A FORWARD-BACKWARD ALGORITHM FOR IMAGE RESTORATION WITH SPARSE REPRESENTATIONS

Caroline Chaux,[1] *Patrick L. Combettes,*[2] *Jean-Christophe Pesquet,*[1] *and Valérie R. Wajs*[2]

[1]Institut Gaspard Monge and UMR–CNRS 8049, Université de Marne la Vallée, 77454 Marne la Vallée Cedex 2, France
[2]Laboratoire Jacques-Louis Lions, UMR-CNRS 7598, Université Pierre et Marie Curie – Paris 6, 75005 Paris, France

ABSTRACT

A number of recent approaches to image deconvolution and denoising in Hilbert spaces consist of minimizing the sum of a residual energy and of a function promoting a sparse decomposition in an orthonormal basis. Using convex-analytical tools, we provide a systematic analysis of this generic problem. We first present general properties of the problem, and then propose a flexible forward-backward algorithm to solve it. These results bring together and extend various results found in the literature and make it possible to devise new sparsity-promoting restoration schemes. Such schemes are illustrated via numerical simulations.

1. PROBLEM FORMULATION

We consider the standard problem of restoring an image $\overline{x}$ in a real Hilbert space $\mathcal{H}$ from the observation of an image

$$z = T\overline{x} + w, \tag{1}$$

where $T\colon \mathcal{H} \to \mathcal{H}$ is a bounded linear operator and where $w \in \mathcal{H}$ stands for an additive noise perturbation. In many applications, it is important to obtain a compact representation of an image, e.g., for such purposes as coding, storing, or processing. A standard method for obtaining a compact representation of an image $x \in \mathcal{H}$ is to decompose it with respect to a family of vectors $(e_k)_{k\in K\subset\mathbb{N}}$, and to retain those coefficients in $(\langle x \mid e_k\rangle)_{k\in K}$ with large magnitude [2, 5]. This will indeed provide a good approximation to x if the sequence $(|\langle x \mid e_k\rangle|)_{k\in K}$ is sparse. As observed in [4], sparsity can conveniently be induced in a variational setting by penalizing the standard residual term $\|Tx - z\|^2$ by certain separable functions of the magnitude of the coefficients $(\langle x \mid e_k\rangle)_{k\in K}$. This observation motivates the following problem formulation.

Problem 1 Let

(i) $T\colon \mathcal{H} \to \mathcal{H}$ be a nonzero bounded linear operator;

(ii) $(e_k)_{k\in K\subset\mathbb{N}}$ be an orthonormal basis of $\mathcal{H}$;

(iii) $(\phi_k)_{k\in K}$ be lower semicontinuous convex functions from $\mathbb{R}$ to $]-\infty,+\infty]$ such that $(\forall k \in K)\ \phi_k \geq 0$ and $\phi_k(0) = 0$.

The objective is to

$$\underset{x\in\mathcal{H}}{\text{minimize}}\ \ \frac{1}{2}\|Tx - z\|^2 + \sum_{k\in K} \phi_k(\langle x \mid e_k\rangle). \tag{2}$$

The set of solutions to this problem is denoted by G.

The case when $\phi_k \equiv \omega|\cdot|^2$ with $\omega \in]0,+\infty[$ in (2), yields a standard Tykhonov formulation; the case when $\phi_k \equiv \omega|\cdot|$ has been of special interest as a tool to induce sparseness; in [4], the case when, for every $k \in K$, $\phi_k = \omega_k|\cdot|^p$, with $p \in [1,2]$ and $\omega_k \in]0,+\infty[$ is investigated in detail in [4], where an algorithm was also proposed to solve the resulting problem.

The objective of the present paper is to provide a theoretical and numerical investigation of Problem 1. Unlike the above cited works, our investigation will exploit convex-analytical tools, which will allow us to obtain in a synthetic fashion existence, uniqueness, characterization and stability results on Problem 1, but also more general solution methods than existing ones. In turn, these results will allow us to tackle variational formulations that cannot be handled by current methods, and to obtain more efficient algorithms for existing formulations.

In Section 2, we discuss proximity operators, which will play a central role in our discussion. Section 3 is devoted to deriving some basic properties of Problem 1 and presenting a numerical algorithm to solve it, along with convergence results. Numerical applications to sparse image restoration problems are presented in Section 4.

2. PROXIMITY OPERATORS

Throughout, the underlying image space is a real Hilbert space $\mathcal{H}$ with scalar product $\langle\cdot \mid \cdot\rangle$, norm $\|\cdot\|$, and distance d. $\Gamma_0(\mathcal{H})$ is the class of all convex lower semicontinuous functions from $\mathcal{H}$ to $]-\infty,+\infty]$ that are not identically $+\infty$.

The distance from an image $x \in \mathcal{H}$ to a nonempty set $C \subset \mathcal{H}$ is $d_C(x) = \inf \|x - C\|$; if C is closed and convex then, for every $x \in \mathcal{H}$, there exists a unique point $P_C x \in C$, called the projection of x onto C such that $\|x - P_C x\| = d_C(x)$. The conjugate of a function $f \in \Gamma_0(\mathcal{H})$ is the function $f^* \in \Gamma_0(\mathcal{H})$ defined by

$$(\forall u \in \mathcal{H}) \quad f^*(u) = \sup_{x \in \mathcal{H}} \langle x \mid u \rangle - f(x). \tag{3}$$

The Moreau envelope of index $\gamma \in]0, +\infty[$ of a function $f \in \Gamma_0(\mathcal{H})$ is the continuous convex function

$$^{\gamma}f \colon \mathcal{H} \to \mathbb{R} \colon x \mapsto \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma}\|x - y\|^2. \tag{4}$$

For every $x \in \mathcal{H}$, the infimum in (4) is achieved at a unique point denoted by $\operatorname{prox}_{\gamma f} x$. The operator

$$\operatorname{prox}_f \colon \mathcal{H} \to \mathcal{H} \colon x \mapsto \operatorname*{arg\,min}_{y \in \mathcal{H}} f(y) + \frac{1}{2}\|x - y\|^2 \tag{5}$$

is called the proximity operator of f. The reader is referred to [3] for details on these operators and the proofs of the following results. First, we observe that proximity operators generalize the notion of a projection operator.

Example 2 *Let f be the indicator function of a nonempty closed convex set $C \subset \mathcal{H}$: $(\forall x \in \mathcal{H})\ f(x) = 0$ if $x \in C$; $f(x) = +\infty$ if $x \notin C$. Then $^{\gamma}f = d_C^2/2\gamma$ and $\operatorname{prox}_f = P_C$.*

The next result, known as Moreau's decomposition principle, provides a powerful nonlinear image decomposition scheme which extends in particular the standard linear decomposition with respect to two orthogonal vector subspaces. An illustration of this decomposition principle is shown in Fig. 1.

Proposition 3 *Let $f \in \Gamma_0(\mathcal{H})$, $\gamma \in]0, +\infty[$, and $x \in \mathcal{H}$. Then $\|x\|^2 = 2\gamma\big({}^{\gamma}f(x) + {}^{1/\gamma}(f^*)(x/\gamma)\big)$ and*

$$x = x_\gamma^{\oplus} + x_\gamma^{\ominus}, \text{ where } \begin{cases} x_\gamma^{\oplus} = \operatorname{prox}_{\gamma f} x \\ x_\gamma^{\ominus} = \gamma \operatorname{prox}_{f^*/\gamma}(x/\gamma). \end{cases} \tag{6}$$

Moreover, $f(x_\gamma^{\oplus}) + f^*(x_\gamma^{\ominus}/\gamma) = \langle x_\gamma^{\oplus} \mid x_\gamma^{\ominus} \rangle/\gamma$.

Example 4 *Let $(p_k)_{k \in K}$ be a sequence in $[1, +\infty[$, let $(\omega_k)_{k \in K}$ be a sequence in $]0, +\infty[$, let*

$$f \colon \mathcal{H} \to]-\infty, +\infty] \colon x \mapsto \sum_{k \in K} \omega_k |\langle x \mid e_k \rangle|^{p_k}, \tag{7}$$

and let $x \in \mathcal{H}$. Then $\operatorname{prox}_f x = \sum_{k \in K} \pi_k e_k$ where, for every $k \in K$, π_k is the unique solution of

$$(\forall \eta \in \mathbb{R}) \quad (\eta - \pi_k)(\xi_k - \pi_k)/\omega_k + |\pi_k|^{p_k} \leq |\eta|^{p_k}, \tag{8}$$

Fig. 1. Proximal decomposition (6) with $f \colon (\xi_k)_k \mapsto \sum_k |\xi_k|^{3/2}$ and $\gamma = 7$. Top: Original 256×256 SPOT5 image $\overline{x}$; middle: $\overline{x}_\gamma^{\oplus}$; bottom: $\overline{x}_\gamma^{\ominus}$.

where $\xi_k = \langle x \mid e_k \rangle$; in particular, π_k is given by

$$\begin{cases} \operatorname{sgn}(\xi_k)\max\{|\xi_k| - \omega_k, 0\}, & \text{if } p_k = 1; \\ \xi_k + \dfrac{4\omega_k}{3\cdot 2^{1/3}}\Big((\eta_k - \xi_k)^{1/3} - (\eta_k + \xi_k)^{1/3}\Big), & \\ \quad \text{where } \eta_k = \sqrt{\xi_k^2 + 256\omega_k^3/729}, & \text{if } p_k = \dfrac{4}{3}; \\ \xi_k + \dfrac{9\omega_k^2 \operatorname{sgn}(\xi_k)}{8}\left(1 - \sqrt{1 + \dfrac{16|\xi_k|}{9\omega_k^2}}\right), & \text{if } p_k = \dfrac{3}{2}; \\ \xi_k/(1 + 2\omega_k), & \text{if } p_k = 2; \\ \operatorname{sgn}(\xi_k)\dfrac{\sqrt{1 + 12\omega_k|\xi_k|} - 1}{6\omega_k}, & \text{if } p_k = 3. \end{cases}$$

Let us note that the formula for $p_k = 1$ gives the well-known scalar soft-thresholding operation. This type of thresholder, used in particular in [4, 5] and the references therein, is therefore a proximity operator.

3. ANALYSIS AND NUMERICAL SOLUTION

We provide some properties of Problem 1 and propose an algorithm to solve it; details will be found in [3].

Proposition 5

(i) *Problem 1 possesses at least one solution if there exists $\rho \in]0, +\infty[$ and a function $c\colon [0, +\infty[\to [0, +\infty[$ such that $c(0) = 0$, $\lim_{t \to +\infty} c(t) = +\infty$ and, for every $(\xi_k)_{k\in K}$ in $\ell^2(K)$ such that $\sum_{k\in K} |\xi_k|^2 \geq \rho$,*

$$\sum_{k\in K} \phi_k(\xi_k) \geq c\left(\sum_{k\in K} |\xi_k|^2\right). \tag{9}$$

(ii) *Problem 1 possesses at most one solution if the functions $(\phi_k)_{k\in K}$ are strictly convex or if T is injective.*

(iii) *Problem 1 possesses exactly one solution if*
$$(\exists\, \kappa \in]0, +\infty[)(\forall x \in \mathcal{H})\ \ \|Tx\| \geq \kappa\|x\|.$$

(iv) *Let $\gamma \in]0, +\infty[$. Then $x \in \mathcal{H}$ solves Problem 1 if and only if, for every $k \in K$,*
$$\langle x \mid e_k \rangle = \operatorname{prox}_{\gamma\phi_k}\langle x + \gamma T^*(z - Tx) \mid e_k \rangle.$$

Theorem 6 *Suppose that $G \neq \emptyset$. Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0, +\infty[$ such that $0 < \inf_{n\in\mathbb{N}} \gamma_n \leq \sup_{n\in\mathbb{N}} \gamma_n < 2/\|T\|^2$ and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0, 1]$ such that $\inf_{n\in\mathbb{N}} \lambda_n > 0$. Moreover, for every $n \in \mathbb{N}$, let $(\alpha_{n,k})_{k\in K}$ be a square-summable sequence and suppose that $\sum_{n\in\mathbb{N}} \sqrt{\sum_{k\in K} |\alpha_{n,k}|^2} < +\infty$. Fix $x_0 \in \mathcal{H}$ and, for every $n \in \mathbb{N}$, set*

$$x_{n+1} = x_n + \lambda_n\left(\sum_{k\in K} (\pi_{n,k} + \alpha_{n,k})e_k - x_n\right), \tag{10}$$

where $\pi_{n,k} = \operatorname{prox}_{\gamma_n\phi_k}\langle x_n + \gamma_n(T^*(z - Tx_n)) \mid e_k \rangle$. Then:

(i) $\sum_{n\in\mathbb{N}} \big\|T^*T(x_n - x)\big\|^2 < +\infty$.

(ii) $\sum_{n\in\mathbb{N}} \big\|\operatorname{prox}_{\gamma_n f}\big(x_n + \gamma_n T^*(z - Tx_n)\big) - x_n\big\|^2 < +\infty$, *where* $f\colon y \mapsto \sum_{k\in K} \phi_k(\langle y \mid e_k \rangle)$.

(iii) $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point $x \in G$; the convergence is strong if and only if $\underline{\lim}\, d_G(x_n) = 0$.*

In (10), $\alpha_{n,k}$ stands for some tolerance in the computation of $\operatorname{prox}_{\gamma_n\phi_k}\langle x_n + \gamma_n(T^*(z - Tx_n)) \mid e_k \rangle$.

An important instance of the above framework is when f is given by (7), where $(p_k)_{k\in K}$ lies in $[1, 2]$. Indeed, when $(e_k)_{k\in K}$ is a wavelet basis, there exists a strong connection between Problem 2 and maximum *a posteriori* techniques for estimating $\overline{x}$ in the presence of white Gaussian noise. In this context, using suitably subband-adapted values of p_k amounts to fitting an appropriate generalized Gaussian prior distribution to the wavelet coefficients in each subband. Such a statistical modeling is commonly used in wavelet-based estimation, where values of p_k close to 2 provide a good model at coarse resolution levels, whereas values close to 1 should preferably be used at finer resolutions.

Proposition 7 *Suppose that f is given by* (7) *and, if $\dim \mathcal{H} = +\infty$, that $\inf_{k\in K} \omega_k > 0$ and $\sup_{k\in K} p_k \leq 2$. Then*

(i) *Problem 1 possesses at least one solution, and exactly one if $(\forall k \in K)\ p_k > 1$.*

(ii) *Every sequence $(x_n)_{n\in\mathbb{N}}$ constructed by* (10) *converges strongly to a solution to Problem 1 provided that $\inf\big\{p_k \;\big|\; k \in K,\ p_k > 1\big\} > 1$ (see Example 4 for the computation of $(\pi_{n,k})_{k\in K}$).*

Let us note that in the special case when $\lambda_n \equiv 1$, $\|T\| < 1$, $\gamma_n \equiv 1$, $p_k \equiv p \in [1, 2]$, and $\alpha_{n,k} \equiv 0$, Proposition 7(ii) appears in [4, Theorem 3.1]. We shall exploit the added flexibility afforded by our results in the next section.

4. NUMERICAL RESULTS

We apply Proposition 7 to the problem of restoring an $N \times N$ image, where $N = 256$. The underlying Hilbert space $\mathcal{H}$ is the Euclidean space $\mathbb{R}^{N^2}$. In (1), the original image $\overline{x}$ is that shown in Fig. 1 (top), the degraded image z is that shown in Fig. 2 (top), the operator T represents convolution

with a 7×7 uniform blur such that $\|T\| = 1$, and w is a realization of a zero mean white Gaussian noise with known variance. The blurred-image to noise ratio is 30.28 dB.

Fig. 2. Top: degraded image; bottom: restored image.

In Problem 1, $(e_k)_{1 \leq k \leq N^2}$ is a two-dimensional separable orthonormal wavelet basis. More precisely, we use a 4-band wavelet decomposition with the filter bank coefficients defined in [1, Table VI], over 2 resolution levels. Moreover, p_k takes its values in $\{1, 4/3, 3/2, 2\}$ in (7).

To illustrate the convergence of the algorithm, we compute the normalized error $||x_n - x_\infty||/||x_\infty||$. We consider two scenarii: $\gamma_n \equiv 1$ and $\gamma_n \equiv 1.99$. In each scenario, two subcases are considered: first $\omega_k \equiv \omega$ and $p_k \equiv 1$; then, for each subband, an adapted value of (ω_k, p_k) is selected. Fig. 3 shows a faster decrease of the normalized error for $\gamma_n \equiv 1.99$ than for $\gamma_n \equiv 1$, regardless of the strategy for choosing $(\omega_k, p_k)_{1 \leq k \leq N^2}$. This means that the flexibility afforded by Proposition 7 in the choice of these parameters can be exploited to accelerate the convergence of the algorithm. We also observe that the choice of variable values of ω_k and p_k over the subbands leads to slower convergence. However, it brings an improvement of 0.25 dB in terms of signal-to-noise ratio (SNR). The restored image shown in Fig. 2 (bottom) has been obtained with $\gamma_n \equiv 1.99$ and subband-adapted values of $(\omega_k, p_k)_{1 \leq k \leq N^2}$. Its relative error is 14.76 dB, whereas that of the degraded image z is 11.05 dB (the decibel value of the relative error of an image y is $20 \log_{10} (\|\overline{x}\|/\|y - \overline{x}\|)$).

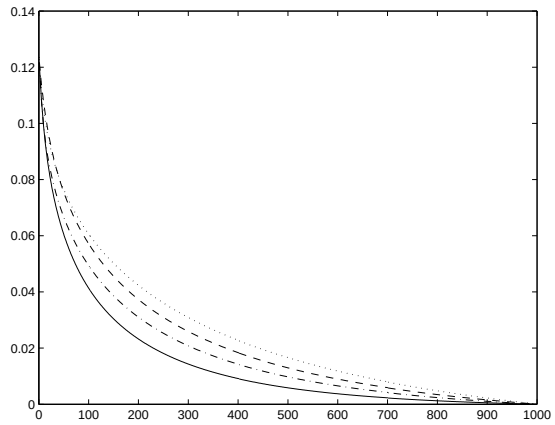

Fig. 3. Normalized error $||x_n - x_\infty||/||x_\infty||$ where (i) $\gamma_n \equiv 1$ and $(\omega_k, p_k) \equiv (\omega, 1)$ (dashed line) or takes on subband-adapted values (dotted line); (ii) $\gamma_n \equiv 1.99$ and $(\omega_k, p_k) \equiv (\omega, 1)$ (solid line) or takes on subband-adapted values (dash-dot line).

5. REFERENCES

[1] O. Alkin and H. Caglar, "Design of efficient M-band coders with linear-phase and perfect-reconstruction properties," *IEEE Trans. Signal Process.*, vol. 43, pp. 1579–1590, 1995.

[2] A. Cohen, *Numerical Analysis of Wavelet Methods.* New York: Elsevier, 2003.

[3] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.,* to appear.

[4] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, vol. 57, pp. 1413–1457, 2004.

[5] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet shrinkage: Asymptopia?," *J. R. Statist. Soc. B.,* vol. 57, pp. 301–369, 1995.