# A GENERAL FRAMEWORK FOR NONLINEAR APPROXIMATIONS WITH APPLICATIONS TO IMAGE RESTORATION*

V. CANDELA† AND A. FALCÓ, PANTALEÓN D. ROMERO‡

**Abstract.** In this paper we establish sufficient conditions for the existence and uniqueness of optimal nonlinear approximations. Most nonlinear problems do not only have difficulties in order to implement good projection algorithms, but also the subsets where we project the functions do not have the geometric properties necessary for classic existence results (such as convexity, for instance). The theoretical results we show here overcome some of these difficulties. We illustrate these results by applying them to a fractional model for image deconvolution, where we can ensure existence and uniqueness of the solution and convergence of the computational algorithms.

**Key words.** Hilbert Space, Pure Greedy Algorithm, Fractional Deconvolution, Image Restoration.

**AMS subject classifications.** 65D15,40A25

**1. Introduction.** Unlike linear approximation, where there exists a solid theoretical background establishing conditions for existence, uniqueness and algorithmic issues, nonlinear approximation is a field with not so deep a knowledge. Nonlinearity is such a wide concept that most fundamental concepts in linear spaces cannot be generalized without losing their strength.

Lack of the vectorial structure of the spaces, of course, is one of the main difficulties to obtain adequate results, but there are other features which are often lost, some of them geometrical (convexity, for instance) and other ones algebraic (linear bases). In the last years, sparsity helped the devise of nonlinear models, by introducing the concept of *dictionaries* as generators of the space sacrificing linear independence. This, optimal approximation can be reduced to the search of a good dictionary first, and, then, the best fitted subset of the dictionary to get an adequate approximation. This is the principle lying under the so called *greedy algorithms*.

However, dictionaries cannot overcome in general structural, mostly geometric, problems. Convexity is usually one of these problems: in practice, spaces generated by any element of the dictionary are not convex, thus eliminating one of the sufficient conditions to ensure the existence of optimal approximations.

In this paper, we provide theoretical results related to nonlinear approximation with milder conditions to give a ground to the general topic of nonlinear approximation for a large class of settings. These conditions are verified in many practical problems.

Image processing, and, particularly, image restoration is one of these fields where this theory can be applied. Both deconvolution and denoising of images are best dealt with adaptive nonlinear models than linear ones (leaving aside computational aspects), in the sense that the process must work in a different way not only for different type of convolution kernels, or different noises, but also depending on the image itself. In [2, 5], the authors propose a blind deconvolution model based on an iterative fractional decomposition of the kernel, with a fractional parameter obtained by the properties of

*Supported by grant PRCEU-UCH30/10, GVA PRE 2010/066 and MTM 2088-03597

†Departamento de Matemática Aplicada, València Burjassot 46100, Valencia (Spain) (vicente.candela@uv.es)

‡Departamento de Ciencias, Físicas, Matemáticas y de la Computación Universidad CEU Cardenal Herrera San Bartolomé 55 46115 Alfara del Patriarca (Valencia), Spain ((afalco/pantaleon.romero)@uch.ceu.es)

the image itself. Though, in practice, this model converges with high quality results, its convergence has been theoretically proved in the context introduced in this paper.

The structure of this paper is as follows: in §2, we introduce the notation, definitions and basic preliminary results in order to simplify the main results in the paper, which are shown in §3. In §4, we consider the blind deconvolution model in [2, 5] as a particular case of this result, and we display some examples of the application of the model in §5. Finally, we will draw conclusions in §6.

**2. Definitions and preliminary results.** Let $V$ be a Hilbert space; we denote by $(\cdot,\cdot)$ and $\|\cdot\|$ a general inner product on $V$ and its associated norm. Let $C$ be a nonempty subset in $V$ such that

(A1) $C$, is a cone, that is, if $v \in C$ then $\lambda v \in C$ for all $\lambda \geq 0$, and

(A2) $C$ is weakly closed in $V$.

EXAMPLE 2.1. *Clearly, every closed and convex cone in $V$ satisfy (A1) and (A2).*

EXAMPLE 2.2. *Falcó and Hackbusch [1] proved that for each $\mathbf{r} \in \mathbb{N}^d$ the set $\mathcal{T}_{\mathbf{r}}$, of tensors in Tucker format which is a non-convex cone, is a weakly closed set in any tensor Banach space with a norm not weaker than the injective norm.*

EXAMPLE 2.3. *Let us consider the non-convex cone*

$$C = \left\{u \in L_2[0,1] : u(x) = \alpha x^{\beta} \text{ where } \alpha \geq 0 \text{ and } \beta \in [0,2]\right\}.$$

*Then the map $\Phi : \mathbb{R}_+ \times [0,2] \to L_2[0,1]$, given by $\Phi(\alpha,\beta)(x) = \alpha x^{\beta}$, is continuous because*

$$\|\Phi(\alpha,\beta) - \Phi(\alpha',\beta')\|_{L_2[0,1]}^2 \leq \frac{(\alpha-\alpha')^2}{\min(\beta,\beta')^2+1}.$$

*Now, assume that $\{u_n(x) = \alpha_n x^{\beta_n}\}_{n\in\mathbb{N}} \subset C$ weakly converges to $u$ in $L_2[0,1]$. Then $\{u_n\}_{n\in\mathbb{N}}$ is bounded:*

$$\|u_n\|_{L_2[0,1]}^2 = \frac{\alpha_n^2}{\beta_n^2+1} \leq C$$

*for some $C \geq 0$ and for all $n \in \mathbb{N}$. As a consequence, $\{(\alpha_n,\beta_n)\}_{n\in\mathbb{N}}$ is a bounded sequence in the closed set $\mathbb{R}_+ \times [0,2]$. Then, there exists a convergent subsequence, also denoted by $\{(\alpha_n,\beta_n)\}_{n\in\mathbb{N}}$, to some $(\alpha,\beta) \in \mathbb{R}_+ \times [0,2]$.*

*Since*

$$\lim_{n\to\infty} \|\Phi(\alpha_n,\beta_n) - \Phi(\alpha,\beta)\|_{L_2[0,1]} = 0,$$

*we have that $u_n - v$, where $v(x) = \alpha x^{\beta}$, converges to zero in $L_2[0,1]$. Thus $v = u$, and $C$ is weakly closed in $L_2[0,1]$.*

Now we want to characterize a projection on $C$ with respect to a given inner product $(\cdot,\cdot)$ on $V$, with associated norm $\|\cdot\|$.

A $C$-projection with respect to inner product $(\cdot,\cdot)$, with associated norm is a map $\Pi(\cdot|C) : z \in V \mapsto \Pi(z|C) \subset C$ defined by

$$\Pi(z|C) = \arg\min_{v\in C} \|z-v\|^2. \tag{2.1}$$

Let be the map $\sigma(\cdot|C) : V \to \mathbb{R}$ defined by

$$\sigma(z|C) = \max_{\substack{w\in C\\ \|w\|=1}} |(z,w)|, \tag{2.2}$$

The following Proposition 2.4 proves that (A3) is a sufficient condition on the inner product $(\cdot,\cdot)$ for the maps $\Pi(\cdot|C)$ and $\sigma(\cdot|C)$ be well defined.

PROPOSITION 2.4. *For each $z \in V$, there exists $v^* \in C$ such that*

$$\|z - v^*\|^2 = \min_{v\in C} \|z - v\|^2 = \|z\|^2 - \sigma(z|C)^2. \tag{2.3}$$

*Moreover, $\sigma(z|C) = \|v^*\|$, and*

$$(z - v^*, v^*) = 0. \tag{2.4}$$

*Proof.* Let us choose any sequence $w_n \in C$ with $\|z - w_n\| \searrow \inf\{\|z - w\| : w \in C\}$. Since $(w_n)_{n\in\mathbb{N}}$ is a bounded sequence in $V$, there exists a subsequence $w_{n_i} \rightharpoonup v \in V$. $v$ belongs to $C$ because $w_{n_i} \in C$ and $C$ is weakly closed. Since also $z - w_{n_i}$ weakly converges to $x - v$, then $\|z - v\| \le \liminf \|z - w_{n_i}\| \le \inf\{\|z - w\| : w \in C\}$. Thus, there exists $v^* \in C$ such that $\|z - v^*\|^2 = \min_{v\in C} \|z - v\|^2$. The second equality of (2.3) follows from

$$\min_{v\in C} \|z - v\|^2 = \min_{\substack{\lambda\in\mathbb{R}_+, w\in C\\ \|w\|=1}} \|z - \lambda w\|^2 \tag{2.5}$$

$$= \min_{\lambda\in\mathbb{R}_+,} \|z\|^2 - 2\lambda(w, z) + \lambda^2 \tag{2.6}$$

$$= \min_{\substack{w\in C\\ \|w\|=1}} \|z\|^2 - (z, w)^2 \tag{2.7}$$

$$= \|z\|^2 - \max_{\substack{w\in C\\ \|w\|=1}} (z, w)^2 \tag{2.8}$$

$$= \|z\|^2 - \sigma(z|C)^2. \tag{2.9}$$

To prove the second part we consider the equality

$$\frac{1}{2}(v, v) - (z, v) = \frac{1}{2}\|z - v\|^2 - \frac{1}{2}\|z\|^2.$$

This implies that for $J_z(v) = \frac{1}{2}(v, v) - (z, v)$ the minimization problem

$$J_z(v^*) = \min_{v\in C} J_z(v). \tag{2.10}$$

is equivalent to

$$\min_{v\in C} \|z - v\|^2, \tag{2.11}$$

and

$$\min_{v\in C} J_z(v) = \frac{1}{2} \min_{v\in C} \|z - v\|^2 - \frac{1}{2}\|z\|^2. \tag{2.12}$$

If $z = 0$ then $v^* = 0$ and the theorem clearly holds. Now, assume that $z \neq 0$. From (2.12) and (2.8) we deduce

$$\min_{v\in C} J_z(v) = -\frac{1}{2} \max_{\substack{w\in C\\ \|w\|=1}} (z, w)^2. \tag{2.13}$$

Thus, $v^* \in C$ solves (2.10) if and only if $v^* = \sigma(z|C)w^*$ for some $w^* \in C$ with $\|w^*\| = 1$. Therefore, the first statement follows. To prove the second one, from (2.13) follows

$$J_z(v^*) = -\frac{1}{2}\sigma(z|C)^2 = -\frac{1}{2}\|v^*\|^2, \tag{2.14}$$

and by using (2.12) we obtain (2.3). Finally, from (2.14) we have that

$$(v^*, v^*) - (z, v^*) = 0,$$

and this follows (2.4). □

A first consequence is the following.

COROLLARY 2.5. *The map $\sigma(\cdot|C)$ defines a seminorm on $V$.*

**3. Main Result.** From now on, we will denote by $U(C) = \overline{\operatorname{span} C}^{\|\cdot\|}$ the closed linear subspace generated by $C$. Now, we introduce the set

$$\mathcal{V}(z|C) = \{w \in C : \|w\| = 1 \ \text{ and } \ \sigma(z|C) = |(z, w)|\}. \tag{3.1}$$

Then the projector $\Pi(\cdot|C)$ can be written as

$$\Pi(z|C) = \sigma(z|C)\mathcal{V}(z|C), \tag{3.2}$$

which means that for $v^* \in \Pi(z|C)$, there exists $w^* \in \mathcal{V}(z|C)$ such that $v^* = \sigma(z|C)w^*$.

Also Proposition 2.4 allows to construct a sequence $\{e_n\}_{n\geq 0} \subset V$ by means of the following iterative scheme. Let $z_0 = 0$, and, for each $n \geq 1$, take

$$e_{n-1} = z - z_{n-1}, \ \text{and update} \tag{3.3}$$
$$z_n = z_{n-1} + z^{(n)} \ \text{where} \ z^{(n)} \in \Pi(e_{n-1}|C). \tag{3.4}$$

Observe that for $n \geqslant 1$,

$$z_n = \sum_{i=1}^{n} z^{(i)}, \quad z^{(i)} \in \Pi(z - z_{i-1}|C) \tag{3.5}$$

or, equivalently, by using Proposition 2.4,

$$z_n = \sum_{i=1}^{n} \sigma(e_{i-1}|C)w^{(i)}, \quad w^{(i)} \in \mathcal{V}(e_{i-1}|C). \tag{3.6}$$

We introduce the following definition of the $C$- *rank*

DEFINITION 3.1. *We define the $C$-rank of an element $z \in V$, denoted by* $\operatorname{rank}(z|C)$*, as follows:*

$$\operatorname{rank}(z|C) = \min\{n : \sigma(e_n|C) = 0\}, \tag{3.7}$$

*where by convention* $\min(\emptyset) = \infty$.

Now, we state the main result of this paper:

THEOREM 3.2. *For $z \in V$, the sequence $\{e_n\}_{n\geqslant 0}$ constructed in (3.3) satisfies that* $\lim_{n\to\infty} e_n = e^*$ *and $e^* \in U(C)^\perp$. Moreover,*

$$P_{U(C)}(z) = z - e^* = \sum_{i=1}^{\operatorname{rank}(z|C)} \sigma(e_{i-1}|C)w^{(i)},$$

*where $P_{U(C)}$ is the orthogonal projection over $U(C)$, and*

$$\|e_n\|^2 = \|z\|^2 - \sum_{i=1}^{n} \sigma(e_{i-1}|C)^2 = \sum_{i=n+1}^{\operatorname{rank}(z|C)} \sigma(e_{i-1}|C)^2.$$

*In consequence,*

$$\|z - P_U(z)\|^2 = \|z\|^2 - \sum_{i=1}^{\operatorname{rank}(z|C)} \sigma(e_{i-1}|C)^2.$$

*Proof.* In order to simplify notation, in this proof we will use $\sigma_i = \sigma(e_{i-1}|C)$, for all $i \geq 0$. Let us first note that it holds for $1 \leq n \leq \operatorname{rank}(z|C)$ that $z^{(n)} \neq 0$ since for such $n$, $\sigma(z - z_{n-1}|C) > 0$ by definition of $C$-rank. We have

$$\begin{aligned} \|e_n\|^2 &= \|e_{n-1} - z^{(n)}\|^2 && (3.8)\\ &= \|e_{n-1}\|^2 - \|z^{(n)}\|^2 \quad \text{(by using (2.3))} && (3.9)\\ &= \|e_{n-1}\|^2 - \sigma_n^2 && (3.10) \end{aligned}$$

Thus $\{\|e_n\|\}_{n=0}^{\operatorname{rank}(z|C)}$ is a strictly decreasing sequence of non-negative real numbers.

We first assume that $\operatorname{rank}(z|C) = r < \infty$. Then, $\sigma_r = \sigma(z - z_r|C) = 0$ and $z^{(r+1)} = 0$ since

$$\|z - z_r - z^{(r+1)}\|^2 = \|z - z_r\|^2 - \sigma_r^2 = \|z - z_r\|^2$$

We have

$$\|z - z_r\|^2 = \min_{v \in C} \|z - z_r - v\|^2 \leq \|z - z_r - \lambda v\|^2$$

for all $\lambda \in \mathbb{R}$ and $v \in C$. This implies that

$$(z - z_r, v) = 0$$

for all $v \in C$. Thus $z - z_r \in U^{\perp}$ and the first statement of the theorem follows.
On the other hand, we assume that $\operatorname{rank}_{\sigma}(z) = \infty$. Then $\{\|e_n\|\}_{n=0}^{\infty}$ is a strictly decreasing sequence of non-negative real numbers, and there exists

$$\lim_{n\to\infty} \|e_n\| = \lim_{n\to\infty} \|z - z_n\| = R \geq 0.$$

Proceeding from (3.10) and using that $e_0 = z$, we obtain

$$\|e_n\|^2 = \|z\|^2 - \sum_{k=1}^{n} \sigma_k^2. \tag{3.11}$$

In consequence, $\sum_{k=1}^{\infty} \sigma_k^2$ is a convergent series and $\lim_{n\to\infty} \sigma_n^2 = 0$. Thus, we obtain also

$$\lim_{n\to\infty} \sigma_n = \lim_{n\to\infty} \|z^{(n)}\| = 0. \tag{3.12}$$

For all $n \geq 1$ and $v \in C$ with $\|v\| = 1$, we have

$$(e_{n-1}, v)^2 \leq \max_{w \in C:\, \|w\|=1} (e_{n-1}, w)^2 = \sigma_n^2 \tag{3.13}$$

and then

$$\lim_{n\to\infty} (e_{n-1}, v)^2 = 0 \tag{3.14}$$

Assuming that $\{e_n\}_{n=0}^{\infty}$ is convergent in the $\|\cdot\|$-norm to some $e^* \in V$, since the sequence is also weakly convergent to $e^*$ we obtain from (3.14) that

$$(e^*, v) = 0$$

for all $v \in C$ with $\|v\| = 1$. Thus, $e^* \in U^\perp$. To conclude the proof we only need to show that $\{e_n\}_{n=1}^{\infty}$ is a Cauchy sequence in $V$ in the $\|\cdot\|$-norm. The following lemmas will be useful.

LEMMA 3.3. *For each $n, m \geq 1$, it follows that*

$$\left|\left(e_{m-1}, z^{(n)}\right)\right| \leq \sigma_m \sigma_n$$

*Proof.* We have

$$\left|\left(e_{m-1}, z^{(n)}\right)\right| = \left|\left(e_{m-1}, \sigma_n w^{(n)}\right)\right| = \left|\left(e_{m-1}, w^{(n)}\right)\right| \sigma_n \leq \sigma_m \sigma_n$$

where we have used

$$\sigma_m = \left|\left(e_{m-1}, w^{(m)}\right)\right| = \max_{w\in C:\|w\|=1} |(e_{m-1}, w)| \geq \left|\left(e_{m-1}, w^{(n)}\right)\right|,$$

□

LEMMA 3.4. *For every $\varepsilon > 0$ and every $N \in \mathbb{N}$ there exists $\tau \geq N$ such that*

$$\sigma_\tau \sum_{k=1}^{\tau} \sigma_k \leq \varepsilon. \tag{3.15}$$

*Proof.* Since $\sum_{j=1}^{\infty} \sigma_j^2 < \infty$, for a given $\varepsilon > 0$ and $N \in \mathbb{N}$, we choose $n \geq N$ such that

$$\sum_{j=n+1}^{\infty} \sigma_j^2 \leq \varepsilon/2$$

Since $\lim_{j\to\infty} \sigma_j = 0$, we construct $\tau : \mathbb{N} \longrightarrow \mathbb{N}$ defined inductively by $\tau(1) = 1$ and for all $k \geq 1$,

$$\tau(k+1) = \min_{j>\tau(k)} \left\{\sigma_j \leq \sigma_{\tau(k)}\right\},$$

such that $\tau$ is strictly increasing and $\lim_{k\to\infty} \tau(k) = \infty$. Observe that for all $k \geq 1$ and $j$ satisfying $\tau(k) \leq j < \tau(k+1)$, it follows that

$$\sigma_{\tau(k+1)} \leq \sigma_{\tau(k)} \leq \sigma_j.$$

Thus, for all $1 \leq j < \tau(k+1)$, we have

$$\sigma_{\tau(k+1)} \leq \sigma_j$$

Now, since $\lim_{k\to\infty} \sigma_{\tau(k)} = 0$, we can choose $\tau = \tau(k+1) > n$ large enough satisfying

$$\sigma_\tau \sum_{j=1}^{n} \sigma_j \le \varepsilon/2.$$

Then

$$\begin{aligned}\sigma_\tau \sum_{j=1}^{\tau} \sigma_j = \sigma_\tau \sum_{j=1}^{n} \sigma_j + \sigma_\tau \sum_{j=n+1}^{\tau} \sigma_j \quad &\le \quad \varepsilon/2 + \sigma_\tau \sum_{j=n+1}^{\tau} \sigma_j \\ \le \varepsilon/2 + \sum_{j=n+1}^{\tau} \sigma_j^2 \quad &\le \quad \varepsilon/2 + \sum_{j=n+1}^{\infty} \sigma_j^2 \\ \le \varepsilon \end{aligned}$$

This proves the lemma. □

LEMMA 3.5. *For all $M > N > 0$, it follows that*

$$\|e_{N-1} - e_{M-1}\|^2 \le \|e_{N-1}\|^2 - \|e_{M-1}\|^2 + 2\,\sigma_M \sum_{k=1}^{M} \sigma_k$$

*Proof.* We have

$$\begin{aligned}\|e_{N-1} - e_{M-1}\|^2 &= \|e_{N-1}\|^2 + \|e_{M-1}\|^2 - 2\,(e_{M-1}, e_{N-1}) \\ &= \|e_{N-1}\|^2 + \|e_{M-1}\|^2 - 2\left(e_{M-1}, e_{M-1} + \sum_{k=N}^{M-1} z^{(k)}\right) \\ &= \|e_{N-1}\|^2 - \|e_{M-1}\|^2 - 2\sum_{k=N}^{M-1}\left(e_{M-1}, z^{(k)}\right) \\ &\le \|e_{N-1}\|^2 - \|e_{M-1}\|^2 + 2\,\sigma_M \sum_{k=N}^{M-1} \sigma_k \quad \text{(by using Lemma 3.3)} \\ &\le \|e_{N-1}\|^2 - \|e_{M-1}\|^2 + 2\,\sigma_M \sum_{k=1}^{M} \sigma_k \quad \text{(by adding positive terms.)}\end{aligned}$$

This ends the proof of lemma. □

Since the limit of $\|e_n\|^2$ goes to $R^2$ as $n \to \infty$, and it is a decreasing sequence, for a given $\varepsilon > 0$ there exists $k_\varepsilon > 0$ such that

$$R^2 \le \|e_{m-1}\|^2 \le R^2 + \varepsilon^2/2$$

for all $m > k_\varepsilon$. Now, we assume that $m > k_\varepsilon$. From Lemma 3.4, for each $m+p$ there exists $\tau > m+p$ such that

$$\sigma_\tau \sum_{k=1}^{\tau} \sigma_k \le \varepsilon^2/4.$$

Now, we would to estimate

$$\|e_{m-1} - e_{m+p-1}\| \le \|e_{m-1} - e_{\tau-1}\| + \|e_{\tau-1} - e_{m+p-1}\|.$$

By using Lemma 3.5 with $M = \tau$ and $N = m$ and $m + p$, we obtain that

$$\|e_{m-1} - e_{\tau-1}\|^2 \leq R^2 + \varepsilon^2/2 - R^2 + \varepsilon^2/2 = \varepsilon^2,$$

and

$$\|e_{m+p-1} - e_{\tau-1}\|^2 \leq R^2 + \varepsilon^2/2 - R^2 + \varepsilon^2/2 = \varepsilon^2,$$

respectively. In consequence $\{e_n\}_{n=0}^{\infty}$ is a Cauchy sequence in the $\|\cdot\|$-norm and it converges to $e^*$. □

**4. An Application to a Fractional Blind Deconvolution Model.** The above results provide a theoretical ground for a large class of nonlinear approximation problems. In this section, we will illustrate this through a particular example modelling blind deconvolution, introduced in [2]. In order to self completeness, we outline the main aspects of the model and refer the interested reader to [2] for details.

**4.1. The Blind Deconvolution Problem.** It is well known that an image $u(x, y)$ gets degraded due to different causes, which can usually be mathematically formulated as follows:

$$u_0(x, y) = (K * u)(x, y) + n(x, y) \tag{4.1}$$

where $K(x, y)$ is an operator representing the deterministic degrading of the image, and $n(x, y)$, the stochastic additive errors (noise). In this section we are going to consider a linear and shift invariant convolution operator, defined as usual:

$$(K * u)(x, y) = \int_{\mathbb{R}^2} K(x - \alpha, y - \beta) \mathrm{d}\alpha \mathrm{d}\beta$$

For simplicity, we will consider *pure* convolution models (i.e., without noise, $n(x, y) = 0$).

Physical convolution features, such as blurring or diffusion, come from smooth operator kernels, $K$.

Deconvolution problems consist of recovering the original image $u$ from the convolved, observed, one ($u_0$). The problem should be solved in the context of Fourier transforms, due to the fundamental theorem of convolution:

$$\widehat{K * u}(\xi, \eta) = \widehat{K}(\xi, \eta)\widehat{u}(\xi, \eta)$$

A naive way to deconvolve is thus to obtain $\widehat{u}$ by a simple division. In practice, regularity of $K$ implies that its Fourier transform decays fast, and this direct deconvolution is unstable, not allowing the recovery of high frequences of $\widehat{u}$). A regularizing term must be included in order to stabilize the problem.

The problem gets even more complicated when the kernel $K$ is not known (blind deconvolution). We must worry about the stability, as before, but we also have to estimate the kernel. The model to analize, which we explain here, makes the assumption that the kernel is a Gaussian. This is not a risked assumption because blurring is produced by that kind of kernels). Fourier transform of Gaussians are also Gaussians, and, hence:

$$\widehat{K}(\xi,\eta) = c\mathrm{e}^{-\gamma(\xi^2+\eta^2)} := \widehat{G_\gamma}(\xi,\eta)$$

being $c$ a constant normalizing the function, and $\gamma$ a positive parameter related to the amount of diffusion. A direct approach to the problem is to fit $\gamma$. However, once again, this naive approach is wrong. The problem is extremely ill posed, because there exist infinite solutions: any Gaussian with a diffusion smaller than $\gamma$ ($G_\lambda$ for $\lambda \leq \gamma$) is also a convolution kernel of the image, and, at least, any $w = G_\lambda * u$ is a solution of the deconvolution problem. There are many other solutions. In fact, additional conditions must be required in order to, at least, expect for a good resolution.

In any case, it is not a right strategy to look for the parameter $\gamma$. In the first place, the own image can distort this parameter. On the other side, knowledge of the kernel is directly related to the amount of regularization needed: the most accuracy for the kernel implies the least amount of frequences recovered.

**4.2. The Fractional Deconvolution Model.** The above remarks lead us to consider a model based on an iterative fractional decomposition of the kernel. Decomposition will be obtained by logarithmic approximation, as follows:

$$\widehat{u_0}(\xi,\eta) = \exp(-\gamma(\xi^2+\eta^2)) \cdot \widehat{u}(\xi,\eta)$$

Thus,

$$\log(|\widehat{u_0}(\xi,\eta)|) \approx -\gamma(\xi^2+\eta^2) + \log(|\widehat{u}(\xi,\eta)|)$$

Assumptions on the image (see [2]) let us fix one of the frequences, for instance, $\eta = 0$, and the function $v(\xi) = \log(|\widehat{u}(\xi,0)|)$ is decreasing. By radial symmetry, $u$ can be recovered from $v$. Analogously, we denote $v_0(\xi) = \log(|\widehat{u_0}(\xi,0)|)$. Without loss of generality, let us assume that $v_0(\xi) = 0$ (in any other case, the arguments below are valid for $w(\xi) = v_0(\xi) - v_0(0)$.

Among all the possible choices of $v$ we select the one given by the following decomposition:

$$v_0(\xi) = v(\xi) - h(\xi)\ ;\ (h(\xi) = \sum_j \alpha_j |\xi|^{\beta_j} \tag{4.2}$$

where $\alpha_j$ and $0 < \beta_j \leq 2$ are such that $h(\xi)$ is the optimal approximation (the projection) to $v_0$ in the space generated by the dictionary:

$\xi^\beta$, with $0 < \beta \leq 2$. Therefore, we are in the setting establlished in example (2.3). Of course, nonlinearity comes from the powers $\beta$, and this is the main difficulty to fit the problem in the approximation frame.

This is equivalent to decompose the kernel $G_\gamma = G_{\alpha_1}^{\beta_1/2} * G_{\alpha_2}^{\beta_2/2} * \cdots * G_{\alpha_k}^{\beta_k/2} * \cdots$, where $G_{\alpha_k}^{\beta_k/2}$ are quasigaussian kernels (those whose Fourier transform is a fractional power of the Gaussian).

**4.3. Convergence of the Algorithm.** Leaving aside stability and regularization aspects (that could be inclded in a wider class of methods known as *weak greedy* ones), the actual algorithm is as follows:

1. Pivot $v_0$, $h_0 = 0$.
2. Given $v_k$, $h_k$, get the projection of $v_k$ on the cone, finding the coefficients $\alpha_k$ and $\beta_k$.
3. Deconvolve $u_k$ from $G_{\alpha_j}^{\beta_j}$, which is equivalent to obtain $v_{k+1}(\xi) = v_k(\xi) - \alpha_k|\xi|^{\beta_k}$; $h_{k+1} = h_k + \alpha_k|\xi|^{\beta_k}$.
4. Stop the process when $||v_{k+1}|| > ||v_k||$, or $||\alpha_{k+1}|\xi|^{\beta_{k+1}} - \alpha_k|\xi|^{\beta_k}|| <$ tolerance.
5. Return $v = v_k$, $h = h_k$.

The limit case (tolerance= 0) leads to two possibilities: if the process is finite, the projection $v$ lays in the space generated by the cone; in the infinite case, it is in the closure of such space. In any of these cases, the theoretical results in the previous section show that this projection exists and it is the desired optimal approximation. As a late remark we remark that $v$ is the residual of the projection, and it consists of the part of the function that cannot be expressed by fractional powers of $\xi$. Thus, the function $v$ is the least regular of the possible ones, and the corresponding image $u$ is the one with most edges among all of the solutions of the deconvolution problem.

Although deconvolution techniques at each step are linear in the sense that they are not adaptive (do not depend on the image) and they are computed *via* fast Fourier transforms, the overall algorithm is nonlinear and adaptive. Decomposition does not depend only on the kernel, but on the observed image. In the following section, we will remark these features of the algorithm.

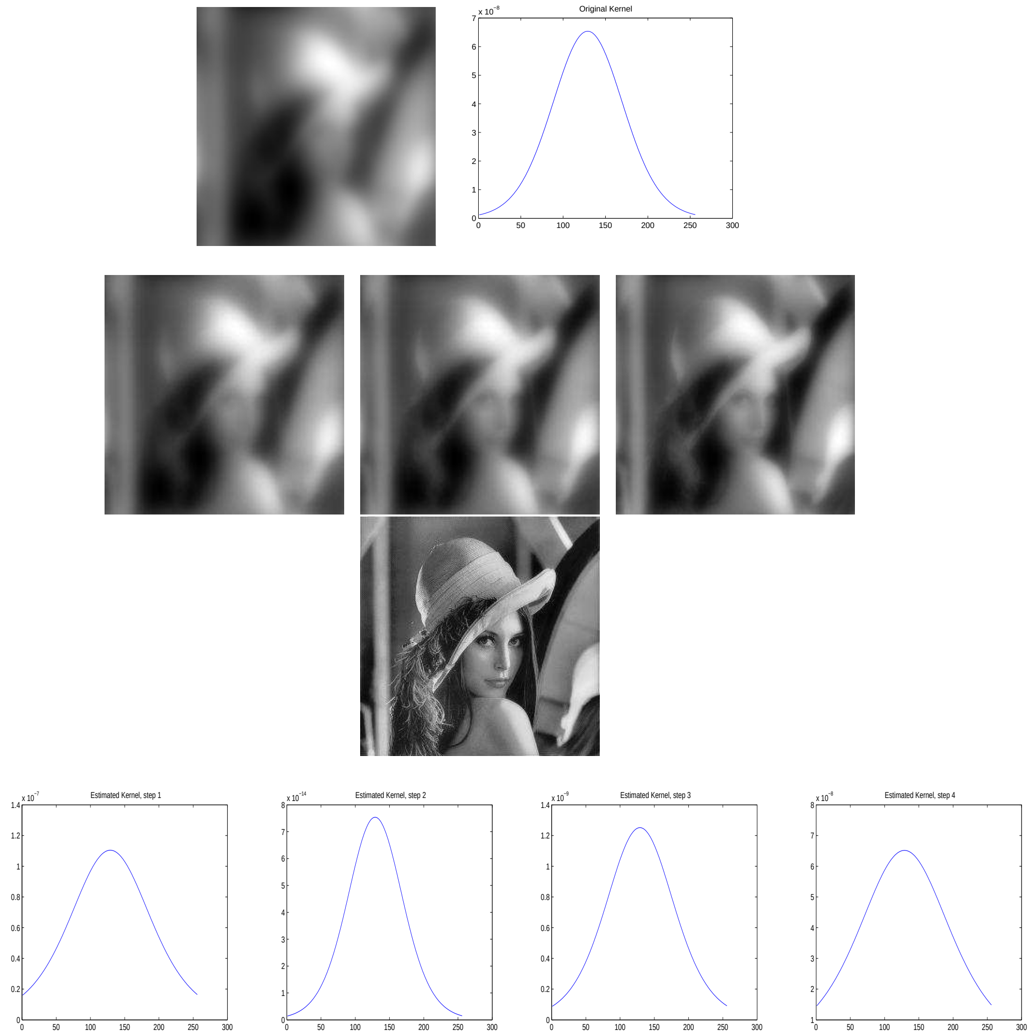
**5. Examples.** In this section we will display some examples in order to illustrate the above results.

The first three figures show the deconvolution process: the detected quasi-Gaussian kernels (below) and their corresponding partial deconvolved images (above). The first image is the original (blurred) one, and the blurring kernel which has to be detected. In the shown examples, good quality deconvolved image is obtained after three or four steps, even for kernels with a large variance. The deconvolved image recovers many of the original details which cannot be recovered by other models (we refer the reader to [2, 5] for comparison purposes).

Adaptiveness of the process is shown in the first two figures: the blurring kernel is the same for both images, but it can be seen that the decomposition of the kernel is different in both cases, due to the features of the original, blurred, image. Fractional regularity of both images is not the same, and the model detects different powers. We may see that, in Lena image, details are recovered in a more gradual way than in the satellite, because the first one is richer in texture than the other one.

The third example shows that dependence on the kernel is not linear: it is interesting to observe that different diffusion parameters ($\gamma$) give place to different decomposition, though it could seem natural to think that, as $G_\gamma * G_\rho = G_{\gamma+\rho}$, the decompositions should follow a similar sequence. As it happens, the kernel in the second example is $G_\gamma * G_\gamma$ (being $G_\gamma$ the blurring kernel in the third example). Nevertheless, the decomposition in the second example does not relates (at least directly) to the decomposition in the third example.

Finally, we show a, non academic, real life application (restoration of a baroque painting). It is important to ensure that fractional decomposition (that is, fractional projection) works in a multichannel context as it is that of color images (see [4]). It is also interesting to remark that, in this example, as it is a real one, the only assumption we can make is to consider the image was blurred by natural causes, and, then, the kernel is a Gaussian or quasi-Gaussian blur. Let us notice that the image shown is very spoiled and it has some other added difficulties besides blurring. Most other models are not able to obtain good deconvolution because the scratching of the image. As we see, in our fractional model, the image can be deconvolved while keeping this scratches, which is often important in order to apply particular and local models for arranging them.




**6. Conclusion and Final Remarks.** In this paper we presented some theoretical results for nonlinear approximation which is underlying many greedy algorithms. In particular, our results show an alternative to convexity, which is a property often lost in the nonlinearity context.

In order to illustrate the strength of the results, we displayed an example related to image restoration. In general, image restoration and denoising are situations where this theory applies: most models consist of finding good approximations to the image under some restrictions. In the shown example, the theory is not only useful to prove consistence and existence of the deconvolution, but it also explains some features of the deconvolved image (recovery of edges, for instance).

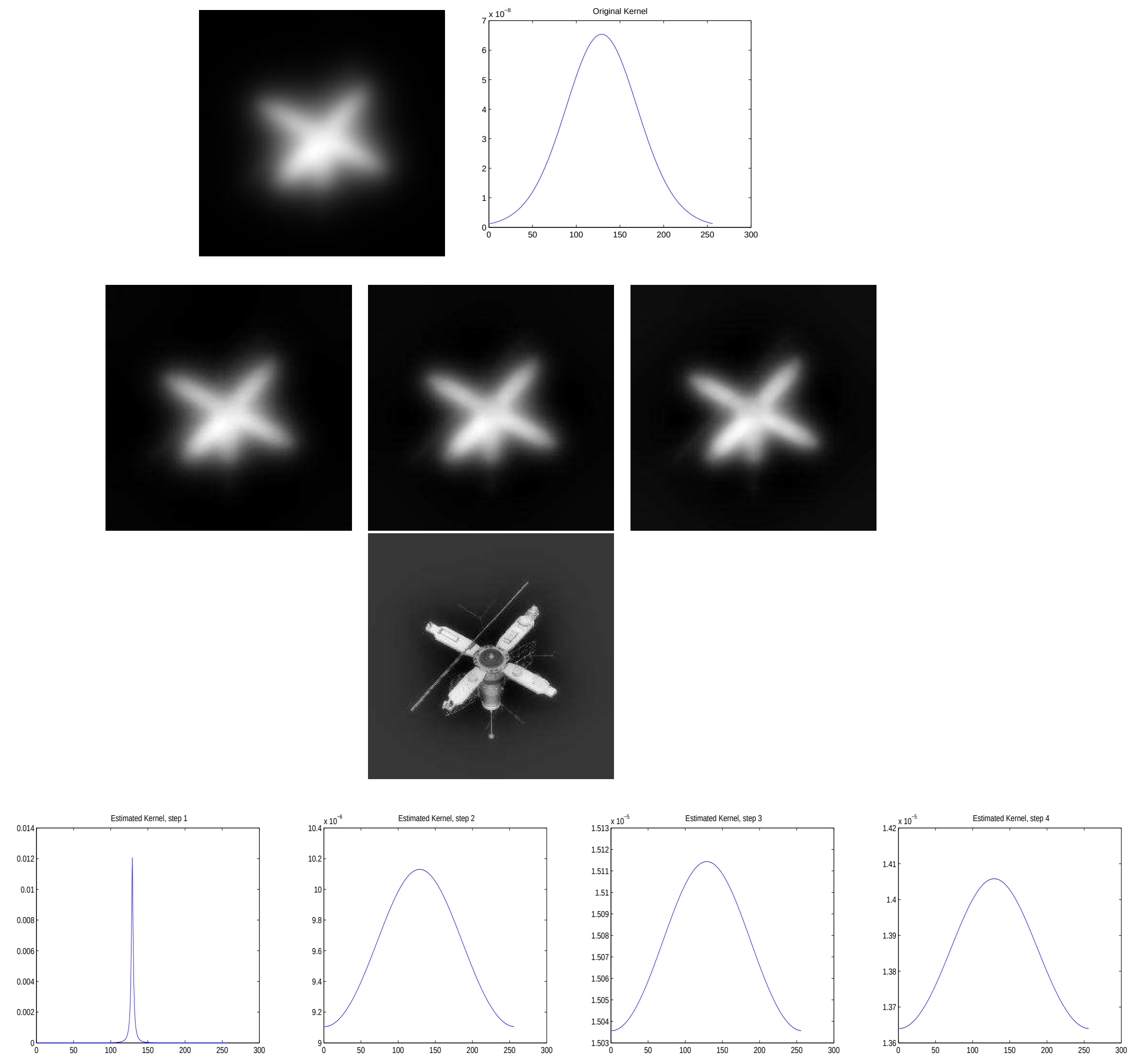



Fig. 5.1: Top: Blurred image and kernel (satellite). Bottom: Blind deconvolution sequence with the corresponding detected kernels.

The computational examples show how partial (iterative) projections work: the first steps are smooth and it is in the last ones, which correspond to the lower exponents when details are obtained. The observed image is thus decomposed in a quasigaussian kernel and versions of deconvolutions with different degrees of smoothing (exponents). The model, hence, propose a frequency-regularity analysis of the observed image, which is a nonlinear multiresolution scheme.

The fractional model introduced in this paper also shows some of the nonlinear features one can wait: adaptiveness to the initial conditions, and joint detection of both the deconvolved image and the blurring kernel. Such as it is proposed, the deconvolved image is the residual of the process. Thus, when the blurred image is

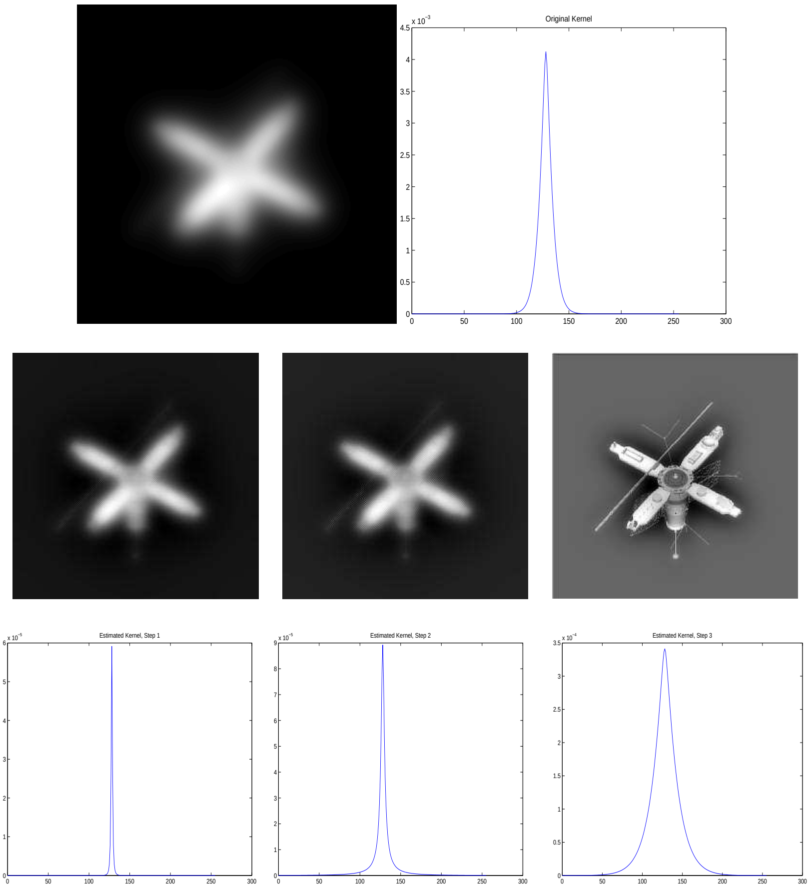



Fig. 5.2: Top: Blurred image and kernel (satellite). Bottom: Blind deconvolution sequence with the corresponding detected kernels. (40 pasos de la ecuación del calor)

in the subspace we are projecting, the deconvolution we obtain is just the null image (a black one). In other words, in general, the image we obtain is the one with most details, among all the possible ones.

REFERENCES

[1] A. Falcó, W. Hackbusch. On minimal subspaces in tensor representations, to appear in Found. Comput. Math. available as *Preprint 70/2010* Max Planck Institute for Mathematics in the Sciences 2010.

[2] P.D. Romero, V.F. Candela, Blind deconvolution models regularized by fractional powers of the Laplacian, J. Math Imaging Vis., 32, 181-191 (2008).

[3] P.D. Romero, V.F. Candela, Mathematical models for restoration of Baroque paintings, Lecture Notes in Computer Sciences, 4179, 24-34 (2006).

[4] P.D. Romero, V.F. Candela, Color Preserving Fractional Models for Blind Deconvolution, submitted

[5] P. D. Romero, V.F. Candela. Modelos de deconvolución ciega fraccionaria. Aplicaciones a la restauración de obras pictóricas. Servici de Publicacions de la Universitat de Valéncia, 2009, ISBN: 978-84-370-7562-4.

[6] R. Hilfer (ed.), Applications of fractional calculus in physics, World Scientific, River Edge (2000).
[7] A. Carasso, Direct blind deconvolution, SIAM J. Numer. Anal., 61, 1980-2007 (2001).
[8] S. Didas, B. Burgeth, A. Imiya, J. Weickert, Regularity and scale-space properties of fractional high order linear filtering, Lecture Notes in Computer Sciences, 3459, 13-25,(2005).
[9] T. Chan, C.K. Wong, Total variation blind deconvolution, IEEE Trans. Image Process, 7(3), 370-375 (1998).
[10] L. Rudin, S. Osher, Total variation based image restoration with free local constraints, Proc. IEEE Int. Conf. Image Process., 31-35 (1994).
[11] L. He, A. Marquina, S. Osher, Blind deconvolution using TV regularization and Bregman iteration. Int. J. Imag. Systems and Techn., 15, 74-83 (2005).

(a) Original image: Epiphany

(b) Fractional Blind Deconvolution

Fig. 5.3: Color Fractional Blind deconvolution. Theme: Epiphany, from the altarpiece of Saint Bartholomew Church, Bienservida (Spain)