Variation Filtering of Time Series: New Results

Paolo d'Alessandro
Former Professor, Math Dept, Third University of Rome

November 6, 2025

Abstract

In this paper we illustrate the theory and applications of the technique of filtering by reducing the variation norm or variation seminorm of a signal, adding many improvements and new results.

1 Introduction

We initiate with a brief overview of the topic for selfcontainedness but also to introduce some improvements, and then move on to present some new results.

The terms "smoothing" and "smoother" are extensively used in the literature on stochastic processes. Our use of the term smoothing here is arguably appropriate in view of the approach that we propose, but there are profound differences with stochastic filtering. In probabilistic terms smoothing is the effect of projections on an increasing family of closed subspace of a Hilbert space.

Here first of all smoothing starts as a deterministic concept. It arises as well from a projection but on a compact convex body, and hence we pass from a linear to a nonlinear projection. In addition, while in the stochastic context the projection is uniquely defined, in our case we have an infinity of possible projections that correspond to the degree of the smoothing effect. In the application of noise filtering, for example, we propose to optimize such a degree maximizing a suitable statistic. In general, some rule of optimization of the strength of the smoothing effect is required in any other application.

Filtering by variation is achieved projecting, with respect to the Hilbert space $||.||_2$ norm, a signal $s \in R^N$ with a variation $v(s)$ (where v is a norm to be defined momentarily) on a closed sphere around the origin of v with radius $r < v(s)$.

Of course, the celebrated Projection Theorem for Hilbert spaces is in force for studying this problem.

A first crucial point is that the closed spheres of the variation norm are polytopes.

There are in the literature specific methods for projecting a point on a polytope, such as the Wolfe algorithm (see [9]).

But our analysis of the structure of the polytopic spheres of v uncover the possibility of using a more efficient technique of projection valid for polytopic spheres belonging to both classes M and $\mathbb{P}$ (to be defined momentarily, and the spheres of v will be proved to be in such family). We call this method fast projection. In practice it will be clear such method provides the projection instantaneously, and this makes feasible to efficiently implement many interesting numerical applications.

This line of research started in 1981 with [3], subsequently published in a Operation Research Conference in 1982 (see [4]). These papers illustrate a smoothing technique based on the variation norm.

The mathematical theory of M norms and the definition of M norms for R^N and then of the subclass $\mathbb{P}$ were given in [2]. In the present paper we define the class $\mathbb{P}$ more in general and then apply our analysis to the intersection of these two classes.

Moreover, in [2] a main result was established, which shows that, for the (polytopic) spheres of norms in class $M \cap \mathbb{P}$, a fast projection technique is applicable. The results of these papers were later on published in a O.R. Journal (see [5]), where, in addition, a proof that v is in the above category was given based on the present author implementation (in Modula-2) of the fast projection method for v, which was developed in 1980.

A couple numerical applications had been published in a research report only [6] (filtering and seasonal adjustment). We present here an overview of these applications, and add a third one (related to portfolio management).

Here are the original contributions of the present paper.

First we give a new and improved proof of the fact that v is a M norm of class $\mathbb{P}$. The new proof is mathematical in nature, rather then based on the code developed for implementing the fast projection. Moreover, it also directly translates into a new algorithm to implement the fast projection, which looks simpler and more efficient that the original one.

The second original contribution is that, thanks to some new preliminary results, we will develop a further mathematical technique of projection on the spheres of the norm v, which is even simpler and faster than the fast projection, and which we call the superfast projection.

Finally, the third original contribution consists in showing that the mathematical techniques developed to project a signal on a sphere defined by the variation norm v, can be used to obtain the projection of a signal on the spheres defined by the pure variation seminorm w. This hugely simplify projections on the w spheres and, in addition, opens the way to an easy comparison between the two approaches.

Ultimately, while the projections on spheres of the variation norm v is simpler than the projection on w spheres, it is a good approximation for this latter, especially for low values of the ratio $|s(1)|/s(w)$. This will give some motivation for the use of v, in view of its greater simplicity and efficiency. However, whenever the use of the pure variation w is advisable, our results show that the fast techniques we have developed for v also apply, by means of an algorithm we introduce, to the pure variation seminorm w.

In the final two sections first we illustrate some practical applications like filtering, seasonal adjusting and portfolio management, which can all be formulated in terms of either v or w, and then we conclude looking at research perspectives.

2 The Variation Norm and Its Polytopic Spheres

In the rest of the paper we will use without detailed recalls many concepts and results of convex analysis, including elementary computations, extreme points and faces (in particular exposed faces) the Krein Milman Theorem and its specialization for polytopes (which are always compact) and so on, as well as basic concepts and result of vector topology, like the fundamental theorem for vector topology bases the collapse of all vector topologies to a unique one in R^N, the fact that R^N is locally compact and so on.

The present section and the next two we review and improve the basic theory as given in [2].

Consider a time series or signal $ts(.) : \mathfrak{N} \to R$, where $\mathfrak{N}$ denotes the natural numbers (representing instant of times from 1 to ∞) and R the real field. We are interested to its restriction to the interval $[1, N]$ (where $N > 1$ is an integer) and denote it with s. Of course, $s \in R^N$.

In this space R^N we work with many norms and seminorms, in the first place the (pure) variation seminorm w and the variation norm v norm defined by:

$$w(s) = \sum_{i=1}^{N-1} |s(i+1) - s(i)|$$

$$v(s) = |s(1)| + \sum_{i=1}^{N-1} |s(i+1) - s(i)| = |s(1)| + w(s)$$

We do not explicitly indicate the dependence of these seminorms and norms on N, unless needed for clarity. We leave as an exercise to verify that $w(.)$ is a bonafide seminorm on R^N and $v(.)$ a bona fide norm for R^N.

If we consider the $N \times N$ matrix:

$$T = \begin{pmatrix} 1 & 0 & .. & .. & 0 & 0 \\ -1 & 1 & .. & .. & 0 & 0 \\ .. & .. & .. & .. & .. & .. \\ 0 & 0 & .. & -1 & 1 & 0 \\ 0 & 0 & .. & .. & -1 & 1 \end{pmatrix}$$

then the following fundamental relation holds (composition of functions is denoted by juxtaposition):

$$v = p_1 T$$

where p_1 denotes for simplicity the $||.||_1$ norm.

We will exploit later on the following well known result where we use the notations e_i for the ith versor on R^N and $\mathcal{C}(.)$ for the convex extension of a set. Moreover, given a set C, $ex(C)$ will denote the set of extreme points of C. Finally given two vectors x and y, $[x:y]$ denotes the segment joining x and y.

Lemma 1 *The following formula holds:*

$$S^1_{p_1} = \mathcal{C}(\{e_1, .., e_N, -e_1, .., -e_N\})$$

Moreover $ex(S^1_{p_1}) = \{e_1, .., e_N, -e_1, .., -e_N\}$.

Proof. From $\{e_1, .., e_N, -e_1, .., -e_N\} \subset S^1_{p_1}$ and the fact that $S^1_{p_1}$ is convex it follows

$$\mathcal{C}(\{e_1, .., e_N, -e_1, .., -e_N\}) \subset S^1_{p_1}$$

Denote the simplex by $Sm = \mathcal{C}(\{e_1, .., e_N\})$. Then, by elementary computations for convex sets

$$\mathcal{C}(\{e_1, .., e_N, -e_1, .., -e_N\}) = \mathcal{C}(\{e_1, .., e_N\} \cup \{-e_1, .., -e_N\}) =$$

$$\cup\{[x:y] : x \in Sm \ \text{ and } y \in -Sm\}$$

Now $z \in S^1_{p_1} \Leftrightarrow \sum |z_i| \leq 1$. Denote by k the indexes of the positive components of z and by j the indexes of the negative ones. Then

$$z = \sum z_k e_k + \sum z_j(-e_j) =$$

$$= (\sum z_k) \sum \frac{z_k}{\sum z_k} e_k + (\sum |z_j|) \sum \frac{|z_j|}{\sum |z_j|}(-e_j) \in$$

$$\cup\{[x:y] : x \in Sm \ \text{ and } y \in -Sm\}$$

and therefore

$$S^1_{p_1} \subset \mathcal{C}(\{e_1, .., e_N, -e_1, .., -e_N\})$$

The variant of this argument if there are only positive or only negative components is obvious and so the first statement is established. Next notice that each singleton $\{e_i\}$ and $\{-e_i\}$ is, by an immediate computation an exposed face of $S^1_{p_1}$ and hence all e_i and all $-e_i$ are extreme points of $S^1_{p_1}$. Since, by the first statement any point of $S^1_{p_1}$ is a convex combination of these vectors, it follows that there cannot be other extreme points. This completes the proof. ∎

If we denote by S^1_v the closed unit sphere around the origin of norm v, from $v = p_1 T$, we have:

$$S^1_v = v^{-1}([-1,1]) = T^{-1} p_1^{-1}([-1,1]) = T^{-1} S^1_{p_1} =$$

$$= T^{-1}\mathcal{C}(\{e_1, .., e_N, -e_1, .., -e_N\}) = \mathcal{C}(\{T^{-1}e_1, .., T^{-1}e_N, -T^{-1}e_1, .., -T^{-1}e_N\})$$

where we have used an elementary computation in the last passage.

Consequently S_v^1 is a polytope, that is, the convex extension of a finite set. Next:

$$T^{-1} = \begin{pmatrix} 1 & 0 & .. & .. & 0 & 0 \\ 1 & 1 & 0 & .. & 0 & 0 \\ .. & .. & .. & .. & .. & .. \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \quad (1)$$

that is, the lower triangular matrix. Hence the polytope S_v^1 is explicitly:

$$S_v^1 = \mathcal{C}(\{ \begin{pmatrix} 1 \\ 1 \\ .. \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ .. \\ 1 \\ 1 \end{pmatrix}, .., \begin{pmatrix} 0 \\ 0 \\ .. \\ 0 \\ 1 \end{pmatrix}, -\begin{pmatrix} 1 \\ 1 \\ .. \\ 1 \\ 1 \end{pmatrix}, -\begin{pmatrix} 0 \\ 1 \\ .. \\ 1 \\ 1 \end{pmatrix}, .., -\begin{pmatrix} 0 \\ 0 \\ .. \\ 0 \\ 1 \end{pmatrix} \}) \ (2)$$

We denote by J the generating set, argument of convex extension in this formula, so that $S_v^1 = \mathcal{C}(J)$. We stress that each element of J represents a jump in the signal, positive or negative according to its sign, so that a signal $s \in S_v^1$ is decomposed in the sequence of its jumps. Indeed the signal s can be represented in the form:

$$s = s(1)T_1^{-1} + \sum_{i=2}^{N}(s(i) - s(i-1))T_i^{-1} =$$

$$= |s(1)|sign(s(1))T_1^{-1} + \sum_{i=2}^{N}|s(i) - s(i-1)|sign(s(i) - s(i-1))T_i^{-1} \quad (3)$$

where T_i^{-1} is the ith columns of T^{-1}. In a more compact form, if we define I_π to be the matrix obtained from the identity matrix attributing to the sequence of 1 the signs of the sequence of signs of jumps $sign(s(1))$, $sign(s(2) - s(1))$, .., and setting to zero the the components corresponding to a null jump we can write:

$$s = T^{-1} I_\pi \begin{pmatrix} |s(1)| \\ |s(2) - s(1)| \\ .. \\ .. \\ \\ |s(N) - s(N-1)| \end{pmatrix} = T^{-1} \begin{pmatrix} s(1) \\ s(2) - s(1) \\ .. \\ .. \\ \\ s(N) - s(N-1) \end{pmatrix}$$

This formula expresses the signal in terms of its jumps in value and hence connects its expression to the variation norm.

3 Properties of Polytopic and M norm Spheres

When we consider a closed sphere S_p^r around the origin of a norm in R^N, many of its properties are shared by S_p^1 since S_p^r is obtained by inflating S_p^1 according to the relation $S_p^r = rS_p^1$. Thus in the sequel we will mostly refer to closed unit spheres. In particular, all the properties that we will introduce for the variation closed unit sphere S_v^1 around the origin hold good for any close sphere S_v^r of radius $r > 0$.

From vector topology we know that the closed variation sphere S_v^1 is a closed convex body, it is circled and hence symmetric and is radial at the origin. From convex analysis we know a number of further interesting properties. As a first example, any polytope is compact and hence S_v^1 is compact (this confirms the well know result that the unique vector topology of R^N is locally compact).

Another result of convex analysis that we can invoke states that, if C is a closed convex body, then the boundary of C is the union of the proper closed exposed faces of C. Thus the boundary of S_v^1 is the union of its proper (exposed) closed faces. And since the extreme point of a closed exposed face are also extreme for S_v^1 (by another well known result of Convex Analysis), all closed exposed faces are polytopes generated by a subset of the generating set of $ex(S_v^1)$ (we will soon characterize such subsets), that is, all closed exposed faces are subpolytopes of S_v^1.

But who is $ex(S_v^1)$?

The following result, which exploits the linear isomorphism connecting any two bases and Lemma 1, tells us, in particular, that $ex(S_v^1) = J$. A direct proof is also possible and is given in the cited reference.

Theorem 2 *Suppose that a set of $\Omega \subset R^N$ is given by $\Omega = \mathcal{C}(B \cup -B)$, where $B = \{b_i\}$ is a linear basis. Then Ω is the closed unit sphere around the origin of a norm and $ex(\Omega) = B \cup -B$.*

Proof. Consider the linear isomorphism defined by the bases $\{e_i\}$ and $\{b_i\}$ represented by the matrix Ψ whose columns are given by the vectors b_i. Then, by Lemma 1,

$$\mathcal{C}(B \cup -B) = \mathcal{C}(\{\Psi e_i\} \cup \{\Psi(-e_i)\}) = \Psi\mathcal{C}(\{e_i\} \cup \{(-e_i)\}) = \Psi S_{p_1}^1$$

The linear isomorphism preserves all the properties for $\mathcal{C}(B \cup -B)$ to qualify as the unit closed sphere of a norm and the extreme points of $\mathcal{C}(B \cup -B)$ are readily seen to be the images under Ψ of the extreme points of $S_{p_1}^1$. The theorem is thereby established. $\blacksquare$

Moving on, we need to introduce the notation $\mathcal{L}(C)$ for the linear extension of the set C.

Theorem 3 *Let $\Omega \subset R^N$ be a polytope so that $\Omega = \mathcal{C}(\Upsilon)$ for some finite set Υ. By the Krein Milman (briefly KM) Theorem for polytopes $\Upsilon \supset ex(\Omega)$. Then Ω is the closed unit sphere around the origin of a norm iff $ex(\Omega)$ contains a basis B for R^N and is symmetric (and hence iff it contains $B \cup -B$). Consequently, it is impossible to specify a polytopic sphere of a norm with less that $2N$ extreme points.*

Proof. Necessity. Because a polytope is always compact the Krein Milman Theorem holds, and since the convex extension of a finite set is compact $\Omega =$

$\mathcal{C}(\Upsilon) = \mathcal{C}(ex(\Omega))$. If $ex(\Omega)$ does not contain a base, then $\mathcal{C}(ex(\Omega)) \subset \mathcal{L}(ex(\Omega))$ which would be a proper subspace of R^N, so that Ω would have no interior, which is a contradiction. The symmetry condition is also necessary in view of the fundamental Theorem on bases for vector topologies (see [1]). In fact, as it is easily verified, $\mathcal{C}(ex(\Omega))$ is symmetric iff $ex(\Omega)$ is symmetric. Sufficiency. By the hypotheses

$$\Omega \supset \mathcal{C}(B \cup -B)$$

Thus Ω is convex symmetric (and hence convex and circled) and radial at zero, so that its Minkowski functional define a norm and Ω is its closed unit sphere. This completes the proof, since the final statement is by now obvious. $\blacksquare$

Remark 4 *Note that under the hypotheses of this theorem $B \cup -B$ is not in general the whole set of extreme points of Ω, as in the case of the preceding theorem. This justifies singling out the case where the sphere of a polytopic norm has the minimum number $2N$ of extreme points because $ex(\Omega) = B \cup -B$ according to the definition that follows.*

Definition 5 *Any norm for R^N, whose closed spheres around the origin are polytopes with $2N$ extreme points (which by the preceding Theorem form a set $B \cup -B$, where B is a linear basis) is called an M norm.*

Examples of M norms are evidently v and p_1. The class of M norms is an equivalence class under linear isomorphism. We are interested here mainly in the study of M norms, and more precisely in a subclass of these norms, which will be soon introduced.

Finally we recall from the same sources the following crucial result (as usual we state it for the unit sphere but the result is true in general). The proof is obtained exploiting the linear isomorphism defined by the bases $\{e_i\}$ and B and the consequent relation between the closed unit sphere of the M norm and the closed unit sphere of $||.||_1$.

Theorem 6 *Let S be the closed unit sphere around the origin of an M norm. Then a point $x \in \mathfrak{B}(S)$ belong to the boundary of S iff $x \in \mathcal{C}(\Xi)$ where $\Xi \subset ex(S) = B \cup -B$ and Ξ does not contain any pair of opposite points.*

Since as recalled earlier $\mathfrak{B}(S)$ is the the union of the proper closed exposed faces of S, we can state the following:

Corollary 7 *Let S be the closed unit sphere around the origin of an M norm. The the family of closed exposed proper faces of S is given by the family of polytopes generated by all subsets of $ex(S)$, which do not contain pairs of opposite points.*

Another important consequence, that follows at once from the theory of M norms developed so far, and the fact that a sphere is radial at the origin is the following:

Theorem 8 *Let S be the closed unit sphere around the origin of an M norm. Then any vector $x \neq 0$ belongs to the relative interior of one of the cones generated by subsets of $ex(S)$, which do not contain pairs of opposite points.*

In the next section we will illustrate a property defining a special class of norms, and apply it to M norms, showing that it makes possible to project (with respect to the $||.||_2$ norm) a point $y \notin S_p^r$ on S_p^r (with the usual notations and where p is any such M norm), by means of an extremely fast computation. Naturally, in this respect, the context is that of the celebrated Projection Theorem for Hilbert spaces.

4 A Special Class of Norms and a Fast Projection Technique

Thanks to the theory developed hitherto, we can now pass on to deal with the central concept of fast projection for the spheres of a special class of M norms.

We start recalling various facts of convex analysis and of vector topology.

This fast projection technique (where projection is intended with respect to the $||.||_2$ Hilbert space norm for R^N) was introduced in [2].

Since R^N has an unique vector topology, any norm which we specify for R^N is necessarily continuous and its closed spheres are necessarily closed sets. Since we can specify norms with polytopic closed spheres, which are compact, any norm for R^N has necessarily compact closed spheres (because by norm equivalence they must be contained in a polytope). Notice that this is another way to confirm that R^N is locally compact.

Since any norm for R^N has compact closed spheres, by the KM Theorem, a closed sphere S_p^r for a norm p in R^N is the closed convex extension of the set of its extreme points $ex(S_p^r)$.

By the Riesz Theorem all continuous linear functionals on R^N have the form $(n, .)$, where the vector n is uniquely defined by the functional.

An exposed face of S_p^r is defined to be the intersection of a supporting hyperplane to S_p^r and S_p^r. Here a supporting hyperplane to S_p^r at $x \in S_p^r$ is an hyperplane $\{y : (n, y) = (n, x)\}$ such that:

$$(n, z) \leq (n, x) \ \forall z \in S_p^r$$

Consequently all exposed faces are closed compact and contained in $\mathcal{B}(S_p^r)$. By the Projection Theorem:

$$P_{S_p^r}(x + r(n)) = x$$

The vector n is a normal to S_p^r at x, and the set of all normals to S_p^r at x is a convex cone, called the normal cone to S_p^r at x and denoted by $\mathfrak{N}_{S_p^r}(x)$. A vector n is normal to the closed exposed face $\mathcal{F}$ if:

$$n \in \mathfrak{N}_{S_p^r}(x) \ \forall x \in \mathcal{F}$$

The set of $\mathfrak{N}_{\mathcal{F}}$ normals to $\mathcal{F}$ is also a convex cone, the normal cone to $\mathcal{F}$. By a a well known Separation Theorem, if $x \in \mathcal{B}(S_p^r)$, then there is a continuous linear functional which separates $\{x\}$ and S_p^r, and this implies that $\mathcal{B}(S_p^r)$ is the union of closed exposed faces of S_p^r.

With these premises we define a special class of norms that play a central role in our investigation. Recall that an extreme point of a closed exposed face is also an extreme point of S_p^r. Thus each closed exposed face is the convex extension of a subset of $ex(S_p^r)$.

Definition 9 *An norm p for R^N is said to be of class $\mathbb{P}$ if the closed unit sphere of p around the origin has the property that each of its closed exposed face $\mathcal{F}$ has a normal which belongs to the cone generated by $ex(\mathcal{F})$. In other words, $\mathfrak{N}_{\mathcal{F}} \cap \mathcal{C}o(ex(\mathcal{F})) \neq \phi$.*

Our next aim is to show how for norms of class $M \cap \mathbb{P}$ the projection operator can be given an extremely simple expression

Before getting to this, we state in general another simple but important result, connecting projection on spheres of arbitrary radius $r > 0$ to projection on the unit spheres. Such result should be borne in mind throughout the sequel. Consider a norm p and a point $x \notin S_p^r$, where S_p^r is the closed sphere of p around the origin with radius r.

Then we can state the following simple but important theorem, where P_C denotes the projection function on the closed convex set C.

Theorem 10 *The following formula holds:*

$$P_{S_p^\rho} x = \rho P_{S_p^1} \frac{x}{\rho}$$

Thus if $r = p(x)$ and we project x on a sphere of radius cr with $0 < c < 1$ we obtain

$$P_{S_p^{cr}} x = cr P_{S_p^1} \frac{x}{cr}$$

Consequently the projection depends continuously on the factor c.

Proof. Clearly $S_p^1 = \frac{1}{\rho} S_p^\rho$. Hence if we project $\frac{x}{\rho}$ on S_p^1, all distances in terms of the $||.||_2$ norm are divided by ρ with respect to the projection of x on S_p^ρ, since norms are absolutely homogeneous. Thus the minimal distance of $\frac{x}{\rho}$ from S_p^1 is $\frac{1}{\rho}$ the minimal distance of x from S_p^ρ. By the same argument, multiplying $P_{S_p^1} \frac{x}{\rho}$ by ρ we infer that x is the point of minimal distance of x from S_p^ρ. ■

We call the application of this result, that is, the procedure of using the rhs to compute the lhs, *scaling and descaling.*

Of course this Theorem is a further reason why we will develop our theory making constant reference to unit spheres.

4.1 Fast Projection Technique

Let p be a M norm of class $\mathbb{P}$, S_p^1 the unit closed sphere around the origin of p and x a vector outside S_p^1, so that $p(x) > 1$.

Let's move on the ray generated by x. Since $x/p(x) \in \mathfrak{B}(S_p^1)$, it is in the relative interior of a proper closed exposed face of S_p^1. This means that x is in the relative interior of the cone $\mathcal{C}o(\{v_1, ..v_k\})$ generated by set of vertexes $\{v_1, ..v_k\}$ of such a face.

Let $\{v_1, ..v_k\}$ the set of such vertexes so that:

$$\frac{x}{p(x)} = \sum \zeta_i v_i, \text{ with } \zeta_i > 0, \forall i, \ \sum \zeta_i = 1$$

and

$$x = p(x) \sum \zeta_i v_i = \sum p(x) \zeta_i v_i = \sum \lambda_i v_i$$

with $\lambda_i = p(x)\zeta_i$. By hypothesis, the face in question has a normal n_1 given by:

$$n_1 = \sum \gamma_i v_i, \text{ with } \gamma_i \geq 0, \forall i \text{ and } n_1 \neq 0$$

The idea of the method is to move toward the sphere along $-n_1$ until either we met the boundary of S_p^1, which means that we have found the projection and the procedure can be arrested, or we hit the relative boundary of the cone, meaning that at least one (or more) of the coefficients $\lambda_i - \beta\gamma_i$ zeroes (while all the others stand positive), for some $\beta > 0$, and hence we arrive to a vector which is in the cone generated by a subface of the face we started with. At this point we can start another similar step. Clearly the procedure is arrested in a finite number of steps, smaller than the dimension of the space. Also it is clear that the computation is practically as fast as the computation of the required normals.

Formally, coming back to the first step, we define two positive constants:

$$\eta_1 = \min\{\lambda_i/\gamma_i : \gamma_i \neq 0\}$$

which gives the value of β at which $x - \beta n_1$ hits the relative boundary of $\mathcal{C}o(\{v_1, ..v_k\})$, and

$$\delta_1 = \frac{\sum \lambda_i - 1}{\sum \gamma_i}$$

which gives the value of β at which $x - \beta n_1$ hits $\mathfrak{B}(S_p^1)$. In fact:

$$x - \delta_1 n_1 = \sum \lambda_i v_i - \frac{\sum \lambda_i - 1}{\sum \gamma_i} \sum \gamma_i v_i = \sum (\lambda_i - \frac{\sum \lambda_i - 1}{\sum \gamma_i} \gamma_i) v_i$$

and

$$\sum (\lambda_i - \frac{\sum \lambda_i - 1}{\sum \gamma_i} \gamma_i) = \sum \lambda_i - \sum \lambda_i + 1 = 1$$

Let

$$\alpha_1 = \min\{\eta_1, \delta_1\}$$

and $x_1 = x - \alpha_1 n_1$. If:

$$\alpha_1 = \delta_1 \Rightarrow p(x_1) = p(x - \alpha_1 n_1) = 1$$

and thus the procedure is arrested because x_1 is the sought projection. If instead $\alpha_1 = \eta_1$, then $x_1$is in the relative boundary of the cone $\mathcal{C}o(\{v_1, ..v_k\})$ and $p(x_1) > 1$.

Note that if x_1 is proportional to a vertex v_j, put $x_2 = v_j$ and, by hypothesis, the procedure can be arrested because x_2 is the sought projection. If x_1 is not proportional to any vertex, perform another identical step, replacing x by x_1.

By construction, the procedure is arrested in $m \leq N$ steps in a point x_m such that $p(x_m) = 1$, and

$$x = x_m + \sum_{i=1}^{m} \alpha_1 n_1$$

where $\sum_{i=1}^{m} \alpha_1 n_1$ is a normal to S_p^1 at x_m, and therefore x_m is the sought projection.

5 Fast Projection on Variation Spheres

As mentioned in [4], in 1980 the author implemented in Pascal the projection procedure described in the preceding Section, for the norm v. Indeed the implementation, devising a diagonalization technique for the computation of normals to the faces of the spheres S_v^r, confirmed that the variation M norm v is of class $\mathbb{P}$. A Modula2 version of the program was subsequently written by the author and the procedure normal was published in 1988 in [5] to give a constructive proof of this circumstance.

However, while the round of ideas behind the diagonalization technique used for the computation of normals turns out to be correct, we will momentarily give a new, better and more readable proof, of purely mathematical nature. This new proof in turn allows to develop simpler and more efficient implementations.

There is a crucial remark connected to this method applied to the variation norm v. Since we always move along the opposite of external normals and since such normals are in the cone of the vertexes (which represent jumps of the signal) *the projected signal has all jumps reduced in absolute size and possibly a jump is reduced to zero, but its sign is never inverted.*

This remark is of paramount importance for us and we will formalize it momentarily. Indeed this remark is one of the premises, on which leans the superfast projection method. The further required premises and the method itself will be introduced after giving a new proof that the variation norm is of class $\mathbb{P}$.

6 A New Proof That Norm v Is of Class $\mathbb{P}$.

The result that insures that the norm v is of Class $\mathbb{P}$ is clearly central in our theory, because it allows to extend the just described fast projection technique to to the norm v, thereby making it possible an extremely efficient implementation of the concept that the projection of a time series s on a sphere S_v^r with $r < v(s)$ can be seen as a smoothing filter. Indeed the proof will also implicitly indicate a technique to calculate the normals, showing that the complexity of such calculation is minimal. The superfast method of projection introduced in the next section further improves efficiency in respect of computation of normals, because it shows that the computation can be carried out exclusively for the case of maximal proper faces.

We start with a Lemma.

Observe that the matrix TT^* has the following structure::

$$TT^* = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & .. & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & .. & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & .. & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & .. & 0 & 0 \\ .. & .. & .. & .. & .. & .. & .. & .. \\ 0 & 0 & 0 & 0 & 0 & .. & -1 & 2 \end{pmatrix}$$

Here is for example and for more clarity the case $N = 6$:

$$TT^* = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix}$$

The structure of T^*T is similar. We give an example for $N = 6$,since examples may be clearer than the general case and immediately extrapolable to any dimension.

$$T^*T = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

Both matrices have nonzero entries only in the main diagonal and the upper and lower subdiagonals. Both these subdiagonals have all the entries equal to -1. TT^* has the main diagonal of all 2 except the first entry which is 1, while T^*T has the main diagonal of all 2 except he last entry which is 1. Also notice that all rows except the first and the last are the same in the two matrices. Of course both these matrices are selfadjoint and nonsingular.

Remark 11 *It is interesting to note that both this matrices have a self-reproducing property in the following sense. If in* TT^* *we delete the last* $k < N$ *rows and columns we obtain a matrix of the same form but of order* $N - k$*. A similar property holds for* T^*T *but deleting the first* k *rows and columns. Even more than that: if in* TT^* *we delete the first* k *rows and columns, we obtain a matrix of the same form as* T^*T*, with the only exception that the last element of the main diagonal is a* 2 *instead of a* 1*. A similar statement arise deleting the last* k *rows of* T^*T*.*

Recall that π denotes a sequence of N signs ($+$ and $-$) and I_π is the matrix obtained modifying the identity matrix in such a way that the sequence of signs on the main diagonal is π, and let e denote the vector with all 1 entries in R^N. Then we can state the following

Lemma 12 *The vector* TT^*e *is nonnegative and the vector* $I_\pi TT^* I_\pi e$ *is nonnegative for arbitrary* π*. The same statements are true for the matrices* T^*T *and* $I_\pi T^*T I_\pi e$*.*

Proof. The first statement is obvious. As to the second we note that the effect of the two matrices I_π is to leave unchanged the signs of the entries on the main diagonal, while all other entries on the rows and the columns of the same index as a -1 in I_π change signs. But clearly this can only affect the entries of each row of TT^* in the sense that they can only possibly pass from a negative value to a positive value. Hence the sum of entries of each row stands non negative. The analysis leading to the remaining statements is analogous and can be safely omitted. ■

Theorem 13 *The* M *norm* v *for* R^N *is of class* $\mathbb{P}$*.*

Proof. The first step of the proof consists in showing that the thesis is true for the maximal proper faces of S_v^1. In a second step we will eliminate this restriction. That a proper face is maximal means, as we know, that a signal in its relative interior has a nonzero jump at all times. We exploit the facts that the norm p_1 is of class $\mathbb{P}$ (a fact that can be ascertained by inspection), that the normal cone to the maximal faces is a ray (and so there is no problem of choosing the right normal) and the link between normal cones of the maximal faces of the spheres $S_{p_1}^1$ and S_v^1. We know that:

$$S_{p_1}^1 = \mathcal{C}(\{e_i\} \cup \{-e_i\})$$

and

$$S_v^1 = \mathcal{C}(\{T^{-1}e_i\} \cup \{-T^{-1}e_i\})$$

where T^{-1} is the lower triangular matrix in (1). In view of Theorem 6 a maximal face F_{p_1} of $S_{p_1}^1$ has the expression:

$$= \mathcal{C}(\pm e_1, .., \pm e_N) \quad (3)$$

If we form the sequence of vectors that are the difference of the successive extreme points of the polytope in (3), we see that a normal n to F_{p_1} has all ± 1 components and maintains the same sequence of sign that are found in (3), and hence such signs correspond to the signs of the jumps in the signal. With this proviso n has the form:

$$n = \begin{pmatrix} \pm 1 \\ .. \\ \pm 1 \end{pmatrix}$$

Now we want to compute the normal z to the corresponding maximal face of S_v^1:

$$F_v = \mathcal{C}(\pm T^{-1}e_1, .., \pm T^{-1}e_N)$$

Thus for $i = 2, ..N$:

$$(\pm T^{-1}e_i - \pm T^{-1}e_{i-1}, z) = 0 \Leftrightarrow ((\pm e_i - \pm e_{i-1}, T^{-1*}z) = 0 \Leftrightarrow$$

$$\Leftrightarrow ((\pm e_i - \pm e_{i-1}, T^{*-1}z) = 0$$

and hence

$$z = T^*n \quad (4)$$

where T^* is the matrix that has with all 1 in the main diagonal, all -1 in the first upper diagonal and all other entries 0:

$$T^* = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ .. & .. & .. & .. & .. \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

We can express (4) in the following way. Consider an arbitrary sequence π of signs ($+$ and $-$) then all possible vectors n are given by $I_\pi e$. Then we can write $z = T^* I_\pi e$. On the other hand we want to show that z is in the cone of vertexes of the F_v. This means that it should be $z = T^{-1} I_\pi \alpha$, where α is a nonnegative vector and $T^{-1} I_\pi$ is the matrix of vertexes of F_v. Hence from $z = T^{-1} I_\pi \alpha = T^* I_\pi e$, we infer, taking into account that $I_\pi^{-1} = I_\pi$, for any π that we must see if:

$$\alpha = I_\pi T T^* I_\pi e \geq 0, \, \forall \pi$$

But this is insured by the Lemma, and so the first part of the thesis (the one regarding the case of all non zero jumps in the signal) is established. Next, if at a certain time index k the normal n has a zero, we know that at the same index the vector α must have a zero, because the corresponding vertex does not appear in z. Thus in the equation $\alpha = I_\pi T T^* I_\pi e$ we can suppress the kth row and the kth column and write the equation:

$$\alpha_{N-1} = G e_{N-1}$$

where α_{N-1} is the vector obtained from α eliminating the kth component and similarly for e_{N-1} and e, and the matrix G is obtained from the matrix $I_\pi T T^* I_\pi$

eliminating the kth row and the kth column. Hence in G, aside from the kth row that has disappeared, we have deleted in $I_\pi TT^* I_\pi$ only elements outside the main diagonal. This implies that $\alpha_{N-1} \geq 0$, thereby completing the proof, by a straightforward iteration arguments if n has more than one zero. Finally notice that, if the deletion regards the first or the last k rows and columns, the Lemma and subsequent Remark 11 would be in force, simplifying even further our argument. ■

Remark 14 *The second part of the proof can also be achieved looking at the behavior of normal cone to the faces, moving down the lattice of faces. The smaller is the set of vertexes the larger is the normal cone. We illustrate this mechanism considering two adjacent maximal faces, which differ because only one of their extreme points have opposite sign. The normal cone of the face intersection of these two faces contains the normal cones of both the maximal faces. It follows that there is a conical combination of the normals to the two maximal faces, which is in the cone of the vertexes of the intersection face. Clearly this mechanism works in general.*

Remark 15 *The above proof offers one possible approach to the implementation of the fast projection, which is more advanced than the technique used in the cited original implementation. Notice also that, in the case of maximal faces, thanks to this theorem, there is a third alternative, which consists in exploiting the Gram Schmidt procedure on the set of consecutive vertexes differences and using as last vector the last vertex (naturally the sign of the normal must be adjusted to select the external normal).*

It may be useful to give some examples to see how this result works in practice. The first example is for a maximal face. The second for a non maximal face.

Example 16 *We give an example for $N = 5$ in which the signal has nonzero jumps at each time. Let*

$$n = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ 1 \end{pmatrix}$$

so that:

$$z = T^* n = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \\ 2 \\ -2 \\ 1 \end{pmatrix}$$

the matrix of vertexes is:

$$V = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 \\ 1 & -1 & 1 & -1 & 0 \\ 1 & -1 & 1 & -1 & 1 \end{pmatrix}$$

and the matrix of differences of subsequent vertexes is:

$$\Delta V = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & -1 & 0 & 0 \\ 2 & -2 & 1 & 0 \\ 2 & -2 & 2 & -1 \\ 2 & -2 & 2 & -2 \end{pmatrix}$$

so that actually $\Delta V^* z = 0$. Finally, as to the vector α of conical coefficients:

$$\alpha = I_\pi T T^* I_\pi e = \begin{pmatrix} 2 \\ 4 \\ 4 \\ 4 \\ 3 \end{pmatrix}$$

and indeed $V\alpha = z$.

Example 17 *We take n so that one of the jumps is lacking.*

$$n = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

In this case:

$$z = T^* n = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

The matrix of vertexes is now:

$$V = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & -1 & 1 & 0 \\ 1 & -1 & 1 & 1 \end{pmatrix}$$

and:

$$\Delta V = \begin{pmatrix} 1 & 0 & 0 \\ 2 & -1 & 0 \\ 2 & -1 & 0 \\ 2 & -2 & -1 \\ 2 & -2 & 0 \end{pmatrix}$$

so that actually $\Delta V^* z = 0$. Finally, we write the system $\alpha_{N-1} = Ge_{N-1}$:

$$\alpha_{N-1} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 1 \\ 1 \end{pmatrix}$$

As we can see the effect of the matrices I_π not only preserves but improves the property that the sum of entries of each row is non negative. so that:

$$\alpha_{N-1} = \begin{pmatrix} 2 \\ 3 \\ 1 \\ 1 \end{pmatrix}$$

and indeed $V\alpha_{N-1} = z$.

7 The Superfast Version of Projection

In 2009 the author implemented in MatLab an even faster version of the algorithm without publishing it. We give here an account of this version with a proof based on the theory of fast projection (exposed in Session 5) and, of course on Theorem 13.

The method is a simple variant of the fast projection technique. The idea is that of projecting always on maximal faces because the axes where normals have zero components are eliminated shrinking the space. In a final step we get back to the original space and the sought projection.

SUPERFARST PROJECTION METHOD

1- If, at a certain time t, $s(t+1) = s(t)$, delete the sample $s(t+1)$. Repeat until at each t there is a jump.

2- Memorize the position of each deleted sample staring from the last.

The resulting signal is on a space R^{n_1}, with $n_1 < N$.

3- Apply the first step of the fast projection method, where now the signal is necessarily in the cone of vertexes of a maximal face of the sphere S^1_v. In this respect note that we of course apply a scaling descaling step. Since the normal cone to the face is a ray, it is possible to calculate the normal either as illustrated in the proof of Theorem 13 or with a simple Gram Schmidt application, rather then a diagonalization procedure illustrated in [5]. If the resulting signal is in $\mathcal{B}(S^1_v)$, go to step 4. If not, then at one or more times the signal is constant.

Proceed as is step 1 and 2 to delete repeated samples and register their position. Repeat this step until the signal is in $\mathcal{B}(S_v^1)$.

4- Restore the repeated samples starting from the last up to the first.

Before proving that the superfast procedure calculates the actual projection of the signal to S_v^1, we state some helpful lemmas.

Lemma 18 *Consider a signal $s \notin S_v^1$ and project it on S_v^1 with the fast projection method obtaining the unique signal fsi. Than the projection on S_v^1 of any intermediate signal produced by the method at the end of each step is fsi.*

Proof. This is clearly built in the method. ■

The next two lemmas regard signal which have some equal consecutive samples. It is important to bear in mind that these samples have no effect on the variation of the signal.

Lemma 19 *Suppose that the signal $s \notin S_v^1$ has some points of no jump, that is, for some t it is true that $s(t+1) = s(t)$.Then the same relation $fsi(t+1) = fsi(t)$ holds for the projection on S_v^1 for all the same $t's$.*

Proof. In fact, under the hypothesis, the vertexes generating the cone where the signal sits does not include the corresponding vectors in (2) (with either a $+$ and $-$ signs). Thus the normal, which is in the cone of these vertexes, has zeros at each position where a value is repeated. Going on in the procedure, the successive normals belong to subcones of this cone, with the consequence that at each step the values in these positions in fsi never differ from the preceding values. ■

Lemma 20 *Suppose that the signal $s \notin S_v^1$ has some points of no jump, that is, for some t it is true that $s(t+1) = s(t)$. Let's consider one of these at time t, so that $s(t+1) = s(t)$. We assume, without restriction of generality that $s(t+2) \neq s(t+1)$. If we construct a new signal sa, inserting in s at $t+2, .., t+k$ further samples all equal to $s(t+1)$, then the projection of sa on S_v^1 can be obtained from fsi, inserting at $t+2, .., t+k$ further samples all equal to $fsi(t+1)$. Similarly, under the same assumption, if construct a signal sd deleting from s the sample $s(t+1)$, its projection on S_v^1, can be obtained deleting from fsi the sample $fsi(t+1)$.*

Proof. This is a rather direct consequence of the same argument used in the proof of the preceding lemma. ■

We are now in a position that allows to establish the following main

Theorem 21 *The superfast projection procedure provides the projection of s on S_v^1.*

Proof. At this point the thesis of this theorem is a rather immediate consequence of the last lemma. ■

8 Projecting On w Spheres By Means of Projections On v Spheres

We start with a general analysis regarding continuous seminorms in real Hilbert spaces and the projection operator to closed spheres defined by a seminorm.

So let H be a real Hilbert space and let p be a continuous seminorm for H. As usual we denote by S_p^r, the closed sphere around the origin of radius r. In the next theorem we adapt from [2] some basic facts about continuous seminorms and give the expression of the projection of a point of the space $x \notin S_p^r$ on S_p^r itself.

Theorem 22 *The following statements hold for a continuous seminorm p defined on a real Hilbert space H.* ***(i)***

$$\mathcal{N}(p) = p^{-1}(\{0\}) = \{0\}^{-} \text{ is a closed linear subspace of } H$$

(this subspace will be denoted by N_p). ***(ii)*** $p|_{N^\perp}$ *is a continuous norm for the Hilbert space $N^\perp$ and*

$$S^r_{p|_{N^\perp}} = S_p^r \cap N_p^\perp$$

. ***(iii)*** *For $x \in H \backslash N$ (so that we can write $x = x_N + x_{N^\perp}$):*

$$P_{S_p^r} x = P_{S^r_{p|_{N_p^\perp}}} P_{N_p^\perp} x + x_N = P_{S^r_{p|_{N_p^\perp}}} x_{N^\perp} + x_N$$

Proof. In the proof for simplicity we denote $N_p^\perp$ by $N^\perp$. Statement (**i**) is a well known fact of vector topology (see e.g. [1] and for more details [7]). Statement (**ii**) is straightforward because $H = N \oplus N^\perp$, where N and $N^\perp$ are closed complementary linear subspaces. Consequently:

$$x \in N^\perp \text{ and } p(x) = 0 \Rightarrow x \in N \Rightarrow x = 0$$

To prove (**iii**) we apply the Projection Theorem and calculate that, for arbitrary $y \in S_p^r$:

$$(x - P_{S^r_{p|_{N^\perp}}} x_{N^\perp} - x_N, P_{S^r_{p|_{N^\perp}}} x_{N^\perp} + x_N - y) =$$

$$= (x_{N^\perp} - P_{S^r_{p|_{N^\perp}}} x_{N^\perp}, P_{S^r_{p|_{N^\perp}}} x_{N^\perp} + x_N - y_N - y_{N^\perp}) =$$

$$= (x_{N^\perp} - P_{S^r_{p|_{N^\perp}}} x_{N^\perp}, P_{S^r_{p|_{N^\perp}}} x_{N^\perp} - y_{N^\perp})$$

On the other hand $p(y_{N^\perp}) = p(y)$ (this is immediate but anyway proved in [7]) so that $y_{N^\perp} \in S_p^r \cap N^\perp$ and $S^r_{p|_{N^\perp}} = S_p^r \cap N^\perp$. Consequently, applying the Projection Theorem we can affirm that the last term in this chain of equalities is nonpositive, and therefore the theorem is established. ■

We can see from this theorem that the case of a seminorm is reduced to computing first the projection of x on N and on $N^\perp$, then projecting this latter projection on $S^r_{p|_{N^\perp}} = S_p^r \cup N^\perp$ and, finally, adding to the result the projection of x on N.

Let's see how this general result specializes in the specific case of the seminorm w defined on R^N. In this respect we can state the following, where for $s \in R^N$ we define $Av(s) = \frac{\sum s_i}{N}$.

Theorem 23 *Consider the w seminorm on R^N. Then $\mathcal{N}(w)$ is the one dimensional linear subspace of constant functions, that is, $\mathcal{L}(\{e\})$. The orthogonal projection on $\mathcal{N}(w)$ is:*

$$P_{\mathcal{N}(w)}(s) = Av(s)e$$

so that

$$P_{\mathcal{N}(w)^\perp}(s) = s - Av(s)e$$

and

$$\mathcal{N}(w)^\perp = \{s : Av(s) = 0\}$$

Proof. It is obvious that $s \in \mathcal{N}(w)$ iff s is a constant function. Hence to project orthogonally a vector s on the linear subspace $\mathcal{N}(w)$ we have to minimize with respect to the constant c the function

$$\sum (s(i) - c)^2$$

with respect to c. We leave this as an exercise, in which once the first and second derivatives with respect to c are calculated, one can verify that the solution is $Av(s)e$. To prove the last statement notice that for a signal to be orthogonal to any constant nonzero signal it must be

$$\sum s(i)c = 0 \Leftrightarrow \sum s(i) = 0 \Leftrightarrow Av(s) = 0$$

■

Notice that we can calculate:

$$(s - Av(s)e, Av(s)e) = NAv(s)^2 - NAv(s)^2 = 0$$

Also recall that, in view of Theorem 22 $w|_{\mathcal{N}(w)^\perp}$ is a norm.

The consequence of these results, ignoring for simplicity the scaling and descaling procedure which is clearly also valid for seminorms, is that to project a signal s on S^1_w, we have first to calculate $Av(s)e$ and $s - Av(s)e$. Next we have to project $s - Av(s)e$ on $S^1_w \cap N^\perp_w$ where $N^\perp_w = \{s : Av(s) = 0\}$ and, finally, add to the result of this projection the signal $Av(s)e$.

Note that, given the invariance of w under translations by constant signal we can write:

$$S^1_w = \{s : w(s) \leq 1\} = \{s : w(s - s(1)e) \leq 1\} = \{s : v(s - s(1)e\} \leq 1$$

Our main theorem relating projection on spheres of the seminorm w to projections on spheres of the norm v will derive from elaborating on the set S^1_w and from two Lemmas on projections.

Lemma 24 *The sphere S^1_w can be expressed as follows*

$$S^1_w = (S^1_w \cap N^\perp_w) \oplus N_w$$

Proof. Consider $z \in S^1_w$ and write with obvious meaning of symbols

$$z = z_{N^\perp_w} + z_{N_w}$$

Notice that the properties of seminorms imply

$$w(z_{N^\perp_w}) \le w(z) = w(z_{N^\perp_w} + z_{N_w}) \le w(z_{N^\perp_w}) + w(z_{N_w}) = w(z_{N^\perp_w})$$

and hence $w(z) = w(z_{N^\perp_w})$. It follows

$$w(z) \le 1 \Leftrightarrow w(z_{N^\perp_w}) \le 1$$

so that, if $z \in S^1_w$, then $z_{N^\perp_w} \in S^1_w \cap N^\perp_w$. In this way we have proved:

$$S^1_w \subset (S^1_w \cap N^\perp_w) \oplus N_w$$

Conversely, consider a vector y of the form $x+w$ with $x \in S^1_w \cap N^\perp_w$ and $w \in N_w$. Then

$$w(y) = w(y_{N^\perp_w}) = w(x) \le 1$$

This shows that

$$(S^1_w \cap N^\perp_w) \oplus N_w \subset S^1_w$$

thereby completing the proof. ■

Lemma 25 *Suppose that a vector $w \in N^\perp_w$ then*

$$P_{S^1_w} w \in N^\perp_w$$

Proof. In fact, if we consider the projection $P_{S^1_w \cap N^\perp_w} w$ any other point of S^1_w has, in view of the preceding lemma, larger distance from w than $P_{S^1_w \cap N^\perp_w} w$. ■

Lemma 26 *Let H be a real Hilbert space, F a closed subspace and C a closed convex set contained in F. Then, $\forall\ y$,*

$$P_C(y) = P_C P_F y$$

Proof. take any point $z \in C$, and write:

$$y - z = y - P_F y + P_F y - z$$

and hence, by the Pythagorean Theorem,

$$||y - z||^2 = ||y - P_F y||^2 + ||P_F y - z||^2$$

Now $||P_F y - z||^2$ is minimized by $||P_F y - P_C P_F y||^2$. Hence $P_C P_F y$ is the sought projection. ■

We are now ready to state our main result.

Theorem 27 *In the already proved expression:*

$$P_{S_w^1} s = P_{S_w^1 \cap N_w^\perp}(s - Av(s)e) + Av(s)e$$

we have, letting $F = \{s : s(1) = 0\}$:

$$P_{S_w^1 \cap N_w^\perp}(s - Av(s)e) = P_{S_v^1} P_F(s - Av(s)e)$$

Proof. We have already shown that

$$S_w^1 = \{s : s \in R^N, \, v(s - s(1)e\} \leq 1$$

Next we claim that

$$F = \{z = s - s(1)e, \, s \in R^N\}$$

In fact clearly if $z = s - s(1)e$ then $z(1) = 0$, so that

$$\{z = s - s(1)e, \, s \in R^N\} \subset F$$

Conversely, consider $z \in F$, so that $z(1) = 0$ and, for an arbitrary real σ_1, define $s = z + \sigma_1 e$, so that $z = s - s(1)e$. This means that

$$F \subset \{z = s - s(1)e, \, s \in R^N\}$$

completing the proof of our claim. Consequently:

$$S_w^1 = \{s : v(s - s(1)e\} \leq 1 = S_v^1 \cap F \quad (*)$$

Next note that since $s - Av(s)e \in N_w^\perp$, in view of Lemma 25

$$P_{S_w^1}(s - Av(s)e) \in N_w^\perp$$

and hence:

$$P_{S_w^1 \cap N_w^\perp}(s - Av(s)e) = P_{S_w^1}(s - Av(s)e)$$

and also, by (*) and applying Lemma 26:

$$P_{S_w^1 \cap N_w^\perp}(s - Av(s)e) = P_{S_w^1}(s - Av(s)e) = P_{S_v^1 \cap F}(s - Av(s)e) = P_{S_v^1 \cap F} P_F(s - Av(s)e)$$

Now projecting on F consists in zeroing the first component. By Lemma 19, the projection on S_v^1 leaves this first component equal to zero. Thus in the last term, projecting on $S_v^1 \cap F$ or S_v^1 is the same. Hence we have proved that

$$P_{S_w^1 \cap N_w^\perp}(s - Av(s)e) = P_{S_v^1} P_F(s - Av(s)e)$$

thereby establishing the theorem. ■

This theorem shows that projecting on S_w^1 can be achieved by the following algorithm, which uses only projections on S_v^1:

Step 1 - Compute $Av(s)$ and $s - Av(s)e$

Step 2 - Compute the projection $P_{S_v^1} P_F(s - Av(s)e)$

Step 3 - Compute $P_{S_v^1} P_F(s - Av(s)e) + Av(s)e$.

9 Examples of Possible Applications of the Fast Projection Techniques

First of all it is very important to note that, thanks to the results of the penultimate section, all applications can not only be formulated both in terms of the norm v or in terms of the seminorm w, but also efficiently implemented in either case.

It is thus only for simplicity and to fix the ideas that we will refer to the norm v.

In [4] the author gave various numerical examples where a time series is smoothed projecting on a v sphere of radius given by various percentages < 1 of the norm of the given signal.

One can think of a large variety of practical applications. Among these there are e.g. filtering of a noise corrupted signal, seasonal adjustments, portfolio management. We will briefly illustrate them.

In [6] the proposed technique of seasonal adjustment used a step for noise filtering.

9.1 Noise Filtering

Suppose that a signal s is corrupted by noise, that is, we can only observe a signal $os = s + n$, where n is a noise. The noise is assumed to be zero mean Gaussian distributed and uncorrelated. Typically the signal varies more slowly than the noise and the noise increase the variation as observed with respect to the original signal s.

The proposed technique of filtering consists in projecting os on a sequence of variation spheres of radius $r(c) = c \times v(os)$, where c is a percentage, which can be taken for example 99%, 98%,... This is made possible by the fact that the projection is extremely fast. At each step we compute the residual signal and evaluate its von Neumann Ratio, to select the percentage which gives us the maximum of such ratio. In other words, we do a brute force maximization of the von Neumann ratio (briefly VN ratio) of the residual signal:

$$os - P_{S_v^{r(c)}}$$

Once the maximum at c_o has been determined, our estimate of the corrupted signal s will be $P_{S_v^{r(c_o)}}$. Naturally one can think of many variant of this technique varying the assumptions and the selected statistic.

9.2 Seasonal Adjustment

There is a large literature on seasonal adjustment, which is based on linear autoregressive models, that is models in which each value of a time series depends linearly on the preceding p values. There are many established techniques and also institutions tend to adopt standardized algorithms. We cite for example [13]. A relatively recent survey with a large bibliography is [14].

Clearly our methods are profoundly different from those of main stream literature. Moreover, we have an exclusive research intent and not that of proposing an alternative to the current state of art.

However, the techniques illustrated in these references might complement the present one for forecasting purposes. That is, one may fit one of those linear models to the trend series obtained with the present techniques and use an autoregressive linear model to forecast subsequent values.

We model the series as a sum of a trend signal plus a noise signal plus a signal representing seasonal variations.

The author could not find a formal definition of the seasonal component. In [6] a rather ad hoc definition was adopted, while here we would like to make an attempt to more generality.

Definition 28 *Consider a reference interval of time for example a year. To fix the ideas we refer to this case. Then a yearly signal is defined to be a seasonal signal if its time domain can be divided in intervals on which the signal has alternating signs. That is, for example it is positive in the first interval, negative on the second and so on (if the intervals are more than two).*

In what follows we continue to refer to the case where the reference period is the year.

Consider a time series that we want to deseasonalize. We assume that we have monthly data for a number of years. We dispose the series in a matrix, whose rows are the monthly data of each successive year.

The three components of the observed series are determined by a two steps procedure.

In the first step we apply the noise filter of the preceding subsection on the columns of the matrix, in this way getting rid of the noise component.

In the second step we consider each row, that is each year, where now the signal in question derives from Step 1.

For each row (or year) $\sigma(.)$ we apply the smoothing filter for a grid of parameters $0 < c < 1$ and look for values of c for which the residual signal is a seasonal signal. Then we select the value of c_o of c, such that the residual signal is a seasonal signal and its norm $||.||_2$ is maximal with respect to all values of c that produce a residual seasonal signal.

At this point we define the signal $P_{S_v^{c_o v(\sigma)}}\sigma$ to be the trend component of the time series under investigation and $\sigma - P_{S_v^{c_o v(\sigma)}}\sigma$ its seasonal component.

9.3 Portfolio Management

Consider a stock for which it is available a price time series in an interval of time, which represent our signal s. It is assumed that the stock is volatile and the management method aims to take advantage of its volatility to produce a profit.

We successively project s on the spheres $S_v^{c_i v(s)}$, where c_i is a percentage. As usual we consider a grid for the parameter c_i. For example it takes values in an interval say $0,95$ to 0.50 in steps of $0,05$.

For each projection we record the maximum M_{c_i} and minimum m_{c_i} of the filtered signal $fsi(c_i)$. We divide M_{c_i} and m_{c_i} in intervals, defined by the same percentage δ_j for the maximum and minimum, say from $0,95M_{c_i}$ to to $0,50M_{c_i}$and from $0,95m_{c_i}$ to $0,50m_{c_i}$.

For each percentage δ_j we make the following computation. Assume for example that we start with a certain amount in cash (such amount is irrelevant for our purposes, but naturally we can assume that it is the exact amount to buy 100 shares at a price that will be given shortly). When $fsi(c_i)$ intersect the level $\delta_j m_{c_i}$ we buy our 100 shares at the price $\delta_j m_{c_i}$. When the price intersect $\delta_j M_{c_j}$, we sell our shares making a profit. At the next intersection of of $fsi(c_i)$ with δm_{c_j} we will invest all the cash obtained, hence buying more than 100 shares (of course there is a granulosity and it is not granted that we will obtain more than the initial amount of shares) and we go on in this way all along the time series. Assume that we stop with the last possible sale. We register the final amount $R(c_i, \delta_j)$ of cash C at hand divided by the invested cash as a function of c_i and δ_j.

Then we compute the maximum of $R(c_i, \delta_j)$, let it be $R(c_o, \delta_o)$.

For the next future buying and selling operations we fix as buying level $\delta_o m_{c_o}$ and the selling level $\delta_o M_{c_o}$. Then we repeat the computation with the new samples available.

Naturally, one can conceive obvious variant of the procedure changing its free parameters and experiment to see what setting appear to be more satisfactory, and adapting the parameters as suggested by the experience that accumulates managing the portfolio.

10 Research Perspectives

The results in the penultimate section introduce a very efficient technique to smooth signal with respect to the pure seminorm variation w using the variation norm v. Moreover, they open the way to simple numerical experiment aiming at comparing the technique of smoothing using the v variation norm with the technique of smoothing using the pure variation seminorm w. We expect that in practice the results are similar, and the more so the lower is the ratio $|s(1)|/w(s)$, justifying in certain cases the possibility of a simpler procedure in terms of v that approximates the analogous results in terms of the pure variation w. In cases where the use of w is mandatory, our results allow to use w with extreme efficiency, dispite the general difficoulties of handling seminorms.

In the stochastic model mentioned in the section on filtering, the l_2 approach consist in projecting the observation on the increasing family Hilbert subspaces H_t of signal that are measurable with respect to the σ algebra generated by the observations up to the current time. It is obvious that recursivity is built in, because at each time $t+1$ the projection is given by the sum of the projection on the subspace corresponding to H_t plus the the projection on the orthogonal complement of H_t in H_{t+1}. Note that all these projections are linear continuous operators.

We note in passing that the filtered signal is a martingale stochastic process adapted to the sequence of σ algebra generated by the observations.

The mechanism behind the Kalman filter, where the model include a dynamic system driven by noise and with noisy observations, is similarly based on calculating the projections on the closed subspaces of signals measurable with respect to the increasing in time family of σ algebras generated by the observation stochastic process. And the stochastic process of projections is a martingale.

This suggests that martingale theory is an appropriate setting for the theory of Kalman filtering. A literature going in this direction originated from [10] (see for example [11]).

In our case the variation filter, although deterministic, has a stochastic interpretation because, in view of the Projection Theorem, the projection on any closed convex set is a continuous and hence measurable function. However, the calculations on probability measures may be difficult because of the high non linearity of the relevant projection. In this perspective it may interesting to begin with, to apply the variation norm filter to the same model assumed by the Kalman filter and perform a numerical comparison to have a first sense of the difference.

Can we hope to arrive to a recursive version of the variation based filter? The question has in our opinion only a theoretical conceptual interest, because the filter is so fast that recursion would add little on the numerical side for the application of the filter.

We doubt that this is possible at all given the high non linearity of the filter (projection on a polytope) and the difficulty of finding a relation between the projections obtained maximizing the VN ratio of the residue in successive instants of time.

Another interesting challenge would be the extension of the theory presented here in continuous time, with reference to continuous time functions of bounded variation.

We believe that there is a strong motivation for a such a theory, given the well known fact that in the theory of stochastic differential equations, that they be a la Ito or based on weak distributions, not only for the in noise term, for example Brownian motion in the case of Ito differential equation, sample functions have non bounded variation with probability one. In fact the stochastic equation fails to have a smoothing effect, in the sense that the same fact (sample functions have non bounded variation with probability one) is true for the solutions of differential equations both in the Ito sense and in the white noise (weak distribution) sense, see [8].

References

[1] J.L. Kelley and I. Namioka, *Linear Topological Spaces*, Springer, New York, 1963.

[2] P. d'Alessandro, *A Finite Convergence Method Solving Certain Projection Problems*, Research Report 02-81, Feb. 1981, Istituto di Elettrotecnica, Universita' dell'Aquila. (available for download on prof-paolo-dalessandro.com).

[3] P. d'Alessandro, *Optimal Smoothings*, Research Report 01-81, Feb. 1981, Istituto di Elettrotecnica, Universita' dell'Aquila.

[4] P. d'Alessandro, *Optimal Smoothings*, Proceedings of the Symposium ber Operation Research, St. Gallen (Switzerland), Athenaum Printers, August 1982.

[5] P. d'Alessandro and M. Dalla Mora, *Fast Projection Method For a Special Class of Polytopes with Applications*, RAIRO, Vol22, No 4, 1988, pp 347-361.

[6] P. d'Alessandro and E. De Luca, *A new Technique of seasonal adjustment with an Application to Epidemiology data*, Research Report No. 12.82, Oct 1982, Istituto di Elettrotecnica, Universita' dell'Aquila. (available for download on the author's site prof-paolo-dalessandro.com.

[7] P. d'Alessandro, *Vector Topologies and the Representation of Random Variables*, Appied Mathematics and Optimization, Vol.4, No.1, 1977.

[8] P. d'Alessandro, A Germani and M. Piccioni, *Relationships Between the Measures induced by the Ito and White Noise Linear Equations*, Mathematics and Computers in Simulation XXVI, 1984, pp. 368-372.

[9] P. Wolfe, *Finding the Nearest Point to a politope*, Mathematical programming, Vol 11, 1976, pp 128-149.

[10] A.V. Balakrishnan, *A Martingale approach to linear recursive state estimation*, SIAM J. Control (1972).

[11] Arunabha Bagchi, *A new martingale approach to Kalman Filtering*, Information Sciences, Vol. 10, Issue 3, 1976, pp. 187-192.

[12] R.E. Kalman *A new Approach to Linear Filtering and Prediction Problems*, Transactions of the ASME–Journal of Basic Engineering, 82 (Series D): 35-45, 1960.

[13] J. P. Burman, *Seasonal Adjustment by Signal Extraction*, J. R. Statist. Soc. A (1980), 143, Part 3, pp. 321-337.

[14] O. Darne, *Les Méthodes et Logiciels de Désaisonnalisation des Séries Economiques : une Revue de la Littérature*, Journal de la société française de statistique, tome 145, no 4 (2004), p. 79-102