Tutorial

Laura Rebollo-Neira

September 9, 2011

This tutorial is intended to be a pedagogical self-contained introduction to the techniques developed within the EPSRC founded project: "Highly nonlinear approximations for sparse signal representation".

List of Symbols

The following notations and symbols will be used without defining them explicitly:

$$
\begin{aligned}
\cup &: \text{ union} \\
\cap &: \text{ intersection} \\
\subseteq &: \text{ subset of} \\
\subset &: \text{ proper subset of} \\
\in &: \text{ belong(s)} \\
\mathbb{N} &: \text{ set of all positive integers} \\
\mathbb{Z} &: \text{ set of all integers} \\
\mathbb{R} &: \text{ field of all real numbers} \\
\mathbb{C} &: \text{ field of all complex numbers} \\
\mathbb{F} &: \text{ field of real or complex numbers} \\
:= &: \text{ is defined by} \\
\Longrightarrow &: \text{ imply (implies)} \\
\Longleftrightarrow &: \text{ if and only if} \\
\to &: \text{ maps to}
\end{aligned}
$$

The Kronecker symbol is given by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

The characteristic function χ_S of a set S is defined as

$$\chi_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{otherwise.} \end{cases}$$

For $n \in \mathbb{N}$ the factorial $n!$ is defined as $n! = n(n-1)\cdots 2 \cdot 1$. The absolute value of a number $a \in \mathbb{F}$ is indicated as $|a|$. If $a \in \mathbb{R}$

$$|a| = \begin{cases} a & \text{if } a \geq 0 \\ -a & \text{if } a < 0. \end{cases}$$

If $a \in \mathbb{C}$ its complex conjugate is denoted by $\overline{a}$ and $|a|^2 = a\overline{a}$.

Elementary Definitions

Note: **Bold face** is used when a terminology is defined. *Italics* are used to emphasize a terminology or statement.

Vector Space

A **vector space** over a field $\mathbb{F}$ is a set $\mathcal{V}$ together with two operations vector addition, denoted $v + w \in \mathcal{V}$ for $v, w \in \mathcal{V}$ and scalar multiplication, denoted $av \in \mathcal{V}$ for $a \in \mathbb{F}$ and $v \in \mathcal{V}$, such that the following axioms are satisfied:

1. $v + w = w + v,\ v, w \in \mathcal{V}$.
2. $u + (v + w) = (u + v) + w,\ u, v, w \in \mathcal{V}$.
3. There exists an element $0 \in \mathcal{V}$, called the zero vector, such that $v + 0 = v,\ v \in \mathcal{V}$.
4. There exists an element $\tilde{v} \in \mathcal{V}$, called the additive inverse of v, such that $v+\tilde{v} = 0$, $v \in \mathcal{V}$.
5. $a(bv) = (ab)v,\ a, b \in \mathbb{F}$ and $v \in \mathcal{V}$.
6. $a(v + w) = av + aw,\ a \in \mathbb{F}$ and $v, w \in \mathcal{V}$.
7. $(a + b)v = av + bv,\ a, b \in \mathbb{F}$ and $v \in \mathcal{V}$.
8. $1v = v,\ v \in \mathcal{V}$, where 1 denotes the multiplicative identity in $\mathbb{F}$.

The elements of a vector space are called **vectors**.

Subspaces - Direct sum

A subset $\mathcal{S}$ of a vector space $\mathcal{V}$ is a **subspace** of $\mathcal{V}$ if it is a vector space with respect to the vector space operations on $\mathcal{V}$. A subspace which is a proper subset of the whole space is called a **proper subspace**. Two subspaces $\mathcal{V}_1$ and $\mathcal{V}_2$ are **complementary** or **disjoint** if $\mathcal{V}_1 \cap \mathcal{V}_2 = \{0\}$.

The sum of two subspaces $\mathcal{V}_1$ and $\mathcal{V}_2$ is the subspace $\mathcal{V} = \mathcal{V}_1+\mathcal{V}_2$ of elements $v = v_1+v_2$, $v_1 \in \mathcal{V}_1$, $v_2 \in \mathcal{V}_2$. If the subspaces $\mathcal{V}_1$ and $\mathcal{V}_2$ are *complementary* $\mathcal{V} = \mathcal{V}_1 + \mathcal{V}_2$ is called **direct sum** and indicated as $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2$. This implies that each element $v \in \mathcal{V}$ has a unique decomposition $v = v_1 + v_2$, $v_1 \in \mathcal{V}_1$, $v_2 \in \mathcal{V}_2$.

For the sets V_1 and V_2, the set $\{v \in V_1 \,:\, v \notin V_2\}$ is denoted by $V_1 \setminus V_2$.

Linear operators and linear functionals

Let $\mathcal{V}_1$ and $\mathcal{V}_2$ be vectors spaces. A mapping $\hat{A} : \mathcal{V}_1 \to \mathcal{V}_2$ is a **linear operator** if

$$\hat{A}(v + w) = \hat{A}v + \hat{A}w, \quad \hat{A}(av) = a\hat{A}v,$$

for all $v, w \in \mathcal{V}_1$ and $a \in F$. $\mathcal{V}_1$ is called the **domain** of $\hat{A}$ and $\mathcal{V}_2$ its **codomain** or **image**. If the codomain of a linear operator is a scalar field, the operator is called a **linear functional** on $\mathcal{V}_1$. The set of all linear functionals on $\mathcal{V}_1$ is called the **dual space** of $\mathcal{V}_1$.

The **adjoint** of an operator $\hat{A} : \mathcal{V}_1 \to \mathcal{V}_2$ is the unique operator $\hat{A}^*$ satisfying that

$$\langle \hat{A}g_1, g_2\rangle = \langle g_1, \hat{A}^* g_2\rangle.$$

If $\hat{A}^* = \hat{A}$ the operator is **self-adjoint** or **Hermitian**

An operator $\hat{A} : \mathcal{V}_1 \to \mathcal{V}_2$ has an **inverse** if there exists $\hat{A}^{-1} : \mathcal{V}_2 \to \mathcal{V}_1$ such that

$$\hat{A}^{-1}\hat{A} = \hat{I}_{\mathcal{V}_1} \quad \text{and} \quad \hat{A}\hat{A}^{-1} = \hat{I}_{\mathcal{V}_2},$$

where $\hat{I}_{\mathcal{V}_1}$ and $\hat{I}_{\mathcal{V}_2}$ denote the identity operators in $\mathcal{V}_1$ and $\mathcal{V}_2$, respectively. By a **generalised inverse** we shall mean an operator $\hat{A}^\dagger$ satisfying the following conditions

$$\begin{aligned} \hat{A}\hat{A}^\dagger\hat{A} &= \hat{A} \\ \hat{A}^\dagger\hat{A}\hat{A}^\dagger &= \hat{A}^\dagger. \end{aligned}$$

The unique generalized inverse satisfying

$$\begin{aligned} (\hat{A}\hat{A}^\dagger)^* &= \hat{A}\hat{A}^\dagger \\ (\hat{A}^\dagger\hat{A})^* &= \hat{A}^\dagger\hat{A}. \end{aligned}$$

is known as the *Moore Penrose pseudoinverse*.

Frames and bases for finite a dimension vector space

If $v_1, \ldots, v_n$ are some elements of a vector space $\mathcal{V}$, by a **linear combination** of $v_1, \ldots, v_n$ we mean an element in $\mathcal{V}$ of the form $a_1v_1 + \cdots + a_nv_n$, with $a_i \in \mathbb{F}, i = 1, \ldots, n$.

Let S be a subset of element of $\mathcal{V}$. The set of all *linear combinations* of elements of S is called the **span** of S and is denoted by $\operatorname{span} S$.

A subset $S = \{v_i\}_{i=1}^n$ of $\mathcal{V}$ is said to be **linearly independent** if and only if

$$a_1v_1 + \cdots + a_nv_n = 0, \qquad \Longrightarrow \qquad a_i = 0, i = 1, \ldots, n.$$

A subset is said to be **linearly dependent** if it is not linearly independent.

$\mathcal{S}$ is said to be a **basis** of $\mathcal{V}$ if it is linearly independent and $\operatorname{span} S = \mathcal{V}$. The **dimension** of a finite dimensional vector space $\mathcal{V}$ is the number of elements in a basis for $\mathcal{V}$. The number of elements in a set is termed the **cardinality** of the set.

If the number of elements spanning a finite dimensional vector space $\mathcal{V}$ is larger than the number of elements of a basis for the same space, the set is called a **redundant frame**. In other words a redundant frame (henceforth called simply a frame) for a finite dimensional vector space $\mathcal{V}$ is a linearly dependent set of vectors spanning $\mathcal{V}$.

Let $\{v_i\}_{i=1}^n$ be a spanning set for $\mathcal{V}$. Then every $v \in \mathcal{V}$ can be expressed as

$$v = a_1v_1 + \cdots, a_nv_n, \quad \text{with,} \quad a_i \in \mathbb{F}, i = 1, \ldots, n.$$

If the spanning set $\{v_i\}_{i=1}^n$ is a basis for $\mathcal{V}$ the numbers $a_i \in \mathbb{F}, i = 1, \ldots, n$ in the above decomposition are unique. In the case of a redundant frame, however, these numbers are not longer unique. For further discussion about redundant frames in finite dimension see [1].

Normed vector space

A **norm** $\|\cdot\|$ on a vector space $\mathcal{V}$ is a function from $\mathcal{V}$ to $\mathbb{R}$ such that for every $v, w \in \mathcal{V}$ and $a \in \mathbb{F}$ the following three properties are fulfilled

1. $\|v\| \geq 0$, and $\|v\| = 0 \Longleftrightarrow v = 0$.
2. $\|av\| = |a|\|v\|$.
3. $\|v + w\| \leq \|v\| + \|w\|$.

A vector space $\mathcal{V}$ together with a norm is called a **normed vector space**.

Inner product space

An **inner product** on a vector space $\mathcal{V}$ is a map from $\mathcal{V}$ to $\mathbb{F}$ which satisfies the following axioms

1. $\langle v, v\rangle \geq 0,\ v \in \mathcal{V}$, and $\langle v, v\rangle = 0 \Longleftrightarrow v = 0$.
2. $\langle v + w, z\rangle = \langle v, z\rangle + \langle w, z\rangle,\ \ v, w, z \in \mathcal{V}$.
3. $\langle v, az\rangle = a\langle v, z\rangle,\ v, z \in \mathcal{V}$ and $a \in \mathbb{F}$.
4. $\langle v, w\rangle = \overline{\langle w, v\rangle},\ v, w \in \mathcal{V}$.

A vector space $\mathcal{V}$ together with an inner product $\langle \cdot, \cdot\rangle$ is called an **inner product space**.

Orthogonality

Two vectors v and w in an inner product space are said to be **orthogonal** if $\langle v, w\rangle = 0$. If, in addition, $\|v\| = \|w\| = 1$ they are **orthonormal**.

Two subspaces $\mathcal{V}_1$ and $\mathcal{V}_2$ are orthogonal if $\langle v_1, v_2\rangle = 0$ for all $v_1 \in \mathcal{V}_1$ and $v_2 \in \mathcal{V}_1$. The sum of two orthogonal subspaces $\mathcal{V}_1$ and $\mathcal{V}_2$ is termed **orthogonal sum** and will be indicated as $\mathcal{V} = \mathcal{V}_1 \oplus^{\perp} \mathcal{V}_2$. The subspace $\mathcal{V}_2$ is called the **orthogonal complement** of $\mathcal{V}_1$ in $\mathcal{V}$. Equivalently, $\mathcal{V}_1$ is the orthogonal complement of $\mathcal{V}_2$ in $\mathcal{V}$.

The spaces $\mathbb{R}^n$, $L^2[a, b]$ and $C^k[a, b]$

The **Euclidean space** $\mathbb{R}^n$ is an inner product space with inner product defined by

$$\langle \mathbf{x}, \mathbf{y}\rangle = x_i\overline{y}_i + \ldots x_n\overline{y}_n,$$

with $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$. The norm $\|\mathbf{x}\|$ is induced by the inner product

$$\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x}\rangle^{\frac{1}{2}} = (x_i\overline{x}_i + \ldots x_n\overline{x}_n)^{\frac{1}{2}} = (|x_i|^2 + \ldots + |x_n|^2)^{\frac{1}{2}}.$$

The **space** $L^2[a, b]$ is an inner product space of functions on $[a, b]$ with inner product defined by

$$\langle f, g\rangle = \int_a^b f(t)\overline{g}(t)\, dt$$

and norm

$$\|f\| = \langle f, f\rangle^{\frac{1}{2}} = \left(\int_a^b |f(t)|^2\, dt\right)^{\frac{1}{2}}.$$

The **space** $C^k[a, b]$ is the space of functions on $[a, b]$ having continuous derivatives up to order $k \in \mathbb{N}$. The space of continuous functions on $[a, b]$ is denoted as $C^0[a, b]$.

Signal Representation, Reconstruction, and Projection

Regardless of its informational content, in this tutorial we consider that a **signal** is an element of an inner product space $\mathcal{H}$ with norm induced by the inner product, $\|\cdot\| = \langle \cdot, \cdot\rangle^{\frac{1}{2}}$. Moreover, we assume that all the signals of interest belong to some finite dimensional subspace $\mathcal{V}$ of

$\mathcal{H}$ spanned by the set $\{v_i \in \mathcal{H}\}_{i=1}^M$. Hence, a signal f can be expressed by a finite linear superposition

$$f = \sum_{i=1}^{M} c_i v_i,$$

where the coefficients c_i, $i = 1, \ldots, M$, are in $\mathbb{F}$.

We call **measurement** or **sampling** to the process of transforming a signal into a number. Hence a **measure** or **sample** is a *functional*. Because we restrict considerations to linear measures the associated functional is linear and can be expressed as

$$m = \langle w, f \rangle \quad \text{for some} \quad w \in \mathcal{H}.$$

We refer the vector w to as **measurement vector**.

Considering M measurements m_i, $i = 1, \ldots, M$, each of which is obtained by a measurement vector w_i, we have a numerical representation of f as given by

$$m_i = \langle w_i, f \rangle, \quad i = 1, \ldots, M.$$

Now we want to answer the question as to whether it is possible to reconstruct $f \in \mathcal{V}$ from these measurements. More precisely, we wish to find the requirements we need to impose upon the measurement vectors w_i, $i = 1, \ldots, M$, so as to use the concomitant measures $\langle w_i, f \rangle$, $i = 1, \ldots, M$, as coefficients for the signal reconstruction, i.e., we wish to have

$$f = \sum_{i=1}^{M} c_i v_i = \sum_{i=1}^{M} \langle w_i, f \rangle v_i. \tag{1}$$

By denoting

$$\hat{E} = \sum_{i=1}^{M} v_i \langle w_i, \cdot \rangle, \tag{2}$$

where the operation $\langle w_i, \cdot \rangle$ indicates that $\hat{E}$ acts by taking inner products, (1) is written as

$$f = \hat{E} f.$$

As will be discussed next, the above equation tells us that the measurement vectors w_i, $i = 1, \ldots, M$, should be such that the operator $\hat{E}$ is a projector onto $\mathcal{V}$.

Projectors

An operator $\hat{E} : \mathcal{H} \to \mathcal{V}$ is a projector if it is *idempotent*, i.e.,

$$\hat{E}^2 = \hat{E}.$$

As a consequence, the projection is onto $\mathcal{R}(\hat{E})$, the range of the operator, and along $\mathcal{N}(\hat{E})$, the null space of the operator. Let us recall that

$$\mathcal{R}(\hat{E}) = \{f, \text{ such that } f = \hat{E} g, \ g \in \mathcal{H}\}.$$

Thus, if $f \in \mathcal{R}(\hat{E})$,

$$\hat{E} f = \hat{E}^2 g = \hat{E} g = f.$$

This implies that $\hat{E}$ behaves like the identity operator for all $f \in \mathcal{R}(\hat{E})$, regardless of $\mathcal{N}(E)$, which is defined as

$$\mathcal{N}(E) = \{g, \text{ such that } \hat{E}g = 0, \ g \in \mathcal{H}\}.$$

It is clear then that to reconstruct a signal $f \in \mathcal{V}$ by means of (1) the involved measurement vectors w_i, $i = 1, \ldots, M$, that we shall also called henceforth **duals**, should give rise to an operator of the form (2), which must be a projector onto $\mathcal{V}$. Notice that the required operator is not unique, because there exist many projectors onto $\mathcal{V}$ having different $\mathcal{N}(\hat{E})$. Thus, for reconstructing signals in the range of the projector its null space can be chosen arbitrarily. However, the null space becomes extremely important when the projector acts on signals outside its range. A popular projector is the orthogonal one.

Oblique and Orthogonal Projector

When $\mathcal{N}(\hat{E})$ happens to be equal to $\mathcal{R}(\hat{E})^\perp$, which indicates the orthogonal complement of $\mathcal{R}(\hat{E})$, the projector is called **orthogonal projector** onto $\mathcal{R}(\hat{E})$. This is the case if and only if the projector is *self adjoint*.

A projector which is not orthogonal is called an **oblique projector** and we need two subscripts to represent it. One subscript to indicate the range of the projector and another to represent the subspace along which the projection is performed. Hence the projector onto $\mathcal{V}$ along $\mathcal{W}^\perp$ is indicated as $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$.

The particular case $\hat{E}_{\mathcal{V}\mathcal{V}^\perp}$ corresponds to an *orthogonal projector* and we use the special notation $\hat{P}_\mathcal{V}$ to indicate such a projector. When a projector onto $\mathcal{V}$ is used for signal processing, $\mathcal{W}^\perp$ can be chosen according to the processing task. The examples below illustrate two different situations.

Example 1

Let us assume that the signal processing task is to approximate a signal $f \in \mathcal{H}$ by a signal $f_\mathcal{V} \in \mathcal{V}$. In this case, one normally would choose $f_\mathcal{V} = \hat{P}_\mathcal{V} f$ because this is the unique signal in $\mathcal{V}$ minimizing the distance $\|f - f_\mathcal{V}\|$. Indeed, let us take another signal g in $\mathcal{V}$ and write it as $g = g + \hat{P}_\mathcal{V} f - \hat{P}_\mathcal{V} f$. Since $f - \hat{P}_\mathcal{V} f$ is orthogonal to every signal in $\mathcal{V}$ we have

$$\|f - g\|^2 = \|f - g + \hat{P}_\mathcal{V} f - \hat{P}_\mathcal{V} f\|^2 = \|f - \hat{P}_\mathcal{V} f\|^2 + \|\hat{P}_\mathcal{V} f - g\|^2.$$

Hence $\|f - g\|$ is minimized if $g = \hat{P}_\mathcal{V} f$.

Remark 1. *Any other projection would yield a distance $\|f - \hat{E}_{\mathcal{V}\mathcal{W}^\perp} f\|$ which satisfies [2]*

$$\|f - \hat{P}_\mathcal{V} f\| \leq \|f - \hat{E}_{\mathcal{V}\mathcal{W}^\perp} f\| \leq \frac{1}{\cos(\theta)} \|f - \hat{P}_\mathcal{V} f\|,$$

where θ is the minimum angle between the subspaces $\mathcal{V}$ and $\mathcal{W}$. The equality is attained for $\mathcal{V} = \mathcal{W}$, which corresponds to the orthogonal projection.

Example 2

Assume that the signal f to be analyzed here is the superposition of two signals, $f = f_1 + f_2$, each component being produced by a different phenomenon we want to discriminate. Let us assume further that $f_1 \in \mathcal{V}$ and $f_2 \in \mathcal{W}^\perp$ with $\mathcal{V}$ and $\mathcal{W}^\perp$ disjoint subspaces. Thus, we can obtain, f_1 say, from f, by an oblique projector onto $\mathcal{V}$ and along $\mathcal{W}^\perp$. The projector will map to zero the component f_2 to produce

$$f_1 = \hat{E}_{\mathcal{V}\mathcal{W}^\perp} f.$$

Construction of Oblique Projectors for signal splitting

Oblique projectors in the context of signal reconstruction were introduced in [2] and further analyzed in [3]. The application to *signal splitting*, also termed *structured noise filtering*, amongst a number of other applications, is discussed in [4]. For advanced theoretical studies of oblique projector operators in infinite dimensional spaces see [5, 6]. We restrict our consideration to numerical constructions in finite dimension, with the aim of addressing the problem of signal splitting when the problem is ill posed.

Given $\mathcal{V}$ and $\mathcal{W}^\perp$ disjoint, i.e., such that $\mathcal{V} \cap \mathcal{W}^\perp = \{0\}$, in order to provide a prescription for constructing $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$ we proceed as follows. Firstly we define $\mathcal{S}$ as the direct sum of $\mathcal{V}$ and $\mathcal{W}^\perp$, which we express as

$$\mathcal{S} = \mathcal{V} \oplus \mathcal{W}^\perp.$$

Let $\mathcal{W} = (\mathcal{W}^\perp)^\perp$ be the orthogonal complement of $\mathcal{W}^\perp$ in $\mathcal{S}$. Thus we have

$$\mathcal{S} = \mathcal{V} \oplus \mathcal{W}^\perp = \mathcal{W} \oplus^\perp \mathcal{W}^\perp.$$

The operations $\oplus$ and $\oplus^\perp$ are termed direct and orthogonal sum, respectively.

Considering that $\{v_i\}_{i=1}^M$ is a spanning set for $\mathcal{V}$ a spanning set for $\mathcal{W}$ is obtained as

$$u_i = v_i - \hat{P}_{\mathcal{W}^\perp} v_i = \hat{P}_{\mathcal{W}} v_i, \quad i = 1, \ldots, M.$$

Denoting as $\{\mathbf{e}_i\}_{i=1}^M$ the standard orthonormal basis in $\mathbb{F}^M$, i.e., the Euclidean inner product $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij}$, we define the operators $\hat{V} : \mathbb{F}^M \to \mathcal{V}$ and $\hat{U} : \mathbb{F}^M \to \mathcal{W}$ as

$$\hat{V} = \sum_{i=1}^{M} v_i \langle \mathbf{e}_i, \cdot \rangle, \qquad \hat{U} = \sum_{i=1}^{M} u_i \langle \mathbf{e}_i, \cdot \rangle.$$

Thus the adjoint operators $\hat{U}^*$ and $\hat{V}^*$ are

$$\hat{V}^* = \sum_{i=1}^{M} \mathbf{e}_i \langle v_i, \cdot \rangle, \qquad \hat{U}^* = \sum_{i=1}^{M} \mathbf{e}_i \langle u_i, \cdot \rangle.$$

It follows that $\hat{P}_{\mathcal{W}} \hat{V} = \hat{U}$ and $\hat{U}^* \hat{P}_{\mathcal{W}} = \hat{U}^*$ hence, $\hat{G} : \mathbb{C}^M \to \mathbb{C}^M$ defined as:

$$\hat{G} = \hat{U}^* \hat{V} = \hat{U}^* \hat{U}$$

is self-adjoint and its matrix representation, G, has elements

$$g_{i,j} = \langle u_i, v_j \rangle = \langle \hat{P}_{\mathcal{W}} u_i, v_j \rangle = \langle u_i, \hat{P}_{\mathcal{W}} u_j \rangle = \langle u_i, u_j \rangle, \quad i, j = 1, \ldots, M.$$

From now on we restrict our signal space to be $\mathcal{S}$, since we would like to build the oblique projector $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$ onto $\mathcal{V}$ and along $\mathcal{W}^\perp$ having the form

$$\hat{E}_{\mathcal{V}\mathcal{W}^\perp} = \sum_{i=1}^{M} v_i \langle w_i, \cdot \rangle. \tag{3}$$

Clearly for the operator to map to zero every vector in $\mathcal{W}^\perp$ the dual vectors $\{w_i\}_{i=1}^M$ must span $\mathcal{W} = (\mathcal{W}^\perp)^\perp = \operatorname{span}\{u_i\}_{i=1}^M$. This entails that for each w_i there exists a set of coefficients $\{b_{i,j}\}_{j=1}^M$ such that

$$w_i = \sum_{i=1}^{M} b_{i,j} u_i, \tag{4}$$

which guarantees that every w_i is orthogonal to all vectors in $\mathcal{W}^\perp$ and therefore $\mathcal{W}^\perp$ is included in the null space of $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$. Moreover, since every signal, g say, in $\mathcal{S}$ can be written as $g = g_{\mathcal{V}} + g_{\mathcal{W}^\perp}$ with $g_{\mathcal{V}} \in \mathcal{V}$ and $g_{\mathcal{W}^\perp} \in \mathcal{W}^\perp$, the fact that $\hat{E}_{\mathcal{V}\mathcal{W}^\perp} g = 0$ implies $g_{\mathcal{V}} = 0$. Hence, $g = g_{\mathcal{W}^\perp}$, which implies that the null space of $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$ restricted to $\mathcal{S}$ is $\mathcal{W}^\perp$.

In order for $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$ to be a projector it is necessary that $\hat{E}^2_{\mathcal{V}\mathcal{W}^\perp} = \hat{E}_{\mathcal{V}\mathcal{W}^\perp}$. As will be shown in the next proposition, if the coefficients $b_{i,j}$ are the matrix elements of a *generalised inverse* of the matrix G this condition is fulfilled.

Proposition 1. *If the coefficients $b_{i,j}$ in* (4) *are the matrix elements of a generalised inverse of the matrix G, which has elements $g_{i,j} = \langle v_i, u_j \rangle$, $i, j = 1, \ldots, M$, the operator in* (3) *is a projector.*

Proof. For the measurement vectors in (4) to yield a projector of the form (3), the corresponding operator should be idempotent, i.e.,

$$\sum_{n=1}^{M}\sum_{m=1}^{M}\sum_{i=1}^{M}\sum_{j=1}^{M} v_i b_{i,j} \langle u_j, v_n \rangle \overline{b_{n,m}} \langle u_m, \cdot \rangle = \sum_{i=1}^{M}\sum_{j=1}^{M} v_i \overline{b_{i,j}} \langle u_j, \cdot \rangle. \tag{5}$$

Defining

$$\hat{B} = \sum_{i=1}^{M}\sum_{j=1}^{M} \mathbf{e}_i \overline{b_{i,j}} \langle \mathbf{e}_j, \cdot \rangle \tag{6}$$

and using the operators $\hat{V}$ and $\hat{U}^*$, as given above, the left hand side in (5) can be expressed as

$$\hat{V}\hat{B}^*\hat{U}^*\hat{V}\hat{B}^*\hat{U}^* \tag{7}$$

and the right hand side as

$$\hat{V}\hat{B}^*\hat{U}^*. \tag{8}$$

Assuming that $\hat{B}^*$ is a generalised inverse of $(\hat{U}^*\hat{V})$ indicated as $\hat{B}^* = (\hat{U}\hat{V})^\dagger$ it satisfies (c.f. Section)

$$\hat{B}^*(\hat{U}^*\hat{V})\hat{B}^* = \hat{B}^*, \tag{9}$$

and therefore, from (7), the right hand side of (5) follows. Since $\hat{B}^* = (\hat{U}^*\hat{V})^\dagger = \hat{G}^\dagger$ and $\hat{G}^* = \hat{G}$, we have $\hat{B} = \hat{G}^\dagger$. Hence, if the elements $b_{i,j}$ determining $\hat{B}$ in (6) are the matrix elements of a generalised inverse of the matrix representation of $\hat{G}$, the corresponding vectors $\{w_i\}_{i=1}^n$ obtained by (4) yield an operator of the form (3), which is an oblique projector. □

Property 1. *Let $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$ be the oblique projector onto $\mathcal{V}$ and along $\mathcal{W}^\perp$ and $\hat{P}_{\mathcal{W}}$ the orthogonal projector onto $\mathcal{W} = (\mathcal{W}^\perp)^\perp$. Then the following relation holds*

$$\hat{P}_{\mathcal{W}} \hat{E}_{\mathcal{V}\mathcal{W}^\perp} = \hat{P}_{\mathcal{W}}.$$

Proof. $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$ given in (3) can be recast, in terms of operator $\hat{V}$ and $\hat{U}^*$, as:

$$\hat{E}_{\mathcal{V}\mathcal{W}^\perp} = \hat{V}(\hat{U}^*\hat{V})^\dagger \hat{U}^*.$$

Applying $\hat{P}_{\mathcal{W}}$ both sides of the equation we obtain:

$$\hat{P}_{\mathcal{W}} \hat{E}_{\mathcal{V}\mathcal{W}^\perp} = \hat{P}_{\mathcal{W}} \hat{V}(\hat{U}^*\hat{V})^\dagger \hat{U}^* = \hat{U}(\hat{U}^*\hat{V})^\dagger \hat{U}^* = \hat{U}(\hat{U}^*\hat{U})^\dagger \hat{U}^*,$$

which is a well known form for the orthogonal projector onto $\mathcal{R}(\hat{U}) = \mathcal{W}$. $\square$

Remark 2. *Notice that the operative steps for constructing an oblique projector are equivalent to those for constructing an orthogonal one. The difference being that in general the spaces* $\mathrm{span}\{v_i\}_{i=1}^M = \mathcal{V}$ *and* $\mathrm{span}\{w_i\}_{i=1}^M = \mathcal{W}$ *are different. For the special case $u_i = v_i$, $i = 1, \ldots, M$, both sets of vectors span $\mathcal{V}$ and we have an* orthogonal projector *onto $\mathcal{V}$ along $\mathcal{V}^\perp$.*

Example 3

Suppose that the chirp signal in the first graph of Fig. 1 is corrupted by impulsive noise belonging to the subspace

$$\mathcal{W}^\perp = \mathrm{span}\{y_j(t) = e^{-100000(t-0.05j)^2}, \quad t \in [0,1]\}_{j=1}^{200}.$$

The chirp after being corrupted by a realization of the noise consisting of 95 pulses taken randomly from elements of $\mathcal{W}^\perp$ is plotted in the second graph of Fig. 1.

Consider that the signal subspace is well represented by $\mathcal{V}$ given by

$$\mathcal{V} = \mathrm{span}\{v_{i+1}(t) = \cos \pi i t, \quad t \in [0,1]\}_{i=0}^{M=99}.$$

In order to eliminate the impulsive noise from the chirp we have to compute the measurement vectors $\{w_i\}_{i=1}^{100}$, here functions of $t \in [0,1]$, determining the appropriate projector. For this we first need a representation of $\hat{P}_{\mathcal{W}^\perp}$, which is obtained simply by transforming the set $\{y_j\}_{j=1}^{200}$ into an orthonormal set $\{o_j\}_{j=1}^{200}$ to have

$$\hat{P}_{\mathcal{W}^\perp} = \sum_{j=1}^{200} o_j \langle o_j, \cdot \rangle.$$

The function for constructing an orthogonal projector in a number of different ways is OrthProj.m.

With $\hat{P}_{\mathcal{W}^\perp}$ we construct vectors

$$u_{i+1}(t) = \cos \pi i t - \sum_{j=1}^{200} o_j(t) \langle o_j(t), \cos \pi i t \rangle, \quad i = 0, \ldots, 99, \quad t \in [0,1].$$

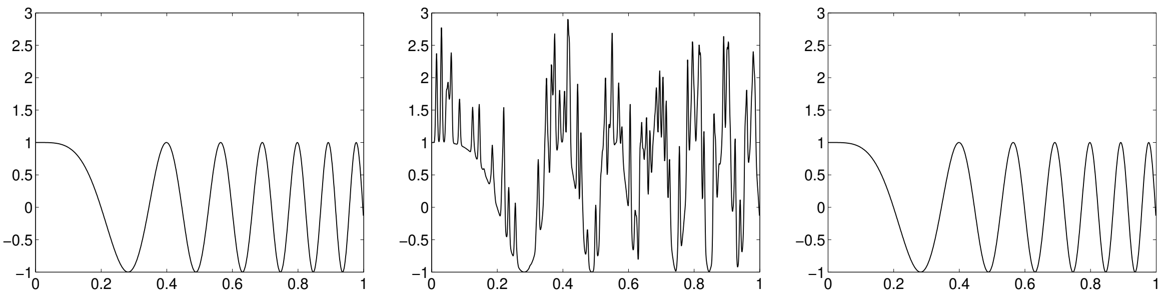

Figure 1: Chirp signal (first graph). Chirp corrupted by 95 randomly taken pulses (middle graph). Chirp denoised by oblique projection (last graph).

The inner products involved in the above equations and in the elements, $g_{i,j}$, of matrix G

$$g_{i+1,j+1} = \int_0^1 u_{i+1}(t) \cos \pi j t \, dt, \quad i = 0, \ldots, 99, \; j = 0, \ldots, 99$$

are computed numerically. This matrix has an inverse, which is used to obtain functions $\{w_i(t), \, t \in [0,1]\}_{i=1}^{100}$ giving rise to the required oblique projector. The chirp filtered by such a projector is depicted in the last graph of Fig. 1.

Example 4

Here we deal with an image of a poster in memory of the Spanish Second Republic shown in the first picture of Fig. 2. This image is an array of 609×450 pixels that we have processed row by row. Each row is a vector $\mathbf{I}_i \in \mathbb{R}^{450}$, $i = 1, \ldots, 609$. The image is affected by structured noise produced when random noise passes through a channel characterized by a given matrix A having 160 columns and 450 rows. The model for each row $\mathbf{I}_i^o \in \mathbb{R}^{609}$ of the noisy image is

$$\mathbf{I}_i^o = \mathbf{I}_i + A\mathbf{h}_i, \quad i = 1, \ldots, 450,$$

where each $\mathbf{h}_i$ is a random vector in $\mathbb{R}^{160}$. The image plus noise is represented in the middle graph of Fig. 2. In order to denoised the image we consider that every row $\mathbf{I}_i \in \mathbb{R}^{450}$ is well represented in a subspace $\mathcal{V}$ spanned by discrete cosines. More precisely, we consider $\mathbf{I}_i \in \mathcal{V}$ for $i = 1, \ldots, 450$, where

$$\mathcal{V} = \operatorname{span} \left\{ \mathbf{x}_i = \cos\left(\frac{\pi(2j-1)(i-1)}{2L} \right), \quad j = 1, \ldots, 609 \right\}_{i=1}^{290}.$$

The space of the noise is spanned by the 160 vectors in $\mathbb{R}^{450}$ corresponding to the columns of the given matrix A. The image, after being filtered row by row by the oblique projector onto $\mathcal{V}$ and along the space of the noise, is depicted in the last graph of Fig. 2.

Possible constructions of oblique projector

Notice that the oblique projector onto $\mathcal{V}$ is independent of the selection of the spanning set for $\mathcal{W}$. The possibility of using different spanning sets yields a number of different ways of computing $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$, all of them theoretically equivalent, but not necessarily numerically equivalent when the problem is ill posed.

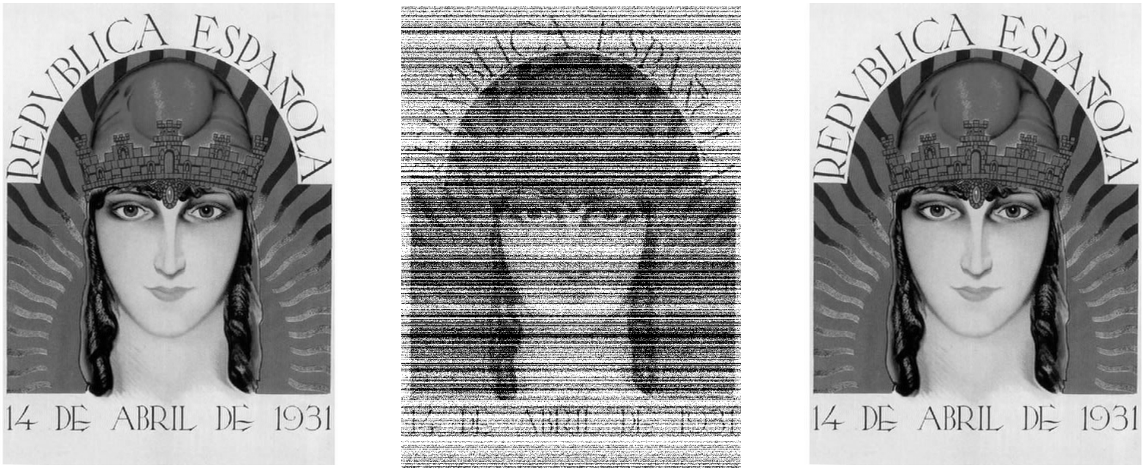

Figure 2: Image of a poster in memory of the Spanish Second Republic (first picture). Image plus structured noise (middle picture). The image obtained from the middle picture by an oblique projection (last picture).

Given the sets $\{v_i\}_{i=1}^M$ and $\{u_i\}_{i=1}^M$ we have considered the following theoretically equivalent ways of computing vectors $\{w_i\}_{i=1}^M$.

i) $w_i = \sum_{j=1}^M \tilde{g}_{i,j} u_j$, where $\tilde{g}_{i,j}$ is the (i,j)-th element of the inverse of the matrix G having elements $g_{i,j} = \langle u_i, v_j \rangle$, $i,j = 1, \ldots, M$.

ii) Vectors $\{w_i\}_{i=1}^M$ are as in i) but the matrix elements of G are computed as $g_{i,j} = \langle u_i, u_j \rangle$, $i,j = 1, \ldots, M$.

iii) Orthonormalising $\{u_i\}_{i=1}^M$ to obtain $\{q_i\}_{i=1}^{M'}$ $M' \leq M$ vectors $\{w_i\}_{i=1}^M$ are then computed as

$$w_i = \sum_{j=1}^{M} \tilde{g}_{i,j} q_j,$$

with $g_{i,j} = \langle q_i, v_j \rangle$, $i,j = 1, \ldots, M$.

iv) Same as in iii) but $g_{i,j} = \langle q_i, u_j \rangle$, $i,j = 1, \ldots, M$.

Moreover, considering that $\psi_n \in \mathbb{F}^M$, $n = 1, \ldots, M$, are the eigenvectors of $\hat{G}$ and assuming that there exist N nonzero eigenvalues on ordering these eigenvalues in descending order λ_n, $n = 1, \ldots, N$, we can express the matrix elements of the Moore-Penrose pseudo inverse of G as:

$$g^{\dagger}_{i,j} = \sum_{n=1}^{N} \psi_n(i) \frac{1}{\lambda_n} \psi_n^*(j), \tag{10}$$

with $\psi_n(i)$ the i-th component of ψ_n. Moreover, the orthonormal vectors

$$\xi_n = \frac{\hat{W} \psi_n}{\sigma_n}, \quad \sigma_n = \sqrt{\lambda_n}, \quad n = 1, \ldots, N \tag{11}$$

are singular vectors of $\hat{W}$, which satisfies $\hat{W}^*\xi_n = \sigma_n\psi_n$, as it is immediate to verify. By defining now the vectors η_n, $n = 1, \ldots, N$ as

$$\eta_n = \frac{\hat{V}\psi_n}{\sigma_n}, \quad n = 1, \ldots, N, \tag{12}$$

the projector $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$ in (3) is recast

$$\hat{E}_{\mathcal{V}\mathcal{W}^\perp} = \sum_{n=1}^{N} \eta_n \langle \xi_n, \cdot \rangle. \tag{13}$$

Inversely, the representation (3) of $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$ arises from (13), since

$$w_i = \sum_{n=1}^{N} \xi_n \frac{1}{\sigma_n} \psi_n^*(i), \quad i = 1, \ldots, M. \tag{14}$$

Proposition 2. *The vectors $\xi_n \in \mathcal{W}$, $n = 1, \ldots, N$ and $\eta_n \in \mathcal{V}$, $n = 1, \ldots, N$ given in* (11) *and* (12) *are biorthogonal to each other and span $\mathcal{W}$ and $\mathcal{V}$, respectively.*

The proof of this proposition can be found in [7] Appendix A.

All the different numerical computations for an oblique projector discussed above can be realized with the routine ObliProj.m.

The need of nonlinear approaches

Even when the subspaces $\mathcal{V}$ and $\mathcal{W}^\perp$ are 'theoretically' complementary, in practice, due to the calculations being performed in finite precision arithmetics, the inaccuracy in the computation of the corresponding projector may cause the failure to correctly separate signals in $\mathcal{V}$ and $\mathcal{W}^\perp$. The next example illustrates the situation.

Example 5

Let $\mathcal{V}$ be the cardinal cubic spline space with distance 0.01 between consecutive knots, on the interval $[0, 1]$. This is a subspace of dimension $M = 103$, which we span using a B-spline basis

$$B = \{B_i(x),\, x \in [0, 1]\}_{i=1}^{103}$$

The functions $B_i(x)$ in V are obtained by translations of a prototype function and the restriction to the interval $[0, 1]$. A few of such functions are plotted in the left graph of Fig. 3.

Taking, randomly, 30 B-splines $\{B_{\ell_i}\}_{i=1}^{30}$ from B we simulate a signal by a weighted superposition of such functions, i.e.,

$$f_{\mathcal{V}}(x) = \sum_{i=1}^{30} c_{\ell_i} B_{\ell_i}(x), \quad x \in [0, 1], \tag{15}$$

with the coefficients c_{ℓ_i} randomly chosen from $[0, 1]$.

We simulate a background by considering it belongs to the subspace $\mathcal{W}^\perp$ spanned by the set of functions

$$Y = \{y_j(x) = (x + 0.01j)^{-0.01j},\, x \in [0, 1]\}_{j=1}^{50}.$$

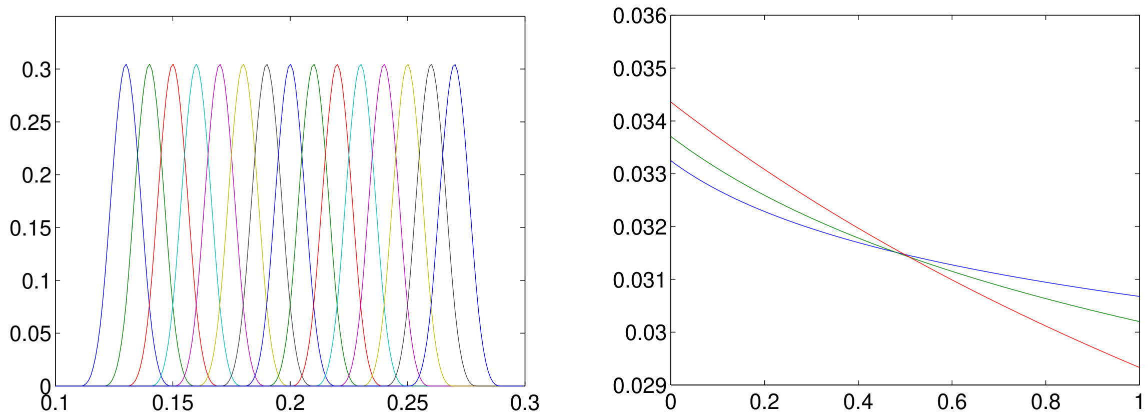

Figure 3: Left graph: cubic B spline functions, in the rage $x \in [0.1, 0.3]$, from the set spanning the space of the signal response. Right graph: tree of the functions spanning the space of the background.

A few functions from this set are plotted in the right graph of Fig. 3 (normalized to unity on $[0, 1]$). The background, $g(x)$ is generated by the linear combination

$$g(x) = \sum_{j=1}^{50} j^4 e^{-0.05j} y_j(x) \tag{16}$$

To simulate the data we have perturbed the superposition of (15) and (16), by 'very small' Gaussian errors (of variance up to 0.00001% the value of each data point). The simulated data are plotted in the left graph of Fig. 4.

This example is very illustrative of how sensitive to errors the oblique projection is. The subspace we are dealing with are disjoint: the last five singular values of operator $\hat{W}^*$ (c.f. (11)) are:

$$0.3277, 0.3276, 1.0488 \times 10^{-4}, 6.9356 \times 10^{-8}, 2.3367 \times 10^{-10},$$

while the first is $\sigma_1 = 1.4493$. The smallest singular value cannot be considered a numerical representation of zero, when the calculations are being carried out in double precision arithmetic. Hence, one can assert that the condition $\mathcal{V} \cap \mathcal{W}^\perp = \{0\}$ is fulfilled. However, due to the three small singular values the oblique projector along $\mathcal{W}^\perp$ onto the whole subspace $\mathcal{V}$ is very unstable, which causes the failure to correctly separate signals in $\mathcal{V}$ from the background. The result of applying the oblique projector onto the signal of the left graph is represented by the broken line in the right graph. As can be observed, the projection does not yield the required signal, which is represented by the continuous dark line in the same graph. Now, since the spectrum of singular values has a clear jump (the last three singular values are far from the previous ones) it might seem that one could regularize the projection by truncation of singular values. Nevertheless, such a methodology turns out to be not appropriate for the present problem, as it does not yield the correct separation. Propositions 3 below analyzes the effect that regularization by truncation of singular values produces in the resulting projection.

Proposition 3. *Truncation of the expansion* (13) *to consider up to r terms, produces an oblique projector along* $\tilde{\mathcal{W}}^\perp = \tilde{\mathcal{W}}_r^\perp + \tilde{\mathcal{W}}_0 + \tilde{\mathcal{V}}_0$, *with* $\tilde{\mathcal{W}}_r^\perp = \mathrm{span}\{|\xi_i\rangle\}_{i=1}^r, \tilde{\mathcal{W}}_0 = \mathrm{span}\{|\xi_i\rangle\}_{i=r+1}^N$ *and* $\tilde{\mathcal{V}}_0 = \mathrm{span}\{|\eta_i\rangle\}_{i=r+1}^N$ *onto* $\tilde{\mathcal{V}}_r = \mathrm{span}\{|\eta_i\rangle\}_{i=1}^r$.

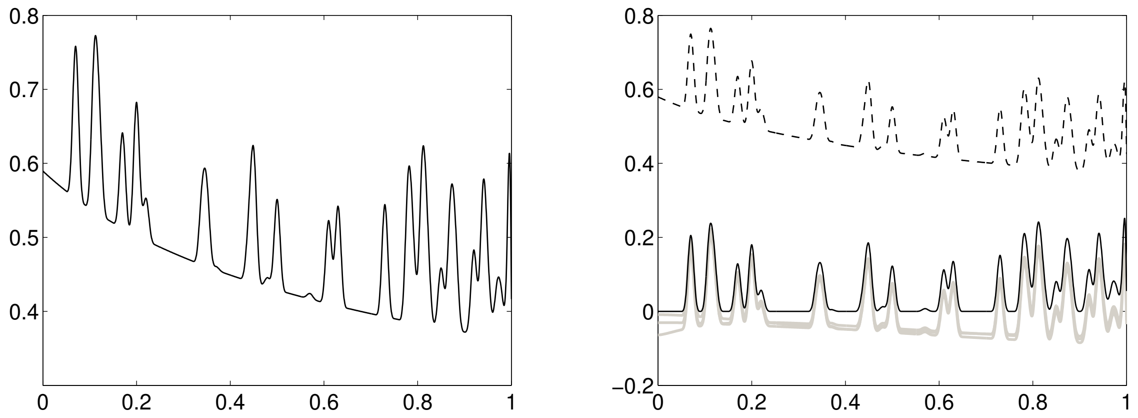

Figure 4: Left graph: signal plus background. Right graph: the dark continuous line corresponds to the signal to be discriminated from the one in the left graph. The broken line corresponds to the approximation resulting from the oblique projection. The three closed lines correspond to the approximations obtained by truncation of one, two, and three singular values.

The proof of these propositions can be found in [7] Appendix B. For complete studies of a projector when $\mathcal{V} \cap \mathcal{W}^\perp \neq \{0\}$ see [8] and [9].

This example illustrates, very clearly, the need for nonlinear approaches: We know that a unique and stable solution does exist, since the signal which is to be discriminated from the background actually belongs to a subspace of the given spline space, and the construction of the oblique projectors onto such a subspace is well posed. Whoever, the lack of knowledge about the subspace prevents the separation of the signal components by a linear operation. A possibility to tackle the problem is to transform it into the one of finding the subspace to which the sought signal component belongs to. In this way the problem can be addressed by nonlinear techniques.

Recursive updating/downdating of oblique projectors

Here we provide the equations for updating and downdating oblique projectors in order to account for the following situations:

Let us consider that the oblique projector $\hat{E}_{\mathcal{V}_k\mathcal{W}^\perp}$ onto the subspace $\mathcal{V}_k = \text{span}\{v_i\}_{i=1}^k$ along a given subspace $\mathcal{W}^\perp$ is known. If the subspace $\mathcal{V}_k$ is enlarged to $\mathcal{V}_{k+1}$ by the inclusion of one element, i.e., $\mathcal{V}_{k+1} = \text{span}\{v_i\}_{i=1}^{k+1}$, we wish to construct $\hat{E}_{\mathcal{V}_{k+1}\mathcal{W}^\perp}$ from the availability of $\hat{E}_{\mathcal{V}_k\mathcal{W}^\perp}$. On the other hand, if the subspace $\mathcal{V}_k = \text{span}\{v_i\}_{i=1}^k$ is reduced by the elimination of one element, say the j-th one, we wish to construct the corresponding oblique projector $\hat{E}_{\mathcal{V}_{k\setminus j}\mathcal{W}^\perp}$ from the knowledge of $\hat{E}_{\mathcal{V}_k\mathcal{W}^\perp}$. The subspace $\mathcal{W}^\perp$ is assumed to be fixed. Its orthogonal complement $\mathcal{W}_k$ in $\mathcal{H}_k = \mathcal{V}_k \oplus \mathcal{W}^\perp$ changes with the index k to satisfy $\mathcal{H}_k = \mathcal{W}_k \oplus^\perp \mathcal{W}^\perp$.

Updating the oblique projector $\hat{E}_{\mathcal{V}_k\mathcal{W}^\perp}$ to $\hat{E}_{\mathcal{V}_{k+1}\mathcal{W}^\perp}$

We assume that $\hat{E}_{\mathcal{V}_k\mathcal{W}^\perp}$ is known and write it in the explicit form

$$\hat{E}_{\mathcal{V}_k\mathcal{W}^\perp} = \sum_{i=1}^{k} v_i \langle w_i^k, \cdot \rangle. \tag{17}$$

In order to inductively construct the duals w_i^{k+1}, $i = 1, \ldots, k+1$ we have to discriminate two possibilities

i) $\mathcal{V}_{k+1} = \mathrm{span}\{v_i\}_{i=1}^{k+1} = \mathrm{span}\{v_i\}_{i=1}^{k} = \mathcal{V}_k$, i.e., $v_{k+1} \in \mathcal{V}_k$.

ii) $\mathcal{V}_{k+1} = \mathrm{span}\{v_i\}_{i=1}^{k+1} \supset \mathrm{span}\{v_i\}_{i=1}^{k} = \mathcal{V}_k$, i.e. $v_{k+1} \notin \mathcal{V}_k$.

Case i)

Proposition 4. *Let $v_{k+1} \in \mathcal{V}_k$ and vectors w_i^k in* (17) *be given. For an arbitrary vector $y_{k+1} \in \mathcal{H}$ the dual vectors w_i^{k+1} computed as*

$$w_i^{k+1} = w_i^k - \langle u_{k+1}, w_i^k \rangle y_{k+1} \tag{18}$$

for $i = 1, \ldots, k$ and $w_{k+1}^{k+1} = y_{k+1}$ produce the identical oblique projector as the dual vectors $w_i^k, i = 1, \ldots, k$.

Case ii)

Proposition 5. *Let vector $v_{k+1} \notin \mathcal{V}_k$ and vectors $w_i^k, i = 1, \ldots, k$ in* (17) *be given. Thus the dual vectors w_i^{k+1} computed as*

$$w_i^{k+1} = w_i^k - w_{k+1}^{k+1} \langle u_{k+1}, w_i^k \rangle, \tag{19}$$

where $w_{k+1}^{k+1} = \frac{\gamma_{k+1}}{\|\gamma_{k+1}\|^2}$ with $\gamma_{k+1} = u_{k+1} - \hat{P}_{\mathcal{W}_k} u_{k+1}$, provide us with the oblique projector $\hat{E}_{\mathcal{V}_{k+1}\mathcal{W}^\perp}$.

The proof these propositions are given in [10]. The codes for updating the dual vectors are FrInsert.m and FrInsertBlock.m.

Property 2. *If vectors $\{v_i\}_{i=1}^k$ are linearly independent they are also biorthogonal to the dual vectors arising inductively from the recursive equation* (19)*.*

The proof of this property is in [10].

Remark 3. *If vectors $\{v_i\}_{i=1}^k$ are not linearly independent the oblique projector $\hat{E}_{\mathcal{V}_k\mathcal{W}^\perp}$ is not unique. Indeed, if $\{w_i^k\}_{i=1}^k$ are dual vectors giving rise to $\hat{E}_{\mathcal{V}_k\mathcal{W}^\perp}$ then one can construct infinitely many duals as:*

$$\tilde{y}_i = w_i^k + y_i - \sum_{j=1}^{k} y_j \langle v_j, w_i^k \rangle \quad i = 1, \ldots, k, \tag{20}$$

where y_i, $i = 1, \ldots, k$ are arbitrary vectors in $\mathcal{H}$ (see [10]).

Downdating the oblique projector $\hat{E}_{\mathcal{V}_k\mathcal{W}^\perp}$ to $\hat{E}_{\mathcal{V}_{k\setminus j}\mathcal{W}^\perp}$

Let us suppose that by the elimination of the element j the subspace $\mathcal{V}_k$ is reduced to $\mathcal{V}_{k\setminus j} = \text{span}\{v_i\}_{\substack{i=1\\i\neq j}}^k$. In order to give the equations for adapting the corresponding dual vectors generating the oblique projector $\hat{E}_{\mathcal{V}_{k\setminus j}\mathcal{W}^\perp}$ we need to consider two situations:

i) $\mathcal{V}_{k\setminus j} = \text{span}\{v_i\}_{\substack{i=1\\i\neq j}}^k = \text{span}\{v_i\}_{i=1}^k = \mathcal{V}_k$ i.e., $v_j \in \mathcal{V}_{k\setminus j}$.

ii) $\mathcal{V}_{k\setminus j} = \text{span}\{v_i\}_{\substack{i=1\\i\neq j}}^k \subset \text{span}\{v_i\}_{i=1}^k = \mathcal{V}_k$, i.e., $v_j \notin \mathcal{V}_{k\setminus j}$.

Case i)

Proposition 6. *Let $\hat{E}_{\mathcal{V}_k\mathcal{W}^\perp}$ be given by* (17) *and let us assume that removing vector v_j from the spanning set of $\mathcal{V}_k$ leaves the identical subspace, i.e., $v_j \in \mathcal{V}_{k\setminus j}$. Hence, if the remaining dual vectors are modified as follows:*

$$w_i^{k\setminus j} = w_i^k + \frac{\langle u_j, w_i^k\rangle w_j^k}{1 - \langle u_j, w_j^k\rangle}, \tag{21}$$

the corresponding oblique projector does not change, i.e. $\hat{E}_{\mathcal{V}_{k\setminus j}\mathcal{W}^\perp} = \hat{E}_{\mathcal{V}_k\mathcal{W}^\perp}$.

Case ii)

Proposition 7. *Let $\hat{E}_{\mathcal{V}_k\mathcal{W}^\perp}$ be given by* (17) *and let us assume that the vector v_j to be removed from the spanning set of $\mathcal{V}_k$ is not in $\mathcal{V}_{k\setminus j}$. In order to produce the oblique projector $\hat{E}_{\mathcal{V}_{k\setminus j}\mathcal{W}^\perp}$ the appropriate modification of the dual vectors can be achieved by means of the following equation*

$$w_i^{k\setminus j} = w_i^k - \frac{w_j^k\langle w_j^k, w_i^k\rangle}{\|w_j^k\|^2}. \tag{22}$$

The proof these propositions are given in [10]. The code for updating the vectors are FrDelete.m.

Signals discrimination by subspace selection

We discuss now the possibility of extracting the signal $f_\mathcal{V}$, from a mixture $f = f_\mathcal{V} + f_\mathcal{V}$ when the subspaces $\mathcal{V}$ and $\mathcal{W}^\perp$ are not well separated and the oblique projector onto $\mathcal{V}$ along $\mathcal{W}^\perp$ fails to yield the right signals separation. For this we introduce the following hypothesis on the class of signals to be considered:

We assume that the signal of interest is K-sparse in a spanning set for $\mathcal{V}$

This implies that given $\{v_i\}_{i=1}^M$ there exists a subset of elements characterized by the set of indices J, of cardinality $K < M$, spanning the subspace $\tilde{\mathcal{V}} = \text{span}\{v_\ell\}_{\ell\in J}$ and such that $f_\mathcal{V} = \hat{E}_{\tilde{\mathcal{V}}\mathcal{W}^\perp} f$. Thus, the hypothesis generates an, in general, intractable problem because the number of possible subspaces spanned by K vectors out of M is a combinatorial number $\binom{M}{K} = \frac{M!}{(M-K)!K!}$.

The techniques developed within the project aim at reducing complexity by making the search for the right subspace signal dependent.

Given a signal f, and assuming that the subspaces $\mathcal{W}^\perp$ and $\mathcal{V} = \mathrm{span}\{v_i\}_{i=1}^M$, are known, the goal is to find $\{v_\ell\}_{\ell \in J} \subset \{v_i\}_{i=1}^M$ spanning $\tilde{\mathcal{V}}$ and such that $\hat{E}_{\tilde{\mathcal{V}}\mathcal{W}^\perp} f = \hat{E}_{\mathcal{V}\mathcal{W}^\perp} f$. The cardinality of the subset of indices J should be such that construction of $\hat{E}_{\tilde{\mathcal{V}}\mathcal{W}^\perp}$ is well posed. This assumption characterizes the class of signals that can be handled by the proposed approaches.

Under the stated hypothesis, if the subspace $\tilde{\mathcal{V}}$ were known, one would have

$$\hat{E}_{\mathcal{V}\mathcal{W}} f = \hat{E}_{\tilde{\mathcal{V}}\mathcal{W}^\perp} f = \sum_{\ell \in J} v_\ell \langle w_\ell, f \rangle. \tag{23}$$

However, if the computation of $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$ is an ill posed problem, which is the situation we are considering, $\hat{E}_{\mathcal{V}\mathcal{W}^\perp} f$ is not available. In order to look for the subset of indices J yielding $\tilde{\mathcal{V}}$ one may proceed as follows: Applying $\hat{P}_{\mathcal{W}}$ on every term of (23) and using the properties $\hat{P}_{\mathcal{W}} \hat{E}_{\mathcal{V}\mathcal{W}^\perp} = \hat{P}_{\mathcal{W}}$ and $\hat{P}_{\mathcal{W}} \hat{E}_{\tilde{\mathcal{V}}\mathcal{W}^\perp} = \hat{P}_{\tilde{\mathcal{W}}}$, where $\tilde{\mathcal{W}} = \mathrm{span}\{u_\ell\}_{\ell \in J}$, (23) becomes

$$\hat{P}_{\mathcal{W}} f = \hat{P}_{\tilde{\mathcal{W}}} f = \sum_{\ell \in J} u_\ell \langle w_\ell, f \rangle. \tag{24}$$

Since $\mathcal{W}^\perp$ is given and $\hat{P}_{\mathcal{W}} f = f - \hat{P}_{\mathcal{W}^\perp} f$, the left hand side of (23) is available and we can search for the set $\{u_\ell\}_{\ell \in J} \subset \{u_i\}_{i=1}^M$, in a stepwise manner by adaptive pursuit approaches.

Oblique Matching Pursuit (OBMP)

The criterion we use for the forward recursive selection of the set $\{v_\ell\}_{\ell \in J} \subset \{v_i\}_{i=1}^M$ yielding the right signal separation is in line with the *consistency principle* introduced in [2] and extended in [3]. Furthermore, it happens to coincide with the Optimize Orthogonal Matching Pursuit (OOMP) [12] approach applied to find the sparse representation of the projected signal $\hat{P}_{\mathcal{W}} f$ using the dictionary $\{u_i\}_{i=1}^M$

By fixing $\hat{P}_{\mathcal{W}_k}$, at iteration $k+1$ we select the index ℓ_{k+1} such that $\|\hat{P}_{\mathcal{W}} |f\rangle - \hat{P}_{\mathcal{W}_{k+1}} |f\rangle\|^2$ is minimized.

Proposition 8. *Let us denote by J the set of indices $\{\ell_1, \ldots, \ell_k\}$. Given $\mathcal{W}_k = \mathrm{span}\{|u_{\ell_i}\rangle\}_{i=1}^k$, the index ℓ_{k+1} corresponding to the atom $|u_{\ell_{k+1}}\rangle$ for which $\|\hat{P}_{\mathcal{W}} f - \hat{P}_{\mathcal{W}_{k+1}} f\|^2$ is minimal is to be determined as*

$$\ell_{k+1} = \arg \max_{n \in J \setminus J_k} \frac{|\langle \gamma_n | f \rangle|}{\|\gamma_n\|}, \ \|\gamma_n\rangle\| \neq 0, \tag{25}$$

with $\gamma_n = u_n - \hat{P}_{\mathcal{W}_k} u_n$ and J_k the set of indices that have been previously chosen to determine $\mathcal{W}_k$.

Proof. It readily follows since $\hat{P}_{\mathcal{W}_{k+1}} f = \hat{P}_{\mathcal{W}_k} f + \frac{\gamma_n \langle \gamma_n, f \rangle}{\|\gamma_n\|^2}$ and hence

$$\|\hat{P}_{\mathcal{W}} f - \hat{P}_{\mathcal{W}_{k+1}} f\|^2 = \|\hat{P}_{\mathcal{W}} f\|^2 - \|\hat{P}_{\mathcal{W}_k} f\|^2 - \frac{|\langle \gamma_n, f \rangle|^2}{\|\gamma_n\|^2}.$$

Because $\hat{P}_{\mathcal{W}} f$ and $\hat{P}_{\mathcal{W}_k} f$ are fixed, $\|\hat{P}_{\mathcal{W}} f - \hat{P}_{\mathcal{W}_{k+1}} f\|^2$ is minimized if $\frac{|\langle \gamma_n, f \rangle|}{\|\gamma_n\|}$, $\|\gamma_n\| \neq 0$ is maximal over all $n \in J \setminus J_k$. ∎

The original OBMP selection criterion proposed in [11] selects the index ℓ_{k+1} as the maximizer over $n \in J \setminus J_k$ of

$$\frac{|\langle \gamma_n, f \rangle|}{\|\gamma_n\|^2}, \quad \|\gamma_n\| \neq 0.$$

This condition was proposed in [11] based on the consistency principle which states that the reconstruction of a signal should be self consistent in the sense that, if the approximation is measured with the same vectors, the same measures should be obtained (see [2,3]). Accordingly, the above OBMP criterion was derived in [11] in order to select the measurement vector w_{k+1}^{k+1} producing the maximum *consistency error* $\Delta = |\langle w_{k+1}^{k+1}, f - \hat{E}_{\mathcal{V}_k\mathcal{W}^\perp} f\rangle|$, with regard to a new measurement w_{k+1}^{k+1}. However, since the measurement vectors are not normalized to unity, it is sensible to consider the consistency error relative to the corresponding vector norm $\|w_{k+1}^{k+1}\|$, and select the index so as to maximize over $k+1 \in J \setminus J_k$ the *relative consistency error*

$$\tilde{\Delta} = \frac{|\langle w_{k+1}^{k+1}, f - \hat{E}_{\mathcal{V}_k\mathcal{W}^\perp} f\rangle|}{\|w_{k+1}^{k+1}\|}, \quad \|w_{k+1}^{k+1}\| \neq 0. \tag{26}$$

In order to cancel this error, the new approximation is constructed accounting for the concomitant measurement vector.

Property 3. *The index ℓ_{k+1} satisfying (25) maximizes over $k+1 \in J \setminus J_k$ the relative consistency error* (26)

Proof. Since for all vector w_{k+1}^{k+1} given in (19) $\langle w_{k+1}^{k+1}, \hat{E}_{\mathcal{V}_k\mathcal{W}^\perp} f\rangle = 0$ and $\|w_{k+1}^{k+1}\| = \|\gamma_{k+1}\|^{-1}$ we have

$$\tilde{\Delta} = \frac{|\langle w_{k+1}^{k+1}, f\rangle|}{\|w_{k+1}^{k+1}\|} = \frac{|\langle \gamma_{k+1}, f\rangle|}{\|\gamma_{k+1}\|}.$$

Hence, maximization of $\tilde{\Delta}$ over $k+1 \in J \setminus J_k$ is equivalent to (25). □

It is clear at this point that the forward selection of indices prescribed by proposition (25) is equivalent to selecting the indices by applying OOMP [12] on the projected signal $\hat{P}_{\mathcal{W}} f$ using the dictionary $\{u_i\}_{i=1}^M$. The routine for implementing the pursuit strategy for subspace selection according to criterion (25) is OBMP.m. An example of application is given in .

Implementing corrections

Let us discuss now the possibility of correcting bad moves in the forward selection, which is specially necessary when dealing with ill posed problems. Indeed, assume we are trying to approximate a signal which is K-sparse in a given dictionary, and the search for the right atoms become ill posed after the iteration, r, say, with $r > K$. The r-value just indicates that it is not possible to continue with the forward selection, because the computations would become inaccurate and unstable. Hence, if the right solution was not yet found, one needs to implement a strategy accounting for the fact that it is not feasible to compute more than r measurement vectors. A possibility is provided by the swapping-based refinement to the OOMP approach introduced in [13]. It consists of interchanging already selected atoms with nonselected ones.

Consider that at iteration r the correct subspace has not appeared yet and the selected indices are labeled by the r indices $\ell_1, \ldots, \ell_r$. In order to choose the index of the atom that minimizes the norm of the residual error as passing from approximation $\hat{P}_{\mathcal{W}_r} f$ to approximation $\hat{P}_{\mathcal{W}_{r\setminus j}} f$ we should fix the index of the atom to be deleted, ℓ_j say, as the one for which the quantity

$$\frac{|c_i^r|}{\|w_i^r\|}, \; i = 1, \ldots, r. \tag{27}$$

is minimized [13, 14].

The process of eliminating one atom from the atomic decomposition is called *backward step* while the process of adding one atom is called *forward step.* The forward selection criterion to choose the atom to replace the one eliminated in the previous step is accomplished by finding the index $\ell_i, i = 1, \dots, r$ for which the functional

$$e_n = \frac{|\langle \nu_n, f \rangle|}{\|\nu_n\|}, \quad \text{with} \quad \nu_n = u_n - \hat{P}_{\mathcal{W}_{r \setminus j}} u_n, \quad \|\nu_n\| \neq 0 \tag{28}$$

is maximized. In our framework, using (22), the projector $\hat{P}_{\mathcal{W}_{r\setminus j}}$ is computed as

$$\hat{P}_{\mathcal{W}_{r\setminus j}} = \hat{P}_{\mathcal{W}_r} - \frac{\langle w_i^r, w_j^r \rangle \langle w_j^r, \cdot \rangle}{\|w_j^r\|^2}.$$

The swapping of pairs of atoms is repeated until the swapping operation, if carried out, would not decrease the approximation error. The implementation details for an effective realization of this process are given in [13]. Moreover, the process has been extended to include the swapping of more than a pair of atoms. This possibility is of assistance in the application of signal splitting, see [15]. A number of codes for implementing correction to Pursuit Startegis can be found in Pursuits.

Sparse representation by minimization of the $q-$norm like quantity - Handling the ill posed case.

The problem of finding the sparsest representation of a signal for a given dictionary is equivalent to minimization of the zero norm $\|c\|_0$ (or counting measure) which is defined as:

$$\|c\|_0 = \sum_{i=1}^{M} |c_i|^0,$$

where c_i are the coefficients of the atomic decomposition

$$f_{\mathcal{V}} = \sum_{i=1}^{M} c_i v_i. \tag{29}$$

Thus, $\|c\|_0$ is equal to the number of nonzero coefficients in (29). However, sice minimization of $\|c\|_0$ is numerically intractable, the minimization of $\sum_{i=1}^{M} |c_i|^q$, for $0 < q \leq 1$ has been considered [16]. Because the minimization of $\sum_{i=1}^{M} |c_i|^q$, $0 < q < 1$ does not lead to a convex optimization problem, the most popular norm to minimize, when a sparse solution is required, is the 1-norm $\sum_{i=1}^{M} |c_i|$. Minimization of the 1-norm is considered the best convex approximant to the minimizer of $\|c\rangle\|_0$ [17, 18]. In the context of signals splitting already stated, we are not particularly concerned about convexity so we have considered the minimization of $\sum_{i=1}^{M} |c_i|^q$, $0 < q \leq 1$, allowing for uncertainty in the available data [7]. This was implemented by a recursive process for incorporating constrains, which is equivalent to the procedure introduced in [19] and applied in [20].

Managing the constraints

Without loss of generality we assume here that the measurements on the signal in hand are given by the values the signal takes at the sampling points x_j, $j = 1, \ldots, N$. Thus, the measures are obtained from $f_{\mathcal{W}}$ as $f_{\mathcal{W}}(x_j)$, $j = 1, \ldots, N$ and the functionals $u_i(x_j)$, $j = 1, \ldots, N$ from vectors u_i, $i = 1, \ldots, M$. Since the values $f_{\mathcal{W}}(x_j)$, $j = 1, \ldots, N$ arise from observations or experiments they are usually affected by errors therefore we use the notation $f^o_{\mathcal{W}}(x_j)$, $j = 1, \ldots, N$ for the observed data and request that the model given by the r.h.s. of (29) satisfies the restriction

$$\sum_{j=1}^{N} (f^o_{\mathcal{W}}(x_j) - f_{\mathcal{W}}(x_j))^2 \leq \delta, \tag{30}$$

δ accounting for the data's error. Nevertheless, rather than using directly this restriction as constraints of the $q-$norm^q we handle the available information using an idea introduced much earlier, in [19], and applied in [20] for transforming the constraint (30) into linear equality constraints. Replacing $f_{\mathcal{W}}(x_j)$ by (29), the condition of minimal square distance $\sum_{j=1}^{N}(f^o_{\mathcal{W}}(x_j) - f_{\mathcal{W}}(x_j))^2$ leads to the so called *normal equations*:

$$\langle u_n, f^o_W \rangle = \sum_{i=1}^{M} c_i \langle u_n, u_i \rangle, \quad n = 1 \ldots, M. \tag{31}$$

Of course, since we are concerned with ill posed problems we cannot use all these equations to find the coefficients c_i, $i = 1, \ldots, M$. However, as proposed in [19], we could use 'some' of these equations as constraints of our optimization process. The number of such equations being the necessary to reach the condition (30). We have then transformed the original problem into the one of minimizing $\sum_{i=1}^{M} |c_i|^q$, $0 < q \leq 1$, subject to a number of equations selected from (31), the ℓ_n-th, $n = 1 \ldots, r$ ones say. In line with [19] we select the subset of equations (31) in an iterative fashion. We start by the initial estimation $c_i = C$, $i = 1, \ldots, M$, where the constant C is determined by minimizing the distant between the model and the data. Thus,

$$C = \frac{\sum_{n=1}^{M} \langle u_n, f^o_{\mathcal{W}} \rangle}{\sum_{i=1}^{M} \sum_{n=1}^{M} \langle u_i, u_n \rangle}. \tag{32}$$

With this initial estimation we 'predict' the normal equations (31) and select as our first constraint the worst predicted by the initial solution, let this equation be the ℓ_1-th one. We then minimize $\sum_{i=1}^{M} |c_i|^q$ subject to the constraint

$$\langle u_{\ell_1}, f^o_{\mathcal{W}} \rangle = \sum_{i=1}^{M} c_i \langle u_{\ell_1}, u_i \rangle, \tag{33}$$

and indicate the resulting coefficients as $c_i^{(1)}$, $i = 1, \ldots, M$. With these coefficients we predict equations (31) and select the worst predicted as a new constraint to obtain $c_i^{(2)}$, $i = 1, \ldots, M$ and so on. The iterative process is stopped when the condition (30) is reached.

The numerical example discussed next has been solved by recourse to the method for minimization of the (q-norm)^q published in [16]. Such an iterative method, called FOCal Underdetermined System Solver (FOCUSS) in that publication, is straightforward implementable. It evolves by computation of pseudoinverse matrices, which under the given hypothesis of our problem, and within our recursive strategy for feeding the constraints, are guaranteed to be numerically stable (for a detailed explanation of the method see [16]). The routine for implementing the proposed strategy is ALQMin.m.

Numerical Simulation

We test the proposed approaches, first on the simulation of Example and then extend that simulation to consider a more realistic level of uncertainty in the data. Let us remark that the signal is meant to represent an emission spectrum consisting of the superposition of spectral lines (modeled by B-spline functions of support 0.04) which are centered at the positions $(n-1)\Delta$, $n = 0, \ldots, 102$, with $\Delta = 0.01$. Since the errors in the data in Example are not significant, both OBMP and the procedure outlined in the previous section accurately recovers the spectrum from the background, with any positive value of the q-parameter less than or equal to one. The result (coinciding with the theoretical one) is shown in the right hand top graph of Fig. 5.

Now we transform the example into a more realistic situation by adding larger errors to the data. In this case, the data set is perturbed by Gaussian errors of variance up to 1% of each data point. Such a piece of data is plotted in the left middle graph of Fig. 3 and the spectrum extracted by the $q-$norm like approach (for $q = 0.8$) is represented by the broken line in the right middle graph of Fig. 5. The corresponding OBMP approach is plotted in the first graph of Fig. 6 and is slightly superior.

Finally we increase the data's error up to 3% of each data point (left bottom graph of Fig. 5) and, in spite of the perceived significant distortion of the signal, we could still recover a spectrum which, as shown by the broken line in the right bottom graph of Fig.5 is a fairly good approximation of the true one (continuous line). The OBMP approach is again superior, as can be observed in the second graph of Fig. 6. Experiments for different realization of the errors (with the same variance) have produced results essentially equivalent. The same is true for other realizations of the signal.

Adaptive non-uniform B-spline dictionaries

In this work, with Zhiqiang Xu [21],we consider the sparse representation matter for the large class of signals which are amenable to satisfactory approximation in spline spaces [**?**,22]. Given a signal, we have the double goal of a) finding a spline space for approximating the signal and b) constructing those dictionaries for the space which are capable of providing a sparse representation of such a signal. In order to achieve both aims we have first discussed the construction of dictionaries of B-spline functions for non-uniform partitions.

Background and notations

We refer to the fundamental books [23,25] for a complete treatment of splines. Here we simply introduce the notation and basic definitions which are needed in the present context.

Definition 1. *Given a finite closed interval $[c, d]$ we define a* **partition** *of $[c, d]$ as the finite set of points*

$$\Delta := \{x_i\}_{i=0}^{N+1}, N \in \mathbb{N}, \text{ such that } c = x_0 < x_1 < \cdots < x_N < x_{N+1} = d. \tag{34}$$

We further define N subintervals $I_i, i = 0, \ldots, N$ as: $I_i = [x_i, x_{i+1}), i = 0, \ldots, N-1$ and $I_N = [x_N, x_{N+1}]$.

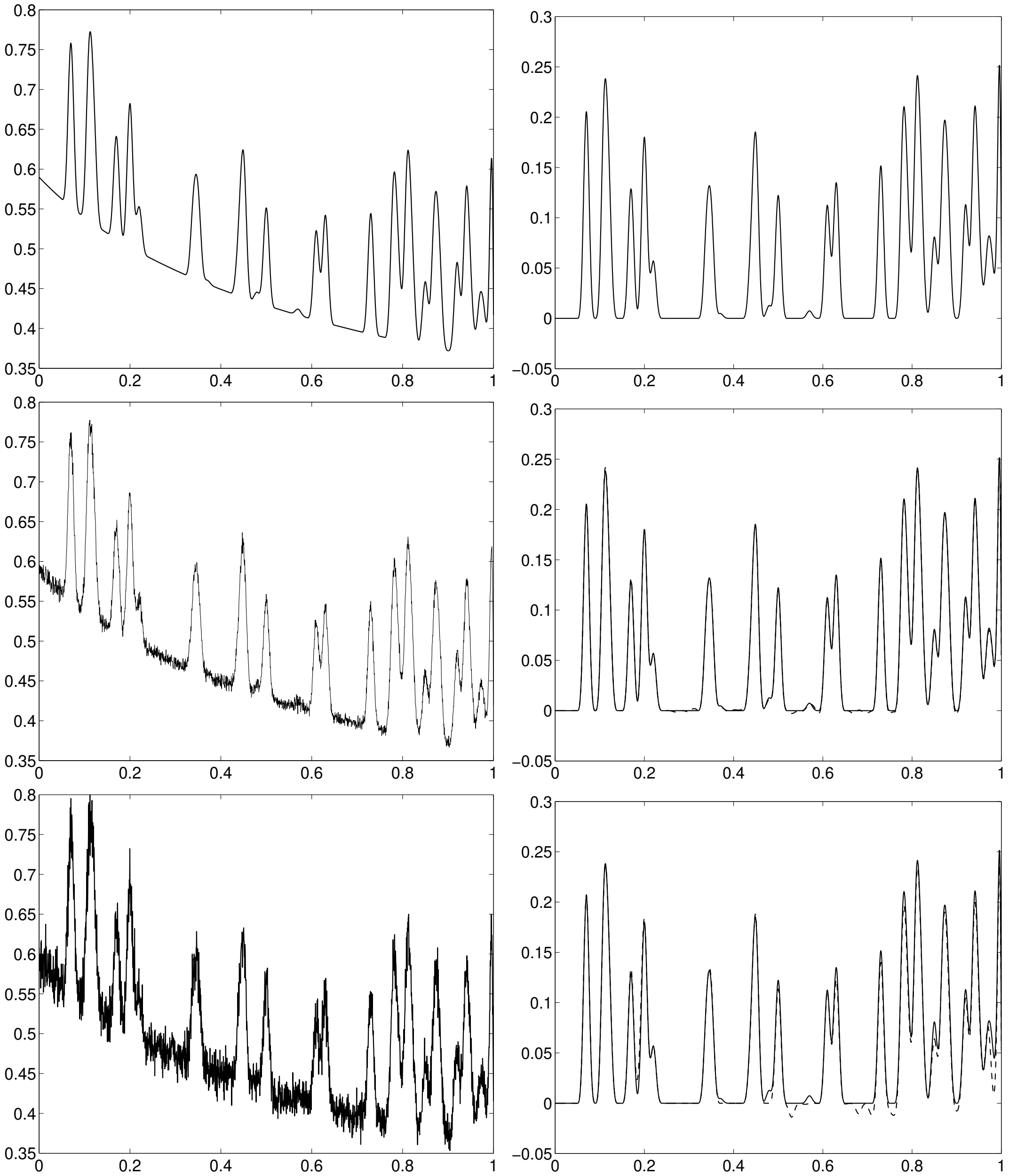

Figure 5: Top left graph: signal plus background. Top right graph: Recovered signal Middle left graph: signal distorted up to to 1%. Middle right graph: q-norm like approach approximation (broken line) Bottom graphs: Same description as in the previous graphs but the data distorted up to 3%.

Definition 2. *Let Π_m be the space of polynomials of degree smaller than or equal to $m \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$, i.e.,*

$$\{\Pi_m(x) : \Pi_m(x) = \sum_{i=0}^{m} a_i x^i, \;\; x \in \mathbb{R}\Big\},$$

and define

$$S_m(\Delta) = \{f \in C^{m-2}[c,d] \;\; : \;\; f|_{I_i} \in \Pi_{m-1}, i = 0, \dots, N\}, \tag{35}$$

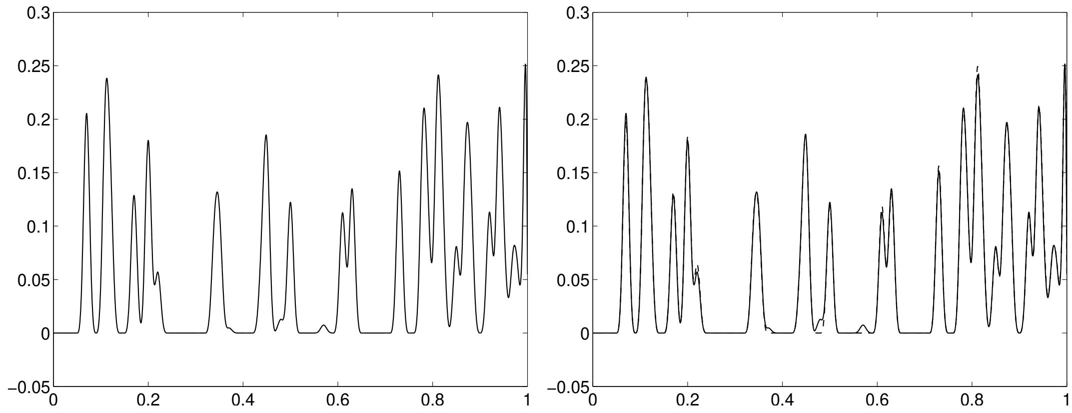

Figure 6: Same description as the right middle and bottom graphs of Fig. 5 but applying the BOMP method discussed in Section

where $f|_{I_i}$ indicates the restriction of the function f on the interval I_i.

The construction of nonuniform dictionaries arises from the next proposition [21]

Theorem 1. *Suppose that $\Delta_j, j = 1, \dots, n$ are partitions of $[c, d]$. Then*

$$S_m(\Delta_1) + \cdots + S_m(\Delta_n) \;=\; S_m(\cup_{j=1}^n \Delta_j).$$

Building B-spline dictionaries

Let us start by recalling that an **extended partition** with single inner knots associated with $S_m(\Delta)$ is a set $\tilde{\Delta} = \{y_i\}_{i=1}^{2m+N}$ such that

$$y_{m+i} = x_i,\; i = 1, \dots, N,\; x_1 < \cdots < x_N$$

and the first and last m points $y_1 \leq \cdots \leq y_m \leq c, \quad d \leq y_{m+N+1} \leq \cdots \leq y_{2m+N}$ can be arbitrarily chosen. With each fixed extended partition $\tilde{\Delta}$ there is associated a unique B-spline basis for $S_m(\Delta)$, that we denote as $\{B_{m,j}\}_{j=1}^{m+N}$. The B-spline $B_{m,j}$ can be defined by the recursive formulae [23]:

$$\begin{aligned} B_{1,j}(x) \;&=\; \begin{cases} 1, & t_j \leq x < t_{j+1}, \\ 0, & \text{otherwise}, \end{cases} \\ B_{m,j}(x) \;&=\; \frac{x - y_j}{y_{j+m-1} - y_j} B_{m-1,j}(x) + \frac{y_{j+m} - x}{y_{j+m} - y_{j+1}} B_{m-1,j+1}(x). \end{aligned}$$

The following theorem paves the way for the construction of dictionaries for $S_m(\Delta)$. We use the symbol # to indicate the cardinality of a set.

Theorem 2. *Let Δ_j, $j = 1, \dots, n$ be partitions of $[c, d]$ and $\Delta = \cup_{j=1}^n \Delta_j$. We denote the B-spline basis for $S_m(\Delta_j)$ as $\{B_{m,k}^{(j)} : k = 1, \dots, m + \#\Delta_j\}$. Accordingly, a dictionary, $\mathcal{D}_m(\Delta : \cup_{j=1}^n \Delta_j)$, for $S_m(\Delta)$ can be constructed as*

$$\mathcal{D}_m(\Delta : \cup_{j=1}^n \Delta_j) \;:=\; \cup_{j=1}^n \{B_{m,k}^{(j)} : k = 1, \dots, m + \#\Delta_j\},$$

so as to satisfy

$$\mathrm{span}\{\mathcal{D}_m(\Delta : \cup_{j=1}^n \Delta_j)\} \;=\; S_m(\Delta).$$

When $n = 1$, $\mathcal{D}_m(\Delta : \Delta_1)$ *is reduced to the B-spline basis of* $S_m(\Delta)$.

Remark 4. *Note that the number of functions in the above defined dictionary is equal to* $n \cdot m + \sum_{j=1}^n \#\Delta_j$, *which is larger than* $\dim S_m(\Delta) = m + \#\Delta$. *Hence, excluding the trivial case* $n = 1$, *the dictionary constitutes a redundant dictionary for* $S_m(\Delta)$.

According to Theorem 2, to build a dictionary for $S_m(\Delta)$ we need to choose n-subpartitions $\Delta_j \in \Delta$ such that $\cup_{j=1}^n \Delta_j = \Delta$. This gives a great deal of freedom for the actual construction of a non-uniform B-spline dictionary. Fig.7 shows some examples which are produced by generating a random partition of $[0, 4]$ with 6 interior knots. From an arbitrary partition

$$\Delta \;:=\; \{0 = x_0 < x_1 < \cdots < x_6 < x_7 = 4\},$$

we generate two subpartitions as

$$\Delta_1 \;:=\; \{0 = x_0 < x_1 < x_3 < x_5 < x_7 = 4\}, \quad \Delta_2 \;:=\; \{0 = x_0 < x_2 < x_4 < x_6 < x_7 = 4\}$$

and join together the B-spline basis for Δ_1 (light lines in the right graphs of Fig.7) and Δ_2 (dark lines in the same graphs)

Application to sparse signal representation

Given a signal, f say, we address now the issue of determining a partition Δ, and sub-partitions $\Delta_j, j = 1, \ldots, n$, such that: a) $\cup_{j=1}^n \Delta_j = \Delta$ and b) the partitions are suitable for generating a sparse representation of the signal in hand. As a first step we propose to tailor the partition to the signal f by setting Δ taking into account the critical points of the curvature function of the signal, i.e.,

$$T \;:=\; \{t : \left(\frac{f''}{(1 - f'^2)^{3/2}}\right)'(t) = 0\}.$$

Usually the entries in T are chosen as the initial knots of Δ. In order to obtain more knots we apply subdivision between consecutive knots in T thereby obtaining a partition Δ with the decided number of knots. An algorithm for implementing such procedure can be found in [21]. According to Theorem 2, in order to build a dictionary for $S_m(\Delta)$ we need to choose n-subpartitions $\Delta_j \in \Delta$ such that $\cup_{j=1}^n \Delta_j = \Delta$. As an example we suggest a simple method for producing n-subpartitions $\Delta_j \in \Delta$, which is used in the numerical simulations of the next section. Considering the partition $\Delta = \{x_0, x_1, \ldots, x_{N+1}\}$ such that $c = x_0 < x_1 < \cdots < x_{N+1} = d$, for each integer j in $[1, n]$ we set

$$\Delta_j \;:=\; \{c, d\} \cup \{x_k : k \in [1, N] \text{ and } k \mod n = j - 1\},$$

e.g. if $N = 10$ and $n = 3$, we have $\Delta_1 = \{c, x_3, x_6, x_9, d\}$, $\Delta_2 = \{c, x_1, x_4, x_7, x_{10}, d\}$ $\Delta_3 = \{c, x_2, x_5, x_8, d\}$. The codes for creating a partition adapted to a given signal are ProducePartition.m and FinalProducePartition.m and the one code for creating the dictionaries for the space is CutDic.m.

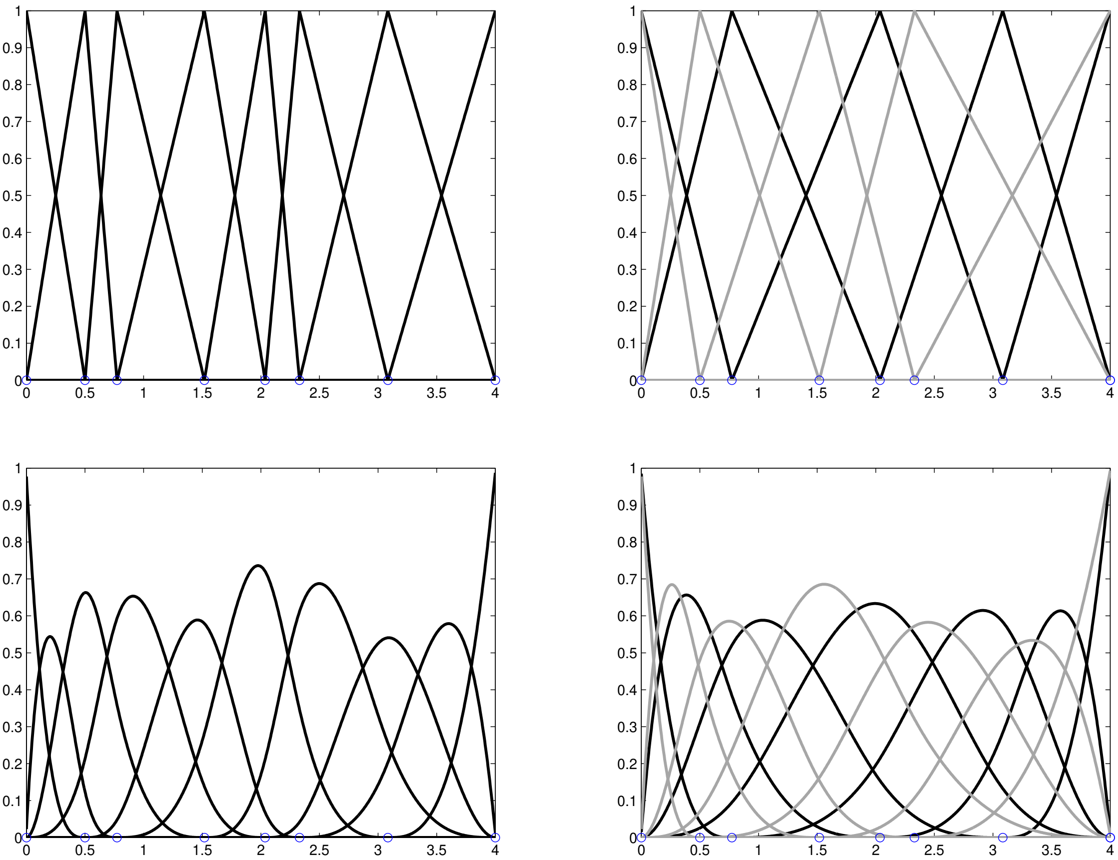

Figure 7: Examples of bases (graphs on the left) and the corresponding dictionaries (right graphs) for a random partition The top graphs correspond to linear B-splines ($m = 2$). The bottom graphs involve cubic B-splines ($m = 4$).

Numerical examples

We produce here an example illustrating the potentiality of the proposed nonuniform dictionaries for achieving sparse representations by nonlinear techniques. The signals we consider are the chirp signal f_1 and the seismic signal f_2 plotted in the Fig. 8.

We deal with the chirp signal f_1 on the interval $[0, 8]$, by discretizing it into $L = 2049$ samples and applying Algorithm 1 to produce the partition. The resulting number of knots is 1162, which is enough to approximate the signal, by a cubic B-spline basis for the space, within the precision $\text{tol}_1 = 0.01\|\text{f}_1\|$. A dictionary $\mathcal{D}_4(\Delta : \cup_{j=1}^{10} \Delta_j)$ for the identical space is constructed by considering 10 subpartitions, which yield $N_1 = 1200$ functions.

The signal f_2 is a piece of $L = 513$ data. A partition of cardinality 511 is obtained as $\Delta = T(f_1, 8)$ and the dictionary of cubic splines we have used arises by considering 3 subpartitions, which yields a dictionary $\mathcal{D}_4(\Delta : \cup_{j=1}^{3} \Delta_j)$ of cardinality $N_2 = 521$.

Denoting by α_n^i, $n = 1, \ldots, N_i$ the atoms of the ith-dictionary, we look now for the subsets of indices Γ_i, $i = 1, 2$ of cardinality K_i, $i = 1, 2$ providing us with a sparse representation of the

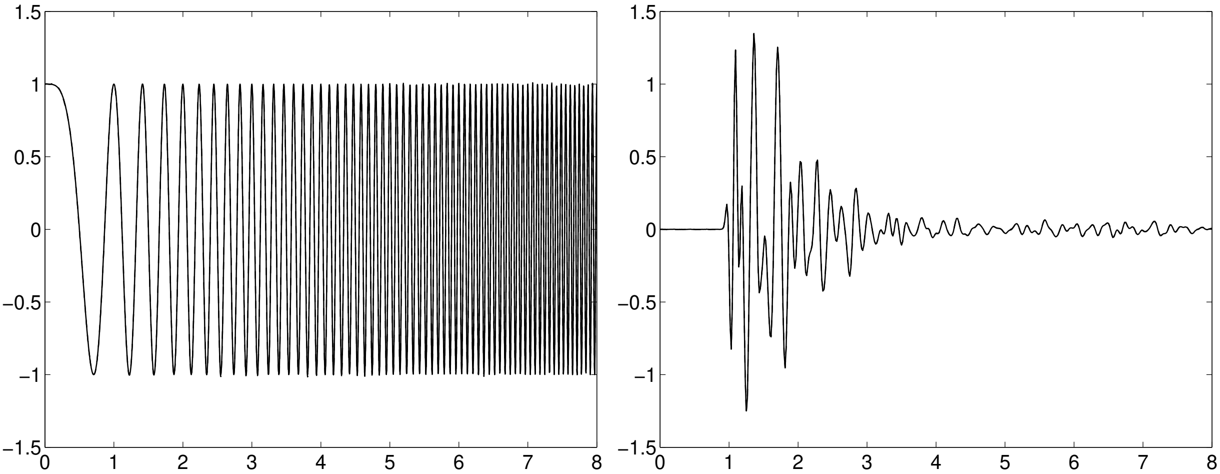

Figure 8: Chirp signal f_1 (left). Seismic signal f_2 (right).

signals. In other words, we are interested in the approximations

$$f_i^{K_i} \;=\; \sum_{n\in\Gamma_i} c_n^i \alpha_n^i, \quad i = 1, 2.$$

such that $\|f_i^{K_i} - f_i\| \le \mathrm{tol}_i,\ i = 1, 2$ and the values K_i, $i = 1, 2$ are satisfactory small for the approximation to be considered sparse. Since the problem of finding the sparsest solution is intractable, for all the signals we look for a satisfactory sparse representation using the same greedy strategy, which evolves by selecting atoms through stepwise minimization of the residual error as follows.

i) The atoms are selected one by one according to the Optimized Orthogonal Matching Pursuit (OOMP) method [12] until the above defined tolerance for the norm of the residual error is reached.

ii) The previous approximation is improved, without greatly increasing the computational cost, by a 'swapping refinement' which at each step interchanges one atom of the atomic decomposition with a dictionary atom, provided that the operation decreases the norm of the residual error [13].

iii) A Backward-Optimized Orthogonal Matching Pursuit (BOOMP) method [14] is applied to disregard some coefficients of the atomic decomposition, in order to produce an approximation up to the error of stage i). The last two steps are repeated until no further swapping is possible. The routine implementing the steps is OOMPFinalRefi.m.

The described technique is applied to all the non-orthogonal dictionaries we have considered for comparison with the proposed approach. The results are shown in Table 1. In the first column we place the dictionaries to be compared. These are: 1) the spline basis for the space adapted to the corresponding signal. 2) The dictionary for the identical spaces consisting of functions of larger support. 3) The orthogonal cosine bases used by the discrete cosine transform (dct). 4) The semi-orthogonal cardinal Chui-Wang spline wavelet basis [26] and 5) the Chui-Wang cardinal spline dictionary for the same space [27]. Notice that whilst the non-uniform spline

Table 1: Comparison of sparsity performance achieved by selecting atoms from the non-uniform bases and dictionaries for adapted spline space (1st and 2nd rows), dft (3rd row), cardinal wavelet bases and dictionaries (4th and 5th rows).

Dictionaries	K^1 (signal f_1)	K^2 (signal f_2)
Non-uniform spline basis	1097	322
Non-uniform spline dictionary	173	129
Discreet cosine transform	263	208
Cardinal Chui-Wang wavelet basis	246	201
Cardinal Chui-Wang wavelet dictionary	174	112

space is adapted to the corresponding signal, only the dictionary for the space achieves the sparse representation. Moreover the performance is superior to that of the Chui-Wang spline wavelet basis [26] and similar to the cardinal Chui-Wang dictionary, which is known to render a very good representation for these signals [27]. However, whilst the Chui-Wang cardinal spline wavelet dictionaries introduced in [27] are significantly redundant with respect to the corresponding basis (about twice as larger) the non-uniform B-spline dictionaries introduced here contain a few more functions than the basis. Nevertheless, as the examples indicate, the improvement in the sparseness of the approximation a dictionary may yield with respect to the B-spline basis is enormous.

Application to filtering low frequency noise from a seismic signal

A common interference with broadband seismic signals is produced by long waves, generated by known or unknown sources, called *infragravity waves* [28–30]. Such an interference is refereed to as low frequency noise, because it falls in a frequency range of up to 0.05 Hz. Thus, in [31] we consider that the model of the subspace of this type of structured noise, on a signal given by $L = 403$ samples, is

$$\mathcal{W}^{\perp} = \text{span}\{e^{\imath \frac{2\pi n(i-1)}{L}}, \, i = 1, \ldots, L\}_{n=-21}^{21}. \tag{36}$$

The particular realization of the noise we have simulated is plotted in the left graph of Fig 9. However, the success of correct filtering does not depend on the actual form of the noise (as long as it belongs to the subspace given in (36)) because the approach we consider guarantees the suppression of the whole subspace $\mathcal{W}^{\perp}$. The seismic signal to be discriminated from the noisy one is a piece of the signal in the right graph of Fig 8. The middle graph in Fig 9 depicts the signal f which is formed by adding both components. Constructing $\hat{P}_{\mathcal{W}^{\perp}}$ we obtain $f_{\mathcal{W}} = f - \hat{P}_{\mathcal{W}^{\perp}} f$ and use the routine ProducePartition.m to find a partition adapted to this projection. In this case, to succeed in modelling a signal subspace complementary to $\mathcal{W}^{\perp}$ we needed a basis for the nonuniform spline space and a regularized version of the FOCUSS algorithm [32]. The result is plotted in the right graph of Fig 9. The implementation details can be found in [31].

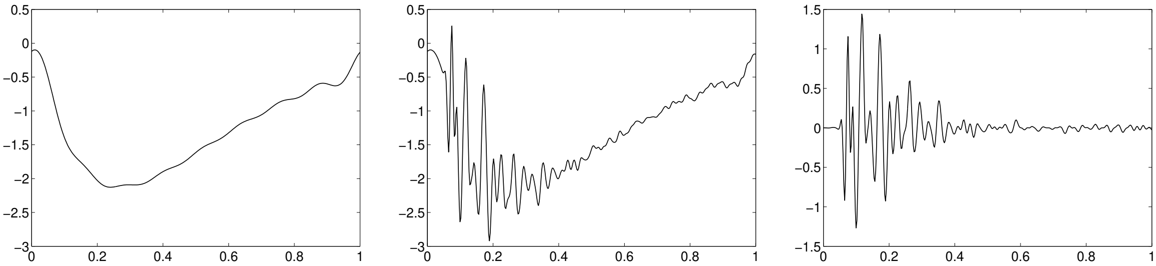

Figure 9: Simulated low frequency noise (left) Signal plus noise (middle) Approximation recovered from the middle graph as explained in [31]

Sparsity and 'something else'

We present here a 'bonus' of sparse representations by alerting that they can be used for embedding information. Certainly, since a sparse representation entails a projection onto a subspace of lower dimensionality, it generates a null space. This feature suggests the possibility of using the created space for storing data. In particular, in a work with James Bowley [35], we discuss an application involving the null space yielded by the sparse representation of an image, for storing part of the image itself. We term this application 'Image Folding'.

Consider that by an appropriate technique one finds a sparse representation of an image. Let $\{v_{\ell_i}\}_{i=1}^K$ be the K-dictionary's atoms rendering such a representation and $\mathcal{V}_K$ the space they span. The sparsity property of a representation implies that $\mathcal{V}_K$ is a subspace considerably smaller than the image space $\mathbb{R}^N \otimes \mathbb{R}^N$. We can then construct a complementary subspace $\mathcal{W}^\perp$, such that $\mathbb{R}^N \otimes \mathbb{R}^N = \mathcal{V}_K \oplus \mathcal{W}^\perp$, and compute the dual vectors $\{w_i^K\}_{i=1}^K$ yielding the oblique projection onto $\mathcal{V}_K$ along $\mathcal{W}^\perp$. Thus, the coefficients of the sparse representation can be calculated as:

$$c_i = \langle w_i^K, I\rangle,\ i = 1, \ldots, K. \tag{37}$$

Now, if we take a vector in $F \in \mathcal{W}^\perp$ and add it to the image forming the vector $G = I + F$ to replace I in (37), since F is orthogonal to the duals $\{w_i^K\}_{i=1}^K$, we still have

$$\langle w_i^K, G\rangle = \langle w_i^K, I + F\rangle = \langle w_i^K, I\rangle = c_i,\ i = 1, \ldots, K. \tag{38}$$

This suggests the possibility of using the sparse representation of an image to embed the image with additional information stored in the vector $F \in \mathcal{W}^\perp$. In order to do this, we apply the earlier proposed scheme to embed redundant representations [34], which in this case operates as described below.

Embedding Scheme
We can embed H numbers h_i, $i = 1, \ldots, H$ into a vectors $F \in \mathcal{W}^\perp$ as follows.

- Take an orthonormal basis z_i, $i = 1, \ldots, H$ for $\mathcal{W}^\perp$ and form vector F as the linear combination
$$F = \sum_{i=1}^{H} h_i z_i.$$
- Add F to I^K to obtain $G = I^K + F$

Information Retrieval
Given G retrieve the numbers h_i, $i = 1, \ldots, H$ as follows.

Figure 10: The small picture at the top is the folded Image. The left picture below is the unfolded image without knowledge of the private key to initialize the permutation. The next is the unfolded picture when the correct key is used.

- Use G to compute the coefficients of the sparse representation of I as in (38). Use this coefficients to reconstruct the image $\tilde{I}^K = \sum_{i=1}^{K} c_i^K v_i^K$
- Obtain F from the given G and the reconstructed $\tilde{I}^K$ as $F = G - \tilde{I}^K$. Use F and the orthonormal basis z_i, $i = 1, \ldots, H$ to retrieve the embedded numbers h_i, $i = 1, \ldots, H$ as

$$h_i = \langle z_i, F \rangle,\ i = 1, \ldots, H$$

One can encrypt the embedding procedure simply by randomly controlling the order of the orthogonal basis z_i, $i = 1, \ldots, H$ or by applying some random rotation to the basis, requiting a *key* for generating it. An example is given in the next section.

Application to Image Folding

In order to apply the embedding scheme to fold an image we simply process the image by dividing it into, say q blocks, I_q of $N_q \otimes N_q$ pixels. We find the representation of each block with a combination of discrete cosine and spline based subdictionaries constructed as explained in [35]. With this dictionaries the image of Fig 10 has a highly sparse representation at PSNR=40 dB (the CR is approximately 11) so that we are able to store the whole image embedding the few pixels shown in the top picture of Fig 10. The next picture is the unfolded image without using the security *key* and the right picture the one obtained with the correct *key*. For further implementation details see [33]. An example of how to call the available codes for sparse image approximation with cosine and spline based dictionaries can be found here.

References

[1] O. Christensen, Fames and Bases an introductory course, Birkhauser (2008).

[2] M. Unser, A. Aldroubi, A general sampling theory for nonideal acquisition devices, IEEE Trans. Signal Processing 42 (1994) 2915–2925.

[3] Y. Eldar, Sampling with arbitrary sampling and reconstruction spaces and oblique dual frame vectors, Journal of Fourier Analysis and Applications 9 (2003) 77–96.

[4] R. Behrens, L. Scharf, Signal processing applications of oblique projection operators, IEEE Transactions on Signal Processing 42 (1994) 1413–1424.

[5] G. Corach, A. Maestripieri, D Stojanoff, A Classification of Projectors, Banach Center Publ. 67 (2005), 145–160.

[6] G. Corach, A. Maestripieri, D Stojanoff, Projections in operators ranges, Proc. Amer. Math. Soc. 134 (2005), no 3, 765–788.

[7] L. Rebollo-Neira, A. Plastino, Nonlinear non-extensive approach for identification of structured information, Phyica A, in press (2009).

[8] A. Hirabayashi, M. Unser, Consistent sampling and signal recovery, IEEE Transactions on Signal Processing 55 (2007) 4104–4115.

[9] A. Cerny, Characterization of the oblique projector $U(VU)^{\dagger}V$ with application to constrained least squares, Linear Algebra and Its Applications, (2009)

[10] L. Rebollo-Neira, Constructive updating/downdating of oblique projectors: a generalization of the Gram–Schmidt process, Journal of Physics A: Mathematical and Theoritcal 40 (2007) 6381–6394.

[11] L. Rebollo-Neira, Oblique matching pursuit, IEEE Signal Processing Letters 14 (10) (2007) 703–706.

[12] L. Rebollo-Neira, D. Lowe, Optimized orthogonal matching pursuit approach, IEEE Signal Processing Letters 9 (2002) 137–140.

[13] M. Andrle, L. Rebollo-Neira, A swapping-based refinement of orthogonal matching pursuit strategies, Signal Processing 86 (2006) 480–495.

[14] M. Andrle, L. Rebollo-Neira, E. Sagianos, Backward-optimized orthogonal matching pursuit approach, IEEE Signal Proc. Let. 11 (2004) 705–708.

[15] L. Rebollo-Neira, Measurements design and phenomena discrimination, J. Phys. A: Math. Theor. 42 (2009) 165210.

[16] B.D. Rao and K. Kreutz-Delgado, An Affine Scaling Methodology for Best Basis Selection, IEEE Trans. Sig. Proc. 47 (1999) 187–200.

[17] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM Journal on Scientific Computing 20 (1998) 33–61.

[18] S.S. Chen and D.L. Donoho and M.A. Saunders, Atomic Decomposition by Basis Pursuit, SIAM Rev. 43 (2001) 129–156.

[19] L. Rebollo-Neira, A. Constantinides, A. Plastino, F. Zyserman, A. Alvarez, R. Bonetto, H. Viturro, Statistical inference, state distribution, and noisy data, Physica A 198 (1993) 514–537

[20] L. Rebollo-Neira, A. G. Constantinides, A. Plastino, A. Alvarez, Bonetto, M. Iñiguez Rodriguez, Statistical analysis of a mixed-layer x-ray diffraction peak, Journal of Physics D, 30, 17 (1997) 2462–2469.

[21] L. Rebollo-Neira, Z. Xu, Adaptive non-uniform B-spline dictionaries on a compact interval, arXiv:0908.0691v1 [math.FA]

[22] M. Unser, Splines. A perfect fit for signal and image processing, *IEEE Signal Processing Magazine*, 22–38 (1999).

[23] L. L. Schumaker, Spline Functions: Basic Theory, Wiley, New-York, 1981.

[24] C. K. Chui, Multivariate splines, SIAM, Philadelphia, 1988.

[25] Carl De Boor, A Practical Guide to Splines, Springer, New York, 2001.

[26] C. Chui, J. Wang, On compactly supported spline wavelets and a duality principle, Trans. Amer. Math. Soc. 330 (1992) 903–915.

[27] M. Andrle, L. Rebollo-Neira, From cardinal spline wavelet bases to highly coherent dictionaries, Journal of Physics A 41 (2008) 172001.

[28] R. D. Kosian, N. V. B. Pykhov, B. L. Edge, Coastal processes in tideless seas, ASCE Publications, 2000.

[29] D. Dolenc, B. Romanowicz, B. Uhrhammer, P. McGill, D. Neuhauser, D. Stakes, Identifying and removing noise from the Monterey ocean bottom broadband seis mic station (MOBB) data, Geochem. Geophys. Geosyst., 8, (2007) Q02005, doi:10.1029/2006GC001403.

[30] W. C. Crawford, S. C. Webb, Identifying and removing tilt noise from low-frequency (< 0.1 Hz) seafloor vertical seismic data, Bull. Seism. Soc. Am., 90, 952-963, 2000.

[31] Z. Xu, L. Rebollo-Neira, A. Plastino, Subspace modeling for structured noise suppression, arXiv:0908.1020v1 [math-ph] (2009).

[32] B.D. Rao, K. Engan, S. F. Cotter, J. Palmer, K. Kreutz-Delgado, Subset selection in noise based on diversity measure minimization, IEEE Transactions on Signal Processing, 51, 3 (2003) 760– 770, 10.1109/TSP.2002.808076.

[33] J. Bowley, L. Rebollo-Neira, Sparsity and 'something else', (2009).

[34] J.Miotke, L. Rebollo-Neira Oversampling of Fourier Coefficients for Hiding Messages, Applied and Computational Harmonic Analysis , 16,3, (2004) 203–207.

[35] J. Bowley, L. Rebollo-Neira, Sparse image representation by discrete cosine/spline based dictionaries, (2009).