



HAL
open science

Convergence rates of Gibbs measures with degenerate minimum

Pierre Bras

► **To cite this version:**

Pierre Bras. Convergence rates of Gibbs measures with degenerate minimum. Bernoulli, 2022, 28 (4), pp.2431-2458. 10.3150/21-BEJ1424 . hal-03183003

HAL Id: hal-03183003

<https://hal.science/hal-03183003>

Submitted on 26 Mar 2021

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

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



Distributed under a Creative Commons Attribution 4.0 International License

Convergence rates of Gibbs measures with degenerate minimum

Pierre Bras*

Abstract

We study convergence rates of Gibbs measures, with density proportional to $e^{-f(x)/t}$, as $t \rightarrow 0$ where $f : \mathbb{R}^d \rightarrow \mathbb{R}$ admits a unique global minimum at x^* . We focus on the case where the Hessian is not definite at x^* . We assume instead that the minimum is strictly polynomial and give a higher order nested expansion of f at x^* , which depends on every coordinate. We give an algorithm yielding such an expansion if the polynomial order of x^* is no more than 8, in connection with Hilbert's 17th problem. However, we prove that the case where the order is 10 or higher is fundamentally different and that further assumptions are needed. We then give the rate of convergence of Gibbs measures using this expansion. Finally we adapt our results to the multiple well case.

1 Introduction

Gibbs measures and their convergence properties are often used in stochastic optimization to minimize a function defined on \mathbb{R}^d . That is, let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a measurable function and let $x^* \in \mathbb{R}^d$ be such that f admits a global minimum at x^* . It is well known [Hwa80] that under standard assumptions, the associated Gibbs measure with density proportional to $e^{-f(x)/t}$ for $t > 0$, converges weakly to the Dirac mass at x^* , δ_{x^*} , when $t \rightarrow 0$. The Langevin equation $dX_s = -\nabla f(X_s)ds + \sigma dW_s$ consists in a gradient descent with Gaussian noise. For $\sigma = \sqrt{2t}$, its invariant measure has a density proportional to $e^{-f(x)/t}$ (see for example [Kha12], Lemma 4.16), so for small t we can expect it to converge to $\operatorname{argmin}(f)$ [Dal14] [Bar20]. The simulated annealing algorithm [LA87] builds a Markov chain from the Gibbs measure where the parameter t converges to zero over the iterations. This idea is also used in [GM90], giving a stochastic gradient descent algorithm where the noise is gradually decreased to zero. Adding a small noise to the gradient descent allows to explore the space and to escape from traps such as local minima and saddle points which appear in non-convex optimization problems [Laz92] [DPG⁺14]. Such methods have been recently brought up to light again with SGLD (Stochastic Gradient Langevin Dynamics) algorithms [WT11] [LCCC15], especially for Machine Learning and calibration of artificial neural networks, which is a high-dimensional non-convex optimization problem.

The rates of convergence of Gibbs measures have been studied in [Hwa80], [Hwa81] and [AH10] under differentiability assumptions on f . It turns out to be of order $t^{1/2}$ as soon as the Hessian matrix $\nabla^2 f(x^*)$ is positive definite. Furthermore, in the multiple well case i.e. if the minimum of f is attained at finitely many points x_1^*, \dots, x_m^* , [Hwa80] proves that the limit distribution is a sum of Dirac masses $\delta_{x_i^*}$ with coefficients proportional to $\det(\nabla^2 f(x_i^*))^{-1/2}$ as soon as all the Hessian matrices are positive definite. If such is not the case, we can conjecture that the limit distribution is concentrated around the x_i^* where the degeneracy is of the highest order.

The aim of this paper is to provide a rate of convergence in this degenerate setting, i.e. when x^* is still a strict global minimum but $\nabla^2 f(x^*)$ is no longer definite, which extends the range of applications of Gibbs measure-based algorithms where positive definiteness is generally assumed. A general framework is given in [AH10], which provides rates of convergence based on dominated convergence. However a strong and rather technical assumption on f is needed and checking it seems, to some

*Sorbonne Université, Laboratoire de Probabilités, Statistique et Modélisation, UMR 8001, case 188, 4 pl. Jussieu, F-75252 Paris Cedex 5, France. E-mail: pierre.bras@sorbonne-universite.fr

extent, more demanding than proving the result. To be more precise, the assumption reads as follows: there exists a function $g : \mathbb{R}^d \rightarrow \mathbb{R}$ with $e^{-g} \in L^1(\mathbb{R}^d)$ and $\alpha_1, \dots, \alpha_d \in (0, +\infty)$ such that

$$\forall h \in \mathbb{R}^d, \quad \frac{1}{t} [f(x^* + (t^{\alpha_1} h_1, \dots, t^{\alpha_d} h_d)) - f(x^*)] \xrightarrow[t \rightarrow 0]{} g(h_1, \dots, h_d). \quad (1)$$

Our objective is to give conditions on f such that (1) is fulfilled and then to elucidate the expression of g depending on f and its derivatives by studying the behaviour of f at x^* in every direction. Doing so we can apply the results from [AH10] yielding the convergence rate of the corresponding Gibbs measures. The orders $\alpha_1, \dots, \alpha_d$ must be chosen carefully and not too big, as the function g needs to depend on every of its variables h_1, \dots, h_d , which is a necessary condition for e^{-g} to be integrable. We also extend our results to the multiple well case.

We generally assume f to be coercive, i.e. $f(x) \rightarrow +\infty$ as $\|x\| \rightarrow +\infty$, \mathcal{C}^{2p} in a neighbourhood of x^* for some $p \in \mathbb{N}$ and we assume that the minimum is polynomial strict, i.e. the function f is bounded below in a neighbourhood of x^* by some non-negative polynomial function, null only at x^* . Thus we can apply a multi-dimensional Taylor expansion to f at x^* , where the successive derivatives of $f : \mathbb{R}^d \rightarrow \mathbb{R}$ are seen as symmetric tensors of \mathbb{R}^d . The idea is then to consider the successive subspaces where the derivatives of f are null up to some order ; using that the Taylor expansion of $f(x^* + h) - f(x^*)$ is non-negative, some cross derivative terms are null. However a difficulty arises at orders 6 and higher, as the set where the derivatives of f are null up to some order is no longer a vector subspace in general. This difficulty is linked with Hilbert's 17th problem [Hil88], stating that a non-negative multivariate polynomial cannot be written as the sum of squares of polynomials in general. We thus need to change the definition of the subspaces we consider. Following this, we give a recursive algorithm yielding an adapted decomposition of \mathbb{R}^d into vector subspaces and a function g satisfying (1) up to a change of basis, giving a canonical higher order nested decomposition of f at x^* in degenerate cases. An interesting fact is that the case where the polynomial order of x^* is 10 or higher fundamentally differs from those of orders 2, 4, 6 and 8, owing to the presence of even cross terms which may be not null. The algorithm we provide works at the orders 10 or higher only under the assumption that all such even cross terms are null. In general, it is more difficult to get a general expression of g for the orders 10 and higher. We then apply our results to [AH10], where we give conditions such that the hypotheses of [AH10], especially (1), are satisfied so as to infer rates of convergence of Gibbs measures in the degenerate case where $\nabla^2 f(x^*)$ is not necessarily positive definite. The function g given by our algorithm is a non-negative polynomial function and non-constant in any of its variables, however it needs to be assumed to be coercive to be applied to [AH10]. We study the case where g is not coercive and give a method to deal with simple generic non-coercive cases, where our algorithm seems to be a first step to a more general procedure. However, we do not give a general method in this case.

Our results are applied to Gibbs measures but they can also be applied to more general contexts, as we give a canonical higher order nested expansion of f at a minimum, in the case where some derivatives are degenerate.

For general properties of symmetric tensors we refer to [CGLM08]. In the framework of stochastic approximation, [FP99] Section 3.1 introduced the notion of strict polynomial local extremum and investigated their properties as higher order "noisy traps".

The paper is organized as follows. In Section 3, we recall convergence properties of Gibbs measures and revisit the main theorem from [AH10]. This theorem requires, as an hypothesis, to find an expansion of f at its global minimum ; we properly state this problem in Section 3.2 under the assumption of strict polynomial minimum. In Section 3.3, we state our main result for both single well and multiple well cases, as well as our algorithm. In Section 4, we detail the expansion of f at its minimum for each order and provide the proof. We give the general expression of the canonical higher order nested expansion at any order in Section 4.1, where we distinguish the orders 10 and higher from the lower ones. We then provide the proof for each order 2, 4, 6 and 8 in Sections 4.3, 4.4, 4.6 and 4.7 respectively. We need to prove that, with the exponents $\alpha_1, \dots, \alpha_d$ we specify, the convergence

in (1) holds ; we do so by proving that, using the non-negativity of the Taylor expansion, some cross derivative terms are zero. Because of Hilbert's 17th problem, we need to distinguish the orders 6 and 8 from the orders 2 and 4, as emphasized in Section 4.5. For orders 10 and higher, such terms are not necessarily zero and must then be assumed to be zero. In Section 4.8, we give a counter-example if this assumption is not satisfied before proving the result. In Section 4.9, we prove that for every order the resulting function g is constant in none of its variables and that the convergence in (1) is uniform on every compact set. In Section 4.10, we study the case where the function g is not coercive and give a method to deal with the simple generic case. In Section 5, we prove our main theorems stated in Section 3.3 using the expansion of f established in Section 4. Finally, in Section 6, we deal with a "flat" example where all the derivatives in the local minimum are zero and where we cannot apply our main theorems.

2 Definitions and notations

We give a brief list of notations that are used throughout the paper.

We endow \mathbb{R}^d with its canonical basis (e_1, \dots, e_d) and the Euclidean norm denoted by $\|\cdot\|$. For $x \in \mathbb{R}^d$ and $r > 0$ we denote by $\mathcal{B}(x, r)$ the Euclidean ball of \mathbb{R}^d of center x and radius r . For E a vector subspace of \mathbb{R}^d , we denote by $p_E : \mathbb{R}^d \rightarrow E$ the orthogonal projection on E . For a decomposition of \mathbb{R}^d into orthogonal subspaces, $\mathbb{R}^d = E_1 \oplus \dots \oplus E_p$, we say that an orthogonal transformation $B \in \mathcal{O}_d(\mathbb{R})$ is adapted to this decomposition if for all $j \in \{1, \dots, p\}$,

$$\forall i \in \{\dim(E_1) + \dots + \dim(E_{j-1}) + 1, \dots, \dim(E_1) + \dots + \dim(E_j)\}, B \cdot e_i \in E_j.$$

For $a, b \in \mathbb{R}^d$, we denote by $a * b$ the element-wise product, i.e.

$$\forall i \in \{1, \dots, d\}, (a * b)_i = a_i b_i.$$

For v^1, \dots, v^k vectors in \mathbb{R}^d and T a tensor of order k of \mathbb{R}^d , we denote the tensor product

$$T \cdot (v^1 \otimes \dots \otimes v^k) = \sum_{i_1, \dots, i_k \in \{1, \dots, d\}} T_{i_1 \dots i_k} v_{i_1}^1 \dots v_{i_k}^k.$$

More generally, if $j \leq k$ and v^1, \dots, v^j are j vectors in \mathbb{R}^d , then $T \cdot (v^1 \otimes \dots \otimes v^j)$ is a tensor of order $k - j$ such that:

$$T \cdot (v^1 \otimes \dots \otimes v^j)_{i_{j+1} \dots i_k} = \sum_{i_1, \dots, i_j \in \{1, \dots, d\}} T_{i_1 \dots i_k} v_{i_1}^1 \dots v_{i_j}^j.$$

For $h \in \mathbb{R}^d$, $h^{\otimes k}$ denotes the tensor of order k such that

$$h^{\otimes k} = (h_{i_1} \dots h_{i_k})_{i_1, \dots, i_k \in \{1, \dots, d\}}.$$

For a function $f \in \mathcal{C}^p(\mathbb{R}^d, \mathbb{R})$, we denote $\nabla^k f(x)$ the differential of order $k \leq p$ of f at x , as $\nabla^k f(x)$ is the tensor of order k defined by:

$$\nabla^k f(x) = \left(\frac{\partial^k f(x)}{\partial x_{i_1} \dots \partial x_{i_k}} \right)_{i_1, i_2, \dots, i_k \in \{1, \dots, d\}}.$$

By Schwarz's theorem, this tensor is symmetric, i.e. for all permutation $\sigma \in \mathfrak{S}_k$,

$$\frac{\partial^k f(x)}{\partial x_{i_{\sigma(1)}} \dots \partial x_{i_{\sigma(k)}}} = \frac{\partial^k f(x)}{\partial x_{i_1} \dots \partial x_{i_k}}.$$

We recall the Taylor-Young formula in any dimension, and the Newton multinomial formula.

Theorem 1 (Taylor-Young formula). *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be \mathcal{C}^p and let $x \in \mathbb{R}^d$. Then:*

$$f(x+h) \underset{h \rightarrow 0}{=} \sum_{k=0}^p \frac{1}{k!} \nabla^k f(x) \cdot h^{\otimes k} + \|h\|^p o(1).$$

We denote by $\binom{k}{i_1, \dots, i_p}$ the p -nomial coefficient, defined as:

$$\binom{k}{i_1, \dots, i_p} = \frac{k!}{i_1! \dots i_p!}.$$

Theorem 2 (Newton multinomial formula). *Let $h_1, \dots, h_p \in \mathbb{R}^d$, then*

$$(h_1 + h_2 + \dots + h_p)^{\otimes k} = \sum_{\substack{i_1, \dots, i_p \in \{0, \dots, k\} \\ i_1 + \dots + i_p = k}} \binom{k}{i_1, \dots, i_p} h_1^{\otimes i_1} \otimes \dots \otimes h_p^{\otimes i_p}. \quad (2)$$

For T a tensor of order k , we say that T is non-negative (resp. positive) if

$$\forall h \in \mathbb{R}^d, T \cdot h^{\otimes k} \geq 0 \text{ (resp. } T \cdot h^{\otimes k} > 0). \quad (3)$$

We denote $L^1(\mathbb{R}^d)$ the set of measurable functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$ that are integrable with respect to the Lebesgue measure on \mathbb{R}^d . We denote by λ_d the Lebesgue measure on \mathbb{R}^d . For $f : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $e^{-f} \in L^1(\mathbb{R}^d)$, we define for $t > 0$, $C_t := (\int_{\mathbb{R}^d} e^{-f/t})^{-1}$ and π_t the Gibbs measure

$$\pi_t(x) dx := C_t e^{-f(x)/t} dx.$$

For a family of random variables $(Y_t)_{t \in (0,1]}$ and Y a random variable, we write $Y_t \xrightarrow[t \rightarrow 0]{\mathcal{L}} Y$ meaning that (Y_t) weakly converges to Y .

We give the following definition of a strict polynomial local minimum of f :

Definition 1. *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be \mathcal{C}^{2p} for $p \in \mathbb{N}$ and let x^* be a local minimum of f . We say that f has a strict polynomial local minimum at x^* of order $2p$ if p is the smallest integer such that:*

$$\exists r > 0, \forall h \in \mathcal{B}(x^*, r) \setminus \{0\}, \sum_{k=2}^{2p} \frac{1}{k!} \nabla^k f(x^*) \cdot h^{\otimes k} > 0. \quad (4)$$

Remarks :

1. A local minimum x^* of f is not necessarily strictly polynomial, for example, $f : x \mapsto e^{-\|x\|^{-2}}$ and $x^* = 0$.
2. If x^* is polynomial strict, then the order is necessarily even, because if x^* is not polynomial strict of order $2l$ for some $l \in \mathbb{N}$, then we have $h_n \rightarrow 0$ such that the Taylor expansion in h_n up to order $2l$ is zero ; by the minimum condition, the Taylor expansion in h_n up to order $2l+1$ must be non-negative, so we also have $\nabla^{2l+1} f(x^*) \cdot h_n^{\otimes 2l+1} = 0$.

For $f : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $\min_{\mathbb{R}^d}(f)$ exists, we denote by $\operatorname{argmin}(f)$ the arguments of the minima of f , i.e.

$$\operatorname{argmin}(f) = \left\{ x \in \mathbb{R}^d : f(x) = \min_{\mathbb{R}^d}(f) \right\}.$$

Without ambiguity, we write "minimum" or "local minimum" to designate $f(x^*)$ as well as x^* . Finally, we define, for $x^* \in \mathbb{R}^d$ and $p \in \mathbb{N}$:

$$\mathcal{A}_p(x^*) := \left\{ f \in \mathcal{C}^{2p}(\mathbb{R}^d, \mathbb{R}) : f \text{ admits a local minimum at } x^* \right\}.$$

$$\mathcal{A}_p^*(x^*) := \left\{ f \in \mathcal{C}^{2p}(\mathbb{R}^d, \mathbb{R}) : f \text{ admits a strict polynomial local minimum at } x^* \text{ of order } 2p \right\}.$$

3 Convergence of Gibbs measures

3.1 Properties of Gibbs measures

Let us consider a Borel function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ with $e^{-f} \in L^1(\mathbb{R}^d)$. We study the asymptotic behaviour of the probability measures of density for $t \in (0, \infty)$:

$$\pi_t(x)dx = C_t e^{-\frac{f(x)}{t}} dx$$

when $t \rightarrow 0$. When t is small, the measure π_t tends to the set $\operatorname{argmin}(f)$. The following proposition makes this statement precise.

Proposition 1. *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a Borel function such that*

$$f^* := \operatorname{ess\,inf}(f) = \inf\{y : \lambda_d\{f \leq y\} > 0\} > -\infty,$$

and $e^{-f} \in L^1(\mathbb{R}^d)$. Then

$$\forall \varepsilon > 0, \pi_t(\{f \geq f^* + \varepsilon\}) \xrightarrow[t \rightarrow 0]{} 0.$$

Proof. As $f^* > -\infty$, we may assume without loss of generality that $f^* = 0$ by replacing f by $f - f^*$. Let $\varepsilon > 0$. It follows from the assumptions that $f \geq 0$ λ_d -a.e. and $\lambda_d\{f \leq \varepsilon\} > 0$ for every $\varepsilon > 0$. As $e^{-f} \in L^1(\mathbb{R}^d)$, we have

$$\lambda_d\{f \leq \varepsilon/3\} \leq e^{\varepsilon/3} \int_{\mathbb{R}^d} e^{-f} d\lambda_d < +\infty.$$

Moreover by dominated convergence, it is clear that

$$C_t^{-1} \downarrow \lambda_d\{f = 0\} < +\infty.$$

We have

$$C_t \leq \left(\int_{f \leq \varepsilon/3} e^{-\frac{f(x)}{t}} dx \right)^{-1} \leq \left(e^{-\frac{\varepsilon}{3t}} \underbrace{\lambda_d\{f \leq \frac{\varepsilon}{3}\}}_{>0} \right)^{-1}.$$

Then

$$\pi_t\{f \geq \varepsilon\} = C_t \int_{f \geq \varepsilon} e^{-\frac{f(x)}{t}} dx \leq \frac{e^{\varepsilon/3t} \int_{f \geq \varepsilon} e^{-f(x)/t} dx}{\lambda_d\{f \leq \frac{\varepsilon}{3}\}} \leq \frac{e^{-\varepsilon/3t} C_{3t}^{-1}}{\lambda_d\{f \leq \frac{\varepsilon}{3}\}} \xrightarrow[t \rightarrow 0]{} 0,$$

because if $f(x) \geq \varepsilon$, then $e^{-\frac{f(x)}{t}} \leq e^{-\frac{2\varepsilon}{3t}} e^{-\frac{f(x)}{3t}}$, and where we used that $C_{3t}^{-1} \leq C_1^{-1}$ if $t \leq 1/3$ \square

Now, let us assume that $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is continuous, $e^{-f} \in L^1(\mathbb{R}^d)$ and f admits a unique global minimum at x^* so that $\operatorname{argmin}(f) = \{x^*\}$. In [AH10] is proved the weak convergence of π_t to δ_{x^*} and a rate of convergence depending on the behaviour of $f(x^* + h) - f(x^*)$ for small enough h . Let us recall this result in detail ; we may assume without loss of generality that $x^* = 0$ and $f(x^*) = 0$.

Theorem 3 (Athreya-Hwang, 2010). *Let $f : \mathbb{R}^d \rightarrow [0, \infty)$ be a Borel function such that :*

1. $e^{-f} \in L^1(\mathbb{R}^d)$.
2. For all $\delta > 0$, $\inf\{f(x), \|x\| > \delta\} > 0$.
3. There exist $\alpha_1, \dots, \alpha_d > 0$ such that for all $(h_1, \dots, h_d) \in \mathbb{R}^d$,

$$\frac{1}{t} f(t^{\alpha_1} h_1, \dots, t^{\alpha_d} h_d) \xrightarrow[t \rightarrow 0]{} g(h_1, \dots, h_d) \in \mathbb{R}.$$

$$4. \int_{\mathbb{R}^d} \sup_{0 < t < 1} e^{-\frac{f(t^{\alpha_1} h_1, \dots, t^{\alpha_d} h_d)}{t}} dh_1 \dots dh_d < \infty.$$

For $0 < t < 1$, let X_t be a random vector with distribution π_t . Then $e^{-g} \in L^1(\mathbb{R}^d)$ and

$$\left(\frac{(X_t)_1}{t^{\alpha_1}}, \dots, \frac{(X_t)_d}{t^{\alpha_d}} \right) \xrightarrow{\mathcal{L}} X \quad \text{as } t \rightarrow 0 \quad (5)$$

where the distribution of X has a density proportional to $e^{-g(x_1, \dots, x_d)}$.

Remark: Hypothesis 2. is verified as soon as f is continuous, coercive (i.e. $f(x) \rightarrow +\infty$ when $\|x\| \rightarrow +\infty$) and that $\operatorname{argmin}(f) = \{0\}$.

To study the rate of convergence of the measure π_t when $t \rightarrow 0$ using Theorem 3, we need to identify $\alpha_1, \dots, \alpha_d$ and g such that the condition (5) holds, up to a possible change of basis. Since x^* is a local minimum, the Hessian $\nabla^2 f(x^*)$ is positive semi-definite. Moreover, if $\nabla^2 f(x^*)$ is positive definite, then choosing $\alpha_1 = \dots = \alpha_d = \frac{1}{2}$, we have:

$$\frac{1}{t} f(t^{1/2} h) \xrightarrow{t \rightarrow 0} \frac{1}{2} h^T \cdot \nabla^2 f(x^*) \cdot h := g(x).$$

And using an orthogonal change of variable:

$$\int_{\mathbb{R}^d} e^{-g(x)} dx = \int_{\mathbb{R}^d} e^{-\frac{1}{2} \sum_{i=1}^d \beta_i y_i^2} dy_1 \dots dy_d < \infty,$$

where the eigenvalues β_i are positive. However, if $\nabla^2 f(x^*)$ is not positive definite, then some of the β_i are zero and the integral does not converge.

3.2 Statement of the problem

We still consider the function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ and assume that $f \in \mathcal{A}_p^*(x^*)$ for some $x^* \in \mathbb{R}^d$ and some integer $p \geq 1$. Then our objective is to find $\alpha_1 \geq \dots \geq \alpha_d \in (0, +\infty)$ and an orthogonal transformation $B \in \mathcal{O}_d(\mathbb{R})$ such that:

$$\forall h \in \mathbb{R}^d, \quad \frac{1}{t} [f(x^* + B \cdot (t^\alpha * h)) - f(x^*)] \xrightarrow{t \rightarrow 0} g(h_1, \dots, h_d), \quad (6)$$

where t^α denotes the vector $(t^{\alpha_1}, \dots, t^{\alpha_d})$ and where $g : \mathbb{R}^d \rightarrow \mathbb{R}$ is a measurable function which is not constant in any h_1, \dots, h_d , i.e. for all $i \in \{1, \dots, d\}$, there exist $h_1, \dots, h_{i-1}, h_{i+1}, \dots, h_d \in \mathbb{R}^d$ such that

$$h_i \mapsto g(h_1, \dots, h_d) \text{ is not constant.} \quad (7)$$

Then we say that $\alpha_1, \dots, \alpha_d, B$ and g are a solution of the problem (6). The hypothesis that g is not constant in any of its variables is important ; otherwise, we could simply take $\alpha_1 = \dots = \alpha_d = 1$ and obtain, by the first order condition:

$$\frac{1}{t} [f(x^* + t(h_1, \dots, h_d)) - f(x^*)] \xrightarrow{t \rightarrow 0} 0.$$

3.3 Main results : rate of convergence of Gibbs measures

Theorem 4 (Single well case). *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be \mathcal{C}^{2p} with $p \in \mathbb{N}$ and such that:*

1. f is coercive, i.e. $f(x) \rightarrow +\infty$ when $\|x\| \rightarrow +\infty$.
2. $\operatorname{argmin}(f) = 0$.
3. $f \in \mathcal{A}_p^*(0)$ and $f(0) = 0$.

4. $e^{-f} \in L^1(\mathbb{R}^d)$.

Let $(E_k)_k$, $(\alpha_i)_i$, B and g to be defined as in Algorithm 1 stated right after, so that for all $h \in \mathbb{R}^d$,

$$\frac{1}{t} [f(x^* + B \cdot (t^\alpha * h)) - f(x^*)] \xrightarrow[t \rightarrow 0]{} g(h),$$

and where g is not constant in any of its variables. Moreover, assume that g is coercive and the following technical hypothesis if $p \geq 5$:

$$\begin{aligned} \forall h \in \mathbb{R}^d, \forall (i_1, \dots, i_p) \in \{0, 2, \dots, 2p\}^p, \\ \frac{i_1}{2} + \dots + \frac{i_p}{2p} < 1 \implies \nabla^{i_1 + \dots + i_p} f(x^*) \cdot p_{E_1}(h)^{\otimes i_1} \otimes \dots \otimes p_{E_p}(h)^{\otimes i_p} = 0. \end{aligned} \quad (8)$$

Then the conclusion of Theorem 3 holds, with:

$$\left(\frac{1}{t^{\alpha_1}}, \dots, \frac{1}{t^{\alpha_d}} \right) * (B^{-1} \cdot X_t) \xrightarrow{\mathcal{L}} X \text{ as } t \rightarrow 0,$$

where X has a density proportional to $e^{-g(x)}$.

Algorithm 1. Let $f \in \mathcal{A}_p^*(x^*)$ for $p \in \mathbb{N}$.

1. Define $(F_k)_{0 \leq k \leq p-1}$ recursively as:

$$\begin{cases} F_0 = \mathbb{R}^d \\ F_k = \{h \in F_{k-1} : \forall h' \in F_{k-1}, \nabla^{2k} f(x^*) \cdot h \otimes h'^{\otimes 2k-1} = 0\}. \end{cases}$$

2. For $1 \leq k \leq p-1$, define the subspace E_k as the orthogonal complement of F_k in F_{k-1} . By abuse of notation, define $E_p := F_{p-1}$.

3. Define $B \in \mathcal{O}_d(\mathbb{R})$ as an orthogonal transformation adapted to the decomposition

$$\mathbb{R}^d = E_1 \oplus \dots \oplus E_p.$$

4. Define for $1 \leq i \leq d$,

$$\alpha_i := \frac{1}{2^j} \text{ for } i \in \{\dim(E_1) + \dots + \dim(E_{j-1}) + 1, \dots, \dim(E_1) + \dots + \dim(E_j)\}. \quad (9)$$

5. Define $g : \mathbb{R}^d \rightarrow \mathbb{R}$ as

$$g(h) = \sum_{k=2}^{2p} \frac{1}{k!} \sum_{\substack{i_1, \dots, i_p \in \{0, \dots, k\} \\ i_1 + \dots + i_p = k \\ \frac{i_1}{2} + \dots + \frac{i_p}{2p} = 1}} \binom{k}{i_1, \dots, i_p} \nabla^k f(x^*) \cdot p_{E_1}(B \cdot h)^{\otimes i_1} \otimes \dots \otimes p_{E_p}(B \cdot h)^{\otimes i_p}. \quad (10)$$

Remarks :

1. The function g is not unique, as we can choose any base B adapted to the decomposition $\mathbb{R}^d = E_1 \oplus \dots \oplus E_p$.
2. The case $p \geq 5$ is fundamentally different from the case $p \leq 4$, since Algorithm 1 may fail to provide such $(E_k)_k$, $(\alpha_i)_i$, B and g if the technical hypothesis (8) is not fulfilled, as explained in Section 4.8. This yields fewer results for the case $p \geq 5$.
3. For $p \in \{1, 2, 3, 4\}$, the detail the expression of g in (17), (18), (20) and (23) respectively.

4. The function g has the following general properties : g is a non-negative polynomial of order $2p$; $g(0) = 0$ and $\nabla g(0) = 0$.
5. The condition on g to be coercive may seem not natural. We give more details about the case where g is not coercive in Section 4.10 and give a way to deal with the simple generic case of non-coercivity. However dealing with the general case where g is not coercive goes beyond the scope of our work.
6. The hypothesis that g is coercive is a necessary condition for $e^{-g} \in L^1(\mathbb{R}^d)$. We actually prove in Proposition 3 that it is a sufficient condition.

Still following [AH10], we study the multiple well case, i.e. the global minimum is attained in a finite number of points in \mathbb{R}^d , say $\{x_1^*, \dots, x_m^*\}$ for some $m \in \mathbb{N}$. In this case, the limiting measure of π_t will have its support in $\{x_1^*, \dots, x_m^*\}$, with different weights.

Theorem 5 (Athreya-Hwang, 2010). *Let $f : \mathbb{R}^d \rightarrow [0, \infty)$ measurable such that:*

1. $e^{-f} \in L^1(\mathbb{R}^d)$.
2. For all $\delta > 0$, $\inf\{f(x), \|x - x_i^*\| > \delta, 1 \leq i \leq m\} > 0$.
3. There exist $(\alpha_{ij})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq d}}$ such that for all i, j , $\alpha_{ij} \geq 0$ and for all i :

$$\frac{1}{t} f(x_i^* + (t^{\alpha_{i1}} h_1, \dots, t^{\alpha_{id}} h_d)) \xrightarrow[t \rightarrow 0]{} g_i(h_1, \dots, h_d) \in [0, \infty).$$

4. For all $i \in \{1, \dots, m\}$,

$$\int_{\mathbb{R}^d} \sup_{0 < t < 1} e^{-\frac{f(x_i^* + (t^{\alpha_{i1}} h_1, \dots, t^{\alpha_{id}} h_d))}{t}} dh_1 \dots dh_d < \infty.$$

Then, let $\alpha := \min_{1 \leq i \leq m} \left\{ \sum_{j=1}^d \alpha_{ij} \right\}$ and let $J := \left\{ i \in \{1, \dots, m\} : \sum_{j=1}^d \alpha_{ij} = \alpha \right\}$. For $0 < t < 1$, let X_t be a random vector with distribution π_t . Then:

$$X_t \xrightarrow[t \rightarrow 0]{\mathcal{L}} \frac{1}{\sum_{j \in J} \int_{\mathbb{R}^d} e^{-g_j(x)} dx} \sum_{i \in J} \int_{\mathbb{R}^d} e^{-g_i(x)} dx \cdot \delta_{x_i^*}.$$

Theorem 6 (Multiple well case). *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be \mathcal{C}^{2p} for $p \in \mathbb{N}$ and such that:*

1. f is coercive i.e. $f(x) \rightarrow +\infty$ when $\|x\| \rightarrow +\infty$.
2. $\operatorname{argmin}(f) = \{x_1^*, \dots, x_m^*\}$ and for all i , $f(x_i^*) = 0$.
3. For all $i \in \{1, \dots, m\}$, $f \in \mathcal{A}_{p_i}^*(x_i^*)$ for some $p_i \leq p$.
4. $e^{-f} \in L^1(\mathbb{R}^d)$.

Then, for every $i \in \{1, \dots, m\}$, we consider $(E_{ik})_k$, $(\alpha_{ij})_j$, B_i and g_i as defined in Algorithm 1, where we consider f to be in $\mathcal{A}_{p_i}^*(x_i^*)$, so that for every $h \in \mathbb{R}^d$:

$$\frac{1}{t} f(x_i^* + B_i \cdot (t^{\alpha_i} * h)) \xrightarrow[t \rightarrow 0]{} g_i(h_1, \dots, h_d) \in [0, \infty),$$

where t^{α_i} is the vector $(t^{\alpha_{i1}}, \dots, t^{\alpha_{id}})$ and where g_i is not constant in any of its variables. Furthermore, we assume that for all i , g_i is coercive and the following technical hypothesis for every i such that $p_i \geq 5$:

$$\begin{aligned} \forall h \in \mathbb{R}^d, \forall (i_1, \dots, i_{p_i}) \in \{0, 2, \dots, 2p_i\}^{p_i}, \\ \frac{i_1}{2} + \dots + \frac{i_{p_i}}{2p} < 1 \implies \nabla^{i_1 + \dots + i_{p_i}} f(x_i^*) \cdot p_{E_{i_1}}(h)^{\otimes i_1} \otimes \dots \otimes p_{E_{i_{p_i}}}(h)^{\otimes i_{p_i}} = 0. \end{aligned}$$

Let $\alpha := \min_{1 \leq i \leq m} \left\{ \sum_{j=1}^d \alpha_{ij} \right\}$ and let $J := \left\{ i \in \{1, \dots, m\} : \sum_{j=1}^d \alpha_{ij} = \alpha \right\}$. Then:

$$X_t \xrightarrow{t \rightarrow 0} \frac{1}{\sum_{j \in J} \int_{\mathbb{R}^d} e^{-g_j(x)} dx} \sum_{i \in J} \int_{\mathbb{R}^d} e^{-g_i(x)} dx \cdot \delta_{x_i^*}.$$

Moreover, let $\delta > 0$ be small enough so that the balls $\mathcal{B}(x_i^*, \delta)$ are disjoint, and define the random vector X_{it} to have the law of X_t conditionally to the event $\|X_t - x_i^*\| < \delta$. Then:

$$\left(\frac{1}{t^{\alpha_{i1}}}, \dots, \frac{1}{t^{\alpha_{id}}} \right) * (B_i^{-1} \cdot X_{it}) \xrightarrow{\mathcal{L}} X_i \quad \text{as } t \rightarrow 0,$$

where X_i has a density proportional to $e^{-g_i(x)}$.

4 Expansion of f at a local minimum with degenerate derivatives

In this section, we aim at answering to the problem stated in (6) in order to devise conditions to apply Theorem 3. This problem can also be considered in a more general setting, independently of the study of the convergence of Gibbs measures. It provides a non degenerate higher order nested expansion of f at a local minimum when some of the derivatives of f are degenerate. Note here that we only need x^* to be a local minimum instead of a global minimum, since we only give local properties.

For $k \leq p$, we define the tensor of order k , $T_k := \nabla^k f(x^*)$.

4.1 Expansion of f for any order p

In this section, we state our result in a synthetic form. The proofs of the cases $p = 1, 2, 3, 4$ are individually detailed in Sections 4.3, 4.4, 4.6 and 4.7 respectively.

Theorem 7. *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be \mathcal{C}^{2p} for some $p \in \mathbb{N}$ and assume that $f \in \mathcal{A}_p^*(x^*)$ for some $x^* \in \mathbb{R}^d$.*

1. *If $p \in \{1, 2, 3, 4\}$, then there exists orthogonal subspaces of \mathbb{R}^d , E_1, \dots, E_p such that*

$$\mathbb{R}^d = E_1 \oplus \dots \oplus E_p,$$

and satisfying for every $h \in \mathbb{R}^d$:

$$\frac{1}{t} \left[f \left(x^* + t^{1/2} p_{E_1}(h) + \dots + t^{1/(2p)} p_{E_p}(h) \right) - f(x^*) \right] \quad (11)$$

$$\xrightarrow{t \rightarrow 0} \sum_{k=2}^{2p} \frac{1}{k!} \sum_{\substack{i_1, \dots, i_p \in \{0, \dots, k\} \\ i_1 + \dots + i_p = k \\ \frac{i_1}{2} + \dots + \frac{i_p}{2p} = 1}} \binom{k}{i_1, \dots, i_p} T_k \cdot p_{E_1}(h)^{\otimes i_1} \otimes \dots \otimes p_{E_p}(h)^{\otimes i_p}. \quad (12)$$

The convergence is uniform with respect to h on every compact set. Moreover, let $B \in \mathcal{O}_d(\mathbb{R})$ be an orthogonal transformation adapted to the decomposition $E_1 \oplus \dots \oplus E_p$, then

$$\frac{1}{t} \left[f \left(x^* + B \cdot (t^\alpha * h) \right) - f(x^*) \right] \xrightarrow{t \rightarrow 0} g(h), \quad (13)$$

where

$$g(h) = \sum_{k=2}^{2p} \frac{1}{k!} \sum_{\substack{i_1, \dots, i_p \in \{0, \dots, k\} \\ i_1 + \dots + i_p = k \\ \frac{i_1}{2} + \dots + \frac{i_p}{2p} = 1}} \binom{k}{i_1, \dots, i_p} T_k \cdot p_{E_1}(B \cdot h)^{\otimes i_1} \otimes \dots \otimes p_{E_p}(B \cdot h)^{\otimes i_p} \quad (14)$$

is not constant in any of its variables h_1, \dots, h_d and

$$\alpha_i := \frac{1}{2j} \quad \text{for } i \in \{\dim(E_1) + \dots + \dim(E_{j-1}) + 1, \dots, \dim(E_1) + \dots + \dim(E_j)\}. \quad (15)$$

2. If $p \geq 5$ and if there exist orthogonal subspaces of \mathbb{R}^d , E_1, \dots, E_p such that

$$\mathbb{R}^d = E_1 \oplus \dots \oplus E_p$$

and satisfying the following additional assumption

$$\begin{aligned} \forall h \in \mathbb{R}^d, \forall (i_1, \dots, i_p) \in \{0, 2, \dots, 2p\}^p, \\ \frac{i_1}{2} + \dots + \frac{i_p}{2} < 1 \implies T_{i_1 + \dots + i_p} \cdot p_{E_1}(h)^{\otimes i_1} \otimes \dots \otimes p_{E_p}(h)^{\otimes i_p} = 0, \end{aligned} \quad (16)$$

then (12) stills holds true, as well as the uniform convergence on every compact set. Moreover, if $B \in \mathcal{O}_d(\mathbb{R})$ is an orthogonal transformation adapted to the previous decomposition, then (13) still hold true. However, depending on the function f , such subspaces do not necessarily exist.

Remarks:

1. The limit (12) can be rewritten as:

$$\sum_{k=2}^{2p} \sum_{\substack{i_1, \dots, i_p \in \{0, \dots, k\} \\ i_1 + \dots + i_p = k \\ \frac{i_1}{2} + \dots + \frac{i_p}{2} = 1}} T_k \cdot \frac{p_{E_1}(h)^{\otimes i_1}}{i_1!} \otimes \dots \otimes \frac{p_{E_p}(h)^{\otimes i_p}}{i_p!}.$$

2. For $p \in \{1, 2, 3, 4\}$, we explicitly give the expression of the sum (12) and the p -tuples (i_1, \dots, i_p) such that $\frac{i_1}{2} + \dots + \frac{i_p}{2} = 1$, in (17), (18), (20) and (23) respectively.
3. For $p \in \{1, 2, 3, 4\}$, we give in Algorithm 1 an explicit construction of the orthogonal subspaces E_1, \dots, E_p as complementaries of annulation sets of some derivatives of f .
4. The case $p \geq 5$ is fundamentally different from the case $p \in \{1, 2, 3, 4\}$. The strategy of proof developed for $p \in \{1, 2, 3, 4\}$ fails if the assumption (16) is not satisfied. In 4.8 a counter-example is detailed. The case $p \geq 5$ yields fewer results than for $p \leq 4$, as the assumption (16) is strong.
5. For $p \geq 5$, such subspaces E_1, \dots, E_p may also be obtained from Algorithm 1, however (16) is not necessarily true in this case.

The proof of Theorem 7 is given first individually for each $p \in \{1, 2, 3, 4\}$, in Sections 4.3, 4.4, 4.6, 4.7 respectively. The proof for $p \geq 5$ is given in Section 4.8. The proof of the uniform convergence and of the fact that g is not constant is given in Section 4.9.

4.2 Review of the one dimensional case

We review the case $d = 1$, as it guides us for the proof in the case $d \geq 2$. The strategy is to find the first derivative $f^{(m)}(x^*)$ which is non zero and then to choose $\alpha_1 = 1/m$.

Proposition 2. *Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be C^p for some $p \in \mathbb{N}$ and let x^* be a strict polynomial local minimum of f . Then :*

1. *The order of the local minimum m is an even number and $f^{(m)}(x^*) > 0$.*
2. $f(x^* + h) \underset{h \rightarrow 0}{=} f(x^*) + \frac{f^{(m)}(x^*)}{m!} h^m + o(h^m)$

Then $\alpha_1 := 1/m$ is the solution of (6) and

$$\frac{1}{t} (f(x^* + t^{1/m} h) - f(x^*)) \xrightarrow{t \rightarrow 0} \frac{f^{(m)}(x^*)}{m!} h^m$$

which is a non-constant function of h , since $f^{(m)}(x^*) \neq 0$. The direct proof using the Taylor formula is left to the reader.

4.3 Proof of Theorem 7 for $p = 1$

Let $f \in \mathcal{A}_1^*(x^*)$. The assumption that x^* is a strict polynomial local minimum at order 2 implies that $\nabla^2 f(x^*)$ is positive definite. Let us denote $(\beta_i)_{1 \leq i \leq d}$ its positive eigenvalues. By the spectral theorem, let us write $\nabla^2 f(x^*) = B \text{Diag}(\beta_{1:d}) B^T$ for some $B \in \mathcal{O}_d(\mathbb{R})$. Then:

$$\frac{1}{t} (f(x^* + t^{1/2} B \cdot h) - f(x^*)) \xrightarrow{t \rightarrow 0} \frac{1}{2} \sum_{i=1}^d \beta_i h_i^2. \quad (17)$$

Thus, a solution of (6) is $\alpha_1 = \dots = \alpha_d = \frac{1}{2}$, B , and $g(h_1, \dots, h_d) = \frac{1}{2} \sum_{i=1}^d \beta_i h_i^2$, which is a non-constant function of every h_1, \dots, h_d , since for all i , β_i is positive.

In the following, our objective is to establish a similar result when $\nabla^2 f(x^*)$ is not necessarily positive definite.

4.4 Proof of Theorem 7 for $p = 2$

Theorem 8. *Let $f \in \mathcal{A}_2(x^*)$. Then there exist orthogonal subspaces E and F such that $\mathbb{R}^d = E \oplus F$, and that for all $h \in \mathbb{R}^d$:*

$$\begin{aligned} & \frac{1}{t} \left[f(x^* + t^{1/2} p_E(h) + t^{1/4} p_F(h)) - f(x^*) \right] \\ & \xrightarrow{t \rightarrow 0} \frac{1}{2} \nabla^2 f(x^*) \cdot p_E(h)^{\otimes 2} + \frac{1}{2} \nabla^3 f(x^*) \cdot p_E(h) \otimes p_F(h)^{\otimes 2} + \frac{1}{4!} \nabla^4 f(x^*) \cdot p_F(h)^{\otimes 4}. \end{aligned} \quad (18)$$

Moreover, if $f \in \mathcal{A}_2^*(x^*)$, then this is a solution to the problem (6), with $E_1 = E$, $E_2 = F$, α defined in (15), B adapted to the previous decomposition and g defined in (14).

Remark: The set of 2-tuples (i_1, i_2) such that $\frac{i_1}{2} + \frac{i_2}{4} = 1$, are $(2, 0)$, $(1, 2)$ and $(0, 4)$, which gives the terms appearing in the sum in (12).

Proof. Let $F := \{h \in \mathbb{R}^d : \nabla^2 f(x^*) \cdot h^{\otimes 2} = 0\}$. By the spectral theorem and since $\nabla^2 f(x^*)$ is positive semi-definite, $F = \{h \in \mathbb{R}^d : \nabla^2 f(x^*) \cdot h = 0^{\otimes 1}\}$ is a vector subspace of \mathbb{R}^d . Let E be the orthogonal complement of F in \mathbb{R}^d .

For $h \in \mathbb{R}^d$ we expand the left term of (18) using the Taylor formula up to order 4 and the multinomial formula (2), giving

$$\sum_{k=2}^4 \frac{1}{k!} \sum_{\substack{i_1, i_2 \in \{0, \dots, k\} \\ i_1 + i_2 = k}} \binom{k}{i_1, i_2} t^{\frac{i_1}{2} + \frac{i_2}{4} - 1} T_k \cdot p_E(h)^{\otimes i_1} \otimes p_F(h)^{\otimes i_2} + o(1).$$

The terms with coefficient t^a , $a > 0$, are $o(1)$ as $t \rightarrow 0$. By definition of F we have $\nabla^2 f(x^*) \cdot p_F(h) = 0^{\otimes 1}$, so we also have

$$\nabla^3 f(x^*) \cdot p_F(h)^{\otimes 3} = 0$$

by the local minimum condition. This yields the convergence stated in (18).

Moreover, if x^* is a local minimum of polynomial order 4, then by the local minimum condition, $\nabla^4 f(x^*) > 0$ on F in the sense of (3). Moreover, since $\nabla^2 f(x^*) > 0$ on E , then the limit is not constant in any h_1, \dots, h_d . \square

Remark: The cross odd term is not necessarily null. For example, consider

$$\begin{aligned} f : \quad \mathbb{R}^2 & \longrightarrow \mathbb{R} \\ (x, y) & \longmapsto x^2 + y^4 + xy^2. \end{aligned}$$

Then f admits a global minimum at $x^* = 0$ since $|xy^2| \leq \frac{1}{2}(x^2 + y^4)$. We have $E_1 = \mathbb{R}(1, 0)$, $E_2 = \mathbb{R}(0, 1)$ and for all $(x, y) \in \mathbb{R}^2$, $T_3 \cdot (xe_1) \otimes (ye_2)^{\otimes 2} = 2xy^2$ is not identically null.

4.5 Difficulties beyond the 4th order and Hilbert's 17th problem

If we do not assume as in the previous section that $\nabla^4 f(x^*)$ is not positive on F , then we carry on the development of $f(x^*+h)$ up to higher orders. A first idea is to consider $F_2 := \{h \in F : \nabla^4 f(x^*) \cdot h^{\otimes 4} = 0\} \subseteq F$ and E_2 a complement subspace of F_2 in F , and to continue this process by induction as in Section 4.4. However, F_2 is not necessarily a subspace of F .

Indeed, let T be a symmetric tensor defined on $\mathbb{R}^{d'}$ of order $2k$ with $k \in \mathbb{N}$. As T is symmetric, there exist vectors $v^1, \dots, v^q \in \mathbb{R}^{d'}$, and scalars $\lambda_1, \dots, \lambda_q \in \mathbb{R}$ such that $T = \sum_i \lambda_i (v^i)^{\otimes 2k}$ (see [CGLM08], Lemma 4.2.), so

$$\forall h \in \mathbb{R}^{d'}, T \cdot h^{\otimes 2k} = \sum_{i=1}^q \lambda_i (v^i)^{\otimes 2k} \cdot h^{\otimes 2k} = \sum_{i=1}^q \lambda_i \langle v^i, h \rangle^{2k}.$$

For $k = 2$ and $T = \nabla^{2k} f(x^*)|_F$, since x^* is a local minimum, we have, identifying F and $\mathbb{R}^{d'}$,

$$\forall h \in \mathbb{R}^{d'}, T \cdot h^{\otimes 2k} \geq 0$$

Then, we could think it implies that for all i , $\lambda_i \geq 0$, and then

$$T \cdot h^{\otimes 2k} = 0 \implies \forall i, \langle v^i, h \rangle = 0$$

which would give a linear characterization of $\{h \in \mathbb{R}^{d'} : T \cdot h^{\otimes 2k} = 0\}$ and in this case, F_2 would be a subspace of F . However this reasoning is not correct in general as we do not have necessarily that for all i , $\lambda_i \geq 0$.

We can build counter-examples as follows. Since T is a non-negative symmetric tensor, T can be seen as a non-negative homogeneous polynomial of degree $2k$ with d' variables. A counter-example at order $2k = 4$ is $T(X, Y, Z) = ((X - Y)(X - Z))^2$, which is a non-negative polynomial of order 4, but $\{T = 0\} = \{X = Y \text{ or } X = Z\}$, which is not a vector space.

Another counterexample given in [Mot67] at order $2k = 6$ is the following. We define

$$T(X, Y, Z) = Z^6 + X^4 Y^2 + X^2 Y^4 - 3X^2 Y^2 Z^2$$

By the arithmetic-geometric mean inequality and its equality case, T is non-negative and $T(x, y, z) = 0$ if and only if $z^6 = x^4 y^2 = x^2 y^4$, so that

$$\{T = 0\} = \mathbb{R} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \cup \mathbb{R} \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \cup \mathbb{R} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \cup \mathbb{R} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}.$$

Hence, $\{T = 0\}$ is not a subspace of \mathbb{R}^3 . In particular T cannot be written as $\sum_i \lambda_i (v^i)^{\otimes 2k}$ with $\lambda_i \geq 0$.

In fact, this problem is linked with the Hilbert's seventeenth problem that we recall below.

Problem 1 (Hilbert's seventeenth problem). *Let P be a non-negative polynomial with d' variables, homogeneous of even degree $2k$. Find polynomials P_1, \dots, P_r with d' variables, homogeneous of degree k , such that $P = \sum_{i=1}^r P_i^2$*

Hilbert proved in 1888 [Hil88] that there does not always exist a solution. In general $\{T = 0\}$ is not even a submanifold of $\mathbb{R}^{d'}$. Indeed, taking $T : h \mapsto \nabla^{2k} f(x^*) \cdot h^{\otimes 2k}$, we have $\partial_h T \cdot h = 2k \nabla^{2k} f(x^*) \cdot h^{\otimes 2k-1}$ is not surjective in $h = 0$, so the surjectivity condition for $\{T = 0\}$ to be a submanifold is not fulfilled.

4.6 Proof of Theorem 7 for $p = 3$

We slightly change our strategy of proof developed in Section 4.4. For $k \geq 2$, we define F_k recursively as

$$F_k := \{h \in F_{k-1} : \forall h' \in F_{k-1}, \nabla^{2k} f(x^*) \cdot h \otimes h'^{\otimes 2k-1} = 0\}, \quad (19)$$

instead of $\{h \in F_{k-1} : \nabla^{2k} f(x^*) \cdot h^{\otimes 2k} = 0\}$. Then, by construction, F_k is a vector subspace of \mathbb{R}^d .

Theorem 9. *Let $f \in \mathcal{A}_3(x^*)$. Then there exist orthogonal subspaces of \mathbb{R}^d , E_1 , E_2 and F_2 , such that*

$$\mathbb{R}^d = E_1 \oplus E_2 \oplus F_2,$$

and such that for all $h \in \mathbb{R}^d$,

$$\begin{aligned} & \frac{1}{t} \left[f(x^* + t^{1/2} p_{E_1}(h) + t^{1/4} p_{E_2}(h) + t^{1/6} p_{F_2}(h)) - f(x^*) \right] \\ & \xrightarrow{t \rightarrow 0} \frac{1}{2} \nabla^2 f(x^*) \cdot p_{E_1}(h)^{\otimes 2} + \frac{1}{2} \nabla^3 f(x^*) \cdot p_{E_1}(h) \otimes p_{E_2}(h)^{\otimes 2} + \frac{1}{4!} \nabla^4 f(x^*) \cdot p_{E_2}(h)^{\otimes 4} \\ & \quad + \frac{4}{4!} \nabla^4 f(x^*) \cdot p_{E_1}(h) \otimes p_{F_2}(h)^{\otimes 3} + \frac{10}{5!} \nabla^5 f(x^*) \cdot p_{E_2}(h)^{\otimes 2} \otimes p_{F_2}(h)^{\otimes 3} + \frac{1}{6!} \nabla^6 f(x^*) \cdot p_{F_2}(h)^{\otimes 6}. \end{aligned} \quad (20)$$

Moreover, if $f \in \mathcal{A}_3^*(x^*)$, then this is a solution to the problem (6), with $E_3 = F_2$, α defined in (15), B adapted to the previous decomposition and g defined in (14).

Remark: The set of 3-tuples (i_1, i_2, i_3) such that $\frac{i_1}{2} + \frac{i_2}{4} + \frac{i_3}{6} = 1$, are $(2, 0, 0)$, $(1, 2, 0)$, $(0, 4, 0)$, $(1, 0, 3)$, $(0, 2, 3)$, $(0, 0, 6)$, which gives the terms appearing in (12).

Proof. We consider the subspace

$$F_1 := \{h \in \mathbb{R}^d : T_2 \cdot h^{\otimes 2} = 0\} = \{h \in \mathbb{R}^d : T_2 \cdot h = 0^{\otimes 1}\},$$

since $T_2 \geq 0$. Then, let E_1 be the orthogonal complement of F_1 in \mathbb{R}^d and consider the vector subspace of F_1 defined by

$$F_2 = \{h \in F_1 : \forall h' \in F_1, T_4 \cdot h \otimes h'^{\otimes 3} = 0\}.$$

Let E_2 be the orthogonal complement of F_2 in F_1 . Then we have

$$\mathbb{R}^d = E_1 \oplus F_1 = E_1 \oplus E_2 \oplus F_2.$$

For $h \in \mathbb{R}^d$ we expand the left term of (20) using the Taylor formula up to order 6 and the multinomial formula (2), giving

$$\sum_{k=2}^6 \frac{1}{k!} \sum_{\substack{i_1, i_2, i_3 \in \{0, \dots, k\} \\ i_1 + i_2 + i_3 = k}} \binom{k}{i_1, i_2, i_3} t^{\frac{i_1}{2} + \frac{i_2}{4} + \frac{i_3}{6} - 1} T_k \cdot p_{E_1}(h)^{\otimes i_1} \otimes p_{E_2}(h)^{\otimes i_2} \otimes p_{F_2}(h)^{\otimes i_3} + o(1),$$

and we prove the convergence stated in (20).

All the terms with coefficient t^a where $a > 0$ are $o(1)$ as $t \rightarrow 0$.

Order 2: we have $T_2 \cdot p_{E_2}(h) = 0^{\otimes 1}$ and $T_2 \cdot p_{F_2}(h) = 0^{\otimes 1}$ so the only term for $k = 2$ is $\frac{1}{2} T_2 \cdot p_{E_1}(h)^{\otimes 2}$.

Order 3: \triangleright Since x^* is a local minimum and $T_2 \cdot p_{F_1}(h)^{\otimes 2} = 0$, we have $T_3 \cdot p_{F_1}(h)^{\otimes 3} = 0$. Then, using property Proposition 6, if the factor $p_{E_1}(h)$ does not appear as an argument in T_3 , then the corresponding term is zero.

\triangleright Let us prove that

$$T_3 \cdot p_{E_1}(h) \otimes p_{F_2}(h)^{\otimes 2} = 0.$$

Using Theorem 8 with $E = E_1$, $F = E_2 \oplus F_2$, we have in particular that for all $h \in \mathbb{R}^d$,

$$\frac{1}{2} T_2 \cdot p_E(h)^{\otimes 2} + \frac{1}{2} T_3 \cdot p_E(h) \otimes p_F(h)^{\otimes 2} + \frac{1}{4!} T_4 \cdot p_F(h)^{\otimes 4} \geq 0. \quad (21)$$

Then taking $h \in E_1 \oplus F_2$ so that $h = p_{E_1}(h) + p_{F_2}(h)$ and with

$$\left[T_4 \cdot p_{F_2}(h) \right]_{|F_1} \equiv 0^{\otimes 3}, \quad (22)$$

we may rewrite (21) as

$$\frac{1}{2}T_2 \cdot p_{E_1}(h)^{\otimes 2} + \frac{1}{2}T_3 \cdot p_{E_1}(h) \otimes p_{F_2}(h)^{\otimes 2} \geq 0.$$

Now, considering $h' = \lambda h$, we have that for all $\lambda \in \mathbb{R}$,

$$\lambda^2 \left(\frac{1}{2}T_2 \cdot p_{E_1}(h)^{\otimes 2} + \frac{\lambda}{2}T_3 \cdot p_{E_1}(h) \otimes p_{F_2}(h)^{\otimes 2} \right) \geq 0,$$

so that necessarily $T_3 \cdot p_{E_1}(h) \otimes p_{F_2}(h)^{\otimes 2} = 0$.

▷ Let us prove that

$$T_3 \cdot p_{E_1}(h) \otimes p_{E_2}(h) \otimes p_{F_2}(h) = 0.$$

We use again (21), with $p_F(h) = p_{E_2}(h) + p_{F_2}(h)$, so that

$$\frac{1}{2}T_2 \cdot p_{E_1}(h)^{\otimes 2} + \frac{1}{2}T_3 \cdot p_{E_1}(h) \otimes (p_{E_2}(h) + p_{F_2}(h))^{\otimes 2} + \frac{1}{4!}T_4 \cdot (p_{E_2}(h) + p_{F_2}(h))^{\otimes 4} \geq 0.$$

But using (22) and that $T_3 \cdot p_{E_1}(h) \otimes p_{F_2}(h)^{\otimes 2} = 0$, we obtain

$$\frac{1}{2}T_2 \cdot p_{E_1}(h)^{\otimes 2} + \frac{1}{2}T_3 \cdot p_{E_1}(h) \otimes p_{E_2}(h)^{\otimes 2} + T_3 \cdot p_{E_1}(h) \otimes p_{E_2}(h) \otimes p_{F_2}(h) + \frac{1}{4!}T_4 \cdot p_{E_2}(h)^{\otimes 4} \geq 0.$$

Now, considering $h' = p_{E_1}(h) + p_{E_2}(h) + \lambda p_{F_2}(h)$, we have that for all $\lambda \in \mathbb{R}$,

$$\frac{1}{2}T_2 \cdot p_{E_1}(h)^{\otimes 2} + \frac{1}{2}T_3 \cdot p_{E_1}(h) \otimes p_{E_2}(h)^{\otimes 2} + \lambda T_3 \cdot p_{E_1}(h) \otimes p_{E_2}(h) \otimes p_{F_2}(h) + \frac{1}{4!}T_4 \cdot p_{E_2}(h)^{\otimes 4} \geq 0,$$

so necessarily $T_3 \cdot p_{E_1}(h) \otimes p_{E_2}(h) \otimes p_{F_2}(h) = 0$.

▷ The last remaining term for $k = 3$ is $\frac{1}{2}T_3 \cdot p_{E_1}(h) \otimes p_{E_2}(h)^{\otimes 2}$.

Order 4: If the factor $p_{E_1}(h)$ does not appear and if the factor $p_{F_2}(h)$ appears at least once, then using (22) the corresponding term is zero. If $p_{E_1}(h)$ appears, the only term with a non-positive exponent of t is $\frac{4}{4!}T_4 \cdot p_{E_1}(h) \otimes p_{F_2}(h)^{\otimes 3}$. So the only terms for $k = 4$ are $\frac{1}{4!}T_4 \cdot p_{E_2}(h)^{\otimes 4}$ and $\frac{4}{4!}T_4 \cdot p_{E_1}(h) \otimes p_{F_2}(h)^{\otimes 3}$.

Order 5: ▷ The terms where $p_{E_1}(h)$ appears at least once have a coefficient t^a with $a > 0$ so are $o(1)$ when $t \rightarrow 0$.

▷ We have $T_2 \cdot p_{F_2}(h)^{\otimes 2} = 0$, $T_3 \cdot p_{F_2}(h)^{\otimes 3} = 0$, $T_4 \cdot p_{F_2}(h)^{\otimes 4} = 0$ and since x^* is a local minimum, we have

$$T_5 \cdot p_{F_2}(h)^{\otimes 5} = 0.$$

▷ Let us prove that

$$T_5 \cdot p_{E_2}(h) \otimes p_{F_2}(h)^{\otimes 4} = 0.$$

Let $h \in \mathbb{R}^d$. We have

$$\frac{1}{t^{11/12}} \left[f(x^* + t^{1/4}p_{E_2}(h) + t^{1/6}p_{F_2}(h)) - f(x^*) \right] \xrightarrow{t \rightarrow 0} \frac{1}{4!}T_5 \cdot p_{E_2}(h) \otimes p_{F_2}(h)^{\otimes 4} \geq 0.$$

Hence, considering $h' = \lambda h$, we have for every $\lambda \in \mathbb{R}$,

$$\lambda^5 T_5 \cdot p_{E_2}(h) \otimes p_{F_2}(h)^{\otimes 4} \geq 0,$$

which yields the desired result.

Order 2	(2, 0, 0, 0)
Order 3	(2, 1, 0, 0)
Order 4	(0, 4, 0, 0), (1, 1, 0, 2), (1, 0, 3, 0)
Order 5	(1, 0, 0, 4), (0, 2, 3, 0), (0, 3, 0, 2)
Order 6	(0, 1, 3, 2), (0, 2, 0, 4), (0, 0, 6, 0)
Order 7	(0, 1, 0, 6), (0, 0, 3, 4)
Order 8	(0, 0, 0, 8)

Table 1: Terms expressed as 4-tuples in the development (23)

▷ The only remaining term for $p = 5$ is

$$\frac{10}{5!} T_5 \cdot p_{E_2}(h)^{\otimes 2} \otimes p_{F_2}(h)^{\otimes 3}.$$

Order 6: The only term for $k = 6$ is $\frac{1}{6!} T_6 \cdot p_{F_2}(h)^{\otimes 6}$; the other terms have a coefficient t^a with $a > 0$, so are $o(1)$ when $t \rightarrow 0$. \square

Remark : As in Theorem 8 and the remark that follows, the remaining odd cross-terms cannot be proved to be zero using the same method of proof, and may be actually not zero. For example, consider:

$$f : \quad \mathbb{R}^2 \quad \longrightarrow \quad \mathbb{R} \\ (x, y) \quad \longmapsto \quad x^4 + y^6 + x^2 y^3,$$

which satisfies $h \mapsto \nabla^5 f(x^*) \cdot p_{E_2}(h)^{\otimes 2} \otimes p_{F_2}(h)^{\otimes 3} \neq 0$.

4.7 Proof of Theorem 7 for $p = 4$

Theorem 10. *Let $f \in \mathcal{A}_4(x^*)$. Then there exist orthogonal subspaces of \mathbb{R}^d , E_1 , E_2 , E_3 and F_3 such that*

$$\mathbb{R}^d = E_1 \oplus E_2 \oplus E_3 \oplus F_3,$$

and for all $h \in \mathbb{R}^d$,

$$\begin{aligned} & \frac{1}{t} \left[f(x^* + t^{1/2} p_{E_1}(h) + t^{1/4} p_{E_2}(h) + t^{1/6} p_{E_3}(h) + t^{1/8} p_{F_3}(h)) - f(x^*) \right] \\ \xrightarrow{t \rightarrow 0} & \sum_{k=2}^8 \frac{1}{k!} \sum_{\substack{i_1, \dots, i_4 \in \{0, \dots, k\} \\ i_1 + \dots + i_4 = k}} \binom{k}{i_1, \dots, i_4} T_k \cdot p_{E_1}(h)^{\otimes i_1} \otimes p_{E_2}(h)^{\otimes i_2} \otimes p_{E_3}(h)^{\otimes i_3} \otimes p_{F_3}(h)^{\otimes i_4}. \end{aligned} \quad (23)$$

These terms are summarized as tuples (i_1, \dots, i_4) in Table 1. Moreover, if $f \in \mathcal{A}_4^*(x^*)$, then this is a solution to (6), with $E_4 = F_3$, α defined in (15), B adapted to the previous decomposition and g defined in (14).

Proof. As before, we define the subspaces $F_0 := \mathbb{R}^d$ and by induction:

$$F_k = \{h \in F_{k-1} : \forall h' \in F_{k-1}, T_{2k} \cdot h \otimes h'^{\otimes 3} = 0\}$$

for $k = 1, 2, 3$. We define E_k as the orthogonal complement of F_k in F_{k-1} for $k = 1, 2, 3$, so that

$$\mathbb{R}^d = E_1 \oplus E_2 \oplus E_3 \oplus F_3.$$

E_1	F_1		
$T_2 \geq 0$	$T_2 = 0$		
	E_2	F_2	
	$T_4 \geq 0$	$T_4 = 0$	
		E_3	F_3
	$T_6 \geq 0$	$T_6 = 0$	

Table 2: Illustration of the subspaces

Then we apply a Taylor expansion up to order 8 to the left side of (23) and the multinomial formula (2), which reads

$$\sum_{k=2}^8 \frac{1}{k!} \sum_{\substack{i_1, \dots, i_4 \in \{0, \dots, k\} \\ i_1 + \dots + i_4 = k}} \binom{k}{i_1, \dots, i_4} t^{\frac{i_1}{2} + \dots + \frac{i_4}{8} - 1} T_k \cdot p_{E_1}(h)^{\otimes i_1} \otimes p_{E_2}(h)^{\otimes i_2} \otimes p_{E_3}(h)^{\otimes i_3} \otimes p_{F_3}(h)^{\otimes i_4} + o(1).$$

▷ If $\frac{i_1}{2} + \dots + \frac{i_4}{8} > 1$ then the corresponding term is in $o(1)$ when $t \rightarrow 0$.

▷ If $\frac{i_1}{2} + \dots + \frac{i_4}{8} < 1$ then the corresponding term diverges when $t \rightarrow 0$, so we need to prove that actually

$$T_k \cdot p_{E_1}(h)^{\otimes i_1} \otimes p_{E_2}(h)^{\otimes i_2} \otimes p_{E_3}(h)^{\otimes i_3} \otimes p_{F_3}(h)^{\otimes i_4} = 0. \quad (24)$$

– If $\frac{i_1}{2} + \frac{i_2}{4} + \frac{i_3}{6} + \frac{i_4}{8} < 1$ but if we also have $\frac{i_1}{2} + \frac{i_2}{4} + \frac{i_3}{6} + \frac{i_4}{6} < 1$, then by applying the property at the order 6 (Theorem 9) with the 3-tuple $(i_1, i_2, i_3 + i_4)$, we get (24).

– So we only need to consider 4-tuples such that $\frac{i_1}{2} + \frac{i_2}{4} + \frac{i_3}{6} + \frac{i_4}{8} < 1$ and $\frac{i_1}{2} + \frac{i_2}{4} + \frac{i_3}{6} + \frac{i_4}{6} \geq 1$. We can remove all the terms which are null by the definitions of the subspaces E_1, E_2, E_3, F_3 . The remaining terms are:

For $k = 4$: $\frac{t^{21/24}}{6} T_4 \cdot p_{E_1}(h) \otimes p_{F_3}(h)^{\otimes 3}$, $\frac{t^{11/12}}{2} T_4 \cdot p_{E_1}(h) \otimes p_{E_3}(h) \otimes p_{F_3}(h)^{\otimes 2}$, $\frac{t^{23/24}}{2} T_4 \cdot p_{E_1}(h) \otimes p_{E_3}(h)^{\otimes 2} \otimes p_{F_3}(h)$.

For $k = 5$: $\frac{t^{21/24}}{12} T_5 \cdot p_{E_2}(h)^{\otimes 2} \otimes p_{F_3}(h)^{\otimes 3}$, $\frac{t^{11/12}}{4} T_5 \cdot p_{E_2}(h)^{\otimes 2} \otimes p_{E_3}(h) \otimes p_{F_3}(h)^{\otimes 2}$, $\frac{t^{23/24}}{4} T_5 \cdot p_{E_2}(h)^{\otimes 2} \otimes p_{E_3}(h)^{\otimes 2} \otimes p_{F_3}(h)$.

For $k = 6$: $\frac{t^{21/24}}{5!} T_6 \cdot p_{E_2}(h) \otimes p_{F_3}(h)^{\otimes 5}$, $\frac{t^{11/12}}{4!} T_6 \cdot p_{E_2}(h) \otimes p_{E_3}(h) \otimes p_{F_3}(h)^{\otimes 4}$, $\frac{t^{23/24}}{12} T_6 \cdot p_{E_2}(h) \otimes p_{E_3}(h)^{\otimes 2} \otimes p_{F_3}(h)^{\otimes 3}$.

First, we note that

$$\begin{aligned} & \frac{1}{t^{21/24}} \left[f(x^* + t^{1/2} p_{E_1}(h) + t^{1/4} p_{E_2}(h) + t^{1/6} p_{E_3}(h) + t^{1/8} p_{F_3}(h)) - f(x^*) \right] \\ & \xrightarrow{t \rightarrow 0} \frac{1}{6} T_4 \cdot p_{E_1}(h) \otimes p_{F_3}(h)^{\otimes 3} + \frac{1}{12} T_5 \cdot p_{E_2}(h)^{\otimes 2} \otimes p_{F_3}(h)^{\otimes 3} + \frac{1}{5!} T_6 \cdot p_{E_2}(h) \otimes p_{F_3}(h)^{\otimes 5} \geq 0. \end{aligned}$$

Then, considering $h' = \lambda p_{E_1}(h) + p_{E_2}(h) + p_{E_3}(h) + p_{F_3}(h)$, we have that for all $\lambda \in \mathbb{R}$,

$$\frac{\lambda}{6} T_4 \cdot p_{E_1}(h) \otimes p_{F_3}(h)^{\otimes 3} + \frac{1}{12} T_5 \cdot p_{E_2}(h)^{\otimes 2} \otimes p_{F_3}(h)^{\otimes 3} + \frac{1}{5!} T_6 \cdot p_{E_2}(h) \otimes p_{F_3}(h)^{\otimes 5} \geq 0,$$

so necessarily

$$T_4 \cdot p_{E_1}(h) \otimes p_{F_3}(h)^{\otimes 3} = 0.$$

Then, considering $h' = p_{E_2}(h) + \lambda p_{F_3}(h)$ for $\lambda \in \mathbb{R}$, we get successively that the two other terms are null.

Likewise, we prove successively that the terms in $t^{11/12}$ are null, and then that the terms in $t^{23/24}$ are null. This yields the convergence stated in (23). \square

4.8 Counter-example and proof of Theorem 7 with $p \geq 5$ under the hypothesis (16)

Algorithm 1 may fail to yield such expansion of f for orders no lower than 10 if the hypothesis (16) is not fulfilled. Indeed for $p \geq 5$, there exist p -tuples (i_1, \dots, i_p) such that $\frac{i_1}{2} + \dots + \frac{i_p}{2p} < 1$ and i_1, \dots, i_p are all even. Such tuples do not appear at orders 8 and lower, but they do appear at orders 10 and higher, for example $(0, 2, 0, 0, 4)$ for $k = 6$. In such a case, we cannot use the positiveness argument to prove that the corresponding term $T_k \cdot p_{E_1}(h)^{\otimes i_1} \otimes \dots \otimes p_{E_p}(h)^{\otimes i_p}$ is zero, and in fact, it may be not zero.

Let us give a counter example. Consider

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R} \\ (x, y) \longmapsto x^4 + y^{10} + x^2 y^4.$$

Then $f \in \mathcal{A}_5^*(0)$ and we have $E_1 = \{0\}$, $E_2 = \mathbb{R} \cdot (1, 0)$, $E_3 = \{0\}$, $E_4 = \{0\}$, $F_4 = \mathbb{R} \cdot (0, 1)$. But

$$\frac{1}{t} f(t^{1/4}, t^{1/10}) = \frac{1}{t} (t + t + t^{9/10})$$

goes to $+\infty$ when $t \rightarrow 0$.

Now, let us give the proof of Theorem 7 for $p \geq 5$. In this proof, we assume that the subspaces E_1, \dots, E_p given in Algorithm 1 satisfy the hypothesis (16).

Proof. We develop (11), which reads:

$$\sum_{k=2}^{2p} \frac{1}{k!} \sum_{\substack{i_1, \dots, i_p \in \{0, \dots, k\} \\ i_1 + \dots + i_p = k}} \binom{k}{i_1, \dots, i_p} t^{\frac{i_1}{2} + \dots + \frac{i_p}{2p} - 1} T_k \cdot p_{E_1}(h)^{\otimes i_1} \otimes \dots \otimes p_{E_p}(h)^{\otimes i_p} + o(1) =: S.$$

The terms such that $\frac{i_1}{2} + \dots + \frac{i_p}{2p} < 1$ may diverge when $t \rightarrow 0$, so let us prove that they are in fact null. Let

$$\alpha := \inf \left\{ \frac{i_1}{2} + \dots + \frac{i_p}{2p} : h \longmapsto \sum_{k=2}^{2p} \frac{1}{k!} \sum_{\substack{i_1, \dots, i_p \in \{0, \dots, k\} \\ i_1 + \dots + i_p = k \\ \frac{i_1}{2} + \dots + \frac{i_p}{2p} = \alpha}} \binom{k}{i_1, \dots, i_p} T_k \cdot p_{E_1}(h)^{\otimes i_1} \otimes \dots \otimes p_{E_p}(h)^{\otimes i_p} \neq 0 \right\},$$

and assume by contradiction that $\alpha < 1$. Then we have for all $h \in \mathbb{R}^d$:

$$t^{1-\alpha} S \xrightarrow{t \rightarrow 0} \left(\sum_{k=2}^{2p} \frac{1}{k!} \sum_{\substack{i_1, \dots, i_p \in \{0, \dots, k\} \\ i_1 + \dots + i_p = k \\ \frac{i_1}{2} + \dots + \frac{i_p}{2p} = \alpha}} \binom{k}{i_1, \dots, i_p} T_k \cdot p_{E_1}(h)^{\otimes i_1} \otimes \dots \otimes p_{E_p}(h)^{\otimes i_p} \right) \geq 0,$$

by the local minimum property. Then, considering $h' = \lambda_1 p_{E_1}(h) + \dots + \lambda_p p_{E_p}(h)$, we have, for all $h \in \mathbb{R}^d$ and $\lambda_1, \dots, \lambda_d \in \mathbb{R}$,

$$\sum_{k=2}^{2p} \frac{1}{k!} \sum_{\substack{i_1, \dots, i_p \in \{0, \dots, k\} \\ i_1 + \dots + i_p = k \\ \frac{i_1}{2} + \dots + \frac{i_p}{2p} = \alpha}} \lambda_1^{i_1} \dots \lambda_p^{i_p} \binom{k}{i_1, \dots, i_p} T_k \cdot p_{E_1}(h)^{\otimes i_1} \otimes \dots \otimes p_{E_p}(h)^{\otimes i_p} \geq 0. \quad (25)$$

Now, we fix $h \in \mathbb{R}^d$ such that the polynomial in (25) in the variables $\lambda_1, \dots, \lambda_p$ is not identically zero, and we consider k_{\max} its highest homogeneous degree, so that we have

$$\sum_{\substack{i_1, \dots, i_p \in \{0, \dots, k_{\max}\} \\ i_1 + \dots + i_p = k_{\max} \\ \frac{i_1}{2} + \dots + \frac{i_p}{2} = \alpha}} \lambda_1^{i_1} \dots \lambda_p^{i_p} \binom{k_{\max}}{i_1, \dots, i_p} T_{k_{\max}} \cdot p_{E_1}(h)^{\otimes i_1} \otimes \dots \otimes p_{E_p}(h)^{\otimes i_p} \geq 0.$$

If k_{\max} is odd, this yields a contradiction, taking $\lambda_1 = \dots = \lambda_p =: \lambda \rightarrow \pm\infty$. If k_{\max} is even, we consider the index l_1 such that $i_{l_1} =: a_1$ is maximal and the coefficients in the above sum with $i_{l_1} = a_1$ are not all zero. Then fixing all the λ_l for $l \neq l_1$ and taking $\lambda_{l_1} \rightarrow \infty$, we have

$$\sum_{\substack{i_1, \dots, i_p \in \{0, \dots, k_{\max}\} \\ i_1 + \dots + i_p = k_{\max} \\ \frac{i_1}{2} + \dots + \frac{i_p}{2} = \alpha \\ i_{l_1} = a_1}} \lambda_1^{i_1} \dots \lambda_p^{i_p} \binom{k_{\max}}{i_1, \dots, i_p} T_{k_{\max}} \cdot p_{E_1}(h)^{\otimes i_1} \otimes \dots \otimes p_{E_p}(h)^{\otimes i_p} \geq 0.$$

Thus, if a_1 is odd, this yields a contradiction. If a_1 is even, we carry on this process by induction : knowing l_1, \dots, l_r , we choose the index l_{r+1} such that $l_{r+1} \notin \{l_1, \dots, l_r\}$, the corresponding term

$$\sum_{\substack{i_1, \dots, i_p \in \{0, \dots, k_{\max}\} \\ i_1 + \dots + i_p = k_{\max} \\ \frac{i_1}{2} + \dots + \frac{i_p}{2} = \alpha \\ i_{l_1} = a_1, \dots, i_{l_{r+1}} = a_{r+1}}} \lambda_1^{i_1} \dots \lambda_p^{i_p} \binom{k_{\max}}{i_1, \dots, i_p} T_{k_{\max}} \cdot p_{E_1}(h)^{\otimes i_1} \otimes \dots \otimes p_{E_p}(h)^{\otimes i_p}$$

is not identically null and such that $i_{l_{r+1}} =: a_{r+1}$ is maximal. Necessarily, a_{r+1} is even. In the end we will find a non-zero term whose exponents i_ℓ are all even which contradicts assumption (16). \square

4.9 Proofs of the uniform convergence and of the non-constant property

In this section we prove the additional properties claimed in Theorem 7 : the uniform convergence with respect to h on every compact set and the fact that the function g is not constant in any of its variables h_1, \dots, h_d .

Proof. First, let us prove that the convergence is uniform with respect to h on every compact set. Let $\varepsilon > 0$ and let $R > 0$. By the Taylor formula at order $2p$, there exists $\delta > 0$ such that for $\|h\| < \delta$,

$$\left| f(x^* + h) - f(x^*) - \sum_{k=2}^{2p} \frac{1}{k!} \sum_{i_1 + \dots + i_p = k} \binom{k}{i_1, \dots, i_p} T_k \cdot p_{E_1}(h)^{\otimes i_1} \otimes \dots \otimes p_{E_p}(h)^{\otimes i_p} \right| \leq \varepsilon \|h\|^{2p}.$$

Now, let us consider $t \rightarrow 0$ and $h \in \mathbb{R}^d$ with $\|h\| \leq R$. Then we have:

$$\forall t \leq \max \left(1, \left(\frac{\delta}{R} \right)^{1/(2p)} \right), \|t^{1/2} p_{E_1}(h) + \dots + t^{1/(2p)} p_{E_p}(h)\| \leq \delta,$$

so that

$$\left| \frac{1}{t} \left[f(x^* + t^{1/2} p_{E_1}(h) + \dots + t^{1/(2p)} p_{E_p}(h)) - f(x^*) \right] - \sum_{k=2}^{2p} \frac{1}{k!} \sum_{i_1 + \dots + i_p = k} \binom{k}{i_1, \dots, i_p} \cdot t^{\frac{i_1}{2} + \dots + \frac{i_p}{2} - 1} T_k \cdot p_{E_1}(h)^{\otimes i_1} \otimes \dots \otimes p_{E_p}(h)^{\otimes i_p} \right| \leq \frac{\varepsilon}{t} \|t^{1/2} p_{E_1}(h) + \dots + t^{1/(2p)} p_{E_p}(h)\|^{2p}.$$

We proved or assumed that the terms such that $\frac{i_1}{2} + \dots + \frac{i_p}{2p} < 1$ are zero. We denote by $g_1(h)$ the sum in the last equation with the terms such that $\frac{i_1}{2} + \dots + \frac{i_p}{2p} = 1$ and by $g_2(h)$ the sum with the terms such that $\frac{i_1}{2} + \dots + \frac{i_p}{2p} > 1$. We also define a as the smallest exponent of t appearing in $g_2(h)$:

$$a := \min \left\{ \frac{i_1}{2} + \dots + \frac{i_p}{2p} : i_1, \dots, i_p \in \{0, \dots, 2p\}, i_1 + \dots + i_p \leq 2p, \frac{i_1}{2} + \dots + \frac{i_p}{2p} > 1 \right\} > 1.$$

So that:

$$\begin{aligned} & \left| \frac{1}{t} \left[f(x^* + t^{1/2} p_{E_1}(h) + \dots + t^{1/(2p)} p_{E_p}(h)) - f(x^*) \right] - g_1(h) \right| \\ & \leq t^{a-1} |g_2(h)| + \frac{\varepsilon}{t} \|t^{1/2} p_{E_1}(h) + \dots + t^{1/(2p)} p_{E_p}(h)\|^{2p}. \end{aligned} \quad (26)$$

We remark that $h \mapsto g_2(h)$ is a polynomial function so is bounded on every compact set. We also have:

$$\frac{\varepsilon}{t} \|t^{1/2} p_{E_1}(h) + \dots + t^{1/(2p)} p_{E_p}(h)\|^{2p} \leq \frac{\varepsilon (t^{1/(2p)})^{2p}}{t} \|h\|^{2p} = \varepsilon \|h\|^{2p}.$$

So (26) converges to 0 as $t \rightarrow 0$, uniformly with respect to h on every compact set.

Now let us assume that $f \in \mathcal{A}_p^*(x^*)$; we prove that the function g defined in (10) is not constant in any of its variables in the sense of (7). Let $B \in \mathcal{O}_d(\mathbb{R})$ adapted to the decomposition $\mathbb{R}^d = E_1 \oplus \dots \oplus E_p$. We have:

$$\frac{1}{t} [f(x^* + B \cdot (t^\alpha * h)) - f(x^*)] \xrightarrow{t \rightarrow 0} g(h).$$

Let $i \in \{1, \dots, p\}$ and k such that $v_i := B \cdot e_i \in E_k$. Let us assume by contradiction that g does not depend on the i^{th} coordinate. Considering the expression of g in (10) and setting all the variables outside E_k to 0, we have:

$$\forall h \in E_k, \lambda \in \mathbb{R} \mapsto T_{2k} \cdot (h + \lambda v_i)^{\otimes 2k}$$

is constant. Then applying (2), we have:

$$\forall h \in E_k, T_{2k} \cdot v_i \otimes h^{\otimes 2k-1} = 0.$$

Moreover, for $h \in F_{k-1}$, let us write $h = h' + h''$ where $h' \in E_k$ and $h'' \in F_k$, so that

$$T_{2k} \cdot v_i \otimes h^{\otimes 2k-1} = T_{2k} \cdot v_i \otimes h'^{\otimes 2k-1} = 0,$$

where we used that

$$\forall h^{(3)} \in F_{k-1}, T_{2k} \cdot h'' \otimes (h^{(3)})^{\otimes 2k-1} = 0$$

following (19), and Proposition 6. Considering the definition of E_k as the orthogonal complement of F_k , which is defined in (19), the last equation contradicts that $v_i \in E_k$. \square

4.10 Non coercive case

The function g we obtain in Algorithm 1 is a non-negative polynomial function which is constant in none of its variables. However, this does not always guarantee that $e^{-g} \in L^1(\mathbb{R}^d)$, or even that g is coercive. Indeed, g can be null on an unbounded continuous polynomial curve, while the polynomial degree of the minimum x^* of f is higher than the degree of g in these variables. For example, let us consider

$$\begin{aligned} f: \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (x, y) &\mapsto (x - y^2)^2 + x^6. \end{aligned} \quad (27)$$

Then $f \in \mathcal{A}_3^*(0)$ and using Algorithm 1, we get

$$g(x, y) = (x - y^2)^2,$$

which does not satisfy $e^{-g} \in L^1(\mathbb{R}^d)$. In fact this case is highly degenerate, as, with

$$f_\varepsilon(x, y) := f(x, y) + \varepsilon xy^2 = x^2 + y^4 - (2 - \varepsilon)xy^2 + x^6,$$

we have that $g_\varepsilon(x, y) = x^2 + y^4 - (2 - \varepsilon)xy^2$ satisfies $e^{-g_\varepsilon} \in L^1(\mathbb{R}^d)$ for every $\varepsilon \in (0, 4)$ and that x^* is not the global minimum of f_ε for every $\varepsilon \in (-\infty, 0) \cup (4, \infty)$.

We now prove that instead of assuming $e^{-g} \in L^1(\mathbb{R}^d)$, we can only assume that g is coercive, which is justified in the following proposition. More specific conditions for g to be coercive can be found in [BS15] and [BS19].

Proposition 3. *Let $g : \mathbb{R}^d \rightarrow \mathbb{R}$ be the polynomial function obtained from Algorithm 1. If g is coercive, then $e^{-g} \in L^1(\mathbb{R}^d)$.*

Proof. Let

$$A_k := \text{Span}(e_i : i \in \{\dim(E_1) + \dots + \dim(E_{k-1}) + 1, \dots, \dim(E_1) + \dots + \dim(E_k)\})$$

for $k \in \{1, \dots, p\}$. By construction of g , note that for all $t \in [0, +\infty)$,

$$g\left(\sum_{k=1}^p t^{1/2k} p_{A_k}(h)\right) = tg(h).$$

Since g is coercive, there exists $R \geq 1$ such that for every h with $\|h\| \geq R$, $g(h) \geq 1$. Then, for every $h \in \mathbb{R}^d$, we have:

$$\begin{aligned} g(h) &= g\left(\sum_{k=1}^p p_{A_k}(h)\right) = g\left(\sum_{k=1}^p \frac{\|h\|^{1/2k}}{R^{1/2k}} p_{A_k}\left(R^{1/2k} \frac{h}{\|h\|^{1/2k}}\right)\right) \\ &= \frac{\|h\|}{R} g\left(\sum_{k=1}^p p_{A_k}\left(R^{1/2k} \frac{h}{\|h\|^{1/2k}}\right)\right). \end{aligned}$$

Then, for $\|h\| \geq R$,

$$\left\|\sum_{k=1}^p p_{A_k}\left(R^{1/2k} \frac{h}{\|h\|^{1/2k}}\right)\right\|^2 = \sum_{k=1}^p \frac{R^{1/k}}{\|h\|^{1/k}} \|p_{A_k}(h)\|^2 \geq \frac{R}{\|h\|} \|h\|^2 = R\|h\| \geq R^2 \geq R,$$

so that $g(h) \geq \frac{\|h\|}{R}$ which in turn implies $e^{-g} \in L^1(\mathbb{R}^d)$. \square

We now deal with the simplest configuration where the function g is not coercive, as described in (28), by dealing with the case where f is given by (27), which is an archetype of such configuration. However, dealing with the general case is more complicated and to give a general formula for the rate of convergence of the measure π_t in this case is not our current objective.

Proposition 4. *Let the function f be given by (27). Then, if $(X_t, Y_t) \sim C_t e^{-f(x,y)/t} dx dy$, we have:*

$$\left(\frac{X_t}{t^{1/6}}, \frac{Y_t^2 - X_t}{t^{1/2}}\right) \xrightarrow{t \rightarrow 0} C \frac{e^{-x^6}}{\sqrt{x}} \frac{e^{-y^2}}{\sqrt{\pi}} \mathbb{1}_{x \geq 0} dx dy,$$

where $C = \left(\int_0^\infty \frac{e^{-x^6}}{\sqrt{x}} dx\right)^{-1}$.

Proof. First, let us consider the normalizing constant C_t . We have :

$$\begin{aligned} C_t^{-1} &= \int_{\mathbb{R}^2} e^{-\frac{(x-y)^2+x^6}{t}} dx dy = 2t^{3/4} \int_{-\infty}^{\infty} e^{-t^2 x^6} \int_0^{\infty} e^{-(y^2-x)^2} dy dx \\ &= t^{3/4} \int_{-\infty}^{\infty} e^{-t^2 x^6} \int_{-x}^{\infty} \frac{e^{-u^2}}{\sqrt{u+x}} dy dx = t^{7/12} \int_{-\infty}^{\infty} e^{-x^6} \int_{-t^{-1/3}x}^{\infty} \frac{e^{-u^2}}{\sqrt{t^{1/3}u+x}} du dx \\ &\underset{t \rightarrow 0}{\sim} t^{7/12} \int_0^{\infty} \frac{e^{-x^6}}{\sqrt{x}} \int_{-\infty}^{\infty} e^{-u^2} du dx, \end{aligned}$$

where the convergence is obtained by dominated convergence and where we performed the change of variables $x' = t^{-1/6}x$ and $u = t^{-1/2}(y^2 - x)$. Then we consider, for $a_1 < b_1$ and $a_2 < b_2$,

$$\mathbb{P} \left(\frac{X_t}{t^{1/6}} \in [a_1, b_1], \frac{Y^2 - X}{t^{1/2}} \in [a_2, b_2] \right).$$

Performing the same changes of variables and using the above equivalent of C_t completes the proof. \square

More generally, if the function g is not coercive and if we can write, up to a change of basis,

$$g(h_1, \dots, h_d) = Q_1(h_1, h_2)^2 + Q_2(h_3, h_4)^2 + \dots + Q_r(h_{2r-1}, h_{2r})^2 + \tilde{g}(h_{2r+1}, \dots, h_d), \quad (28)$$

where the Q_i are polynomials with two variables null on an unbounded curve (for example, $Q_i(x, y) = (x - y^2)$, $Q_i(x, y) = (x^2 - y^3)$, $Q_i(x, y) = x^2 y^2$), and where \tilde{g} is a non-negative coercive polynomial, then

$$\begin{aligned} &\left(a_1((X_t)_1, (X_t)_2, t), \dots, a_r((X_t)_{2r-1}, (X_t)_{2r}, t), \left(\frac{1}{t^{\alpha_{2r+1}}}, \dots, \frac{1}{t^{\alpha_d}} \right) * \left(\tilde{B} \cdot ((X_t)_{2r+1}, \dots, (X_t)_d) \right) \right) \\ &\xrightarrow{t \rightarrow 0} b_1(x_1, x_2) \dots b_r(x_{2r-1}, x_{2r}) C e^{-\tilde{g}(x_{2r+1}, \dots, x_d)} dx_1 \dots dx_{2r} dx_{2r+1} \dots dx_d, \end{aligned}$$

where C is a normalization constant, $\tilde{B} \in \mathcal{O}_{d-2r-1}(\mathbb{R})$ is an orthogonal transformation and for all $k = 1, \dots, r$, $a_k : \mathbb{R}^2 \times (0, +\infty) \rightarrow \mathbb{R}^2$ and b_k is a density on \mathbb{R}^2 . Such a_k and b_k can be obtained by applying the same method as in Proposition 4. Algorithm 1 yields the first change of variable for this method, given by the exponents (α_i) (in the proof of Proposition 4, the first change of variable is $t^{-1/2}x$ and $t^{-1/4}y$) and thus seems to be the first step of a more general procedure in this case. However, we do not give a general formula as the general case is cumbersome. Moreover, we do not give a method where the non coercive polynomials Q_i depend on more than two variables, like

$$Q(x, y, z) = (x - y^2)^2 + (x - z^2)^2.$$

The method sketched in Proposition 4 cannot be directly applied to this case.

5 Proofs of Theorem 4 and Theorem 6 using Theorem 7

5.1 Single well case

We now prove Theorem 4.

Proof. Using Theorem 7, we have for all $h \in \mathbb{R}^d$:

$$\frac{1}{t} f(B \cdot (t^\alpha * h)) \xrightarrow{t \rightarrow 0} g(h).$$

To simplify the notations, assume that there is no need of a change of basis i.e. $B = I_d$. We want to apply Theorem 3 to the function f . However the condition

$$\int_{\mathbb{R}^d} \sup_{0 < t < 1} e^{-\frac{f(t^{\alpha_1} h_1, \dots, t^{\alpha_d} h_d)}{t}} dh_1 \dots dh_d < \infty$$

is not necessarily true. Instead, let $\varepsilon > 0$ and we apply Theorem 3 to \tilde{f} , where \tilde{f} is defined as:

$$\tilde{f}(h) = \begin{cases} f(h) & \text{if } h \in \mathcal{B}(0, \delta) \\ \|h\|^2 & \text{else,} \end{cases}$$

and where $\delta > 0$ will be fixed later. Then \tilde{f} satisfies the hypotheses of Theorem 3. The only difficult point to prove is the last condition of Theorem 3. If $t \in (0, 1]$ and $h \in \mathbb{R}^d$ are such that $(t^{\alpha_1}h_1, \dots, t^{\alpha_d}h_d) \notin \mathcal{B}(0, \delta)$, then

$$\frac{\tilde{f}(t^{\alpha_1}h_1, \dots, t^{\alpha_d}h_d)}{t} = \frac{\|(t^{\alpha_1}h_1, \dots, t^{\alpha_d}h_d)\|^2}{t} \geq \|h\|^2,$$

because for all i , $\alpha_i \leq \frac{1}{2}$. If t and h are such that $(t^{\alpha_1}h_1, \dots, t^{\alpha_d}h_d) \in \mathcal{B}(0, \delta)$, then choosing δ such that for all $(t^{\alpha_1}h_1, \dots, t^{\alpha_d}h_d) \in \mathcal{B}(0, \delta)$,

$$\left| \frac{f(t^{\alpha_1}h_1, \dots, t^{\alpha_d}h_d)}{t} - g(h) \right| \leq \varepsilon,$$

which is possible because of the uniform convergence on every compact set (see Section 4.9), we derive that

$$\frac{f(t^{\alpha_1}h_1, \dots, t^{\alpha_d}h_d)}{t} \geq g(h) - \varepsilon.$$

Hence

$$\int_{\mathbb{R}^d} \sup_{0 < t < 1} e^{-\frac{\tilde{f}(t^{\alpha_1}h_1, \dots, t^{\alpha_d}h_d)}{t}} dh_1 \dots dh_d \leq \int_{\mathbb{R}^d} e^{-\|h\|^2} dh + e^\varepsilon \int_{\mathbb{R}^d} e^{-g(h)} dh.$$

Since g is coercive, using Proposition 3 we have $e^{-g} \in L^1(\mathbb{R}^d)$ and it follows from Theorem 3 that if \tilde{X}_t has density $\tilde{\pi}_t(x) := \tilde{C}_t e^{-\tilde{f}(x)/t}$, then

$$\left(\frac{(\tilde{X}_t)_1}{t^{\alpha_1}}, \dots, \frac{(\tilde{X}_t)_d}{t^{\alpha_d}} \right) \xrightarrow{\mathcal{L}} X \text{ as } t \rightarrow 0,$$

where X has density proportional to $e^{-g(x)}$.

Now, let us prove that if X_t has density proportional to $e^{-f(x)/t}$, then we also have

$$\left(\frac{(X_t)_1}{t^{\alpha_1}}, \dots, \frac{(X_t)_d}{t^{\alpha_d}} \right) \xrightarrow{\mathcal{L}} X \text{ as } t \rightarrow 0. \quad (29)$$

Let $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ be continuous with compact support. Then

$$\begin{aligned} & \mathbb{E} \left[\varphi \left(\frac{(X_t)_1}{t^{\alpha_1}}, \dots, \frac{(X_t)_d}{t^{\alpha_d}} \right) - \varphi \left(\frac{(\tilde{X}_t)_1}{t^{\alpha_1}}, \dots, \frac{(\tilde{X}_t)_d}{t^{\alpha_d}} \right) \right] \\ &= \int_{\mathbb{R}^d} \varphi \left(\frac{x_1}{t^{\alpha_1}}, \dots, \frac{x_d}{t^{\alpha_d}} \right) \left(C_t e^{-\frac{f(x_1, \dots, x_d)}{t}} - \tilde{C}_t e^{-\frac{\tilde{f}(x_1, \dots, x_d)}{t}} \right) dx_1 \dots dx_d =: I_1 + I_2, \end{aligned}$$

where I_1 is the integral on the set $\mathcal{B}(0, \delta)$ and I_2 on $\mathcal{B}(0, \delta)^c$. We have then:

$$|I_2| \leq \|\varphi\|_\infty (\pi_t(\mathcal{B}(0, \delta)^c) + \tilde{\pi}_t(\mathcal{B}(0, \delta)^c)) \xrightarrow{t \rightarrow 0} 0,$$

where we used Proposition 1. On the other hand, we have $f = \tilde{f}$ on $\mathcal{B}(0, \delta)$, so that

$$|I_1| \leq \|\varphi\|_\infty |C_t - \tilde{C}_t| \int_{\mathcal{B}(0, \delta)} e^{-\frac{f(x)}{t}} dx \leq \|\varphi\|_\infty \left| 1 - \frac{\tilde{C}_t}{C_t} \right|.$$

And we have:

$$\frac{\tilde{C}_t}{C_t} = \frac{\int e^{-\frac{f(x)}{t}} dx}{\int e^{-\frac{\tilde{f}(x)}{t}} dx} = \frac{\int_{\mathcal{B}(0,\delta)} e^{-\frac{f(x)}{t}} dx + \int_{\mathcal{B}(0,\delta)^c} e^{-\frac{f(x)}{t}} dx}{\int_{\mathcal{B}(0,\delta)} e^{-\frac{\tilde{f}(x)}{t}} dx + \int_{\mathcal{B}(0,\delta)^c} e^{-\frac{\tilde{f}(x)}{t}} dx}.$$

By Proposition 1, we have when $t \rightarrow 0$

$$\begin{aligned} \int_{\mathcal{B}(0,\delta)^c} e^{-\frac{\tilde{f}(x)}{t}} dx &= o\left(\int_{\mathcal{B}(0,\delta)} e^{-\frac{\tilde{f}(x)}{t}} dx\right) \\ \int_{\mathcal{B}(0,\delta)^c} e^{-\frac{f(x)}{t}} dx &= o\left(\int_{\mathcal{B}(0,\delta)} e^{-\frac{f(x)}{t}} dx\right), \end{aligned}$$

so that $\tilde{C}_t/C_t \rightarrow 1$, so $I_1 \rightarrow 0$, which then implies (29). \square

5.2 Multiple well case

We now prove Theorem 6.

Proof. The first point is a direct application of Theorem 5. For the second point, we remark that X_{it} has a density proportional to $e^{-f_i(x)/t}$, where

$$f_i(x) := \begin{cases} f(x) & \text{if } x \in \mathcal{B}(x_i, \delta) \\ +\infty & \text{else.} \end{cases}$$

We then consider \tilde{f}_i as in Section 5.1:

$$\tilde{f}_i(x) = \begin{cases} f_i(x) & \text{if } x \in \mathcal{B}(x_i, \delta) \\ \|x\|^2 & \text{else.} \end{cases}$$

and still as in Section 5.1, we apply Theorem 3 to \tilde{f}_i and then prove that random variables with densities proportional to $e^{-\tilde{f}_i(x)/t}$ and $e^{-f_i(x)/t}$ respectively have the same limit in law. \square

6 Infinitely flat minimum

In this section, we deal with an example of infinitely flat global minimum, where we cannot use a Taylor expansion.

Proposition 5. *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ such that*

$$\forall x \in \mathcal{B}(0, 1), f(x) = e^{-\frac{1}{\|x\|^2}}$$

and

$$\forall x \notin \mathcal{B}(0, 1), f(x) > a$$

for some $a > 0$. Furthermore, assume that f is coercive and $e^{-f} \in L^1(\mathbb{R}^d)$. Then, if X_t has density π_t ,

$$\log^{1/2}\left(\frac{1}{t}\right) \cdot X_t \xrightarrow{\mathcal{L}} X \quad \text{as } t \rightarrow 0,$$

where $X \sim \mathcal{U}(\mathcal{B}(0, 1))$.

Proof. Noting that $\int_{\|x\|>1} e^{-f(x)/t} dx \rightarrow 0$ as $t \rightarrow 0$ by dominated convergence, we have

$$C_t \underset{t \rightarrow 0}{\sim} \left(\int_{\mathcal{B}(0,1)} e^{-e^{-\frac{1}{\|x\|^2}}/t} dx \right)^{-1} = \log^{d/2} \left(\frac{1}{t} \right) \left(\underbrace{\int_{\mathcal{B}(0, \sqrt{\log(1/t)})} e^{-t \frac{1}{\|x\|^2 - 1}} dx}_{\xrightarrow{t \rightarrow 0} \text{Vol}(\mathcal{B}(0,1))} \right)^{-1},$$

where the convergence of the integral is obtained by dominated convergence. Then we have, for $-1 < a_i < b_i < 1$ and $\sum_i a_i^2 < 1$, $\sum_i b_i^2 < 1$:

$$\mathbb{P} \left(\log^{1/2} \left(\frac{1}{t} \right) \cdot X_t \in \prod_{i=1}^d [a_i, b_i] \right) = \frac{C_t}{\log^{d/2} \left(\frac{1}{t} \right)} \int_{(a_i)}^{(b_i)} e^{-t \frac{1}{|x|^2} - 1} dx \xrightarrow{t \rightarrow 0} \frac{\prod_{i=1}^d (b_i - a_i)}{\text{Vol}(\mathcal{B}(0, 1))}.$$

□

Acknowledgements

I would like to thank Gilles Pagès for insightful discussions.

References

- [AH10] Krishna B. Athreya and Chii-Ruey Hwang. Gibbs measures asymptotics. *Sankhya A*, 72(1):191–207, February 2010.
- [Bar20] Gerardo Barrera. Limit behavior of the invariant measure for Langevin dynamics. *arXiv e-prints*, page arXiv:2006.06808, 2020.
- [BS15] Tomas Bajbar and Oliver Stein. Coercive Polynomials and Their Newton Polytopes. *SIAM Journal on Optimization*, 25:1542–1570, January 2015.
- [BS19] Tomas Bajbar and Oliver Stein. Coercive polynomials: stability, order of growth, and Newton polytopes. *Optimization*, 68(1):99–124, 2019.
- [CGLM08] Pierre Comon, Gene Golub, Lek-Heng Lim, and Bernard Mourrain. Symmetric tensor and symmetric tensor rank. *SIAM Journal on matrix analysis and applications*, 2008.
- [Dal14] Arnak S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *arXiv e-prints*, page arXiv:1412.7392, 2014.
- [DPG⁺14] Yann Dauphin, Razvan Pascanu, Caglar Gulcehre, Kyunghyun Cho, Surya Ganguli, and Yoshua Bengio. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. *arXiv e-prints*, page arXiv:1406.2572, 2014.
- [FP99] Jean-Claude Fort and Gilles Pagès. Asymptotic Behavior of a Markovian Stochastic Algorithm with Constant Step. *SIAM Journal on Control and Optimization*, 37(5):1456–1482, 1999.
- [GM90] Saul B. Gelfand and Sanjoy K. Mitter. Recursive stochastic algorithms for global optimization in \mathbb{R}^d . In *29th IEEE Conference on Decision and Control*, pages 220–221 vol.1, 1990.
- [Hil88] David Hilbert. Ueber die Darstellung definiter Formen als Summe von Formenquadraten. *Mathematische Annalen*, 1888.

- [Hwa80] Chii-Ruey Hwang. Laplace’s method revisited: Weak convergence of probability measures. *Ann. Probab.*, 8(6):1177–1182, December 1980.
- [Hwa81] Chii-Ruey Hwang. A Generalization of Laplace’s Method. *Proceedings of the American Mathematical Society*, 82(3):446–451, 1981.
- [Kha12] Rafail Khasminskii. *Stochastic stability of differential equations. With contributions by G. N. Milstein and M. B. Nevelson. 2nd completely revised and enlarged ed*, volume 66. Springer-Verlag Berlin Heidelberg, 2012.
- [LA87] Peter J. M. Laarhoven and Emile H. L. Aarts. *Simulated Annealing: Theory and Applications*. Kluwer Academic Publishers, USA, 1987.
- [Laz92] Vladimir A. Lazarev. Convergence of Stochastic-Approximation procedures in the case of a Regression Equation with Several Roots. *Probl. Peredachi Inf.*, 28(1):75–88, 1992.
- [LCCC15] Chunyuan Li, Changyou Chen, David Carlson, and Lawrence Carin. Preconditioned Stochastic Gradient Langevin Dynamics for Deep Neural Networks. *arXiv e-prints*, page arXiv:1512.07666, 2015.
- [Mot67] Theodore S. Motzkin. The arithmetic-geometric inequality. In *Inequalities (Proc. Sympos. Wright-Patterson Air Force Base, Ohio, 1965)*, pages 205–224. Academic Press, New York, 1967.
- [WT11] Max Welling and Yee Whye Teh. Bayesian Learning via Stochastic Gradient Langevin Dynamics. In *Proceedings of the 28th International Conference on International Conference on Machine Learning, ICML’11*, page 681–688. Omnipress, 2011.

A Properties of tensors

Proposition 6. *Let T_k be a symmetric tensor of order k in \mathbb{R}^d . Let E be a subspace of \mathbb{R}^d . Assume that*

$$\forall h \in E, T_k \cdot h^{\otimes k} = 0.$$

Then we have

$$\forall h_1, \dots, h_k \in E, T_k \cdot h_1 \otimes \dots \otimes h_k = 0.$$

Proof. Using (2), we have for $h_1, \dots, h_k \in E$ and $\lambda_1, \dots, \lambda_k \in \mathbb{R}$,

$$T_k \cdot (\lambda_1 h_1 + \dots + \lambda_k h_k)^{\otimes k} = \sum_{i_1 + \dots + i_k = k} \binom{k}{i_1, \dots, i_k} \lambda_1^{i_1} \dots \lambda_k^{i_k} T_k \cdot h_1^{\otimes i_1} \otimes \dots \otimes h_k^{\otimes i_k} = 0,$$

which is an identically null polynomial in the variables $\lambda_1, \dots, \lambda_k$, so every coefficient is null, in particular

$$\forall h_1, \dots, h_k \in E, T_k \cdot h_1 \otimes \dots \otimes h_k = 0.$$

□