The Annals of Statistics
2017, Vol. 45, No. 2, 591–614
DOI: 10.1214/16-AOS1460

ONLINE ESTIMATION OF THE GEOMETRIC MEDIAN IN HILBERT SPACES: NONASYMPTOTIC CONFIDENCE BALLS

BY HERVÉ CARDOT, PEGGY CÉNAC AND ANTOINE GODICHON-BAGGIONI

Université de Bourgogne Franche Comté

Estimation procedures based on recursive algorithms are interesting and powerful techniques that are able to deal rapidly with very large samples of high dimensional data. The collected data may be contaminated by noise so that robust location indicators, such as the geometric median, may be preferred to the mean. In this context, an estimator of the geometric median based on a fast and efficient averaged nonlinear stochastic gradient algorithm has been developed by [*Bernoulli* **19** (2013) 18–43]. This work aims at studying more precisely the nonasymptotic behavior of this nonlinear algorithm by giving nonasymptotic confidence balls in general separable Hilbert spaces. This new result is based on the derivation of improved L^2 rates of convergence as well as an exponential inequality for the nearly martingale terms of the recursive nonlinear Robbins–Monro algorithm.

1. Introduction. Dealing with large samples of observations taking values in high dimensional spaces, such as functional spaces, is not unusual nowadays. In this context, simple estimators of location such as the arithmetic mean can be greatly influenced by a small number of outlying values and robust indicators of location may be preferred to the mean. We focus in this work on the estimation of the geometric median, also called L^1-median or spatial median. It is a multivariate generalization of the real median introduced by [13] that can be defined in general metric spaces.

Let H be a separable Hilbert space, we denote by $\langle \cdot, \cdot \rangle$ its inner product and by $\| \cdot \|$ the associated norm. Let X be a random variable taking values in H, the geometric median m of X is defined by

$$m := \arg\min_{h \in H} \mathbb{E}\big[\|X - h\| - \|X\|\big]. \tag{1.1}$$

Many properties of this median in the general setting of separable Banach spaces, such as existence and uniqueness, as well as robustness are given in [14] (see also the review [25]). Recently, this median has received much attention in the literature. For example, [18] suggests to consider, in various statistical contexts, the geometric median of independent estimators in order to obtain much tighter concentration bounds. In functional data analysis, [15] consider resistant estimators of

Received January 2015; revised February 2016.
MSC2010 subject classifications. Primary 62G05; secondary 62L20.
Key words and phrases. Functional data analysis, martingales in Hilbert spaces, recursive estimation, robust statistics, spatial median, stochastic gradient algorithms.

the covariance operator based on the geometric median in order to derive a robust test of equality of the second-order structure for two samples. The geometric median is also chosen to be the central location indicator in various types of robust functional principal components analysis (see [12, 17] and [4]). The posterior geometric median of estimators has also been used in a robust Bayesian context by [19]. Finally, a general definition of the geometric median on manifolds is given in [11] and [1] with signal processing issues in mind.

Consider a sequence of i.i.d. copies $X_1, X_2, \dots, X_n, \dots$ of X. A natural estimator $\widehat{m}_n$ of m, based on $X_1, \dots, X_n$, is obtained by minimizing the empirical risk

$$\widehat{m}_n := \arg\min_{h \in H} \sum_{i=1}^{n} [\|X_i - h\| - \|X_i\|]. \tag{1.2}$$

Convergence properties of the empirical estimator $\widehat{m}_n$ are reviewed in [20] when the dimension of H is finite whereas the recent work of [9] proposes a deep asymptotic study for random variables taking values in separable Banach spaces. Given a sample $X_1, \dots, X_n$ the computation of $\widehat{m}_n$ generally relies on a variant of the Weiszfeld's algorithm (see, e.g., [28] and [16]) introduced by [27]. This iterative algorithm is relatively fast (see [6] for an improved version) but it is not adapted to handle very large datasets of high-dimensional data since it requires to store all the data in memory.

However, huge datasets are not unusual anymore with the development of automatic sensors and smart meters. In this context, [8] have developed a much faster algorithm, which thanks to its recursive nature does not require to store all the data and can be updated automatically when the data arrive sequentially. The estimation procedure is based on the simple following recursive scheme:

$$Z_{n+1} = Z_n + \gamma_n \frac{X_{n+1} - Z_n}{\|X_{n+1} - Z_n\|}, \tag{1.3}$$

where the sequence of steps (γ_n) controls the convergence of the algorithm and satisfy the usual conditions for the convergence of Robbins–Monro algorithms (see Section 3). The averaged version of the algorithm is defined as follows:

$$\overline{Z}_{n+1} = \overline{Z}_n + \frac{1}{n+1}(Z_{n+1} - \overline{Z}_n), \tag{1.4}$$

with $\overline{Z}_0 = 0$, so that $\overline{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$. The averaging step described in (1.4), and first studied in [23], allows a considerable improvement of the convergence compared to the initial Robbins–Monro algorithm described in (1.3). It is shown in [8] that the recursive averaged estimator $\overline{Z}_n$ and the empirical estimator $\widehat{m}_n$ have the same Gaussian limiting distribution.

However, the asymptotic normality shown in [8] does not give any clue of how far the distribution of the estimator is from its asymptotic law for any fixed sample size n. The aim of this work is to give new arguments in favor of the averaged stochastic estimator of the geometric median by providing a sharp control of its deviations around the true median, for finite samples. Indeed the obtention of finite sample guarantees with high probability is always desirable for the statisticians who have to study real data, since the samples under study will always have a finite sample size. Nice arguments for considering nonasymptotic properties of estimators are given, for example, in [24]. The obtention of such results generally requires much more mathematical efforts compared to more classical weak convergence results as well as more restrictive conditions on the existence of all the moments of the variable (see, e.g., [29] or [26]). Note also that, as far as we know, there are only very few results in the literature on nonasymptotic bounds for nonlinear recursive algorithms (see, however, [5] for recursive PCA or [3]).

The construction of our nonasymptotic confidence balls (see Theorem 4.1 and Theorem 4.2) rely on the obtention of the optimal rate of convergence in quadratic mean (see Theorem 3.1) of the Robbins–Monro algorithm used for estimating the geometric median as well as new exponential inequalities for "near" martingale sequences in Hilbert spaces (see Proposition 4.1), similar to the seminal result of [22] for martingales. These properties do not require any additional conditions on the moments of the data to hold. The proof of Theorem 3.1 is based on a new approach which consists in obtaining first, relations between the L^2 and the L^4 estimation errors and then make an induction using these relations to get the optimal rate of convergence in quadratic mean of Robbins–Monro algorithms. This new approach may give keys to obtain nonasymptotic results when the objective function only possesses locally strong convexity properties.

The paper is organized as follows. Section 2 recalls some convexity properties of the geometric median as well as the basic assumptions ensuring the uniqueness of the geometric median. In Section 3, the rates of convergence of the stochastic gradient algorithm are derived in quadratic mean as well as in L^4. In Section 4, an exponential inequality is derived borrowing ideas from [26]. It enables us to build nonasymptotic confidence balls for the Robbins–Monro algorithm as well as its averaged version. The most innovative part of the proofs is given in Section 5 whereas the other technical details are gathered in the Supplementary Material [7].

2. Assumptions on the median and convexity properties. Let us first state basic assumptions on the median.

(A1) The random variable X is not concentrated on a straight line: for all $h \in H$, there exists $h' \in H$ such that $\langle h, h' \rangle = 0$ and

$$\operatorname{Var}(\langle h', X \rangle) > 0.$$

(A2) X is not concentrated around single points: there is a constant $C > 0$ such that for all $h \in H$:

$$\mathbb{E}[\|X - h\|^{-1}] \leq C.$$

Assumption (A1) ensures that the median m is uniquely defined (see [14]). Assumption (A2) is closely related to small ball probabilities and to the dimension of H. It was proved in [10] that when $H = \mathbb{R}^d$, assumption (A2) is satisfied when $d \geq 2$ under classical assumptions on the density of X. A detailed discussion on assumption (A2) and its connection with small balls probabilities can be found in [8].

We now recall some results about convexity and robustness of the geometric median. We denote by $G : H \longrightarrow \mathbb{R}$ the convex function we would like to minimize, defined for all $h \in H$ by

$$G(h) := \mathbb{E}\big[\|X - h\| - \|X\|\big]. \tag{2.1}$$

This function is Fréchet differentiable on H, we denote by Φ its Fréchet derivative, and for all $h \in H$:

$$\Phi(h) := \nabla_h G = -\mathbb{E}\left[\frac{X - h}{\|X - h\|}\right].$$

Under previous assumptions, m is the unique zero of Φ.

Let us define $U_{n+1} := -\frac{X_{n+1} - Z_n}{\|X_{n+1} - Z_n\|}$ and let us introduce the sequence of σ-algebra $\mathcal{F}_n := \sigma(Z_1, \dots, Z_n) = \sigma(X_1, \dots, X_n)$. For all integers $n \geq 1$,

$$\mathbb{E}[U_{n+1} | \mathcal{F}_n] = \Phi(Z_n). \tag{2.2}$$

The sequence $(\xi_n)_n$ defined by $\xi_{n+1} := \Phi(Z_n) - U_{n+1}$ is a martingale difference sequence with respect to the filtration $(\mathcal{F}_n)$. Moreover, we have for all n, $\|\xi_{n+1}\| \leq 2$ and

$$\mathbb{E}\big[\|\xi_{n+1}\|^2 | \mathcal{F}_n\big] \leq 1 - \|\Phi(Z_n)\|^2 \leq 1. \tag{2.3}$$

Algorithm (1.3) can be written as a Robbins–Monro or a stochastic gradient algorithm:

$$Z_{n+1} - m = Z_n - m - \gamma_n \Phi(Z_n) + \gamma_n \xi_{n+1}. \tag{2.4}$$

We now consider the Hessian of G, which is denoted by $\Gamma_h : H \longrightarrow H$. It satisfies (see [12])

$$\Gamma_h = \mathbb{E}\left[\frac{1}{\|X - h\|}\left(I_H - \frac{(X - h) \otimes (X - h)}{\|X - h\|^2}\right)\right],$$

where I_H is the identity operator in H and $u \otimes v(h) = \langle u, h \rangle v$ for all $u, v, h \in H$. The following (local) strong convexity properties will be useful (see [8] for proofs).

PROPOSITION 2.1. *Under assumptions* (A1) *and* (A2), *for any real number $A > 0$, there is a positive constant c_A such that for all $h \in H$ with $\|h\| \leq A$, and for all $h' \in H$:*

$$c_A \|h'\|^2 \leq \langle h', \Gamma_h h' \rangle \leq C \|h'\|^2.$$

As a particular case, there is a positive constant c_m *such that for all* $h' \in H$:

$$c_m \|h'\|^2 \leq \langle h', \Gamma_m h' \rangle \leq C \|h'\|^2. \tag{2.5}$$

The following corollary recall some properties of the spectrum of the Hessian of G, in particular on the spectrum of Γ_m.

COROLLARY 2.1. *Under assumptions* (A1) *and* (A2), *for all* $h \in H$, *there is an increasing sequence of nonnegative eigenvalues* $(\lambda_{j,h})$ *and an orthonormal basis* $(v_{j,h})$ *of eigenvectors of* Γ_h *such that*

$$\Gamma_h v_{j,h} = \lambda_{j,h} v_{j,h},$$
$$\sigma(\Gamma_h) = \{\lambda_{j,h}, j \in \mathbb{N}\},$$
$$\lambda_{j,h} \leq C.$$

Moreover, if $\|h\| \leq A$, *for all* $j \in \mathbb{N}$ *we have* $c_A \leq \lambda_{j,h} \leq C$.

As a particular case, the eigenvalues $\lambda_{j,m}$ *of* Γ_m *satisfy,* $c_m \leq \lambda_{j,m} \leq C$, *for all* $j \in \mathbb{N}$.

The bounds are an immediate consequence of Proposition 2.1. Remark that with these different convexity properties of the geometric median, we are close to the framework of [2]. The difference comes from the fact that G does not satisfy the generalized self-concordance assumption which is central in the latter work.

3. Rates of convergence of the Robbins–Monro algorithms. If the sequence $(\gamma_n)_n$ of step-sizes fulfills the classical following assumptions:

$$\sum_{n \geq 1} \gamma_n^2 < \infty \quad \text{and} \quad \sum_{n \geq 1} \gamma_n = \infty,$$

and (A1) and (A2) hold, the recursive estimator Z_n is strongly consistent (see [8], Theorem 3.1). The first condition on the step-sizes ensures that the recursive algorithm converges toward some value in H whereas the second condition forces the algorithm to converge to m, the unique minimizer of G.

From now on, Z_1 is chosen so that it is bounded (consider, e.g., $Z_1 = X_1 \mathbb{1}_{\{\|X\| \leq M'\}}$ for some nonnegative constant M'). Consequently, there is a positive constant M such that for all $n \geq 1$:

$$\mathbb{E}\left[\|Z_n - m\|^2\right] \leq M.$$

Let us consider now sequences $(\gamma_n)_n$ of the form $\gamma_n = c_\gamma n^{-\alpha}$ where c_γ is a positive constant, and $\alpha \in (1/2, 1)$. Note that considering $\alpha = 1$ would be possible, with a suitable constant c_γ which is unknown in practice, in order to obtain the optimal parametric rate of convergence. The algorithm can be very sensitive to the values c_γ. That is why we prefer to introduce an averaging step with $\alpha < 1$, which is in practice and theoretically more efficient, since it has the same asymptotic variance as the empirical risk minimizer ([8], Theorem 3.4).

In order to get confidence balls for the median, the following additional assumption is supposed to hold.

(A3) There is a positive constant C such that for all $h \in H$:

$$\mathbb{E}\big[\|X-h\|^{-2}\big] \leq C.$$

This assumption ensures that the remainder term in the Taylor approximation to the gradient is bounded. Note that this assumption is also required to get the asymptotic normality in [8]. It is also assumed in [9] for deriving the asymptotic normality of the empirical median estimator. Remark that for the sake of simplicity, we have considered the same constant C in (A2) and (A3). As in (A2), Assumption (A3) is closely related to small ball probabilities and when $H=\mathbb{R}^d$, this assumption is satisfied when $d \geq 3$ under weak conditions.

We state now the first new and important result on the rates of convergence in quadratic mean of the Robbins–Monro algorithm. A comparison with Proposition 3.2 in [8] reveals that the term $\log n$ has disappeared as well as the constant C_N that was related to a sequence $(\Omega_N)_N$ of events whose probability was tending to one. This is a significant improvement which is crucial to get a deep study of the estimators and to get nonasymptotic results.

THEOREM 3.1. *Assuming* (A1)–(A3) *hold*, *the algorithm* (Z_n) *defined by* (1.3), *with* $\gamma_n = c_\gamma n^{-\alpha}$, *converges in quadratic mean*, *for all* $\alpha \in (1/2, 1)$ *and for all* $\alpha < \beta < 3\alpha - 1$, *with the following rate*:

$$\mathbb{E}\big[\|Z_n - m\|^2\big] = O\left(\frac{1}{n^\alpha}\right), \tag{3.1}$$

$$\mathbb{E}\big[\|Z_n - m\|^4\big] = O\left(\frac{1}{n^\beta}\right). \tag{3.2}$$

Upper bounds for the rates of convergence at order four are also given because they will be useful in several proofs. Remark that obtaining better rates of convergence at the order four would also be possible at the expense of longer proofs, and since it is not necessary here, it is not given.

The proof of this theorem relies on a new approach which consists in an induction on n using two decompositions of the algorithm which enables us to obtain an upper bound of the quadratic mean error and the L^4 error. Note that this approach can be used in several cases when the function we would like to minimize is only locally strongly convex.

LEMMA 3.1. *Assuming* (A1)–(A3) *hold*, *there are positive constants* C_1, C_2, C_3, C_4 *such that for all* $n \geq 1$:

$$\mathbb{E}\big[\|Z_n - m\|^2\big] \leq C_1 e^{-C_4 n^{1-\alpha}} + \frac{C_2}{n^\alpha} + C_3 \sup_{n/2-1 \leq k \leq n} \mathbb{E}\big[\|Z_k - m\|^4\big]. \tag{3.3}$$

The proof of Lemma 3.1 is given in Section 5. In order to get a rate of convergence of the last term in previous inequality, we use a second decomposition [see equation (2.4)], to get a bound of the fourth moment.

LEMMA 3.2. *Assuming the three assumptions* (A1) *to* (A3), *for all* $\alpha \in (1/2, 1)$, *there are a rank* n_α *and positive constants* C_1', C_2' *such that for all* $n \geq n_\alpha$:

$$\mathbb{E}[\|Z_{n+1} - m\|^4] \leq \left(1 - \frac{1}{n}\right)^2 \mathbb{E}[\|Z_n - m\|^4] + \frac{C_1'}{n^{3\alpha}} + C_2' \frac{1}{n^{2\alpha}} \mathbb{E}[\|Z_n - m\|^2]. \tag{3.4}$$

The proof of Lemma 3.2 is given in Section 5. The next result gives the exact rate of convergence in quadratic mean and states that it is not possible to get the parametric rates of convergence with the Robbins–Monro algorithm when $\alpha \in (1/2, 1)$.

PROPOSITION 3.1. *Assume* (A1)–(A3) *hold*, *for all* $\alpha \in (1/2, 1)$, *there is a positive constant* C' *such that for all* $n \geq 1$,

$$\mathbb{E}[\|Z_n - m\|^2] \geq \frac{C'}{n^\alpha}.$$

The proof of Proposition 3.1 is given in the Supplementary Material [7].

4. Nonasymptotic confidence balls.

4.1. *Nonasymptotic confidence balls for the Robbins–Monro algorithm.* The aim is now to derive an upper bound for $\mathbb{P}[\|Z_n - m\| \geq t]$, for $t > 0$. A simple and first result can be obtained by applying Markov's inequality and Theorem 3.1. We give below a sharper bound that relies on exponential inequalities that are close to the ones given in Theorem 3.1 in [22]. The following theorem gives nonasymptotic confidence balls for the Robbins–Monro algorithm.

THEOREM 4.1. *Assume that* (A1)–(A3) *hold. There is a positive constant* C *such that for all* $\delta \in (0, 1)$, *there is a rank* n_δ *such that for all* $n \geq n_\delta$,

$$\mathbb{P}\left[\|Z_n - m\| \leq \frac{C}{n^{\alpha/2}} \ln\left(\frac{4}{\delta}\right)\right] \geq 1 - \delta.$$

The proof is given in the Supplementary Material [7]. This result is obtained via the study of a linearized version of the gradient (2.4),

$$Z_{n+1} - m = Z_n - m - \gamma_n \Gamma_m (Z_n - m) + \gamma_n \xi_{n+1} - \gamma_n \delta_n, \tag{4.1}$$

where $\delta_n := \Phi(Z_n) - \Gamma_m(Z_n - m)$. Introducing for all $n \geq 1$, the following operators:

$$\alpha_n := I_H - \gamma_n \Gamma_m,$$
$$\beta_n := \prod_{k=1}^{n} \alpha_k = \prod_{k=1}^{n} (I_H - \gamma_k \Gamma_k),$$
$$\beta_0 := I_H,$$

by induction, (4.1) yields

$$Z_n - m = \beta_{n-1}(Z_1 - m) + \beta_{n-1} M_n - \beta_{n-1} R_n, \tag{4.2}$$

with $R_n := \sum_{k=1}^{n-1} \gamma_k \beta_k^{-1} \delta_k$ and $M_n := \sum_{k=1}^{n-1} \gamma_k \beta_k^{-1} \xi_{k+1}$. Note that (M_n) is a martingale sequence adapted to the filtration $(\mathcal{F}_n)$. Moreover,

$$\begin{aligned}
\mathbb{P}\big[\|Z_n - m\| \geq t\big] &\leq \mathbb{P}\left[\|\beta_{n-1} M_n\| \geq \frac{t}{2}\right] + \mathbb{P}\left[\|\beta_{n-1} R_n\| \geq \frac{t}{4}\right] \\
&\quad + \mathbb{P}\left[\|\beta_{n-1}(Z_1 - m)\| \geq \frac{t}{4}\right] \\
&\leq \mathbb{P}\left[\|\beta_{n-1} M_n\| \geq \frac{t}{2}\right] + 4\frac{\mathbb{E}[\|\beta_{n-1} R_n\|]}{t} \\
&\quad + 16\frac{\mathbb{E}[\|\beta_{n-1}(Z_1 - m)\|^2]}{t^2}.
\end{aligned} \tag{4.3}$$

Then we must get upper bounds for each term on the right-hand side of previous inequality. As explained in Remark 4.1 below, it is not possible to directly apply Theorem 3.1 of [22] to the quasi martingale term but the following proposition gives an analogous exponential inequality in the case where we do not have exactly a sequence of martingale differences.

PROPOSITION 4.1. *Let $(\beta_{n,k})_{(k,n)\in\mathbb{N}\times\mathbb{N}}$ be a sequence of linear operators on H and (ξ_n) be a sequence of H-valued martingale differences adapted to a filtration $(\mathcal{F}_n)$. Moreover, let (γ_n) be a sequence of positive real numbers. Then, for all $r > 0$ and for all $n \geq 1$,*

$$\begin{aligned}
&\mathbb{P}\left[\left\|\sum_{k=1}^{n-1} \gamma_k \beta_{n-1,k} \xi_{k+1}\right\| \geq r\right] \\
&\quad \leq 2e^{-r}\left\|\prod_{j=2}^{n}\big(1 + \mathbb{E}\big[e^{\|\gamma_{j-1}\beta_{n-1,j-1}\xi_j\|} - 1 - \|\gamma_{j-1}\beta_{n-1,j-1}\xi_j\| \,|\, \mathcal{F}_{j-1}\big]\big)\right\| \\
&\quad \leq 2\exp\left(-r + \left\|\sum_{j=2}^{n} \mathbb{E}\big[e^{\|\gamma_{j-1}\beta_{n-1,j-1}\xi_j\|} - 1 - \|\gamma_{j-1}\beta_{n-1,j-1}\xi_j\| \,|\, \mathcal{F}_{j-1}\big]\right\|\right).
\end{aligned}$$

The proof of Proposition 4.1 is in the Supplementary Material [7]. As in [26], it enables to give a sharp upper bound for $\mathbb{P}[\| \sum_{k=1}^{n-1} \gamma_k \beta_{n-1,k} \xi_{k+1} \| \geq t]$.

COROLLARY 4.1. *Let $(\beta_{n,k})$ be sequence of linear operators on H, (ξ_n) be a sequence of H-valued martingale differences adapted to a filtration $(\mathcal{F}_n)$ and (γ_n) be a sequence of positive real numbers. Let (N_n) and (σ_n^2) be two deterministic sequences such that*

$$N_n \geq \sup_{k \leq n-1} \| \gamma_k \beta_{n-1,k} \xi_{k+1} \| \qquad a.s. \quad and \quad \sigma_n^2 \geq \sum_{k=1}^{n-1} \mathbb{E}[\| \gamma_k \beta_{n-1,k} \xi_{k+1} \| | \mathcal{F}_n].$$

For all $t > 0$ and all $n \geq 1$,

$$\mathbb{P}\left[\left\| \sum_{k=1}^{n-1} \gamma_k \beta_{n-1,k} \xi_{k+1} \right\| \geq t \right] \leq 2 \exp\left(-\frac{t^2}{2(\sigma_n^2 + t N_n/3)} \right).$$

In our context, Corollary 4.1 can be written as follows.

COROLLARY 4.2. *Let $(N_n)_{n \geq 1}$ and $(\sigma_n^2)_{n \geq 1}$ be two deterministic sequences such that*

$$N_n \geq \sup_{k \leq n-1} \| \gamma_k \beta_{n-1} \beta_k^{-1} \xi_{k+1} \| \qquad a.s. \quad and \quad \sigma_n^2 \geq \sum_{k=1}^{n-1} \mathbb{E}[\| \gamma_k \beta_{n-1} \beta_k^{-1} \xi_{k+1} \| | \mathcal{F}_n].$$

Then, for all $t > 0$ and for all $n \geq 1$,

$$\mathbb{P}\left[\left\| \sum_{k=1}^{n-1} \gamma_k \beta_{n-1} \beta_k^{-1} \xi_{k+1} \right\| \geq t \right] \leq 2 \exp\left(-\frac{t^2}{2(\sigma_n^2 + t N_n/3)} \right).$$

REMARK 4.1. Note that $(\beta_{n-1} M_n)$ is not a martingale sequence. Then a first idea could be to apply Theorem 3.1 in [22] to the martingale term $M_n = \sum_{k=1}^{n-1} \beta_k^{-1} \gamma_k \xi_{k+1}$ but this does not work. Indeed, although there is a positive constant M such that $\| \beta_{n-1} M_n \| \leq M$ for all $n \geq 1$, the sequence $\| \beta_{n-1} \| \| M_n \|$ may not be convergent ($\| \beta_{n-1} \|$ denotes the usual spectral norm of operator β_{n-1}). Then it is possible to exhibit sequences (ξ_n) such that for all $t > 0$,

$$\lim_{n \to \infty} \mathbb{P}[\| \beta_{n-1} \| \| M_n \| \geq t] = 1,$$

$$\lim_{n \to \infty} \mathbb{P}[\| \beta_{n-1} M_n \| \geq t] = 0.$$

Indeed, let $\lambda_{\min}$ and $\lambda_{\max}$ be the lim inf and lim sup of the eigenvalues of the hessian Γ_m and suppose that $\lambda_{\min} < \lambda_{\max}$ and suppose $\gamma_n \lambda_{\max} \leq 1$ for all $n \geq 1$. Then $\| \beta_{n-1} \| = \prod_{k=1}^{n-1} (1 - \lambda_{\min} \gamma_k)$. Moreover, there exists a sequence $(h_n)_{n \geq 1}$ such that $\| h_n \| = 1$ for all $n \geq 1$, and a positive constant λ such that $\lambda_{\min} < \lambda \leq \lambda_{\max}$, and

$$\left\| \sum_{k=1}^{n-1} \gamma_k \beta_k^{-1} h_k \right\| = \sum_{k=1}^{n-1} \gamma_k \prod_{j=1}^{k} (1 - \lambda \gamma_j)^{-1}.$$

Thus,

$$\|\beta_{n-1}\| \left\| \sum_{k=1}^{n-1} \gamma_k \beta_k^{-1} h_k \right\| \xrightarrow[n\to\infty]{} +\infty.$$

4.2. *Nonasymptotic confidence balls for the averaged algorithm.* The following theorem, which is one of the most important result of this paper, provides nonasymptotic confidence balls for the averaged algorithm.

THEOREM 4.2. *Assume that* (A1)–(A3) *hold. For all* $\delta \in (0,1)$, *there is a rank* n_δ *such that for all* $n \geq n_\delta$,

$$\mathbb{P}\left[\|\overline{Z}_n - m\| \leq \frac{4}{\lambda_{\min}} \left(\frac{2}{3n} + \frac{1}{\sqrt{n}} \right) \ln\left(\frac{4}{\delta} \right) \right] \geq 1 - \delta.$$

The proof heavily relies on the following decomposition, which is obtained, as in [8] and [21], using decomposition (4.1). Indeed, summing and applying Abel's transform, we get

$$\begin{aligned} \Gamma_m(\overline{Z}_n - m) = {} & \frac{Z_1 - m}{\gamma_1 n} - \frac{Z_{n+1} - m}{\gamma_n n} + \frac{1}{n} \sum_{k=2}^{n} \left[\frac{1}{\gamma_k} - \frac{1}{\gamma_{k-1}} \right] (Z_k - m) \\ & - \frac{1}{n} \sum_{k=1}^{n} \delta_k + \frac{1}{n} \sum_{k=1}^{n} \xi_{k+1}. \end{aligned} \tag{4.4}$$

Noting that $\sum_{k=1}^{n} \xi_{k+1}$ is a martingale term adapted to the filtration $(\mathcal{F}_n)$, the proof of Theorem 4.2 relies on the application of Pinelis–Bernstein's lemma (see [26], Appendix A) to this term and on the fact that, thanks to Theorem 3.1, it can be shown that the other terms at the right-hand side of (4.4) are negligible.

REMARK 4.2. We can also have a more precise form of the rank n_δ (see the proof of Theorem 4.2):

$$n_\delta := \max\left\{ \left(\frac{6C_1'}{\delta \ln(\frac{4}{\delta})} \right)^{\frac{1}{1/2-\alpha/2}}, \left(\frac{6C_2'}{\delta \ln(\frac{4}{\delta})} \right)^{\frac{1}{\alpha-1/2}}, \left(\frac{6C_3'}{\delta \ln(\frac{4}{\delta})} \right)^{\frac{1}{2}} \right\}, \tag{4.5}$$

where C_1', C_2' and C_3' are constants. We can remark that the first two terms are the leading ones and if the rate α is chosen equal to $2/3$, they are of the same order that is $n_\delta = O((\frac{-1}{\delta \ln \delta})^6)$.

REMARK 4.3. We can make an informal comparison of previous result with the central limit theorem stated in ([8], Theorem 3.4), even if the latter result is only of asymptotic nature. Under assumptions (A1)–(A3), it has been shown that

$$\sqrt{n}(\overline{Z}_n - m) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \Gamma_m^{-1} \Sigma \Gamma_m^{-1}),$$

with

$$\Sigma = \mathbb{E}\left[\frac{(X-m)}{\|X-m\|} \otimes \frac{(X-m)}{\|X-m\|}\right].$$

This implies, with the continuity of the norm in H, that for all $t > 0$,

$$\lim_{n\to\infty} \mathbb{P}[\|\sqrt{n}(\overline{Z}_n - m)\| \geq t] = \mathbb{P}[\|V\| \geq t],$$

where V is a centered H-valued Gaussian random vector with covariance operator $\Delta_V = \Gamma_m^{-1}\Sigma\Gamma_m^{-1}$. Operator Δ_V is self-adjoint and nonnegative, so that it admits a spectral decomposition $\Delta_V = \sum_{j\geq 1} \eta_j v_j \otimes v_j$, where $\eta_1 \geq \eta_2 \geq \cdots \geq 0$ is the sequence of ordered eigenvalues associated to the orthonormal eigenvectors v_j, $j \geq 1$. Using the Karhunen–Loève's expansion of V, we directly get that

$$\|V\|^2 = \sum_{j\geq 1} \eta_j^2 V_j^2,$$

where $V_1, V_2, \ldots$ are i.i.d. centered Gaussian variables with unit variance. Thus, the distribution of $\|V\|^2$ is a mixture of independent Chi-square random variables with one degree of freedom. Computing the quantiles of $\|V\|$ to build confidence balls would require to know, or to estimate, all the (leading) eigenvalues of the rather complicated operator Δ_V and this is not such an easy task. Indeed, it would be necessary to project on a finite dimensional space to get the inverse of the Hessian before extracting the leading eigenvectors of the covariance. Finally, the last issue with the use of the central limit theorem to get confidence balls is that, to our knowledge, its rate of convergence is not known.

On the other hand, the use of the confidence balls given in Theorem 4.2 only requires the knowledge of $\lambda_{\min}$. This eigenvalue is not difficult to estimate since it can also be written as

$$\lambda_{\min} = \mathbb{E}\left[\frac{1}{\|X-m\|}\right] - \lambda_{\max}\left(\mathbb{E}\left[\frac{1}{\|X-m\|^3}(X-m)\otimes(X-m)\right]\right),$$

where $\lambda_{\max}(A)$ denotes the largest eigenvalue of operator A.

REMARK 4.4. Under previous assumptions, with analogous calculus to the ones in the proof of Theorem 4.2 and applying Theorem 3.1, it can be shown that there is a positive constant C' such that for all $n \geq 1$,

$$\mathbb{E}[\|\overline{Z}_n - m\|] \leq \frac{C'}{\sqrt{n}}.$$

Moreover, assuming the additional condition $\alpha > 2/3$, it can be shown that there is a positive constant C'' such that

$$\mathbb{E}[\|\overline{Z}_n - m\|^2] \leq \frac{C''}{n}.$$

The averaged algorithm converges at the parametric rate of convergence in quadratic mean.

5. Proofs.

5.1. *Proof of Theorem* 3.1. As explained in Section 3, the proof of Theorem 3.1 is based on Lemma 3.1, which allows to obtain an upper bound of the quadratic mean error, and on Lemma 3.2, which gives an upper bound of the L^4 error. We first prove Lemma 3.1. In order to do so, we have to introduce a new technical lemma which gives a bound of the rest in the Taylor's expansion of the gradient. This will enable us to bound the rest term $\beta_{n-1}R_n$ in decomposition (4.2).

LEMMA 5.1. *Assuming assumption* (A3), *there is a constant C_m such that for all $n \geq 1$*:

$$\|\delta_n\| \leq C_m \|Z_n - m\|^2, \tag{5.1}$$

where $\delta_n := \Phi(Z_n) - \Gamma_m(Z_n - m)$ is the second order term in the Taylor's decomposition of $\Phi(Z_n)$.

The proof is given in the Supplementary Material [7]. We can now prove Lemma 3.1.

PROOF OF LEMMA 3.1. Let us study the asymptotic behavior of the sequence of operators (β_n). Since Γ_m admits a spectral decomposition, we have $\|\alpha_k\| \leq \sup_j |1 - \gamma_k \lambda_j|$ where (λ_j) is the sequence of eigenvalues of Γ_m. Since for all $j \geq 1$ we have $0 < c_m \leq \lambda_j \leq C$, there is a rank n_0 such that for all $n \geq n_0$, $\gamma_n C < 1$. In particular, for all $n \geq n_0$ we have $\|\alpha_n\| \leq 1 - \gamma_n c_m$. Thus, there is a positive constant c_1 such that for all $n \geq 1$:

$$\|\beta_{n-1}\| \leq c_1 \exp\left(-\lambda_{\min} \sum_{k=1}^{n-1} \gamma_k\right) \leq c_1 \exp\left(-c_m \sum_{k=1}^{n-1} \gamma_k\right), \tag{5.2}$$

where $\lambda_{\min} > 0$ is the smallest eigenvalue of Γ_m. Similarly, there is a positive constant c_2 such that for all integer n and for all integer $k \leq n-1$:

$$\|\beta_{n-1}\beta_k^{-1}\| \leq c_2 \exp\left(-c_m \sum_{j=k+1}^{n-1} \gamma_j\right). \tag{5.3}$$

Moreover, for all $n > n_0$, $k \geq n_0$ such that $k \leq n-1$ (see [8] for more details),

$$\|\beta_{n-1}\beta_k^{-1}\| \leq \exp\left(-c_m \sum_{j=k+1}^{n-1} \gamma_j\right). \tag{5.4}$$

Using decomposition (4.2) again, we get

$$\begin{aligned} \mathbb{E}[\|Z_n - m\|^2] \leq {} & 3\mathbb{E}[\|\beta_{n-1}(Z_1 - m)\|^2] + 3\mathbb{E}[\|\beta_{n-1}M_n\|^2] \\ & + 3\mathbb{E}[\|\beta_{n-1}R_n\|^2]. \end{aligned} \tag{5.5}$$

We now bound each term on the right-hand side of previous inequality.

Step 1: *The quasi-deterministic term*: Using inequality (5.2), with help of an integral test for convergence, for all $n \geq 1$:

$$\begin{aligned}\mathbb{E}[\|\beta_{n-1}(Z_1 - m)\|^2] &\leq c_1^2 \exp\left(-2c_m \sum_{k=1}^{n-1} \gamma_k\right)\mathbb{E}[\|Z_1 - m\|^2] \\ &\leq c_1^2\left(-2c_m c_\gamma \int_1^n t^{-\alpha}\,dt\right)\mathbb{E}[\|Z_1 - m\|^2] \\ &\leq c_1^2 M \exp\left(2\frac{c_m c_\gamma}{1-\alpha}\right)\exp\left(-2\frac{c_m c_\gamma}{1-\alpha} n^{1-\alpha}\right).\end{aligned}$$

Since $\alpha < 1$, this term converges exponentially to 0.

Step 2: *The martingale term*: We have

$$\begin{aligned}\|\beta_{n-1}M_n\|^2 &= \left\|\sum_{k=1}^{n-1}\gamma_k \beta_{n-1}\beta_k^{-1}\xi_{k+1}\right\|^2 \\ &\leq \sum_{k=1}^{n-1}\gamma_k^2\|\beta_{n-1}\beta_k^{-1}\|^2\|\xi_{k+1}\|^2 \\ &\quad + 2\sum_{k=1}^{n-1}\sum_{k'<k}\gamma_k\gamma_{k'}\langle\beta_{n-1}\beta_k^{-1}\xi_{k+1}, \beta_{n-1}\beta_{k'}^{-1}\xi_{k'+1}\rangle.\end{aligned}$$

Since (ξ_n) is a sequence of martingale differences, for all $k' < k$ we have $\mathbb{E}[\langle\xi_{k+1}, \xi_{k'+1}\rangle] = 0$. Thus,

$$\mathbb{E}[\|\beta_{n-1}M_n\|^2] \leq \sum_{k=1}^{n-1}\gamma_k^2\|\beta_{n-1}\beta_k^{-1}\|^2, \tag{5.6}$$

because for all $k \in \mathbb{N}$, $\mathbb{E}[\|\xi_{k+1}\|^2] \leq 1$. The term $\|\beta_{n-1}\beta_k^{-1}\|$ converges exponentially to 0 when k is lower enough than n. We denote by $E(\cdot)$ the integer function and we isolate the dominating term. Let us split the sum into two parts:

$$\sum_{k=1}^{n-1}\gamma_k^2\|\beta_{n-1}\beta_k^{-1}\|^2 = \sum_{k=1}^{E(n/2)-1}\gamma_k^2\|\beta_{n-1}\beta_k^{-1}\|^2 + \sum_{k=E(n/2)}^{n-1}\gamma_k^2\|\beta_{n-1}\beta_k^{-1}\|^2. \tag{5.7}$$

We shall show that the first term on the right-hand side in (5.7) converges exponentially to 0 and that the second term on the right-hand side, which is the dominating one, converges at the rate $\frac{1}{n^\alpha}$. Indeed, we deduce from inequality (5.3):

$$\sum_{k=1}^{E(n/2)-1}\gamma_k^2\|\beta_{n-1}\beta_k^{-1}\|^2 \leq c_2\sum_{k=1}^{E(n/2)-1}\gamma_k^2 e^{-2c_m\frac{n}{2}\frac{c_\gamma}{n^\alpha}} \leq c_2 e^{-c_m c_\gamma n^{1-\alpha}}\sum_{k=1}^{E(n/2)-1}\gamma_k^2.$$

Since $\sum \gamma_k^2 < \infty$, we get $\sum_{k=1}^{E(n/2)-1} \gamma_k^2 \|\beta_{n-1}\beta_k^{-1}\|^2 = O(e^{-c_m c_\gamma n^{1-\alpha}})$.

We now bound the second term on the right-hand side of equality (5.7). Using inequality (5.4), for all $n > 2n_0$:

$$\begin{aligned}
\sum_{k=E(n/2)}^{n-1} \gamma_k^2 \|\beta_{n-1}\beta_k^{-1}\|^2 &\leq \sum_{k=E(n/2)}^{n-2} \gamma_k^2 e^{-2c_m \sum_{j=k+1}^{n-1}\gamma_j} + \gamma_{n-1}^2 \\
&\leq c_\gamma \left(\frac{1}{E(n/2)}\right)^\alpha \sum_{k=E(n/2)}^{n-2} \gamma_k e^{-2c_m \sum_{j=k+1}^{n-1}\gamma_j} + \gamma_{n-1}^2 \\
&\leq \frac{2^\alpha c_\gamma}{n^\alpha} \sum_{k=E(n/2)}^{n-2} \gamma_k e^{-2c_m \sum_{j=k+1}^{n-1}\gamma_j} + \gamma_{n-1}^2.
\end{aligned}$$

Moreover, for all $n > 2n_0$ and $k \leq n-2$,

$$\sum_{j=k+1}^{n-1} \gamma_j \leq \int_{k+1}^{n} \frac{c_\gamma}{s^\alpha} ds = \frac{c_\gamma}{1-\alpha}[n^{1-\alpha} - (k+1)^{1-\alpha}],$$

and hence $e^{-2c_m \sum_{j=k+1}^{n-1}\gamma_j} \leq e^{-2c_m \frac{c_\gamma}{1-\alpha}[n^{1-\alpha}-(k+1)^{1-\alpha}]}$. Since $\frac{1}{k^\alpha} \leq \frac{2}{(k+1)^\alpha}$,

$$\begin{aligned}
\sum_{k=E(n/2)}^{n-2} \gamma_k e^{2c_m \frac{c_\gamma}{1-\alpha}(k+1)^{1-\alpha}} &\leq 2^\alpha c_\gamma \sum_{k=E(n/2)}^{n-2} \frac{1}{(k+1)^\alpha} e^{2c_m \frac{c_\gamma}{1-\alpha}(k+1)^{1-\alpha}} \\
&\leq 2^\alpha c_\gamma \int_{E(n/2)}^{n-1} \frac{1}{(t+1)^\alpha} e^{2c_m \frac{c_\gamma}{1-\alpha}(t+1)^{1-\alpha}} dt \\
&\leq \frac{2^{\alpha-1}}{c_m} e^{2c_m \frac{c_\gamma}{1-\alpha} n^{1-\alpha}}.
\end{aligned}$$

Note that the integral test for convergence is valid because there is a rank $n_0' \in \mathbb{N}$ such that the function $t \longmapsto \frac{1}{(t+1)^\alpha} e^{2c_m \frac{c_\gamma}{1-\alpha}(t+1)^{1-\alpha}}$ is increasing on $[n_0', \infty)$. Let $n_1 := \max\{2n_0+1, n_0'\}$, for all $n \geq n_1$:

$$\sum_{k=E(n/2)}^{n-1} \gamma_k^2 \|\beta_{n-1}\beta_k^{-1}\|^2 \leq \frac{2^{2\alpha-1}c_\gamma}{c_m}\frac{1}{n^\alpha} + c_\gamma 2^{2\alpha}\frac{1}{n^{2\alpha}}. \tag{5.8}$$

Consequently, there is a positive constant C_2 such that for all $n \geq 1$,

$$3\mathbb{E}[\|\beta_{n-1}M_n\|^2] \leq C_2 \frac{1}{n^\alpha}. \tag{5.9}$$

REMARK 5.1. Note that splitting the sum in equation (5.7) is really crucial to get the good rate of convergence of the martingale term. Remark that a different split was considered in [8], which leads to a nonoptimal bound of the form

$$\mathbb{E}\big[\|\beta_{n-1}M_n\|^2\big] \leq \frac{C_2 \ln n}{n^{\alpha}}.$$

Step 3: *The rest term*: In the same way, we split the sum into two parts:

$$\sum_{k=1}^{n-1} \gamma_k \beta_{n-1}\beta_k^{-1}\delta_k = \sum_{k=1}^{E(n/2)-1} \gamma_k \beta_{n-1}\beta_k^{-1}\delta_k + \sum_{k=E(n/2)}^{n-1} \gamma_k \beta_{n-1}\beta_k^{-1}\delta_k. \tag{5.10}$$

One can check (see the proof of Lemma 5.3 for more details) that there is a positive constant M such that for all $n \geq 1$,

$$\mathbb{E}\big[\|Z_n - m\|^4\big] \leq M. \tag{5.11}$$

Moreover, by Lemma 5.1, $\|\delta_n\| \leq C_m \|Z_n - m\|^2$. Thus, for all $k, k' \geq 1$, the application of Cauchy–Schwarz's inequality gives us

$$\begin{aligned}
\mathbb{E}\big[\|\delta_k\|\|\delta_{k'}\|\big] &\leq C_m^2 \mathbb{E}\big[\|Z_k - m\|^2 \|Z_{k'} - m\|^2\big] \\
&\leq C_m^2 \sup_{n\geq 1} \mathbb{E}\big[\|Z_n - m\|^4\big] \leq C_m^2 M.
\end{aligned}$$

As a particular case, we also have $\mathbb{E}[|\langle \delta_k, \delta_{k'}\rangle|] \leq C_m^2 M$. Applying this result to the term on the right-hand side in (5.10),

$$\begin{aligned}
\mathbb{E}\left[\left\|\sum_{k=1}^{E(n/2)-1} \gamma_k \beta_{n-1}\beta_k^{-1}\delta_k\right\|^2\right] &\leq C_m^2 M \left[\sum_{k=1}^{E(n/2)-1} \gamma_k \|\beta_{n-1}\beta_k^{-1}\|\right]^2 \\
&\leq c_2 C_m^2 M e^{-2c_m c_\gamma n^{1-\alpha}} \left(\sum_{k=1}^{E(n/2)-1} \gamma_k\right)^2 \\
&\leq C_1' e^{-2c_m c_\gamma n^{1-\alpha}} n^{2-2\alpha}.
\end{aligned}$$

This term converges exponentially to 0. To bound the second term, we use the same idea as for the martingale term. Applying previous inequalities for the terms $\mathbb{E}[\|\delta_k\|\|\delta_{k'}\|]$ which appear in the double products, we get

$$\begin{aligned}
&\mathbb{E}\left[\left\|\sum_{k=E(n/2)}^{n-1} \gamma_k \beta_{n-1}\beta_k^{-1}\delta_k\right\|^2\right] \\
&\quad\leq C_m^2 \sup_{E(n/2)\leq k\leq n-1} \mathbb{E}\big[\|Z_k - m\|^4\big] \left[\sum_{k=E(n/2)}^{n-1} \gamma_k \|\beta_{n-1}\beta_k^{-1}\|\right]^2 \\
&\quad\leq C_3 \sup_{E(n/2)\leq k\leq n-1} \mathbb{E}\big[\|Z_k - m\|^4\big],
\end{aligned}$$

since $[\sum_{k=E(n/2)}^{n-1} \gamma_k \|\beta_{n-1}\beta_k^{-1}\|]^2$ is bounded. Indeed, one can check it with similar calculus to the ones in the proof of inequality (5.9). We put together the terms which converge exponentially to 0. □

To prove Lemma 3.2, we introduce two technical lemmas. The first one gives a bound of the decomposition in the particular case when $\|Z_n - m\|$ is not too large.

LEMMA 5.2. *If assumptions* (A1) *and* (A2) *holds*, *there are a rank n_α and a constant c such that for all $n \geq n_\alpha$, $\|Z_n - m\| \leq cn^{1-\alpha}$ yields*

$$\langle \Phi(Z_n), Z_n - m \rangle \geq \frac{1}{c_\gamma n^{1-\alpha}} \|Z_n - m\|^2. \tag{5.12}$$

As a corollary, *there is also a deterministic rank n'_α such that for all $n \geq n'_\alpha$, $\|Z_n - m\| \leq cn^{1-\alpha}$ yields*

$$\|Z_n - m - \gamma_n \Phi(Z_n)\|^2 \leq \left(1 - \frac{1}{n}\right)\|Z_n - m\|^2. \tag{5.13}$$

PROOF. We suppose that $\|Z_n - m\| \leq cn^{1-\alpha}$. We must consider two cases.

If $\|Z_n - m\| \leq 1$, then we have in particular $\|Z_n\| \leq \|m\| + 1$. Consequently, we get with Corollary 2.2 in [8] that there is a positive constant c_1 such that $\langle \Phi(Z_n), Z_n - m\rangle \geq c_1 \|Z_n - m\|^2$.

If $\|Z_n - m\| \geq 1$, since $\Phi(Z_n) = \int_0^1 \Gamma_{m+t(Z_n-m)}(Z_n - m)\, dt$,

$$\langle \Phi(Z_n), Z_n - m \rangle = \int_0^1 \langle Z_n - m, \Gamma_{m+t(Z_n-m)}(Z_n - m)\rangle\, dt.$$

Moreover, operators Γ_h are nonnegative for all $h \in H$. Applying Proposition 2.1 of [8], and since for all $t \in [0, \frac{1}{\|Z_n-m\|}]$ we have $\|m + t(Z_n - m)\| \leq \|m\| + 1$, there is a positive constant c_2 such that

$$\begin{aligned}
\langle \Phi(Z_n), Z_n - m \rangle &= \int_0^1 \langle Z_n - m, \Gamma_{m+t(Z_n-m)}(Z_n - m)\rangle\, dt \\
&\geq \int_0^{1/\|Z_n-m\|} \langle Z_n - m, \Gamma_{m+t(Z_n-m)}(Z_n - m)\rangle\, dt \\
&\geq \int_0^{1/\|Z_n-m\|} c_2 \|Z_n - m\|^2\, dt \\
&\geq \frac{c_2}{cn^{1-\alpha}} \|Z_n - m\|^2.
\end{aligned}$$

We can choose a rank n_α such that for all $n \geq n_\alpha$ we have $c_1 \geq \frac{1}{c_\gamma n^{1-\alpha}}$ which concludes the proof of inequality (5.12) with $c = c_2 c_\gamma$.

We now prove inequality (5.13). For all $n \geq n_\alpha$, $\|Z_n - m\| \leq cn^{1-\alpha}$ yields

$$\begin{aligned}\|Z_n - m - \gamma_n \Phi(Z_n)\|^2 &\leq \|Z_n - m\|^2 - \frac{2}{c_\gamma n^{1-\alpha}} \frac{c_\gamma}{n^\alpha} \|Z_n - m\|^2 \\ &\quad + \gamma_n^2 C^2 \|Z_n - m\|^2 \\ &= \left(1 - \frac{2}{n} + C^2 \frac{c_\gamma^2}{n^{2\alpha}}\right) \|Z_n - m\|^2.\end{aligned}$$

Thus, we can choose a rank $n'_\alpha \geq n_\alpha$ such that for all $n \geq n'_\alpha$ we have $C^2 \frac{c_\gamma^2}{n^{2\alpha}} \leq \frac{1}{n}$. Note that this is possible since $\alpha > 1/2$. $\square$

The second lemma shows that the probability for $\|Z_n - m\|$ to be large is very small as n increases.

LEMMA 5.3. *There is a positive constant C_α such that for all $n \geq 1$,*

$$\mathbb{P}[\|Z_n - m\| \geq cn^{1-\alpha}] \leq \frac{C_\alpha}{n^{4-\alpha}},$$

where c has been defined in the previous lemma.

The proof is given in the Supplementary Material [7].

PROOF OF LEMMA 3.2. For all $n \geq 1$,

$$\begin{aligned}\mathbb{E}[\|Z_{n+1} - m\|^4] &= \mathbb{E}[\|Z_{n+1} - m\|^4 \mathbb{1}_{\|Z_n - m\| \geq cn^{1-\alpha}}] \\ &\quad + \mathbb{E}[\|Z_{n+1} - m\|^4 \mathbb{1}_{\|Z_n - m\| < cn^{1-\alpha}}],\end{aligned} \tag{5.14}$$

with c defined in Lemma 5.2. Let us bound the first term in (5.14). Since $\|Z_{n+1} - m\| \leq \|Z_n - m\| + \gamma_n \leq \|Z_1 - m\| + \sum_{k=1}^n \gamma_k$ and since Z_1 is bounded or deterministic, there is a constant C'_α such that for all integer $n \geq 1$,

$$\|Z_n - m\| \leq C'_\alpha n^{1-\alpha}.$$

Consequently,

$$\begin{aligned}\mathbb{E}[\|Z_{n+1} - m\|^4 \mathbb{1}_{\|Z_n - m\| \geq cn^{1-\alpha}}] &\leq \mathbb{E}[(C'_\alpha (n+1)^{1-\alpha})^4 \mathbb{1}_{\|Z_n - m\| \geq cn^{1-\alpha}}] \\ &\leq (C'_\alpha (n+1)^{1-\alpha})^4 \mathbb{P}[\|Z_n - m\| \geq cn^{1-\alpha}].\end{aligned}$$

Thus, applying Lemma 5.3, we get

$$(C'_\alpha (n+1)^{1-\alpha})^4 \mathbb{P}[\|Z_n - m\| \geq cn^{1-\alpha}] \leq \frac{C'^4_\alpha C_\alpha (n+1)^{4-4\alpha}}{n^{4-\alpha}} \leq 2^{4-4\alpha} \frac{C'^4_\alpha C_\alpha}{n^{3\alpha}}.$$

We now bound the second term. Suppose that $\|Z_n - m\| \le cn^{1-\alpha}$. Since $\|\xi_{n+1}\| \le 2$, using Lemma 5.2, there is a rank n_α such that for all $n \ge n_\alpha$,

$$\begin{aligned}
&\|Z_{n+1} - m\|^2 \mathbb{1}_{\|Z_n - m\| < cn^{1-\alpha}} \\
&\quad = \big(\|Z_n - m - \gamma_n \Phi(Z_n)\|^2 + \gamma_n^2 \|\xi_{n+1}\|^2 \\
&\qquad + 2\gamma_n \langle Z_n - m - \gamma_n \Phi(Z_n), \xi_{n+1}\rangle\big) \mathbb{1}_{\|Z_n - m\| < cn^{1-\alpha}} \\
&\quad \le \left(\left(1 - \frac{1}{n}\right)\|Z_n - m\|^2 + 4\gamma_n^2 \right. \\
&\qquad \left. + 2\gamma_n \langle Z_n - m - \gamma_n \Phi(Z_n), \xi_{n+1}\rangle\right) \mathbb{1}_{\|Z_n - m\| < cn^{1-\alpha}}.
\end{aligned}$$

Moreover, since (ξ_{n+1}) is a sequence of martingale differences for the filtration $(\mathcal{F}_n)$,

$$\mathbb{E}\big[\langle Z_n - m - \gamma_n \Phi(Z_n), \xi_{n+1}\rangle \mathbb{1}_{\|Z_n - m\| \le cn^{1-\alpha}} | \mathcal{F}_n\big] = 0,$$

$$\mathbb{E}\big[\langle Z_n - m - \gamma_n \Phi(Z_n), \xi_{n+1}\rangle \|Z_n - m\|^2 \mathbb{1}_{\|Z_n - m\| \le cn^{1-\alpha}} | \mathcal{F}_n\big] = 0.$$

Applying Cauchy–Schwarz's inequality,

$$\begin{aligned}
&\mathbb{E}\big[\|Z_{n+1} - m\|^4 \mathbb{1}_{\|Z_n - m\| \le cn^{1-\alpha}}\big] \\
&\quad \le \left(1 - \frac{1}{n}\right)^2 \mathbb{E}\big[\|Z_n - m\|^4 \mathbb{1}_{\|Z_n - m\| \le cn^{1-\alpha}}\big] + 16\gamma_n^4 \\
&\qquad + 8\gamma_n^2 \left(1 - \frac{1}{n}\right) \mathbb{E}\big[\|Z_n - m\|^2 \mathbb{1}_{\|Z_n - m\| \le cn^{1-\alpha}}\big] \\
&\qquad + 4\gamma_n^2 \mathbb{E}\big[\langle Z_n - m - \gamma_n \Phi(Z_n), \xi_{n+1}\rangle^2 \mathbb{1}_{\|Z_n - m\| \le cn^{1-\alpha}}\big] \\
&\quad \le \left(1 - \frac{1}{n}\right)^2 \mathbb{E}\big[\|Z_n - m\|^4\big] + 16\gamma_n^4 + 8\gamma_n^2 \mathbb{E}\big[\|Z_n - m\|^2\big] \\
&\qquad + 4\gamma_n^2 \mathbb{E}\big[\|Z_n - m - \gamma_n \Phi(Z_n)\|^2 \mathbb{E}\big[\|\xi_{n+1}\|^2 | \mathcal{F}_n\big] \mathbb{1}_{\|Z_n - m\| \le cn^{1-\alpha}}\big].
\end{aligned}$$

Finally, since $\mathbb{E}[\|\xi_{n+1}\|^2 | \mathcal{F}_n] \le 1$, applying Lemma 5.3 we get

$$\begin{aligned}
&\mathbb{E}\big[\|Z_{n+1} - m\|^4 \mathbb{1}_{\|Z_n - m\| \le cn^{1-\alpha}}\big] \\
&\quad \le \left(1 - \frac{1}{n}\right)^2 \mathbb{E}\big[\|Z_n - m\|^4\big] + 16\gamma_n^4 \\
&\qquad + 8\gamma_n^2 \mathbb{E}\big[\|Z_n - m\|^2\big] + 4\gamma_n^2 \left(1 - \frac{1}{n}\right) \mathbb{E}\big[\|Z_n - m\|^2\big] \\
&\quad \le \left(1 - \frac{1}{n}\right)^2 \mathbb{E}\big[\|Z_n - m\|^4\big] + 16\gamma_n^4 + 12\gamma_n^2 \mathbb{E}\big[\|Z_n - m\|^2\big].
\end{aligned}$$

Since $\gamma_n^4 = o(\frac{1}{n^{3\alpha}})$, there are positive constants C_1', C_2' such that for all $n \geq n_\alpha$,

$$\begin{aligned}
&\mathbb{E}[\|Z_{n+1} - m\|^4] \\
&\quad = \mathbb{E}[\|Z_{n+1} - m\|^4 \mathbb{1}_{\|Z_n - m\| \geq cn^{1-\alpha}}] + \mathbb{E}[\|Z_{n+1} - m\|^4 \mathbb{1}_{\|Z_n - m\| \leq cn^{1-\alpha}}] \\
&\quad \leq \frac{2^{4-4\alpha} C_\alpha'^4 C_\alpha}{n^{3\alpha}} + \left(1 - \frac{1}{n}\right)^2 \mathbb{E}[\|Z_n - m\|^4] + 16\gamma_n^4 + 12\gamma_n^2 \mathbb{E}[\|Z_n - m\|^2] \\
&\quad \leq \left(1 - \frac{1}{n}\right)^2 \mathbb{E}[\|Z_n - m\|^4] + C_1' \frac{1}{n^{3\alpha}} + C_2' \frac{1}{n^{2\alpha}} \mathbb{E}[\|Z_n - m\|^2]. \qquad \square
\end{aligned}$$

PROOF OF THEOREM 3.1. Let $\beta \in (\alpha, 3\alpha - 1)$, there is a rank $n_\beta \geq n_\alpha$ (n_α is defined in Lemma 3.2) such that for all $n \geq n_\beta$ we have $(1 - \frac{1}{n})^2 (\frac{n+1}{n})^\beta + (C_1' + C_2') 2^{3\alpha} \frac{1}{(n+1)^{3\alpha - \beta}} \leq 1$ (C_1', C_2' are defined in Lemma 3.2). Indeed, since $\beta < 3\alpha - 1 < 2$,

$$\begin{aligned}
&\left(1 - \frac{1}{n}\right)^2 \left(\frac{n+1}{n}\right)^\beta + (C_1' + C_2') 2^{3\alpha} \frac{1}{(n+1)^{3\alpha - \beta}} \\
&\qquad = 1 - (2 - \beta) \frac{1}{n} + o\left(\frac{1}{n}\right).
\end{aligned}$$

We now prove by induction that there are positive constants C', C'' such that $2C' \geq C'' \geq C' \geq 1$ and such that for all $n \geq n_\beta$,

$$\begin{aligned}
\mathbb{E}[\|Z_n - m\|^2] &\leq \frac{C'}{n^\alpha}, \\
\mathbb{E}[\|Z_n - m\|^4] &\leq \frac{C''}{n^\beta}.
\end{aligned}$$

Let us choose $C' \geq n_\beta \mathbb{E}[\|Z_{n_\beta} - m\|^2]$ and $C'' \geq n_\beta \mathbb{E}[\|Z_{n_\beta} - m\|^4]$. This is possible since there is a positive constant M such that for all $n \geq 1$, $\sup\{\mathbb{E}[\|Z_n - m\|^2], \mathbb{E}[\|Z_n - m\|^4]\} \leq M$. Let $n \geq n_\beta$, using Lemma 3.2 and by induction,

$$\begin{aligned}
\mathbb{E}[\|Z_{n+1} - m\|^4] &\leq \left(1 - \frac{1}{n}\right)^2 \mathbb{E}[\|Z_n - m\|^4] + \frac{C_1'}{n^{3\alpha}} + \frac{C_2'}{n^{2\alpha}} \mathbb{E}[\|Z_n - m\|^2] \\
&\leq \left(1 - \frac{1}{n}\right)^2 \frac{C''}{n^\beta} + \frac{C_1'}{n^{3\alpha}} + \frac{C_2' C'}{n^{3\alpha}}.
\end{aligned}$$

Moreover, since $C' \leq C''$ and since $C'' \geq 1$,

$$\mathbb{E}[\|Z_{n+1} - m\|^4] \leq \left(1 - \frac{1}{n}\right)^2 \frac{C''}{n^\beta} + \frac{C_1' C''}{n^{3\alpha}} + \frac{C_2' C''}{n^{3\alpha}}.$$

Factorizing by $\frac{C''}{(n+1)^{\beta}}$, we get

$$\begin{aligned}
&\mathbb{E}\big[\|Z_{n+1}-m\|^4\big] \\
&\quad\le \left(1-\frac{1}{n}\right)^2\left(1+\frac{1}{n}\right)^{\beta}\frac{C''}{(n+1)^{\beta}} \\
&\qquad + (C_1'+C_2')\left(1+\frac{1}{n}\right)^{3\alpha}\frac{1}{(n+1)^{3\alpha-\beta}}\frac{C''}{(n+1)^{\beta}} \\
&\quad\le \left(\left(1-\frac{1}{n}\right)^2\left(1+\frac{1}{n}\right)^{\beta} + (C_1'+C_2')2^{3\alpha}\frac{1}{(n+1)^{3\alpha-\beta}}\right)\frac{C''}{(n+1)^{\beta}}.
\end{aligned}$$

By definition of n_{β},

$$\mathbb{E}\big[\|Z_{n+1}-m\|^4\big] \le \frac{C''}{(n+1)^{\beta}}. \tag{5.15}$$

We now prove that $\mathbb{E}[\|Z_{n+1}-m\|^2] \le \frac{C'}{(n+1)^{\alpha}}$. Since $C'' \le 2C'$, by Lemma 3.1 and by induction, there is a constant $C''' > 0$ such that

$$\begin{aligned}
\mathbb{E}\big[\|Z_{n+1}-m\|^2\big] &\le \frac{C'''}{(n+1)^{\alpha}} + C_3 \sup_{n/2+1\le k\le n+1}\mathbb{E}\big[\|Z_k-m\|^4\big] \\
&\le \frac{C'''}{(n+1)^{\alpha}} + 2^{\beta+1}C_3\frac{1}{(n+1)^{\beta-\alpha}}\frac{C'}{(n+1)^{\alpha}}.
\end{aligned}$$

To get $\mathbb{E}[\|Z_{n+1}-m\|^2] \le \frac{C'}{(n+1)^{\alpha}}$, we choose $C' \ge C''' + 2^{\beta+1}C_3\frac{1}{(n+1)^{\beta-\alpha}}$, which concludes the induction. The proof is complete for all $n \ge 1$ by taking $C' \ge \max_{n\le n_{\beta}}\{n^{\alpha}\mathbb{E}[\|Z_n-m\|^2]\}$ and $C'' \ge \max_{n\le n_{\beta}}\{n^{\beta}\mathbb{E}[\|Z_n-m\|^4]\}$. $\square$

5.2. *Proof of Theorem* 4.2. Let us recall the decomposition of the averaged algorithm

$$\begin{aligned}
\Gamma_m(\overline{Z}_n-m) = \frac{Z_1-m}{n\gamma_1} - \frac{Z_{n+1}-m}{n\gamma_n} + \sum_{k=2}^{n}\left(\frac{1}{\gamma_k}-\frac{1}{\gamma_{k-1}}\right)(Z_k-m) \\
- \frac{1}{n}\sum_{k=1}^{n}\delta_k + \frac{1}{n}\sum_{k=1}^{n}\xi_{k+1}.
\end{aligned}$$

We now bound each term on the right-hand side of previous inequality. Note that since $\mathbb{E}[\|Z_n-m\|^2] \le \frac{C'}{n^{\alpha}}$, applying Cauchy–Schwarz's inequality, we have $\mathbb{E}[\|Z_n-m\|] \le \sqrt{\frac{C'}{n^{\alpha}}}$. Then

$$\mathbb{E}\left[\left\|\frac{Z_{n+1}-m}{n\gamma_n}\right\|^2\right] \le \frac{n^{2\alpha}}{c_{\gamma}n^2}\mathbb{E}\big[\|Z_{n+1}-m\|^2\big] \le \frac{2^{\alpha}C'}{c_{\gamma}}\frac{1}{n^{2-\alpha}}.$$

Since $\alpha < 1$, remark that $\frac{2-\alpha}{2} > \frac{1}{2}$. Moreover, since $\gamma_k^{-1} - \gamma_{k-1}^{-1} \leq 2\alpha c_\gamma^{-1} k^{\alpha-1}$, there is a positive constant C_1 such that

$$\begin{aligned}\mathbb{E}\left[\left\|\frac{1}{n}\sum_{k=2}^{n}(Z_k - m)\left(\gamma_k^{-1} - \gamma_{k-1}^{-1}\right)\right\|\right] &\leq \frac{2\alpha c_\gamma^{-1}}{n}\sum_{k=2}^{n}\mathbb{E}\left[\|Z_k - m\|\right]k^{\alpha-1} \\ &\leq \frac{2\alpha c_\gamma^{-1}\sqrt{C'}}{n}\sum_{k=2}^{n/2-1} k^{\alpha/2-1} \\ &\leq \frac{C_1}{n^{1-\alpha/2}}.\end{aligned}$$

Note also that since $\alpha < 1$, we have $1 - \alpha/2 \geq 1/2$. Moreover, since $\|\delta_n\| \leq C_m\|T_n\|^2$, there is a positive constant C_2 such that

$$\mathbb{E}\left[\left\|\frac{1}{n}\sum_{k=1}^{n}\delta_k\right\|\right] \leq \frac{C_m}{n}\sum_{k=1}^{n}\mathbb{E}\left[\|Z_k - m\|^2\right] \leq \frac{C_m C'}{n}\sum_{k=1}^{n}k^{-\alpha} \leq C_2\frac{1}{n^\alpha}.$$

Finally, there is a positive constant C_3 such that $\mathbb{E}[\|\frac{Z_1-m}{\gamma_1 n}\|] \leq \frac{C_3}{n}$.

We now study the martingale term. Let M be a constant and (σ_n) be a sequence of positive real numbers defined by $M := 2 \geq \sup_i \|\xi_i\|$ and $\sigma_n^2 := n \geq \sum_{k=1}^{n}\mathbb{E}[\|\xi_k\|^2|\mathcal{F}_{k-1}]$. Applying Pinelis–Bernstein's lemma, we have for all $t > 0$,

$$\mathbb{P}\left(\sup_{1\leq k\leq n}\left\|\sum_{j=1}^{k}\xi_{j+1}\right\| \geq t\right) \leq 2\exp\left[-\frac{t^2}{2(\sigma_n^2 + Mt/3)}\right].$$

Consequently,

$$\begin{aligned}\mathbb{P}\left(\frac{\|\sum_{k=1}^{n}\xi_{k+1}\|}{n} \geq t\right) &\leq \mathbb{P}\left(\sup_{1\leq k\leq n}\left\|\sum_{j=1}^{k}\xi_{j+1}\right\| \geq tn\right) \\ &\leq 2\exp\left[-\frac{t^2}{2(\sigma_n'^2 + N_n' t/3)}\right],\end{aligned}$$

with $\sigma_n'^2 := 1/n$ and $N_n' := 2/n$. As in the proof of Theorem 4.1, there are positive constants C_1', C_2', C_3' such that for all $t > 0$,

$$\begin{aligned}&\mathbb{P}\left[\|\Gamma_m(\overline{Z}_n - m)\| \geq t\right] \\ &\quad \leq 2\exp\left[-\frac{(t/2)^2}{2(\sigma_n'^2 + N_n' t/6)}\right] + \frac{C_1'}{n^{1-\alpha/2}} + \frac{C_2'}{n^\alpha} + \frac{C_3'}{n} \\ &\quad =: g(t,n).\end{aligned}$$

We search values of t such that $g(t,n) \leq \delta$ and we must solve the following system of inequalities:

$$2\exp\left[-\frac{(t/2)^2}{2(\sigma_n'^2 + N_n t/6)}\right] \leq \delta/2, \qquad \frac{C_1'}{tn^{1-\alpha/2}} \leq \delta/6,$$
$$\frac{C_2'}{tn^{\alpha}} \leq \delta/6, \qquad \frac{C_3'}{tn} \leq \delta/6.$$

We get that t must satisfy (see [26], Appendix A, for the martingale term):

$$t \geq 4\left(\frac{N_n'}{3} + \sigma_n'\right)\ln\left(\frac{4}{\delta}\right), \qquad t \geq \frac{6C_1'}{\delta}\frac{1}{n^{1-\alpha/2}},$$
$$t \geq \frac{6C_2'}{\delta}\frac{1}{n^{\alpha}}, \qquad t \geq \frac{6C_3'}{\delta}\frac{1}{n}.$$

Since $(\frac{N_n'}{3} + \sigma_n') = \frac{2}{3n} + \frac{1}{\sqrt{n}}$, the other terms are negligible for n large enough and we can choose

$$n_\delta := \max\left\{\left(\frac{6C_1'}{\delta\ln(\frac{4}{\delta})}\right)^{\frac{1}{1/2-\alpha/2}}, \left(\frac{6C_2'}{\delta\ln(\frac{4}{\delta})}\right)^{\frac{1}{\alpha-1/2}}, \left(\frac{6C_3'}{\delta\ln(\frac{4}{\delta})}\right)^{\frac{1}{2}}\right\}. \tag{5.16}$$

Acknowledgments. We thank the anonymous reviewers for helpful and constructive comments on an earlier draft.

SUPPLEMENTARY MATERIAL

Supplement to "Online estimation of the geometric median in Hilbert spaces: Nonasymptotic confidence balls" (DOI: 10.1214/16-AOS1460SUPP; .pdf). We provide the proofs of some technical ancillary lemmas and propositions.

REFERENCES

[1] ARNAUDON, M., DOMBRY, C., PHAN, A. and YANG, L. (2012). Stochastic algorithms for computing means of probability measures. *Stochastic Process. Appl.* **122** 1437–1455. MR2914758

[2] BACH, F. (2014). Adaptivity of averaged stochastic gradient descent to local strong convexity for logistic regression. *J. Mach. Learn. Res.* **15** 595–627. MR3190851

[3] BACH, F. and MOULINES, E. (2013). Non-strongly-convex smooth stochastic approximation with convergence rate $o(1/n)$. In *Advances in Neural Information Processing Systems* 773–781. Curran Associates, Inc., Red Hook, NY.

[4] BALI, J. L., BOENTE, G., TYLER, D. E. and WANG, J.-L. (2011). Robust functional principal components: A projection-pursuit approach. *Ann. Statist.* **39** 2852–2882. MR3012394

[5] BALSUBRAMANI, A., DASGUPTA, S. and FREUND, Y. (2013). The fast convergence of incremental PCA. In *Advances in Neural Information Processing Systems* 3174–3182. Curran Associates, Inc., Red Hook, NY.

[6] BECK, A. and SABACH, S. (2015). Weiszfeld's method: Old and new results. *J. Optim. Theory Appl.* **164** 1–40. MR3296283

[7] CARDOT, H., CÉNAC, P. and GODICHON-BAGGIONI, A. (2016). Supplement to "Online estimation of the geometric median in Hilbert spaces: Nonasymptotic confidence balls." DOI:10.1214/16-AOS1460SUPP.

[8] CARDOT, H., CÉNAC, P. and ZITT, P.-A. (2013). Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli* **19** 18–43. MR3019484

[9] CHAKRABORTY, A. and CHAUDHURI, P. (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *Ann. Statist.* **42** 1203–1231. MR3224286

[10] CHAUDHURI, P. (1992). Multivariate location estimation using extension of R-estimates through U-statistics type approach. *Ann. Statist.* **20** 897–916. MR1165598

[11] FLETCHER, P. T., VENKATASUBRAMANIAN, S. and JOSHI, S. (2009). The geometric median on Riemannian manifolds with application to robust atlas estimation. *NeuroImage* **45** S143–S152.

[12] GERVINI, D. (2008). Robust functional estimation using the median and spherical principal components. *Biometrika* **95** 587–600. MR2443177

[13] HALDANE, J. B. S. (1948). Note on the median of a multivariate distribution. *Biometrika* **35** 414–417.

[14] KEMPERMAN, J. H. B. (1987). The median of a finite measure on a Banach space. In *Statistical Data Analysis Based on the L_1-Norm and Related Methods* (*Neuchâtel*, 1987) 217–230. North-Holland, Amsterdam. MR0949228

[15] KRAUS, D. and PANARETOS, V. M. (2012). Dispersion operators and resistant second-order functional data analysis. *Biometrika* **99** 813–832.

[16] KUHN, H. W. (1973). A note on Fermat's problem. *Math. Program.* **4** 98–107. MR0316102

[17] LOCANTORE, N., MARRON, J. S., SIMPSON, D. G., TRIPOLI, N., ZHANG, J. T. and COHEN, K. L. (1999). Robust principal component analysis for functional data. *TEST* **8** 1–73. MR1707596

[18] MINSKER, S. (2015). Geometric median and robust estimation in Banach spaces. *Bernoulli* **21** 2308–2335. MR3378468

[19] MINSKER, S., SRIVASTAVA, S., LIN, L. and DUNSON, D. (2014). Scalable and robust Bayesian inference via the median posterior. In *Proceedings of the* 31*st International Conference on Machine Learning* (*ICML*-14) 1656–1664.

[20] MÖTTÖNEN, J., NORDHAUSEN, K. and OJA, H. (2010). Asymptotic theory of the spatial median. In *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in Honor of Professor Jana Jurečková. Inst. Math. Stat. Collect.* **7** 182–193. IMS, Beachwood, OH. MR2808378

[21] PELLETIER, M. (2000). Asymptotic almost sure efficiency of averaged stochastic algorithms. *SIAM J. Control Optim.* **39** 49–72 (electronic). MR1780908

[22] PINELIS, I. (1994). Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.* **22** 1679–1706. MR1331198

[23] POLYAK, B. T. and JUDITSKY, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.* **30** 838–855. MR1167814

[24] RUDELSON, M. and VERSHYNIN, R. (2010). Non-asymptotic theory of random matrices: Extreme singular values. In *Proceedings of the International Congress of Mathematicians. Volume III* 1576–1602. Hindustan Book Agency, New Delhi. MR2827856

[25] SMALL, C. G. (1990). A survey of multidimensional medians. *Int. Stat. Rev.* **58** 263–277.

[26] TARRÈS, P. and YAO, Y. (2014). Online learning as stochastic approximation of regularization paths: Optimality and almost-sure convergence. *IEEE Trans. Inform. Theory* **60** 5716–5735. MR3252416

[27] VARDI, Y. and ZHANG, C.-H. (2000). The multivariate L_1-median and associated data depth. *Proc. Natl. Acad. Sci. USA* **97** 1423–1426 (electronic). MR1740461

[28] WEISZFELD, E. (1937). Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tohoku Math. J.* (2) **43** 355–386.

[29] WOODROOFE, M. (1972). Normal approximation and large deviations for the Robbins–Monro process. *Z. Wahrsch. Verw. Gebiete* **21** 329–338. MR0314098

H. CARDOT
P. CÉNAC
A. GODICHON-BAGGIONI
INSTITUT DE MATHÉMATIQUES DE BOURGOGNE
UNIVERSITÉ DE BOURGOGNE FRANCHE COMTÉ
UMR CNRS 5584
9 AV. A. SAVARY
21000 DIJON
FRANCE
E-MAIL: herve.cardot@u-bourgogne.fr
peggy.cenac@u-bourgogne.fr
antoine.godichon@u-bourgogne.fr