INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

The restricted isometry property meets nonlinear approximation with redundant frames

Rémi Gribonval — Morten Nielsen

N° 7548

February 2011

Audio, Speech, and Language Processing

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--7548--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
RENNES - BRETAGNE ATLANTIQUE

The restricted isometry property meets nonlinear approximation with redundant frames

Rémi Gribonval * , Morten Nielsen †

Theme : Audio, Speech, and Language Processing
Équipe-Projet METISS

Rapport de recherche n° 7548 — February 2011 — 19 pages

Abstract: It is now well known that sparse or compressible vectors can be stably recovered from their low-dimensional projection, provided the projection matrix satisfies a Restricted Isometry Property (RIP). We establish new implications of the RIP with respect to nonlinear approximation in a Hilbert space with a redundant frame. The main ingredients of our approach are: a) Jackson and Bernstein inequalities, associated to the characterization of certain approximation spaces with interpolation spaces; b) a new proof that for overcomplete frames which satisfy a Bernstein inequality, these interpolation spaces are nothing but the collection of vectors admitting a representation in the dictionary with compressible coefficients; c) the proof that the RIP implies Bernstein inequalities. As a result, we obtain that in most overcomplete random Gaussian dictionaries with fixed aspect ratio, just as in any orthonormal basis, the error of best m-term approximation of a vector decays at a certain rate if, and only if, the vector admits a compressible expansion in the dictionary. Yet, for mildly overcomplete dictionaries with a one-dimensional kernel, we give examples where the Bernstein inequality holds, but the same inequality fails for even the smallest perturbation of the dictionary.

Key-words: Bernstein inequality, random dictionaries, restricted isometry condition

This work was supported in part by the European Union through the project SMALL (Sparse Models, Algorithms and Learning for Large-Scale data). The project SMALL acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 225913.

* Rémi Gribonval is with INRIA, Centre Inria Rennes - Bretagne Atlantique, Campus de Beaulieu, F-35042 Rennes Cedex, Rennes, France. Phone: +33 2 99 84 25 06. Fax: +33 2 99 84 71 71. Email: remi.gribonval@inria.fr.

† Morten Nielsen is with the Department of Mathematical Sciences, Aalborg University, Frederik Bajersvej 7G, DK - 9220 Aalborg East, Denmark. Email: mnielsen@math.aau.dk

Centre de recherche INRIA Rennes – Bretagne Atlantique
IRISA, Campus universitaire de Beaulieu, 35042 Rennes Cedex
Téléphone : +33 2 99 84 71 00 — Télécopie : +33 2 99 84 71 71

Sur la propriété d'isométrie restreinte et l'approximation non-linéaire avec un dictionnaire redondant

Résumé : Il est maintenant bien établi que les vecteurs parcimonieux ou compressibles peuvent être estimés de façon stable à partir de leur projection en petite dimension, dès que la matrice de projection satisfait une Propriété d'Isométrie Restreinte (RIP). Nous établissons de nouvelles conséquences de la RIP vis à vis de l'approximation non-linéaire dans un espace de Hilbert avec un repére oblique (ou *frame*) redondant. Les principaux ingrédients de notre approche sont : a) des inégalités de Jackson et de Bernstein, associées à des caractérisations de certains espaces d'approximation en termes d'espaces d'interpolation ; b) la preuve que pour des repères obliques satisfaisant ladite inégalité de Bernstein, les espaces d'interpolation en question ne sont rien d'autres que l'ensemble des vecteurs ayant une représentation à coefficients compressibles dans le dictionnaire ; c) la preuve que la RIP implique des inégalités de Bernstein. Une conséquence de ces résultats est que la plupart des dictionnaires Gaussiens aléatoires de facteur de redondance fixé se comportent comme une base orthogonale du point de vue des espaces d'approximation : l'erreur optimale d'approximation à m termes d'un vecteur décroît à une certaine vitesse en fonction de m si, et seulement si, le vecteur admet une représentation compressible dans le dictionnaire. Toutefois, nous montrons qu'il existe aussi des dictionnaires de redondance minimale, dont le noyau est de dimension un, pour lesquels l'inégalité de Bernstein, bien que vérifiée, peut être mise en défaut lorsque le dictionnaire subit une perturbation arbitrairement petite.

Mots-clés : Inégalité de Bernstein, dictionnaire aléatoire, condition d'isométrie restreinte

1 Introduction

Data approximation using sparse linear expansions from overcomplete dictionaries has become a central theme in signal and image processing with applications ranging from data acquisition (compressed sensing) to denoising and compression. Dictionaries can be seen as collections of vectors $\{\varphi_j\}$ from a Banach space X equipped with a norm $\|\cdot\|_X$, and one wishes to approximate data vectors f using k-term expansions $\sum_{j\in I} c_j\varphi_j$ where I is an index set of size k. Formally, using the matrix notation $\mathbf{\Phi}c = \sum_j c_j\varphi_j$ and denoting $\|c\|_0 = \sharp\{j, c_j \neq 0\}$ the number of nonzero components in the vector c , we can define the (nonlinear) set of all such k-term expansions

$$\Sigma_k(\mathbf{\Phi}) := \{\mathbf{\Phi}c, \quad \|c\|_0 \leq k\}.$$

1.1 Best k-term approximation

A first question we may want to answer, for each data vector f, is: *how well can we approximate it using elements of* $\Sigma_k(\mathbf{\Phi})$? The error of best k-term approximation is a quantitative answer for a fixed k:

$$\sigma_k(f,\mathbf{\Phi}) := \inf_{y\in\Sigma_k(\mathbf{\Phi})} \|f-y\|_X.$$

A more global view is given by the largest approximation rate $s > 0$ such that[1]

$$\sigma_k(f,\mathbf{\Phi}) \lesssim k^{-s}, \quad \forall k \geq 1.$$

To measure more finely the rate of approximation, one defines for $0 < q < \infty$ [8, Chapter 7, Section 9]

$$|f|_{\mathcal{A}^s_q(\mathbf{\Phi})} := \left(\sum_{k\geq 1} [k^s\sigma_k(f,\mathbf{\Phi})]^q \, k^{-1}\right)^{1/q} \asymp \left(\sum_{j\geq 0} \left[2^{js}\sigma_{2^j}(f,\mathbf{\Phi})\right]^q\right)^{1/q}. \tag{1.1}$$

and the associated *approximation spaces*

$$\mathcal{A}^s_q(\mathbf{\Phi}) := \{f \in \mathcal{H}, |f|_{\mathcal{A}^s_q(\mathbf{\Phi})} < \infty\} \tag{1.2}$$

1.2 Sparse or compressible representations

Alternatively, we may be interested in *sparse / compressible representations* of f in the dictionary. Suppose the vectors forming $\mathbf{\Phi}$ are quasi-normalized in X: for all j, $0 < c \leq \|\varphi_j\|_X \leq C < \infty$. Then using ℓ^τ (quasi)-norms (in particular, $0 < \tau \leq 1$) one defines[2]

$$\|f\|_{\ell^\tau(\mathbf{\Phi})} := \inf_{c|\mathbf{\Phi}c=f} \|c\|_\tau \tag{1.3}$$

and the associated *sparsity spaces* (also called *smoothness spaces*, for when $\mathbf{\Phi}$ is, e.g., a wavelet frame, they indeed characterize smoothness on the Besov scale)

$$\ell^\tau(\mathbf{\Phi}) := \{f, \|f\|_{\ell^\tau(\mathbf{\Phi})} < \infty\}.$$

[1] The notation $a \lesssim b$ indicates the existence of a finite constant C such that $a \leq C \cdot b$. The notation $a \asymp b$ means that we have both $a \lesssim b$ and $b \lesssim a$. As usual, C will denote a generic finite constant, independent from the other quantities of interest. Different occurences of this notation in the paper may correspond to different values of the constant.

[2] It has been shown in [11] that under mild assumptions on the dictionary, such as Eq. (1.4), the definition (1.3) is fully equivalent to the more general topological definition of $\|f\|_{\ell^\tau(\mathbf{\Phi})}$ introduced in [8].

1.3 Direct and inverse estimates

Interestingly, the above defined concepts are related. In a Hilbert space $X = \mathcal{H}$, when $\mathbf{\Phi}$ satisfies the upper bound

$$\|\mathbf{\Phi}c\|_{\mathcal{H}}^2 \leq B \cdot \|c\|_2^2, \quad \forall c \in \ell^2, \tag{1.4}$$

the sparsity spaces for $0 < \tau < 2$ are characterized as

$$\ell^\tau(\mathbf{\Phi}) = \{f, \exists c, \|c\|_{\ell^\tau} < \infty, f = \mathbf{\Phi}c\} = \mathbf{\Phi}\ell^\tau,$$

and for any $s > 0$ we have the so-called *Jackson inequality*

$$\sigma_k(f, \mathbf{\Phi}) \leq C_\tau(B) \cdot \|f\|_{\ell^\tau(\mathbf{\Phi})} \cdot k^{-s}, \quad s = \frac{1}{\tau} - \frac{1}{2}, \quad \forall f \in \ell^\tau(\mathbf{\Phi}), \forall k \in \mathbb{N} \tag{1.5}$$

where, as indicated by the notation, the constant $C_\tau(B)$ only depends on τ and the upper bound B in (1.4). Note that the upper bound (1.4) holds true whenever the dictionary is a *frame*: B is then called the *upper frame bound*, and we will use this terminology.

When $\mathbf{\Phi}$ is an orthogonal basis, a converse result is true: if $\sigma_k(f, \mathbf{\Phi})$ decays as k^{-s} then $\|f\|_{\ell^\tau_w(\mathbf{\Phi})} < \infty$, where ℓ^τ_w is a weak ℓ^τ space [8] and $s = 1/\tau - 1/2$. More generally, inverse estimates are related to a *Bernstein inequality* [7, 6].

$$\|f_k\|_{\ell^\tau(\mathbf{\Phi})} \leq C \cdot k^r \cdot \|f_k\|_{\mathcal{H}}, \quad \forall f_k \in \Sigma_k(\mathbf{\Phi}), \forall k. \tag{1.6}$$

The inequality (1.6) is related to the so-called Bernstein-Nikolsky inequality, we refer the reader to [1, 6] for more information.

1.4 When approximation spaces are sparsity spaces

When a Jackson inequality holds together with a Bernstein inequality with matching exponent $r = 1/\tau - 1/2$, it is possible to characterize (with equivalent (quasi)-norms) the approximation spaces $\mathcal{A}^s_q(\mathbf{\Phi})$ as real interpolation spaces [6, Chapter 7] between $\mathcal{H}$, denoted $(\mathcal{H}, \ell^\tau(\mathbf{\Phi}))_{\theta,q}$, where $s = \theta r$, $0 < \theta < 1$. The definition of real interpolation spaces will be recalled in Section 2. Let us just mention here that it is based on decay properties of the *K-functional*

$$K(f, t; \mathcal{H}, \ell^p(\mathbf{\Phi})) = \inf_{c \in \ell^p} \left\{ \|f - \mathbf{\Phi}c\|_{\mathcal{H}} + t\|c\|_p \right\}. \tag{1.7}$$

A priori, without a more explicit description of real interpolation spaces, the characterization of approximation spaces as interpolation spaces may seem just a sterile pedantic rewriting. Fortunately, we show in Section 2 (*Theorem 2.1*) that the Bernstein inequality (1.6), together with the upper frame bound (1.4), allows to directly identify approximation spaces with sparsity spaces, with equivalent (quasi)-norms, for certain ranges of parameters. In particular, the following result can be obtained as a consequence of Theorem 2.1.

Theorem 1.1. *Suppose that $\mathbf{\Phi}$ satisfies the upper frame bound* (1.4) *with constant B as well as the Bernstein inequality* (1.6) *with some $0 < \tau \leq 1$, with exponent $r = 1/\tau - 1/2$ and constant C. Then we have*

$$\mathcal{A}^r_\tau(\mathbf{\Phi}) = \ell^\tau(\mathbf{\Phi}) \tag{1.8}$$

with equivalent norms, i.e.

$$c_1(B, C) \cdot \|f\|_{\ell^\tau(\mathbf{\Phi})} \leq \|f\|_{\mathcal{A}^r_\tau(\mathbf{\Phi})} := \|f\|_{\mathcal{H}} + |f|_{\mathcal{A}^r_\tau(\mathbf{\Phi})} \leq c_2(B, C) \cdot \|f\|_{\ell^\tau(\mathbf{\Phi})}$$

where the constants only depend on B and C.

In other words, under the assumptions of Theorem 1.1, a data vector $f \in \mathcal{H}$ can be approximated by k-term expansions at rate k^{-r} (in the sense $f \in \mathcal{A}^r_\tau(\mathbf{\Phi})$, where $r = 1/\tau - 1/2$) if, *and only if*, f admits a sparse *representation* $f = \sum_j c_j \varphi_j$ with $\sum_j |c_j|^\tau < \infty$.

1.5 Ideal *vs* practical approximation algorithms

Consider a function f that can be approximated at rate k^{-r} using k-term expansions from $\mathbf{\Phi}$: $\sigma_k(f, \mathbf{\Phi}) \lesssim k^{-r}, \forall k \geq 1$. Under the assumptions of the above Theorem, we can conclude that the function f indeed admits a *representation* $f = \sum_j c_j \varphi_j$ with $\sum_j |c_j|^\tau < \infty$. Suppose that we know how to compute such a representation (e.g., that we can solve the optimization problem $\min \|c\|_\tau$ subject to $f = \mathbf{\Phi} c$). Then, sorting the coefficients in decreasing order of magnitude $|c_{j_m}| \geq |c_{j_{m+1}}|$, one can build a simple sequence of k-term approximants $f_m := \sum_{m=1}^k c_{j_m} \varphi_{j_m}$ which converge to f at the rate r: $\|f - f_k\|_{\mathcal{H}} \lesssim k^{-r}$. Note that one may not however be able to guarantee that $\|f - f_k\|_{\mathcal{H}} \leq C \sigma_k(f, \mathbf{\Phi})$ for a fixed constant $C < \infty$.

A special case of interest is $\tau = 1$, where the optimization problem

$$\min \|c\|_1 \text{ subject to } f = \mathbf{\Phi} c$$

is convex, and the unit ball in $\ell^1(\mathbf{\Phi})$ is simply the convex hull of the symmetrized dictionary $\{\pm \varphi_j\}_j$ with φ_j the atoms of the dictionary $\mathbf{\Phi}$. Therefore, under the assumptions of the above Theorem for $\tau = 1$, if a function can be approximated at rate $k^{-1/2}$ then, after proper rescaling, it belongs to the convex hull of the symmetrized dictionary, and there exists constructive algorithms such as Orthonormal Matching Pursuit [13, 14] which are guaranteed to provide the rate of approximation $k^{-1/2}$ [8, Theorem 3.7].

1.6 Null Space Properties and fragility of Bernstein inequalities

On the one hand, it is known [11] that Jackson inequalities are always satisfied provided that the dictionary is a *frame*, i.e.,

$$A\|f\|^2_{\mathcal{H}} \leq \|\mathbf{\Phi}^T f\|^2_2 \leq B\|f\|^2_{\mathcal{H}}, \quad \forall f \in \mathcal{H}. \tag{1.9}$$

The upper bound B is actually equivalent to the upper frame bound (1.4) and therefore sufficient for a Jackson inequality to hold.

On the other hand, Bernstein inequalities are known to be much more subtle and seemingly fragile: they may be satisfied for certain structured dictionaries, but not for arbitrarily small perturbations thereof [10].

In Section 3, for the sake of simplicity we restrict our attention to the case $\tau = 1$ when the dictionary $\mathbf{\Phi}$ forms a frame for a general Hilbert space $\mathcal{H}$. We show that the Bernstein inequality for $\ell^1(\mathbf{\Phi})$,

$$\|\mathbf{\Phi} c\|_{\ell^1(\mathbf{\Phi})} \leq C m^{1/2} \|\mathbf{\Phi} c\|_{\mathcal{H}}, \qquad c : \|c\|_0 \leq m, \forall m \geq 1, \tag{1.10}$$

is closely linked to *properties of the kernel of* $\mathbf{\Phi}$ given by

$$N(\mathbf{\Phi}) := \{z \in \ell^2 : \mathbf{\Phi} z = 0\}.$$

The seemingly simple case where we have a one dimensional null space for the dictionary, $N(\mathbf{\Phi}) = \text{span}\{z\}$ for some fixed sequence z, is particularly useful to demonstrate the fragility of the Bernstein estimates as the following example shows.

Example 1.2. Given any infinite dictionary $\mathbf{\Phi}$ with $N(\mathbf{\Phi}) = \text{span}\{z\}$, where $z = (z_j)_{j=1}^{\infty} \in \ell^p$, for some $0 < p \leq 1$. Then for each $\varepsilon > 0$, there is a vector $\tilde{z}$ with $\|z - \tilde{z}\|_p < \varepsilon$ such that the Bernstein inequality (1.10) fails for any dictionary $\tilde{\mathbf{\Phi}}$ with $N(\tilde{\mathbf{\Phi}}) = \text{span}\{\tilde{z}\}$.

A specific case is given by $\mathbf{\Phi} = \mathcal{B} \cup \{g\}$, with $\mathcal{B}$ the Dirac basis for ℓ^2 and $g \in \ell^p$ for some $0 < p < 1$. Then we can find an arbitrarily small perturbation $\tilde{g}$ of g in ℓ^p such that the Bernstein inequality fails for the "perturbed" dictionary $\tilde{\mathbf{\Phi}} = \mathcal{B} \cup \{\tilde{g}\}$.

Notice that in the preceeding example, nothing was asssumed about the Bernstein inequality for the dictionary $\mathbf{\Phi}$ itself. Thus, arbitrarily close to any dictionary with a reasonable one dimensional null space, there is a "bad" dictionary.

However, it is possible to find good dictionaries with a one dimensional null space for which (1.6) holds. The following example of such a dictionary.

Example 1.3. Suppose $\mathbf{\Phi}$ satisfies $N(\mathbf{\Phi}) = \text{span}\{z\}$, where $z = (z_j)_{j=1}^{\infty}$ is such that there is a constant $C < \infty$ satisfying

$$\forall k \in \mathbb{N} : \sum_{j=k}^{\infty} |z_j| \leq C|z_k|.$$

Then the Bernstein inequality (1.10) holds true.

An explicit implementation of this example is given by $\mathbf{\Phi} = \mathcal{B} \cup \{g\}$, with $\mathcal{B} = \{e_k\}_{k\in\mathbb{N}}$ an orthonormal basis for ℓ^2 and $g = -\sum_{k=1}^{\infty} a^k e_k$ for some fixed $0 < a < 1$.

Examples 1.2 and 1.3 combined show that one can always perturb a nice dictionary $\mathbf{\Phi}$ for which (1.6) holds ever so slightly as to make (1.6) collapse.

We justify the two examples in Section 3 by performing a careful analysis of the Bernstein inequality (1.10) when $\mathbf{\Phi}$ is a frame. In Section 3.1 we study the general frame dictionary and derive a sufficient condition stated in Proposition 3.1 for (1.10) to hold. Then in Section 3.2 we present a more refined analysis (Proposition 3.2) in the special case where the kernel $N(\mathbf{\Phi})$ is one-dimensional. The proof of Proposition 3.2 is based on an application of the general results in Section 3.1.

1.7 Incoherence and the Restricted Isometry Property

The above examples illustrate that the Bernstein inequality (and its nice consequences such as Theorem 1.1) can be fairly fragile. However, this could be misleading, and we will now show that in a certain sense "most" dictionaries satisfy the inequality in a robust manner.

In a previous work we showed that *incoherent* frames [12] satisfy a "robust" Bernstein inequality, although with an exponent $r = 2(1/\tau - 1/2)$ instead of the exponent $s = 1/\tau - 1/2$ that would match the Jackson inequality. This inequality is then robust, because small enough perturbations of incoherent dictionaries remain incoherent.

In the last decade, a very intense activity related to Compressed Sensing [9] has lead to the emergence of the concept of Restricted Isometry Property (RIP) [3, 4], which generalizes the notion of coherence. A dictionary $\mathbf{\Phi}$ is said to satisfy the RIP of order k with constant δ if, for any coefficient sequence c satisfying $\|c\|_0 \leq k$, we have

$$(1 - \delta) \cdot \|c\|_2^2 \leq \|\mathbf{\Phi} c\|_{\mathcal{H}}^2 \leq (1 + \delta) \cdot \|c\|_2^2. \tag{1.11}$$

The RIP has been widely studied for random dictionaries, and used to relate the minimum ℓ^1 norm solution $c^\star$ of an inverse linear problem $f = \mathbf{\Phi} c$ to a "ground truth" solution c_0 which is assumed to be ideally sparse (or approximately sparse).

In this paper, we are a priori not interested in "recovering" a coefficient vector c_0 from the observation $f = \mathbf{\Phi} c_0$. Instead, we wish to understand how the rate of ideal (but NP-hard) k-term approximation of f using $\mathbf{\Phi}$ is related to the existence of a representation with small ℓ^τ norm.

In Section 4, we study finite-dimensional dictionaries, where it turns out that the lower bound in the RIP (1.11) provides an appropriate tool to obtain Bernstein inequalities with controlled constant[3]. Namely, we say that the dictionary $\mathbf{\Phi}$ satisfies LRIP(k,δ) with a constant $\delta < 1$ provided that

$$\|\mathbf{\Phi} c\|_{\mathcal{H}}^2 \geq (1-\delta)\cdot \|c\|_2^2, \tag{1.12}$$

for any sequence c satisfying $\|c\|_0 \leq k$. We prove (*Lemma 4.1*) that in $\mathcal{H} = \mathbb{R}^N$ the lower frame bound $A > 0$ and the LRIP$(\kappa N, \delta)$, imply a Bernstein inequality for $0 < \tau \leq 2$ with exponent $r = 1/\tau - 1/2$. As a result we have:

Theorem 1.4. *Let $\mathbf{\Phi}$ be an $m \times N$ frame with frame bounds $0 < A \leq B < \infty$. Assume that $\mathbf{\Phi}$ satisfies* LRIP$(\kappa N, \delta)$, *where $\delta < 1$ and $0 < \kappa < 1$. Then*

- *for $0 < \tau \leq 2$, the Bernstein inequality* (1.6) *holds with exponent $r = 1/\tau - 1/2$ and a constant $C_\tau(A,\delta,\kappa) < \infty$ (*cf *Eq.* (4.1)*)*
- *for $0 < \tau \leq 1$, $0 < \theta < 1$, we have, with equivalent norms,*

$$\mathcal{A}_\tau^r(\mathbf{\Phi}) = \ell^\tau(\mathbf{\Phi}) = (\mathcal{H}, \ell^p(\mathbf{\Phi}))_{\theta,\tau}, \quad \frac{1}{\tau} = \frac{\theta}{p} + \frac{1-\theta}{2}.$$

The constant $C_\tau(A,\delta,\kappa)$ and the constants in norm equivalences may depend on A, B, δ, and κ, but they do not depend on the dimension N.

For random Gaussian dictionaries, the typical order of magnitude of $A, \delta(\kappa)$ is known and governed by the aspect ratio $R := N/m$ of the dictionary, provided that it is sufficiently high dimensional (its number of rows should be above a threshold $m(R)$ implicitly defined in Section 4). We obtain the following theorem.

Theorem 1.5. *Let $\mathbf{\Phi}$ be an $m \times N$ matrix with i.i.d. Gaussian entries $\mathcal{N}(0, 1/m)$. Let $R := N/m$ be the redundancy of the dictionary. If $m \geq m(R)$ then, except with probability at most $10R^2 \cdot \exp(-\gamma(R)m)$, we have for all $0 < \tau \leq 1$ the equality*

$$\mathcal{A}_\tau^r(\mathbf{\Phi}) = \ell^\tau(\mathbf{\Phi}) = (\mathcal{H}, \ell^p(\mathbf{\Phi}))_{\theta,\tau}, \quad r = 1/\tau - 1/2 = \theta(1/p - 1/2). \tag{1.13}$$

with equivalent norms.

The constants driving the equivalence of the norms are universal: they only depend on τ and the redundancy factor R but not *on the individual dimensions m and N. Similarly $\gamma(R)$ and $m(R)$ only depend on R.*

For $R \geq 1.27$ we have $\gamma(R) > 7 \cdot 10^{-6}$. For large R we have $\gamma(R) \approx 0.002$.

Indeed, for random Gaussian dictionaries in high-dimension, with high probability, the Bernstein inequality holds for all $0 < \tau \leq 2$ with constants driven by the aspect ratio $R := N/m$ but otherwise *independent of the dimension N*.Using the notion of decomposable dictionary [12, Theorem 3.3], this finite dimensional result can be easily adapted to build arbitrarily overcomplete dictionaries in infinite dimension that satisfy the equality (1.8).

[3]The control of constants is the crucial part, since in finite dimension all norms are equivalent, which implies that the Bernstein inequality is always trivially satisfied.

The result of Theorem 1.5 should be compared to our earlier result for incoherent frames obtained in [12]. In [12] we found an incoherent dictionary with aspect ratio (approximately) 2 for which the Bernstein inequality (1.6) can be shown to hold only for the exponent $r = 2(1/\tau - 1/2)$, i.e., for r twice as large as the Jackson exponent $s = 1/\tau - 1/2$. Theorem 1.5 illustrates that the result in [12] really corresponds to a "worst case" behaviour and there are indeed many dictionaries (according to the Gaussian measure: the overwhelming majority of dictionaries) with a much better behaviour with respect to Bernstein estimates. This holds true even for aspect ratios R that can be arbitrarily large.

1.8 Conclusion and discussion

The restricted isometry property is a concept that has been essentially motivated by the understanding of sparse regularization for linear inverse problems such as compressed sensing. Beyond this traditional use of the concept, we have shown new connections between the RIP and nonlinear approximation.

The main result we obtained is that, from the point of view of nonlinear approximation, a frame which satisfies a nontrivial restricted property $\delta_k < 1$ (i.e., in the regime $k \propto N$) behaves like an orthogonal basis: the optimal rate of m-term approximation can be achieved with an approach that does not involve solving a (potentially) NP-hard problem to compute the best m-term approximation for each m. In such nice dictionaries, near optimal k-term approximation can be achieved in two steps, like in an orthonormal basis:

- decompose the data vector $f = \sum_j c_j \varphi_j$, with coefficients as sparse as possible in the sense of minimum ℓ^τ norm;
- keep the m largest coefficients to build the approximant $f_m := \sum_{j \in I_m} c_j \varphi_j$.

The second main result is that redundant dictionaries with the above property are not the exception, but rather the rule. While it is possible to build nasty overcomplete dictionaries either directly or by arbitrarily small perturbations of some "nice" dictionaries", in a certain sense the vast majority of overcomplete dictionaries are nice.

One should note that several results of this paper are expressed in finite dimension, where all norms are equivalent. The strength of the results is therefore not the mere existence of inequalities between norms, but in the fact that the involved constants do not depend on the dimension. From a numerical perspective, the control of these constants has essentially an impact in (very) large dimension, and it is not clear whether the constants numerically computed for random dictionaries are useful for dimensions less than a few millions.

A few key questions remains open. For a given data vector f, it is generally not known in advance to which $\ell^\tau(\mathbf{\Phi})$ space f belongs: under which conditions is it possible to efficiently compute a sparse decomposition $f = \sum_j c_j \varphi_j$ which is guaranteed to be near optimal in the sense that $\|c\|_\tau$ is almost minimum whenever $f \in \ell^\tau(\mathbf{\Phi})$? Can ℓ^1 minimization (which is convex) be used and provide near best performance under certain conditions ? This is left to future work.

2 Interpolation spaces

We recall the definition of the K-functional. Let Y be a (quasi-)normed space continuously embedded in a Hilbert space $\mathcal{H}$. For $f \in \mathcal{H}$, the K-functional is defined by

$$K(f,t) = K(f,t;\mathcal{H},Y) := \inf_{g\in Y}\{\|f-g\|_{\mathcal{H}} + t\|g\|_Y\}$$

and the norm defining the interpolation spaces $(\mathcal{H},Y)_{\theta,q}$, $0<\theta<1$, $0<q<\infty$, is given by:

$$\|f\|^q_{(\mathcal{H},Y)_{\theta,q}} := \int_0^\infty [t^{-\theta}K(f,t)]^q \frac{dt}{t} \asymp \sum_{j\geq 0} 2^{j\theta q}K(f,2^{-j})^q.$$

The interpolation space $(\mathcal{H},Y)_{\theta,q}$ is simply the set of f for which the norm is finite. In our case we consider a frame dictionary $\mathbf{\Phi}$ and $Y=\ell^p(\mathbf{\Phi})$, which is continuously embedded in $\mathcal{H}$ for $0<p\leq 2$. We have the following result.

Theorem 2.1. *Suppose $\mathbf{\Phi}$ is a frame dictionary for a Hilbert space $\mathcal{H}$. Let $0<\tau\leq 1$ and suppose the Bernstein inequality for $\ell^\tau(\mathbf{\Phi})$ holds with exponent r:*

$$\|f_k\|_{\ell^\tau(\mathbf{\Phi})} \leq C\cdot k^r \cdot \|f_k\|_{\mathcal{H}}, \quad \forall f_k\in\Sigma_k(\mathbf{\Phi}), \forall k.$$

Define $\beta := r/(1/\tau-1/2)$. Then, for all $0<\theta<1$, we have the embedding

$$(\mathcal{H},\ell^p(\mathbf{\Phi}))_{\theta,\tau} \hookrightarrow \ell^\tau(\mathbf{\Phi}), \quad \frac{1}{\tau} = \frac{\theta/\beta}{p} + \frac{(1-\theta/\beta)}{2}; \tag{2.1}$$

Moreover, we have

$$\mathcal{A}^r_\tau(\mathbf{\Phi}) \hookrightarrow \ell^\tau(\mathbf{\Phi}),$$

and if in addition $r = 1/\tau - 1/2$ (i.e., $\beta = 1$), then

$$\mathcal{A}^r_\tau(\mathbf{\Phi}) = \ell^\tau(\mathbf{\Phi})$$

with equivalent norms. The constants in the norm inequalities depend only on p, on the Bernstein constant for $\ell^\tau(\mathbf{\Phi})$, and on the upper frame bound for $\mathbf{\Phi}$.

Proof. We use the general technique proposed by DeVore and Popov [7], and adapt it to the special structure of the considered function spaces. One can check that with $Y=\ell^p(\mathbf{\Phi})$ we have

$$K(f,t) = \inf_c\left\{\|f-\mathbf{\Phi}c\|_{\mathcal{H}} + t\|c\|_p\right\}.$$

For each j we consider c_j an (almost) minimizer of the right hand side above for $t=2^{-j}$. Fix $0<\theta<1$ and define $s:=r/\theta$ and $p<2$ such that $s=1/p-1/2$, and set $m_j = \lfloor 2^{j/s}\rfloor \asymp 2^{j/s}$. Define $\tilde{c}_j$ to match c_j on its m_j largest coordinates, and be zero anywhere else. Finally, define $f_0 := 0$, $f_j := \mathbf{\Phi}\tilde{c}_j$, $j\in\mathbb{N}$. We can observe that

$$\begin{aligned}
\|f-f_j\|_{\mathcal{H}} &\leq \|f-\mathbf{\Phi}c_j\|_{\mathcal{H}} + \|\mathbf{\Phi}(c_j-\tilde{c}_j)\|_{\mathcal{H}} \\
&\overset{(a)}{\lesssim} \|f-\mathbf{\Phi}c_j\|_{\mathcal{H}} + \|c_j-\tilde{c}_j\|_2 \\
&\lesssim \|f-\mathbf{\Phi}c_j\|_{\mathcal{H}} + \|c_j\|_p\cdot m_j^{-s} \lesssim \|f-\mathbf{\Phi}c_j\|_{\mathcal{H}} + \|c_j\|_p\cdot 2^{-j} \\
&\lesssim K(f,2^{-j}),
\end{aligned}$$

where in (a) we used the upper frame bound B of $\mathbf{\Phi}$. Accordingly we get

$$\|f_{j+1}-f_j\|_{\mathcal{H}} \leq \|f-f_j\|_{\mathcal{H}} + \|f-f_{j+1}\|_{\mathcal{H}} \leq C\cdot K(f,2^{-j})$$

where the constant only depends on p and the upper frame bound B of $\mathbf{\Phi}$. Since $\tau \leq 1$, we have the quasi-triangle inequality

$$\|u+v\|^{\tau}_{\ell^{\tau}(\mathbf{\Phi})} \leq \|u\|^{\tau}_{\ell^{\tau}(\mathbf{\Phi})} + \|v\|^{\tau}_{\ell^{\tau}(\mathbf{\Phi})}.$$

Since $f = \lim_{j\to\infty} f_j = \sum_{j=0}^{\infty}(f_{j+1} - f_j)$ we obtain

$$\begin{aligned}
\|f\|^{\tau}_{\ell^{\tau}(\mathbf{\Phi})} &\leq \sum_{j\geq 0} \|f_{j+1} - f_j\|^{\tau}_{\ell^{\tau}(\mathbf{\Phi})} \\
&\overset{(b)}{\lesssim} \sum_{j\geq 0} \left[(2^{j/s})^{r} \|f_{j+1} - f_j\|_{\mathcal{H}}\right]^{\tau} \\
&\lesssim \sum_{j\geq 0} \left[2^{j\theta} K(f, 2^{-j})\right]^{\tau} \asymp \|f\|^{\tau}_{(\mathcal{H},\ell^{p}(\mathbf{\Phi}))_{\theta,\tau}}.
\end{aligned}$$

In (b) we used the fact that $f_{j+1} - f_j \in \Sigma_m(\mathbf{\Phi})$ with $m = m_j + m_{j+1} \lesssim 2^{j/s}$, and the assumption that the Bernstein inequality with exponent r holds for $\ell^{\tau}(\mathbf{\Phi})$. To summarize we obtain $(\mathcal{H}, \ell^{p}(\mathbf{\Phi}))_{\theta,\tau} \subset \ell^{\tau}(\mathbf{\Phi})$, together with the norm inequality

$$\|f\|_{\ell^{\tau}(\mathbf{\Phi})} \leq C \cdot \|f\|_{(\mathcal{H},\ell^{p}(\mathbf{\Phi}))_{\theta,\tau}}$$

where the constant only depends on the Bernstein constant for $\ell^{\tau}(\mathbf{\Phi})$, on p, and on the upper frame bound B for $\mathbf{\Phi}$. We have $1/\tau - 1/2 = r/\beta = (r/\beta s)s = (\theta/\beta)(1/p - 1/2)$, i.e., $1/\tau = (\theta/\beta)/p + (1 - \theta/\beta)/2$.

Similarly, we can define $f_0 = 0$ and f_j a (near)best m_j-term approximation to f with $m_j = 2^{j-1}$, $j \geq 1$ and obtain $\|f_{j+1} - f_j\|_{\mathcal{H}} \leq 2\sigma_{2^{j-1}}(f, \mathbf{\Phi}), j \geq 1$. Using the Bernstein inequality and derivations essentially identical to the previous lines we get

$$\|f\|^{\tau}_{\ell^{\tau}(\mathbf{\Phi})} \lesssim \|f_1\|^{\tau}_{\mathcal{H}} + \sum_{j\geq 1} \left[2^{(j-1)r} \sigma_{2^{j-1}}(f, \mathbf{\Phi})\right]^{\tau} \lesssim \|f\|^{\tau}_{\mathcal{A}^{r}_{\tau}(\mathbf{\Phi})}.$$

The constant only depends on the Bernstein constant for $\ell^{\tau}(\mathbf{\Phi})$.

Using [11, Theorem 6], the upper frame bound implies the continuous embedding $\ell^{\tau}(\mathbf{\Phi}) \hookrightarrow \mathcal{A}^{s}_{\tau}(\mathbf{\Phi})$ with $s = 1/\tau - 1/2$. Hence, when the Berstein exponent is $r = 1/\tau - 1/2 = s$ we have equality that is to say $\mathcal{A}^{r}_{\tau}(\mathbf{\Phi}) = \ell^{\tau}(\mathbf{\Phi})$ with equivalent norms. ☐

Remark 2.2. A consequence of Theorem 2.1 is a partial answer to an open question raised in [12], where "blockwise incoherent dictionaries" are considered and a Bernstein inequality is proved, with exponent $r = \beta(1/\tau - 1/2)$, $\beta = 2$, for all $0 < \tau \leq 2$, yielding the two-sided embedding [12, Theorem 3.2]:

$$\ell^{\tau}(\mathbf{\Phi}) \hookrightarrow \mathcal{A}^{s}_{q}(\mathbf{\Phi}) \hookrightarrow (\mathcal{H}, \ell^{\tau}(\mathbf{\Phi}))_{1/2,q}, \quad 0 < \tau \leq 2, \quad \tau \leq q < \infty, \quad s = 1/\tau - 1/2.$$

By Theorem 2.1, for $0 < q \leq 1$, the Bernstein inequality with exponent $r = \beta(1/q - 1/2)$ further implies the embedding $(\mathcal{H}, \ell^{\tau}(\mathbf{\Phi}))_{1/2,q} \hookrightarrow \ell^{q}(\mathbf{\Phi})$ where $1/q = 1/(4\tau) + 3/8$, i.e., $q = 8\tau/(3\tau + 2)$. As a result we have

$$\ell^{\tau}(\mathbf{\Phi}) \hookrightarrow \mathcal{A}^{s}_{q}(\mathbf{\Phi}) \hookrightarrow \ell^{q}(\mathbf{\Phi}), \quad 0 < \tau \leq 2/5, \quad q = 8\tau/(3\tau + 2), \quad s = 1/\tau - 1/2.$$

We know from [12] an example of blockwise incoherent dictionary where the exponent of the Berstein inequality cannot be improved, hence the above embedding is also sharp for this class of dictionaries.

3 Bernstein estimates for frame dictionaries

In this section we are interested in the Bernstein inequality (1.10) in the general case where the dictionary $\mathbf{\Phi}$ forms a frame for a Hilbert space $\mathcal{H}$. The dimension of $\mathcal{H}$ may be finite or infinite. We will show that the Bernstein inequality is closely linked to properties of the kernel of $\mathbf{\Phi}$ given by

$$N = N(\mathbf{\Phi}) := \{z \in \ell^2 : \mathbf{\Phi} z = 0\}.$$

In fact, the frame property ensures that $\|\mathbf{\Phi} c\|_{\mathcal{H}} \asymp \inf_{z\in N} \|c+z\|_2$ for any sequence $c \in \ell^2$. Hence, the Bernstein inequality (1.10) holds if and only if the quantity

$$C(\mathbf{\Phi}) := \sup_{m\in\mathbb{N}} \sup_{c:\|c\|_0\le m} \sup_{z\in N} \inf_{v\in N} m^{-1/2} \cdot \frac{\|c+v\|_1}{\|c+z\|_2} \tag{3.1}$$

is finite.

We split our analysis in two parts. In Section 3.1 we derive an upper bound on $C(\mathbf{\Phi})$ that results in a sufficient condition for (1.10) to hold for a general frame dictionary (Proposition 3.1). In Section 3.2 we specialize to the case where the kernel $N(\mathbf{\Phi})$ is one-dimensional.

The analysis in Section 3.2 is used to justify Examples 1.2 and 1.3.

3.1 Bernstein constant for general dictionaries

Here we derive an upper estimate of the quantity $C(\mathbf{\Phi})$, given by (3.1), for general frame dictionaries in a Hilbert space. This estimate leads to the following sufficient condition for a Bernstein inequality for such dictionaries.

Proposition 3.1. *Suppose the dictionary $\mathbf{\Phi}$ forms a frame for the Hilbert space $\mathcal{H}$, and $\mathbf{\Phi}$ has kernel $N := N(\mathbf{\Phi})$. Then the Bernstein inequality* (1.10) *holds provided that*

$$\sup_{z\in N} \sup_{m\in\mathbb{N}} \sup_{I:|I|\le m} m^{-1/2} \cdot \frac{\|z_{I^c}\|_1}{\|z_{I^c}\|_2} < \infty. \tag{3.2}$$

Moreover, in the case where $N \cap \ell^1 = \{0\}$, the Bernstein inequality (1.10) *holds if and only if*

$$C_1(\mathbf{\Phi}) := \sup_{z\in N} \sup_{m\in\mathbb{N}} \frac{\|z_m\|_1}{m^{1/2}\sigma_m(z)_2} < \infty, \tag{3.3}$$

where z_m is the vector containing the m largest entries in z.

Proof. We prove the sufficient condition for (1.10) by deriving an upper bound for $C(\mathbf{\Phi})$ given by (3.1). For any $m \in \mathbb{N}$,

$$\begin{aligned}
\sup_{c:\|c\|_0\le m} \sup_{z\in N} \inf_{v\in N} \frac{\|c+v\|_1}{m^{1/2}\|c+z\|_2} &= \sup_{I:|I|\le m} \sup_{c:\operatorname{supp}(c)\subseteq I} \sup_{z\in N} \inf_{v\in N} \frac{\|c+v\|_1}{m^{1/2}\|c+z\|_2} \\
&= \sup_{I:|I|\le m} \sup_{c:\operatorname{supp}(c)\subseteq I} \sup_{z\in N} \inf_{v\in N} \frac{\|c+v_I\|_1 + \|v_{I^c}\|_1}{m^{1/2}\sqrt{\|c+z_I\|_2^2 + \|z_{I^c}\|_2^2}} \\
&\le \sup_{I:|I|\le m} \sup_{c:\operatorname{supp}(c)\subseteq I} \sup_{z\in N} \frac{\|c+z_I\|_1 + \|z_{I^c}\|_1}{m^{1/2}\sqrt{\|c+z_I\|_2^2 + \|z_{I^c}\|_2^2}}.
\end{aligned} \tag{3.4}$$

For a given support I and $z \in N$, we introduce $\gamma_I^z := \|z_{I^c}\|_1 / \|z_{I^c}\|_2$. Notice that for $|I| \leq m$ and c with $\mathrm{supp}(c) \subseteq I$, we have

$$\begin{aligned}
\|c + z_I\|_1 + \|z_{I^c}\|_1 &\leq m^{1/2}\|c + z_I\|_2 + \gamma_I^z \|z_{I^c}\|_2 \\
&\leq \max\{m^{1/2}, \gamma_I^z\}\left(\|c + z_I\|_2 + \|z_{I^c}\|_2\right) \\
&\leq \sqrt{2}\max\{m^{1/2}, \gamma_I^z\}\sqrt{\|c + z_I\|_2^2 + \|z_{I^c}\|_2^2}. \qquad (3.5)
\end{aligned}$$

We let $\gamma_m^z := \sup_{I:|I|\leq m} \gamma_I^z$. Hence, from (3.4) we deduce that

$$C(\mathbf{\Phi}) \leq \sqrt{2}\max\left\{1, \sup_{z\in N}\sup_{m\in\mathbb{N}} \frac{\gamma_m^z}{m^{1/2}}\right\},$$

which shows that condition (3.2) implies $C(\mathbf{\Phi}) < \infty$.

Let us now consider the case $N \cap \ell^1 = \{0\}$. Notice that the infimum over $v \in N$ in (3.1) is attained for $v = 0$. Hence, $C(\mathbf{\Phi}) = \sup_{z\in N} B_z$, with

$$B_z := \sup_{m\in\mathbb{N}} \sup_{I:|I|\leq m} \sup_{c:\mathrm{supp}(c)\subseteq I} \frac{\|c\|_1}{m^{1/2}\sqrt{\|c + z_I\|_2^2 + \|z_{I^c}\|_2^2}}. \qquad (3.6)$$

For a fixed support I, standard estimates show that

$$\frac{\|c\|_1}{m^{1/2}\sqrt{\|c + z_I\|_2^2 + \|z_{I^c}\|_2^2}}$$

is maximal for choices of the type $c = -[z_I + \lambda \mathrm{sign}(z)\mathbf{1}_I]$, $\lambda > 0$. This choice of c leads to the corresponding (squared) optimization problem

$$\sup_{\lambda\in\mathbb{R}} \frac{(\lambda m + \|z_I\|_1)^2}{m(\lambda^2 m^2 + \|z_{I^c}\|_2^2)} = \frac{1}{m}\left(\frac{\|z_I\|_1^2}{\|z_{I^c}\|_2^2} + 1\right).$$

Notice that

$$\sup_{I:|I|\leq m} \frac{\|z_I\|_1}{\|z_{I^c}\|_2} = \frac{\|z_m\|_1}{\sigma_m(z)_2},$$

so we deduce that

$$C_1(\mathbf{\Phi}) \leq C(\mathbf{\Phi}) \leq C_1(\mathbf{\Phi}) + 1, \qquad (3.7)$$

which completes the proof. □

3.2 Dictionaries with one dimensional null-spaces

We now turn to the simplified case where the dictionary $\mathbf{\Phi}$ has a one-dimensional null-space. In this case, we derive necessary conditions for the Bernstein inequality (1.10) to hold that is valid even when $N(\mathbf{\Phi}) \subset \ell^1$, a case not covered by the necessary condition of Proposition 3.1. We prove the following:

Proposition 3.2. *Suppose the dictionary $\mathbf{\Phi}$ is a frame for the Hilbert space $\mathcal{H}$ and has a one-dimensional null-space, $N(\mathbf{\Phi}) = \mathrm{span}\{z\}$. Also suppose the Bernstein inequality* (1.10) *holds. Then*

$$C_2(z) := \sup_{m\in\mathbb{N}} \sup_{I:|I|\leq m} \min\left(\frac{\|z_I\|_1}{m^{1/2}\|z_{I^c}\|_2}, \frac{\|z_{I^c}\|_1}{m^{1/2}\|z_{I^c}\|_2}\right) < \infty. \qquad (3.8)$$

Moreover, if $z \in \ell^p$ for some $0 < p < 1$, and the Bernstein inequality (1.10) *holds for $\mathbf{\Phi}$, then*

$$\sup_{m\in\mathbb{N}} \frac{\sigma_m(z)_1}{m^{1/2}\sigma_m(z)_2} < \infty.$$

Proof of Proposition 3.2. In this setting, the Bernstein inequality (1.10) holds if and only if the quantity

$$C(\mathbf{\Phi}) := \sup_{m\in\mathbb{N}} \sup_{c:\|c\|_0\le m} \sup_{\mu\in\mathbb{R}} \inf_{\lambda\in\mathbb{R}} \frac{\|c+\lambda z\|_1}{m^{1/2}\|c+\mu z\|_2}$$

is finite. By rescaling, we have

$$\begin{aligned}
C(\mathbf{\Phi}) &= \sup_{m\in\mathbb{N}} \sup_{c:\|c\|_0\le m} \sup_{\mu\in\mathbb{R}} \inf_{\lambda\in\mathbb{R}} \frac{\|\frac{\lambda}{\mu}z+\frac{1}{\mu}c\|_1}{m^{1/2}\|\frac{1}{\mu}c+z\|_2}\\
&= \sup_{m\in\mathbb{N}} \sup_{\tilde{c}:\|\tilde{c}\|_0\le m} \inf_{\delta\in\mathbb{R}} \frac{\|\delta z+\tilde{c}\|_1}{m^{1/2}\|\tilde{c}+z\|_2}\\
&= \sup_{m\in\mathbb{N}} \sup_{I:|I|\le m} \sup_{c:\mathrm{supp}(c)\subseteq I} \inf_{\delta\in\mathbb{R}} \frac{\|\delta z+c\|_1}{m^{1/2}\|c+z\|_2}\\
&= \sup_{m\in\mathbb{N}} \sup_{I:|I|\le m} \sup_{c:\mathrm{supp}(c)\subseteq I} \inf_{\delta\in\mathbb{R}} \frac{\|\delta z_I+c\|_1+\|\delta z_{I^c}\|_1}{m^{1/2}\sqrt{\|c+z_I\|_2^2+\|z_{I^c}\|_2^2}}.
\end{aligned} \tag{3.9}$$

To get a lower estimate for $C(\mathbf{\Phi})$, we simply chose $c = -z_I$ in (3.9) to obtain

$$\begin{aligned}
C(\mathbf{\Phi}) &\ge \sup_{m\in\mathbb{N}} \sup_{I:|I|\le m} \inf_{\delta\in\mathbb{R}} \frac{|\delta-1|\|z_I\|_1+|\delta|\|z_{I^c}\|_1}{m^{1/2}\|z_{I^c}\|_2}\\
&= \sup_{m\in\mathbb{N}} \sup_{I:|I|\le m} \min\left(\frac{\|z_I\|_1}{m^{1/2}\|z_{I^c}\|_2}, \frac{\|z_{I^c}\|_1}{m^{1/2}\|z_{I^c}\|_2}\right)\\
&\ge \sup_{m\in\mathbb{N}} \min\left(\frac{\|z_m\|_1}{m^{1/2}\sigma_m(z)_2}, \frac{\sigma_m(z)_1}{m^{1/2}\sigma_m(z)_2}\right).
\end{aligned} \tag{3.10}$$

Then clearly $C(\mathbf{\Phi}) < \infty$ implies condition (3.8).

If, in addition, we have $z \in \ell^p$ for some $0 < p < 1$, then it follows from standard results on nonlinear approximation with bases in ℓ^2, see [8], that $m^{1/2}\sigma_m(z)_2 \to 0$ as $m \to \infty$. Thus

$$\frac{\|z_m\|_1}{m^{1/2}\sigma_m(z)_2} \to \infty,$$

and we conclude from (3.10) that

$$\sup_{m\in\mathbb{N}} \frac{\sigma_m(z)_1}{m^{1/2}\sigma_m(z)_2} < \infty.$$

□

We now turn to a justification of Examples 1.2 and 1.3 using Propositions 3.1 and 3.2.

3.2.1 Examples 1.2 and 1.3 revisited

We first verify the claim made in Example 1.2. Given any dictionary $\mathbf{\Phi}$ with $N(\mathbf{\Phi}) = \text{span}\{z\}$, where $z = (z_j)_{j=1}^{\infty} \in \ell^p$, for some $0 < p \le 1$. For any $\varepsilon > 0$, we modify z as follows

1. Choose $m_0 \ge 2$ such that $\sum_{j=m_0}^{\infty} |z_j|^p < \varepsilon/2$.
2. Choose a sequence $\{m_\ell\}_{\ell=1}^{\infty}$ satisfying $m_{\ell+1}/m_\ell \to \infty$. Notice that the sequence will necessarily have super-exponential growth.
3. Fix $\beta > 1/p \ge 1$, and choose $\{\gamma_j\}_{j=0}^{\infty}$ such that $\gamma_\ell := C(m_{\ell+1} - m_\ell)^{-\beta}$, with the constant C defined by the equation $\sum_{\ell=0}^{\infty} \gamma_\ell^p [m_{\ell+1} - m_\ell] = \sum_{j=m_0+1}^{\infty} |z_j|^p$.
4. Now define $\tilde{z} = (\tilde{z}_j)_{j=0}^{\infty}$ by

$$\tilde{z}_j := \begin{cases} z_j, & 0 \le j \le m_0 \\ \gamma_\ell, & j \in [m_\ell + 1, m_{\ell+1}], \quad \ell \in \mathbb{N}_0. \end{cases}$$

It is easy to verify (using 1. and 3.) that $\|z - \tilde{z}\|_p^p < \varepsilon$.

Let us consider the index set $I = [1, m_\ell]$, $\ell \ge 1$. We have,

$$\begin{aligned} \frac{\|\tilde{z}_{I^c}\|_1^2}{\|\tilde{z}_{I^c}\|_2^2} &\ge \frac{[(m_{\ell+1} - m_\ell)\gamma_\ell]^2}{C^2 \sum_{k=\ell}^{\infty} (m_{k+1} - m_k)^{1-2\beta}} \\ &\ge \frac{(m_{\ell+1} - m_\ell)^{2-2\beta}}{(m_{\ell+1} - m_\ell)^{1-2\beta}} \\ &\ge m_{\ell+1} - m_\ell. \end{aligned}$$

Thus,

$$\frac{\|\tilde{z}_{I^c}\|_1^2}{m_\ell \|\tilde{z}_{I^c}\|_2^2} \ge \frac{m_{\ell+1} - m_\ell}{m_\ell} = \frac{m_{\ell+1}}{m_\ell} - 1 \to \infty,$$

as $\ell \to \infty$. Also, since $\beta > 1$,

$$\begin{aligned} \frac{\|\tilde{z}_I\|_1^2}{m_\ell \|\tilde{z}_{I^c}\|_2^2} &\ge \frac{C'}{C^2 \sum_{k=\ell}^{\infty} (m_{k+1} - m_k)^{1-2\beta}} \\ &\ge \frac{C'}{m_\ell (m_{k+1} - m_k)^{1-2\beta}} \\ &\ge \frac{m_{\ell+1} - m_\ell}{m_\ell} \to \infty, \end{aligned}$$

as $\ell \to \infty$. We conclude that $C_2(\tilde{z}) = \infty$, with $C_2(\tilde{z})$ given in Proposition 3.2.

To verify the claim made in Example 1.3, suppose $\mathbf{\Phi}$ satisfies $N(\mathbf{\Phi}) = \text{span}\{z\}$, where $z = (z_j)_{j=1}^{\infty}$ is such that there is a constant $C < \infty$ satisfying

$$\forall k \in \mathbb{N} : \sum_{j=k}^{\infty} |z_j| \le C|z_k|.$$

Then for any finite index set $I \subset \mathbb{N}$, we let $m_I := \min\{k : k \notin I\}$, and notice that $\|z_{I^c}\|_1 \leq \sum_{j=m_I}^{\infty} |z_j| \leq C|z_{m_I}|$, while $\|z_{I^c}\|_2 \geq |z_{m_I}|$. Hence,

$$\frac{\|z_{I^c}\|_1}{\|z_{I^c}\|_2} \leq C\frac{|z_{m_I}|}{|z_{m_I}|} = C,$$

and

$$\sup_{m \in \mathbb{N}} \sup_{I:|I| \leq m} \frac{\|z_{I^c}\|_1}{m^{1/2}\|z_{I^c}\|_2} \leq C < \infty,$$

so (3.2) is satisfied and the Bernstein inequality (1.10) holds by Proposition 3.1.

4 Bernstein inequality and the RIP

For certain incoherent dictionaries studied in [12], the Berstein inequality *cannot* match the Jackson inequality, but it still holds with a sharp exponent $r = 2(1/\tau - 1/2)$ for any $\tau \leq 2$, i.e. the sharp factor that can be used in Theorem 2.1 is $\beta = 2$. This result exploits incoherence [12, Lemma 2.3] to prove that the lower bound in the RIP is satisfied for k of the order of $\sqrt{N}$. Below we prove that the lower frame bound (1.9), together with the lower bound in the RIP (1.12) with *k of the order of N*, implies the Bernstein inequality (1.6) with controlled constant and exponent matching that of the Jackson inequality (1.5). This Lemma therefore extends our previous result based on incoherence [12, Theorem 2.1].

Lemma 4.1. *Let $\mathbf{\Phi}$ be an $m \times N$ dictionary. Suppose $\mathbf{\Phi}$ has lower frame bound $A > 0$ and satisfies* LRIP$(\kappa N, \delta)$*, where $\delta < 1$ and $0 < \kappa < 1$. Then for $0 < \tau \leq 2$, the Bernstein inequality* (1.6) *holds with exponent $r = 1/\tau - 1/2$ and constant*

$$C_\tau(A, \delta, \kappa) := \max\left\{(1-\delta)^{-1/2}, A^{-1/2}\kappa^{1/2-1/\tau}\right\}. \tag{4.1}$$

Proof. First, suppose $1 \leq k \leq \kappa N$. Take $f \in \Sigma_k(\mathbf{\Phi})$, and write $f = \mathbf{\Phi}c$ with $\|c\|_0 \leq k$. Then, by the LRIP$(\kappa N, \delta)$ condition,

$$\|f\|_{\ell^\tau(\mathbf{\Phi})} \leq \|c\|_\tau \leq k^{1/\tau-1/2}\|c\|_2 \leq (1-\delta)^{-1/2}k^{1/\tau-1/2} \cdot \|\mathbf{\Phi}c\|_{\mathcal{H}}.$$

For $\kappa N \leq k \leq N$, take $f \in \Sigma_k(\mathbf{\Phi})$. We express f in terms of its canonical frame expansion relative to $\mathbf{\Phi}$,

$$f = \sum_{j=1}^{N} \langle f, \tilde{\varphi}_j \rangle \varphi_j. \tag{4.2}$$

We recall that the dual frame $\{\tilde{\varphi}_j\}$ has an upper frame bound A^{-1}. Hence, we can use the expansion (4.2) to deduce that

$$\begin{aligned}\|f\|_{\ell^\tau(\mathbf{\Phi})} &\leq \|\{\langle f, \tilde{\varphi}_j \rangle\}\|_\tau \leq N^{1/\tau-1/2}\|\{\langle f, \tilde{\varphi}_j \rangle\}\|_2 \\ &\leq A^{-1/2}N^{1/\tau-1/2} \cdot \|\mathbf{\Phi}c\|_{\mathcal{H}} \\ &\leq [A^{-1/2}\kappa^{1/2-1/\tau}]k^{1/\tau-1/2} \cdot \|\mathbf{\Phi}c\|_{\mathcal{H}}.\end{aligned}$$

The Bernstein inequality and its constant now follow at once from the two separate estimates. □

Lemma 4.1 proves half of Theorem 1.4. Let us complete the proof now.

Proof of Theorem 1.4. As we have seen, the lower frame bound and the LRIP$(\kappa N, \delta)$ property imply the Bernstein inequality for all $0 < \tau \leq 2$. Moreover the upper frame bound implies a Jackson inequality. For $0 < p < \tau \leq 1$ both the Jackson and Bernstein inequalities hold for $\ell^p(\mathbf{\Phi})$ with exponent $1/p - 1/2$, hence [6, Chapter 7] we have with equivalent norms

$$\mathcal{A}^r_\tau(\mathbf{\Phi}) = (\mathcal{H}, \ell^p(\mathbf{\Phi}))_{\theta,\tau}, \quad r = \theta(1/p - 1/2), \quad 0 < \theta < 1.$$

The Bernstein inequality also holds for $\ell^\tau(\mathbf{\Phi})$ with exponent $r = 1/\tau - 1/2$, hence by Theorem 1.1 $\mathcal{A}^r_\tau(\mathbf{\Phi}) = \ell^\tau(\mathbf{\Phi})$ with equivalent norms. ☐

Next we wish to estimate A, B, δ, κ when $\mathbf{\Phi}$ is a random Gaussian dictionary. The following Lemma summarizes well known facts (see e.g. [4, 2]).

Lemma 4.2. *Let $\mathbf{\Phi}$ be an $m \times N$ matrix with i.i.d. Gaussian entries $\mathcal{N}(0, 1/m)$. For any $\varepsilon > 0$ and $1 \leq k < m$, it satisfies the* LRIP(k, δ) *with $1 - \delta = (1 - \eta)^2$, where*

$$\eta := \sqrt{\frac{k}{m}} \cdot \left(1 + (1+\varepsilon) \cdot \sqrt{2 \cdot \left(1 + \log \frac{N}{k}\right)}\right) \tag{4.3}$$

except with probability at most

$$\exp\left(-2\varepsilon k \cdot (1 + \log \frac{N}{k})\right). \tag{4.4}$$

Moreover, except with probability at most $\exp(-\varepsilon^2 m/2)$*, it has the lower frame bound*

$$A \geq (\sqrt{N/m} - 1 - \varepsilon)^2 \tag{4.5}$$

and, except with probability at most $\exp(-\varepsilon^2 m/2)$*, it has the upper frame bound*

$$B \leq (\sqrt{N/m} + 1 + \varepsilon)^2 \tag{4.6}$$

Proof. First, for a given index set Λ of cardinality $k < m$, we observe that the restricted matrix $\mathbf{\Phi}_\Lambda$ is $m \times k$ with i.i.d. Gaussian entries $\mathcal{N}(0, 1/m)$, hence its smallest singular value exceeds $1 - \sqrt{k/m} - t$ except with probability at most $\exp(-mt^2/2)$ [5, Theorem II.13]. By a union bound, the smallest singular values among all submatrices $\mathbf{\Phi}_\Lambda$ associated to the $\binom{N}{k}$ possible index sets Λ of cardinality k exceeds $1 - \sqrt{k/m} - t$, except with probability at most $p(t) := \binom{N}{k} \cdot \exp(-mt^2/2)$. Since for all N, k we have $\binom{N}{k} \leq (Ne/k)^k = \exp\left(k \cdot (1 + \log \frac{N}{k})\right)$, it follows that

$$p(t) \leq \exp\left(k \cdot (1 + \log \frac{N}{k}) - mt^2/2\right).$$

For $\varepsilon > 0$ we set

$$t := (1+\varepsilon) \cdot \sqrt{\tfrac{2k}{m} \cdot \left(1 + \log \tfrac{N}{k}\right)}$$

and obtain that, except with probability at most

$$p(\varepsilon) \quad \leq \quad \exp\left(k \cdot (1 + \log \frac{N}{k}) \cdot \left(1 - (1+\varepsilon)^2\right)\right) \leq \exp\left(-2\varepsilon k \cdot (1 + \log \frac{N}{k})\right)$$

we have: for all k-sparse vector c with $\|c\|_0 = k$,

$$\|\mathbf{\Phi} c\|_2 \quad \geq \quad \left[1 - \sqrt{k/m} \cdot \left(1 + (1+\varepsilon) \cdot \sqrt{2 \cdot \left(1 + \log \frac{N}{k}\right)}\right)\right] \cdot \|c\|_2.$$

To control the frame bounds we consider the random matrix $\mathbf{\Psi} := \sqrt{\frac{m}{N}}\mathbf{\Phi}^T$. Since $\mathbf{\Psi}$ is $N \times m$ with i.i.d. Gaussian entries $\mathcal{N}(0, 1/N)$, for any $t > 0$ all its singular values exceed $1 - \sqrt{m/N} - t$, except with probability at most $\exp(-Nt^2/2)$ [5, Theorem II.13]. Setting $t = \varepsilon \cdot \sqrt{m/N}$, since $\|\mathbf{\Phi}^T x\|_2^2 = \frac{N}{m}\|\mathbf{\Psi} x\|_2^2$, we obtain that $\mathbf{\Phi}$ has lower frame bound

$$\sqrt{A} \geq \sqrt{\frac{N}{m}} \cdot \Big(1 - (1+\varepsilon) \cdot \sqrt{m/N}\Big) = \sqrt{N/m} - 1 - \varepsilon$$

except with probability at most $\exp(-\varepsilon^2 m/2)$. We proceed identically for the upper frame bound, using the fact that for any $t > 0$, no singular value of $\mathbf{\Psi}$ exceeds $1 + \sqrt{m/N} + t$, except with probability at most $\exp(-Nt^2/2)$ [5, Theorem II.13]. □

We now obtain our first main theorem (Theorem 1.1) by controlling the constant δ from below when k/m is bounded from above, given the redundancy $R = N/m$ of the dictionary $\mathbf{\Phi}$.

Proof of Theorem 1.1. In Appendix A we exhibit a threshold $t(R) \in (0,1)$ such that if $N/m = R$ and $t = k/m \leq t(R)$ then

$$\eta := \sqrt{\frac{k}{m}} \cdot \left(1 + 2 \cdot \sqrt{2 \cdot \left(1 + \log \frac{N}{k}\right)}\right) \leq 1/2.$$

Consider $k := \lfloor t(R) m \rfloor$. By Lemma 4.2 the dictionary $\mathbf{\Phi}$ satisfies the LRIP(k, δ) with $(1-\delta)^{-1/2} = (1-\eta)^{-1} = 2$ except with probability at most

$$\begin{aligned} p_1 = \exp\left(-2k \cdot \left(1 + \log \frac{N}{k}\right)\right) &\leq \exp\left(-2\left[t(R)m - 1\right] \cdot (1 + \log R)\right) \\ &\leq e^{2(1+\log R)} \cdot \exp\big(-2t(R)(1+\log R)m\big) \end{aligned}$$

Moreover, setting $\varepsilon := (\sqrt{R}-1)/2$, it has lower frame bound $\sqrt{A} \geq \sqrt{N/m} - 1 - \varepsilon \geq (\sqrt{R} - 1)/2$ except with probability at most $p_2 = \exp(-\varepsilon^2 m/2)$. For $m \geq m(R) := 2/t(R)$ we have

$$\kappa = \frac{k}{N} = \frac{k}{m} \cdot \frac{m}{N} \geq \frac{t(R) - \frac{1}{m}}{R} \geq \frac{t(R)}{2R}.$$

and by Lemma 4.1, we obtain (except with probability at most $p_1 + p_2$) that the Bernstein inequality holds for each $0 < \tau \leq 2$ with constant

$$C_\tau(R) \leq \max\left(2, 2(\sqrt{R}-1)^{-1} \cdot \left[\frac{t(R)}{2R}\right]^{1/2 - 1/\tau}\right). \tag{4.7}$$

Since we also have the upper frame bound $\sqrt{B} \leq \sqrt{R} + 1 + \varepsilon'$ except with probability at most $p_3 = \exp(-(\varepsilon')^2 m/2)$ we obtain with $\varepsilon' = 1$ that the upper frame bound $\sqrt{R} + 2$ together with the Bernstein inequality with constant $C_\tau(R)$ jointly hold, except with probability at most $p_1 + p_2 + p_3 \leq \beta \exp(-\gamma m)$ where

$$\begin{aligned} \beta &= e^{2+2\log R} + 2 = e^2 R^2 + 2 \leq (e^2+2) R^2 \leq 10\, R^2; \\ \gamma &\geq \min\Big(2t(R) \cdot (1 + \log R), (\sqrt{R}-1)^2/8, 1/2\Big) =: \gamma(R). \end{aligned}$$

As shown in Appendix A, $\lim_{R\to\infty} \gamma(R) \approx 0.002$, and $\gamma(R) \geq 7 \cdot 10^{-6}$ when $R \geq 1.28$. □

A

For $u \in (0,1)$ we have $u \log 1/u \leq e$, hence for $u \in (0,1)$ and $0 < p \leq 1$:

$$\log 1/u = (1/p) \log 1/u^p \leq (1/p) e/u^p = (1/u)^p (e/p).$$

Therefore, for $a > 1$, $b \geq e$, $0 < t < 1$, using $u = t/b$, we obtain

$$\begin{aligned} \eta(t) \quad &:= \quad \sqrt{t} \cdot \left(1 + a\sqrt{\log(b/t)}\right) \leq \sqrt{t} \cdot \left(1 + a\sqrt{(b/t)^p (e/p)}\right) \\ &\leq \quad t^{\frac{1}{2}-\frac{p}{2}} \cdot \left(t^{\frac{p}{2}} + ae^{\frac{1}{2}}\sqrt{b^p/p}\right) \leq t^{\frac{1}{2}-\frac{p}{2}} \cdot 2ae^{\frac{1}{2}}\sqrt{b^p/p} \end{aligned}$$

where in the last inequality we used the fact that $ae^{\frac{1}{2}}\sqrt{b^p/p} > 1$ (all the factors exceed one) and $t^{\frac{1}{p}} < 1$. For $p = 1/\log b$ we have $b^p/p = e \log b$ hence

$$\eta(t) \leq 2ae\sqrt{\log b} \cdot t^{\frac{1}{2}\left(1-\frac{1}{\log b}\right)}$$

The definition of $\eta(t)$ can be identified with (4.3) for $\varepsilon = 1$ with $t = k/m$, $a = 2\sqrt{2}$ and $b = eN/m = eR \geq e$. Denoting $c = 4ae = 8\sqrt{2}e$, we have just proved

$$\eta(t) \leq (c/2) \cdot \sqrt{1 + \log R} \cdot t^{\frac{1}{2}\left(\frac{\log R}{1+\log R}\right)}, \quad \forall 0 < t < 1.$$

Defining

$$t(R) := \left[c^2 \cdot (1 + \log R)\right]^{-1-\frac{1}{\log R}} \in (0,1), \tag{A.1}$$

we have the guarantee $\eta\big(t(R)\big) \leq 1/2$ as well as the identity

$$2t(R) \cdot (1 + \log R) = 2c^{-2} \cdot \left[c^2 \cdot (1 + \log R)\right]^{-\frac{1}{\log R}}.$$

The right hand side is an increasing function of R, with limit zero when $R \to 1$ and limit $2c^{-2}$ when $R \to \infty$. When $R \geq R_0 := (1 + 4/c)^2$ we have $(\sqrt{R} - 1)^2/8 \geq 2c^{-2}$ hence

$$\gamma(R) := \min\left(2t(R) \cdot (1 + \log R), (\sqrt{R} - 1)^2/8\right) = 2t(R) \cdot (1 + \log R).$$

Since $c = 8\sqrt{2}e = 2^{7/2}e$, we have $c^2 = 2^7 e^2$ hence

$$2c^{-2} = 2^{-6}e^{-2} \approx 0.0021 > 0.002, \text{ and } R_0 = (1 + \frac{1}{\sqrt{8}e})^2 \approx 1.277 > 1.27.$$

For $R \geq R_0$,

$$\gamma(R) \geq 2c^{-2}[c^2 \cdot (1 + \log R_0)]^{-\frac{1}{\log R_0}} \approx 7.8 \cdot 10^{-6} > 7 \cdot 10^{-6}.$$

and $\lim_{R\to\infty} \gamma(R) = 2c^{-2} > 2 \cdot 10^{-3}$. Finally, when $R \geq R_0$ we have $m(R) = 2/t(R) = 4(1 + \log R)/\gamma(R) \leq 6 \cdot 10^5 \cdot (1 + \log R)$, and in the limit of large R we obtain $m(R) \asymp 2c^2(1 + \log R) \lesssim 2000 \cdot (1 + \log R)$.

References

[1] H. H. Bang and M. Morimoto. On the Bernstein-Nikolsky inequality. *Tokyo J. Math.*, 14(1):231–238, 1991.

[2] R. Baraniuk, M. Davenport, R. A. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, Dec. 2008.

[3] E. J. Candes and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.

[4] E. J. Candes and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.

[5] K. R. Davidson and S. J. Szarek. Local operator theory, random matrices and banach spaces. In W. B. Johnson and J. Lindenstrauss, editors, *Handbook on the Geometry of Banach spaces*, volume 1, pages 317–366. Elsevier Science, 2001, 2003.

[6] R. A. DeVore and G. G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993.

[7] R. A. DeVore and V. A. Popov. Interpolation spaces and nonlinear approximation. In *Function spaces and applications (Lund, 1986)*, volume 1302 of *Lecture Notes in Math.*, pages 191–205. Springer, Berlin, 1988.

[8] R. A. DeVore and V. N. Temlyakov. Some remarks on greedy algorithms. *Adv. Comput. Math.*, 5(2-3):173–187, 1996.

[9] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[10] R. Gribonval and M. Nielsen. Letter to the editor: on a problem of Gröchenig about nonlinear approximation with localized frames. *J. Fourier Anal. Appl.*, 10(4):433–437, 2004.

[11] R. Gribonval and M. Nielsen. Nonlinear approximation with dictionaries. I. Direct estimates. *J. Fourier Anal. Appl.*, 10(1):51–71, 2004.

[12] R. Gribonval and M. Nielsen. Nonlinear approximation with dictionaries. II. Inverse estimates. *Constr. Approx.*, 24(2):157–173, 2006.

[13] S. Mallat and Z. Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41(12):3397–3415, Dec. 1993.

[14] Y. Pati, R. Rezaiifar, and P. Krishnaprasad. Orthonormal matching pursuit : recursive function approximation with applications to wavelet decomposition. In *Proceedings of the* 27^{th} *Annual Asilomar Conf. on Signals, Systems and Computers*, Nov. 1993.

Centre de recherche INRIA Rennes – Bretagne Atlantique
IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

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