T E C H N I C A L
R E P O R T

0357

INVERSE IMAGING WITH MIXED PENALTIES

DE MOL, C. and M. DEFRISE

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

INVERSE IMAGING WITH MIXED PENALTIES

C. De Mol
Dept Math. Université Libre de Bruxelles, Campus Plaine CPI 217, 1050 Brussels, Belgium.
M. Defrise
Dept Nuclear Medicine, AZ, Vrije Universiteit Brussel, Laarbeeklaan 101, 1090 Brussels, Belgium.

Abstract: This paper proposes new iterative algorithms for solving linear inverse problems when the solution can be written as the sum of a smooth part and of a part which is sparse in pixel space or in terms of the coefficients of its expansion on an arbitrary orthonormal basis.

SPARSITY VERSUS SMOOTHNESS CONSTRAINTS

Many linear or linearized inverse imaging or scattering problems can be cast in the following form: solve for the array $\mathbf{f}$ (which contains M values representing the unknown *object*, e.g. pixel values or characteristics of the probed sample) the following linear equation

$$\mathbf{Af} = \mathbf{g} \tag{1}$$

where $\mathbf{g}$ is the *image* or measurement vector containing N data values and $\mathbf{A}$ is the $N \times M$ matrix modeling the imaging process ($\mathbf{A}$ is assumed to be known). For simplicity, we have used a single index for labeling the object and data arrays, but the formulation obviously applies to 2D or 3D imaging. The usual approach to deal with noisy data is to minimize the least-squares discrepancy (data misfit) or, in the case of ill-conditioned matrices (typical for inverse problems), to solve the following penalized least-squares problem

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \Phi(\mathbf{f}) \quad \text{with} \quad \Phi(\mathbf{f}) = \|\mathbf{Af} - \mathbf{g}\|^2 + \mu\|\mathbf{f}\|^2 \tag{2}$$

where $\|\mathbf{f}\|^2 = \sum_{m=1}^{M} |f_m|^2$ denotes the squared l^2-norm of $\mathbf{f}$ and μ is a small positive regularization parameter controlling the balance between stability and fidelity to the data. The corresponding minimizer $\mathbf{f}^* = (\mathbf{A}^*\mathbf{A} + \mu\mathbf{I})^{-1}\mathbf{A}^*\mathbf{g}$ is usually referred to as the Tikhonov regularized solution of (1) ($\mathbf{I}$ is the identity matrix and $\mathbf{A}^*$ the adjoint of $\mathbf{A}$). An alternative to matrix inversion is provided by iterative schemes, such as the so-called *damped Landweber iteration*

$$\mathbf{f}^{(0)}\text{arbitrary} \; ; \qquad \mathbf{f}^{(k+1)} = \mathbf{T}\,\mathbf{f}^{(k)} \quad \text{for} \quad k = 0, 1, \ldots \tag{3}$$

where the iteration mapping $\mathbf{T}$ is given by $\mathbf{T} = (1+\mu)^{-1}\mathbf{L}$ with $\mathbf{Lf} \equiv \mathbf{f} + \mathbf{A}^*(\mathbf{g} - \mathbf{Af})$. Let us assume that the imaging matrix is renormalized so that $\|\mathbf{A}\| < 1$. Then $\forall \mathbf{f}, \mathbf{h}$, $\|\mathbf{Lf} - \mathbf{Lh}\| \le \|\mathbf{f} - \mathbf{h}\|$ ($\mathbf{L}$ is non-expansive); hence, for strictly positive μ, the mapping $\mathbf{T}$ is a contraction. This ensures the convergence of the iteration (3) to the unique fixed point of $\mathbf{T}$, which is the unique minimizer of (2).

Linear estimates of this sort, however, may not be optimal whenever the object to be restored is known a priori to be sparse, i.e. to have many zero entries. Indeed, even if the original object is sparse, the Tikhonov solution restored from a noisy image will not in general be so. Therefore, it has been advocated [1,2] that the l^2-penalty in (2) could be advantageously replaced by a penalty on the l^1-norm of $\mathbf{f}$: $|||\mathbf{f}||| = \sum_{m=1}^{M} |f_m|$. This modification increases the penalty on components $|f_m| < 1$ and simultaneously decreases the penalty on larger components, thus favouring the restoration of objects with few but large components (as we shall see, the components below some threshold value are even set to zero, a fact which promotes sparsity in the reconstructed object). This strategy leads to the following penalized least-squares problem

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \Phi(\mathbf{f}) \quad \text{with} \quad \Phi(\mathbf{f}) = \|\mathbf{Af} - \mathbf{g}\|^2 + 2\tau|||\mathbf{f}||| \qquad (\tau > 0) \; . \tag{4}$$

Notice that as for (2) this functional is convex. For $\mathbf{A} = \mathbf{I}$ (and $N = M$), the minimizer $\mathbf{f}^*$ is easily seen to be equal to the *soft-thresholded* data vector

$$(\mathbf{S}_\tau \mathbf{g})_n = \begin{cases} g_n - \tau \,\text{sign}(g_n) & \text{if} \quad |g_n| \ge \tau \\ 0 & \text{if} \quad |g_n| < \tau \; . \end{cases} \tag{5}$$

(Note that, when implemented on wavelet coefficients, (5) is a simple *denoising* scheme as proposed in [3]). When $\mathbf{A} \neq \mathbf{I}$, the operator couples all object components and therefore problem (4) becomes a complicated quadratic programming optimization problem. As an alternative, we have proposed in [4] to use an optimization transfer method [5] and to define the *surrogate* functional $\Phi^{\text{SUR}}(\mathbf{f};\mathbf{a}) = \|\mathbf{Af}-\mathbf{g}\|^2 - \|\mathbf{Af}-\mathbf{Aa}\|^2 + \|\mathbf{f}-\mathbf{a}\|^2 + 2\tau|\!|\!|\mathbf{f}|\!|\!|$ (since $\|\mathbf{A}\| < 1$, this functional is strictly convex for any choice of $\mathbf{a}$, and we have $\Phi^{\text{SUR}}(\mathbf{f};\mathbf{a}) \geq \Phi(\mathbf{f})$ while $\Phi^{\text{SUR}}(\mathbf{f};\mathbf{f}) = \Phi(\mathbf{f})$), the minimizer of (4) being approached through the following iterative scheme:

$$\mathbf{f}^{(0)}\text{arbitrary ;} \qquad \mathbf{f}^{(k+1)} = \arg\min_{\mathbf{f}} \Phi^{\text{SUR}}(\mathbf{f};\mathbf{f}^{(k)}) \quad \text{for} \quad k=0,1,\ldots \tag{6}$$

At each iteration, the minimization problem is decoupled for each pixel value and can be solved explicitly, yielding an iterative algorithm similar to (3) with $\mathbf{T}$ given now by $\mathbf{T} = \mathbf{S}_\tau\mathbf{L}$, namely a Landweber scheme with a soft-thresholding performed at each iteration. When the null-space of $\mathbf{A}$ is reduced to zero (i.e. when $\mathbf{Af} = 0$ implies $\mathbf{f} = 0$), the convergence of this scheme can be easily established since $\mathbf{L}$ is then a contraction whereas the nonlinear thresholding operator $\mathbf{S}_\tau$ is easily seen to be non-expansive. The product of a contractive by a nonexpansive mapping being contractive, $\mathbf{T}$ is a contraction. This ensures convergence of the iteration to the fixed point of $\mathbf{T}$, which can be shown to coincide with the minimizer of (4) – unique in this case. The convergence of this iterative scheme to a solution of the minimization problem (4) has been proved in [6] to hold under much more general assumptions. Even when $\mathbf{A}$ has a non-zero null-space, strong convergence can be established in an infinite-dimensional setting where $\|g\|$ is the Hilbert L^2-norm of g and $|\!|\!|\mathbf{f}|\!|\!|$ is a l^p-norm, with $1 \leq p \leq 2$, on the sequence of coefficients of $\mathbf{f}$ on any orthogonal basis in L^2, such as e.g. a wavelet basis or a Fourier basis. Moreover, it has been shown in [6] that problem (4) yields a properly regularized solution which coincides in the limit case $p = 2$ with Tikhonov's solution (2) (with $\mu = 2\tau$).

A MIXED-PENALTY APPROACH

In the present paper, we extend the previous framework to cope with the case where the object is known a priori to be the sum $\mathbf{f} = \mathbf{u} + \mathbf{v}$ of a sparse part $\mathbf{u}$ and of a smooth part or diffuse background $\mathbf{v}$, a situation met in several applications e.g. in astronomy, spectroscopy, medical imaging, etc. Hence we consider the mixed-penalty (convex) functional

$$\Phi(\mathbf{u},\mathbf{v}) = \|\mathbf{A}(\mathbf{u}+\mathbf{v})-\mathbf{g}\|^2 + 2\tau|\!|\!|\mathbf{u}|\!|\!| + \mu\|\mathbf{v}\|^2 \tag{7}$$

and the following surrogate

$$\Phi^{\text{SUR}}(\mathbf{u},\mathbf{v};\mathbf{a}) = \Phi(\mathbf{u},\mathbf{v}) + \|\mathbf{u}+\mathbf{v}-\mathbf{a}\|^2 - \|\mathbf{A}(\mathbf{u}+\mathbf{v})-\mathbf{Aa}\|^2 \ . \tag{8}$$

We minimize (7) iteratively as follows ($k = 0,1,\ldots$ and $\mathbf{f}^{(k)} = \mathbf{u}^{(k)} + \mathbf{v}^{(k)}$) :

$$\mathbf{u}^{(0)},\mathbf{v}^{(0)}\text{arbitrary; } \mathbf{u}^{(k+1)} = \arg\min_{\mathbf{u}} \Phi^{\text{SUR}}(\mathbf{u},\mathbf{v}^{(k)};\mathbf{f}^{(k)}); \ \mathbf{v}^{(k+1)} = \arg\min_{\mathbf{v}} \Phi^{\text{SUR}}(\mathbf{u}^{(k)},\mathbf{v};\mathbf{f}^{(k)}) \tag{9}$$

getting explicit expressions for the minimizers

$$\mathbf{u}^{(k+1)} = \mathbf{S}_\tau\left(\mathbf{u}^{(k)} + \mathbf{A}^*(\mathbf{g}-\mathbf{Af}^{(k)})\right); \quad \mathbf{v}^{(k+1)} = (1+\mu)^{-1}\left(\mathbf{v}^{(k)} + \mathbf{A}^*(\mathbf{g}-\mathbf{Af}^{(k)})\right) \ . \tag{10}$$

The convergence of this scheme to a minimizer of (7) can be established using an extension of the proofs in [6] (paper in preparation). Let us note that a potentially faster converging variant of this algorithm is obtained by replacing the update for $\mathbf{v}$ by $\mathbf{v}^{(k+1)} = (1+\mu)^{-1}\left(\mathbf{v}^{(k)} + \mathbf{A}^*(\mathbf{g}-\mathbf{Au}^{(k+1)}-\mathbf{Av}^{(k)})\right)$.

A more substantial variation on the same theme is derived by viewing the functionals (7) and (8) as depending on the unknowns $\mathbf{u}$ and $\mathbf{f}$ instead of $\mathbf{u}$ and $\mathbf{v}$, namely by considering

$$\Phi(\mathbf{u},\mathbf{f}) = \|\mathbf{Af}-\mathbf{g}\|^2 + 2\tau|\!|\!|\mathbf{u}|\!|\!| + \mu\|\mathbf{f}-\mathbf{u}\|^2 \tag{11}$$

with the surrogate

$$\Phi^{\text{SUR}}(\mathbf{u},\mathbf{f};\mathbf{a}) = \Phi(\mathbf{u},\mathbf{f}) + \|\mathbf{f}-\mathbf{a}\|^2 - \|\mathbf{Af}-\mathbf{Aa}\|^2 \ . \tag{12}$$

The iterative minimization of (11) can then be done as follows ($k = 0,1,\ldots$) :

$$\mathbf{f}^{(0)}\text{arbitrary;} \quad \mathbf{u}^{(k)} = \arg\min_{\mathbf{u}} \Phi^{\text{SUR}}(\mathbf{u},\mathbf{f}^{(k)};\mathbf{f}^{(k)}); \quad \mathbf{f}^{(k+1)} = \arg\min_{\mathbf{f}} \Phi^{\text{SUR}}(\mathbf{u}^{(k)},\mathbf{f};\mathbf{f}^{(k)}) \tag{13}$$

with the following explicit expressions for the minimizers

$$\mathbf{u}^{(k)} = \mathbf{S}_{\tau/\mu}\left(\mathbf{f}^{(k)}\right); \quad \mathbf{f}^{(k+1)} = \mathbf{u}^{(k)} + (1+\mu)^{-1}\left(\mathbf{f}^{(k)} + \mathbf{A}^*(\mathbf{g}-\mathbf{Af}^{(k)}) - \mathbf{u}^{(k)}\right) \ . \tag{14}$$

Noticing that the above algorithm can be rewritten in terms of $\mathbf{f}$ alone as

$$\mathbf{f}^{(k+1)} = (1+\mu)^{-1}\left(\mathbf{f}^{(k)} + \mathbf{A}^*(\mathbf{g} - \mathbf{A}\mathbf{f}^{(k)}) - \mu\ \mathbf{S}_{\tau/\mu}(\mathbf{f}^{(k)})\right) \equiv \mathbf{T}\ \mathbf{f}^{(k)} \tag{15}$$

we can give a simple proof of convergence for the scheme (13-14) by showing that $\mathbf{T}$ is a contraction in the case where $\mathbf{A}$ has a zero null-space. Indeed, we have for arbitrary $\mathbf{f}$ and $\mathbf{h}$

$$\|\mathbf{T}\ \mathbf{f} - \mathbf{T}\ \mathbf{h}\| \le \frac{1}{1+\mu}\ \|(\mathbf{I} - \mathbf{A}^*\mathbf{A})(\mathbf{f} - \mathbf{h})\| + \frac{1}{1+\mu}\ \|\mu\ \mathbf{S}_{\tau/\mu}(\mathbf{f}) - \mu\ \mathbf{S}_{\tau/\mu}(\mathbf{h})\| \le \rho\ \|\mathbf{f} - \mathbf{h}\| \tag{16}$$

with $\rho < 1$ since $\|\mathbf{I} - \mathbf{A}^*\mathbf{A}\| < 1$ and $\mathbf{S}_{\tau/\mu}$ is non-expansive.

An interesting insight into the previous algorithm is gained by remarking that we can minimize the functional (11) for $\mathbf{u}$, given $\mathbf{f}$, by solving a simple denoising problem, getting as solution $\mathbf{u} = \mathbf{S}_{\tau/\mu}(\mathbf{f})$, and by reinserting this solution into the functional. This amounts to solving

$$\mathbf{f}^* = \arg\min_{\mathbf{f}}\ \tilde{\Phi}(\mathbf{f}) \quad \text{with} \quad \tilde{\Phi}(\mathbf{f}) = \|\mathbf{A}\mathbf{f} - \mathbf{g}\|^2 + \mu \sum_m \Pi(f_m) \tag{17}$$

$$\text{and} \quad \Pi(f_m) = \begin{cases} |f_m|^2 & \text{if} \quad |f_m| \le \tau/\mu \\ (2\tau/\mu)|f_m| - (\tau/\mu)^2 & \text{if} \quad |f_m| > \tau/\mu\ . \end{cases} \tag{18}$$

We recognize in (18) the Huber function used for robust estimation from data containing outliers. The rationale behind this Huber prior is to penalize small values of $\mathbf{f}$, likely due to noise, with a l^2-type penalty and to penalize to a lesser extent, with a l^1-type penalty, larger values of $\mathbf{f}$ which are expected to reflect some true structure. The minimization of (17) can again be done iteratively, using the same surrogate as above for the data misfit.

CONCLUDING REMARKS

We have devised several iterative algorithms for the restoration of objects which are a sum of a sparse and a smooth part. Although we used for simplicity a finite-dimensional setting, the schemes can easily be generalized to an infinite-dimensional framework, similar to the one used in [6]. Also, the l^2-norm penalty $\|v\|^2$ used for the smooth part can be replaced by other quadratic norms, e.g. of the type $\|Wv\|^2$, where W is a linear operator with bounded inverse, involving a few derivatives. Moreover, as said, the l^1-penalty $|\!|\!|u|\!|\!|$ can be replaced by a l^1- or l^p-norm (with $1 < p \le 2$) on the coefficients of the expansion of u on a given arbitrary orthonormal basis in L^2. This can be exploited to enforce sparsity in the wavelet or in the Fourier domain rather than in pixel space, the thresholding being applied on the wavelet or Fourier coefficients of each iterate. Moreover, if positive reconstructions are needed, positivity can be enforced at each iteration, and since the projector on the cone of positive objects is a non-expansive operator, this does not impair the non-expansivity or contractivity of the iteration mapping. Extensions of the present algorithms to deal with nonlinear imaging operators $\mathbf{A}$ may also be desirable, although convergence results are probably hard to get in such a case.

REFERENCES

[1] Tibshirani, R., "Regression shrinkage and selection via the lasso," *J. Royal Statist. Soc. B*, **vol. 58**, 1996, pp. 267-288.

[2] Chen, S., Donoho, D., and M. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM Review*, **vol. 43**, 2001, 129–159.

[3] Donoho, D., and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, **vol. 81**, 1994, pp. 425-455.

[4] De Mol, C., and M. Defrise, "A note on wavelet-based inversion methods," in: *"Inverse Problems, Image Analysis and Medical Imaging,"* M. Z. Nashed and O. Scherzer eds, Series "Contemporary Mathematics" **vol. 313**, pp. 85–96, American Mathematical Society, 2002.

[5] Lange, K., Hunter, D.R., and I. Yang, "Optimization transfer algorithms using surrogate objective functions", *J. Comp. Graph. Stat.*, **vol. 9**, 2000, pp. 1-59.

[6] Daubechies, I., Defrise, M., and C. De Mol, "Least-Squares Problems with Sparsity Constraints," Preprint, June 2003.