HAL
open science

A probabilistic approach to high-dimensional least-squares approximations

Nicolas Champagnat, Christophe Chipot, Erwan Faou

To cite this version:

Nicolas Champagnat, Christophe Chipot, Erwan Faou. A probabilistic approach to high-dimensional least-squares approximations. 2008. inria-00345411v1

HAL Id: inria-00345411

https://inria.hal.science/inria-00345411v1

Preprint submitted on 9 Dec 2008 (v1), last revised 25 Jun 2010 (v2)

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

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

A probabilistic approach to high-dimensional least-squares approximations.

Nicolas Champagnat, Christophe Chipot and Erwan Faou

November 24, 2008

Abstract

Least-squares problems set on high-dimensional spaces are considered. The solution is approximated employing a Monte Carlo method that takes the average of a random variable defined as the solutions of random small least-squares problems drawn as subsystems of the original problem. The conditions that ensure convergence and consistency of the method are discussed, along with an analysis of the computational cost in specific instances. The proposed algorithms generalize and improve the statistical analysis of distributed multipoles (SADM) approach put forth by Chipot *et al* in [6] for the derivation of distributed atomic multipoles from the quantum-mechanical electrostatic potential.

MSC numbers: 65C05, 93E24, 41A45, 41A63

Keywords: Least-squares approximation, Monte Carlo methods, high dimensional problems.

1 Introduction

The main objective of this work is to derive and analyze new schemes for the numerical approximation of least-squares problems set on high dimensional spaces. This work originates from the statistical analysis of distributed multipoles (SADM) algorithm introduced by Chipot *et al.* in [6] for the derivation of atomic multipoles from the quantum-mechanical electrostatic potential mapped on a grid of points surrounding a molecule of interest. The central idea consists in drawing subsystems of the original large least-squares problem and compute the average of the corresponding distribution of solutions as an approximation of the original

solution. This method not only provides a numerical approximation of the solution, but also a global statistical distribution that reflects the accuracy of the physical model being utilized.

Interestingly enough, it turns out that this kind of approach can be extended to many situations arising in computational mathematics and physics. The principle of the SADM algorithm is in fact very general, and can be adapted to derive efficient algorithms that are robust with the dimension of the underlying space of approximation. This in turn provides new numerical methods of practical interest for high dimensional least-squares problems, where traditional methods are impossible to implement.

The goal of the present contribution is twofold:

- Introduce a general mathematical framework, and analyze the consistency, convergence and cost of the proposed algorithms in an abstract setting and in specific situations where calculations can be made explicit (Wishart or subgaussian distribution). The main outcome is that the subsystems drawn from the original system have to be chosen rectangular and not square (as initially proposed in the SADM method) to yield convergent and efficient algorithms.
- Apply these results to revisit and improve the SADM method. This is mainly achieved in Section 5 by considering a simple, three-point charge model of water.

Let us now describe more precisely the problematic:

Let (Ω, μ) be a probability space Ω equipped with a measure μ. For a given arbitrary function $f \in \mathrm{L}^2(\Omega)$ and n given functions $\gamma_j(x) \in \mathrm{L}^2(\Omega)$, $j = 1, \ldots, n$ all taking values in $\mathbb{R}$, we consider the problem of approximating $f(x)$ by a linear combination of the functions $\gamma_j(x)$, $j = 1, \ldots, n$.

We seek an approximation of $f(x)$ in the sense that ideally, we would like to solve the problem of finding $\alpha = (\alpha_j)_{j=1}^n \in \mathbb{R}^n$, minimizing the function

$$\mathbb{R}^n \ni a \mapsto \|f(x) - \sum_{j=1}^n a_j \gamma_j(x)\|^2_{\mathrm{L}^2(\Omega)}. \tag{1.1}$$

The actual quality of the least-squares approximation is given by the size of the residue $\|\rho(\alpha)\|_{\mathrm{L}^2(\Omega)}$ where for $a = (a_j)_{j=1}^n \in \mathbb{R}^n$,

$$\rho(a)(x) = f(x) - \sum_{j=1}^n a_j \gamma_j(x). \tag{1.2}$$

Many minimization problems arising in mathematics and in physics can be stated under this form, for instance:

(a) $\Omega = [a,b]^n$ with two real numbers a and $b > a$, and equipped with the measure $\mathrm{d}\mu(x) = (b-a)^{-n}\mathrm{d}x$ where $\mathrm{d}x$ is the Lebesgue measure on $\mathbb{R}^n$. Taking $\gamma : \Omega \to \mathbb{R}^{n+1}$ defined by $\gamma_i(x) = x_i$ for all $i \in \{1,\dots,n\}$ and $\gamma_{n+1} \equiv 1$, the problem is equivalent to finding $\beta \in \mathbb{R}$ and $\alpha \in \mathbb{R}^n$ minimizing the function

$$\|f(x) - \beta - \langle \alpha, x\rangle\|^2_{\mathrm{L}^2([a,b]^n)}$$

where $\langle\,\cdot\,,\cdot\,\rangle$ is the standard Euclidean product in $\mathbb{R}^n$. This is nothing else than a multivariate linear interpolation.

Similarly, any polynomial approximation problem in $\mathrm{L}^2([a,b]^n,\mu)$, where μ is a weight function, can be written in the form (1.1) by taking as γ_j a basis of polynomials in dimension n.

(b) Taking $\Omega = \mathbb{R}^n$ equipped with a given n-dimensional Gaussian measure leads to many different situations: The approximation by Hermite functions in $\mathbb{R}^n$ if γ_j are polynomials, the approximation of f by Gaussian chirps signal [11] in the case where $\gamma_j(x)$ are oscillating functions of x, or alternatively approximation by Gaussian wavepackets functions [10] in the context of molecular dynamics.

(c) Consider $\Omega = \{1,\dots,M\}$ with $M >> n$ equipped with the uniform probability measure $M^{-1}\sum_{i=1}^M \delta_i$. In this case, an application f is represented by a vector $b \in \mathbb{R}^M$, whereas γ is represented by a matrix A with n columns and M lines. The problem is then equivalent to the problem of finding $\alpha \in \mathbb{R}^n$ that minimizes

$$\|A\alpha - b\|_2^2$$

where $\|\cdot\|_2$ is the Euclidean norm on $\mathbb{R}^M$.

(d) Consider $\Omega = \mathbb{R}^n \times \Omega'$ equipped with the measure $\mu \otimes \nu$ where μ and ν are probability measures on $\mathbb{R}^n$ and Ω' respectively. Taking $f(x,\omega') = h(x) + X(\omega')$ where $X(\omega')$ is a given random variable on Ω', and $\gamma_j(x,\omega') = x_j$ for $j = 1,\dots,n$ yields the problem of minimizing

$$\min_{\alpha\in\mathbb{R}^n} \mathbb{E}\Big[\|\langle\alpha,x\rangle - b(x,\omega')\|^2_{\mathrm{L}^2(\mathbb{R}^n)}\Big] \tag{1.3}$$

which corresponds to the linear regression of a function observed with some independent noise. For example, if $h(x) = \langle a, x\rangle$ with $a \in \mathbb{R}^n$ and $\mathbb{E}(X) = 0$, the solution of the problem (1.3) is $\alpha = a$. This can also be interpreted as a linear filtering problem, in the specific case where f is exactly a linear combination of the γ_j, $1 \le j \le n$.

The problem (1.1) is equivalent to solving the linear equation

$$\langle \gamma, \gamma^T\rangle_{\mathrm{L}^2}\alpha = \langle \gamma, f\rangle_{\mathrm{L}^2}$$

where $\alpha = (\alpha_i)_{i=1}^n$ and $\langle \gamma, \gamma^T \rangle_{\mathrm{L}^2}$ is the $n \times n$ matrix with coefficients $\langle \gamma_i, \gamma_j \rangle_{\mathrm{L}^2}$, $i, j = 1, \ldots, n$.

If the family $(\gamma_i(x))_{i=1}^n$ defines a full rank set of elements of $\mathrm{L}^2(\Omega)$, the matrix $\langle \gamma, \gamma^T \rangle_{\mathrm{L}^2}$ is invertible, and the solution of the previous equation reads

$$\alpha = \langle \gamma, \gamma^T \rangle_{\mathrm{L}^2}^{-1} \cdot \langle \gamma, f \rangle_{\mathrm{L}^2}. \tag{1.4}$$

Apart from specific situations, where, for instance, the γ_j can be assumed orthogonal, the numerical approximation of (1.4) is extremely costly with respect to the dimension of Ω and n (see for instance [4]). Typically, discretization of problems of the form (**a**) yields a problem of the form (**c**) with $m = N^n$ where N is the number of interpolation points in $[a, b]$ needed to approximate the L^2 integrals. For $n = 30$, this method is not tractable in practice, even if $N = 2$.

To avoid this *curse of dimensionality*, an alternative would consist in approximating the integrals in the formula (1.4) by using Monte Carlo methods. In large dimension, the matrix $\langle \gamma, \gamma^T \rangle_{\mathrm{L}^2}$ is, however, often ill-conditioned, and obtaining a correct approximation of the inverse of this matrix might require in practice a very large number of draws to minimize the error in the value of α.

In [6], Chipot *et al.* propose an alternative algorithm coined statistical analysis of distributed multipoles (SADM). The aim of their work is to derive atomic multipoles from the quantum-mechanical electrostatic potential mapped on a grid of points surrounding a molecule of interest, following a different pathway than the conventional least-squares scheme embodied in (1.1). In the setting presented above, this corresponds to a problem of the form **(c)**, where $(\alpha_j)_{j=1}^n$ represent the unknown multipoles borne by the n particles, and $\gamma_j(x)$, the electrostatic potential decaying with respect to $||x - x_j||$, where $x_1, \ldots, x_n$ denote the positions of the particles. The space Ω is made of M points in the three-dimensional space (located away from the atomic positions) with $M >> n$. Instead of solving directly the problem (1.1), Chipot *et al.* draw n points $x^{(i)}$ amongst the M points in Ω, solve the $n \times n$ problem $f(x^{(i)}) = \sum_{j=1}^n \gamma_j(x^{(i)})\alpha_j$, $i = 1, \ldots, n$ and subsequently plot the distributions corresponding to each α_j. They notice that the latter are Cauchy-like distributions (with seemingly infinite expectation) centered around the exact solution of the original least-squares problem.

Retaining the idea of drawing subsystems of the original problem, we develop and analyze in this work a new probabilistic method for the approximation of the solution to the least-squares problem (1.1). Roughly speaking, it can be described as follows:

- Draw m points $x^{(i)}$, $i = 1, \ldots, m$ in (Ω, μ). Typically $m = n + 2$ or $2n$.
- Solve the $m \times n$ least-squares sub-problem by minimizing the function

$$\mathbb{R}^n \ni \beta \mapsto \sum_{i=1}^m |f(x^{(i)}) - \sum_{j=1}^m \alpha_j \gamma_j(x^{(i)})|^2 \tag{1.5}$$

- Take the average of the random variable β.

Let us mention that taking $m = n$ in the previous algorithm leads in all the examples that were considered to a random variable β with infinite expectation, which might partly explain the Cauchy distributions observed in [6] with the SADM method.

In the following Section 2, we introduce a rigorous definition of this algorithm. We then discuss the conditions of convergence and consistency of the algorithm, and examine the numerical cost of the latter.

In Section 3, we first consider the case where the problems (1.5) are solvable almost everywhere. This will be typically the case of the linear interpolation in the situation **(b)**. In such a situation, we show that the convergence of the algorithm relies on integrability conditions on the inverse of the smallest eigenvalue of the random matrices appearing in the subproblems (1.5). Strikingly, apart from very specific situations — *viz.* the Wishart case studied in Subsection 3.5 and the sub-gaussian case studied in Subsection 4.2, the validity of such conditions turns out to represent very difficult open problems on random matrices, see for example [13] and references therein.

In Section 4, we consider the more general and realistic situations where the random matrices associated with (1.5) are not assumed to be invertible almost everywhere, and one is expected to condition the solution of (1.5) to avoid poorly conditioned scenarios (for example, this is always the case in situation **(c)**). We then study the cost of the algorithm in the specific instances where the random variable β can be expressed in terms of random matrices with sub-Gaussian entries.

Note that in the previous cases, the cost of the algorithm depends only on n and m, but not on the dimension on Ω. Moreover, in contrast with the classical Monte Carlo approximation, we show that its numerical cost depends on the residue $\|\rho(\alpha)\|_{\mathrm{L}^2}$, where α is the solution of the problem (1.1) and where ρ has been defined in (1.2). In other words, it is cheaper for a higher quality of the least-squares approximation (1.1). Our method, therefore, appears to be of noteworthy practical interest in situations where Ω is "large" and the matrix $\langle \gamma, \gamma^T \rangle_{\mathrm{L}^2}$ is ill-conditioned, but the quality of approximation given by the least-squares problem remains good.

We conclude by supplying numerical examples in Section 5. In particular, Section 5.2 is devoted to showing that the SADM method of [6] can be improved by considering sub-problems with $m > n$, in lieu of $m = n$ as was initially introduced. This is illustrated in the paradigmatic case of the three-point charge model of water.

2 Principle of the algorithms

We propose to replace the problem (1.1) by the following new problem. For a given number $m \geq n$, let $X^{(i)}$ be m idenpendent and identically distributed (i.i.d.) random variables in Ω with common law μ and let X denote the vector $(X^{(1)}, \ldots, X^{(m)})$. We define the functions $F : \Omega^m \to \mathbb{R}^m$ and $\Gamma : \Omega^m \to \mathscr{L}(\mathbb{R}^m, \mathbb{R}^n)$ by the formulae

$$\forall i = 1, \ldots, m, \quad F_i(x^{(1)}, \ldots, x^{(m)}) = f(x^{(i)}) \tag{2.1}$$

and

$$\forall i = 1, \ldots, m, \quad \forall\, j = 1, \ldots, n, \quad \Gamma_{ij}(x^{(1)}, \ldots, x^{(m)}) = \gamma_j(x^{(i)}). \tag{2.2}$$

We then define the random vector β in $\mathbb{R}^n$ as the point where the function

$$\beta \mapsto \|F(X) - \Gamma(X)\beta\|_2^2$$

reaches its minimum, where $\|\cdot\|_2$ is the standard Euclidean norm on $\mathbb{R}^m$. β is, therefore, the solution of the least-squares problem associated with vector $F(X)$ and matrix $\Gamma(X)$. Equivalently, if $\Gamma^T(X)\Gamma(X)$ is invertible,

$$\beta = R(X)F(X) := ((\Gamma^T\Gamma)^{-1}\Gamma^T)(X)F(X). \tag{2.3}$$

The expectation of β is then given by the formula

$$\bar{\beta} := \mathbb{E}\beta = \int_{\Omega^m} ((\Gamma^T\Gamma)^{-1}\Gamma^T F)(x^{(1)}, \ldots, x^{(m)})\, \mathrm{d}\mu(x^{(1)}) \otimes \cdots \otimes \mathrm{d}\mu(x^{(m)}). \tag{2.4}$$

Our algorithm consists in approximating the previous expectation value by a Monte Carlo method. More specifically, we approximate $\bar{\beta}$ by

$$\bar{\beta}_N = \frac{1}{N}\sum_{i=1}^{N} \beta_i, \tag{2.5}$$

where $\beta_i, i \geq 1$ are i.i.d. realizations of the random vector $\beta \in \mathbb{R}^n$, obtained by (2.3) from i.i.d. realizations of the random $n \times m$ matrix X.

In practice, the invertibility of $\Gamma^T(X)\Gamma(X)$ cannot be guaranteed — and obviously not for problems of the form (**c**), where we have to consider only random variables $X^{(1)}, \ldots, X^{(m)}$ that are almost surely (a.s.) distinct, or can lead to ill-conditioned problems.

In a more general setting, we, hence, restrict ourselves to realizations of X, such that matrix $\Gamma(X)$ is sufficiently well conditioned, in the following sense: Let $s_1(\Gamma(X))$ denote the smallest eigenvalue of the symmetric positive matrix $\Gamma(X)^T\Gamma(X)$. Next, we will only consider realizations of X, such that $s_1(\Gamma(X))$

is greater than some threshold σ, which may depend on n and m. In this case, rather than approximating (2.4), we will estimate the conditional expectation

$$\bar{\beta}^\sigma := \mathbb{E}^\sigma \beta = \mathbb{E}[\, \beta \mid s_1(\Gamma(X)) > \sigma \,] \tag{2.6}$$

by

$$\bar{\beta}^\sigma_N = \frac{1}{N} \sum_{i=1}^{N} \beta^\sigma_i, \tag{2.7}$$

where the β^σ_i are obtained from a sequence of i.i.d. realizations of the random vector $\beta \in \mathbb{R}^n$ in (2.3), from which have been removed all realizations such that $s_1(\Gamma(X)) \leq \sigma$. Note that (2.5) is a particular case of (2.7) for $\sigma = 0$, provided that $\mathbb{P}(s_1(\Gamma(X)) = 0) = 0$.

Such a method will be of interest in terms of computational cost if m is on the order of magnitude of n (in all the applications considered herein, $m = n+2$ or $m = 2n$ will be sufficient) and if $\mathbb{P}(s_1(\Gamma(X)) > \sigma)$ is not too small — because drawing a realization of X such that $s_1(\Gamma(X)) > \sigma$ requires in average a number $\mathbb{P}(s_1(\Gamma(X)) > \sigma)^{-1}$ of realizations of X.

From the perspective of precision, this method will perform well if the variance of β conditionally on $\{s_1(\Gamma(X)) > \sigma\}$ has an appropriate behavior with respect to n and m, and if $\bar{\beta}^\sigma$ defined in (2.6) provides a good approximation of the solution of the original least-squares problem. We will give various conditions that warrant the latter requirements.

For the ease of presentation, we first consider in Section 3 the case where we can take $\sigma = 0$, *i.e.* where the matrix $\Gamma(X)^T\Gamma(X)$ is invertible almost everywhere. We study the consistency of the algorithm, give the condition of convergence of the Monte Carlo method, and analyze the cost of the numerical scheme. In the particular instance where matrix $\Gamma(X)^T\Gamma(X)$ has the Wishart distribution, we provide precise estimates, and an optimal choice of the parameter m. In Section 4, we extend these results to the more general case where $\sigma > 0$.

3 The invertible case

In this section, we assume that $\Gamma(X)^T\Gamma(X)$ is a.s. invertible. In this case, the random variable β given by (2.3) is a.s. well defined.

3.1 Notations

For a given $m \times n$ matrix $A = (a_{ij})_{1\leq i\leq m, 1\leq j\leq n}$, we define $s_1(A)$ as the least singular value of A, *i.e.* the smallest eigenvalue of A^TA. Note that the eigenvalues

of A^TA are nonnegative. We denote by $\|\cdot\|_F$ the Schur-Frobenius norm on $n\times m$ matrices

$$\|B\|_F^2 = \sum_{i=1}^m \sum_{j=1}^n b_{ij}^2.$$

where $B = (b_{ij})_{1\le i\le n, 1\le j\le m}$. With this notation, we have for any $m\times n$ matrix $A = (a_{ij})_{1\le i\le m, 1\le j\le n}$

$$\|(A^TA)^{-1}A^T\|_F^2 = \mathrm{Tr}((A^TA)^{-1}A^TA(A^TA)^{-1}) = \mathrm{Tr}((A^TA)^{-1}) \le \frac{n}{s_1(A)}. \quad (3.1)$$

We also write $\|\cdot\|_p$, $p\in[1,+\infty]$ the usual L^p norm on $\mathbb{R}^m$. For any $n\times m$ matrix $B = (b_{ij})_{1\le i\le n, 1\le j\le m}$ and any $v = (v_1,\dots,v_m)\in\mathbb{R}^m$, we have

$$\begin{aligned}\|Bv\|_2^2 &= \sum_{i=1}^n \sum_{1\le j,k\le m} b_{ij}v_jb_{ik}v_k \\ &\le \frac{1}{2}\Big(\sum_{i,j,k} b_{ij}^2|v_jv_k| + \sum_{i,j,k} b_{ik}^2|v_jv_k|\Big) \\ &\le \|B\|_F^2\,\|v\|_1\,\|v\|_\infty \le \|B\|_F^2\,\|v\|_1^2, \end{aligned} \quad (3.2)$$

where we used the inequality $|b_{ij}b_{ik}|\le \frac{1}{2}(b_{ij}^2+b_{ik}^2)$.

Let $\Gamma(X)$ be the random matrix defined by (2.2). For $q\ge 2$ we define

$$K_q(\Gamma) := \left[\mathbb{E}\frac{1}{s_1(\Gamma(X))^{\frac{q}{2}}}\right]^{\frac{2}{q}}. \quad (3.3)$$

Note that $K_q(\Gamma)$ depends on n and m.

Armed with this set of definitions, we easily get the following estimate:

Lemma 3.1 *Let $p\in[1,\infty]$ and assume that $g\in\mathrm{L}^p(\Omega)$. Let G be the function associated with g by (2.1) and $R(X)$ be the random matrix defined in (2.3).*

(a) *Assume that $p\in(1,\infty]$ and that $K_q(\Gamma)<+\infty$ where q is such that $q^{-1}+p^{-1}=1$. Then we have*

$$\mathbb{E}\|R(X)G(X)\|_2 \le \sqrt{n}m\sqrt{K_q(\Gamma)}\|g\|_{\mathrm{L}^p(\Omega)}. \quad (3.4)$$

(b) *Assume that $p\in(2,\infty]$ and that $K_q(\Gamma)<+\infty$ where q is such that $2q^{-1}+2p^{-1}=1$. Then we have*

$$\mathbb{E}\|R(X)G(X)\|_2^2 \le nm^2K_q(\Gamma)\|g\|_{\mathrm{L}^p(\Omega)}^2. \quad (3.5)$$

Proof. Using (3.2) and (3.1), we have

$$\|R(X)G(X)\|_2 \leq \sqrt{n}\, s_1(\Gamma(X))^{-1/2}\|G(X)\|_1 .$$

Taking the expectation and using Hölder's inequality, we get

$$\mathbb{E}\|R(X)G(X)\|_2 \leq \sqrt{n}\sqrt{K_q(\Gamma)}\Big(\mathbb{E}\|G(X)\|_1^p\Big)^{1/p}.$$

Now, $Y \mapsto (\mathbb{E}|Y|^p)^{1/p}$ defines a norm on the set of random vectors on Ω with finite p-th-order moment. We, hence, obtain

$$\begin{aligned}\big(\mathbb{E}\|G(X)\|_1^p\big)^{1/p} &= \Big(\mathbb{E}\Big(\sum_{i=1}^m |g(X^{(i)})|\Big)^p\Big)^{1/p} \\ &\leq \sum_{i=1}^m \big(\mathbb{E}|g(X^{(i)})|^p\big)^{1/p} = m\|g\|_{\mathrm{L}^p(\Omega)} ,\end{aligned}$$

and this yields Point (a). Point (b) is obtained from similar computations. ■

3.2 Consistency

Formula (2.4) proposes an alternative solution $\bar{\beta} = \mathbb{E}\beta$ to the solution α given by (1.4) of the least-squares problem (1.1). We now provide estimates between these two solutions.

A precise general error estimate depends on the tackled problem. Here, we give a very general result. We recall that $\rho(a)$ denotes the residue (1.2) associated with the function f and the coefficients a_j, $j = 1, \ldots, n$.

Proposition 3.2 *Let $a = (a_j)_{j=1}^n \in \mathbb{R}^n$ be given. Let $f \in \mathrm{L}^2(\Omega)$ and $\bar{\beta}$ given by (2.4). Assume that $\rho(a) \in \mathrm{L}^p(\Omega)$ with $p \in (1, +\infty]$, and that $K_q(\Gamma) < +\infty$ where $K_q(\Gamma)$ is defined by (3.3) and $q^{-1} + p^{-1} = 1$. We then have the following estimate*

$$\|\bar{\beta} - a\|_2 \leq \mathbb{E}\|\beta - a\|_2 \leq \sqrt{n}m\sqrt{K_q(\Gamma)}\|\rho(a)\|_{\mathrm{L}^p(\Omega)} . \tag{3.6}$$

Proof. Let $g(x) = \sum_{j=1}^n a_j\gamma_j(x)$ and G the function defined by (2.1). As $\Gamma(X)^T\Gamma(X)$ is invertible, it is clear that

$$R(X)G(X) = a.$$

Hence we can write

$$\beta - a = R(X)\rho(a)(X). \tag{3.7}$$

Since $\|\bar{\beta} - a\|_2 \leq \mathbb{E}\|\beta - a\|_2$, the inequality (3.6) is then a consequence of Lemma 3.1 (a). ■

Note that this result implies that $\mathbb{E}\beta$ is well defined as soon as $\rho(0) = f \in \mathrm{L}^p(\Omega)$ and $K_q(\Gamma) < \infty$ for some p and q, such that $p^{-1} + q^{-1} = 1$. With this prerequisite, the Monte Carlo method converges.

Note also that the result with $p = 2$ applied to the solution α given by (1.4) of the original least-squares problem (1.1) shows that the error between $\bar{\beta}$ given by our method and the original solution α is on the order of magnitude of the residue $\|\rho(\alpha)\|_{\mathrm{L}^2(\Omega)}$ that determines the quality of the least-squares approximation itself.

Another consistency result can be inferred in the limit $m \to +\infty$. Let us assume that f and γ_j, $1 \leq j \leq n$ belong to $\mathrm{L}^2(\Omega)$. We observe that

$$\frac{1}{m}(\Gamma(X)^T\Gamma(X))_{ij} = \frac{1}{m}\sum_{k=1}^{m}\gamma_i(X^{(k)})\gamma_j(X^{(k)}) \tag{3.8}$$

converges $\mathbb{P}$-a.s. to $(\langle\gamma, \gamma^T\rangle_{\mathrm{L}^2})_{ij}$ when $m \to +\infty$. Similarly,

$$\frac{1}{m}(\Gamma(X)^T F(X))_i = \frac{1}{m}\sum_{k=1}^{m}\gamma_i(X^{(k)})f(X^{(k)}) \tag{3.9}$$

converges $\mathbb{P}$-a.s. to $(\langle\gamma, f\rangle_{\mathrm{L}^2})_i$. Consequently, if matrix $\langle\gamma, \gamma^T\rangle_{\mathrm{L}^2}$ is invertible,

$$\beta = \left(\frac{1}{m}\Gamma^T(X)\Gamma(X)\right)^{-1}\frac{1}{m}\Gamma(X)^T F(X) \tag{3.10}$$

converges $\mathbb{P}$-a.s. to α given by (1.4) when $m \to +\infty$.

3.3 Convergence

For Monte Carlo approximations of an n-dimensional vector Y having finite variance, the confidence region corresponding to the asymptotic confidence level η is known to be an ellipsoid (see for instance [1]), the diameter of which is for the Euclidean norm of the form

$$C\sqrt{\lambda/N}$$

where the constant C depends on η and n, N is the number of draws in the Monte Carlo approximation and λ is the largest eigenvalue of the covariance matrix of Y

$$\mathrm{Cov}(Y) := \mathbb{E}[(Y - \mathbb{E}Y)(Y - \mathbb{E}Y)^T].$$

Though less accurate but simpler, it is also possible to construct confidence regions of asymptotic level less than η of the form $[a_1, b_1] \times \ldots \times [a_n, b_n]$, by

taking each $[a_i, b_i]$ as a confidence interval of asymptotic level η/n for the i-th coordinate Y_i of Y. In particular, one can choose

$$b_i - a_i = 2x(n,\eta)\sqrt{\mathrm{Var}(Y_i)/N}, \quad \forall i \in \{1,\dots,n\}$$

where $x(n,\eta) > 0$ is such that

$$\frac{1}{\sqrt{2\pi}} \int_{x(n,\eta)}^{+\infty} e^{-u^2/2} du = \frac{\eta}{2n}. \tag{3.11}$$

Since, for all $x > 0$,

$$\int_x^{+\infty} e^{-u^2/2} du \leq \frac{1}{x} \int_x^{+\infty} u e^{-u^2/2} du = \frac{e^{-x^2/2}}{x}, \tag{3.12}$$

we have

$$x(n,\eta)^2 \leq \left(2 \log \frac{n\sqrt{2}}{\eta\sqrt{\pi}}\right) \vee 1. \tag{3.13}$$

In this case, the diameter for the Euclidean norm of the confidence region is, therefore, bounded by

$$2x(n,\eta)\sqrt{\mathrm{Tr}(\mathrm{Cov}(Y))/N}. \tag{3.14}$$

The confidence regions obtained via this route are only asymptotic of level η for large N. Note that obtaining non-asymptotic estimates would be possible upon using Berry-Essen-type inequalities — see for instance [12].

As has been seen above, the efficiency of the Monte Carlo method for the estimation of β is controlled by $\mathrm{Tr}(\mathrm{Cov}(\beta))$. The next result gives bounds on this quantity.

Theorem 3.3 *Let $p \in (2, +\infty]$ and assume that the residue $\rho(\bar{\beta}) \in \mathrm{L}^p(\Omega)$. Assume moreover that $K_q(\Gamma) < +\infty$, see (3.3), where q is such that $2p^{-1} + 2q^{-1} = 1$. Then we have*

$$\mathrm{Tr}(\mathrm{Cov}(\beta)) \leq nm^2 K_q(\Gamma) \|\rho(\bar{\beta})\|^2_{\mathrm{L}^p(\Omega)} \tag{3.15}$$

Proof. Define $g = \rho(\bar{\beta})$ and denote by G the function defined by (2.1). Using (3.7), we calculate that

$$\mathrm{Tr}(\mathrm{Cov}(\beta)) = \mathbb{E}\|\beta - \mathbb{E}\beta\|_2^2 = \mathbb{E}\|R(X)G(X)\|_2^2.$$

The result, hence, follows from Lemma 3.1 (b). ■

Note that in the case $p = +\infty$, we have $q = 2$.

In Proposition 3.2 and Theorem 3.3, assumptions of the form $K_q(\Gamma) < +\infty$ were used, where $K_q(\Gamma)$ is given by (3.3). These assumptions correspond to the integrability condition on functions $s_1(\Gamma(X))^{\frac{q}{2}}$. The following proposition provides a condition to ensure the integrability of these functions.

Proposition 3.4 *Let Y be a random variable satisfying the following estimate: There exist constants $\delta > 0$ and $\gamma > 0$ such that*

$$\forall \epsilon \geq 0, \quad \mathbb{P}(Y \leq \epsilon) \leq (\delta\epsilon)^{\gamma}. \tag{3.16}$$

Then, for any $0 < r < \gamma$,

$$\mathbb{E}(Y^{-r}) \leq \frac{\delta^r}{1 - r/\gamma}. \tag{3.17}$$

Proof. This result follows from the following integration by parts, where $\int_0^\infty h(x)\mathrm{d}\mathbb{P}(Y \in [0,x))$ denotes the Stieltjes integral of the measurable function h with respect to the Stieltjes measure on $[0,\infty)$ associated with the non-decreasing function $x \mapsto \mathbb{P}(Y \in [0,x))$.

$$\begin{aligned}
\mathbb{E}(Y^{-r}) &= \int_0^\infty x^{-r}\mathrm{d}\mathbb{P}(Y \in [0,x)) \\
&= \int_0^\infty r x^{-r-1}\mathbb{P}(Y \in [0,x))\mathrm{d}x \\
&\leq r\int_0^\infty x^{-r-1}((\delta x)^{\gamma} \wedge 1)\mathrm{d}x.
\end{aligned}$$

If $r < \gamma$,

$$\int_0^\infty x^{-r-1}((\delta x)^{\gamma} \wedge 1)\mathrm{d}x = \frac{1}{r}\Big(\frac{\delta^r}{1 - r/\gamma}\Big),$$

which implies (3.17). ■

In subsection 3.5, we show that condition (3.16) holds for the random variable $Y = s_1(\Gamma(X))$ and for some γ and δ depending on n and m, in the case where $\Gamma(X)^T\Gamma(X)$ is a Wishart matrix. In the end of section 4, we also present results in the case where $\Gamma(X)$ is a matrix with independent sub-Gaussian entries, using recent results of Rudelson & Vershynin — see [13]. In more general situations, this condition is related to difficult problems on random matrices, which, to our knowledge, have not been solved yet.

3.4 Computational cost of the method

Let ε be a required precision for the approximation of $\bar{\beta} = \mathbb{E}\beta$ by the Monte Carlo simulation (2.5). For large N, using (3.14), we must take

$$N \sim 4x(n,\eta)^2\varepsilon^{-2}\mathrm{Tr}(\mathrm{Cov}(\beta)).$$

Moreover, at each step, the algorithm requires that matrix $\Gamma(X)^T\Gamma(X)$ and vector $\Gamma(X)^T F(X)$ be evaluated and matrix $\Gamma(X)^T\Gamma(X)$ inverted. The cost of these operations is of order Cn^2m, where C is independent of m and n.

Using (3.13), we see that the cost of the algorithm if of order

$$C\varepsilon^{-2}mn^2 \log n \operatorname{Tr}(\operatorname{Cov}(\beta)).$$

Under the hypothesis of Theorem 3.3, this computational cost can be written

$$C\varepsilon^{-2}m^3n^3 \log n \, K_q(\Gamma)\|\rho(\bar{\beta})\|^2_{\mathrm{L}^p(\Omega)} \tag{3.18}$$

for $2p^{-1} + 2q^{-1} = 1$. It may be observed that this cost depends only on n and m — and not the dimension of Ω. Moreover, it depends on the least-squares residue of the problem (1.1). In the event where the f is close to a linear combination of functions γ_j, the algorithm is, therefore, cheaper.

Remark 3.5 *Note that if α denotes the solution of the least-squares problem (1.1), we have*

$$\|\rho(\mathbb{E}\beta)\|_{\mathrm{L}^p(\Omega)} \leq \|\rho(\alpha)\|_{\mathrm{L}^p(\Omega)} + \|\sum_{j=1}^{n}(\alpha_j - \mathbb{E}\beta_j)\gamma_j\|_{\mathrm{L}^p(\Omega)}.$$

Using (3.6), it can be seen that if the functions $\gamma_j \in \mathrm{L}^p(\Omega)$ with $p > 2$, and if $K_q(\Gamma) < +\infty$ for $p^{-1} + q^{-1} = 1$, then there exists a constant C_n depending on n such that

$$\|\rho(\bar{\beta})\|_{\mathrm{L}^p(\Omega)} \leq C_n\|\rho(\alpha)\|_{\mathrm{L}^p(\Omega)}.$$

As a consequence, the cost of our algorithm is driven by the quality of the original least-squares approximation in Problem (1.1).

3.5 The case of Wishart matrices

Let us now consider the case where $\Omega = \mathbb{R}^n$,

$$\mathrm{d}\mu(x) = (2\pi)^{-n/2}\exp(-\|x\|_2^2/2)\mathrm{d}x_1 \ldots \mathrm{d}x_n$$

and $\gamma_j(x) = x_j$ for $j \in \{1, \ldots, n\}$ — *i.e.* linear interpolation. In this case, random vectors $X^{(i)}$ are standard n-dimensional Gaussian vectors, the matrix $\Gamma(X)$ is a $m \times n$ matrix with i.i.d. standard Gaussian entries and the law of the matrix $\Gamma(X)^T\Gamma(X)$ is the so-called Wishart distribution — see *e.g.* [1].

The joint distribution of its eigenvalues is known explicitly and can be found for example in [1, p.534]. In particular, $\Gamma(X)^T\Gamma(X)$ is a.s. invertible if $m \geq n$. The explicit density of the eigenvalues has been used to obtain estimates on the law of the smallest eigenvalue of such matrices in [7, 8]. The particular result we are going to use is summarized in Lemma 3.3 of [5], and reads with our notation

as follows. For all $m \geq n \geq 2$, let $k = m - n + 1$. The density $p(x)$ of $s_1(\Gamma(X))$ then satisfies

$$L_{n,m} e^{-nx/2} x^{\frac{k}{2}-1} \leq p(x) \leq L_{n,m} e^{-x/2} x^{\frac{k}{2}-1}, \quad \forall x > 0, \tag{3.19}$$

where

$$L_{n,m} = \frac{2^{\frac{k}{2}-1} \Phi(\frac{m+1}{2})}{\Phi(\frac{n}{2})\Phi(k)}, \tag{3.20}$$

where Φ is the well-known Gamma function, defined for all $x > 0$ by

$$\Phi(x) = \int_0^{+\infty} e^{-t} t^{x-1} \mathrm{d}t.$$

Lemma 3.6 *For all $m \geq n \geq 2$, the random variable $Y = s_1(\Gamma(X))$ satisfies (3.16) for*

$$\gamma = \frac{m-n+1}{2} = \frac{k}{2}$$
$$\textit{and} \quad \delta = e^2 \frac{m}{k^2}.$$

Moreover, the constant γ above is the smallest such that (3.16) holds for all $\varepsilon > 0$ for some constant δ.

Proof. The proof of this result makes use of the following bounds for the Gamma function [5, Lemma 2.7]. For all $x > 0$,

$$\sqrt{2\pi}\, x^{x+\frac{1}{2}} e^{-x} < \Phi(x+1) = x\Phi(x) < \sqrt{2\pi}\, x^{x+\frac{1}{2}} e^{-x+\frac{1}{12x}}.$$

These inequalities can be plugged into (3.20) to get that, for all $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}(s_1(\Gamma(X)) \leq \varepsilon) &\leq \frac{n\, 2^{\gamma-1} \Phi(\frac{m-1}{2}+1)}{\Phi(\frac{n}{2}+1) 2\gamma \Phi(2\gamma)}\, \gamma \int_0^{\varepsilon} x^{\gamma-1} \mathrm{d}x \\
&\leq \frac{\frac{n}{2} 2^{\gamma} \sqrt{2\pi} (\frac{m-1}{2})^{\frac{m}{2}} e^{-\frac{m-1}{2}+\frac{1}{6(m-1)}}}{\sqrt{2\pi}(\frac{n}{2})^{\frac{n+1}{2}} e^{-\frac{n}{2}} \sqrt{2\pi} (2\gamma)^{2\gamma+\frac{1}{2}} e^{-2\gamma}}\, \varepsilon^{\gamma} \\
&\leq \frac{e^{1+\frac{1}{6(m-1)}}}{2\sqrt{\pi\gamma}}\, \frac{(\frac{m-1}{2})^{\frac{m}{2}}}{(\frac{n}{2})^{\frac{n-1}{2}}} \left(\frac{2e^2\varepsilon}{4\gamma^2 e}\right)^{\gamma}.
\end{aligned}$$

Now,

$$\begin{aligned}
\frac{(\frac{m-1}{2})^{\frac{m}{2}}}{(\frac{n}{2})^{\frac{n-1}{2}}} = \left(\frac{m-1}{n}\right)^{\frac{n-1}{2}} \left(\frac{m-1}{2}\right)^{\gamma} &\leq \left(1 + \frac{k-2}{n}\right)^{\frac{n}{2}} \left(\frac{m-1}{2}\right)^{\gamma} \\
&\leq e^{\frac{k-2}{2}} \left(\frac{m-1}{2}\right)^{\gamma} = \frac{1}{e} \left(\frac{e(m-1)}{2}\right)^{\gamma}.
\end{aligned}$$

Combining this inequality with the facts that $m-1 \geq 1$ and $\gamma \geq 1/2$ yields

$$\mathbb{P}(s_1(\Gamma(X)) \leq \varepsilon) \leq \frac{e^{1/6}}{\sqrt{2\pi}}\Big(\frac{e^2(m-1)}{4\gamma^2}\,\varepsilon\Big)^{\gamma} \leq \Big(\frac{e^2 m}{k^2}\,\varepsilon\Big)^{\gamma}.$$

Because of (3.19), we have that $p(x) \sim L_{n,m}x^{\gamma-1}$ as $x \to 0$. Therefore, one easily sees that $\gamma = k/2$ is the minimal value of γ for (3.16) to holds. ■

Using this result and Proposition 3.4, we immediately obtain the following:

Proposition 3.7 *Let $m > n$ be given and assume that the random matrix $\Gamma(X)^T\Gamma(X)$ associated with the function Γ defined in (2.2) follows a Wishart distribution. Let q be such that*

$$1 \leq q < k = m - n + 1. \tag{3.21}$$

Then we have

$$K_q(\Gamma) \leq \frac{e^2 m}{k^2}\Big(1 - \frac{q}{k}\Big)^{-\frac{2}{q}} \tag{3.22}$$

where K_q is defined in (3.3).

Combining this result and the result of Theorem 3.3, if $\rho(\bar{\beta}) \in \mathrm{L}^p(\Omega)$ with $p > 2$, the convergence of the algorithm is ensured if $K_q < \infty$ in (3.15) with $2p^{-1} + 1q^{-1} = 1$. This means, (see (3.21))

$$2 \leq \frac{2p}{p-2} < m - n + 1$$

or equivalently

$$m > n + \frac{p+2}{p-2}.$$

If for example $\rho(\bar{\beta}) \in \mathrm{L}^\infty(\Omega)$, $m \geq n+2$ will be sufficient.

Assume still that $\rho(\bar{\beta}) \in \mathrm{L}^\infty(\Omega)$. Using (3.22) with $q = 2$, it can be seen in view of (3.18) that the cost of the algorithm is bounded by

$$C\varepsilon^{-2}n^3m^3 \log n\, \frac{m}{k^2}\Big(1 - \frac{2}{k}\Big)^{-1}\|\rho(\bar{\beta})\|^2_{\mathrm{L}^\infty}.$$

for some constant C independent of n and m. Using the notation $\gamma = k/2$, we can rewrite this cost in term of γ as

$$C'\varepsilon^{-2}n^3 \log n\, \frac{(n+2\gamma-1)^4}{\gamma(\gamma-1)}\|\rho(\bar{\beta})\|^2_{\mathrm{L}^\infty}.$$

To determine the optimal choice of m, let us now try to find the optimal number γ that minimizes this cost. The derivative of this expression with respect to γ has the same sign as

$$8\gamma(\gamma-1) - (n+2\gamma-1)(2\gamma-1) = 4\gamma^2 - 2(n+2)\gamma + n - 1.$$

Since this quantity is negative if $\gamma = 1/2$, the only root of this polynomial greater than 1 is given by

$$\gamma^* = \frac{n + 2 + \sqrt{n^2 + 8}}{4},$$

which is the optimal choice of γ in terms of computational effort. This yields an optimal choice $m^* \sim 2\gamma^* + n - 1$. Note that for large n, we have $\gamma^* \sim n/2$ and $m^* \sim 2n$.

With this optimal choice, the computational cost of the algorithm can be written as

$$C_n \varepsilon^{-2} \left\|\rho(\bar{\beta})\right\|_{\mathrm{L}^\infty}^2 \quad \text{with} \quad C_n \sim C n^5 \log n \quad \text{as} \quad n \to +\infty. \tag{3.23}$$

Considering a similar calculation with $q = 1$, we can easily see that the consistency error of Proposition 3.2 for this choice of parameters can be bounded by

$$C'_n \|\rho(a)\|_{\mathrm{L}^\infty} \quad \text{with} \quad C'_n \sim C' n \quad \text{as} \quad n \to +\infty.$$

4 The general case

Let us now consider the general case where $\Gamma(X)^T \Gamma(X)$ is not assumed to be a.s. invertible.

Let $\sigma > 0$ be given. We denote by $\mathbb{E}^\sigma$ (resp. Cov^σ) the expectation (resp. covariance matrix) conditionally on the event $\{s_1(\Gamma(X)) > \sigma\}$. As an approximation of the solution of the least-squares problem, we will examine the conditional expectation

$$\bar{\beta}^\sigma = \mathbb{E}^\sigma(\beta). \tag{4.1}$$

4.1 Consistency, convergence and computational cost

We first generalize the consistency results given in Proposition 3.2:

Proposition 4.1 *Let a_j, $j = 1, \ldots, n$ be n numbers a_j. Assume that $\rho(a) \in \mathrm{L}^p(\Omega)$ for some $p \in (1, +\infty]$, then*

$$\|\bar{\beta}^\sigma - a\|_2 \leq \mathbb{E}^\sigma \|\beta - a\|_2 \leq \frac{\sqrt{n} m}{\mathbb{P}(s_1(\Gamma(X)) \geq \sigma)^{1/p}} \left(\mathbb{E}^\sigma \frac{1}{s_1(\Gamma(X))^{q/2}}\right)^{1/q} \|\rho(a)\|_{\mathrm{L}^p(\Omega)}$$

where q is such that $q^{-1} + p^{-1} = 1$. If $\rho(a) \in \mathrm{L}^1(\Omega)$, then we have

$$\|\bar{\beta}^\sigma - a\|_2 \leq \mathbb{E}^\sigma \|\beta - a\|_2 \leq \frac{\sqrt{n} m}{\sqrt{\sigma}\, \mathbb{P}(s_1(\Gamma(X)) \geq \sigma)} \|\rho(a)\|_{\mathrm{L}^1(\Omega)}.$$

Proof. Noticing that

$$\mathbb{E}^\sigma \|\rho(a)(X)\|_1^p \leq \frac{\mathbb{E}\|\rho(a)(X)\|_1^p}{\mathbb{P}(s_1(\Gamma(X)) \geq \sigma)},$$

the proof is exactly the same as that put forth in Lemma 3.1 and Proposition 3.2. ■

The previous result implies in particular that conditional expectation (4.1) is always well defined for $\sigma > 0$ as soon as $f \in \mathrm{L}^p(\Omega)$ for some $p \in [1, +\infty]$. Indeed, $\rho(0) = f$ and

$$\mathbb{E}^\sigma \frac{1}{s_1(\Gamma(X))^{q/2}} \leq \sigma^{-q/2}$$

for any $q \geq 1$.

The following result generalizes Theorem 3.3 to the case where $\sigma > 0$. Its proof is very similar to that of Theorem 3.3. We will, hence, omit it here.

Theorem 4.2 *Assume that the function $\rho(\bar{\beta}^\sigma) \in \mathrm{L}^p(\Omega)$ for $p \in (2, +\infty]$. We have*

$$\mathrm{Tr}(\mathrm{Cov}^\sigma(\beta)) \leq \frac{nm^2}{\mathbb{P}(s_1(\Gamma(X)) \geq \sigma)^{2/p}} \Big(\mathbb{E}^\sigma \frac{1}{s_1(\Gamma(X))^{p/(p-2)}}\Big)^{(p-2)/p} \|\rho(\bar{\beta}^\sigma)\|^2_{\mathrm{L}^p(\Omega)} \tag{4.2}$$

with the conventions $p/(p-2) = 1$ if $p = +\infty$.

If $\rho(\bar{\beta}^\sigma) \in \mathrm{L}^2(\Omega)$, we have the estimate

$$\mathrm{Tr}(\mathrm{Cov}^\sigma(\beta)) \leq \frac{nm^2}{\sigma\, \mathbb{P}(s_1(\Gamma(X)) \geq \sigma)} \|\rho(\bar{\beta}^\sigma)\|^2_{\mathrm{L}^2(\Omega)}. \tag{4.3}$$

It is always possible to resort to the trivial inequality

$$\forall\, r, \sigma > 0, \quad \mathbb{E}^\sigma(s_1(\Gamma(X))^{-r}) \leq \sigma^{-r}$$

to infer explicit bounds from the previous two results. To attain a more accurate estimate, we may use the following easy generalization of Proposition 3.4, given here without proof.

Proposition 4.3 *Fix $\sigma \geq 0$ and assume that random variable Y satisfies the following estimate: There exist constants δ and γ such that*

$$\forall\, \epsilon \geq \sigma, \quad \mathbb{P}(Y \leq \epsilon) \leq (\delta\epsilon)^\gamma. \tag{4.4}$$

Then, for any $r \neq \gamma$, if $0 < \sigma < \delta^{-1}$,

$$\mathbb{E}(Y^{-r} \mid Y \geq \sigma) \leq \frac{\delta^r}{1 - (\delta\sigma)^\gamma} \Big(\frac{1}{1 - r/\gamma} + \frac{(\delta\sigma)^{-(r-\gamma)}}{1 - \gamma/r}\Big). \tag{4.5}$$

We now consider the cost of the algorithm: Let $\beta^{(\sigma)}$ denote a random variable having the law of β conditioned on $\{s_1(\Gamma(X)) \geq \sigma\}$. The cost of the algorithm is determined by

- the number N of simulations of $\beta^{(\sigma)}$ needed to ensure that the diameter of the confidence region for the Monte Carlo method for the estimation of $\mathbb{E}(\beta^{(\sigma)}) = \mathbb{E}^\sigma(\beta) = \bar{\beta}^\sigma$ is smaller than a given precision ε. To control this, we use the upper bound on the confidence region diameter given by (3.14), where η is the level of confidence of the approximation;
- the average number of draws of the random variable X needed to simulate a realization of $\beta^{(\sigma)}$, which is $1/\mathbb{P}(\pi(X) \geq \sigma)$. Note that a draw corresponds to simulating a nm-dimensional random variable, which are considered to bear a cost proportional to nm.
- the computation of the $n \times n$ matrix $\Gamma(X)\Gamma(X)^T$, which is of order n^2m — all other computational costs, including the cost of the computation of $s_1(\Gamma(X))$ or the inversion of $\Gamma(X)^T\Gamma(X)$, are of a smaller order with respect to the dimension n of the problem, provided that $m \geq n$.

Consequently, the cost of the algorithm is bounded by

$$CN\mathbb{P}(s_1(\Gamma(X)) \geq \sigma)^{-1}(nm + n^2m)$$

for some constants $C > 0$. As

$$N \sim 4x(n,\eta)^2\varepsilon^{-2}\mathrm{Tr}(\mathrm{Cov}^\sigma(\beta)),$$

because of (3.13), the cost can be bounded by

$$C\varepsilon^{-2}n^2m\log n\mathbb{P}(s_1(\Gamma(X)) \geq \sigma)^{-1}\mathrm{Tr}(\mathrm{Cov}^\sigma(\beta)).$$

Thus, if $\rho(\bar{\beta}^\sigma) \in \mathrm{L}^p(\Omega)$ for $p \in (2, +\infty]$, because of Theorem 4.2, the cost is bounded by

$$C\varepsilon^{-2}n^3m^3\log n\,\mathbb{P}(s_n(\Gamma(X)) \geq \sigma)^{-1-\frac{2}{p}}\left(\mathbb{E}^\sigma\frac{1}{s_1(\Gamma(X))^{\frac{p}{p-2}}}\right)^{\frac{p-2}{p}}\|\rho(\bar{\beta}^\sigma)\|^2_{\mathrm{L}^p(\Omega)}$$

for some constant $C > 0$. Similarly, if $\rho(\bar{\beta}^\sigma) \in \mathrm{L}^2(\Omega)$, it is bounded by

$$C\varepsilon^{-2}n^3m^3\log n\,\mathbb{P}(s_n(\Gamma(X)) \geq \sigma)^{-2}\sigma^{-1}\|\rho(\bar{\beta}^\sigma)\|^2_{\mathrm{L}^p(\Omega)}.$$

We, hence, can see that the choice of an optimal threshold σ has to be balanced to optimize the ratio between $\mathbb{E}^\sigma s_1(\Gamma(X))^{-r}$ and probability $\mathbb{P}(s_n(\Gamma(X)) \geq \sigma)$ at some appropriate powers.

4.2 The sub-Gaussian case

We conclude this section by considering the case where $\Omega = \mathbb{R}^n$,

$$\mathrm{d}\mu(x) = \otimes_{i=1}^n \mathrm{d}\nu(x_i)$$

for some probability measure ν on $\mathbb{R}$, and $\gamma_j(x) = h(x_j)$ for $j \in \{1, \dots, n\}$ for some function h on $\mathbb{R}$. This is tantamount to the case of an approximation of the function f on $\mathbb{R}^n$ by a linear combination of functions depending on only one variable.

In this case, it is clear that all the entries of matrix $\Gamma(X)$ are i.i.d. Let us assume that these random variables are *sub-Gaussian*, *i.e.*

$$\forall t > 0, \quad \nu(\{x \in \mathbb{R} : |h(x)| > t\}) \leq 2\exp(-t^2/R^2)$$

for some $R > 0$. Such is the case, in particular if h is bounded or if ν has compact support and h is continuous on the support of ν. In [13, Thm. 1.1], Rudelson & Vershynin have recently proved that, under this assumption, there exist constants A and B depending only on R such that, for all $m \geq n$ and all $\epsilon > 0$,

$$\mathbb{P}\Big(s_1(\Gamma(X)) \leq \epsilon(\sqrt{m} - \sqrt{n-1})^2\Big) \leq (A\epsilon)^{(m-n+1)/2} + e^{-Bm}. \tag{4.6}$$

Writing just like in Subsection 3.5 k for $m - n + 1$, it can be seen that

$$\begin{aligned} \mathbb{P}(s_1(\Gamma(X)) \leq \varepsilon) &\leq \Big(\frac{\sqrt{A\varepsilon}}{\sqrt{m} - \sqrt{n-1}}\Big)^k + \big(e^{-Bm/k}\big)^k \\ &\leq \Big(\frac{\sqrt{A\varepsilon}}{\sqrt{m} - \sqrt{n-1}} + e^{-Bm/k}\Big)^k. \end{aligned}$$

(4.4), therefore, holds for $Y = s_1(\Gamma(X))$ and

$$\sigma \geq \sigma_0 := \frac{B^2 m^2 (\sqrt{m} - \sqrt{n-1})^2}{k^2 A} e^{-2Bm/k},$$
$$\delta = \frac{(1 + k/Bm)^2 A}{(\sqrt{m} - \sqrt{n-1})^2}$$
$$\text{and} \quad \gamma = \frac{k}{2}.$$

Note that, since $\delta\sigma_0 = (1 + Bm/k)^2 e^{-2Bm/k} < 1$, the inequality in (4.4) is not trivial and supplies some information on the law of $s_1(\Gamma(X))$.

As in Subsection 3.5, the inequality (4.5) can be combined with the results of Theorem 4.2 and Proposition 4.1 to obtain a precise error estimate and convergence bounds in this case.

Such computations are, however, cumbersome because the optimal choice of σ cannot be determined explicitly. The result in terms of order of computational cost in n, nevertheless, appears to be relatively unaffected by the choice of σ — *i.e.* it acts only on the constant factor. For instance, under the assumption that $\rho(\bar{\beta}^\sigma) \in \mathrm{L}^\infty(\Omega)$, because of Proposition 4.3, the computational cost is smaller than

$$C\varepsilon^{-2}n^3 \log n \|\rho(\bar{\beta}^\sigma)\|^2_{\mathrm{L}^\infty} \frac{m^3\delta}{(1-(\delta\sigma)^\gamma)^2}\Big(\frac{1}{1-1/\gamma} + \frac{(\delta\sigma)^{\gamma-1}}{1-\gamma}\Big)$$

If one chooses $\sigma = \sigma_0$ and if one assumes that $(\sigma_0\delta)^\gamma \to 0$ as $n \to +\infty$, observing that

$$\delta = \frac{(\sqrt{m}+\sqrt{n-1})^2(1+k/Bm)^2 A}{k^2} \leq \frac{Cm}{k^2},$$

the cost is bounded from above by

$$C\varepsilon^{-2}n^3 \log n \frac{m^4}{\gamma(\gamma-1)} \|\rho(\bar{\beta}^\sigma)\|^2_{\mathrm{L}^\infty}$$

for $\gamma > 1$. We recognize the same cost as in Subsection 3.5. The optimal choice of γ, therefore, behaves as $n/2$ as $n \to +\infty$ — and for this choice we actually have $(\sigma_0\delta)^\gamma \to 0$). The cost is bounded by

$$C\varepsilon^{-2}n^5 \log n \|\rho(\bar{\beta}^\sigma)\|^2_{\mathrm{L}^\infty}$$

for some constant $C > 0$. One can check that any other choice of σ yields the same order in n as $n \to +\infty$, should one choose $\gamma \sim n/2$.

It ought to be noted that these bounds do not allow one to pick $\sigma = 0$. As far as we know, this seems to be an open and difficult question to prove that (4.6) holds without the right-hand-side, additive term e^{-Bm}. In particular, it requires additional assumptions to hold — *e.g.* random variable $h(Y)$, where Y has law ν, has no atom, *i.e.* that $\nu(\{h = y\}) = 0$ for all $y \in \mathbb{R}$ (otherwise, the matrix $\Gamma(X)$ could have $m - n$ identical rows, and, thus, have a rank less than n, with non-zero probability).

5 Numerical examples

Let us now illustrate the method through numerical examples, considering first the case of linear approximation in a domain of the form $[0,1]^n$. We will then compare our approach with the original SADM method developed in [6].

5.1 Toy problem

We first consider here the case where $\Omega = [0,1]^n$, and

$$f(x) = \sum_{i=1}^{n} x_i(1 + rx_i).$$

where $r > 0$ is a parameter. We seek an approximation $\sum_{i=1}^{n} a_j x_j$ of $f(x)$, and use the algorithm described above with $\sigma = 0$ and $m = n + 2$. In Figures 1 and 2, we plot the evolution of the first component of $\bar{\beta}_N$ for $N = 1$ to $N = 2.10^5$, in those cases where $n = 3$, $n = 10$, $n = 30$ and $n = 60$. We observe that the rate of convergence is similar in these four cases.

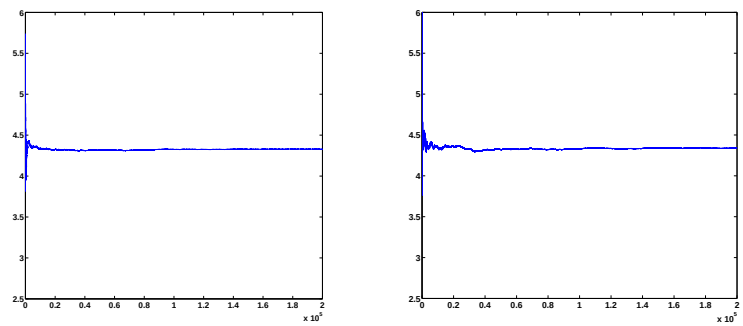

Figure 1: Convergence of β for $n = 3$ (left) and $n = 10$ (right).

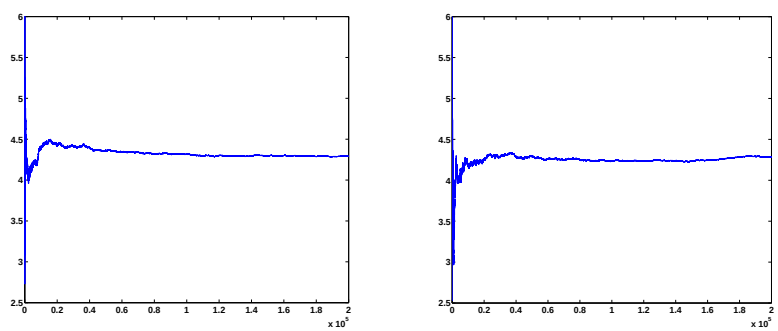

Figure 2: Convergence of β for $n = 30$ (left) and $n = 60$ (right).

In Figure 3, we set $n = 10$. On the left, we plot the evolution of the first component of $\bar{\beta}_N$ defined with $m = n + 2$ for $N = 1$ to $N = 10^4$ for $r = 5$, $r = 1$

and $r = 0.1$. As expected, this first component tends to 1, and with a better rate of convergence for small r. On the right, we plot the evolution of $\bar{\beta}_N$ in the case where $r = 0.1$, but with $m = n$. Convergence is not reached.

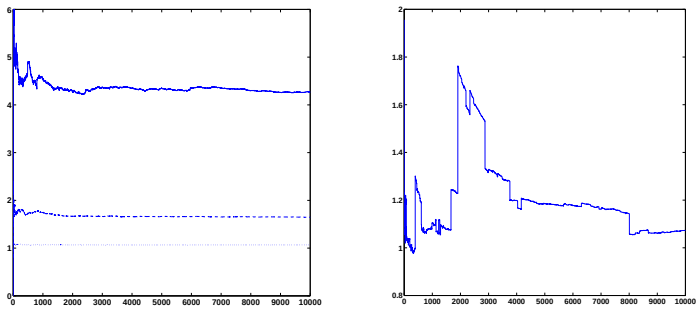

Figure 3: Left: Convergence of β for $r = 5$ (line), $r = 1$ (dashed line) and $r = 0.1$ (dotted line). Right: Result for $m = n$.

5.2 Improvement of the SADM method

The statistical analysis of distributed multipoles (SADM) algorithm put forth in [6] corresponds to a problem of the form **(c)**, where $(\alpha_j)_{j=1}^n$ represent the unknown multipoles borne by the n particles, and $\gamma_j(x)$ the electrostatic potential functions decaying with respect to $||x - x_j||$ where $x_1, \ldots, x_n$ denote the positions of the particles. The space Ω is made of M points in the three-dimensional Cartesian space, lying away from the atomic positions, with $M >> n$.

However more computationally intensive than the least-squares scheme, this pictorial approach provides a valuable information as to whether the atomic multipoles are appropriately defined, depending on how spread out the corresponding distributions are. For instance, description of the molecular electrostatic potential of dichlorodifluoromethane (CCl_2F_2) by means of a simple point-charge model yields a counterintuitive $C^{\delta-}$—$X^{\delta+}$ bond polarity — where X = Cl or F, blatantly violating the accepted rules of electronegativity differences. Whereas the least-squares route merely supplies crude values of the charge borne by the participating atoms, the SADM method offers a diagnosis of pathological scenarios, like that of dichlorodifluoromethane. In the latter example, the charge centered on the carbon atom is indeterminate, as mirrored by its markedly spread distribution [6]. The crucial issue of buried atoms illustrated here in the particular instance of CCl_2F_2 can be tackled by enforcing artificially the correct bond polarity by means of hyperbolic restraints [3]. Violations of the classical rules

of electronegativity differences may, however, often reflect the incompleteness of the electrostatic model — *e.g.* describing an atomic quadrupole by a mere point charge. Addition of atomic dipoles to the rudimentary point-charge model restores the expected, intuitive $C^{\delta+}$—$X^{\delta-}$ bond polarity [6].

In this section, we revisit the prototypical example of the three-point charge model of water. The molecular geometry was optimized at the MP2/6-311++G(d,p) level of approximation. The electrostatic potential was subsequently mapped on a grid of 2,106 points surrounding the molecule, at the same level of theory, including inner-shell orbitals. All the calculations were carried out with the Gaussian 03 suite of programs [9]. Brute-force solution of the least-squares problem (1.1), employing the Opep code [2], yields a net charge of -0.782 electron-charge unit (e.c.u.) on the oxygen atom — hence, a charge of $+0.391$ e.c.u. borne by the two hydrogen atoms, with a root-mean square deviation between the point-charge model regenerated and the quantum-mechanical electrostatic potential of 1.09 atomic units, and a mean signed error of 51.1 %. This notoriously large error reflects the incompleteness of the model — a simple point charge assigned to the oxygen atom being obviously unable to describe in a satisfactory fashion the large quadrupole borne by the latter.

On account of the $\mathscr{C}_{2v}$ space-group symmetry of water, only one net atomic charge would, in principle, need to be determined — the point charges borne by the two hydrogen atoms being inferred from that of the oxygen atom. Inasmuch as the SADM scheme is concerned, this symmetry relationship translates to a single equation to be solved per realization or experiment. Without loss of generality, two independent parameters will, however, be derived from the electrostatic potential, the point charges borne by the two hydrogen being assumed to be equal. Furthermore, in lieu of solving the individual $C^2_{2,106}$ systems of 2×2 linear equations, incommensurable with the available computational resources, it was chosen to select randomly 500,000 such systems.

The running averages of the charge borne by the oxygen atom are shown in Figure 4 as a function of the number of individual realizations, for the SADM algorithm with $n = N_s$ points and its proposed enhancement, using 2, 4 and 8 additional grid points per realization — with the notations utilized in the previous section, the latter translates to $m = N_s + 2$, $N_s + 4$ and $N_s + 8$. From the onset, it can be seen that the SADM scheme yields the worst agreement with the target value derived from the least-squares problem (1.1), and that inclusion of supplementary equations to the SADM algorithm rapidly improves the accord. However minute, this improvement is perceptible as new grid points are added to the independent realizations. Equally perceptible is the convergence property of the running average, reaching faster an asymptotic value upon addition of grid points. Congruent with what was established previously, the present set of results emphasizes that the SADM method cannot recover the value derived from

the least-squares equations. They further suggest that convergence towards the latter value will only be achieved in the limit where the number of added points coincides with the total number of grid points minus the number of parameters to be determined — *i.e.* one unique realization.

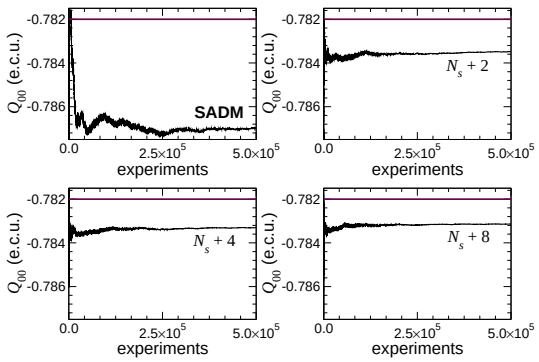

Figure 4: Running average of the point charge, Q_{00}, borne by the oxygen atom of water ($N_s = 2$ parameters) as a function of the number of independent realizations, wherein systems of 2×2 (SADM), 4×2, 6×2 and 10×2 linear equations are solved. The thick, dark horizontal line at $Q_{00} = -0.782$ e.c.u. corresponds to the solution of the least-squares problem.

Not too surprisingly, closer examination of the corresponding charge distributions in Figure 5 reveals that as additional grid points are added to the individual realizations, not only does the width of these distributions narrow down, but the latter are progressively reshaped. As was conjectured in [6], the SADM algorithm yields Cauchy distributions, which is apparent from Figure 5. Improvement of the method alters the form of the probability function, now closer to a normal distribution. Interestingly enough, the slightly skewed shape of the distributions, particularly visible on their left-hand side — as a probable manifestation of the incompleteness of the electrostatic model, precludes perfect enveloping by the model distributions, either Cauchy– or Gaussian–like.

Put together, the present computations reinforce the conclusions drawn hitherto, contradicting in particular the illegitimate assumption that the SADM and the least-squares solutions might coincide [6]. From a numerical standpoint, however, the results obtained from both strategies appear to be reasonably close, thereby warranting that the SADM algorithm should not be obliterated, as it constitutes a valuable pedagogical tool for assessing the appropriateness of electrostatic models.

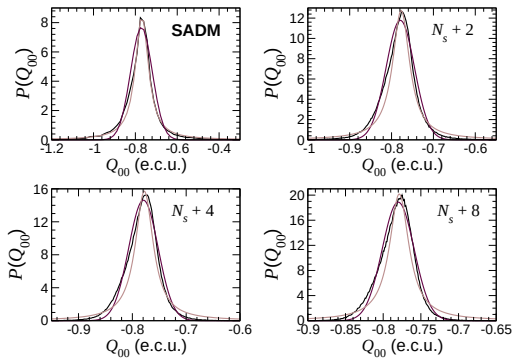

Figure 5: Normalized distributions of the charge, Q_{00}, borne by the oxygen atom of water ($N_s = 2$ parameters) obtained from 500,000 independent realizations, wherein systems of 2×2 (SADM), 4×2, 6×2 and 10×2 linear equations are solved (black curves). The light and dark curves correspond, respectively, to numerically fitted Cauchy and Gaussian distributions.

6 Conclusion

In this work, a probabilistic approach to high-dimensional least-squares approximations has been developed. Originally inspired by the SADM method introduced for the derivation of distributed atomic multipoles from the quantum-mechanical electrostatic potential, this novel approach can be generalized to a wide class of least-squares problems, yielding convergent and efficient numerical schemes in those cases where the space of approximation is very large.

This novel approach constitutes a marked improvement over the SADM method. Complete analysis of the numerical algorithm in general cases, in terms of both computational effort and optimal error estimatation, relies on open and difficult issues prevalent to random matrix problems.

References

[1] T. W. Anderson. *An introduction to multivariate statistical analysis*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, second edition, 1984.

[2] J. G. Ángyán, C. Chipot, F. Dehez, C. Hättig, G. Jansen, and C. Millot. Opep: A tool for the optimal partitioning of electric properties. *J. Comput. Chem.*, 24:997–1008, 2003.

[3] C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman. A well–behaved electrostatic potential based method using charge restraints for deriving atomic charges: The resp model. *J. Phys. Chem.*, 97:10269–10280, 1993.

[4] A. Bjorck. *Numerical Methods for Least Squares Problems*. SIAM, Philadelphia, PA, 1996.

[5] Zizhong Chen and Jack J. Dongarra. Condition numbers of Gaussian random matrices. *SIAM J. Matrix Anal. Appl.*, 27(3):603–620 (electronic), 2005.

[6] C. Chipot, J. G. Ángyán, and C. Millot. Statistical analysis of distributed multipoles derived from the molecular electrostatic potential. *Mol. Phys.*, 94:881–895, 1998.

[7] Alan Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4):543–560, 1988.

[8] Alan Edelman. The distribution and moments of the smallest eigenvalue of a random matrix of Wishart type. *Linear Algebra Appl.*, 159:55–80, 1991.

[9] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03 *Revision C.02*. Gaussian Inc., Wallingford, CT, 2004.

[10] E. J. Heller. Time dependent approach to semiclassical dynamics. *J. Chem. Phys.*, 62:1544–1555, 1975.

[11] Yves Meyer and Hong Xu. Wavelet analysis and chirps. *Appl. Comput. Harmon. Anal.*, 4(4):366–379, 1997.

[12] V. V. Petrov. *Sums of independent random variables*. Springer-Verlag, New York, 1975. Translated from the Russian by A. A. Brown, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82.

[13] M. Rudelson and R. Vershynin. The smallest singular value of a random rectangular matrix. Preprint, 2008.