International Journal of Pure and Applied Mathematics
Volume 81 No. 2 2012, 267-304
ISSN: 1311-8080 (printed version)
url: http://www.ijpam.eu

ijpam.eu

INCORPORATING NONNEGATIVITY AND SPATIAL REGULARIZATION CONSTRAINTS USING PROXIMAL PROJECTIONS

Joe Qranfal
Department of Mathematics
Simon Fraser University
8888, University Drive, Burnaby, B.C., CANADA, V5A 1S6

Abstract: It happens that a solution of a problem lacks the property of being positive or of being spatially smooth. Based on proximal projections, we explore techniques in how to incorporate spatial regularization, including the nonnegativity, into that solution. We introduce the general algorithm and more subsequent ones that impose the nonnegativity into a solution and enforce the spatial regularization while using both norms, the 2-norm and the 1-norm, and also via segmentation. The new constrained nonnegative solution possesses theoretical properties that are stated and proved. We validate numerically these algorithms to solve an inverse medical imaging problem.

AMS Subject Classification: 49N45, 49N60, 62G05, 65K10, 90C06, 90C40, 93C41

Key Words: estimation, stochastic filtering, Kalman filter, optimal filtering, convex optimization, proximal, projection, state constraint, recursive, optimal, nonnegative, regularization, reconstruction, iterative algorithm, median, segmentation, Holder, Tikhonov, medical image, dynamic SPECT, time varying, time dependent, cross entropy, nonnegative reconstruction, Markov, expectation maximization, maximum likelihood, temporal regularization

Received: May 20, 2012

1. Introduction

We propose in this paper, grounded on proximal projections, how to incorporate the nonnegativity constraint and more spatial regularization into a not necessarily positive estimator $\hat{x}$ of an unknown nonnegative vector x. Setting negative values of $\hat{x}$ to zero or taking their absolute value would not give in general an acceptable solution. Furthermore, it might happen that the solution $\hat{x}$ includes temporal smoothness, however it would perhaps lack the spatial one specially when we have at hand a spatially ill-posed problem.

The remainder of the paper is organized as follows. First we describe in Section 2 the general setting of our approach and introduce the general algorithm that gives the nonnegative solution and more of the spatial regularization we desire. Subsequently, the next five sections detail this algorithm in different scopes. Section 3 leads to the algorithm that gives the nonnegative solution only. Section 4 details the Tikhonov regularization, which is based on a 2-norm, to impose spatial smoothness while Section 5 details the spatial regularization based on a 1-norm. Section 6 puts the two previous sections in a more general setting when Section 7 brings yet another spatial regularization via segmentation. Then Section 8 states and proves properties of the new introduced nonnegative proximal projected estimator $x^{\star}$. We validate numerically the proposed approaches in Section 9 and we finally conclude in Section 10 summing up our findings.

2. Proximal Projection Approach

Notation. The following notation is used throughout the paper. We denote by $\mathbb{R}^p$ and $\mathbb{R}^p_+$ the p-dimensional Euclidean space and the nonnegative orthant, respectively. The set of all $n \times p$ matrices with real entries is denoted by $\mathbb{R}^{n \times p}$. I denotes the identity matrix; its size is always clear from the context. The operator $\mathrm{Tr}(B)$ denotes the trace of the matrix B, which is the sum of its diagonal components. For a vector u, the Euclidean norm is denoted by $\| \cdot \|$ and $u^\top$ denotes the transpose vector. The $\imath^{th}$ component of a vector $u \in \mathbb{R}^p$ is denoted by $u_\imath$. Let x and y be two random vectors; $\mathbb{E}(x)$ and $\mathbb{E}(x|y)$ denote the expectation of x and the conditional expectation of x given y. The conditional expectation of x_k given $y_1, \cdots, y_s$ and its variance/covariance matrix $\mathbb{E}[(x_k - \hat{x}_{k|s})(x_k - \hat{x}_{k|s})^\top]$ are denoted $\hat{x}_{k|s} = \mathbb{E}(x_k|y_1, \cdots, y_s)$ and $P_{k|s}$. We also refer to $\hat{x}_{k|k}$ and $P_{k|k}$ as simply $\hat{x}_k$ and P_k respectively. We denote by int $\mathbb{B}$ the relative interior of the set $\mathbb{B}$; it is the set that contains all points which are not on the "edge" of $\mathbb{B}$, relative to the smallest subspace in which $\mathbb{B}$ lies.

It so happens that sometimes an estimate or a solution $\hat{x}$ of a problem lacks the property of being positive and/or being spatially smooth. We could have for instance that some or all of the components of the vector $\hat{x}$ are negative while all the ones of the unknown vector x we are solving for are positive. We propose in this paper techniques in how to incorporate the nonnegativity and more of the spatial regularization into $\hat{x}$. We inject a-priori information about the unknown x by restricting the function domain to be nonnegative. This is fundamentally a statistical cornerstone to some of the regularization approaches. The cost function that we aim to minimize is

$$(x-\hat{x})^\top W(x-\hat{x}),$$

where W is a symmetric positive definite weighting matrix. This is equivalent to minimizing

$$\|x-\hat{x}\|_W^2 \, .$$

Given $\hat{x}$ we analyze then a numerical algorithm to compute a nonnegative minimum of the convex function

$$\varphi(x) = \frac{1}{2}\|\hat{x}-x\|_W^2 + \alpha\psi(x) \tag{1}$$

where $\alpha > 0$ is a parameter. In the case where $\psi = 0$, many authors [1, 2, 3] have proposed nonnegative minimization techniques using active set method, Newton method or quasi-Newton method involving a line search strategy which is computationally expensive. We propose here a method based on a proximal approach.

Let us start with the following unconstrained optimization problem

$$\min_{x\in\mathbb{R}^N} \varphi(x) \tag{2}$$

The convex optimization problem (2) has a unique minimizer that we denote by $x^\star = \mathrm{prox}_\psi^\alpha(\hat{x})$; that is

$$\mathrm{prox}_\psi^\alpha(\hat{x}) = \arg\min_{x\in\mathbb{R}^N}\left(\frac{1}{2}\|\hat{x}-x\|_W^2 + \alpha\psi(x)\right)$$

If $W = I$, $\mathrm{prox}_\psi^\alpha$ is the Moreau's proximity operator [4] of index $\alpha \in \left]0,+\infty\right[$ of the function ψ. These operators generalize the projection onto convex sets. In the particular case when ψ is the indicator function of a convex set $\mathbb{B}$, $\psi(v) =$

$\chi_{\mathbb{B}}(v)$ where it is zero if v is in the closed convex set $\mathbb{B}$ and $+\infty$ otherwise, $\mathrm{prox}_{\psi}^{\alpha}$ is the weighted projection of $\hat{x}$ onto the set $\mathbb{B}$ and the orthogonal one if $W = I$. When the set $\mathbb{B}$ is the nonnegative orthant $\mathbb{R}_{+}^{N}$, choosing $\hat{x}^{+} = \max\{\hat{x}, 0\}$ as a projection of $\hat{x}$ may give a good result and this is commonly used [5]. However, due to the weighted norm, such approach is not recommended.

From now on we take $\mathbb{B} = \mathbb{R}_{+}^{N}$. Our approach could be referred to as a generalized proximal method. It is an iterative approach for minimizing the convex and differentiable function φ. Thus we obtained the following algorithm, partial results are here [6] and more here [7],

Algorithm 2.1. Choose $\gamma > 0$ and choose $\alpha \in]0, +\infty[$. Start with $x^0 \in \operatorname{int}\mathbb{B}$. For $\ell = 0,\ 1,\ \ldots$ compute

$$x_{\imath}^{\ell+1} = x_{\imath}^{\ell} \exp\Big(-\gamma(\nabla\varphi)_{\imath}(x^{\ell})\Big), \quad \imath = 1, \ldots, N$$

until convergence.

In the limit, algorithm (2.1) finds an approximate solution $x^{*} = \mathrm{prox}_{\psi}^{\alpha}(\hat{x})$. Derivation and convergence of algorithm 2.1 and choice of the parameter γ are discussed in detail in [7].

This is a generalized proximal approach that computes $x^{\star}$, the weighted proximal projection of $\hat{x}$, onto $\mathbb{R}_{+}^{N}$. It is a proximal method that generalizes the projection operator; and a distinctive iterative algorithm that requires a relatively simple calculation executed repeatedly. The iterations give rise to a sequence of approximate answers that converges to the solution of the problem, $x^{\star}$, regardless of the starting point $x^0 \in \operatorname{int}\mathbb{B}$.

Remark 2.2. Iterative approaches, as it is the case with ours, could serve as a regularization of ill-posed problems. The number of iterations plays the role of the regularization parameter since semi-convergence happens when we deal with noisy images/solutions as it was observed for the EM (expectation maximization) algorithm [5]. An iterative method gives rise to a sequence of approximate answers that, in the best case, converges to the solution of the problem. As the number of iterations increases the iterates get at first closer to the desired solution and then move away. An obvious remedy is to stop the iterations earlier.

Remark 2.3. In case someone is not concerned with the nonnegativity and only interested with other spatial regularization, it suffices to set $\varphi(x) = \alpha\psi(x)$ and start with $x^0 \in \mathbb{R}^N$ instead.

3. Nonnegativity Constraint

We have just covered in Section 2 how we impose nonnegativity while using an iterative algorithm 2.1 with $\psi = 0$ or $\alpha = 0$ to achieve that goal. Using $W = P^{-1}$, where P is the covariance matrix of the error $x - \hat{x}$, gives the optimal nonnegative estimator; more on optimality is covered in Section 8.4. Letting then

$$\begin{aligned}\varphi(x) &= \frac{1}{2}\|x - \hat{x}\|^2_{P^{-\mathbf{1}}}\\ &= \frac{1}{2}(x - \hat{x})^\top P^{-1}(x - \hat{x})\end{aligned}$$

we have

$$\nabla\varphi(x) = P^{-1}(x - \hat{x})$$

and algorithm 2.1 reduces to

Algorithm 3.1. Choose $\gamma > 0$ and start with $x^0 \in \operatorname{int}\mathbb{B}$. For $\ell = 0,\ 1,\ \ldots$ compute

$$x_\imath^{\ell+1} = x_\imath^\ell \exp\left(-\gamma(P^{-1}(x^\ell - \hat{x}))_\imath\right), \quad \imath = 1, \ldots, N$$

until convergence.

The clustering point $x^\star$ of algorithm 3.1 is the nonnegative solution we desire, obtained as a proximal projection of the initial estimate/solution $\hat{x}$. Including the prior knowledge of the nonnegativity into the solution is one way of enforcing a spatial regularization. Next we add another one.

4. Tikhonov Regularization

Having the ability to use prior knowledge concerning x could stabilize the algorithm. Tikhonov regularization [8, 9], known as ridge regression in the statistical community, is our second remedy to help cure ill-posedness. It has been introduced in various settings. It is also known in the literature as Tikhonov-Phillips regularization due to the work of D. L. Phillips [10]. Recall that we are minimizing the function of (1)

$$\varphi(x) = \frac{1}{2}\|\hat{x} - x\|^2_W + \alpha\psi(x) \tag{3}$$

When we are interested in a nonnegative solution only, we set $\psi = 0$. Since we aim for more regularization, we must impose $\psi > 0$. To enforce a Tikhonov regularization type, we choose

$$\psi(x) = \frac{1}{2}\|L(x-\bar{x})\|^2 \tag{4}$$

where $\bar{x}$ is some target value of x and L is some appropriately selected Tikhonov matrix. For instance, if we choose $\bar{x} = 0$ and $L = I$, we are then concerned with a minimum norm solution. If we take $\bar{x} = 0$ and L to be some differential operator, we are then interested in a spatially smooth outcome. Choosing α to be high implies we are relying more in our prior information while having it extremely small means that we are not really interested in a regularized answer. Hence there is a risk of ending up with a solution that is shaped more with our prior and there is a also a risk of ending with an undesired solution in case we forsake our prior. Attaining an equilibrium between including prior information and working with the data only is our goal, while accomplishing it is not an easy endeavor. Next we see our proximal approach at work to achieve regularization of Tikhonov type.

Using equations (3)-(4), we have

$$\nabla\varphi(x) = P^{-1}(x-\hat{x}) + \alpha L(x-\bar{x})$$

and algorithm 2.1 becomes

Algorithm 4.1. Choose $\gamma > 0$ and choose $\alpha \in]0,+\infty[$. Start with $x^0 \in \operatorname{int}\mathbb{B}$. For $\ell = 0,\, 1,\, \ldots$ compute

$$x_\imath^{\ell+1} = x_\imath^\ell \exp\left(-\gamma\Big(\big(P^{-1}(x^\ell-\hat{x})\big)_\imath + \alpha\big(L(x^\ell-\bar{x})\big)_\imath\Big)\right), \quad \imath = 1,\ldots,N$$

until convergence.

Alternative methods exist for regularization of imaging problems. In addition to Tikhonov regularization, we propose next a regularization by a function with better edge preserving properties.

5. Energy Function and Approximation

A cost function that involves a 2-norm as a regularizer, à la Tikhonov-Philips, is usually unsatisfying because we know that many images/unknowns are not globally smooth. They have region boundaries across which the image values change sharply. The quadratic regularization causes the edges to become blurred.

Little variations between neighboring locations are due to noise while large variations are due to the presence of edges. This premise is the basis of most edge preserving regularization schemes including applications to tomography [11, 12, 13]. We need a cost function that favors local smoothness with well defined boundaries. We propose to use a 1-norm instead of the 2-norm in the penalty cost function ψ. Both penalty cost functions are convex functions. However, the 1-norm based one has the advantage that it increases less rapidly than the quadratic function for sufficiently large arguments since it is a linear increase instead of a quadratic one. Thus large differences between neighboring locations are penalized less than with the quadratic penalty. This uses local information to detect if an edge is present or not.

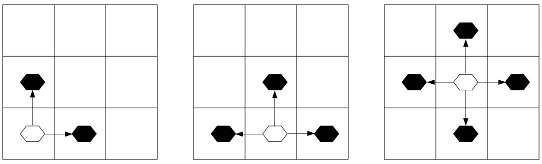

Figure 1: First order neighborhood configuration.

Let $\mathcal{N}_\imath$ denote the set of indexes of voxels/pixels which are neighbors of pixel $\imath$. Define the energy function ψ by

$$\psi(x,m) = \sum_{\imath} \sum_{\jmath \in \mathcal{N}_\imath} w_{\imath\jmath} |x_\imath - m_\jmath| \tag{5}$$

where $w_{\imath\jmath} \geq 0$ are the neighborhood weights, m is a target image. From now on, we assume that $w_{\imath\jmath} = 1 \, \forall \jmath \in \mathcal{N}_\imath$ and is zero otherwise, in the sense that all neighboring locations have the same contribution. Therefore

$$\psi(x,m) = \sum_{\imath} \sum_{\jmath \in \mathcal{N}_\imath} |x_\imath - m_\jmath| \tag{6}$$

The function ψ is related to the Gibbs distribution in the Bayesian imaging context [14, 21]. This energy function does not penalize large differences between locations in the same neighborhood. We adopt a first order neighborhood, see figure 1 for a 2D example. We refer the reader to [14, 16] for higher order neighborhoods. The absolute value function preserves the edges, e.g. abrupt changes in the image texture. This function penalizes deviations within uniform regions without necessarily penalizing the larger differences which occurs at the boundary between two different regions of the image. This is an advantage over the Tikhonov based method. A connection exists between $|x|$ and the median as follow

$$\text{median}\{z_1, \cdots, z_m\} = \arg\min_{s\in\mathbb{R}} \sum_{\imath} |s - z_\imath| \tag{7}$$

For instance, median$\{1, 1, 7\} = 1 = \arg\min_{s\in\mathbb{R}}(|s-1| + |s-1| + |s-7|)$. The result (7) has been proven, that is, the median minimizes the sum of the absolute deviations [17, 18, 19]. Said otherwise, given a set of values $z_1, z_2, \cdots, z_m$, the sum of absolute deviations is minimal when deviations are calculated from the median.

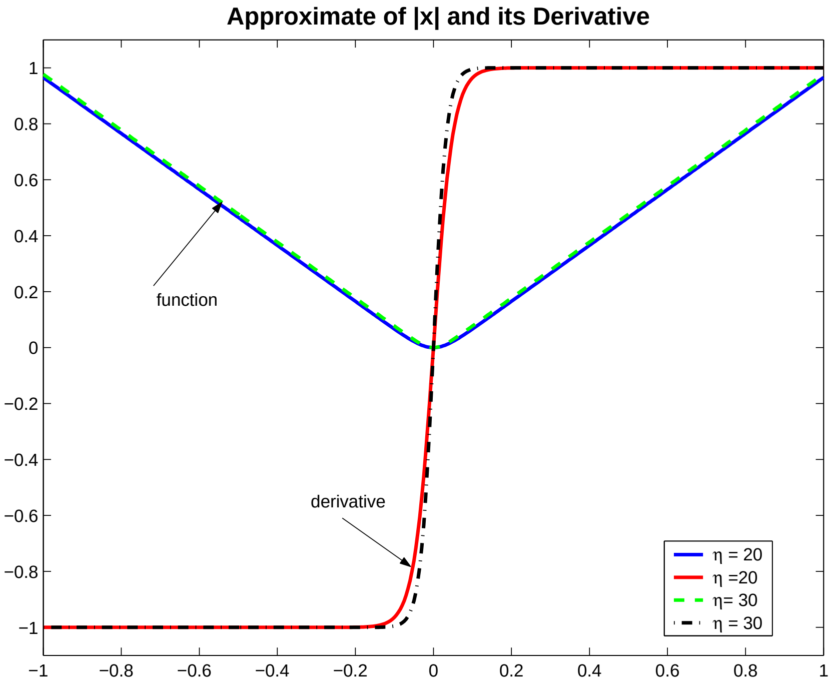

Figure 2: log cosh function and its derivative with two values of η.

We call "Median" the regularization involving the function ψ in (6). This regularization function is particularly suited to recover blocky images with sharp faces and edges. Nonetheless, the absolute value function is convex but not differentiable where the value is zero. Therefore, the optimization problem becomes non differentiable which is computationally impracticable. To circumvent this difficulty we approximate the absolute value with the function

$$\varphi_\eta(x) = \frac{1}{\eta}\log\cosh(\eta\, x) \tag{8}$$

which goes back to Green [20] as an extension of the Geman and McClure potential function [21]. There exists $\delta > 0$ such that when η is close to δ, then $\varphi_\eta(x) \to |x|$. Thus an appropriate choice of η may give a better approximation with numerical advantages in optimization. Note that $\varphi_\eta(x)$ is differentiable and its first derivative is given by $\varphi'_\eta(x) = \tanh(\eta\, x)$, see figure 2. There exist other differentiable functions which approximate the absolute value quite well; take for instance $\varphi_\eta(x) = \sqrt{\eta^2 + x^2}$.

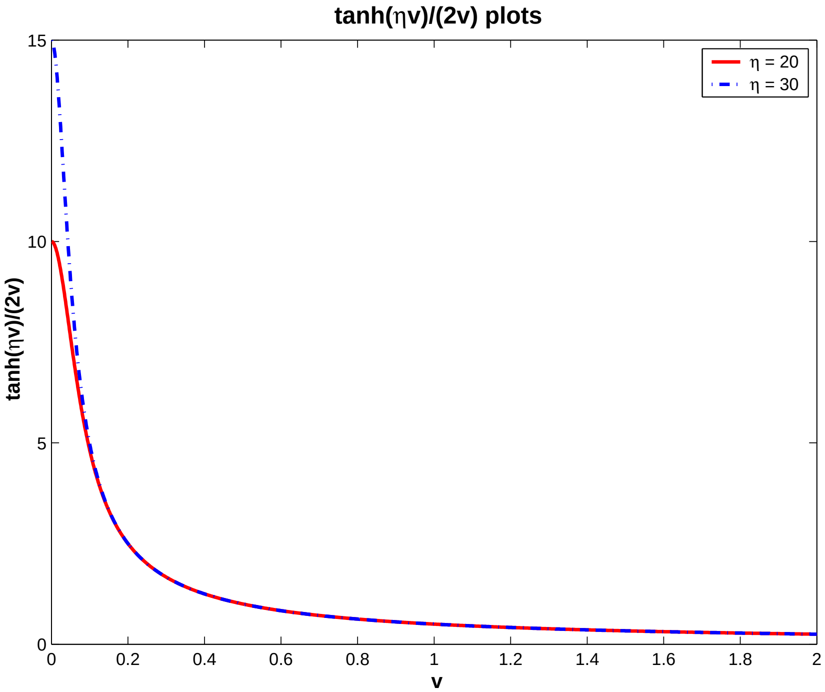

Figure 3: $\tanh(\eta v)/(2v)$ function with two values of η.

In order to encourage smoothing within a region and discourage smoothing across boundaries, Charbonnier et al [22] have suggested three conditions on the weighting function $\psi'(v)/(2v)$, namely

1. $0 < \lim_{v\to 0} \psi'(v)/(2v) = M$ to ensure isotropic smoothing in homogeneous areas.
2. $\lim_{v\to\infty} \psi'(v)/(2v) = 0$ to ensure preservation of edges.
3. $\psi'(v)/(2v)$ is strictly decreasing to avoid instabilities.

where M is finite. Tikhonov regularization is associated with $\psi(v) = v^2$ which is convex. However, $\psi'(v)/(2v) = 1$ does not satisfy the second and third conditions. Total variation is related to $\psi(v) = v$ which is convex, but $\psi'(v)/(2v) = 1/(2v)$ does not satisfy the first condition. Geman and McClure [21] use $\psi(v) = v^2/1 + v^2$ which verifies the three conditions but is not convex. To approximate the convex function we employ, $\psi(v) = |v|$, with the convex and differentiable function $\varphi_\eta(v) = \frac{1}{\eta}\log\cosh(\eta v)$. It has $\psi'(v)/(2v) = \tanh(\eta v)/(2v)$ which satisfies the three conditions with $M = \eta/2$, figure 3 illustrates these facts.

Now we have assembled all the ingredients to study the numerical solution of problem (1). With ψ given in (6), we get

$$\varphi(x,m) = \frac{1}{2}\|\hat{x} - x\|^2_{P^{-1}} + \alpha \sum_{\imath} \sum_{\jmath\in\mathcal{N}_\imath} |x_\imath - m_\jmath| \tag{9}$$

We seek to minimize (9), hence

$$(x^\star, m^\star) = \arg \min_{x\geq 0, m} \varphi(x,m) \tag{10}$$

The function φ is continuous, nonnegative, convex, and coercive so that it has a global minimum $(x^\star, m^\star)$. This is a *joint estimation* of vectors x and m that we solve iteratively via an *alternating algorithm* as follow

$$x^{\ell+1} = \arg\min_{x\geq 0} \varphi(x, m^\ell) \tag{11}$$

$$m^{\ell+1} = \arg\min_{m} \sum_{\imath} \sum_{\jmath\in\mathcal{N}_\imath} |x_\imath^{\ell+1} - m_\jmath| \tag{12}$$

Applying the result (7) and rearranging the double sums in (12), we have

$$m_\jmath^{\ell+1} = \text{median}\{x_\imath^{\ell+1}, \imath \in \mathcal{N}_\jmath\} \quad \jmath = 1, \ldots, N$$

where $\mathcal{N}_\jmath$ is the set of indexes of locations which are neighbors of location $m_\jmath$. Thus (11) becomes

$$x^{\ell+1} = \arg\min_{x\geq 0} \frac{1}{2}\|\hat{x} - x\|^2_{P^{-1}} + \alpha \sum_{\imath} \sum_{\jmath\in\mathcal{N}_\imath} |x_\imath - m_\jmath^\ell| \tag{13}$$

In order to make the minimization problem differentiable, we employ the approximation (8), so that

$$x^{\ell+1} = \arg\min_{x\geq 0} \frac{1}{2}\|\hat{x}-x\|^2_{P^{-\mathbf{1}}} + \frac{\alpha}{\eta}\sum_{\imath}\sum_{\jmath\in\mathcal{N}_\imath} \log\cosh\Big(\eta(x_\imath - m^\ell_\jmath)\Big) \tag{14}$$

Let

$$\varphi(x) = \frac{1}{2}\|\hat{x}-x\|^2_{P^{-\mathbf{1}}} + \frac{\alpha}{\eta}\sum_{\imath}\sum_{\jmath\in\mathcal{N}_\imath} \log\cosh\Big(\eta(x_\imath - m^\ell_\jmath)\Big) \tag{15}$$

Fixing the index $\imath$ and taking the partial derivative of φ w.r.t. $x_\imath$, we obtain

$$\frac{\partial\varphi}{\partial x_\imath}(x) = P^{-1}(x-\hat{x})_\imath + \alpha\sum_{\jmath\in\mathcal{N}_\imath} \tanh\Big(\eta(x_\imath - m^\ell_\jmath)\Big) \tag{16}$$

The general algorithm 2.1 yields the following alternating algorithm.

Algorithm 5.1. Choose $\gamma > 0$ and choose $\alpha \in]0,+\infty[$. Start with $x^0 \in \operatorname{int}\mathbb{B}$ and $m^0_\jmath = \text{median}\{x^0_\imath, \imath \in \mathcal{N}_\jmath\}$ $\jmath = 1,\ldots,N$. For $\ell = 0, 1, \ldots$ compute

$$x^{\ell+1}_\imath \;=\; x^\ell_\imath \exp\left(-\gamma\Big(\big(P^{-1}(x^\ell-\hat{x})\big)_\imath - \alpha\sum_{\jmath\in\mathcal{N}_\imath}\tanh\Big(\eta(x^\ell_\imath - m^\ell_\jmath)\Big)\Big)\right),$$

$\imath = 1,\cdots,N$

$$m^{\ell+1}_\jmath \;=\; \text{median}\{x^{\ell+1}_\imath, \imath\in\mathcal{N}_\jmath\} \quad \jmath = 1,\ldots,N$$

until convergence.

Tikhonov regularization and median regularization differ only in the norm they use, the former uses the 2-norm and the latter uses the 1-norm. Next, we introduce a more generalized context of regularization.

6. Weighted Hölder Filter

Now, we describe some regularization schemes depending on a different choice of $\psi(v)$. When using Tikhonov regularization, we have $\psi(v) = \|Lv\|$, where L is a differential operator. The median regularization function is $\psi(v) = \sum_\imath\sum_{\jmath\in\mathcal{N}_\imath}|v_\imath - \hat{x}_\jmath|$; whose approximation is $\psi_\eta(v) = \sum_\imath\sum_{\jmath\in\mathcal{N}_\imath}\varphi_\eta(v_\imath - \hat{x}_\jmath)$. Another type of regularization function close to the Tikhonov family is the mean regularization, where $\psi(v) = \sum_i\sum_{\jmath\in\mathcal{N}_\imath}|v_\imath - \hat{x}_\jmath|^2$. The two former regularization operators belong to a more general family of filter operators that we call *weighted Hölder filter*. Typically, the weighted Hölder filter replaces each location by the weighted Hölder mean of its neighborhood. That is,

$$\tilde{v}_\imath = \arg\min_{z\in\mathbb{R}} \sum_{\jmath\in\mathcal{N}_\imath} w_{\imath\jmath}|z - v_\jmath|^p$$

where $1 \leq p < +\infty$. The weighting factors $w_{\imath\jmath}$ are usually taken to be equal to 1; that is all the neighboring pixels/voxels/locations/components contribute the same way. Choosing p such that $1 \leq p < +\infty$ makes the above functional convex and differentiable except the case $p = 1$ at the origin. This latter instance was dealt with in Section 5 when $w_{ij} = 1$. The weighted Hölder mean filter transforms an image v to a new one given by $\tilde{v} = \mathrm{M}^p(v)$ such that

$$\mathrm{M}^p(v) = \arg\min_{u\in\mathbb{R}^N} \sum_\imath \sum_{\jmath\in\mathcal{N}_\imath} w_{\imath\jmath}|u_\imath - v_\jmath|^p$$

It is straightforward to show that $\tilde{v}_\imath = \mathrm{M}^p_\imath(v)$. For the sake of clarity, we write $\tilde{v}_\imath = \mathrm{M}^p(v_\jmath,\ \jmath \in \mathcal{N}_\imath)$. There are two special cases with $w_{\imath\jmath} = 1$, $p = 1$ where the weighted Hölder mean coincides with the median, and $p = 2$ where we find the arithmetic mean. The median and mean regularization methods are based on the following alternating minimization formulation,

$$\min_{v,u\in\mathbb{R}^N_+} \left(\frac{1}{2}\|\hat{x} - v\|^2_{P^{-1}} + \alpha \sum_\imath \sum_{\jmath\in\mathcal{N}_\imath} |u_\imath - v_\jmath|^p\right) \tag{17}$$

An optimal solution (v^*, u^*) is such that $v^* = \mathrm{prox}^\alpha_{\psi_p}(u^*)$, and $u^* = \mathrm{M}^p(v^*)$, which can be computed by using the following iterations

$$v^{m+1} = \mathrm{prox}^\alpha_{\psi^m_p}(\hat{x}), \quad u^m = \mathrm{M}^p(v^m), \quad m = 1,\, 2,\, \ldots$$

where the starting guess is $v^0 > 0$, $\psi^m_p(v) = \sum_\imath \sum_{\jmath\in\mathcal{N}_\imath} |u^m_\imath - v_\jmath|^p$, and $p \in \{1,\, 2\}$. We could utilize our algorithm 2.1 to solve $v^{m+1} = \mathrm{prox}^\alpha_{\psi^m_p}(\hat{x})$. Notice that the incorporated target value in the regularization function is not fixed; it is updated during the iteration. Foundations of this alternating minimization, including convergence, are provided in [23].

7. Segmentation Regularization

In medical imaging for instance, we are sometimes not interested in individual intensities/values of each and every pixel/voxel/component of an image (vector solution) but rather on some segment/ROI (region of interest) intensities.

We are then more concerned with a segmented reconstruction [24, 25]. A CT scan for instance might give us an idea about the ROI. In case we have this prior knowledge about the selection of ROI before hand, we could include this constraint, reduce the size of our problem, and have by the same token a better spatial regularization. A commonly used approach is to proceed through a change of variable, see for instance [26]. Let ξ be the image vector of the disjoint p ROIs. Let E represent the $N \times p$ belonging matrix of each pixel to a unique ROI. It has therefore only one 1 in every column and row and the rest of the entries are zeros. A 1 in row $\imath$ and column $\jmath$ implies the $\imath^{th}$ pixel belongs to the $\jmath^{th}$ ROI. Consequently we have the following relation

$$x = E\xi \tag{18}$$

Therefore, in lieu of solving for a bigger size x, we solve for a much smaller size ξ. algorithm 3.1 becomes

Algorithm 7.1. Choose $\gamma > 0$ and start with $\xi^0 \in \operatorname{int} \mathbb{R}^p_+$. For $\ell =$ 0, 1, ... compute

$$\xi_\imath^{\ell+1} = \xi_\imath^\ell \exp\left(-\gamma(P^{-1}(\xi^\ell - \hat{\xi}))_\imath\right), \quad \imath = 1, \dots, p$$

until convergence to $\xi^\star$. Then set

$$x^\star = E\xi^\star$$

8. Properties of the Proximal Projection

We deal with an operator that we refer to as a proximal projection and we would like to assess this estimator's goodness in terms of known properties of parameter estimation. Next, we establish that it is a ML (maximum likelihood) estimator.

8.1. Maximum Likelihood

The nonnegative unknown x gives rise to an estimator $\hat{x}$. Making the assumption that $\tilde{x} = x - \hat{x}$ follows a normal distribution of centre 0 and covariance matrix P, $\tilde{x} \sim \mathcal{N}(0, P)$, the pdf is

$$g(\tilde{x}) = \frac{1}{(2\pi)^{\frac{N}{2}}\sqrt{\det(P)}} \exp\left(-\frac{1}{2}(x - \hat{x})^\top P^{-1}(x - \hat{x})\right) \tag{19}$$

We have

$$\hat{x} = x + \tilde{x} \tag{20}$$

We reverse the process. Knowing or observing $\hat{x}$, we aim to select the nonnegative parameter value $x^\star$ which realizes the largest possible pdf $g(\tilde{x})$. In other words, we look for a constrained ML estimator of (20). Hence we need to minimize

$$\frac{1}{2}(x - \hat{x})^\top P^{-1}(x - \hat{x}) \tag{21}$$

This is equivalent to minimizing $\|x - \hat{x}\|^2_{P^{-\mathbf{1}}}$ in $\mathbb{R}^N_+$. The minimizer is an oblique/weighted projection onto the nonnegative orthant. We obtain the same quantity to minimize while using the likelihood terminology. The log-likelihood function, that is the log of the pdf $g(\tilde{x})$, to be maximized with respect to x is

$$LL(x) = -\frac{1}{2}(x - \hat{x})^\top P^{-1}(x - \hat{x}) + c$$

for some constant c. Because the logarithm is a continuous strictly increasing function over the range of the likelihood, the values which maximize the likelihood will also maximize its logarithm. Since maximizing the logarithm requires simpler algebra, it is the logarithm which is maximized.

Maximum likelihood estimation is based on the assumptions that the distribution of the data is known and the expectation model is correct. ML methods have desirable mathematical and optimality properties. Recall the invariance property of ML estimators [27].

Lemma 8.1. *Suppose that $\hat{\theta}$ is the ML estimator of θ in $\mathbb{R}^n$. Consider the (not necessarily injective) vector mapping $\varrho : \mathbb{R}^n \to \mathbb{R}^m$. Then $\varrho(\hat{\theta})$ is the ML of $\varrho(\theta)$ in $\mathbb{R}^m$.*

We set $C = \mathbb{R}^N_+$, the nonnegative orthant where N is the size of the unknown we are solving for. For the symmetric positive definite matrix P^{-1}, define

$$\begin{aligned} \mathrm{proj}^{P^{-\mathbf{1}}} &: \mathbb{R}^N \to C \\ \hat{z} &\mapsto z^\star = \arg\min_{z \in \mathbb{B}} \|z - \hat{z}\|^2_{P^{-\mathbf{1}}} \end{aligned} \tag{22}$$

where

$$\|z - \hat{z}\|^2_{P^{-\mathbf{1}}} = (z - \hat{z})^\top P^{-1}(z - \hat{z}) \tag{23}$$

Refer to figure 4 that illustrates this projection in a 2D setting.

Assume that the estimator $\hat{x}$ of the nonnegative unknown x is a ML of x in $\mathbb{R}^N$ within the framework of a Gaussian pdf and a linear model; the KF (Kalman filter) estimator of the next Section 9 is a famous example [28, 29].

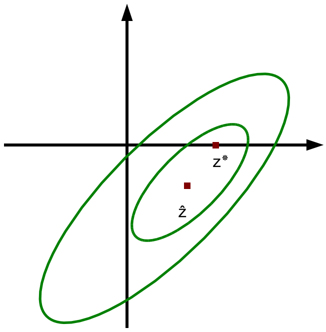

Figure 4: Illustrating the oblique projection for 2D. $\hat{z}$ is the initial estimate and $z^\star$ is the nonnegative estimate as an oblique projection onto the first quadrant.

Moreover, since $\mathbb{B}$ is a closed and convex set and the quadratic form $(z - \hat{z})^\top P^{-1}(z - \hat{z})$ is convex in the variable $z \in \mathbb{R}^N$, we conclude that $\mathrm{proj}^{P^{-\mathbf{1}}}(\hat{z})$ exists and is unique [30], so that the mapping (22) is well defined. Furthermore, $\mathrm{proj}^{P^{-\mathbf{1}}}(v) = v$ if and only if $v \in \mathbb{B}$ because the matrix P^{-1} is positive definite. Recall that x is nonnegative and $\hat{x}$ is its ML in $\mathbb{R}^N$. Apply lemma 8.1 with $\varrho = \mathrm{proj}^{P^{-\mathbf{1}}}$, we deduce that $x^\star_{P^{-\mathbf{1}}} = \mathrm{proj}^{P^{-\mathbf{1}}}(\hat{x})$ is the ML estimator of $x = \mathrm{proj}^{P^{-\mathbf{1}}}(x)$ with respect to the matrix P^{-1}. We have just proved the following theorem.

Theorem 8.2. *Let x be a nonnegative unknown and $\hat{x}$ be its ML estimator in $\mathbb{R}^N$. Let P^{-1} be the symmetric positive definite matrix inverse of the covariance matrix of the error $\tilde{x} = x - \hat{x}$. While $\hat{x}$ is the ML estimator of x in $\mathbb{R}^N$, the proximal projected estimator $x^\star_{P^{-\mathbf{1}}}$ is the constrained ML estimator of x in $\mathbb{B} = \mathbb{R}^N_+$ within the framework of a linear model and Gaussian pdf given by equation* (19).

For ease of notation, we will drop from now on the subscript P^{-1}. Hence we would refer to $x^\star_{P^{-\mathbf{1}}}$ as $x^\star$ and it is assumed that the projection is done with respect to P^{-1}.

8.2. Consistency

The ideal case is to have $\tilde{x} = 0$, however, this is close to impossible because of the Cramér-Rao inequality that set a lower bound on $\mathrm{Cov}(\tilde{x})$ [31]. Hence we must settle for less and require that at least this error converges, with probability 1, to 0 when we augment the size n of the data.

$$\lim_{n\to\infty} P(|\hat{x} - x| < \epsilon) = 1 \quad \forall\epsilon > 0 \tag{24}$$

Besides the invariance property, a ML estimator possesses a second property ([32] and references therein).

Theorem 8.3. *A ML estimator is asymptotically consistent in the sense of equation* (24)*.*

If the estimator $\hat{x}$ is a ML, then so is $x^\star$ by theorem 8.2 and therefore $x^\star$ is also asymptotically consistent by theorem 8.3. Thus we have another property of our estimator $x^\star$ giving the following result.

Corollary 8.4. *When $\hat{x} \in \mathbb{R}^N$ is a ML then the estimator $x^\star \in \mathbb{B}$ is an asymptotic consistent ML estimator of the nonnegative unknown $x \in \mathbb{B}$, that is to say*

$$\lim_{n\to\infty} P(|x^\star - x| < \epsilon) = 1 \quad \forall\epsilon > 0 \tag{25}$$

where n is the size of the data.

The set $\mathbb{B} = \mathbb{R}^N_+$ not being a subspace of $\mathbb{R}^N$ and $\mathrm{proj}^{P^{-\mathbf{1}}}$ being a projection, the mapping (22) is then a nonlinear operator that maps $\hat{x}$ into $x^\star$. Hence $x^\star$ is not a linear function of x even when $\hat{x}$ is a linear one, as it is the case with KF [33]. Usually, there will be no unbiased and optimal nonlinear estimator of x even in the event of normally distributed data. Nonetheless, ML estimators could exhibit asymptotic behavior, to the extent they could be unbiased and optimal for a fixed number of data. In addition, the convex set $\mathbb{B}$, being a cone, has the salient property of "almost" linearity referred to as the nonnegative homogeneity [34], $\mathrm{Proj}_{\mathbb{B}}(\beta z) = \beta\,\mathrm{Proj}_{\mathbb{B}}(z) \quad \forall\,\beta > 0$.

8.3. Unbiasedness

Theorem 8.5. *If the unconstrained estimator $\hat{x}$ of the nonnegative unknown x is an unbiased, $\mathbb{E}(\hat{x}) = \mathbb{E}(x)$, then $x^\star$ is an unbiased constrained estimator of x, meaning*

$$\mathbb{E}(x^\star) = \mathbb{E}(x) \tag{26}$$

Remark 8.6. The case of the unconstrained KF estimator, covered in the next Section 9, was proven to be the BLUE (best linear unbiased estimator) [28, 29, 33]. Simon et. al. [35] proved also the unbiasedness property of the KF and the optimality one with state equality constraints $Dx = d$; that is when the state x is known to belong to a hyperplane. Simon et.al. [36] then proved both properties in the case of state variable inequality constraints $Dx \leq d$. They notice that almost all algorithms for solving such optimization problems belong to the active set methods. They base their argument on this fact assuming that the correct set of active constraints is known a-priori to them. We do not use an active set method to solve the general constrained problem, refer to Section 3, and we do not even know if there are any null components nor where they are located in the unknown x. We are instead solving for lower bounds $x \geq 0$ constraints. Their arguments are therefore not useful to us.

Proof. We seek to find the oblique projection of $\hat{x}$ on the positive orthant $\mathbb{B}$,

$$\min_z \frac{1}{2}\|z - \hat{x}\|_{P^{-\mathbf{1}}} \quad \text{subject to } z \in \mathbb{B}$$

The Lagrangian of the constrained problem is,

$$\mathcal{L}(z, \lambda) = \frac{1}{2}(z - \hat{x})^\top P^{-1}(z - \hat{x}) - \lambda^\top z \tag{27}$$

We formulate the first order Karush-Kuhn-Tucker (KKT) necessary conditions [37, 38, 39],

Stationarity:

$$\nabla \mathcal{L}(x^\star, \lambda^\star) = 0 \tag{28}$$

that is,

$$P^{-1}(x^\star - \hat{x}) - \lambda^\star = 0$$

or

$$\lambda^\star = P^{-1}(x^\star - \hat{x}) \tag{29}$$

Primal feasibility:

$$x^\star \geq 0 \tag{30}$$

Dual feasibility:

$$\lambda^\star \geq 0 \tag{31}$$

Complementary slackness:

$$(\lambda^\star)^\top x^\star = 0 \tag{32}$$

Even though we are interested in a general case of an oblique projection, the argument in the case of the orthogonal projection is interesting in itself. So let us consider the particular case when $P^{-1} = I$. Equation (29) combined with equation (31) gives,

$$\lambda^\star = x^\star - \hat{x} \geq 0$$

so that $x^\star \geq \hat{x}$. But we have $x^\star \geq 0$, equation (30), and $(x^\star - \hat{x})^\top x^\star = 0$, equation (32), which implies that $x^\star = \max(\hat{x}, 0)$.

Recall Jensen's Inequality [34]. If x is a random variable such that

$$x \in \{z \in S \subseteq \mathbb{R}^m \mid g(z) < +\infty\}$$

with probability one, and g is convex, then we have

$$g(\mathbb{E}(x)) \leq \mathbb{E}(g(x)) \tag{33}$$

provided the expectations exists. We apply Jensen's Inequality (33) with the convex function $\max(y, 0)$ to get,

$$0 \leq \max(\mathbb{E}(\hat{x}), 0) \leq \mathbb{E}(\max(\hat{x}, 0))$$

$$0 \leq \mathbb{E}(\hat{x}) \leq \mathbb{E}(x^\star) \tag{34}$$

Using equation (32) and the fact that the error and the state are uncorrelated [29], we have

$$\mathbb{E}(\lambda_\imath^\star x_\imath^\star) = \mathbb{E}(\lambda_\imath^\star)\mathbb{E}(x_\imath^\star) = 0 \quad \forall \imath \tag{35}$$

Recall that we assume that the output $\hat{x}$, as an unconstrained estimator of the positive unknown x, to be unbiased; that is $\mathbb{E}(\hat{x}) = \mathbb{E}(x) \geq 0$. There are two cases to consider for equation (35). On one hand, if $\mathbb{E}(\lambda_\imath^\star) = \mathbb{E}(x_\imath^\star - \hat{x}_\imath) = 0$ for some $\imath$, then $\mathbb{E}(x_\imath^\star) = \mathbb{E}(\hat{x}_\imath)$. On the other hand, if $\mathbb{E}(x_\imath^\star) = 0$ for some $\imath$, then using inequality (34) we have $0 \leq \mathbb{E}(\hat{x}_\imath) \leq \mathbb{E}(x_\imath^\star) = 0$; that is $\mathbb{E}(\hat{x}_\imath) = 0$. Both cases sum up to

$$\mathbb{E}(x^\star) = \mathbb{E}(\hat{x})$$

We therefore conclude the proof of the unbiasedness in the case of orthogonal projection, $\mathbb{E}(x^\star) = \mathbb{E}(x)$.

Let us now proceed to the more general case when the projection is oblique. The complementary slackness says that $(x^\star)_\imath \lambda_\imath^\star = 0 \; \forall \imath$. Since the error and the state are uncorrelated [29], we have $\mathbb{E}(x_\imath^\star \lambda_\imath^\star) = \mathbb{E}(x_\imath^\star)\mathbb{E}(\lambda_\imath^\star) = 0$. There are two cases to consider. On one hand, if $\mathbb{E}(\lambda_\imath^\star) = 0$ for some $\imath$, then $P^{-1}\mathbb{E}(x_\imath^\star - \hat{x}_\imath) = 0$ so that $\mathbb{E}(x_\imath^\star) = \mathbb{E}(\hat{x}_\imath)$, since P^{-1} is invertible and its inverse is P. On the other hand, if $\mathbb{E}(x_\imath^\star) = 0$ for some $\imath$, let $\mathcal{I}$ be the set of these $\imath$.

Since $\forall \imath \in \mathcal{I}$, $\mathbb{E}(x^\star_\imath) = 0$, then $x^\star_\imath = 0$ since $x^\star \geq 0$; which implies $\lambda^\star_\imath = (P^{-1}(x^\star - \hat{x}))_\imath > 0$. Since $\forall \imath \in \overline{\mathcal{I}}$, $x^\star_\imath > 0$, then $(P^{-1}(x^\star - \hat{x}))_\imath = 0$. The matrix P^{-1} is symmetric positive definite, therefore

$$\begin{aligned}\sum_{\imath \in \mathcal{I} \cup \overline{\mathcal{I}}} (x^\star - \hat{x})_\imath (P^{-1}(x^\star - \hat{x}))_\imath &= \sum_{\imath \in \mathcal{I}} (x^\star - \hat{x})_\imath (P^{-1}(x^\star - \hat{x}))_\imath \\ &\quad + \sum_{\imath \in \overline{\mathcal{I}}} (x^\star - \hat{x})_\imath (P^{-1}(x^\star - \hat{x}))_\imath \\ &= \sum_{\imath \in \mathcal{I}} (x^\star - \hat{x})_\imath (P^{-1}(x^\star - \hat{x}))_\imath \\ &= (x^\star - \hat{x})^\top P^{-1} (x^\star - \hat{x}) \\ &\geq 0\end{aligned}$$

Consequently,

$$\sum_{\imath \in \mathcal{I}} (x^\star - \hat{x})_\imath (P^{-1}(x^\star - \hat{x}))_\imath \geq 0$$

or

$$-\sum_{\imath \in \mathcal{I}} \hat{x}_\imath (P^{-1}(x^\star - \hat{x}))_\imath \geq 0$$

Passing to the expectation we have,

$$-\sum_{\imath \in \mathcal{I}} \mathbb{E}(\hat{x}_\imath) \mathbb{E}((P^{-1}(x^\star - \hat{x}))_\imath) \geq 0$$

We know that $\mathbb{E}(\hat{x}_\imath) \geq 0$ and $\mathbb{E}((P^{-1}(x^\star - \hat{x}))_\imath) > 0$; this implies $\mathbb{E}(\hat{x}_\imath) = 0$ for all $\imath \in \mathcal{I}$. The two cases that we considered here allow us to conclude that $\mathbb{E}(x^\star) = \mathbb{E}(\hat{x}) = \mathbb{E}(x)$. The proximal projected $x^\star$ is an unbiased constrained estimator of the nonnegative unknown x. □

8.4. Optimality

When the unconstrained estimator $\hat{x}$ of the unknown x is unbiased and a ML then we have seen that $x^\star \in \mathbb{B}$ is a ML, consistent, and unbiased constrained estimator of the nonnegative unknown x. However the KF estimator in $\mathbb{R}^N$ for instance, discussed in Section 9, is shown to be an unbiased and optimal estimator of $x \in \mathbb{R}^N_+$ while being linear. Such an estimator is not anymore optimal when we seek the estimate to be positive as the unknown x. Does then the positive $x^\star$ perform better than the initial solution $\hat{x}$ since $x^\star$ is rather a constrained ML in $\mathbb{B}$? The answer is yes in the following sense.

Theorem 8.7. *Assume that $\hat{x} \in \mathbb{R}^N$ is unbiased unconstrained estimator of the nonnegative unknown $x \in \mathbb{R}_+^N = \mathbb{B}$. The constrained estimator $x^\star \in \mathbb{B}$ of $x \in \mathbb{B}$ performs better than $\hat{x}$ in the sense that the mean square error of $x^\star$ is smaller than the mean square error of $\hat{x}$,*

$$\operatorname{MSE}(x^\star) \leq \operatorname{MSE}(\hat{x}) \tag{36}$$

which is equivalent to say

$$\operatorname{Tr}\big(\operatorname{Cov}(x - x^\star)\big) \leq \operatorname{Tr}\big(\operatorname{Cov}(x - \hat{x})\big)$$

Recall the definition of the operator MSE (mean square error) giving by

$$\begin{aligned} \operatorname{MSE}(x^\star) &= \mathbb{E}\Big((x - x^\star)^\top (x - x^\star)\Big) && (37) \\ \operatorname{MSE}(\hat{x}) &= \mathbb{E}\Big((x - \hat{x})^\top (x - \hat{x})\Big) && (38) \end{aligned}$$

Before giving a proof, we need a definition of nonexpansive mappings [5].

Definition 8.8. Let χ be a Hilbert space and $\|.\|$ a vector norm. An operator T, not necessary linear, in χ is a nonexpansive mapping if $\forall\, z_1, z_2 \in \chi$, then

$$\|T(z_2) - T(z_1)\| \leq \|z_2 - z_1\| \tag{39}$$

Recollect a classical result in convex optimization about projection onto closed convex sets [40].

Proposition 8.1. *Let $u, v \in \chi$, a Hilbert space with $\|.\|$ as a vector norm, and let $\operatorname{proj}_F(u), \operatorname{proj}_F(v)$ be the corresponding projections onto any closed and convex set F, then*

$$\|\operatorname{proj}_F(u) - \operatorname{proj}_F(v)\| \leq \|u - v\| \tag{40}$$

This property of the projection onto closed convex set states that the projection operator is nonexpansive (definition 8.8), see figure 5 for a geometrical intuition of this result.

Proof. In the case of unbiased estimators, like both $x^\star$ and $\hat{x}$, it is known [27] that equations (37) and (38) entail

$$\begin{aligned} \operatorname{MSE}(x^\star) &= \operatorname{Var}(x^\star) \\ \operatorname{MSE}(\hat{x}) &= \operatorname{Var}(\hat{x}) \end{aligned}$$

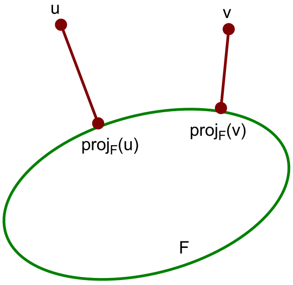

Figure 5: Illustrating the inequality (40) for projection onto closed convex set.

Apply proposition 8.1 to $\chi = \mathbb{R}^N, F = \mathbb{B} = \mathbb{R}^N_+$ with $\text{proj}_{\mathbb{B}}$ the oblique projection onto $\mathbb{B}$ as defined in the operation (22), so that $x^\star = \text{proj}_{\mathbb{B}}(\hat{x})$. Since $x \in \mathbb{B}$ means $x = \text{proj}_{\mathbb{B}}(x)$, we have

$$\|x - x^\star\|^2 \leq \|x - \hat{x}\|^2 \tag{41}$$

The expectation operator is positive, $Y \geq 0 \Rightarrow \mathbb{E}(Y) \geq 0$. Thus inequality (41) implies

$$\mathbb{E}\left((x - x^\star)^\top (x - x^\star)\right) \leq \mathbb{E}\left((x - \hat{x})^\top (x - \hat{x})\right)$$

that is

$$\text{MSE}(x^\star) \leq \text{MSE}(\hat{x})$$

Note

$$\text{MSE}(Y) = \text{Tr}\left(\text{Cov}(Y)\right)$$

Hence theorem 8.7 is also saying

$$\text{Tr}\left(\text{Cov}(x^\star - x)\right) \leq \text{Tr}\left(\text{Cov}(\hat{x} - x)\right) \tag{42}$$

The estimator $x^\star \in \mathbb{B}$ performs better than $\hat{x}$ in the minimum variance sense.

□

8.5. Summary

We are solving for $x \in \mathbb{R}^N_+$ and we have at hand an estimator $\hat{x}$ of x but in $\mathbb{R}^N$. We know also that the estimator $\hat{x}$ is a ML, unbiased, and optimal. In this section, we established general properties of the proximal projected estimator $x^\star \in \mathbb{R}^N_+$ of $\hat{x}$. We have shown that using a weighted projection, with respect to the symmetric positive definite matrix P^{-1}, the proximal projection $x^\star$ as an estimator in $\mathbb{R}_+^N$ performs better than $\hat{x} \in \mathbb{R}^N$. Not only does the constrained estimator $x^\star$ conserve the same properties as $\hat{x}$ of being unbiased, consistent, and ML, but in addition it is optimal in the sense that the MSE($x^\star$) is smaller than MSE($\hat{x}$).

Next, we test numerically the four algorithms 3.1, 4.1, 5.1, and 7.1 that we have developed so far.

9. Numerical Experiments

We validate numerically in this section the four algorithms 3.1, 4.1, 5.1, and 7.1 to solve a medical imaging reconstruction problem arising in a nuclear medicine modality, see Subsections 9.1 and 9.2. The original image/unknown x, we are solving for, has only nonnegative values. The initial solution $\hat{x} \in \mathbb{R}^N$ is obtained via the KF algorithm, this is shown in Subsection 9.3. However the vector solution $\hat{x}$ is meaningless since it has some negative components. We use algorithm 3.1 to get the nonnegative solution $x^\star$, details are given in Subsection 9.4. The KF ensures by its nature a temporal regularization but fails to ensure a spatial one. Tikhnonov spatial regularization is implemented in Subsection 9.5, the Median one is shown in Subsection 9.6, and regularization via segmentation follows in Subsection 9.7. Subsection 9.8 concludes this Section 9.

9.1. Stochastic Problem Setting

Single photon emission computed tomography (SPECT) is a nuclear medicine technique where a radiopharmaceutical, a judiciously designed chemical tagged with a radioactive isotope, is administered to the patient, usually by intravenous injection. It is chosen to amass in a targeted organ or region of the body, the heart or the brain for instance. The radioactive isotope emits photons which are detected by an external device, the gamma camera, at several angular positions around the patient body. Data from these 2D angular views/projections are reconstructed into a 2D or 3D image. This reconstructed radionuclide distribution from measured data is a useful tool to clinically interpret and diagnose unhealthy tissue.

We are then given the nonnegative vector $y(t)$, the data collected at time t. The unknown vector $x(t)$ is not directly observable and usually refers to a 2D or 3D spatial object/image that varies over time. The goal is to reconstruct the object/image $x(t)$ from the measured data $y(t)$. We discretize both vectors $x(t)$

and $y(t)$ as follows. Let $t_k,\ k = 1, \ldots, S$, be a sequence of data acquisition times, N the total number of locations of x and M the equal number of measurements at each time t_k. We denote by $x_k \in \mathbb{R}_+^N$ and $y_k \in \mathbb{R}_+^M$ the spatial distribution and the measured data at the k^{th} instant of time. The observations $y_1, y_2 \ldots, y_S$ are independent random vectors. Furthermore, each observation y_k depends on x_k only. We describe next an optimal linear real-time reconstruction of a dynamic image.

On one hand, the activities sequence $x_1, x_2 \ldots, x_S$ satisfy a linear evolution property with a given time varying nonnegative transition/evolution matrix $A_k \in \mathbb{R}^{N \times N}$. That is

$$x_k = A_k x_{k-1} + \mu_k \tag{43}$$

where μ_k is the error random vector in modeling the transition from x_{k-1} to x_k with $\mathbb{E}(\mu_k)$ zero and covariance matrix Q_k. On the other hand, we assume that we are given a nonnegative system matrix C_k such that the observation y_k and activity x_k vectors are related by the following

$$y_k = C_k x_k + \nu_k \tag{44}$$

where ν_k is the noise vector in recording the data with $\mathbb{E}(\nu_k)$ zero and covariance matrix R_k. In practice, high noise level makes the problem very challenging if no prior information is available. Each acquisition time constitutes a separate reconstruction problem. We start off with initial guesses $\hat{x}_0$ and P_0, we obtain for $k = 1, \cdots, S$ the subsequent KF solution $\hat{x}_k$ recursively as

Predicting Step Assume we have an initial estimate activity $\hat{x}_{0|0}$ and its covariance matrix $P_{0|0}$. For $k = 1, \cdots, S$, compute the following steps that yield the predicted variance and activity,

$$\begin{aligned} P_{k|k-1} &= A_k P_{k-1|k-1} A_k^\top + Q_k && (45)\\ \hat{x}_{k|k-1} &= A_k \hat{x}_{k-1|k-1} && (46) \end{aligned}$$

Correcting Step Then compute the following correcting steps that yield the filtered variance and activity,

$$\begin{aligned} K_k &= P_{k|k-1} C_k^\top (C_k P_{k|k-1} C_k^\top + R_k)^{-1} && (47)\\ P_{k|k} &= (I - K_k C_k) P_{k|k-1} (I - K_k C_k)^\top + K_k R_k K_k^\top && (48)\\ \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k (y_k - C_k \hat{x}_{k|k-1}) && (49) \end{aligned}$$

where K_k is the Kalman gain.

Smoothing Step The recursive algorithm that calculates the estimate $\hat{x}_{k|S}$, where S denotes the total number of measurements, is called the *Kalman smoother*. We refer to $\hat{x}_{k|S}$ as $\hat{x}_k$ too. That is to get smoothed values, run the following backward recursion for $k = S - 1, \cdots, 1$:

$$\begin{aligned} J_k &= P_{k|k} A_k^\top P_{k+1|k}^{-1} && (50) \\ P_{k|S} &= P_{k|k} + J_k (P_{k+1|S} - P_{k+1|k}) J_k^\top && (51) \\ \hat{x}_{k|S} &= \hat{x}_{k|k} + J_k (\hat{x}_{k+1|S} - \hat{x}_{k+1|k}) && (52) \end{aligned}$$

where J_k is called the backward gain. In spite of the nonnegativity of A_k, P_{k-1}, Q_k, C_k, R_k, and y_k, the update equation for $\hat{x}_k$ in (49) can not guarantee its nonnegativity. Inversion of a matrix and subtraction of vectors are involved in equations (47) and (49) which may introduce negative elements. This is not feasible in nuclear medicine and also gives unidentifiable images. Setting negative values of the reconstructed activity to zero or taking the absolute value does not give an acceptable solution. The proposed solution here is to project $\hat{x}_k$ onto $\mathbb{R}_+^N$ to have $x_k^\star$ using a proximal approach. From now on we drop the index k from $\hat{x}_k$ and $x_k^\star$ and we refer to them as $\hat{x}$ and $x^\star$ respectively.

Assume we are provided with system projection matrices $C_1, \cdots, C_S$ and data $y_1, \cdots, y_S$. We give a systematic method on how we can implement our proximal projection approach.

Procedure 9.1.

step 1 Start with an initial guess vector $\hat{x}_{0|0}$ and an initial covariance matrix $P_{0|0}$.
For $k = 1, \cdots, S$, execute step 2, 3, and 4

step 2 At the k^{th} recursion, choose Q_k and R_k

step 3 Use the Kalman filter algorithm, Eqs. (45) to (49), to calculate $\hat{x}_{k|k-1}$ and $\hat{x}_{k|k}$

step 4 Use either algorithm 3.1, 4.1, 5.1, or 7.1 to achieve nonnegativity and/or smoothness of $\hat{x}_{k|k}$ if necessary

step 5 For $k = S - 1, \cdots, 1$, use the Kalman smoother algorithm, Eqs. (50) to (52), to calculate the estimate $\hat{x}_{k|S}$ and

step 6 Use algorithm 3.1 or 7.1 to achieve nonnegativity of $\hat{x}_{k|S}$ if necessary.

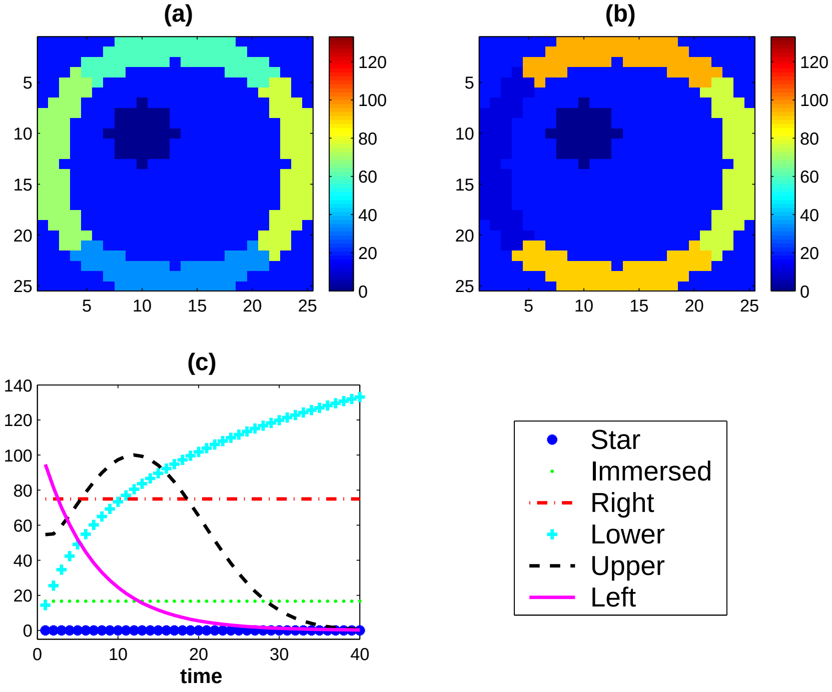

Figure 6: Simulated annulus with its different ROIs and their TACs: (a) simulated activity at time 3, (b) at time 15, and (c) TACs of the 6 different ROIs.

9.2. Simulation

Our simulated phantom is composed of six regions of interest (ROI) or segments. Each ROI has a different time activity curve (TAC), see figure 6. The example investigated in this work is based on the teboroxime dynamics in the body during first hour post injection. The choice of the time activity curves (TACs) is motivated by the behavior of liver, healthy myocardium, muscles, stenotic myocardium, and lungs, respectively. Only one slice is modeled; that is we simulate a 2D object. The star-like shape placed on the left ensures that the phantom is not entirely symmetrical. We simulate 120 projections over 360°, one projection for every 3° with attenuation and a 2D Gaussian detector response.

There are three camera heads consisting of 64 square bins each measuring 0.625 cm in each side, see figure 7. The distance from the annulus to the camera head rotation axis is 30 cm. We simulate 40 time instances for three heads; that

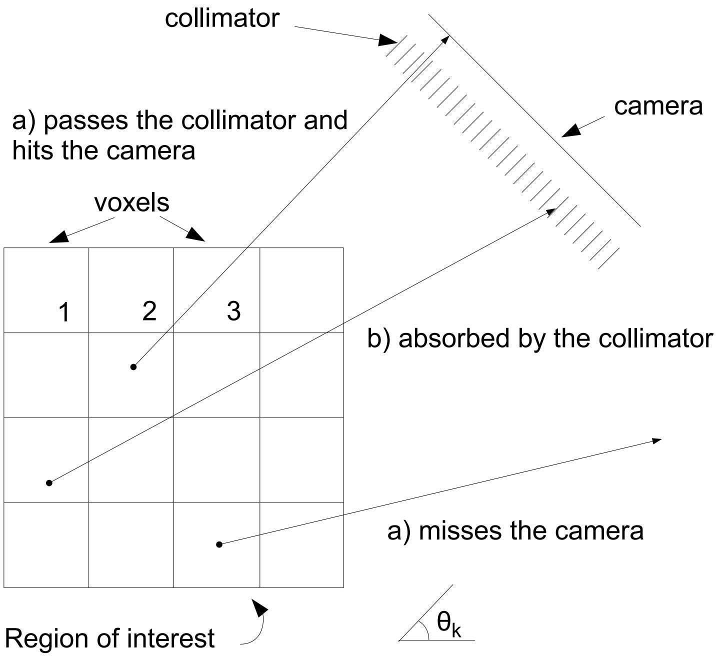
collimator
camera
a) passes the collimator and hits the camera
voxels
1
2
3
b) absorbed by the collimator
a) misses the camera
Region of interest
θ_k

Figure 7: Photon radiating from the region of interest. a) passes the collimator and hits the camera, b) absorbed by the collimator, c) misses the camera.

is we have 3 × 40 = 120 projections for a camera rotating clock wise (CW) in a circular orbit. Head 1 starts at $-60°$, head 2 at $60°$, and head 3 at $180°$. A low energy high resolution (LEHR) collimator is used with a full width at half maximum (fwhm). We determine the blurred parallel strip/beam geometry system matrices for all projections with resolution recovery and attenuation correction [41].

We have 64 projection/measurement values for each head, which amounts to a total of 192 observations at each time frame. The size of the image we aim to reconstruct is $625 = 25 \times 25$ dixels; this is an under-determined problem with a ratio of 1:3.25 of data to unknowns. It is an ill-posed problem. We have six kinds of TACs that are very representative for clinical applications. The annulus has four arcs that we name “Left”, “Upper”, “Right”, and “Lower” according to their location. The activity is decreasing in the Left arc, increasing-decreasing in the Upper arc, constant in the Right arc, and increasing in the Lower arc; see figure 6. The star-like shape has zero activity within it and is called the “Star” region; we refer to it as “Background” too. The annulus is immersed within a region that has the sky-blue color in our figure. It is called “Immersed” and has a constant activity. We have six ROIs in total.

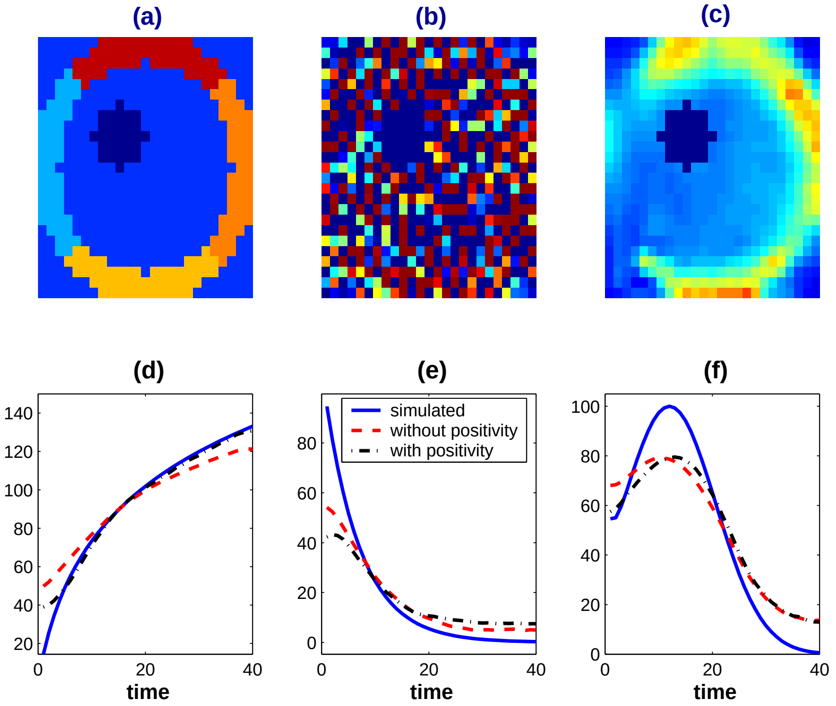

Figure 8: Without and with positivity reconstructed images and TACs: (a) simulated at time 9, (b) reconstructed without enforcing positivity, (c) reconstructed with enforcing positivity, (d), (e) and (f) TACs of Lower, Left, and Upper regions respectively.

9.3. Preliminary Tests

Right after each time step k of KF, we took the absolute value of the reconstructed activity, $\text{abs}(\hat{x}_k)$, and, in a second experiment, we set to zero its negative values; that is we took $\max(\hat{x}_k, 0)$ where $\hat{x}_k$ is the Kalman output. We repeated the experiment in applying $\text{abs}(\hat{x})$ and $\max(\hat{x}, 0)$ only once at the end of the algorithm. The images were unidentifiable in all these four cases. Recall that using $\max(\hat{x}_k, 0)$ means we apply the orthogonal projection. Consequently the fact that we did not get any meaningful image confirms theorem 8.2; refer also to Section 2 for more details. We ran KF without enforcing the positivity; that is we applied procedure 9.1 without steps 4 and 6. We then ran KF with positivity; that is we applied procedure 9.1 with algorithm 3.1 in steps 4 and 6. The averaged TACs in every region look similar without and with positivity which is in accordance with theorem 8.5 emphasizing the unbiasedness of $x^\star$ and $\hat{x}$, see figure 8. This is explained by the fact that KF gives an optimal estimate on average at each time k for every region but not for every dixel. However, we got only meaningless images without positivity. It is clear that our approach of enforcing positivity in the output images, that come from the classical KF algorithm, is better than using the *abs* and *max* functions or than just doing nothing. General theoretical results about these observations have been established in Section 8. The proximal approach to enforce nonnegativity is indeed an efficient tool to enforce some spatial regularization, refer to Section 3 for more details and to theorem 8.7 stating the fact that the estimator $x^\star$ performs better than $\hat{x}$ in the MSE sense.

9.4. Non Regularized Kalman

The Kalman algorithm gives two reconstructed images, one after the filtering step that we call “filtered”, another one after the smoothing step, that we call “smoothed”, refer to Section 9.1. We are interested in the behavior of dynamic regions. Therefore, from now on we only show the TACs of the lower, left, and upper arcs. Figure 9 depicts images of the simulated/true annulus at various times together with the reconstructed filtered and smoothed ones when we enforce the positivity.

Figure 10 displays the averaged TACs over three different regions. We plot the true ones, shown in blue, the reconstructed filtered ones, shown in red, and the reconstructed smoothed ones, shown in black. Note that both reconstructed TACs look pretty close to the true ones shape-wise and in quantity/intensity/color. This is very interesting since we use only a basic approximation, namely first-order random walk, to describe the evolution model. Smoothed TACs look indeed “smoother” than the filtered ones as promised by the smoothing step in the Kalman algorithm. There are however some differences in intensity between pixels within the same region, see for instance the right and lower arcs. Smoothing is done over time with KF; that is we have a temporal regularization but not a spatial one. Tikhonov spatial regularization is the topic of the next section.

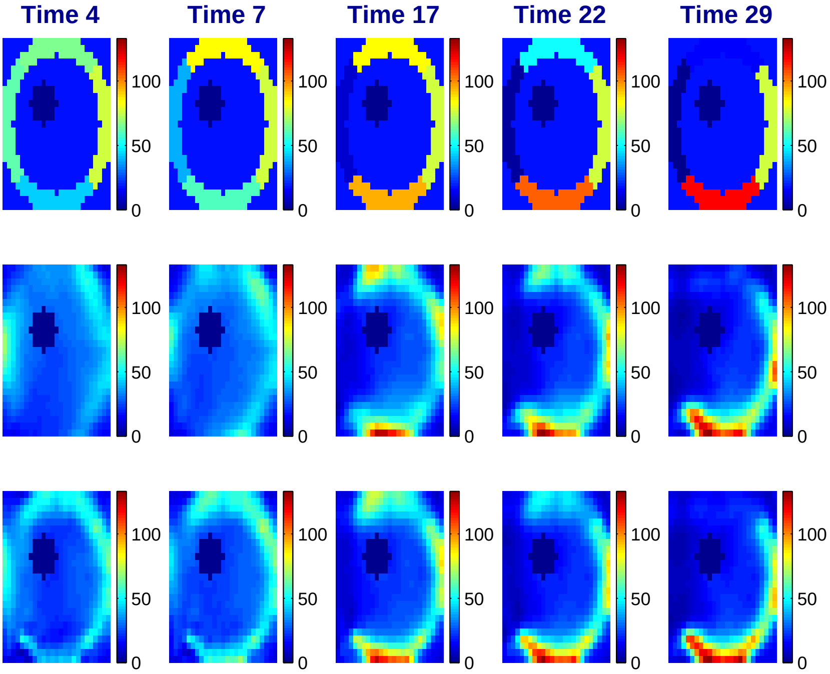

Figure 9: Nonnegative images at various times: Truth in 1^{st} row, filtered image in 2^{nd} row, smoothed image in 3^{rd} row.

9.5. Tikhonov Regularization

In addition to incorporating the nonnegativity, another approach to remedy to our ill-posed inverse problem is the use of spatial regularization. We first experiment with Tikhonov regularization developed in Section 4. The operator L is an appropriately chosen regularization operator, $\hat{x}$ is the output activity of Kalman algorithm. We apply procedure 9.1 with algorithm 4.1 in step 4. We tried $L = I$; that is we preferred a solution with smaller norm. We also chose L to be the second order differential operator that we note as Diff2 where the neighboring system is shown in figure 1. We did not notice any significant change or improvement from one setting to another. Images presented here are from the setting $L =$ Diff2. We observe that there are some pixels' grouping if we compare the reconstructed image to the one done with enforcing the non-negativity constraint only; compare for instance the lower and upper regions in figure 9 and figure 11. As it is known in the regularization literature, Tikhonov tends to over-smooth. We observe the same effect here where we see that images

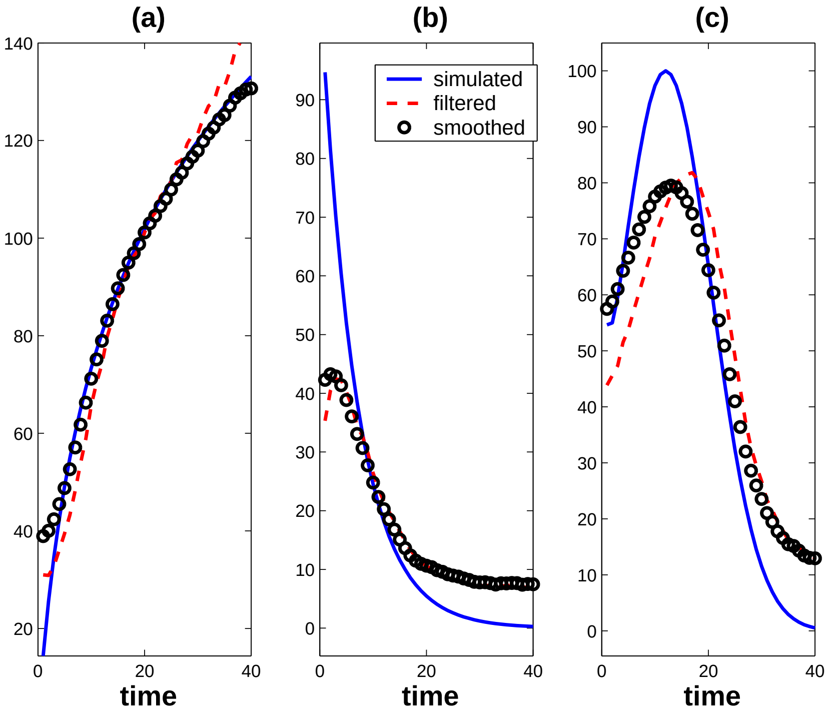

Figure 10: Averaged TACs for each region with positivity: Blue TACs for truth, red TACs for reconstructed filtered, and black TACs for smoothed. (a), (b), and (c) TACs for lower, left, and upper arcs respectively.

look blurred.

9.6. Median

To avoid the over smoothing of Tikhonov type regularization, we introduced in Section 5 an edge preserving regularization as an alternative. We apply procedure 9.1 with algorithm 5.1 in step 4. Figure 12 exhibits the reconstructed filtered and smoothed images together with the simulated/true ones at different instances of time when we applied the median regularization. Notice the blocky segments and edge-preserving at the borders of the regions. As in the previous approaches, reconstructed smoothed images are "smoother" and somehow "better" than the filtered ones. The median approach groups pixels together within a certain ROI. Furthermore, the too much smoothing effect of Tikhonov

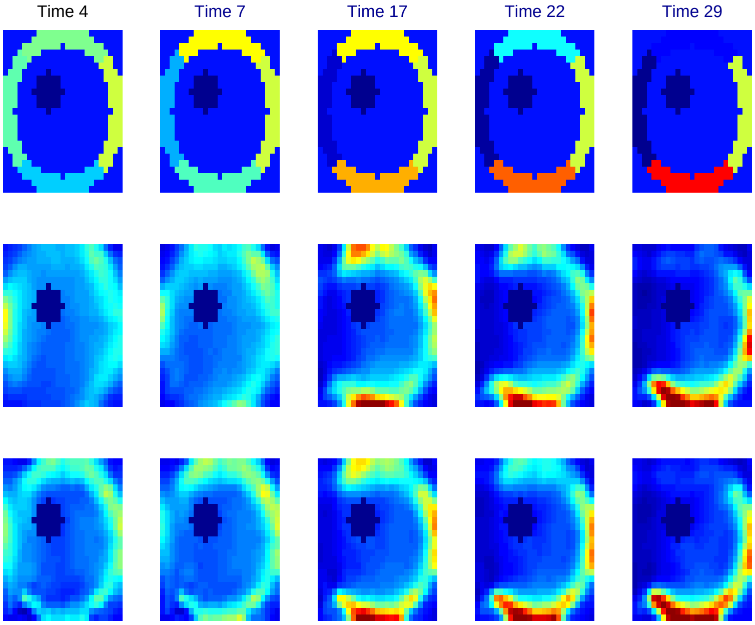

Figure 11: Tikhonov regularized images at various times: Truth in 1^{st} row, filtered image in 2^{nd} row, smoothed image in 3^{rd} row.

regularization, for instance in the middle of the images and at the borders of the arcs, has diminished with the median regularization.

9.7. Spatial Regularization via Segmentation

Finally we include spatial regularization via segmentation. When we know how to partition the digital phantom into nonintersecting regions, we can include this additional constraint into the problem and reduce its size at the same time. Instead of a transition and an evolution models for the activity x, we have rather similar ones for the activity ξ. Hence equations (43) and (44) write

$$\xi_k = \widetilde{A}_k \xi_{k-1} + \tilde{\mu}_k \tag{53}$$

$$y_k = \widetilde{C}_k \xi_k + \tilde{\nu}_k \tag{54}$$

where $\widetilde{C}_k = C_k E$. We offer here an experiment done when we have perfect information about the six ROIs. The size of the image we aim to reconstruct is $4096 = 64 \times 64$ dixels. We utilize procedure 9.1 where the variable is now ξ instead of x as we mentioned in Section 7. Procedure 9.1 employs algorithm 7.1

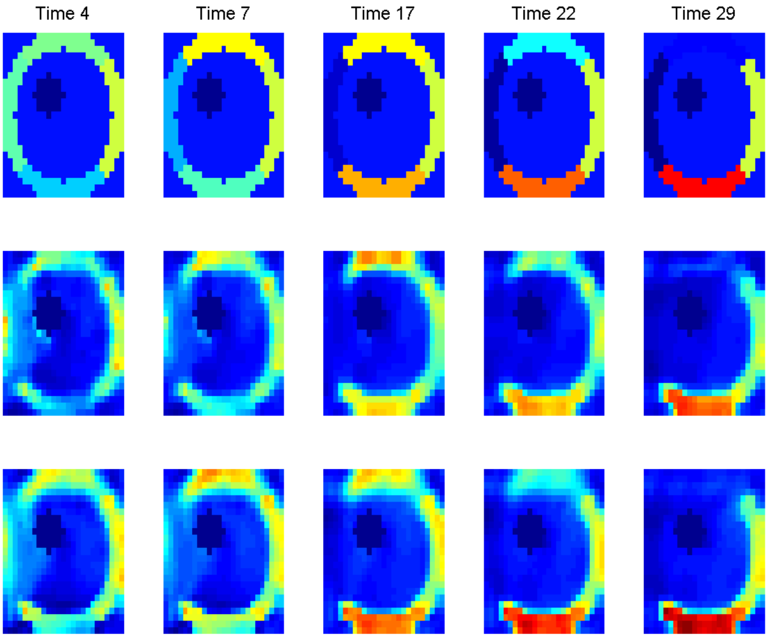

Figure 12: Median regularized images at different times: Truth in 1^{st} row, filtered image in 2^{nd} row, smoothed image in 3^{rd} Row.

in step 4 and 6 while the matrix $C_k E$ substitutes the matrix C_k in steps 4 and 6. The computation takes 1.72 seconds, 2% of the one of 25 × 25 phantom. Figure 13 shows the reconstructed filtered and smoothed images together with the simulated one while figure 14 exhibits the reconstructed TACs being very close to the true time activity curve. We get a much better reconstruction since we have more data. This sustains the asymptotic consistency property, as the observations number increases, of the estimator $x^\star$ given in corollary 8.4. Furthermore, the estimator seems to confirm that it is sufficient, acceptable, unbiased, and efficient thus verifying the criteria stipulated in Section 8.

9.8. Summary

We have presented in this Section 9 validations of our four algorithms 3.1, 4.1, 5.1, and 7.1 to incorporate nonnegativity and more spatial regularization

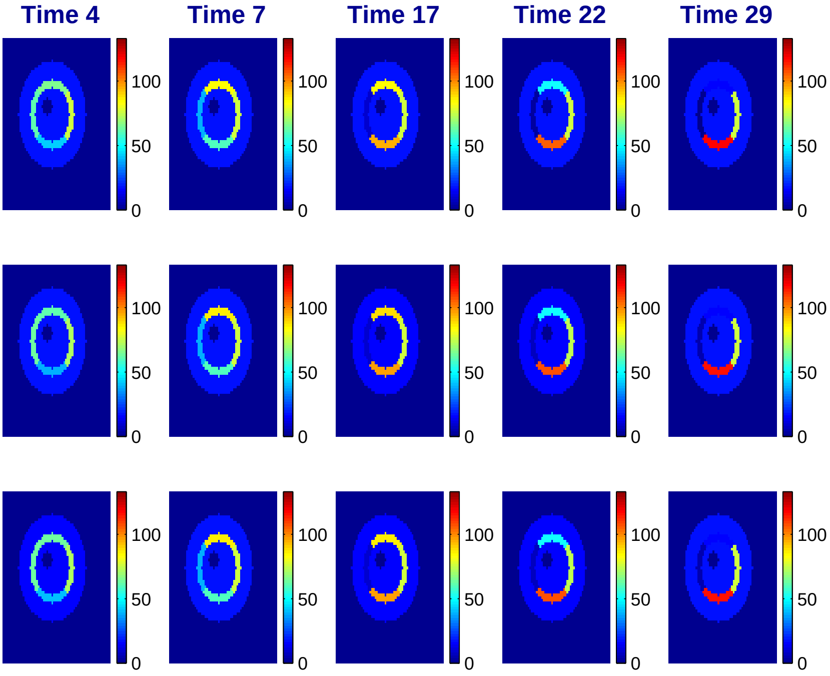

Figure 13: Regularization via segmentation images at different times: Truth in 1^{st} row, filtered image in 2^{nd} row, smoothed image in 3^{rd} Row.

constraints to solve the reconstruction problem of a nuclear medicine modality. Initial solution $\hat{x}$ was found using the Kalman algorithm. However this solution is meaningless because it fails to be nonnegative. We remedy this shortcoming by projecting this solution $\hat{x}$ into the positive octant via a proximal approach obtaining $x^{\star}$; which by the same token helps to minimize more the effect of the ill-posedness of the problem. Analysis of images and TACs showed a net improvement compare to without this added proximal projection.

Kalman in its nature takes care of the temporal regularization. However, we do not have a spatial smoothness especially among pixels within the same region. We used then three more spatial regularization approaches, Tikhonov and Median developed in Sections 4 and 5 and regularization via segmentation covered in Section 7. Numerical results confirm the effectiveness of our methods, specially with the "Median" approach that preserves the edges, while keeping the temporal smoothness feature. Much better performance is even more pronounced with regularization via segmentation. The numerical outperforming of incorporating the nonnegativity constraint into the initial KF solution $\hat{x}$ to

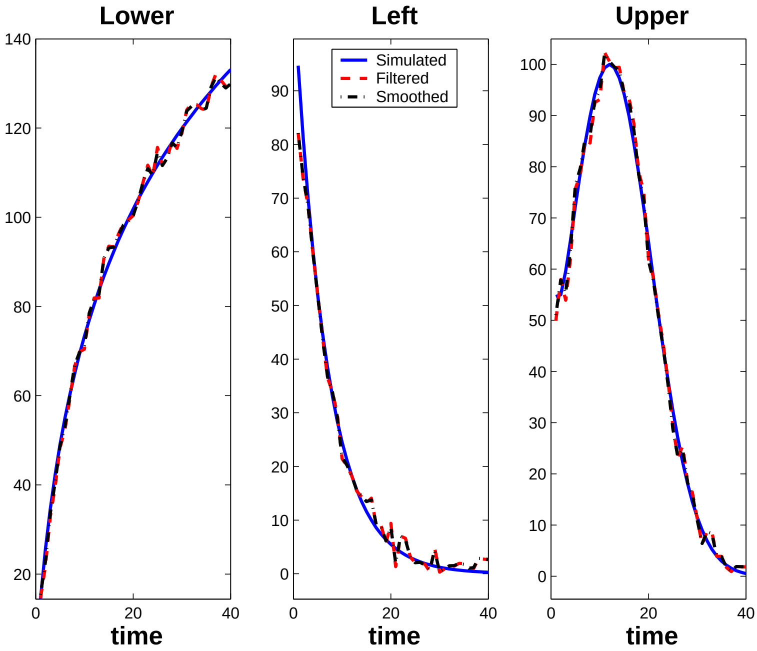

Figure 14: TACs of three regions: Blue TAC for truth, red TAC for reconstructed filtered, and black TAC for smoothed.

get $x^\star$ compare to without is not a "coincidence" but rather was explained theoretically. The proximal projection $x^\star$ possesses indeed nice general properties that were stated and proven in Section 8 and confirmed here.

As a side note about Kalman filter is that KF has been known to be very time consuming for large scale systems as it is the case here in medical imaging; data is in the thousands and even in the millions. Remedies have been used such as ensemble Kalman; which is a a Monte Carlo implementation of the Bayesian update problem. Two new alternatives to KF as yet, SMART Filter and EM Filter, have been reported here [42] and there [43] respectively; where cpu times have been reduced from hours to just minutes.

10. Conclusion

In this paper we have shown, based on proximal projection techniques, how to enforce nonnegativity constraint and three more spatial regularization based on both norms, the 2-norm and the 1-norm, and segmentation. When we are seeking a positive estimator of an unknown vector $x \in \mathbb{R}_+^N$ and we have rather a not necessarily positive estimator $\hat{x} \in \mathbb{R}^N$, we obtain a better estimator $x^\star \in \mathbb{R}_+^N$ as an oblique proximal projection of $\hat{x}$ with respect to a weighting matrix. We proposed four algorithms of imposing nonnegativity and more spatial regularization into $\hat{x}$. We have also stated and proved properties of the estimator $x^\star$ that it inherits from its parent $\hat{x}$ in terms of ML, consistency, and unbiasedness. In addition we proved that $x^\star$ performs better than $\hat{x} \in \mathbb{R}^N$; that is it is optimal in $\mathbb{R}_+^N$ in the mean square error sense. We validate our algorithms with an application to an inverse problem of reconstructing a medical image within the framework of SPECT, a modality in nuclear medicine. Numerical results confirm that our proximal projection approaches not only render the original solution meaningful but also enhance more its spatial appearance as an image, specially with the Median and segmentation regularization.

References

[1] R. Bro and S. Jong, A fast non-negativity-constrained least squares algorithm, *Jour. of Chemo.*, **11** (1997), 393-401.

[2] C.L. Lawson and R.J. Hanson, Solving Least Squares Problems, *SIAM*, Philadelphia, PA (1995).

[3] M.H. Van Benthem and M.R. Keenan, Fast algorithm for the solution of large-scale non-negativity-constrained least squares problems, *J.O.C., J. Chemometrics*, **18** (2004), 441-450.

[4] P.L. Combettes and V.R. Wajs, Signal Recovery by Proximal Forward-Backward Splitting, *Multiscale Model. Simul.*, **4** (2005), 1168-1200.

[5] M. Bertero and P. Boccacci, Introduction to Inverse Problems in Imaging, *IOP Publishing*, Bristol (1998).

[6] J. Qranfal, G. Tanoh, Regularized Kalman filtering for dynamic SPECT, In: *J. Phys.: Conf. Ser.*, **124** (2008).

[7] J. Qranfal, *Optimal Recursive Estimation Techniques for Dynamic Medical Image Reconstruction*, Simon Fraser University (2009).

[8] A.N. Tikhonov, On the stability of inverse problems, *Dokl. Akad. Nauk SSSR*, **39** (1943), 195-198.

[9] A.N. Tikhonov and V.Y. Arsenin, Solutions of Ill-Posed Problems, *Winston, New York* (1977).

[10] D.L. Phillips, A technique for the Numerical Solution of Certain Integral Equations of the First Kind, *J. Assoc. Comput. Mach.*, **9** (1962), 84-97.

[11] S. Alenius and U. Ruotsalainen, Bayesian Image Reconstruction for Emission Tomography Based Median Root Prior, *E.J. Nucl. Med.*, **24** (1997), 258-265.

[12] S. Alenius, U. Ruotsalainen, and J. Astola, Using Local Median as the Location of the Prior Distribution in Iterative Emission Tomography Image Reconstruction, *IEEE Trans. Nucl. Sci.*, **45** (1998), 3097-3107.

[13] I.T. Hsiao, A. Rangarajan, and G. Gindi, A new convex edge-preserving median prior with applications to tomography, *IEEE Trans. Med. Imaging*, **22** (2008), 580-585.

[14] S. Geman and D. Geman, Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6** (1984), 721-741.

[15] S. Geman and D. McClure, Statistical methods for Tomographic Image Reconstruction, *Bulletin of the International Statistical Institute*, **4** (1987), 5-21.

[16] T. Hebert and R. Leahy, A generalized EM Algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors, *IEEE Transactions on Medical Imaging*, **8** (1989).

[17] H. Cramér, Mathematical Methods of Statistics, *University Press, Princeton, N.J.* (1946).

[18] J.A. Hanley, L. Joseph, R.W. Platt, M.K. Chung, and P. Bélisle, Visualizing the median as the minimum-deviation location, *Amer. Statist.*, **55** (2001), 150-152.

[19] N.C. Schwertman, A.J. Gilks, and J. Cameron, A simple noncalculus proof that the median minimizes the sum of the absolute deviations, *Amer. Statist.*, **44** (1990), 38-39.

[20] P.G. Green, Bayesian reconstruction from emission tomography data using a modified EM algorithm, *IEEE Trans. Med. Imaging*, **9** (1990), 84-93.

[21] S. Geman and D. McClure, Bayesian Image Analysis: An Application to Single Photon Emission Tomography, *Statist. Comput. Sect., Amer. Statist. Assoc.* (1985), 12-18.

[22] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, Deterministic edge-preserving regularization in computed imaging, *IEEE Trans. on Image Processing*, **5** (1997), 298-311.

[23] H.H. Bauschke, P.L. Combettes, and D. Noll, Joint minimization with alternating Bregman proximity operators, *Pacific Journal of Optimization*, **2** (2006), 401-424.

[24] D. Hochbaum, J. Qranfal, and G. Tanoh, Experimental Analysis of the MRF Algorithm for Segmentation of Noisy Medical Images, *Algorithmic Operations Research*, **6** (2011), 79-90.

[25] B.W. Reutter, G.T. Gullberg, R.H. Huesman, Direct least Squares Estimation of Spatiotemporal Distribution from Dynamic SPECT Projections using Spatial Segmentation and Temporal B-splines, *IEEE trans. med. imaging*, **19** (2000), 434-450.

[26] R. Carson, A Maximum Likelihood Method for Region-of-Interest Evaluation In Emission Tomography, *J. Comput. Assit. Tomogr.*, **10** (1986), 654-663.

[27] H.W. Sorenson, Parameter Estimation, Principles and Problems, *Marcel Dekker Inc.*, New York, NY (1980).

[28] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*, Wiley-Interscience (2006).

[29] B.D. O. Anderson, J.B. Moore, *Optimal Filtering*, Printice-Hall, Englewood, Ciffs, NJ (1979).

[30] http://cialab.ee.washington.edu/REPRINTS/1997-AlternatingProjectio ns.pdf (Last visited: May 28, 2011).

[31] A.V. Bos, *Parameter Estimation for Scientists and Engineers*, Cambridge University Press, Hoboken, NJ (2007).

[32] C.H. Franklin, *Properties of ML Estimators*, http://users.polisci.wisc.edu/franklin/Content/MLE/Lecs/MLELec04p 4up.pdf (2005).

[33] R.E. Kalman, A new approach to linear filtering and prediction problems, *Trans. of the ASME-Jour. of Basic Eng.*, **82** (1960).

[34] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge, New York, NY (2004).

[35] D. Simon and T. Chia, Kalman Filtering with State Equality Constraints, *IEEE Trans. Aero. and Electr. Systs.*, **39** (2002), 128-136.

[36] D. Simon and D.L. Simon, Kalman Filtering with Inequality Constraints for Turbofan Engine Health Estimation, *IEE Proceedings*, **153** (2006), 371-378.

[37] H.W. Kuhn and A.W. Tucker, *Nonlinear programming*, http://projecteuclid.org/DPubS/Repository/1.0/Disseminate?view=body&id=pdf_1&handle=euclid.bsmsp/1200500249 (1951), Last visited: May 28, 2011.

[38] W. Karush, *Minima of Functions of Several Variables with Inequalities as Side Constraints*, University of Chicago (1939).

[39] T.H. Kjeldsen, A contextualized historical analysis of the Kuhn-Tucker theorem in nonlinear programming: the impact of World War II, *Historia Math*, **4** (2000), 331-361.

[40] H. Stark, *Image Recovery: Theory and Applications*, Academic, New York, NY (1987).

[41] G. Tanoh, *Algorithmes du point intérieur pour l'optimisation en tomographie dynamique et en mécanique du contact*, Université Paul Sabatier (2004).

[42] J. Qranfal and C. Byrne, SMART Filter and its Application to Dynamic SPECT Reconstruction, *International Journal of Pure and Applied Mathematics*, **73** (2011), 405-434.

[43] J. Qranfal and C. Byrne, EM Filter for Time-Varying SPECT Reconstruction, *International Journal of Pure and Applied Mathematics*, **73** (2011), 379-403.