THEORETICAL ANALYSIS OF SOME REGULARIZED IMAGE DENOISING METHODS

Patrick L. Combettes and Valérie R. Wajs

Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie – Paris 6, 75005 Paris, France

ABSTRACT

Regularization techniques have been in use in signal recovery for over four decades. In this paper, we propose a new, synthetic approach to the study of regularization methods in image denoising problems based on Moreau's proximity operators. We exploit the remarkable properties enjoyed by these operators to establish in a systematic fashion a variety of properties of regularized denoising problems and to propose new numerical schemes to solve them.

1. INTRODUCTION

In many image processing applications, the image $\overline{x}$ of interest must be recovered from a noisy measurement

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

where w represents a realization of a random process. In most practical situations, such a problem can be cast in a real Hilbert space $(\mathcal{H}, \|\cdot\|)$. A widespread approach to this type of inverse problems is to seek an estimate x_γ of $\overline{x}$ as a solution to the variational problem

$$\text{Minimize } \; f(y) + \frac{1}{2\gamma}\|x-y\|^2 \; \text{ for } \; y \in \mathcal{H}, \tag{2}$$

where $\gamma \in \left]0,+\infty\right[$ and $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ is a lower semicontinuous convex function. Among the numerous formulations that fit this setting, let us mention the following.

- If f is the indicator function of a closed convex set C, i.e.,

  $$f(y) = \iota_C(y) = 0 \text{ if } \; y \in C; +\infty \text{ if } \; y \notin C, \tag{3}$$

  then (2) amounts to finding the best approximation to x from C. This setting is discussed in [6].

- A common situation in analog models is to take $\mathcal{H} = \mathrm{H}^1(\Omega)$ and f to be an integral functional of the form

  $$f(y) = \int_\Omega \varphi\big(\omega, y(\omega), \nabla y(\omega)\big) d\omega. \tag{4}$$

  This setting covers various approaches, including total variation, least-squares, Fisher information, and entropic denoising, e.g., [1, 2, 3, 9, 13, 14, 16, 17].

- In discrete models, images are $N \times N$ matrices and $\mathcal{H} = \mathbb{R}^{N\times N}$. If w is a zero mean Gaussian noise, then (2) covers maximum *a posteriori* models with an *a priori* Gibbs density $p \propto \exp(-f)$. See [10, 19] for further special instances of (2) in discrete models.

Under the present hypotheses, Problem (2) is well-posed in the sense that, for every data $x \in \mathcal{H}$, it possesses a unique solution x_γ (a stronger property will be established in Proposition 1). From a strict mathematical programming viewpoint, the regularization approach associated with (2) consists of letting the regularization parameter $\lambda = 1/(2\gamma)$ go to zero, i.e., $\gamma \to +\infty$. In this regard, it is known from standard optimization theory that, under mild assumptions, x_γ converges to the projection of x onto the set of minimizers of f as $\gamma \to +\infty$ [9]. In image processing, however, (2) is usually solved for a fixed value of γ which is believed to be appropriate in terms of balancing the contribution of the functional f, which reflects some *a priori* knowledge, and the quadratic "data fidelity" term, which incorporates the contribution of the model (1). For instance, suppose that a bound δ^2 on the noise power $\|w\|^2$ is known (see [4] for the estimation of δ). Then, under suitable qualification conditions, Lagrangian theory [12] asserts that there exists a value of γ for which the solution to (2) minimizes f subject to the constraint $\|x-y\| \leq \delta$. In practice, however, this Lagrange multiplier may be hard to determine and γ is often chosen in a heuristic fashion.

In order to study regularization techniques in a unified framework, it is useful to abstract the generic problem (2) into a device that maps an observed image x into a "filtered" image x_γ, defined as the solution to (2). We shall write this relation as

$$x_\gamma = \operatorname{prox}_{\gamma f} x. \tag{5}$$

The so-called *proximity operator* $\operatorname{prox}_{\gamma f}$ thus defined was introduced in 1962 by Moreau [15]. It turns out to enjoy remarkable properties that will allow us to establish in a systematic fashion general results on the behavior and the properties of the regularization problem (2) and of its solutions. The objective of the present paper is to give a brief overview of this program; further developments and numerical results are contained in [9]. Let us note that the idea

0-7803-8554-3/04/$20.00 ©2004 IEEE.

of using proximity operators in the modeling, analysis, and solution of signal processing problems goes back to [5].

In Section 2, we provide the necessary mathematical background. In Section 3, we present some basic properties of the proximal denoising scheme (5). Section 4 is devoted to a powerful proximal image decomposition principle that generalizes the orthogonal decomposition principle in linear image processing. Finally, in Section 5, we present splitting methods for solving (2).

2. PROXIMITY OPERATORS

We provide the necessary background in convex analysis and proximity operators. For a detailed account, see [12]. Throughout, the underlying image space is a real Hilbert space $\mathcal{H}$ with scalar product $\langle \cdot \mid \cdot \rangle$ and norm $\|\cdot\|$; $\Gamma_0(\mathcal{H})$ is the class of all convex lower semicontinuous functions from $\mathcal{H}$ to $]-\infty,+\infty]$ which 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 $P_C x \in C$ such that $\|x - P_C x\| = d_C(x)$. The point $P_C x$ is the *projection* of x onto C. The conjugate of a function $f \in \Gamma_0(\mathcal{H})$ is the function $f^* \in \Gamma_0(\mathcal{H})$ defined by

$$(\forall u \in \mathcal{H}) \quad f^*(u) = \sup_{x\in\mathcal{H}} \langle x \mid u\rangle - f(x). \tag{6}$$

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

$${}^{\gamma}\!f\colon x \mapsto \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|x-y\|^2. \tag{7}$$

For every $x \in \mathcal{H}$, the infimum in (7) is achieved at a unique point denoted by $\operatorname{prox}_{\gamma f}(x)$. The operator

$$\operatorname{prox}_f\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \operatorname*{arg\,min}_{y\in\mathcal{H}} f(y) + \frac{1}{2}\|x-y\|^2 \tag{8}$$

thus defined is called the *proximity operator* of f. Let us note that, if $f = \iota_C$ (see (3)), then

$${}^{\gamma}\iota_C = d_C^2/2\gamma \quad \text{and} \quad \operatorname{prox}_{\gamma f} = P_C. \tag{9}$$

Proximity operators are therefore a generalization of projection operators.

3. PROXIMAL IMAGE DENOISING: BASIC PROPERTIES

We fix a function $f \in \Gamma_0(\mathcal{H})$ and, for the sake of convenience, we shall sometimes use the notation (5) when no ambiguity is possible.

As noted earlier, $\operatorname{prox}_{\gamma f}$ acts as a (generally nonlinear) denoising filter. Our first result states that this filtering operation is very robust in the Hadamard sense in that there is a Lipschitz-continuous dependence between the observed data and the denoised images.

Proposition 1 [9] *Let x and y be two images in $\mathcal{H}$ and $\gamma \in \,]0,+\infty[$. Then $\|\operatorname{prox}_{\gamma f} x - \operatorname{prox}_{\gamma f} y\| \le \|x-y\|$, in short, $\|x_\gamma - y_\gamma\| \le \|x-y\|$.*

Another robustness issue of interest is the behavior of the denoised image when the function f varies. The following results provides answers for certain types of perturbations.

Proposition 2 [9] *Let x be an image in $\mathcal{H}$ and $\gamma \in \,]0,+\infty[$. Then:*

i) *Affine perturbation: Let $g = f + \langle \cdot \mid u\rangle + \alpha$, where $u \in \mathcal{H}$ and $\alpha \in \mathbb{R}$. Then*

$$\operatorname{prox}_{\gamma g} x = \operatorname{prox}_{\gamma f}(x - \gamma u). \tag{10}$$

ii) *Quadratic perturbation: Let $g = f + \alpha\|\cdot\|^2/2$, where $\alpha \in [0,+\infty[$. Then*

$$\operatorname{prox}_{\gamma g} x = \operatorname{prox}_{\gamma f/(1+\alpha\gamma)} \big(x/(1+\alpha\gamma)\big). \tag{11}$$

iii) *Translation: Let $g = f(\cdot - u)$, where $u \in \mathcal{H}$. Then*

$$\operatorname{prox}_{\gamma g} x = u + \operatorname{prox}_{\gamma f}(x-u). \tag{12}$$

iv) *Unitary transformation: Let $g = f \circ L$, where $L\colon \mathcal{H} \to \mathcal{H}$ is a unitary operator. Then*

$$\operatorname{prox}_{\gamma g} x = L^{-1}\big(\operatorname{prox}_{\gamma f}(Lx)\big). \tag{13}$$

v) *Moreau envelope: Let $g = {}^{\gamma}\!f$. Then*

$$\operatorname{prox}_g x = x + \frac{1}{1+\gamma}\big(\operatorname{prox}_{(1+\gamma)f} x - x\big). \tag{14}$$

In the special case when $f = \iota_C$ for some nonempty closed convex set $C \subset \mathcal{H}$ then, in view of (9), (14) becomes an underrelaxed projection operation as

$$\operatorname{prox}_g x = x + \frac{1}{1+\gamma}\big(P_C x - x\big) \tag{15}$$

lies on the segment between x and $P_C x$. Along the same lines, proximity operators can also be used to perform various thresholding operations. Here is a simple example.

Proposition 3 [9] *Take a nonempty closed convex set $C \subset \mathcal{H}$ and set $f = d_C$. Then*

$$x_\gamma = \begin{cases} x + \dfrac{\gamma}{d_C(x)}(P_C x - x) & \text{if } d_C(x) > \gamma \\ P_C x & \text{if } d_C(x) \le \gamma. \end{cases} \tag{16}$$

Here, C can be thought of as some set of images possessing a certain property (see [4, 18] for examples of closed convex sets modeling pertinent constraints in image recovery). If the image x is close enough to satisfying the property in question, then the denoised image x_γ is simply the projection of x onto C; otherwise, x_γ is obtained through a nonstationary underrelaxation of this projection. Let us now consider the special case when $C = \{0\}$. Then (16) becomes a basic soft thresholding operation

$$x_\gamma = \begin{cases} \big(1 - \gamma/\|x\|\big)x & \text{if } \|x\| > \gamma \\ 0 & \text{if } \|x\| \leq \gamma, \end{cases} \tag{17}$$

which can be taken to various transform domains through (13). After simple manipulations, we can thus recover standard wavelet shrinkage techniques [11] for image denoising and express them in the format (5) [9].

4. PROXIMAL IMAGE DECOMPOSITION

By way of introduction, let us consider a basic denoising problem in $\mathcal{H} = L^2$. Suppose that the Fourier transform $\widehat{w}$ of the noise in (1) lies mostly in some high-frequency band B whereas $\widehat{\overline{x}}$ lies mostly in a lower frequency range. Then it is natural to obtain the denoised image $x^\oplus$ by low-pass filtering x. Now let V be the closed vector subspace of $\mathcal{H}$ of images whose Fourier transform is zero on B and let P_V be the projection operator onto V. Then it is well-known [4, 18] that the above filtering operation can be written as $x^\oplus = P_V x$. Upon invoking the standard orthogonal decomposition theorem, we thus obtain a decomposition of the energy of the image x as

$$\|x\|^2 = d_V^2(x) + d_{V^\perp}^2(x) \tag{18}$$

and a decomposition of x into a low-frequency component $x^\oplus$ and a high-frequency component $x^\ominus$, namely

$$x = x^\oplus + x^\ominus, \quad \text{where} \quad \begin{cases} x^\oplus = P_V x \\ x^\ominus = P_{V^\perp} x. \end{cases} \tag{19}$$

Now let $f = \iota_V$. Then, since (6) yields $f^* = \iota_{V^\perp}$, we deduce from (9) that (18) can be rewritten as

$$\|x\|^2 = 2\gamma\big({}^{\gamma}f(x) + {}^{1/\gamma}(f^*)(x/\gamma)\big) \tag{20}$$

and (19) as

$$x = x_\gamma^\oplus + x_\gamma^\ominus, \quad \text{where} \quad \begin{cases} x_\gamma^\oplus = \operatorname{prox}_{\gamma f} x \\ x_\gamma^\ominus = \gamma \operatorname{prox}_{f^*/\gamma}(x/\gamma), \end{cases} \tag{21}$$

whence

$$f(x_\gamma^\oplus) + f^*(x_\gamma^\ominus/\gamma) = \big\langle x_\gamma^\oplus \mid x_\gamma^\ominus \big\rangle/\gamma. \tag{22}$$

A remarkable result due to Moreau [15] states that the decompositions (20), (21), and (22) are actually true for *any* function $f \in \Gamma_0(\mathcal{H})$, thus yielding a powerful nonlinear signal decomposition principle. So far, the denoising problem (2) has been viewed only as an operation $x \mapsto x_\gamma^\oplus$. The above decomposition principle tells us that the component of the image that it left out, namely $x_\gamma^\ominus$, actually possesses a "dual structure" since, up to a factor γ, it is obtained by applying the same type of proximal operation to x, except that the function f is now replaced by its conjugate f^*. We therefore obtain a decomposition of the image

$$x \mapsto (x_\gamma^\oplus, x_\gamma^\ominus) \tag{23}$$

which is entirely parametrized by f. Deeper insights into the properties of the denoising procedure can be gained from the availability of the two components in the decomposition [5, 9]. For instance, in the special case when $\mathcal{H} = L^2(\Omega)$ and f is the total variation, i.e.,

$$f\colon x \mapsto \int_\Omega |\nabla x(\omega)| d\omega, \tag{24}$$

then the denoising problem (2) was proposed in [17]. In this case, the above proximal decomposition principle appears implicitly in [14], where $x_\gamma^\oplus$ was described as a bounded variation component of the image x carrying most of its structure, while $x_\gamma^\ominus$ was described as a texture/noise component. One will find in [14] a detailed and insightful analysis of this Rudin-Osher-Fatemi decomposition model.

Another benefit of the decomposition (21) is to provide an alternative means to compute $x_\gamma^\oplus$. Indeed, in some problems, it may be difficult to compute $x_\gamma^\oplus = \operatorname{prox}_{\gamma f} x$ directly but the dual problem of applying $\operatorname{prox}_{f^*/\gamma}$ may be easier. One can then obtain the denoised image as

$$x_\gamma^\oplus = x - \gamma \operatorname{prox}_{f^*/\gamma}(x/\gamma). \tag{25}$$

As an illustration, suppose that f is defined as

$$f\colon x \mapsto \sup_{\varphi \in G} \langle x \mid L\varphi \rangle, \tag{26}$$

where L is a bounded linear operator from some real Hilbert space $\mathcal{G}$ to $\mathcal{H}$, and where G is a nonempty convex subset of $\mathcal{G}$. Now let $C = \overline{L(G)}$. Then we can write

$$f\colon x \mapsto \sup_{u \in C} \langle x \mid u \rangle = \sigma_C(x), \tag{27}$$

where σ_C denotes the support function of the closed convex set C. Since $f^* = \sigma_C^* = \iota_C$, (9) asserts that we can calculate the denoised image $x_\gamma^\oplus$ through a projection operation, as (25) becomes $x_\gamma^\oplus = x - \gamma P_C(x/\gamma)$. In the case when f is the total variation functional, this scheme is implicitly used in [2].

5. SPLITTING AND CONSTRAINED SOLUTIONS

In some problems, solving numerically (2) may be difficult due to the nature of f. In other words, the evaluation of prox_f may be hard for some functions. Let us now suppose that f can be split into the sum of two functions, say

$$f = f_1 + f_2, \quad (f_1, f_2) \in \Gamma_0(\mathcal{H})^2. \tag{28}$$

In view of (2), the problem becomes

$$\text{Minimize } \; f_1(y) + f_2(y) + \frac{1}{2\gamma}\|x-y\|^2 \; \text{ for } \; y \in \mathcal{H}. \tag{29}$$

Depending on the properties of f_1 and f_2, several schemes can be devised to compute prox_f by using the (usually simpler) operators $\operatorname{prox}_{f_1}$ and/or $\operatorname{prox}_{f_2}$ iteratively [7]. For instance, suppose that

i) $\operatorname{prox}_{f_1}$ is relatively easy to implement;

ii) f_2 is differentiable and ∇f_2 is β-Lipschitz on $\mathcal{H}$ (see, e.g., [8] for examples).

Then the following algorithm can be used, which requires proximal steps with respect to f_1 only.

Theorem 4 [9] *Fix* $\varepsilon \in \left]0, \gamma/(\beta\gamma+1)\right[$ *and let* $(\gamma_n)_{n\in\mathbb{N}}$ *be a sequence in* $[\varepsilon, 2\gamma/(\beta\gamma+1)-\varepsilon]$. *Then, for every* $y_0 \in \mathcal{H}$, *the sequence* $(y_n)_{n\in\mathbb{N}}$ *constructed according to the recursion*

$$y_{n+1} = \operatorname{prox}_{\gamma_n f_1}\left[\left(1 - \frac{\gamma_n}{\gamma}\right)y_n + \frac{\gamma_n}{\gamma}\big(x - \gamma\nabla f_2(y_n)\big)\right]$$

converges to $x_\gamma = \operatorname{prox}_{\gamma f}\, x$.

A special case of interest is when $f_1 = \iota_C$, for some nonempty closed convex set $C \subset \mathcal{H}$. In this case, we seek the optimal solution to the denoising problem relative to f_2 over the feasibility set C, i.e.,

$$\text{Minimize } \; f_2(y) + \frac{1}{2\gamma}\|x-y\|^2 \; \text{ for } \; y \in C. \tag{30}$$

This formulation makes it possible to incorporate more *a priori* information in terms of hard constraints on the solution (see [4, 18] for examples of pertinent constraints). In this case, the iteration in Theorem 4 assumes the form

$$y_{n+1} = P_C\left[\left(1 - \frac{\gamma_n}{\gamma}\right)y_n + \frac{\gamma_n}{\gamma}\big(x - \gamma\nabla f_2(y_n)\big)\right].$$

We conclude by noting that certain formulations, e.g., those described in [20], can be recast as the problem of minimizing $f_1 + {}^{\mu}\varphi$, where $\varphi \in \Gamma_0(\mathcal{H})$ [9]. This problem can be solved via the iteration $y_{n+1} = \operatorname{prox}_{\mu f_1} \operatorname{prox}_{\mu\varphi} y_n$. The reader is referred to [7] for a more general iteration and convergence results, and to [9] for image denoising applications.

6. REFERENCES

[1] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Englewood Cliffs, NJ: Prentice-Hall, 1977.

[2] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imaging Vision,* vol. 20, pp. 89–97, Mar. 2004.

[3] A. Chambolle and P. L. Lions, "Image recovery via total variation minimization and related problems," *Numer. Math.,* vol. 76, pp. 167–188, 1997.

[4] P. L. Combettes, "The convex feasibility problem in image recovery," in *Advances in Imaging and Electron Physics*, vol. 95, pp. 155-270. New York: Academic, 1996.

[5] P. L. Combettes, "Convexité et signal," *Actes du Congrès de Mathématiques Appliquées et Industrielles SMAI'01,* pp. 6–16. Pompadour, France, May 28–June 1, 2001.

[6] P. L. Combettes, "A block-iterative surrogate constraint splitting method for quadratic signal recovery," *IEEE Trans. Signal Processing,* vol. 51, pp. 1771-1782, July 2003.

[7] P. L. Combettes, "Solving monotone inclusions via compositions of nonexpansive averaged operators," *Optimization,* to appear.

[8] P. L. Combettes and P. Bondon, "Hard-constrained inconsistent signal feasibility problems," *IEEE Trans. Signal Processing,* vol. 47, pp. 2460-2468, Sept. 1999.

[9] P. L. Combettes and V. R. Wajs, "Proximal image recovery," Technical report, 2004.

[10] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 37, pp. 2024–2036, Dec. 1989.

[11] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[12] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems.* Philadelphia, PA: SIAM, 1999.

[13] S. L. Keeling, "Total variation based convex filters for medical imaging," *Appl. Math. Comput.,* vol. 139, pp. 101–119, 2003.

[14] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations.* Providence, RI: AMS, 2001.

[15] J. J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *C. R. Acad. Sci. Paris Sér. A Math.,* vol. 255, pp. 2897–2899, Nov. 1962.

[16] D. Noll, "Reconstruction with noisy data: An approach via eigenvalue optimization," *SIAM J. Optim.,* vol. 8, pp. 82–104, Feb. 1998.

[17] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D,* vol. 60, pp. 259–268, 1992.

[18] H. Stark and Y. Yang, *Vector Space Projections : A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics.* New York: Wiley, 1998.

[19] D. M. Titterington, "General structure of regularization procedures in image reconstruction," *Astron. Astrophys.,* vol. 144, pp. 381–387, Mar. 1985.

[20] L. A. Vese and S. J. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *J. Sci. Comput.,* vol. 19, pp. 553–572, Dec. 2003.