Fixed Point Strategies in Data Science

Patrick L. Combettes, *Fellow, IEEE*, and Jean-Christophe Pesquet, *Fellow, IEEE*

***Abstract*—The goal of this paper is to promote the use of fixed point strategies in data science by showing that they provide a simplifying and unifying framework to model, analyze, and solve a great variety of problems. They are seen to constitute a natural environment to explain the behavior of advanced convex optimization methods as well as of recent nonlinear methods in data science which are formulated in terms of paradigms that go beyond minimization concepts and involve constructs such as Nash equilibria or monotone inclusions. We review the pertinent tools of fixed point theory and describe the main state-of-the-art algorithms for provenly convergent fixed point construction. We also incorporate additional ingredients such as stochasticity, block-implementations, and non-Euclidean metrics, which provide further enhancements. Applications to signal and image processing, machine learning, statistics, neural networks, and inverse problems are discussed.**

***Index Terms*—Convex optimization, fixed point, game theory, monotone inclusion, image recovery, inverse problems, machine learning, neural networks, nonexpansive operator, signal processing.**

I. Introduction

Attempts to apply mathematical methods to the extraction of information from data can be traced back to the work of Boscovich [34], Gauss [136], Laplace [166], and Legendre [170]. Thus, in connection with the problem of estimating parameters from noisy observations, Boscovich and Laplace invented the least-deviations data fitting method, while Legendre and Gauss invented the least-squares data fitting method. On the algorithmic side, the gradient method was invented by Cauchy [59] to solve a data fitting problem in astronomy, and more or less heuristic methods have been used from then on. The early work involving provenly convergent numerical solutions methods was focused mostly on quadratic minimization problems or linear programming techniques, e.g., [6], [151], [154], [233], [238]. Nowadays, general convex optimization methods have penetrated virtually all branches of data science [11], [54], [67], [82], [100], [140], [219], [228]. In fact, the optimization and data science communities have never been closer, which greatly facilitates technology transfers towards applications. Reciprocally, many of the recent advances in convex optimization algorithms have been motivated by data processing problems in signal recovery, inverse problems, or machine learning. At the same time, the design and the convergence analysis of some of the most potent splitting methods in highly structured or large-scale optimization are based on concepts that are not found in the traditional optimization toolbox but reach deeper into nonlinear analysis. Furthermore, an increasing number of problem formulations go beyond optimization in the sense that their solutions are not optimal in the classical sense of minimizing a function but, rather, satisfy more general notions of equilibria. Among the formulations that fall outside of the realm of standard minimization methods, let us mention variational inequality and monotone inclusion models, game theoretic approaches, neural network structures, and plug-and-play methods.

Given the abundance of activity described above and the increasingly complex formulations of some data processing problems and their solution methods, it is essential to identify general structures and principles in order to simplify and clarify the state of the art. It is the objective of the present paper to promote the viewpoint that fixed point theory constitutes an ideal technology towards this goal. Besides its unifying nature, the fixed point framework offers several advantages. On the algorithmic front, it leads to powerful convergence principles that demystify the design and the asymptotic analysis of iterative methods. Furthermore, fixed point methods can be implemented using stochastic perturbations, as well as block-coordinate or block-iterative strategies which reduce the computational load and memory requirements of the iterations.

Historically, one of the first uses of fixed point theory in signal recovery is found in the bandlimited reconstruction method of [165], which is based on the iterative Banach-Picard contraction process

$$x_{n+1} = Tx_n, \tag{1}$$

where the operator T has Lipschitz constant $\delta < 1$. The importance of dealing with the more general class of nonexpansive operators, i.e., those with Lipschitz constant $\delta = 1$, was emphasized by Youla in [247] and [249]; see also [213], [230], [239]. Since then, many problems in data science have been modeled and solved using nonexpansive operator theory; see for instance [20], [54], [82], [106], [108], [117], [176], [197], [220], [229].

The outline of the paper is as follows. In order to make the paper as self-contained as possible, we present in Section II the essential tools and results from nonlinear analysis on which fixed point approaches are grounded. These include notions of convex analysis, monotone operator theory, and averaged operator theory. Section III provides an overview of basic fixed point principles and methods. Section IV addresses the broad class of monotone inclusion problems and their fixed point modeling. Using the tools of Section III, various splitting strategies are described, as well as block-iterative and block-coordinate algorithms. Section V discusses

The work of P. L. Combettes was supported by the National Science Foundation under grant CCF-1715671. The work of J.-C. Pesquet was supported by Institut Universitaire de France and the ANR Chair in Artificial Intelligence BRIGEABLE.

P. L. Combettes (corresponding author) is with North Carolina State University, Department of Mathematics, Raleigh, NC 27695-8205, USA and J.-C. Pesquet is with Université Paris-Saclay, Inria, CentraleSupélec, Centre de Vision Numérique, 91190 Gif sur Yvette, France. E-mail: plc@math.ncsu.edu, jean-christophe@pesquet.eu.

applications of splitting methods to a large panel of techniques for solving structured convex optimization problems. Moving beyond traditional optimization, algorithms for Nash equilibria are investigated in Section VI. Section VII shows how fixed point strategies can be applied to four additional categories of data science problems that have no underlying minimization interpretation. Some brief conclusions are drawn in Section VIII. For simplicity, we have adopted a Euclidean space setting. However, most results remain valid in general Hilbert spaces up to technical adjustments.

II. Notation and Mathematical Foundations

We review the basic tools and principles from nonlinear analysis that will be used throughout the paper. Unless otherwise stated, the material of this section can be found in [21]; for convex analysis see also [206].

A. Notation

Throughout, $\mathcal{H}$, $\mathcal{G}$, $(\mathcal{H}_i)_{1\leqslant i\leqslant m}$, and $(\mathcal{G}_k)_{1\leqslant k\leqslant q}$ are Euclidean spaces. We denote by $2^{\mathcal{H}}$ the collection of all subsets of $\mathcal{H}$ and by $\boldsymbol{\mathcal{H}}=\mathcal{H}_1\times\cdots\times\mathcal{H}_m$ and $\boldsymbol{\mathcal{G}}=\mathcal{G}_1\times\cdots\times\mathcal{G}_q$ the standard Euclidean product spaces. A generic point in $\boldsymbol{\mathcal{H}}$ is denoted by $\boldsymbol{x}=(x_i)_{1\leqslant i\leqslant m}$. The scalar product of a Euclidean space is denoted by $\langle\cdot\mid\cdot\rangle$ and the associated norm by $\|\cdot\|$. The adjoint of a linear operator L is denoted by L^*. Let C be a subset of $\mathcal{H}$. Then the *distance function* to C is $d_C\colon x\mapsto\inf_{y\in C}\|x-y\|$ and the *relative interior* of C, denoted by $\operatorname{ri}C$, is its interior relative to its affine hull.

B. Convex analysis

The central notion in convex analysis is that of a convex set: a subset C of $\mathcal{H}$ is *convex* if it contains all the line segments with end points in the set, that is,

$$(\forall x\in C)(\forall y\in C)(\forall\alpha\in\,]0,1[\,)\quad \alpha x+(1-\alpha)y\in C. \tag{2}$$

The projection theorem is one of the most important results of convex analysis.

Theorem 1 (projection theorem) *Let C be a nonempty closed convex subset of $\mathcal{H}$ and let $x\in\mathcal{H}$. Then there exists a unique point $\operatorname{proj}_C x\in C$, called the* projection *of x onto C, such that $\|x-\operatorname{proj}_C x\|=d_C(x)$. In addition, for every $p\in\mathcal{H}$,*

$$p=\operatorname{proj}_C x\quad\Leftrightarrow\quad\begin{cases}p\in C\\(\forall y\in C)\;\langle y-p\mid x-p\rangle\leqslant 0.\end{cases} \tag{3}$$

Convexity for functions is inherited from convexity for sets as follows. Consider a function $f\colon\mathcal{H}\to\,]-\infty,+\infty]$. Then f is *convex* if its *epigraph*

$$\operatorname{epi}f=\big\{(x,\xi)\in\mathcal{H}\times\mathbb{R}\;\big|\;f(x)\leqslant\xi\big\} \tag{4}$$

is a convex set. This is equivalent to requiring that

$$(\forall x\in\mathcal{H})(\forall y\in\mathcal{H})(\forall\alpha\in\,]0,1[\,)\quad f\big(\alpha x+(1-\alpha)y\big)\leqslant\alpha f(x)+(1-\alpha)f(y). \tag{5}$$

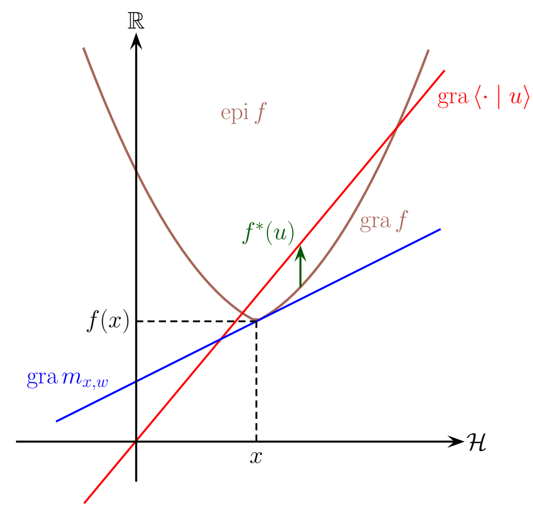

Fig. 1: The graph of a function $f\in\Gamma_0(\mathcal{H})$ is shown in brown. The area above the graph is the closed convex set $\operatorname{epi}f$ of (4). Let $u\in\mathcal{H}$ and let the red line be the graph of the linear function $\langle\cdot\mid u\rangle$. In view of (12), the value of $f^*(u)$ (in green) is the maximum signed difference between the red line and the brown line. Now fix $x\in\mathcal{H}$ and $w\in\partial f(x)$. By (13), the affine function $m_{x,w}\colon y\mapsto\langle y-x\mid w\rangle+f(x)$ satisfies $m_{x,w}\leqslant f$ and it coincides with f at x. Its graph is represented in blue. Every subgradient w gives such an affine minorant.

If $\operatorname{epi}f$ is closed, then f is *lower semicontinuous* in the sense that, for every sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ and $x\in\mathcal{H}$,

$$x_n\to x\quad\Rightarrow\quad f(x)\leqslant\varliminf f(x_n). \tag{6}$$

Finally, we say that $f\colon\mathcal{H}\to\,]-\infty,+\infty]$ is *proper* if $\operatorname{epi}f\neq\varnothing$, which is equivalent to

$$\operatorname{dom}f=\big\{x\in\mathcal{H}\;\big|\;f(x)<+\infty\big\}\neq\varnothing. \tag{7}$$

The class of functions $f\colon\mathcal{H}\to\,]-\infty,+\infty]$ which are proper, lower semicontinuous, and convex is denoted by $\Gamma_0(\mathcal{H})$. The following result is due to Moreau [185].

Theorem 2 (proximation theorem) *Let $f\in\Gamma_0(\mathcal{H})$ and let $x\in\mathcal{H}$. Then there exists a unique point $\operatorname{prox}_f x\in\mathcal{H}$, called the* proximal point *of x relative to f, such that*

$$f\big(\operatorname{prox}_f x\big)+\frac{1}{2}\|x-\operatorname{prox}_f x\|^2=\min_{y\in\mathcal{H}}\bigg(f(y)+\frac{1}{2}\|x-y\|^2\bigg). \tag{8}$$

In addition, for every $p\in\mathcal{H}$,

$$p=\operatorname{prox}_f x\;\Leftrightarrow\;(\forall y\in\mathcal{H})\;\langle y-p\mid x-p\rangle+f(p)\leqslant f(y). \tag{9}$$

The above theorem defines an operator prox_f called the *proximity operator* of f (see [100] for a tutorial, and [21, Chapter 24] and [89] for a detailed account with various properties). Now let C be a nonempty closed convex subset of $\mathcal{H}$. Then its *indicator function* ι_C, defined by

$$\iota_C\colon \mathcal{H} \to \left]-\infty,+\infty\right] : x \mapsto \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{if } x \notin C, \end{cases} \tag{10}$$

lies in $\Gamma_0(\mathcal{H})$ and it follows from (3) and (9) that

$$\mathrm{prox}_{\iota_C} = \mathrm{proj}_C. \tag{11}$$

This shows that Theorem 2 generalizes Theorem 1. Let us now introduce basic convex analytical tools (see Fig. 1). The *conjugate* of $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ is

$$f^*\colon \mathcal{H} \to [-\infty,+\infty] : u \mapsto \sup_{x\in\mathcal{H}} \big(\langle x \mid u\rangle - f(x)\big). \tag{12}$$

The *subdifferential* of a proper function $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ is the set-valued operator $\partial f\colon \mathcal{H} \to 2^{\mathcal{H}}$ which maps a point $x \in \mathcal{H}$ to the set (see Fig. 2)

$$\partial f(x) = \big\{u \in \mathcal{H} \;\big|\; (\forall y \in \mathcal{H})\; \langle y - x \mid u\rangle + f(x) \leqslant f(y)\big\}. \tag{13}$$

A vector in $\partial f(x)$ is a *subgradient* of f at x. If C is a nonempty closed convex subset of $\mathcal{H}$, $N_C = \partial\iota_C$ is the *normal cone* operator of C, that is, for every $x \in \mathcal{H}$,

$$N_C x = \begin{cases} \big\{u \in \mathcal{H} \;\big|\; (\forall y \in C)\; \langle y - x \mid u\rangle \leqslant 0\big\}, & \text{if } x \in C; \\ \varnothing, & \text{otherwise.} \end{cases} \tag{14}$$

Let us denote by $\mathrm{Argmin}\, f$ the set of minimizers of a function $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ (the notation $\mathrm{Argmin}_{x\in\mathcal{H}} f(x)$ will also be used). The most fundamental result in optimization is actually the following immediate consequence of (13).

Theorem 3 (Fermat's rule) *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be a proper function. Then* $\mathrm{Argmin}\, f = \big\{x \in \mathcal{H} \;\big|\; 0 \in \partial f(x)\big\}$.

Theorem 4 (Moreau) *Let $f \in \Gamma_0(\mathcal{H})$. Then $f^* \in \Gamma_0(\mathcal{H})$, $f^{**} = f$, and* $\mathrm{prox}_f + \mathrm{prox}_{f^*} = \mathrm{Id}$.

A function $f \in \Gamma_0(\mathcal{H})$ is *differentiable* at $x \in \mathrm{dom}\, f$ if there exists a vector $\nabla f(x) \in \mathcal{H}$, called the *gradient* of f at x, such that

$$(\forall y \in \mathcal{H}) \quad \lim_{\alpha\downarrow 0} \frac{f(x+\alpha y) - f(x)}{\alpha} = \langle y \mid \nabla f(x)\rangle. \tag{15}$$

Example 5 Let C be a nonempty closed convex subset of $\mathcal{H}$. Then $\nabla d_C^2/2 = \mathrm{Id} - \mathrm{proj}_C$.

Proposition 6 *Let $f \in \Gamma_0(\mathcal{H})$, let $x \in \mathrm{dom}\, f$, and suppose that f is differentiable at x. Then $\partial f(x) = \{\nabla f(x)\}$.*

We close this section by examining fundamental properties of a canonical convex minimization problem.

Proposition 7 *Let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{G})$, and let $L\colon \mathcal{H} \to \mathcal{G}$ be linear. Suppose that $L(\mathrm{dom}\, f) \cap \mathrm{dom}\, g \neq \varnothing$ and set $S = \mathrm{Argmin}\,(f + g \circ L)$. Then the following hold:*

i) *Suppose that $\lim_{\|x\|\to+\infty} f(x) + g(Lx) = +\infty$. Then $S \neq \varnothing$.*
ii) *Suppose that $\mathrm{ri}(L(\mathrm{dom}\, f)) \cap \mathrm{ri}(\mathrm{dom}\, g) \neq \varnothing$. Then*

$$\begin{aligned} S &= \big\{x \in \mathcal{H} \;\big|\; 0 \in \partial f(x) + L^*\big(\partial g(Lx)\big)\big\} \\ &= \big\{x \in \mathcal{H} \;\big|\; (\exists v \in \partial g(Lx)) \; -L^*v \in \partial f(x)\big\}. \end{aligned}$$

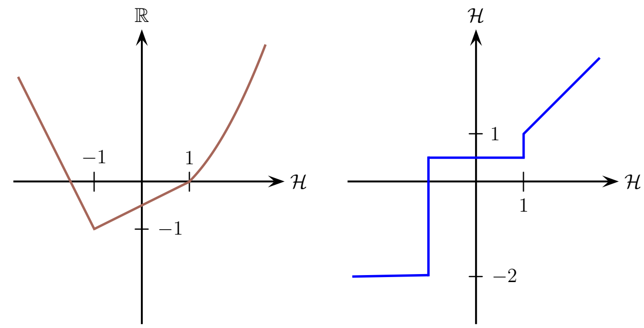

Fig. 2: Left: Graph of a function defined on $\mathcal{H} = \mathbb{R}$. Right: Graph of its subdifferential.

C. Nonexpansive operators

We introduce the main classes of operators pertinent to our discussion. First, we need to define the notion of a relaxation for an operator.

Definition 8 Let $T\colon \mathcal{H} \to \mathcal{H}$ and let $\lambda \in \left]0,+\infty\right[$. Then the operator $R = \mathrm{Id} + \lambda(T - \mathrm{Id})$ is a *relaxation* of T. If $\lambda \leqslant 1$, then R is an *underrelaxation* of T and, if $\lambda \geqslant 1$, R is an *overrelaxation* of T; in particular, if $\lambda = 2$, R is the *reflection* of T.

Definition 9 Let $\alpha \in \left]0,1\right]$. An α-relaxation sequence is a sequence $(\lambda_n)_{n\in\mathbb{N}}$ in $\left]0,1/\alpha\right[$ such that $\sum_{n\in\mathbb{N}} \lambda_n(1-\alpha\lambda_n) = +\infty$.

Example 10 Let $\alpha \in \left]0,1\right]$ and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0,+\infty\right[$. Then $(\lambda_n)_{n\in\mathbb{N}}$ is an α-relaxation sequence in each of the following cases:

i) $\alpha < 1$ and $(\forall n \in \mathbb{N})\; \lambda_n = 1$.
ii) $(\forall n \in \mathbb{N})\; \lambda_n = \lambda \in \left]0,1/\alpha\right[$.
iii) $\inf_{n\in\mathbb{N}} \lambda_n > 0$ and $\sup_{n\in\mathbb{N}} \lambda_n < 1/\alpha$.
iv) There exists $\varepsilon \in \left]0,1\right[$ such that $(\forall n \in \mathbb{N})\; \varepsilon/\sqrt{n+1} \leqslant \lambda_n \leqslant 1/\alpha - \varepsilon/\sqrt{n+1}$.

An operator $T\colon \mathcal{H} \to \mathcal{H}$ is *Lipschitzian* with constant $\delta \in \left]0,+\infty\right[$ if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|Tx - Ty\| \leqslant \delta\|x - y\|. \tag{16}$$

If $\delta < 1$ above, then T is a *Banach contraction* (also called a *strict contraction*). If $\delta = 1$, that is,

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|Tx - Ty\| \leqslant \|x - y\|, \tag{17}$$

then T is *nonexpansive*. On the other hand, T is *cocoercive* with constant $\beta \in \left]0,+\infty\right[$ if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x - y \mid Tx - Ty\rangle \geqslant \beta\|Tx - Ty\|^2. \tag{18}$$

If $\beta = 1$ in (18), then T is *firmly nonexpansive*. Alternatively, T is firmly nonexpansive if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \; \|Tx - Ty\|^2 \leqslant \|x - y\|^2 - \|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2. \tag{19}$$

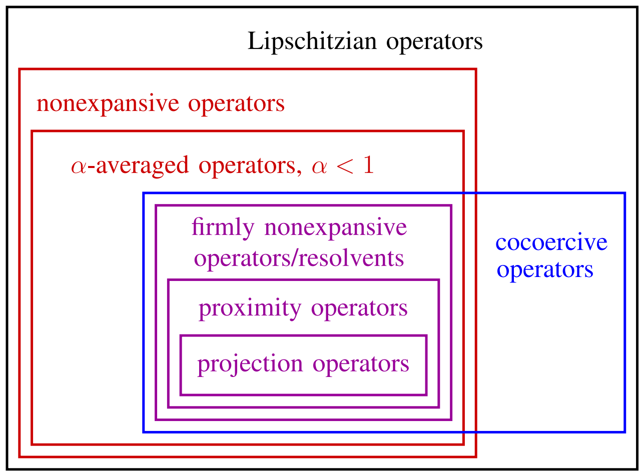

Fig. 3: Classes of nonlinear operators.

Equivalently, T is firmly nonexpansive if the reflection

$$\mathrm{Id} + 2(T - \mathrm{Id}) \ \text{ is nonexpansive.} \tag{20}$$

More generally, let $\alpha \in \left]0,1\right]$. Then T is α-*averaged* if the overrelaxation

$$\mathrm{Id} + \alpha^{-1}(T - \mathrm{Id}) \ \text{ is nonexpansive} \tag{21}$$

or, equivalently, if there exists a nonexpansive operator $Q\colon \mathcal{H} \to \mathcal{H}$ such that T can be written as the underrelaxation

$$T = \mathrm{Id} + \alpha(Q - \mathrm{Id}). \tag{22}$$

An alternative characterization of α-averagedness is

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \ \ \|Tx - Ty\|^2 \leqslant \|x - y\|^2 \\ - \frac{1-\alpha}{\alpha}\|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2. \tag{23}$$

Averaged operators will be the most important class of nonlinear operators we use in this paper. They were introduced in [12] and their central role in many nonlinear analysis algorithms was pointed out in [87], with further refinements in [111], [153]. Note that

$$\begin{aligned} T \text{ is firmly nonexpansive} &\Leftrightarrow \mathrm{Id} - T \text{ is firmly nonexpansive} \\ &\Leftrightarrow T \text{ is } 1/2\text{-averaged} \\ &\Leftrightarrow T \text{ is } 1\text{-cocoercive.} \end{aligned} \tag{24}$$

Here is an immediate consequence of (9) and (24).

Example 11 Let $f \in \Gamma_0(\mathcal{H})$. Then prox_f and $\mathrm{Id} - \mathrm{prox}_f$ are firmly nonexpansive. In particular, if C is a nonempty closed convex subset of $\mathcal{H}$, then (11) implies that proj_C and $\mathrm{Id} - \mathrm{proj}_C$ are firmly nonexpansive.

The relationships between the different types of nonlinear operators discussed so far are depicted in Fig. 3. The next propositions provide further connections between them.

Proposition 12 *Let $\delta \in \left]0,1\right[$, let $T\colon \mathcal{H} \to \mathcal{H}$ be δ-Lipschitzian, and set $\alpha = (\delta+1)/2$. Then T is α-averaged.*

Proposition 13 *Let $T\colon \mathcal{H} \to \mathcal{H}$, let $\beta \in \left]0,+\infty\right[$, and let $\gamma \in \left]0,2\beta\right[$. Then T is β-cocoercive if and only if $\mathrm{Id} - \gamma T$ is $\gamma/(2\beta)$-averaged.*

It follows from the Cauchy-Schwarz inequality that a β-cocoercive operator is β^{-1}-Lipschitzian. In the case of gradients of convex functions, the converse is also true.

Proposition 14 (Baillon-Haddad) *Let $f\colon \mathcal{H} \to \mathbb{R}$ be a differentiable convex function such that ∇f is β^{-1}-Lipschitzian for some $\beta \in \left]0,+\infty\right[$. Then ∇f is β-cocoercive.*

We now describe operations that preserve averagedness and cocoercivity.

Proposition 15 *Let $T\colon \mathcal{H} \to \mathcal{H}$, let $\alpha \in \left]0,1\right[$, and let $\lambda \in \left]0,1/\alpha\right[$. Then T is α-averaged if and only if $(1-\lambda)\mathrm{Id} + \lambda T$ is $\lambda\alpha$-averaged.*

Proposition 16 *For every $i \in \{1,\ldots,m\}$, let $\alpha_i \in \left]0,1\right[$, let $\omega_i \in \left]0,1\right]$, and let $T_i\colon \mathcal{H} \to \mathcal{H}$ be α_i-averaged. Suppose that $\sum_{i=1}^m \omega_i = 1$ and set $\alpha = \sum_{i=1}^m \omega_i\alpha_i$. Then $\sum_{i=1}^m \omega_i T_i$ is α-averaged.*

Example 17 For every $i \in \{1,\ldots,m\}$, let $\omega_i \in \left]0,1\right]$ and let $T_i\colon \mathcal{H} \to \mathcal{H}$ be firmly nonexpansive. Suppose that $\sum_{i=1}^m \omega_i = 1$. Then $\sum_{i=1}^m \omega_i T_i$ is firmly nonexpansive.

Proposition 18 *For every $i \in \{1,\ldots,m\}$, let $\alpha_i \in \left]0,1\right[$ and let $T_i\colon \mathcal{H} \to \mathcal{H}$ be α_i-averaged. Set*

$$T = T_1 \circ \cdots \circ T_m \quad \text{and} \quad \alpha = \cfrac{1}{1 + \cfrac{1}{\displaystyle\sum_{i=1}^m \frac{\alpha_i}{1-\alpha_i}}}. \tag{25}$$

Then T is α-averaged.

Example 19 Let $\alpha_1 \in \left]0,1\right[$, let $\alpha_2 \in \left]0,1\right[$, let $T_1\colon \mathcal{H} \to \mathcal{H}$ be α_1-averaged, and let $T_2\colon \mathcal{H} \to \mathcal{H}$ be α_2-averaged. Set

$$T = T_1 \circ T_2 \quad \text{and} \quad \alpha = \frac{\alpha_1 + \alpha_2 - 2\alpha_1\alpha_2}{1 - \alpha_1\alpha_2}. \tag{26}$$

Then T is α-averaged.

Proposition 20 ([118]) *Let $T_1\colon \mathcal{H} \to \mathcal{H}$ and $T_2\colon \mathcal{H} \to \mathcal{H}$ be firmly nonexpansive, let $\alpha_3 \in \left]0,1\right[$, and let $T_3\colon \mathcal{H} \to \mathcal{H}$ be α_3-averaged. Set $\alpha = 1/(2-\alpha_3)$ and*

$$T = T_1 \circ (T_2 - \mathrm{Id} + T_3 \circ T_2) + \mathrm{Id} - T_2. \tag{27}$$

Then T is α-averaged.

Proposition 21 *For every $k \in \{1,\ldots,q\}$, let $0 \neq L_k\colon \mathcal{H} \to \mathcal{G}_k$ be linear, let $\beta_k \in \left]0,+\infty\right[$, and let $T_k\colon \mathcal{G}_k \to \mathcal{G}_k$ be β_k-cocoercive. Set*

$$T = \sum_{k=1}^q L_k^* \circ T_k \circ L_k \quad \text{and} \quad \beta = \frac{1}{\displaystyle\sum_{k=1}^q \frac{\|L_k\|^2}{\beta_k}}. \tag{28}$$

Then the following hold:

i) *T is β-cocoercive* [21].
ii) *Suppose that $\sum_{k=1}^q \|L_k\|^2 \leqslant 1$ and that the operators $(T_k)_{1\leqslant k\leqslant q}$ are firmly nonexpansive. Then T is firmly nonexpansive* [21].
iii) *Suppose that $\sum_{k=1}^q \|L_k\|^2 \leqslant 1$ and that $(T_k)_{1\leqslant k\leqslant q}$ are proximity operators. Then T is a proximity operator* [89].

Remark 22 The statement of Proposition 21iii) can be made more precise [89]. To wit, for every $k \in \{1,\ldots,q\}$, let $\omega_k \in]0,+\infty[$, let $0 \neq L_k\colon \mathcal{H} \to \mathcal{G}_k$ be linear, let $g_k \in \Gamma_0(\mathcal{G}_k)$, and let $h_k\colon v \mapsto \inf_{w\in\mathcal{G}_k}(g_k^*(w) + \|v-w\|^2/2)$ be the *Moreau envelope* of g_k^*. Then, if $\sum_{k=1}^q \omega_k\|L_k\|^2 \leqslant 1$, we have

$$\sum_{k=1}^{q} \omega_k\big(L_k^* \circ \operatorname{prox}_{g_k} \circ L_k\big) = \operatorname{prox}_f, \quad \text{where}$$
$$f = \left(\sum_{k=1}^{q} \omega_k h_k \circ L_k\right)^* - \frac{\|\cdot\|^2}{2}. \tag{29}$$

Let $T\colon \mathcal{H} \to \mathcal{H}$ and let

$$\operatorname{Fix} T = \big\{x \in \mathcal{H} \;\big|\; Tx = x\big\} \tag{30}$$

be its set of *fixed points*. If T is a Banach contraction, then it admits a unique fixed point. However, if T is merely nonexpansive, the situation is quite different. Indeed, a nonexpansive operator may have no fixed point (take $T\colon x \mapsto x+z$, with $z \neq 0$), exactly one (take $T = -\mathrm{Id}$), or infinitely many (take $T = \mathrm{Id}$). Even those operators which are firmly nonexpansive can fail to have fixed points.

Example 23 $T\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto (x+\sqrt{x^2+4})/2$ is firmly nonexpansive and $\operatorname{Fix} T = \varnothing$.

Proposition 24 *Let $T\colon \mathcal{H} \to \mathcal{H}$ be nonexpansive. Then* $\operatorname{Fix} T$ *is closed and convex.*

Proposition 25 *Let $(T_i)_{1\leqslant i\leqslant m}$ be nonexpansive operators from $\mathcal{H}$ to $\mathcal{H}$, and let $(\omega_i)_{1\leqslant i\leqslant m}$ be real numbers in $]0,1]$ such that $\sum_{i=1}^m \omega_i = 1$. Suppose that $\bigcap_{i=1}^m \operatorname{Fix} T_i \neq \varnothing$. Then* $\operatorname{Fix}\big(\sum_{i=1}^m \omega_i T_i\big) = \bigcap_{i=1}^m \operatorname{Fix} T_i$.

Proposition 26 *For every $i \in \{1,\ldots,m\}$, let $\alpha_i \in]0,1[$ and let $T_i\colon \mathcal{H} \to \mathcal{H}$ be α_i-averaged. Suppose that $\bigcap_{i=1}^m \operatorname{Fix} T_i \neq \varnothing$. Then* $\operatorname{Fix}(T_1 \circ \cdots \circ T_m) = \bigcap_{i=1}^m \operatorname{Fix} T_i$.

D. Monotone operators

Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a set-valued operator. Then A is described by its graph

$$\operatorname{gra} A = \big\{(x,u) \in \mathcal{H}\times\mathcal{H} \;\big|\; u \in Ax\big\}, \tag{31}$$

and its inverse A^{-1}, defined by the relation

$$(\forall (x,u) \in \mathcal{H}\times\mathcal{H}) \quad x \in A^{-1}u \quad \Leftrightarrow \quad u \in Ax, \tag{32}$$

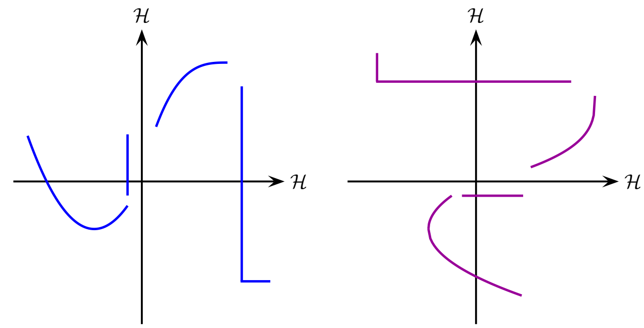

Fig. 4: Left: Graph of a (nonmonotone) set-valued operator. Right: Graph of its inverse.

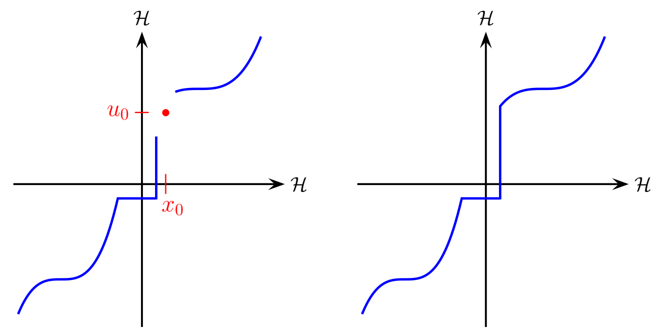

Fig. 5: Left: Graph of a monotone operator which is not maximally monotone: we can add the point (x_0,u_0) to its graph and still get a monotone graph. Right: Graph of a maximally monotone operator: adding any point to this graph destroys its monotonicity.

always exists (see Fig. 4). The operator A is *monotone* if

$$\big(\forall (x,u) \in \operatorname{gra} A\big)\big(\forall (y,v) \in \operatorname{gra} A\big) \quad \langle x-y \mid u-v\rangle \geqslant 0, \tag{33}$$

in which case A^{-1} is also monotone.

Example 27 Let $f\colon \mathcal{H} \to]-\infty,+\infty]$ be a proper function, let $(x,u) \in \operatorname{gra}\partial f$, and let $(y,v) \in \operatorname{gra}\partial f$. Then (13) yields

$$\begin{cases} \langle x-y \mid u\rangle + f(y) \geqslant f(x)\\ \langle y-x \mid v\rangle + f(x) \geqslant f(y). \end{cases} \tag{34}$$

Adding these inequality yields $\langle x-y \mid u-v\rangle \geqslant 0$, which shows that ∂f is monotone.

A natural question is whether the operator obtained by adding a point to the graph of a monotone operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ is still monotone. If it is not, then A is said to be *maximally monotone*. Thus, A is maximally monotone if, for every $(x,u) \in \mathcal{H}\times\mathcal{H}$,

$$(x,u) \in \operatorname{gra} A \;\Leftrightarrow\; (\forall (y,v) \in \operatorname{gra} A)\;\; \langle x-y \mid u-v\rangle \geqslant 0. \tag{35}$$

These notions are illustrated in Fig. 5. Let us provide some basic examples of maximally monotone operators, starting with the subdifferential of (13) (see Fig. 2).

Example 28 (Moreau) Let $f \in \Gamma_0(\mathcal{H})$. Then ∂f is maximally monotone and $(\partial f)^{-1} = \partial f^*$.

Example 29 Let $T\colon \mathcal{H} \to \mathcal{H}$ be monotone and continuous. Then T is maximally monotone. In particular, if T is cocoercive, it is maximally monotone.

Example 30 Let $T\colon \mathcal{H} \to \mathcal{H}$ be nonexpansive. Then $\mathrm{Id} - T$ is maximally monotone.

Example 31 Let $T\colon \mathcal{H} \to \mathcal{H}$ be linear (hence continuous) and *positive* in the sense that $(\forall x \in \mathcal{H})\ \langle x \mid Tx\rangle \geqslant 0$. Then T is maximally monotone. In particular, if T is *skew*, i.e., $T^* = -T$, then it is maximally monotone.

Given $A\colon \mathcal{H} \to 2^{\mathcal{H}}$, the *resolvent* of A is the operator $J_A = (\mathrm{Id} + A)^{-1}$, that is,

$$(\forall (x,p) \in \mathcal{H} \times \mathcal{H}) \quad p \in J_A x \quad \Leftrightarrow \quad x - p \in Ap. \tag{36}$$

In addition, the *reflected resolvent* of A is

$$R_A = 2J_A - \mathrm{Id}. \tag{37}$$

A profound result which connects monotonicity and nonexpansiveness is Minty's theorem [180]. It implies that if, $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone, then J_A is single-valued, defined everywhere on $\mathcal{H}$, and firmly nonexpansive.

Theorem 32 (Minty) *Let $T\colon \mathcal{H} \to \mathcal{H}$. Then T is firmly nonexpansive if and only if it is the resolvent of a maximally monotone operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$.*

Example 33 Let $f \in \Gamma_0(\mathcal{H})$. Then $J_{\partial f} = \mathrm{prox}_f$.

Let f and g be functions in $\Gamma_0(\mathcal{H})$ which satisfy the constraint qualification $\mathrm{ri}(\mathrm{dom}\, f) \cap \mathrm{ri}(\mathrm{dom}\, g) \neq \varnothing$. In view of Proposition 7ii) and Example 28, the minimizers of $f + g$ are precisely the solutions to the inclusion $0 \in Ax + Bx$ involving the maximally monotone operators $A = \partial f$ and $B = \partial g$. Hence, it may seem that in minimization problems the theory of subdifferentials should suffice to analyze and solve problems without invoking general monotone operator theory. As discussed in [89], this is not the case and monotone operators play an indispensable role in various aspects of convex minimization. We give below an illustration of this fact in the context of Proposition 7.

Example 34 ([44]) Given $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and a linear operator $L\colon \mathcal{H} \to \mathcal{G}$, the objective is to

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + g(Lx) \tag{38}$$

using f and g separately by means of their respective proximity operators. To this end, let us bring into play the Fenchel-Rockafellar dual problem

$$\underset{v \in \mathcal{G}}{\text{minimize}} \;\; f^*(-L^*v) + g^*(v). \tag{39}$$

We derive from [21, Theorem 19.1] that, if $(x, v) \in \mathcal{H} \times \mathcal{G}$ solves the inclusion

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \underbrace{\begin{bmatrix} \partial f & 0 \\ 0 & \partial g^* \end{bmatrix}}_{\text{subdifferential}} \begin{bmatrix} x \\ v \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & L^* \\ -L & 0 \end{bmatrix}}_{\text{skew}} \begin{bmatrix} x \\ v \end{bmatrix}, \tag{40}$$

then x solves (38) and v solves (39). Now introduce the variable $\boldsymbol{z} = (x, v)$, the function $\Gamma_0(\mathcal{H} \times \mathcal{G}) \ni \boldsymbol{h}\colon \boldsymbol{z} \mapsto f(x) + g^*(v)$, the operator $\boldsymbol{A} = \partial \boldsymbol{h}$, and the skew operator $\boldsymbol{B}\colon \boldsymbol{z} \mapsto (L^*v, -Lx)$. Then it follows from Examples 28 and 31 that (40) can be written as the maximally monotone inclusion $\boldsymbol{0} \in \boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{z}$, which does not correspond to a minimization problem since $\boldsymbol{B}$ is not a gradient [21, Proposition 2.58]. As a result, genuine monotone operator splitting methods were employed in [44] to solve (40) and, thereby, (38) and (39). Applications of this framework can be found in image restoration [191] and in empirical mode decomposition [198].

Example 35 The primal-dual pair (38)–(39) can be exploited in various ways; see for instance [66], [90], [91], [164]. A simple illustration is found in sparse signal recovery and machine learning, where one often aims at solving (38) by choosing g to be a norm $|||\cdot|||$ [5], [11], [96], [123], [178]. Now let $|||\cdot|||_*\colon \mathcal{G} \to \mathbb{R}\colon v \mapsto \sup_{|||y||| \leqslant 1} \langle y \mid v\rangle$ be the dual norm and let $B_* = \big\{ v \in \mathcal{G} \;\big|\; |||v|||_* \leqslant 1 \big\}$ be the associated unit ball. Then (39) is the constrained optimization problem

$$\underset{v \in B_*}{\text{minimize}} \;\; f^*(-L^*v). \tag{41}$$

This dual formulation underlies several investigations, e.g., [127], [188].

III. Fixed point algorithms

We review the main fixed point construction algorithms.

A. Basic iteration schemes

First, we recall that finding the fixed point of a Banach contraction is relatively straightforward via the standard Banach-Picard iteration scheme (1).

Theorem 36 ([21]) *Let $\delta \in \left]0,1\right[$, let $T\colon \mathcal{H} \to \mathcal{H}$ be δ-Lipschitzian, and let $x_0 \in \mathcal{H}$. Set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Tx_n. \tag{42}$$

Then T has a unique fixed point $\overline{x}$ and $x_n \to \overline{x}$. More precisely, $(\forall n \in \mathbb{N})\ \|x_n - \overline{x}\| \leqslant \delta^n \|x_0 - \overline{x}\|$.

If T is merely nonexpansive (i.e., $\delta = 1$) with $\mathrm{Fix}\, T \neq \varnothing$, Theorem 36 fails. For instance, let $T \neq \mathrm{Id}$ be a rotation in the Euclidean plane. Then it is nonexpansive with $\mathrm{Fix}\, T = \{0\}$ but the sequence $(x_n)_{n\in\mathbb{N}}$ constructed by the successive approximation process (42) does not converge. Such scenarios can be handled via the following result.

Theorem 37 ([21]) *Let $\alpha \in]0,1]$, let $T\colon \mathcal{H} \to \mathcal{H}$ be an α-averaged operator such that $\operatorname{Fix} T \neq \varnothing$, let $(\lambda_n)_{n\in\mathbb{N}}$ be an α-relaxation sequence. Set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n\big(Tx_n - x_n\big). \tag{43}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges to a point in $\operatorname{Fix} T$.

Remark 38 In connection with Theorems 36 and 37, let us make the following observations.

i) If $\alpha < 1$ in Theorem 37, choosing $\lambda_n = 1$ in (43) (see Example 10i)) yields (42).
ii) In contrast with Theorem 36, the convergence in Theorem 37 is not linear in general [23], [32].
iii) When $\alpha = 1$, (43) is known as the *Krasnosel'skiĭ-Mann iteration*.

Next, we present a more flexible fixed point theorem which involves iteration-dependent composite averaged operators.

Theorem 39 ([111]) *Let $\varepsilon \in]0,1/2[$ and let $x_0 \in \mathcal{H}$. For every $n \in \mathbb{N}$, let $\alpha_{1,n} \in]0,1/(1+\varepsilon)]$, let $\alpha_{2,n} \in]0,1/(1+\varepsilon)]$, let $T_{1,n}\colon \mathcal{H} \to \mathcal{H}$ be $\alpha_{1,n}$-averaged, and let $T_{2,n}\colon \mathcal{H} \to \mathcal{H}$ be $\alpha_{2,n}$-averaged. In addition, for every $n \in \mathbb{N}$, let*

$$\lambda_n \in \big[\varepsilon, (1-\varepsilon)(1+\varepsilon\alpha_n)/\alpha_n\big], \tag{44}$$

where $\alpha_n = (\alpha_{1,n} + \alpha_{2,n} - 2\alpha_{1,n}\alpha_{2,n})/(1-\alpha_{1,n}\alpha_{2,n})$, and set

$$x_{n+1} = x_n + \lambda_n\big(T_{1,n}(T_{2,n}x_n) - x_n\big). \tag{45}$$

Suppose that $S = \bigcap_{n\in\mathbb{N}} \operatorname{Fix}(T_{1,n} \circ T_{2,n}) \neq \varnothing$. Then the following hold:

i) $(\forall x \in S)\ \sum_{n\in\mathbb{N}} \|T_{2,n}x_n - x_n - T_{2,n}x + x\|^2 < +\infty$.
ii) *Suppose that a subsequence of $(x_n)_{n\in\mathbb{N}}$ converges to a point in S. Then $(x_n)_{n\in\mathbb{N}}$ converges to a point in S.*

Remark 40 The assumption in Theorem 39ii) holds in particular when, for every $n \in \mathbb{N}$, $T_{1,n} = T_1$ and $T_{2,n} = T_2$.

Below, we present a variant of Theorem 37 obtained by considering the composition of m operators. In the case of firmly nonexpansive operators, this result is due to Martinet [177].

Theorem 41 ([87]) *For every $i \in \{1,\ldots,m\}$, let $\alpha_i \in]0,1[$ and let $T_i\colon \mathcal{H} \to \mathcal{H}$ be α_i-averaged. Let $x_0 \in \mathcal{H}$, suppose that $\operatorname{Fix}(T_1 \circ \cdots \circ T_m) \neq \varnothing$, and iterate*

$$\begin{array}{l} \text{for } n = 0,1,\ldots \\ \left\lfloor \begin{array}{ll} x_{mn+1} &= T_m x_{mn} \\ x_{mn+2} &= T_{m-1} x_{mn+1} \\ &\ \vdots \\ x_{mn+m-1} &= T_2 x_{mn+m-2} \\ x_{mn+m} &= T_1 x_{mn+m-1}. \end{array} \right. \end{array} \tag{46}$$

Then $(x_{mn})_{n\in\mathbb{N}}$ converges to a point $\overline{x}_1$ in $\operatorname{Fix}(T_1 \circ\cdots\circ T_m)$. Now set $\overline{x}_m = T_m\overline{x}_1$, $\overline{x}_{m-1} = T_{m-1}\overline{x}_m$, …, $\overline{x}_2 = T_2\overline{x}_3$. Then, for every $i \in \{1,\ldots,m-1\}$, $(x_{mn+i})_{n\in\mathbb{N}}$ converges to $\overline{x}_{m+1-i}$.

B. Algorithms for fixed point selection

The algorithms discussed so far construct an unspecified fixed point of a nonexpansive operator $T\colon \mathcal{H} \to \mathcal{H}$. In some applications, one may be interested in finding a specific fixed point, for instance one of minimum norm or, more generally, one that minimizes some quadratic function [6], [86]. One will find in [86] several algorithms to minimize convex quadratic functions over fixed point sets, as well as signal recovery applications. Beyond quadratic selection, one may wish to minimize a strictly convex function $g \in \Gamma_0(\mathcal{H})$ over the closed convex set (see Proposition 24) $\operatorname{Fix} T$, i.e.,

$$\underset{x\in\operatorname{Fix} T}{\text{minimize}}\ \ g(x). \tag{47}$$

Instances of such formulations can be found in signal interpolation [193] and machine learning [186]. Algorithms to solve (47) have been proposed in [84], [152], [241] under various hypotheses. Here is an example.

Proposition 42 ([241]) *Let $T\colon \mathcal{H} \to \mathcal{H}$ be nonexpansive, let $g\colon \mathcal{H} \to \mathbb{R}$ be strongly convex and differentiable with a Lipschitzian gradient, let $x_0 \in \mathcal{H}$, and let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1]$ such that $\alpha_n \to 0$, $\sum_{n\in\mathbb{N}} \alpha_n = +\infty$, and $\sum_{n\in\mathbb{N}} |\alpha_{n+1} - \alpha_n| < +\infty$. Suppose that (47) has a solution and iterate*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Tx_n - \alpha_n\nabla g(Tx_n). \tag{48}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges to the solution to (47).

C. A fixed point method with block operator updates

We turn our attention to a composite fixed point problem.

Problem 43 Let $(\omega_i)_{1\leqslant i\leqslant m}$ be real numbers in $]0,1]$ such that $\sum_{i=1}^m \omega_i = 1$. For every $i \in \{0,\ldots,m\}$, let $T_i\colon \mathcal{H} \to \mathcal{H}$ be α_i-averaged for some $\alpha_i \in]0,1[$. The task is to find a fixed point of $T_0 \circ \sum_{i=1}^m \omega_i T_i$, assuming that such a point exists.

A simple strategy to solve Problem 43 is to set $R = \sum_{i=1}^m \omega_i T_i$, observe that R is averaged by Proposition 16, and then use Theorem 39 and Remark 40 to find a fixed point of $T_0 \circ R$. This, however, requires the activation of the m operators $(T_i)_{1\leqslant i\leqslant m}$ to evaluate R at each iteration, which is a significant computational burden when m is sizable. In the degenerate case when the operators $(T_i)_{0\leqslant i\leqslant m}$ have common fixed points, Problem 43 amount to finding such a point (see Propositions 25 and 26) and this can be done using the strategies devised in [17], [22], [82], [163] which require only the activation of blocks of operators at each iteration. Such approaches fail in our more challenging setting, which assumes only that $\operatorname{Fix}(T_0 \circ \sum_{i=1}^m \omega_i T_i) \neq \varnothing$. However, with a strategy based on tools from mean iteration theory [93], it is possible to devise an algorithm which operates by updating only a block of operators $(T_i)_{i\in I_n}$ at iteration n.

Theorem 44 ([95]) *Consider the setting of Problem 43. Let M be a strictly positive integer and let $(I_n)_{n\in\mathbb{N}}$ be a sequence of nonempty subsets of $\{1,\ldots,m\}$ such that*

$$(\forall n \in \mathbb{N}) \quad \bigcup_{k=n}^{n+M-1} I_k = \{1,\ldots,m\}. \tag{49}$$

Let $x_0 \in \mathcal{H}$, let $(t_{i,-1})_{1\leqslant i\leqslant m} \in \mathcal{H}^m$, and iterate

$$\begin{array}{l} \text{for } n = 0,1,\ldots \\ \left\lfloor \begin{array}{l} \text{for every } i \in I_n \\ \lfloor \; t_{i,n} = T_i x_n \\ \text{for every } i \in \{1,\ldots,m\} \smallsetminus I_n \\ \lfloor \; t_{i,n} = t_{i,n-1} \\ x_{n+1} = T_0\big(\sum_{i=1}^m \omega_i t_{i,n}\big). \end{array} \right. \end{array} \tag{50}$$

Then the following hold:

i) *Let x be a solution to Problem 43 and let $i \in \{1,\ldots,m\}$. Then $x_n - T_i x_n \to x - T_i x$.*
ii) *$(x_n)_{n\in\mathbb{N}}$ converges to a solution to Problem 43.*
iii) *Suppose that, for some $i \in \{0,\ldots,m\}$, T_i is a Banach contraction. Then $(x_n)_{n\in\mathbb{N}}$ converges linearly to the unique solution to Problem 43.*

At iteration n, I_n is the set of indices of operators to be activated. The remaining operators are not used and their most recent evaluations are recycled to form the update x_{n+1}. Condition (49) imposes the mild requirement that each operator in $(T_i)_{1\leqslant i\leqslant m}$ be evaluated at least once over the course of any M consecutive iterations. The choice of M is left to the user.

D. Perturbed fixed point methods

For various modeling or computational reasons, exact evaluations of the operators in fixed point algorithms may not be possible. Such perturbations can be modeled by deterministic additive errors [87], [162], [177] but also by stochastic ones [102], [129]. Here is a stochastically perturbed version of Theorem 37, which is a straightforward variant of [102, Corollary 2.7].

Theorem 45 *Let $\alpha \in]0,1]$, let $T\colon \mathcal{H} \to \mathcal{H}$ be an α-averaged operator such that $\operatorname{Fix} T \neq \varnothing$, and let $(\lambda_n)_{n\in\mathbb{N}}$ be an α-relaxation sequence. Let x_0 and $(e_n)_{n\in\mathbb{N}}$ be $\mathcal{H}$-valued random variables. Set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n\big(Tx_n + e_n - x_n\big). \tag{51}$$

Suppose that $\sum_{n\in\mathbb{N}} \lambda_n \sqrt{\mathsf{E}(\|e_n\|^2 \,|\, \mathcal{X}_n)} < +\infty$ a. s., where $\mathcal{X}_n$ is the σ-algebra generated by $(x_0,\ldots,x_n)$. Then $(x_n)_{n\in\mathbb{N}}$ converges a. s. to a $(\operatorname{Fix} T)$-valued random variable.

E. Random block-coordinate fixed point methods

We have seen in Section III-C that the computational cost per iteration could be reduced in certain fixed point algorithms by updating only some of the operators involved in the model. In this section, we present another approach to reduce the iteration cost by considering scenarios in which the underlying Euclidean space $\boldsymbol{\mathcal{H}}$ is decomposable in m factors $\boldsymbol{\mathcal{H}} = \mathcal{H}_1 \times \cdots \times \mathcal{H}_m$. In the spirit of the Gauss-Seidel algorithm, one can explore the possibility of activating only some of the coordinates of certain operators at each iteration of a fixed point method. The potential advantages of such a procedure are a reduced computational cost per iteration, reduced memory requirements, and an increased implementation flexibility.

In the product space $\boldsymbol{\mathcal{H}}$, consider the basic update process

$$\boldsymbol{x}_{n+1} = \boldsymbol{T}_n \boldsymbol{x}_n, \tag{52}$$

under the assumption that the operator $\boldsymbol{T}_n$ is decomposable explicitly as

$$\boldsymbol{T}_n\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}\colon \boldsymbol{x} \mapsto (T_{1,n}\boldsymbol{x},\ldots,T_{m,n}\boldsymbol{x}), \tag{53}$$

with $T_{i,n}\colon \boldsymbol{\mathcal{H}} \to \mathcal{H}_i$. Updating only some coordinates is performed by modifying iteration (52) as

$$(\forall i \in \{1,\ldots,m\}) \quad x_{i,n+1} = x_{i,n} + \varepsilon_{i,n}\big(T_{i,n}x_n - x_{i,n}\big), \tag{54}$$

where $\varepsilon_{i,n} \in \{0,1\}$ signals the activation of the i-th coordinate of $\boldsymbol{x}_n$. If $\varepsilon_{i,n} = 1$, the i-th component is updated whereas, if $\varepsilon_{i,n} = 0$, it is unchanged. The main difficulty facing such an approach is that the nonexpansiveness property of an operator is usually destroyed by coordinate sampling. To remove this roadblock, a possibility is to make the activation variables random, which results in a stochastic algorithm for which almost sure convergence holds [102], [155].

Theorem 46 ([102]) *Let $\alpha \in]0,1]$, let $\epsilon \in]0,1/2[$, and let $\boldsymbol{T}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}\colon \boldsymbol{x} \mapsto (T_i\,\boldsymbol{x})_{1\leqslant i\leqslant m}$ be an α-averaged operator where $T_i\colon \boldsymbol{\mathcal{H}} \to \mathcal{H}_i$. Let $(\lambda_n)_{n\in\mathbb{N}}$ be in $[\epsilon, \alpha^{-1} - \epsilon]$, set $D = \{0,1\}^m \smallsetminus \{\boldsymbol{0}\}$, let $\boldsymbol{x}_0$ be an $\boldsymbol{\mathcal{H}}$-valued random variable, and let $(\boldsymbol{\varepsilon}_n)_{n\in\mathbb{N}}$ be identically distributed D-valued random variables. Iterate*

$$\begin{array}{l} \text{for } n = 0,1,\ldots \\ \left\lfloor \begin{array}{l} \text{for } i = 1,\ldots,m \\ \lfloor \; x_{i,n+1} = x_{i,n} + \varepsilon_{i,n}\lambda_n\big(T_i\boldsymbol{x}_n - x_{i,n}\big). \end{array} \right. \end{array} \tag{55}$$

In addition, assume that the following hold:

i) $\operatorname{Fix} \boldsymbol{T} \neq \varnothing$.
ii) *For every $n \in \mathbb{N}$, $\boldsymbol{\varepsilon}_n$ and $(\boldsymbol{x}_0,\ldots,\boldsymbol{x}_n)$ are mutually independent.*
iii) $(\forall i \in \{1,\ldots,m\})$ $\mathsf{Prob}\,[\varepsilon_{i,0} = 1] > 0$.

Then $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges a. s. to a $\operatorname{Fix} \boldsymbol{T}$-valued random variable.

Further results in this vein for iterations involving nonstationary compositions of averaged operators can be found in [102]. Mean square convergence results are also available under additional assumptions on the operators $(\boldsymbol{T}_n)_{n\in\mathbb{N}}$ [104].

IV. Fixed point modeling of monotone inclusions

A. Splitting sums of monotone operators

Our first basic model is that of finding a zero of the sum of two monotone operators. It will be seen to be central in understanding and solving data science problems in optimization form (see also Example 34 for a special case) and beyond.

Problem 47 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone operators. The task is to

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Bx, \tag{56}$$

under the assumption that a solution exists.

A classical method for solving Problem 47 is the *Douglas-Rachford* algorithm, which was first proposed in [173] (see also [126]; the following relaxed version is from [85]).

Proposition 48 (Douglas-Rachford splitting) *Let $(\lambda_n)_{n\in\mathbb{N}}$ be a $1/2$-relaxation sequence, let $\gamma \in \left]0,+\infty\right[$, and let $y_0 \in \mathcal{H}$. Iterate*

$$\begin{array}{l}\text{for } n = 0,1,\ldots\\ \left\lfloor \begin{array}{l} x_n = J_{\gamma B} y_n \\ z_n = J_{\gamma A}(2x_n - y_n) \\ y_{n+1} = y_n + \lambda_n(z_n - x_n). \end{array}\right.\end{array} \tag{57}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges to a solution to Problem 47.

The Douglas-Rachford algorithm requires the ability to evaluate two resolvents at each iteration. However, if one of the operators is single-valued and Lipschitzian, it is possible to apply it explicitly, hence requiring only one resolvent evaluation per iteration. The resulting algorithm, proposed by Tseng [232], is often called the *forward-backward-forward* splitting algorithm since it involves two explicit (forward) steps using B and one implicit (backward) step using A.

Proposition 49 (Tseng splitting) *In Problem 47, assume that B is δ-Lipschitzian for some $\delta \in \left]0,+\infty\right[$. Let $x_0 \in \mathcal{H}$, let $\varepsilon \in \left]0,1/(\delta+1)\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be in $[\varepsilon,(1-\varepsilon)/\delta]$, and iterate*

$$\begin{array}{l}\text{for } n = 0,1,\ldots\\ \left\lfloor \begin{array}{l} y_n = x_n - \gamma_n B x_n \\ z_n = J_{\gamma_n A} y_n \\ r_n = z_n - \gamma_n B z_n \\ x_{n+1} = x_n - y_n + r_n. \end{array}\right.\end{array} \tag{58}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges to a solution to Problem 47.

As noted in Section II-C, if B is cocoercive, then it is Lipschitzian, and Proposition 49 is applicable. However, in this case it is possible to devise an algorithm which requires only one application of B per iteration, as opposed to two in (58). To see this, let $\gamma_n \in \left]0,2\beta\right[$ and $x \in \mathcal{H}$. Then it follows at once from (36) that x solves Problem 47 $\Leftrightarrow -\gamma_n Bx \in \gamma_n Ax$ $\Leftrightarrow (x-\gamma_n Bx) - x \in \gamma_n Ax \Leftrightarrow x = J_{\gamma_n A}(x-\gamma_n Bx) \Leftrightarrow x \in \operatorname{Fix}(T_{1,n}\circ T_{2,n})$, where $T_{1,n} = J_{\gamma_n A}$ and $T_{2,n} = \mathrm{Id}-\gamma_n B$. As seen in Theorem 32, $T_{1,n}$ is $1/2$-averaged. On the other hand, we derive from Proposition 13 that, if $\alpha_{2,n} = \gamma_n/(2\beta)$, then $T_{2,n}$ is $\alpha_{2,n}$-averaged. With these considerations, we invoke Theorem 39 to obtain the following algorithm, which goes back to [179].

Proposition 50 (forward-backward splitting [111]) *Suppose that, in Problem 47, B is β-cocoercive for some $\beta \in \left]0,+\infty\right[$. Let $\varepsilon \in \left]0,\min\{1/2,\beta\}\right[$, let $x_0 \in \mathcal{H}$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be in $[\varepsilon, 2\beta/(1+\varepsilon)]$. Let*

$$(\forall n \in \mathbb{N}) \quad \lambda_n \in \big[\varepsilon, (1-\varepsilon)\big(2+\varepsilon-\gamma_n/(2\beta)\big)\big]. \tag{59}$$

Iterate

$$\begin{array}{l}\text{for } n = 0,1,\ldots\\ \left\lfloor \begin{array}{l} u_n = x_n - \gamma_n B x_n \\ x_{n+1} = x_n + \lambda_n\big(J_{\gamma_n A} u_n - x_n\big). \end{array}\right.\end{array} \tag{60}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges to a solution to Problem 47.

We now turn our attention to a more structured version of Problem 47, which includes an additional Lipschitzian monotone operator.

Problem 51 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone operators, let $\delta \in \left]0,+\infty\right[$, and let $C\colon \mathcal{H} \to \mathcal{H}$ be monotone and δ-Lipschitzian. The task is to

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Bx + Cx, \tag{61}$$

under the assumption that a solution exists.

The following approach provides also a dual solution.

Proposition 52 (splitting three operators I [101]) *Consider Problem 51 and let $\varepsilon \in \left]0,1/(2+\delta)\right[$. Let $(\gamma_n)_{n\in\mathbb{N}}$ be in $[\varepsilon,(1-\varepsilon)/(1+\delta)]$, let $x_0 \in \mathcal{H}$, and let $u_0 \in \mathcal{H}$. Iterate*

$$\begin{array}{l}\text{for } n = 0,1,\ldots\\ \left\lfloor \begin{array}{l} y_n = x_n - \gamma_n(Cx_n + u_n) \\ p_n = J_{\gamma_n A}\, y_n \\ q_n = u_n + \gamma_n\big(x_n - J_{B/\gamma_n}(u_n/\gamma_n + x_n)\big) \\ x_{n+1} = x_n - y_n + p_n - \gamma_n(Cp_n + q_n) \\ u_{n+1} = q_n + \gamma_n(p_n - x_n). \end{array}\right.\end{array} \tag{62}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges to a solution to Problem 51 and $(u_n)_{n\in\mathbb{N}}$ converges to a solution u to the dual problem, i.e., $0 \in -(A+C)^{-1}(-u) + B^{-1}u$.

When C is β-cocoercive in Problem 51, we can take $\delta = 1/\beta$. In this setting, an alternative algorithm is obtained as follows. Let us fix $\gamma \in \left]0,+\infty\right[$ and define

$$T = J_{\gamma A} \circ \big(2J_{\gamma B} - \mathrm{Id} - \gamma C \circ J_{\gamma B}\big) + \mathrm{Id} - J_{\gamma B}. \tag{63}$$

By setting $T_1 = J_{\gamma A}$, $T_2 = J_{\gamma B}$, and $T_3 = \mathrm{Id} - \gamma C$ in Proposition 20, we deduce from Proposition 13 that, if $\gamma \in \left]0,2\beta\right[$ and $\alpha = 2\beta/(4\beta-\gamma)$, then T is α-averaged. Now take $y \in \mathcal{H}$ and set $x = J_{\gamma B} y$, hence $y - x \in \gamma Bx$ by (36). Then $y \in \operatorname{Fix} T \Leftrightarrow J_{\gamma A}(2x - y - \gamma Cx) + y - x = y \Leftrightarrow J_{\gamma A}(2x-y-\gamma Cx) = x \Leftrightarrow x - y - \gamma Cx \in \gamma Ax$ by (36). Thus, $0 = (x-y)+(y-x) \in \gamma(Ax+Bx+Cx)$, which shows that x solves Problem 51. Altogether, since y can be constructed via Theorem 37, we obtain the following convergence result.

Proposition 53 (splitting three operators II [118]) *In Problem 51, assume that C is β-cocoercive for some $\beta \in \left]0,+\infty\right[$. Let $\gamma \in \left]0,2\beta\right[$ and set $\alpha = 2\beta/(4\beta-\gamma)$. Furthermore, let $(\lambda_n)_{n\in\mathbb{N}}$ be an α-relaxation sequence and let $y_0 \in \mathcal{H}$. Iterate*

$$\begin{array}{l}\text{for } n = 0,1,\ldots\\ \left\lfloor \begin{array}{l} x_n = J_{\gamma B}\, y_n \\ r_n = y_n + \gamma C x_n \\ z_n = J_{\gamma A}(2x_n - r_n) \\ y_{n+1} = y_n + \lambda_n(z_n - x_n). \end{array}\right.\end{array} \tag{64}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges to a solution to Problem 51.

Remark 54

i) Work closely related to Proposition 53 can be found in [41], [47], [200]. See also [199], which provides further developments and a discussion of [41], [118], [200].
ii) Unlike algorithm (62), (64) imposes constant proximal parameters and requires the cocoercivity of C, but it involves only one application of C per iteration. An extension of (64) appears in [242] in the context of minimization problems.

B. Splitting sums of composite monotone operators

The monotone inclusion problems of Section IV-A are instantiations of the following formulation, which involves an arbitrary number of maximally monotone operators and compositions with linear operators.

Problem 55 Let $\delta \in \left]0,+\infty\right[$ and let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. For every $k \in \{1,\ldots,q\}$, let $B_k\colon \mathcal{G}_k \to 2^{\mathcal{G}_k}$ be maximally monotone, let $0 \neq L_k\colon \mathcal{H} \to \mathcal{G}_k$ be linear, and let $C_k\colon \mathcal{G}_k \to \mathcal{G}_k$ be monotone and δ-Lipschitzian. The task is to

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + \sum_{k=1}^{q} L_k^*\big((B_k + C_k)(L_k x)\big), \quad (65)$$

under the assumption that a solution exists.

In the context of Problem 55, the principle of a splitting algorithm is to involve all the operators individually. In the case of a set-valued operator A or B_k, this means using the associated resolvent, whereas in the case of a single-valued operator C_k or L_k, a direct application can be considered. An immediate difficulty one faces with (65) is that it involves many set-valued operators. However, since inclusion is a binary relation, for reasons discussed in [44], [88] and analyzed in more depth in [210], it is not possible to deal with more than two such operators. To circumvent this fundamental limitation, a strategy is to rephrase Problem 55 as a problem involving at most two set-valued operators in a larger space. This strategy finds its root in convex feasibility problems [196] and it was first adapted to the problem of finding a zero of the sum of m operators in [142], [218]. In [44], it was used to deal with the presence of linear operators (see in particular Example 34), with further developments in [35], [36], [90], [101], [237]. In the same spirit, let us reformulate Problem 55 by introducing

$$\begin{cases} \boldsymbol{L}\colon \mathcal{H} \to \boldsymbol{\mathcal{G}}\colon x \mapsto (L_1 x,\ldots,L_q x)\\ \boldsymbol{B}\colon \boldsymbol{\mathcal{G}} \to 2^{\boldsymbol{\mathcal{G}}}\colon (y_k)_{1\leqslant k\leqslant q} \mapsto \times_{k=1}^{q} B_k y_k\\ \boldsymbol{C}\colon \boldsymbol{\mathcal{G}} \to \boldsymbol{\mathcal{G}}\colon (y_k)_{1\leqslant k\leqslant q} \mapsto (C_k y_k)_{1\leqslant k\leqslant q}\\ \boldsymbol{V} = \operatorname{range} \boldsymbol{L}. \end{cases} \quad (66)$$

Note that $\boldsymbol{L}$ is linear, $\boldsymbol{B}$ is maximally monotone, and $\boldsymbol{C}$ is monotone and δ-Lipschitzian. In addition, the inclusion (65) can be rewritten more concisely as

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + \boldsymbol{L}^*\big((\boldsymbol{B}+\boldsymbol{C})(\boldsymbol{L}x)\big). \quad (67)$$

In particular, suppose that $A = 0$. Then, upon setting $\boldsymbol{y} = \boldsymbol{L}x \in \boldsymbol{V}$, we obtain the existence of a point $\boldsymbol{u} \in (\boldsymbol{B}+\boldsymbol{C})\boldsymbol{y}$ in $\ker \boldsymbol{L}^* = \boldsymbol{V}^\perp$. In other words,

$$\boldsymbol{0} \in N_{\boldsymbol{V}}\boldsymbol{y} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{C}\boldsymbol{y}. \quad (68)$$

Solving this inclusion is equivalent to solving a problem similar to Problem 51, formulated in $\boldsymbol{\mathcal{G}}$. Thus, applying Proposition 53 to (68) leads to the following result.

Proposition 56 *In Problem 55, suppose that $A = 0$, that the operators $(C_k)_{1\leqslant k\leqslant q}$ are β-cocoercive for some $\beta \in \left]0,+\infty\right[$, and that $Q = \sum_{k=1}^{q} L_k^* \circ L_k$ is invertible. Let $\gamma \in \left]0,2\beta\right[$, set $\alpha = 2\beta/(4\beta-\gamma)$, and let $(\lambda_n)_{n\in\mathbb{N}}$ be an α-relaxation sequence. Further, let $\boldsymbol{y}_0 \in \boldsymbol{\mathcal{G}}$, set $s_0 = Q^{-1}\Big(\sum_{k=1}^{q} L_k^* y_{0,k}\Big)$, and iterate*

$$\begin{array}{l} \text{for } n = 0,1,\ldots\\ \left\lfloor \begin{array}{l} \text{for } k = 1,\ldots,q\\ \left\lfloor\ p_{n,k} = J_{\gamma B_k} y_{n,k}\right.\\ x_n = Q^{-1}\big(\sum_{k=1}^{q} L_k^* p_{n,k}\big)\\ c_n = Q^{-1}\big(\sum_{k=1}^{q} L_k^* C_k p_{n,k}\big)\\ z_n = x_n - s_n - \gamma c_n\\ \text{for } k = 1,\ldots,q\\ \left\lfloor\ y_{n+1,k} = y_{n,k} + \lambda_n(x_n + z_n - p_{n,k})\right.\\ s_{n+1} = s_n + \lambda_n z_n. \end{array}\right. \end{array} \quad (69)$$

Then $(x_n)_{n\in\mathbb{N}}$ converges to a solution to (65).

A strategy for handling Problem 55 in its general setting consists of introducing an auxiliary variable $\boldsymbol{v} \in \boldsymbol{B}(\boldsymbol{L}x)$ in (67), which can then be rewritten as

$$\begin{cases} 0 \in Ax + \boldsymbol{L}^*\boldsymbol{v} + \boldsymbol{L}^*\big(\boldsymbol{C}(\boldsymbol{L}x)\big)\\ \boldsymbol{0} \in -\boldsymbol{L}x + \boldsymbol{B}^{-1}\boldsymbol{v}. \end{cases} \quad (70)$$

This results in an instantiation of Problem 47 in $\boldsymbol{\mathcal{K}} = \mathcal{H} \times \boldsymbol{\mathcal{G}}$ involving the maximally monotone operators

$$\begin{cases} \boldsymbol{\mathcal{A}}_1\colon \boldsymbol{\mathcal{K}} \to 2^{\boldsymbol{\mathcal{K}}}\colon (x,\boldsymbol{v}) & \mapsto \begin{bmatrix} A & 0\\ 0 & \boldsymbol{B}^{-1} \end{bmatrix} \begin{bmatrix} x\\ \boldsymbol{v} \end{bmatrix}\\ \boldsymbol{\mathcal{B}}_1\colon \boldsymbol{\mathcal{K}} \to \boldsymbol{\mathcal{K}}\colon (x,\boldsymbol{v}) & \mapsto \begin{bmatrix} \boldsymbol{L}^* \circ \boldsymbol{C} \circ \boldsymbol{L} & \boldsymbol{L}^*\\ -\boldsymbol{L} & 0 \end{bmatrix} \begin{bmatrix} x\\ \boldsymbol{v} \end{bmatrix}. \end{cases} \quad (71)$$

We observe that, in $\boldsymbol{\mathcal{K}}$, $\boldsymbol{\mathcal{B}}_1$ is Lipschitzian with constant $\chi = \|\boldsymbol{L}\|(1+\delta\|\boldsymbol{L}\|)$. By applying Proposition 49 to (70), we obtain the following algorithm.

Proposition 57 ([101]) *Consider Problem 55. Set*

$$\chi = \sqrt{\textstyle\sum_{k=1}^{q}\|L_k\|^2}\Big(1+\delta\sqrt{\textstyle\sum_{k=1}^{q}\|L_k\|^2}\Big). \quad (72)$$

Let $x_0 \in \mathcal{H}$, let $\boldsymbol{v}_0 \in \boldsymbol{\mathcal{G}}$, let $\varepsilon \in \left]0, 1/(\chi+1)\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be in $[\varepsilon, (1-\varepsilon)/\chi]$, and iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
u_n = x_n - \gamma_n \sum_{k=1}^q L_k^*(C_k(L_k x_n) + v_{n,k}) \\
p_n = J_{\gamma_n A} u_n \\
\text{for } k = 1, \ldots, q \\
\left\lfloor
\begin{array}{l}
y_{n,k} = v_{n,k} + \gamma_n L_k x_n \\
z_{n,k} = y_{n,k} - \gamma_n J_{\gamma^{-1} B_k}\big(y_{n,k}/\gamma_n\big) \\
s_{n,k} = z_{n,k} + \gamma_n L_k p_n \\
v_{n+1,k} = v_{n,k} - y_{n,k} + s_{n,k}
\end{array}
\right. \\
r_n = p_n - \gamma_n \sum_{k=1}^q L_k^*(C_k(L_k p_n) + z_{n,k}) \\
x_{n+1} = x_n - u_n + r_n.
\end{array}
\right.
\end{array}
\tag{73}
$$

Then $(x_n)_{n\in\mathbb{N}}$ converges to a solution to Problem 55.

An alternative approach consists of reformulating (70) in the form of Problem 47 with the maximally monotone operators

$$
\begin{cases}
\boldsymbol{\mathcal{A}}_2 \colon \boldsymbol{\mathcal{K}} \to 2^{\boldsymbol{\mathcal{K}}} \colon (x, \boldsymbol{v}) & \mapsto \begin{bmatrix} A & \boldsymbol{L}^* \\ -\boldsymbol{L} & \boldsymbol{B}^{-1} \end{bmatrix} \begin{bmatrix} x \\ \boldsymbol{v} \end{bmatrix} \\[2ex]
\boldsymbol{\mathcal{B}}_2 \colon \boldsymbol{\mathcal{K}} \to \boldsymbol{\mathcal{K}} \colon (x, \boldsymbol{v}) & \mapsto \begin{bmatrix} \boldsymbol{L}^* \circ \boldsymbol{C} \circ \boldsymbol{L} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ \boldsymbol{v} \end{bmatrix}.
\end{cases}
\tag{74}
$$

Instead of working directly with these operators, it may be judicious to use preconditioned versions $\boldsymbol{V} \circ \boldsymbol{\mathcal{A}}_2$ and $\boldsymbol{V} \circ \boldsymbol{\mathcal{B}}_2$, where $\boldsymbol{V} \colon \boldsymbol{\mathcal{K}} \to \boldsymbol{\mathcal{K}}$ is a self-adjoint strictly positive linear operator. If $\boldsymbol{\mathcal{K}}$ is renormed with

$$
\|\cdot\|_{\boldsymbol{V}} \colon (x, \boldsymbol{v}) \mapsto \sqrt{\langle (x, \boldsymbol{v}) \mid \boldsymbol{V}^{-1}(x, \boldsymbol{v}) \rangle},
\tag{75}
$$

then $\boldsymbol{V} \circ \boldsymbol{\mathcal{A}}_2$ is maximally monotone in the renormed space and, if $\boldsymbol{C}$ is cocoercive in $\boldsymbol{\mathcal{G}}$, then $\boldsymbol{V} \circ \boldsymbol{\mathcal{B}}_2$ is cocoercive in the renormed space. Thus, setting

$$
\boldsymbol{V} = \begin{bmatrix} W & 0 \\ 0 & (\sigma^{-1}\mathbf{Id} - \boldsymbol{L} \circ W \circ \boldsymbol{L}^*)^{-1} \end{bmatrix},
\tag{76}
$$

where $W \colon \mathcal{H} \to \mathcal{H}$, and applying Proposition 50 in this context yields the following result (see [90]).

Proposition 58 *Suppose that, in Problem 55, $A = 0$ and $(C_k)_{1\leqslant k\leqslant q}$ are β-cocoercive for some $\beta \in \left]0, +\infty\right[$. Let $W \colon \mathcal{H} \to \mathcal{H}$ be a self-adjoint strictly positive linear operator and let $\sigma \in \left]0, +\infty\right[$ be such that $\kappa = \|\boldsymbol{L} \circ W \circ \boldsymbol{L}^*\| <$ $\min\{1/\sigma, 2\beta\}$. Let $\varepsilon \in \left]0, \min\{1/2, \beta/\kappa\}\right[$, let $x_0 \in \mathcal{H}$, and let $\boldsymbol{v}_0 \in \boldsymbol{\mathcal{G}}$. For every $n \in \mathbb{N}$, let*

$$
\lambda_n \in \big[\varepsilon, (1-\varepsilon)\big(2 + \varepsilon - \kappa/2\beta\big)\big].
\tag{77}
$$

Iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\text{for } k = 1, \ldots, q \\
\left\lfloor\; s_{n,k} = C_k(L_k x_n) \right. \\[1ex]
z_n = x_n - W\big(\sum_{k=1}^q L_k^*(s_{n,k} + v_{n,k})\big) \\
\text{for } k = 1, \ldots, q \\
\left\lfloor
\begin{array}{l}
w_{n,k} = v_{n,k} + \sigma L_k z_n \\
y_{n,k} = w_{n,k} - \sigma J_{\sigma^{-1} B_k}\left(w_{n,k}/\sigma\right) \\
v_{n+1,k} = v_{n,k} + \lambda_n (y_{n,k} - v_{n,k})
\end{array}
\right. \\
u_n = x_n - W\big(\sum_{k=1}^q L_k^*(s_{n,k} + y_{n,k})\big) \\
x_{n+1} = x_n + \lambda_n (u_n - x_n).
\end{array}
\right.
\end{array}
\tag{78}
$$

Then $(x_n)_{n\in\mathbb{N}}$ converges to a solution to Problem 55.

Other choices of the metric operator $\boldsymbol{V}$ are possible, which lead to different primal-dual algorithms [107], [112], [164], [237]. An advantage of (73) and (78) over (69) is that the first two do not require the inversion of linear operators.

C. Block-iterative algorithms

As will be seen in Problems 84 and 86, systems of inclusions arise in multivariate optimization problems (they will also be present in Nash equilibria; see, e.g., (153) and (183)). We now focus on general systems of inclusions involving maximally monotone operators as well as linear operators coupling the variables.

Problem 59 For every $i \in I = \{1, \ldots, m\}$ and $k \in K = \{1, \ldots, q\}$, let $A_i \colon \mathcal{H}_i \to 2^{\mathcal{H}_i}$ and $B_k \colon \mathcal{G}_k \to 2^{\mathcal{G}_k}$ be maximally monotone, and let $L_{k,i} \colon \mathcal{H}_i \to \mathcal{G}_k$ be linear. The task is to

$$
\text{find } \overline{x}_1 \in \mathcal{H}_1, \ldots, \overline{x}_m \in \mathcal{H}_m \text{ such that } (\forall i \in I)
$$
$$
0 \in A_i \overline{x}_i + \sum_{k\in K} L_{k,i}^*\bigg(B_k\bigg(\sum_{j\in I} L_{k,j}\overline{x}_j\bigg)\bigg),
\tag{79}
$$

under the assumption that the *Kuhn-Tucker set*

$$
\boldsymbol{Z} = \bigg\{ (\overline{\boldsymbol{x}}, \overline{\boldsymbol{v}}) \in \boldsymbol{\mathcal{H}} \times \boldsymbol{\mathcal{G}} \;\bigg|\; (\forall i \in I)\; -\sum_{k\in K} L_{k,i}^* \overline{v}_k \in A_i \overline{x}_i
$$
$$
\text{and } (\forall k \in K) \; \sum_{i\in I} L_{k,i}\overline{x}_i \in B_k^{-1}\overline{v}_k \bigg\}
\tag{80}
$$

is nonempty.

We can regard m as the number of coordinates of the solution vector $\overline{\boldsymbol{x}} = (\overline{x}_i)_{1\leqslant i\leqslant m}$. In large-scale applications, m can be sizable and so can the number of terms q, which is often associated with the number of observations. We have already discussed in Sections III-C and III-E techniques in which not all the indices i or k need to be activated at a given iteration. Below, we describe a block-iterative method proposed in [92] which allows for partial activation of both the families $(A_i)_{1\leqslant i\leqslant m}$ and $(B_k)_{1\leqslant k\leqslant q}$, together with individual, iteration-dependent proximal parameters for each operator. The method displays an unprecedented level of flexibility and it does not require the inversion of linear operators or knowledge of their norms.

The principle of the algorithm is as follows. Denote by $I_n \subset I$ and $K_n \subset K$ the blocks of indices of operators to be updated at iteration n. We impose the mild condition that there exist $M \in \mathbb{N}$ such that each operator index i and k is used at least once within any M consecutive iterations, i.e., for every $n \in \mathbb{N}$,

$$
\bigcup_{j=n}^{n+M-1} I_j = \{1, \ldots, m\} \text{ and } \bigcup_{j=n}^{n+M-1} K_j = \{1, \ldots, q\}.
\tag{81}
$$

For each $i \in I_n$ and $k \in K_n$, we select points $(a_{i,n}, a_{i,n}^*) \in \operatorname{gra} A_i$ and $(b_{k,n}, b_{k,n}^*) \in \operatorname{gra} B_k$ and use them to construct a closed half-space $\boldsymbol{H}_n \subset \boldsymbol{\mathcal{H}} \times \boldsymbol{\mathcal{G}}$ which contains $\boldsymbol{Z}$. The primal variable $\boldsymbol{x}_n$ and the dual variable $\boldsymbol{v}_n$ are updated as $(\boldsymbol{x}_{n+1}, \boldsymbol{v}_{n+1}) = \operatorname{proj}_{\boldsymbol{H}_n}(\boldsymbol{x}_n, \boldsymbol{v}_n)$. The resulting algorithm can also be implemented with relaxations and in an asynchronous fashion [92]. For simplicity, we present the unrelaxed synchronous version.

Proposition 60 ([92]) *Consider the setting of Problem 59. Take sequences $(I_n)_{n\in\mathbb{N}}$ in I and $(K_n)_{n\in\mathbb{N}}$ in K satisfying (81), with $I_0 = I$ and $K_0 = K$. Let $\varepsilon \in]0,1[$ and, for every $i \in I$ and every $k \in K$, let $(\gamma_{i,n})_{n\in\mathbb{N}}$ and $(\mu_{k,n})_{n\in\mathbb{N}}$ be sequences in $[\varepsilon, 1/\varepsilon]$. Let $\boldsymbol{x}_0 \in \boldsymbol{\mathcal{H}}$, let $\boldsymbol{v}_0 \in \boldsymbol{\mathcal{G}}$, and iterate*

$$
\begin{array}{l}
\text{for } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
\text{for every } i \in I_n\\
\left\lfloor
\begin{array}{l}
l_{i,n}^* = \sum_{k\in K} L_{k,i}^* v_{k,n}\\
a_{i,n} = J_{\gamma_{i,n} A_i}\big(x_{i,n} - \gamma_{i,n} l_{i,n}^*\big)\\
a_{i,n}^* = \gamma_{i,n}^{-1}(x_{i,n} - a_{i,n}) - l_{i,n}^*
\end{array}
\right.\\
\text{for every } i \in I \smallsetminus I_n\\
\left\lfloor\ (a_{i,n}, a_{i,n}^*) = (a_{i,n-1}, a_{i,n-1}^*)\right.\\
\text{for every } k \in K_n\\
\left\lfloor
\begin{array}{l}
l_{k,n} = \sum_{i\in I} L_{k,i} x_{i,n}\\
b_{k,n} = J_{\mu_{k,n} B_k}\big(l_{k,n} + \mu_{k,n} v_{k,n}\big)\\
b_{k,n}^* = v_{k,n} + \mu_{k,n}^{-1}(l_{k,n} - b_{k,n})
\end{array}
\right.\\
\text{for every } k \in K \smallsetminus K_n\\
\left\lfloor\ (b_{k,n}, b_{k,n}^*) = (b_{k,n-1}, b_{k,n-1}^*)\right.\\
\text{for every } i \in I\\
\left\lfloor\ t_{i,n}^* = a_{i,n}^* + \sum_{k\in K} L_{k,i}^* b_{k,n}^*\right.\\
\text{for every } k \in K\\
\left\lfloor\ t_{k,n} = b_{k,n} - \sum_{i\in I} L_{k,i} a_{i,n}\right.\\
\tau_n = \sum_{i\in I} \|t_{i,n}^*\|^2 + \sum_{k\in K} \|t_{k,n}\|^2\\
\text{if } \tau_n > 0\\
\left\lfloor
\begin{array}{l}
\theta_n = \dfrac{1}{\tau_n} \max\Big\{0, \sum_{i\in I}\big(\langle x_{i,n} \mid t_{i,n}^*\rangle - \langle a_{i,n} \mid a_{i,n}^*\rangle\big)\\
\qquad\qquad + \sum_{k\in K}\big(\langle t_{k,n} \mid v_{k,n}\rangle - \langle b_{k,n} \mid b_{k,n}^*\rangle\big)\Big\}
\end{array}
\right.\\
\text{else } \theta_n = 0\\
\text{for every } i \in I\\
\left\lfloor\ x_{i,n+1} = x_{i,n} - \theta_n t_{i,n}^*\right.\\
\text{for every } k \in K\\
\left\lfloor\ v_{k,n+1} = v_{k,n} - \theta_n t_{k,n}.\right.
\end{array}
\right.
\end{array}
\tag{82}
$$

Then $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges to a solution to Problem 59.

Recent developments on splitting algorithms for Problem 59 as well as variants and extensions thereof can be found in [48], [50], [138], [159], [160].

V. Fixed point modeling of minimization problems

We present key applications of fixed point models in convex optimization.

A. Convex feasibility problems

The most basic convex optimization problem is the convex feasibility problem, which asks for compliance with a finite number of convex constraints the object of interest is known to satisfy. This approach was formalized by Youla [247], [249] in signal recovery and it has enjoyed a broad success [80], [82], [150], [220], [227], [231].

Problem 61 Let $(C_i)_{1\leqslant i\leqslant m}$ be nonempty closed convex subsets of $\mathcal{H}$. The task is to

$$\text{find } x \in \bigcap_{i=1}^{m} C_i. \tag{83}$$

Suppose that Problem 61 has a solution and that each set C_i is modeled as the fixed point set of an α_i-averaged operator $T_i\colon \mathcal{H} \to \mathcal{H}$ for some $\alpha_i \in]0,1[$. Then, applying Theorem 37 with $T = T_1 \circ \cdots \circ T_m$ (which is averaged by Proposition 18) and $\lambda_n = 1$ for every $n \in \mathbb{N}$, we obtain that the sequence $(x_n)_{n\in\mathbb{N}}$ constructed via the iteration

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = (T_1 \circ \cdots \circ T_m)x_n \tag{84}$$

converges to a fixed point x of $T_1 \circ \cdots \circ T_m$. In view of Proposition 26, x is a solution to (83). In particular, if each T_i is the projection operator onto C_i (which was seen to be $1/2$-averaged), we obtain the classical POCS (Projection Onto Convex Sets) algorithm [39], [128]

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \big(\operatorname{proj}_{C_1} \circ \cdots \circ \operatorname{proj}_{C_m}\big)x_n \tag{85}$$

popularized in [249] and which goes back to [161] in the case of affine hyperplanes. In this algorithm, the projection operators are used sequentially. Another basic projection method for solving (83) is the barycentric projection algorithm

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \frac{1}{m}\sum_{i=1}^{m} \operatorname{proj}_{C_i} x_n, \tag{86}$$

which uses the projections simultaneously and goes back to [78] in the case of affine hyperplanes. Its convergence is proved by applying Theorem 37 to $T = m^{-1}\sum_{i=1}^{m} \operatorname{proj}_{C_i}$ which is $1/2$-averaged by Example 17. More general fixed point methods are discussed in [17], [22], [83], [163].

B. Split feasibility problems

The so-called *split feasibility problem* is just a convex feasibility problem involving a linear operator [53], [60], [61].

Problem 62 Let $C \subset \mathcal{H}$ and $D \subset \mathcal{G}$ be closed convex sets and let $0 \neq L\colon \mathcal{H} \to \mathcal{G}$ be linear. The task is to

$$\text{find } x \in C \text{ such that } Lx \in D, \tag{87}$$

under the assumption that a solution exists.

In principle, we can reduce this problem to a 2-set version of (83) with $C_1 = C$ and $C_2 = L^{-1}(D)$. However the projection onto C_2 is usually not tractable, which makes projection algorithms such as (85) or (86) not implementable. To work around this difficulty, let us define $T_1 = \operatorname{proj}_C$ and

$T_2 = \mathrm{Id} - \gamma G_2$, where $G_2 = L^* \circ (\mathrm{Id} - \mathrm{proj}_D) \circ L$ and $\gamma \in \left]0, +\infty\right[$. Then $(\forall x \in \mathcal{H})\ Lx \in D \Leftrightarrow G_2 x = 0$.[1] Hence,

$$\mathrm{Fix}\, T_1 = C \quad \text{and} \quad \mathrm{Fix}\, T_2 = \big\{x \in \mathcal{H} \;\big|\; Lx \in D\big\}. \tag{88}$$

Furthermore, T_1 is α_1-averaged with $\alpha_1 = 1/2$. In addition, $\mathrm{Id} - \mathrm{proj}_D$ is firmly nonexpansive by (24) and therefore 1-cocoercive. It follows from Proposition 21 that G_2 is cocoercive with constant $1/\|L\|^2$. Now let $\gamma \in \left]0, 2/\|L\|^2\right[$ and set $\alpha_2 = \gamma\|L\|^2/2$. Then Proposition 13 asserts that $\mathrm{Id} - \gamma G_2$ is α_2-averaged. Altogether, we deduce from Example 19 that $T_1 \circ T_2$ is α-averaged. Now let $(\lambda_n)_{n\in\mathbb{N}}$ be an α-relaxation sequence. According to Theorem 37 and Proposition 26, the sequence produced by the iterations

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad x_{n+1} &= x_n + \lambda_n \\ &\quad \cdot \Big(\mathrm{proj}_C\big(x_n - \gamma L^*(Lx_n - \mathrm{proj}_D(Lx_n))\big) - x_n\Big) \\ &= x_n + \lambda_n\big(T_1(T_2 x_n) - x_n\big)\end{aligned} \tag{89}$$

converges to a point in $\mathrm{Fix}\, T_1 \cap \mathrm{Fix}\, T_2$, i.e., in view of (88), to a solution to Problem 62. In particular, if we take $\lambda_n = 1$, the update rule in (89) becomes

$$x_{n+1} = \mathrm{proj}_C\Big(x_n - \gamma L^*\big(Lx_n - \mathrm{proj}_D(Lx_n)\big)\Big). \tag{90}$$

C. Convex minimization

We deduce from Fermat's rule (Theorem 3) and Proposition 6 the fact that a differentiable convex function $f\colon \mathcal{H} \to \mathbb{R}$ admits $x \in \mathcal{H}$ as a minimizer if and only if $\nabla f(x) = 0$. Now let $\gamma \in \left]0, +\infty\right[$. Then this property is equivalent to $x = x - \gamma\nabla f(x)$, which shows that

$$\mathrm{Argmin}\, f = \mathrm{Fix}\, T, \quad \text{where} \quad T = \mathrm{Id} - \gamma\nabla f. \tag{91}$$

If we add the assumption that ∇f is δ-Lipschitzian, then it is $1/\delta$-cocoercive by Proposition 14. Hence, if $0 < \gamma < 2/\delta$, it follows from Proposition 13, that T in (91) is α-averaged with $\alpha = \gamma\delta/2$. We then derive from Theorem 37 the convergence of the steepest-descent method.

Proposition 63 (steepest-descent) *Let $f\colon \mathcal{H} \to \mathbb{R}$ be a differentiable convex function such that $\mathrm{Argmin}\, f \neq \varnothing$ and ∇f is δ-Lipschitzian for some $\delta \in \left]0, +\infty\right[$. Let $\gamma \in \left]0, 2/\delta\right[$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a $\gamma\delta/2$-relaxation sequence, and let $x_0 \in \mathcal{H}$. Set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma\lambda_n\nabla f(x_n). \tag{92}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges to a point in $\mathrm{Argmin}\, f$.

Now, let us remove the smoothness assumption by considering a general function $f \in \Gamma_0(\mathcal{H})$. Then it is clear from (9) that $(\forall x \in \mathcal{H})\ x = \mathrm{prox}_f x \Leftrightarrow (\forall y \in \mathcal{H})\ f(x) \leqslant f(y)$. In other words, we obtain the fixed point characterization

$$\mathrm{Argmin}\, f = \mathrm{Fix}\, T, \quad \text{where} \quad T = \mathrm{prox}_f. \tag{93}$$

[1]Set $T = \mathrm{Id} - \mathrm{proj}_D$ and fix $\overline{x} \in \mathcal{H}$ such that $L\overline{x} \in D$. Then $T(L\overline{x}) = 0$ and thus $G_2\overline{x} = 0$. Conversely, take $x \in \mathcal{H}$ such that $G_2 x = 0$. Since T is firmly nonexpansive by Example 11, applying (18) with $\beta = 1$ yields $0 = \langle 0 \mid x - \overline{x}\rangle = \langle G_2 x - G_2\overline{x} \mid x - \overline{x}\rangle = \langle L^*(T(Lx) - T(L\overline{x})) \mid x - \overline{x}\rangle = \langle T(Lx) - T(L\overline{x}) \mid Lx - L\overline{x}\rangle \geqslant \|T(Lx) - T(L\overline{x})\|^2 = \|T(Lx)\|^2$. So $T(Lx) = 0$ and therefore $Lx = \mathrm{proj}_D(Lx) \in D$.

In turn, since prox_f is firmly nonexpansive (see Example 11), we derive at once from Theorem 37 the convergence of the proximal point algorithm.

Proposition 64 (proximal point algorithm) *Let $f \in \Gamma_0(\mathcal{H})$ be such that $\mathrm{Argmin}\, f \neq \varnothing$. Let $\gamma \in \left]0, +\infty\right[$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a $1/2$-relaxation sequence, and let $x_0 \in \mathcal{H}$. Set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n\big(\mathrm{prox}_{\gamma f} x_n - x_n\big). \tag{94}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges to a point in $\mathrm{Argmin}\, f$.

Remark 65 We can interpret the barycentric projection algorithm (86) as an unrelaxed instance of the proximal point algorithm (94) with $\gamma = 1$ by applying Remark 22 with $q = m$ and, for every $k \in \{1, \ldots, q\}$, $\omega_k = 1/q$, $\mathcal{G}_k = \mathcal{H}$, $L_k = \mathrm{Id}$, and $g_k = \iota_{C_k}$.

A more versatile minimization model is the following instance of the formulation discussed in Proposition 7.

Problem 66 Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$ be such that $(\mathrm{ri}\,\mathrm{dom}\, f) \cap (\mathrm{ri}\,\mathrm{dom}\, g) \neq \varnothing$ and $\lim_{\|x\|\to+\infty} f(x) + g(x) = +\infty$. The task is to

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x) + g(x). \tag{95}$$

It follows from Proposition 7i) that Problem 66 has a solution and from Proposition 7ii) that it is equivalent to Problem 47 with $A = \partial f$ and $B = \partial g$. It then remains to invoke Proposition 48 and Example 33 to obtain the following algorithm, which employs the proximity operators of f and g separately.

Proposition 67 (Douglas-Rachford splitting) *Let $(\lambda_n)_{n\in\mathbb{N}}$ be a $1/2$-relaxation sequence, let $\gamma \in \left]0, +\infty\right[$, and let $y_0 \in \mathcal{H}$. Iterate*

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} x_n = \mathrm{prox}_{\gamma g} y_n \\ z_n = \mathrm{prox}_{\gamma f}(2x_n - y_n) \\ y_{n+1} = y_n + \lambda_n(z_n - x_n). \end{array} \right. \end{array} \tag{96}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges to a solution to Problem 66.

The Douglas-Rachford algorithm was first employed in signal and image processing in [99] and it has since been applied to various problems, e.g., [73], [172], [194], [221], [250]. For a recent application to joint scale/regression estimation in statistical data analysis involving several product space reformulations, see [97]. We now present two applications to matrix optimization problems. Along the same lines, the Douglas-Rachford algorithm is also used in tensor decomposition [135].

Example 68 Let $\mathcal{H}$ be the space of $N \times N$ real symmetric matrices equipped with the Frobenius norm. We denote by $\xi_{i,j}$ the ijth component of $X \in \mathcal{H}$. Let $O \in \mathcal{H}$. The *graphical lasso problem* [133], [201] is to

$$\underset{X\in\mathcal{H}}{\text{minimize}}\;\; f(X) + \ell(X) + \mathrm{trace}(OX), \tag{97}$$

where

$$f(X) = \chi \sum_{i=1}^{N} \sum_{j=1}^{N} |\xi_{i,j}|, \quad \text{with } \chi \in [0, +\infty[, \tag{98}$$

and

$$\ell(X) = \begin{cases} -\ln\det X, & \text{if } X \text{ is positive definite;} \\ +\infty, & \text{otherwise.} \end{cases} \tag{99}$$

Problem (97) arises in the estimation of a sparse precision (i.e., inverse covariance) matrix from an observed matrix O and it has found applications in graph processing. Since $\ell \in \Gamma_0(\mathcal{H})$ is a symmetric function of the eigenvalues of its arguments, by [21, Corollary 24.65], its proximity operator at X is obtained by performing an eigendecomposition $[U, (\mu_i)_{1\leqslant i\leqslant N}] = \operatorname{eig}(X) \Leftrightarrow X = U\operatorname{Diag}(\mu_1, \ldots, \mu_N)U^\top$. Here, given $\gamma \in]0, +\infty[$, [21, Example 24.66] yields

$$\operatorname{prox}_{\gamma\ell} X = U\operatorname{Diag}\big((\operatorname{prox}_{-\gamma\ln}\mu_1, \ldots, \operatorname{prox}_{-\gamma\ln}\mu_N)\big)U^\top, \tag{100}$$

where $\operatorname{prox}_{-\gamma\ln}\colon \xi \mapsto (\xi + \sqrt{\xi^2 + 4\gamma})/2$. Let $(\lambda_n)_{n\in\mathbb{N}}$ be a $1/2$-relaxation sequence, let $\gamma \in]0, +\infty[$, and let $Y_0 \in \mathcal{H}$. Upon setting $g = \ell + \langle\cdot \mid O\rangle$, the Douglas-Rachford algorithm of (96) for solving (97) becomes

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} [U_n, (\mu_{i,n})_{1\leqslant i\leqslant N}] = \operatorname{eig}(Y_n - \gamma O) \\ X_n = U_n\operatorname{Diag}\big((\operatorname{prox}_{-\gamma\ln}\mu_{i,n})_{1\leqslant i\leqslant N}\big)U_n^\top \\ Z_n = \operatorname{soft}_{\gamma\chi}(2X_n - Y_n) \\ Y_{n+1} = Y_n + \lambda_n(Z_n - X_n), \end{array} \right. \end{array} \tag{101}$$

where $\operatorname{soft}_{\gamma\chi}$ denotes the soft-thresholding operator on $[-\gamma\chi, \gamma\chi]$ applied componentwise. Applications of (101) as well as variants with other choices of ℓ and g are discussed in [27].

Example 69 (robust PCA) Let M and N be integers such that $M \geqslant N > 0$, and let $\mathcal{H}$ be the space of $N \times M$ real matrices equipped with the Frobenius norm. The robust Principal Component Analysis (PCA) problem [56], [234] is to

$$\underset{\substack{X\in\mathcal{H}, Y\in\mathcal{H}\\ X+Y=O}}{\operatorname{minimize}} \;\; \|Y\|_{\mathrm{nuc}} + \chi\|X\|_1, \tag{102}$$

where $\|\cdot\|_1$ is the componentwise ℓ_1-norm, $\|\cdot\|_{\mathrm{nuc}}$ is the nuclear norm, and $\chi \in]0, +\infty[$. Let $X = U\operatorname{Diag}(\sigma_1, \ldots, \sigma_N)V^\top$ be the singular value decomposition of $X \in \mathcal{H}$. Then $\|X\|_{\mathrm{nuc}} = \sum_{i=1}^N \sigma_i$ and, by [21, Example 24.69],

$$\operatorname{prox}_{\chi\|\cdot\|_{\mathrm{nuc}}} X = U\operatorname{Diag}\big(\operatorname{soft}_\chi\sigma_1, \ldots, \operatorname{soft}_\chi\sigma_N\big)V^\top. \tag{103}$$

An implementation of the Douglas-Rachford algorithm in the product space $\mathcal{H} \times \mathcal{H}$ to solve (102) is detailed in [21, Example 28.6].

By combining Propositions 50, 6, and 14, together with Example 33, we obtain the convergence of the forward-backward splitting algorithm for minimization. The broad potential of this algorithm in data science was evidenced in [108]. Inertial variants are presented in [4], [8], [24], [31], [65], [93].

Proposition 70 (forward-backward splitting) *Suppose that, in Problem 66, g is differentiable everywhere and that its gradient is δ-Lipschitzian for some $\delta \in]0, +\infty[$. Let $\varepsilon \in]0, \min\{1/2, 1/\delta\}[$, let $x_0 \in \mathcal{H}$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be in $[\varepsilon, 2/(\delta(1+\varepsilon))]$, and let*

$$(\forall n \in \mathbb{N}) \quad \lambda_n \in \big[\varepsilon, (1-\varepsilon)\big(2 + \varepsilon - \delta\gamma_n/2\big)\big]. \tag{104}$$

Iterate

$$\begin{array}{l} \textit{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} u_n = x_n - \gamma_n\nabla g(x_n) \\ x_{n+1} = x_n + \lambda_n\big(\operatorname{prox}_{\gamma_n f} u_n - x_n\big). \end{array} \right. \end{array} \tag{105}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges to a solution to Problem 66.

Example 71 Let M and N be integers such that $M \geqslant N > 0$, and let $\mathcal{H}$ be the space of $N \times M$ real-valued matrices equipped with the Frobenius norm. The task is to reconstruct a low-rank matrix given its projection O onto a vector space $V \subset \mathcal{H}$. Let $L = \operatorname{proj}_V$. The problem is formulated as

$$\underset{X\in\mathcal{H}}{\operatorname{minimize}} \;\; \frac{1}{2}\|O - LX\|^2 + \chi\|X\|_{\mathrm{nuc}}, \tag{106}$$

where $\chi \in]0, +\infty[$. As seen in Example 69, the proximity operator of the nuclear norm has a closed form expression. In addition, $g\colon X \mapsto \|O - LX\|^2/2$ is convex and its gradient $\nabla g\colon X \mapsto L^*(LX - O) = LX - O$ is nonexpansive. Problem (106) can thus be solved by algorithm (105) where $f = \chi\|\cdot\|_{\mathrm{nuc}}$ and $\delta = 1$. A particular case of (106) is the matrix completion problem [57], [58], where only some components of the sought matrix are observed. If $\mathbb{K}$ denotes the set of indices of the unknown matrix components, we have $V = \big\{X \in \mathcal{H} \;\big|\; (\forall(i,j) \in \mathbb{K})\; \xi_{i,j} = 0\big\}$.

Example 72 Let X and W be mutually independent $\mathbb{R}^N$-valued random vectors. Assume that X is absolutely continuous and square-integrable, and that its probability density function is log-concave. Further, assume that W is Gaussian with zero-mean and covariance $\sigma^2\mathrm{I}_N$, where $\sigma \in]0, +\infty[$. Let $Y = X + W$. For every $y \in \mathbb{R}^N$, $Qy = \mathsf{E}(X \mid Y = y)$ is the minimum mean square error (MMSE) denoiser for X given the observation y. The properties of Q have been investigated in [145]. It can be shown that Q is the proximity operator of the conjugate of $h = (-\sigma^2\log p)^* - \|\cdot\|^2/2 \in \Gamma_0(\mathbb{R}^N)$, where p is the density of Y. Let $g\colon \mathbb{R}^N \to \mathbb{R}$ be a differentiable convex function with a δ-Lipschitzian gradient for some $\delta \in]0, +\infty[$, and let $\gamma \in]0, 2/\delta[$. The iteration

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Q\big(x_n - \gamma\nabla g(x_n)\big) \tag{107}$$

therefore turns out to be a special case of the forward-backward algorithm (105), where $f = h^*/\gamma$ and $(\forall n \in \mathbb{N})$ $\lambda_n = 1$. This algorithm is studied in [240] from a different perspective.

The projection-gradient method goes back to the classical papers [141], [171]. A version can be obtained by setting $f = \iota_C$ in Proposition 70, where C is the constraint set. Below, we describe the simpler formulation resulting from the application of Theorem 37 to $T = \operatorname{proj}_C \circ (\mathrm{Id} - \gamma\nabla g)$.

Example 73 (projection-gradient) Let C be a nonempty closed convex subset of $\mathcal{H}$ and let $g\colon \mathcal{H} \to \mathbb{R}$ be a differentiable convex function, with a δ-Lipschitzian gradient for some $\delta \in \left]0,+\infty\right[$. The task is to

$$\underset{x\in C}{\text{minimize}}\;\; g(x), \tag{108}$$

under the assumption that $\lim_{\|x\|\to+\infty} g(x) = +\infty$ or C is bounded. Let $\gamma \in \left]0,2/\delta\right[$ and set $\alpha = 2/(4-\gamma\delta)$. Furthermore, let $(\lambda_n)_{n\in\mathbb{N}}$ be an α-relaxation sequence and let $x_0 \in \mathcal{H}$. Iterate

$$\begin{array}{l}\text{for } n=0,1,\ldots\\ \left\lfloor\begin{array}{l} y_n = x_n - \gamma\nabla g(x_n)\\ x_{n+1} = x_n + \lambda_n\big(\text{proj}_C y_n - x_n\big).\end{array}\right.\end{array} \tag{109}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges to a solution to (108).

As a special case of Example 73, we obtain the convergence of the *alternating projections* algorithm [70], [171].

Example 74 (alternating projections) Let C_1 and C_2 be nonempty closed convex subsets of $\mathcal{H}$, one of which is bounded. Given $x_0 \in \mathcal{H}$, iterate

$$(\forall n\in\mathbb{N})\quad x_{n+1} = \text{proj}_{C_1}\big(\text{proj}_{C_2} x_n\big). \tag{110}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges to a solution to the constrained minimization problem

$$\underset{x\in C_1}{\text{minimize}}\;\; d_{C_2}(x). \tag{111}$$

This follows from Example 73 applied to $g = d_{C_2}^2/2$. Note that $\nabla g = \text{Id} - \text{proj}_{C_2}$ has Lipschitz constant $\delta = 1$ (see Example 5) and hence (110) is the instance of (109) obtained by setting $\gamma = 1$ and $(\forall n\in\mathbb{N})$ $\lambda_n = 1$ (see Example 10i)).

The following version of Problem 66 involves m smooth functions.

Problem 75 Let $(\omega_i)_{1\leqslant i\leqslant m}$ be real numbers in $\left]0,1\right]$ such that $\sum_{i=1}^m \omega_i = 1$. Let $f_0 \in \Gamma_0(\mathcal{H})$ and, for every $i \in \{1,\ldots,m\}$, let $\delta_i \in \left]0,+\infty\right[$ and let $f_i\colon \mathcal{H}\to\mathbb{R}$ be a differentiable convex function with a δ_i-Lipschitzian gradient. Suppose that

$$\lim_{\|x\|\to+\infty} f_0(x) + \sum_{i=1}^m \omega_i f_i(x) = +\infty. \tag{112}$$

The task is to

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f_0(x) + \sum_{i=1}^m \omega_i f_i(x). \tag{113}$$

To solve Problem 75, an option is to apply Theorem 44 to obtain a forward-backward algorithm with block-updates.

Proposition 76 ([95]) *Consider the setting of Problem 75. Let $(I_n)_{n\in\mathbb{N}}$ be a sequence of nonempty subsets of $\{1,\ldots,m\}$ such that (49) holds for some $M \in \mathbb{N}\smallsetminus\{0\}$. Let $\gamma \in \left]0, 2/\max_{1\leqslant i\leqslant m}\delta_i\right[$, let $x_0 \in \mathcal{H}$, let $(t_{i,-1})_{1\leqslant i\leqslant m} \in \mathcal{H}^m$, and iterate*

$$\begin{array}{l}\text{for } n=0,1,\ldots\\ \left\lfloor\begin{array}{l} \text{for every } i\in I_n\\ \left\lfloor\; t_{i,n} = x_n - \gamma\nabla f_i(x_n)\right.\\ \text{for every } i\in\{1,\ldots,m\}\smallsetminus I_n\\ \left\lfloor\; t_{i,n} = t_{i,n-1}\right.\\ x_{n+1} = \text{prox}_{\gamma f_0}\big(\sum_{i=1}^m \omega_i t_{i,n}\big).\end{array}\right.\end{array} \tag{114}$$

Then the following hold:

i) *Let x be a solution to Problem 75 and let $i \in \{1,\ldots,m\}$. Then $\nabla f_i(x_n) \to \nabla f_i(x)$.*
ii) *$(x_n)_{n\in\mathbb{N}}$ converges to a solution to Problem 75.*
iii) *Suppose that, for some $i \in \{0,\ldots,m\}$, f_i is strongly convex. Then $(x_n)_{n\in\mathbb{N}}$ converges linearly to the unique solution to Problem 75.*

A method related to (114) is proposed in [181]; see also [183] for a special case. Here is a data analysis application.

Example 77 Let $(e_k)_{1\leqslant k\leqslant N}$ be an orthonormal basis of $\mathcal{H}$ and, for every $k \in \{1,\ldots,N\}$, let $\psi_k \in \Gamma_0(\mathbb{R})$. For every $i \in \{1,\ldots,m\}$, let $0 \neq a_i \in \mathcal{H}$, let $\mu_i \in \left]0,+\infty\right[$, and let $\phi_i\colon \mathbb{R}\to\left[0,+\infty\right[$ be a differentiable convex function such that ϕ_i' is μ_i-Lipschitzian. The task is to

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; \sum_{k=1}^N \psi_k(\langle x \mid e_k\rangle) + \frac{1}{m}\sum_{i=1}^m \phi_i(\langle x \mid a_i\rangle). \tag{115}$$

As shown in [95], (115) is an instantiation of (113) and, given $\gamma \in \left]0, 2/(\max_{1\leqslant i\leqslant m}\mu_i\|a_i\|^2)\right[$ and subsets $(I_n)_{n\in\mathbb{N}}$ of $\{1,\ldots,m\}$ such that (49) holds, it can be solved by (114), which becomes

$$\begin{array}{l}\text{for } n=0,1,\ldots\\ \left\lfloor\begin{array}{l} \text{for every } i\in I_n\\ \left\lfloor\; t_{i,n} = x_n - \gamma\phi_i'(\langle x_n \mid a_i\rangle)a_i\right.\\ \text{for every } i\in\{1,\ldots,m\}\smallsetminus I_n\\ \left\lfloor\; t_{i,n} = t_{i,n-1}\right.\\ y_n = \sum_{i=1}^m \omega_i t_{i,n}\\ x_{n+1} = \sum_{k=1}^N \big(\text{prox}_{\gamma\psi_k}\langle y_n \mid e_k\rangle\big)e_k.\end{array}\right.\end{array} \tag{116}$$

A popular setting is obtained by choosing $\mathcal{H} = \mathbb{R}^N$ and $(e_k)_{1\leqslant k\leqslant N}$ as the canonical basis, $\alpha \in \left]0,+\infty\right[$, and, for every $k \in \{1,\ldots,K\}$, $\psi_k = \alpha|\cdot|$. This reduces (115) to

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\;\; \alpha\|x\|_1 + \sum_{i=1}^m \phi_i(\langle x \mid a_i\rangle). \tag{117}$$

Choosing, for every $i \in \{1,\ldots,m\}$, $\phi_i\colon t\mapsto |t-\eta_i|^2$ where $\eta_i \in \mathbb{R}$ models an observation, yields the lasso formulation, whereas choosing $\phi_i\colon t\mapsto \ln(1+\exp(t)) - \eta_i t$, where $\eta_i \in \{0,1\}$ models a label, yields the penalized logistic regression framework [148].

Next, we extend Problem 66 to a flexible composite minimization problem. See [36], [71], [72], [74], [75], [90], [94], [97], [182], [194], [195], [202] for concrete instantiations of this model in data science.

Problem 78 Let $\delta \in]0,+\infty[$ and let $f \in \Gamma_0(\mathcal{H})$. For every $k \in \{1,\ldots,q\}$, let $g_k \in \Gamma_0(\mathcal{G}_k)$, let $0 \neq L_k\colon \mathcal{H} \to \mathcal{G}_k$ be linear, and let $h_k\colon \mathcal{G}_k \to \mathbb{R}$ be a differentiable convex function, with a δ-Lipschitzian gradient. Suppose that $\lim_{\|x\|\to+\infty} f(x) + \sum_{k=1}^q (g_k(L_kx) + h_k(L_kx)) = +\infty$ and that

$$(\exists\, z \in \operatorname{ri}\operatorname{dom} f)(\forall k \in \{1,\ldots,q\}) \quad L_kz \in \operatorname{ri}\operatorname{dom} g_k. \tag{118}$$

The task is to

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x) + \sum_{k=1}^{q} \big(g_k(L_kx) + h_k(L_kx)\big). \tag{119}$$

Thanks to the qualification condition (118), Problem 78 is an instance of Problem 55 where $A = \partial f$ and, for every $k \in \{1,\ldots,q\}$, $B_k = \partial g_k$ and $C_k = \nabla g_k$. Since the operators $(C_k)_{1\leqslant k\leqslant q}$ are $1/\delta$-cocoercive, the iterative algorithms from Propositions 56, 57, and 58 are applicable. For example, Proposition 58 with the substitution $J_{\sigma^{-1}B_k} = \operatorname{prox}_{\sigma^{-1}g_k}$ (see Example 33) allows us to solve the problem. In particular, the resulting algorithm was proposed in [69], [175] in the case when $W = \tau\mathrm{Id}$ with $\tau \in]0,+\infty[$. See also [66], [107], [112], [113], [131], [149], [164], [237] for related work.

Example 79 Let $o \in \mathbb{R}^N$ and let $M \in \mathbb{R}^{K\times N}$ be such that $\mathrm{I}_N - M^\top M$ is positive semidefinite. Let $\varphi \in \Gamma_0(\mathbb{R}^N)$ and let C be a nonempty closed convex subset of $\mathbb{R}^N$. The denoising problem of [216] is cast as

$$\underset{x\in C}{\text{minimize}}\;\; \psi(x) + \frac{1}{2}\|x-o\|^2, \tag{120}$$

where the function

$$\psi\colon x \mapsto \varphi(x) - \inf_{y\in\mathcal{H}} \Big(\varphi(y) + \frac{1}{2}\|M(x-y)\|^2\Big) \tag{121}$$

is generally nonconvex. However, (120) is a convex problem. Further developments can be found in [1]. Note that (120) is actually equivalent to Problem 78 with $q = 2$, $\mathcal{H} = \mathbb{R}^N \times \mathbb{R}^N$, $\mathcal{G}_1 = \mathcal{H}$, $\mathcal{G}_2 = \mathbb{R}^N$, $f\colon (x,y) \mapsto \varphi(x)$, $h_1\colon (x,y) \mapsto \iota_C(x)$, $g_1\colon (x,y) \mapsto x^\top(\mathrm{I}_N - M^\top M)x/2 - \langle x \mid o\rangle + \|My\|^2/2$, $g_2 = \varphi^*$, $L_1 = \mathrm{Id}$, $L_2\colon (x,y) \mapsto M^\top M(x-y)$, and $h_2 = 0$.

Remark 80 (ADMM) Let us revisit the composite minimization problem of Proposition 7 and Example 34. Let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{G})$, and let $L\colon \mathcal{H} \to \mathcal{G}$ be linear. Suppose that $\lim_{\|x\|\to+\infty} f(x) + g(Lx) = +\infty$ and $\operatorname{ri}(L(\operatorname{dom} f)) \cap \operatorname{ri}(\operatorname{dom} g) \neq \varnothing$. Then the problem

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x) + g(Lx) \tag{122}$$

is a special case of Problem 78 and it can therefore be solved by any of the methods discussed above. Now let $\gamma \in]0,+\infty[$ and let us make the following additional assumptions:

i) $L^* \circ L$ is invertible.
ii) The operator

$$\operatorname{prox}^L_{\gamma f}\colon \mathcal{G} \to \mathcal{H}\colon y \mapsto \underset{x\in\mathcal{H}}{\operatorname{argmin}}\Big(f(x) + \frac{\|Lx-y\|^2}{2}\Big)$$

is easy to implement.

Then, given $y_0 \in \mathcal{G}$ and $z_0 \in \mathcal{G}$, the alternating-direction method of multipliers (ADMM) constructs a sequence $(x_n)_{n\in\mathbb{N}}$ that converges to a solution to (122) via the iterations [37], [126], [134], [139]

$$\begin{array}{l} \text{for } n = 0,1,\ldots \\ \left\lfloor \begin{array}{l} x_n = \operatorname{prox}^L_{\gamma f}(y_n - z_n) \\ d_n = Lx_n \\ y_{n+1} = \operatorname{prox}_{\gamma g}(d_n + z_n) \\ z_{n+1} = z_n + d_n - y_{n+1}. \end{array} \right. \end{array} \tag{123}$$

This iteration process can be viewed as an application of the Douglas-Rachford algorithm (96) to the Fenchel dual of (122) [134], [126]. Variants of this algorithm are discussed in [14], [100], [125], and applications to image recovery in [2], [3], [132], [137], [143], [217].

D. Inconsistent feasibility problems

We consider a more structured variant of Problem 61 which can also be considered as an extension of Problem 62.

Problem 81 Let C be a nonempty closed convex subset of $\mathcal{H}$ and, for every $i \in \{1,\ldots,m\}$, let $L_i\colon \mathcal{H} \to \mathcal{G}_i$ be a nonzero linear operator and let D_i be a nonempty closed convex subset of $\mathcal{G}_i$. The task is to

$$\text{find}\;\; x \in C \;\; \text{such that } (\forall i \in \{1,\ldots,m\})\; L_ix \in D_i. \tag{124}$$

To address the possibility that this problem has no solution due to modeling errors [62], [81], [248], we fix weights $(\omega_i)_{1\leqslant i\leqslant m}$ in $]0,1]$ such that $\sum_{i=1}^m \omega_i = 1$ and consider the surrogate problem

$$\underset{x\in C}{\text{minimize}}\;\; \frac{1}{2}\sum_{i=1}^{m} \omega_i d^2_{D_i}(L_ix), \tag{125}$$

where C acts as a hard constraint. This is a valid relaxation of (124) in the sense that, if (124) does have solutions, then those are the only solutions to (125). Now set $f_0 = \iota_C$. In addition, for every $i \in \{1,\ldots,m\}$, set $f_i\colon x \mapsto (1/2)d^2_{D_i}(L_ix)$ and notice that f_i is differentiable and that its gradient $\nabla f_i = L_i^* \circ (\mathrm{Id} - \operatorname{proj}_{D_i}) \circ L_i$ has Lipschitz constant $\delta_i = \|L_i\|^2$. Furthermore, (112) holds as long as C is bounded or, for some $i \in \{1,\ldots,m\}$, D_i is bounded and L_i is invertible. We have thus cast (125) as an instance of Problem 75 [95]. In view of (114), a solution is found as the limit of the sequence $(x_n)_{n\in\mathbb{N}}$ produced by the block-update algorithm

$$\begin{array}{l} \text{for } n = 0,1,\ldots \\ \left\lfloor \begin{array}{l} \text{for every } i \in I_n \\ \left\lfloor\; t_{i,n} = x_n + \gamma L_i^*\big(\operatorname{proj}_{D_i}(L_ix_n) - L_ix_n\big) \right. \\ \text{for every } i \in \{1,\ldots,m\} \smallsetminus I_n \\ \left\lfloor\; t_{i,n} = t_{i,n-1} \right. \\ x_{n+1} = \operatorname{proj}_C\big(\sum_{i=1}^m \omega_i t_{i,n}\big), \end{array} \right. \end{array} \tag{126}$$

where γ and $(I_n)_{n\in\mathbb{N}}$ are as in Proposition 76.

E. Stochastic forward-backward method

Consider the minimization of $f+g$, where $f\in\Gamma_0(\mathcal{H})$ and $g\colon\mathcal{H}\to\mathbb{R}$ is a differentiable convex function. In certain applications, it may happen that only stochastic approximations to f or g are available. A generic stochastic form of the forward-backward algorithm for such instances is [103]

$$(\forall n\in\mathbb{N})\quad x_{n+1}=x_n+\lambda_n\big(\mathrm{prox}_{\gamma_n f_n}(x_n-\gamma_n u_n)+a_n-x_n\big), \tag{127}$$

where $\gamma_n\in\,]0,+\infty[$, $\lambda_n\in\,]0,1]$, $f_n\in\Gamma_0(\mathcal{H})$ is an approximation to f, u_n is a random variable approximating $\nabla g(x_n)$, and a_n is a random variable modeling a possible additive error. When $f=f_n=0$, $\lambda_n=1$, and $a_n=0$, we recover the standard stochastic gradient method for minimizing g, which was pioneered in [129], [130].

Example 82 As in Problem 75, let $f\in\Gamma_0(\mathcal{H})$ and let $g=m^{-1}\sum_{i=1}^m g_i$, where each $g_i\colon\mathcal{H}\to\mathbb{R}$ is a differentiable convex function. The following specialization of (127) is obtained by setting, for every $n\in\mathbb{N}$, $f_n=f$ and $u_n=\nabla g_{\mathrm{i}(n)}(x_n)$, where $\mathrm{i}(n)$ is a $\{1,\ldots,m\}$-valued random variable. This leads to the incremental proximal stochastic gradient algorithm described by the update equation

$$x_{n+1}=x_n+\lambda_n\Big(\mathrm{prox}_{\gamma_n f}\big(x_n-\gamma_n\nabla g_{\mathrm{i}(n)}(x_n)\big)-x_n\Big). \tag{128}$$

For related algorithms, see [30], [119], [120], [158], [214].

Various convergence results have been established for algorithm (127). If ∇g is Lipschitzian, (127) is closely related to the fixed point iteration in Theorem 45. The almost sure convergence of $(x_n)_{n\in\mathbb{N}}$ to a minimizer of $f+g$ can be guaranteed in several scenarios [7], [103], [208]. Fixed point strategies allow us to derive convergence results such as the following.

Theorem 83 ([103]) *Let $f\in\Gamma_0(\mathcal{H})$, let $\delta\in\,]0,+\infty[$, and let $g\colon\mathcal{H}\to\mathbb{R}$ be a differentiable convex function such that ∇g is δ-Lipschitzian and $S=\mathrm{Argmin}\,(f+g)\neq\varnothing$. Let $\gamma\in\,]0,2/\delta[$ and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,1]$ such that $\sum_{n\in\mathbb{N}}\lambda_n=+\infty$. Let x_0, $(u_n)_{n\in\mathbb{N}}$, and $(a_n)_{n\in\mathbb{N}}$ be $\mathcal{H}$-valued random variables with finite second-order moments. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence produced by (127) with $\gamma_n=\gamma$ and $f_n=f$. For every $n\in\mathbb{N}$, let $\mathcal{X}_n$ be the σ-algebra generated by $(x_0,\ldots,x_n)$ and set $\zeta_n=\mathsf{E}(\|u_n-\mathsf{E}(u_n\,|\,\mathcal{X}_n)\|^2\,|\,\mathcal{X}_n)$. Assume that the following are satisfied a. s.:*

i) $\sum_{n\in\mathbb{N}}\lambda_n\sqrt{\mathsf{E}(\|a_n\|^2\,|\,\mathcal{X}_n)}<+\infty$.
ii) $\sum_{n\in\mathbb{N}}\sqrt{\lambda_n}\|\mathsf{E}(u_n\,|\,\mathcal{X}_n)-\nabla g(x_n)\|<+\infty$.
iii) $\sup_{n\in\mathbb{N}}\zeta_n<+\infty$ *and* $\sum_{n\in\mathbb{N}}\sqrt{\lambda_n\zeta_n}<+\infty$.

Then $(x_n)_{n\in\mathbb{N}}$ converges a. s. to an S-valued random variable.

Extensions of these stochastic optimization approaches can be designed by introducing an inertial parameter [207] or by bringing into play primal-dual formulations [103].

F. Random block-coordinate optimization algorithms

We design block-coordinate versions of optimization algorithms presented in Section V-C, in which blocks of variables are updated randomly.

Problem 84 For every $i\in\{1,\ldots,m\}$ and $k\in\{1,\ldots,q\}$, let $f_i\in\Gamma_0(\mathcal{H}_i)$, let $g_k\in\Gamma_0(\mathcal{G}_k)$, and let $0\neq L_{k,i}\colon\mathcal{H}_i\to\mathcal{G}_k$ be linear. Suppose that

$$(\exists\,\boldsymbol{z}\in\boldsymbol{\mathcal{H}})(\exists\,\boldsymbol{w}\in\boldsymbol{\mathcal{G}})(\forall i\in\{1,\ldots,m\})(\forall k\in\{1,\ldots,q\})$$
$$-\sum_{j=1}^q L_{j,i}^*w_j\in\partial f_i(z_i)\;\text{ and }\;\sum_{j=1}^m L_{k,j}z_j\in\partial g_k^*(w_k). \tag{129}$$

The task is to

$$\underset{\boldsymbol{x}\in\boldsymbol{\mathcal{H}}}{\text{minimize}}\;\;\sum_{i=1}^m f_i(x_i)+\sum_{k=1}^q g_k\Bigg(\sum_{i=1}^m L_{k,i}x_i\Bigg). \tag{130}$$

Let $\gamma\in\,]0,+\infty[$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,2[$, and set

$$\boldsymbol{V}=\Bigg\{(x_1,\ldots,x_m,y_1,\ldots,y_q)\in\boldsymbol{\mathcal{H}}\times\boldsymbol{\mathcal{G}}\;\Bigg|\;(\forall k\in\{1,\ldots,q\})\;y_k=\sum_{i=1}^m L_{k,i}x_i\Bigg\} \tag{131}$$

Let us decompose the projection operator $\mathrm{proj}_{\boldsymbol{V}}$ as $\mathrm{proj}_{\boldsymbol{V}}\colon\boldsymbol{x}\mapsto(Q_j\boldsymbol{x})_{1\leqslant j\leqslant m+q}$. A random block-coordinate form of the Douglas-Rachford algorithm for solving Problem 84 is [102]

$$\begin{array}{l}
\text{for } n=0,1,\ldots\\
\left\lfloor\begin{array}{l}
\text{for } i=1,\ldots,m\\
\left\lfloor\begin{array}{l}
z_{i,n+1}=z_{i,n}+\varepsilon_{i,n}\big(Q_i(\boldsymbol{x}_n,\boldsymbol{y}_n)-z_{i,n}\big)\\
x_{i,n+1}=x_{i,n}\\
\qquad+\varepsilon_{i,n}\lambda_n\big(\mathrm{prox}_{\gamma f_i}(2z_{i,n+1}-x_{i,n})-z_{i,n+1}\big)
\end{array}\right.\\
\text{for } k=1,\ldots,q\\
\left\lfloor\begin{array}{l}
w_{k,n+1}=w_{k,n}+\varepsilon_{m+k,n}\big(Q_{m+k}(\boldsymbol{x}_n,\boldsymbol{y}_n)-w_{k,n}\big)\\
y_{k,n+1}=y_{k,n}\\
\quad+\varepsilon_{m+k,n}\lambda_n\big(\mathrm{prox}_{\gamma g_k}(2w_{k,n+1}-y_{k,n})-w_{k,n+1}\big),
\end{array}\right.
\end{array}\right.
\end{array} \tag{132}$$

where $\boldsymbol{x}_n=(x_{i,n})_{1\leqslant i\leqslant m}$ and $\boldsymbol{y}_n=(y_{k,n})_{1\leqslant k\leqslant q}$. Moreover, $(\varepsilon_{j,n})_{1\leqslant j\leqslant m+q,n\in\mathbb{N}}$ are binary random variables signaling the activated components.

Proposition 85 ([102]) *Let $\boldsymbol{S}$ be the set of solutions to Problem 84 and set $D=\{0,1\}^{m+q}\smallsetminus\{\boldsymbol{0}\}$. Let $\gamma\in\,]0,+\infty[$, let $\epsilon\in\,]0,1[$, let $(\lambda_n)_{n\in\mathbb{N}}$ be in $[\epsilon,2-\epsilon]$, let $\boldsymbol{x}_0$ and $\boldsymbol{z}_0$ be $\boldsymbol{\mathcal{H}}$-valued random variables, let $\boldsymbol{y}_0$ and $\boldsymbol{w}_0$ be $\boldsymbol{\mathcal{G}}$-valued random variables, and let $(\boldsymbol{\varepsilon}_n)_{n\in\mathbb{N}}$ be identically distributed D-valued random variables. In addition, suppose that the following hold:*

i) *For every $n\in\mathbb{N}$, $\boldsymbol{\varepsilon}_n$ and $(\boldsymbol{x}_0,\ldots,\boldsymbol{x}_n,\boldsymbol{y}_0,\ldots,\boldsymbol{y}_n)$ are mutually independent.*
ii) $(\forall j\in\{1,\ldots,m+q\})\;\mathsf{Prob}\,[\varepsilon_{j,0}=1]>0$.

Then the sequence $(\boldsymbol{z}_n)_{n\in\mathbb{N}}$ generated by (132) converges a. s. to an $\boldsymbol{S}$-valued random variable.

Applications based on Proposition 85 appear in the areas of machine learning [96] and binary logistic regression [42].

If the functions $(g_k)_{1\leqslant k\leqslant q}$ are differentiable in Problem 84, a block-coordinate version of the forward-backward algorithm can also be employed, namely,

$$\begin{array}{l}\text{for } n=0,1,\ldots\\ \left\lfloor \begin{array}{l}\text{for } i=1,\ldots,m\\ \left\lfloor \begin{array}{l} r_{i,n}=\varepsilon_{i,n}\Big(x_{i,n}-\\ \qquad \gamma_{i,n}\sum_{k=1}^{q}L_{k,i}^{*}\Big(\nabla g_k\big(\sum_{j=1}^{m}L_{k,j}x_{j,n}\big)\Big)\Big)\\ x_{i,n+1}=x_{i,n}+\varepsilon_{i,n}\lambda_n\big(\operatorname{prox}_{\gamma_{i,n}f_i}r_{i,n}-x_{i,n}\big),\end{array}\right.\end{array}\right.\end{array} \tag{133}$$

where $\gamma_{i,n}\in\,]0,+\infty[$ and $\lambda_n\in\,]0,1]$. The convergence of (133) has been investigated in various settings in terms of the expected value of the cost function [189], [203], [204], [211], the mean square convergence of the iterates [104], [203], [204], or the almost sure convergence of the iterates [102], [211]. It is shown in [211] that algorithms such as the so-called *random Kaczmarz method* to solve standard linear systems are special cases of (133).

A noteworthy feature of the block-coordinate forward-backward algorithm (133) is that, at iteration n, it allows for the use of distinct parameters $(\gamma_{i,n})_{1\leqslant i\leqslant m}$ to update each component. This was observed to be beneficial to the convergence profile in several applications [76], [203]. See also [211] for further developments along these lines.

G. Block-iterative multivariate minimization algorithms

We investigate a specialization of a primal-dual version of the multivariate inclusion Problem 59 in the context of Problem 84.

Problem 86 Consider the setting of Problem 84. The task is to solve the primal minimization problem

$$\underset{\boldsymbol{x}\in\boldsymbol{\mathcal{H}}}{\text{minimize}}\;\sum_{i=1}^{m}f_i(x_i)+\sum_{k=1}^{q}g_k\Bigg(\sum_{i=1}^{m}L_{k,i}x_i\Bigg), \tag{134}$$

along with its dual problem

$$\underset{\boldsymbol{v}\in\boldsymbol{\mathcal{G}}}{\text{minimize}}\;\sum_{i=1}^{m}f_i^{*}\Bigg(-\sum_{k=1}^{q}L_{k,i}^{*}v_k\Bigg)+\sum_{k=1}^{q}g_k^{*}(v_k). \tag{135}$$

We solve Problem 86 with algorithm (82) by replacing $J_{\gamma_{i,n}A_i}$ by $\operatorname{prox}_{\gamma_{i,n}f_i}$ and $J_{\mu_{k,n}B_k}$ by $\operatorname{prox}_{\mu_{k,n}g_k}$. This block-iterative method then produces a sequence $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ which converges to a solution to (134) and a sequence $(\boldsymbol{v}_n)_{n\in\mathbb{N}}$ which converges to a solution to (135) [92].

Examples of problems that conform to the format of Problems 84 or 86 are encountered in image processing [28], [43], [46] as well as in machine learning [5], [11], [96], [156], [157], [178], [236], [251].

H. Splitting based on Bregman distances

The notion of a Bregman distance goes back to [40] and it has been used since the 1980s in signal recovery; see [55], [64]. Let $\varphi\in\Gamma_0(\mathcal{H})$ be strictly convex, and differentiable on $\operatorname{int}\operatorname{dom}\varphi\neq\varnothing$ (more precisely, we require a Legendre function, see [18], [19] for the technical details). The associated *Bregman distance* between two points x and y in $\mathcal{H}$ is

$$D_\varphi(x,y)=\begin{cases}\varphi(x)-\varphi(y)-\langle x-y\mid\nabla\varphi(y)\rangle, & \\ & \text{if } \; y\in\operatorname{int}\operatorname{dom}\varphi;\\ +\infty, & \text{otherwise.}\end{cases} \tag{136}$$

This construction captures many interesting discrepancy measures in data analysis such as the Kullback-Leibler divergence. Another noteworthy instance is when $\varphi=\|\cdot\|^2/2$, which yields $D_\varphi(x,y)=\|x-y\|^2/2$ and suggests extending standard tools such as projection and proximity operators (see Theorems 1 and 2) by replacing the quadratic kernel by a Bregman distance [18], [19], [40], [63], [124], [225]. For instance, under mild conditions on $f\in\Gamma_0(\mathcal{H})$ [19], the *Bregman proximal point* of $y\in\operatorname{int}\operatorname{dom}\varphi$ relative to f is the unique point $\operatorname{prox}_f^{\varphi}y$ which solves

$$\underset{p\in\operatorname{int}\operatorname{dom}\varphi}{\text{minimize}}\;f(p)+D_\varphi(p,y). \tag{137}$$

The *Bregman projection* $\operatorname{proj}_C^{\varphi}y$ of y onto a nonempty closed convex set C in $\mathcal{H}$ is obtained by setting $f=\iota_C$ above. Various algorithms such as the POCS algorithm (85) or the proximal point algorithm (94) have been extended in the context of Bregman distances [18], [19]. For instance [18] establishes the convergence to a solution to Problem 61 of a notable extension of POCS in which the sets are Bregman-projected onto in arbitrary order, namely

$$(\forall n\in\mathbb{N})\quad x_{n+1}=\operatorname{proj}_{C_{\mathrm{i}(n)}}^{\varphi}x_n, \tag{138}$$

where $\mathrm{i}\colon\mathbb{N}\to\{1,\ldots,m\}$ is such that, for every $p\in\mathbb{N}$ and every $j\in\{1,\ldots,m\}$, there exists $n\geqslant p$ such that $\mathrm{i}(n)=j$.

A motivation for such extensions is that, for certain functions, proximal points are easier to compute in the Bregman sense than in the standard quadratic sense [16], [98], [190]. Some work has also focused on monotone operator splitting using Bregman distances as an extension of standard methods [98]. The Bregman version of the basic forward-backward minimization method of Proposition 70, namely,

$$\begin{array}{l}\text{for } n=0,1,\ldots\\ \left\lfloor \begin{array}{l} u_n=\nabla\varphi(x_n)-\gamma_n\nabla g(x_n)\\ x_{n+1}=\big(\nabla\varphi+\gamma_n\partial f\big)^{-1}u_n\end{array}\right.\end{array} \tag{139}$$

has also been investigated in [16], [49], [190] (note that the standard quadratic kernel corresponds to $\nabla\varphi=\mathrm{Id}$). In these papers, it was shown to converge in instances when (105) cannot be used because ∇g is not Lipschitzian.

VI. Fixed point modeling of Nash equilibria

In addition to the notation of Section II-A, given $i\in\{1,\ldots,m\}$, $x_i\in\mathcal{H}_i$, and $\boldsymbol{y}\in\boldsymbol{\mathcal{H}}$, we set

$$\begin{cases}\boldsymbol{\mathcal{H}}_{\smallsetminus i}=\mathcal{H}_1\times\cdots\times\mathcal{H}_{i-1}\times\mathcal{H}_{i+1}\times\cdots\times\mathcal{H}_m\\ \boldsymbol{y}_{\smallsetminus i}=(y_j)_{1\leqslant j\leqslant m,j\neq i}\\ (x_i;\boldsymbol{y}_{\smallsetminus i})=(y_1,\ldots,y_{i-1},x_i,y_{i+1},\ldots,y_m).\end{cases} \tag{140}$$

In various problems arising in signal recovery [9], [10], [28], [43], [46], [114], [115], [121], telecommunications [168],

[215], machine learning [38], [116], network science [244], [246], and control [26], [33], [254], the solution is not a single vector but a collections of vectors $\boldsymbol{x} = (x_1, \ldots, x_m) \in \boldsymbol{\mathcal{H}}$ representing the actions of m competing players. Oftentimes, such solutions cannot be modeled via a standard minimization problem of the form

$$\underset{\boldsymbol{x}\in\boldsymbol{\mathcal{H}}}{\text{minimize}} \;\; \boldsymbol{h}(\boldsymbol{x}) \tag{141}$$

for some function $\boldsymbol{h}\colon \boldsymbol{\mathcal{H}} \to \left]-\infty,+\infty\right]$, but rather as a Nash equilibrium [187]. In this game-theoretic setting [167], player i aims at minimizing his individual loss (or negative payoff) function $\boldsymbol{h}_i\colon \boldsymbol{\mathcal{H}} \to \left]-\infty,+\infty\right]$, that incorporates the actions of the other players. An action profile $\overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}}$ is called a *Nash equilibrium* if unilateral deviations from it are not profitable, i.e.,

$$(\forall i \in \{1,\ldots,m\}) \quad \boldsymbol{h}_i(\overline{x}_i; \overline{\boldsymbol{x}}_{\smallsetminus i}) = \min_{x_i\in\mathcal{H}_i} \boldsymbol{h}_i(x_i; \overline{\boldsymbol{x}}_{\smallsetminus i}). \tag{142}$$

In other words, if

$$\begin{aligned}&\text{best}_i\colon \boldsymbol{\mathcal{H}}_{\smallsetminus i} \to 2^{\mathcal{H}_i}\colon \boldsymbol{x}_{\smallsetminus i} \mapsto \\ &\big\{x_i \in \mathcal{H}_i \;\big|\; (\forall y_i \in \mathcal{H}_i)\; \boldsymbol{h}_i(y_i; \boldsymbol{x}_{\smallsetminus i}) \geqslant \boldsymbol{h}_i(x_i; \boldsymbol{x}_{\smallsetminus i})\big\}\end{aligned} \tag{143}$$

denotes the *best response operator* of player i, $\overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}}$ is a Nash equilibrium if and only if

$$(\forall i \in \{1,\ldots,m\}) \quad \overline{x}_i \in \text{best}_i(\overline{\boldsymbol{x}}_{\smallsetminus i}). \tag{144}$$

This property can also be expressed in terms of the set-valued operator

$$\boldsymbol{B}\colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}\colon \boldsymbol{x} \mapsto \text{best}_1(\boldsymbol{x}_{\smallsetminus 1}) \times \cdots \times \text{best}_m(\boldsymbol{x}_{\smallsetminus m}). \tag{145}$$

Thus, a point $\overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}}$ is a Nash equilibrium if and only if it is a fixed point of $\boldsymbol{B}$ in the sense that $\overline{\boldsymbol{x}} \in \boldsymbol{B}\overline{\boldsymbol{x}}$.

A. Cycles in the POCS algorithm

Let us go back to feasibility and Problem 61. The POCS algorithm (85) converges to a solution to the feasibility problem (83) when one exists. Now suppose that Problem 61 is inconsistent, with C_1 bounded. Then, as seen in Example 74, in the case of $m = 2$ sets, the sequence $(x_{2n})_{n\in\mathbb{N}}$ produced by the alternating projection algorithm (110), written as

$$\begin{aligned}&\text{for } n = 0,1,\ldots\\ &\left\lfloor\begin{aligned} x_{2n+1} &= \text{proj}_{C_2} x_{2n}\\ x_{2n+2} &= \text{proj}_{C_1} x_{2n+1},\end{aligned}\right.\end{aligned} \tag{146}$$

converges to a point $\overline{x}_1 \in \text{Fix}\,(\text{proj}_{C_1} \circ \text{proj}_{C_2})$, i.e., to a minimizer of d_{C_2} over C_1. More precisely [70], if we set $\overline{x}_2 = \text{proj}_{C_2}\overline{x}_1$, then $\overline{x}_1 = \text{proj}_{C_1}\overline{x}_2$ and $(\overline{x}_1, \overline{x}_2)$ solves

$$\underset{x_1\in C_1,\, x_2\in C_2}{\text{minimize}} \;\; \|x_1 - x_2\|. \tag{147}$$

An extension of the alternating projection method (146) to m sets is the POCS algorithm (85), which we write as

$$\begin{aligned}&\text{for } n = 0,1,\ldots\\ &\left\lfloor\begin{aligned} x_{mn+1} &= \text{proj}_{C_m} x_{mn}\\ x_{mn+2} &= \text{proj}_{C_{m-1}} x_{mn+1}\\ &\;\;\vdots\\ x_{mn+m} &= \text{proj}_{C_1} x_{mn+m-1}.\end{aligned}\right.\end{aligned} \tag{148}$$

As first shown in [146] (this is also a consequence of Theorem 41), for every $i \in \{1,\ldots,m\}$, $(x_{mn+i})_{n\in\mathbb{N}}$ converges to a point $\overline{x}_{m+1-i} \in C_{m+1-i}$; in addition $(\overline{x}_i)_{1\leqslant i\leqslant m}$ forms a *cycle* in the sense that (see Fig. 6)

$$\begin{aligned}\overline{x}_1 = \text{proj}_{C_1}\overline{x}_2,\; \ldots,\; \overline{x}_{m-1} = \text{proj}_{C_{m-1}}\overline{x}_m,&\\ \text{and} \quad \overline{x}_m = \text{proj}_{C_m}\overline{x}_1.&\end{aligned} \tag{149}$$

As shown in [13], in stark contrast with the case of $m = 2$

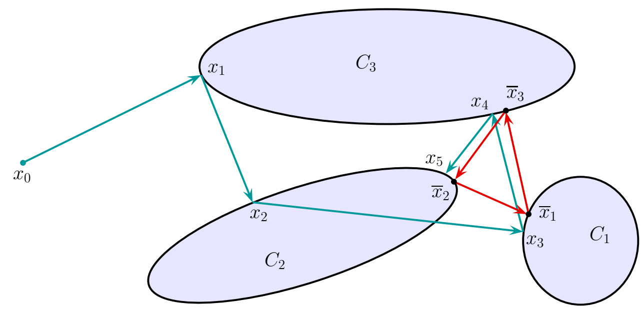

Fig. 6: The POCS algorithm with $m = 3$ sets and initialized at x_0 produces the cycle $(\overline{x}_1, \overline{x}_2, \overline{x}_3)$.

sets and (147), there exists no function $\Phi\colon \mathcal{H}^m \to \mathbb{R}$ such that cycles solve the minimization problem

$$\underset{x_1\in C_1,\ldots,\, x_m\in C_m}{\text{minimize}} \;\; \Phi(x_1,\ldots,x_m), \tag{150}$$

which deprives cycles of a minimization interpretation. Nonetheless, cycles are equilibria in a more general sense, which can be described from three different perspectives.

- Fixed point theory: Define two operators $\boldsymbol{P}$ and $\boldsymbol{L}$ from $\mathcal{H}^m$ to $\mathcal{H}^m$ by

$$\begin{cases}\boldsymbol{P}\colon \boldsymbol{x} \mapsto (\text{proj}_{C_1}x_1, \ldots, \text{proj}_{C_m}x_m)\\ \boldsymbol{L}\colon \boldsymbol{x} \mapsto (x_2, \ldots, x_m, x_1).\end{cases} \tag{151}$$

  Then, in view of (149), the set of cycles is precisely the set of fixed points of $\boldsymbol{P} \circ \boldsymbol{L}$, which is also the set of fixed points of $\boldsymbol{T} = \boldsymbol{P} \circ \boldsymbol{F}$, where $\boldsymbol{F} = (\mathbf{Id} + \boldsymbol{L})/2$ (see [21, Corollary 26.3]). Since Example 11 implies that $\boldsymbol{P}$ is firmly nonexpansive and since $\boldsymbol{L}$ is nonexpansive, $\boldsymbol{F}$ is firmly nonexpansive as well. It thus follows from Example 19, that the cycles are the fixed points of the $2/3$-averaged operator $\boldsymbol{T}$.
- Game theory: Consider a game in $\mathcal{H}^m$ in which the goal of player i is to minimize the loss

$$\boldsymbol{h}_i\colon (x_i; \boldsymbol{x}_{\smallsetminus i}) \mapsto \iota_{C_i}(x_i) + \frac{1}{2}\|x_i - x_{i+1}\|^2, \tag{152}$$

  i.e., to be in C_i and as close as possible to the action of player $i+1$ (with the convention $x_{m+1} = x_1$). Then a cycle $(\overline{x}_1, \ldots, \overline{x}_m)$ is a solution to (142) and therefore a Nash equilibrium. Let us note that the best response operator of player i is $\text{best}_i\colon \boldsymbol{x}_{\smallsetminus i} \mapsto \text{proj}_{C_i} x_{i+1}$.

- Monotone inclusion: Applying Fermat's rule to each line of (142) in the setting of (152), and using (14), we obtain

$$\begin{cases} 0 \in N_{C_1}\overline{x}_1 + \overline{x}_1 - \overline{x}_2 \\ \quad\vdots \\ 0 \in N_{C_{m-1}}\overline{x}_{m-1} + \overline{x}_{m-1} - \overline{x}_m \\ 0 \in N_{C_m}\overline{x}_m + \overline{x}_m - \overline{x}_1. \end{cases} \tag{153}$$

In terms of the maximally monotone operator $\boldsymbol{A} = N_{C_1\times\cdots\times C_m}$ and the cocoercive operator

$$\boldsymbol{B}\colon \boldsymbol{x} \mapsto (x_1 - x_2, \ldots, x_{m-1} - x_m, x_m - x_1), \tag{154}$$

(153) can be rewritten as an instance of Problem 47 in $\mathcal{H}^m$, namely, $\mathbf{0} \in \boldsymbol{A}\overline{\boldsymbol{x}} + \boldsymbol{B}\overline{\boldsymbol{x}}$.

B. Proximal cycles

We have seen in Section VI-A a first example of a Nash equilibrium. This setting can be extended by replacing the indicator function ι_{C_i} in (152) by a general function $\varphi_i \in \Gamma_0(\mathcal{H})$ modeling the self-loss of player i, i.e.,

$$\boldsymbol{h}_i\colon (x_i; \boldsymbol{x}_{\smallsetminus i}) \mapsto \varphi_i(x_i) + \frac{1}{2}\|x_i - x_{i+1}\|^2. \tag{155}$$

The solutions to the resulting problem (142) are *proximal cycles*, i.e., m-tuples $(\overline{x}_i)_{1\leqslant i\leqslant m} \in \mathcal{H}^m$ such that

$$\overline{x}_1 = \operatorname{prox}_{\varphi_1}\overline{x}_2,\ \ldots,\ \overline{x}_{m-1} = \operatorname{prox}_{\varphi_{m-1}}\overline{x}_m, \quad \text{and} \quad \overline{x}_m = \operatorname{prox}_{\varphi_m}\overline{x}_1. \tag{156}$$

Furthermore, the equivalent monotone inclusion and fixed point representations of the cycles in Section VI-A remain true with

$$\boldsymbol{P}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}\colon \boldsymbol{x} \mapsto \big(\operatorname{prox}_{\varphi_1}x_1, \ldots, \operatorname{prox}_{\varphi_m}x_m\big) \tag{157}$$

and $\boldsymbol{A} = \partial\boldsymbol{f}$, where $\boldsymbol{f}\colon \boldsymbol{x} \mapsto \sum_{i=1}^m \varphi_i(x_i)$. Here, the best response operator of player i is $\operatorname{best}_i\colon \boldsymbol{x}_{\smallsetminus i} \mapsto \operatorname{prox}_{\varphi_i}x_{i+1}$. Examples of such cycles appear in [43], [108].

C. Construction of Nash equilibria

A more structured version of the Nash equilibrium formulation (142), which captures (155) and therefore (152), is provided next.

Problem 87 For every $i \in \{1, \ldots, m\}$, let $\psi_i \in \Gamma_0(\mathcal{H}_i)$, let $\boldsymbol{f}_i\colon \boldsymbol{\mathcal{H}} \to \left]-\infty, +\infty\right]$, let $\boldsymbol{g}_i\colon \boldsymbol{\mathcal{H}} \to \left]-\infty, +\infty\right]$ be such that, for every $\boldsymbol{x} \in \boldsymbol{\mathcal{H}}$, $\boldsymbol{f}_i(\cdot; \boldsymbol{x}_{\smallsetminus i}) \in \Gamma_0(\mathcal{H}_i)$ and $\boldsymbol{g}_i(\cdot; \boldsymbol{x}_{\smallsetminus i}) \in \Gamma_0(\mathcal{H}_i)$. The task is to

$$\text{find } \overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}} \quad \text{such that} \quad (\forall i \in \{1, \ldots, m\}) \quad \overline{x}_i \in \underset{x_i\in\mathcal{H}_i}{\operatorname{Argmin}}\ \psi_i(x_i) + \boldsymbol{f}_i(x_i; \overline{\boldsymbol{x}}_{\smallsetminus i}) + \boldsymbol{g}_i(x_i; \overline{\boldsymbol{x}}_{\smallsetminus i}). \tag{158}$$

Under suitable assumptions on $(\boldsymbol{f}_i)_{1\leqslant i\leqslant m}$ and $(\boldsymbol{g}_i)_{1\leqslant i\leqslant m}$, monotone operator splitting strategies can be contemplated to solve Problem 87. This approach was initiated in [79] in a special case of the following setting, which reduces to that investigated in [45] when $(\forall i \in \{1, \ldots, m\})\ \psi_i = 0$.

Assumption 88 In Problem 87, the functions $(\boldsymbol{f}_i)_{1\leqslant i\leqslant m}$ coincide with a function $\boldsymbol{f} \in \Gamma_0(\boldsymbol{\mathcal{H}})$. For every $i \in \{1, \ldots, m\}$ and every $\boldsymbol{x} \in \boldsymbol{\mathcal{H}}$, $\boldsymbol{g}_i(\cdot; \boldsymbol{x}_{\smallsetminus i})$ is differentiable on $\mathcal{H}_i$ and $\nabla_i\,\boldsymbol{g}_i(\boldsymbol{x})$ denotes its derivative relative to x_i. Moreover,

$$(\forall \boldsymbol{x} \in \boldsymbol{\mathcal{H}})(\forall \boldsymbol{y} \in \boldsymbol{\mathcal{H}}) \quad \sum_{i=1}^m \langle \nabla_i\,\boldsymbol{g}_i(\boldsymbol{x}) - \nabla_i\,\boldsymbol{g}_i(\boldsymbol{y}) \mid x_i - y_i\rangle \geqslant 0, \tag{159}$$

and

$$(\exists\, \boldsymbol{z} \in \boldsymbol{\mathcal{H}}) \quad -\big(\nabla_1\,\boldsymbol{g}_1(\boldsymbol{z}), \ldots, \nabla_m\,\boldsymbol{g}_m(\boldsymbol{z})\big) \in \partial\boldsymbol{f}(\boldsymbol{z}) + \underset{i=1}{\overset{m}{\times}} \partial\psi_i(z_i). \tag{160}$$

In the context of Assumption 88, let us introduce the maximally monotone operators on $\boldsymbol{\mathcal{H}}$

$$\begin{cases} \boldsymbol{A} = \partial\boldsymbol{f} \\ \boldsymbol{B}\colon \boldsymbol{x} \mapsto \times_{i=1}^m \partial\psi_i(x_i) \\ \boldsymbol{C}\colon \boldsymbol{x} \mapsto \big(\nabla_1\,\boldsymbol{g}_1(\boldsymbol{x}), \ldots, \nabla_m\,\boldsymbol{g}_m(\boldsymbol{x})\big). \end{cases} \tag{161}$$

Then the solutions to the inclusion problem (see Problem 51) $\mathbf{0} \in \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{x} + \boldsymbol{C}\boldsymbol{x}$ solve Problem 87 [45]. In turn, applying the splitting scheme of Proposition 52 leads to the following implementation.

Proposition 89 *Consider the setting of Assumption 88 with the additional requirement that, for some* $\delta \in \left]0, +\infty\right[$,

$$(\forall \boldsymbol{x} \in \boldsymbol{\mathcal{H}})(\forall \boldsymbol{y} \in \boldsymbol{\mathcal{H}}) \quad \sum_{i=1}^m \|\nabla_i\,\boldsymbol{g}_i(\boldsymbol{x}) - \nabla_i\,\boldsymbol{g}_i(\boldsymbol{y})\|^2 \leqslant \delta^2 \sum_{i=1}^m \|x_i - y_i\|^2. \tag{162}$$

Let $\varepsilon \in \left]0, 1/(2+\delta)\right[$, *let* $(\gamma_n)_{n\in\mathbb{N}}$ *be in* $[\varepsilon, (1-\varepsilon)/(1+\delta)]$, *let* $\boldsymbol{x}_0 \in \boldsymbol{\mathcal{H}}$, *and let* $\boldsymbol{v}_0 \in \boldsymbol{\mathcal{H}}$. *Iterate*

$$\begin{array}{l} \textit{for}\ n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} \textit{for}\ i = 1, \ldots, m \\ \lfloor\ y_{i,n} = x_{i,n} - \gamma_n\big(\nabla_i\,\boldsymbol{g}_i(\boldsymbol{x}_n) + v_{i,n}\big) \\ \boldsymbol{p}_n = \operatorname{prox}_{\gamma_n\boldsymbol{f}}\,\boldsymbol{y}_n \\ \textit{for}\ i = 1, \ldots, m \\ \left\lfloor \begin{array}{l} q_{i,n} = v_{i,n} + \gamma_n\big(x_{i,n} - \operatorname{prox}_{\psi_i/\gamma_n}(v_{i,n}/\gamma_n + x_{i,n})\big) \\ x_{i,n+1} = x_{i,n} - y_{i,n} + p_{i,n} - \gamma_n\big(\nabla_i\,\boldsymbol{g}_i(\boldsymbol{p}_n) + q_{i,n}\big) \\ v_{i,n+1} = q_{i,n} + \gamma_n(p_{i,n} - x_{i,n}). \end{array}\right. \end{array}\right. \end{array} \tag{163}$$

Then there exists a solution $\overline{\boldsymbol{x}}$ *to Problem 87 such that, for every* $i \in \{1, \ldots, m\}$, $x_{i,n} \to \overline{x}_i$.

Example 90 Let $\varphi_1\colon \mathcal{H}_1 \to \mathbb{R}$ be convex and differentiable with a δ_1-Lipschitzian gradient, let $\varphi_2\colon \mathcal{H}_2 \to \mathbb{R}$ be convex and differentiable with a δ_2-Lipschitzian gradient, let $L\colon \mathcal{H}_1 \to \mathcal{H}_2$ be linear, and let $C_1 \subset \mathcal{H}_1$, $C_2 \subset \mathcal{H}_2$, and $\boldsymbol{D} \subset \mathcal{H}_1 \times \mathcal{H}_2$ be nonempty closed convex sets. Suppose that there exists $\boldsymbol{z} \in \mathcal{H}_1 \times \mathcal{H}_2$ such that $-(\nabla\varphi_1(z_1) +$

$L^*z_2, \nabla\varphi_2(z_2) - Lz_1) \in N_{\boldsymbol{D}}(z_1, z_2) + N_{C_1}z_1 \times N_{C_2}z_2$. Then the 2-player game

$$\begin{cases} \overline{x}_1 \in \underset{x_1 \in C_1}{\text{Argmin}} \;\; \iota_{\boldsymbol{D}}(x_1, \overline{x}_2) + \varphi_1(x_1) + \langle Lx_1 \mid \overline{x}_2\rangle \\ \overline{x}_2 \in \underset{x_2 \in C_2}{\text{Argmin}} \;\; \iota_{\boldsymbol{D}}(\overline{x}_1, x_2) + \varphi_2(x_2) - \langle L\overline{x}_1 \mid x_2\rangle \end{cases} \tag{164}$$

is an instance of Problem 87 with $\boldsymbol{f}_1 = \boldsymbol{f}_2 = \iota_{\boldsymbol{D}}$, $\psi_1 = \iota_{C_1}$, $\psi_2 = \iota_{C_2}$, and

$$\begin{cases} \boldsymbol{g}_1 \colon (x_1, x_2) \mapsto \varphi_1(x_1) + \langle Lx_1 \mid x_2\rangle \\ \boldsymbol{g}_2 \colon (x_1, x_2) \mapsto \varphi_2(x_2) - \langle Lx_1 \mid x_2\rangle. \end{cases} \tag{165}$$

In addition, Assumption 88 is satisfied, as well as (162) with $\delta = \max\{\delta_1, \delta_2\} + \|L\|$. Moreover, in view of (11), algorithm (163) becomes

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} y_{1,n} = x_{1,n} - \gamma_n\big(\nabla\varphi_1(x_{1,n}) + L^*x_{2,n} + v_{1,n}\big) \\ y_{2,n} = x_{2,n} - \gamma_n\big(\nabla\varphi_2(x_{2,n}) - Lx_{1,n} + v_{2,n}\big) \\ \boldsymbol{p}_n = \text{proj}_{\boldsymbol{D}}\, \boldsymbol{y}_n \\ q_{1,n} = v_{1,n} + \gamma_n\big(x_{1,n} - \text{proj}_{C_1}(v_{1,n}/\gamma_n + x_{1,n})\big) \\ q_{2,n} = v_{2,n} + \gamma_n\big(x_{2,n} - \text{proj}_{C_2}(v_{2,n}/\gamma_n + x_{2,n})\big) \\ x_{1,n+1} = x_{1,n} - y_{1,n} + p_{1,n} \\ \qquad\qquad -\gamma_n\big(\nabla\varphi_1(p_{1,n}) + L^*p_{2,n} + q_{1,n}\big) \\ x_{2,n+1} = x_{2,n} - y_{2,n} + p_{2,n} \\ \qquad\qquad -\gamma_n\big(\nabla\varphi_2(p_{2,n}) - Lp_{1,n} + q_{2,n}\big) \\ v_{1,n+1} = q_{1,n} + \gamma_n(p_{1,n} - x_{1,n}) \\ v_{2,n+1} = q_{2,n} + \gamma_n(p_{2,n} - x_{2,n}). \end{array} \right. \end{array} \tag{166}$$

Condition (162) means that the operator $\boldsymbol{C}$ of (161) is δ-Lipschitzian. The stronger assumption that it is cocoercive, allows us to bring into play the three-operator splitting algorithm of Proposition 53 to solve Problem 87.

Proposition 91 *Consider the setting of Assumption 88 with the additional requirement that, for some $\beta \in \left]0, +\infty\right[$,*

$$(\forall \boldsymbol{x} \in \boldsymbol{\mathcal{H}})(\forall \boldsymbol{y} \in \boldsymbol{\mathcal{H}}) \quad \sum_{i=1}^m \langle x_i - y_i \mid \nabla_i\, \boldsymbol{g}_i(\boldsymbol{x}) - \nabla_i\, \boldsymbol{g}_i(\boldsymbol{y})\rangle \geqslant \beta \sum_{i=1}^m \|\nabla_i\, \boldsymbol{g}_i(\boldsymbol{x}) - \nabla_i\, \boldsymbol{g}_i(\boldsymbol{y})\|^2. \tag{167}$$

Let $\gamma \in \left]0, 2\beta\right[$ and set $\alpha = 2\beta/(4\beta - \gamma)$. Furthermore, let $(\lambda_n)_{n\in\mathbb{N}}$ be an α-relaxation sequence and let $\boldsymbol{y}_0 \in \boldsymbol{\mathcal{H}}$. Iterate

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} \text{for } i = 1, \ldots, m \\ \left\lfloor \begin{array}{l} x_{i,n} = \text{prox}_{\gamma\psi_i}\, y_{i,n} \\ r_{i,n} = y_{i,n} + \gamma\nabla_i\, \boldsymbol{g}_i(\boldsymbol{x}_n) \end{array} \right. \\ \boldsymbol{z}_n = \text{prox}_{\gamma \boldsymbol{f}}(2\boldsymbol{x}_n - \boldsymbol{r}_n) \\ \boldsymbol{y}_{n+1} = \boldsymbol{y}_n + \lambda_n(\boldsymbol{z}_n - \boldsymbol{x}_n). \end{array} \right. \end{array} \tag{168}$$

Then there exists a solution $\overline{\boldsymbol{x}}$ to Problem 87 such that, for every $i \in \{1, \ldots, m\}$, $x_{i,n} \to \overline{x}_i$.

Example 92 For every $i \in \{1, \ldots, m\}$, let $C_i \subset \mathcal{H}_i$ be a nonempty closed convex set, let $L_i \colon \mathcal{H}_i \to \mathcal{G}$ be linear, and let $o_i \in \mathcal{G}$. The task is to solve the Nash equilibrium (with the convention $L_{m+1}\overline{x}_{m+1} = L_1\overline{x}_1$)

$$\text{find } \overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}} \quad \text{such that} \quad (\forall i \in \{1, \ldots, m\}) \quad \overline{x}_i \in \underset{x_i \in C_i}{\text{Argmin}} \;\; \psi_i(x_i) + \frac{\|L_ix_i + L_{i+1}\overline{x}_{i+1} - o_i\|^2}{2}. \tag{169}$$

Here, the action of player i must lie in C_i, and it is further penalized by ψ_i and the proximity of the linear mixture $L_ix_i + L_{i+1}\overline{x}_{i+1}$ to some vector o_i. For instance if, for every $i \in \{1, \ldots, m\}$, $C_i = \mathcal{H}_i$, $o_i = 0$, and $L_i = (-1)^i\text{Id}$, we recover the setting of Section VI-B. The equilibrium (169) is an instantiation of Problem 87 with $\boldsymbol{f}_1 = \boldsymbol{f}_2 \colon \boldsymbol{x} \mapsto \sum_{i=1}^m \iota_{C_i}(x_i)$ and, for every $i \in \{1, \ldots, m\}$, $\boldsymbol{g}_i \colon \boldsymbol{x} \mapsto \|L_ix_i + L_{i+1}x_{i+1} - o_i\|^2/2$. In addition, as in [45, Section 9.4.3], (167) holds with $\beta = (2\max_{1\leqslant i\leqslant m} \|L_i\|^2)^{-1}$. Finally, (168) reduces to (with the convention $L_{m+1}x_{m+1,n} = L_1x_{1,n}$)

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} \text{for } i = 1, \ldots, m \\ \left\lfloor \begin{array}{l} x_{i,n} = \text{prox}_{\gamma\psi_i}\, y_{i,n} \\ r_{i,n} = y_{i,n} + \gamma L_i^*(L_ix_{i,n} + L_{i+1}x_{i+1,n} - o_i) \\ z_{i,n} = \text{proj}_{C_i}(2x_{i,n} - r_{i,n}) \\ y_{i,n+1} = y_{i,n} + \lambda_n(z_{i,n} - x_{i,n}). \end{array} \right. \end{array} \right. \end{array} \tag{170}$$

Remark 93

i) As seen in Example 90, the functions of (165) satisfy the Lipschitz condition (162). However the cocoercivity condition (167) does not hold. For instance, if $\varphi_1 = 0$ and $\varphi_2 = 0$ then, for every $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathcal{H}_1 \times \mathcal{H}_2$,

$$\langle \nabla_1\, \boldsymbol{g}_1(\boldsymbol{x}) - \nabla_1\, \boldsymbol{g}_1(\boldsymbol{y}) \mid x_1 - y_1\rangle + \langle \nabla_2\, \boldsymbol{g}_2(\boldsymbol{x}) - \nabla_2\, \boldsymbol{g}_2(\boldsymbol{y}) \mid x_2 - y_2\rangle = 0. \tag{171}$$

ii) Distributed splitting algorithms for finding Nash equilibria are discussed in [25], [26], [244], [245].

iii) An asynchronous block-iterative decomposition algorithm to solve Nash equilibrium problems involving a mix of nonsmooth and smooth functions acting on linear mixtures of actions is proposed in [51].

VII. Fixed point modeling of other non-minimization problems

A. Neural network structures

A feedforward neural network (see Fig. 7) consists of the composition of nonlinear activation operators and affine

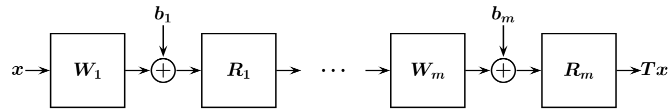

Fig. 7: Feedforward neural network: the ith layer involves a linear weight operator W_i, a bias vector b_i, and an activation operator R_i, which is assumed to be an averaged nonexpansive operator.

operators. More precisely, such an m-layer network can be modeled as

$$T = T_m \circ \cdots \circ T_1, \tag{172}$$

where $T_i = R_i \circ (W_i \cdot + b_i)$, with $W_i \in \mathbb{R}^{N_i \times N_{i-1}}$, $b_i \in \mathbb{R}^{N_i}$, and $R_i \colon \mathbb{R}^{N_i} \to \mathbb{R}^{N_i}$ (see Fig. 7). If the i-th layer is convolutional, then the corresponding weight matrix W_i has a Toeplitz (or block-Toeplitz) structure. Many common activation operators are separable, i.e.,

$$R_i \colon (\xi_k)_{1\leqslant k\leqslant N_i} \mapsto \big(\varrho_{i,k}(\xi_k)\big)_{1\leqslant k\leqslant N_i}, \tag{173}$$

where $\varrho_{i,k} \colon \mathbb{R} \to \mathbb{R}$. For example, the ReLU activation function is given by

$$\varrho_{i,k} \colon \xi \mapsto \begin{cases} \xi, & \text{if } \xi > 0; \\ 0, & \text{if } \xi \leqslant 0, \end{cases} \tag{174}$$

and the unimodal sigmoid activation function is

$$\varrho_{i,k} \colon \xi \mapsto \frac{1}{1+e^{-\xi}} - \frac{1}{2}. \tag{175}$$

An example of a nonseparable operator is the softmax activator

$$R_i \colon (\xi_k)_{1\leqslant k\leqslant N_i} \mapsto \left(e^{\xi_k} \Big/ \sum_{j=1}^{N_i} e^{\xi_j} \right)_{1\leqslant k\leqslant N_i}. \tag{176}$$

It was observed in [106] that almost all standard activators are actually averaged operators in the sense of (21). In particular, as discussed in [105], many activators are proximity operators in the sense of Theorem 2. In this case, in (173), there exist functions $(\phi_k)_{1\leqslant k\leqslant N_i}$ in $\Gamma_0(\mathbb{R})$ such that

$$R_i \colon (\xi_k)_{1\leqslant k\leqslant N_i} \mapsto \big(\operatorname{prox}_{\phi_k} \xi_k\big)_{1\leqslant k\leqslant N_i}. \tag{177}$$

For ReLU, ϕ_k reduces to $\iota_{[0,+\infty[}$ whereas, for the unimodal sigmoid, it is the function

$$\xi \mapsto \begin{cases} (\xi+1/2)\ln(\xi+1/2) + (1/2-\xi)\ln(1/2-\xi) \\ \qquad -(|\xi|^2+1/4)/2, & \text{if } |\xi| < 1/2; \\ -1/4, & \text{if } |\xi| = 1/2; \\ +\infty, & \text{if } |\xi| > 1/2. \end{cases} \tag{178}$$

For softmax, we have $R_i = \operatorname{prox}_{\varphi_i}$ where

$$\varphi_i \colon (\xi_k)_{1\leqslant k\leqslant N_i} \mapsto \begin{cases} \sum_{i=1}^{N_i} \big(\xi_k \ln \xi_k - |\xi_k|^2/2\big), \\ \qquad \text{if } \min\limits_{1\leqslant k\leqslant N_i} \xi_k \geqslant 0 \text{ and } \displaystyle\sum_{k=1}^{N_i} \xi_k = 1; \\ +\infty, \text{ otherwise.} \end{cases} \tag{179}$$

The weight matrices $(W_i)_{1\leqslant i\leqslant m}$ play a crucial role in the overall nonexpansiveness of the network. Indeed, under suitable conditions on these matrices, the network T is averaged. For example, let $W = W_m \cdots W_1$ and let

$$\theta_m = \|W\| + \sum_{\ell=1}^{m-1} \sum_{0\leqslant j_1 < \cdots < j_\ell \leqslant m-1} \|W_m \cdots W_{j_\ell+1}\| \times \|W_{j_\ell} \cdots W_{j_{\ell-1}+1}\| \cdots \|W_{j_1} \cdots W_0\|. \tag{180}$$

Then, if there exists $\alpha \in [1/2, 1]$ such that

$$\|W - 2^m(1-\alpha)\mathrm{Id}\| - \|W\| + 2\theta_m \leqslant 2^m\alpha, \tag{181}$$

T is α-averaged. Other sufficient conditions have been established in [105]. These results pave the way to a theoretical analysis of neural networks from the standpoint of fixed point methods. In particular, assume that $N_m = N_0$ and consider a recurrent network of the form

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = (1-\lambda_n)x_n + \lambda_n T x_n, \tag{182}$$

where $\lambda_n \in \,]0,+\infty[$ models a skip connection. Then, according to Theorem 37, the convergence of $(x_n)_{n\in\mathbb{N}}$ to a fixed point of T is guaranteed under condition (181) provided that $(\lambda_n)_{n\in\mathbb{N}}$ is an α-relaxation sequence. As shown in [105], when for every $i \in \{1,\ldots,m\}$, R_i is the proximity operator of some function $\varphi_i \in \Gamma_0(\mathbb{R}^{N_i})$, the recurrent network delivers asymptotically a solution to the system of inclusions

$$\begin{cases} b_1 \in \overline{x}_1 - W_1\overline{x}_m + \partial\varphi_1(\overline{x}_1) \\ b_2 \in \overline{x}_2 - W_2\overline{x}_1 + \partial\varphi_2(\overline{x}_2) \\ \quad \vdots \\ b_m \in \overline{x}_m - W_m\overline{x}_{m-1} + \partial\varphi_m(\overline{x}_m), \end{cases} \tag{183}$$

where $\overline{x}_m \in \operatorname{Fix} T$ and, for every $i \in \{2,\ldots,m\}$, $\overline{x}_i = T_i\overline{x}_{i-1}$. Alternatively, (183) is a Nash equilibrium of the form (142) where (we set $\overline{x}_0 = \overline{x}_m$)

$$\boldsymbol{h}_i \colon (x_i; \overline{\boldsymbol{x}}_{\smallsetminus i}) \mapsto \varphi_i(x_i) + \frac{1}{2}\|x_i - b_i - W_i\overline{x}_{i-1}\|^2. \tag{184}$$

Fixed point theory also allows us to provide conditions for T to be Lipschitzian and to calculate an associated Lipschitz constant. Such results are useful to evaluate the robustness of the network to adversarial perturbations of its input [223]. As shown in [106], if θ_m is given by (180), $\theta_m/2^{m-1}$ is a Lipschitz constant of T and

$$\|W\| \leqslant \frac{\theta_m}{2^{m-1}} \leqslant \|W_1\| \cdots \|W_m\|. \tag{185}$$

This bound is thus more accurate than the product of the individual bounds corresponding to each layer used in [223]. Tighter estimations can also be derived, especially when the activation operators are separable [106], [169], [212]. Note that the lower bound in (185) would correspond to a linear network where all the nonlinear activation operators would be removed. Interestingly, when all the weight matrices have components in $[0,+\infty[$ and the activation operators are separable, $\|W\|$ is a Lipschitz constant of the network [106].

Special cases of the neural network model of [105] are investigated in [147], [224]. Another special case of interest is when the operator T in (172) corresponds to the *unrolling* (or *unfolding*) of a fixed point algorithm [184], that is, each operator T_i corresponds to one iteration of such an algorithm [15], [144], [243], [253]. The algorithm parameters, as well as possible hyperparameters of the problem, can then be optimized from a training set by using differentiable programming. Let us note that the results of [105], [106] can be used to characterize the nonexpansiveness properties of the resulting neural network [29].

B. Plug-and-play methods

The principle of the so-called *plug-and-play* (PnP) methods [52], [192], [205], [209], [222], [235] is to replace a proximity operator appearing in some proximal minimization algorithm by another operator Q. The rationale is that, since a proximity operator can be interpreted as a denoiser [108], one can consider replacing this proximity operator by a more sophisticated denoiser Q, or even learning it in a supervised manner from a database of examples. Example 72 described implicitly a PnP algorithm that can be interpreted as a minimization problem. Here are some techniques that go beyond the optimization setting.

Algorithm 94 (PnP forward-backward) Let $f\colon \mathcal{H} \to \mathbb{R}$ be a differentiable convex function, let $Q\colon \mathcal{H} \to \mathcal{H}$, let $\gamma \in \left]0,+\infty\right[$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0,+\infty\right[$, and let $x_0 \in \mathcal{H}$. Iterate

$$\begin{array}{l}\text{for } n = 0,1,\ldots\\ \left\lfloor \begin{array}{l} y_n = x_n - \gamma\nabla f(x_n)\\ x_{n+1} = x_n + \lambda_n(Qy_n - x_n). \end{array}\right.\end{array} \tag{186}$$

The convergence of $(x_n)_{n\in\mathbb{N}}$ in (186) is related to the properties of $T = Q \circ (\mathrm{Id} - \gamma\nabla f)$. Suppose that T is α-averaged with $\alpha \in \left]0,1\right]$, and that $S = \operatorname{Fix} T \neq \varnothing$. Then it follows from Theorem 37 that, if $(\lambda_n)_{n\in\mathbb{N}}$ is an α-relaxation sequence, then $(x_n)_{n\in\mathbb{N}}$ converges to a point in S.

Algorithm 95 (PnP Douglas-Rachford) Let $f \in \Gamma_0(\mathcal{H})$, let $Q\colon \mathcal{H} \to \mathcal{H}$, let $\gamma \in \left]0,+\infty\right[$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0,+\infty\right[$, and let $x_0 \in \mathcal{H}$. Iterate

$$\begin{array}{l}\text{for } n = 0,1,\ldots\\ \left\lfloor \begin{array}{l} x_n = \operatorname{prox}_{\gamma f} y_n\\ y_{n+1} = y_n + \lambda_n\big(Q(2x_n - y_n) - x_n\big). \end{array}\right.\end{array} \tag{187}$$

In view of (187),

$$(\forall n \in \mathbb{N})\quad y_{n+1} = \Big(1 - \frac{\lambda_n}{2}\Big)y_n + \frac{\lambda_n}{2}Ty_n, \tag{188}$$

where $T = (2Q - \mathrm{Id}) \circ (2\operatorname{prox}_{\gamma f} - \mathrm{Id})$. Now assume that Q is such that T is α-averaged for some $\alpha \in \left]0,1\right]$ and $\operatorname{Fix} T \neq \varnothing$. Then it follows from Theorem 37 that, if $(\lambda_n/2)_{n\in\mathbb{N}}$ is an α-relaxation sequence, then $(y_n)_{n\in\mathbb{N}}$ converges to a point in $\operatorname{Fix} T$ and we deduce that $(x_n)_{n\in\mathbb{N}}$ converges to a point in $S = \operatorname{prox}_{\gamma f}(\operatorname{Fix} T)$. Conditions for T to be a Banach contraction in the two previous algorithms are given in [209].

Applying the Douglas-Rachford algorithm to the dual of Problem 66 leads to a simple form of the alternating direction method of multipliers. Thus, consider algorithm 95, where f, γ, and Q are replaced by f^*, $1/\gamma$ and $\mathrm{Id} + \gamma^{-1}Q(-\gamma\cdot)$, respectively, and $(\forall n \in \mathbb{N})$ $\lambda_n = 1$. Then we obtain the following algorithm [68], which is applied to image fusion in [226].

Algorithm 96 (PnP ADMM) Let $f \in \Gamma_0(\mathcal{H})$, let $Q\colon \mathcal{H} \to \mathcal{H}$, let $\gamma \in \left]0,+\infty\right[$, let $y_0 \in \mathcal{H}$, let $z_0 \in \mathcal{H}$, and let $\gamma \in \left]0,+\infty\right[$. Iterate

$$\begin{array}{l}\text{for } n = 0,1,\ldots\\ \left\lfloor \begin{array}{l} x_n = Q(y_n - z_n)\\ y_{n+1} = \operatorname{prox}_{\gamma f}(x_n + z_n)\\ z_{n+1} = z_n + x_n - y_{n+1}. \end{array}\right.\end{array} \tag{189}$$

Note that, beyond the above fixed point descriptions of S, the properties of the solutions in plug-and-play methods are elusive in general.

C. Adjoint mismatch problem

A common inverse problem formulation is to

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x) + \frac{1}{2}\|Hx - y\|^2 + \frac{\kappa}{2}\|x\|^2, \tag{190}$$

where $f \in \Gamma_0(\mathcal{H})$, $y \in \mathcal{G}$ models the observation, $H\colon \mathcal{H} \to \mathcal{G}$ is a linear operator, and $\kappa \in \left[0,+\infty\right[$. This is a particular case of Problem 66 where

$$g = \frac{1}{2}\|H\cdot - y\|^2 + \frac{\kappa}{2}\|\cdot\|^2, \tag{191}$$

has Lipschitzian gradient $\nabla g\colon x \mapsto H^*(Hx - y) + \kappa x$. It can therefore be solved via Proposition 70, which therefore requires the application of the adjoint operator H^* at each iteration. Due to both physical and computational limitations in certain applications, this adjoint may be hard to implement and it is replaced by a linear approximation $K\colon \mathcal{G} \to \mathcal{H}$ [174], [252]. This leads to a surrogate of the proximal-gradient scheme (105) of the form

$$\begin{aligned}&(\forall n \in \mathbb{N})\quad x_{n+1} = x_n +\\ &\lambda_n\Big(\operatorname{prox}_{\gamma f}\big((1-\gamma\kappa)x_n - \gamma K(Hx_n - y)\big) - x_n\Big),\end{aligned} \tag{192}$$

with $\gamma \in \left]0,+\infty\right[$ and $\{\lambda_n\}_{n\in\mathbb{N}} \subset \left]0,1\right]$. Let us assume that $L = K \circ H + \kappa\mathrm{Id}$ is a cocoercive operator. Then the above algorithm is an instance of the forward-backward splitting algorithm introduced in Proposition 50 to solve Problem 47 where $A = \partial f$ and $B = L\cdot - Ky$. This means that a solution produced by algorithm (192) no longer solves a minimization problem since L is not a gradient in general [21, Proposition 2.58]. However, suppose that g is ν-strongly convex with $\nu \in \left]0,+\infty\right[$, let $\zeta_{\min}$ be the minimum eigenvalue of $L + L^*$, set $\chi = 1/(\nu + \zeta_{\min})$, let $\widehat{x}$ be the solution to Problem 66, and let $\widetilde{x}$ be the solution to Problem 47. Then, as shown in [77],

$$\|\widetilde{x} - \widehat{x}\| \leqslant \chi\,\|(H^* - K)(H\widehat{x} - y)\|. \tag{193}$$

A sufficient condition ensuring that L is cocoercive is that $\zeta_{\min} > 0$. The problem of adjoint mismatch when $f = 0$ is studied in [122].

D. Problems with nonlinear observations

We describe the framework presented in [109], [110] to address the problem of recovering an ideal object $\overline{x} \in \mathcal{H}$ from linear and nonlinear transformations $(r_k)_{1 \leqslant k \leqslant q}$ of it.

Problem 97 For every $k \in \{1, \ldots, q\}$, let $R_k \colon \mathcal{H} \to \mathcal{G}_k$ and let $r_k \in \mathcal{G}_k$. The task is to

$$\text{find } x \in \mathcal{H} \text{ such that } (\forall k \in \{1, \ldots, q\}) \;\; R_k x = r_k. \tag{194}$$

In the case when $q = 2$, $\mathcal{G}_1 = \mathcal{G}_2 = \mathcal{H}$, and R_1 and R_2 are projectors onto vector subspaces, Problem 97 reduces to the classical linear recovery framework of [247] which can be solved by projection methods. We can also express Problem 61 as a special case of Problem 97 by setting $m = q$ and

$$(\forall k \in \{1, \ldots, q\}) \quad r_k = 0 \quad \text{and} \quad R_k = \mathrm{Id} - \mathrm{proj}_{C_k}. \tag{195}$$

In the presence of more general nonlinear operators, however, projection techniques are not applicable to solve (194). Furthermore, standard minimization approaches such as minimizing the least-squares residual $\sum_{k=1}^{q} \|R_k x - r_k\|^2$ typically lead to an intractable nonconvex problem. Yet, we can employ fixed point arguments to approach the problem and design a provenly convergent method to solve it. To this end, assume that (194) has a solution and that each operator R_k is *proxifiable* in the sense that there exists $S_k \colon \mathcal{G}_k \to \mathcal{H}$ such that

$$\begin{cases} S_k \circ R_k \text{ is firmly nonexpansive} \\ (\forall x \in \mathcal{H}) \quad S_k(R_k x) = S_k r_k \quad \Rightarrow \quad R_k x = r_k. \end{cases} \tag{196}$$

Clearly, if R_k is firmly nonexpansive, e.g., a projection or proximity operator (see Fig. 3), then it is proxifiable with $S_k = \mathrm{Id}$. Beyond that, many transformations found in data analysis, including discontinuous operations such as wavelet coefficients hard-thresholding, are proxifiable [109], [110]. Now set

$$(\forall k \in \{1, \ldots, q\}) \quad T_k = S_k r_k + \mathrm{Id} - S_k \circ R_k. \tag{197}$$

Then the operators $(T_k)_{1 \leqslant k \leqslant q}$ are firmly nonexpansive and Problem 97 reduces finding one of their common fixed points. In view of Propositions 18 and 26, this can be achieved by applying Theorem 37 with $T = T_1 \circ \cdots \circ T_q$. The more sophisticated block-iterative methods of [22], [110] are also applicable.

Let us observe that the above model is based purely on a fixed point formalism which does not involve monotone inclusions or optimization concepts. See [109], [110] for data science applications.

VIII. Concluding remarks

We have shown that fixed point theory provides an essential set of tools to efficiently model, analyze, and solve a broad range of problems in data science, be they formulated as traditional minimization problems or in more general forms such as Nash equilibria, monotone inclusions, or nonlinear operator equations. Thus, as illustrated in Section VII, nonlinear models that would appear to be predestined to nonconvex minimization methods can be effectively solved with the fixed point machinery. The prominent role played by averaged operators in the construction of provenly convergent fixed point iterative methods has been highlighted. Also emphasized is the fact that monotone operators are the backbone of many powerful modeling approaches. We believe that fixed point strategies are bound to play an increasing role in future advances in data science.

Acknowledgment. The authors thank Minh N. Bùi and Zev C. Woodstock for their careful proofreading of the paper.

References

[1] J. Abe, M. Yamagishi, and I. Yamada, Linearly involved generalized Moreau enhanced models and their proximal splitting algorithm under overall convexity condition, *Inverse Problems*, vol. 36, art. 035012, 2020.

[2] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, Fast image recovery using variable splitting and constrained optimization, *IEEE Trans. Image Process.*, vol. 19, pp. 2345–2356, 2010.

[3] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems, *IEEE Trans. Image Process.*, vol. 20, pp. 681–695, 2011.

[4] V. Apidopoulos, J.-F. Aujol, and C. Dossal, Convergence rate of inertial forward-backward algorithm beyond Nesterov's rule, *Math. Program.*, vol. A180, pp. 137–156, 2020.

[5] A. Argyriou, R. Foygel, and N. Srebro, Sparse prediction with the k-support norm, *Proc. Adv. Neural Inform. Process. Syst. Conf.*, vol. 25, pp. 1457–1465, 2012.

[6] E. Artzy, T. Elfving, and G. T. Herman, Quadratic optimization for image reconstruction II, *Comput. Graph. Image Process.*, vol. 11, pp. 242–261, 1979.

[7] Y. F. Atchadé, G. Fort, and E. Moulines, On perturbed proximal gradient algorithms, *J. Machine Learn. Res.*, vol. 18, pp. 1–33, 2017.

[8] H. Attouch and A. Cabot, Convergence of a relaxed inertial forward-backward algorithm for structured monotone inclusions, *Appl. Math. Optim.*, vol. 80, pp. 547–598, 2019.

[9] J.-F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle, Image decomposition into a bounded variation component and an oscillating component, *J. Math. Imaging Vision,* vol. 22, pp. 71–88, 2005.

[10] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher, Structure-texture image decomposition – modeling, algorithms, and parameter selection. *Int. J. Comput. Vision,* vol. 67, pp. 111–136, 2006.

[11] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, Optimization with sparsity-inducing penalties, *Found. Trends Machine Learn.*, vol. 4, pp. 1–106, 2012.

[12] J.-B. Baillon, R. E. Bruck, and S. Reich, On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces, *Houston J. Math.*, vol. 4, pp. 1–9, 1978.

[13] J.-B. Baillon, P. L. Combettes, and R. Cominetti, There is no variational characterization of the cycles in the method of periodic projections, *J. Funct. Anal.*, vol. 262, pp. 400–408, 2012.

[14] S. Banert, R. I. Boţ, and E. R. Csetnek, Fixing and extending some recent results on the ADMM algorithm, *Numer. Algorithms*, vol. 86, pp. 1303–1325, 2021.

[15] S. Banert, A. Ringh, J. Adler, J. Karlsson, and O. Öktem, Data-driven nonsmooth optimization, *SIAM J. Optim.*, vol. 30, pp. 102–131, 2020.

[16] H. H. Bauschke, J. Bolte, and M. Teboulle, A descent lemma beyond Lipschitz gradient continuity: First-order methods revisited and applications, *Math. Oper. Res.*, vol. 42, pp. 330–348, 2017.

[17] H. H. Bauschke and J. M. Borwein, On projection algorithms for solving convex feasibility problems, *SIAM Rev.*, vol. 38, pp. 367–426, 1996.

[18] H. H. Bauschke and J. M. Borwein, Legendre functions and the method of random Bregman projections, *J. Convex Anal.*, vol. 4, pp. 27–67, 1997.

[19] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, Bregman monotone optimization algorithms, *SIAM J. Control Optim.*, vol. 42, pp. 596–636, 2003.

[20] H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. New York: Springer, 2011.

[21] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed., corrected printing. New York: Springer, 2019.

[22] H. H. Bauschke, P. L. Combettes, and S. G. Kruk, Extrapolation algorithm for affine-convex feasibility problems, *Numer. Algorithms,* vol. 41, pp. 239–274, 2006.

[23] H. H. Bauschke, F. Deutsch, and H. Hundal, Characterizing arbitrarily slow convergence in the method of alternating projections, *Int. Trans. Oper. Res.*, vol. 16, pp. 413–425, 2009.

[24] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imaging Sci.*, vol. 2, pp. 183–202, 2009.

[25] G. Belgioioso and S. Grammatico, A Douglas-Rachford splitting for semi-decentralized equilibrium seeking in generalized aggregative games, *Proc. IEEE Conf. Decision Control*. Miami, USA, Dec. 17–19, 2018, pp. 3541–3546.

[26] G. Belgioioso, A. Nedich, and S. Grammatico, Distributed generalized Nash equilibrium seeking in aggregative games on time-varying networks, *IEEE Trans. Automat. Control*, vol. 66, pp. 2061–2075, 2021.

[27] A. Benfenati, E. Chouzenoux, and J.-C. Pesquet, Proximal approaches for matrix optimization problems: Application to robust precision matrix estimation, *Signal Process.*, vol. 169, art. 107417, 2020.

[28] M. Bergounioux, Mathematical analysis of a inf-convolution model for image processing, *J. Optim. Theory Appl.*, vol. 168, pp. 1–21, 2016.

[29] C. Bertocchi, E. Chouzenoux, M.-C. Corbineau, J.-C. Pesquet, and M. Prato, Deep unfolding of a proximal interior point method for image restoration, *Inverse Problems*, vol. 36, art. 034005, 2020.

[30] D. P. Bertsekas, Incremental proximal methods for large scale convex optimization, *Math. Program.*, vol. B129, pp. 163–195, 2011.

[31] J. M. Bioucas-Dias and M. A. T. Figueiredo, A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration, *IEEE Trans. Image Process.*, vol. 16, pp. 2992–3004, 2007.

[32] J. M. Borwein, B. Sims, and M. K. Tam, Norm convergence of realistic projection and reflection methods, *Optimization*, vol. 64, pp. 161–178, 2015.

[33] A. Borzì and C. Kanzow, Formulation and numerical solution of Nash equilibrium multiobjective elliptic control problems, *SIAM J. Control Optim.*, vol. 51, pp. 718–744, 2013.

[34] R. J. Boscovich, De literaria expeditione per pontificiam ditionem et synopsis amplioris operis..., *Bononiensi Scientiarum et Artum Instituto atque Academia Commentarii,* vol. 4, pp. 353–396, 1757.

[35] R. I. Boţ and C. Hendrich, A Douglas–Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators, *SIAM J. Optim.*, vol. 23, pp. 2541–2565, 2013.

[36] R. I. Boţ and C. Hendrich, Convergence analysis for a primal-dual monotone + skew splitting algorithm with applications to total variation minimization, *J. Math. Imaging Vision*, vol. 49, pp. 551–568, 2014.

[37] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Found. Trends Machine Learn.*, vol. 3, pp. 1–122, 2010.

[38] M. Bravo, D. Leslie, and P. Mertikopoulos, Bandit learning in concave N-person games, *Proc. Adv. Neural Inform. Process. Syst. Conf.*, vol. 31, pp. 5661–5671, 2018.

[39] L. M. Brègman, The method of successive projection for finding a common point of convex sets, *Soviet Math. – Doklady,* vol. 6, pp. 688–692, 1965.

[40] L. M. Brègman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, *USSR Comput. Math. Math. Phys.*, vol. 7, pp. 200–217, 1967.

[41] L. M. Briceño-Arias, Forward-Douglas-Rachford splitting and forward-partial inverse method for solving monotone inclusions, *Optimization*, vol. 64, pp. 1239–1261, 2015.

[42] L. M. Briceño-Arias, G. Chierchia, E. Chouzenoux, and J.-C. Pesquet, A random block-coordinate Douglas-Rachford splitting method with low computational complexity for binary logistic regression, *Comput. Optim. Appl.*, vol. 72, pp. 707–726, 2019.

[43] 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.*, vol. 2, pp. 485–508, 2009.

[44] L. M. Briceño-Arias and P. L. Combettes, A monotone+skew splitting model for composite monotone inclusions in duality, *SIAM J. Optim.*, vol. 21, pp. 1230–1250, 2011.

[45] L. M. Briceño-Arias and P. L. Combettes, Monotone operator methods for Nash equilibria in non-potential games, in: *Computational and Analytical Mathematics*, (D. Bailey et al., Eds.) New York: Springer, 2013, pp. 143–159.

[46] L. M. Briceño-Arias, P. L. Combettes, J.-C. Pesquet, and N. Pustelnik, Proximal algorithms for multicomponent image recovery problems, *J. Math. Imaging Vision*, vol. 41, pp. 3–22, 2011.

[47] L. M. Briceño-Arias and D. Davis, Forward-backward-half forward algorithm for solving monotone inclusions, *SIAM J. Optim.*, vol. 28, pp. 2839–2871, 2018.

[48] M. N. Bùi and P. L. Combettes, Warped proximal iterations for monotone inclusions, *J. Math. Anal. Appl.*, vol. 491, art. 124315, 2020.

[49] M. N. Bùi and P. L. Combettes, Bregman forward-backward operator splitting, *Set-Valued Var. Anal.*, published online 2020-11-28.

[50] M. N. Bùi and P. L. Combettes, Multivariate monotone inclusions in saddle form, *Math. Oper. Res.*, to appear.

[51] M. N. Bùi and P. L. Combettes, A warped resolvent algorithm to construct Nash equilibria, 2021. https://arxiv.org/abs/2101.00532

[52] G. T. Buzzard, S. H. Chan, S. Sreehari, and C. A. Bouman, Plug-and-play unplugged: Optimization-free reconstruction using consensus equilibrium, *SIAM J. Imaging Sci.*, vol. 11, pp. 2001–2020, 2018.

[53] C. Byrne, Iterative oblique projection onto convex sets and the split feasibility problem, *Inverse Problems*, vol. 18, pp. 441–453, 2002.

[54] C. L. Byrne, *Iterative Optimization in Inverse Problems.* Boca Raton, FL: CRC Press, 2014.

[55] C. Byrne and Y. Censor, Proximity function minimization using multiple Bregman projections, with applications to split feasibility and Kullback-Leibler distance minimization, *Ann. Oper. Res.*, vol. 105, pp. 77–98, 2001.

[56] E. J. Candès, X. Li, Y. Ma, and J. Wright, Robust principal component analysis?, *J. ACM*, vol. 58, art. 11, 2011.

[57] E. J. Candès and B. Recht, Exact matrix completion via convex optimization, *Found. Comput. Math.*, vol. 9, pp. 717–772, 2009.

[58] E. J. Candès and T. Tao, The power of convex relaxation: Near-optimal matrix completion, *IEEE Trans. Inform. Theory*, vol. 56, pp. 2053–2080, 2010.

[59] A. Cauchy, Méthode générale pour la résolution des systèmes d'équations simultanées, *C. R. Acad. Sci. Paris*, vol. 25, pp. 536–538, 1847.

[60] Y. Censor and T. Elfving, A multiprojection algorithm using Bregman projections in a product space, *Numer. Algorithms*, vol. 8, pp. 221–239, 1994.

[61] Y. Censor, T. Elfving, N. Kopf, and T. Bortfeld, The multiple-sets split feasibility problem and its applications for inverse problems, *Inverse Problems*, vol. 21, pp. 2071–2084, 2005.

[62] Y. Censor and M. Zaknoon, Algorithms and convergence results of projection methods for inconsistent feasibility problems: A review, *Pure Appl. Funct. Anal.*, vol. 3, pp. 565–586, 2018.

[63] Y. Censor and S. A. Zenios, Proximal minimization algorithm with D-functions, *J. Optim. Theory Appl.*, vol. 73, pp. 451–464, 1992.

[64] Y. Censor and S. A. Zenios, *Parallel Optimization – Theory, Algorithms and Applications.* New York: Oxford University Press, 1997.

[65] A. Chambolle and C. Dossal, On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm," *J. Optim. Theory Appl.*, vol. 166, pp. 968–982, 2015.

[66] A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, *J. Math. Imaging Vision*, vol. 40, pp. 120–145, 2011.

[67] A. Chambolle and T. Pock, An introduction to continuous optimization for imaging, *Acta Numer.*, vol. 25, pp. 161–319, 2016.

[68] S. H. Chan, X. Wang, and O. A. Elgendy, Plug-and-play ADMM for image restoration: Fixed-point convergence and applications, *IEEE Trans. Comput. Imaging*, vol. 3, pp. 84–98, 2017.

[69] P. Chen, J. Huang, and X. Zhang, A primal-dual fixed point algorithm for convex separable minimization with applications to image restoration, *Inverse Problems*, vol. 29, art. 025011, 2013.

[70] W. Cheney and A. A. Goldstein, Proximity maps for convex sets, *Proc. Amer. Math. Soc.*, vol. 10, pp. 448–450, 1959.

[71] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, Epigraphical splitting for solving constrained convex optimization problems with proximal tools, *Signal Image Video Process.*, vol. 9, pp. 1737–1749, 2015.

[72] G. Chierchia, N. Pustelnik, B. Pesquet-Popescu, and J.-C. Pesquet, A non-local structure tensor-based approach for multicomponent image recovery problems, *IEEE Trans. Image Process.*, vol. 23, pp. 5531–5544, 2014.

[73] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard, An interpolating distance between optimal transport and Fisher-Rao metrics, *Found. Comput. Math.*, vol. 18, pp. 1–44, 2018.

[74] E. Chouzenoux, M.-C. Corbineau, and J.-C. Pesquet, A proximal interior point algorithm with applications to image processing, *J. Math. Imaging Vision*, vol. 62, pp. 919–940, 2020.

[75] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot, A convex approach for image restoration with exact Poisson-Gaussian likelihood, *SIAM J. Imaging Sci.*, vol. 8, pp. 2662–2682, 2015.

[76] E. Chouzenoux, J.-C. Pesquet, and A. Repetti. A block coordinate variable metric forward-backward algorithm, *J. Global Optim.*, vol. 66, pp. 457–485, 2016.

[77] E. Chouzenoux, J.-C. Pesquet, C. Riddell, M. Savanier, and Y. Trousset, Convergence of proximal gradient algorithm in the presence of adjoint mismatch, *Inverse Problems*, 2021.

[78] G. Cimmino, Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari, *La Ricerca Scientifica (Roma)*, vol. 1, pp. 326–333, 1938.

[79] G. Cohen, Nash equilibria: Gradient and decomposition algorithms, *Large Scale Syst.*, vol. 12, pp. 173–184, 1987.

[80] P. L. Combettes, The foundations of set theoretic estimation, *Proc. IEEE*, vol. 81, pp. 182–208, 1993.

[81] P. L. Combettes, Inconsistent signal feasibility problems: Least-squares solutions in a product space, *IEEE Trans. Signal Process.*, vol. 42, pp. 2955–2966, 1994.

[82] P. L. Combettes, The convex feasibility problem in image recovery, in: *Advances in Imaging and Electron Physics* (P. Hawkes, Ed.), vol. 95, pp. 155–270. New York: Academic Press, 1996.

[83] P. L. Combettes, Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections, *IEEE Trans. Image Process.*, vol. 6, pp. 493–506, 1997.

[84] P. L. Combettes, Strong convergence of block-iterative outer approximation methods for convex optimization, *SIAM J. Control Optim.*, vol. 38, 538–565, 2000.

[85] P. L. Combettes, Fejér-monotonicity in convex optimization, in: *Encyclopedia of Optimization*, (C. A. Floudas and P. M. Pardalos, Eds.), vol. 2. New York: Springer, 2001, pp. 106–114. (Also available in 2nd ed., pp. 1016–1024, 2009.)

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

[87] P. L. Combettes, Solving monotone inclusions via compositions of nonexpansive averaged operators, *Optimization*, vol. 53, pp. 475–504, 2004.

[88] P. L. Combettes, Can one genuinely split $m > 2$ monotone operators? *Workshop on Algorithms and Dynamics for Games and Optimization*, Playa Blanca, Tongoy, Chile, October 14-18, 2013. https://pcombet.math.ncsu.edu/2013open-pbs1.pdf

[89] P. L. Combettes, Monotone operator theory in convex optimization, *Math. Program.*, vol. B170, pp. 177–206, 2018.

[90] P. L. Combettes, L. Condat, J.-C. Pesquet, and B. C. Vũ, A forward-backward view of some primal-dual optimization methods in image recovery, *Proc. IEEE Int. Conf. Image Process.* Paris, France, Oct. 27-30, 2014, pp. 4141–4145.

[91] P. L. Combettes, Đinh Dũng, and B. C. Vũ, Dualization of signal recovery problems, *Set-Valued Anal.*, vol. 18, pp. 373–404, 2010.

[92] P. L. Combettes and J. Eckstein, Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions, *Math. Program.*, vol. B168, pp. 645–672, 2018.

[93] P. L. Combettes and L. E. Glaudin, Quasinonexpansive iterations on the affine hull of orbits: From Mann's mean value algorithm to inertial methods, *SIAM J. Optim.*, vol. 27, pp. 2356–2380, 2017.

[94] P. L. Combettes and L. E. Glaudin, Proximal activation of smooth functions in splitting algorithms for convex image recovery, *SIAM J. Imaging Sci.*, vol. 12, pp. 1905–1935, 2019.

[95] P. L. Combettes and L. E. Glaudin, Solving composite fixed point problems with block updates, *Adv. Nonlinear Anal.*, vol. 10, pp. 1154–1177, 2021.

[96] P. L. Combettes, A. M. McDonald, C. A. Micchelli, and M. Pontil, Learning with optimal interpolation norms, *Numer. Algorithms*, vol. 81, pp. 695–717, 2019.

[97] P. L. Combettes and C. L. Müller, Perspective maximum likelihood-type estimation via proximal decomposition, *Electron. J. Stat.*, vol. 14, pp. 207–238, 2020.

[98] P. L. Combettes and Q. V. Nguyen, Solving composite monotone inclusions in reflexive Banach spaces by constructing best Bregman approximations from their Kuhn-Tucker set, *J. Convex Anal.*, vol. 23, pp. 481–510, 2016.

[99] P. L. Combettes and J.-C. Pesquet, A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery, *IEEE J. Select. Topics Signal Process.*, vol. 1, pp. 564–574, 2007.

[100] P. L. Combettes and J.-C. Pesquet, Proximal splitting methods in signal processing, in [20], pp. 185–212.

[101] P. L. Combettes and J.-C. Pesquet, Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators, *Set-Valued Var. Anal.*, vol. 20, pp. 307–330, 2012.

[102] P. L. Combettes and J.-C. Pesquet, Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping, *SIAM J. Optim.*, vol. 25, pp. 1221–1248, 2015.

[103] P. L. Combettes and J.-C. Pesquet, Stochastic approximations and perturbations in forward-backward splitting for monotone operators, *Pure Appl. Funct. Anal.*, vol. 1, pp. 13–37, 2016.

[104] P. L. Combettes and J.-C. Pesquet, Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping II: Mean-square and linear convergence, *Math. Program.*, vol. B174, pp. 433–451, 2019.

[105] P. L. Combettes and J.-C. Pesquet, Deep neural network structures solving variational inequalities, *Set-Valued Var. Anal.*, vol. 28, pp. 491–518, 2020.

[106] P. L. Combettes and J.-C. Pesquet, Lipschitz certificates for layered network structures driven by averaged activation operators, *SIAM J. Math. Data Sci.*, vol. 2, pp. 529–557, 2020.

[107] P. L. Combettes and B. C. Vũ, Variable metric forward-backward splitting with applications to monotone inclusions in duality, *Optimization*, vol. 63, pp. 1289–1318, 2014.

[108] P. L. Combettes and V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.*, vol. 4, pp. 1168–1200, 2005.

[109] 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. Amsterdam, The Netherlands, Jan. 18-22, 2021.

[110] P. L. Combettes and Z. C. Woodstock, Reconstruction of functions from prescribed proximal points, 2021. http://arxiv.org/abs/2101.04074

[111] P. L. Combettes and I. Yamada, Compositions and convex combinations of averaged nonexpansive operators, *J. Math. Anal. Appl.*, vol. 425, pp. 55–70, 2015.

[112] L. Condat, A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms, *J. Optim. Theory Appl.*, vol. 158, pp. 460–479, 2013.

[113] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi, Proximal splitting algorithms: A tour of recent advances, with new twists, 2020. https://arxiv.org/abs/1912.00137

[114] A. Danielyan, V. Katkovnik, and K. Egiazarian, BM3D frames and variational image deblurring, *IEEE Trans. Image Process.*, vol. 21, pp. 1715–1728, 2012.

[115] J. Darbon and T. Meng, On decomposition models in imaging sciences and multi-time Hamilton–Jacobi partial differential equations, *SIAM J. Imaging Sci.*, vol. 13, pp. 971–1014, 2020.

[116] P. Dasgupta and J. B. Collins, A survey of game theoretic approaches for adversarial machine learning in cybersecurity tasks, *AI Magazine*, vol. 40, pp. 31–43, 2019.

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

[118] D. Davis and W. Yin, A three-operator splitting scheme and its optimization applications, *Set-Valued Var. Anal.*, vol. 25, pp. 829–858, 2017.

[119] A. J. Defazio, T. S. Caetano, and J. Domke, Finito: A faster, permutable incremental gradient method for big data problems, *Proc. Int. Conf. Machine Learn.*, pp. 1125–1133. Beijing, China, June 22-24, 2014.

[120] A. Defazio, F. Bach, and S. Lacoste-Julien, SAGA: A fast incremental gradient method with support for non-strongly convex composite objectives, *Proc. Adv. Neural Inform. Process. Syst. Conf.* vol. 27, pp. 1646–1654, 2014.

[121] C. De Mol and M. Defrise, Inverse imaging with mixed penalties, *ULB Institutional Repository*, pp. 798–800, 2004.

[122] Y. Dong, P. C. Hansen, M. E. Hochstenbach, and N. A. Brogaard Riis, Fixing nonconvergence of algebraic iterative reconstruction with an unmatched backprojector, *SIAM J. Sci. Comput.*, vol. 41, pp. A1822–A1839, 2019.

[123] D. L. Donoho and M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization, *Proc. Nat. Acad. Sci.*, vol. 100, pp. 2197–2202, 2003.

[124] J. Eckstein, Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming, *Math. Oper. Res.*, vol. 18, pp. 202–226, 1993.

[125] J. Eckstein, Parallel alternating direction multiplier decomposition of convex programs, *J. Optim. Theory Appl.*, vol. 80, pp. 39–62, 1994.

[126] J. Eckstein and D. P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Math. Program.*, vol. 55, pp. 293–318, 1992.

[127] L. El Ghaoui, V. Viallon, and T. Rabbani, Safe feature elimination in sparse supervised learning, *Pac. J. Optim.*, vol. 8, pp. 667–698, 2012.

[128] I. I. Eremin, Generalization of the relaxation method of Motzkin-Agmon, *Uspekhi Mat. Nauk*, vol. 20, pp. 183–187, 1965.

[129] Yu. M. Ermol'ev, On the method of generalized stochastic gradients and quasi-Fejér sequences, *Cybernetics*, vol. 5, pp. 208–220, 1969.

[130] Yu. M. Ermol'ev and Z. V. Nekrylova, Some methods of stochastic optimization, *Kibernetika (Kiev)*, vol. 1966, pp. 96–98, 1966.

[131] E. Esser, X. Zhang, and T. Chan, A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science, *SIAM J. Imaging Sci.*, vol. 3, pp. 1015–1046, 2010.

[132] M. A. T. Figueiredo and J. M. Bioucas-Dias, Restoration of Poissonian images using alternating direction optimization, *IEEE Trans. Image Process.*, vol. 19, pp. 3133–3145, 2010.

[133] J. Friedman, T. Hastie, and R. Tibshirani, Sparse inverse covariance estimation with the graphical lasso, *Biostatistics*, vol. 9, pp. 432–441, 2008.

[134] D. Gabay, Applications of the method of multipliers to variational inequalities, in: M. Fortin and R. Glowinski (Eds.), *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary Value Problems,* pp. 299–331. Amsterdam: North-Holland, 1983.

[135] S. Gandy, B. Recht, and I. Yamada, Tensor completion and low-n-rank tensor recovery via convex optimization, *Inverse Problems*, vol. 27, art. 025010, 2011.

[136] C. F. Gauss, *Theoria Motus Corporum Coelestium.* Hamburg: Perthes and Besser, 1809.

[137] J.-F. Giovannelli and A. Coulais, Positive deconvolution for superimposed extended source and point sources, *Astron. Astrophys.*, vol. 439, pp. 401–412, 2005.

[138] P. Giselsson, Nonlinear forward-backward splitting with projection correction, 2019. https://arxiv.org/abs/1908.07449

[139] R. Glowinski and P. Le Tallec (Eds.), *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*. Philadelphia: SIAM, 1989.

[140] R. Glowinski, S. J. Osher, and W. Yin (Eds.), *Splitting Methods in Communication, Imaging, Science, and Engineering*. New York: Springer, 2016.

[141] A. A. Goldstein, Convex programming in Hilbert space, *Bull. Amer. Math. Soc.* vol. 70, pp. 709–710, 1964.

[142] E. G. Gol'shtein, A general approach to decomposition of optimization systems, *Sov. J. Comput. Syst. Sci.*, vol. 25, pp. 105–114, 1987.

[143] T. Goldstein and S. Osher, The split Bregman method for L1-regularized problems, *SIAM J. Imaging Sci.*, vol. 2, pp. 323–343, 2009.

[144] K. Gregor and Y. LeCun, Learning fast approximations of sparse coding, *Proc. Int. Conf. Machine Learn.*, pp. 399–406. Haifa, Israel, June 21-24, 2010.

[145] R. Gribonval and P. Machart, Reconciling "priors" & "priors" without prejudice?, *Proc. Adv. Neural Inform. Process. Syst. Conf.*, vol. 26, pp. 2193–2201, 2013.

[146] 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.*, vol. 7, pp. 1–24, 1967.

[147] M. Hasannasab, J. Hertrich, S. Neumayer, G. Plonka, S. Setzer, and G. Steidl, Parseval proximal neural networks, *J. Fourier Anal. Appl.*, vol. 26, art. 59, 2020.

[148] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd. ed. New York: Springer, 2009.

[149] B. He and X. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, *SIAM J. Imaging Sci.*, vol. 5, pp. 119–149, 2012.

[150] G. T. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, 2nd ed. New York: Springer, 2009.

[151] G. T. Herman, A. Lent, and S. W. Rowland, ART: Mathematics and applications, *J. Theor. Biology*, vol. 42, pp. 1–18, 1973.

[152] S. A. Hirstoaga, Iterative selection methods for common fixed point problems, *J. Math. Anal. Appl.*, vol. 324, pp. 1020–1035, 2006.

[153] X. Huang, E. K. Ryu, and W. Yin, Tight coefficients of averaged operators via scaled relative graph, *J. Math. Anal. Appl.*, vol. 490, art. 124211, 12 pp., 2020.

[154] B. R. Hunt, The inverse problem of radiography, *Math. Biosciences,* vol. 8, pp. 161–179, 1970.

[155] F. Iutzeler, P. Bianchi, P. Ciblat, and W. Hachem, Asynchronous distributed optimization using a randomized alternating direction method of multipliers, *Proc. 52nd Conf. Decision Control*, pp. 3671–3676. Florence, Italy, Dec. 10–13, 2013.

[156] L. Jacob, G. Obozinski, and J.-Ph. Vert, Group lasso with overlap and graph lasso, *Proc. Int. Conf. Machine Learn.*, pp. 433–440. Montréal, Canada, June 14–18, 2009.

[157] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach, Proximal methods for hierarchical sparse coding, *J. Machine Learn. Res.*, vol. 12, pp. 2297–2334, 2011.

[158] R. Johnson and T. Zhang, Accelerating stochastic gradient descent using predictive variance reduction, *Proc. Adv. Neural Inform. Process. Syst. Conf.*, vol. 26, pp. 315–323, 2013.

[159] P. R. Johnstone and J. Eckstein, Projective splitting with forward steps, *Math. Program.*, published online 2020-09-30.

[160] P. R. Johnstone and J. Eckstein, Single-forward-step projective splitting: Exploiting cocoercivity, *Comput. Optim. Appl.* vol. 78, pp. 125–166, 2021.

[161] S. Kaczmarz, Angenäherte Auflösung von Systemen linearer Gleichungen, *Bull. Acad. Sci. Pologne,* vol. A35, pp. 355–357, 1937.

[162] B. Lemaire, Stability of the iteration method for nonexpansive mappings, *Serdica Math. J.*, vol. 22, pp. 331–340, 1996.

[163] K. C. Kiwiel and B. Łopuch, Surrogate projection methods for finding fixed points of firmly nonexpansive mappings, *SIAM J. Optim.*, vol. 7, pp. 1084–1102, 1997.

[164] N. Komodakis and J.-C. Pesquet, Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems, *IEEE Signal Process. Mag.*, vol. 32, pp. 31–54, 2015.

[165] H. J. Landau and W. L. Miranker, The recovery of distorted band-limited signals, *J. Math. Anal. Appl.*, vol. 2, pp. 97–104, 1961.

[166] P. S. Laplace, Sur quelques points du système du monde, *Mémoires Acad. Royale Sci. Paris,* pp. 1–87, 1789.

[167] R. Laraki, J. Renault, and S. Sorin, *Mathematical Foundations of Game Theory.* New York: Springer, 2019.

[168] S. Lasaulce and H. Tembine, *Game Theory and Learning for Wireless Networks: Fundamentals and Applications*. Amsterdam: Elsevier, 2011.

[169] F. Latorre, P. T. Rolland, and V. Cevher, Lipschitz constant estimation of neural networks via sparse polynomial optimization, *Proc. Int. Conf. Learn. Represent.* Addis Abeba, Ethiopia, April 26-30, 2020.

[170] A. M. Legendre, *Nouvelles Méthodes pour la Détermination des Orbites des Comètes.* Firmin Didot, Paris, 1805.

[171] E. S. Levitin and B. T. Polyak, Convergence of minimizing sequences in conditional extremum problems, *Soviet Math. Dokl.*, vol. 7, pp. 764–767, 1966.

[172] S. B. Lindstrom and B. Sims, Survey: Sixty years of Douglas–Rachford, *J. Aust. Math. Soc.*, published online 2020-02-20.

[173] P.-L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.*, vol. 16, pp. 964–979, 1979.

[174] D. A. Lorenz, S. Rose, and F. Schöpfer, The randomized Kaczmarz method with mismatched adjoint, *Bit. Numer. Math.*, vol. 58, pp. 1079–1098, 2018.

[175] I. Loris and C. Verhoeven, On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty, *Inverse Problems*, vol. 27, art. 125007, 2011.

[176] C. P. Mariadassou and B. Yegnanarayana, Image reconstruction from noisy digital holograms, *IEE Proceedings-F*, vol. 137, pp. 351–356, 1990.

[177] B. Martinet, *Algorithmes pour la Résolution de Problèmes d'Optimisation et de Minimax.* Thèse, Université de Grenoble, France, 1972. http://hal.archives-ouvertes.fr/

[178] A. M. McDonald, M. Pontil, and D. Stamos, New perspectives on k-support and cluster norms, *J. Machine Learn. Res.*, vol. 17, pp. 1–38, 2016.

[179] B. Mercier, *Lectures on Topics in Finite Element Solution of Elliptic Problems*. Lectures on Mathematics and Physics, vol. 63. Bombay: Tata Institute of Fundamental Research, 1979.

[180] G. J. Minty, Monotone (nonlinear) operators in Hilbert space, *Duke Math. J.*, vol. 29, pp. 341–346, 1962.

[181] K. Mishchenko, F. Iutzeler, and J. Malick, A distributed flexible delay-tolerant proximal gradient algorithm, *SIAM J. Optim.*, vol. 30, pp. 933–959, 2020.

[182] G. Moerkotte, M. Montag, A. Repetti, and G. Steidl, Proximal operator of quotient functions with application to a feasibility problem in query optimization, *J. Comput. Appl. Math.*, vol. 285, pp. 243–255, 2015.

[183] A. Mokhtari, M. Gürbüzbalaban, and A. Ribeiro, Surpassing gradient descent provably: A cyclic incremental method with linear convergence rate, *SIAM J. Optim.*, vol. 28, pp. 1420–1447, 2018.

[184] V. Monga, Y. Li, and Y. C. Eldar, Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing, *IEEE Signal Process. Mag.*, vol. 38, pp. 18–44, 2021.

[185] J. J. Moreau, Fonctions convexes duales et points proximaux dans un espace hilbertien, *C. R. Acad. Sci. Paris*, vol. A255, pp. 2897–2899, 1962.

[186] Y. Nakayama, M. Yamagishi, and I. Yamada, A hierarchical convex optimization for multiclass SVM achieving maximum pairwise margins with least empirical hinge-loss, 2020. https://arxiv.org/abs/2004.08180

[187] J. Nash, Non-cooperative games, *Ann. Math.*, vol. 54, pp. 286–295, 1951.

[188] E. Ndiaye, O. Fercoq, A. Gramfort, and J. Salmon, Gap safe screening rules for sparsity enforcing penalties, *J. Machine Learn. Res.*, vol. 18, pp. 1–33, 2017.

[189] I. Necoara and D. Clipici, Parallel random coordinate descent method for composite minimization: Convergence analysis and error bounds, *SIAM J. Optim.*, vol. 26, pp. 197–226, 2016.

[190] Q. V. Nguyen, Forward-backward splitting with Bregman distances, *Vietnam J. Math.*, vol. 45, pp. 519–539, 2017.

[191] D. O'Connor and L. Vandenberghe, Primal-dual decomposition by operator splitting and applications to image deblurring, *SIAM J. Imaging Sci.*, vol. 7, pp. 1724–1754, 2014.

[192] S. Ono, Primal-dual plug-and-play image restoration, *IEEE Signal Process. Lett.*, vol. 24, pp. 1108–1112, 2017.

[193] S. Ono and I. Yamada, Hierarchical convex optimization with primal-dual splitting, *IEEE Trans. Signal Process.*, vol. 63, pp. 373–388, 2015.

[194] N. Papadakis, G. Peyré, and E. Oudet, Optimal transport with proximal splitting, *SIAM J. Imaging Sci.*, vol. 7, pp. 212–238, 2014.

[195] M. Q. Pham, L. Duval, C. Chaux, and J. Pesquet A primal-dual proximal algorithm for sparse template-based adaptive filtering: Application to seismic multiple removal, *IEEE Trans. Signal Process.*, vol. 62, pp. 4256–4269, 2014.

[196] G. Pierra, Decomposition through formalization in a product space, *Math. Program.*, vol. 28, pp. 96–115, 1984.

[197] L. C. Potter and K. S. Arun, A dual approach to linear inverse problems with convex constraints, *SIAM J. Control Optim.*, vol. 31, pp. 1080–1092, 1993.

[198] N. Pustelnik, P. Borgnat, and P. Flandrin, Empirical mode decomposition revisited by multicomponent non-smooth convex optimization, *Signal Process.*, vol. 102, pp. 313–331, 2014.

[199] H. Raguet, A note on the forward-Douglas-Rachford splitting for monotone inclusion and convex optimization, *Optim. Lett.*, vol. 13, pp. 717–740, 2019.

[200] H. Raguet, J. Fadili, and G. Peyré, A generalized forward-backward splitting, *SIAM J. Imaging Sci.*, vol. 6, pp. 1199–1226, 2013.

[201] P. Ravikumar, M. J. Wainwright, G. Raskutti, and B. Yu, High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence, *Electron. J. Statist.*, vol. 5, pp. 935–980, 2011.

[202] A Repetti, M. Pereyra, and Y. Wiaux, Scalable Bayesian uncertainty quantification in imaging inverse problems via convex optimization, *SIAM J. Imaging Sci.*, vol. 12, pp. 87–118, 2019.

[203] P. Richtárik and M. Takáč, Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function, *Math. Program.*, vol. A144, pp. 1–38, 2014.

[204] P. Richtárik and M. Takáč, On optimal probabilities in stochastic coordinate descent methods, *Optim. Lett.*, vol. 10, pp 1233–1243, 2016.

[205] A. Rond, R. Giryes, and M. Elad, Poisson inverse problems by the plug-and-play scheme, *J. Vis. Commun. Image Repres.*, vol. 41, pp. 96–108, 2016.

[206] R. T. Rockafellar, *Convex Analysis.* Princeton, NJ: Princeton University Press, 1970.

[207] L. Rosasco, S. Villa, and B. C. Vũ, A stochastic inertial forward-backward splitting algorithm for multivariate monotone inclusions, *Optimization*, vol. 65, 1293–1314, 2016.

[208] L. Rosasco, S. Villa, and B. C. Vũ, Convergence of stochastic proximal gradient algorithm, *Appl. Math. Optim.*, vol. 82, pp. 891–917, 2020.

[209] E. K. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin, Plug-and-play methods provably converge with properly trained denoisers, *Proc. Int. Conf. Machine Learn.*, pp. 5546–5557. Miami, USA, June 11-13, 2019.

[210] E. K. Ryu, Uniqueness of DRS as the 2 operator resolvent-splitting and impossibility of 3 operator resolvent-splitting, *Math. Program.*, vol. A182, pp. 233–273, 2020.

[211] S. Salzo and S. Villa, Parallel random block-coordinate forward-backward algorithm: A unified convergence analysis, *Math. Program.*, published online 2021-04-11.

[212] K. Scaman and A. Virmaux, Lipschitz regularity of deep neural networks: Analysis and efficient estimation, *Proc. Adv. Neural Inform. Process. Syst. Conf.*, vol. 31, pp. 3839–3848, 2018.

[213] R. W. Schafer, R. M. Mersereau, and M. A. Richards, Constrained iterative restoration algorithms, *Proc. IEEE*, vol. 69, pp. 432–450, 1981.

[214] M. Schmidt, N. Le Roux, and F. Bach, Minimizing finite sums with the stochastic average gradient, *Math. Program.*, vol. A162, pp. 83–112, 2017.

[215] G. Scutari, D. P. Palomar, F. Facchinei, and J.-S. Pang, Convex optimization, game theory, and variational inequality theory, *IEEE Signal Process. Mag.*, vol 27, pp. 35–49, 2010.

[216] I. Selesnick, A. Lanza, S. Morigi, and F. Sgallari, Non-convex total variation regularization for convex denoising of signals, *J. Math. Imaging Vision*, vol. 62, pp. 825–841, 2020.

[217] S. Setzer, G. Steidl, and T. Teuber, Deblurring Poissonian images by split Bregman techniques, *J. Visual Commun. Image Represent.*, vol. 21, pp. 193–199, 2010.

[218] J. E. Spingarn, Partial inverse of a monotone operator, *Appl. Math. Optim.*, vol. 10, pp. 247–265, 1983.

[219] S. Sra, S. Nowozin, and S. J. Wright, *Optimization for Machine Learning.* MIT Press, Cambridge, MA, 2012.

[220] H. Stark (Ed.), *Image Recovery: Theory and Application.* San Diego, CA: Academic Press, 1987.

[221] G. Steidl and T. Teuber, Removing multiplicative noise by Douglas-Rachford splitting methods, *J. Math. Imaging Vis.*, vol. 36, pp. 168–184, 2010.

[222] Y. Sun, B. Wohlberg, and U. S. Kamilov, An online plug-and-play algorithm for regularized image reconstruction, *IEEE Trans. Comput. Imaging*, vol. 5, pp. 395–408, 2019.

[223] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. J. Goodfellow, and R. Fergus, Intriguing properties of neural networks, 2013. https://arxiv.org/pdf/1312.6199

[224] W. Tang, E. Chouzenoux, J.-C. Pesquet, and H. Krim, Deep transform and metric learning network: Wedding deep dictionary learning and neural networks, 2020, https://arxiv.org/abs/2002.07898

[225] M. Teboulle, Entropic proximal mappings with applications to nonlinear programming, *Math. Oper. Res.*, vol. 17, pp. 670–690, 1992.

[226] A. M. Teodoro, J. M. Bioucas-Dias, and M. A. T. Figueiredo, A convergent image fusion algorithm using scene-adapted Gaussian-mixture-based denoising, *IEEE Trans. Image Process.*, vol. 28, pp. 451–463, 2019.

[227] N. T. Thao and D. Rzepka, Time encoding of bandlimited signals: Reconstruction by pseudo-inversion and time-varying multiplierless FIR filtering, *IEEE Trans. Signal Process.*, vol. 69, pp. 341–356, 2021.

[228] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*, 2nd ed. Amsterdam: Elsevier, 2020.

[229] S. Theodoridis, K. Slavakis, and I. Yamada, Adaptive learning in a world of projections, *IEEE Signal Process. Mag.*, vol. 28, pp. 97–123, 2011.

[230] V. Tom, T. Quatieri, M. Hayes, and J. McClellan, Convergence of iterative nonexpansive signal reconstruction algorithms, *IEEE Trans. Acoust. Speech Signal Process.*, vol. 29, pp. 1052–1058, 1981.

[231] H. J. Trussell and M. R. Civanlar, The feasible solution in signal restoration, *IEEE Trans. Acoust. Speech Signal Process.*, vol. 32, pp. 201–212, 1984.

[232] P. Tseng, A modified forward-backward splitting method for maximal monotone mappings, *SIAM J. Control Optim.*, vol. 38, pp. 431–446, 2000.

[233] S. Twomey, The application of numerical filtering to the solution of integral equations encountered in indirect sensing measurements, *J. Franklin Inst.*, vol. 279, pp. 95–109, 1965.

[234] N. Vaswani, Y. Chi, and T. Bouwmans, Rethinking PCA for modern data sets: Theory, algorithms, and applications, *Proc. IEEE*, vol. 106, pp. 1274–1276, 2018.

[235] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, Plug-and-play priors for model based reconstruction, *Proc. IEEE Global Conf. Signal Inform. Process.* Austin, TX, USA, December 3–5, 2013, pp. 945–948.

[236] S. Villa, L. Rosasco, S. Mosci, and A. Verri, Proximal methods for the latent group lasso penalty, *Comput. Optim. Appl.*, vol. 58, pp. 381–407, 2014.

[237] B. C. Vũ, A splitting algorithm for dual monotone inclusions involving cocoercive operators, *Adv. Comput. Math.*, vol. 38, pp. 667–681, 2013.

[238] H. M. Wagner, Linear programming techniques for regression analysis, *J. Amer. Stat. Assoc.*, vol. 54, pp. 206–212, 1959.

[239] R. G. Wiley, On an iterative technique for recovery of bandlimited signals, *Proc. IEEE*, vol. 66, pp. 522–523, 1978.
[240] X. Xu, Y. Sun, J. Liu, B. Wohlberg, and U. S. Kamilov, Provable convergence of plug-and-play priors with MMSE denoisers, *IEEE Signal Process. Lett.*, vol. 27, pp. 1280–1284, 2020.
[241] I. Yamada, The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings, in: *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*, (D. Butnariu, Y. Censor, and S. Reich, Eds.). Amsterdam: Elsevier, 2001, pp. 473–504.
[242] M. Yan, A new primal-dual algorithm for minimizing the sum of three functions with a linear operator, *J. Sci. Comput.*, vol. 76, pp. 1698–1717, 2018.
[243] Y. Yang, J. Sun, H. Li, and Z. Xu, Deep ADMM-Net for compressive sensing MRI, *Proc. Adv. Neural Inform. Process. Syst. Conf.*, vol. 29, pp. 10–19, 2016.
[244] P. Yi and L. Pavel, Distributed generalized Nash equilibria computation of monotone games via double-layer preconditioned proximal-point algorithms, *IEEE Trans. Control. Netw. Syst.*, vol. 6, pp. 299–311, 2019.
[245] P. Yi and L. Pavel, An operator splitting approach for distributed generalized Nash equilibria computation, *Automatica*, vol. 102, pp. 111–121, 2019.
[246] H. Yin, U. V. Shanbhag, and P. G. Mehta, Nash equilibrium problems with scaled congestion costs and shared constraints, *IEEE Trans. Autom. Control.*, vol. 56, pp. 1702–1708, 2011.
[247] D. C. Youla, Generalized image restoration by the method of alternating orthogonal projections, *IEEE Trans. Circuits Syst.*, vol. 25, pp. 694–702, 1978.
[248] 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.*, vol. 33, pp. 465–468, 1986.
[249] D. C. Youla and H. Webb, Image restoration by the method of convex projections: Part 1 – theory, *IEEE Trans. Med. Imaging*, vol. 1, pp. 81–94, 1982.
[250] Y. Yu, J. Peng, X. Han, and A. Cui, A primal Douglas–Rachford splitting method for the constrained minimization problem in compressive sensing, *Circuits Syst. Signal Process.*, vol. 36, pp. 4022–4049, 2017.
[251] M. Yuan and Y. Lin, Model selection and estimation in regression with grouped variables, *J. Roy. Stat. Soc.*, vol. B68, pp. 49–67, 2006.
[252] G. L. Zeng and G. T. Gullberg, Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, *IEEE Trans. Med. Imaging*, vol. 19, pp. 548–555, 2000.
[253] J. Zhang and B. Ghanem, ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1828–1837. Salt Lake City, USA, June 18–22, 2018.
[254] R. Zhang and L. Guo, Controllability of stochastic game-based control systems, *SIAM J. Control Optim.*, vol. 57, pp. 3799–3826, 2019.

Patrick L. Combettes (Fellow, IEEE 2006) joined the faculty of the City University of New York (City College and Graduate Center) in 1990 and the Laboratoire Jacques-Louis Lions of Université Pierre et Marie Curie–Paris 6 (now Sorbonne Université) in 1999. He has been a Distinguished Professor of Mathematics at North Carolina State University since 2016. He is the founding director (2009-2013) of the CNRS research consortium MOA on mathematical optimization and its applications.

Jean-Christophe Pesquet (Fellow, IEEE 2012) received the engineering degree from Supélec, Gif-sur-Yvette, France, in 1987, the Ph.D. and HDR degrees from Université Paris-Sud in 1990 and 1999, respectively. From 1991 to 1999, he was an Assistant Professor at Université Paris-Sud, and a Research Scientist at the Laboratoire des Signaux et Systèmes (CNRS). From 1999 to 2016, he was a Full Professor at Université Paris-Est and from 2012 to 2016, he was the Deputy Director of the Laboratoire d'Informatique of the university (CNRS). He is currently a Distinguished Professor at CentraleSupélec, Université Paris-Saclay, and the Director of the Center for Visual Computing and OPIS Inria group. His research interests include statistical signal/image processing and optimization methods with applications to data science. He has also been a Senior Member of the Institut Universitaire de France since 2016.