A Variational Inequality Model for the Construction of Signals from Inconsistent Nonlinear Equations*

Patrick L. Combettes† and Zev C. Woodstock‡

Abstract. Building up on classical linear formulations, we posit that a broad class of problems in signal synthesis and in signal recovery are reducible to the basic task of finding a point in a closed convex subset of a Hilbert space that satisfies a number of nonlinear equations involving firmly nonexpansive operators. We investigate this formalism in the case when, due to inaccurate modeling or perturbations, the nonlinear equations are inconsistent. A relaxed formulation of the original problem is proposed in the form of a variational inequality. The properties of the relaxed problem are investigated, and a provenly convergent block-iterative algorithm, whereby only blocks of the underlying firmly nonexpansive operators are activated at a given iteration, is devised to solve it. Numerical experiments illustrate robust recoveries in several signal and image processing applications.

Key words. Firmly nonexpansive operator, nonlinear equality constraint, nonlinear signal recovery, proximal point, variational inequality.

1. Introduction.

Signal construction encompasses forward problems such as image synthesis, holography, filter design, time-frequency distribution synthesis, and radiation therapy planning, as well as inverse problems such as density estimation, signal denoising, image interpolation, signal extrapolation, audio declipping, image reconstruction, or deconvolution; see, e.g., [4, 16, 19, 29, 31, 32, 45, 47, 48, 51, 58]. Essential components in the mathematical modeling of signal construction problems are equations tying the ideal solution $\overline{x}$ in a space $\mathcal{H}$ to given prescriptions in a space $\mathcal{G}$, say $W\overline{x} = p$, where W is an operator mapping $\mathcal{H}$ to $\mathcal{G}$. The prescription p can be a design specification in forward problems, or an observation in inverse problems.

In 1978, Youla [60] elegantly brought to light the simple geometry that underlies many classical problems in signal construction by reducing them to the following formulation: Given closed vector subspaces C and D in a real Hilbert space $\mathcal{H}$, and a point $p \in D$,

$$\text{find} \quad x \in C \quad \text{such that} \quad \operatorname{proj}_D x = p, \tag{1.1}$$

where proj_D denotes the projection operator onto D. In the context of signal recovery, the original signal of interest $\overline{x}$ is known to lie in C, and some observation p of it is available in the form of its projection onto D. A natural nonlinear extension of this setting is obtained by considering nonempty closed convex sets C in $\mathcal{H}$ and D in a real Hilbert space $\mathcal{G}$, a bounded linear operator $L\colon \mathcal{H} \to \mathcal{G}$, and a point $p \in D$, and setting as an objective to

$$\text{find} \quad x \in C \quad \text{such that} \quad \operatorname{proj}_D(Lx) = p. \tag{1.2}$$

*The work of P. L. Combettes was supported by the National Science Foundation under grant CCF-1715671 and the work of Z. C. Woodstock was supported by the National Science Foundation under grant DGE-1746939.

†North Carolina State University, Department of Mathematics, Raleigh, NC 27695, USA (plc@math.ncsu.edu)

‡North Carolina State University, Department of Mathematics, Raleigh, NC 27695, USA (zwoodst@ncsu.edu)

An early instance of this model appears in [1], where C is a set of bandlimited signals and p is an observation of N clipped samples of the original signal. Thus, $L\colon \mathcal{H} \to \mathbb{R}^N$ is the sampling operator and $D = \big\{y \in \mathbb{R}^N \mid \|y\|_\infty \leqslant \rho\big\}$ for some $\rho \in \left]0, +\infty\right[$. A key property of projectors onto closed convex sets is their firm nonexpansiveness. Recall that an operator $F\colon \mathcal{G} \to \mathcal{G}$ is firmly nonexpansive if [6]

$$(\forall x \in \mathcal{G})(\forall y \in \mathcal{G}) \quad \langle x - y \mid Fx - Fy\rangle \geqslant \|Fx - Fy\|^2. \tag{1.3}$$

In [26, 27], it was shown that many nonlinear observation processes found in signal processing, machine learning, and inference problems can be represented through such operators. This prompts us to consider the following formulation, whereby the prescriptions are modeled via Wiener systems (see Figure 1.1).

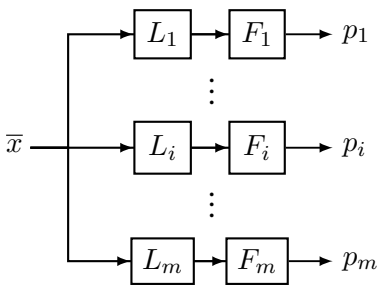

Figure 1.1. *Illustration of Problem 1.1 with m prescriptions $(p_i)_{1\leqslant i\leqslant m}$. The ith prescription p_i is the output produced when the ideal signal $\overline{x}$ is input to a Wiener system $W_i = F_i \circ L_i$, i.e., the concatenation of a linear system L_i and a nonlinear system F_i [49]. In the proposed model, F_i is a firmly nonexpansive operator.*

Problem 1.1. Let I be a nonempty finite set and let C be a nonempty closed convex subset of a real Hilbert space $\mathcal{H}$. For every $i \in I$, let $\mathcal{G}_i$ be a real Hilbert space, let $p_i \in \mathcal{G}_i$, let $L_i\colon \mathcal{H} \to \mathcal{G}_i$ be a nonzero bounded linear operator, and let $F_i\colon \mathcal{G}_i \to \mathcal{G}_i$ be a firmly nonexpansive operator. The task is to

$$\text{find } \; x \in C \; \text{ such that } \; (\forall i \in I) \;\; F_i(L_i x) = p_i. \tag{1.4}$$

The work of [26, 27] assumes that the prescription equations in Problem 1.1 are exact and hence that a solution exists. In many instances, however, the prescription operators may be imperfectly known or the model may be corrupted by perturbations, so that Problem 1.1 may not have solutions, e.g., [17, 18, 31]. A dramatic consequence of this lack of feasibility is that the algorithms proposed [26, 27] are known to diverge in such situations. To deal robustly with possibly inconsistent equations, one must therefore come up with an appropriate relaxed formulation of Problem 1.1, i.e., one that seeks a point in C that satisfies the nonlinear equations in an approximate sense, and coincides with the original problem (1.4) if it happens to be consistent. To guide our design of a relaxed problem, let us consider a classical instantiation of Problem 1.1.

Example 1.2. Specialize Problem 1.1 by setting, for every $i \in I$,

$$p_i = 0 \;\text{ and }\; F_i = \mathrm{Id} - \mathrm{proj}_{D_i}, \;\text{ where } D_i \text{ is a nonempty closed convex subset of } \mathcal{G}_i, \tag{1.5}$$

and note that the operators $(F_i)_{i\in I}$ are firmly nonexpansive [6, Corollary 4.18]. In this context, (1.4) reduces to the convex feasibility problem [15, 19, 62]

$$\text{find } x \in C \text{ such that } (\forall i \in I)\ L_i x \in D_i. \tag{1.6}$$

Let $(\omega_i)_{i\in I}$ be real numbers in $]0,1]$ such that $\sum_{i\in I}\omega_i = 1$ and, for every $i \in I$, let d_{D_i} be the distance function to D_i. As seen in [23] (see also [16, 17, 18, 22, 32, 61] for special cases), a relaxation of (1.6) when it may be inconsistent is the least-squares problem

$$\underset{x\in C}{\text{minimize}}\ f(x), \quad \text{where} \quad f\colon x \mapsto \frac{1}{2}\sum_{i\in I}\omega_i d^2_{D_i}(L_i x) = \frac{1}{2}\sum_{i\in I}\omega_i\|L_i x - \text{proj}_{D_i}(L_i x)\|^2. \tag{1.7}$$

An important property of this formulation is that f is a smooth convex function since [6, Corollary 12.31] asserts that

$$(\forall i \in I) \quad \nabla\frac{d^2_{D_i}\circ L_i}{2} = L_i^*\circ(\text{Id} - \text{proj}_{D_i})\circ L_i = L_i^*\circ F_i\circ L_i - L_i^* p_i. \tag{1.8}$$

It can therefore be solved by the projection-gradient algorithm [6, Corollary 28.10]. Let us also note that (1.7) is a valid relaxation of (1.6). Indeed, if the latter has solutions, then f vanishes on C at those points only, and (1.7) is therefore equivalent to (1.6). Historically, the first instance of the above relaxation process seems to be Legendre's least-squares methods [37]. There, $\mathcal{H} = \mathbb{R}^N = C$ and, for every $i \in I$, $\mathcal{G}_i = \mathbb{R}$, $D_i = \{\beta_i\}$, and $L_i = \langle\cdot \mid a_i\rangle$, where $\beta_i \in \mathbb{R}$ and $0 \neq a_i \in \mathbb{R}^N$. Set $b = (\beta_i)_{i\in I}$, let A be the matrix with rows $(a_i)_{i\in I}$, and let $(\forall i \in I)\ \omega_i = 1/\text{card}\, I$. Then (1.6) consists of solving the linear system $Ax = b$ and (1.7) of minimizing the function $x \mapsto \|Ax - b\|^2$.

In general, there is no suitable relaxation of Problem 1.1 in the form of a tractable convex minimization problem such as (1.7). For instance, in Example 1.2, we can rewrite (1.7) as

$$\underset{x\in C}{\text{minimize}}\ f(x), \quad \text{where} \quad f\colon x \mapsto \frac{1}{2}\sum_{i\in I}\omega_i\|F_i(L_i x) - p_i\|^2. \tag{1.9}$$

However, beyond the special case (1.5), f is typically a nonconvex and nondifferentiable function [4, 43, 64], which makes it impossible to guarantee the construction of solutions. Another plausible formulation that captures (1.7) would be to introduce in Problem 1.1 the closed convex sets $(\forall i \in I)\ D_i = \big\{y \in \mathcal{G}_i \mid F_i y_i = p_i\big\}$. However, the resulting minimization problem (1.7) is intractable because we typically do not know how to evaluate the operators $(\text{proj}_{D_i})_{i\in I}$, and therefore cannot evaluate f and its gradient.

Our strategy to relax (1.4) is to forgo the optimization approach in favor of the broader framework of *variational inequalities*. To motivate this approach, let us go back to Example 1.2. Then it follows from Lemma 2.4 below and (1.8) that (1.7) is equivalent to finding $x \in C$ such that $(\forall y \in C)\ \sum_{i\in I}\omega_i\langle L_i(y - x) \mid F_i(L_i x) - p_i\rangle \geqslant 0$. We shall show that this variational inequality constitutes an appropriate relaxed formulation of Problem 1.1 in the presence of general firmly nonexpansive operators $(F_i)_{i\in I}$, and that it can be solved iteratively through an efficient block-iterative fixed point algorithm. Here is a precise formulation of our relaxed problem.

Problem 1.3. Let I be a nonempty finite set, let $(\omega_i)_{i\in I}$ be real numbers in $]0,1]$ such that $\sum_{i\in I}\omega_i = 1$, and let C be a nonempty closed convex subset of a real Hilbert space $\mathcal{H}$. For every $i \in I$, let $\mathcal{G}_i$ be a real Hilbert space, let $p_i \in \mathcal{G}_i$, let $L_i\colon \mathcal{H} \to \mathcal{G}_i$ be a nonzero bounded linear operator, and let $F_i\colon \mathcal{G}_i \to \mathcal{G}_i$ be a firmly nonexpansive operator. The task is to

$$\text{find } x \in C \text{ such that } (\forall y \in C) \sum_{i\in I}\omega_i\langle L_i(y-x) \mid F_i(L_ix) - p_i\rangle \geqslant 0. \tag{1.10}$$

The paper is organized as follows. Section 2 provides the notation and the necessary background, as well as preliminary results. It covers in particular the basics of monotone operator theory, which will play an essential role in the paper. In Section 3, we illustrate the flexibility and the breadth the proposed firmly nonexpansive Wiener model. In Section 4, we analyze various properties of Problem 1.3, in particular as a relaxation of Problem 1.1. We also provide in that section a block-iterative algorithm to solve Problem 1.3. Section 5 is devoted to numerical experiments in the area of signal and image processing.

2. Notation, background, and preliminary results.

2.1. Notation. Our notation follows [6], to which one can refer for background on monotone operators and convex analysis. Let $\mathcal{H}$ be a real Hilbert space with scalar product $\langle\cdot \mid \cdot\rangle$, associated norm $\|\cdot\|$, and identity operator Id. The family of all subsets of $\mathcal{H}$ is denoted by $2^{\mathcal{H}}$. The Hilbert direct sum of a family of real Hilbert spaces $(\mathcal{H}_i)_{i\in I}$ is denoted by $\bigoplus_{i\in I}\mathcal{H}_i$.

Let $T\colon \mathcal{H} \to \mathcal{H}$. Then T is *cocoercive* if there exists $\beta \in]0,+\infty[$ such that

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x-y \mid Tx-Ty\rangle \geqslant \beta\|Tx-Ty\|^2, \tag{2.1}$$

and *firmly nonexpansive* if $\beta = 1$ above. The set of *fixed points* of T is $\operatorname{Fix} T = \big\{x \in \mathcal{H} \mid Tx = x\big\}$.

Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$. The *graph* of A is $\operatorname{gra} A = \big\{(x,x^*) \in \mathcal{H}\times\mathcal{H} \mid x^* \in Ax\big\}$, the *domain* of A is $\operatorname{dom} A = \big\{x \in \mathcal{H} \mid Ax \neq \varnothing\big\}$, the *range* of A is $\operatorname{ran} A = \big\{x^* \in \mathcal{H} \mid (\exists\, x \in \mathcal{H})\ x^* \in Ax\big\}$, the set of zeros of A is $\operatorname{zer} A = \big\{x \in \mathcal{H} \mid 0 \in Ax\big\}$, the *inverse* of A is $A^{-1}\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x^* \mapsto \big\{x \in \mathcal{H} \mid x^* \in Ax\big\}$, and the *resolvent* of A is $J_A = (\mathrm{Id} + A)^{-1}$. Further, A is *monotone* if

$$\big(\forall (x,x^*) \in \operatorname{gra} A\big)\big(\forall (y,y^*) \in \operatorname{gra} A\big) \quad \langle x-y \mid x^*-y^*\rangle \geqslant 0, \tag{2.2}$$

and *maximally monotone* if, for every $(x,x^*) \in \mathcal{H}\times\mathcal{H}$,

$$(x,x^*) \in \operatorname{gra} A \quad \Leftrightarrow \quad \big(\forall (y,y^*) \in \operatorname{gra} A\big)\ \langle x-y \mid x^*-y^*\rangle \geqslant 0. \tag{2.3}$$

If A is maximally monotone, then J_A is a single-valued firmly nonexpansive operator defined on $\mathcal{H}$. If A is monotone and satisfies

$$(\forall (x,x^*) \in \operatorname{dom} A \times \operatorname{ran} A)\ \ \sup\big\{\langle x-y \mid y^*-x^*\rangle \mid (y,y^*) \in \operatorname{gra} A\big\} < +\infty, \tag{2.4}$$

then it is 3^* *monotone*.

$\Gamma_0(\mathcal{H})$ is the class of all lower semicontinuous convex functions from $\mathcal{H}$ to $]-\infty,+\infty]$ which are proper in the sense that they are not identically $+\infty$. Let $f \in \Gamma_0(\mathcal{H})$. The *domain* of f is $\operatorname{dom} f = \big\{x \in \mathcal{H} \mid f(x) < +\infty\big\}$, the *conjugate* of f is the function

$$\Gamma_0(\mathcal{H}) \ni f^*\colon x^* \mapsto \sup_{x\in\mathcal{H}}\big(\langle x \mid x^*\rangle - f(x)\big), \tag{2.5}$$

and the *subdifferential* of f is the maximally monotone operator

$$\partial f\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \big\{x^* \in \mathcal{H} \mid (\forall y \in \mathcal{H})\ \langle y - x \mid x^*\rangle + f(x) \leqslant f(y)\big\}. \tag{2.6}$$

The *Moreau envelope* of f is

$$\widetilde{f}\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \inf_{y\in\mathcal{H}} \left(f(y) + \frac{\|x-y\|^2}{2}\right). \tag{2.7}$$

For every $x \in \mathcal{H}$, the infimum in (2.7) is achieved at a unique point, which is denoted by $\operatorname{prox}_f x$. This defines the *proximity operator* $\operatorname{prox}_f = J_{\partial f}$ of f.

Let C be a nonempty closed and convex subset of $\mathcal{H}$. The *distance* from $x \in \mathcal{H}$ to C is $d_C(x) = \inf_{y\in C}\|x-y\|$, the *indicator function* of C is

$$\iota_C\colon \mathcal{H} \to \left]-\infty,+\infty\right]\colon x \mapsto \begin{cases} 0, & \text{if } x \in C;\\ +\infty, & \text{if } x \notin C, \end{cases} \tag{2.8}$$

the *normal cone* to C at $x \in \mathcal{H}$ is

$$N_C x = \partial \iota_C(x) = \begin{cases} \big\{x^* \in \mathcal{H} \mid (\forall y \in C)\ \ \langle y - x \mid x^*\rangle \leqslant 0\big\}, & \text{if } x \in C;\\ \varnothing, & \text{otherwise,} \end{cases} \tag{2.9}$$

and the *projection operator* onto C is $\operatorname{proj}_C = \operatorname{prox}_{\iota_C} = J_{N_C}$.

The following facts will also come into play.

Lemma 2.1. *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\mu \in \left]0,+\infty\right[$, and let $\gamma \in \left]0, 1/\mu\right[$. Set $B = A - \mu\mathrm{Id}$ and $\beta = 1 - \gamma\mu$. Then $J_{\gamma B}\colon \mathcal{H} \to \mathcal{H}$ is β-cocoercive. Furthermore, $J_{\gamma B} = J_{\beta^{-1}\gamma A} \circ (\beta^{-1}\mathrm{Id})$.*

Proof. Let x and q be in $\mathcal{H}$. Since $\beta^{-1}\gamma A$ is maximally monotone, its resolvent is single-valued with domain $\mathcal{H}$. Therefore,

$$\begin{aligned} q \in J_{\gamma B}x &\Leftrightarrow x - q \in \gamma B q\\ &\Leftrightarrow x - \beta q \in \gamma A q\\ &\Leftrightarrow \beta^{-1}x - q \in \beta^{-1}\gamma A q\\ &\Leftrightarrow q = J_{\beta^{-1}\gamma A}\big(\beta^{-1}x\big), \end{aligned} \tag{2.10}$$

which shows that $J_{\gamma B} = J_{\beta^{-1}\gamma A} \circ (\beta^{-1}\mathrm{Id})$ is single-valued with domain $\mathcal{H}$. Finally, since $M = \beta\gamma A$ is maximally monotone, it follows from [6, Corollary 23.26] that $J_{\gamma B} = J_{\beta^{-2}M} \circ (\beta^{-1}\mathrm{Id})$ is β-cocoercive. ∎

Lemma 2.2 ([6, Proposition 24.68]). *Let $\mathcal{H}$ be the real Hilbert space of $N \times M$ matrices under the Frobenius norm, and set $s = \min\{N, M\}$. Denote the singular value decomposition of $x \in \mathcal{H}$ by $x = U_x \operatorname{diag}(\sigma_1(x), \ldots, \sigma_s(x))V_x^\top$. Let $\phi \in \Gamma_0(\mathbb{R})$ be even, and set*

$$F\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto U_x \operatorname{diag}\Big(\operatorname{prox}_\phi\big(\sigma_1(x)\big), \ldots, \operatorname{prox}_\phi\big(\sigma_s(x)\big)\Big)V_x^\top. \tag{2.11}$$

Then F is firmly nonexpansive.

2.2. Variational inequalities. The following notion of a *variational inequality* was formulated in [12] (see Figure 2.1).

Definition 2.3. *Let C be a nonempty closed convex subset of $\mathcal{H}$ and let $B\colon \mathcal{H} \to \mathcal{H}$ be a monotone operator. The associated* variational inequality *problem is to*

$$\text{find } x \in C \text{ such that } (\forall y \in C)\ \langle y - x \mid Bx\rangle \geqslant 0. \tag{2.12}$$

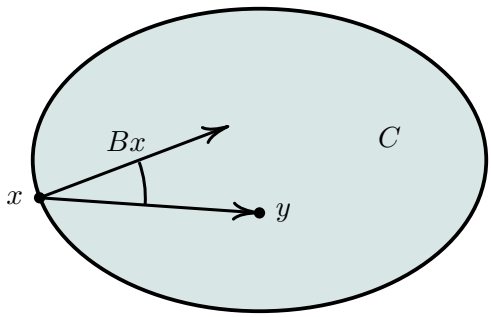

Figure 2.1. *Illustration of the variational inequality principle. The point x solves (2.12) because it lies in C and, for every $y \in C$, the angle between $y - x$ and Bx is acute.*

Variational inequalities are used in various areas of mathematics and its applications [8, 30, 35, 65]. They are also central in constrained minimization problems.

Lemma 2.4. [6, Proposition 27.8] *Let $f\colon \mathcal{H} \to \mathbb{R}$ be a differentiable convex function, let C be a nonempty closed convex subset of $\mathcal{H}$, and let $x \in \mathcal{H}$. Then x minimizes f over C if and only if it satisfies the variational inequality*

$$x \in C \text{ and } (\forall y \in C)\ \langle y - x \mid \nabla f(x)\rangle \geqslant 0. \tag{2.13}$$

2.3. Composite sums of monotone operators. We shall require the following Brézis–Haraux-type theorem, which remains valid in general reflexive Banach spaces (see [10, Théorème 3] for the special case of the sum of two monotone operators).

Lemma 2.5. *Let $\mathcal{H}$ be a real Hilbert space and let $(\mathcal{G}_i)_{i\in I}$ be a finite family of real Hilbert spaces. Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a 3^* monotone operator and, for every $i \in I$, let $B_i\colon \mathcal{G}_i \to 2^{\mathcal{G}_i}$ be a 3^* monotone operator and let $L_i\colon \mathcal{H} \to \mathcal{G}_i$ be a bounded linear operator. Suppose that $A + \sum_{i\in I} L_i^* \circ B_i \circ L_i$ is maximally monotone. Then*

$$\begin{cases} \operatorname{int}\big(\operatorname{ran} A + \sum_{i\in I} L_i^*(\operatorname{ran} B_i)\big) = \operatorname{int}\operatorname{ran}\big(A + \sum_{i\in I} L_i^* \circ B_i \circ L_i\big) \\ \overline{\operatorname{ran} A + \sum_{i\in I} L_i^*(\operatorname{ran} B_i)} = \overline{\operatorname{ran}}\,\big(A + \sum_{i\in I} L_i^* \circ B_i \circ L_i\big). \end{cases} \tag{2.14}$$

Proof. Clearly, $\operatorname{ran}(A + \sum_{i\in I} L_i^* \circ B_i \circ L_i) \subset (\operatorname{ran} A + \sum_{i\in I} L_i^*(\operatorname{ran} B_i))$. It is therefore enough to show that

$$\begin{cases} \operatorname{int}\big(\operatorname{ran} A + \sum_{i\in I} L_i^*(\operatorname{ran} B_i)\big) \subset \operatorname{ran}\big(A + \sum_{i\in I} L_i^* \circ B_i \circ L_i\big) \\ \operatorname{ran} A + \sum_{i\in I} L_i^*(\operatorname{ran} B_i) \subset \overline{\operatorname{ran}}\,\big(A + \sum_{i\in I} L_i^* \circ B_i \circ L_i\big). \end{cases} \tag{2.15}$$

Without loss of generality, set $I = \{1, \dots, m\}$ and introduce the Hilbert direct sum $\boldsymbol{\mathcal{H}} = \mathcal{H} \oplus \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_m$. Furthermore, introduce the bounded linear operator $\boldsymbol{L} \colon \mathcal{H} \to \boldsymbol{\mathcal{H}} \colon x \mapsto (x, L_1 x, \dots, L_m x)$ and the operator $\boldsymbol{M} \colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}} \colon (x, y_1, \dots, y_m) \mapsto Ax \times B_1 y_1 \times \cdots \times B_m y_m$, which is 3^* monotone since A and $(B_i)_{i \in I}$ are. Note also that, since $\boldsymbol{L}^* \colon \boldsymbol{\mathcal{H}} \to \mathcal{H} \colon (x, y_1, \dots, y_m) \mapsto x + \sum_{i \in I} L_i^* y_i$, the operator

$$\boldsymbol{L}^* \circ \boldsymbol{M} \circ \boldsymbol{L} = A + \sum_{i \in I} L_i^* \circ B_i \circ L_i \tag{2.16}$$

is maximally monotone. We can therefore apply [42, Theorem 5] to obtain

$$\begin{cases} \operatorname{int} \boldsymbol{L}^*(\operatorname{ran} \boldsymbol{M}) \subset \operatorname{ran} (\boldsymbol{L}^* \circ \boldsymbol{M} \circ \boldsymbol{L}) \\ \boldsymbol{L}^*(\operatorname{ran} \boldsymbol{M}) \subset \overline{\operatorname{ran}} (\boldsymbol{L}^* \circ \boldsymbol{M} \circ \boldsymbol{L}), \end{cases} \tag{2.17}$$

which is precisely (2.15). ∎

We consider below a monotone inclusion problem involving several operators.

Problem 2.6. Let $(\omega_i)_{i \in I}$ be a finite family of real numbers in $]0, 1]$ such that $\sum_{i \in I} \omega_i = 1$, let $A_0 \colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and, for every $i \in I$, let $\beta_i \in]0, +\infty[$ and let $A_i \colon \mathcal{H} \to \mathcal{H}$ be β_i-cocoercive. The task is to find $x \in \mathcal{H}$ such that $0 \in A_0 x + \sum_{i \in I} \omega_i A_i x$.

Proposition 2.7. [24, Proposition 4.9] *Consider the setting of Problem 2.6 under the assumption that it has a solution. Let K be a strictly positive integer and let $(I_n)_{n \in \mathbb{N}}$ be a sequence of nonempty subsets of I such that $(\forall n \in \mathbb{N}) \bigcup_{k=0}^{K-1} I_{n+k} = I$. Let $\gamma \in]0, 2 \min_{1 \leqslant i \leqslant m} \beta_i[$, let $x_0 \in \mathcal{H}$, and let $(\forall i \in I)\ t_{i,-1} \in \mathcal{H}$. Iterate*

$$\begin{aligned} &\text{for } n = 0, 1, \dots \\ &\quad \text{for every } i \in I_n \\ &\quad\quad t_{i,n} = x_n - \gamma A_i x_n \\ &\quad \text{for every } i \in I \smallsetminus I_n \\ &\quad\quad t_{i,n} = t_{i,n-1} \\ &\quad x_{n+1} = J_{\gamma A_0}\Big(\sum_{i \in I} \omega_i t_{i,n} \Big). \end{aligned} \tag{2.18}$$

Then $(x_n)_{n \in \mathbb{N}}$ converges weakly to a solution to Problem 2.6.

3. Firmly nonexpansive Wiener models. The proposed Wiener model (see Figure 1.1) involves a linear operator followed by a firmly nonexpansive operator acting on a real Hilbert space $\mathcal{G}$. Typical examples of linear transformations in the context of signal construction include the Fourier transform, the Radon transform, wavelet decompositions, frame decompositions, audio effects, or blurring operators. We show that firmly nonexpansive operators model many useful nonlinearities in this context. Key examples based on those of [27] are recalled and new ones are proposed. Following [27], we call $p \in \mathcal{G}$ a *proximal point* of $y \in \mathcal{G}$ relative to a firmly nonexpansive operator $F \colon \mathcal{G} \to \mathcal{G}$ if $Fy = p$.

3.1. Projection operators. As seen in Section 2.1, the projection operator onto a nonempty closed convex set is firmly nonexpansive.

Example 3.1. For every $j \in \{1, \dots, m\}$, let G_j be a real Hilbert space and let $\mathsf{D}_j \subset \mathsf{G}_j$ be nonempty closed and convex. Suppose that $\mathcal{G} = \bigoplus_{1 \leqslant j \leqslant m} \mathsf{G}_j$. The operator

$$F \colon (\mathsf{y}_j)_{1 \leqslant j \leqslant m} \mapsto (\operatorname{proj}_{\mathsf{D}_j} \mathsf{y}_j)_{1 \leqslant j \leqslant m}, \tag{3.1}$$

which is also the projection onto the closed convex set $D = \bigtimes_{1 \leqslant j \leqslant m} \mathsf{D}_j$, is the hard clipper of [27, Example 2.11]. If we specialize to the case when, for every $j \in \{1, \dots, m\}$, $\mathsf{G}_j = \mathbb{R}$, we obtain the standard hard clipping operators of [1, 31, 55].

Example 3.2. Let $K \subset \mathcal{G}$ be a nonempty closed convex cone. The operator $F = \operatorname{proj}_K$ is used as a distortion model when K is the positive orthant [53, Section 10.4.1]. Another instance of a conic projection operator arises in isotonic regression [5], where $K = \big\{(\xi_i)_{1 \leqslant i \leqslant N} \in \mathbb{R}^N \mid \xi_1 \leqslant \cdots \leqslant \xi_N\big\}$.

Example 3.3. Compression schemes such as downsampling project a high-dimensional object of interest onto a closed convex subset of a low-dimensional subspace of $\mathcal{G}$ [41].

3.2. Proximity operators. As seen in Section 2.1, the proximity operator of a function in $\Gamma_0(\mathcal{G})$ is firmly nonexpansive. The following construction will be particularly useful.

Example 3.4. For every $j \in \{1, \dots, m\}$, let G_j be a real Hilbert space and let $\mathsf{g}_j \in \Gamma_0(\mathsf{G}_j)$. Suppose that $\mathcal{G} = \bigoplus_{1 \leqslant j \leqslant m} \mathsf{G}_j$ and set $F \colon \mathcal{G} \to \mathcal{G} \colon (\mathsf{y}_j)_{1 \leqslant j \leqslant m} \mapsto (\operatorname{prox}_{\mathsf{g}_i} \mathsf{y}_j)_{1 \leqslant j \leqslant m}$. Then [6, Proposition 24.11] asserts that

$$F = \operatorname{prox}_g, \quad \text{where} \quad g \colon \mathcal{G} \to \left]-\infty, +\infty\right] : (\mathsf{y}_j)_{1 \leqslant j \leqslant m} \mapsto \sum_{j=1}^{m} \mathsf{g}_j(\mathsf{y}_j). \tag{3.2}$$

Example 3.5. In Example 3.4 suppose that, for every $j \in \{1, \dots, m\}$, $\mathsf{g}_j = \phi_j \circ \|\cdot\|$, where ϕ_j is an even function in $\Gamma_0(\mathbb{R})$ such that $\phi_j(0) = 0$ and $\phi_j \neq \iota_{\{0\}}$. Set $(\forall j \in \{1, \dots, m\})$ $\rho_j = \max \partial \phi_j(0)$. Then we derive from [11, Proposition 2.1] that

$$F \colon \mathcal{G} \to \mathcal{G} \colon (\mathsf{y}_j)_{1 \leqslant j \leqslant m} \mapsto \Big(\big(\operatorname{prox}_{\phi_j} \|\mathsf{y}_j\|\big) \lfloor \mathsf{y}_j \rfloor_{\rho_j} \Big)_{1 \leqslant j \leqslant m}, \tag{3.3}$$

$$\text{where} \quad \lfloor \mathsf{y}_j \rfloor_{\rho_j} = \begin{cases} \mathsf{y}_j / \|\mathsf{y}_j\|, & \text{if } \|\mathsf{y}_j\| > \rho_j; \\ 0, & \text{if } \|\mathsf{y}_j\| \leqslant \rho_j. \end{cases}$$

Example 3.6. Consider the special case of Example 3.5 in which, for some $j \in \{1, \dots, m\}$, ϕ_j is not differentiable at the origin, which implies that $\rho_j > 0$. Then $\operatorname{prox}_{\mathsf{g}_j}$ acts as a thresholder with respect to the jth variable in the sense that, if $\|\mathsf{y}_j\| \leqslant \rho_j$, then the jth coordinate of Fy is zero. For instance, suppose that, for every $j \in \{1, \dots, m\}$, $\phi_j = \rho_j |\cdot|$, hence $\partial \phi_j(0) = [-\rho_j, \rho_j]$ and $\mathsf{g}_j = \rho_j \|\cdot\|$. Then $Fy = p$ is acquired though the group-shrinkage operation [63]

$$p = \left(\left(1 - \frac{\rho_j}{\max\{\|\mathsf{y}_j\|, \rho_j\}} \right) \mathsf{y}_j \right)_{1 \leqslant j \leqslant m}. \tag{3.4}$$

Example 3.7. In contrast to the hard clipping operations of Example 3.1, soft clipping operators are not projection operators in general, but many turn out to be proximity operators [27] (see Figure 3.1). For instance, consider the setting of Example 3.5 with

$$(\forall j \in \{1,\dots,m\}) \quad \phi_j \colon \eta \mapsto \begin{cases} -|\eta| - \ln(1-|\eta|) - \dfrac{\eta^2}{2}, & \text{if } |\eta| < 1; \\ +\infty, & \text{if } |\eta| \geqslant 1. \end{cases} \tag{3.5}$$

Then we obtain the soft clipping operator

$$F \colon (\mathsf{y}_j)_{1\leqslant j\leqslant m} \mapsto \left(\frac{\mathsf{y}_j}{1+\|\mathsf{y}_j\|}\right)_{1\leqslant j\leqslant m} \tag{3.6}$$

used in [39]. Soft clipping operators model sensors in signal processing [4, 39, 53] and activation functions in neural networks [25].

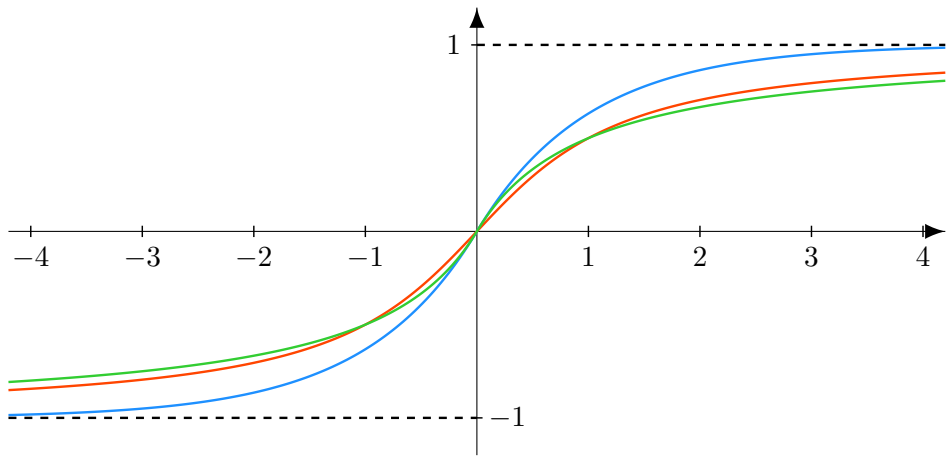

Figure 3.1. *Proximal soft clipping operators on $\mathbb{R}$ with saturation at ± 1: $\eta \mapsto \operatorname{sign}(\eta)(1-\exp(-|\eta|))$ [53, Section 10.6.3] (blue), $\eta \mapsto 2\arctan(\eta)/\pi$ [25] (red), and $\eta \mapsto \eta/(1+|\eta|)$ [39] (green).*

3.3. General firmly nonexpansive operators. Not all firmly nonexpansive operators are proximity operators [21].

Example 3.8. Let $(R_j)_{1\leqslant j\leqslant m}$ be nonexpansive operators on $\mathcal{G}$. Then the operator

$$F = \frac{\mathrm{Id} + R_1 \circ \cdots \circ R_m}{2} \tag{3.7}$$

is firmly nonexpansive [6, Proposition 4.4] but it is not a proximity operator [21, Example 3.5]. A concrete instance of (3.7) is found in audio signal processing. Consider a distortion $p \in \mathcal{G}$ of a linearly degraded audio signal $L\overline{x} \in \mathcal{G}$ modeled by

$$F(L\overline{x}) = p, \tag{3.8}$$

where L produces effects such as echo or reverberation [53, Chapter 11], and F comprises several simpler operations $(R_j)_{1\leqslant j\leqslant m}$ which turn out to be firmly nonexpansive (see, e.g.,

Example 3.2, [27], and [53, Section 10.6.2]). These simpler distortion operators are then used in series and blended with a proportion of the input signal [53, Section 10.9], so that the overall process is described by (3.7) (see Figure 3.2). More generally F remains firmly nonexpansive when $R_1 \circ \cdots \circ R_m$ is replaced by any nonexpansive operator.

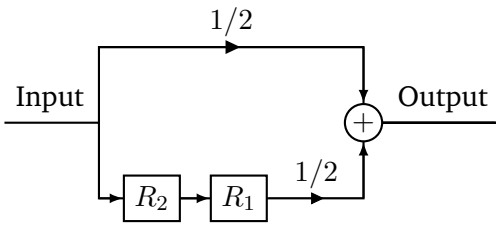

Figure 3.2. *The distortion operator F in Example 3.8 for $m = 2$.*

3.4. Proximation. In some instances, a prescription $q \in \mathcal{G}$ may be given by an equation of the form $Qy = q$, where $Q\colon \mathcal{G} \to \mathcal{G}$ is not firmly nonexpansive. In this section, we provide constructive examples of *proximation*, by which we mean the replacement of the equality $Qy = q$ with an equivalent equality $Fy = p$, where $p \in \mathcal{G}$ and $F\colon \mathcal{G} \to \mathcal{G}$ is firmly nonexpansive.

Definition 3.9. *Let $Q\colon \mathcal{G} \to \mathcal{G}$ and let $q \in \operatorname{ran} Q$. Then (Q, q) is* proximable *if there exists a firmly nonexpansive operator $F\colon \mathcal{G} \to \mathcal{G}$ and $p \in \operatorname{ran} F$ such that $(\forall y \in \mathcal{G})\ Qy = q \Leftrightarrow Fy = p$. In this case (F, p) is a* proximation *of (Q, q).*

We begin with a necessary condition describing when this technique is possible.

Proposition 3.10. *Let $Q\colon \mathcal{G} \to \mathcal{G}$ and $q \in \operatorname{ran} Q$ be such that (Q, q) is proximable. Then*

$$Q^{-1}(\{q\}) = \{y \in \mathcal{G} \mid Qy = q\} \text{ is closed and convex.} \tag{3.9}$$

Proof. The proximation assumption means that there exists a firmly nonexpansive operator $F\colon \mathcal{G} \to \mathcal{G}$ and $p \in \operatorname{ran} F$ such that $Q^{-1}(\{q\}) = F^{-1}(\{p\})$. Now set $T = \operatorname{Id} - F + p$. Then it follows from [6, Proposition 4.4] that T is firmly nonexpansive, and therefore from [6, Corollary 4.24] that $Q^{-1}(\{q\}) = F^{-1}(\{p\}) = \operatorname{Fix} T$ is closed and convex. ∎

Interestingly, condition (3.9) is also assumed in various nonlinear recovery problems [45, 46, 56]. However, the solution techniques of these papers require the ability to project onto $Q^{-1}(\{q\})$ – a capability which rarely occurs when $\dim \mathcal{G} > 1$. The numerical approach proposed in Section 4 will circumvent this requirement and lead to provenly convergent algorithms which instead rely on evaluating the associated firmly nonexpansive operator $F\colon \mathcal{G} \to \mathcal{G}$.

Example 3.11 ([27, Proposition 2.14]). For every $j \in \{1, \ldots, m\}$, let G_j be a real Hilbert space, let D_j be a nonempty closed convex subset of G_j, let $\gamma_j \in]0, +\infty[$, and set

$$\mathsf{Q}_j\colon \mathsf{G}_j \to \mathsf{G}_j\colon \mathsf{y}_j \mapsto \begin{cases} \mathsf{y}_j, & \text{if } d_{\mathsf{D}_j}(\mathsf{y}_j) > \gamma_j; \\ \operatorname{proj}_{\mathsf{D}_j} \mathsf{y}_j, & \text{if } d_{\mathsf{D}_j}(\mathsf{y}_j) \leqslant \gamma_j \end{cases} \tag{3.10}$$

and

$$\mathsf{S}_j\colon \mathsf{G}_j \to \mathsf{G}_j\colon \mathsf{y}_j \mapsto \begin{cases} \mathsf{y}_j + \dfrac{\gamma_j}{d_{\mathsf{D}_j}(\mathsf{y}_j)}(\operatorname{proj}_{\mathsf{D}_j}\mathsf{y}_j - \mathsf{y}_j), & \text{if } \mathsf{y}_j \notin \mathsf{D}_j; \\ \mathsf{y}_j, & \text{if } \mathsf{y}_j \in \mathsf{D}_j. \end{cases} \tag{3.11}$$

Suppose that $\mathcal{G} = \bigoplus_{1\leqslant j\leqslant m} \mathsf{G}_j$, set $Q\colon \mathcal{G} \to \mathcal{G}\colon (\mathsf{y}_j)_{1\leqslant j\leqslant m} \mapsto (\mathsf{Q}_j\mathsf{y}_j)_{1\leqslant j\leqslant m}$, and let $q \in \operatorname{ran} Q$. Even though Q is discontinuous, (Q, q) is proxifiable. Indeed, set $S\colon \mathcal{G} \to \mathcal{G}\colon (\mathsf{y}_j)_{1\leqslant j\leqslant m} \mapsto (\mathsf{S}_j\mathsf{y}_j)_{1\leqslant j\leqslant m}$, $F\colon \mathcal{G} \to \mathcal{G}\colon (\mathsf{y}_j)_{1\leqslant j\leqslant m} \mapsto (\mathsf{S}_j(\mathsf{Q}_j\mathsf{y}_j))_{1\leqslant j\leqslant m}$, and $p = Sq$. Then (F, p) is a proxification of (Q, q). In particular if, for every $j \in \{1, \ldots, m\}$, $\mathsf{D}_j = \{0\}$, then Q is the block thresholding estimation operator of [34, Section 2.3].

Example 3.12. Consider Example 3.11 with, for every $j \in \{1, \ldots, m\}$, $\mathsf{G}_j = \mathbb{R}$, $\mathsf{D}_j = \{0\}$, and $\gamma_j = \gamma \in]0, +\infty[$. Then each operator Q_j in (3.10) reduces to the hard thresholder

$$\operatorname{hard}_\gamma : \eta \mapsto \begin{cases} \eta, & \text{if } |\eta| > \gamma; \\ 0, & \text{if } |\eta| \leqslant \gamma, \end{cases} \tag{3.12}$$

$\mathsf{S}_j : \eta \mapsto \eta - \gamma \operatorname{sign}(\eta)$, and

$$\mathsf{S}_j \circ \operatorname{hard}_\gamma \;=\; \operatorname{soft}_\gamma : \eta \mapsto \operatorname{sign}(\eta) \max\{|\eta| - \gamma, 0\} \tag{3.13}$$

is the soft thresholder on $[-\gamma, \gamma]$. Furthermore, it follows from Example 3.11 that (F, p) is a proxification of (Q, q). The resulting transformation Q is used for signal compression in [28, 54], and as a sensing model in [9].

Next, we combine Example 3.12 with Lemma 2.2 to address low rank matrix approximation. Note the properties of ϕ in Lemma 2.2 imply that $\operatorname{prox}_\phi 0 = 0$. Therefore, firmly nonexpansive operators of the form (2.11) cannot increase the rank of a matrix.

Example 3.13. Let $\mathcal{G}$ be the real Hilbert space of $N \times M$ matrices under the Frobenius norm, set $s = \min\{N, M\}$, and let us denote the singular value decomposition of $y \in \mathcal{G}$ by $y = U_y \operatorname{diag}(\sigma_1(y), \ldots, \sigma_s(y)) V_y^\top$. Let $\rho \in]0, +\infty[$, let hard_ρ be given by (3.12), set $\mathsf{S}\colon \mathbb{R} \to \mathbb{R}\colon \eta \mapsto \eta - \rho \operatorname{sign}(\eta)$, and set

$$\begin{cases} Q\colon \mathcal{G} \to \mathcal{G}\colon y \mapsto U_y \operatorname{diag}\Big(\operatorname{hard}_\rho\big(\sigma_1(y)\big), \ldots, \operatorname{hard}_\rho\big(\sigma_s(y)\big)\Big) V_y^\top \\ S\colon \mathcal{G} \to \mathcal{G}\colon y \mapsto U_y \operatorname{diag}\Big(\mathsf{S}\big(\sigma_1(y)\big), \ldots, \mathsf{S}\big(\sigma_s(y)\big)\Big) V_y^\top. \end{cases} \tag{3.14}$$

Let $q \in \operatorname{ran} Q$, and set $F = S \circ Q$ and $p = Sq$. Since $\operatorname{soft}_\rho = \operatorname{prox}_{\rho|\cdot|}$ and $\rho|\cdot|$ is even, it follows from Example 3.12 and Lemma 2.2 that (F, p) is a proxification of (Q, q). The operator Q is used in image compression to produce low rank approximations [3, 36, 44, 59], and the associated firmly nonexpansive operator F soft-thresholds singular values at level ρ.

Remark 3.14. In the setting of Example 3.13, consider the compression technique performed by the nonconvex projection operator $R\colon \mathcal{G} \to \mathcal{G}$ [13], which truncates singular values at a given rank $r \in \{1, \ldots, s-1\}$, i.e., $R\colon y \mapsto U_y \operatorname{diag}\big(\sigma_1(y), \ldots, \sigma_r(y), 0, \ldots, 0\big) V_y^\top$. Let $y \in \mathcal{G}$ and set $q = Ry$. Then, for every $\rho \in]\sigma_{r+1}(y), \sigma_r(y)[$, $Qy = q$. Therefore, knowledge of the low rank approximation q to y can be exploited in our framework by proxifying (Q, q) using Example 3.13. Note that ρ can be estimated from q since one has access to $\sigma_r(q) = \sigma_r(y)$.

Our last example illustrates how proxification can be used to handle a prescription arising from an extension of the notion of a proximity operator for nonconvex functions.

Example 3.15. Let $\mu \in]0, +\infty[$, let $\gamma \in]0, 1/\mu[$, set $\beta = 1 - \gamma\mu$, and let $g\colon \mathcal{G} \to]-\infty, +\infty]$ be proper, lower semicontinuous, and μ-*weakly convex* in the sense that $g + \mu\|\cdot\|^2/2$ is convex. For every $y \in \mathcal{G}$, $g + \|y - \cdot\|^2/(2\gamma)$ is a strongly convex function in $\Gamma_0(\mathcal{G})$ and, by [6, Corollary 11.17], it therefore admits a unique minimizer $Q_{\gamma g}y$, which defines the operator $Q_{\gamma g}\colon \mathcal{G} \to \mathcal{G}$. Now let $q \in \operatorname{ran} Q_{\gamma g}$ and set $A = \partial(g + \mu\|\cdot\|^2/2)$, $B = A - \mu\mathrm{Id}$, $F = \beta Q_{\gamma g}$, and $p = \beta q$. Then A is maximally monotone but in general, since g is not convex, $Q_{\gamma g}$ is not firmly nonexpansive. However,

$$
\begin{aligned}
(\forall (y,p) \in \mathcal{G} \times \mathcal{G}) \quad Q_{\gamma g}y = p &\Leftrightarrow p \in \operatorname{zer}\Big(\partial\Big(\gamma g + \frac{\gamma\mu}{2}\|\cdot\|^2 - \frac{\gamma\mu}{2}\|\cdot\|^2 + \frac{1}{2}\|y - \cdot\|^2\Big)\Big) \\
&\Leftrightarrow p \in \operatorname{zer}(\gamma A + \beta\mathrm{Id} - y) = \operatorname{zer}(\mathrm{Id} + \gamma B - y) \\
&\Leftrightarrow J_{\gamma B}y = p,
\end{aligned} \tag{3.15}
$$

so Lemma 2.1 implies that $Q_{\gamma g} = J_{\gamma B}$ is β-cocoercive. Thus, (F, p) is a proxification of $(Q_{\gamma g}, q)$. Operators of the form $Q_{\gamma g}$ are used for shrinkage in [7, 38, 50] in the same spirit as in Example 3.6. For instance, for $\mathcal{G} = \mathbb{R}$ and $\rho \in]0, +\infty[$, the penalty $g = \ln(\rho + |\cdot|)$ of [38, 50] is ρ^{-2}-weakly convex and yields

$$
Q_{\gamma g}\colon y \mapsto \begin{cases} \dfrac{1}{2}\big(y - \rho + \sqrt{|y+\rho|^2 - 4\gamma}\big), & \text{if } \ y > \dfrac{\gamma}{\rho}; \\ 0, & \text{if } \ |y| \leqslant \dfrac{\gamma}{\rho}; \\ \dfrac{1}{2}\big(y + \rho - \sqrt{|y-\rho|^2 - 4\gamma}\big), & \text{if } \ y < -\dfrac{\gamma}{\rho}. \end{cases} \tag{3.16}
$$

3.5. Operators arising from monotone equilibria. The property that the object of interest is a zero of the sum of two monotone operators can be modeled in our framework as follows.

Example 3.16. Let $A\colon \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone, let $\beta \in]0, +\infty[$, and let $B\colon \mathcal{G} \to \mathcal{G}$ be β-cocoercive. Let $\gamma \in]0, 2\beta[$ and set

$$
F = \left(1 - \frac{\gamma}{4\beta}\right)\big(\mathrm{Id} - J_{\gamma A} \circ (\mathrm{Id} - \gamma B)\big) \quad \text{and} \quad p = 0. \tag{3.17}
$$

Then F is firmly nonexpansive and, for every $y \in \mathcal{G}$, $Fy = p \Leftrightarrow y \in \operatorname{zer}(A + B)$. Indeed, set $R = J_{\gamma A} \circ (\mathrm{Id} - \gamma B)$. By [6, Proposition 26.1(iv)], R is $(2 - \gamma/2\beta)^{-1}$-averaged and $\operatorname{zer} F = \operatorname{Fix} R = \operatorname{zer}(A + B)$. It follows from [6, Proposition 4.39] that $\mathrm{Id} - R$ is $(1 - \gamma/(4\beta))$-cocoercive, which makes F firmly nonexpansive.

Example 3.17. Let $f \in \Gamma_0(\mathcal{G})$, let $\beta \in]0, +\infty[$, and let $g\colon \mathcal{G} \to \mathbb{R}$ be a convex and differentiable function such that ∇g is β^{-1}-Lipschitzian. Consider the task of enforcing the property

$$
y \in \operatorname{Argmin}(f + g). \tag{3.18}
$$

Set $A = \partial f$ and $B = \nabla g$. Then B is β-cocoercive [6, Corollary 18.17], and (3.18) holds if and only if $y \in \operatorname{zer}(A + B)$. Therefore, Example 3.16 yields a proximal point representation (F, p) of (3.18).

4. Analysis and numerical solution of Problem 1.3. We first show that Problem 1.3 is an appropriate relaxation of Problem 1.1.

Proposition 4.1. *Suppose that the set of solutions to Problem 1.1 is nonempty. Then it coincides with the set of solutions to Problem 1.3.*

Proof. Let $\overline{x}$ be a solution to Problem 1.1. Then it is clear that $\overline{x}$ solves Problem 1.3. Now let x be a solution to Problem 1.3. Then $x \in C$ and

$$(\forall y \in C) \quad \sum_{i\in I} \omega_i \langle L_i(x-y) \mid F_i(L_i x) - p_i\rangle \leqslant 0. \tag{4.1}$$

Therefore, since $\overline{x} \in C$ and, for every $i \in I$, $F_i(L_i\overline{x}) = p_i$, we obtain

$$\sum_{i\in I} \omega_i \langle L_i x - L_i\overline{x} \mid F_i(L_i x) - F_i(L_i\overline{x})\rangle \leqslant 0 \tag{4.2}$$

and, by firm nonexpansiveness of the operators $(F_i)_{i\in I}$,

$$\sum_{i\in I} \omega_i \|F_i(L_i x) - F_i(L_i\overline{x})\|^2 \leqslant \sum_{i\in I} \omega_i \langle L_i x - L_i\overline{x} \mid F_i(L_i x) - F_i(L_i\overline{x})\rangle \leqslant 0. \tag{4.3}$$

We conclude that $(\forall i \in I)\ F_i(L_i x) = F_i(L_i\overline{x}) = p_i$. ∎

Remark 4.2. Consider the setting of Problem 1.3 and set $\boldsymbol{\mathcal{G}} = \bigoplus_{i\in I} \mathcal{G}_i$, $\boldsymbol{L}\colon \mathcal{H} \to \colon \boldsymbol{\mathcal{G}}\colon x \mapsto (L_i x)_{i\in I}$, $\boldsymbol{F}\colon \boldsymbol{\mathcal{G}} \to \boldsymbol{\mathcal{G}}\colon (y_i)_{i\in I} \mapsto (F_i y_i)_{i\in I}$, and $\boldsymbol{p} = (p_i)_{i\in I}$. Note that

$$\text{Problem 1.1 admits a solution if and only if } \boldsymbol{p} \in \boldsymbol{F}\big(\boldsymbol{L}(C)\big). \tag{4.4}$$

Thus, the quantity $d_{\boldsymbol{F}(\boldsymbol{L}(C))}(\boldsymbol{p})$ provides a measure of inconsistency of Problem 1.1. We can actually use a solution to Problem 1.3 to estimate it. Indeed, suppose that $\overline{x}_1$ and $\overline{x}_2$ are solutions to (1.10). Then (1.3) yields

$$\begin{aligned}
\sum_{i\in I} \omega_i \|F_i(L_i\overline{x}_1) - F_i(L_i\overline{x}_2)\|^2 &\leqslant \sum_{i\in I} \omega_i \langle L_i\overline{x}_1 - L_i\overline{x}_2 \mid F_i(L_i\overline{x}_1) - F_i(L_i\overline{x}_2)\rangle \\
&= \sum_{i\in I} \omega_i \langle L_i(\overline{x}_1 - \overline{x}_2) \mid F_i(L_i\overline{x}_1) - p_i\rangle \\
&\quad + \sum_{i\in I} \omega_i \langle L_i(\overline{x}_2 - \overline{x}_1) \mid F_i(L_i\overline{x}_2) - p_i\rangle \\
&\leqslant 0.
\end{aligned} \tag{4.5}$$

Hence, for every $i \in I$, there exists a unique $\overline{p}_i \in \mathcal{G}_i$ such that every solution $\overline{x}$ to Problem 1.3 satisfies

$$F_i(L_i\overline{x}) = \overline{p}_i. \tag{4.6}$$

In turn, if $\overline{x}$ is any solution to Problem 1.3, then

$$d_{\boldsymbol{F}(\boldsymbol{L}(C))}(\boldsymbol{p}) = \inf_{x\in C} \|\boldsymbol{p} - \boldsymbol{F}(\boldsymbol{L}x)\| \leqslant \|\boldsymbol{p} - \boldsymbol{F}(\boldsymbol{L}\overline{x})\| = \sqrt{\sum_{i\in I} \|p_i - \overline{p}_i\|^2}. \tag{4.7}$$

Next, we turn to the existence of solutions.

Proposition 4.3. *Problem 1.3 admits a solution in each of the following instances:*

(i) $\sum_{i\in I}\omega_i L_i^* p_i \in \operatorname{ran}(N_C + \sum_{i\in I}\omega_i L_i^* \circ F_i \circ L_i)$.
(ii) C *is bounded.*
(iii) $\operatorname{ran} N_C + \sum_{i\in I}\omega_i L_i^*(\operatorname{ran} F_i) = \mathcal{H}$.
(iv) *For some* $i \in I$, L_i^* *is surjective and one of the following holds:*
  a) $L_i^*(\operatorname{ran} F_i) = \mathcal{H}$.
  b) F_i *is surjective.*
  c) $\|F_i(y)\| \to +\infty$ *as* $\|y\| \to +\infty$.
  d) $\operatorname{ran}(\operatorname{Id} - F_i)$ *is bounded.*
  e) *There exists a continuous convex function* $g_i\colon \mathcal{G}_i \to \mathbb{R}$ *such that* $F_i = \operatorname{prox}_{g_i}$.

Proof. Set $A = N_C$ and $(\forall i \in I)\ B_i = \omega_i F_i$. Then the operators $(B_i)_{i\in I}$ are cocoercive. Now define

$$M = A + \sum_{i\in I} L_i^* \circ B_i \circ L_i. \tag{4.8}$$

It follows from [6, Proposition 4.12] that $B = \sum_{i\in I} L_i^* \circ B_i \circ L_i$ is cocoercive and hence maximally monotone by [6, Example 20.31], with $\operatorname{dom} B = \mathcal{H}$. On the other hand, [6, Example 20.26] asserts that A is maximally monotone. We therefore derive from [6, Corollary 25.5(i)] that

$$M \text{ is maximally monotone.} \tag{4.9}$$

(i): Let $x \in \mathcal{H}$. In view of (2.9), x solves Problem 1.3 if and only if

$$-\sum_{i\in I}\omega_i L_i^*\big(F_i(L_i x) - p_i\big) \in N_C x, \tag{4.10}$$

that is, $\sum_{i\in I}\omega_i L_i^* p_i \in Mx$.

(ii): Since $\operatorname{dom} M = \operatorname{dom} A = C$ is bounded, it follows from (4.9) and [6, Corollary 21.25] that M is surjective, so (i) holds.

(iii): It follows from [6, Example 25.14] that A is 3^* monotone and from [6, Example 25.20(i)] that the operators $(B_i)_{i\in I}$ are likewise. Hence, in view of (4.9) we invoke Lemma 2.5 to get

$$\operatorname{int}\operatorname{ran} M = \operatorname{int}\operatorname{ran}\Big(A + \sum_{i\in I} L_i^* \circ B_i \circ L_i\Big) = \operatorname{int}\Big(\operatorname{ran} A + \sum_{i\in I} L_i^*(\operatorname{ran} B_i)\Big) = \mathcal{H}. \tag{4.11}$$

So M is surjective and (i) holds.

(iv)b)⇒(iv)a)⇒(iii): Clear.

(iv)c)⇒(iv)b): Since F_i is maximally monotone by [6, Example 20.30], this follows from [6, Corollary 21.24].

(iv)d)⇒(iv)c): Set $\rho = \sup_{y\in\mathcal{G}_i}\|y - F_i y\|$. Then $\|F_i y\| \geqslant \|y\| - \|y - F_i y\| \geqslant \|y\| - \rho \to +\infty$ as $\|y\| \to +\infty$.

(iv)e)⇒(iv)b): We derive from [6, Proposition 16.27] that $\mathcal{G}_i = \operatorname{int}\operatorname{dom} g_i \subset \operatorname{dom}\partial g_i = \operatorname{dom}(\operatorname{Id} + \partial g_i) = \operatorname{ran}(\operatorname{Id} + \partial g_i)^{-1} = \operatorname{ran}\operatorname{prox}_{g_i}$. ∎

Example 4.4. A simple instance when Problem 1.1 has no solution, while the relaxed Problem 1.3 does, is the following. Take disjoint nonempty closed convex subsets C and D of $\mathcal{H}$ such that C is bounded, and let $I = \{1\}$, $\mathcal{G}_1 = \mathcal{H}$, $L_1 = \mathrm{Id}$, $F_1 = \mathrm{Id} - \mathrm{proj}_D$, and $p_1 = 0$. Then the solution set of Problem 1.1 is $C \cap D = \varnothing$, while that of Problem 1.3 is $\mathrm{Fix}\,(\mathrm{proj}_C \circ \mathrm{proj}_D) \neq \varnothing$ [33].

We have described in Example 1.2 an instance of the relaxed Problem 1.3 which is in fact a minimization problem. The next proposition describes a general setting in which a minimization problem underlies Problem 1.3. It involves the Moreau envelope of (2.7).

Proposition 4.5. *Consider the setting of Problem 1.3 and suppose that, for every $i \in I$, there exists $g_i \in \Gamma_0(\mathcal{G}_i)$ such that $F_i = \mathrm{prox}_{g_i}$. Then the objective of Problem 1.3 is to*

$$\underset{x\in C}{\text{minimize}}\;\; f(x), \quad \text{where} \quad f\colon x \mapsto \sum_{i\in I} \omega_i\Big(\widetilde{g_i^*}(L_i x) - \langle L_i x \mid p_i\rangle\Big). \tag{4.12}$$

Proof. We derive from [6, Proposition 24.4] that $(\forall i \in I)\ \nabla\widetilde{g_i^*} = \mathrm{prox}_{g_i}$. In turn, f is differentiable and

$$(\forall x \in \mathcal{H}) \quad \nabla f(x) = \sum_{i\in I} \omega_i L_i^*\big(\mathrm{prox}_{g_i}(L_i x) - p_i\big) = \sum_{i\in I} \omega_i L_i^*\big(F_i(L_i x) - p_i\big). \tag{4.13}$$

Consequently, (1.10) is equivalent to finding a solution to (2.13), i.e., by Lemma 2.4, to minimizing f over C. ∎

Next, we present a block-iterative algorithm for solving Problem 1.3.

Proposition 4.6. *Consider the setting of Problem 1.3 under the assumption that it has a solution. Let K be a strictly positive integer and let $(I_n)_{n\in\mathbb{N}}$ be a sequence of nonempty subsets of I such that*

$$(\forall n \in \mathbb{N}) \quad \bigcup_{k=0}^{K-1} I_{n+k} = I. \tag{4.14}$$

Let $x_0 \in \mathcal{H}$, let $\gamma \in\,]0, 2[$, and, for every $i \in I$, let $t_{i,-1} \in \mathcal{H}$ and set $\gamma_i = \gamma/\|L_i\|^2$. Iterate

$$\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\quad \text{for every } i \in I_n \\
\quad\quad t_{i,n} = x_n - \gamma_i L_i^*\big(F_i(L_i x_n) - p_i\big) \\
\quad \text{for every } i \in I \smallsetminus I_n \\
\quad\quad t_{i,n} = t_{i,n-1} \\
\quad x_{n+1} = \mathrm{proj}_C\left(\displaystyle\sum_{i=1}^{m} \omega_i t_{i,n}\right).
\end{array} \tag{4.15}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a solution to Problem 1.3.

Proof. Set $A_0 = N_C$ and $(\forall i \in I)\ A_i = \|L_i\|^{-2}(L_i^* \circ F_i \circ L_i - L_i^* p_i)$. For every $i \in I$, since F_i is firmly nonexpansive, it follows from [6, Proposition 4.12] that A_i is firmly nonexpansive, i.e., cocoercive with $\beta_i = 1$. Thus, (4.15) is a special case of (2.18), and the conclusion follows from Proposition 2.7. ∎

An attractive feature of (4.15) is its ability to activate only a subblock of operators $(F_i)_{i\in I_n}$ at iteration n, as opposed to all of them as in classical algorithms dealing with inconsistent common fixed point problems [16, 17, 18, 20, 22]. This flexibility is of the utmost relevance for very-large-scale applications. It will also be seen in Section 5 to lead to more efficient implementations. Condition (4.14) regulates the frequency of activation of the operators. Since K can be chosen arbitrarily, it is actually quite mild.

5. Numerical experiments.

In this section, we illustrate the ability of the proposed framework to efficiently model and solve various signal and image recovery problems with inconsistent nonlinear prescriptions. Each instance will use the block-iterative algorithm (4.15) which was shown in Proposition 4.6 to produce an exact solution of Problem 1.3 from any initial point in $\mathcal{H}$. Here, we implement it with $x_0 = 0$.

Remark 5.1. In the modeling of signal construction problems as minimization problems, it is common practice to add a function g to the objective in order to promote desirable properties in the solutions. Several functions are thus averaged and contribute collectively to defining solutions. A prominent example is the promotion of sparsity through the addition of a penalty such as the ℓ^1 norm in $\mathbb{R}^N$ [14, 57]. In the more general variational inequality setting of Problem 1.3, this can be mimicked by adding the prescription $Fy = 0$, where $F = \mathrm{Id} - \mathrm{prox}_g$, i.e., by Moreau's decomposition, $F = \mathrm{prox}_{g^*}$ [6, Remark 14.4]. Note that exact satisfaction of the equality $Fy = 0$ would just mean that one minimizes g since $\operatorname{Fix}\mathrm{prox}_g = \operatorname{Argmin} g$. In general, when incorporated to Problem 1.3, the pair $(F, p) = (\mathrm{Id} - \mathrm{prox}_g, 0)$ is intended to promote the properties g would in a standard minimization problem. We investigate in Sections 5.3 and 5.4 this technique to encourage sparsity in $\mathbb{R}^N$ through the incorporation of the operator $F = \mathrm{proj}_{B_\infty(0;\rho)} = \mathrm{Id} - \mathrm{prox}_{\rho\|\cdot\|_1}$, where $B_\infty(0;\rho)$ is the ℓ^∞ ball of $\mathbb{R}^N$ centered at the origin and with radius $\rho \in \left]0, +\infty\right[$.

5.1. Image recovery from phase.

The goal is to recover the original image $\overline{x} \in \mathcal{H} = \mathbb{R}^N$ ($N = 256^2$) shown in Figure 5.1(a) from the following:

- Bounds on pixel values: $\overline{x} \in C = [0, 255]^N$.
- The degraded image $p_1 \in \mathcal{G}_1 = \mathcal{H}$ shown in Figure 5.1(b), which is modeled as follows. The image $\overline{x}$ is blurred by $L_1 \colon \mathcal{H} \to \mathcal{G}_1$, which performs discrete convolution with a 15×15 Gaussian kernel with standard deviation of 3.5, then corrupted by an additive noise $w_1 \in \mathcal{G}_1$. The blurred image-to-noise ratio is $20\log_{10}(\|L_1\overline{x}\|/\|w_1\|) = 24.0$ dB. Pixel values beyond 60 are then clipped. Altogether, $p_1 = \mathrm{proj}_{D_1}(L_1\overline{x} + w_1)$, where $D_1 = [0, 60]^N$. This process models a low-quality image acquired by a device which saturates at photon counts beyond a certain threshold. We therefore use $F_1 = \mathrm{proj}_{D_1}$ in (1.10).
- An approximation $\rho_2 = 138$ of the mean pixel value of $\overline{x}$. To enforce this information, following Example 1.2, we set $\mathcal{G}_2 = \mathcal{H}$, $L_2 = \mathrm{Id}$, $p_2 = 0$, and

$$F_2 \colon (\eta_k)_{1\leqslant k\leqslant N} \mapsto \left(\frac{\sum_{k=1}^N \eta_k}{N} - \rho_2\right)\mathbf{1}. \tag{5.1}$$

- The phase $\theta \in [-\pi, \pi]^N$ of the 2-D discrete Fourier transform of a noise-corrupted version of $\overline{x}$, i.e., $\theta = \angle\mathrm{DFT}\,(\overline{x} + w_3)$, where $w_3 \in \mathcal{H}$ yields an image-to-noise ratio

$20\log_{10}(\|\overline{x}\|/\|w_3\|) = 49.0$ dB. To model this information, we set $\mathcal{G}_3 = \mathcal{H}$, $L_3 = \mathrm{Id}$, $p_3 = 0$, and, following Example 1.2, we employ

$$F_3 \colon y \mapsto y - \mathrm{IDFT}\left(\left|\mathrm{DFT}\, y\right| \max\left\{\cos\left(\angle(\mathrm{DFT}\, y) - \theta\right), 0\right\} \exp(\imath\theta)\right). \tag{5.2}$$

Due to the noise present in p_1 and θ, and the inexact estimation ρ_2 of the pixel mean, this instance of Problem 1.1 ($I = \{1, 2, 3\}$) is inconsistent. We thus arrive at the relaxed Problem 1.3 by setting $\omega_1 = \omega_2 = \omega_3 = 1/3$. By Proposition 4.3(ii), since C is bounded, Problem 1.3 is guaranteed to possess a solution. The solution shown in Figure 5.1(c) is computed using algorithm (4.15) with $\gamma = 1.9$ and $(\forall n \in \mathbb{N})\ I_n = I$. This experiment illustrates a nonlinear recovery scenario with inconsistent measurements which nonetheless produces realistic solutions obtained by exploiting all available information.

(a) (b) (c)

Figure 5.1. *Experiment of Section 5.1: (a) Original image $\overline{x}$. (b) Degraded image p_1. (c) Recovered image.*

5.2. Signal recovery. The goal is to recover the original signal $\overline{x} \in \mathcal{H} = C = \mathbb{R}^N$ ($N = 1024$) shown in Figure 5.2(a) from the following:

- A piecewise constant approximation p_1 of $\overline{x}$, given by $p_1 = \mathrm{proj}_{D_1}(\overline{x} + w_1)$, where $w_1 \in \mathcal{G}_1 = \mathcal{H}$ represents noise and D_1 is the subspace of signals in $\mathcal{G}_1$ which are constant by blocks along each of the 16 sets of 64 consecutive indices in $\{1, \ldots, N\}$ (see Figure 5.2(b)). The signal-to-noise ratio is $20\log_{10}(\|\overline{x}\|/\|w_1\|) = -2.3$ dB. We model this observation by setting $L_1 = \mathrm{Id}$ and $F_1 = \mathrm{proj}_{D_1}$.
- A bound $\rho_2 = 0.025$ on the magnitude of the finite differences of $\overline{x}$. To enforce this information, following Example 1.2, we set $\mathcal{G}_2 = \mathbb{R}^{N-1}$, $L_2 \colon \mathcal{H} \to \mathcal{G}_2 \colon (\xi_i)_{1\leqslant i\leqslant N} \mapsto (\xi_{i+1} - \xi_i)_{1\leqslant i\leqslant N-1}$, $p_2 = 0$, and $F_2 = \mathrm{Id} - \mathrm{proj}_{D_2}$, where $D_2 = \left\{y \in \mathcal{G}_2 \mid \|y\|_\infty \leqslant \rho_2\right\}$, that is, using (3.13),

$$F_2 \colon (\eta_k)_{1\leqslant k\leqslant N-1} \mapsto \left(\mathrm{soft}_\gamma\left(\eta_k\right)\right)_{1\leqslant k\leqslant N-1}. \tag{5.3}$$

- A collection of $m = 1200$ noisy thresholded scalar observations $r_3 = (\chi_j)_{j\in J} \in \mathbb{R}^m$ of $\overline{x}$, where $J = \{3, \ldots, m+2\}$. The true data formation model is

$$(\forall j \in J) \quad \chi_j = R(\langle \overline{x} \mid e_j \rangle) + \nu_j, \tag{5.4}$$

where $(e_j)_{j\in J}$ is a dictionary of random vectors in $\mathbb{R}^N$ with zero-mean i.i.d. entries, the noise vector $w_3 = (\nu_j)_{j\in J}$ yields a signal-to-noise ratio of $20\log_{10}(\|r_3\|/\|w_3\|) = 17.8$ dB, and R is the thresholding operator of the type found in [2, 52] ($\rho = 0.05$), namely

$$R\colon \mathbb{R}\to\mathbb{R}\colon \eta \mapsto \begin{cases} \operatorname{sign}(\eta)\sqrt[4]{\eta^4-\rho^4}, & \text{if } |\eta|>\rho; \\ 0, & \text{if } |\eta|\leqslant\rho. \end{cases} \tag{5.5}$$

We assume that R is misspecified and that the presence of noise is unknown, so that the data acquisition process is incorrectly modeled as

$$(\forall j\in J)\quad \chi_j = Q(\langle \overline{x} \mid e_j\rangle), \tag{5.6}$$

where

$$Q\colon \mathbb{R}\to\mathbb{R}\colon \eta \mapsto \begin{cases} \operatorname{sign}(\eta)\sqrt{\eta^2-\rho^2}, & \text{if } |\eta|>\rho; \\ 0, & \text{if } |\eta|\leqslant\rho. \end{cases} \tag{5.7}$$

Note that Q is not Lipschitzian. Nonetheless, with

$$S\colon \mathbb{R}\to\mathbb{R}\colon \eta\mapsto \operatorname{sign}(\eta)\left(\sqrt{\eta^2+\rho^2}-\rho\right), \tag{5.8}$$

it is straightforward to verify that $S\circ Q = \operatorname{soft}_\rho$ and that, for every $j\in J$, $(F_j, p_j) = (\operatorname{soft}_\rho, S\chi_j)$ is a proxification of (Q,χ_j). Also, for every $j\in J$, we set $\mathcal{G}_j = \mathbb{R}$ and $L_j = \langle\cdot \mid e_j\rangle$.

We thus consider the instantiation of Problem 1.3 in which $I = \{1,2\}\cup J$ and, for every $i\in I$, $\omega_i = 1/(\operatorname{card} I)$. Since $(e_j)_{j\in J}$ is overcomplete and, for every $j\in J$, F_j is surjective, it follows that $\mathcal{H} = \big\{\sum_{j\in J}\omega_j\eta_j e_j \mid \eta_j\in\operatorname{ran}F_j\big\} = \sum_{j\in J}\omega_j L_j^*(\operatorname{ran}F_j) \subset \sum_{i\in I}\omega_i L_i^*(\operatorname{ran}F_i)$, so Problem 1.3 is guaranteed to possess a solution by Proposition 4.3(iii). Algorithm (4.15) produces the signal shown in Figure 5.2(c) with $\gamma = 1.9$ and the following activation strategy. At every iteration, F_1 and F_2 are activated, while we partition J into four blocks of 300 elements and cyclically activate one block per iteration, i.e.,

$$(\forall n\in\mathbb{N})(\forall j\in\{0,1,2,3\})\quad I_{4n+j} = \{1,2,3+300j,\dots,2+300(j+1)\}, \tag{5.9}$$

which satisfies condition (4.14) with $K = 4$. The results show that, even when the data are noisy and poorly modeled, Problem 1.3 produces quite robust recoveries. The execution time savings resulting from the use of (5.9) compared to the full activation strategy (i.e., $I_n = I$ for every $n\in\mathbb{N}$) are displayed in Figure 5.3. Note that in very-large-scale scenarios in which all data cannot be simultaneously loaded into memory, activation strategies such as (5.9) make algorithm (4.15) implementable.

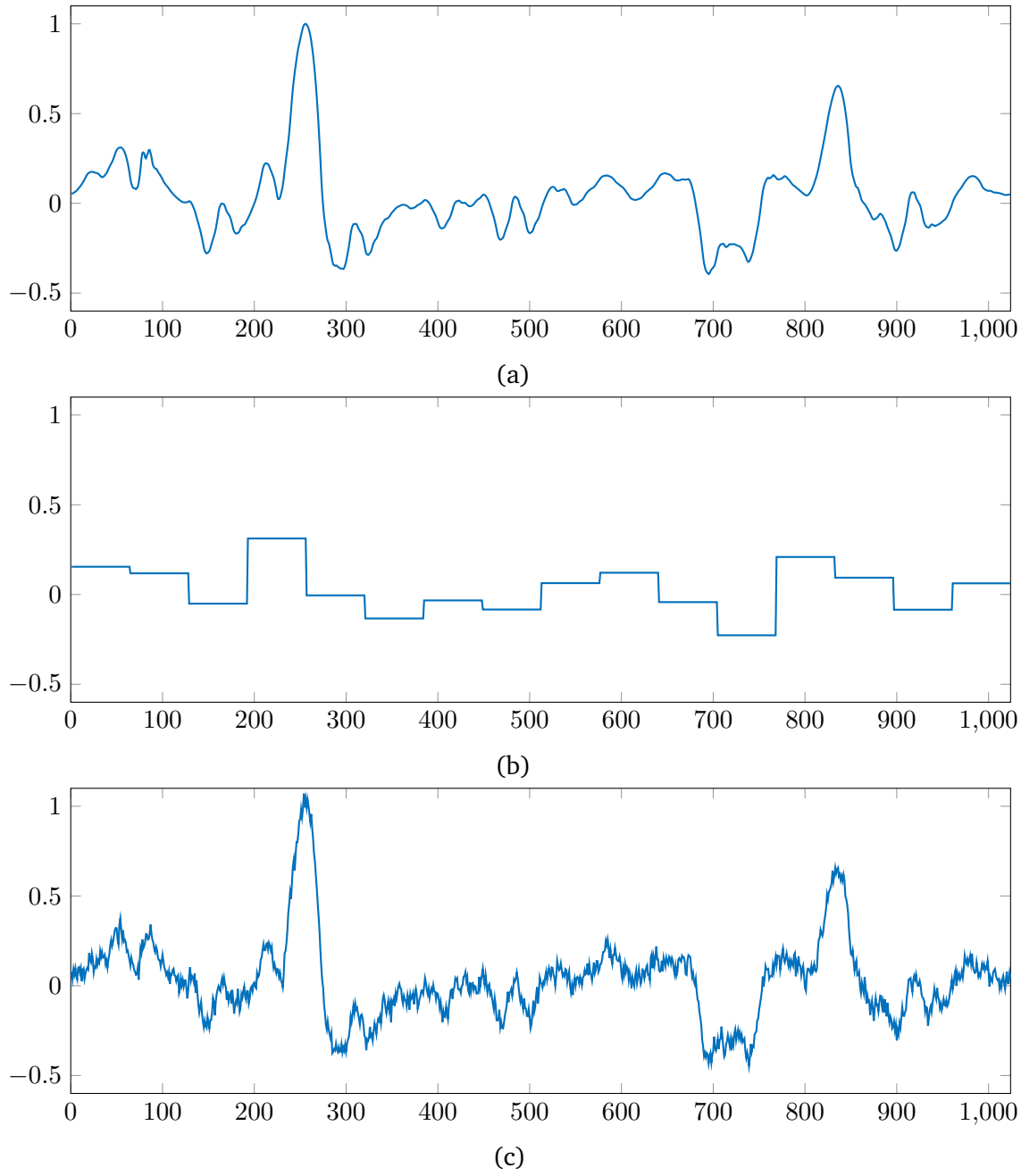

Figure 5.2. *Experiment of Section 5.2: (a): Original signal $\overline{x}$. (b): Piecewise constant approximation p_1. (c): Recovered signal.*

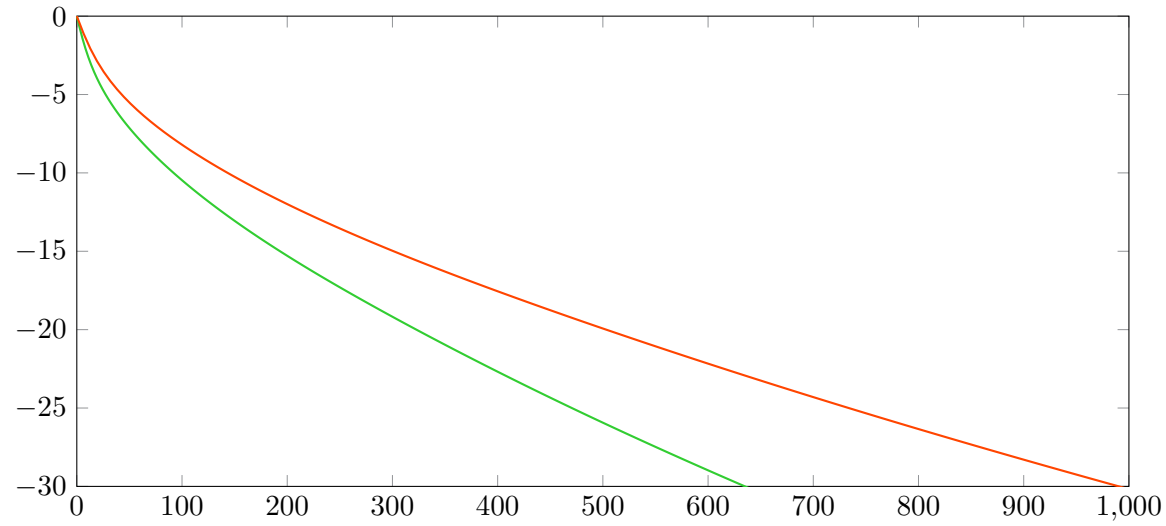

Figure 5.3. *Experiment of Section 5.2: Relative error* $20\log_{10}(\|x_n - x_\infty\|/\|x_0 - x_\infty\|)$ *(dB) versus execution time (seconds) for full activation (red) and cyclic activation (5.9) (green).*

(a) (b) (c)

Figure 5.4. *Experiment of Section 5.3: (a) Original image* $\overline{x}$*. (b) Degraded image* q_1*. (c) Recovered image.*

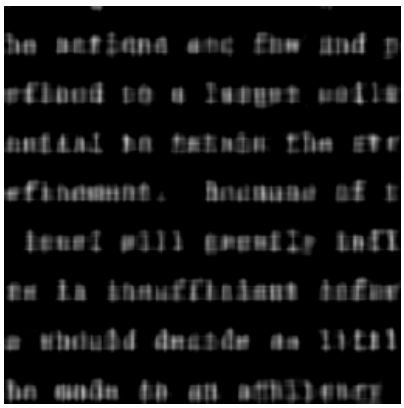

Figure 5.5. *Experiment of Section 5.3: Recovered image with logarithmic thresholding instead of soft thresholding.*

5.3. Sparse image restoration. The goal is to recover the original image $\overline{x} \in \mathcal{H} = \mathbb{R}^N$ ($N = 256^2$) shown in Figure 5.4(a) from the following:

- Bounds on pixel values: $x \in C = [0, 255]^N$.
- The low rank approximation $q_1 \in \mathcal{G}_1 = \mathcal{H}$ displayed in Figure 5.4(b) of a blurred noisy version of $\overline{x}$ modeled as follows. The blurring operator $L_1 \colon \mathcal{H} \to \mathcal{G}_1$ applies a discrete convolution with a uniform 7×7 kernel, and the operators Q and S are as in Example 3.13, with threshold $\rho = 500$. Then $q_1 = Q(L_1\overline{x} + w_1)$ is a rank-85 compression, where $w_1 \in \mathcal{G}_1$ induces a blurred image-to-noise ratio of $20 \log_{10}(\|L_1\overline{x}\|/\|w_1\|) = 17.6$ dB. By Example 3.13, we obtain a proximation of (Q, q_1) with $(F_1, p_1) = (S \circ Q, Sq_1)$.
- $\overline{x}$ is sparse. To promote this property in the solutions to (1.10), following Remark 5.1, we set $\mathcal{G}_2 = \mathcal{H}$, $L_2 = \mathrm{Id}$, $p_2 = 0$, $\rho_2 = 1.5$, and $F_2 = \mathrm{proj}_{B_\infty(0;\rho_2)}$.

We therefore arrive at an instance of Problem 1.3 with $I = \{1, 2\}$ and $\omega_1 = \omega_2 = 1/2$. Since C is bounded, Proposition 4.3(ii) guarantees that a solution exists. Algorithm (4.15) with $\gamma = 1$ yields the recovery in Figure 5.4(c). Even though computing F_1 requires only one singular value decomposition (not two, as (3.14) may suggest), it is the most numerically expensive operator in this problem. Therefore, we choose to activate F_1 only every 5 iterations, i.e.,

$$I_n = \begin{cases} \{2\}, & \text{if } n \not\equiv 0 \mod 5; \\ \{1, 2\}, & \text{if } n \equiv 0 \mod 5. \end{cases} \tag{5.10}$$

Figure 5.6 displays the time savings resulting from the use of (5.10) compared to full activation (both activation strategies yield visually indistinguishable recoveries). Notice that, while the observation in Figure 5.4(b) is virtually illegible, many of the words in the recovery of Figure 5.4(c) can be discerned.

Finally, we examine the use of the non-firmly nonexpansive sparsity-promoting operator of Example 3.15. Specifically, $Q_{\gamma g}$ is given by (3.16), which is induced by the logarithmic penalty with parameters $\rho = \rho_2$ and $\gamma = 0.05/\rho_2^2$. This implies that $0.95Q_{\gamma g}$ is firmly nonexpansive and hence that $\mathrm{Id} - 0.95Q_{\gamma g}$ is likewise. Figure 5.5 displays the result when F_2 is replaced by componentwise applications of $\mathrm{Id} - 0.95Q_{\gamma g}$. In this experiment, the ℓ^1 penalty-based operator

F_2 yields a sharper recovery in Figure 5.4(c) than the recovery in Figure 5.5, which is induced by the logarithmic penalty.

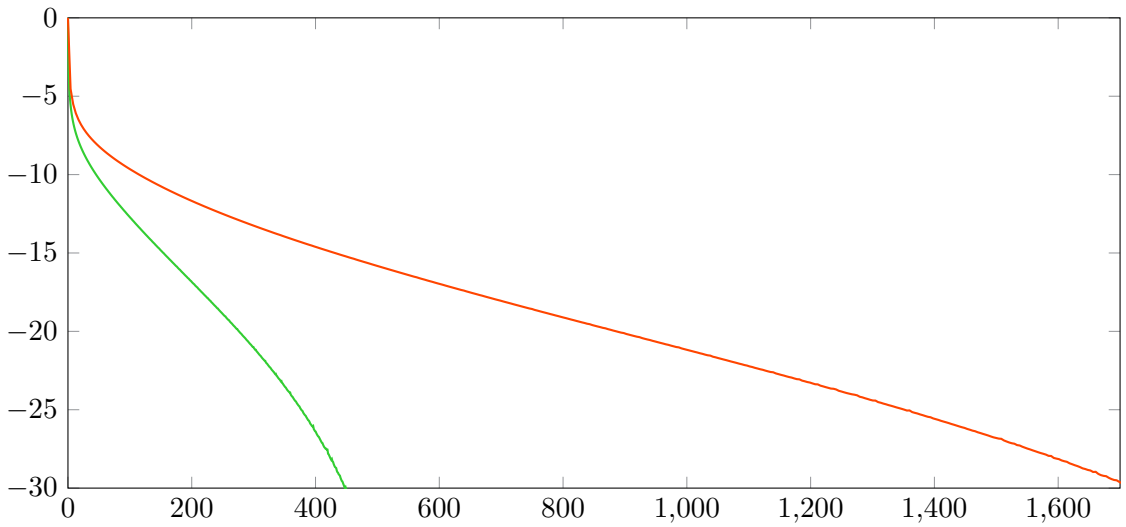

Figure 5.6. *Experiment of Section 5.3: Relative error* $20\log_{10}(\|x_n - x_\infty\|/\|x_0 - x_\infty\|)$ *(dB) versus execution time (seconds) for full activation (red) and block activation* (5.10) *(green).*

5.4. Source separation. This experiment incorporates nonlinear compression to a problem in astronomy, which seeks to separate a background image $\overline{x}_1 \in \mathbb{R}^N$ ($N = 600^2$) of stars from a galaxy image $\overline{x}_2 \in \mathbb{R}^N$ [40]. The goal is to construct the image pair $(\overline{x}_1, \overline{x}_2) \in \mathcal{H} = \mathbb{R}^N \times \mathbb{R}^N$ given the following:

- Bounds on pixel values: $(\overline{x}_1, \overline{x}_2) \in C = [0, 255]^N \times [0, 255]^N$.
- The low rank approximation $q_1 \in \mathcal{G}_1 = \mathbb{R}^N$ shown in Figure 5.7(b) of the original superposition $\overline{x}_1 + \overline{x}_2$ shown in Figure 5.7(a), which is modeled as follows. Set $L_1 \colon \mathcal{H} \to \mathcal{G}_1 \colon (x_1, x_2) \mapsto x_1 + x_2$, and let Q and S be as in Example 3.13 with $\rho = 1500$. The resulting rank-22 approximation of $\overline{x}_1 + \overline{x}_2$ is given by $q_1 = Q(L_1(\overline{x}_1, \overline{x}_2))$. It follows from Example 3.13 that $(F_1, p_1) = (S \circ Q, Sq_1)$ is a proximation of (Q, q_1).
- $\overline{x}_1$ is sparse, and $\overline{x}_2$ admits a sparse representation relative to the 2-D discrete cosine transform $L \colon \mathbb{R}^N \to \mathbb{R}^N$ [40]. To encourage these properties, as discussed in Remark 5.1, we set $\mathcal{G}_2 = \mathcal{H}$, $L_2 \colon (x_1, x_2) \mapsto (x_1, Lx_2)$, $p_2 = 0$, and $F_2 \colon (y_1, y_2) \mapsto (\mathrm{proj}_{B_\infty(0;10)} y_1, \mathrm{proj}_{B_\infty(0;45)} y_2)$. In view of Example 3.1, F_2 is firmly nonexpansive.

Thus, we arrive at an instance of Problem 1.3 with $I = \{1, 2\}$ and $\omega_1 = \omega_2 = 1/2$. By Proposition 4.3(ii) this problem is guaranteed to possess a solution since C is bounded. Algorithm (4.15) with $\gamma = 1$ provides the solution shown in Figure 5.7(c)–(d). To improve algorithmic performance, we adopt the activation strategy (5.10); see Figure 5.8 for time savings compared to the full activation strategy. As can be seen from Figure 5.7, this approach produces effective recoveries. Even though this problem involves a discontinuous observation process, we can nonetheless solve it with algorithm (4.15), which exploits all of the information at hand.

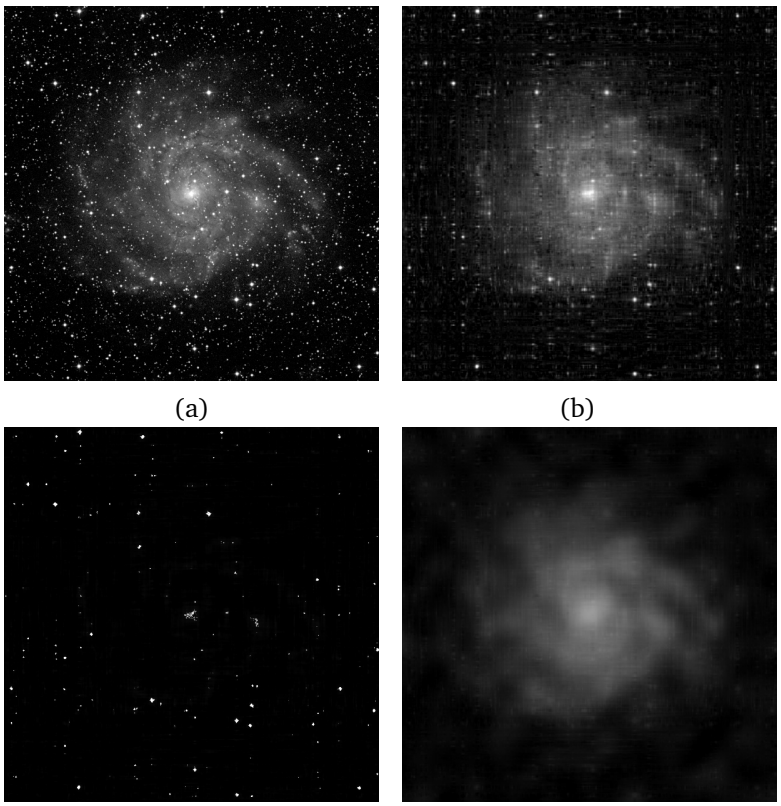

Figure 5.7. *Experiment of Section 5.4: (a) Original image $\overline{x}_1 + \overline{x}_2$. (b) Low-rank compression of $\overline{x}_1 + \overline{x}_2$. (c) Recovered background (stars). (d) Recovered foreground (galaxy).*

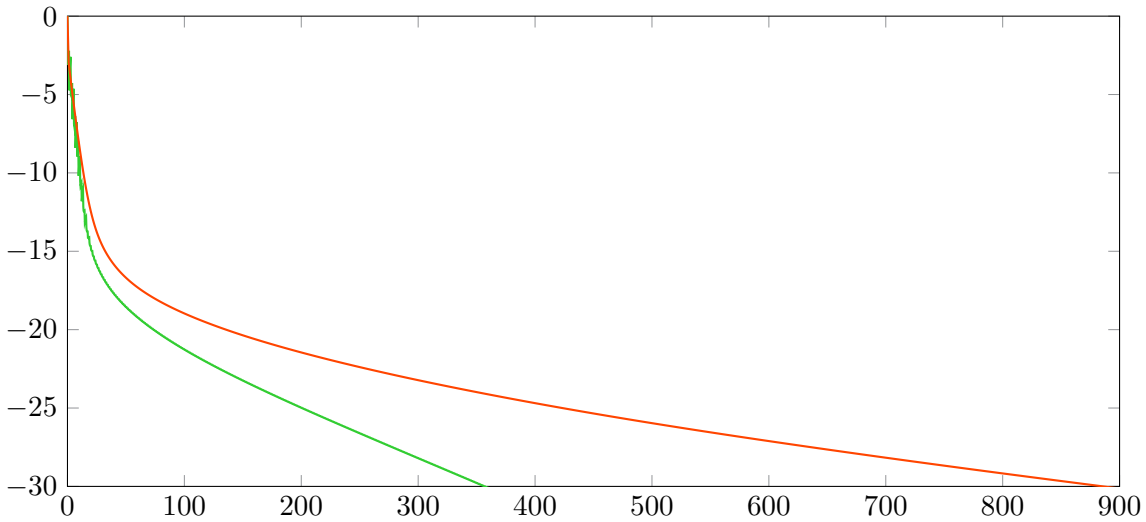

Figure 5.8. *Experiment of Section 5.4: Relative error* $20\log_{10}(\|x_n - x_\infty\|/\|x_0 - x_\infty\|)$ *(dB) versus execution time (seconds) for full activation (red) and block activation* (5.10) *(green).*

REFERENCES

[1] J. S. Abel and J. O. Smith, *Restoring a clipped signal,* Proc. Int. Conf. Acoust. Speech Signal Process., 3 (1991), pp. 1745–1748.

[2] F. Abramovich, T. Sapatinas, and B. W. Silverman, *Wavelet thresholding via a Bayesian approach,* J. R. Stat. Soc. Ser. B Stat. Methodol., 60 (1998), pp. 725–749.

[3] H. Andrews and C. Patterson, *Singular value decomposition (SVD) image decoding,* IEEE Trans. Commun., 24 (1976), pp. 425–432.

[4] F. R. Ávila, M. P. Tcheou, and L. W. P. Biscainho, *Audio soft declipping based on constrained weighted least squares,* IEEE Signal Process. Lett., 24 (2017), pp. 1348–1352.

[5] R. E. Barlow and H. D. Brunk, *The isotonic regression problem and its dual,* J. Amer. Stat. Assoc., 67 (1972), pp. 140–147.

[6] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces,* 2nd ed., Springer, New York, 2017.

[7] I. Bayram and S. Bulek, *A penalty function promoting sparsity within and across groups,* IEEE Trans. Signal Process., 65 (2017), pp. 4238–4251.

[8] A. Bensoussan and J.-L. Lions, *Applications des Inéquations Variationnelles en Contrôle Stochastique,* Bordas, Paris, 1978. English translation: *Applications of Variational Inequalities in Stochastic Control,* North Holland, New York, 1982.

[9] H. Boche, M. Guillemard, G. Kutyniok, and F. Philipp, *Signal recovery from thresholded frame measurements,* Proc. 15th SPIE Wavelets Sparsity Conf., 8858 (2013), pp. 80–86.

[10] H. Brézis and A. Haraux, *Image d'une somme d'opérateurs monotones et applications,* Israel J. Math., 23 (1976), pp. 165–186.

[11] L. M. Briceño-Arias and P. L. Combettes, *Convex variational formulation with smooth coupling for multicomponent signal decomposition and recovery,* Numer. Math. Theory Methods Appl., 2 (2009), pp. 485–508.

[12] F. E. Browder, *Nonlinear monotone operators and convex sets in Banach spaces,* Bull. Amer. Math. Soc., 71 (1965), pp. 780–785.

[13] J. A. Cadzow, *Signal enhancement – A composite property mapping algorithm,* IEEE Trans.

Acoust., Speech, Signal Process., 36 (1988), pp. 49–62.
[14] E. CANDÈS AND T. TAO, *Near-optimal signal recovery from random projections: Universal encoding strategies?* IEEE Trans. Inform. Theory, 52 (2006), pp. 5406–5425.
[15] Y. CENSOR AND T. ELFVING, *A multiprojection algorithm using Bregman projections in a product space,* Numer. Algorithms, 8 (1994), pp. 221–239.
[16] Y. CENSOR, T. ELFVING, N. KOPF, AND T. BORTFELD, *The multiple-sets split feasibility problem and its applications for inverse problems,* Inverse Problems, 21 (2005), pp. 2071–2084.
[17] Y. CENSOR AND M. ZAKNOON, *Algorithms and convergence results of projection methods for inconsistent feasibility problems: A review,* Pure Appl. Funct. Anal., 3 (2018), pp. 565–586.
[18] P. L. COMBETTES, *Inconsistent signal feasibility problems: Least-squares solutions in a product space,* IEEE Trans. Signal Process., 42 (1994), pp. 2955–2966.
[19] P. L. COMBETTES, *The convex feasibility problem in image recovery*, in Advances in Imaging and Electron Physics, P. Hawkes, ed., vol. 95, Academic Press, New York, 1996, pp. 155–270.
[20] P. L. COMBETTES, *Systems of structured monotone inclusions: Duality, algorithms, and applications,* SIAM J. Optim., 23 (2013), pp. 2420–2447.
[21] P. L. COMBETTES, *Monotone operator theory in convex optimization,* Math. Program., B170 (2018), pp. 177–206.
[22] P. L. COMBETTES AND P. BONDON, *Hard-constrained inconsistent signal feasibility problems,* IEEE Trans. Signal Process., 47 (1999), pp. 2460–2468.
[23] P. L. COMBETTES AND L. E. GLAUDIN, *Proximal activation of smooth functions in splitting algorithms for convex image recovery,* SIAM J. Imaging Sci., 12 (2019), pp. 1905–1935.
[24] P. L. COMBETTES AND L. E. GLAUDIN, *Solving composite fixed point problems with block updates,* Adv. Nonlinear Anal., 10 (2021), pp. 1154–1177.
[25] P. L. COMBETTES AND J.-C. PESQUET, *Deep neural network structures solving variational inequalities,* Set-Valued Var. Anal., 28 (2020), pp. 491–518.
[26] P. L. COMBETTES AND Z. C. WOODSTOCK, *A fixed point framework for recovering signals from nonlinear transformations,* Proc. Europ. Signal Process. Conf., pp. 2120–2124, 2020.
[27] P. L. COMBETTES AND Z. C. WOODSTOCK, *Reconstruction of functions from prescribed proximal points,* J. Approx. Theory, 268 (2021), art. 105606.
[28] S. J. DILWORTH, N. J. KALTON, D. KUTZAROVA, AND V. N. TEMLYAKOV, *The thresholding greedy algorithm, greedy bases, and duality,* Constr. Approx., 19 (2003), pp. 575–597.
[29] M. EHRGOTT, Ç. GÜLER, H. M. HAMACHER, AND L. SHAO, *Mathematical optimization in intensity modulated radiation therapy,* Ann. Oper. Res., 175 (2010), pp. 309–365.
[30] F. FACCHINEI F AND J.-S. PANG, *Finite-Dimensional Variational Inequalities and Complementarity Problems,* Springer, New York, 2003.
[31] S. FOUCART AND J. LI, *Sparse recovery from inaccurate saturated measurements,* Acta Appl. Math., 158 (2018), pp. 49–66.
[32] M. GOLDBURG AND R. J. MARKS II, *Signal synthesis in the presence of an inconsistent set of constraints,* IEEE Trans. Circuits Syst., 32 (1985), pp. 647–663.
[33] L. G. GUBIN, B. T. POLYAK, AND E. V. RAIK, *The method of projections for finding the common point of convex sets,* Comput. Math. Math. Phys., 7 (1967), pp. 1–24.
[34] P. HALL, G. KERKYACHARIAN, AND D. PICARD, *Block threshold rules for curve estimation using kernel and wavelet methods,* Ann. Statist., 26 (1998), pp. 922–942.
[35] D. KINDERLEHRER AND G. STAMPACCHIA, *An Introduction to Variational Inequalities and Their Applications,* Academic Press, New York, 1980.
[36] L. KNOCKAERT, B. DE BACKER AND D. DE ZUTTER, *SVD compression, unitary transforms, and computational complexity,* IEEE Trans. Signal Process., 47 (1999), pp. 2724–2729.
[37] A. M. LEGENDRE, *Nouvelles Méthodes pour la Détermination des Orbites des Comètes,* Firmin Didot, Paris, 1805.
[38] D. MALIOUTOV AND A. ARAVKIN, *Iterative log thresholding,* Proc. Int. Conf. Acoust. Speech Signal Process., pp. 7198–7202, 2014.

[39] A. MARMIN, A. JEZIERSKA, M. CASTELLA, AND J.-C. PESQUET, *Global optimization for recovery of clipped signals corrupted with Poisson-Gaussian noise,* IEEE Signal Process. Lett., 27 (2020), pp. 970–974.
[40] M. B. MCCOY, V. CEVHER, Q. T. DINH, A. ASAEI, AND L. BALDASSARRE, *Convexity in source separation: Models, geometry, and algorithms,* IEEE Signal Process. Mag., 31 (2014), pp. 87–95.
[41] K. NASROLLAHI AND T. B. MOESLUND, *Super-resolution: A comprehensive survey,* Mach. Vis. Appl., 25 (2014), pp. 1423–1468.
[42] T. PENNANEN, *On the range of monotone composite mappings,* J. Nonlinear Convex Anal., 2 (2001), pp. 193–202.
[43] B. PETERS, B. R. SMITHYMAN, AND F. J. HERRMANN, *Projection methods and applications for seismic nonlinear inverse problems with multiple constraints,* Geophys., 84 (2019), pp. R251–R269.
[44] A. RANADE, S. S. MAHABALARAO, AND S. KALE, *A variation on SVD based image compression,* Image Vis. Comput., 25 (2007), pp. 771–777.
[45] L. RENCKER, F. BACH, W. WANG, AND M. D. PLUMBLEY, *Sparse recovery and dictionary learning from nonlinear compressive measurements,* IEEE Trans. Signal Process., 67 (2019), pp. 5659–5670.
[46] D. RZEPKA, M. MIŚKOWICZ, D. KOŚCIELNIK, AND N. T. THAO, *Reconstruction of signals from level-crossing samples using implicit information,* IEEE Access, 6 (2018), pp. 35001–35011.
[47] B. E. A. SALEH, *Image synthesis: Discovery instead of recovery*, in: H. Stark (ed.) *Image Recovery: Theory and Application*, pp. 463–498, Academic Press, San Diego, CA, 1987.
[48] R. J. SAMWORTH, *Recent progress in log-concave density estimation,* Statist. Sci., 33 (2018), pp. 493–509.
[49] M. SCHETZEN, *Nonlinear system modeling based on the Wiener theory,* Proc. IEEE, 69 (1981), pp. 1557–1573.
[50] I. SELESNICK AND M. FARSHCHIAN, *Sparse signal approximation via nonseparable regularization,* IEEE Trans. Signal Process., 65 (2017), pp. 2561–2575.
[51] N. T. SHAKED AND J. ROSEN, *Multiple-viewpoint projection holograms synthesized by spatially incoherent correlation with broadband functions,* J. Opt. Soc. Amer. A, 25 (2008), pp. 2129–2138.
[52] T. TAO AND B. VIDAKOVIC, *Almost everywhere behavior of general wavelet shrinkage operators,* Appl. Comput. Harmon. Anal., 9 (2000), pp. 72–82.
[53] E. TARR, *Hack Audio*, Routledge, New York, 2019.
[54] V. N. TEMLYAKOV, *The best m-term approximation and greedy algorithms,* Adv. Comput. Math., 8 (1998), pp. 249–265.
[55] T. TESHIMA, M. XU, I. SATO, AND M. SUGIYAMA, *Clipped matrix completion: A remedy for ceiling effects,* Proc. AAAI Conf. Artif. Intell., pp. 5151–5158, 2019.
[56] N. T. THAO AND M. VETTERLI, *Deterministic analysis of oversampled A/D conversion and decoding improvement based on consistent estimates,* IEEE Trans. Signal Process., 42 (1994), pp. 519–531.
[57] R. TIBSHIRANI, *Regression shrinkage and selection via the lasso: A retrospective,* J. R. Statist. Soc. B, 73 (2011), pp. 273–282.
[58] L. B. WHITE, *The wide-band ambiguity function and Altes' Q-distribution: Constrained synthesis and time-scale filtering,* IEEE Trans. Inform. Theory, 38 (1992), pp. 886–892.
[59] J.-F. YANG AND C.-L. LU, *Combined techniques of singular value decomposition and vector quantization for image coding,* IEEE Trans. Image Process., 4 (1995), pp. 1141–1146.
[60] D. C. YOULA, *Generalized image restoration by the method of alternating orthogonal projections,* IEEE Trans. Circuits Syst., 25 (1978), pp. 694–702.
[61] D. C. YOULA AND V. VELASCO, *Extensions of a result on the synthesis of signals in the presence of inconsistent constraints,* IEEE Trans. Circuits Syst., 33 (1986), pp. 465–468.

[62] D. C. YOULA AND H. WEBB, *Image restoration by the method of convex projections: Part 1 – theory,* IEEE Trans. Med. Imaging, 1 (1982), pp. 81–94.
[63] M. YUAN AND Y. LIN, *Model selection and estimation in regression with grouped variables,* J. R. Stat. Soc. Ser. B Stat. Methodol., 68 (2006), pp. 49–67.
[64] C. A. ZARZER, *On Tikhonov regularization with non-convex sparsity constraints,* Inverse Problems, 15 (2009), art. 025006.
[65] E. ZEIDLER, *Nonlinear Functional Analysis and Its Applications II/B: Nonlinear Monotone Operators,* Springer, New York, 1990.