Chapter 1
Proximal Splitting Methods in Signal Processing*

Patrick L. Combettes and Jean-Christophe Pesquet

Abstract The proximity operator of a convex function is a natural extension of the notion of a projection operator onto a convex set. This tool, which plays a central role in the analysis and the numerical solution of convex optimization problems, has recently been introduced in the arena of inverse problems and, especially, in signal processing, where it has become increasingly important. In this paper, we review the basic properties of proximity operators which are relevant to signal processing and present optimization methods based on these operators. These proximal splitting methods are shown to capture and extend several well-known algorithms in a unifying framework. Applications of proximal methods in signal recovery and synthesis are discussed.

Key words: Alternating-direction method of multipliers, backward-backward algorithm, convex optimization, denoising, Douglas-Rachford algorithm, forward-backward algorithm, frame, Landweber method, iterative thresholding, parallel computing, Peaceman-Rachford algorithm, proximal algorithm, restoration and reconstruction, sparsity, splitting.

AMS 2010 Subject Classification: 90C25, 65K05, 90C90, 94A08

P. L. Combettes
UPMC Université Paris 06, Laboratoire Jacques-Louis Lions – UMR CNRS 7598, 75005 Paris, France. e-mail: plc@math.jussieu.fr

J.-C. Pesquet
Laboratoire d'Informatique Gaspard Monge, UMR CNRS 8049, Université Paris-Est, 77454 Marne la Vallée Cedex 2, France. e-mail: pesquet@univ-mlv.fr

* This work was supported by the Agence Nationale de la Recherche under grants ANR-08-BLAN-0294-02 and ANR-09-EMER-004-03.

1.1 Introduction

Early signal processing methods were essentially linear, as they were based on classical functional analysis and linear algebra. With the development of nonlinear analysis in mathematics in the late 1950s and early 1960s (see the bibliographies of [6,142]) and the availability of faster computers, nonlinear techniques have slowly become prevalent. In particular, convex optimization has been shown to provide efficient algorithms for computing reliable solutions in a broadening spectrum of applications.

Many signal processing problems can *in fine* be formulated as convex optimization problems of the form

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; f_1(x)+\cdots+f_m(x), \tag{1.1}$$

where $f_1,\ldots,f_m$ are convex functions from $\mathbb{R}^N$ to $]-\infty,+\infty]$. A major difficulty that arises in solving this problem stems from the fact that, typically, some of the functions are not differentiable, which rules out conventional smooth optimization techniques. In this paper, we describe a class of efficient convex optimization algorithms to solve (1.1). These methods proceed by *splitting* in that the functions $f_1, \ldots, f_m$ are used individually so as to yield an easily implementable algorithm. They are called *proximal* because each nonsmooth function in (1.1) is involved via its proximity operator. Although proximal methods, which can be traced back to the work of Martinet [98], have been introduced in signal processing only recently [46, 55], their use is spreading rapidly.

Our main objective is to familiarize the reader with proximity operators, their main properties, and a variety of proximal algorithms for solving signal and image processing problems. The power and flexibility of proximal methods will be emphasized. In particular, it will be shown that a number of apparently unrelated, well-known algorithms (e.g., iterative thresholding, projected Landweber, projected gradient, alternating projections, alternating-direction method of multipliers, alternating split Bregman) are special instances of proximal algorithms. In this respect, the proximal formalism provides a unifying framework for analyzing and developing a broad class of convex optimization algorithms. Although many of the subsequent results are extendible to infinite-dimensional spaces, we restrict ourselves to a finite-dimensional setting to avoid technical digressions.

The paper is organized as follows. Proximity operators are introduced in Section 1.2, where we also discuss their main properties and provide examples. In Sections 1.3 and 1.4, we describe the main proximal splitting algorithms, namely the forward-backward algorithm and the Douglas-Rachford algorithm. In Section 1.5, we present a proximal extension of Dykstra's projection method which is tailored to problems featuring strongly convex objectives. Composite problems involving linear transformations of the variables are addressed in Section 1.6. The algorithms discussed so far are designed for $m = 2$ functions. In Section 1.7, we discuss parallel variants of these algorithms for problems involving $m \geq 2$ functions. Concluding remarks are given in Section 1.8.

Notation. We denote by $\mathbb{R}^N$ the usual N-dimensional Euclidean space, by $\|\cdot\|$ its norm, and by I the identity matrix. Standard definitions and notation from convex analysis will be used [13, 87, 114]. The domain of a function $f\colon \mathbb{R}^N \to \,]-\infty,+\infty]$ is $\operatorname{dom} f = \{x \in \mathbb{R}^N \mid f(x) < +\infty\}$. $\Gamma_0(\mathbb{R}^N)$ is the class of lower semicontinuous convex functions from $\mathbb{R}^N$ to $]-\infty,+\infty]$ such that $\operatorname{dom} f \neq \varnothing$. Let $f \in \Gamma_0(\mathbb{R}^N)$. The conjugate of f is the function $f^* \in \Gamma_0(\mathbb{R}^N)$ defined by

$$f^*\colon \mathbb{R}^N \to \,]-\infty,+\infty]\colon u \mapsto \sup_{x\in\mathbb{R}^N} x^\top u - f(x), \tag{1.2}$$

and the subdifferential of f is the set-valued operator

$$\partial f\colon \mathbb{R}^N \to 2^{\mathbb{R}^N}\colon x \mapsto \big\{u \in \mathbb{R}^N \mid (\forall y \in \mathbb{R}^N)\ (y-x)^\top u + f(x) \leq f(y)\big\}. \tag{1.3}$$

Let C be a nonempty subset of $\mathbb{R}^N$. The indicator function of C is

$$\iota_C\colon x \mapsto \begin{cases} 0, & \text{if } x \in C;\\ +\infty, & \text{if } x \notin C, \end{cases} \tag{1.4}$$

the support function of C is

$$\sigma_C = \iota_C^*\colon \mathbb{R}^N \to \,]-\infty,+\infty]\colon u \mapsto \sup_{x\in C} u^\top x, \tag{1.5}$$

the distance from $x \in \mathbb{R}^N$ to C is $d_C(x) = \inf_{y\in C}\|x-y\|$, and the relative interior of C (i.e., interior of C relative to its affine hull) is the nonempty set denoted by $\operatorname{ri} C$. If C is closed and convex, the projection of $x \in \mathbb{R}^N$ onto C is the unique point $P_C x \in C$ such that $d_C(x) = \|x - P_C x\|$.

1.2 From projection to proximity operators

One of the first widely used convex optimization splitting algorithms in signal processing is POCS (Projection Onto Convex Sets) [31, 42, 141]. This algorithm is employed to recover/synthesize a signal satisfying simultaneously several convex constraints. Such a problem can be formalized within the framework of (1.1) by letting each function f_i be the indicator function of a nonempty closed convex set C_i modeling a constraint. This reduces (1.1) to the classical *convex feasibility problem* [31, 42, 44, 86, 93, 121, 122, 128, 141]

$$\text{find } x \in \bigcap_{i=1}^m C_i. \tag{1.6}$$

The POCS algorithm [25, 141] activates each set C_i individually by means of its projection operator P_{C_i}. It is governed by the updating rule

$$x_{n+1} = P_{C_1} \cdots P_{C_m} x_n. \tag{1.7}$$

When $\bigcap_{i=1}^m C_i \neq \varnothing$ the sequence $(x_n)_{n\in\mathbb{N}}$ thus produced converges to a solution to (1.6) [25]. Projection algorithms have been enriched with many extensions of this basic iteration to solve (1.6) [10, 43, 45, 90]. Variants have also been proposed to solve more general problems, e.g., that of finding the projection of a signal onto an intersection of convex sets [22,47,137]. Beyond such problems, however, projection methods are not appropriate and more general operators are required to tackle (1.1). Among the various generalizations of the notion of a convex projection operator that exist [10, 11, 44, 90], proximity operators are best suited for our purposes.

The projection $P_C x$ of $x \in \mathbb{R}^N$ onto the nonempty closed convex set $C \subset \mathbb{R}^N$ is the solution to the problem

$$\underset{y\in\mathbb{R}^N}{\text{minimize}} \;\; \iota_C(y) + \frac{1}{2}\|x-y\|^2. \tag{1.8}$$

Under the above hypotheses, the function ι_C belongs to $\Gamma_0(\mathbb{R}^N)$. In 1962, Moreau [101] proposed the following extension of the notion of a projection operator, whereby the function ι_C in (1.8) is replaced by an arbitrary function $f \in \Gamma_0(\mathbb{R}^N)$.

Definition 1.2.1 (Proximity operator) *Let $f \in \Gamma_0(\mathbb{R}^N)$. For every $x \in \mathbb{R}^N$, the minimization problem*

$$\underset{y\in\mathbb{R}^N}{\text{minimize}} \;\; f(y) + \frac{1}{2}\|x-y\|^2 \tag{1.9}$$

admits a unique solution, which is denoted by $\operatorname{prox}_f x$*. The operator* $\operatorname{prox}_f \colon \mathbb{R}^N \to \mathbb{R}^N$ *thus defined is the* proximity operator *of f.*

Let $f \in \Gamma_0(\mathbb{R}^N)$. The proximity operator of f is characterized by the inclusion

$$(\forall (x,p) \in \mathbb{R}^N \times \mathbb{R}^N) \quad p = \operatorname{prox}_f x \quad \Leftrightarrow \quad x - p \in \partial f(p), \tag{1.10}$$

which reduces to

$$(\forall (x,p) \in \mathbb{R}^N \times \mathbb{R}^N) \quad p = \operatorname{prox}_f x \quad \Leftrightarrow \quad x - p = \nabla f(p) \tag{1.11}$$

if f is differentiable. Proximity operators have very attractive properties that make them particularly well suited for iterative minimization algorithms. For instance, prox_f is firmly nonexpansive, i.e.,

$$(\forall x \in \mathbb{R}^N)(\forall y \in \mathbb{R}^N) \quad \|\operatorname{prox}_f x - \operatorname{prox}_f y\|^2 + \|(x - \operatorname{prox}_f x) - (y - \operatorname{prox}_f y)\|^2 \leq \|x-y\|^2, \tag{1.12}$$

and its fixed point set is precisely the set of minimizers of f. Such properties allow us to envision the possibility of developing algorithms based on the proximity operators $(\mathrm{prox}_{f_i})_{1\le i\le m}$ to solve (1.1), mimicking to some extent the way convex feasibility algorithms employ the projection operators $(P_{C_i})_{1\le i\le m}$ to solve (1.6). As shown in Table 1.1, proximity operators enjoy many additional properties. One will find in Table 1.2 closed-form expressions of the proximity operators of various functions in $\Gamma_0(\mathbb{R})$ (in the case of functions such as $|\cdot|^p$, proximity operators implicitly appear in several places, e.g., [3, 4, 35]).

From a signal processing perspective, proximity operators have a very natural interpretation in terms of denoising. Let us consider the standard denoising problem of recovering a signal $\overline{x}\in\mathbb{R}^N$ from an observation

$$y=\overline{x}+w, \tag{1.13}$$

where $w\in\mathbb{R}^N$ models noise. This problem can be formulated as (1.9), where $\|\cdot-y\|^2/2$ plays the role of a data fidelity term and where f models a priori knowledge about $\overline{x}$. Such a formulation derives in particular from a Bayesian approach to denoising [21, 124, 126] in the presence of Gaussian noise and of a prior with a log-concave density $\exp(-f)$.

1.3 Forward-backward splitting

In this section, we consider the case of $m=2$ functions in (1.1), one of which is smooth.

Problem 1.3.1 Let $f_1\in\Gamma_0(\mathbb{R}^N)$, let $f_2\colon\mathbb{R}^N\to\mathbb{R}$ be convex and differentiable with a β-Lipschitz continuous gradient ∇f_2, i.e.,

$$(\forall(x,y)\in\mathbb{R}^N\times\mathbb{R}^N)\quad \|\nabla f_2(x)-\nabla f_2(y)\|\le\beta\|x-y\|, \tag{1.14}$$

where $\beta\in\,]0,+\infty[$. Suppose that $f_1(x)+f_2(x)\to+\infty$ as $\|x\|\to+\infty$. The problem is to

$$\underset{x\in\mathbb{R}^N}{\mathrm{minimize}}\;\; f_1(x)+f_2(x). \tag{1.15}$$

It can be shown [55] that Problem 1.3.1 admits at least one solution and that, for any $\gamma\in\,]0,+\infty[$, its solutions are characterized by the fixed point equation

$$x=\mathrm{prox}_{\gamma f_1}\big(x-\gamma\nabla f_2(x)\big). \tag{1.16}$$

This equation suggests the possibility of iterating

$$x_{n+1}=\underbrace{\mathrm{prox}_{\gamma_n f_1}}_{\text{backward step}}\;\underbrace{\big(x_n-\gamma_n\nabla f_2(x_n)\big)}_{\text{forward step}} \tag{1.17}$$

Table 1.1 Properties of proximity operators [27,37,53–55,102]: $\varphi \in \Gamma_0(\mathbb{R}^N)$; $C \subset \mathbb{R}^N$ is nonempty, closed, and convex; $x \in \mathbb{R}^N$.

	Property	$f(x)$	$\operatorname{prox}_f x$
i	translation	$\varphi(x-z)$, $z \in \mathbb{R}^N$	$z + \operatorname{prox}_\varphi(x-z)$
ii	scaling	$\varphi(x/\rho)$, $\rho \in \mathbb{R} \smallsetminus \{0\}$	$\rho \operatorname{prox}_{\varphi/\rho^2}(x/\rho)$
iii	reflection	$\varphi(-x)$	$-\operatorname{prox}_\varphi(-x)$
iv	quadratic perturbation	$\varphi(x) + \alpha\|x\|^2/2 + u^\top x + \gamma$ $u \in \mathbb{R}^N$, $\alpha \geq 0$, $\gamma \in \mathbb{R}$	$\operatorname{prox}_{\varphi/(\alpha+1)}\big((x-u)/(\alpha+1)\big)$
v	conjugation	$\varphi^*(x)$	$x - \operatorname{prox}_\varphi x$
vi	squared distance	$\frac{1}{2} d_C^2(x)$	$\frac{1}{2}(x + P_C x)$
vii	Moreau envelope	$\widetilde{\varphi}(x) = \inf_{y \in \mathbb{R}^N} \varphi(y) + \frac{1}{2}\|x-y\|^2$	$\frac{1}{2}\big(x + \operatorname{prox}_{2\varphi} x\big)$
viii	Moreau complement	$\frac{1}{2}\|\cdot\|^2 - \widetilde{\varphi}(x)$	$x - \operatorname{prox}_{\varphi/2}(x/2)$
ix	decomposition in an orthonormal basis $(b_k)_{1\leq k\leq N}$	$\sum_{k=1}^N \phi_k(x^\top b_k)$ $\phi_k \in \Gamma_0(\mathbb{R})$	$\sum_{k=1}^N \operatorname{prox}_{\phi_k}(x^\top b_k) b_k$
x	semi-orthogonal linear transform	$\varphi(Lx)$ $L \in \mathbb{R}^{M\times N}$, $LL^\top = \nu I$, $\nu > 0$	$x + \nu^{-1} L^\top\big(\operatorname{prox}_{\nu\varphi}(Lx) - Lx\big)$
xi	quadratic function	$\gamma\|Lx-y\|^2/2$ $L \in \mathbb{R}^{M\times N}$, $\gamma > 0$, $y \in \mathbb{R}^M$	$(I + \gamma L^\top L)^{-1}(x + \gamma L^\top y)$
xii	indicator function	$\iota_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise} \end{cases}$	$P_C x$
xiii	distance function	$\gamma d_C(x)$, $\gamma > 0$	$\begin{cases} x + \gamma(P_C x - x)/d_C(x) & \text{if } d_C(x) > \gamma \\ P_C x & \text{otherwise} \end{cases}$
xv	function of distance	$\phi(d_C(x))$ $\phi \in \Gamma_0(\mathbb{R})$ even, differentiable at 0 with $\phi'(0) = 0$	$\begin{cases} x + \left(1 - \dfrac{\operatorname{prox}_\phi d_C(x)}{d_C(x)}\right)(P_C x - x) & \text{if } x \notin C \\ x & \text{otherwise} \end{cases}$
xv	support function	$\sigma_C(x)$	$x - P_C x$
xvii	thresholding	$\sigma_C(x) + \phi(\|x\|)$ $\phi \in \Gamma_0(\mathbb{R})$ even and not constant	$\begin{cases} \dfrac{\operatorname{prox}_\phi d_C(x)}{d_C(x)}(x - P_C x) & \text{if } d_C(x) > \max \operatorname{Argmin} \phi \\ x - P_C x & \text{otherwise} \end{cases}$

for values of the step-size parameter γ_n in a suitable bounded interval. This type of scheme is known as a *forward-backward* splitting algorithm for, using the terminology used in discretization schemes in numerical analysis [132], it can be broken up into a forward (explicit) gradient step using the function f_2, and a backward (implicit) step using the function f_1. The forward-backward algorithm finds its roots in the projected gradient method [94] and in decomposition methods for solving variational inequalities [99, 119]. More recent forms of the algorithm and refinements can be found in [23, 40, 48, 85, 130]. Let us note that, on the one hand, when $f_1 = 0$, (1.17) reduces to the *gradient method*

$$x_{n+1} = x_n - \gamma_n \nabla f_2(x_n) \tag{1.18}$$

for minimizing a function with a Lipschitz continuous gradient [19, 61]. On the other hand, when $f_2 = 0$, (1.17) reduces to the *proximal point algorithm*

$$x_{n+1} = \operatorname{prox}_{\gamma_n f_1} x_n \tag{1.19}$$

for minimizing a nondifferentiable function [26, 48, 91, 98, 115]. The forward-backward algorithm can therefore be considered as a combination of these two basic schemes. The following version incorporates relaxation parameters $(\lambda_n)_{n\in\mathbb{N}}$.

Algorithm 1.3.2 (Forward-backward algorithm)

Fix $\varepsilon \in \left]0, \min\{1, 1/\beta\}\right[$, $x_0 \in \mathbb{R}^N$

For $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} \gamma_n \in [\varepsilon, 2/\beta - \varepsilon] \\ y_n = x_n - \gamma_n \nabla f_2(x_n) \\ \lambda_n \in [\varepsilon, 1] \\ x_{n+1} = x_n + \lambda_n(\operatorname{prox}_{\gamma_n f_1} y_n - x_n). \end{array} \right. \tag{1.20}$$

Proposition 1.3.3 [55] *Every sequence $(x_n)_{n\in\mathbb{N}}$ generated by Algorithm 1.3.2 converges to a solution to Problem 1.3.1.*

The above forward-backward algorithm features varying step-sizes $(\gamma_n)_{n\in\mathbb{N}}$ but its relaxation parameters $(\lambda_n)_{n\in\mathbb{N}}$ cannot exceed 1. The following variant uses constant step-sizes and larger relaxation parameters.

Algorithm 1.3.4 (Constant-step forward-backward algorithm)

Fix $\varepsilon \in \left]0, 3/4\right[$ and $x_0 \in \mathbb{R}^N$

For $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} y_n = x_n - \beta^{-1} \nabla f_2(x_n) \\ \lambda_n \in [\varepsilon, 3/2 - \varepsilon] \\ x_{n+1} = x_n + \lambda_n(\operatorname{prox}_{\beta^{-1} f_1} y_n - x_n). \end{array} \right. \tag{1.21}$$

Proposition 1.3.5 [13] *Every sequence $(x_n)_{n\in\mathbb{N}}$ generated by Algorithm 1.3.4 converges to a solution to Problem 1.3.1.*

Although they may have limited impact on actual numerical performance, it may be of interest to know whether linear convergence rates are available for the forward-backward algorithm. In general, the answer is negative: even in the simple setting of Example 1.3.11 below, linear convergence of the iterates $(x_n)_{n\in\mathbb{N}}$ generated by Algorithm 1.3.2 fails [9, 139]. Nonetheless it can be achieved at the expense of additional assumptions on the problem [10, 24, 40, 61, 92, 99, 100, 115, 119, 144].

Another type of convergence rate is that pertaining to the objective values $(f_1(x_n)+f_2(x_n))_{n\in\mathbb{N}}$. This rate has been investigated in several places [16, 24, 83] and variants of Algorithm 1.3.2 have been developed to improve it [15, 16, 84, 104, 105, 131, 136] in the spirit of classical work by Nesterov [106]. It is important to note that the convergence of the sequence of iterates $(x_n)_{n\in\mathbb{N}}$, which is often crucial in practice, is no longer guaranteed in general in such variants. The proximal gradient method proposed in [15, 16] assumes the following form.

Algorithm 1.3.6 (Beck-Teboulle proximal gradient algorithm)
Fix $x_0 \in \mathbb{R}^N$, set $z_0 = x_0$ and $t_0 = 1$
For $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} y_n = z_n - \beta^{-1}\nabla f_2(z_n) \\ x_{n+1} = \operatorname{prox}_{\beta^{-1}f_1} y_n \\ t_{n+1} = \dfrac{1+\sqrt{4t_n^2+1}}{2} \\ \lambda_n = 1 + \dfrac{t_n - 1}{t_{n+1}} \\ z_{n+1} = x_n + \lambda_n(x_{n+1} - x_n). \end{array} \right. \tag{1.22}$$

While little is known about the actual convergence of sequences produced by Algorithm 1.3.6, the $O(1/n^2)$ rate of convergence of the objective function they achieve is optimal [103], although the practical impact of such property is not always manifest in concrete problems (see Figure 1.2 for a comparison with the Forward-Backward algorithm).

Proposition 1.3.7 [16] *Assume that, for every $y \in \operatorname{dom} f_1$, $\partial f_1(y) \neq \varnothing$, and let x be a solution to Problem 1.3.1. Then every sequence $(x_n)_{n\in\mathbb{N}}$ generated by Algorithm 1.3.6 satisfies*

$$(\forall n \in \mathbb{N} \smallsetminus \{0\}) \quad f_1(x_n) + f_2(x_n) \leq f_1(x) + f_2(x) + \frac{2\beta\|x_0 - x\|^2}{(n+1)^2}. \tag{1.23}$$

Other variations of the forward-backward algorithm have also been reported to yield improved convergence profiles [20, 70, 97, 134, 135].

Problem 1.3.1 and Proposition 1.3.3 cover a wide variety of signal processing problems and solution methods [55]. For the sake of illustration, let us provide a few examples. For notational convenience, we set $\lambda_n \equiv 1$ in Algorithm 1.3.2, which reduces the updating rule to (1.17).

Example 1.3.8 (projected gradient) In Problem 1.3.1, suppose that $f_1 = \iota_C$, where C is a closed convex subset of $\mathbb{R}^N$ such that $\{x \in C \mid f_2(x) \leq \eta\}$ is nonempty and bounded for some $\eta \in \mathbb{R}$. Then we obtain the constrained minimization problem

$$\underset{x\in C}{\text{minimize}} \;\; f_2(x). \tag{1.24}$$

Table 1.2 Proximity operator of $\phi \in \Gamma_0(\mathbb{R})$; $\alpha \in \mathbb{R}$, $\kappa > 0$, $\underline{\kappa} > 0$, $\overline{\kappa} > 0$, $\omega > 0$, $\underline{\omega} < \overline{\omega}$, $q > 1$, $\tau \geq 0$ [37, 53, 55].

	$\phi(x)$	$\operatorname{prox}_\phi x$
i	$\iota_{[\underline{\omega},\overline{\omega}]}(x)$	$P_{[\underline{\omega},\overline{\omega}]}\, x$
ii	$\sigma_{[\underline{\omega},\overline{\omega}]}(x) = \begin{cases} \underline{\omega}x & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ \overline{\omega}x & \text{otherwise} \end{cases}$	$\operatorname{soft}_{[\underline{\omega},\overline{\omega}]}(x) = \begin{cases} x - \underline{\omega} & \text{if } x < \underline{\omega} \\ 0 & \text{if } x \in [\underline{\omega},\overline{\omega}] \\ x - \overline{\omega} & \text{if } x > \overline{\omega} \end{cases}$
iii	$\psi(x) + \sigma_{[\underline{\omega},\overline{\omega}]}(x)$ $\psi \in \Gamma_0(\mathbb{R})$ differentiable at 0 $\psi'(0) = 0$	$\operatorname{prox}_\psi\big(\operatorname{soft}_{[\underline{\omega},\overline{\omega}]}(x)\big)$
iv	$\max\{\|x\| - \omega, 0\}$	$\begin{cases} x & \text{if } \|x\| < \omega \\ \operatorname{sign}(x)\omega & \text{if } \omega \leq \|x\| \leq 2\omega \\ \operatorname{sign}(x)(\|x\| - \omega) & \text{if } \|x\| > 2\omega \end{cases}$
v	$\kappa\|x\|^q$	$\operatorname{sign}(x)p$, where $p \geq 0$ and $p + q\kappa p^{q-1} = \|x\|$
vi	$\begin{cases} \kappa x^2 & \text{if } \|x\| \leq \omega/\sqrt{2\kappa} \\ \omega\sqrt{2\kappa}\|x\| - \omega^2/2 & \text{otherwise} \end{cases}$	$\begin{cases} x/(2\kappa+1) & \text{if } \|x\| \leq \omega(2\kappa+1)/\sqrt{2\kappa} \\ x - \omega\sqrt{2\kappa}\operatorname{sign}(x) & \text{otherwise} \end{cases}$
vii	$\omega\|x\| + \tau\|x\|^2 + \kappa\|x\|^q$	$\operatorname{sign}(x)\operatorname{prox}_{\kappa\|\cdot\|^q/(2\tau+1)} \dfrac{\max\{\|x\| - \omega, 0\}}{2\tau+1}$
viii	$\omega\|x\| - \ln(1 + \omega\|x\|)$	$(2\omega)^{-1}\operatorname{sign}(x)\Big(\omega\|x\| - \omega^2 - 1 + \sqrt{\big\|\omega\|x\| - \omega^2 - 1\big\|^2 + 4\omega\|x\|}\Big)$
ix	$\begin{cases} \omega x & \text{if } x \geq 0 \\ +\infty & \text{otherwise} \end{cases}$	$\begin{cases} x - \omega & \text{if } x \geq \omega \\ 0 & \text{otherwise} \end{cases}$
x	$\begin{cases} -\omega x^{1/q} & \text{if } x \geq 0 \\ +\infty & \text{otherwise} \end{cases}$	$p^{1/q}$, where $p > 0$ and $p^{2q-1} - xp^{q-1} = q^{-1}\omega$
xi	$\begin{cases} \omega x^{-q} & \text{if } x > 0 \\ +\infty & \text{otherwise} \end{cases}$	$p > 0$ such that $p^{q+2} - xp^{q+1} = \omega q$
xii	$\begin{cases} x\ln(x) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ +\infty & \text{otherwise} \end{cases}$	$W(e^{x-1})$, where W is the Lambert W-function
xiii	$\begin{cases} -\ln(x - \underline{\omega}) + \ln(-\underline{\omega}) & \text{if } x \in \,]\underline{\omega}, 0] \\ -\ln(\overline{\omega} - x) + \ln(\overline{\omega}) & \text{if } x \in \,]0, \overline{\omega}[\\ +\infty & \text{otherwise} \end{cases}$ $\underline{\omega} < 0 < \overline{\omega}$	$\begin{cases} \frac{1}{2}\Big(x + \underline{\omega} + \sqrt{\|x - \underline{\omega}\|^2 + 4}\Big) & \text{if } x < 1/\underline{\omega} \\ \frac{1}{2}\Big(x + \overline{\omega} - \sqrt{\|x - \overline{\omega}\|^2 + 4}\Big) & \text{if } x > 1/\overline{\omega} \\ 0 & \text{otherwise} \end{cases}$ (see Figure 1.1)
xiv	$\begin{cases} -\kappa\ln(x) + \tau x^2/2 + \alpha x & \text{if } x > 0 \\ +\infty & \text{otherwise} \end{cases}$	$\dfrac{1}{2(1+\tau)}\Big(x - \alpha + \sqrt{\|x - \alpha\|^2 + 4\kappa(1+\tau)}\Big)$
xv	$\begin{cases} -\kappa\ln(x) + \alpha x + \omega x^{-1} & \text{if } x > 0 \\ +\infty & \text{otherwise} \end{cases}$	$p > 0$ such that $p^3 + (\alpha - x)p^2 - \kappa p = \omega$
xvi	$\begin{cases} -\kappa\ln(x) + \omega x^q & \text{if } x > 0 \\ +\infty & \text{otherwise} \end{cases}$	$p > 0$ such that $q\omega p^q + p^2 - xp = \kappa$
xvii	$\begin{cases} -\underline{\kappa}\ln(x - \underline{\omega}) - \overline{\kappa}\ln(\overline{\omega} - x) & \\ & \text{if } x \in \,]\underline{\omega}, \overline{\omega}[\\ +\infty & \text{otherwise} \end{cases}$	$p \in \,]\underline{\omega}, \overline{\omega}[$ such that $p^3 - (\underline{\omega} + \overline{\omega} + x)p^2 +$ $\big(\underline{\omega}\overline{\omega} - \underline{\kappa} - \overline{\kappa} + (\underline{\omega} + \overline{\omega})x\big)p = \underline{\omega}\overline{\omega}x - \underline{\omega}\overline{\kappa} - \overline{\omega}\underline{\kappa}$

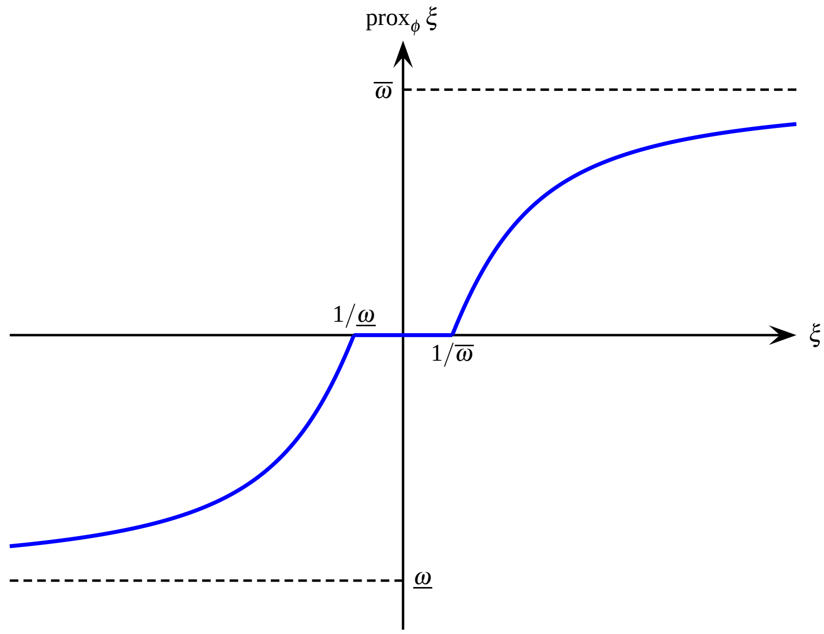


Fig. 1.1 Proximity operator of the function

$$\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right] : \xi \mapsto \begin{cases} -\ln(\xi-\underline{\omega})+\ln(-\underline{\omega}) & \text{if } \xi \in \left]\underline{\omega},0\right] \\ -\ln(\overline{\omega}-\xi)+\ln(\overline{\omega}) & \text{if } \xi \in \left]0,\overline{\omega}\right[\\ +\infty & \text{otherwise.} \end{cases}$$

The proximity operator thresholds over the interval $[1/\underline{\omega}, 1/\overline{\omega}]$, and saturates at $-\infty$ and $+\infty$ with asymptotes at $\underline{\omega}$ and $\overline{\omega}$, respectively (see Table 1.2.xiii and [53]).

Since $\mathrm{prox}_{\gamma f_1} = P_C$ (see Table 1.1.xii), the forward-backward iteration reduces to the *projected gradient method*

$$x_{n+1} = P_C\big(x_n - \gamma_n \nabla f_2(x_n)\big), \quad \varepsilon \leq \gamma_n \leq 2/\beta - \varepsilon. \tag{1.25}$$

This algorithm has been used in numerous signal processing problems, in particular in total variation denoising [34], in image deblurring [18], in pulse shape design [50], and in compressed sensing [73].

Example 1.3.9 (projected Landweber) In Example 1.3.8, setting $f_2\colon x \mapsto \|Lx - y\|^2/2$, where $L \in \mathbb{R}^{M\times N} \smallsetminus \{0\}$ and $y \in \mathbb{R}^M$, yields the constrained least-squares problem

$$\underset{x\in C}{\text{minimize}}\ \frac{1}{2}\|Lx-y\|^2. \tag{1.26}$$

Since $\nabla f_2\colon x\mapsto L^\top(Lx-y)$ has Lipschitz constant $\beta=\|L\|^2$, (1.25) yields the *projected Landweber method* [68]

$$x_{n+1}=P_C\big(x_n+\gamma_n L^\top(y-Lx_n)\big),\ \ \varepsilon\le\gamma_n\le 2/\|L\|^2-\varepsilon. \tag{1.27}$$

This method has been used in particular in computer vision [89] and in signal restoration [129].

Example 1.3.10 (backward-backward algorithm) Let f and g be functions in $\Gamma_0(\mathbb{R}^N)$. Consider the problem

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\ \ f(x)+\widetilde{g}(x), \tag{1.28}$$

where $\widetilde{g}$ is the Moreau envelope of g (see Table 1.1.vii), and suppose that $f(x)+\widetilde{g}(x)\to+\infty$ as $\|x\|\to+\infty$. This is a special case of Problem 1.3.1 with $f_1=f$ and $f_2=\widetilde{g}$. Since $\nabla f_2\colon x\mapsto x-\operatorname{prox}_g x$ has Lipschitz constant $\beta=1$ [55, 102], Proposition 1.3.3 with $\gamma_n\equiv 1$ asserts that the sequence $(x_n)_{n\in\mathbb{N}}$ generated by the *backward-backward algorithm*

$$x_{n+1}=\operatorname{prox}_f(\operatorname{prox}_g x_n) \tag{1.29}$$

converges to a solution to (1.28). Detailed analyses of this scheme can be found in [1, 14, 48, 108].

Example 1.3.11 (alternating projections) In Example 1.3.10, let f and g be respectively the indicator functions of nonempty closed convex sets C and D, one of which is bounded. Then (1.28) amounts to finding a signal x in C at closest distance from D, i.e.,

$$\underset{x\in C}{\text{minimize}}\ \frac{1}{2}d_D^2(x). \tag{1.30}$$

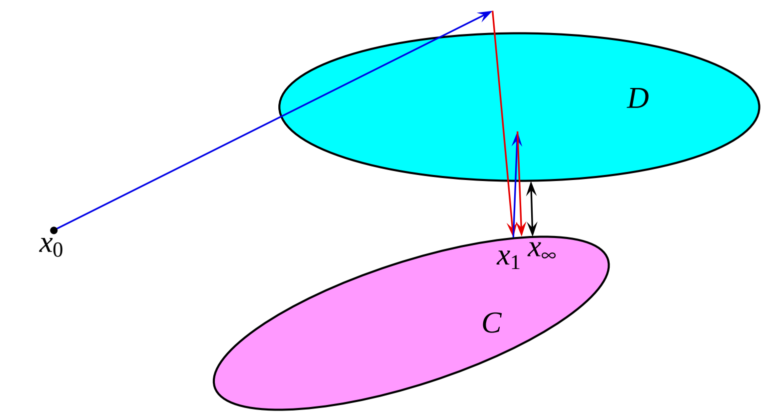
Moreover, since $\operatorname{prox}_f=P_C$ and $\operatorname{prox}_g=P_D$, (1.29) yields the *alternating projection method*

$$x_{n+1}=P_C(P_D x_n), \tag{1.31}$$

which was first analyzed in this context in [41]. Signal processing applications can be found in the areas of spectral estimation [80], pulse shape design [107], wavelet construction [109], and signal synthesis [140].

Example 1.3.12 (iterative thresholding) Let $(b_k)_{1\le k\le N}$ be an orthonormal basis of $\mathbb{R}^N$, let $(\omega_k)_{1\le k\le N}$ be strictly positive real numbers, let $L\in\mathbb{R}^{M\times N}\smallsetminus\{0\}$, and let $y\in\mathbb{R}^M$. Consider the ℓ^1–ℓ^2 problem

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\ \ \sum_{k=1}^N \omega_k|x^\top b_k|+\frac{1}{2}\|Lx-y\|^2. \tag{1.32}$$



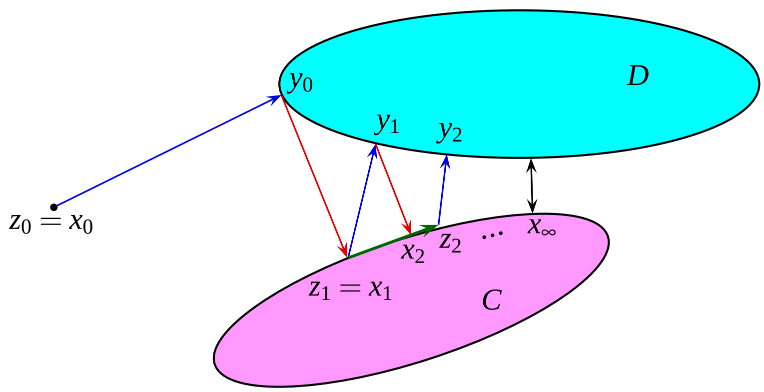


Fig. 1.2 Forward-backward versus Beck-Teboulle : As in Example 1.3.11, let C and D be two closed convex sets and consider the problem (1.30) of finding a point x_∞ in C at minimum distance from D. Let us set $f_1 = \iota_C$ and $f_2 = d_D^2/2$. Top: The forward-backward algorithm with $\gamma_n \equiv 1.9$ and $\lambda_n \equiv 1$. As seen in Example 1.3.11, it reduces to the alternating projection method (1.31). Bottom: The Beck-Teboulle algorithm.

This type of formulation arises in signal recovery problems in which y is the observed signal and the original signal is known to have a sparse representation in the basis $(b_k)_{1\leq k\leq N}$, e.g., [17, 20, 56, 58, 72, 73, 125, 127]. We observe that (1.32) is a special case of (1.15) with

$$\begin{cases} f_1\colon x\mapsto \sum_{1\leq k\leq N}\omega_k|x^\top b_k| \\ f_2\colon x\mapsto \|Lx-y\|^2/2. \end{cases} \tag{1.33}$$

Since $\operatorname{prox}_{\gamma f_1}\colon x\mapsto \sum_{1\leq k\leq N}\operatorname{soft}_{[-\gamma\omega_k,\gamma\omega_k]}(x^\top b_k)\,b_k$ (see Table 1.1.viii and Table 1.2.ii), it follows from Proposition 1.3.3 that the sequence $(x_n)_{n\in\mathbb{N}}$ generated by the *iterative thresholding algorithm*

$$x_{n+1}=\sum_{k=1}^N \xi_{k,n}b_k,\quad \text{where}$$
$$\begin{cases} \xi_{k,n}=\operatorname{soft}_{[-\gamma_n\omega_k,\gamma_n\omega_k]}\big(x_n+\gamma_n L^\top(y-Lx_n)\big)^\top b_k \\ \varepsilon\leq\gamma_n\leq 2/\|L\|^2-\varepsilon, \end{cases} \tag{1.34}$$

converges to a solution to (1.32).

Additional applications of the forward-backward algorithm in signal and image processing can be found in [28–30, 32, 36, 37, 53, 55, 57, 74].

1.4 Douglas-Rachford splitting

The forward-backward algorithm of Section 1.3 requires that one of the functions be differentiable, with a Lipschitz continuous gradient. In this section, we relax this assumption.

Problem 1.4.1 Let f_1 and f_2 be functions in $\Gamma_0(\mathbb{R}^N)$ such that

$$(\operatorname{ri}\operatorname{dom} f_1)\cap(\operatorname{ri}\operatorname{dom} f_2)\neq\varnothing \tag{1.35}$$

and $f_1(x)+f_2(x)\to+\infty$ as $\|x\|\to+\infty$. The problem is to

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\;\; f_1(x)+f_2(x). \tag{1.36}$$

What is nowadays referred to as the *Douglas-Rachford algorithm* goes back to a method originally proposed in [60] for solving matrix equations of the form $u=Ax+Bx$, where A and B are positive-definite matrices (see also [132]). The method was transformed in [95] to handle nonlinear problems and further improved in [96] to address monotone inclusion problems. For further developments, see [48, 49, 66].

Problem 1.4.1 admits at least one solution and, for any $\gamma\in\,]0,+\infty[$, its solutions are characterized by the two-level condition [52]

$$\begin{cases} x = \operatorname{prox}_{\gamma f_2} y \\ \operatorname{prox}_{\gamma f_2} y = \operatorname{prox}_{\gamma f_1}(2\operatorname{prox}_{\gamma f_2} y - y), \end{cases} \tag{1.37}$$

which motivates the following scheme.

Algorithm 1.4.2 (Douglas-Rachford algorithm)
Fix $\varepsilon \in]0,1[$, $\gamma > 0$, $y_0 \in \mathbb{R}^N$
For $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} x_n = \operatorname{prox}_{\gamma f_2} y_n \\ \lambda_n \in [\varepsilon, 2-\varepsilon] \\ y_{n+1} = y_n + \lambda_n\big(\operatorname{prox}_{\gamma f_1}(2x_n - y_n) - x_n\big). \end{array} \right. \tag{1.38}$$

Proposition 1.4.3 [52] *Every sequence $(x_n)_{n\in\mathbb{N}}$ generated by Algorithm 1.4.2 converges to a solution to Problem 1.4.1.*

Just like the forward-backward algorithm, the Douglas-Rachford algorithm operates by splitting since it employs the functions f_1 and f_2 separately. It can be viewed as more general in scope than the forward-backward algorithm in that it does not require that any of the functions have a Lipschitz continuous gradient. However, this observation must be weighed against the fact that it may be more demanding numerically as it requires the implementation of two proximal steps at each iteration, whereas only one is needed in the forward-backward algorithm. In some problems, both may be easily implementable (see Fig. 1.3 for an example) and it is not clear a priori which algorithm may be more efficient.

Applications of the Douglas-Rachford algorithm to signal and image processing can be found in [38, 52, 62, 63, 117, 118, 123].

The limiting case of the Douglas-Rachford algorithm in which $\lambda_n \equiv 2$ is the *Peaceman-Rachford algorithm* [48, 66, 96]. Its convergence requires additional assumptions (for instance, that f_2 be strictly convex and real-valued) [49].

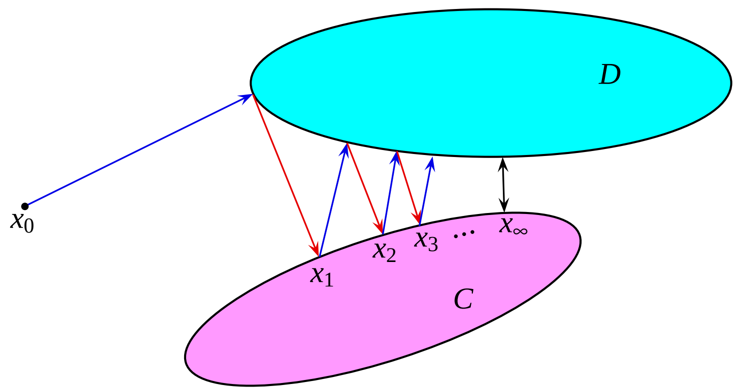
1.5 Dykstra-like splitting

In this section we consider problems involving a quadratic term penalizing the deviation from a reference signal r.

Problem 1.5.1 Let f and g be functions in $\Gamma_0(\mathbb{R}^N)$ such that $\operatorname{dom} f \cap \operatorname{dom} g \neq \varnothing$, and let $r \in \mathbb{R}^N$. The problem is to

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; f(x) + g(x) + \frac{1}{2}\|x - r\|^2. \tag{1.39}$$

It follows at once from (1.9) that Problem 1.5.1 admits a unique solution, namely $x = \operatorname{prox}_{f+g} r$. Unfortunately, the proximity operator of the sum of two functions is usually intractable. To compute it iteratively, we can observe that (1.39) can be



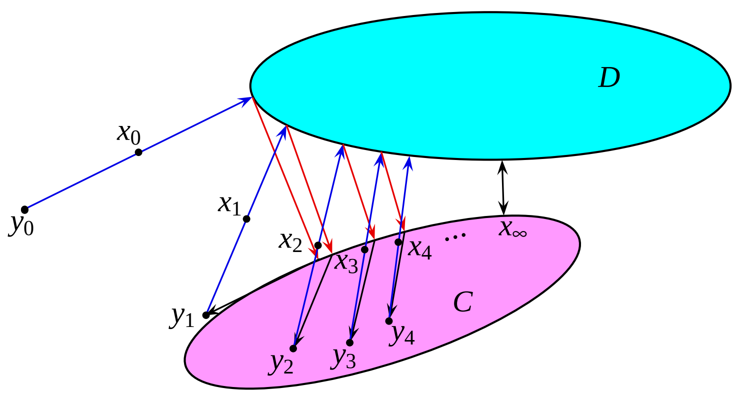


Fig. 1.3 Forward-backward versus Douglas-Rachford: As in Example 1.3.11, let C and D be two closed convex sets and consider the problem (1.30) of finding a point x_∞ in C at minimum distance from D. Let us set $f_1 = \iota_C$ and $f_2 = d_D^2/2$. Top: The forward-backward algorithm with $\gamma_n \equiv 1$ and $\lambda_n \equiv 1$. As seen in Example 1.3.11, it assumes the form of the alternating projection method (1.31). Bottom: The Douglas-Rachford algorithm with $\gamma = 1$ and $\lambda_n \equiv 1$. Table 1.1.xii yields $\operatorname{prox}_{f_1} = P_C$ and Table 1.1.vi yields $\operatorname{prox}_{f_2} \colon x \mapsto (x + P_D x)/2$. Therefore the updating rule in Algorithm 1.4.2 reduces to $x_n = (y_n + P_D y_n)/2$ and $y_{n+1} = P_C(2x_n - y_n) + y_n - x_n = P_C(P_D y_n) + y_n - x_n$.

viewed as an instance of (1.36) in Problem 1.4.1 with $f_1 = f$ and $f_2 = g + \|\cdot - r\|^2/2$. However, in this Douglas-Rachford framework, the additional qualification condition (1.35) needs to be imposed. In the present setting we require only the minimal feasibility condition $\operatorname{dom} f \cap \operatorname{dom} g \neq \varnothing$.

Algorithm 1.5.2 (Dykstra-like proximal algorithm)
Set $x_0 = r$, $p_0 = 0$, $q_0 = 0$
For $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} y_n = \operatorname{prox}_g(x_n + p_n) \\ p_{n+1} = x_n + p_n - y_n \\ x_{n+1} = \operatorname{prox}_f(y_n + q_n) \\ q_{n+1} = y_n + q_n - x_{n+1}. \end{array} \right. \tag{1.40}$$

Proposition 1.5.3 [12] *Every sequence $(x_n)_{n\in\mathbb{N}}$ generated by Algorithm 1.5.2 converges to the solution to Problem 1.5.1.*

Example 1.5.4 (best approximation) Let f and g be the indicator functions of closed convex sets C and D, respectively, in Problem 1.5.1. Then the problem is to find the best approximation to r from $C \cap D$, i.e., the projection of r onto $C \cap D$. In this case, since $\operatorname{prox}_f = P_C$ and $\operatorname{prox}_g = P_D$, the above algorithm reduces to Dykstra's projection method [22, 64].

Example 1.5.5 (denoising) Consider the problem of recovering a signal $\overline{x}$ from a noisy observation $r = \overline{x} + w$, where w models noise. If f and g are functions in $\Gamma_0(\mathbb{R}^N)$ promoting certain properties of $\overline{x}$, adopting a least-squares data fitting objective leads to the variational denoising problem (1.39).

1.6 Composite problems

We focus on variational problems with $m = 2$ functions involving explicitly a linear transformation.

Problem 1.6.1 Let $f \in \Gamma_0(\mathbb{R}^N)$, let $g \in \Gamma_0(\mathbb{R}^M)$, and let $L \in \mathbb{R}^{M\times N} \smallsetminus \{0\}$ be such that $\operatorname{dom} g \cap L(\operatorname{dom} f) \neq \varnothing$ and $f(x) + g(Lx) \to +\infty$ as $\|x\| \to +\infty$. The problem is to

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; f(x) + g(Lx). \tag{1.41}$$

Our assumptions guarantee that Problem 1.6.1 possesses at least one solution. To find such a solution, several scenarios can be contemplated.

1.6.1 Forward-backward splitting

Suppose that in Problem 1.6.1 g is differentiable with a τ-Lipschitz continuous gradient (see (1.14)). Now set $f_1 = f$ and $f_2 = g \circ L$. Then f_2 is differentiable and its gradient

$$\nabla f_2 = L^\top \circ \nabla g \circ L \tag{1.42}$$

is β-Lipschitz continuous, with $\beta = \tau\|L\|^2$. Hence, we can apply the forward-backward splitting method, as implemented in Algorithm 1.3.2. As seen in (1.20), it operates with the updating rule

$$\left\lfloor \begin{array}{l} \gamma_n \in [\varepsilon, 2/(\tau\|L\|^2) - \varepsilon] \\ y_n = x_n - \gamma_n L^\top \nabla g(Lx_n) \\ \lambda_n \in [\varepsilon, 1] \\ x_{n+1} = x_n + \lambda_n(\operatorname{prox}_{\gamma_n f} y_n - x_n). \end{array} \right. \tag{1.43}$$

Convergence is guaranteed by Proposition 1.3.3.

1.6.2 Douglas-Rachford splitting

Suppose that in Problem 1.6.1 the matrix L satisfies

$$LL^\top = \nu I, \quad \text{where} \quad \nu \in \left]0, +\infty\right[\tag{1.44}$$

and $(\operatorname{ri} \operatorname{dom} g) \cap \operatorname{ri} L(\operatorname{dom} f) \neq \varnothing$. Let us set $f_1 = f$ and $f_2 = g \circ L$. As seen in Table 1.1.x, $\operatorname{prox}_{f_2}$ has a closed-form expression in terms of prox_g and we can therefore apply the Douglas-Rachford splitting method (Algorithm 1.4.2). In this scenario, the updating rule reads

$$\left\lfloor \begin{array}{l} x_n = y_n + \nu^{-1} L^\top \big(\operatorname{prox}_{\gamma\nu g}(Ly_n) - Ly_n\big) \\ \lambda_n \in [\varepsilon, 2 - \varepsilon] \\ y_{n+1} = y_n + \lambda_n \big(\operatorname{prox}_{\gamma f}\big(2x_n - y_n\big) - x_n\big). \end{array} \right. \tag{1.45}$$

Convergence is guaranteed by Proposition 1.4.3.

1.6.3 Dual forward-backward splitting

Suppose that in Problem 1.6.1 $f = h + \|\cdot - r\|^2/2$, where $h \in \Gamma_0(\mathbb{R}^N)$ and $r \in \mathbb{R}^N$. Then (1.41) becomes

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \;\; h(x) + g(Lx) + \frac{1}{2}\|x - r\|^2, \tag{1.46}$$

which models various signal recovery problems, e.g., [33, 34, 51, 59, 112, 138]. If (1.44) holds, $\operatorname{prox}_{g\circ L}$ is decomposable, and (1.46) can be solved with the Dykstra-like method of Section 1.5, where $f_1 = h + \|\cdot - r\|^2/2$ (see Table 1.1.iv) and $f_2 = g\circ L$ (see Table 1.1.x). Otherwise, we can exploit the nice properties of the Fenchel-Moreau-Rockafellar dual of (1.46), solve this dual problem by forward-backward splitting, and recover the unique solution to (1.46) [51].

Algorithm 1.6.2 (Dual forward-backward algorithm)
Fix $\varepsilon \in \left]0, \min\{1, 1/\|L\|^2\}\right[$, $u_0 \in \mathbb{R}^M$
For $n = 0, 1, \ldots$

$$
\left\lfloor
\begin{array}{l}
x_n = \operatorname{prox}_h(r - L^\top u_n)\\
\gamma_n \in \left[\varepsilon, 2/\|L\|^2 - \varepsilon\right]\\
\lambda_n \in [\varepsilon, 1]\\
u_{n+1} = u_n + \lambda_n\big(\operatorname{prox}_{\gamma_n g^*}(u_n + \gamma_n L x_n) - u_n\big).
\end{array}
\right.
\tag{1.47}
$$

Proposition 1.6.3 [51] *Assume that $(\operatorname{ri}\operatorname{dom} g) \cap \operatorname{ri} L(\operatorname{dom} h) \neq \varnothing$. Then every sequence $(x_n)_{n\in\mathbb{N}}$ generated by the dual forward-backward algorithm 1.6.2 converges to the solution to (1.46).*

1.6.4 Alternating-direction method of multipliers

Augmented Lagrangian techniques are classical approaches for solving Problem 1.6.1 [77, 78] (see also [75, 79]). First, observe that (1.41) is equivalent to

$$
\underset{\substack{x\in\mathbb{R}^N,\, y\in\mathbb{R}^M\\ Lx=y}}{\operatorname{minimize}} \; f(x) + g(y). \tag{1.48}
$$

The *augmented Lagrangian* of index $\gamma \in \left]0, +\infty\right[$ associated with (1.48) is the saddle function

$$
\begin{aligned}
\mathscr{L}_\gamma \colon \mathbb{R}^N \times \mathbb{R}^M \times \mathbb{R}^M &\to \left]-\infty, +\infty\right]\\
(x, y, z) &\mapsto f(x) + g(y) + \frac{1}{\gamma} z^\top (Lx - y) + \frac{1}{2\gamma}\|Lx - y\|^2.
\end{aligned}
\tag{1.49}
$$

The alternating-direction method of multipliers consists in minimizing $\mathscr{L}_\gamma$ over x, then over y, and then applying a proximal maximization step with respect to the Lagrange multiplier z. Now suppose that

$$
L^\top L \text{ is invertible and } (\operatorname{ri}\operatorname{dom} g) \cap \operatorname{ri} L(\operatorname{dom} f) \neq \varnothing. \tag{1.50}
$$

By analogy with (1.9), if we denote by prox_f^L the operator which maps a point $y \in \mathbb{R}^M$ to the unique minimizer of $x \mapsto f(x) + \|Lx - y\|^2/2$, we obtain the following implementation.

Algorithm 1.6.4 (Alternating-direction method of multipliers (ADMM))
Fix $\gamma > 0$, $y_0 \in \mathbb{R}^M$, $z_0 \in \mathbb{R}^M$
For $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} x_n = \operatorname{prox}^L_{\gamma f}(y_n - z_n) \\ s_n = L x_n \\ y_{n+1} = \operatorname{prox}_{\gamma g}(s_n + z_n) \\ z_{n+1} = z_n + s_n - y_{n+1}. \end{array} \right. \tag{1.51}$$

The convergence of the sequence $(x_n)_{n\in\mathbb{N}}$ thus produced under assumption (1.50) has been investigated in several places, e.g., [75, 77, 79]. It was first observed in [76] that the ADMM algorithm can be derived from an application of the Douglas-Rachford algorithm to the dual of (1.41). This analysis was pursued in [66], where the convergence of $(x_n)_{n\in\mathbb{N}}$ to a solution to (1.41) is shown. Variants of the method relaxing the requirements on L in (1.50) have been proposed [5, 39].

In image processing, ADMM was applied in [81] to an ℓ_1 regularization problem under the name “alternating split Bregman algorithm.” Further applications and connections are found in [2, 69, 117, 143].

1.7 Problems with $m \geq 2$ functions

We return to the general minimization problem (1.1).

Problem 1.7.1 Let $f_1, \ldots, f_m$ be functions in $\Gamma_0(\mathbb{R}^N)$ such that

$$(\operatorname{ri}\operatorname{dom} f_1) \cap \cdots \cap (\operatorname{ri}\operatorname{dom} f_m) \neq \varnothing \tag{1.52}$$

and $f_1(x) + \cdots + f_m(x) \to +\infty$ as $\|x\| \to +\infty$. The problem is to

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; f_1(x) + \cdots + f_m(x). \tag{1.53}$$

Since the methods described so far are designed for $m = 2$ functions, we can attempt to reformulate (1.53) as a 2-function problem in the m-fold product space

$$\mathcal{H} = \mathbb{R}^N \times \cdots \times \mathbb{R}^N \tag{1.54}$$

(such techniques were introduced in [110, 111] and have been used in the context of convex feasibility problems in [10, 43, 45]). To this end, observe that (1.53) can be rewritten in $\mathcal{H}$ as

$$\underset{\substack{(x_1,\ldots,x_m)\in\mathcal{H} \\ x_1=\cdots=x_m}}{\text{minimize}} \;\; f_1(x_1) + \cdots + f_m(x_m). \tag{1.55}$$

If we denote by $x = (x_1, \ldots, x_m)$ a generic element in $\mathcal{H}$, (1.55) is equivalent to

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; \iota_D(x) + f(x), \tag{1.56}$$

where

$$\begin{cases} D = \big\{(x,\ldots,x) \in \mathscr{H} \mid x \in \mathbb{R}^N\big\} \\ f\colon x \mapsto f_1(x_1)+\cdots+f_m(x_m). \end{cases} \tag{1.57}$$

We are thus back to a problem involving two functions in the larger space $\mathscr{H}$. In some cases, this observation makes it possible to obtain convergent methods from the algorithms discussed in the preceding sections. For instance, the following parallel algorithm was derived from the Douglas-Rachford algorithm in [54] (see also [49] for further analysis and connections with Spingarn's splitting method [120]).

Algorithm 1.7.2 (Parallel proximal algorithm (PPXA))

Fix $\varepsilon \in \,]0,1[$, $\gamma > 0$, $(\omega_i)_{1\le i\le m} \in \,]0,1]^m$ such that
$\sum_{i=1}^m \omega_i = 1$, $y_{1,0} \in \mathbb{R}^N, \ldots, y_{m,0} \in \mathbb{R}^N$
Set $x_0 = \sum_{i=1}^m \omega_i y_{i,0}$
For $n = 0,1,\ldots$

$$\left\lfloor \begin{array}{l} \text{For } i = 1,\ldots,m \\ \left\lfloor p_{i,n} = \operatorname{prox}_{\gamma f_i/\omega_i} y_{i,n} \right. \\ p_n = \displaystyle\sum_{i=1}^m \omega_i p_{i,n} \\ \varepsilon \le \lambda_n \le 2-\varepsilon \\ \text{For } i = 1,\ldots,m \\ \left\lfloor y_{i,n+1} = y_{i,n} + \lambda_n\big(2p_n - x_n - p_{i,n}\big) \right. \\ x_{n+1} = x_n + \lambda_n(p_n - x_n). \end{array} \right.$$

Proposition 1.7.3 [54] *Every sequence $(x_n)_{n\in\mathbb{N}}$ generated by Algorithm 1.7.2 converges to a solution to Problem 1.7.1.*

Example 1.7.4 (image recovery) In many imaging problems, we record an observation $y \in \mathbb{R}^M$ of an image $\overline{z} \in \mathbb{R}^K$ degraded by a matrix $L \in \mathbb{R}^{M\times K}$ and corrupted by noise. In the spirit of a number of recent investigations (see [37] and the references therein), a tight frame representation of the images under consideration can be used. This representation is defined through a synthesis matrix $F^\top \in \mathbb{R}^{K\times N}$ (with $K \le N$) such that $F^\top F = \nu I$, for some $\nu \in \,]0,+\infty[$. Thus, the original image can be written as $\overline{z} = F^\top \overline{x}$, where $\overline{x} \in \mathbb{R}^N$ is a vector of frame coefficients to be estimated. For this purpose, we consider the problem

$$\underset{x\in C}{\text{minimize}}\;\; \frac{1}{2}\|LF^\top x - y\|^2 + \Phi(x) + \operatorname{tv}(F^\top x), \tag{1.58}$$

where C is a closed convex set modeling a constraint on $\overline{z}$, the quadratic term is the standard least-squares data fidelity term, Φ is a real-valued convex function on $\mathbb{R}^N$ (e.g., a weighted ℓ^1 norm) introducing a regularization on the frame coefficients, and tv is a discrete total variation function aiming at preserving piecewise smooth areas and sharp edges [116]. Using appropriate gradient filters in the computation of tv, it is possible to decompose it as a sum of convex functions $(\mathrm{tv}_i)_{1\le i\le q}$, the proximity operators of which can be expressed in closed form [54, 113]. Thus, (1.58) appears as a special case of (1.53) with $m = q+3$, $f_1 = \iota_C$, $f_2 = \|LF^\top \cdot - y\|^2/2$, $f_3 = \Phi$, and $f_{3+i} = \mathrm{tv}_i(F^\top \cdot)$ for $i \in \{1,\ldots,q\}$. Since a tight frame is employed, the proximity operators of f_2 and $(f_{3+i})_{1\le i\le q}$ can be deduced from Table 1.1.x. Thus, the PPXA algorithm is well suited for solving this problem numerically.

A product space strategy can also be adopted to address the following extension of Problem 1.5.1.

Problem 1.7.5 Let $f_1, \ldots, f_m$ be functions in $\Gamma_0(\mathbb{R}^N)$ such that $\operatorname{dom} f_1 \cap \cdots \cap \operatorname{dom} f_m \neq \varnothing$, let $(\omega_i)_{1\le i\le m} \in \left]0,1\right]^m$ be such that $\sum_{i=1}^m \omega_i = 1$, and let $r \in \mathbb{R}^N$. The problem is to

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; \sum_{i=1}^m \omega_i f_i(x) + \frac{1}{2}\|x - r\|^2. \tag{1.59}$$

Algorithm 1.7.6 (Parallel Dykstra-like proximal algorithm)

$$\begin{array}{l}
\text{Set } x_0 = r,\ z_{1,0} = x_0,\ \ldots,\ z_{m,0} = x_0\\
\text{For } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
\text{For } i = 1,\ldots,m\\
\left\lfloor\; p_{i,n} = \operatorname{prox}_{f_i} z_{i,n}\right.\\
x_{n+1} = \sum_{i=1}^m \omega_i p_{i,n}\\
\text{For } i = 1,\ldots,m\\
\left\lfloor\; z_{i,n+1} = x_{n+1} + z_{i,n} - p_{i,n}.\right.
\end{array}
\right.
\end{array} \tag{1.60}$$

Proposition 1.7.7 [49] *Every sequence $(x_n)_{n\in\mathbb{N}}$ generated by Algorithm 1.7.6 converges to the solution to Problem 1.7.5.*

Next, we consider a composite problem.

Problem 1.7.8 For every $i \in \{1,\ldots,m\}$, let $g_i \in \Gamma_0(\mathbb{R}^{M_i})$ and let $L_i \in \mathbb{R}^{M_i\times N}$. Assume that

$$(\exists q \in \mathbb{R}^N) \quad L_1 q \in \operatorname{ri}\operatorname{dom} g_1, \ldots, L_m q \in \operatorname{ri}\operatorname{dom} g_m, \tag{1.61}$$

that $g_1(L_1 x) + \cdots + g_m(L_m x) \to +\infty$ as $\|x\| \to +\infty$, and that $Q = \sum_{1\le i\le m} L_i^\top L_i$ is invertible. The problem is to

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; g_1(L_1 x) + \cdots + g_m(L_m x). \tag{1.62}$$

Proceeding as in (1.55) and (1.56), (1.62) can be recast as

$$\underset{\substack{x\in\mathcal{H},\, y\in\mathcal{G}\\ y = Lx}}{\text{minimize}} \;\; \iota_D(x) + g(y), \tag{1.63}$$

where

$$\begin{cases} \mathscr{H} = \mathbb{R}^N \times \cdots \times \mathbb{R}^N,\ \mathscr{G} = \mathbb{R}^{M_1} \times \cdots \times \mathbb{R}^{M_m} \\ L\colon \mathscr{H} \to \mathscr{G} : x \mapsto (L_1 x_1, \ldots, L_m x_m) \\ g\colon \mathscr{G} \to \left]-\infty, +\infty\right] : y \mapsto g_1(y_1) + \cdots + g_m(y_m). \end{cases} \tag{1.64}$$

In turn, a solution to (1.62) can be obtained as the limit of the sequence $(x_n)_{n\in\mathbb{N}}$ constructed by the following algorithm, which can be derived from the alternating-direction method of multipliers of Section 1.6.4 (alternative parallel offsprings of ADMM exist, see for instance [65]).

Algorithm 1.7.9 (Simultaneous-direction method of multipliers (SDMM))
Fix $\gamma > 0$, $y_{1,0} \in \mathbb{R}^{M_1}, \ldots, y_{m,0} \in \mathbb{R}^{M_m}$, $z_{1,0} \in \mathbb{R}^{M_1}, \ldots, z_{m,0} \in \mathbb{R}^{M_m}$

$$\begin{array}{l} \text{For } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} x_n = Q^{-1} \sum_{i=1}^m L_i^\top (y_{i,n} - z_{i,n}) \\ \text{For } i = 1, \ldots, m \\ \left\lfloor \begin{array}{l} s_{i,n} = L_i x_n \\ y_{i,n+1} = \operatorname{prox}_{\gamma g_i}(s_{i,n} + z_{i,n}) \\ z_{i,n+1} = z_{i,n} + s_{i,n} - y_{i,n+1} \end{array} \right. \end{array} \right. \end{array} \tag{1.65}$$

This algorithm was derived from a slightly different viewpoint in [118] with a connection with the work of [71]. In these papers, SDMM is applied to deblurring in the presence of Poisson noise. The computation of x_n in (1.65) requires the solution of a positive-definite symmetric system of linear equations. Efficient methods for solving such systems can be found in [82]. In certain situations, fast Fourier diagonalization is also an option [2, 71].

In the above algorithms, the proximal vectors, as well as the auxiliary vectors, can be computed simultaneously at each iteration. This parallel structure is useful when the algorithms are implemented on multicore architectures. A parallel proximal algorithm is also available to solve multicomponent signal processing problems [27]. This framework captures in particular problem formulations found in [7, 8, 80, 88, 133]. Let us add that an alternative splitting framework applicable to (1.53) was recently proposed in [67].

1.8 Conclusion

We have presented a panel of convex optimization algorithms sharing two main features. First, they employ proximity operators, a powerful generalization of the notion of a projection operator. Second, they operate by splitting the objective to be minimized into simpler functions that are dealt with individually. These methods are applicable to a wide class of signal and image processing problems ranging from restoration and reconstruction to synthesis and design. One of the main advantages of these algorithms is that they can be used to minimize nondifferentiable objectives, such as those commonly encountered in sparse approximation and compressed sensing, or in hard-constrained problems. Finally, let us note that the variational problems described in (1.39), (1.46), and (1.59), consist of computing a proximity operator. Therefore the associated algorithms can be used as a subroutine to compute approximately proximity operators within a proximal splitting algorithm, provided the latter is error tolerant (see [48, 49, 51, 66, 115] for convergence properties under approximate proximal computations). An application of this principle can be found in [38].

References

1. Acker, F., Prestel, M.A.: Convergence d'un schéma de minimisation alternée. Ann. Fac. Sci. Toulouse V. Sér. Math. **2**, 1–9 (1980)
2. Afonso, M.V., Bioucas-Dias, J.M., Figueiredo, M.A.T.: Fast image recovery using variable splitting and constrained optimization (2009). http://arxiv.org/abs/0910.4887
3. Antoniadis, A., Fan, J.: Regularization of wavelet approximations. J. Amer. Statist. Assoc. **96**, 939–967 (2001)
4. Antoniadis, A., Leporini, D., Pesquet, J.C.: Wavelet thresholding for some classes of non-Gaussian noise. Statist. Neerlandica **56**, 434–453 (2002)
5. Attouch, H., Soueycatt, M.: Augmented Lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces – applications to games, PDEs and control. Pacific J. Optim. **5**, 17–37 (2009)
6. Aubin, J.P.: Optima and Equilibria – An Introduction to Nonlinear Analysis, 2nd edn. Springer-Verlag, New York (1998)
7. Aujol, J.F., Aubert, G., Blanc-Féraud, L., Chambolle, A.: Image decomposition into a bounded variation component and an oscillating component. J. Math. Imaging Vision **22**, 71–88 (2005)
8. Aujol, J.F., Chambolle, A.: Dual norms and image decomposition models. Int. J. Computer Vision **63**, 85–104 (2005)
9. Bauschke, H.H., Borwein, J.M.: Dykstra's alternating projection algorithm for two sets. J. Approx. Theory **79**, 418–443 (1994)
10. Bauschke, H.H., Borwein, J.M.: On projection algorithms for solving convex feasibility problems. SIAM Rev. **38**, 367–426 (1996)
11. Bauschke, H.H., Combettes, P.L.: A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces. Math. Oper. Res. **26**, 248–264 (2001)
12. Bauschke, H.H., Combettes, P.L.: A Dykstra-like algorithm for two monotone operators. Pacific J. Optim. **4**, 383–391 (2008)
13. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer-Verlag (2011). To appear
14. Bauschke, H.H., Combettes, P.L., Reich, S.: The asymptotic behavior of the composition of two resolvents. Nonlinear Anal. **60**, 283–301 (2005)
15. Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE Trans. Image Process. **18**, 2419–2434 (2009)
16. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**, 183–202 (2009)
17. Bect, J., Blanc-Féraud, L., Aubert, G., Chambolle, A.: A ℓ^1 unified variational framework for image restoration. Lecture Notes in Comput. Sci. **3024**, 1–13 (2004)
18. Benvenuto, F., Zanella, R., Zanni, L., Bertero, M.: Nonnegative least-squares image deblurring: improved gradient projection approaches. Inverse Problems **26**, 18 (2010). Art. 025004
19. Bertsekas, D.P., Tsitsiklis, J.N.: Parallel and Distributed Computation: Numerical Methods. Athena Scientific, Belmont, MA (1997)

20. Bioucas-Dias, J.M., Figueiredo, M.A.T.: A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration. IEEE Trans. Image Process. **16**, 2992–3004 (2007)
21. Bouman, C., Sauer, K.: A generalized Gaussian image model for edge-preserving MAP estimation. IEEE Trans. Image Process. **2**, 296–310 (1993)
22. Boyle, J.P., Dykstra, R.L.: A method for finding projections onto the intersection of convex sets in Hilbert spaces. Lecture Notes in Statist. **37**, 28–47 (1986)
23. Bredies, K.: A forward-backward splitting algorithm for the minimization of non-smooth convex functionals in Banach space. Inverse Problems **25**, 20 (2009). Art. 015005
24. Bredies, K., Lorenz, D.A.: Linear convergence of iterative soft-thresholding. J. Fourier Anal. Appl. **14**, 813–837 (2008)
25. Brègman, L.M.: The method of successive projection for finding a common point of convex sets. Soviet Math. Dokl. **6**, 688–692 (1965)
26. Brézis, H., Lions, P.L.: Produits infinis de résolvantes. Israel J. Math. **29**, 329–345 (1978)
27. Briceño-Arias, L.M., Combettes, P.L.: Convex variational formulation with smooth coupling for multicomponent signal decomposition and recovery. Numer. Math. Theory Methods Appl. **2**, 485–508 (2009)
28. Cai, J.F., Chan, R.H., Shen, L., Shen, Z.: Convergence analysis of tight framelet approach for missing data recovery. Adv. Comput. Math. **31**, 87–113 (2009)
29. Cai, J.F., Chan, R.H., Shen, L., Shen, Z.: Simultaneously inpainting in image and transformed domains. Numer. Math. **112**, 509–533 (2009)
30. Cai, J.F., Chan, R.H., Shen, Z.: A framelet-based image inpainting algorithm. Appl. Comput. Harm. Anal. **24**, 131–149 (2008)
31. Censor, Y., Zenios, S.A.: Parallel Optimization: Theory, Algorithms and Applications. Oxford University Press, New York (1997)
32. Chaâri, L., Pesquet, J.C., Ciuciu, P., Benazza-Benyahia, A.: An iterative method for parallel MRI SENSE-based reconstruction in the wavelet domain (2009). http://www-syscom.univ-mlv.fr/∼chaari/downloads/MEDIA_Chaari.pdf
33. Chambolle, A.: An algorithm for total variation minimization and applications. J. Math. Imaging Vision **20**, 89–97 (2004)
34. Chambolle, A.: Total variation minimization and a class of binary MRF model. Lecture Notes in Comput. Sci. **3757**, 136–152 (2005)
35. Chambolle, A., DeVore, R.A., Lee, N.Y., Lucier, B.J.: Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage. IEEE Trans. Image Process. **7**, 319–335 (1998)
36. Chan, R.H., Setzer, S., Steidl, G.: Inpainting by flexible Haar-wavelet shrinkage. SIAM J. Imaging Sci. **1**, 273–293 (2008)
37. Chaux, C., Combettes, P.L., Pesquet, J.C., Wajs, V.R.: A variational formulation for frame-based inverse problems. Inverse Problems **23**, 1495–1518 (2007)
38. Chaux, C., Pesquet, J.C., Pustelnik, N.: Nested iterative algorithms for convex constrained image recovery problems. SIAM J. Imaging Sci. **2**, 730–762 (2009)
39. Chen, G., Teboulle, M.: A proximal-based decomposition method for convex minimization problems. Math. Programming **64**, 81–101 (1994)
40. Chen, G.H.G., Rockafellar, R.T.: Convergence rates in forward-backward splitting. SIAM J. Optim. **7**, 421–444 (1997)
41. Cheney, W., Goldstein, A.A.: Proximity maps for convex sets. Proc. Amer. Math. Soc. **10**, 448–450 (1959)
42. Combettes, P.L.: The foundations of set theoretic estimation. Proc. IEEE **81**, 182–208 (1993)
43. Combettes, P.L.: Inconsistent signal feasibility problems: Least-squares solutions in a product space. IEEE Trans. Signal Process. **42**, 2955–2966 (1994)
44. Combettes, P.L.: The convex feasibility problem in image recovery. In: P. Hawkes (ed.) Advances in Imaging and Electron Physics, vol. 95, pp. 155–270. Academic Press, New York (1996)
45. Combettes, P.L.: Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections. IEEE Trans. Image Process. **6**, 493–506 (1997)

46. Combettes, P.L.: Convexité et signal. In: Proc. Congrès de Mathématiques Appliquées et Industrielles SMAI'01, pp. 6–16. Pompadour, France (2001)
47. Combettes, P.L.: A block-iterative surrogate constraint splitting method for quadratic signal recovery. IEEE Trans. Signal Process. **51**, 1771–1782 (2003)
48. Combettes, P.L.: Solving monotone inclusions via compositions of nonexpansive averaged operators. Optimization **53**, 475–504 (2004)
49. Combettes, P.L.: Iterative construction of the resolvent of a sum of maximal monotone operators. J. Convex Anal. **16**, 727–748 (2009)
50. Combettes, P.L., Bondon, P.: Hard-constrained inconsistent signal feasibility problems. IEEE Trans. Signal Process. **47**, 2460–2468 (1999)
51. Combettes, P.L., Dũng, D., Vũ, B.C.: Dualization of signal recovery problems (2009). http://arxiv.org/abs/0907.0436
52. Combettes, P.L., Pesquet, J.C.: A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. IEEE J. Selected Topics Signal Process. **1**, 564–574 (2007)
53. Combettes, P.L., Pesquet, J.C.: Proximal thresholding algorithm for minimization over orthonormal bases. SIAM J. Optim. **18**, 1351–1376 (2007)
54. Combettes, P.L., Pesquet, J.C.: A proximal decomposition method for solving convex variational inverse problems. Inverse Problems **24**, 27 (2008). Art. 065014
55. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. Multiscale Model. Simul. **4**, 1168–1200 (2005)
56. Daubechies, I., Defrise, M., De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Comm. Pure Appl. Math. **57**, 1413–1457 (2004)
57. Daubechies, I., Teschke, G., Vese, L.: Iteratively solving linear inverse problems under general convex constraints. Inverse Probl. Imaging **1**, 29–46 (2007)
58. De Mol, C., Defrise, M.: A note on wavelet-based inversion algorithms. Contemp. Math. **313**, 85–96 (2002)
59. Didas, S., Setzer, S., Steidl, G.: Combined ℓ_2 data and gradient fitting in conjunction with ℓ_1 regularization. Adv. Comput. Math. **30**, 79–99 (2009)
60. Douglas, J., Rachford, H.H.: On the numerical solution of heat conduction problems in two or three space variables. Trans. Amer. Math. Soc. **82**, 421–439 (1956)
61. Dunn, J.C.: Convexity, monotonicity, and gradient processes in Hilbert space. J. Math. Anal. Appl. **53**, 145–158 (1976)
62. Dupé, F.X., Fadili, M.J., Starck, J.L.: A proximal iteration for deconvolving Poisson noisy images using sparse representations. IEEE Trans. Image Process. **18**, 310–321 (2009)
63. Durand, S., Fadili, J., Nikolova, M.: Multiplicative noise removal using L1 fidelity on frame coefficients. J. Math. Imaging Vision **36**, 201–226 (2010)
64. Dykstra, R.L.: An algorithm for restricted least squares regression. J. Amer. Stat. Assoc. **78**, 837–842 (1983)
65. Eckstein, J.: Parallel alternating direction multiplier decomposition of convex programs. J. Optim. Theory Appl. **80**, 39–62 (1994)
66. Eckstein, J., Bertsekas, D.P.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Programming **55**, 293–318 (1992)
67. Eckstein, J., Svaiter, B.F.: General projective splitting methods for sums of maximal monotone operators. SIAM J. Control Optim. **48**, 787–811 (2009)
68. Eicke, B.: Iteration methods for convexly constrained ill-posed problems in Hilbert space. Numer. Funct. Anal. Optim. **13**, 413–429 (1992)
69. Esser, E.: Applications of Lagrangian-based alternating direction methods and connections to split Bregman (2009). ftp://ftp.math.ucla.edu/pub/camreport/cam09-31.pdf
70. Fadili, J., Peyré, G.: Total variation projection with first order schemes (2009). http://hal.archives-ouvertes.fr/hal-00380491
71. Figueiredo, M.A.T., Bioucas-Dias, J.M.: Deconvolution of Poissonian images using variable splitting and augmented Lagrangian optimization. In: Proc. IEEE Workshop Statist. Signal Process. Cardiff, UK (2009). http://arxiv.org/abs/0904.4872
72. Figueiredo, M.A.T., Nowak, R.D.: An EM algorithm for wavelet-based image restoration. IEEE Trans. Image Process. **12**, 906–916 (2003)

73. Figueiredo, M.A.T., Nowak, R.D., Wright, S.J.: Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. IEEE J. Selected Topics Signal Process. **1**, 586–597 (2007)
74. Fornasier, M.: Domain decomposition methods for linear inverse problems with sparsity constraints. Inverse Problems **23**, 2505–2526 (2007)
75. Fortin, M., Glowinski, R. (eds.): Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems. North-Holland, Amsterdam (1983)
76. Gabay, D.: Applications of the method of multipliers to variational inequalities. In: M. Fortin, R. Glowinski (eds.) Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, pp. 299–331. North-Holland, Amsterdam (1983)
77. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite elements approximations. Comput. Math. Appl. **2**, 17–40 (1976)
78. Glowinski, R., Marrocco, A.: Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. RAIRO Anal. Numer. **2**, 41–76 (1975)
79. Glowinski, R., Tallec, P.L. (eds.): Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics. SIAM, Philadelphia (1989)
80. Goldburg, M., Marks II, R.J.: Signal synthesis in the presence of an inconsistent set of constraints. IEEE Trans. Circuits Syst. **32**, 647–663 (1985)
81. Goldstein, T., Osher, S.: The split Bregman method for $L1$-regularized problems. SIAM J. Imaging Sci. **2**, 323–343 (2009)
82. Golub, G.H., Van Loan, C.F.: Matrix Computations, 3rd edn. Johns Hopkins University Press, Baltimore, MD (1996)
83. Güler, O.: On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. **20**, 403–419 (1991)
84. Güler, O.: New proximal point algorithms for convex minimization. SIAM J. Optim. **2**, 649–664 (1992)
85. Hale, E.T., Yin, W., Zhang, Y.: Fixed-point continuation for l_1-minimization: methodology and convergence. SIAM J. Optim. **19**, 1107–1130 (2008)
86. Herman, G.T.: Fundamentals of Computerized Tomography – Image Reconstruction from Projections, 2nd edn. Springer-Verlag, London (2009)
87. Hiriart-Urruty, J.B., Lemaréchal, C.: Fundamentals of Convex Analysis. Springer-Verlag, New York (2001)
88. Huang, Y., Ng, M.K., Wen, Y.W.: A fast total variation minimization method for image restoration. Multiscale Model. Simul. **7**, 774–795 (2008)
89. Johansson, B., Elfving, T., Kozlovc, V., Censor, Y., Forssén, P.E., Granlund, G.: The application of an oblique-projected Landweber method to a model of supervised learning. Math. Comput. Modelling **43**, 892–909 (2006)
90. Kiwiel, K.C., Łopuch, B.: Surrogate projection methods for finding fixed points of firmly nonexpansive mappings. SIAM J. Optim. **7**, 1084–1102 (1997)
91. Lemaire, B.: The proximal algorithm. In: J.P. Penot (ed.) New Methods in Optimization and Their Industrial Uses, *International Series of Numerical Mathematics*, vol. 87, pp. 73–87. Birkhäuser, Boston, MA (1989)
92. Lemaire, B.: Itération et approximation. In: Équations aux Dérivées Partielles et Applications, pp. 641–653. Gauthiers-Villars, Paris (1998)
93. Lent, A., Tuy, H.: An iterative method for the extrapolation of band-limited functions. J. Math. Anal. Appl. **83**, 554–565 (1981)
94. Levitin, E.S., Polyak, B.T.: Constrained minimization methods. U.S.S.R. Comput. Math. Math. Phys. **6**, 1–50 (1966)
95. Lieutaud, J.: Approximation d'Opérateurs par des Méthodes de Décomposition. Thèse, Université de Paris (1969)
96. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**, 964–979 (1979)

97. Loris, I., Bertero, M., De Mol, C., Zanella, R., Zanni, L.: Accelerating gradient projection methods for ℓ^1-constrained signal recovery by steplength selection rules. Appl. Comput. Harm. Anal. **27**, 247–254 (2009)
98. Martinet, B.: Régularisation d'inéquations variationnelles par approximations successives. Rev. Française Informat. Rech. Opér. **4**, 154–158 (1970)
99. Mercier, B.: Topics in Finite Element Solution of Elliptic Problems. No. 63 in Lectures on Mathematics. Tata Institute of Fundamental Research, Bombay (1979)
100. Mercier, B.: Inéquations Variationnelles de la Mécanique. No. 80.01 in Publications Mathématiques d'Orsay. Université de Paris-XI, Orsay, France (1980)
101. Moreau, J.J.: Fonctions convexes duales et points proximaux dans un espace hilbertien. C. R. Acad. Sci. Paris Sér. A Math. **255**, 2897–2899 (1962)
102. Moreau, J.J.: Proximité et dualité dans un espace hilbertien. Bull. Soc. Math. France **93**, 273–299 (1965)
103. Nemirovsky, A.S., Yudin, D.B.: Problem Complexity and Method Efficiency in Optimization. Wiley, New York (1983)
104. Nesterov, Yu.: Smooth minimization of non-smooth functions. Math. Program. **103**, 127–152 (2005)
105. Nesterov, Yu.: Gradient methods for minimizing composite objective function. CORE discussion paper 2007076, Université Catholique de Louvain, Center for Operations Research and Econometrics (2007)
106. Nesterov, Yu.E.: A method of solving a convex programming problem with convergence rate $o(1/k^2)$. Soviet Math. Dokl. **27**, 372–376 (1983)
107. Nobakht, R.A., Civanlar, M.R.: Optimal pulse shape design for digital communication systems by projections onto convex sets. IEEE Trans. Communications **43**, 2874–2877 (1995)
108. Passty, G.B.: Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. J. Math. Anal. Appl. **72**, 383–390 (1979)
109. Pesquet, J.C., Combettes, P.L.: Wavelet synthesis by alternating projections. IEEE Trans. Signal Process. **44**, 728–732 (1996)
110. Pierra, G.: Éclatement de contraintes en parallèle pour la minimisation d'une forme quadratique. Lecture Notes in Comput. Sci. **41**, 200–218 (1976)
111. Pierra, G.: Decomposition through formalization in a product space. Math. Programming **28**, 96–115 (1984)
112. Potter, L.C., Arun, K.S.: A dual approach to linear inverse problems with convex constraints. SIAM J. Control Optim. **31**, 1080–1092 (1993)
113. Pustelnik, N., Chaux, C., Pesquet, J.C.: Parallel proximal algorithm for image restoration using hybrid regularization (2009). http://arxiv.org/abs/0911.1536
114. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton, NJ (1970)
115. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
116. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**, 259–268 (1992)
117. Setzer, S.: Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage. Lecture Notes in Comput. Sci. **5567**, 464–476 (2009)
118. Setzer, S., Steidl, G., Teuber, T.: Deblurring Poissonian images by split Bregman techniques. J. Vis. Commun. Image Represent. **21**, 193–199 (2010)
119. Sibony, M.: Méthodes itératives pour les équations et inéquations aux dérivées partielles non linéaires de type monotone. Calcolo **7**, 65–183 (1970)
120. Spingarn, J.E.: Partial inverse of a monotone operator. Appl. Math. Optim. **10**, 247–265 (1983)
121. Stark, H. (ed.): Image Recovery: Theory and Application. Academic Press, San Diego, CA (1987)
122. Stark, H., Yang, Y.: Vector Space Projections : A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics. Wiley, New York (1998)
123. Steidl, G., Teuber, T.: Removing multiplicative noise by Douglas-Rachford splitting methods. J. Math. Imaging Vision **36**, 168–184 (2010)

124. Thompson, A.M., Kay, J.: On some Bayesian choices of regularization parameter in image restoration. Inverse Problems **9**, 749–761 (1993)
125. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. Royal. Statist. Soc. B **58**, 267–288 (1996)
126. Titterington, D.M.: General structure of regularization procedures in image reconstruction. Astronom. and Astrophys. **144**, 381–387 (1985)
127. Tropp, J.A.: Just relax: Convex programming methods for identifying sparse signals in noise. IEEE Trans. Inform. Theory **52**, 1030–1051 (2006)
128. Trussell, H.J., Civanlar, M.R.: The feasible solution in signal restoration. IEEE Trans. Acoust., Speech, Signal Process. **32**, 201–212 (1984)
129. Trussell, H.J., Civanlar, M.R.: The Landweber iteration and projection onto convex sets. IEEE Trans. Acoust., Speech, Signal Process. **33**, 1632–1634 (1985)
130. Tseng, P.: Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. SIAM J. Control Optim. **29**, 119–138 (1991)
131. Tseng, P.: On accelerated proximal gradient methods for convex-concave optimization (2008). http://www.math.washington.edu/∼tseng/papers/apgm.pdf
132. Varga, R.S.: Matrix Iterative Analysis, 2nd edn. Springer-Verlag, New York (2000)
133. Vese, L.A., Osher, S.J.: Image denoising and decomposition with total variation minimization and oscillatory functions. J. Math. Imaging Vision **20**, 7–18 (2004)
134. Vonesh, C., Unser, M.: A fast thresholded Landweber algorithm for wavelet-regularized multidimensional deconvolution. IEEE Trans. Image Process. **17**, 539–549 (2008)
135. Vonesh, C., Unser, M.: A fast multilevel algorithm for wavelet-regularized image restoration. IEEE Trans. Image Process. **18**, 509–523 (2009)
136. Weiss, P., Aubert, G., Blanc-Féraud, L.: Efficient schemes for total variation minimization under constraints in image processing. SIAM J. Sci. Comput. **31**, 2047–2080 (2009)
137. Yamada, I., Ogura, N., Yamashita, Y., Sakaniwa, K.: Quadratic optimization of fixed points of nonexpansive mappings in Hilbert space. Numer. Funct. Anal. Optim. **19**, 165–190 (1998)
138. Youla, D.C.: Generalized image restoration by the method of alternating orthogonal projections. IEEE Trans. Circuits Syst. **25**, 694–702 (1978)
139. Youla, D.C.: Mathematical theory of image restoration by the method of convex projections. In: H. Stark (ed.) Image Recovery: Theory and Application, pp. 29–77. Academic Press, San Diego, CA (1987)
140. Youla, D.C., Velasco, V.: Extensions of a result on the synthesis of signals in the presence of inconsistent constraints. IEEE Trans. Circuits Syst. **33**, 465–468 (1986)
141. Youla, D.C., Webb, H.: Image restoration by the method of convex projections: Part 1 – theory. IEEE Trans. Medical Imaging **1**, 81–94 (1982)
142. Zeidler, E.: Nonlinear Functional Analysis and Its Applications, vol. I–V. Springer-Verlag, New York (1985–1990)
143. Zhang, X., Burger, M., Bresson, X., Osher, S.: Bregmanized nonlocal regularization for deconvolution and sparse reconstruction (2009). ftp://ftp.math.ucla.edu/pub/camreport/cam09-03.pdf
144. Zhu, C.Y.: Asymptotic convergence analysis of the forward-backward splitting algorithm. Math. Oper. Res. **20**, 449–464 (1995)