanks to Properties (2.1-(c)-(b)). The global cost function MMK in (2.12) is thus convex and continuous in $\mathbf{p}_0, \dots, \mathbf{p}_T \in \Sigma_N^{T+1}$ and the state space Σ_N^{T+1} is convex and compact which ensures the existence of a solution. Additionally, according to [21, Lemma VII-1.0.1.], the set of solutions is closed and convex. □

Replacing all the $\mathbf{MK}_t(\mathbf{p}_{t-1}, \mathbf{p}_t)$ by their dual definitions (2.11), (2.12) rewrites as:

$$\begin{array}{ccc} \min \\ \mathbf{p}_t \in \Sigma_N, \\ t \in \{0, \dots, T\} \end{array} \quad \begin{array}{ccc} \max \\ f_t, g_t, \\ f_{t-1,i} + g_{t,j} \leq c_t^{i,j} \end{array} \quad \sum_{t=0}^{T+1} \lambda_t \left(\langle f_{t-1}, \mathbf{p}_{t-1} \rangle + \langle g_t, \mathbf{p}_t \rangle \right) + \sum_{t=0}^{T} \varrho(L_{t+1}(\mathbf{p}_t)), \tag{2.13}$$

or equivalently, after rearranging the summation, as:

$$\begin{array}{ccc} \min \\ \mathbf{p}_t \in \Sigma_N, \\ t \in \{0, \dots, T\} \end{array} \quad \begin{array}{ccc} \max \\ f_t, g_t, \\ f_{t,i} + g_{t+1,j} \leq c_{t+1}^{i,j} \end{array} \quad \left[\sum_{t=0}^{T} \Big(\langle \mathbf{p}_t, \lambda_{t+1} f_t + \lambda_t g_t \rangle + \varrho(L_{t+1}(\mathbf{p}_t)) \Big) \right. \tag{2.14}$$

$$+ \lambda_0\langle f_{-1}, \mathbf{p}_{\text{now}}\rangle + \lambda_{T+1}\langle g_{T+1}, \mathbf{p}_{\text{targ}}\rangle\Big].$$

Owing to Theorem 2.3, the minimum problem has a solution $(\mathbf{p}_t)_{t=0}^T$. On the other hand, since the primal of $\mathbf{MK}_t$ (2.10) is clearly bounded from below, the dual problem also has a solution $(f_t, g_{t+1})_{t=-1}^T$. Actually the $(\mathbf{p}_t, f_t, g_t)_t$-solution set of problem (2.13) or (2.14) is unbounded. Indeed, for each given $(\mathbf{p}_t)_{t=0}^T$ with a respective maximal solution $(f_t, g_{t+1})_{t=-1}^T$, we claim that $(f_t - \mathbf{v}_t, g_{t+1} + \mathbf{v}_t)_{t=-1}^T$ is also solution for any component-wise constant vector $\mathbf{v}_t = (\nu_t, \dots, \nu_t)$. To see this, observe that taking $(f_{t-1} - \mathbf{v}_{t-1}, g_t + \mathbf{v}_{t-1})$ does not modify the t-term of the first sum in (2.13) (using that $\mathbf{p}_t$ and $\mathbf{p}_{t-1}$ are probability measures) and that it still fulfills the constraints $f_{t-1,i} + g_{t,j} \le c_t^{i,j}$.

Therefore, a small transformation on the feasible set provides us a simpler solution set (and will allow us to achieve more accurate results, in particular in Theorem 2.6): for each t, one can choose ν_{t-1} such that the sum of components of $g_t + \mathbf{v}_{t-1}$ equals 0, and then replace (f_{t-1}, g_t) by $(f_{t-1} - \mathbf{v}_{t-1}, g_t + \mathbf{v}_{t-1})$. As seen before, it does not modify the MMK problem optimal value (2.14) nor the optimal solutions $(\mathbf{p}_t)_{t=0}^T$. Thus we have proved the following result.

Theorem 2.4. *The MMK problem* (2.12) *has the same optimal value than*

$$\begin{array}{c}\min\\ \mathbf{p}_t \in \Sigma_N,\\ t \in \{0, ..., T\}\end{array} \quad \begin{array}{c}\max\\ f_t, g_t,\\ f_{t,i} + g_{t+1,j} \le c_{t+1}^{i,j},\\ \sum g_{t+1,j} = 0\end{array} \quad \Bigg[\sum_{t=0}^{T}\Big(\langle \mathbf{p}_t, \lambda_{t+1} f_t + \lambda_t g_t\rangle + \varrho(L_{t+1}(\mathbf{p}_t))\Big) \tag{2.15}$$

$$+ \lambda_0\langle f_{-1}, \mathbf{p}_{\text{now}}\rangle + \lambda_{T+1}\langle g_{T+1}, \mathbf{p}_{\text{targ}}\rangle\Bigg].$$

Moreover, the solutions set $\{(\mathbf{p}_t)_{t=0}^T : (\mathbf{p}_t)_{t=0}^T$ *solution of problem* (2.4)$\}$ *is the equal to the solutions set* $\{(\mathbf{p}_t)_{t=0}^T : (\mathbf{p}_t)_{t=0}^T$ *solution of problem* (2.15)$\}$.

2.5 Saddle-Point problem

It is important to notice that the above problem can be thought of as a bilinear saddle-point problem, which will provide special properties and numerical methods. We recall that a bilinear saddle-point problem is an optimization problem written under the generic form

$$\min_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}} \Big(\langle x, A^\top y\rangle + \varphi(x) - \psi(y)\Big) \tag{2.16}$$

where $\mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{Y} \subset \mathbb{R}^m$ are convex sets, $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear operator and φ, ψ are convex functions.

Theorem 2.5 (Bilinear Saddle-Point Representation)**.** *The MMK problem* (2.15) *is a Bilinear Saddle-Point Problem of the form* (2.16)*, where* $x := (\mathbf{p}_0, \dots, \mathbf{p}_T)^\top \in \mathbb{R}^n$ *with* $n := (T+1)N$*,* $y := (f_{-1}, f_0, g_0, \dots, f_T, g_T, g_{T+1})^\top \in \mathbb{R}^m$ *with* $m := 2(T+2)N$*, with a linear function* ψ*.*

In (2.16), the choice of denoting the matrix $A^\top$ and not just A is made in several references in the context of bilinear saddle-point problems. This choice is usually inspired by a primal-dual problem, a minmax problem where the objective function is the Lagrangian of a convex optimization problem of the form $\min_{x\in\mathcal{X},\ Ax=b}\varphi(x)$, where y represents the dual variable. In general, saddle-point problems as (2.16) (also called minmax problems) have a large presence in the literature with several variations on convexity or smoothness of functions φ, ψ (see [40, 7]). These problems also pop-up in various applications, like image processing [7], resource allocation [18].

Proof of Theorem 2.5. With the notation for x, y given in the Theorem statement, the sum in (2.15) writes as

$$\sum_{t=0}^{T}\langle \mathbf{p}_t, A_t^\top \begin{pmatrix} f_t \\ g_t \end{pmatrix}\rangle + \varphi(x) - \psi(y) \tag{2.17}$$

where $\varphi(x) := \sum_{t=0}^{T}\varrho(L_{t+1}(\mathbf{p}_t))$, $\psi(y) := -\lambda_0\langle f_{-1}, \mathbf{p}_{\text{now}}\rangle - \lambda_{T+1}\langle g_{T+1}, \mathbf{p}_{\text{targ}}\rangle$, $A_t^\top \begin{pmatrix} f_t \\ g_t \end{pmatrix} := \lambda_{t+1}f_t + \lambda_t g_t$; in other words, $A_t^\top = (\lambda_{t+1}\mathrm{Id}_{\mathbb{R}^N}, \lambda_t\mathrm{Id}_{\mathbb{R}^N}) \in \mathbb{R}^N \otimes \mathbb{R}^{2N}$. Therefore, the sum (2.17) takes the required saddle-point form $\langle x, A^\top y\rangle + \varphi(x) - \psi(y)$ where the matrix $A^\top$ is a block matrix

$$A^\top := \begin{pmatrix} 0 & A_0^\top & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & A_1^\top & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & A_{T-1}^\top & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & A_T^\top & 0 \end{pmatrix}. \tag{2.18}$$

The convexity of φ stems from Properties (2.1)-(c). The state spaces of x and y are readily convex. We are done. □

We refer the reader to [21] and references therein for properties of saddle-point problems. In our case, we gather in the next result the following caracterisations of the solutions set. It extends the compactness and convexity properties of Theorem 2.3 to the dual variables of (2.15).

Theorem 2.6 (Solution Set of the Multidate MMK Problem)**.** *In any case, the solutions set $\mathcal{S}$ of the saddle-point problem (2.16) is non empty and is a product product $\mathcal{S} = \bar{X}\times\bar{Y} \subset \mathcal{X}\times\mathcal{Y}$. If, in addition, $\mathbf{p}_{\text{targ}}$ and $\mathbf{p}_{\text{now}}$ do not vanish ($\mathbf{p}_{\text{targ}}(\mathfrak{o}) > 0, \mathbf{p}_{\text{now}}(\mathfrak{o}) > 0$ for any $\mathfrak{o}\in\mathcal{O}$), then the sets $\bar{X}$ and $\bar{Y}$ are compact and convex.*

Proof. We are in a position to apply Sion's theorem (Theorem A.1), in particular owing to the compacity of $\mathcal{X}$ (as a tensor product of probability simplices). Therefore $\inf_x\sup_y\cdots = \sup_y\inf_x\ldots$, and as a consequence of Theorem A.2-(iii), $\mathcal{S}$ is nonempty and has a product form $\bar{X}\times\bar{Y}$ for some sets $\bar{X}, \bar{Y}$.

Under the extra set of conditions, we claim that it is enough to prove that there exists $x_0\in\mathcal{X}$ such that:

$$l(x_0, y) := \langle x_0, A^\top y\rangle - \psi(y) \xrightarrow[y\in\mathcal{Y}, \|y\|\to\infty]{} -\infty. \tag{2.19}$$

Indeed, with (2.19) at hand, the compactness and convexity of $\mathcal{S}$ follow by [21, Theorem VII-4.3.1] (since $\mathcal{Y}$ is unbounded). But since $\mathcal{S}$ is a product space, $P_1(\mathcal{S}) = \bar{X}$ where P_1 is the projection onto the first n coordinates: P_1 being linear, $\bar{X}$ is also convex and compact. The same argument applies for $\bar{Y}$ and we are done with the compactness and convexity of $\bar{X}$ and $\bar{Y}$.

We now prove (2.19). In view of (2.13) and the notations of Theorem 2.5, we have $l(x,y) = \sum_{t=0}^{T+1} S_t$ where

$$S_t := \lambda_t \left(\langle f_{t-1}, \mathbf{p}_{t-1} \rangle + \langle g_t, \mathbf{p}_t \rangle \right).$$

First, note that (2.19) is equivalent to show that for every sequence $(y^k = (f_{-1}^k, f_0^k, g_0^k, \dots, f_T^k, g_T^k, g_{T+1}^k))_k$ such that $\|y^k\| \to +\infty$ there is a subsequence $(y^{k'}) \subset (y^k)$ such that $l(x_0, y^{k'}) \to -\infty$, provided that x_0 has been well chosen once for all.

$\triangleright$ *Let us justify that S_t is uniformly upper bounded.* Indeed,

$$S_t \leq \lambda_t \left(\sum_i (c_t^{i,j_t^*} - g_{t,j_t^*}) \mathbf{p}_{t-1,i} + \sum_j g_{t,j_t^*} \mathbf{p}_{t,j} \right) = \lambda_t \sum_i c_t^{i,j_t^*} \mathbf{p}_{t-1,i} \leq \sup_t \lambda_t \sup_{i,j,t} c_t^{i,j} =: \bar{S},$$

where j_t^* is such that $g_{t,j_t^*} = \max_j g_{t,j}$.

$\triangleright$ *We now prove that at least one S_t goes to $-\infty$, for some x_0;* then, combined with the above, we will be done with (2.19).

We choose $\mathbf{p}_t := (\frac{1}{N}, \frac{1}{N}, ..., \frac{1}{N}) \in \Sigma_N$, for $t \in \{0, ..., T\}$ for defining x_0. Since we consider $\|y^k\| \to +\infty$, as $k \to \infty$, either $\|g^k\| \to +\infty$ (passing through a subsequence if necessary) or $\|g^k\|$ is bounded and $\|f^k\| \to +\infty$ (again along a possible subsequence). We split the proof according to these two cases.

Case 1: $\|g^k\| = \|(g_t^k)_{t=0}^{T+1}\| \to +\infty$ as $k \to +\infty$.

Because $\sum_j g_{t,j}^k = 0$ for all $t \in \{0, ..., T+1\}$, there is a component j^* and a time t^* such that $g_{t^*,j^*}^k \to +\infty$. Particularly, we can choose j^* such that $g_{j*,t*}^k = \max_j g_{j,t^*}^k \to +\infty$ (initially the component depends on the index k but we can, passing by a subsequence if necessary, choose a fixed component j^* and t^*). Then

$$\begin{aligned} S_{t^*} &= \lambda_{t^*} \left(\sum_i f_{t^*-1,i} \mathbf{p}_{t^*-1,i} + \sum_j g_{t^*,j} \mathbf{p}_{t^*,j} \right) \\ &\leq \lambda_{t^*} \left(\sum_i (c_t^{i,j^*} - g_{t^*,j^*}) \mathbf{p}_{t^*-1,i} + \sum_j g_{t^*,j} \mathbf{p}_{t^*,j} \right) \\ &\leq \bar{S} + \lambda_{t^*} \sum_j (g_{t^*,j} - g_{t^*,j^*}) \mathbf{p}_{t^*,j} \end{aligned}$$

simply using that the sums of $\mathbf{p}_{t^*-1,i}$ over i and $\mathbf{p}_{t^*,j}$ over j are both equal to 1. Each term in the j-sum is non-positive; because the sum of $\sum_j g_{t,j}^k = 0$ and $g_{j*,t*}^k \to +\infty$, at least one component $g_{t,j}^k$ must be negative, which implies that

$$S_{t^*} \leq \bar{S} - \lambda_{t^*} g_{t^*,j^*} \inf_j \mathbf{p}_{t^*,j}.$$

The above upper bound goes to $-\infty$ since in any case of $t^* \in \{0, \dots, T+1\}$, $\mathbf{p}_{t^*,j} > 0$ ($\mathbf{p}_{t^*}$ is either the uniform distribution if $t^* \in \{0, \dots, T\}$ or equal to $\mathbf{p}_{\text{targ}}$ of $t^* = T+1$). We have proved $S_{t^*} \to -\infty$.

Case 2: $\sup_k \|g_k\| < +\infty$ and $\|f_k\| = \|(f_t^k)_{t=-1}^T\| \to +\infty$ as $k \to +\infty$.
Since $f_{t-1,i}^k \le c_t^{i,j} - g_{t,j}^k$ for any i, j, $f_{t-1,i}^k$ must be bounded from above by a constant uniform in k. In addition, there must be i^*, t^* such that

$$f_{t^*-1,i^*}^k \to -\infty \text{ as } k \to +\infty$$

(up to subsequence extraction). Then, clearly

$$\begin{aligned} S_{t^*} &= \lambda_{t^*} \left(\sum_i f_{t^*-1,i} \mathbf{p}_{t^*-1,i} + \sum_j g_{t^*,j} \mathbf{p}_{t^*,j} \right) \\ &\le \lambda_{t^*} f_{t^*-1,i^*} \inf_i \mathbf{p}_{t^*-1,i} + \sup_{t,i,k}(\lambda_t f_{t-1,i}^k) + \sup_{t,j,k}(\lambda_t g_{t,j}^k). \end{aligned}$$

It follows that $S_{t^*} \to -\infty$, because $\mathbf{p}_{t^*-1,i} > 0$ (recall that $\mathbf{p}_{t^*-1}$ is either the uniform distribution if $t^* \in \{1, \dots, T+1\}$ or equal to $\mathbf{p}_{\text{now}}$ of $t^* = 0$). We are done.

□

Solutions of saddle-point problems (2.16) need not to be a singleton. Particularly, when $\mathcal{Y} = \mathbb{R}^m$ and $\psi(y) = \langle b, y \rangle$, the problem (2.16) is equivalent to the optimization problem $\min_{Ax=b} \varphi(x)$ that can have infinitely many solutions if φ is not strictly convex. By strengthening conditions on φ, the solution in $\mathbf{p}_t$ is unique.

Proposition 2.1. *If φ is strictly convex, then there is an unique solution $(\mathbf{p}_t)_{t=0}^T$ for the problem (2.15). For the MMK problem, saying that φ is strictly convex is equivalent assaying that ϱ is strictly convex with respect to $\mathbf{p}_t$.*

Proof. Consider the function

$$\Phi(x) := \max_{y \in \mathcal{Y}} \langle x, A^T y \rangle - \psi(y);$$

Φ is convex as it is the maximum of linear functions (with respect to x). Adding the strictly convex function φ gives a strictly convex function $x \mapsto \Phi(x) + \varphi(x)$. Then problem (2.16) (or equivalently (2.15)) boils down to a strictly convex optimization problem in a compact set, therefore with a unique minimum. □

3 Numerical experiments

Due to the important role of saddle-point Problems in recent applications, notably in image processing, statistics and convex otimization ([10],[23],[41], a large range of methods are available to solve this type of problem. Several convex optimization methods as of Frank-Wolfe Method and Mirror methods [15, 25] were adapted to these maxmin problems, specially in the case of strictly convex

functions φ and ψ. Nevertheless, one of the most useful and comprehensive methods in applications are still the Primal-Dual method, first presented by Arrow and al. [2] and then its further variations [44, 10].

3.1 Primal-Dual algorithm for the bilinear saddle-point problem

The Primal-Dual method (PD) treats general coupling functions with a segregated scheme (isolating variables x and y) which is particularly interesting to our case of study. A way of understanding the intuition behind the Primal-Dual algorithm is to rewrite problem (2.16) in the form of its first order optimality conditions. Then, we want to find saddle points $(\bar{x}, \bar{y}) \in \mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^n \times \mathbb{R}^m$ such that:

$$0 \in \partial(\varphi + \mathbf{0}^e_{\mathcal{X}})(\bar{x}) + A^T\bar{y}, \quad 0 \in \partial(\psi + \mathbf{0}^e_{\mathcal{Y}})(\bar{y}) - A\bar{x} \tag{3.1}$$

where

$$\mathbf{0}^e_{\mathcal{C}} = \begin{cases} 0, & \text{if } x \in \mathcal{C}, \\ +\infty, & \text{if } x \notin \mathcal{C} \end{cases}$$

is a proper convex function (for non-empty convex set $\mathcal{C}$). For any proper convex function $g : \mathbb{R}^n \to (-\infty, +\infty]$, we have that the subgradient (see [21, Definition 1.1.4, Chapter VI], or [34] for the extension of the definition for proper functions):

$$\partial g(x) : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$$

is a well defined set-valued mapping. In particular, since g is convex, for every $x \in \mathbb{R}^n$ and every $r > 0$, the optimization problem

$$g_r(x) := \min_z \left(g(z) + \frac{1}{2r}\|z - x\|^2 \right) \tag{3.2}$$

defines the Moreau-Yosida regularization of g [20, Chapter XV, Section 4], and it has a unique minimizer $z \in \mathbb{R}^n$. In addition, the argmins of g_r and g coincide, see [20, Chapter XV, Theorem 4.1.7]. By the first order necessary and sufficient optimality condition, the optimal z in (3.2) is such that:

$$0 \in \partial g(z) + \frac{1}{r}(z - x) \Leftrightarrow x \in (\mathrm{Id} + r\partial g)(z) \Leftrightarrow z = (\mathrm{Id} + r\partial g)^{-1}(x)$$

and in particular, the so-called *Resolvent Operator*:

$$(\mathrm{Id} + r\partial g)^{-1} : \mathbb{R}^n \to \mathbb{R}^n$$

is well defined; the argmin of g is a fixed point of this operator. This result is a particular case of an extensive theory of Maximal Monotone Operators first developed by Minty (see [26, 27]). Applying this result to the operators in (3.1) we have that the Resolvent Operators

$$\left(\mathrm{Id} + \tau\partial(\varphi + \mathbf{0}^e_{\mathcal{X}} + \langle A^\top\bar{y}, \cdot\rangle)\right)^{-1} : \mathbb{R}^n \to \mathbb{R}^n \text{ and } \left(\mathrm{Id} + \sigma\partial(\psi + \mathbf{0}^e_{\mathcal{Y}} - \langle A\bar{x}, \cdot\rangle)\right)^{-1} : \mathbb{R}^m \to \mathbb{R}^m \tag{3.3}$$

are well defined for any $\tau, \sigma > 0$. Then, finding zeros in maximal monotone operators (3.1) is equivalent to finding fixed points of the Resolvent operators (3.3). This is the intuition behind the Arrow-Hurwicz algorithm [2] which alternates operators on x and y in the form

$$\begin{cases} x^{k+1} = (\mathrm{Id} + \tau\partial(\varphi + \mathbf{0}^e_{\mathcal{X}} + \langle A^\top y^k, \cdot\rangle))^{-1}(x^k), \\ y^{k+1} = (\mathrm{Id} + \sigma\partial(\psi + \mathbf{0}^e_{\mathcal{Y}} - \langle Ax^{k+1}, \cdot\rangle))^{-1}(y^k), \end{cases}$$

or in its classical form, for the same first order optimality conditions:

$$\begin{cases} x^{k+1} = (\mathrm{Id} + \tau\partial(\varphi + \mathbf{0}^e_{\mathcal{X}}))^{-1}(x^k - \tau A^\top y^k), \\ y^{k+1} = (\mathrm{Id} + \sigma\partial(\psi + \mathbf{0}^e_{\mathcal{Y}}))^{-1}(y^k + \sigma A x^{k+1}). \end{cases}$$

On this basis, the Primal-Dual algorithm extrapolates x^{k+1} and x^k defining an extra variable $\bar{x}^{k+1} = x^{k+1} + \theta(x^{k+1} - x^k)$. The most useful case in applications is $\theta = 1$, as we see in Algorithm 1 (taken from [10, Algorithm 1]). The pseudo-code of PD is presented as follows, where τ and σ are going to be specified in section 3.2.

Algorithm 1 Basic PD Algorithm

Result: $(\bar{x}, \bar{y})$ minimizer of (2.16)
Step 0: Choose $(x^0, y^0) \in \mathcal{X} \times \mathcal{Y}$
Step 1: $\hat{x}^{k+1} = x^k - \tau A^\top y^k$
Step 2 $x^{k+1} = \arg\min_{x \in \mathcal{X}} \left(\varphi(x) + \frac{1}{2\tau}\|x - \hat{x}^{k+1}\|^2 \right)$
Step 3: $\bar{x}^{k+1} = x^{k+1} + (x^{k+1} - x^k)$
Step 4: $\hat{y}^{k+1} = y^k + \sigma A \bar{x}^{k+1}$
Step 5: $y^{k+1} = \arg\min_{y \in \mathcal{Y}} \left(\psi(y) + \frac{1}{2\sigma}\|y - \hat{y}^{k+1}\|^2 \right)$.
Go back to **Step 1.**

3.2 Primal-Dual algortithm for the MMK problem

The proper functioning of the Basic PD algorithm depends on the realization of Steps 2 and 5. The structure of the *Multidate Monge Kantorovich* problem, when written as (2.13), allows a simple decomposition in time $t \in \{0, ..., T\}$ of these steps. Steps 2 and 5 become T minimization convex problems of the same type. The specific classification of the problem depends on the functions φ and ψ.

Algorithm 2 Basic PD Algorithm for the MMK problem

Result: $(\mathbf{p}_t)_{t=0}^T$ and $(f_t, g_{t+1})_{t=-1}^T$ minimizer of (2.16)

Step 0: Define $\mathbf{p}_{-1} = \mathbf{p}_{\text{now}}$, $\mathbf{p}_{T+1} = \mathbf{p}_{\text{targ}}$ and $\mathbf{p}_t^0 = (\frac{1}{N}, ..., \frac{1}{N})$ for $t \in \{0, ..., T\}$. Define:

$$(f_t^0, g_{t+1}^0) = \underset{f_{t,i}+g_{t+1,j} \le c_{t+1}^{i,j} \ \sum g_{t+1,j}=0}{\arg\min} \sum f_{t,i}\mathbf{p}_{t,i}^0 + \sum g_{t+1,j}\mathbf{p}_{t+1,j}^0,$$

for $t \in \{-1, ..., T\}$. Define τ and σ (see Algorithm 5).

Step 1: $\hat{\mathbf{p}}_t^{k+1} = \mathbf{p}_t^k - \tau(\lambda_{t+1} f_t^k + \lambda_t g_t^k)$, for all $t \in \{0, ..., T\}$

Step 2: $\mathbf{p}_t^{k+1} = \underset{p \in \Sigma_N}{\arg\min} \left(\varrho(L_{t+1}(p)) + \frac{1}{2\tau}\|p - \hat{\mathbf{p}}_t^{k+1}\|^2 \right)$, for all $t \in \{0, ..., T\}$

Step 3: $\bar{\mathbf{p}}_t^{k+1} = \mathbf{p}_t^{k+1} + (\mathbf{p}_t^{k+1} - \mathbf{p}_t^k)$, for all $t \in \{0, ..., T\}$

Step 4: $(\hat{f}_t^{k+1}, \hat{g}_t^{k+1}) = (f_t^k, g_t^k) + \sigma(\lambda_{t+1}\bar{\mathbf{p}}_t^{k+1}, \lambda_t \bar{\mathbf{p}}_t^{k+1})$, for all $t \in \{0, ..., T\}$

$\hat{f}_{-1}^{k+1} = f_{-1}^k$, $\hat{g}_{T+1}^{k+1} = g_{T+1}^k$

Step 5:

$$(f_t^{k+1}, g_{t+1}^{k+1})_{t=-1}^T = \underset{\substack{(f_t, g_{t+1})_{t=-1}^T, \\ f_{t,i} + g_{t+1,j} \le c_{t+1}^{i,j}}}{\arg\min} \psi(f_{-1}, g_{T+1}) + \frac{1}{2\sigma}\|(f_t, g_{t+1})_{t=-1}^T - (\hat{f}_t^{k+1}, \hat{g}_{t+1}^{k+1})_{t=-1}^T\|^2$$

where by a light abuse of notation, we insist on the dependence of ψ with respect to the active variables f_{-1}, g_{T+1}

Go back to **Step 1.**

Step 5 in Algorithm 2 consists in a naive application of the same step in Algorithm 1. It is possible to break step 5 in $T+2$ smaller problems of the same type. Indeed, we have:

$$\underset{\substack{(f_t, g_{t+1})_{t=-1}^T, \\ f_{t,i} + g_{t+1,j} \le c_{t+1}^{i,j}}}{\arg\min} \psi(f_{-1}, g_{T+1}) + \frac{1}{2\sigma}\|(f_t, g_{t+1})_{t=-1}^T - (\hat{f}_t^{k+1}, \hat{g}_{t+1}^{k+1})_{t=-1}^T\|^2 =$$

$$\underset{\substack{(f_t, g_{t+1})_{t=-1}^T, \\ f_{t,i} + g_{t+1,j} \le c_{t+1}^{i,j}}}{\arg\min} \psi(f_{-1}, g_{T+1}) + \frac{1}{2\sigma}\sum_{t=0}^{T-1}\|(f_t, g_{t+1}) - (\hat{f}_t^{k+1}, \hat{g}_{t+1}^{k+1})\|^2 + \frac{1}{2\sigma}\|(f_{-1}, g_0) - (\hat{f}_{-1}^{k+1}, \hat{g}_0^{k+1})\|^2$$

$$+ \frac{1}{2\sigma}\|(f_T, g_{T+1}) - (\hat{f}_T^{k+1}, \hat{g}_{T+1}^{k+1})\|^2.$$

Then, remembering that: $\psi(f_{-1}, g_{T+1}) = -\lambda_0\langle \mathbf{p}_{\text{now}}, f_{-1}\rangle - \lambda_{T+1}\langle \mathbf{p}_{\text{targ}}, g_{T+1}\rangle$, Step 5 is equivalent to the following

Algorithm 3 Step 5 in time steps

Step 5:

$$(f_t^{k+1}, g_{t+1}^{k+1}) = \underset{f_{t,i}+g_{t+1,j}\le c_{t+1}^{i,j}}{\arg\min} \|(f_t, g_{t+1}) - (\hat{f}_t^{k+1}, \hat{g}_{t+1}^{k+1})\|^2, \quad t \in \{0, ..., T-1\}$$

$$(f_{-1}^{k+1}, g_0^{k+1}) = \underset{f_{-1,i}+g_{0,j}\le c_0^{i,j}}{\arg\min} -\lambda_0\langle \mathbf{p}_{\text{now}}, f_{-1}\rangle + \frac{1}{2\sigma}\|(f_{-1}, g_0) - (\hat{f}_{-1}^{k+1}, \hat{g}_0^{k+1})\|^2$$

$$(f_T^{k+1}, g_{T+1}^{k+1}) = \underset{f_{T,i}+g_{T+1,j}\le c_{T+1}^{i,j}}{\arg\min} -\lambda_{T+1}\langle \mathbf{p}_{\text{targ}}, g_{T+1}\rangle + \frac{1}{2\sigma}\|(f_T, g_{T+1}) - (\hat{f}_T^{k+1}, \hat{g}_{T+1}^{k+1})\|^2$$

Another possible simplification is to replace in Step 4: $\hat{f}_{-1}^{k+1} = f_{-1}^k + \sigma\lambda_0\mathbf{p}_{\text{now}}$, and $\hat{g}_{T+1}^{k+1} = g_{T+1}^k + \sigma\lambda_{T+1}\mathbf{p}_{\text{targ}}$, providing a short version of Steps 4 and 5:

Algorithm 4 Short version of steps 4 and 5

Step 4:

$$(\hat{f}_t^{k+1}, \hat{g}_{t+1}^{k+1}) = (f_t^k, g_{t+1}^k) + \sigma(\lambda_{t+1}\bar{\mathbf{p}}_t^{k+1}, \lambda_{t+1}\bar{\mathbf{p}}_{t+1}^{k+1}), \quad t \in \{-1, ..., T\}$$

Step 5:

$$(f_t^{k+1}, g_{t+1}^{k+1}) = \underset{f_{t,i}+g_{t+1,j}\le c_{t+1}^{i,j}}{\arg\min} \|(f_t, g_{t+1}) - (\hat{f}_t^{k+1}, \hat{g}_{t+1}^{k+1})\|^2, \quad t \in \{-1, ..., T\}$$

When $\varrho(L_{t+1}(\mathbf{p}_t))$ is linear in $\mathbf{p}_t$, Steps 2 and 5 are $T+1$ and $T+2$ quadratic problems of size N and $2N$ respectively. This means that the complexity of the algorithm grows linearly in T. In the next paragraph we discuss an important part of Step 0 in Algorithm 2: the definition of parameters τ and σ.

Parameters of adaptive Primal-Dual algorithm. One of the primary difficulties in the implementation of PD algorithm is the careful choice of step-size parameters. The speed of the method depends on how these parameters are adapted to characteristics of the problem.

The main convergence theorem for Primal-Dual algorithm (Algorithm 1) states that the condition $\tau\sigma|||A|||^2 < 1$ on the parameters (τ, σ) is sufficient to ensure convergence (see [10, Theorem 1] or Proposition 3.1). Here A is defined in (2.18) and $|||\cdot|||$ is the matrix norm subordinated to the Euclidean norm $\|\cdot\|$, i.e. $|||A||| = \max\{\|Ax\| : x \in \mathbb{R}^{2N}, \|x\| = 1\}$. It is well-known that the norm of A equals the maximal eigenvalue of the matrix $A^\top A$, i.e. $|||A||| = \sqrt{\lambda_{\max}(A^\top A)}$. Because of the definition of $A^\top$ as a block-matrix, the matrix $A^\top A$ can be written as a block of matrices of the

form

$$A^\top A = \begin{pmatrix} A_0^\top A_0 & \cdots & \cdots & \cdots & \vdots \\ \vdots & A_1^\top A_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & A_{T-1}^\top A_{T-1} & \vdots \\ \vdots & \cdots & \cdots & \cdots & A_T^\top A_T \end{pmatrix}.$$

Since $A_t = \begin{pmatrix} \lambda_{t+1}\mathrm{Id}_{\mathbb{R}^N} \\ \lambda_t \mathrm{Id}_{\mathbb{R}^N} \end{pmatrix}$, we have that $A_t^T A_t = (\lambda_t^2 + \lambda_{t+1}^2)\mathrm{Id}_{\mathbb{R}^N}$, and therefore

$$|||A||| = \sqrt{\max_{t\in\{0,\dots,T\}}(\lambda_t^2 + \lambda_{t+1}^2)}.$$

This simple analysis allows us to obtain a sufficient upper bound to the product $\tau\sigma$:

$$\tau\sigma < \frac{1}{\max_{t\in\{0,\dots,T\}}(\lambda_t^2 + \lambda_{t+1}^2)}. \tag{3.4}$$

Inequality (3.4) partially solves the problem of choosing (τ, σ). To properly define the parameters, we study the proportion between τ and σ in our particular case. We note that $\mathbf{p}_t$ in Algorithm 2 is such that $\|\mathbf{p}_t\|_1 = 1$ (where $\|\cdot\|_1$ is the ℓ_1 norm), and $\|(f_t, g_t)\|_1$ may vary depending on the cost. Since the Primal-Dual algorithm sums the terms $\mathbf{p}_t$ and $\tau(\lambda_{t+1}f_t + \lambda_t g_t)$ (Step 1 in Algorithm 2), and the terms (f_t, g_t) and $\sigma(\lambda_{t+1}\mathbf{p}_t, \lambda_t\mathbf{p}_t)$ (Step 4 in Algorithm 2), it is important to re-adjust the order of magnitude of each term by the definition of parameters τ and σ. On this basis, we estimate that the ideal parameters τ, σ should have the proportions (for some $C > 0$)

$$\tau \approx C\frac{1}{\|\lambda_{t+1}f_t + \lambda_t g_t\|_1} \quad \text{and} \quad \sigma \approx C\frac{\|(f_t, g_t)\|_1}{\|(\lambda_{t+1}p_t, \lambda_t p_t)\|_1} = \frac{\|(f_t, g_t)\|_1}{\lambda_t + \lambda_{t+1}}, \qquad \text{for } t \in \{0, ..., T\},$$

then:

$$\frac{\sigma}{\tau} \approx \frac{\|\lambda_{t+1}f_t + \lambda_t g_t\|_1 \|(f_t, g_t)\|_1}{\lambda_t + \lambda_{t+1}}, \qquad \text{for } t \in \{0, ..., T\}. \tag{3.5}$$

Since τ, σ do not depend on t, the estimation (3.5) depends on the fact that the order of magnitude of the right hand side do not depend on t. To overcome this problem, we can define the proportion between σ and τ as the mean in t of the right-hand side in (3.5). Combining this rule with the inequality (3.4), we are able to define precisely the Step 0 of Algorithm 2, this is given in Algorithm 5.

Algorithm 5 Definition of τ and σ (Step 0 for Algorithm 2)

Step 0: Define $\mathbf{p}_{-1} = \mathbf{p}_{\text{now}}$, $\mathbf{p}_{T+1} = \mathbf{p}_{\text{targ}}$ and $\mathbf{p}_t^0 = (\frac{1}{N}, ..., \frac{1}{N})$ for $t \in \{0, ..., T\}$. Define:

$$(f_t^0, g_{t+1}^0) = \underset{f_{t,i}+g_{t+1,j} \leq c_{t+1}^{i,j},\ \sum g_{t+1,j}=0}{\arg\min} \sum f_{t,i}\mathbf{p}_{t,i}^0 + \sum g_{t+1,j}\mathbf{p}_{t+1,j}^0,$$

for $t \in \{-1, ..., T\}$. Define τ and σ such that:

$$\sigma\tau = \frac{1}{2\max_{t\in\{0,...,T\}}(\lambda_t^2 + \lambda_{t+1}^2)} \qquad \text{and} \qquad \frac{\sigma}{\tau} = \frac{1}{T+1}\sum_{t=0}^{T} \frac{\|\lambda_{t+1}f_t^0 + \lambda_{t+1}g_t^0\|_1 \|(f_t^0, g_t^0)\|_1}{\lambda_t + \lambda_{t+1}}.$$

We observe in the numerical tests that we are actually close to the ideal parameters τ, σ (see Section 3.4).

3.3 Convergence analysis

In this subsection we present the convergence definition of the bilinear saddle-point problem (2.16) consistently with definitions and results from A. Chambolle and T. Pock [10]. Since the above article defines the saddle-point problem in a vector space, we shall rewrite problem (2.16) as:

$$\min_{x\in\mathbb{R}^n} \max_{y\in\mathbb{R}^m} \left(\langle x, A^\top y\rangle - \psi^e(y) + \varphi^e(x)\right), \tag{3.6}$$

where $m = 2(T+2)N$ and $n = (T+1)N$, and where $\psi^e : \mathbb{R}^m \to [-\infty, +\infty)$, $\varphi^e : \mathbb{R}^n \to (-\infty, +\infty]$:

$$\psi^e(y) = \psi(y) - \mathbf{0}_{\mathcal{Y}}^e, \qquad \varphi^e(x) = \varphi(x) + \mathbf{0}_{\mathcal{X}}^e.$$

Then, ψ^e, φ^e are proper convex and lower semi-continuous functions. With this definition, saddle points for problems (2.16) and (3.6) are the same.

The most important definition of convergence for the Primal-Dual algorithm measures convergence from the gap between the min and the max parts of the bilinear problem.

Definition 3.1 (Convergence criteria of Primal-Dual gap)**.** *Consider $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$, and two subsets $B_1 \times B_2 \subset \mathbb{R}^n \times \mathbb{R}^m$, then the partial Primal-Dual gap $\mathcal{G}_{B_1,B_2}^e$ is defined as:*

$$\mathcal{G}_{B_1,B_2}^e(\bar{x}, \bar{y}) := \max_{y\in B_2}\{\langle \bar{x}, A^\top y\rangle - \psi^e(y) + \varphi^e(\bar{x})\} - \min_{x\in B_1}\{\langle x, A^\top \bar{y}\rangle - \psi^e(\bar{y}) + \varphi^e(x)\}.$$

The Primal-Dual algorithm (1) is said to converge by the Primal-Dual gap criteria if there are sets $B_1 \subset \mathbb{R}^n$ and $B_2 \subset \mathbb{R}^m$, such that there is a saddle-point $(\hat{x}, \hat{y})$ for problem (3.6) that lies in the interior of the set $B_1 \times B_2$ and the partial gap along the algorithm sequence converges to 0:

$$\mathcal{G}_{B_1,B_2}^e(x^k, y^k) \to 0.$$

We observe that, with this definition, if a saddle point $(\hat{x}, \hat{y})$ lies in $B_1 \times B_2$, then $\mathcal{G}_{B_1,B_2}^e(\bar{x}, \bar{y}) \geq 0$ for all $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$.

Remark 3.1. *We note that if* $(\bar{x},\bar{y}) \notin \mathcal{X}\times\mathcal{Y}$, *then* $\mathcal{G}^e_{B_1,B_2}(\bar{x},\bar{y}) = +\infty$. *For* $(\bar{x},\bar{y}) \in \mathcal{X}\times\mathcal{Y}$ *the definition 3.1 can also be written as:*

$$\mathcal{G}_{B_1,B_2}(\bar{x},\bar{y}) := \max_{y\in B_2\cap\mathcal{Y}}\{\langle \bar{x}, A^\top y\rangle - \psi(y)+\varphi(\bar{x})\} - \min_{x\in B_1\cap\mathcal{X}}\{\langle x, A^\top \bar{y}\rangle - \psi(\bar{y}) + \varphi(x)\}. \tag{3.7}$$

From now on, when $(\bar{x},\bar{y}) \in \mathcal{X}\times\mathcal{Y}$, we use equation (3.7) and the notation $\mathcal{G}_{B_1,B_2}$ to define the Primal-Dual gap.

Remark 3.2. *Note that the sequence* (x^k, y^k) *generated by the Primal-Dual Algorithm 1 lies in the set* $\mathcal{X}\times\mathcal{Y}$.

Proposition 3.1 (Convergence rate of the Primal-Dual gap)**.** *Let* $(\hat{x},\hat{y}) \in \mathbb{R}^n\times\mathbb{R}^m$ *and let* $B_1 \subset \mathbb{R}^n$ *and* $B_2 \subset \mathbb{R}^m$ *be such that* $(\hat{x},\hat{y})$ *lies in the interior of* $B_1\times B_2$. *Then, if the Primal-Dual gap*

$$\mathcal{G}^e_{B_1,B_2}(\hat{x},\hat{y}) = 0,$$

then $(\hat{x},\hat{y})$ *is a saddle point of problem (2.16) (or equivalently problem (3.6)).*

Besides, if $\tau\sigma < \frac{1}{\|A\|^2}$, *the sequence* (x^k, y^k) *generated by Algorithm 1 is convergent, and* $(x^k, y^k) \to (\hat{x},\hat{y})$, *for a saddle point* $(\hat{x},\hat{y})$. *In addition, the convergence order for the sequence of means:* $\tilde{x}^K = \frac{1}{K}\sum_{k=1}^{K} x^k$, $\tilde{y}^K = \frac{1}{K}\sum_{k=1}^{K} y^k$ *is* $\mathcal{O}(\frac{1}{K})$ *as* $K \to +\infty$, *in the sense*

$$\mathcal{G}_{B_1,B_2}(\tilde{x}^K, \tilde{y}^K) = \mathcal{O}(\frac{1}{K})$$

for any bounded sets $B_1 \subset \mathbb{R}^n$ *and* $B_2 \subset \mathbb{R}^m$.

Proof. See [10, Theorem 1]. We observe that in the notation of [10, Theorem 1] the function F^* is the convex conjugate of a function F, that is convex lower semi-continuous. Since ψ^e is convex and lower semi-continuous, we have that $\psi^{e,**} = \psi^e$ (see [22, Corollary 1.3.6, Appendix E]) which means that ψ^e can be viewed as the conjugate function of a convex lower semi-continuous function $\psi^{e,*}$ (see [22, Theorem 1.1.2, Appendix E]). The statement of [10, Theorem 1] assumes the existence of a saddle point $(\hat{x},\hat{y})$ for problem (3.6). This assumption is not necessary in our case since we have proved the existence of a saddle point for problem (2.16) (see Theorem 2.3).

We note that, since any element of the sequence (x^k, y^k) generated by the Primal-Dual algorithm 1 lies in the set $\mathcal{X}\times\mathcal{Y}$, so does the sequence of means $(\tilde{x}_K, \tilde{y}_K)$ (remind that $\mathcal{X}$ and $\mathcal{Y}$ are convex). This justifies the use of the notation $\mathcal{G}_{B_1,B_2}$ instead of the notation $\mathcal{G}^e_{B_1,B_2}$ (see remark 3.1). □

If we choose $B_1\times B_2$ to be the entire set $\mathbb{R}^m\times\mathbb{R}^n$, the global Primal-Dual gap is written as:

$$\mathcal{G}_{\mathcal{X},\mathcal{Y}}(\bar{x},\bar{y}) = \max_{y\in\mathcal{Y}}\{\langle \bar{x}, A^\top y\rangle - \psi(y) + \varphi(\bar{x})\} - \min_{x\in\mathcal{X}}\{\langle x, A^\top \bar{y}\rangle - \psi(\bar{y}) + \varphi(x)\},$$

and Proposition 3.1 states that:

$$\mathcal{G}_{\mathcal{X},\mathcal{Y}}(\bar{x},\bar{y}) = 0 \iff (\bar{x},\bar{y}) \text{ is a saddle point.}$$

For the MMK problem (2.13) we have that $\mathcal{X} = \{(\mathbf{p}_t)_{t=0}^T : p_t^{(i)} \geq 0,\ \sum_{i=1}^N p_t^{(i)} = 1\}$ and $\mathcal{Y} = \{(f_t, g_{t+1})_{t=-1}^T : f_{t,i} + g_{t+1,j} \leq c_{t+1}^{i,j}, \sum g_{t+1,j} = 0\}$, then $\mathcal{Y}$ is not a bounded set. Nevertheless, if we consider a sequence (x^k, y^k) defined by Algorithm 1 and if we verify that $\mathcal{G}_{\mathcal{X}\times\mathcal{Y}}(x^k, y^k) \to 0$, then it implies that for any bounded set $B_1 \times B_2 \subset \mathbb{R}^n \times \mathbb{R}^m$ such that a there is a saddle point $(\hat{x}, \hat{y}) \in B_1 \times B_2$ that

$$0 \leq \mathcal{G}_{B_1,B_2}(x^k, y^k) \leq \mathcal{G}_{\mathcal{X},\mathcal{Y}}(x^k, y^k) \to 0.$$

Then, verifying the convergence of the global Primal-Dual gap on $\mathcal{X} \times \mathcal{Y}$ implies the convergence of the partial Primal-Dual gap for any bounded set $B_1 \times B_2$ containing a saddle point.

By proposition 3.1, for any fixed τ and σ, such that $\tau\sigma|||A|||^2 < 1$, the algorithm converges with convergence order $\mathcal{O}(1/K)$ in the sequence of means. A faster convergence in practice of this method, if employed with well-chosen stepsize policies, is studied in [10, Section 5] and depends on the regularity of φ and ψ. Particularly, if φ or ψ are strictly convex, the Primal-Dual algorithm can achieve a convergence order $\mathcal{O}(1/K^2)$. These strict convexity conditions are satisfied in our framework if and only if the function φ is stricly convex (see Theorem 2.5).

3.4 Numerical analysis of convergence

The objective of this simulation study is twofold: to numerically verify convergence of the Primal-Dual algorithm according to Definition 3.1 for different parameters τ and σ, on a toy example; second, to assess the computational time as N and T increase.

3.4.1 Toy example

We consider a toy problem fixing all elements of the MMK problem (2.13) as follows:

- $T = 3$, $\lambda = 1/4$.
- The discretized region for the attributes $\mathfrak{o} = (\mathfrak{e}, \mathfrak{g})$ is $\mathcal{O} := [0, 15] \times [0, 15]$.
- The quantity of points on the grid $\mathbb{G}$ is $N = N_{\mathfrak{e}} \times N_{\mathfrak{g}} \in \{25, 100\}$, points are equispaced in each direction ($\mathfrak{e}$ or $\mathfrak{g}$) and $N_{\mathfrak{e}} = N_{\mathfrak{g}}$.
- The cost c_t is the squared Euclidean norm for $t \in \{0, 1, ..., 3\}$.
- $\mathbf{p}_{\text{now}}$ and $\mathbf{p}_{\text{targ}}$ are discrete probabilities on the grid $\mathbb{G}$. Considering the Voronoi cells of each $\mathfrak{o}_i \in \mathbb{G} : \mathcal{C}(\mathfrak{o}_i) = \{\mathfrak{o} : \arg\inf_j \|\mathfrak{o} - \mathfrak{o}_j\| = \mathfrak{o}_i\}$, for $i \in \{1, ..., N\}$, then set $\mathbf{p}_{\text{now},i} := \mathbf{P}_{\text{now}}(\mathcal{C}(\mathfrak{o}_i))$, and $\mathbf{p}_{\text{targ},i} := \mathbf{P}_{\text{target}}(\mathcal{C}(\mathfrak{o}_i))$ for some continuous distributions $\mathbf{P}_{\text{now}}$ and $\mathbf{P}_{\text{target}}$ specified below.
- $\mathbf{P}_{\text{now}}$ is defined by a *Gaussian mixture model*. Namely, consider the two Gaussian distributions $Z_{\text{now}}^1 \sim \mathcal{N}(\mu_1, C_1)$ and $Z_{\text{now}}^2 \sim \mathcal{N}(\mu_2, C_2)$, then:

$$\mathbf{P}_{\text{now}} \mid I \sim Z_{\text{now}}^I$$

  where $\mathbb{P}(I = 1) = \mathbb{P}(I = 2) = 1/2$. The parameters values are $\mu_1 = (3, 11)^\top$, $C_1 = 1.5\ \text{Id}_{\mathbb{R}^2}$, $\mu_2 = (10, 3)^\top$ and $C_2 = 1.7\ \text{Id}_{\mathbb{R}^2}$, where $\text{Id}_{\mathbb{R}^2}$ is the two-dimensional identity matrix.

- For $\mathbf{P}_{\text{target}}$ we choose another Gaussian distribution:

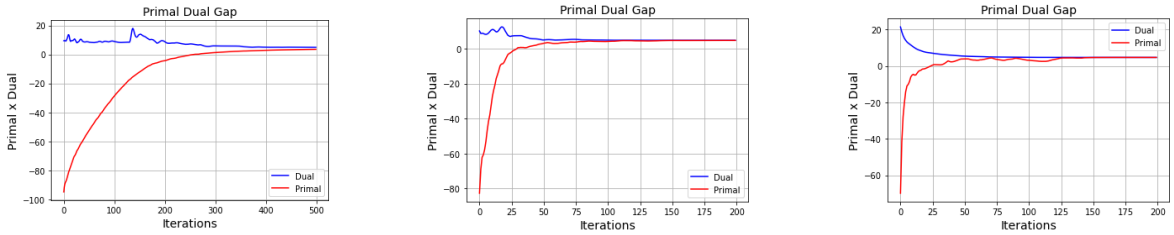
$$\mathbf{P}_{\text{target}} \sim \mathcal{N}(\mu, C)$$

where $\mu = (3,7)^\top$ and $C = 2\ \mathrm{Id}_{\mathbb{R}^2}$.

- $\rho = \mathtt{VaR}_\alpha$ with $\alpha = 99\%$.

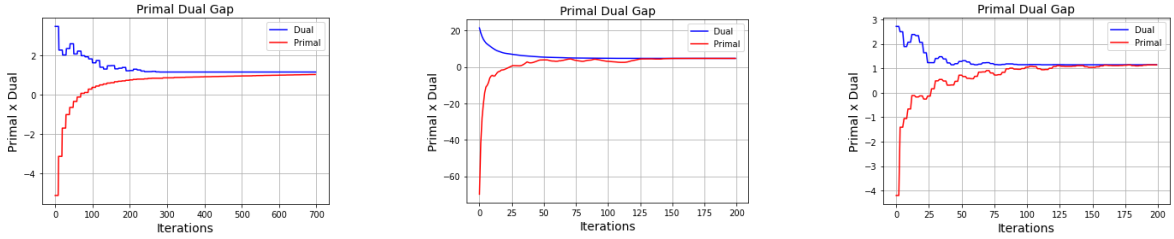
- The parameter θ depends exclusively of the $\mathfrak{g}$ coordinate, that is discretized on the grid $\mathbb{G}$ in values $\mathfrak{g}_k \in [0, 15]$, $k \in \{1, .., N_{\mathfrak{g}}\}$. The parameter $\theta = (\beta_1, \gamma)$ in the one-dimensional Gaussian copula model (2.2) (with $d = 1$) is distributed as $\theta \sim \text{Beta}(a_k, b_k) \otimes \text{LogNormal}(\gamma_k, \nu_k)$, where $a_k \in [0.18, 0.21]$, $b_k = 0.01$, $\gamma_k = [1.18, 2.1]$, $\nu_k = 0.1$; for each k, the values of a_k and γ_k are randomly chosen (uniformly on their respective intervals). These set of parameters is chosen in order to obtain realistic values of loss $L_{t+1}(\mathbf{P}_t) \approx 0.02$.

In the following plots, we verify the Primal-Dual gap convergence and compare different choices of algorithm parameters (τ, σ). The ideal parameter computed by Algorithm 5 is $\tau = 3/400, \sigma = 200$ for $N = 5$ and $\tau = 3/1000, \sigma = 500$ for $N = 10$, which results are reported in the center of Figures 1 and 2.



(a) $\tau = \frac{3}{40}, \sigma = 20$ (b) $\tau = \frac{3}{400}, \sigma = 200$ (c) $\tau = \frac{3}{4000}, \sigma = 2000$

Figure 1: Primal-Dual Gap for different values of parameters (τ, σ), $N = 25$



(a) $\tau = \frac{3}{100}, \sigma = 50$ (b) $\tau = \frac{3}{1000}, \sigma = 500$ (c) $\tau = \frac{3}{10000}, \sigma = 5000$

Figure 2: Primal-Dual Gap for different values of parameters (τ, σ), $N = 100$

When we move away from the natural proportion (3.5), even when respecting the parameter inequality in Proposition 3.1, the convergence is clearly slower. Different values of τ and σ, close to those that have the best performance in convergence, have been also tested, we do not report

the graphs of performance, they are very close to those presented in Figures 1 and 2. These tests confirm that the algorithm 5 for choosing (τ, σ) not only ensures convergence, but also yields it at a competitive speed.

3.4.2 Execution time

On this section, we argue that the computational time for the MMK problem (2.12) in the previous toy example remains reasonable for large numbers N of points in the discretization grid $\mathbb{G} \subset \mathcal{O}$ (the number of different obligor attributes in the portfolio) and time T. Observe that thanks to the characteristics of Primal-Dual Algorithm when applied to MMK problem (2), Step 2 and Step 5 can be performed in parallel in time $t \in \{0, ..., T\}$, meaning that in that case the execution time will be little affected by the number of time steps T (provided that the number of parallel threads is large enough).

Step 0 in Algo 5 solves 3 linear problems of size N that can be costly with respect to N. However, this has to be done only once, with the main purpose of computing the parameters (τ, σ). Moreover, a previous expert knowledge in time magnitude order of (f, g) would be likely sufficient to avoid these linear problems.

In the Table 1 we report the execution time of one iteration (Step 1 to Step 5) for different values of N and T when Steps 1 to 5 of Algorithm 2-Algorithm 4 are executed sequentially. The processor is Intel(R) Core(TM) i7-6600U CPU @2.60GHz, 2.81 GHz and we use Python langage and standard Python libraries for the implementation.

$T \backslash N$	25	100	125	200
2	1.3s	3.2s	55s	439s
5	1.6s	5.5s	130s	602s
10	1.8s	9.6s	170s	887s
20	2.3s	17s	416s	1120s

Table 1: Execution time of the Primal-Dual Algorithm 4.

The verification of the Primal-Dual gap requires a linear problem of size $2N$ for each iteration, which is costly for $N > 100$. On the other hand, it is known that Primal-Dual algorithm converges with order $\mathcal{O}(1/K)$ (see proposition 3.1) which can be used to reduce the necessity of computing the Primal-Dual gap (see Definition 3.1). Because of that, the computational time reported in Table 1 does not include the computation of the Primal-Dual gap.

In our experiments, we have observed that the number of iterations until a convergence with error tolerance of 10^{-4} for the Primal-Dual Gap was between 100 and 300 depending on the N and the difficulty of the problem. All in all, and in view of Table 1, leveraging on parallel computations, a MMK problem with $N = 200, T = 20$ can be solved in approximately $300 \times 1120s/20 \approx 5$ hours. Besides, in our example the risk metric ρ induces a linear dependence of the credit risk term as a function of $\mathbf{p}_t$ (Lemma A.1): presumably, a strictly convex term $\mathbf{p}_t \mapsto \rho(L_{t+1}(\mathbf{p}_t))$ would speed up the convergence of the algorithm by choosing carefully parameters (τ_k, σ_k), depending on the iteration (see [10, Section 5.1]). This is left to further investigation.

4 Financial data

4.1 Overall description

The *Multidate Monge Kantorovich* model presented in Section 2.3 allows a variety of interpretations and data sets. Clearly, the cost c_t is an important parameter of the model and needs to be adjusted according to the desired context. In the current interpretation, c_t is defined based on an economic sector and an environmental index. For the sake of our experiments, we will consider fictitious portfolio distribution ($\mathbf{p}_{\text{now}}$) based on realistic obligors $\mathfrak{o}$, corresponding to large companies.

The data set $\mathcal{D}$ used to statistically calibrate the credit risk model and define the transport cost is composed by a set of companies. These companies can be interpreted as potential obligors in the portfolio. Obligors are identified by ordered pairs of attributes $\mathfrak{o} = (\mathfrak{e}, \mathfrak{g})$, where $\mathfrak{e}$ represents an ESG index, and $\mathfrak{g}$ an economic sector. The cost c_t do not depend on the time t and is defined by a combination of $\mathfrak{e}$ and $\mathfrak{g}$. The sector $\mathfrak{g}$ provides an economic information about the couple return/risk of the obligor, see more details in Section 4.3.

ESG is an environmental, social, and governance criteria that measures a set of standards of a company's operations. Environmental criteria consider how a company performs as a steward of nature. Social criteria examine how it manages relationships with employees, suppliers, customers, and the communities where it operates. Governance deals with a company's leadership, executive pay, audits, internal controls, and shareholder rights. Larger values of ESG (see Section 4.2) mean better scores in social, environmental and governance criteria. For our experiments, the employed ESG index is proposed by Arasbesque Group [2]. With the benefit of being an open data index with last updated in may 2021, this ESG index is a quantitative data tool that analyses the sustainability performance of over 7.000 of the world's largest listed corporations using quantitative models and data scores.

We are aware that for the purpose of portfolio alignment to Paris agreement as mentioned in introduction, using ESG index (measuring environmental impacts, but also other responsibility factors) for the obligor attribute $\mathfrak{e}$ is questionable: having an index reporting more accurately on direct and indirect CO2-emissions (Scopes 1 to 3) would be more in the spirit of the Paris agreement; nevertheless, to the best of our knowledge, such a data base of CO2-emission type index is under construction, and not yet available; alternatively, we have preferred the well-document ESG index which constitutes a reasonable proxy. Hopefully, in the near future (because of the many quantitative initiatives to address climate risk), a more precise dataset will be available, so to improve the description of attribute $\mathfrak{e}$ and the accuracy of the cost c_t, and consequently improves the interpretation of the MMK model outputs.

Our data set is also composed by the Credit Rating of each one of the companies. The database is mainly provided by the open data website Kaggle [3] and concerns the period between 2005 and 2015. The Credit Rating is a classification of the credit risk of a company made by Financial Risk Agencies. The Credit Risk Agencies present in our data set are: DBRS, Egan-Jones Ratings Company, Fitch Ratings, Moody's Investors and Standard & Poor's Ratings Services. The data is

[2] https://www.arabesque.com/s-ray/our-scores/

[3] https://www.kaggle.com/agewerc/corporate-credit-rating

also freely available. The classification comprises indices in the set

$$\{AAA, AA, A, BBB, BB, B, CCC, CC, C\},$$

ranging from the most secure obligors to the least.

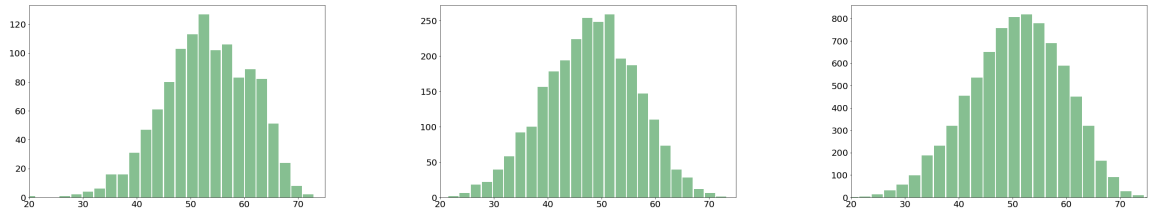
Clearly, credit risk is closely related to the Credit Spread, that is, the difference between the return of a risky debtor and the default-risk return (sovereign debts say). We benefit from Credit Rating to estimate the Credit Spread of companies and use it as one of the factors to define the cost c, see details in Section 4.3.

Each company in the data set is identified with an activity sector $\mathfrak{g}$. The possible sectors are: Consumer Services (Restaurants, Travel Agencies, Lodging), Consumer Non-Durables (Food Companies, Alcohol), Energy Minerals (Oil & gas, energy generation), Non-Energy Minerals (Chemicals), Productor Manufacturing (Industrials), Utilities (Basic Amenities, such as water and electricity,), Retail Trade (Retail, Distribution, e-commerce), Health Technology (Medicine devices, Biotechnology), Health Services (Health Plans), Technology Services (Communication, Social Networks, Softwares), Electronic Technology (Mobile devices, Semiconductors), Transportation (Airplanes and Autos), Process Industries (Tobacco, Textiles) and Finance (Banks, Hedge Funds).

The calibration of the 1-factor Gaussian model (Section 2), is made using the correlation between fluctuations of the stock market prices and the Gross domestic product (GDP in short) in a given period. For more details, see Section 4.3.3.

4.2 ESG and Credit Rating Distributions

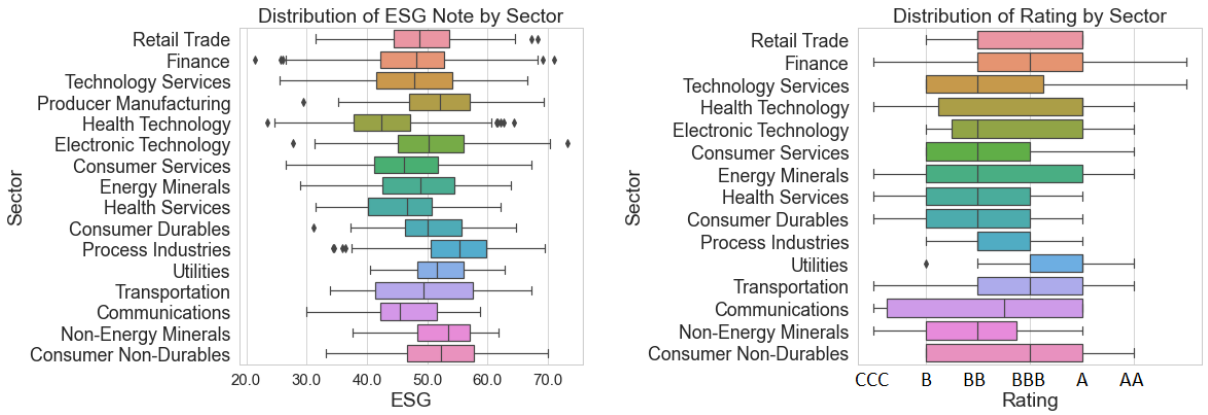
To maximise available data, we select our portfolio from American companies. The geographical concentration makes the comparison between sectors and the 1-factor risk model (see Section 4.3.3) more accurate and, in addition, American companies are a good approximation for the world companies in terms of ESG (according to Arabesque S-Ray), as we can see from the histograms in Figure 3. The selected database relies on 2660 American companies in 16 sectors. Companies are grouped



(a) ESG Europe (b) ESG United States (c) ESG World

Figure 3: ESG distributions for Europe, United States and the World

by activity sector, see Figure 4 for the box-and-whisker plots of ESG and Credit Rating according to activity sector. The cost c between two obligors (enabling us to compute the distance between two portfolios) is based on the ESG index of the obligors and the Credit Spread of obligors. The Credit Spread is computed using the available Credit Rating Data, for more details see Section 4.3.



(a) Distribution of ESG

(b) Distribution of Credit Rating

Figure 4: Distribution of ESG and Credit Rating according to activity sector

4.3 Relations between the Database and the MMK Model

4.3.1 Portfolio

The final data set $\mathcal{D}$ is composed of 470 American companies for which both ESG and Credit Rating are available.

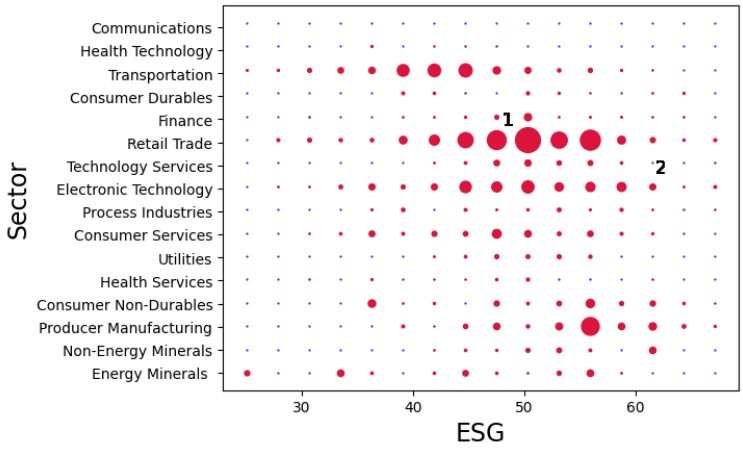


Figure 5: Example of Portfolio

We recall that elements in $\mathcal{D}$ are identified by two attributes $\mathfrak{o} = (\mathfrak{e}, \mathfrak{g})$, where $\mathfrak{e}$ is an environmental indicator - here the ESG score of the company, and $\mathfrak{g}$ is the group - here an economic sector. Consistently with our framework, we consider the set of obligors as a discretized region, i.e. $\mathcal{O} = \{\mathfrak{o}_1, ..., \mathfrak{o}_N\} \subset \mathbb{R}^+ \times \{\text{Activity sectors}\}$: on the horizontal axis, we find the ESG score, and on the vertical axis the sector. The set of obligors $\mathcal{O}$ is a grid of uniformly-distributed points

covering the interval of possible ESG scores and Sectors. A portfolio at time t is a probability vector $(p_t^{(1)}, \dots, p_t^{(N)})$, where $p_t^{(i)} = \mathbf{P}_t(O = \mathfrak{o}_i)$ which reads as the percentage of the obligors in the sector $\mathfrak{g}_i$ with ESG score $\mathfrak{e}_i$.

The data set $\mathcal{D}$ fixes the region of possible obligors. Different portfolios (vectors $\mathbf{p}_t$) can be considered in the whole discretized region $\mathcal{O}$. A portfolio at time t can be represented as circles in the grid for which the size is proportional to the probability $p_t^{(i)}$ of each grid point. For instance, in Figure 5, point 1 represents the companies in the Retail Trade sector that have an ESG index between 45.95 and 49.15. An example of a company that could be in this portfolio is the Walmart.Inc, a Retail and Trade company with an ESG of 46.59. Point 2 represents the companies of the Electronic Technology sector that have an ESG index between 60.25 and 63.05. An example of company that could be in this portfolio is the Lockheed Martin Corporation with an ESG of 61.49.

4.3.2 Definition of the Cost

With the purpose of defining a distance between sectors, we consider the mean of the credit spread in each sector as the incentive for investing (or not) in the sector following a rentability criteria. The European Commission [4] proposes a equivalence between credit rating and credit spread as in Table 2.

Rating	AAA	AA	A	BBB	BB	B	CCC	CC	C
Credit Spread (%)	0.05%	0.15%	0.25%	1%	7.5%	20%	28%	34%	40%

Table 2: Conversion from Rating to Credit Spread

This equivalence is used in our model to compute the cost. The cost $c(\mathfrak{o}, \mathfrak{o}')$ (independent on the time t) is then composed of three parts. The first one considers the difference credit spread of the obligors: $\mathfrak{o}, \mathfrak{o}'$. The second one considers the difference in the ESG score of the obligors. The third one considers a fixed cost c_{fixed} of replacing contracts. All in all, it gives

$$c(\mathfrak{o}, \mathfrak{o}') = c_{\text{fixed}} + \lambda(\overline{\text{CS}}(\mathfrak{g}') - \overline{\text{CS}}(\mathfrak{g})) + |\mathfrak{e}' - \mathfrak{e}|,$$

where λ is a normalization coefficient, and $\overline{\text{CS}}(\mathfrak{g})$ the mean credit spread CS (according to the above Table) over the companies in the group $\mathfrak{g}$. As an example, the Retail Trade sector has a mean credit spread of 4.5%, while the Electronic Technology sector has a mean credit spread of 8.2%.

4.3.3 Credit Parameter Distribution

For the purpose of calibrating the one-factor Gaussian copula model described in Section 2.2 ($d = 1$), we need to define the distribution of the parameter $\theta = (\beta, \gamma)$[5] conditionally on the obligor attributes $\mathfrak{o}$, i.e. $q_{t,\mathfrak{o}}(\mathrm{d}\theta)$. Ideally, this conditional distribution on the parameter θ should be inferred from the data set. Besides, for the convenience in our experiments of having a linear credit risk loss (Lemma

[4] https://eur-lex.europa.eu/legal-content

[5] to simplify notation, we write β for the parameter β_1 in (2.2) when $d = 1$.

A.1), we restrict to 10 sectors for which companies have positive β (the correlation parameter in the Gaussian copula model). The selection of these sectors is made according to the estimated β following the methodology below. The selected sectors are listed in Table 3.

Actually, a direct inference of the conditional distribution $q_{t,\mathfrak{o}}(\mathrm{d}\theta)$ from the data set has not been possible, since unfortunately, although existing, these data are not publicly available (in risk departments of financial institutions, they are usually inferred from the observed historical default rates). Alternatively for our numerical study, we assume that the parameter θ depends mostly on the activity sector $\mathfrak{g}_k, k \in \{1, ..., 10\}$ (and not on the ESG score), reducing the problem to identifying the distributions of (β_k, γ_k) for each $k \in \{1, ..., 10\}$. Since $\beta_k \in (0, 1)$ and $\gamma_k > 0$ it is natural to suppose that $\beta_k \sim \mathrm{Beta}(a_k, b_k)$ for[6] some $(a_k, b_k) \in (0, +\infty)^2$ and $\gamma_k \sim \mathrm{Lognormal}(\mu_k, \sigma_k)$, for some $(\mu_k, \sigma_k) \in \mathbb{R} \times (0, +\infty)$. All in all, we set

$$q_{t,\mathfrak{o}}(\mathrm{d}\theta) = \mathrm{Beta}(a_k, b_k) \otimes \mathrm{Lognormal}(\mu_k, \sigma_k).$$

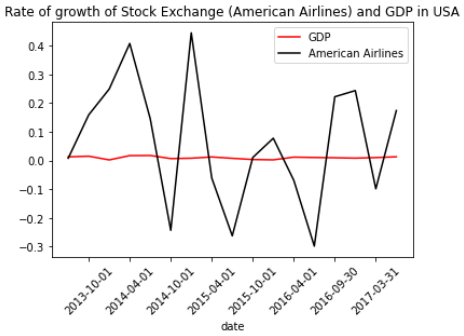
Let us now discuss how we fit the parameters $a_k, b_k, \mu_k, \sigma_k$ to realistic values.

Parameters for the distribution of β_k. The credit-risk parameter β_k represents the correlation between the global economy and the economic sector $\mathfrak{g}_k$. Identifying the parameters a_k, b_k then can be made by matching the two first statistical moments of this correlation among different companies from the sector $\mathfrak{g}_k$, with the two first moments of the $\mathrm{Beta}(a_k, b_k)$-distribution.

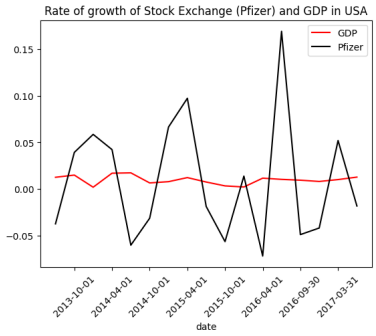
To get samples of β_k, one should compute (for many companies) the correlation between the GDP growth and the stock fluctuations of companies in the database for the sector $\mathfrak{g}_k$. Since we do not dispose of the historical data of the stock price for each company in the database $\mathcal{D}$ (for which ESG and Credit Rating are available), we have replaced the companies in $\mathcal{D}$ by the 500 most important American companies in each sector according to Standard & Poor's agency: the S&P500 companies. Doing so, we assume that for each economic sector $\mathfrak{g}_k$, the correlation between the growth rate of the stock prices of companies in database $\mathcal{D}$ (in sector $\mathfrak{g}_k$) and of the GDP of United States in the same period (2013-2017) has statistically the same distribution as when computed with S&P500 companies; this transfer assumption is coherent with the one-factor Gaussian copula model.

Figure 6 plots the GDP in United States and the stock prices for two companies in S&P500, which serves to get samples of β according to their respective sectors.

[6] the Beta distribution is taken with a support equal to $(0, 1)$



(a) American Airlines, in sector "Transportation"



(b) Pfizer, in sector "Health Technology"

Figure 6: Growth rates of two important American companies in their sectors

Proceeding like this for companies in S&P500, and computing correlation on a quarterly basis, we obtain the two first statistical moments of β_k according to the sector $\mathfrak{g}_k$: we denote by $\widehat{\text{Mean}}(\beta_k)$ and $\widehat{\text{Variance}}(\beta_k)$ these empirical mean and variance, their values are reported in Table 3. The results are coherent with the intuition behind these correlations: sectors like Energy Minerals, Non-Energy Minerals and Producer Manufacturing have globally more important correlations with the GDP than Communications or Electronic Technology.

Sector	Mean	Variance
Transportation	0.0803	0.0322
Electronic Technology	0.0898	0.0570
Health Technology	0.1413	0.0446
Utilities	0.1092	0.0443
Non-Energy Minerals	0.3013	0.0101
Producer Manufacturing	0.1306	0.0535
Health Services	0.2038	0.0366
Energy Minerals	0.2682	0.0968
Consumer Durables	0.1154	0.0886
Communications	0.0876	0.0126

Table 3: First statistics of the correlation between GDP and stock prices for each sector

Now, the parameters a_k, b_k for the distribution of $\beta_k \sim \text{Beta}(a_k, b_k)$ are calibrated so that to match the two first moments of the model with those of the data, i.e. $\frac{a_k}{a_k+b_k} = \mathbb{E}\left(\text{Beta}(a_k, b_k)\right) = \widehat{\text{Mean}}(\beta_k)$ and $\frac{a_k b_k}{(a_k+b_k)^2(a_k+b_k+1)} = \text{Var}\left(\text{Beta}(a_k, b_k)\right) = \widehat{\text{Var}}(\beta_k)$. After simple computations we obtain

$$a_k = \widehat{\text{Mean}}(\beta_k)\Big(\frac{(1-\widehat{\text{Mean}}(\beta_k))\widehat{\text{Mean}}(\beta_k)}{\widehat{\text{Var}}(\beta_k)} - 1\Big) \quad \text{and} \quad b_k = a_k\Big(\frac{1}{\widehat{\text{Mean}}(\beta_k)} - 1\Big).$$

Parameters for the distribution of $\gamma_k \sim$Lognormal(μ_k, σ_k^2). As we have done for the calibration of the β_k-distribution, we are to collect samples of γ_k and then fit the lognormal distribution parameters by matching the two first moments between the lognormal model and the data. To get samples of γ_k from each company in the sector $\mathfrak{g}_k$, we leverage the default probability formula (2.2):

$$\mathbb{P}\left(X_k \geq \gamma_k \mid F\right) = \Phi\left((\beta_k F - \gamma_k)/\sqrt{1-\beta_k^2}\right),$$

and match it with the default probability computed from the credit spread CS for the considered company:

$$\text{Default Probability} = 1 - \exp(-\text{CS} \times H),$$

where H is the amount of years in the time period. We recall that our data set includes the company rating and the quantity CS is computed using the conversion Table 2. Because β_k are not publicly available for these companies, we replace it by its mean $\widehat{\text{Mean}}(\beta_k)$ and we take $F = 0$ as if the economic risk factor were at its medium value. With these samples at hand, it is now easy to match its mean and variance with those of Lognormal(μ_k, σ_k^2), sector by sector. The results are reported in Table 4.

Sector	μ_k	σ_k^2
Transportation	2.5675	0.1885
Electronic Technology	2.4982	0.0721
Health Technology	2.3539	0.0820
Utilities	2.3997	0.0979
Non-Energy Minerals	2.4488	0.1268
Producer Manufacturing	2.4773	0.1131
Health Services	2.6053	0.2017
Energy Minerals	2.3172	0.0647
Consumer Durables	2.6479	0.1882
Communications	2.2346	0.0021

Table 4: Parameters μ_k and σ_k^2 of the distribution of γ_k

4.4 Analysis of the Portfolio Transition

In this final numerical test we compose a fictitious portfolio invested in the 10 selected sectors. We generate $\mathbf{p}_{\text{now}}$ and $\mathbf{p}_{\text{targ}}$ using Gaussians distributions and imagining that $\mathbf{p}_{\text{targ}}$ is a portfolio with a better ESG index mean, to be coherent with the Paris agreement. The portfolios can be visualized in Figure 10b. As a risk metric we use $\rho := \texttt{VaR}_\alpha$ with $\alpha = 99\%$: since the calibrated β's are positive in the selected sectors, it is possible to use Lemma A.1 and then compute $\rho(L_{t+1}(\mathbf{p}_t))$ as a linear function of $\mathbf{p}_t$.

Two criteria are explicitly used to define problem (2.4): Risk of Loss and ESG index. We will analyze the optimal transition according to (2.4) with respect to these two criteria. We compare the results obtained by the MMK solution with a "naive" transition path $\tilde{\mathbf{p}}_t$, interpolating linearly

between $\mathbf{p}_{\text{now}}$ and $\mathbf{p}_{\text{targ}}$:

$$\tilde{\mathbf{p}}_t = \Big(1 - \frac{t+1}{T+1}\Big)\mathbf{p}_{\text{now}} + \Big(\frac{t+1}{T+1}\Big)\mathbf{p}_{\text{targ}}, \quad \text{for } t \in \{-1, ..., T+1\}.$$

In our case of study $T = 2$, meaning that we have three intermediate periods $\{0, 1, 2\}$. The number of points in the grid $\mathbb{G}$ is $N = N_{\mathfrak{e}} \times N_{\mathfrak{g}} = 10 \times 10$, points are equispaced in each direction. We set $\lambda_t = \frac{1}{4}$, for $t \in \{0, ..., 3\}$. The computation time was 15.36 minutes for a processor Intel(R) Core(TM) i7-6600U CPU @ 2.60GHz,2.81 GHz. The number of iterations until convergence is 230 for a tolerance of 10^{-4} with respect to the global primal-dual gap (see Definition 3.1).

Figure 7 shows the time-evolution of the credit risk exposure. We observe that in the "optimal" transition $\mathbf{p}_t$ given by the MMK solution path, the risk keeps always under the risk of the linear path $\tilde{\mathbf{p}}_t$. .

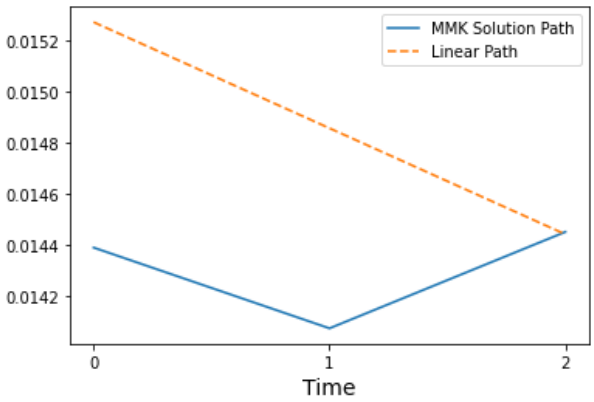


Figure 7: Evolution of the credit risk term $\rho(L_{t+1}(\mathbf{p}_t))$

In terms of objective value, the comparison between the evolution of the linear path and the optimal MMK path shows different trajectories. Figure 8 illustrates the fact that, even if the linear path $\tilde{\mathbf{p}}_t$ has a greater objective value for the MMK problem (2.4), it is not possible to ensure that for each time t in the trajectory the solution $\mathbf{p}_t$ is the best for the pointwise criterion $\mathbf{MK}_t$, reinforcing the importance of considering the whole trajectory and accounting for credit risk.

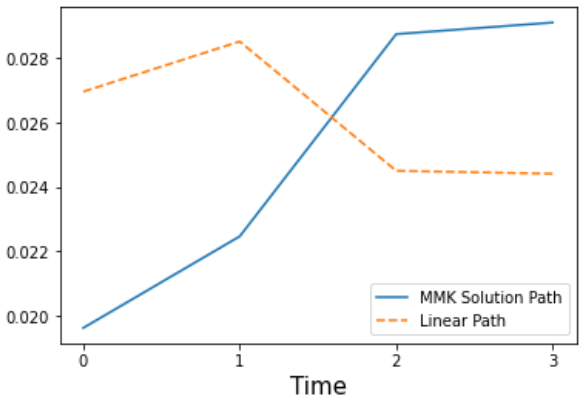


Figure 8: Evolution of the objective value $\mathbf{MK}_t$, for $t \in \{0, 1, 2, 3\}$.

We can verify an important improvment in the global objective value of the whole trajectory: for the linear path, the global objective value is 0.02609, while it is equal to 0.024983 for the MMK path. The difference between values represents approximately 15% of the objective value of the MMK problem. We remember that in the composition of the cost c, ESG score $\mathfrak{e}$ is normalized in percentage scale.

In the point of view of ESG score changes, the expected behavior shows to be very similar to the linear one, see Figure 9. Here the mean ESG is given by $\frac{1}{N}\sum_{o=(\mathfrak{e},\mathfrak{g})\in O}\mathfrak{e}$.

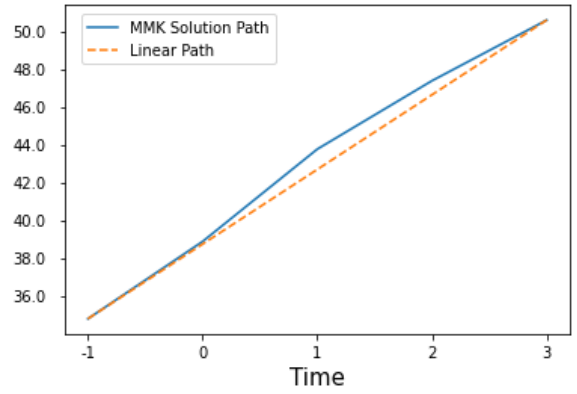
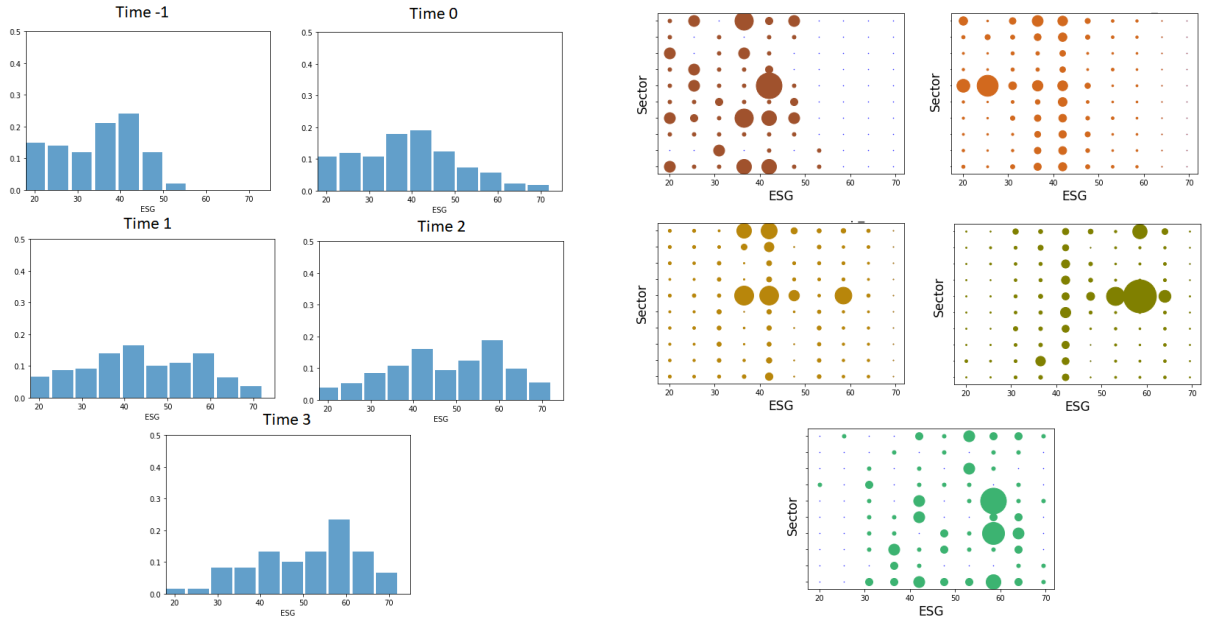


Figure 9: Evolution of ESG score for time $t \in \{-1, 0, 1, 2, 3\}$

The evolution of the ESG distribution on the grid and the distribution path can be graphically visualized in Figures 10a and 10b respectively, confirming the ecological transition of the credit portfolio to a greener situation as expected.



(a) ESG histograms

(b) Graphic evolution of the portfolio transition

Figure 10: On the left, the evolution of ESG histograms for time periods $t \in \{-1, 0, 1, 2, 3\}$. On the right, portfolios $\mathbf{p}_t$ for $t \in \{-1, 0, 1, 2, 3\}$

5 Conclusion

The purpose of this work is to define a quantitative methodology of ecological portfolio transition regarding climate change agreements. The proposed model, Multidate Monge Kantorovich problem (MMK), considers some important parameters for this portfolio transition: ESG of companies in the portfolio (or ecological score), return of investments, and credit risk. Since MMK problem is based on Optimal Transport tools, it aims to quantify not only the adjustment of each portfolio distribution of the trajectory with the selected criteria but also the difficulty in transitioning from one portfolio distribution in time t to the next one in time $t + 1$, keeping an eye on credit risk.

In this work we have demonstrated that the MMK problem is well defined and we have derived some properties of its solutions set. We have proposed a numerical method that converges in available time and that offers important advantages regarding the structure of our problem. In this context, we develop the adaptation of the method to our case of study.

To test our model, we have worked with a variety of financial real data: ESG, credit rating, stock prices and GDP. This data set has allowed us to analyze MMK solutions on realistic data and compare it with a naive form of transition (linear interpolation). Results are encouraging and show important progress with respect to the naive choice: all in all, since the methodology is effective, it can help to get valuable insights on the optimal transition of an institutional actor in finance.

A Background results

A.1 Minimax theorem

The following theorem is a generalization to general ambient spaces of John von Neumann's minimax theorem stated in Euclidean spaces.

Theorem A.1 (Sion's minimax theorem, [39, Corollary 3.3]). *Let $\mathcal{X}$ and $\mathcal{Y}$ be convex spaces one of which is compact. Consider $f : \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}$ such that*

- *$f(x, \cdot)$ is upper semicontinuous and quasi-concave on $\mathcal{Y}$ for any $x \in \mathcal{X}$,*
- *$f(\cdot, y)$ is lower semicontinuous and quasi-convex on $\mathcal{X}$ for any $y \in \mathcal{Y}$,*

then

$$\inf_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} f(x, y) = \sup_{y\in\mathcal{Y}} \inf_{x\in\mathcal{X}} f(x, y).$$

In general, the assumptions of being u.s.c. and l.s.c. cannot be removed nor appreciably weakened, even in finite dimensions. Last, there are some variants for concave-convex like functions, see [39, Theorems 4.1, 4.1', 4.2, 4.2'].

Theorem A.2 (Existence and properties of saddle-points [21, Chapter VII, Proposition 4.1.3, Theorem 4.2.5]). *Let $\mathcal{X}$ and $\mathcal{Y}$ be arbitrary spaces and let $f : \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}$. Let $\mathcal{S}$ be the set of saddle-points:*

$$\mathcal{S} = \{(\bar{x}, \bar{y}) : f(\bar{x}, y) \le f(\bar{x}, \bar{y}) \le f(x, \bar{y}) \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}\}.$$

i) On the set $\mathcal{S}$, the function f is constant;

ii) If $(\bar{x}_1, \bar{y}_1)$ and $(\bar{x}_2, \bar{y}_2)$ are in $\mathcal{S}$, then $(\bar{x}_1, \bar{y}_2)$ and $(\bar{x}_2, \bar{y}_1)$ are also in $\mathcal{S}$;

iii) $\mathcal{S}$ is non-empty if and only if

$$\inf_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} f(x, y) = \sup_{y\in\mathcal{Y}} \inf_{x\in\mathcal{X}} f(x, y).$$

In this case, $\mathcal{S}$ has a product form:

$$\begin{aligned} &\mathcal{S} = \Phi \times \Psi \\ &\quad \textit{where} \quad \Phi := \{\bar{x} \in \mathcal{X} : \sup_{y\in\mathcal{Y}} f(\bar{x}, y) = \inf_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} f(x, y)\}, \\ &\qquad\qquad \Psi := \{\bar{y} \in \mathcal{Y} : \inf_{x\in\mathcal{X}} f(x, \bar{y}) = \sup_{y\in\mathcal{Y}} \inf_{x\in\mathcal{X}} f(x, y)\}. \end{aligned}$$

In particular, under the assumptions of Theorem A.1, the set of saddle-points is not empty.

A.2 Monge-Kantorovich problem

For the sake of self-containedness, we state standard results on Monge-Kantorovich problem, see [42] for a more detailed presentation.

Theorem A.3 (Existence of an optimal coupling for non-negative cost [42, Theorem 4.1]). *Let $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$ be two Polish probability spaces. Let $c : \mathcal{X} \times \mathcal{Y} \mapsto [0, +\infty]$ be a lower semicontinuous cost function. Then there is a coupling of (μ, ν) which minimizes the total cost $\mathbb{E}(c(X, Y))$ among all possible couplings (X, Y), i.e. achieving the bound*

$$\mathbf{MK}(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y)\pi(\mathrm{d}x, \mathrm{d}y)$$

where $\Pi(\mu, \nu)$ is the set of probability measures π on $\mathcal{X} \times \mathcal{Y}$ such that $X_{\#\pi} = \mu$ and $Y_{\#\pi} = \nu$.

Theorem A.4 (Kantorovich duality for non-negative cost [42, Theorem 5.10]). *Let $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$ be two Polish probability spaces. Let $c : \mathcal{X} \times \mathcal{Y} \mapsto [0, +\infty]$ be a lower semicontinuous cost function. Then there is a duality:*

$$\mathbf{MK}(\mu, \nu) = \sup_{(f,g)\in L^1(\mu)\times L^1(\nu): f+g\leq c} \int_{\mathcal{X}} f(x)\mu(\mathrm{d}x) + \int_{\mathcal{Y}} g(y)\nu(\mathrm{d}y).$$

The finite discrete case is notationally simpler, that is in the case where the spaces $\mathcal{X}$ and $\mathcal{Y}$ are identical and made of N elements ($\mathcal{O} := (\mathfrak{o}_1, \cdots, \mathfrak{o}_N)$) and where the cost function is finite on $\mathcal{O}\times\mathcal{O}$. The measures μ and functions f on $\mathcal{X}$ are represented through the scalar values $\mu_j = \mu(\mathfrak{o}_j)$ and $f_j = f(\mathfrak{o}_j)$, and the integral $\int_{\mathcal{X}} f(x)\mu(\mathrm{d}x)$ becomes a scalar product in $\mathbb{R}^N$, of the form $\langle f, \mu\rangle$, where $f = (f_1, \ldots, f_N)^\top$ and $\mu = (\mu_1, \ldots, \mu_N)^\top$. The Kantorovich duality simply rewrites as follows.

Corollary A.1. *In the discrete case $\mathcal{X} = \mathcal{Y} = \mathcal{O} := (\mathfrak{o}_1, \cdots, \mathfrak{o}_N)$, the primal and dual Kantorovich problems have solutions and the dual problem writes*

$$\mathbf{MK}(\mu, \nu) := \max\Big(\langle f, \mu\rangle + \langle g, \nu\rangle\Big),$$

where the set of admissible dual variables (f, g) is s.t. $f_i + g_j \leq c^{i,j} = c(\mathfrak{o}_i, \mathfrak{o}_j)$ for any i, j.

A.3 Admissibility of $\mathtt{VaR}_\alpha$ as risk criterion

Lemma A.1. *In the one-factor ($d = 1$) Gaussian copula model (2.2) with positive correlations ($\beta_1 > 0$), the risk metric $\rho = \mathtt{VaR}_\alpha$ satisfies all the Properties 2.1, for any $\alpha \in (0, 1)$.*

Proof. Let $\mathbf{P}_t$ be a discrete probability measure given by the vector $\mathbf{p}_t$ with components $p_t^i = \mathbf{P}_t(O = \mathfrak{o}_i)$. We claim that the function $\mathbf{p}_t \mapsto \rho(L_{t+1}(\mathbf{p}_t))$ given in (2.9) is a linear function of $\mathbf{p}_t$, hence all Properties 2.1 are fulfilled. To justify the linearity, combine (2.9) and (2.2) to write

$$f(F_{t+1}) = L_{t+1}(\mathbf{p}_t) = \sum_{i=1}^{N} p_t^{(i)} \int_\Theta \Phi\left(\left(\beta_1 F_{t+1}^{(1)} - \gamma\right)/\beta_0\right) q_{t+1,\mathfrak{o}_i}(\mathrm{d}(\beta_0, \beta_1, \gamma)).$$

Under the assumption that the probability measure $q_{t+1,\mathfrak{o}_i}(\cdot)$ puts mass only to points such that $\beta_1 > 0, \beta_0 > 0$, the above function f is increasing, and thus $\mathtt{VaR}_\alpha(f(F_{t+1})) = f(\mathtt{VaR}_\alpha(F_{t+1}))$. The function f being linear in $\mathbf{p}_t$, the result is proved. □

References

[1] M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

[2] K. J. Arrow, L. Hurwicz, and H. Uzawa. Studies in linear and non-linear programming. *Stanford Mathematical Studies in the Social Sciences*, 2:7–50, 1958.

[3] P. Artzner, F. Delbaen, J. Eber, and D. Heath. Coherent measures of risk. *Math. Finance*, 9(3):203–228, 1999.

[4] Basel Committee on Banking Supervision. Climate-related risk drivers and their transmission channels. *Bank for International Settlements, https://www.bis.org/bcbs/publ/d517.htm*, d517, 2021.

[5] S. Battiston, A. Mandel, I. Monasterolo, F. Schütze, and G. Visentin. A climate stress-test of the financial system. *Nature Climate Change*, 7(4):283–288, 2017.

[6] J. Bender, T. A. Bridges, and K. Shah. Reinventing climate investing: building equity portfolios for climate risk mitigation and adaptation. *Journal of Sustainable Finance & Investment*, 9(3):191–213, 2019.

[7] R. I. Bot, E. R. Csetnek, and M. Sedlmayer. An accelerated minimax algorithm for convex-concave saddle point problems with nonsmooth coupling function. *arXiv, 2104.06206*, 2021.

[8] D. Burago, Y. Burago, and S. Ivanov. *A Course in Metric Geometry*, volume 33. 01 2001.

[9] M. Carney. Speech: "breaking the tragedy of the horizon", September, 29th 2015.

[10] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[11] I. Cochran and A. Pauthier. A framework for alignment with the Paris agreement: Why, what and how for financial institutions?, 2019.

[12] European Banking Authority. EBA action plan on sustainable finance. *https://eba.europa.eu/sites/default/documents/files/document_library//EBA%20Action%20plan%20on%20sustainable%20finance.pdf*, 2019.

[13] M. Fang, K. S. Tan, and T. S. Wirjanto. Sustainable portfolio management under climate change. *Journal of Sustainable Finance & Investment*, 9(1):45–67, 2019.

[14] A. Galichon. *Optimal transport methods in economics*. Princeton University Press, 2016.

[15] G. Gidel, T. Jebara, and S. Lacoste-Julien. Frank-Wolfe algorithms for saddle point problems. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, pages 362–371. PMLR, 2017.

[16] K. Giesecke, K. Spiliopoulos, and R. B. Sowers. Default clustering in large portfolios: Typical events. *The Annals of Applied Probability*, 23(1), 2013.

[17] P. Glasserman and J. Li. Importance sampling for portfolio credit risk. *Management science*, 51(11):1643–1656, 2005.

[18] D. Goldsztajn and F. Paganini. Proximal regularization for the saddle point gradient dynamics. *IEEE Transactions on Automatic Control*, 66(9):4385–4392, 2021.

[19] S. Gülten and A. Ruszczyński. Two-stage portfolio optimization with higher-order conditional measures of risk. *Annals of Operations Research*, 229:409–427, 2015.

[20] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms. II*, volume 306 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1993.

[21] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms. I*, volume 305 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1996.

[22] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis*. Springer, 2004.

[23] J. Huang, Y. Jiao, B. Jin, J. Liu, X. Lu, and C. Yang. A unified primal dual active set algorithm for nonconvex sparse recovery. *Statistical Science*, 36(2):215 – 238, 2021.

[24] A. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management*. Princeton series in finance. Princeton University Press, 2005.

[25] P. Mertikopoulos, B. Lecouat, H. Zenati, C.-S. Foo, V. Chandrasekhar, and G. Piliouras. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. In *International Conference on Learning Representations (ICLR)*, 2019.

[26] G. J. Minty. On the monotonicity of the gradient of a convex function. *Pacific Journal of Mathematics*, 14(1):243 – 247, 1964.

[27] G. J. Minty. A theorem on maximal monotonic sets in Hilbert space. *Journal of Mathematical Analysis and Applications*, 11:434–439, 1965.

[28] NGFS. A call for action climate change as a source of financial risk, April 2019.

[29] B. of International Settlements. Calculation of RWA for market risk - Internal models approach. *https://www.bis.org/basel_framework/chapter/MAR/30.htm?inforce=20191215*, 2019.

[30] M. Panik. *Fundamentals of convex analysis*, volume 24. Springer, 1993.

[31] B. Petrova. Multistage portfolio optimization with multivariate dominance constraints. *Computational Management Science*, 16(1-2):17–46, 2019.

[32] G. Peyré and M. Cuturi. Computational optimal transport: with applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.

[33] J. Raynau, P. Tankov, and S. Voisin. Portfolio alignment to a 2°C trajectory: science or art? *Available at SSRN, https://ssrn.com/abstract=3644171*, 2020.

[34] R. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific Journal of Mathematics*, 17(3):497 – 510, 1966.

[35] R. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2(1):21–42, 2000.

[36] T. Roncalli. *Handbook of Financial Risk Management*. CRC Press, 2020.

[37] F. Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. Progress in Nonlinear Differential Equations and Their Applications. Springer International Publishing, 2015.

[38] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming: Modeling and theory*. Society for Industrial and Applied Mathematics, 2014.

[39] M. Sion. On general minimax theorems. *Pacific J. Math.*, 8(1):171–176, 1958.

[40] K. K. Thekumparampil, P. Jain, P. Netrapalli, and S. Oh. Efficient algorithms for smooth minimax optimization. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[41] Q. Tran-Dinh, A. Alacaoglu, and O. Fercoq. An adaptive primal-dual framework for nonsmooth convex minimization. *Math. Prog. Comp.*, 12(3):451–491, 2020.

[42] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften, Fundamental Principles of Mathematical Sciences*. Springer-Verlag, Berlin, 2009.

[43] World Bank Group. World bank group climate change action plan 2021–2025 : Supporting green, resilient, and inclusive development. *Open Knowledge Repository Word Bank Group, https://openknowledge.worldbank.org/handle/10986/35799/CCAP-2021-25.pdf*, 2021.

[44] M. Zhu and T. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. Technical report, UCLA CAM Report, 2008.