Bollettino Unione Matematica Italiana

Albert Cohen

Greedy Algorithms for Adaptive Approximation

Bollettino dell'Unione Matematica Italiana, Serie 9, Vol. **2** (2009), n.2, p. 391–402.

Unione Matematica Italiana

<http://www.bdim.eu/item?id=BUMI_2009_9_2_2_391_0>

Bollettino dell'Unione Matematica Italiana, Unione Matematica Italiana, 2009.

Bollettino U. M. I.
(9) **II** (2009), 391-402

Greedy Algorithms for Adaptive Approximation (*)

ALBERT COHEN

Abstract. – *We discuss the performances of greedy algorithms for two problems of numerical approximation. The first one is the best approximation of an arbitrary function by an N-terms linear combination of simple functions adaptively picked within a large dictionary. The second one is the approximation of an arbitrary function by a piecewise polynomial function on an optimally adapted triangulation of cardinality N. Performance is measured in terms of convergence rate with respect to the number of element in the dictionary in the first case and of triangles in the second case.*

1. – Introduction.

Approximation theory is the branch of mathematics which studies the process of approximating an arbitrary function f by simpler functions which typically depend on a finite number N - or $\mathcal{O}(N)$ - of parameters. It plays a pivotal role in the analysis of numerical methods. Typical examples of approximation processes are algebraic or trigonometric polynomials, finite elements or linear combinations of wavelets.

One usually makes the distinction between *linear* and *nonlinear* approximation. In the first case, the simple function is picked from a linear space, such as polynomials of degree N or piecewise polynomial functions on some fixed partition of cardinality N. The approximation is typically computed by projection of f onto this space. In the second case, the simple function is picked from a nonlinear space, yet still characterizable by $\mathcal{O}(N)$ parameters. Such a situation typically occurs when dealing with adaptive or data driven approximations, which makes it relevant for applications as diverse as data compression, statistical estimation or numerical schemes for partial differential or integral equations (see [11] for a general survey). However the notion of projection is not anymore applicable and therefore a critical question is:

How to compute the best possible approximation to a given function f from a nonlinear space?

(*) Conferenza tenuta a Torino il 7 luglio 2006 in occasione del "Joint Meeting S.I.M.A.I. - S.M.A.I. - S.M.F. - U.M.I. sotto gli auspici dell'E.M.S. Mathematics and its Applications".

Let us translate this question in concrete terms for the two specific examples that will be further discussed in this paper:

Adaptive triangulations: given a function f defined on a polygonal domain Ω and given $N > 0$, find a partition of Ω into N triangles such that the L^p-error between f and its projection f_N onto piecewise polynomial functions of some fixed degree m on this partition is minimized.

Best N-term approximation: given a dictionary $\mathcal{D}$ of functions which is normalized and complete in some Hilbert space $\mathcal{H}$, and given $f \in \mathcal{H}$ and $N > 0$, find the combination $f_N = \sum_{k=1}^{N} c_k g_k$ which approximates f at best, where $\{c_1, \cdots, c_N\}$ are real numbers and $\{g_1, \cdots, g_N\}$ are picked from $\mathcal{D}$.

In order to make these problems computationally tractable, one may assume in the first example that the vertices of each triangle are picked within a limited yet large number of locations M, or in the second example that the search is limited to a subset of $\mathcal{D}$ of cardinality M. However the exhaustive search for the optimal solution has the combinatorial order of complexity $\binom{M}{N}$ and both problems are therefore generally not solvable in polynomial time in N and M. A relevant goal is therefore to look for sub-optimal yet acceptable solutions which can be computed in reasonable time.

Greedy algorithms constitute a simple approach for achieving this goal. They rely on stepwise local optimization procedures for picking the parameters in an inductive fashion, with the hope of approaching the globally optimal solution. They are particularly easy to implement, yet the analysis of their approximation performance gives rise to many interesting problems.

We present in § 2 and § 3 two different types of greedy algorithms corresponding respectively to the two above mentionned nonlinear approximation problems. The analysis of their performance shows that although they do not provide with the optimal solution to these problems, they still yield optimal convergence estimates for the error $f - f_N$ measured in the appropriate norm L^p or $\mathcal{H}$.

2. – Greedy algorithms for adaptive triangulations.

In the context of adaptive triangulations, an important distinction is between *isotropic* and *anisotropic* triangulations. In the first case the triangles satisfy a condition which guarantees that they do not differ too much from equilateral triangles. This can either be stated in terms of a minimal value $\theta_0 > 0$ for every angle, or by a uniform bound on the aspect ratio

$$r_T := \frac{h_T}{\rho_T}$$

of each triangle T where h_T and ρ_T respectively denote the diameter of T and of its largest inscribed disc. In the second case, which is in the scope of the present paper, the aspect ratio is allowed to be arbitrarily large, i.e. long and thin triangles are allowed. In summary, adaptive and anisotropic triangulations mean that we do not fix any constraint on the size and shape of the triangles.

Given a function f and a norm $\|\cdot\|_X$ of interest, we can formulate the problem of finding the *optimal triangulation* for f in the X-norm in two related forms:

- For a given N find a triangulation $\mathcal{T}_N$ with N triangles and a piecewise polynomial function f_N of some fixed degree m on $\mathcal{T}_N$ such that the error $\|f - f_N\|_X$ is minimized.
- For a given tolerance $\varepsilon > 0$ find a triangulation $\mathcal{T}_N$ with minimal number of triangles N and a piecewise polynomial function f_N such that $\|f - f_N\|_X \leq \varepsilon$.

In this paper X is the L^p norm for some arbitrary $1 \leq p \leq \infty$.

Concrete mesh generation algorithms have been developped in order to generate in reasonable time triangulations which are “close” to the above described optimal trade-off between error and complexity. They are typically governed by two intuitively desirable features:

1. The triangulation should *equidistribute* the local approximation error between each triangle. This rationale is typically used in local mesh refinement algorithms for numerical PDE’s [21]: a triangle is refined when the local approximation error (estimated by an a-posteriori error indicator) is large.
2. In the case of anisotropic meshes, the local aspect ratio should in addition be optimally adapted to the approximated function f. In the case of piecewise linear approximation ($m = 1$), this is achieved by imposing that the triangles are isotropic with respect to a distorted metric induced by the Hessian d^2f. We refer in particular to [6] where this task is executed using Delaunay mesh generation techniques.

While these last algorithms fastly produce anisotropic meshes which are naturally adapted to the approximated function, they suffer from two intrinsic limitations:

1. They are based on the evaluation of the hessian d^2f, and therefore do not in principle apply to arbitrary functions $f \in L^p(\Omega)$ for $1 \leq p \leq \infty$ or to noisy data.
2. They are non-hierarchical: for $N > M$, the triangulation $\mathcal{T}_N$ is not a refinement of $\mathcal{T}_M$.

The need for hierarchical triangulations is critical in the construction of wavelet bases, which play an important role in applications to image and terrain data processing, in particular data compression [8]. In such applications, the multilevel structure is also of key use for the fast encoding of the information. Hierarchy is also useful in the design of optimally converging adaptive methods for PDE's [13, 18, 5]. However, all these developments are so far mostly restricted to isotropic refinement methods.

A natural objective is therefore to design adaptive algorithmic techniques that combine hierarchy and anisotropy, and that apply to any function $f \in L^p(\Omega)$.

A simple greedy algorithm was introduced in [9] in order to fullfill this objective. The algorithm is based on a local approximation operator $\mathcal{A}_T$ acting from $L^p(T)$ onto Π_m - the set of polynomials of total degree less or equal to m. Here, the parameters $m \geq 0$ and $1 \leq p \leq \infty$ are arbitrary but fixed. We define the local L^p approximation error

$$e_T(f)_p := \|f - \mathcal{A}_T f\|_{L^p(T)}.$$

The most natural choice for $\mathcal{A}_T$ is the operator of best $L^p(T)$ approximation, so that

$$e_T(f)_p := \inf_{\pi \in \Pi_m} \|f - \pi\|_{L^p(T)}.$$

In practice, one might prefer to use an operator which is easier to compute, yet nearly optimal in the sense that

$$\|f - \mathcal{A}_T f\|_{L^p(T)} \leq C \inf_{\pi \in \Pi_m} \|f - \pi\|_{L^p(T)}. \tag{2.1}$$

with C a Lebesgue constant independent of f and T. This is in particular achieved by taking $\mathcal{A}_T = \Pi_T$ the $L^2(T)$-orthogonal projection onto Π_m.

Our algorithm starts with a a coarse triangulation $\mathcal{T}_{N_0}$ with N_0 triangles. Given $\mathcal{T}_N$, the algorithm constructs $\mathcal{T}_{N+1}$ by first selecting the triangle T which maximizes the local approximation error $e_T(f)_p$. This T is then bisected into two sub-triangles of equal area by bisection from one of its three vertices $a_i \in \{a_0, a_1, a_2\}$ towards the mid-point of the opposite edge. We denote by $T_{i,1}$ and $T_{i,2}$ the two resulting triangles. The chosen vertex is

$$i^* := \text{Argmin}_{i=0,1,2} d_T(i,f),$$

where $d_T(i, f)$ is a *decision function* that we describe below, defined for any triangle T. Stopping criterions for the algorithm can be defined in various ways:

- Number of triangles: stop once a prescribed N is attained.
- Local error: stop once $e_T(f)_p \leq \varepsilon$ for all $T \in \mathcal{T}_N$, for some prescribed $\varepsilon > 0$.
- Global error: stop once $\|f - f_N\|_{L^p} \leq \varepsilon$ for some prescribed $\varepsilon > 0$.

The role of selecting the triangle which maximizes the local approximation error is obviously to obtain a triangulation which equidistributes the error. The role of the decision function is to drive the generation of anisotropic triangles according to the local properties of f, in contrast to simpler procedures such as *newest vertex bisection* (i.e. split T from the most recently created vertex) which is independant of f and generates triangulations with isotropic shape constraint.

Therefore, the choice of $d_T(i,f)$ is critical in order to obtain triangles with an optimal aspect ratio. The most natural choice corresponds to the optimal split

$$d_T(i,f) = e_{T_{i,1}}(f)_p^p + e_{T_{i,2}}(f)_p^p, \tag{2.2}$$

i.e. minimize the resulting L^p error after bisection. It was shown in [10] that optimal aspect ratio can be reached in the case of piecewise linear approximation with the use of decision functions which are either based on the L^2 or L^∞ norm, namely

$$d_T(i,f) = \|f - \Pi_{T_{i,1}} f\|^2_{L^2(T_{i,1})} + \|f - \Pi_{T_{i,2}} f\|^2_{L^2(T_{i,2})}, \tag{2.3}$$

with Π_T the local orthogonal projection operator, or

$$d_T(i,f) = \|f - I_{T_{i,1}} f\|_{L^\infty(T_{i,1})} + \|f - I_{T_{i,2}} f\|_{L^\infty(T_{i,2})}, \tag{2.4}$$

with I_T the local interpolation operator.

REMARK 2.1. – The triangulations which are generated by the greedy procedure are in general non-conforming, i.e. exhibit hanging nodes. This is not problematic in the present setting since we consider approximation in the L^p norm which does not require global continuity of the piecewise polynomial functions.

Let us first illustrate numerically the optimal adaptation properties of the refinement procedure in terms of triangle shape. For this purpose, we take $f = \boldsymbol{q}$ a quadratic form i.e. an homogeneous polynomial of degree 2. In this case, all triangles should have the same aspect ratio since the hessian is constant. In order to measure the adequation of the shape of a triangle T with $\boldsymbol{q}$, we introduce the following quantity: if (a, b, c) are the vectors corresponding to the edges of T, we define

$$\sigma_q(T) = \frac{\max\{|\boldsymbol{q}(a)|, |\boldsymbol{q}(b)|, |\boldsymbol{q}(c)|\}}{|T|\sqrt{|\det(\boldsymbol{q})|}},$$

where $\det(f)$ is the determinant of the 2×2 symmetric matrix associated to $\boldsymbol{q}$. Using the reference triangle and an affine change of variable, it is proved in [10] that

$$e_T(\boldsymbol{q})_p \sim |T|^{1+\frac{1}{p}} \sigma_q(T)\sqrt{\det(\boldsymbol{q})},$$

with equivalence constants independent of $\boldsymbol{q}$ and T. Therefore, if T is a triangle of given area, its shape should be designed in order to minimize $\sigma_q(T)$.

In the case where $\boldsymbol{q}$ is positive definite or negative definite, $\sigma_q(T)$ takes small values when T is isotropic *with respect to the metric*

$$|(x,y)|_q := \sqrt{|\boldsymbol{q}(x,y)|},$$

the minimal value $\frac{4}{\sqrt{3}}$ being attained for an equilateral triangle for this metric. Specifically, we choose $\boldsymbol{q}(x,y) := x^2 + 100y^2$ and display on Figure 1 (left) the triangulation $\mathcal{T}_{256}$ obtained after 8 iterations of the refinement procedure, initialized on a triangle which is equilateral for the euclidean metric (and therefore not adapted to $\boldsymbol{q}$). Triangles such that $\sigma_q(T) \leq 4\sqrt{3}$ (at most 3 times the minimal value) are displayed in white, others in grey. We observe that most triangles produced by the refinement procedure are of the first type and therefore have a good aspect ratio.

The case of a quadratic function of mixed signature is illustrated on Figure 1 (right) with $\boldsymbol{q}(x,y) := x^2 - 10y^2$. For such quadratic functions, triangles which are isotropic with respect to the positive quadratic form $\tilde{\boldsymbol{q}}$ corresponding to the absolute value of the symmetric matrix associated to $\boldsymbol{q}$ (here $\tilde{\boldsymbol{q}}(x,y) = x^2 + 10y^2$) have a low value of σ_q. But one can also check that σ_q is left invariant by any linear transformation with eigenvalues $\left(t, \frac{1}{t}\right)$ for any $t > 0$ and eigenvectors corresponding to the the *null cone* of $\boldsymbol{q}$ (here $(\sqrt{10}, 1)$ and $(\sqrt{10}, -1)$). Therefore long and thin triangles which are aligned with this null cone also have a low σ_q. Triangles T such that $\sigma_{\tilde{q}}(T) \leq 4\sqrt{3}$ are displayed in white, those such that $\sigma_q(T) \leq 4\sqrt{3}$ while $\sigma_{\tilde{q}}(T) > 4\sqrt{3}$ - i.e. adapted to $\boldsymbol{q}$ but not to $\tilde{\boldsymbol{q}}$ - are displayed in grey, and the others in dark. We observe that most triangles produced by the refinement procedure are either of the first or second type and therefore have a good aspect ratio.

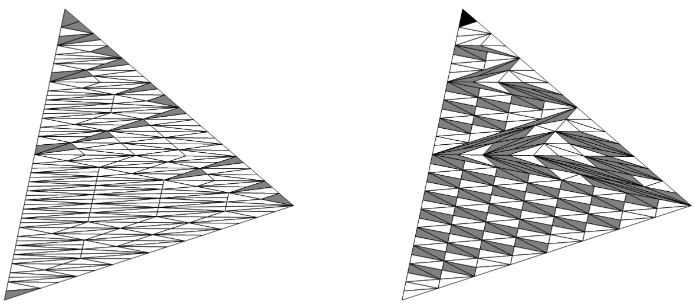

Fig. 1. – $\mathcal{T}_{256}$ for $q(x,y) := x^2 + 100y^2$ (left) and $q(x,y) := x^2 - 10y^2$ (right).

These experimental observations are confirmed by the results in [10] which show in particular that there exists an absolute constant C_0 such that the proportion of triangles such that $\sigma_q(T) \leq C_0$ tends to 1 as the refinement procedure is iterated. Based on these results optimal error estimates have been established in [10] for the approximation of more general smooth functions $f \in C^2$ by piecewise linear functions on adaptive triangulations $\mathcal{T}_N$ generated by the greedy algorithm. The estimates in [10] are of the form

$$\|f - f_N\|_{L^p} \leq CN^{-1}\|\sqrt{\det(|d^2 f|)}\|_{L^q}, \quad \frac{1}{q} = \frac{1}{p} + 1, \tag{2.5}$$

and were proved to be optimal in [7, 1]. However the triangulations proposed in these last papers are non-hierarchical and based on the evaluation of $d^2 f$. On the other hand, the validity of (2.5) for the greedy algorithm is so far limited to strictly convex functions, although conjectured to hold for any smooth f.

We illustrate the adaptive triangulations produced by the greedy algorithm for a function f displaying a sharp transition along a curved edge. Specifically we take

$$f(x, y) = f_\delta(x, y) := g_\delta(\sqrt{x^2 + y^2}),$$

where g_δ is defined by $g_\delta(r) = \frac{5 - r^2}{4}$ for $0 \leq r \leq 1$, $g_\delta(1 + \delta + r) = -\frac{5 - (1 - r)^2}{4}$ for $r \geq 0$, g_δ is a polynomial of degree 5 on $[1, 1 + \delta]$ which is determined by imposing that g_δ is globally C^2. The parameter δ therefore measures the sharpness of the transition.

Figure 2 displays the triangulation $\mathcal{T}_{10000}$ obtained after 10000 steps of the algorithm for $\delta = 0.2$. In particular, triangles T such that $\sigma_q(T) \leq 4$ where for the quadratic form associated to $d^2 f$ measured at the barycenter of T are displayed in white, others in grey. As expected, most triangles are of the first type therefore well adapted to f. We also display on this figure the adaptive isotropic

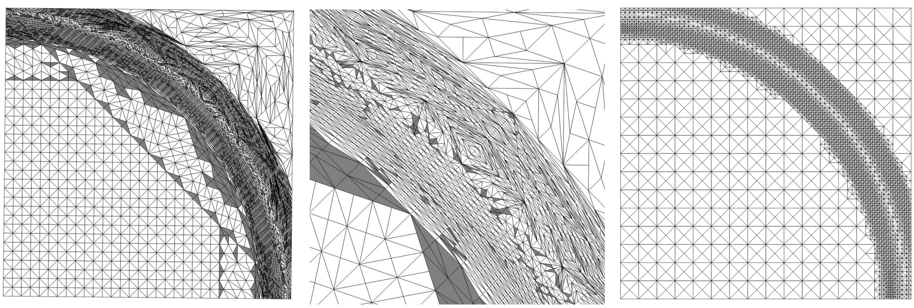

Fig. 2. – $\mathcal{T}_{10000}$ (left), detail (center), isotropic triangulation (right).

triangulation produced by the greedy tree algorithm based on newest vertex bisection for the same number of triangles.

Since f is a C^2 function, approximations by uniform, adaptive isotropic and adaptive anisotropic triangulations all yield the convergence rate $\mathcal{O}(N^{-1})$. However the constant

$$C := \limsup_{N\to+\infty} N\|f - f_N\|_{L^2},$$

strongly differs depending on the algorithm and on the sharpness of the transition, as illustrated on the table below. We denote by C_U, C_I and C_A the values $N\|f - f_N\|_2$ for $N = 8192$, in the uniform, isotropic an anisotropic case respectively. On columns 2, 3 and 4 we observe that C_U and C_I grow as $\delta \to 0$ while C_A remains uniformly bounded. This is in accordance with the fact that the quantity $A(f) = \|\sqrt{\det(|d^2 f|)}\|_{L^{2/3}}$ which appears in (2.5) if $p = 2$ remains uniformly bounded as $\delta \to 0$, as illustrated on column 5.

δ	C_U	C_I	C_A	$A(f)$
0.2	7.87	1.78	0.74	6.74
0.1	23.7	2.98	0.92	8.52
0.05	65.5	4.13	0.92	8.50
0.02	200	6.60	0.92	8.47

3. – Greedy algorithm for N-term approximation.

Greedy algorithms for building N-term approximations were initially introduced in the context of statistical data analysis. Their approximation properties were first explored in [2, 15] in relation with neural network estimation, and in [12] for general dictionaries. Surveys on such algorithms is given in [19, 20].

We only describe here the four most commonly used greedy algorithms:

1. Stepwise Projection (SP): $\{g_1, \cdots, g_{k-1}\}$ being selected we define f_{k-1} as the orthogonal projection onto $\mathrm{Span}\{g_1, \cdots, g_{k-1}\}$. The next g_k is selected so to minimize the distance between f and $\mathrm{Span}\{g_1, \cdots, g_{k-1}, g\}$ among all choices of $g \in \mathcal{D}$.
2. Orthonormal Matching Pursuit (OMP): with the same definition for f_{k-1}, we select g_k so to maximize the inner product $|\langle f - f_{k-1}, g\rangle|$ among all choices of $g \in \mathcal{D}$. In contrast to SP, we do not need to evaluate the anticipated projection error for all choices of $g \in \mathcal{D}$, which makes OMP more attractive from a computational viewpoint.
3. Relaxed Greedy Algorithm (RGA): f_{k-1} being constructed, we define $f_k = \alpha_k f_{k-1} + \beta_k g_k$, where (α_k, β_k, g) are selected so to minimize the distance

between f and $\alpha f_{k-1} + \beta g$ among all choices of (α, β, g). It is often convenient to fix α_k in advance, which leads to selecting g_k which maximizes $|\langle f - \alpha_k f_{k-1}, g\rangle|$ and $\beta_k = \langle f - \alpha_k f_{k-1}, g_k\rangle$. A frequently used choice is $\alpha_k := (1 - c/k)_+$ for some fixed $c > 1$. The intuitive role of the relaxation parameter α_k is to damp the memory of the algorithm which might have been misled in its first steps. Since no orthogonal projection is involved, RGA is even cheaper than OMP.

4. Pure Greedy Algorithm (PGA): this is simply RGA with the particular choice $\alpha_k = 1$. We therefore select g_k so to maximize the inner product $|\langle f - f_{k-1}, g\rangle|$ as in OMP, and then set $f_k = f_{k-1} + \langle f - f_{k-1}, g_k\rangle g_k$.

It should be noted that in the case where $\mathcal{D}$ is an orthonormal basis, SP, OMP and PGA are equivalent to the procedure of retaining the largest coefficients in the expansion of f which is known to produce the best N-term approximation.

For a general dictionary, $\mathcal{D}$ a natural question is wether a similar property holds: if f admits a sparse representation in $\mathcal{D}$, can we derive some corresponding rate of convergence for the greedy algorithm ? By analogy with the case of an orthonormal basis, we could assume that $f = \sum_{g\in\mathcal{D}} c_g g$ for some sequence $(c_g)_{g\in\mathcal{D}} \in w\ell^p$ and ask wether the greedy algorithm converges with rate N^{-s} with $s = \frac{1}{p} - \frac{1}{2}$. However, the condition $(c_g)_{g\in\mathcal{D}} \in w\ell^p$ is not anymore appropriate since it does not generally guarantee the convergence of $\sum_{g\in\mathcal{D}} c_g g$ in $\mathcal{H}$.

A first set of results concerns the case where f admits a summable expansion, i.e. $(c_g)_{g\in\mathcal{D}} \in \ell^1$ or equivalently f belongs to a multiple of the convex hull of $(-\mathcal{D}) \cup \mathcal{D}$. In this case, the series $\sum_{g\in\mathcal{D}} c_g g$ trivially converges in $\mathcal{H}$ since $\|g\|_{\mathcal{H}} = 1$ for all $g \in \mathcal{D}$. We denote as $\mathcal{L}^1$ the space of such f, equiped with the norm

$$\|f\|_{\mathcal{L}^1} := \inf_{f=\sum c_g g} \sum_{g\in\mathcal{D}} |c_g|.$$

Clearly $\mathcal{L}^1 \subset \mathcal{H}$ with continuous embedding. The following result was proved in [15] for SP and RGA with the choice $\alpha_k := (1 - c/k)_+$ and in [12] for OMP.

THEOREM 3.1. – *If $f \in \mathcal{L}^1$, then*

$$\|f - f_N\|_{\mathcal{H}} \le C\|f\|_{\mathcal{L}^1} N^{-\frac{1}{2}},$$

with C a fixed constant.

Note that the exponent $s = 1/2$ is consistent with $p = 1$. The case a more general function $f \in \mathcal{H}$ that does not have a summable expansion can be treated by the following result [3] which again holds for SP, OMP and RGA with the choice $\alpha_k := (1 - c/k)_+$.

THEOREM 3.2. – *If $f \in \mathcal{H}$, then for any $h \in \mathcal{L}^1$, we have*

$$\|f - f_N\|_{\mathcal{H}} \leq \|f - h\|_{\mathcal{H}} + C\|h\|_{\mathcal{L}^1} N^{-\frac{1}{2}}.$$

with C a fixed constant.

This result reveals that the accuracy of the greedy approximant is in some sense stable under perturbation, although the component selection process involved in the algorithm is unstable by nature.

An immediate consequence is that the greedy algorithm is convergent for any $f \in \mathcal{H}$ since we can approximate f to arbitrary accuracy by an $h \in \mathcal{L}^1$ (for example with a finite expansion in $\mathcal{D}$).

We can also use this result in order to identify more precisely the classes of functions which govern the approximation rate of the algorithm. Indeed, since the choice of $h \in \mathcal{L}^1$ is arbitrary, we have

$$\|f - f_N\|_{\mathcal{H}} \leq \inf_{h \in \mathcal{L}^1} \{\|f - h\|_{\mathcal{H}} + C\|h\|_{\mathcal{L}^1} N^{-1/2}\}.$$

The right hand side has the form of a so-called K-functional which is the central tool in the theory of interpolation space. Generally speaking, if X and Y are a pair of Banach function space, the corresponding K-functional is defined for all $f \in X + Y$ and $t > 0$ by

$$K(f, t) = K(f, t, X, Y) := \inf_{g \in X, h \in Y, g + h = f} \{\|g\|_X + t\|h\|_Y\}.$$

One then defines interpolation space by growth conditions on $K(f, t)$. In particular we say that $f \in [X, Y]_{\theta,\infty}$ (with $0 < \theta < 1$) if and only if there is a constant C such that for all $t > 0$,

$$K(f, t) \leq Ct^{\theta}.$$

We refer to [4] for general treatments of interpolation spaces. In our present setting, we see that

$$\|f - f_N\|_{\mathcal{H}} \leq K(f, CN^{-\frac{1}{2}}, \mathcal{H}, \mathcal{L}^1),$$

and we therefore obtain

$$f \in [cH, \mathcal{L}^1]_{\theta,\infty} \Rightarrow \|f - f_N\|_{\mathcal{H}} \leq CN^{-s} \;\; s = \frac{\theta}{2}.$$

This result is consistent with the particular case of an orthonormal basis since in this case $\mathcal{H} \sim \ell^2(\mathcal{D})$ and $\mathcal{L}^1 \sim \ell^2(\mathcal{D})$ so that $[\mathcal{H}, \mathcal{L}^1]_{\theta,\infty} \sim [\ell^2, \ell]_{\theta,\infty}$ which is known to coincide with the space $w\ell^p$ with $\frac{1}{p} = \theta + \frac{(1-\theta)}{2} =$. We therefore recover the fact that $\|f - f_N\|_{\mathcal{H}} \leq CN^{-s}$ when $(c_g)_{g \in \mathcal{D}} \in w\ell^p$ with $\frac{1}{p} = \frac{1}{2} + s$. For a more general dictionary, if we are able characterize the space $\mathcal{L}^1$ by some smoothness condition in $\mathcal{H}$, then $[\mathcal{H}, \mathcal{L}^1]_{\theta,\infty}$ will correspond to some intermediate smoothness condition.

The above results show that greedy algorithms have the convergence rate N^{-s} with $0 \le s \le \frac{1}{2}$ when f has a moderately concentrated expansion in $\mathcal{D}$.

At the other end, one might ask how the algorithm behaves when f has a highly concentrated expansion, i.e. $f = \sum_{g\in\mathcal{D}} c_g g$ with $(c_g)_{g\in\mathcal{D}} \in \ell^p$ for some $p < 1$. The limit case $p = 0$ of a finitely supported expansion corresponds to the *sparse recovery problem*: from the data of f can we recover its exact finite expansion by a fast algorithm?

For a general dictionary, it was proved in [12] that $(c_g)_{g\in\mathcal{D}} \in \ell^p$ with $p < 1$ implies the existence of a sequence f_N of N-terms approximant which converge towards f with the optimal rate N^{-s} with $s = \frac{1}{p} - \frac{1}{2}$. However SP, OMP and RGA may fail to converge faster than $N^{-\frac{1}{2}}$. They may also fail to solve the sparse recovery problem.

On the other hand we know that SP, OMP and PGA are successful in the special case where $\mathcal{D}$ is an orthonormal basis. A natural question is therefore to understand the general conditions on a $\mathcal{D}$ under which the convergence of greedy algorithms might fully benefit of such concentration properties, similar to the case of an orthonormal basis. Important progress has been recently made in this direction, in relation with the topic of *compressed sensing*. We refer in particular to [14] in which it is proved that OMP succeeds with high probability in the sparse recovery problem for randomly generated dictionaries.

Other open questions concern the PGA algorithm for which it was proved in [12] that $f \in \mathcal{L}_1$ implies that

$$\|f - f_N\| \le CN^{-\frac{1}{6}}.$$

This rate was improved to $N^{-\frac{11}{62}}$ in [16], but on the other hand it was shown [17] that for a particular dictionary there exists $f \in \mathcal{L}_1$ such that

$$\|f - f_N\| \ge cN^{-0.27}.$$

The exact best rate N^{-s} achievable for a general dictionary and $f \in \mathcal{L}^1$ is still unknown, but we already see that PGA is sub-optimal in comparison to SP, OMP and RGA. An interesting problem is thus to understand which conditions should be imposed on the dictionary in order to recover an optimal rate of convergence for this particular algorithm.

REFERENCES

[1] V. Babenko - Y. Babenko - A. Ligun - A. Shumeiko, *On Asymptotical Behavior of the Optimal Linear Spline Interpolation Error of C^2 Functions*, East J. Approx., **12**(1) (2006), 71-101.

[2] A. Barron, *Universal approximation bounds for superposition of n sigmoidal functions*, IEEE Trans. Inf. Theory **39** (1993), 930-945

[3] A. Barron - A. Cohen - W. Dahmen - R. DeVore, *Approximation and learning by greedy algorithms*, to appear in Annals of Statistics (2007).

[4] J. Bergh - J. Löfström, *Interpolation spaces*, Springer Verlag, Berlin, 1976.

[5] P. Binev - W. Dahmen - R. DeVore, *Adaptive Finite Element Methods with Convergence Rates*, Numerische Mathematik **97** (2004), 219-268.

[6] H. Borouchaki - P. J. Frey - P. L. George - P. Laug - E. Saltel, *Mesh generation and mesh adaptivity: theory, techniques*, in Encyclopedia of computational mechanics, E. Stein, R. de Borst and T. J. R. Hughes ed., John Wiley & Sons Ltd., 2004.

[7] L. Chen - P. Sun - J. Xu, *Optimal anisotropic meshes for minimizing interpolation error in L^p-norm*, Math. of Comp., **76** (2007), 179-204.

[8] A. Cohen - W. Dahmen - I. Daubechies - R. DeVore, *Tree-structured approximation and optimal encoding*, App. Comp. Harm. Anal., **11** (2001), 192-226.

[9] A. Cohen - N. Dyn - F. Hecht - J. M. Mirebeau, *Adaptive multiresolution analysis based on anisotropic triangulations*, preprint, Laboratoire J.-L. Lions, 2008.

[10] A. Cohen - J. M. Mirebeau, *Greedy bisection generates optimally adapted triangulations*, preprint, Laboratoire J.-L. Lions, 2008.

[11] R. DeVore, *Nonlinear approximation*, Acta Numerica (1997), 51-150.

[12] R. DeVore - V. Temlyakov, *Some remarks on greedy algorithms*, Advances in Computational Mathematics, **5** (1998), 173-187.

[13] W. Dörfler, *A convergent adaptive algorithm for Poisson's equation*, SIAM J. Numer. Anal., **33** (1996), 1106-1124.

[14] A. C. Gilbert - J. A. Tropp, *Signal recovery from random measurements via Orthogonal Matching Pursuit*, IEEE Trans. Info. Theory, **53** (2007), 4655-4666.

[15] L. K. Jones, *A simple lemma on greedy approximation in Hilbert spaces and convergence rates for projection pursuit regression and neural network training*, Ann. Stat., **20** (1992), 608-613.

[16] S. V. Konyagin - V. N. Temlyakov, *Rate of convergence of Pure greedy Algorithm*, East J. Approx. **5** (1999), 493-499.

[17] E. D. Livshitz - V. N. Temlyakov, *Two lower estimates in greedy approximation*, Constr. Approx., **19** (2003), 509-524.

[18] P. Morin - R. Nochetto - K. Siebert, *Convergence of adaptive finite element methods*, SIAM Review, **44** (2002), 631-658.

[19] V. Temlyakov, *Nonlinear methods of approximation*, Journal of FOCM, **3** (2003), 33-107.

[20] V. Temlyakov, *Greedy algorithms*, to appear in Acta Numerica.

[21] R. Verfurth, *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*, Wiley-Teubner, 1996.

Albert Cohen: Laboratoire Jacques-Louis Lions
Université Pierre et Marie Curie, 175, rue du Chevaleret, 75013 Paris, France
e-mail: cohen@ann.jussieu.fr

Received September 9, 2008