Log-linear Convergence of the Scale-invariant $(\mu/\mu\_w, \lambda)$-ES and Optimal mu for Intermediate Recombination for Large Population Sizes

Mohamed Jebalia, Anne Auger

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Log-linear Convergence of the Scale-invariant $(\mu/\mu_w, \lambda)$-ES and Optimal μ for Intermediate Recombination for Large Population Sizes

Mohamed Jebalia — Anne Auger

N° 7275

Mai 2010

Optimization, Learning and Statistical Methods

ISSN 0249-6399 ISRN INRIA/RR--7275--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
SACLAY - ÎLE-DE-FRANCE

Log-linear Convergence of the Scale-invariant $(\mu/\mu_w, \lambda)$-ES and Optimal μ for Intermediate Recombination for Large Population Sizes

Mohamed Jebalia*, Anne Auger*

Theme : Optimization, Learning and Statistical Methods
Applied Mathematics, Computation and Simulation
Équipe-Projet TAO

Rapport de recherche n° 7275 — Mai 2010 — 39 pages

Abstract: Evolution Strategies (ESs) are population-based methods well suited for parallelization. In this report, we study the convergence of the $(\mu/\mu_w, \lambda)$-ES, an ES with weighted recombination, and derive its optimal convergence rate and optimal μ especially for large population sizes. First, we theoretically prove the log-linear convergence of the algorithm using a scale-invariant adaptation rule for the step-size and minimizing spherical objective functions and identify its convergence rate as the expectation of an underlying random variable. Then, using Monte-Carlo computations of the convergence rate in the case of equal weights, we derive optimal values for μ that we compare with previously proposed rules. Our numerical computations show also a dependency of the optimal convergence rate in $\ln(\lambda)$ in agreement with previous theoretical results.

Key-words: Evolution Strategies, Numerical Optimization, Log-linear Convergence, Weighted Recombination, Intermediate Recombination, Selection

* firstname.lastname@inria.fr

Centre de recherche INRIA Saclay – Île-de-France
Parc Orsay Université
4, rue Jacques Monod, 91893 ORSAY Cedex
Téléphone : +33 1 72 92 59 00

Convergence log-linéaire du scale-invariant $(\mu/\mu_w, \lambda)$-ES et μ optimaux pour la recombinaison à poids égaux pour de grande tailles de population

Résumé : Les Stratégies d'Évolution (SE) sont des méthodes à base de populations adaptées à la parallélisation. Dans ce rapport, on étudie le SE avec recombinaison, $(\mu/\mu_w, \lambda)$-ES, particulièrement dans le cas de grandes tailles de population. Nous prouvous théroriquement le comportement log-linéaire de l'algorithme lors de la minimization d'une fonction sphérique et identifions la vitesse de convergence relative à cet algorithme comme étant l'espérance d'une certaine variable aléatoire. En utilisant des échantillonnages de Monte-Carlo pour calculer les vitesses de convergence dans le cas de poids égaux, nous déterminons les valeurs optimales de μ et proposons une nouvelle formule pour choisir μ pour de très grandes valeurs de λ. Cette règle est comparée avec des règles proposées précédemment dans d'autres études. Les calculs numériques montrent aussi une dépendence de la vitesse de convergence en $\ln(\lambda)$ ce qui rejoint des résultats théroriques précédentes.

Mots-clés : Stratégies d'Évolution, Optimisation Numérique, Convergence log-linéaire, Recombinaison, Poids de recombinaison, Sélection

1 Introduction

Evolution Strategies (ESs) are robust stochastic search methods [2, 4] for solving continuous optimization problems where the goal is to minimize[1] a real valued objective function f defined on an open subset of $\mathbb{R}^d$. At each iteration of an ES, new solutions are in general generated by adding Gaussian perturbations (mutations) to some (optionally recombined) current ones. These Gaussian mutations are parameterized by the step-size giving the general scale of the search, and the covariance matrix giving the principal directions of the Gaussian distribution. In state-of-the art ESs, these parameters are adapted at each iteration [1, 2, 4, 5]. We focus on isotropic ESs where the step-size is adapted and the covariance matrix is kept equal to the identity matrix I_d and therefore the search distribution is spherical. Adaptation in ESs allows them to have a log-linear behavior (convergence or divergence) when minimizing spherical objective functions [11, 13, 6, 8]. Log-linear convergence (resp. divergence) means that there exists a constant value $c < 0$ called convergence rate (resp. $c > 0$) such that the distance to the optimum, d_n, at an iteration n satisfies $\lim_n \frac{1}{n} \ln(d_n) = c$. Spherical objective functions are defined as

$$f(x) = g(\|x\|), \tag{1}$$

where $g : [0, \infty[\mapsto \mathbb{R}$ is a strictly increasing function, $x \in \mathbb{R}^d$ and $\|.\|$ denotes the Euclidean norm on $\mathbb{R}^d$. Log-linear behavior holds also when minimizng spherical functions perturbed by noise [12].

In this report, we investigate ESs with weighted recombination, denoted $(\mu/\mu_w, \lambda)$-ES, and used in the state-of-the-art ES, the Covariance Matrix Adaptation-ES (CMA-ES) [5]. The $(\mu/\mu_w, \lambda)$-ES is an ES which evolves a single solution. Let $\mathbf{X}_n$ be the solution (the parent) at iteration n, λ new solutions $\mathbf{Y}_n^i$ (offspring) are then generated using independent Gaussian samplings of mean $\mathbf{X}_n$. Then, the offspring are evaluated, the μ best offspring $(\mathbf{Y}_n^{i:\lambda})_{1 \leq i \leq \mu}$ are selected and the new solution $\mathbf{X}_{n+1}$ is obtained by recombining these selected offspring using recombination weights denoted $\left(w^i\right)_{1 \leq i \leq \mu}$, i.e., $\mathbf{X}_{n+1} = \sum_{i=1}^{\mu} w^i \mathbf{Y}_n^{i:\lambda}$ [2]. We will specifically study the $(\mu/\mu_w, \lambda)$-ES with large (offspring) population size λ compared to the search space dimension d, i.e., $\lambda \gg d$. This is motivated by the increasing possibilities of parallelization with the raise of the number of parallel machines, supercomputers and grids. ESs are population-based methods and then are well suited for parallelization which consists in distributing the number of evaluations λ on the processes available. The performance of the $(\mu/\mu_w, \lambda)$-ES as a function of λ has been theoretically investigated [14, 16]. Under the approximation $d \to +\infty$, the study in [14] investigated the $(\mu/\mu_w, \lambda)$-ES minimizing any spherical function and using an artificial step-size adaptation rule termed scale-invariant which sets the step-size at each iteration proportionally to the distance of the current solution to the optimum. The progress rate φ which measures the one-step expected progress to the optimum verifies $\varphi = O\left(\mu \ln\left(\frac{\lambda}{\mu}\right)\right)$ [14]. This suggests that, if μ is chosen proportional to λ, the progress rate of the $(\mu/\mu_w, \lambda)$-ES can be linear in μ and in λ. The study in [16] is based on a theoretical computations of lower bounds for the convergence ratio which measures the convergence rate in probability of wide classes of ESs. It shows that the convergence ratio of the $(\mu/\mu_w, \lambda)$-ES varies at best linearly with $\ln(\lambda)$ for sufficiently large λ when minimizing any spher-

[1] Without loss of generality, the minimization of a real value function f is equivalent to the maximization of $-f$.

[2] If $\mu = 1$, only the best offspring is taken and then the $(\mu/\mu_w, \lambda)$-ES is simply the $(1, \lambda)$-ES.

ical function [16]. This suggests that the bound found in [14] is not tight for finite dimensions.

A natural question arising when using recombination is how to choose the number of offspring μ to be recombined. Studies based on computations of the progress rate when the search space dimension goes to infinity suggest to use $\mu = \lfloor \frac{\lambda}{4} \rfloor$ [14] or $\mu = \lfloor \frac{\lambda}{2} \rfloor$ [7][3] for two different choices of the (positive) weights $\left(w^i\right)_{1 \leq i \leq \mu}$. CMA-ES which has been designed to work well on small population sizes uses $\mu = \lfloor \frac{\lambda}{2} \rfloor$ as a default parameter. However, when using a large population size λ, the convergence rate of some real-world algorithms tested in [15, 9] using the rules $\mu = \lfloor \frac{\lambda}{4} \rfloor$ or $\mu = \lfloor \frac{\lambda}{2} \rfloor$ as recommended in [14, 7] is worse than the theoretical prediction of [16]. This is due to the fact that the rules used in these tests for choosing μ, are recommended by the studies performed under the approximation $(d \to +\infty)$ [14, 7] and thus under the assumption $\lambda \ll d$. For some values of λ and d such that $\lambda \gg d$, Beyer [18] computed, using some approximations permitted by the assumption $(d \to +\infty)$, optimal choices for μ when minimizing spherical functions. However, no explicit rule for the choice of μ has been proposed when $\lambda \gg d$. Performing experiments with $\lambda \gg d$ on a $(\mu/\mu_w, \lambda)$-ES using equal weights, the so-called self-adaptation rule for the step-size and two variants for the covariance matrix adaptation, Teytaud [10] proposed to choose μ equal to $\min\{d, \lfloor \frac{\lambda}{4} \rfloor\}$.

Since it is in general difficult to appraise whether the effect observed when changing the setting of one parameter on a real algorithm is coming from the fact that the setting of an other parameter may subsequently becomes sub-optimal, we want here to identify independently of any real step-size or covariance matrix update rule the optimal setting for μ especially for lagre λ. This optimal setting can be used to identify a rule for choosing best optimal values μ in real-world algorithms like CMA-ES. We want also to verify whether an optimal choice for μ allows to have a dependency of the convergence rate in $\ln(\lambda)$ and thus reach the lower bounds predicted by [16]. In order to do so, we perform in this report a theoretical and numerical investigation of the convergence and the optimal choice for μ relative to the isotropic $(\mu/\mu_w, \lambda)$-ES. We focus on large population sizes. The objective functions investigated are the spherical functions allowing ESs which do not use recombination to reach optimal convergence rates [6, 8]. In Section 2, we present the mathematical formulation of the algorithm. In Section 3, we identify the optimal step-size adaptation rule of the algorithm when minimizing spherical functions. In Section 4, we theoretically prove the log-linear convergence of the algorithm using the scale-invariant adaptation rule and identify its convergence rate. In Section 5, using Monte-Carlo computations of the convergence rate, optimal μ values and optimal convergence rates are derived for some dimensions and in the specific case of equal weights $(w^i)_{1 \leq i \leq \mu}$. A new rule for choosing μ is proposed based on our results. Proofs of theoretical results are presented in the appendix.

Preliminary Notations In this report $\mathbb{Z}^+$ denotes the set of non-negative integers $\{0, 1, 2, \ldots\}$, $\mathbb{N}$ denotes the set of positive integers $\{1, 2, \ldots\}$ and μ and λ are two positive integers such that $1 \leq \mu \leq \lambda$. The recombination weights w^i, $1 \leq i \leq \mu$ are strictly positive constants which verify $\sum_{i=1}^{\mu} w^i = 1$. The unit vector $(1, 0, \ldots, 0) \in \mathbb{R}^d$ is denoted as e_1. $(\Omega, \mathcal{A}, P)$ is a probability space: Ω is a set, $\mathcal{A}$ a σ-algebra defined on this set and P a probability measure defined on $(\Omega, \mathcal{A})$.

[3] The rule proposed in [7] where negative weights are allowed is rather $\mu = \lambda$, but the study implies that if the weights can be only positive the rule becomes $\mu = \lfloor \frac{\lambda}{2} \rfloor$.

2 Mathematical Formulation of the Isotropic $(\mu/\mu_w, \lambda)$ Evolution Strategy Minimizing Spherical Functions

In this section we will introduce the mathematical formulation of the isotropic $(\mu/\mu_w, \lambda)$-ES for minimizing a spherical function (1). Let $\mathbf{X}_0 \in \mathbb{R}^d$ be the first solution randomly chosen using a law absolutely continuous with respect to the Lebesgue measure. Let σ_0 be a strictly positive variable (possibly) randomly chosen. Let $(\mathbf{N}_n^i)_{i\in[1,\lambda],n\in\mathbb{Z}^+}$, be a sequence of random vectors defined on the probability space $(\Omega, \mathcal{A}, P)$, independent and identically distributed (i.i.d.) with common law the isotropic multivariate normal distribution on $\mathbb{R}^d$ with mean $(0,\ldots,0) \in \mathbb{R}^d$ and covariance matrix identity I_d, which we will simply denote $\mathcal{N}(0, I_d)$. We assume that the sequence $(\mathbf{N}_n^i)_{i\in[1,\lambda],n\in\mathbb{Z}^+}$ is independent of $\mathbf{X}_0$. Let σ_n be the step-size mutation at iteration n such that for all $(i, n) \in [1, \lambda] \times \mathbb{Z}^+$, σ_n and $\mathbf{N}_n^i$ are independent. An offspring $\mathbf{Y}_n^i$ where $i = 1, \ldots, \lambda$ writes as $\mathbf{Y}_n^i := \mathbf{X}_n + \sigma_n \mathbf{N}_n^i$, and its objective function value is $g(\|\mathbf{Y}_n^i\|)$ in our case of minimization of spherical functions. Let $\mathbf{N}_n^{i:\lambda}(\mathbf{X}_n, \sigma_n)$ $(1 \leq i \leq \mu)$ denotes the mutation vector relative to the i^{th} best offspring according to its fitness value. As the function g is increasing, the vectors $\mathbf{N}_n^{i:\lambda}(\mathbf{X}_n, \sigma_n)$ (where, for all i in $\{1,\ldots,\mu\}$, the indices $i:\lambda$ are in $\{1,\ldots,\lambda\}$) verify:

$$\begin{aligned}&\left\|\mathbf{X}_n + \sigma_n \mathbf{N}_n^{1:\lambda}(\mathbf{X}_n, \sigma_n)\right\| \leq \ldots \leq \left\|\mathbf{X}_n + \sigma_n \mathbf{N}_n^{\mu:\lambda}(\mathbf{X}_n, \sigma_n)\right\| \text{ and} \\ &\left\|\mathbf{X}_n + \sigma_n \mathbf{N}_n^{\mu:\lambda}(\mathbf{X}_n, \sigma_n)\right\| \leq \left\|\mathbf{X}_n + \sigma_n \mathcal{N}_n^j\right\| \forall j \in \{1,\ldots,\lambda\}\backslash\{1:\lambda,\ldots,\mu:\lambda\}\,.\end{aligned} \tag{2}$$

Using the fact that $\sum_{i=1}^{\mu} w^i = 1$, the new parent $\mathbf{X}_{n+1} = \sum_{i=1}^{\mu} w^i \mathbf{Y}_n^{i:\lambda}$ can be rewritten as:

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \sigma_n \sum_{i=1}^{\mu} w^i \mathbf{N}_n^{i:\lambda}(\mathbf{X}_n, \sigma_n)\,. \tag{3}$$

In the specific case where the scale-invariant rule is used for the adaptation of $(\sigma_n)_{n\in\mathbb{Z}^+}$, i.e., $\sigma_n = \sigma\|\mathbf{X}_n\|$ (with $\sigma > 0$), the previous equation becomes:

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \sigma\|\mathbf{X}_n\| \sum_{i=1}^{\mu} w^i \mathbf{N}_n^{i:\lambda}(\mathbf{X}_n, \sigma\|\mathbf{X}_n\|)\,. \tag{4}$$

Finally, σ_n is updated, i.e., σ_{n+1} is computed independently of $\mathbf{N}_{n+1}^i$ for all $i \in [1, \lambda]$. Throughout the remainder of this report, we will denote in a general context where $u \in \mathbb{R}^d$, $s \in \mathbb{R}$ and $(\mathbf{N}_n^i)_{i\in[1,\lambda],n\in\mathbb{Z}^+}$ is a sequence of random vectors (i.i.d.) with common law $\mathcal{N}(0, I_d)$ and such that for all $(i, n) \in [1, \lambda] \times \mathbb{Z}^+$, $\mathbf{N}_n^i$ is independent of u and s, $\mathbf{N}_n^{i:\lambda}(u, s)$ the random vector which verifies (2) where $\mathbf{X}_n$ and σ_n are respectively replaced by u and s. For $n = 0$ and $i \in \{1,\ldots,\mu\}$, the notation $\mathbf{N}_0^{i:\lambda}(u, s)$ will be replaced by the notation $\mathbf{N}^{i:\lambda}(u, s)$.

3 Optimal Step-size Adaptation Rule When Minimizing Spherical Functions

The (log-linear) convergence rate of the isotropic scale-invariant $(\mu/\mu_w, \lambda)$-ES minimizing any spherical function and satisfying the recurrence relation (4) is, as will be shown in Section 4, the function V that we will introduce in the following definition.

Definition 1 *Let* e_1 *denotes the unit vector* $(1, 0, \ldots, 0) \in \mathbb{R}^d$. *For* $\sigma \geq 0$, *let* $\mathrm{Z}(\sigma)$ *be the random variable defined as* $\mathrm{Z}(\sigma) := \left\| \mathrm{e}_1 + \sigma \sum_{i=1}^{\mu} w^i \mathbf{N}^{i:\lambda}(\mathrm{e}_1, \sigma) \right\|$ *where the random variables* $\mathbf{N}^{i:\lambda}(\mathrm{e}_1, \sigma)$ *are obtained similarly to (2) but with* $n = 0$ *and* $(\mathbf{X}_n, \sigma_n)$ *replaced by* (e_1, σ). *We introduce the function* V *as the function mapping* $[0, +\infty[$ *into* $\mathbb{R}$ *as follows:*

$$\mathrm{V}(\sigma) := E\left[\ln \mathrm{Z}(\sigma)\right] = E\left[\ln \left\| \mathrm{e}_1 + \sigma \sum_{i=1}^{\mu} w^i \mathbf{N}^{i:\lambda}(\mathrm{e}_1, \sigma) \right\|\right] . \tag{5}$$

Fig. 1 (top, left) represents numerical computations of the function V in some specific settings. In the following proposition, we show that V is well defined and we study its properties. Note that in the following, the notation V will be sometimes replaced by V_μ when we need to stress the dependence of V on μ.

Proposition 1 *The function* V *introduced in (5) has the following properties:*

(i) V *is well defined for* $d \geq 1$, *and continuous for* $d \geq 2$, *on* $[0, +\infty[$.

(ii) *For* $d \geq 2$, $\lim_{\sigma \to +\infty} \mathrm{V}(\sigma) = +\infty$.

(iii) *If* $\mu \leq \frac{\lambda}{2}$, *for* $d \geq 2$, $\exists\, \bar{\sigma} > 0$ *such that* $\mathrm{V}(\bar{\sigma}) < 0$.

(iv) *If* $\mu \leq \frac{\lambda}{2}$, *for* $d \geq 2$, $\exists\, \sigma_{opt} > 0$ *such that* $\inf_{\{\sigma \geq 0\}} \mathrm{V}(\sigma) = \mathrm{V}(\sigma_{opt}) < 0$.

(v) *For* $d \geq 2$ *and* $\lambda \geq 2$, *if* $\mu \leq \lambda/2$, $\exists\, (\sigma_{opt}, \mu_{opt})$ *such that* $\mathrm{V}_{\mu_{opt}}(\sigma_{opt}) = \inf_{\{\sigma \geq 0, \mu \leq \lambda/2\}} \mathrm{V}_\mu(\sigma) < 0$.

Proof: see page 28

Summary of the proof A basic step in the proof of (i) and (ii) is to write V as the sum of $\mathrm{V}^+(\sigma) := E\left[\ln^+ \mathrm{Z}(\sigma)\right]$ and $\mathrm{V}^-(\sigma) := E\left[\ln^- \mathrm{Z}(\sigma)\right]$. Then, for (i), integrands in these quantities are upper bounded by quantities which do not depend on σ and the result follows by the Lebesgue dominated convergence theorem for continuity. For (ii), we show that $\mathrm{V}(\sigma)$ is lower bounded by an expectation of a given random variable which depends on σ. We show using the Monotone convergence theorem that this lower bound converges to infinity when σ goes to infinity and then the result follows. For proving (iii), we prove before, using the concept of uniform integrability of a family of random variables that $d\, \mathrm{V}\left(\frac{\sigma^*}{d}\right)$ ($\sigma^* > 0$ fixed) converges to a certain limit depending on σ^* when d goes to $+\infty$. Using the fact that this limit can be negative for a given σ^* we prove our claim. (iv) is proven using (i), (ii) and (iii) and the intermediate value theorem. (v) follows easily from (iv).

An important point that we can see from this proposition is that, given $\lambda \geq 2$ and $d \geq 2$, and under the condition $\mu \leq \lambda/2$, μ and σ can be chosen such that the relative convergence rate V is optimal (v). We conducted numerical computations of V in the case where $d = 10$, $\lambda = 10$ and equal weights $(w^i)_{1 \leq i \leq \mu}$. The cases with $\mu = 1, 2$ and 5 are represented in Fig. 1 (top, left). It can be seen that the curves are in conformity with (i), (iii), (iv) and (v) of Proposition 1. In particular, for each μ, there exists a σ_{opt} realizing the minimum of V and we can see that the optimal μ (among the represented μ values 1,2 and 5) is 2. In the following theorem, we will see that the optimal value of V is also the optimal convergence rate in expectation that can

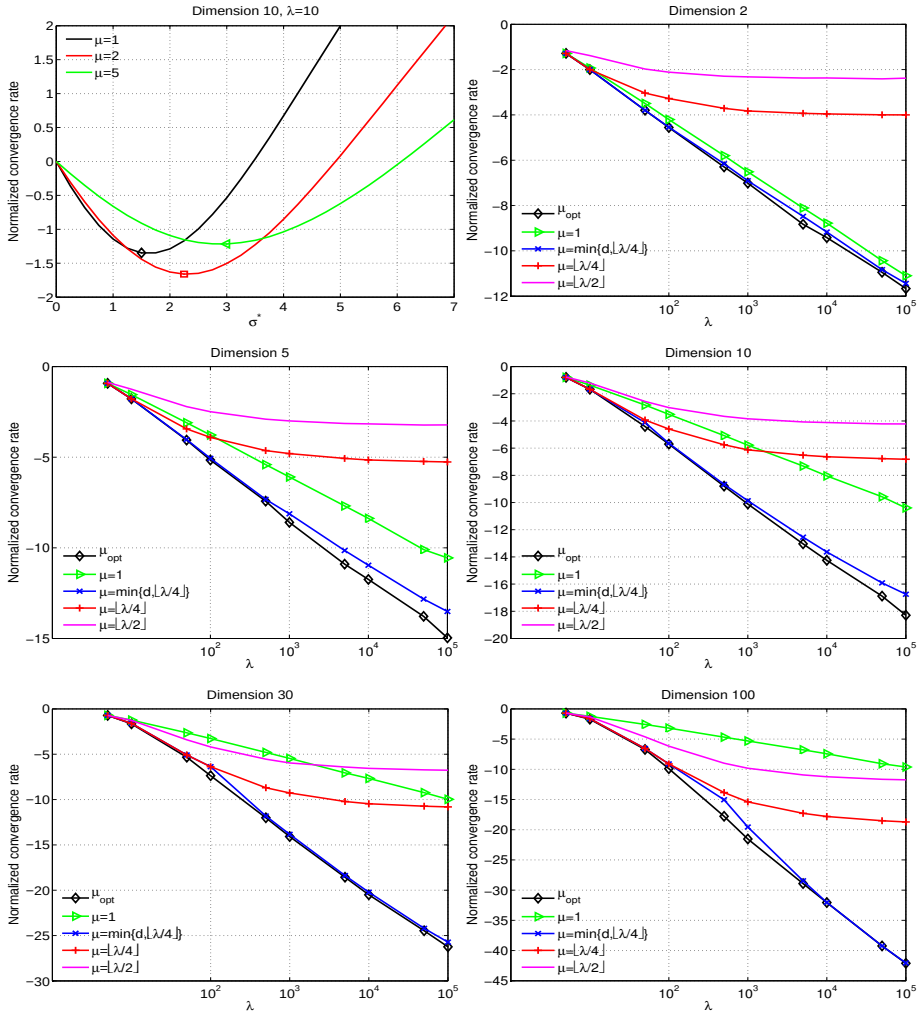

Figure 1: **Top, left:** Plots of the normalized convergence rate $d \times \mathrm{V}_\mu(\frac{\sigma^*}{d})$ where V_μ $(= \mathrm{V})$ is defined in (5) as a function of $\sigma^* > 0$ with $d = 10$, $\lambda = 10$, $w^i = \frac{1}{\mu}$, $\forall\, i = 1, \ldots, \mu$ and $\mu \in \{1, 2, 5\}$. The plots were obtained doing Monte-Carlo estimations of V using 10^6 samples. **Other curves:** Optimal convergence rate $(d \times \mathrm{V}_\mu(\frac{\sigma^*_{\mathrm{opt}}}{d}))$ associated to different choices of μ and μ_{opt} realizing the minimum of $(\sigma^*, \mu) \mapsto \mathrm{V}_\mu(\frac{\sigma^*}{d})$, as a function of λ for dimensions 2, 5, 10, 30, 100.

be reached by the $(\mu/\mu_w, \lambda)$-ES minimizing a spherical function and using any step-size adaptation rule $(\sigma_n)_{n\geq 0}$, or more precisely, the smallest value of $\frac{1}{n} E\left[\ln \frac{\|\mathbf{X}_n\|}{\|\mathbf{X}_0\|}\right]$ that can be reached by the sequence $(\mathbf{X}_n)_{n\geq 0}$ satisfying the recurrence relation (3). This optimal value corresponds also to the smallest value of $\frac{1}{n} E\left[\ln \frac{\|\mathbf{X}_n\|}{\|\mathbf{X}_0\|}\right]$ that can be reached by the isotropic scale-invariant $(\mu/\mu_w, \lambda)$-ES minimizing a spherical function, i.e., where $(\mathbf{X}_n)_{n\geq 0}$ satisfies the recurrence relation (4) with $\sigma = \sigma_{opt}$.

Theorem 1 *Let $(\mathbf{X}_n)_{n\geq 0}$ be the sequence of random vectors satisfying the recurrence relation (3) and relative to the $(\mu/\mu_w, \lambda)$-ES minimizing any spherical function (1).*

Then, for $\lambda \geq 2$ and $d \geq 2$, we have

$$\frac{1}{n} E\left[\ln \frac{\|\mathbf{X}_n\|}{\|\mathbf{X}_0\|}\right] \geq \mathrm{V}(\sigma_{opt}), \tag{6}$$

where σ_{opt} is given in Proposition 1 as $\sigma_{opt} = argmin_{\{\sigma>0\}} \mathrm{V}(\sigma)$ and $\mathrm{V}(\sigma_{opt})$ corresponds to $\frac{1}{n} E\left[\ln\left(\frac{\|\mathbf{X}_n\|}{\|\mathbf{X}_0\|}\right)\right]$ for a $(\mu/\mu_w, \lambda)$-ES using the specific scale-invariant adaptation rule with $\sigma_n = \sigma_{opt}\|\mathbf{X}_n\|$ and minimizing any spherical function (1).

Proof: see page 35

Summary of the proof The first step for proving the theorem is to remark that:

$$E\left[\ln \frac{\|\mathbf{X}_{k+1}\|}{\|\mathbf{X}_k\|}\right] = E\left[E\left[\ln \left\| \frac{\mathbf{X}_k}{\|\mathbf{X}_k\|} + \frac{\sigma_k}{\|\mathbf{X}_k\|} \sum_{i=1}^{\mu} w^i \mathbf{N}_k^{i:\lambda}\left(\frac{\mathbf{X}_k}{\|\mathbf{X}_k\|}, \frac{\sigma_k}{\|\mathbf{X}_k\|}\right)\right\| \mid (\mathbf{X}_k, \sigma_k)\right]\right].$$

By the isotropy of the norm function and of the multivariate normal distribution, the term $\frac{\mathbf{X}_k}{\|\mathbf{X}_k\|}$ in the previous equation can be replaced by e_1. Then $E\left[\ln \frac{\|\mathbf{X}_{k+1}\|}{\|\mathbf{X}_k\|}\right] = E\left[\mathrm{V}\left(\frac{\sigma_k}{\|\mathbf{X}_k\|}\right)\right]$ where $E\left[\mathrm{V}\left(\frac{\sigma_k}{\|\mathbf{X}_k\|}\right)\right]$ is, by Proposition 1, lower bounded by $\mathrm{V}(\sigma_{opt})$. The result follows from summing such inequalities from $k = 0$ to $k = n - 1$.

This theorem states that the artificial scale-invariant adaptation rule with the specific setting $\sigma_n = \sigma_{opt}\|\mathbf{X}_n\|$ is the rule which allows to obtain the best convergence rate of the $(\mu/\mu_w, \lambda)$-ES when minimizing spherical functions. The relative convergence rate is then a tight lower bound that can be reached in this context. Then, for our study on minimization of spherical functions, we will use the $(\mu/\mu_w, \lambda)$-ES with the artificial scale-invariant adaptation rule, i.e., with $\sigma_n = \sigma\|\mathbf{X}_n\|$ where σ is a strictly positive constant. In the specific case where σ equals σ_{opt}, the convergence rate is optimal.

4 Log-Linear Behavior of the Scale-invariant $(\mu/\mu_w, \lambda)$-ES Minimizing Spherical Functions

Log-linear convergence of ESs can be in general shown using the application of different Law of Large Numbers (LLN) such as LLN for independent or orthogonal random variables or LLN for Markov chains. Log-linear behavior has been shown for ESs which do not use recombination [11, 6, 13, 8]. The key idea of the proof is stated in the following proposition.

Proposition 2 *Let $\sigma \geq 0$ and let $(\mathbf{X}_n)_n$ be the sequence of random vectors satisfying the recurrence relation (4). We introduce the sequence of random variables $(\mathrm{Z}_n)_{n\in\mathbb{Z}^+}$ by $\mathrm{Z}_n := \left\| \frac{\mathbf{X}_n}{\|\mathbf{X}_n\|} + \sigma \sum_{i=1}^{\mu} w^i \mathbf{N}_n^{i:\lambda}\left(\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|}, \sigma\right)\right\|$ where $\mathbf{N}_n^{i:\lambda}\left(\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|}, \sigma\right)$ are obtained similarly to (2) but with replacing $(\mathbf{X}_n, \sigma_n)$ by $\left(\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|}, \sigma\right)$. Then for $n \geq 0$, we have*

$$\frac{1}{n} \ln \frac{\|\mathbf{X}_n\|}{\|\mathbf{X}_0\|} = \frac{1}{n} \sum_{k=0}^{n-1} \ln \mathrm{Z}_k \ a.s. \tag{7}$$

Proof: see page 36

Using the isotropy of the norm function and of the multivariate normal distribution, the terms $\ln \mathrm{Z}_k$ appearing in the right hand side of the previous equation are independent identically distributed with a common expectation $\mathrm{V}(\sigma)$ which we have proved to be finite in Proposition 1. The following theorem is then obtained by the application of the LLN for independent identically distributed random variables with a finite expectation to the right hand side of the previous equation.

Theorem 2 (Log-linear Behavior of the Scale-invariant $(\mu/\mu_w, \lambda)$-ES) *The scale-invariant $(\mu/\mu_w, \lambda)$-ES defined in (4) and minimizing any spherical function (1) converges (or diverges) log-linearly in the sense that for $\sigma > 0$ the sequence $(\mathbf{X}_n)_n$ of random vectors given by the recurrence relation (4) verifies*

$$\lim_{n \to +\infty} \frac{1}{n} \ln \|\mathbf{X}_n\| = \mathrm{V}(\sigma) \tag{8}$$

almost surely, where V refers to the quantity defined in (5).

Proof: see page 36

Theorem 2 establishes that, provided that V is non zero, the convergence of the scale-invariant $(\mu/\mu, \lambda)$-ES minimizing any spherical objective function given in (1) is log-linear. This theorem also provides the convergence (or divergence) rate $\mathrm{V}(\sigma)$ of the sequence $\left(\ln\left(\|\mathbf{X}_n\|\right)\right)_n$: If $\mathrm{V}(\sigma) < 0$, the distance to the optimum, $(\|\mathbf{X}_n\|)_{n \geq 0}$, converges log-linearly to zero and if $\mathrm{V}(\sigma) > 0$, the algorithm diverges log-linearly. From Proposition 1, we know that, for all $d \geq 2$, for all $\lambda \geq 2$ and all $\mu \geq 1$ with the condition $\mu \leq \lambda/2$, there exists $\sigma > 0$ such that $\mathrm{V}(\sigma) < 0$ and therefore the algorithm converges. Moreover, by the same proposition, we know that for any $d, \lambda \geq 2$ there is an optimal choice of (σ, μ) such that the optimal convergence rate is reached.

A practical interest of this result is that if someone chooses the optimal value of μ and is able to tune the adaptation rule of his algorithm such that the quantity $\frac{\sigma_n}{\|\mathbf{X}_n\|}$ is (after an adaptation time) stable around the optimal value for σ, a convergence rate close to the optimal convergence rate can be obtained at least for spherical functions. This can be useful especially for choosing μ when the population size λ is large.

The goal is then to compute those optimal values (i.e., μ_{opt} and σ_{opt}) depending on λ and d. Fortunately, another important point of Theorem 2 is that the convergence rate is expressed in terms of the expectation of a given random variable (see Definition 1). Therefore, the convergence rate V can be numerically computed using Monte-Carlo simulations. Numerical computations allowing to derive optimal convergence rate values and relative optimal values of μ will be investigated in the following section.

5 Numerical Experiments

In this section, we numerically compute, for a fixed dimension and λ, values of μ leading to optimal convergence rates. We compare the convergence rate associated to those optimal μ with the ones obtained with previous choices of μ (proportional to $\lfloor \lambda/2 \rfloor, \ldots$). We also investigate how the optimal convergence rate depends on the population size λ in particular for $\lambda \gg d$. The context of our numerical study is

the specific $(\mu/\mu_w, \lambda)$-ES with intermediate recombination, i.e., with equal weights $w^i = \frac{1}{\mu},\ (i = 1, \ldots, \mu)$ which is simply denoted $(\mu/\mu, \lambda)$-ES.

Since V is expressed in terms of expectation of a random variable, we can perform a Monte-Carlo simulation of the normalized convergence rate $d \times \mathrm{V}_\mu \left(\frac{\sigma^*}{d}\right)$ where $\sigma^* > 0$ is called normalized step-size. The values computed are then relative to the scale-invariant $(\mu/\mu, \lambda)$-ES with $\sigma_n = \frac{\sigma^*}{d}\|\mathbf{X}_n\|$ and minimizing a spherical function. Our experimental procedure relies on finding the minimal value of $(\sigma^*, \mu) \mapsto d \times \mathrm{V}_\mu \left(\frac{\sigma^*}{d}\right)$ for μ in a range $\mu_{\mathbf{range}}$ and for values of σ^* taken in a range $\sigma_{\mathbf{range}}$. The minimal value, denoted $d \times \mathrm{V}_{\mu_{opt}} \left(\frac{\sigma^*_{opt}}{d}\right)$, is the normalized optimal convergence rate. However, we will also call 'normalized optimal convergence rate' the minimal value of $\sigma^* \mapsto d \times \mathrm{V}_\mu \left(\frac{\sigma^*}{d}\right)$ for μ fixed which we denote $d \times \mathrm{V}_\mu \left(\frac{\sigma^*_{opt}}{d}\right)$. The difference should be clear within the context.

As a first experiment, we took $\mu_{\mathbf{range}} = \{2^k; k \in \mathbb{Z}^+ \text{and } 2^k \leq \frac{\lambda}{2}\}$ and $\sigma_{\mathbf{range}} = \ln(\mu + 1) * \ln(\lambda) * [0 : 0.1 : 3]$. We experimented discrete values of λ from $\lambda = 5$ to $\lambda = 10^5$ with a number of Monte-Carlo samplings decreasing as a function of λ from 10^4 to 500. These first computations show that for the values of λ and d tested, the approximation

$$\min_{\{\sigma^* \in \sigma_{\mathbf{range}}\}} d \times \mathrm{V}_\mu \left(\frac{\sigma^*}{d}\right) \simeq a(\lambda, d) \ln^2(\mu) + b(\lambda, d) \ln(\mu) + c(\lambda, d) \qquad (9)$$

is reliable (for $\mu > 1$) and we determined numerically the coefficients $a(\lambda, d), b(\lambda, d)$ and $c(\lambda, d)$. Using these quadratic approximations, we performed a second serie of tests where the values of μ were taken around the optimal value of the polynomial approximation, $\sigma_{\mathbf{range}} = m * \ln(\mu + 1) * \ln(\lambda) * [0 : 0.1 : 3]$ (with $m \leq \frac{2}{3}$) and using more Monte-Carlo samplings.

In Fig. 2 (left), we plotted the normalized optimal convergence rate values and the optimal normalized convergence rates relative to the rule $\mu = \min\{\lfloor \frac{\lambda}{4} \rfloor, d\}$ from [10] as a function of λ and for different dimensions. It can be seen that the optimal convergence rate is, for λ sufficiently large, linear as a function of $\ln(\lambda)$. This result is in agreement with the results in [16]. This figure shows also that the rule $\mu = \min\{\lfloor \frac{\lambda}{4} \rfloor, d\}$ provides convergence rates very close to optimal ones. The curves in Fig. 2 (left) are smooth. However, to obtain the exact optimal values of μ (denoted μ_{opt}), we would need a very large number of Monte-Carlo samplings and (in parallel) a very small discretisation in σ^* that is not affordable. Therefore, we plotted in Fig. 2 (right), the ranges of μ values giving the optimal convergence rate up to a precision of 0.2, as a function of λ and for dimensions $d = 2, 10, 30$ and 100. Those ranges are called 0.2-confidence intervals in μ in the sequel. In the same graph, we plotted values of μ computed as the argmin of the polynomial approximation (9) that we denote μ_{th}. It can be seen that μ_{th} values are in the 0.2-confidence interval in μ. However, the values $\mu = \min\{\lfloor \frac{\lambda}{4} \rfloor, d\}$ for $\lambda = 10^4$ and $d \in \{10, 30, 100\}$, are not in the 0.2-confidence interval in μ. In Figure 1, we compare, for dimensions 2, 5, 10, 30 and 100, optimal convergence rates for different choices of μ, namely $\mu = 1$, $\lfloor \frac{\lambda}{4} \rfloor$ ([14]), $\lfloor \frac{\lambda}{2} \rfloor$ ([7]), $\min\{\lfloor \frac{\lambda}{4} \rfloor, d\}$ ([10]) and the optimal rule (i.e., μ_{opt} values). We observe, for all the dimensions tested, that for μ equal $\lfloor \frac{\lambda}{4} \rfloor$ and $\lfloor \frac{\lambda}{2} \rfloor$, the convergences rate do not scale linearly in $\ln(\lambda)$ and are then sub-optimal. For $\mu = 1$ and $\min\{\lfloor \frac{\lambda}{4} \rfloor, d\}$, the scaling is linear in $\ln(\lambda)$ and close to the optimal convergence rate for $\mu = \min\{\lfloor \frac{\lambda}{4} \rfloor, d\}$ for all the dimensions tested.

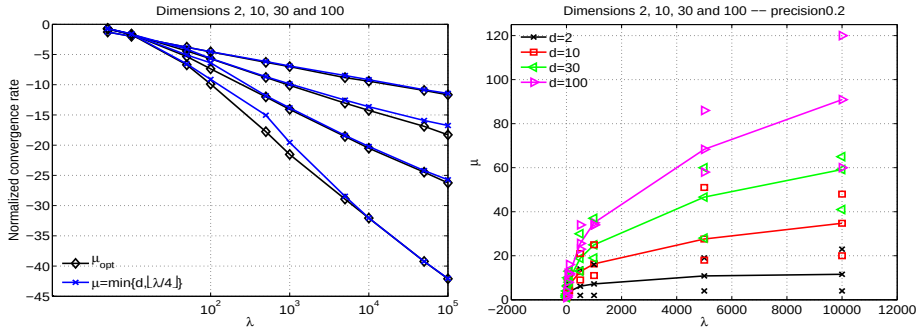

Figure 2: **Left:** Plots of the normalized optimal convergence rate $d \times \mathrm{V}_{\mu_{opt}}\left(\frac{\sigma^*_{opt}}{d}\right)$ where $V_\mu(=\mathrm{V})$ is defined in (5) and optimal convergence rate $d \times \mathrm{V}_{\mu}\left(\frac{\sigma^*_{opt}}{d}\right)$ relative to the rule $\mu = \min\{\lfloor \frac{\lambda}{4} \rfloor, d\}$, as a function of λ (log-scale for λ) for dimensions 2, 10, 30 and 100 (from top to bottom). **Right:** Plots of the values μ_{th} (solid lines with markers) giving the optimal μ relative to the quadratic approximation (9) together with extremity of range of μ values (shown with markers) giving convergence rates up to a precision of 0.2 from the optimal numerical value. The dimensions represented are 2, 10, 30 and 100 (from bottom to top).

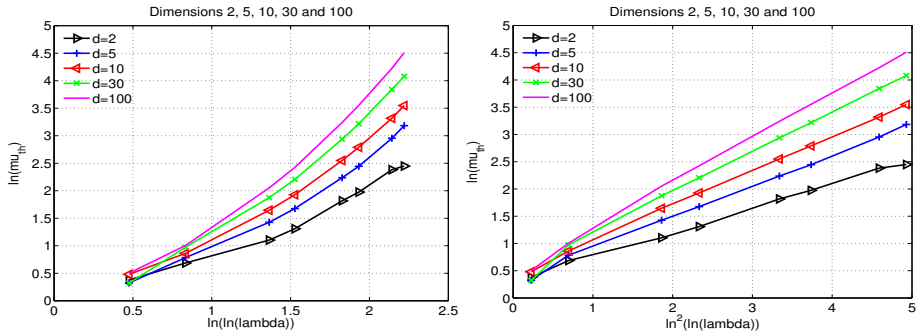

Figure 3: Plots of the logarithm of the argmin μ values (denoted μ_{th}) of the quadratic approximation in (9) as a function of $\ln(\ln(\lambda))$ (left) ($\ln^2(\ln(\lambda))$ (right)) for dimensions 2, 5, 10, 30 and 100.

Fig. 2 (right) suggests also that the values of μ_{th} vary as a function of $\ln(\lambda)$. Some preliminary plots, lead us to make the hypothesis that μ_{th} values varies as a function of $\ln^{(\alpha(d))}(\lambda)$ where $\alpha(d)$ depends on the dimension. To check this hypothesis, we plotted $\ln(\mu_{th})$ as a function of $\ln(\ln(\lambda))$ as represented in Fig. 3 (left). This figure suggests that $\ln(\mu_{th})$ could vary as a function of $\ln^2(\ln(\lambda))$ which has been confirmed by Fig. 3 (right). This suggests that a good setting of μ for λ large is such that $\ln(\mu_{th}) = \alpha(d)\ln^2(\ln(\lambda)) + \beta(d)$ where $\alpha(d), \beta(d) > 0$ are some constants that have to be tuned for each dimension.

6 Conclusion

In this report, we have developed a complementary theoretical/numerical approach in order to investigate the isotropic $(\mu/\mu_w, \lambda)$-ES minimizing spherical functions. First, we have shown the log-linear convergence of this algorithm (provided good choice of parameters) with a scale-invariant adaptation rule for the step-size and we have expressed the convergence rate as the expectation of a given random variable. Second, thanks to the expression of the convergence rate, we have numerically computed, using Monte-Carlo simulations, optimal values for the choice of μ and $\frac{\sigma_n}{d_n}$ and their relative optimal convergence rates. We have investigated in particular large values of λ. Our results suggest that the optimal μ is monotonously increasing in λ as opposed to the rule $\mu = \min\{\lfloor \frac{\lambda}{4} \rfloor, d\}$ proposed in [10] but that however this latter rule gives a convergence rate close to the optimal one. We have confirmed as well that for the rules $\mu = \lfloor \frac{\lambda}{4} \rfloor$ and $\lfloor \frac{\lambda}{2} \rfloor$, the convergence rate does not scale linearly in $\ln(\lambda)$ and is thus sub-optimal.

Acknowledgments The authors would like to thank Nikolaus Hansen for his advises on how to approach the problem tackled in the paper. This work received support by the French national research agency (ANR) within the COSINUS project ANR-08-COSI-007-12.

References

[1] M. Schumer and K. Steiglitz. Adaptive step size random search. *Automatic Control, IEEE Transactions on*, 13:270–276, 1968.

[2] I. Rechenberg. *Evolutionstrategie: Optimierung Technisher Systeme nach Prinzipien des Biologischen Evolution*. Fromman-Hozlboog Verlag, Stuttgart, 1973.

[3] M. Loeve. *Probability theory I*. Springer-Verlag, 1977.

[4] H.-P. Schwefel. Collective phenomena in evolutionary systems. In P. Checkland and I. Kiss, editors, *Problems of Constancy and Change-The Complementarity of Systems Approaches to Complexity, Proc. of 31st Annual Meeting Int'l Soc. for General System Research*,2:1025–1033, Budapest, 1987.

[5] Nikolaus Hansen and Andreas Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.

[6] A. Auger and N. Hansen. Reconsidering the progress rate theory for evolution strategies in finite dimensions. In ACM Press, editor, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2006)*, pages 445–452, 2006.

[7] D.V. Arnold. Optimal weighted recombination. In *Foundations of Genetic Algorithms 8*, pages 215–237. Springer Verlag, 2005.

[8] M. Jebalia, A. Auger, and P. Liardet. Log-linear convergence and optimal bounds for the (1+1)-ES. In N. Monmarché and al., editors, *Proceedings of Evolution Artificielle (EA'07)*, volume 4926 of *LNCS*, pages 207–218. Springer, 2008.

[9] F. Teytaud and O. Teytaud. On the parallel speed-up of Estimation of Multivariate Normal Algorithm and Evolution Strategies. In *Proceedings of EvoStar 2009*, pages 655–664. 2009.

[10] F. Teytaud. A new selection ratio for large population sizes. In *Proceedings of EvoStar. 2010.*

[11] A. Bienvenüe and O. François. Global convergence for evolution strategies in spherical problems: some simple proofs and difficulties. *Th. Comp. Sc.*, 306(1-3):269-289, 2003.

[12] M. Jebalia, A. Auger and N. Hansen. Log-linear convergence and divergence of the scale-invariant (1+1)-ES in noisy environments. *Algorithmica*, (to appear), 2010.

[13] Anne Auger. Convergence results for (1,λ)-SA-ES using the theory of φ-irreducible markov chains. *Theoretical Computer Science*, 334(1–3):35–69, 2005.

[14] H.-G. Beyer. *The Theory of Evolution Strategies*. Nat. Comp. Series. Springer-Verlag, 2001.

[15] H.-G. Beyer and B. Sendhoff. Covariance Matrix Adaptation revisited - The CMSA Evolution Strategy. In Günter Rudolph and al., editors, *Proceedings of Parallel Problem Solving from Nature (PPSN X)*, volume 5199 of *(LNCS)*, pages 123–132. Springer Verlag, 2008.

[16] O. Teytaud and H. Fournier. Lower Bounds for Evolution Strategies Using VC-dimension. In Günter Rudolph and al., editors, *Proceedings of PPSN X*, pages 102–111. Springer, 2008.

[17] D. Williams. *Probability with Martingales*. Cambridge University Press, Cambrdige, 1991.

[18] H.-G. Beyer. *Toward a Theory of Evolution Strategies: On the Benefits of Sex - the (μ/μ,λ) Theory*. *Evolutionary Computation*, 3(1):81–111, 1995.

Appendix

We provide in the appendix the proofs of the theorems stated in the core of the report. The proofs often require intermediate results. These intermediate results are organized in lemmas and propositions that are stated and proven in the following section, before to tackle the proofs of the main results.

Useful Definitions and Preliminary Results

Further definitions and notations We recall here that $\mathbb{Z}^+$ denotes the set of non-negative integers $\{0, 1, 2, \ldots\}$ and $\mathbb{N}$ denotes the set of positive integers $\{1, 2, \ldots\}$. We recall that e_1 is the unit vector $(1, 0, \ldots, 0) \in \mathbb{R}^d$. For $\mu, \lambda \in \mathbb{N}$ such that $1 \leq \mu \leq \lambda$, $\mathcal{P}_\lambda^\mu = \frac{\lambda!}{(\lambda-\mu)!}$ is the number of permutations of μ elements among a set of λ elements For a set A, $x \mapsto \mathbb{1}_A(x)$ denotes the indicator function that is equal to one if $x \in A$ and zero otherwise. $(\Omega, \mathcal{A}, P)$ is a probability space: Ω is a set, $\mathcal{A}$ a σ-algebra defined on this set and P a probability measure defined on $(\Omega, \mathcal{A})$. For $p \in \mathbb{N}$, $\mathbb{R}^p$ is equipped with the Borel σ-algebra denoted $\mathfrak{B}(\mathbb{R}^p)$. For a subset $S \subset \mathbb{R}^p$, $\mathfrak{B}(S)$ will denote the Borel σ-algebra on S. If X is a random variable defined on $(\Omega, \mathcal{A}, P)$, i.e., a measurable function from Ω to $\mathbb{R}$, then, for $B \subset \mathbb{R}$, $B \in \mathfrak{B}(\mathbb{R})$, the indicator function $\mathbb{1}_{\{X \in B\}}$ maps Ω to $\{0, 1\}$ and equals one if and only if $X(\omega) \in B$ for $\omega \in \Omega$: $\omega \in \Omega \mapsto \mathbb{1}_{\{\omega : X(\omega) \in B\}}(\omega)$. $\mathcal{N}(a, b^2)$ denotes a normal distribution with mean a and variance b^2. $\mathcal{N}(0, C)$ denotes the multivariate normal distribution with mean $(0, \ldots, 0) \in \mathbb{R}^d$ and covariance matrix C. We recall that the identity matrix is denoted I_d.

Definition 2
For a given $u \in \mathbb{R}^d$ and $s \geq 0$ fixed, we introduce the map $h_{\{u,s\}} : \mathbb{R}^d \to \mathbb{R}$ such that, for $a \in \mathbb{R}^d$, we have

$$h_{\{u,s\}}(a) = \|u + sa\| \,. \tag{10}$$

Let $(\mathbf{N}_n^i)_{i \in [1,\lambda], n \in \mathbb{Z}^+}$ be a sequence of random vectors defined on $(\Omega, \mathcal{A}, P)$, i.i.d. with common law $\mathcal{N}(0, I_d)$ and independent of u. For all $n \geq 0$, We define the vector $\left(\mathbf{N}_n^{1:\lambda}(u,s), \ldots, \mathbf{N}_n^{\mu:\lambda}(u,s)\right)$ as the random vector containing μ different elements from the set $\{\mathbf{N}_n^i,\ i \in [1, \lambda]\}$ and verifying

$$\begin{aligned} &h_{\{u,s\}}\left(\mathbf{N}_n^{1:\lambda}(u,s)\right) \quad \leq \ldots \leq h_{\{u,s\}}\left(\mathbf{N}_n^{\mu:\lambda}(u,s)\right) \text{ and} \\ &h_{\{u,s\}}\left(\mathbf{N}_n^{\mu:\lambda}(u,s)\right) \leq h_{\{u,s\}}\left(\mathbf{N}_n^{j}(u,s)\right) \forall j \in \{1, \ldots, \lambda\} \backslash \{1\!:\!\lambda, \ldots, \mu\!:\!\lambda\} \,. \end{aligned} \tag{11}$$

For $n \geq 0$, we define the random vector $\mathbf{M}_n(u,s)$ as the recombination of the μ vectors $\mathbf{N}_n^{i:\lambda}(u,s),\ 1 \leq i \leq \mu$ using the weights $w^i,\ i = 1, \ldots, \mu$, i.e.,

$$\mathbf{M}_n(u,s) := \sum_{i=1}^{\mu} w^i \mathbf{N}_n^{i:\lambda}(u,s) \,. \tag{12}$$

Remark 1 *The relation (2) is then a specific case of (11) where (u,s) is replaced by $(\mathbf{X}_n, \sigma_n)$. Moreover, using the previous definition, the relations (3) and (4) will be respectively replaced in the sequel by*

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \sigma_n \mathbf{M}_n(\mathbf{X}_n, \sigma_n) \tag{13}$$

and

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \sigma \|\mathbf{X}_n\| \mathbf{M}_n(\mathbf{X}_n, \sigma \|\mathbf{X}_n\|) \,. \tag{14}$$

For $n = 0$, we respectively replace in the sequel, the notations $\mathbf{N}_0^i$, $\mathbf{N}_0^{i:\lambda}$, $(i = 1, \ldots, \lambda)$ and $\mathbf{M}_0$ by the notations $\mathbf{N}^i$, $\mathbf{N}^{i:\lambda}$ $(i = 1, \ldots, \lambda)$ and $\mathbf{M}$.

The probability density function of the random vector

$$\left(\mathbf{N}^{1:\lambda}(u,s), \ldots, \mathbf{N}^{i:\lambda}(u,s), \ldots, \mathbf{N}^{\mu:\lambda}(u,s)\right)$$

introduced in (11) (where the notation $\left(\mathbf{N}^{1:\lambda}(u,s), \ldots, \mathbf{N}^{i:\lambda}(u,s), \ldots, \mathbf{N}^{\mu:\lambda}(u,s)\right)$ replace the notation $\left(\mathbf{N}_0^{1:\lambda}, \ldots, \mathbf{N}_0^{i:\lambda}(u,s), \ldots, \mathbf{N}_0^{\mu:\lambda}(u,s)\right)$ in (11)) is given in the following lemma.

Lemma 1 *For fixed $u \in \mathbb{R}^d$ and s a positive constant, let $h_{\{u,s\}}$ be the map introduced in (10). The probability density function of the random vector $\left(\mathbf{N}^{1:\lambda}(u,s), \ldots, \mathbf{N}^{\mu:\lambda}(u,s)\right)$ selected according to $h_{\{u,s\}}$ as introduced in (11) is the function $p_{\{u,s\}}$ mapping $\mathbb{R}^{\mu d}$ into $\mathbb{R}^+$:*

$$p_{\{u,s\}}(x^1, \ldots, x^\mu) = \frac{1}{(2\pi)^{\frac{\mu d}{2}}} e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}} \mathcal{P}_\lambda^\mu \times$$
$$\mathbb{1}_{\left\{h_{\{u,s\}}(x^1) \leq \ldots \leq h_{\{u,s\}}(x^\mu)\right\}} \left(x^1, \ldots, x^\mu\right) P^{\,\lambda-\mu}\left[h_{\{u,s\}}\left(x^\mu\right) \leq h_{\{u,s\}}\left(\mathbf{N}\right)\right] ,$$

where $\mathbf{N}$ is an independent random vector following $\mathcal{N}(0, I_d)$.

Proof. Let $(\mathbf{N}^i)_{i\in[1,\lambda]}$ be λ independent samplings of $\mathcal{N}(0, I_d)$. For fixed $u \in \mathbb{R}^d$ and $s > 0$, we recall that $h_{\{u,s\}} : \mathbb{R}^d \to \mathbb{R}$ is the map introduced in (10) by $h_{\{u,s\}}(a) = \|u + sa\|$ for all a in $\mathbb{R}^d$. We are going to compute the probability density function of the random vector $(\mathbf{N}^{1:\lambda}(u,s), \ldots, \mathbf{N}^{i:\lambda}(u,s), \ldots, \mathbf{N}^{\mu:\lambda}(u,s))$. The random vector $\left(\mathbf{N}^{1:\lambda}(u,s), \ldots, \mathbf{N}^{i:\lambda}(u,s), \ldots, \mathbf{N}^{\mu:\lambda}(u,s)\right)$ verify:

$$h_{\{u,s\}}(\mathbf{N}^{1:\lambda}(u,s)) \leq \ldots \leq h_{\{u,s\}}(\mathbf{N}^{i:\lambda}(u,s)) \ldots \leq h_{\{u,s\}}(\mathbf{N}^{\mu:\lambda}(u,s)) . \quad (15)$$

The number of possibilities for the choice of the indexes $1 : \lambda, \ldots, \mu : \lambda$ (i.e., for the choice of the elements having the μ least values according to the function h) is $\mathcal{P}_\lambda^\mu = \frac{\lambda!}{(\lambda-\mu)!}$. Let $(A^1, \ldots, A^\mu) \in \mathfrak{B}(\mathbb{R}^{\mu d})$. As the variables $(\mathbf{N}^i)_{i\in[1,\lambda]}$ play the same role, we have:

$$P\left\{\mathbf{N}^{1:\lambda} \in A^1; \ldots; \mathbf{N}^{i:\lambda} \in A^i; \ldots, \mathbf{N}^{\mu:\lambda} \in A^\mu\right\} = \mathcal{P}_\lambda^\mu P(E_1 \cap E_2 \cap E_3) \quad (16)$$

with

$$E_1 = \left\{\cap_{i=1}^{\mu} \left(\mathbf{N}^i \in A^i\right)\right\} ,$$

$$E_2 = \left\{h_{\{u,s\}}(\mathbf{N}^1) \leq h_{\{u,s\}}(\mathbf{N}^2) \leq \ldots \leq h_{\{u,s\}}(\mathbf{N}^\mu)\right\} ,$$

and

$$E_3 = \left\{h_{\{u,s\}}(\mathbf{N}^{(u)}) \geq h_{\{u,s\}}(\mathbf{N}^\mu),\ \forall\, \mu + 1 \leq u \leq \lambda\right\} .$$

The probability $P(E_1 \cap E_2 \cap E_3)$ can be rewritten as:

$$\begin{aligned} &P(E_1 \cap E_2 \cap E_3) \\ &= E_{\{\mathbf{N}^1, \ldots, \mathbf{N}^\lambda\}}\left[\mathbb{1}_{\{E_1 \cap E_2 \cap E_3\}}\right] \\ &= E_{\{\mathbf{N}^1, \ldots, \mathbf{N}^\mu\}}\left[E_{\{\mathbf{N}^{\mu+1}, \ldots, \mathbf{N}^\lambda\}}\left[\mathbb{1}_{\{E_1 \cap E_2 \cap E_3\}} | \mathbf{N}^1, \ldots, \mathbf{N}^\mu\right]\right] \\ &= E_{\{\mathbf{N}^1, \ldots, \mathbf{N}^\mu\}}\left[\mathbb{1}_{\{E_1\}} \times E_{\{\mathbf{N}^{\mu+1}, \ldots, \mathbf{N}^\lambda\}}\left[\mathbb{1}_{\{E_2 \cap E_3\}} | \mathbf{N}^1, \ldots, \mathbf{N}^\mu\right]\right] \end{aligned} \quad (17)$$

For a given $(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, u, s)$, we define the quantity $\mathrm{H}(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, u, s)$ as

$$\mathrm{H}(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, u, s) := \mathcal{P}_\lambda^\mu E_{\{\mathbf{N}^{\mu+1}, \ldots, \mathbf{N}^\lambda\}} \left[\mathbb{1}_{\{E_2 \cap E_3\}} | \mathbf{N}^1, \ldots, \mathbf{N}^\mu \right] .$$

By (16) and (17), one can write:

$$P\left\{ \mathbf{N}^{1:\lambda} \in A^1, \ldots, \mathbf{N}^{\mu:\lambda} \in A^\mu \right\} = E_{\{\mathbf{N}^1, \ldots, \mathbf{N}^\mu\}} \left[\mathbb{1}_{\{E_1\}} \times \mathrm{H}(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, u, s) \right]$$

Moreover, we have:

$$\mathrm{H}(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, u, s) = \mathcal{P}_\lambda^\mu \mathbb{1}_{\left\{ h_{\{u,s\}}(\mathbf{N}^1) \leq h_{\{u,s\}}(\mathbf{N}^i) \ldots \leq h_{\{u,s\}}(\mathbf{N}^\mu) \right\}} P^{\lambda-\mu} \left\{ h_{\{u,s\}}(\mathbf{N}^\mu) \leq h_{\{u,s\}}(\mathbf{N}) | \mathbf{N}^\mu, y^\mu \right\} ,$$

where $\mathbf{N}$ is an independent random vector following $\mathcal{N}(0, I_d)$. Finally, we get:

$$P\left\{ (\mathbf{N}^1, \ldots, \mathbf{N}^\mu) \in (A^1 \times \ldots \times A^\mu) \right\} = \int_{A_1} \ldots \int_{A_\mu} \left(\frac{1}{(2\pi)^{\frac{\mu d}{2}}} e^{-\frac{\sum_{i=1}^\mu \|x^i\|^2}{2}} \mathrm{H}(x^1, \ldots, x^\mu, u, s) \right) dx^1 \ldots dx^\mu .$$

□

Thanks to the isotropy of the standard d-dimensional normal distribution $\mathcal{N}(0, I_d)$, we can state the following important lemma that will be also used in several results.

Lemma 2 *Let u be a fixed vector in $\mathbb{R}^d$ such that $\|u\| = 1$ and let s be a positive constant. Let $\mathbf{M}(u, s)$ be the random vector introduced in (12). Then for all $t \in \mathbb{R}$, we have*

$$E\left(e^{it(\|u + s\mathbf{M}(u,s)\|)} \right) = E\left(e^{it(\|\mathrm{e}_1 + s\mathbf{M}(\mathrm{e}_1,s)\|)} \right) , \tag{18}$$

and consequently, for $\sigma > 0$, the random variables $\mathrm{Z}(\sigma) = \|\mathrm{e}_1 + \sigma\mathbf{M}(\mathrm{e}_1, \sigma)\|$ introduced in Definition 1[4] and $\|u + \sigma\mathbf{M}(u, \sigma)\|$ follow the same distribution.

Proof. Let $u \in \mathbb{R}^d$ be fixed verifying $\|u\| = 1$ and let s be a positive constant. Let $\mathbf{M}(u, s) = \sum_{i=1}^\mu w^i \mathbf{N}^{i:\lambda}(u, s)$ be the random vector introduced in (12) obtained by the recombination of the vectors $\mathbf{N}^{i:\lambda}(u, s)$ which has been selected according to the criterion $h_{\{u,s\}}$ as expressed in (11). Let also $\mathbf{M}(\mathrm{e}_1, s)$ be the random vector obtained relatively to $h_{\{\mathrm{e}_1,s\}}$. Let $\mathbf{N}$ be an independent random vector following $\mathcal{N}(0, I_d)$. According to Lemma 1, the probability density function of the random vector $\left(\mathbf{N}^{1:\lambda}(u, s), \ldots, \mathbf{N}^{\mu:\lambda}(u, s)\right)$ introduced in (11) is:

$$p_{\{u,s\}}(x^1, \ldots, x^\mu) = \frac{1}{(2\pi)^{\frac{\mu d}{2}}} e^{-\frac{\sum_{i=1}^\mu \|x^i\|^2}{2}} \mathcal{P}_\lambda^\mu \mathbb{1}_{\left\{ h_{\{u,s\}}(x^1) \leq \ldots \leq h_{\{u,s\}}(x^\mu) \right\}} \left(x^1, \ldots, x^\mu\right) P^{\lambda-\mu} \left[h_{\{u,s\}}\left(x^\mu\right) \leq h_{\{u,s\}}\left(\mathbf{N}\right) \right] .$$

for $(x^1, \ldots, x^\mu) \in \mathbb{R}^{d\mu}$. In the expression of $p_{\{u,s\}}$, let us give interest to the quantity

$$P^{\lambda-\mu} \left[h_{\{u,s\}}\left(x^\mu\right) \leq h_{\{u,s\}}\left(\mathbf{N}\right) \right]$$

[4]Note that the expression of Z is reformulated here using the random vector $\mathbf{M}$ introduced in the appendix (12).

where x^μ is fixed. Let R be an orthogonal matrix such that $R\left(u\right) = \mathrm{e}_1$. Since R is an orthogonal matrix, $\|R(y)\| = \|y\|$ for all $y \in \mathbb{R}^d$. Then $\|u + s\mathbf{N}\| = \|R\left(u + s\mathbf{N}\right)\|$ almost surely. Besides $\|R\left(u + s\mathbf{N}\right)\| = \|\mathrm{e}_1 + sR\left(\mathbf{N}\right)\|$ but since $\mathbf{N}$ follows the distribution $\mathcal{N}(0, I_d)$ which is spherical, $R\left(\mathbf{N}\right)$ has the same law, i.e., $\mathcal{N}(0, I_d)$ and thus $\|u + s\mathbf{N}\|$ (with u fixed) and $\|\mathrm{e}_1 + s\mathbf{N}\|$ have the same distribution which gives

$$P^{\,\lambda-\mu}\left[h_{\{u,s\}}\left(x^\mu\right) \leq h_{\{u,s\}}\left(\mathbf{N}\right)\right] = P^{\,\lambda-\mu}\left[h_{\{u,s\}}\left(x^\mu\right) \leq h_{\{\mathrm{e}_1,s\}}\left(\mathbf{N}\right)\right] ,$$

Consequently, we have

$$p_{\{u,s\}}(x^1, \ldots, x^\mu) = \frac{1}{(2\pi)^{\frac{\mu d}{2}}} e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}} \mathcal{P}_\lambda^\mu \times$$
$$\mathbb{1}_{\left\{h_{\{u,s\}}(x^1) \leq \ldots \leq h_{\{u,s\}}(x^\mu)\right\}}\left(x^1, \ldots, x^\mu\right) P^{\,\lambda-\mu}\left[h_{\{u,s\}}\left(x^\mu\right) \leq h_{\{\mathrm{e}_1,s\}}\left(\mathbf{N}\right)\right] .$$

The random vectors $\|u + s\mathbf{M}(u,s)\|$ and $\|\mathrm{e}_1 + s\mathbf{M}(\mathrm{e}_1,s)\|$ have the same distribution if their characteristic functions are identical:

$$E\left\{e^{it(\|u+s\mathbf{M}(u,s)\|)}\right\} = \int_{\mathbb{R}^{\mu d}} e^{it\left(\left\|u+s\sum_{i=1}^{\mu} w^i x^i\right\|\right)} p_{\{u,s\}}(x^1, \ldots, x^\mu) dx^1 \ldots dx^\mu .$$

Again, we use the fact that $R\left(\|u + st\|\right) = \|\mathrm{e}_1 + sR(t)\|$. We get

$$E\left\{e^{it(\|u+s\mathbf{M}(u,s)\|)}\right\} = \int_{\mathbb{R}^{\mu d}} e^{it\left(\left\|\mathrm{e}_1+s\sum_{i=1}^{\mu} w^i R(x^i)\right\|\right)} p_{\{u,s\}}(x^1, \ldots, x^\mu) dx^1 \ldots dx^\mu ,$$

with $p_{\{u,s\}}(x^1, \ldots, x^\mu)$ rewritten as

$$p_{\{u,s\}}(x^1, \ldots, x^\mu) = \mathbb{1}_{\left\{h_{\{\mathrm{e}_1,s\}}(R(x^1)) \leq \ldots \leq h_{\{\mathrm{e}_1,s\}}(R(x^\mu))\right\}}\left(x^1, \ldots, x^\mu\right)$$
$$\mathcal{P}_\lambda^\mu \frac{1}{(2\pi)^{\frac{\mu d}{2}}} e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}} \times P^{\,\lambda-\mu}\left[h_{\{\mathrm{e}_1,s\}}\left(R(x^\mu)\right) \leq h_{\{\mathrm{e}_1,s\}}\left(\mathbf{N}\right)\right] .$$

Then, we apply the change of variables $s^i = R(x^i)$ for all $1 \leq i \leq \mu$. The quantity $e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}} dx^1 \ldots dx^\mu$ is then replaced by $e^{-\frac{\sum_{i=1}^{\mu} \|s^i\|^2}{2}} ds^1 \ldots ds^\mu$ as the rotation (using R) of a μ independent random vectors following the spherical distribution of $\mathcal{N}(0, I_d)$ results in μ independent random vectors having the same $\mathcal{N}(0, I_d)$ distribution. Moreover, for each $i \in \{1, \ldots, \mu\}$, $h_{\{\mathrm{e}_1,s\}}\left(R(x^i)\right) = h_{\{\mathrm{e}_1,s\}}\left(s^i\right)$ such that we get

$$p_{\{u,s\}}\left(x^1, \ldots, x^\mu\right) = p_{\{\mathrm{e}_1,s\}}(s^1, \ldots, s^\mu) .$$

Consequently, for all $u \in \mathbb{R}^d$ fixed with $\|u\| = 1$ and for any positive constant s,

$$E\left\{e^{it(\|u+s\mathbf{M}(u,s)\|)}\right\} = E\left\{e^{it(\|\mathrm{e}_1+s\mathbf{M}(\mathrm{e}_1,s)\|)}\right\} .$$

□

Proposition 3 *Let $(\mathbf{X}_n)_{n\geq 0}$ be the sequence defined by (3). Then for any $\mathbf{x}^* \in \mathbb{R}^d$, for all $n \geq 0$, $\|\mathbf{X}_n - \mathbf{x}^*\| > 0$ almost surely.*

Proof. Without loss of generality, we suppose that $\mathbf{x}^* = (0, \ldots, 0) \in \mathbb{R}^d$. Note also that we will use the equivalent from of the recurrence relation (3), i.e., the recurrence relation (13) which uses the vector $\mathbf{M}$ introduced in (12). We will show inductively that, for all $n \geq 0$, $\|\mathbf{X}_n\| > 0$ almost surely :
1) The vector $\mathbf{X}_0$ is randomly sampled using a law absolutely continuous w.r.t the Lebesgue measure. Therefore $P(\|\mathbf{X}_0\| > 0) = 1$.
2) Suppose that $P(\|\mathbf{X}_n\| > 0) = 1$ for $n \geq 0$. We have:

$$P(\|\mathbf{X}_{n+1}\| > 0) = P(\|\mathbf{X}_{n+1}\| > 0 \cap \|\mathbf{X}_n\| = 0) + P(\|\mathbf{X}_{n+1}\| > 0 \cap \|\mathbf{X}_n\| > 0)\,.$$

As $P(\|\mathbf{X}_{n+1}\| > 0 \cap \|\mathbf{X}_n\| = 0) \leq P(\|\mathbf{X}_n\| = 0)$ and $P(\|\mathbf{X}_n\| = 0) = 0$ (inductive hypothesis), we get

$$\begin{aligned}
&P(\|\mathbf{X}_{n+1}\| > 0) \\
&= P(\|\mathbf{X}_{n+1}\| > 0 \cap \|\mathbf{X}_n\| > 0) \\
&= P(\|\mathbf{X}_{n+1}\| > 0 | \|\mathbf{X}_n\| > 0) P(\|\mathbf{X}_n\| > 0) \\
&= P(\|\mathbf{X}_{n+1}\| > 0 | \|\mathbf{X}_n\| > 0) \\
&= P(\|\mathbf{X}_n + \sigma_n \mathbf{M}_n\left(\mathbf{X}_n, \sigma_n\right)\| > 0 | \|\mathbf{X}_n\| > 0) \\
&= P\left(\left\| \frac{\mathbf{X}_n}{\|\mathbf{X}_n\|} + \frac{\sigma_n}{\|\mathbf{X}_n\|} \mathbf{M}_n\left(\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|}, \frac{\sigma_n}{\|\mathbf{X}_n\|}\right)\right\| > 0 | \sigma_n, \|\mathbf{X}_n\| > 0\right)\,.
\end{aligned} \tag{19}$$

By Lemma 2, we know that the distribution of $\left\| \frac{\mathbf{X}_n}{\|\mathbf{X}_n\|} + \frac{\sigma_n}{\|\mathbf{X}_n\|} \mathbf{M}_n\left(\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|}, \frac{\sigma_n}{\|\mathbf{X}_n\|}\right)\right\|$ conditionally to $\mathbf{X}_n$ and σ_n fixed is equal to that of $\left\| \mathrm{e}_1 + \frac{\sigma_n}{\|\mathbf{X}_n\|} \mathbf{M}\left(\mathrm{e}_1, \frac{\sigma_n}{\|\mathbf{X}_n\|}\right)\right\|$. Then, we have:

$$\begin{aligned}
&P\left(\left\| \frac{\mathbf{X}_n}{\|\mathbf{X}_n\|} + \frac{\sigma_n}{\|\mathbf{X}_n\|} \mathbf{M}_n\left(\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|}, \frac{\sigma_n}{\|\mathbf{X}_n\|}\right)\right\| > 0 | \|\mathbf{X}_n\| > 0\right) \\
&\qquad = P\left(\left\| \mathrm{e}_1 + \frac{\sigma_n}{\|\mathbf{X}_n\|} \mathbf{M}\left(\mathrm{e}_1, \frac{\sigma_n}{\|\mathbf{X}_n\|}\right)\right\| > 0 | \sigma_n, \|\mathbf{X}_n\| > 0\right) \\
&\qquad\qquad = 1 - P\left(\left\| \mathrm{e}_1 + \frac{\sigma_n}{\|\mathbf{X}_n\|} \mathbf{M}\left(\mathrm{e}_1, \frac{\sigma_n}{\|\mathbf{X}_n\|}\right)\right\| = 0 | \sigma_n, \|\mathbf{X}_n\| > 0\right)\,.
\end{aligned}$$

We know that, $\forall s > 0$, $P(\|\mathrm{e}_1 + s\mathbf{M}(\mathrm{e}_1, s)\| = 0) = 0$ as the variable $\|\mathrm{e}_1 + s\mathbf{M}(\mathrm{e}_1, s)\|$ is absolutely continuous w.r.t. the Lebesgue measure. This gives that, conditionnally to $u_n := \frac{\sigma_n}{\|\mathbf{X}_n\|}$ fixed, $P\left(\|\mathrm{e}_1 + u_n \mathbf{M}\left(\mathrm{e}_1, u_n\right)\| = 0 | u_n\right) = 0$. Consequently $P(\|\mathbf{X}_{n+1}\| > 0) = 1$ and then for all $n \geq 0$, $\|\mathbf{X}_n\| > 0$ almost surely. □

Computation of the Limit of the Normalized Convergence Rate

An important step in the proof Proposition 1(iii), is the computation of the limit $\lim_{d \to +\infty} d \times \mathrm{V}\left(\frac{\sigma^*}{d}\right)$ where V is introduced in (5)and $\sigma^* \geq 0$. First, we will need the following Definition.

Definition 3 *Let:*

- σ^* *be a strictly positive constant,*
- $h_d : \mathbb{R}^d \mapsto \mathbb{R}$ *be a sequence of maps ($d \geq 1$) defined by* $h_d(x) = \|\mathrm{e}_1 + \frac{\sigma^*}{d} x\|$, $\forall x \in \mathbb{R}^d$ *with* $\sigma^* > 0$ *fixed,*

- $\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \mathbf{N}$ *be* $\mu+1$ *random vectors independent and identically distributed with* $\mathcal{N}(0, I_d)$ *as a common law, and,*

We define the sequences of random variables $\left\{A_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*)\right\}_{d\geq 1}$ *and* $\left\{\mathrm{H}_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*)\right\}_{d\geq 1}$ *as the following:*

$$A_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*) := \left[1 + \frac{\sigma^*}{d}\sum_{i=1}^{\mu} w^i (\mathbf{N}^i)_1\right]^2 + \left(\frac{\sigma^*}{d}\right)^2 \left[\sum_{i=1}^{\mu}(w^i)^2\left(\sum_{j=2}^{d}(\mathbf{N}^i)_j^2\right) + 2\sum_{1\leq i<k\leq \mu} w^i w^k \sum_{j=2}^{d}(\mathbf{N}^i)_j (\mathbf{N}^k)_j\right],$$

and

$$\mathrm{H}_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*) := \mathcal{P}_\lambda^\mu \mathbb{1}_{\{h_d(\mathbf{N}^1)\leq \ldots \leq h_d(\mathbf{N}^\mu)\}} E^{\lambda-\mu}\left[\mathbb{1}_{\{h_d(\mathbf{N}^\mu)\leq h_d(\mathbf{N})\}} | \mathbf{N}^\mu\right].$$

The sequence of functions $(h_d)_{d\geq 1}$ introduced in Definition 3 verify the following result:

Lemma 3 *Let* σ^* *be a strictly positive constant. Let* $\mathbf{N}$ *be a random vector following the distribution* $\mathcal{N}(0, I_d)$*. Then, the sequence* $(h_d)_{d\geq 1}$ *introduced in Definition 3 verify the following equation*

$$\lim_{d\to+\infty} d\left(h_d(\mathbf{N}) - 1\right) = \sigma^* \mathbf{N}_1{}^5 + \frac{(\sigma^*)^2}{2}. \tag{20}$$

Proof. Let $\sigma^* > 0$ be fixed. Let L_d denotes the sequence of random variables defined, for $d \geq 1$, by:

$$L_d := 2\frac{\sigma^*}{d}\mathbf{N}_1 + \frac{(\sigma^*)^2}{d}\frac{\sum_{j=1}^d (\mathbf{N})_j^2}{d}.$$

As, for a given $x \in \mathbb{R}^d$, $\left\|\mathrm{e}_1 + \frac{\sigma^*}{d}x\right\| = \sqrt{1 + \frac{2\sigma^*}{d}x_1 + \left(\frac{\sigma^*}{d}\right)^2 \sum_{j=1}^d x_j^2}$, we have:

$$d\left(h_d\left(\mathbf{N}\right) - 1\right) = d\left((1 + L_d)^{\frac{1}{2}} - 1\right).$$

By the LLN of independent identically distributed random variables, we have

$$\lim_{d\to+\infty} \frac{\sum_{j=1}^d (\mathbf{N})_j^2}{d} = E((\mathbf{N}_1)^2) = 1.$$

Therefore, we have $\lim_{d\to+\infty} L_d = 0$ almost surely. This gives, using the previous equation and the fact that $(1+y)^{\frac{1}{2}} \sim 1 + \frac{1}{2}y$ that

$$\begin{aligned}\lim_{d\to+\infty} d\left(h_d(\mathbf{N}) - 1\right) &= \lim_{d\to+\infty} d\left((1 + L_d)^{\frac{1}{2}} - 1\right) = \lim_{d\to+\infty} \frac{d}{2} L_d \\ &= \lim_{d\to+\infty} \sigma^* \mathbf{N}_1 + \frac{(\sigma^*)^2}{2}\frac{\sum_{j=1}^d \mathbf{N}_j^2}{d} = \sigma^* \mathbf{N}_1 + \frac{(\sigma^*)^2}{2}.\end{aligned}$$

[5] The notation $\mathbf{N}_l$ where $l \in \{1, \ldots, d\}$ denotes here (and in the proof of the Lemma) the l^{th} variable of the random vector $\mathbf{N}$.

□

In order to compute the limit, when d goes to infinity, of the quantity $\mathrm{V}\left(\frac{\sigma^*}{d}\right)$ we will write this quantity as the expectation of the family of random variables A_d and H_d which has been introduced in Definition 3 and then prove their uniform integrability. After this step, we can compute the limit of $\mathrm{V}\left(\frac{\sigma^*}{d}\right)$, when d goes to infinity.

Proposition 4 *Let:*

- σ^* *be a strictly positive constant,*
- $\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \mathbf{N}$ *be* $\mu+1$ *random vectors independent and identically distributed with* $\mathcal{N}(0, I_d)$ *as a common law, and,*
- $h_d : \mathbb{R}^d \mapsto \mathbb{R}$ *be the sequence of maps* ($d \geq 1$) *introduced in Definition 3,*
- $\left\{A_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*)\right\}_{d\geq 1}$ *and* $\left\{\mathrm{H}_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*)\right\}_{d\geq 1}$ *be the sequences of random variables introduced in Definition 3.*

Therefore, the function V *defined in (5) verify:*

$$d \times \mathrm{V}\left(\frac{\sigma^*}{d}\right) = E\left[\frac{d}{2}\ln\left(A_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*)\right)\mathrm{H}_d\left(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*\right)\right]. \qquad (21)$$

Moreover, the family $\left\{\frac{d}{2}\ln\left(A_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*)\right)\mathrm{H}_d\left(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*\right)\right\}_{d\geq 1}$ *is uniformly integrable.*

Before to prove the proposition, we recall the following result which is a part of the L^r-convergence theorem stated in [3, p. 165].

Theorem 3 (L^r-convergence theorem) *Let* $(\mathrm{U}_n)_{n\geq 0}$ *a sequence of random variables such that* $\int_{\mathbb{R}} |\mathrm{U}_n|^r < +\infty$. *Then the following affirmations are equivalent:*
i) $\lim_{n\to\infty} \int_{\mathbb{R}} |\mathrm{U}_n - \mathrm{U}|^r = 0$ *(*U_n *converges to* U *in the sense of the norm* L^r*)*
ii) U_n *converges in probability to* U *and* $\lim_{n\to\infty} \int_{\mathbb{R}} |\mathrm{U}_n|^r = \int_{\mathbb{R}} |\mathrm{U}|^r < +\infty$.

This theorem will be used in the following proof of the proposition.
Proof. Let us rewrite $\mathrm{V}(\sigma(d))$ in (5) using $\sigma(d) = \frac{\sigma^*}{d}$:

$$d \times \mathrm{V}\left(\frac{\sigma^*}{d}\right) = \frac{\mathcal{P}_\lambda^\mu}{(2\pi)^{\mu d/2}} \int_{\mathbb{R}^{\mu d}} \frac{d}{2}\ln(\|\mathrm{e}_1 + \frac{\sigma^*}{d}\sum_{i=1}^{\mu} w^i x^i\|^2) e^{-\frac{\sum_{i=1}^{\mu}\|x^i\|^2}{2}} P^{\,\lambda-\mu}\left[h_d\left(x^\mu\right) \leq h_d\left(\mathbf{N}\right)\right] \mathbb{1}_{\{h_d(x^1)\leq\ldots\leq h_d(x^\mu)\}}\left(x^1, \ldots, x^\mu\right) dx^1 \ldots dx^\mu .$$

In the remainder of this proof, the positive quantities σ^*, λ and μ are fixed. Let $\{\mathrm{H}_d\}_{d\geq 1}$ be the sequence of measurable functions, which verify, for $d \geq 1$, $\mathrm{H}_d : (\mathbb{R}^d)^\mu \times \mathbb{R}^+ \mapsto \mathbb{R}$ such that:

$$\mathrm{H}_d(x^1, \ldots, x^\mu, \sigma^*) = \mathcal{P}_\lambda^\mu \, \mathbb{1}_{\{h_d(x^1)\leq\ldots\leq h_d(x^\mu)\}}\left(x^1, \ldots, x^\mu\right) P^{\,\lambda-\mu}\left[h_d\left(x^\mu\right) \leq h_d\left(\mathbf{N}\right)\right] .$$

The probability of any event is upper bounded by 1. Therefore, the functions H_d are upper bounded by $\mathcal{P}_\lambda^\mu$ and $d \times \mathrm{V}(\frac{\sigma^*}{d})$ can be rewritten as

$$d \times \mathrm{V}\left(\frac{\sigma^*}{d}\right) = \frac{1}{(2\pi)^{\mu d/2}} \times$$
$$\int_{\mathbb{R}^{\mu d}} \frac{d}{2} \ln\left(\left\| e_1 + \frac{\sigma^*}{d} \sum_{i=1}^{\mu} w^i x^i \right\|^2\right) e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}} \mathrm{H}_d(x^1, \ldots, x^\mu, \sigma^*) dx^1 \ldots dx^\mu \,.$$

Let us now remark that, for $(x^1, \ldots, x^\mu) \in \mathbb{R}^{\mu d}$, we have

$$\left\| \mathrm{e}_1 + \frac{\sigma^*}{d} \sum_{i=1}^{\mu} w^i x^i \right\|^2 = \left[1 + \frac{\sigma^*}{d} \sum_{i=1}^{\mu} w^i (x^i)_1\right]^2 + \left(\frac{\sigma^*}{d}\right)^2 \left[\sum_{i=1}^{\mu} (w^i)^2 \left(\sum_{j=2}^{d} (x^i)_j^2\right) + 2 \sum_{1 \le i < k \le \mu} w^i w^k \sum_{j=2}^{d} (x^i)_j (x_k)_j\right] ,$$

where, for $a \in \{1, \ldots, \mu\}$ and $b \in \{1, \ldots, d\}$, the variable $(x^a)_b$ denotes the b^{th} variable of the vector x^a. For $(\mathbf{N}^1, \ldots, \mathbf{N}^\mu) \in \mathbb{R}^{d\mu}$ and $\sigma^* > 0$, let us define the sequences of random variables $\left\{A_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*)\right\}_{d \ge 1}$ and $\left\{\mathrm{H}_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*)\right\}_{d \ge 1}$, by

$$A_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*) := \left[1 + \frac{\sigma^*}{d} \sum_{i=1}^{\mu} w^i (\mathbf{N}^i)_1\right]^2 + \left(\frac{\sigma^*}{d}\right)^2 \left[\sum_{i=1}^{\mu} (w^i)^2 \left(\sum_{j=2}^{d} (\mathbf{N}^i)_j^2\right) + 2 \sum_{1 \le i < k \le \mu} w^i w^k \sum_{j=2}^{d} (\mathbf{N}^i)_j (\mathbf{N}^k)_j\right] ,$$

and

$$\mathrm{H}_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*) := \mathcal{P}_\lambda^\mu \mathbb{1}_{\{h_d(\mathbf{N}^1) \le \ldots \le h_d(\mathbf{N}^\mu)\}} E^{\lambda-\mu}\left[\mathbb{1}_{\{h_d(\mathbf{N}^\mu) \le h_d(\mathbf{N})\}} | \mathbf{N}^\mu\right] .$$

Therefore, we have

$$d \times \mathrm{V}\left(\frac{\sigma^*}{d}\right) = E\left[\frac{d}{2} \ln\left(A_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*)\right) \mathrm{H}_d\left(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*\right)\right] .$$

Let $(\mathrm{K}_d)_{d \ge 1}$ be the sequence of random variables defined as

$$\mathrm{K}_d\left(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*\right) := \frac{d}{2} \ln\left(A_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*)\right) \mathrm{H}_d\left(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*\right)$$

such that we have $d \times \mathrm{V}\left(\frac{\sigma^*}{d}\right) = E\left(\mathrm{K}_d\right)$. Let K_d^+ and K_d^- be respectively the positive and negative part of the function K_d such that $\mathrm{K}_d = \mathrm{K}_d^+ - \mathrm{K}_d^-$. We have to show that the families of positive random variables $(\mathrm{K}_d^+)_{d \ge 1}$ and $(\mathrm{K}_d^-)_{d \ge 1}$ are uniformly integrable. First, we give interest to the family $(\mathrm{K}_d^+)_{d \ge 1}$. Using the fact that:

- the sequence of random variables $(\mathrm{H}_d)_{d \ge 1}$ is upper bounded by $\mathcal{P}_\lambda^\mu$,
- $\ln^+(1 + \sum_{i=1}^{\mu} x^i) \le \ln^+(1 + |\sum_{i=1}^{\mu} x^i|) \le |\sum_{i=1}^{\mu} x^i| \le \sum_{i=1}^{\mu} |x^i|$ (for any $x^1, \ldots, x^\mu \in \mathbb{R}$ such that $\sum_{i=1}^{\mu} x^i > -1$),

- $0 \leq w^i \leq 1 \; \forall i \in \{1, \ldots, \mu\}$,

we get

$$
\begin{aligned}
&(\mathrm{K})_d^+ \\
&\leq \frac{\mathcal{P}_\lambda^\mu}{2} \left[2\sigma^* \sum_{i=1}^{\mu} |\mathbf{N}^i|_1 + \frac{(\sigma^*)^2}{d} \left(\sum_{i=1}^{\mu} \left(\sum_{j=1}^{d} (\mathbf{N}^i)_j^2 \right) + 2 \sum_{1 \leq i < k \leq \mu} \sum_{j=1}^{d} |\mathbf{N}^i|_j |\mathbf{N}^k|_j \right) \right] \\
&\leq \frac{\mathcal{P}_\lambda^\mu}{2} \left[2\sigma^* \sum_{i=1}^{\mu} |\mathbf{N}^i|_1 + \frac{(\sigma^*)^2}{d} \left(\sum_{i=1}^{\mu} \left(\sum_{j=1}^{d} (\mathbf{N}^i)_j^2 \right) + \sum_{1 \leq i < k \leq \mu} \sum_{j=1}^{d} \left((\mathbf{N}^i)_j^2 + (\mathbf{N}^k)_j^2 \right) \right) \right] \\
&\leq \frac{\mathcal{P}_\lambda^\mu}{2} \left[2\sigma^* \sum_{i=1}^{\mu} |\mathbf{N}^i|_1 + \frac{(\sigma^*)^2}{d} \left(\sum_{i=1}^{\mu} \left(\sum_{j=1}^{d} (\mathbf{N}^i)_j^2 \right) + 2 \sum_{i=1}^{\mu} \sum_{j=1}^{d} (\mathbf{N}^i)_j^2 \right) \right] \\
&\leq \mathcal{P}_\lambda^\mu \left[\sigma^* \sum_{i=1}^{\mu} |\mathbf{N}^i|_1 + \frac{3(\sigma^*)^2}{2d} \sum_{i=1}^{\mu} \left(\sum_{j=1}^{d} (\mathbf{N}^i)_j^2 \right) \right]
\end{aligned}
\tag{22}
$$

According to the last inequality, we have to show that, for each $i = 1, \ldots, \mu$, the families $|(\mathbf{N}^i)_1|$ and $\left(\frac{\sum_{j=1}^{d} (\mathbf{N}^i)_j^2}{d} \right)_{d \geq 1}$ are uniformly integrable. For fixed $i \in \{1, \ldots, \mu\}$,

1. the family $|\mathbf{N}^i|$ contains a unique integrable random variable therefore it is uniformly integrable.

2. The random variable $\left(\frac{\sum_{j=1}^{d} (\mathbf{N}^i)_j^2}{d} \right)_d$ converges almost surely (by the Law of Large Numbers) and therefore in probability to 1. Moreover, the sequence of positive real values $E\left[\frac{\sum_{j=1}^{d} (\mathbf{N}^i)_j^2}{d} \right]_d = 1$ converges to $E\left[|1|\right]$ which gives, by Theorem 3, that $\left(\frac{\sum_{j=1}^{d} (\mathbf{N}^i)_j^2}{d} \right)_d$ converges to 1 in the sense of the norm L^1. Finally, the family $\left(\frac{\sum_{j=1}^{d} (\mathbf{N}^i)_j^2}{d} \right)_{d \geq 1}$ converges in L^1 therefore it is uniformly integrable.

Consequently, the upper bound in the last inequality is uniformly integrable which implies that the family $((\mathrm{K})_d^+)_{d \geq 1}$ is uniformly integrable. Let us now give interest to

the family $((\mathrm{K})_d^-)_{d\geq 2}$. We have

$$\begin{aligned}
(\mathrm{K})_d^- &\leq \frac{\mathcal{P}_\lambda^\mu}{2} d \ln^- \left(A_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*) \right) \\
&\leq \frac{\mathcal{P}_\lambda^\mu}{2} d \ln^- \left(1 - \frac{\left(\sum_{i=1}^\mu w^i (\mathbf{N}^i)_1\right)^2}{\left\|\sum_{i=1}^\mu w^i \mathbf{N}^i\right\|^2} \right) \mathbb{1}_{\left\{\sum_{i=1}^\mu w^i (\mathbf{N}^i)_1 < 0\right\}} \\
&= \frac{\mathcal{P}_\lambda^\mu}{2} \ln^- \left[\left(1 - \frac{\left(\sum_{i=1}^\mu w^i (\mathbf{N}^i)_1\right)^2}{\left\|\sum_{i=1}^\mu w^i \mathbf{N}^i\right\|^2} \right)^d \right] \mathbb{1}_{\left\{\sum_{i=1}^\mu w^i (\mathbf{N}^i)_1 < 0\right\}} \\
&= \frac{\mathcal{P}_\lambda^\mu}{2} \ln \left[\left(\frac{1}{1 - \frac{\left(\sum_{i=1}^\mu w^i (\mathbf{N}^i)_1\right)^2}{\left\|\sum_{i=1}^\mu w^i \mathbf{N}^i\right\|^2}} \right)^d \right] \mathbb{1}_{\left\{\sum_{i=1}^\mu w^i (\mathbf{N}^i)_1 < 0\right\}} \\
&\leq 4\mathcal{P}_\lambda^\mu \left(\frac{1}{1 - \frac{\left(\sum_{i=1}^\mu w^i (\mathbf{N}^i)_1\right)^2}{\left\|\sum_{i=1}^\mu w^i \mathbf{N}^i\right\|^2}} \right)^{\frac{d}{8}} \mathbb{1}_{\left\{\sum_{i=1}^\mu w^i (\mathbf{N}^i)_1 < 0\right\}}
\end{aligned} \tag{23}$$

Let us show that the family $(G_d)_{d\geq 2} := \left(\left(\frac{1}{1 - \frac{\left(\sum_{i=1}^\mu w^i (\mathbf{N}^i)_1\right)^2}{\left\|\sum_{i=1}^\mu w^i \mathbf{N}^i\right\|^2}} \right)^{\frac{d}{8}} \mathbb{1}_{\left\{\sum_{i=1}^\mu w^i (\mathbf{N}^i)_1 < 0\right\}} \right)_{d\geq 2}$ is uniformly integrable. This amounts to show, that the family

$$E\left(G_d^2\right) = \left(E\left[\left(\frac{1}{1 - \frac{\left(\sum_{i=1}^\mu w^i (\mathbf{N}^i)_1\right)^2}{\left\|\sum_{i=1}^\mu w^i \mathbf{N}^i\right\|^2}} \right)^{\frac{d}{4}} \mathbb{1}_{\left\{\sum_{i=1}^\mu w^i (\mathbf{N}^i)_1 < 0\right\}} \right] \right)_{d\geq 1}$$

is uniformly bounded. In the beginning, let us remark that the random variable $\sum_{i=1}^\mu w^i \mathbf{N}^i$ follows the distribution $w\mathcal{N}(0, I_d)$ with $w := \sqrt{\sum_{i=1}^\mu (w^i)^2}$. We recall here that $\mathbf{N}$ is a random vector following $\mathcal{N}(0, I_d)$. For $d \geq 2$, $E\left(G_d^2\right)$ can be rewritten as:

$$\begin{aligned}
E\left(G_d^2\right) &= E\left[\left(\frac{1}{1 - \frac{w^2 (\mathbf{N}_1)^2}{w^2 \|\mathbf{N}\|^2}} \right)^{\frac{d}{4}} \mathbb{1}_{\{\mathbf{N}_1 < 0\}} \right] \\
&= E\left[\left(\frac{1}{1 - \frac{(\mathbf{N}_1)^2}{\|\mathbf{N}\|^2}} \right)^{\frac{d}{4}} \mathbb{1}_{\{\mathbf{N}_1 < 0\}} \right] = \frac{1}{2} E\left[\left(\frac{1}{1 - \frac{(\mathbf{N}_1)^2}{\|\mathbf{N}\|^2}} \right)^{\frac{d}{4}} \right].
\end{aligned}$$

Changing to spherical coordinates (for $d \geq 2$), we get:

$$E\left(G_d^2\right) = \frac{1}{2W_{d-2}} \int_0^{\frac{\pi}{2}} \left(\frac{1}{\sin(\theta)} \right)^{\frac{d}{2}} \sin^{d-2}(\theta)\, d\theta = \frac{1}{2W_{d-2}} \int_0^{\frac{\pi}{2}} \sin^{\frac{d}{2}-2}(\theta)\, d\theta = \frac{W_{\frac{d}{2}-2}}{2W_{d-2}}.$$

Suppose now that $\frac{d}{2}$ is an integer. Then $\exists p \geq 1$ such that $d = 2p$. As $\lim_{n\to\infty} \sqrt{n} W_n = \sqrt{\pi/2}$ then

$$\lim_{d\to\infty} \frac{W_{\frac{d}{2}-2}}{W_{d-2}} = \lim_{p\to\infty} \frac{W_{p-2}}{W_{2p-2}} = \lim_{p\to\infty} \frac{\sqrt{2p-2}}{\sqrt{p-2}} \frac{\lim_{p\to\infty} \sqrt{p-2} W_{p-2}}{\lim_{p\to\infty} \sqrt{2p-2} W_{2p-2}} = \sqrt{2}\,.$$

If $\frac{d}{2}$ is odd, then $\frac{d-1}{2}$ is an integer and $W_{\frac{d}{2}-2} \leq W_{\frac{d-1}{2}-2}$ and we have also

$$\lim_{d\to\infty} \frac{W_{\frac{d-1}{2}-2}}{W_{d-2}} = \sqrt{2}\,.$$

Then for $d \geq d_0$, $E(G_d^2) \leq \frac{\sqrt{2}+1}{2}$. Consequently, the family $\{E(G_d^2)\}_{d\geq d_0}$ is uniformly bounded which means that the family $(\mathrm{K}^-)_{d\geq d_0}$ is uniformly integrable and therefore the family

$$\left\{ \frac{d}{2} \ln\left(A_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*)\right) \mathrm{H}_d\left(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*\right) \right\}_{d\geq 1}$$

is uniformly integrable. □

The following proposition holds as a consequence of Proposition 4.

Proposition 5 *Let* V *be the function introduced in (5). Let σ^* be a strictly positive constant. We have:*

$$\lim_{d\to+\infty} d\mathrm{V}\left(\frac{\sigma^*}{d}\right) = \sigma^* \sum_{i=1}^{\mu} w^i E\left[\mathrm{N}^i \mathrm{H}(\mathrm{N}^1, \ldots, \mathrm{N}^\mu)\right] + \frac{(\sigma^*)^2}{2} \sum_{i=1}^{\mu} (w^i)^2 \qquad (24)$$

where, for $\mathrm{N}^1, \ldots, \mathrm{N}^\mu, \mathrm{N}$, $\mu+1$ *independent random variables following the standard normal distribution* $\mathrm{N}(0,1)$[6] *we have:*

$$\mathrm{H}(\mathrm{N}^1, \ldots, \mathrm{N}^\mu) := \mathcal{P}_\lambda^\mu \quad \mathbb{1}_{\{\mathrm{N}^1 \leq \ldots \leq \mathrm{N}^\mu\}} E^{\lambda-\mu}\left[\mathbb{1}_{\{\mathrm{N}^\mu \leq \mathrm{N}\}} | \mathrm{N}^\mu\right]\,.$$

Proof of Proposition 5

We recall here that the multivariate normal distribution on $\mathbb{R}^d$ with mean $(0, \ldots, 0)$ and covariance matrix the identity I_d is simply denoted $\mathcal{N}(0, I_d)$. By Proposition 4, we have

$$d \times \mathrm{V}\left(\frac{\sigma^*}{d}\right) = E\left[\frac{d}{2} \ln\left(A_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*)\right) \mathrm{H}_d\left(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*\right)\right]\,,$$

where the families $\left(A_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*)\right)_{d\geq 1}$ and $\left(\mathrm{H}_d\left(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*\right)\right)_{d\geq 1}$ are introduced in Defintion 3 and the family

$$\left\{ \frac{d}{2} \ln\left(A_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*)\right) \mathrm{H}_d\left(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*\right) \right\}_{d\geq 1}$$

is uniformly integrable. Therefore, we have

$$\lim_{d\to\infty} d \times \mathrm{V}\left(\frac{\sigma^*}{d}\right) = E\left[\lim_{d\to\infty} \frac{d}{2} \ln\left(A_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*)\right) \mathrm{H}_d\left(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*\right)\right]\,.$$

[6] The standard normal distribution is the normal distribution with mean zero and variance of one.

Then we have to compute $\lim_{d\to\infty} \frac{d}{2} \ln\left(A_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*)\right) \mathrm{H}_d\left(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*\right)$.
Step 1: Computation of $\lim_{d\to\infty} \frac{d}{2} \ln\left(A_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*)\right)$ **:**
We define the sequences $\left\{\mathcal{U}_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*)\right\}_{d\geq 1}$, $\left\{\mathcal{V}_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*)\right\}_{d\geq 1}$ and $\left\{\mathcal{W}_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*)\right\}_{d\geq 1}$ as follows:

$$\mathcal{U}_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*) := 2\frac{\sigma^*}{d}\sum_{i=1}^{\mu} w^i(\mathbf{N}^i)_1\,,$$

$$\mathcal{V}_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*) := \frac{(\sigma^*)^2}{d}\sum_{i=1}^{\mu}(w^i)^2\left(\frac{\sum_{j=1}^d(\mathbf{N}^i)_j^2}{d}\right)\,,$$

and

$$\mathcal{W}_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*) := 2\left(\frac{\sigma^*}{d}\right)^2 \sum_{1\leq i<k\leq \mu} w^i w^k \sum_{j=1}^d (\mathbf{N}^i)_j(\mathbf{N}^k)_j$$

The sequence $\left\{A_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*)\right\}_{d\geq 1}$ introduced in Definition 3 is such that

$$A_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*) - 1 = \\ \mathcal{U}_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*) + \mathcal{V}_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*) + \mathcal{W}_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*)\,,$$

Note that $A_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*) - 1$ converges almost surely to zero when d converges to infinity as:

- $\lim_{d\to\infty} \mathcal{U}_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*) = 2\sigma^* \lim_{d\to\infty} \frac{\sum_{i=1}^\mu w^i(\mathbf{N}^i)_1}{d} = 0.$
- $\forall i \in \{1,\dots,\mu\}$, $\lim_{d\to\infty} \frac{\sum_{j=1}^d (\mathbf{N}^i)_j^2}{d} = 1$ (by the LLN for independent random variables). Then $\lim_{d\to\infty} \mathcal{V}_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*) = 0$.
- $|2\sum_{1\leq i<k\leq\mu} w^i w^k \sum_{j=1}^d (\mathbf{N}^i)_j(\mathbf{N}^k)_j| \leq 2\sum_{j=1}^d (\mathbf{N}^i)_j^2$ which gives that $\lim_{d\to\infty} \mathcal{W}_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*) = 0$

We have $\ln(1+x) \sim x$ when $x \to 0$. Therefore,

$$\lim_{d\to\infty} \frac{d}{2} \ln\left(A_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*)\right) = \lim_{d\to\infty} \frac{d}{2}\left(A_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*) - 1\right)\,.$$

Moreover, we have

$$\lim_{d\to\infty} \frac{d}{2}\mathcal{U}_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*) = \sigma^* \sum_{i=1}^{\mu} w^i(\mathbf{N}^i)_1\,,$$

$$\lim_{d\to\infty} \frac{d}{2}\mathcal{V}_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*) = \frac{(\sigma^*)^2}{2}\sum_{i=1}^{\mu}(w^i)^2 \lim_{d\to\infty}\left(\frac{\sum_{j=1}^d(\mathbf{N}^i)_j^2}{d}\right) = \frac{(\sigma^*)^2}{2}\sum_{i=1}^{\mu}(w^i)^2\,,$$

Therefore, we have

$$\lim_{d\to\infty} \frac{d}{2} \ln\left(A_d(\mathbf{N}^1,\dots,\mathbf{N}^\mu,\sigma^*)\right) = \sigma^* \sum_{i=1}^{\mu} w^i(\mathbf{N}^i)_1 \\ + \frac{(\sigma^*)^2}{2}\sum_{i=1}^{\mu}(w^i)^2 + (\sigma^*)^2\left\{\lim_{d\to\infty} \frac{\sum_{1\leq i<k\leq\mu} w^i w^k \sum_{j=1}^d (\mathbf{N}^i)_j(\mathbf{N}^k)_j}{d}\right\}\,.$$

Step 2: Computation of $\lim_{d\to\infty} \mathrm{H}_d\left(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*\right)$: First, we notice that the random variable $\mathrm{H}_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*)$ introduced in Defintion 3 can be rewritten as

$$\begin{aligned}&\mathrm{H}_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*)\\ &\quad = \mathcal{P}_\lambda^\mu \mathbb{1}_{\{d[h_d(\mathbf{N}^1)-1] \leq \ldots \leq d[h_d(\mathbf{N}^\mu)-1]\}} E^{\lambda-\mu}\left[\mathbb{1}_{\{d[h_d(\mathbf{N}^\mu)-1] \leq d[h_d(\mathbf{N})-1]\}} | \mathbf{N}^\mu\right] .\end{aligned}$$

Using the result of Lemma 3 giving the limit of the quantities $d\{h_d - 1\}$, together with the (almost sure) continuity of the indicator function and the dominated convergence theorem, we have:

$$\lim_{d\to\infty} \mathrm{H}_d(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*) = \mathrm{H}(\mathbf{N}^1, \ldots, \mathbf{N}^\mu)$$

almost surely where the random variable H is defined as:

$$\mathrm{H}(\mathbf{N}^1, \ldots, \mathbf{N}^\mu) := \mathcal{P}_\lambda^\mu \mathbb{1}_{\{(\mathbf{N}^1)_1 \leq \ldots \leq (\mathbf{N}^\mu)_1\}} E^{\lambda-\mu}\left[\mathbb{1}_{\{(\mathbf{N}^\mu)_1 \leq \mathrm{N}\}} | (\mathbf{N}^\mu)_1\right] ,$$

where N is defined as a random variable following the distribution $\mathrm{N}(0,1)$.
As the limit random variable $\mathrm{H}(\mathbf{N}^1, \ldots, \mathbf{N}^\mu)$ depends only on $((\mathbf{N}^1)_1, \ldots, (\mathbf{N}^\mu)_1)$, one can write:

$$\lim_{d\to\infty} \mathrm{H}_d\left(\mathbf{N}^1, \ldots, \mathbf{N}^\mu, \sigma^*\right) = \mathrm{H}((\mathbf{N}^1)_1, \ldots, (\mathbf{N}^\mu)_1) . \tag{25}$$

Note that $\frac{1}{(2\pi)^{\frac{\mu}{2}}} e^{-\frac{1}{2}\sum_{i=1}^\mu \left[(x^i)_1\right]^2} \mathrm{H}\left((x^1)_1, \ldots, (x^\mu)_1\right)$ is the probability density function relative to the μ random variables $(\mathbf{N}^{1:\lambda})_1, \ldots, (\mathbf{N}^{\mu:\lambda})_1$ representing the μ first order statistics among λ random variables $(\mathbf{N}^i)_1$. In fact, this quantity can be seen as an analogue to the probability density function p (given in Lemma 1) and relative to the vector $(\mathbf{N}^{1:\lambda}, \ldots, \mathbf{N}^{i:\lambda}, \ldots, \mathbf{N}^{\mu:\lambda})$. This implies that $E\left[\mathrm{H}\left((\mathbf{N}^1)_1, \ldots, (\mathbf{N}^\mu)_1\right)\right] = 1$.
Collecting the results obtained in **Step 1** and **Step 2**, one gets

$$\begin{aligned}&\lim_{d\to\infty} d \times \mathrm{V}\left(\frac{\sigma^*}{d}\right)\\ &= E\left[\left\{\sigma^* \sum_{i=1}^\mu w^i (\mathbf{N}^i)_1 + \frac{(\sigma^*)^2}{2} \sum_{i=1}^\mu (w^i)^2\right\} \mathrm{H}\left((\mathbf{N}^1)_1, \ldots, (\mathbf{N}^\mu)_1\right)\right]\\ &+ (\sigma^*)^2 E\left[\left\{\lim_{d\to\infty} \frac{\sum_{1\leq i<k\leq\mu} w^i w^k \sum_{j=1}^d (\mathbf{N}^i)_j (\mathbf{N}^k)_j}{d}\right\} \mathrm{H}\left((\mathbf{N}^1)_1, \ldots, (\mathbf{N}^\mu)_1\right)\right] .\end{aligned} \tag{26}$$

In the last equation, let us prove that

$$\begin{aligned}&E(\lambda, \mu)\\ &:= E\left[\left\{\lim_{d\to\infty} \frac{\sum_{1\leq i<k\leq\mu} w^i w^k \sum_{j=1}^d (\mathbf{N}^i)_j (\mathbf{N}^k)_j}{d}\right\} \mathrm{H}\left((\mathbf{N}^1)_1, \ldots, (\mathbf{N}^\mu)_1\right)\right] = 0 .\end{aligned}$$

We have

$$
\begin{aligned}
&E(\lambda,\mu)\\
&= \lim_{d\to\infty} E\left[\left\{\frac{\sum_{1\le i<k\le\mu} w^i w^k \sum_{j=1}^{d} (\mathbf{N}^i)_j (\mathbf{N}^k)_j}{d}\right\} \mathrm{H}\left((\mathbf{N}^1)_1,\ldots,(\mathbf{N}^\mu)_1\right)\right]\\
&= \lim_{d\to\infty} E\left[\left\{\frac{\sum_{1\le i<k\le\mu} w^i w^k (\mathbf{N}^i)_1 (\mathbf{N}^k)_1}{d}\right\} \mathrm{H}\left((\mathbf{N}^1)_1,\ldots,(\mathbf{N}^\mu)_1\right)\right]\\
&+ \lim_{d\to\infty} E\left[\left\{\frac{\sum_{1\le i<k\le\mu} w^i w^k \sum_{j=2}^{d} (\mathbf{N}^i)_j (\mathbf{N}^k)_j}{d}\right\} \mathrm{H}\left((\mathbf{N}^1)_1,\ldots,(\mathbf{N}^\mu)_1\right)\right].
\end{aligned}
$$

But we have

$$
\begin{aligned}
&E\left[\frac{\sum_{1\le i<k\le\mu} w^i w^k \sum_{j=2}^{d} (\mathbf{N}^i)_j (\mathbf{N}^k)_j}{d} \mathrm{H}\left((\mathbf{N}^1)_1,\ldots,(\mathbf{N}^\mu)_1\right)\right]\\
&\qquad = \frac{1}{d} \sum_{1\le i<k\le\mu} w^i w^k \sum_{j=2}^{d} E\left[(\mathbf{N}^i)_j (\mathbf{N}^k)_j \mathrm{H}\left((\mathbf{N}^1)_1,\ldots,(\mathbf{N}^\mu)_1\right)\right]\\
&= \frac{1}{d} E\left[\mathrm{H}\left((\mathbf{N}^1)_1,\ldots,(\mathbf{N}^\mu)_1\right)\right] \sum_{1\le i<k\le\mu} w^i w^k \sum_{j=2}^{d} E\left[(\mathbf{N}^i)_j\right] E\left[(\mathbf{N}^k)_j\right] = 0\,,
\end{aligned}
$$

as $E\left[(\mathbf{N}^i)_j\right] = 0,\ \forall i \in \{1,\ldots,\mu\},\ j \in \{1,\ldots,d\}$. Consequently, we get

$$
E(\lambda,\mu) = \lim_{d\to\infty} E\left[\left\{\frac{\sum_{1\le i<k\le\mu} w^i w^k (\mathbf{N}^i)_1 (\mathbf{N}^k)_1}{d}\right\} \mathrm{H}\left((\mathbf{N}^1)_1,\ldots,(\mathbf{N}^\mu)_1\right)\right].
$$

The random variable $\frac{\sum_{1\le i<k\le\mu} w^i w^k (\mathbf{N}^i)_1 (\mathbf{N}^k)_1}{d} \mathrm{H}\left((\mathbf{N}^1)_1,\ldots,(\mathbf{N}^\mu)_1\right)$ converges almost surely to 0 when d goes to infinity. Moreover, as $\mathrm{H}\left((\mathbf{N}^1)_1,\ldots,(\mathbf{N}^\mu)_1\right) \le \mathcal{P}_\lambda^\mu$, it is upper bounded by the random variable $\mathcal{P}_\lambda^\mu \sum_{1\le i<k\le\mu} w^i w^k (\mathbf{N}^i)_1 (\mathbf{N}^k)_1$. Then, by the dominated convergence Theorem, we have:

$$
E(\lambda,\mu) = E\left[\lim_{d\to\infty} \left\{\frac{\sum_{1\le i<k\le\mu} w^i w^k (\mathbf{N}^i)_1 (\mathbf{N}^k)_1}{d}\right\} \mathrm{H}\left((\mathbf{N}^1)_1,\ldots,(\mathbf{N}^\mu)_1\right)\right] = 0\,.
$$

Finally, (26) simplifies to

$$
\lim_{d\to\infty} d\times \mathrm{V}\left(\frac{\sigma^*}{d}\right) = E\left[\left\{\sigma^* \sum_{i=1}^{\mu} w^i (\mathbf{N}^i)_1 + \frac{(\sigma^*)^2}{2} \sum_{i=1}^{\mu} (w^i)^2\right\} \mathrm{H}\left((\mathbf{N}^1)_1,\ldots,(\mathbf{N}^\mu)_1\right)\right],
$$

and consequently

$$
\lim_{d\to\infty} d \times \mathrm{V}\left(\frac{\sigma^*}{d}\right) = \sigma^* \sum_{i=1}^{\mu} w^i \mathrm{E}^i(\lambda,\mu) + \frac{(\sigma^*)^2}{2} \sum_{i=1}^{\mu} (w^i)^2\,, \tag{27}
$$

where $\mathrm{E}^i(\lambda,\mu) := E\left[\mathrm{N}^i \mathrm{H}(\mathrm{N}^1,\ldots,\mathrm{N}^\mu)\right]$ with

$$
\mathrm{H}(\mathrm{N}^1,\ldots,\mathrm{N}^\mu) := \mathcal{P}_\lambda^\mu \mathbb{1}_{\{\mathrm{N}^1\le\ldots\le\mathrm{N}^\mu\}} E^{\lambda-\mu}\left[\mathbb{1}_{\{\mathrm{N}^\mu\le\mathrm{N}\}} | \mathrm{N}^\mu\right].
$$

where $\mathrm{N}^1,\ldots,\mathrm{N}^\mu,\mathrm{N}$ are $\mu+1$ independent random variables following the standard normal distribution in dimension 1. □

Proof of Proposition 1 (stated page 6)

Let σ be a positive constant. Let $w^i,\ i = 1, \ldots, \mu$ be μ positive constants summing to one. We recall that the random vector $\mathbf{M}(\mathrm{e}_1, \sigma) = \sum_{i=1}^{\mu} w^i \mathbf{N}^{i:\lambda}$ is obtained by recombining the random vectors $\mathbf{N}^{i:\lambda}(\mathrm{e}_1, \sigma)$ selected according to $h_{\{\mathrm{e}_1,\sigma\}}$ (defined in (10)) as expressed in (11). In the following, the function $h_{\{\mathrm{e}_1,\sigma\}}$ will be simply denoted h_σ. By Lemma 1, the probability density function of the vector $(\mathbf{N}^{1:\lambda}(\mathrm{e}_1, \sigma), \ldots, \mathbf{N}^{\mu:\lambda}(\mathrm{e}_1, \sigma))$ is given by the function $p_{\{\mathrm{e}_1,\sigma\}}$, that we will simply denote in the following p and which we rewrite as a non-negative function defined in $\mathbb{R}^{\mu d} \times \mathbb{R}^+$ in the following way:

$$p(x^1, \ldots, x^\mu, \sigma) = \frac{1}{(2\pi)^{\frac{\mu d}{2}}} e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}} \mathrm{H}(x^1, \ldots, x^\mu, \sigma)\,,$$

where the function H is given by

$$\begin{aligned}&\mathrm{H}(x^1, \ldots, x^\mu, \sigma)\\ &= \mathcal{P}_\lambda^\mu \times \mathbb{1}_{\{h_\sigma(x^1) \le h_\sigma(x^i) \ldots \le h_\sigma(x^\mu)\}} \left(x^1, \ldots, x^\mu\right) P^{\lambda-\mu}\left\{h_\sigma(x^\mu) \le h_\sigma(\mathbf{N}) | x^\mu\right\}\,,\end{aligned} \tag{28}$$

where $\mathbf{N}$ is an independent random vector following $\mathcal{N}(0, I_d)$. We define the quantities

$$\mathrm{V}^-(\sigma) := E\left[\ln^-(\|\mathrm{e}_1 + \sigma \mathbf{M}(\mathrm{e}_1, \sigma)\|)\right] \tag{29}$$

and

$$\mathrm{V}^+(\sigma) := E\left[\ln^+(\|\mathrm{e}_1 + \sigma \mathbf{M}(\mathrm{e}_1, \sigma)\|)\right]\,. \tag{30}$$

The quantities V^- and V^+ exist but could be infinite. We suppose that λ and μ are fixed and are such that $\lambda \in \mathbb{N}$ and $\mu \in \mathbb{N}$ with $\mu \le \lambda$.

Proof of (i)

Let $\mathcal{D}$ be the subset of $\mathbb{R}^{\mu d}$ containing the elements $(x^1, \ldots, x^\mu) \in \mathbb{R}^{\mu d}$ such that, for $\sigma > 0$, $\left\|\mathrm{e}_1 + \sigma \sum_{i=1}^{\mu} w^i x^i\right\| \neq 0$. Let $k^+, k^- : \mathcal{D} \times [0, +\infty[$ be defined for $(x^1, \ldots, x^\mu, \sigma)$ in $\mathcal{D} \times [0, +\infty[$ by

$$\begin{aligned}&k^+(x^1, \ldots, x^\mu, \sigma) =\\ &\qquad \frac{1}{(2\pi)^{\mu d/2}} \ln^+\left(\left\|\mathrm{e}_1 + \sigma \sum_{i=1}^{\mu} w^i x^i\right\|\right) e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}} \mathrm{H}(x^1, \ldots, x^\mu, \sigma)\,,\end{aligned}$$

and

$$\begin{aligned}&k^-(x^1, \ldots, x^\mu, \sigma) =\\ &\qquad \frac{1}{(2\pi)^{\mu d/2}} \ln^-\left(\left\|\mathrm{e}_1 + \sigma \sum_{i=1}^{\mu} w^i x^i\right\|\right) e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}} \mathrm{H}(x^1, \ldots, x^\mu, \sigma)\,,\end{aligned}$$

where the function H is defined in (28). The functions k^+ and k^- are well defined on $\mathcal{D} \times [0, +\infty[$ and the quantities V^- and V^+ introduced in (29) and (30) can be rewritten as

$$\mathrm{V}^-(\sigma) = \int_{\mathbb{R}^{\mu d}} k^-(x^1, \ldots, x^\mu, \sigma)\, dx^1 \ldots dx^\mu\,,$$

and

$$\mathrm{V}^+(\sigma) = \int_{\mathbb{R}^{\mu d}} k^+(x^1, \ldots, x^\mu, \sigma)\, dx^1 \ldots dx^\mu\,.$$

We are going to show that the quantities $\mathrm{V}^-(\sigma)$ and $\mathrm{V}^+(\sigma)$ are well defined for all σ positive and are continuous with respect to this variable. The proof rely on the Lebesgue Dominated Convergence Theorem for Continuity. In the beginning, let us remark that, for almost all $(x^1, \ldots, x^\mu) \in \mathcal{D}$, the map $\sigma \mapsto \mathrm{H}(x^1, \ldots, x^\mu, \sigma)$ is continuous on $[0, +\infty[$ thanks to the Lebesgue Dominated Convergence Theorem for Continuity. Therefore, for almost all $(x^1, \ldots, x^\mu) \in \mathcal{D}$, the maps $\sigma \mapsto k^-(x^1, \ldots, x^\mu, \sigma)$ and $\sigma \mapsto k^+(x^1, \ldots, x^\mu, \sigma)$ are continuous on $[0, +\infty[$. Moreover, we have for all $(x^1, \ldots, x^\mu, \sigma)$ in $\mathcal{D} \times [0, +\infty[$, $\mathrm{H}(x^1, \ldots, x^\mu, \sigma) \leq \mathcal{P}^\mu_\lambda$.

Case of V^+:

Let $S > 0$, we then have for almost all $(x^1, \ldots, x^\mu)$ in $\mathcal{D}$ and all $(\sigma) \in [0, S]$,

$$\begin{aligned}
k^+(x^1, \ldots, x^\mu, \sigma) &\leq \frac{\mathcal{P}^\mu_\lambda}{(2\pi)^{\mu d/2}} \ln^+ \left(\left\| \mathrm{e}_1 + \sigma \sum_{i=1}^{\mu} w^i x^i \right\| \right) e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}} \\
&\leq \frac{\mathcal{P}^\mu_\lambda}{(2\pi)^{\mu d/2}} \ln^+ \left(1 + \sigma \sum_{i=1}^{\mu} w^i \|x^i\| \right) e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}} \\
&\leq \frac{\mathcal{P}^\mu_\lambda}{(2\pi)^{\mu d/2}} \sigma \sum_{i=1}^{\mu} w^i \|x^i\| e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}} \\
&\leq \frac{\mathcal{P}^\mu_\lambda}{(2\pi)^{\mu d/2}} \sigma \sum_{i=1}^{\mu} \|x^i\| e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}} \\
&\leq \frac{\mathcal{P}^\mu_\lambda}{(2\pi)^{\mu d/2}} \sigma \sum_{i=1}^{\mu} \|x^i\| e^{-\frac{\|x^i\|^2}{2}} \\
&\leq \frac{\mathcal{P}^\mu_\lambda}{(2\pi)^{\mu d/2}} S \sum_{i=1}^{\mu} \|x^i\| e^{-\frac{\|x^i\|^2}{2}}
\end{aligned} \tag{31}$$

As for almost all $(x^1, \ldots, x^\mu) \in \mathcal{D}$, the map $\sigma \mapsto k^+(x^1, \ldots, x^\mu, \sigma)$ is continuous on $[0, S]$ and that the norm of a random vector following the distribution $\mathcal{N}(0, I_d)$ has a finite expectation (i.e., $\int_{\mathbb{R}^d} \|x\| e^{-\frac{\|x\|^2}{2}} < +\infty$) then $(x^1, \ldots, x^\mu) \mapsto \sum_{i=1}^{\mu} \|x^i\| x e^{-\frac{\|x^i\|^2}{2}}$ has a finite expectation on $\mathcal{D}$, and consequently, thanks to the Lebesgue Dominated Convergence Theorem for Continuity, we can state that $\mathrm{V}^+(\sigma)$ is well defined, finite and continuous on every set $[0, S]$ and consequently well defined, finite and continuous on $[0, +\infty[$.

Case of V^-:

Let us now give interest to the quantity $\mathrm{V}^-(\sigma)$. We have

$$k^-(x^1, \ldots, x^\mu, \sigma) \leq \frac{\mathcal{P}^\mu_\lambda}{2(2\pi)^{\mu d/2}} \ln^- \left(\left\| \mathrm{e}_1 + \sigma \sum_{i=1}^{\mu} w^i x^i \right\|^2 \right) e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}} \tag{32}$$

The right hand-side of the last inequality is equal to 0 if $(\sum_{i=1}^{\mu} w^i x^i)_1 \geq 0$. The study of the function $\sigma \mapsto \|\mathrm{e}_1 + \sigma \sum_{i=1}^{\mu} w^i x^i\|^2$ shows that, for $(x^1, \ldots, x^\mu) \in \mathbb{R}^{\mu d}$ such that

$(\sum_{i=1}^{\mu} w^i x^i)_1 < 0$, this function is lower bounded, for $d \geq 2$, by $1 - \frac{\left(\sum_{i=1}^{\mu} w^i x^i\right)_1^2}{\|\sum_{i=1}^{\mu} w^i x^i\|^2}$. This gives, for $d \geq 2$

$$
\begin{aligned}
&k^-(x^1, \ldots, x^\mu, \sigma) \\
&\leq \frac{\mathcal{P}_\lambda^\mu}{2(2\pi)^{\mu d/2}} \ln^- \left(1 - \frac{\left(\sum_{i=1}^{\mu} w^i x^i\right)_1^2}{\|\sum_{i=1}^{\mu} w^i x^i\|^2}\right) e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}} \mathbb{1}_{\{(\sum_{i=1}^{\mu} w^i x^i)_1 < 0\}} \left(x^1, \ldots, x^\mu\right) \\
&\leq \frac{\mathcal{P}_\lambda^\mu}{2(2\pi)^{\mu d/2}} \ln^- \left(1 - \frac{\left(\sum_{i=1}^{\mu} w^i x^i\right)_1^2}{\|\sum_{i=1}^{\mu} w^i x^i\|^2}\right) e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}}
\end{aligned}
\tag{33}
$$

We have already seen that, for almost all $(x^1, \ldots, x^\mu) \in \mathcal{D}$, the map $\sigma \mapsto k^-(x^1, \ldots, x^\mu, \sigma)$ is continuous on $[0, +\infty[$. In order to apply the Lebesgue Dominated Convergence Theorem for Continuity, we have to check that the upper bound (which is independent of σ) given in (33) has a finite expectation. Let $\mathbf{N}(0, C)$ be a random vector following the distribution $\mathcal{N}(0, C)$ and let $\mathbf{N}_1(0, C)$ its first coordinate. We recall that if $C = I_d$, the random vector $\mathbf{N}(0, C)$ is simply denoted $\mathbf{N}$. We have:

$$
\begin{aligned}
&\int_{\mathcal{D}} \frac{\mathcal{P}_\lambda^\mu}{2(2\pi)^{\mu d/2}} \ln^- \left(1 - \frac{\left(\sum_{i=1}^{\mu} w^i x^i\right)_1^2}{\|\sum_{i=1}^{\mu} w^i x^i\|^2}\right) e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}} dx^1 \ldots dx^\mu \\
&\quad \leq \int_{\mathbb{R}^{\mu d}} \frac{\mathcal{P}_\lambda^\mu}{2(2\pi)^{\mu d/2}} \ln^- \left(1 - \frac{\left(\sum_{i=1}^{\mu} w^i x^i\right)_1^2}{\|\sum_{i=1}^{\mu} w^i x^i\|^2}\right) e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}} dx^1 \ldots dx^\mu \\
&\quad = \frac{\mathcal{P}_\lambda^\mu}{2} E\left[\ln^- \left(1 - \frac{\left(\mathbf{N}_1(0, \sum_{i=1}^{\mu} (w^i)^2 I_d)\right)^2}{\|\mathbf{N}(0, \sum_{i=1}^{\mu} (w^i)^2 I_d)\|^2}\right)\right] \\
&\quad = \frac{\mathcal{P}_\lambda^\mu}{2} E\left[\ln^- \left(1 - \frac{\left(\sum_{i=1}^{\mu} (w^i)^2\right) (\mathbf{N}_1)^2}{\left(\sum_{i=1}^{\mu} (w^i)^2\right) \|\mathbf{N}\|^2}\right)\right] \\
&\quad = \mathcal{P}_\lambda^\mu E\left[\frac{1}{2} \ln^- \left(1 - \frac{(\mathbf{N}_1)^2}{\|\mathbf{N}\|^2}\right)\right] .
\end{aligned}
\tag{34}
$$

The last expectation can be rewritten as:

$$
E\left[\frac{1}{2} \ln^- \left(1 - \frac{\left(\mathbf{N}_1(0, I_d)\right)^2}{\|\mathbf{N}(0, I_d)\|^2}\right)\right] = \frac{1}{2(2\pi)^{d/2}} \int_{\mathbb{R}^d} \ln^- \left(1 - \frac{(x^1)^2}{\|x\|^2}\right) e^{-\frac{\|x\|^2}{2}} dx
$$

Changing to spherical coordinates (with $d \geq 2$) we obtain after partial integration

$$
\begin{aligned}
&E\left[\frac{1}{2} \ln^- \left(1 - \frac{(\mathbf{N}_1)^2}{\|\mathbf{N}\|^2}\right)\right] \\
&= \frac{1}{8 W_{d-2} 2^{\left(\frac{d}{2}-1\right)} \Gamma\left(\frac{d}{2}\right)} \int_0^{+\infty} \int_0^{\pi} \ln^- \left(1 - \cos^2(\theta)\right) r^{d-1} e^{-\frac{r^2}{2}} \sin^{d-2}(\theta) \, dr \, d\theta \,,
\end{aligned}
$$

where for $n \in \mathbb{Z}^+$, $W_n = \int_0^{\pi/2} \sin^n \theta \, d\theta$ is the classical Wallis integral and for $z \in \mathbb{C}$ such that $Re(z) > 0, \Gamma(z) = \int_0^{+\infty} e^{-u} u^{z-1} du$ is the Gamma function. The last equation can be rewritten as:

$$
E\left[\frac{1}{2} \ln^- \left(1 - \frac{(\mathbf{N}_1)^2}{\|\mathbf{N}\|^2}\right)\right] = \frac{1}{8 W_{d-2}} \int_0^{\pi} \ln^- \left(1 - \cos^2(\theta)\right) \sin^{d-2}(\theta) \, d\theta \,,
$$

which,in turn writes as:

$$E\left[\frac{1}{2}\ln^-\left(1-\frac{(\mathbf{N}_1)^2}{\|\mathbf{N}\|^2}\right)\right] = \frac{1}{2W_{d-2}}\int_0^{\frac{\pi}{2}} \ln^-\left(\sin(\theta)\right)\sin^{d-2}(\theta)\,d\theta\,, \tag{35}$$

and finally, we get

$$\begin{aligned} E\left[\frac{1}{2}\ln^-\left(1-\frac{(\mathbf{N}_1)^2}{\|\mathbf{N}\|^2}\right)\right] &= \frac{1}{W_{d-2}}\int_0^{\frac{\pi}{2}} \ln\left(\left(\frac{1}{\sin(\theta)}\right)^{\frac{1}{2}}\right)\sin^{d-2}(\theta)\,d\theta \\ &\leq \frac{1}{W_{d-2}}\int_0^{\frac{\pi}{2}} \sin^{d-\frac{5}{2}}(\theta)\,d\theta = \frac{W_{d-\frac{5}{2}}}{W_{d-2}}\,. \end{aligned} \tag{36}$$

The quantity $\frac{W_{d-\frac{5}{2}}}{W_{d-2}}$ is finite (with $d \geq 2$). Then the right hand-side of (33) is finite and using the Lebesque dominated convergence theorem for continuity, one can state that the quantity $\mathrm{V}^-(\sigma)$ is well defined, finite and continuous with respect to σ when $d \geq 2$.

$\underline{\mathrm{V}^-(\sigma) < +\infty \text{ for } d = 1}$

Let $d = 1$ and let $\mathrm{N}^1, \ldots, \mathrm{N}^\mu, \mathrm{N}$ be $\mu + 1$ random variables i.i.d. with common law $\mathcal{N}(0,1)$. We have:

$$\begin{aligned} \mathrm{V}^-(\sigma) &= \int_{\mathbb{R}^\mu} k^-(x^1, \ldots, x^\mu, \sigma) dx^1 \ldots dx^\mu \\ &\leq \mathcal{P}_\lambda^\mu \frac{1}{(2\pi)^{\mu/2}} \ln^-\left(\left|1+\sigma\sum_{i=1}^\mu w^i x^i\right|\right) e^{-\frac{\sum_{i=1}^\mu |x^i|^2}{2}} dx^1 \ldots dx^\mu \\ &= \mathcal{P}_\lambda^\mu E\left[\ln^-\left(\left|1+\sigma\sum_{i=1}^\mu w^i \mathrm{N}^i\right|\right)\right] \\ &= \mathcal{P}_\lambda^\mu E\left[\ln^-\left(\left|1+\sigma w \mathrm{N}\right|\right)\right] \text{ with } w := \left(\sum_{i=1}^\mu (w^i)^2\right)^{\frac{1}{2}} \geq 0\,. \end{aligned} \tag{37}$$

After a change of variables $y = \sigma w x$, the integrand in $\mathrm{V}^-(\sigma)$ will be dominated by $\frac{e^{-\frac{1}{2}}}{\sqrt{2\pi}}\frac{\ln(|1+y|)}{y}$ for $(y,\sigma) \in]-2,0] \times [0,+\infty[$ which has a finite expectation and therefore $\mathrm{V}^-(\sigma) < +\infty$ for all σ positive. □

Proposition needed for proving (ii) and (iii)

In order to prove (ii) and (iii) of Proposition 1, we will need the following proposition.

Proposition 6 *For $d \geq 2$, the function V introduced in (5) is lower bounded by the strictly negative bound $c(d)$ given by*

$$c(d) := \begin{cases} -\mathcal{P}_\lambda^\mu \dfrac{\int_0^{\frac{\pi}{2}} \sin^{-\frac{1}{2}}(\theta)\,d\theta}{\pi} & \text{if } d = 2 \\ -\mathcal{P}_\lambda^\mu \dfrac{1}{2(d-2)}\left(\dfrac{W_{\frac{1}{2}(d-2)}}{W_{d-2}} - 1\right) & \text{if } d \geq 3. \end{cases}$$

Proof. For $\sigma > 0$, let $\mathrm{V}^+(\sigma)$ and $\mathrm{V}^-(\sigma)$ be the positive quantities defined respectively in (30) and (29). We have $\mathrm{V}(\sigma) = \mathrm{V}^+(\sigma) - \mathrm{V}^-(\sigma)$ where V^+ and V^- are finite for all $\sigma \geq 0$ and all $d \geq 1$ by (i) of Proposition 1. Taking from the proof of (i) of Proposition 1 the first inequality of (33), and integrating over $\mathbb{R}^{\mu d}$, one can write:

$$\begin{aligned}\mathrm{V}^-(\sigma) &\leq \mathcal{P}_\lambda^\mu E\left[\frac{1}{2}\ln^-\left(1 - \frac{(\mathbf{N}_1)^2}{\|\mathbf{N}\|^2}\mathbb{1}_{\{\mathbf{N}_1<0\}}\right)\right] \\ &= \frac{\mathcal{P}_\lambda^\mu}{2} E\left[\frac{1}{2}\ln^-\left(1 - \frac{(\mathbf{N}_1)^2}{\|\mathbf{N}\|^2}\right)\right],\end{aligned}$$

where $\mathbf{N}$ is an independent random vector following $\mathcal{N}(0, I_d)$ and $\mathbf{N}_1$ is its first coordinate. The relation (35) taken from the proof of (i) of Proposition 1 gives, due to the fact that $\ln(t) \leq t - 1$ for $t \geq 1$, and assuming that $d \geq 3$:

$$\begin{aligned}\mathrm{V}^-(\sigma) &\leq \frac{\mathcal{P}_\lambda^\mu}{4}\frac{1}{W_{d-2}}\int_0^{\frac{\pi}{2}} \ln\left(\frac{1}{\sin(\theta)}\right)\sin^{d-2}(\theta)\,d\theta \\ &= \frac{\mathcal{P}_\lambda^\mu}{2(d-2)}\frac{1}{W_{d-2}}\int_0^{\frac{\pi}{2}} \ln\left(\left(\frac{1}{\sin(\theta)}\right)^{\frac{d-2}{2}}\right)\sin^{d-2}(\theta)\,d\theta \\ &\leq \frac{\mathcal{P}_\lambda^\mu}{2(d-2)}\frac{1}{W_{d-2}}\int_0^{\frac{\pi}{2}} \left(\left(\frac{1}{\sin(\theta)}\right)^{\frac{d-2}{2}} - 1\right)\sin^{d-2}(\theta)\,d\theta \\ &= \frac{\mathcal{P}_\lambda^\mu}{2(d-2)}\frac{1}{W_{d-2}}\int_0^{\frac{\pi}{2}} \left(\sin^{\frac{1}{2}(d-2)}(\theta) - \sin^{d-2}(\theta)\right)\,d\theta \\ &= \frac{\mathcal{P}_\lambda^\mu}{2(d-2)}\left(\frac{W_{\frac{1}{2}(d-2)}}{W_{d-2}} - 1\right).\end{aligned}$$

Note that the upper bound of the last inequality is positive as the sequence $(W_d)_{d\geq 0}$ is a decreasing sequence. Moreover, as $\lim_{d\to+\infty} W_d\sqrt{d} = \sqrt{\frac{\pi}{2}}$, we have $\lim_{d\to+\infty} \frac{W_{\frac{1}{2}(d-2)}}{W_{d-2}} = \sqrt{2}$ and then $\frac{W_{\frac{1}{2}(d-2)}}{W_{d-2}}$ is upper bounded. For $d = 2$, we get, replacing in (36) d by 2 and W_0 by $\frac{\pi}{2}$:

$$\mathrm{V}^-(\sigma) \leq \frac{\mathcal{P}_\lambda^\mu}{\pi}\int_0^{\frac{\pi}{2}} \sin^{-\frac{1}{2}}(\theta)\,d\theta < +\infty\,.$$

For $d \geq 2$, let $c(d)$ be the strictly negative quantity given by

$$c(d) := \begin{cases} -\mathcal{P}_\lambda^\mu \dfrac{\int_0^{\frac{\pi}{2}} \sin^{-\frac{1}{2}}(\theta)\,d\theta}{\pi} & \text{if } d = 2 \\ -\mathcal{P}_\lambda^\mu \dfrac{1}{2(d-2)}\left(\dfrac{W_{\frac{1}{2}(d-2)}}{W_{d-2}} - 1\right) & \text{if } d \geq 3. \end{cases}$$

Then we have:

$$\mathrm{V}^-(\sigma) \leq -c(d) < +\infty,\ \forall d \geq 2 \tag{38}$$

Then, using the fact that $\mathrm{V}(\sigma) = \mathrm{V}^+(\sigma) - \mathrm{V}^-(\sigma)$, we get $\mathrm{V}(\sigma) \geq c(d)$ for all $d \geq 2$. □

Proof of (ii)

Let us recall the expression of $\mathrm{V}^+(\sigma)$ introduced in (30). We have:

$$\mathrm{V}^+(\sigma) = \int_{\mathbb{R}^{\mu d}} \frac{1}{(2\pi)^{\mu d/2}} \ln^+ \left(\left\| \mathrm{e}_1 + \sigma \sum_{i=1}^{\mu} w^i x^i \right\| \right) \mathrm{H}(x^1, \ldots, x^\mu, \sigma)\, dx^1 \ldots dx^\mu$$

where the function H is given by:

$$\mathrm{H}(x^1, \ldots, x^\mu, \sigma) =$$
$$\mathcal{P}_\lambda^\mu \times \mathbb{1}_{\{h_\sigma(x^1) \leq h_\sigma(x^i) \ldots \leq h_\sigma(x^\mu)\}} \left(x^1, \ldots, x^\mu\right) P^{\lambda-\mu} \left\{ h_\sigma(x^\mu) \leq h_\sigma(\mathbf{N}) | x^\mu, y^\mu \right\} .$$

with $\mathbf{N}$ is an independent random vector following $\mathcal{N}(0, I_d)$ and $h_\sigma(a) = \|\mathrm{e}_1 + \sigma a\|$, $\forall\, a \in \mathbb{R}^d$. We remark that, for all $a_1, a_2 \in \mathbb{R}^d$ and for $\sigma \geq 0$, if $(a_1)_1 \leq (a_2)_1$, $\|a_1\| \leq \|a_2\|$ then $h_\sigma(a_1) \leq h_\sigma(a_2)$. Therefore, H is lower bounded as the following:

$$\mathrm{H}(x^1, \ldots, x^\mu, \sigma) \geq \mathcal{P}_\lambda^\mu e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}} \mathbb{1}_{\left\{\cap_{i=1}^{\mu-1} (\|x^i\| \leq \|x^{i+1}\|)\right\}} \left(x^1, \ldots, x^\mu\right)$$
$$P^{\lambda-\mu} \left\{ \|x^\mu\| \leq \|\mathbf{N}\| \cap (x^\mu)_1 \leq \mathbf{N}_1 | x^\mu \right\} \mathbb{1}_{\left\{\cap_{i=1}^{\mu-1} ((x^i)_1 \leq (x^{i+1})_1)\right\}} \left(x^1, \ldots, x^\mu\right) .$$

Let J be the function, independent of σ, and mapping $\mathbb{R}^{\mu d}$ into $\mathbb{R}^+$ defined as

$$\mathrm{J}(x^1, \ldots, x^\mu) := \mathcal{P}_\lambda^\mu e^{-\frac{\sum_{i=1}^{\mu} \|x^i\|^2}{2}} \times$$
$$\mathbb{1}_{\left\{\| \sum_{i=1}^{\mu} w^i x^i \| \geq 1\right\}} \left(x^1, \ldots, x^\mu\right) \mathbb{1}_{\left\{\cap_{i=1}^{\mu-1} (\|x^i\| \leq \|x^{i+1}\|)\right\}} \left(x^1, \ldots, x^\mu\right)$$
$$P^{\lambda-\mu} \left\{ \|x^\mu\| \leq \|\mathbf{N}\| \cap (x^\mu)_1 \leq \mathbf{N}_1 | x^\mu \right\} \mathbb{1}_{\left\{\cap_{i=1}^{\mu-1} ((x^i)_1 \leq (x^{i+1})_1)\right\}} \left(x^1, \ldots, x^\mu\right) .$$

Then we have

$$\mathrm{V}^+(\sigma) \geq \int_{\mathbb{R}^{\mu d}} \frac{1}{2(2\pi)^{\mu d/2}} \ln^+ \left(\left\| \mathrm{e}_1 + \sigma \sum_{i=1}^{\mu} w^i x^i \right\|^2 \right) \mathrm{J}(x^1, \ldots, x^\mu)\, dx^1 \ldots dx^\mu .$$

A sufficient condition for the convergence of $\mathrm{V}^+(\sigma)$ to infinity is that the right hand side of the last equation converges to infinity when σ goes to infinity. We are going to prove this fact using the monotone convergence Theorem. A simple study of the function $g(\sigma) = \|\mathrm{e}_1 + \sigma y\|^2$ (for a given $y \in \mathbb{R}^d$ fixed) shows that, for $\sigma > \sigma_0 = -\frac{y_1}{\|y\|^2}$, the function g is increasing as a function of σ. Note that if $\|y\| \geq 1$, then $\sigma_0 \leq 1$ and consequently $\forall y \in \mathbb{R}^d$ such that $\|y\| \geq 1$ g is increasing on $[1, +\infty[$. Using these ideas, we conclude that the integrand of the right hand side of the last equation is increasing as a function of σ when $\sigma \geq 1$. Moreover, the limit of this integrand when σ goes to infinity is $+\infty$.Then, according to the monotone convergence theorem, the right hand side of the last equation converges to infinity if σ goes to infinity. Consequently, the left hand side of the same equation, i.e., $\mathrm{V}^+(\sigma)$, goes to $+\infty$ when σ goes to infinity. Now, by (38), we have that for $d \geq 2$, $-\mathrm{V}^-(\sigma) \geq -c(d)$, where V^- is defined in (29) and $c(d)$ depends only on d and do not depend on σ. Then, using the fact that $\mathrm{V}(\sigma) = \mathrm{V}^+(\sigma) - \mathrm{V}^-(\sigma)$ we have:

$$\mathrm{V}(\sigma) \geq -c(d) + \int_{\mathbb{R}^{\mu d}} \frac{1}{2(2\pi)^{\mu d/2}} \ln^+ \left(\left\| \mathrm{e}_1 + \sigma \sum_{i=1}^{\mu} w^i x^i \right\|^2 \right) \mathrm{J}(x^1, \ldots, x^\mu)\, dx^1 \ldots dx^\mu .$$

The limit of the left hand side of the previous equation is $+\infty$ when σ goes to infinity. Therefore $\mathrm{V}(\sigma)$ goes to $+\infty$ when σ goes to infinity. □

A Result needed for proving (iii)

A result needed to prove (iii) of Proposition 1 is stated in the following lemma.

Lemma 4 *For $i \in \{1, \ldots, \lambda\}$, let $\mathrm{N}^{i:\lambda}(0,1)$ be the i^{th} order statistics among λ independent random variables following the distribution $\mathcal{N}(0,1)$. Then $E\left(\mathrm{N}^{i:\lambda}(0,1)\right) \leq 0$ for $i \leq \frac{\lambda}{2}$.*

Proof. The probability distribution function of $\mathrm{N}^{i:\lambda}(0,1)$ is given by:

$$p_{\mathrm{N}^{i:\lambda}(0,1)}(x) = \frac{\lambda!}{(i-1)!(\lambda-i)!}\phi(x)^{i-1}(1-\phi(x))^{\lambda-i}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\,,$$

where ϕ is the cumulative distribution function of $\mathrm{N}(0,1)$. Then $E\left(\mathrm{N}^{i:\lambda}(0,1)\right)$ writes as:

$$E\left(\mathrm{N}^{i:\lambda}(0,1)\right) = \int_{-\infty}^{0} x p_{\mathrm{N}^{i:\lambda}(0,1)}(x)dx + \int_{0}^{+\infty} x p_{\mathrm{N}^{i:\lambda}(0,1)}(x)dx\,.$$

Using the change of variables in the integral over $\mathbb{R}^-$ $u = -x$ and using the fact that $\phi(-u) = 1 - \phi(u)$, we get

$$\begin{aligned} E\left(\mathrm{N}^{i:\lambda}(0,1)\right) &= -\int_{0}^{+\infty} u p_{\mathrm{N}^{i:\lambda}(0,1)}(-u)du + \int_{0}^{+\infty} x p_{\mathrm{N}^{i:\lambda}(0,1)}(x)dx \\ &= \int_{0}^{+\infty} x \frac{\lambda!}{(i-1)!(\lambda-i)!}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} g(\phi(x))dx \end{aligned}$$

where $g(a) = a^{i-1}(1-a)^{\lambda-i} - (1-a)^{i-1}(a)^{\lambda-i}$ for $a \in [\frac{1}{2}, 1]$. Let us show that $g(a) \leq 0\ \forall a \in [\frac{1}{2}, 1]$. We have $g(a) = a^{i-1}(1-a)^{\lambda-i}\left(1 - (\frac{a}{1-a})^{\lambda-2i+1}\right)$. As $a \geq \frac{1}{2}$ then $\frac{a}{1-a} \geq 1$ which implies that $(\frac{a}{1-a})^{\lambda-2i+1} \geq 1$ (as $i \leq \mu \leq \frac{\lambda}{2}$). Consequently $g(a) \leq 0$ for $a \geq \frac{1}{2}$. Therefore, in the last equation of the computation of $E\left(\mathrm{N}^{i:\lambda}\right)$, $g(\phi(x)) \leq 0$ as $\phi(x) \geq \frac{1}{2}$ when x is positive. Therefore, $E\left(\mathrm{N}^{i:\lambda}(0,1)\right) \leq 0$ for $i \leq \frac{\lambda}{2}$. □

Proof of (iii)

In this proof, we suppose that $\mu \leq \frac{\lambda}{2}$. In order to show that $\exists\ \sigma > 0$ such that $\mathrm{V}(\sigma) < 0$, we suppose that $\forall \sigma > 0$, $\mathrm{V}(\sigma) \geq 0$. This particularly implies that $\forall \sigma^* > 0$ and $\forall d \geq 1$, $d\mathrm{V}\left(\frac{\sigma^*}{d}\right) \geq 0$. Therefore $\lim_{d\to+\infty} d\mathrm{V}\left(\frac{\sigma^*}{d}\right) \geq 0$. On the other hand we have by (24):

$$\lim_{d\to+\infty} d\mathrm{V}\left(\frac{\sigma^*}{d}\right) = \sigma^* \sum_{i=1}^{\mu} w^i E\left(\mathrm{N}^{i:\lambda}(0,1)\right) + \frac{(\sigma^*)^2}{2}\sum_{i=1}^{\mu}(w^i)^2$$

where, for $i \in \{1, \ldots, \lambda\}$, $\mathrm{N}^{i:\lambda}$ is the i^{th} order statistics among λ independent random variables following the distribution of $\mathrm{N}(0,1)$. By Lemma 4, we have $E\left(\mathrm{N}^{i:\lambda}(0,1)\right) \leq 0$ as $i \leq \mu \leq \frac{\lambda}{2}$. Therefore, there exists values of σ^* such that the right hand-side of previous equation is strictly negative. This means that there exists $\sigma^* > 0$ such that $\lim_{d\to+\infty} d\mathrm{V}\left(\frac{\sigma^*}{d}\right) < 0$ which is in contradiction with our hypothesis. Consequently, $\exists\ \bar{\sigma} > 0$ such that $\mathrm{V}(\bar{\sigma}) < 0$. □

Proof of (iv)

Let $\mu \leq \frac{\lambda}{2}$. By Proposition 6, (ii) and (iii) of the Proposition, we know that $\lim_{\sigma\to+\infty} \mathrm{V}(\sigma) = +\infty$, that V is lower bounded by $c(d) < 0$ and that $\exists\ \bar{\sigma} > 0$ such that $\mathrm{V}(\bar{\sigma}) < 0$. Therefore, as $\sigma \mapsto \mathrm{V}(\sigma)$ is continuous ((i) of the Proposition), we have:

$$\mathrm{V}([0, +\infty[) = [\min_{\sigma\geq 0} \mathrm{V}(\sigma), \lim_{\sigma\to+\infty} \mathrm{V}(\sigma)[= [\min_{\sigma\geq 0} \mathrm{V}(\sigma), +\infty[\subset [c(d), +\infty[$$

As $\mathrm{V}(0) = 0$ and $\min_{\sigma\geq 0} \mathrm{V}(\sigma) \leq \mathrm{V}(\bar{\sigma}) < 0$ then

$$\min_{\sigma\geq 0} \mathrm{V}(\sigma) = \min_{\sigma>0} \mathrm{V}(\sigma) = \mathrm{V}(\sigma_{opt}) < 0$$

with $\sigma_{opt} := arg\min_{\sigma>0} \mathrm{V}(\sigma)$. □

Proof of (v)

By (iv) of the proposition, for $d, \lambda \geq 2$, there exists for every μ such that $\mu \leq \lambda/2$, an optimal value of σ, $\sigma_{opt}(\mu)$, such that $\mathrm{V}(\sigma_{opt}(\mu)) = \min_{\sigma>0} \mathrm{V}(\sigma(\mu))$. Let us now look for the minimal value that can be reached by V on the set $\{\sigma > 0; \mu \leq \lambda/2$. Then, it can be defined as: $\min_{\{\sigma>0;\mu\leq\lambda/2\}} \mathrm{V}(\sigma, \mu) = \min\{\mathrm{V}(\sigma(1), \ldots, \mathrm{V}(\sigma(\lambda/2)\} =$. As the set $\{\mathrm{V}(\sigma(1), \ldots, \mathrm{V}(\sigma(\lambda/2))\}$ is finite then we can state that there exists $((\sigma_{opt}, \mu_{opt})$ such that

$$\mathrm{V}(\sigma_{opt}, \mu_{opt}) = arg \min_{\{\sigma>0,\mu\leq\lambda/2\}} \mathrm{V}(\sigma, \mu)\,.$$

□

Proof of Theorem 1 (stated page 7)

Proof. The increase (or decrease) of the distance to the optimum at an iteration n can be measured by the quantity $\frac{\|\mathbf{X}_{n+1}\|}{\|\mathbf{X}_n\|}$[7] which writes, according to (3) and (12), as

$$\frac{\|\mathbf{X}_{n+1}\|}{\|\mathbf{X}_n\|} = \left\| \frac{\mathbf{X}_n}{\|\mathbf{X}_n\|} + \frac{\sigma_n}{\|\mathbf{X}_n\|}\mathbf{M}_n\left(\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|}, \frac{\sigma_n}{\|\mathbf{X}_n\|}\right)\right\|.$$

This implies that

$$\ln\left(\|\mathbf{X}_{n+1}\|\right) - \ln\left(\|\mathbf{X}_n\|\right) = \ln\left(\left\| \frac{\mathbf{X}_n}{\|\mathbf{X}_n\|} + \frac{\sigma_n}{\|\mathbf{X}_n\|}\mathbf{M}_n\left(\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|}, \frac{\sigma_n}{\|\mathbf{X}_n\|}\right)\right\|\right).$$

Thanks to Lemma 2 and Proposition 1, we have

$$\begin{aligned}
&E\left(\ln\left(\frac{\|\mathbf{X}_{n+1}\|}{\|\mathbf{X}_n\|}\right) \mid \boldsymbol{\sigma}(\mathbf{X}_n, \sigma_n)\right) \\
&= E\left(\ln\left(\left\|\mathrm{e}_1 + \frac{\sigma_n}{\|\mathbf{X}_n\|}\mathbf{M}_n\left(\mathrm{e}_1, \frac{\sigma_n}{\|\mathbf{X}_n\|}\right)\right\|\right) \mid \boldsymbol{\sigma}(\mathbf{X}_n, \sigma_n)\right) \\
&\geq \inf_{\sigma\geq 0} E\left(\ln\left(\|\mathrm{e}_1 + \sigma\mathbf{M}\left(\mathrm{e}_1, \sigma\right)\|\right)\right) \\
&= \mathrm{V}(\sigma_{opt})
\end{aligned} \tag{39}$$

[7] It is possible to divide by $\|\mathbf{X}_n\|$ thanks to Proposition 3 with $\mathbf{x}^* = (0, \ldots, 0) \in \mathbb{R}^d$.

Now, let us remark that in the specific scale-invariant adaptation rule with $\sigma_n = \sigma_{opt}\|\mathbf{X}_n\|$, we have:

$$E\left(\ln\left(\frac{\|\mathbf{X}_{n+1}\|}{\|\mathbf{X}_n\|}\right) \mid \boldsymbol{\sigma}(\mathbf{X}_n, \sigma_n)\right) = E\left(\|\mathrm{e}_1 + \sigma_{opt}\mathbf{M}\left(\mathrm{e}_1, \sigma_{opt}\right)\| \mid \boldsymbol{\sigma}(\mathbf{X}_n)\right)$$
$$= E\left(\|\mathrm{e}_1 + \sigma_{opt}\mathbf{M}\left(\mathrm{e}_1, \sigma_{opt}\right)\|\right) = \mathrm{V}(\sigma_{opt})\,.$$

This means that the best distance decrease in expectation is given when using $\sigma_n = \sigma_{opt}\|\mathbf{X}_n\|$. Taking the expectation in (39), we get

$$E\left(\ln\left(\frac{\|\mathbf{X}_{n+1}\|}{\|\mathbf{X}_n\|}\right)\right) \geq \mathrm{V}(\sigma_{opt}, 0)\,.$$

Summing such inequalities from 0 to n and dividing by n, we get

$$\frac{1}{n}E\left(\ln\left(\frac{\|\mathbf{X}_n\|}{\|\mathbf{X}_0\|}\right)\right) \geq \mathrm{V}(\sigma_{opt}, 0)\,.$$

□

Proof of Proposition 2 (stated page 8)

At each iteration n, the recurrence relation (3) and the relation (12) gives

$$\|\mathbf{X}_{n+1}\| = \|\mathbf{X}_n + \sigma\|\mathbf{X}_n\|\mathbf{M}_n\left(\mathbf{X}_n, \sigma\|\mathbf{X}_n\|\right)\|\,.$$

The result obtained in Proposition 3 was shown for a more general algorithm than the one investigated here. Therefore, by this proposition, we can state that for all $n \geq 0$, $\|\mathbf{X}_n\| > 0$ almost surely and we can write

$$\|\mathbf{X}_{n+1}\| = \|\mathbf{X}_n\|\left\|\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|} + \sigma\mathbf{M}_n\left(\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|}, \sigma\right)\right\| \; a.s.$$

Taking the logarithm of the previous equation, we get

$$\ln\left(\|\mathbf{X}_{n+1}\|\right) = \ln\left(\|\mathbf{X}_n\|\right) + \ln\left(\left\|\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|} + \sigma\mathbf{M}_n\left(\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|}, \sigma\right)\right\|\right) \; a.s.$$

and after summing such equalities we obtain

$$\ln\left(\|\mathbf{X}_n\|\right) - \ln\left(\|\mathbf{X}_0\|\right) = \sum_{k=0}^{n-1} \ln\left(\left\|\frac{\mathbf{X}_k}{\|\mathbf{X}_k\|} + \sigma\mathbf{M}_k\left(\frac{\mathbf{X}_k}{\|\mathbf{X}_k\|}, \sigma\right)\right\|\right) \; a.s.$$

□

Proof of Theorem 2 (stated page 9)

in order to prove the Theorem, we will make use of the following Proposition:

Proposition 7 *Let $(\mathbf{X}_n)_n$ be the sequence of random vectors defined in (4), let σ be a positive constant and let $\mathbf{M}$ be the random vector introduced in (12) (with $n = 0$, i.e., $\mathbf{M} = \mathbf{M}^{(0)}$). Then the random variables Z_n introduced in Proposition 2 are independent and follow the same distribution as the variable $\mathrm{Z}(\sigma) = \|\mathrm{e}_1 + \sigma\mathbf{M}(\mathrm{e}_1, \sigma)\|$ (introduced in Lemma 2 and Definition 1) or any $\|u + \sigma\mathbf{M}(u, \sigma)\|$ such that $\|u\| = 1$.*

Proof. Let σ be a positive constant. Let $(\mathbf{X}_n)_{n\in\mathbb{Z}^+}$ be the sequence of random vectors defined in (4) and $(\sigma_n))_{n\in\mathbb{Z}^+}$ the associated sequence of positive random variables such that $\sigma_n = \sigma\|\mathbf{X}_n\|$ for all $n \geq 0$. By Proposition 3, $\|\mathbf{X}_n\| > 0$ almost surely such that the random variable $\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|}$ is well defined. Note that the norm of the random variable $\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|}$ is one. Therefore, at an iteration n, we can apply Lemma 2 for $\left\| \frac{\mathbf{X}_n}{\|\mathbf{X}_n\|} + \sigma\mathbf{M}_n\left(\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|}, \frac{\sigma_n}{\|\mathbf{X}_n\|}\right)\right\|$ where $\mathbf{X}_n$ is fixed:

$$\begin{aligned} E\left(e^{it\mathrm{Z}_n}|\boldsymbol{\sigma}\left(\mathbf{X}_n\right)\right) &= E\left(e^{it\left\|\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|}+\sigma\mathbf{M}_n\left(\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|},\sigma\right)\right\|}|\boldsymbol{\sigma}\left(\mathbf{X}_n\right)\right) \\ &= E\left(e^{it\left\|\mathrm{e}_1+\sigma\mathbf{M}(\mathrm{e}_1,\sigma)\right\|}|\boldsymbol{\sigma}\left(\mathbf{X}_n\right)\right). \end{aligned} \tag{40}$$

Therefore $E\left(e^{it\mathrm{Z}_n}\right) = E\left(E\left(e^{it\mathrm{Z}_n}\mid\boldsymbol{\sigma}(\mathbf{X}_n)\right)\right) = E\left(e^{it\left(\left\|\mathrm{e}_1+\sigma\mathbf{M}(\mathrm{e}_1,\sigma)\right\|\right)}\right)$. Therefore, for all $n \geq 0$, the variables Z_n and $\left\|\mathrm{e}_1 + \sigma\mathbf{M}\left(\mathrm{e}_1, \sigma\right)\right\|$ follow the same distribution. This is also valid when replacing e_1 by any vector unit vector u.
For showing the independence of $(\mathrm{Z}_n)_{n\in\mathbb{Z}^+}$, we will prove that for all n, for all $t_0 \in \mathbb{R}, \ldots, t_n \in \mathbb{R}$, $E\left(e^{it_0\mathrm{Z}_0}\ldots e^{it_n\mathrm{Z}_n}\right) = E\left(e^{it_0\mathrm{Z}_0}\right)\ldots E\left(e^{it_n\mathrm{Z}_n}\right)$. We will proceed by induction and suppose that for all $t_0 \in \mathbb{R}, \ldots, t_{n-1} \in \mathbb{R}$, we have $E\left(e^{it_0\mathrm{Z}_0}\ldots e^{it_{n-1}\mathrm{Z}_{n-1}}\right) = E\left(e^{it_0\mathrm{Z}_0}\right)\ldots E\left(e^{it_{n-1}\mathrm{Z}_{n-1}}\right)$ and prove that for all $t_0 \in \mathbb{R}, \ldots, t_n \in \mathbb{R}$

$$E\left(e^{it_0\mathrm{Z}_0}\ldots e^{it_n\mathrm{Z}_n}\right) = E\left(e^{it_0\mathrm{Z}_0}\right)\ldots E\left(e^{it_n\mathrm{Z}_n}\right).$$

Let ζ_n be the σ-algebra $\boldsymbol{\sigma}(\mathbf{X}_0, \mathbf{M}_0, \mathbf{X}^{(1)}, \mathbf{M}_1, \ldots, \mathbf{X}^{(\mathbf{n-1})}, \mathbf{M}_{n-1}, \mathbf{X}_n)$ and let $t_0, \ldots, t_n \in \mathbb{R}^{n+1}$. Then, $E(e^{it_0\mathrm{Z}_0}\ldots e^{it_n\mathrm{Z}_n}) = E(E(e^{it_0\mathrm{Z}_0}\ldots e^{it_n\mathrm{Z}_n}|\zeta_n))$. Since $e^{it_0\mathrm{Z}_0}\ldots e^{it_{n-1}\mathrm{Z}_{n-1}}$ is bounded and ζ_n-measurable [17, p88, j]

$$E(e^{it_0\mathrm{Z}_0}\ldots e^{it_n\mathrm{Z}_n}|\zeta_n) = e^{it_0\mathrm{Z}_0}\ldots e^{it_{n-1}\mathrm{Z}_{n-1}}E(e^{it_n\mathrm{Z}_n}|\zeta_n)\ . \tag{41}$$

Besides, $E(e^{it_n\mathrm{Z}_n}|\zeta_n) = E\left(e^{it_n\left\|\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|}+\sigma\mathbf{M}_n\left(\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|},\sigma\right)\right\|}|\zeta_n\right)$. The variable Z_n depends only on $\mathbf{X}_n$ and $\mathbf{M}_n$ with $\mathbf{M}_n$ depending only on $\mathbf{X}_n$ and the variables $\mathbf{N}_n^i$ which do not depend on $\left(\mathbf{X}_0, \mathbf{M}_0, \mathbf{X}^{(1)}, \mathbf{M}_1, \ldots, \mathbf{X}^{(\mathbf{n-1})}, \mathbf{M}_{n-1}\right)$. Therefore, we have

$$\begin{aligned} &E\left(e^{it_n\left\|\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|}+\sigma\mathbf{M}_n\left(\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|},\sigma\right)\right\|}|\zeta_n\right) \\ &\qquad = E\left(e^{it_n\left\|\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|}+\sigma\mathbf{M}_n\left(\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|},\sigma\right)\right\|}|\boldsymbol{\sigma}(\mathbf{X}_n)\right). \end{aligned}$$

Moreover, we know from (40) that

$$E\left(e^{it_n\left\|\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|}+\sigma\mathbf{M}_n\left(\frac{\mathbf{X}_n}{\|\mathbf{X}_n\|},\sigma\right)\right\|}|\boldsymbol{\sigma}(\mathbf{X}_n)\right) = E\left(e^{it_n\|\mathrm{e}_1+\sigma\mathbf{M}(\mathrm{e}_1,\sigma)\|}\right).$$

Injecting this in (41), we obtain

$$E\left(e^{it_0\mathrm{Z}_0}\ldots e^{it_n\mathrm{Z}_n}|\zeta_n\right) = e^{it_0\mathrm{Z}_0}\ldots e^{it_{n-1}\mathrm{Z}_{n-1}}E\left(e^{it_n\|\mathrm{e}_1+\sigma\mathbf{M}(\mathrm{e}_1,\sigma)\|}\right). \tag{42}$$

We take now the expectation of both sides of the previous equation and obtain

$$E\left(e^{it_0\mathrm{Z}_0}\dots e^{it_n\mathrm{Z}_n}\right) = E\left(e^{it_0\mathrm{Z}_0}\dots e^{it_{n-1}\mathrm{Z}_{n-1}}E\left(e^{it_n\|\mathrm{e}_1+\sigma\mathbf{M}(\mathrm{e}_1,\sigma)\|}\right)\right). \quad (43)$$

As $E\left(e^{it_n\|\mathrm{e}_1+\sigma\mathbf{M}\|}\right)$ is a constant value, previous equation becomes

$$E\left(e^{it_0\mathrm{Z}_0}\dots e^{it_n\mathrm{Z}_n}\right) = E\left(e^{it_0\mathrm{Z}_0}\dots e^{it_{n-1}\mathrm{Z}_{n-1}}\right)E\left(e^{it_n\|\mathrm{e}_1+\sigma\mathbf{M}(\mathrm{e}_1,\sigma)\|}\right). \quad (44)$$

Moreover by induction hypothesis we know that
$E\left(e^{it_0\mathrm{Z}_0}\dots e^{it_{n-1}\mathrm{Z}_{n-1}}\right) = E\left(e^{it_0\mathrm{Z}_0}\right)\dots E\left(e^{it_{n-1}\mathrm{Z}_{n-1}}\right)$ which thus imply that

$$E\left(e^{it_0\mathrm{Z}_0}\dots e^{it_n\mathrm{Z}_n}\right) = E\left(e^{it_0\mathrm{Z}_0}\right)\dots E\left(e^{it_{n-1}\mathrm{Z}_{n-1}}\right)E\left(e^{it_n\|\mathrm{e}_1+\sigma\mathbf{M}(\mathrm{e}_1,\sigma)\|}\right).$$

To finish the proof, we have to prove that $E\left(e^{it_n\|\mathrm{e}_1+\sigma\mathbf{M}(\mathrm{e}_1,\sigma)\|}\right) = E\left(e^{it_n\mathrm{Z}_n}\right)$. Using again the first part of this proof, we know that $\|\mathrm{e}_1 + \sigma\mathbf{M}(\mathrm{e}_1,\sigma)\|$ and Z_n follow the same distribution such that $E\left(e^{it_n\mathrm{Z}_n}\right) = E\left(e^{it_n\|\mathrm{e}_1+\sigma\mathbf{M}(\mathrm{e}_1,\sigma)\|}\right)$. Injecting this result in the previous equation, we obtain the following equation

$$E\left(e^{it_0\mathrm{Z}_0}\dots e^{it_n\mathrm{Z}_n}\right) = E\left(e^{it_0\mathrm{Z}_0}\right)\dots E\left(e^{it_{n-1}\mathrm{Z}_{n-1}}\right)E\left(e^{it_n\mathrm{Z}_n}\right) \quad (45)$$

which achieves to prove the independence of $(\mathrm{Z}_n)_{n\in\mathbb{Z}^+}$. □

Proof of the Theorem :
The variables $\ln(\mathrm{Z}_n)$ where $(\mathrm{Z}_n)_n$ is introduced in Proposition 2, are identically distributed, independent (Proposition 7), and have a finite expectation (Proposition 1). Therefore, the LLN for independent identically distributed random variables with a finite expectation applies for the sequence $(\ln(\mathrm{Z}_n))_n$ in the sense that the quantity $\frac{1}{n}\sum_{k=1}^{n}\ln\left(\left\|\frac{\mathbf{X}_k}{\|\mathbf{X}_k\|} + \sigma\mathbf{M}_k\left(\frac{\mathbf{X}_k}{\|\mathbf{X}_k\|},\sigma\right)\right\|\right)$ converges almost surely to $\mathrm{V}(\sigma)$ when n goes to infinity. Then, by Proposition 2, we have $\frac{1}{n}\ln\left(\frac{\|\mathbf{X}_n\|}{\|\mathbf{X}_0\|}\right)$ converges almost surely to $\mathrm{V}(\sigma)$ when n goes to infinity. □

Contents

Centre de recherche INRIA Saclay – Île-de-France
Parc Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 Orsay Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399