A General Framework for a Class of Non-linear Approximations with Applications to Image Restoration

V. Candela[a,2], A. Falcó[b,1], Pantaleón D. Romero[1,1,2,*]

[a] *Departament de Matemátiques, Campus de Burjassot, Universidad de Valencia, Carrer del Dr. Moliner 50, 46100 Burjassot (Valencia), Spain*

[b] *Departamento de Departamento de Matemáticas, Física y Ciencias Tecnológicas, Universidad CEU Cardenal Herrera, Carrer San Bartolomé 55, 46115 Alfara del Patriarca (Valencia), Spain.*

Abstract

In this paper, we establish sufficient conditions for the existence of optimal non-linear approximations to a linear subspace generated by a given weakly-closed (non-convex) cone of a Hilbert space. Most non-linear problems have difficulties to implement good projection-based algorithms due to the fact that the subsets, where we would like to project the functions, do not have the necessary geometric properties to use the classical existence results (such as convexity, for instance). The theoretical results given here overcome some of these difficulties. To see this we apply them to a fractional model for image deconvolution. In particular, we reformulate and prove the convergence of a computational algorithm proposed in a previous paper by some of the authors. Finally, some examples are given.

Keywords: Non-linear approximation, Fractional Deconvolution, Image Restoration, Weakly-closed non-convex cone.

1. Introduction

Unlike linear approximation, where there exists a solid theoretical background establishing conditions for existence, uniqueness and algorithmic issues, non-linear approximation is a field with not so deep knowledge about the above related concepts. The main drawback of the non-linear approximation theory is given by the fact that most of the fundamental concepts used in the theory of linear spaces cannot be generalized without losing their strength. Lack of the vectorial structure of the spaces is in fact one of the main difficulties to obtain

* Corresponding author
[1] Supported by grant PRCEU-UCH30/10 GVA PRE 2010/066, PRECEU-UCH 15/03
[2] Supported by grant GVA PRE 2010/066 and MTM 2088-03597

adequate results. There are also other features which are often lost, some of them are geometrical (convexity, for instance) and other are algebraic (the use of linear bases).

In the last years, sparsity helped to devise non-linear models by introducing the concept of *dictionaries* as generators of the space, and then sacrificing the concept of linear independence in the basis functions. The strategy of this kind of optimal approximation consists first in finding a good dictionary and next in searching a best approximation in this fitted subset. This is the principle underlying in the so called *greedy algorithms*.

However, a dictionary is not a useful structure to solve most geometric problems. Lack of convexity is usually one of these problems: usually dictionaries are not closed convex sets and hence results about the existence of best approximations are not available. In image processing non-convexity is not an unusual topic (see for example [8]). A recent result [5] provides an example of an optimisation framework over a class of dictionaries (weakly-closed non-convex cones in reflexive Banach spaces).

In this paper, we provide theoretical results related to a class of non-linear approximation problems with milder conditions to give a better theoretical foundation in order to relax conditions for a class of nonlinear approximation problems. These conditions are verified in many practical problems.

Image processing, and, particularly, image restoration is one of the fields where this theory can be applied. Both deconvolution and denoising of images are better dealt with adaptive non-linear models than linear ones (leaving aside computational aspects). In this context *adaptative* means that, besides the different type of convolution kernels and noises, the process shall depend on the known initial data, which are the own image features [1, 2, 3, 12, 6]. In [9, 11], 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 the image itself. Though, in practice, this model converges with high quality results, however, up to the authors knowledge no proof of its convergence is known. Our purpose in this work is to revisit the model introduced in [9, 11] in order to prove the convergence of the proposed algorithm [9, 11], namely ALFA, by using the theoretical results stated in §3.

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 theorem in the paper, which is shown in §3. In §4, we reformulate the blind deconvolution model given in [9, 11] an also we prove the convergence of the algorithm proposed in the aforementioned paper. We display also some examples of the application of the model in §5. Finally, we will draw conclusions in §6.

2. Definitions and preliminary results

In the following, V is a Hilbert space with inner product $(\cdot, \cdot)$ and associated norm $\|\cdot\|$. While V^* denotes the dual space of functionals with bounded dual norm $\|\cdot\|^*$:

$$\|\varphi\|^* = \sup\{|\varphi(v)| : v \in V \text{ with } \|v\| \leq 1\} = \sup\{|\varphi(v)| / \|v\| : 0 \neq v \in V\}.$$

Next we introduce the weak convergence of a sequence in V. We say that a sequence $(z_n)_{n\in\mathbb{N}}$ in a Hilbert space V *converges weakly to* $v \in V$, if $\lim \varphi(v_n) = \varphi(v)$ for all $\varphi \in V^*$. In this case, we write $v_n \rightharpoonup v$.

A subset $M \subset V$ is called *weakly closed*, if $z_n \in M$ and $z_n \rightharpoonup z$ implies $z \in M$. Note that 'weakly closed' is stronger than 'closed', i.e., M weakly closed $\Rightarrow M$ closed in the norm topology.

In this paper we focus on the following class of weakly-closed subsets of V. From now on, we consider $\mathcal{C}$ a non-empty subset in V such that

(A1) $\mathcal{C}$, is a cone, that is, if $v \in \mathcal{C}$ then $\lambda v \in \mathcal{C}$ for all $\lambda \geq 0$, and
(A2) $\mathcal{C}$ is weakly closed in V.

Example 1. *Clearly, every closed and convex cone in V satisfies (A1) and (A2).*

Example 2. *Falcó and Hackbusch [5] proved that for each $\mathbf{r} \in \mathbb{N}^d$ the set $\mathcal{T}_{\mathbf{r}}$, of tensors in Tucker format is a non-convex cone which is a weakly closed set in any tensor Banach space with a norm not weaker than the injective norm.*

Next, we will introduce a set $\mathcal{C}$ arising in the framework of the fractional deconvolution model introduced in [9, 11]. As we will show below, conditions (A1)-(A2) are the milestone to prove the convergence of a class of non-linear approximation algorithms. In particular, as we will see below, the so-called ALFA algorithm proposed in [9, 11] falls in the aforementioned class.

Let us consider the non-convex cone

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

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

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

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

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

for some $C \geq 0$ and for all $n \in \mathbb{N}$. Also the sequence $\{(\alpha_n, \beta_n)\}_{n\in\mathbb{N}}$ is bounded in the closed set $\mathbb{R}_+ \times [0,2]$. Hence 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 $\mathcal{C}$ is weakly closed in $L_2[0,1]$.

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

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

$$\Pi(z|\mathcal{C}) = \arg\min_{v\in\mathcal{C}} \|z - v\|^2. \tag{1}$$

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

$$\sigma(z|\mathcal{C}) = \max_{\substack{w\in\mathcal{C}\\ \|w\|=1}} |(z, w)|. \tag{2}$$

The following Proposition 1 proves that the maps $\Pi(\cdot|\mathcal{C})$ and $\sigma(\cdot|\mathcal{C})$ are well defined.

Proposition 1. *For each $z \in V$, there exists $v^* \in \mathcal{C}$ such that*

$$\|z - v^*\|^2 = \min_{v\in\mathcal{C}} \|z - v\|^2 = \|z\|^2 - \sigma(z|\mathcal{C})^2. \tag{3}$$

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

$$(z - v^*, v^*) = 0. \tag{4}$$

Proof. For each $z \in V$, we denote $\alpha := \inf\{\|z - w\| : w \in \mathcal{C}\} \geq 0$. Clearly, if $z \in \mathcal{C}$ then $\alpha = 0$ and (3) trivially holds. On the other hand, assume that $z \notin \mathcal{C}$ and hence $\alpha \geq 0$. Let us choose any sequence $w_n \in \mathcal{C}$ with $\|z - w_n\| \geq \|z - w_{n+1}\| > \alpha$ for all $n \in \mathbb{N}$ and

$$\lim_{n\to\infty} \|z - w_n\| = \alpha.$$

Since $(w_n)_{n\in\mathbb{N}}$ is a bounded sequence in V, from the local sequential weak compactness (see Chapter V.2. in [15]), there exists a weakly convergent subsequence $w_{n_i} \rightharpoonup v^* \in V$. Moreover, v^* belongs to $\mathcal{C}$ because $w_{n_i} \in \mathcal{C}$ and $\mathcal{C}$ is weakly closed. Since also $z - w_{n_i}$ weakly converges to $z - v^*$, as a consequence of the Banach-Steinhaus theorem (see Corollary 3.86 in [4]), we have

$$\|z - v^*\| \leq \liminf_{n\to\infty} \|z - w_{n_i}\| \leq \alpha.$$

Thus $v^* \in \mathcal{C}$ together $\|z - v^*\|^2 = \alpha = \min_{v\in\mathcal{C}} \|z - v\|^2$ prove the first equality in (3). The second equality in (3) follows from

$$\begin{aligned}
\min_{v\in\mathcal{C}} \|z-v\|^2 &= \min_{\substack{\lambda\in\mathbb{R}_+, w\in\mathcal{C}\\ \|w\|=1}} \|z-\lambda w\|^2 \\
&= \min_{\lambda\in\mathbb{R}_+,} \|z\|^2 - 2\lambda(w,z) + \lambda^2 \\
&= \min_{\substack{w\in\mathcal{C}\\ \|w\|=1}} \|z\|^2 - (z,w)^2 \\
&= \|z\|^2 - \max_{\substack{w\in\mathcal{C}\\ \|w\|=1}} (z,w)^2 \qquad (5)\\
&= \|z\|^2 - \sigma(z|\mathcal{C})^2.
\end{aligned}$$

To prove the second part we consider the equality

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

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

$$J_z(v^*) = \min_{v\in\mathcal{C}} J_z(v). \tag{6}$$

is equivalent to

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

and

$$\min_{v\in\mathcal{C}} J_z(v) \quad = \quad \frac{1}{2}\min_{v\in\mathcal{C}} \|z-v\|^2 - \frac{1}{2}\|z\|^2. \tag{7}$$

If $z = 0$ then $v^* = 0$ and the theorem clearly holds. Now, assume that $z \neq 0$. From (7) and (5) we deduce

$$\min_{v\in\mathcal{C}} J_z(v) = -\frac{1}{2} \max_{\substack{w\in\mathcal{C}\\ \|w\|=1}} (z,w)^2. \tag{8}$$

Thus, $v^* \in \mathcal{C}$ solves (6) if and only if $v^* = \sigma(z|\mathcal{C})w^*$ for some $w^* \in \mathcal{C}$ with $\|w^*\| = 1$. Therefore, the first statement follows. To prove the second one, from (8) follows

$$J_z(v^*) = -\frac{1}{2}\sigma(z|\mathcal{C})^2 = -\frac{1}{2}\|v^*\|^2, \tag{9}$$

and by using (7) we obtain (3). Finally, from (9) we have that

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

and this follows (4). □

A first consequence is the following:

Corollary 1. *The map $\sigma(\cdot|\mathcal{C})$ defines a seminorm on V.*

3. Main Result

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

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

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

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

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

Proposition 1 allows us 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

$$\begin{aligned} e_{n-1} &= z - z_{n-1}, \text{ and update} && (10)\\ z_n &= z_{n-1} + z^{(n)} \text{ where } z^{(n)} \in \Pi(e_{n-1}|\mathcal{C}). && (11) \end{aligned}$$

We observe that for $n \geqslant 1$,

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

or, equivalently, by using Proposition 1,

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

We introduce the following definition of the $\mathcal{C}$- *rank*

Definition 1. *We define the $\mathcal{C}$-rank of an element $z \in V$, denoted by* $\operatorname{rank}(z|\mathcal{C})$, *as follows:*

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

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

Now, we state the main result of this paper:

Theorem 1. *For $z \in V$, the sequence $\{e_n\}_{n\geqslant 0}$ constructed in (10) satisfies that* $\lim_{n\to\infty} e_n = e^*$ *and $e^* \in U(\mathcal{C})^{\perp}$. Moreover,*

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

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

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

In consequence,

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

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

$$\begin{aligned} \|e_n\|^2 &= \|e_{n-1} - z^{(n)}\|^2 \\ &= \|e_{n-1}\|^2 - \|z^{(n)}\|^2 \quad \text{(by using (3))} \\ &= \|e_{n-1}\|^2 - \sigma_n^2 \end{aligned}$$

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

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

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

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

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

for all $v \in \mathcal{C}$. 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 (12) and using that $e_0 = z$, we obtain

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

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

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

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

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

and then

$$\lim_{n\to\infty} (e_{n-1}, v)^2 = 0. \tag{12}$$

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 (12) that

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

for all $v \in \mathcal{C}$ with $\|v\| = 1$. Thus, $e^* \in U^{\perp}$. To conclude the proof we claim that $\{e_n\}_{n=1}^{\infty}$ is a Cauchy sequence in V in the $\|\cdot\|$-norm. The following three technical lemmas help us to show it.

Lemma 1. *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

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

□

Lemma 2. *For every $\varepsilon > 0$ and every $N \in \mathbb{N}$ there exists $\tau \geq N$ such that*

$$\sigma_\tau \sum_{k=1}^{\tau} \sigma_k \leq \varepsilon. \tag{13}$$

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 τ 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 $1 \le j < \tau(k+1)$, we have

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

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

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

Proof. We have

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

This ends the proof of the lemma. □

To prove the claim: $\{e_n\}_{n=1}^{\infty}$ is a Cauchy sequence in V in the $\|\cdot\|$-norm, we proceed as follows. 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 2, for each $m+p$ there exists $\tau > m+p$ such that

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

Now, we would to estimate

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

By using Lemma 3 with $M = \tau$ and $N = m$ and $m+p$, we obtain that

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

and

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

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 background for a large class of non-linear approximation problems. In this section, we will illustrate this through a particular example modelling blind deconvolution, introduced in [9]. In order to use Theorem 1 to prove the convergence of the blind deconvolution algorithm proposed in [9], we need to rewrite in a more convenient setting the main aspects of the model.

4.1. The Blind Deconvolution Problem

A problem arising frequently in image processing is that of recovering the original image from a degraded one. It is well known that an image $z_0(x,y)$ gets degraded due to different (natural or computational) causes, which can usually be mathematically formulated as follows:

$$z_1(x,y) = (K * z_0)(x,y) + n(x,y), \tag{14}$$

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:

$$(K * z_0)(x,y) = \int_{\mathbb{R}^2} K(x-\alpha, y-\beta) z_0(\alpha,\beta) \mathrm{d}\alpha \, \mathrm{d}\beta.$$

In [10], the authors develop a deconvolution model in the context of images degraded by weather and time conditions (in particular, artistic restoration of paintings). Under these circumstances, stochastic errors can be neglected (because of the focus distance or exposal time for the acquisition of the image) and the kernel can be considered almost quasi-Gaussian (below, we will introduce what it is called quasi-Gaussian kernels), due to weather conditions such as humidity, temperature, or time degradation. Hence, the convolution model is the following:

$$z_1(x,y) = (K * z_0)(x,y). \tag{15}$$

Deconvolution problems consist on recovering the original image z_0 from the convolved (observed) one z_1. The problem should be solved in the context of Fourier transforms, due to the fundamental theorem of convolution:

$$\widehat{K * z_0}(\xi,\eta) = \widehat{K}(\xi,\eta)\widehat{z}_0(\xi,\eta) = \widehat{z}_1(\xi,\eta),$$

where $\widehat{z}$ denotes the Fourier transformation of a function $z \in L_2(\mathbb{R}^2)$ defined by

$$\widehat{z}(\xi,\eta) := \int_{\mathbb{R}^2} e^{i(\xi\cdot x+\eta\cdot y)} z(x,y)\, dx\, dy.$$

A naive way to deconvolve is to obtain $\widehat{z}_0$ 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 frequencies of $\widehat{z}_0$. In consequence, a regularizing term must be included in order to stabilize the problem. It gets even more complicated when the kernel K is not known (blind deconvolution), although some theoretical knowledge can be used. For instance, in the above mentioned paper [9], it is considered that the degrading can be modelled by a Lèvy distribution. This Lèvy distribution corresponds to a Green function of a generalized heat equation of the form:

$$\left.\begin{aligned} \upsilon_t &= -\sum_{i=1}^{n} \lambda_i(-\Delta)^{\beta_i}\upsilon, && 0 < t \le 1, \\ \upsilon(x,y,t=1) &= z_1(x,y), && \lambda_i > 0,\ \beta_i \in [0,2]. \end{aligned}\right\} \tag{16}$$

Carasso proposed the Slow Evolution Backward Constraint for blind deconvolution using the Lévy Distribution [1, 2]. Several authors used the equation (16) in a variety of real blurred images [14, 16, 17, 18].

Let us remark that the operator $(-\Delta)$ is positive and, therefore, its powers $(-\Delta)^\beta$ are well defined by

$$(\widehat{(-\Delta^\beta)}z)(\xi,\eta) = c(\xi^2+\eta^2)^\beta\, \hat{z}(\xi,\eta),$$

where c is a parameter introduced in order to normalize the transformation. Recall that the Fourier transformation is an isomorphism (see VII in [7]). We denote by $\vee$ the inverse Fourier transformation, that is,

$$(\widehat{z})^\vee = \widehat{(z^\vee)} = z.$$

Then the generalized heat equation (16) can be solved by using

$$\upsilon(x,y,t) = \left(K^{\beta_n}_{\alpha_n(1-t)} * \cdots * K^{\beta_1}_{\alpha_1(1-t)} * z_1\right)(x,y),$$

via the fractional powers of the convolution kernel

$$K^{\beta}_{\alpha(1-t)} * z := \left(\widehat{K^{\beta}_{\alpha(1-t)} * z}\right)^{\vee} = \left(\widehat{k}^{\beta}_{\alpha(1-t)} \cdot \widehat{z}\right)^{\vee}, \qquad 0 \le t \le 1,$$

where

$$\widehat{k}^{\beta}_{\alpha(1-t)}(\xi,\eta) = e^{-\alpha(1-t)(\xi^2+\eta^2)^{\beta}} \qquad \alpha \ge 0,\ \beta \in [0,2].$$

In particular when $t = 1$, we have

$$K^{\beta}_{0} * z := \left(\widehat{k}^{\beta}_{0} \cdot \widehat{z}\right)^{\vee} = (\widehat{z})^{\vee} = z,$$

for all $\beta \in [0,2]$. In this framework, we assume that

$$\upsilon(x,y,t=0) = z_0(x,y) = \left(K^{\beta_n}_{\alpha_n} * \cdots * K^{\beta_1}_{\alpha_1} * z_1\right)(x,y). \tag{17}$$

We observe that the function

$$\left(K^{\beta_1}_{\alpha_1} * z_1\right)(x,y),$$

is equal to $\upsilon_1(x,y) := \upsilon_1(x,y,t=0)$ for

$$\upsilon_1(x,y,t) = \left(K^{\beta_1}_{\alpha_1(1-t)} * z_1\right)(x,y),$$

the solution of

$$\begin{aligned}\partial_t \upsilon_1 &= -\lambda_1(-\Delta)^{\beta_1}\upsilon_1,\\ \upsilon_1(x,y,t=1) &= z_1(x,y).\end{aligned}$$

Next, by proceeding inductively, we obtain that $z_0(x,y) = \left(K^{\beta_n}_{\alpha_n} * \upsilon_{n-1}\right)(x,y)$ is equal to $\upsilon_n(x,y) := \upsilon_n(x,y,t=0)$ for

$$\upsilon_n(x,y,t) = \left(K^{\beta_n}_{\alpha_n(1-t)} * \upsilon_{n-1}\right)(x,y),$$

the solution of

$$\begin{aligned}\partial_t \upsilon_n &= -\lambda_n(-\Delta)^{\beta_n}\upsilon_n,\\ \upsilon_n(x,y,t=1) &= z_1(x,y).\end{aligned}$$

Take $\upsilon_0(x,y) := z_1(x,y)$. It is well-known that each equation

$$\upsilon_l(x,y) = \left(K^{\beta_l}_{\alpha_l} * \upsilon_{l-1}\right)(x,y), \tag{18}$$

for $1 \le l \le n$, is ill-conditioned and, hence in order to solve it numerically, it needs to be regularized. Classic regularizations such as Tikhonov [13], or total variation [3, 12], appear widely in the literature. In [9], to solve (18) the authors propose to regularize that equation by using a fractional power of the Laplacian. For each $1 \le l \le n$, the regularized equation appears as

$$(-\Delta)^{\beta_l} \upsilon_l + \psi\, \overline{k}^{\beta_l}_{\alpha_l} * \left(k^{\beta_l}_{\alpha_l} * \upsilon_l - \upsilon_{l-1}\right) = 0, \tag{19}$$

here $\overline{k}$ denotes the complex conjugate of k, and υ_l is obtained from

$$\widehat{\upsilon}_l(\xi,\eta) = \frac{\widehat{k}^{\beta_l}_{\alpha_l}(\xi,\eta)\widehat{\upsilon}_{l-1}(\xi,\eta)}{\epsilon(\xi^2+\eta^2)^{\beta_l} + |\widehat{k}^{\beta_l}_{\alpha_l}(\xi,\eta)|^2}, \tag{20}$$

where $\epsilon = \dfrac{c}{\psi}$ and $\widehat{\upsilon}_l$ was previously computed applying the Fourier transform to the equation (19). The equation (15) is regularized by an operator based on fractional powers (19), where parameters (ψ, β) are crucial: ψ relates high and low frequencies, modulating the stochastic noise, and β controls the smoothing. The solution of the regularize equation (20) allows us to recover high, middle and low frequencies recursively by varying ψ (and, consequently, β), and using each partial deconvolution as the starting image for the next one.

4.2. The Fractional Deconvolution Model

The above remarks lead us to consider a model based on an iterative fractional decomposition of the kernel, in the sense of the greedy algorithms. The decomposition will be obtained by logarithmic approximation, as follows. By construction

$$\widehat{z}_0(\xi,\eta) = \widehat{\upsilon}_n(\xi,\eta) = \exp(-\alpha_n(\xi^2+\eta^2)^{\beta_n}) \cdots \exp(-\alpha_1(\xi^2+\eta^2)^{\beta_1})\widehat{\upsilon}_0(\xi,\eta). \tag{21}$$

Since

$$\widehat{\upsilon}_{l-1}(\xi,\eta) = \exp(-\alpha_l(\xi^2+\eta^2)^{\beta_l}) \cdot \widehat{\upsilon}_l(\xi,\eta),$$

we have

$$\log(|\widehat{\upsilon}_{l-1}(\xi,\eta)|) = -\alpha_l(\xi^2+\eta^2)^{\beta_l} + \log(|\widehat{\upsilon}_l(\xi,\eta)|),$$

for $1 \le l \le n$. Then we can write

$$e_l(r) := \alpha_l r^{\beta_l} = \log(|\widehat{\upsilon}_l(\xi,\eta)|) - \log(|\widehat{\upsilon}_{l-1}(\xi,\eta)|),$$

where $r = \xi^2 + \eta^2$, and hence

$$e \in \mathcal{C} = \left\{ \upsilon \in L_2[0,1] : \upsilon(r) = \gamma' r^{\beta'} \text{ where } \gamma' \ge 0 \text{ and } \beta' \in [0,2] \right\}.$$

Moreover, from (21), we have

$$z(\xi,\eta) := \log|\widehat{z}_1(\xi,\eta)| - \log|\widehat{z}_0(\xi,\eta)|$$
$$= \sum_{l=1}^{n} (\log|\widehat{\upsilon}_l(\xi,\eta)| - \log|\widehat{\upsilon}_{l-1}(\xi,\eta)|)$$
$$= \sum_{l=1}^{n} e_l(r) \in \operatorname{span}\mathcal{C}.$$

Since $\mathcal{C}$ is a weakly closed cone, it allows us to introduce the following version of the fractional deconvolution algorithm. We will call it *ALFA* (*Approximation Laplacian Fractional Algorithm*).

ALFA Approximation Laplacian Fractional Algorithm.

1: Given $\upsilon_0 = z_1$ take $\widehat{\upsilon}_0$ and for each $l \geq 1$ proceed until convergence as follows.

2: Take $\widehat{\upsilon}_l$ given by (20) and then compute α_l, β_l such that

$$\Pi(\log(|\widehat{\upsilon}_l(\xi,\eta)|) - \log(|\widehat{\upsilon}_{l-1}(\xi,\eta)|)|\mathcal{C}) = e_l,$$

where $e_l(r) = \alpha_l r^{\beta_l}$, that is, we solve

$$\min_{(\alpha_l,\beta_l)\in\mathbb{R}_+\times[0,2]} \|\log(|\widehat{\upsilon}_l(\xi,\eta)|) - \log(|\widehat{\upsilon}_{l-1}(\xi,\eta)|) - e_l\|_{L_2[0,1]}$$

3: Set $\widehat{\upsilon}_l(\xi,\eta) := \widehat{\upsilon}_{l-1}(\xi,\eta)\cdot\exp(\alpha_l(\xi^2+\eta^2)^{\beta_l})$ take $l = l+1$ and goto 2.

The interpretation from Theorem 1 of the fractional deconvolution algorithm is the following. Let us consider the approximation of

$$z(\xi,\eta) = \log|\widehat{z}_1(\xi,\eta)| - \log|\widehat{z}_0(\xi,\eta)| \in L_2(\mathbb{D}),$$

where $\mathbb{D} = \{(\xi,\eta) : \xi^2+\eta^2 \leq 1\}$, by a sum of functions $\upsilon \in \mathcal{C}$. The limit case leads to two possibilities: if the process is finite, z lays in the space generated by the cone and hence (21) holds; 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. Theorem 1 allows us to obtain the optimal approximation to z and a stopping criterion: when the residuals are approximately equal, or close to zero (under certain thresholding), the algorithm provides a good approximation to z *via* $P_{U(\mathcal{C})}(\cdot)$*). Therefore, the fractional deconvolution method is convergent (thanks*

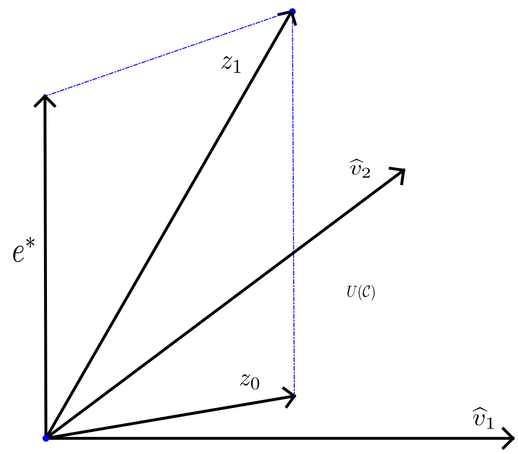


Figure 1: Geometry of the proposed image restoration model.

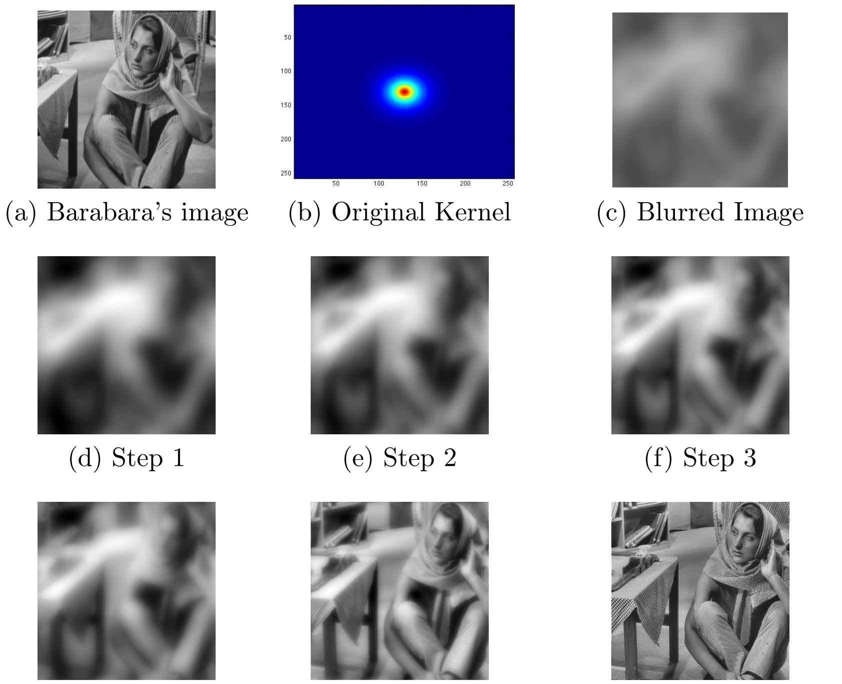
to Theorem 1). From a geometrical point of view, the algorithm behaves as it is shown in Figure 1.

The algorithm begins in a set with the basic structure (A1) and (A2), that is, a weakly-closed cone $\mathcal{C}$. Then the image is projected over $\mathcal{C}$, originating an optimal approach over this non-linear set. We remark that it is not a linear projection in the usual sense, there exists an optimal approximating point over this set, however it is not unique. Once computed the residual between two consecutive iterations, the algorithm tends to the orthogonal projection (Theorem 1), which is the optimal approximation in the $\overline{\operatorname{span}\mathcal{C}}^{\|\cdot\|}$. The approximation to the orthogonal projection up to a given tolerance as a stopping criteria has the advantage that local errors do not get amplified, as it is proved in Lemmas 1, 2 and 3.

5. Examples

In this section we will display some examples in order to illustrate the above results. Figure 2 shows the deconvolution process: the original image, the corresponding partial deconvolved images and their residuals (below). Convolution kernel is obtained after diffusion of Dirac delta function by 80 iterations of the heat equation. In the shown examples, a good quality deconvolved image is obtained after six steps. The deconvolved image recovers many of the original details which cannot be recovered by other models. We refer the reader to [9, 11] for comparison purpose with other type of images and tests.

In Figure 2 (b)-(f) we see the sequence of gradual deblurring. In each step, the algorithm recovers details,and among them, the textures. Adaptiveness of the process is shown in table 1 where we see the evolution of the parameters.



(a) Barabara's image (b) Original Kernel (c) Blurred Image

(d) Step 1 (e) Step 2 (f) Step 3

(g) Step 4 (h) Step 5 (i) Step 6 (deconvolution)

Figure 2: (a) Original image (Barbara). (b)-(g) Deconvolved Image (6 steps)

In Table 1 we see the evolution of the parameters of the decomposed kernels (equation (17)) and relative errors. The algorithm ends when the conditions of Theorem 1 are achieved, that is to say, when the residuals of the greedy algorithm are closer than a given tolerance; in this case the algorithm stops at 0.001281.

Finally, we show a real life application (restoration of a baroque painting) in Figure 3. It is important to ensure that fractional decomposition (that is, fractional projection) works in a multichannel context as it is that of colour images. 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 that 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 can see, in our fractional model, the image can be deconvolved while keeping these scratches, which is often important in order to apply particular and local models for arranging them.

Iteration	α	β	L_2 Relative error
1	0.001796	0.798403	0.028453
2	2.408154	0.157319	0.018320
3	8.080902	0.113034	0.0081933
4	8.823571	0.105422	0.005986
5	10.52049	0.048073	0.001281
6	13.21462	0.01287	0.000295

Table 1: Parameters of the detected kernel, and relative errors for *Barbara*

6. Conclusion and Final Remarks

In this paper we have introduced some theoretical results for the convergence of a class of non-linear approximation problems underlying in many greedy algorithms. In particular, our results provides an alternative tool to the use of convexity, which is usually not present in a non-linear framework.

In order to illustrate the strength of the results, we analyse an algorithm related to image restoration. In general, image restoration and denoising are situations where this theory can be applied: most of the models are used to find good approximations to the original image that is recovered under some restrictions. The theory is not only useful to prove the convergence of the deconvolution algorithm, it also explains some features of the deconvolved image (recovery of edges, for instance).

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

The fractional model introduced in this paper presents some of the non-linear features that one can expect. For example, adaptiveness to the initial conditions, and joint detection of the deconvolved image and the blurring kernel. The deconvolved image is constructed by using the successive residuals of the process. In consequence if the blurred image is in the subspace that we are projecting, then the residual is just the null image (a black one). In other words, in general, the image that we obtain is the best one, among all the possible images.

Acknowledgements: We would like to thank the anonymous referee's for their useful and fruitful comments.

7. References

[1] A. Carasso, Direct blind deconvolution, SIAM Journal on Numerical Analysis, 61 (2001) 1980-2007.

[2] A. Carasso. The APEX method in image sharpening and the use of low exponent Levy stable laws, SIAM J. Appl. Math. 63 (2002) 593–618.

[3] T. Chan, and C.K. Wong, *Total variation blind deconvolution*, IEEE Trans. Image Process., 7 (3) (1998) 370–375

[4] M. Fabian, P. Habala, P. Háyek, V. Montesinos and V. Zizler. *Banach Space Theory: The basis for Linear and Non-Linear Analysis.* CMS Books in Mathematics, Springer-Verlag, New York 2011.

[5] A. Falcó, and W. Hackbusch. *On minimal subspaces in tensor representations*, Found. Comput. Math. Volume 12, Issue 6 (2012) 765–803.

[6] L. He, A. Marquina, and S. Osher, *Blind deconvolution using TV regularization and Bregman iteration.* Int. J. Imag. Systems and Techn., 15 (2005) 74–83.

[7] L. Hörmander. *The Analysis of Linear Partial Differential Equations I: Distribution Theory and Fourier Analysis.* Classics in Mathematics, Springer-Verlag, Berlin 2003.

[8] M. Nicolova *Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares*, Multiscale Model. Simul. Vol 4, No 3 (2005) 960–991.

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

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

[11] P. D. Romero, and V.F. Candela. *Modelos de deconvolución ciega fraccionaria. Aplicaciones a la restauración de obras pictóricas.* Servici de Publicacions de la Universitat de Valéncia 2009.

[12] L. Rudin, and S. Osher, *Total variation based image restoration with free local constraints*, Proc. IEEE Int. Conf. Image Process. (1994) 31–35.

[13] A. Tikhonov, and V. Arsenin, *Solution of Ill-Posed Problems.* Wiley, New York 1977.

[14] T. Wang, Y. Zhang, Y. S. Zhang, and L. X. Lin. *Automatic super-resolution image reconstruction based on hybrid map-pocs*, Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, ICWAPR'07, Beijing, China (2007) Published by IEEE. 426–431.

[15] K. Yosida *Functional Analysis: Fourth Edition.* Springer-Verlag, Berlin 1974.

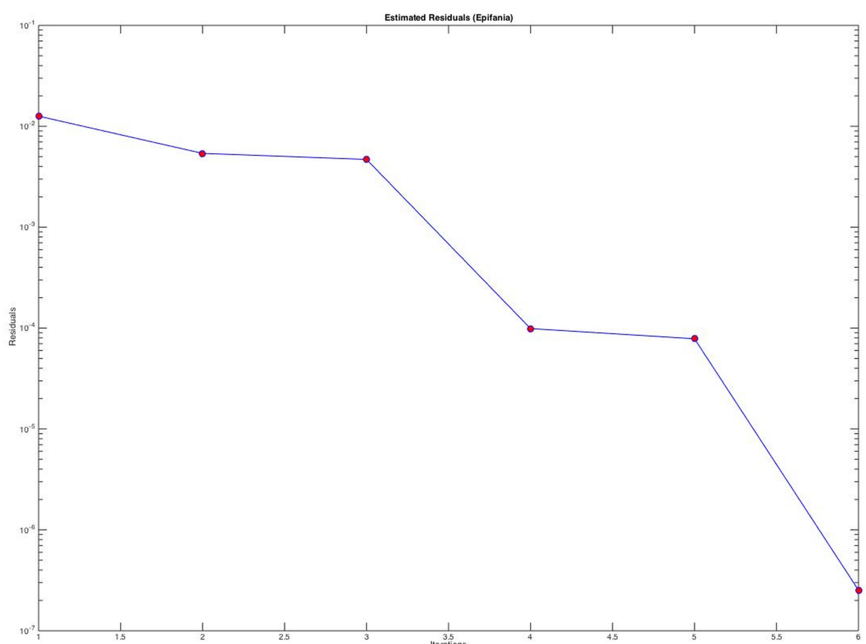
[16] L. Yu, Y. Qi, D. Li, M. Xia, and L. Xuan. *Image restoration of the open-loop adaptive optics retinal imaging system, based on optical transfer function analysis*, Optics Communication, 300 (2013) 178–182.

[17] H. Zhang, Q. Ge, L. Li, Y. Li, and K. Xi. *A new point spread function estimation approach for recovery of atmospheric turbulence degraded photographs*, Proceedings of 4th International Conference on Image and Signal Processing, (2011), Published by IEEE. 774–778.

[18] J. Zhang and Q. Zhang. *Noniterative blind image restoration based on estimation of a significant class of point spread functions*, Optical Engineering 46(7) 077005 (2007).

(a) Original image: Epiphany

(b) Fractional Blind Deconvolution



(c) Estimated Residuals

Figure 3: Color Fractional Blind deconvolution. Theme: Epiphany, from the altarpiece of Saint Bartholomew Church, Bienservida (Spain)