Optimization

A Journal of Mathematical Programming and Operations Research

ISSN: 0233-1934 (Print) 1029-4945 (Online) Journal homepage: http://www.tandfonline.com/loi/gopt20

Fejér-monotone hybrid steepest descent method for affinely constrained and composite convex minimization tasks

Konstantinos Slavakis & Isao Yamada

OPTIMIZATION
https://doi.org/10.1080/02331934.2018.1505885

Taylor & Francis
Taylor & Francis Group

ARTICLE

Fejér-monotone hybrid steepest descent method for affinely constrained and composite convex minimization tasks*

Konstantinos Slavakis [a] and Isao Yamada [b]

[a]Department of Electrical Engineering, University at Buffalo, The State University of New York, Buffalo, NY, USA; [b]Department of Information and Communications Engineering, Tokyo Institute of Technology, Tokyo, Japan

ABSTRACT
This paper introduces the *Fejér-monotone hybrid steepest descent method* (FM-HSDM), a new member to the HSDM family of algorithms, for solving affinely constrained minimization tasks in real Hilbert spaces, where convex smooth and non-smooth losses compose the objective function. FM-HSDM offers sequences of estimates which converge weakly and, under certain hypotheses, strongly to solutions of the task at hand. In contrast to its HSDM's precursors, FM-HSDM enjoys Fejér monotonicity, the step-size parameter stays constant across iterations to promote convergence speed-ups of the sequence of estimates to a minimizer, while only Lipschitzian continuity, and not strong monotonicity, of the derivative of the smooth-loss function is needed to ensure convergence. FM-HSDM utilizes fixed-point theory, variational inequalities and affine-nonexpansive mappings to accommodate affine constraints in a more versatile way than state-of-the-art primal–dual techniques and the alternating direction method of multipliers do. Recursions can be tuned to score low computational footprints, well-suited for large-scale optimization tasks, without compromising convergence guarantees. Results on the rate of convergence to an optimal point are also presented. Finally, numerical tests on synthetic data are used to validate the theoretical findings.

ARTICLE HISTORY
Received 9 August 2017
Accepted 22 July 2018

KEYWORDS
Convex optimization; composite loss; affine constraints; nonexpansive mapping; Fejér monotonicity

2010 MATHEMATICS SUBJECT CLASSIFICATIONS
90C25; 65K15

1. Introduction

1.1. Problem and notation

Problem 1.1: This paper considers the following composite convex minimization task:

$$\min_{x\in\mathcal{A}\subset\mathcal{X}} f(x) + g(x), \tag{1}$$

CONTACT K. Slavakis kslavaki@buffalo.edu Department of Electrical Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260-2500, USA
*Preliminary parts of this study can be found in [1,2].

where $\mathcal{X}$ is a real Hilbert space, the loss functions f,g belong to the class $\Gamma_0(\mathcal{X})$ of all convex, proper and lower semicontinuous functions from $\mathcal{X}$ to $(-\infty, +\infty]$ [3, p. 132], f is everywhere (Fréchet) differentiable with L-Lipschitz-continuous derivative ∇f, i.e. there exists an $L \in \mathbb{R}_{>0}$ such that (s.t.) $\|\nabla f(x_1) - \nabla f(x_2)\| \leq L\|x_1 - x_2\|$, $\forall x_1, x_2 \in \mathcal{X}$, and $\mathcal{A}$ is a closed affine subset of $\mathcal{X}$. Throughout the manuscript, it is assumed that (1) possesses a solution.

Symbols $\mathbb{Z}$ and $\mathbb{R}$ stand for sets of all integer and real numbers, respectively. Moreover, $\mathbb{Z}_{>0} := \{1, 2, \ldots\} \subset \{0, 1, 2, \ldots\} =: \mathbb{Z}_{\geq 0}$, while $\mathbb{R}_{>0} := (0, +\infty)$. The algorithms of this paper are built on a real Hilbert space $\mathcal{X}$, equipped with an inner product $\langle \cdot \mid \cdot \rangle$, with vectors denoted by lower case letters, e.g. x. In the special case where $\mathcal{X}$ is finite dimensional, i.e. Euclidean, vectors of $\mathcal{X}$ are denoted by boldfaced lower case letters, e.g. $\mathbf{x}$, while boldfaced upper case letters are reserved for matrices, e.g. $\mathbf{Q}$. Symbol Id denotes the identity mapping in $\mathcal{X}$, i.e. Id $x = x$, $\forall x \in \mathcal{X}$. In the special case where $\mathcal{X}$ is Euclidean, Id boils down to the identity matrix, denoted by $\mathbf{I}$. Vector/matrix transposition is denoted by the superscript $\top$. For $g \in \Gamma_0(\mathcal{X})$, ∂g denotes the set-valued subdifferential operator which is defined as $x \mapsto \partial g(x) := \{\xi \in \mathcal{X} \mid g(x) + \langle x' - x | \xi \rangle \leq g(x'), \forall x' \in \mathcal{X}\}$.

Let $\mathfrak{B}(\mathcal{X}, \mathcal{X}')$ denote all bounded linear operators from $\mathcal{X}$ to $\mathcal{X}'$ [4], and $\mathfrak{B}(\mathcal{X}) := \mathfrak{B}(\mathcal{X}, \mathcal{X})$. For $Q \in \mathfrak{B}(\mathcal{X}, \mathcal{X}')$, $\|Q\| < \infty$ stands for the norm of Q. Mapping $Q^* \in \mathfrak{B}(\mathcal{X}', \mathcal{X})$ stands for the adjoint of $Q \in \mathfrak{B}(\mathcal{X}, \mathcal{X}')$ [4]. In the case of matrices, the adjoint of a mapping $\mathbf{Q}$ is nothing but the transpose $\mathbf{Q}^\top$. Mapping $Q \in \mathfrak{B}(\mathcal{X})$ is called self-adjoint if $Q^* = Q$. In the case of a symmetric matrix $\mathbf{Q}$, $\lambda(\mathbf{Q})$ denotes an eigenvalue of $\mathbf{Q}$. Furthermore, $\|\mathbf{Q}\| = \sigma_{\max}(\mathbf{Q}) := \lambda_{\max}^{1/2}(\mathbf{Q}^\top\mathbf{Q})$ stands for the (spectral) norm of $\mathbf{Q}$, where $\sigma_{\max}(\cdot) \in \mathbb{R}_{>0}$ denotes the maximum singular value and $\lambda_{\max}(\cdot)$ the maximum eigenvalue of a matrix.

1.2. Background and contributions

1.2.1. The hybrid steepest descent method

To solve (1), this paper extrapolates the paths established by the hybrid steepest descent method (HSDM), which was originally introduced to solve a variational-inequality problem of a strongly monotone operator over the fixed-point set of a nonexpansive mapping [5] (see also, e.g. [6–8] and references therein, for a wider applicability of HSDM in other scenarios). In the context of (1), a version of HSDM solves

$$\min_{x \in \mathrm{Fix}\, T} f(x), \tag{2}$$

where f is a strongly convex function and Fix $T \subset \mathcal{X}$ denotes the fixed-point set of a nonexpansive mapping $T : \mathcal{X} \to \mathcal{X}$ (*cf.* Section 2). For an arbitrarily fixed starting point x_0, HSDM generates the sequence

$$x_{n+1} := Tx_n - \lambda_n \nabla f(Tx_n), \tag{3}$$

which strongly converges to the *unique* minimizer of (2). To secure strong convergence, the step sizes $(\lambda_n)_{n\in\mathbb{Z}_{\geq 0}} \subset \mathbb{R}_{\geq 0}$ satisfy (i) $\sum_{n\in\mathbb{Z}_{\geq 0}} \lambda_n = +\infty$, (ii) $\lim_{n\to\infty} \lambda_n = 0$ and (iii) $\sum_{n\in\mathbb{Z}_{\geq 0}} |\lambda_{n+1} - \lambda_n| < +\infty$. Furthermore, in the case where $\mathcal{X}$ is Euclidean, f is not necessarily strongly convex, and T is attracting nonexpansive [9,10] with bounded Fix T, the requirements on $(\lambda_n)_{n\in\mathbb{Z}_{\geq 0}}$ can be relaxed to (i) $\sum_{n\in\mathbb{Z}_{\geq 0}} \lambda_n = +\infty$, (ii) $\sum_{n\in\mathbb{Z}_{\geq 0}} \lambda_n^2 < +\infty$ for achieving $\lim_{n\to\infty} d_{\mathcal{X}}(\mathbf{x}_n, \operatorname{Arg\,min}_{\operatorname{Fix} T} f) = 0$, where $d_{\mathcal{X}}(\mathbf{x}_n, \operatorname{Arg\,min}_{\operatorname{Fix} T} f)$ stands for the (metric) distance of point $\mathbf{x}_n$ from the set of minimizers of f over Fix T [9]. To speed up HSDM's convergence rate, conjugate-gradient-based variants were introduced in [11–13]. For example, for an arbitrarily fixed starting point $x_0 \in \mathcal{X}$, and $d_0 := -\nabla f(x_0)$, the following recursions (i) $x_{n+1} := T(x_n + \mu\lambda_n d_n)$; (ii) $d_{n+1} := -\nabla f(x_{n+1}) + \beta_{n+1} d_n$, with $\mu > 0$, $\lambda_n \in (0,1]$, $\beta_n \in [0,\infty)$ were introduced in [11]. If $\mu \in (0, 2\eta/L^2)$, $\lim_{n\to\infty} \beta_n = 0$, $(\nabla f(x_n))_{n\in\mathbb{Z}_{\geq 0}}$ is bounded, and (i) $\sum_{n\in\mathbb{Z}_{\geq 0}} \lambda_n = +\infty$, (ii) $\lim_{n\to\infty} \lambda_n = 0$, (iii) $\sum_{n\in\mathbb{Z}_{\geq 0}} |\lambda_{n+1} - \lambda_n| < +\infty$, (iv) $\lambda_n/\lambda_{n+1} \leq \sigma$, $(\sigma \geq 1)$, then $(x_n)_{n\in\mathbb{Z}_{\geq 0}}$ converges strongly to the unique minimizer of (2).

1.2.2. Prior art

To demonstrate the connections of (1) with state-of-the-art methods, it is helpful to notice that the concise description (1) can be unfolded in several ways to describe a large variety of convex composite minimization tasks, e.g.

$$\min_{x\in\mathscr{A}} f(x) + \sum_{j=1}^{J} g_j(H_j x - r_j), \tag{4}$$

where $\{\mathscr{X}_j\}_{j=0}^J$ are real Hilbert spaces, $f \in \Gamma_0(\mathscr{X}_0)$, $g_j \in \Gamma_0(\mathscr{X}_j)$, $H_j \in \mathfrak{B}(\mathscr{X}_0, \mathscr{X}_j)$ and $r_j \in \mathscr{X}_j$, $j \in \{1,\dots,J\}$. Moreover, ∇f is L-Lipschitz continuous and $\mathscr{A}$ is a closed affine subset of $\mathscr{X}_0$. Indeed, it can be verified that (4) can be recast as (1) via $\mathcal{X} := \mathscr{X}_0 \times \mathscr{X}_1 \times \cdots \times \mathscr{X}_J = \{x := (x^{(0)}, x^{(1)}, \dots, x^{(J)}) \mid x^{(j)} \in \mathscr{X}_j, \forall j \in \{0,1,\dots,J\}\}$, $f(x) := f(x^{(0)})$, $g(x) := \sum_{j=1}^J g_j(x^{(j)})$, and the closed affine set $\mathcal{A} := \{x \in \mathcal{X} \mid x^{(0)} \in \mathscr{A}, x^{(j)} = H_j x^{(0)} - r_j, \forall j \in \{1,\dots,J\}\}$. Task (4), in the case where $J=2$, $\mathscr{X} := \mathscr{X}_0 = \mathscr{X}_1$, $H_1 = \mathrm{Id}$, $r_1 = r_2 = 0$, and $\mathscr{A} := \mathscr{X}$, i.e. $\min_{x\in\mathscr{X}} [f(x) + g_1(x) + g_2(H_2 x)]$, has been already studied, e.g. via the primal–dual (PD) algorithmic framework [14–17]. Gradient ∇f, proximal mappings (*cf.* Definition 2.5) $\operatorname{Prox}_{g_1}$ and $\operatorname{Prox}_{g_2^*} = \mathrm{Id} - \operatorname{Prox}_{g_2}$ [3, Rem. 14.4, p. 198], where g_2^* stands for the (Fenchel) conjugate of g_2, as well as adjoint H_2^* are utilized in a computationally efficient way to generate a sequence $(x_n)_{n\in\mathbb{Z}_{\geq 0}} \subset \mathscr{X}$, which converges weakly (and under certain hypotheses, strongly) to a solution of the previous minimization task. Moreover, task (4), in the case where $J=2$, $\mathscr{X} := \mathscr{X}_0 = \mathscr{X}_1 = \mathscr{X}_2$, $H_1 = H_2 = \mathrm{Id}$, $r_1 = r_2 = 0$ and $\mathscr{A} := \mathcal{X}$, i.e. $\min_{x\in\mathscr{X}} [f(x) + g_1(x) + g_2(x)]$, has also attracted attention in the context of the 'three-term operator splitting' framework [18,19]. As in [14–16], ∇f, $\operatorname{Prox}_{g_1}$ and

Prox_{g_2} are employed via computationally efficient recursions in [18,19] to generate a sequence which converges weakly (and under certain hypotheses, strongly) to a solution of the minimization task at hand. All studies in [14–16,18,19] set $\mathscr{A} := \mathscr{X}$. In the case of $\mathscr{A} \subsetneq \mathscr{X}$, one can accommodate the affine constraint $\mathscr{A}$ via the use of the indicator function $\iota_{\mathscr{A}}$ [$\iota_{\mathscr{A}}(x) := 0$, if $x \in \mathscr{A}$, and $\iota_{\mathscr{A}}(x) := +\infty$, if $x \notin \mathscr{A}$] and the additional loss $g_3 := \iota_{\mathscr{A}}$. According to the previous discussion, such an accommodation entails the use of $\mathrm{Prox}_{\iota_{\mathscr{A}}} = P_{\mathscr{A}}$, where $P_{\mathscr{A}}$ denotes the metric projection mapping onto $\mathscr{A}$. Mapping $P_{\mathscr{A}}$ may become computationally demanding, e.g. in the case where $\mathscr{X}$ is a Euclidean space and the affine constraints are described by a matrix of large dimensions (*cf.* Fact A.3), since computing $P_{\mathscr{A}}$ necessitates the costly singular value decomposition of the matrix under query (*cf.* Example A.4). Task (1) in the case where $\mathcal{X}$ is a Euclidean space and $\mathcal{A} := \{\mathbf{x} \in \mathcal{X} \mid \mathbf{a}^\top \mathbf{x} = 0\}$, for some $\mathbf{a} \in \mathcal{X} \setminus \{\mathbf{0}\}$, was treated, within a stochastic setting, in [20].

The celebrated alternating direction method of multipliers (ADMM) [21–25] deals with the task

$$\min_{(x^{(1)},x^{(2)})\in\mathscr{X}_1\times\mathscr{X}_2} g_1(x^{(1)}) + g_2(x^{(2)}) \tag{5a}$$

$$\text{s.to } H_1x^{(1)} + H_2x^{(2)} = r, \tag{5b}$$

where $H_j \in \mathfrak{B}(\mathscr{X}_j, \mathscr{X}_0)$ and $r \in \mathscr{X}_0$. Again, (5) can be recast as (1) under the following setting: $\mathcal{X} := \mathscr{X}_1 \times \mathscr{X}_2 = \{x := (x^{(1)}, x^{(2)}) \mid x^{(1)} \in \mathscr{X}_1, x^{(2)} \in \mathscr{X}_2\}$, $f(x) := 0$, $g(x) := g_1(x^{(1)}) + g_2(x^{(2)})$, and $\mathcal{A} := \{x \in \mathcal{X} \mid H_1x^{(1)} + H_2x^{(2)} = r\}$. Provided that the inverse mappings $(\lambda H_1^*H_1 + \partial g_1)^{-1}$ and $(\lambda H_2^*H_2 + \partial g_2)^{-1}$ exist, the recursive application of $(\lambda H_1^*H_1 + \partial g_1)^{-1}$ and $(\lambda H_2^*H_2 + \partial g_2)^{-1}$ generates a sequence which converges weakly to a solution of (5) [24,25]. ADMM enjoys extremely wide popularity for minimization problems in Euclidean spaces [23], at the expense of the computation of $(\lambda H_1^*H_1 + \partial g_1)^{-1}$ and $(\lambda H_2^*H_2 + \partial g_2)^{-1}$: there may be cases where computing the previous inverse mappings entails the costly task of solving a convex minimization subproblem.

The motivation for the present paper is the algorithmic solution given in the distributed minimization context of [26,27]: for a Euclidean $\mathscr{X}$, and a collection of loss functions $\{f_j, g_j \in \Gamma_0(\mathscr{X})\}_{j=1}^J$, where f_j is everywhere differentiable with an L_j-Lipschitz continuous ∇f_j, $\forall j \in \{1, \ldots, J\}$, nodes $\mathcal{N}$ ($|\mathcal{N}| = J$), connected by edges $\mathcal{E}$ within a network/graph $\mathcal{G} := (\mathcal{N}, \mathcal{E})$, operate in parallel and cooperate to solve

$$\min_{(\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(J)})\in\mathscr{X}^J} \sum_{j=1}^J f_j(\mathbf{x}^{(j)}) + \sum_{j=1}^J g_j(\mathbf{x}^{(j)}) \tag{6a}$$

$$\text{s.to} \quad \mathbf{x}^{(1)} = \cdots = \mathbf{x}^{(J)}. \tag{6b}$$

Each node $j \in \mathcal{N}$ operates only on the pair (f_j, g_j) and communicates the information regarding its updates to its neighbouring nodes to cooperatively solve (6), under the consensus constraint of (6b). Once again, (6) can be seen as a special case of (1) under the following considerations: $\mathcal{X} := \mathscr{X}^J$, $f(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(J)}) := \sum_{j=1}^{J} f(\mathbf{x}^{(j)})$, $g(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(J)}) := \sum_{j=1}^{J} g(\mathbf{x}^{(j)})$ and $\mathcal{A} := \{(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(J)}) \in \mathcal{X} \mid \mathbf{x}^{(1)} = \cdots = \mathbf{x}^{(J)}\}$. Upon defining the $J \times J$ mixing matrices $\mathbf{W} = [w_{ij}]$, $\tilde{\mathbf{W}} = [\tilde{w}_{ij}]$, [27] introduced the following recursions to solve (6): for an arbitrarily fixed starting-point $J \times \dim \mathscr{X}$ matrix $\mathbf{X}_0$, as well as $\mathbf{X}_{1/2} := \mathbf{W}\mathbf{X}_0 - \lambda\nabla f(\mathbf{X}_0)$ and $\mathbf{X}_1 := \mathrm{Prox}_{\lambda g}(\mathbf{X}_{1/2})$, repeat for all $n \in \mathbb{Z}_{\geq 0}$, (i) $\mathbf{X}_{n+3/2} := \mathbf{X}_{n+1/2} + \mathbf{W}\mathbf{X}_{n+1} - \tilde{\mathbf{W}}\mathbf{X}_n - \lambda[\nabla f(\mathbf{X}_{n+1}) - \nabla f(\mathbf{X}_n)]$; (ii) $\mathbf{X}_{n+2} := \mathrm{Prox}_{\lambda g}(\mathbf{X}_{n+3/2})$. If (i) $(i,j) \notin \mathcal{E} \Rightarrow w_{ij} = \tilde{w}_{ij} = 0$, (ii) $\mathbf{W}^\top = \mathbf{W}$, $\tilde{\mathbf{W}}^\top = \tilde{\mathbf{W}}$, (iii) $\ker(\mathbf{W} - \tilde{\mathbf{W}}) = \mathrm{span}\,\mathbf{1} \subset \ker(\mathbf{I} - \tilde{\mathbf{W}})$, (iv) $\tilde{\mathbf{W}} \succ \mathbf{0}$, (v) $(1/2)(\mathbf{I} + \tilde{\mathbf{W}}) \succeq \tilde{\mathbf{W}} \succeq \mathbf{W}$ and (vi) $\lambda \in (0, 2\lambda_{\min}(\tilde{\mathbf{W}})/\max_i L_i)$, then the sequence $(\mathbf{X}_n)_{n\in\mathbb{N}}$ converges to a matrix whose rows provide a solution to (6).

1.2.3. Contributions

Driven by the similarity between the algorithmic solution of [26,27] and HSDM, and aiming at solving (1), this study introduces a new member to the HSDM family of algorithms: the *Fejér-monotone* (FM-)HSDM. Building around the simple recursion of (3) and the concept of a nonexpansive mapping, FM-HSDM's recursions offer sequences which converge weakly and, under certain hypotheses (uniform convexity of loss functions), strongly to a solution of (1); *cf.* Theorems 3.1 and 3.6. Fixed-point theory, variational inequalities and affine-nonexpansive mappings are utilized to accommodate the affine constraint $\mathcal{A}$ in a more flexible way (see, e.g. Proposition 2.10 and Example A.4) than the usage of the indicator function and its associated metric-projection mapping that methods [15,16,18,19] promote. Such flexibility is combined with the first-order information of f and the proximal mapping of g to build recursions of tunable complexity that can score low-computational complexity footprints, well-suited for large-scale minimization tasks. FM-HSDM enjoys Fejér monotonicity, and in contrast to (3) as well as its conjugate gradient-based variants [11–13], only Lipschitzian continuity, and not strong monotonicity, of the derivative of the smooth-part loss is needed to establish convergence of the sequence of estimates. Furthermore, a constant step-size parameter is utilized to effect convergence speed-ups. Finally, as opposed to [11–13], the advocated scheme needs no boundedness assumptions on estimates or gradients to establish weak (or even strong) convergence of the sequence of estimates to a solution of (1). Results on the rate of convergence to an optimal point are also presented. Numerical tests on synthetic data are used to validate the theoretical findings.

2. Affine nonexpansive mappings and variational inequalities

2.1. Nonexpansive mappings and fixed-point sets

Definition 2.1: A self-adjoint mapping $Q \in \mathfrak{B}(\mathcal{X})$ is called *positive* if $\langle Qx|x\rangle \geq 0$, $\forall x \in \mathcal{X}$ [4, Sec. 9.3]. Moreover, the self-adjoint $\Pi \in \mathfrak{B}(\mathcal{X})$ is called *strongly* *positive* if there exists $\delta \in \mathbb{R}_{>0}$ s.t. $\langle \Pi x | x \rangle \geq \delta \|x\|^2$, $\forall x \in \mathcal{X}$. In the context of matrices, $\mathbf{Q}$ is positive iff $\mathbf{Q}$ is positive semidefinite, i.e. $\mathbf{Q} \succeq \mathbf{0}$. Moreover, $\mathbf{\Pi}$ is strongly positive iff $\mathbf{\Pi}$ is positive definite, i.e. $\mathbf{\Pi} \succ \mathbf{0}$, and δ in the previous definition can be taken to be $\lambda_{\min}(\mathbf{\Pi})$.

For a strongly positive Π, $\langle \cdot | \cdot \rangle_\Pi$ stands for the inner product $\langle x | x' \rangle_\Pi := \langle x | \Pi x' \rangle$, $\forall (x, x') \in \mathcal{X}^2$. For a function $\varphi : \mathcal{X} \to \mathbb{R}$, $\nabla \varphi$ and $\nabla \varphi(x)$ stand for the (Gâteaux/Fréchet) derivative and gradient at $x \in \mathcal{X}$, respectively [3, Sec. 2.6, p. 37]. Given $Q \in \mathfrak{B}(\mathcal{X})$, $\ker Q$ stands for the linear subspace $\ker Q := \{x \in \mathcal{X} \mid Qx = 0\}$. Moreover, ran Q denotes the linear subspace ran $Q := Q\mathcal{X} := \{Qx \mid x \in \mathcal{X}\}$. For the case of a matrix $\mathbf{Q}$, ran $\mathbf{Q}$ is the linear subspace spanned by the columns of $\mathbf{Q}$. Finally, the orthogonal complement of a linear subspace is denoted by the superscript $\perp$.

Definition 2.2: The *fixed-point set* of a mapping $T : \mathcal{X} \to \mathcal{X}$ is defined as the set $\operatorname{Fix} T := \{x \in \mathcal{X} \mid Tx = x\}$.

Definition 2.3: Mapping $T : \mathcal{X} \to \mathcal{X}$ is called

(i) *Nonexpansive,* if $\|Tx - Tx'\| \leq \|x - x'\|$, $\forall (x, x') \in \mathcal{X}^2$.
(ii) *Firmly nonexpansive,* if $\|Tx - Tx'\|^2 \leq \langle x - x' | Tx - Tx' \rangle$, $\forall (x, x') \in \mathcal{X}^2$. Any firmly nonexpansive mapping is nonexpansive [3, Sec. 4.1].
(iii) *α-averaged (nonexpansive)*, if there exist an $\alpha \in (0, 1)$ and a nonexpansive mapping $R : \mathcal{X} \to \mathcal{X}$ s.t. $T = \alpha R + (1 - \alpha)\mathrm{Id}$. It can be easily verified that T is nonexpansive with $\operatorname{Fix} R = \operatorname{Fix} T$.

Fact 2.4 ([3, Cor. 4.15, p. 63]): The fixed-point set $\operatorname{Fix} T$ of a nonexpansive mapping T is closed and convex.

Definition 2.5: Given $f \in \Gamma_0(\mathcal{X})$ and $\gamma \in \mathbb{R}_{>0}$, the *proximal* mapping $\operatorname{Prox}_{\gamma f}$ is defined as $\operatorname{Prox}_{\gamma f} : \mathcal{X} \to \mathcal{X} : x \mapsto \arg\min_{z \in \mathcal{X}} (\gamma f(z) + \frac{1}{2}\|x - z\|^2)$.

Example 2.6:

(i) [3, Prop. 4.8, p. 61] Given a non-empty closed convex set $\mathcal{C} \subset \mathcal{X}$, the *metric projection mapping onto* $\mathcal{C}$, defined as $P_{\mathcal{C}} : \mathcal{X} \to \mathcal{C} : x \mapsto P_{\mathcal{C}} x$, with $P_{\mathcal{C}} x$ being the unique minimizer of $\min_{z \in \mathcal{C}} \|x - z\|$, is firmly nonexpansive with $\operatorname{Fix} P_{\mathcal{C}} = \mathcal{C}$.
(ii) [3, Prop. 12.27, p. 176] Given $f \in \Gamma_0(\mathcal{X})$ and $\gamma \in \mathbb{R}_{>0}$, the proximal mapping $\operatorname{Prox}_{\gamma f}$ is firmly nonexpansive with $\operatorname{Fix} \operatorname{Prox}_{\gamma f} = \arg\min f$.
(iii) [3, Prop. 4.2, p. 60] T is firmly nonexpansive iff $\mathrm{Id} - T$ is firmly nonexpansive iff T is $(1/2)$-averaged iff $2T - \mathrm{Id}$ is nonexpansive.
(iv) [28, Prop. 2.2], [9, Thm. 3(b)]. Let $\{T_j\}_{j=1}^J$ be a finite family ($J \in \mathbb{Z}_{>0}$) of nonexpansive mappings from $\mathcal{X}$ to $\mathcal{X}$, and $\{\omega_j\}_{j=1}^J$ be real numbers in $(0, 1]$

s.t. $\sum_{j=1}^{J} \omega_j = 1$. Then, $T := \sum_{j=1}^{J} \omega_j T_j$ is nonexpansive. If $\cap_{j=1}^{J} \mathrm{Fix}\ T_j \neq \emptyset$, then $\mathrm{Fix}\ T = \cap_{j=1}^{J} \mathrm{Fix}\ T_j$. Furthermore, consider real numbers $\{\alpha_j\}_{j=1}^{J} \subset (0, 1)$ s.t. T_j is α_j-averaged, $\forall j$. Define $\alpha := \sum_{j=1}^{J} \omega_j \alpha_j$. Then, T is α-averaged. Hence, if each T_j is firmly nonexpansive, i.e. $(1/2)$-averaged, then T is also firmly nonexpansive.

(v) [28, Prop. 2.5], [9, Thm. 3(b)] Let $\{T_j\}_{j=1}^{J}$ be a finite family ($J \in \mathbb{Z}_{>0}$) of nonexpansive mappings from $\mathcal{X}$ to $\mathcal{X}$. Then, mapping $T := T_1 T_2 \cdots T_J$ is nonexpansive. If $\cap_{j=1}^{J} \mathrm{Fix}\ T_j \neq \emptyset$, then $\mathrm{Fix}\ T = \cap_{j=1}^{J} \mathrm{Fix}\ T_j$. Furthermore, consider real numbers $\{\alpha_j\}_{j=1}^{J} \subset (0, 1)$ s.t. T_j is α_j-averaged, $\forall j$. Define

$$\alpha := \frac{1}{1 + \frac{1}{\sum_{j=1}^{J} \frac{\alpha_j}{1-\alpha_j}}}.$$

Then, T is α-averaged.

In what follows, function $f \in \Gamma_0(\mathcal{X})$ is considered to have an L-Lipschitz continuous ∇f with $\mathrm{dom}\ \nabla f = \mathcal{X}$. By [3, Prop. 16.3(i), p. 224], the previous condition leads to $\mathrm{dom}\ f = \mathcal{X}$, which further implies by [3, Cor. 16.38(iii), p. 234] that $\partial(f + g) = \nabla f + \partial g$.

2.2. Affine nonexpansive mappings

Definition 2.7 ([3, p. 3]): A mapping $T : \mathcal{X} \to \mathcal{X}$ is called *affine* if there exist a linear mapping $Q : \mathcal{X} \to \mathcal{X}$ and a $\pi \in \mathcal{X}$ s.t. $Tx = Qx + \pi, \forall x \in \mathcal{X}$.

Fact 2.8 ([3, Ex. 4.4, p. 72]): Consider the affine mapping $Tx = Qx + \pi, \forall x \in \mathcal{X}$, with Q being linear and $\pi \in \mathcal{X}$. Then, T is nonexpansive iff $\|Q\| \leq 1$.

Define now the following special class of affine-nonexpansive mappings:

$$\mathfrak{T} := \left\{ T : \mathcal{X} \to \mathcal{X} \;\middle|\; \begin{array}{l} Tx = Qx + \pi, \forall x \in \mathcal{X} \\ Q \in \mathfrak{B}(\mathcal{X}); \pi \in \mathcal{X} \\ \|Q\| \leq 1, Q \text{ is positive} \end{array} \right\}. \tag{7}$$

As the following proposition highlights, T is nothing but the class of affine firmly nonexpansive mappings.

Proposition 2.9: *$T \in \mathfrak{T}$ iff $T = Q + \pi$, where $Q \in \mathfrak{B}(\mathcal{X})$ is self-adjoint, $\pi \in \mathcal{X}$, and T is firmly nonexpansive.*

Proof: First, consider $T \in \mathfrak{T}$. Since Q is positive, let $Q^{1/2}$ be the *positive square root* of Q, i.e. the (unique) positive operator which satisfies $Q^{1/2} Q^{1/2} = Q$ [4, Thm. 9.4-2, p. 476]. The positivity of Q yields $\|Q\| =$

$\sup_{x\in\mathcal{X}\setminus\{0\}} |\langle Qx|x\rangle|/\langle x|x\rangle = \sup_{x\in\mathcal{X}\setminus\{0\}} \langle Qx|x\rangle/\langle x|x\rangle$, according to [4, Thm. 9.2-2, p. 466]. Then, $\forall(x,x') \in \mathcal{X}^2$,

$$\begin{aligned}\|Tx - Tx'\|^2 &= \|Qx - Qx'\|^2 = \|Q(x-x')\|^2 = \langle Q(x-x')|Q(x-x')\rangle \\ &= \langle Q^{1/2}(x-x')|QQ^{1/2}(x-x')\rangle \leq \|Q\|\langle Q^{1/2}(x-x')|Q^{1/2}(x-x')\rangle \\ &\leq \langle Q^{1/2}(x-x')|Q^{1/2}(x-x')\rangle = \langle x-x'|Q(x-x')\rangle \\ &= \langle x-x'|Tx-Tx')\rangle,\end{aligned}$$

which suggests that T is firmly nonexpansive.

Now, let $T = Q + \pi$, for a self-adjoint $Q \in \mathfrak{B}(\mathcal{X})$, $\pi \in \mathcal{X}$. Let also T be firmly nonexpansive. Then, $\forall x \in \mathcal{X}$, $\langle x|Qx\rangle = \langle x-0|Q(x-0)\rangle = \langle x-0|Tx - T0)\rangle \geq \|Tx - T0\|^2 \geq 0$; thus Q is positive. By the fact that a firmly nonexpansive mapping is nonexpansive [Definition 2.3(ii)] and Fact 2.8, $\|Q\| \leq 1$. In summary, $T \in \mathfrak{T}$. ■

Proposition 2.10: *Let* $J \in \mathbb{Z}_{>0}$.

(i) *Consider a family* $\{T_j\}_{j=1}^J$ *of members of* $\mathfrak{T}$. *For any set of weights* $\{\omega_j\}_{j=1}^J$ *s.t.* $\omega_j \in (0,1]$ *and* $\sum_{j=1}^J \omega_j = 1$, *mapping* $\sum_{j=1}^J \omega_j T_j x \in \mathfrak{T}$.

(ii) *Consider* $T_0 := Q_0 + \pi_0 \in \mathfrak{T}$. *Moreover, let the self-adjoint* $Q_j \in \mathfrak{B}(\mathcal{X})$, *with* $\|Q_j\| \leq 1$, *and* $\pi_j \in \mathcal{X}, \forall j \in \{1,\ldots,J\}$. *Let now the family* $\{T_j := Q_j + \pi_j\}_{j=1}^J$ *of affine nonexpansive mappings, where each* T_j *does not necessarily belong to* $\mathfrak{T}$, *i.e.* $\{Q_j\}_{j=1}^J$ *might not be positive according to Proposition 2.9. Then, the composition*

$$\begin{aligned}T_J T_{J-1}\ldots T_1 T_0 T_1\ldots T_{J-1}T_J x = Q_J Q_{J-1}\ldots Q_1 Q_0 Q_1\ldots Q_{J-1}Q_J x \\ + \sum_{j=1}^J Q_J Q_{J-1}\ldots Q_1 Q_0 Q_1\ldots Q_{j-1}\pi_j \\ + \sum_{j=1}^J Q_J Q_{J-1}\ldots Q_j \pi_{j-1} + \pi_J, \quad \forall x \in \mathcal{X},\end{aligned}$$

satisfies $T_J T_{J-1}\ldots T_1 T_0 T_1 \ldots T_{J-1}T_J \in \mathfrak{T}$.

Proof: The proof of Proposition 2.10(i) follows easily from Example 2.6(iv) and Proposition 2.9. The formula appearing in Proposition 2.10(ii) can be deduced by mathematical induction on J. Furthermore, $Q_J Q_{J-1}\ldots Q_1 Q_0 Q_1 \ldots Q_{J-1}Q_J$ is self-adjoint, and its positivity follows from the fundamental observation that $\forall x \in \mathcal{X}$, $\langle Q_J Q_{J-1}\cdots Q_1 Q_0 Q_1 \cdots Q_{J-1}Q_J x|x\rangle = \langle Q_0(Q_1\cdots Q_{J-1}Q_J x)|Q_1\cdots Q_{J-1}Q_J x\rangle \geq 0$, due to the positivity of Q_0. Finally, the claim of Proposition 2.10(ii) is established by $\|Q_J\ldots Q_1 Q_0 Q_1 \ldots Q_J\| \leq \|Q_0\| \prod_{j=1}^J \|Q_j\|^2 \leq 1$. ■

Proposition 2.11: *Given the closed affine set $\mathcal{A} \subset \mathcal{X}$, define the following family of mappings*:

$$\mathfrak{T}_{\mathcal{A}} := \{T \in \mathfrak{T} \mid \operatorname{Fix} T = \mathcal{A}\}\,. \tag{8}$$

Then, $\mathfrak{T}_{\mathcal{A}}$ is non-empty.

Proof: The metric projection mapping $P_{\mathcal{A}}$ onto $\mathcal{A}$ is not only firmly nonexpansive with $\operatorname{Fix} P_{\mathcal{A}} = \mathcal{A}$ [*cf.* Example 2.6(i)] but also affine, according also to [3, Cor. 3.20(ii), p. 48]. Hence, by virtue of Proposition 2.9, $P_{\mathcal{A}} \in \mathfrak{T}_{\mathcal{A}} \neq \emptyset$. ∎

It can be verified that the fixed-point set $\operatorname{Fix} T$ of an affine mapping T is affine. However, more can be said about the members of $\mathfrak{T}_{\mathcal{A}}$.

Proposition 2.12: *For any $T \in \mathfrak{T}_{\mathcal{A}}$,*

$$\mathcal{A} = \operatorname{Fix} T = \ker(\mathrm{Id} - Q) + w_* = \ker U + w_*,$$

where w_* is any vector of $\mathcal{A}$ and U is the positive square root of $\mathrm{Id} - Q$, *i.e. the (unique) positive operator which satisfies $U^2 = \mathrm{Id} - Q$ [4, Thm. 9.4-2, p. 476].*

Proof: Since $\|Q\| = \sup_{x\in\mathcal{X}\setminus\{0\}} |\langle Qx|x\rangle|/\|x\|^2$ [4, Thm. 9.2-2, p. 466] and $\|Q\| \leq 1$, it can be easily verified that $\forall x \in \mathcal{X}$, $\langle(\mathrm{Id} - Q)x|x\rangle = \|x\|^2 - \langle Qx|x\rangle \geq \|x\|^2 - \|Q\| \cdot \|x\|^2 \geq \|x\|^2 - \|x\|^2 = 0$, i.e. $\mathrm{Id} - Q$ is positive. Interestingly, the positivity of Q suggests that $\forall x \in \mathcal{X}$, $\langle(\mathrm{Id} - Q)x|x\rangle = \|x\|^2 - \langle Qx|x\rangle \leq \|x\|^2$, which implies, via [4, Thm. 9.2-2, p. 466], that $\|\mathrm{Id} - Q\| \leq 1$. Moreover, by the definition of T, it follows that for any arbitrarily fixed $w_* \in \operatorname{Fix} T$,

$$\begin{aligned}
\operatorname{Fix} T &= \{x \mid Tx = x\} = \{x \mid (\mathrm{Id} - T)x = 0\}\\
&= \{x \mid (\mathrm{Id} - Q)x = \pi\} = \{x \mid (\mathrm{Id} - Q)x = (\mathrm{Id} - Q)w_*\}\\
&= \{x \mid (\mathrm{Id} - Q)(x - w_*) = 0\} = \{x' + w_* \mid (\mathrm{Id} - Q)x' = 0\}\\
&= \ker(\mathrm{Id} - Q) + w_*\,.
\end{aligned}$$

Finally, the characterization $\operatorname{Fix} T = \ker U + w_*$ follows from the previous arguments and $x' \in \ker(\mathrm{Id} - Q) \Leftrightarrow (\mathrm{Id} - Q)x' = 0 \Rightarrow U^2x' = 0 \Rightarrow U^*Ux' = 0 \Rightarrow \langle x'|U^*Ux'\rangle = \langle Ux'|Ux'\rangle = \|Ux'\|^2 = 0 \Rightarrow Ux' = 0 \Leftrightarrow x' \in \ker U \Rightarrow U^2x' = 0 \Rightarrow (\mathrm{Id} - Q)x' = 0 \Leftrightarrow x' \in \ker(\mathrm{Id} - Q)$, which establishes $\ker(\mathrm{Id} - Q) = \ker U$. ∎

Several examples of $\mathfrak{T}_{\mathcal{A}}$ members playing important roles in convex minimization tasks can be found in Appendix 1.

2.3. Variational-inequality problems

Definition 2.13 (Variational-inequality problem): For a nonexpansive mapping $T : \mathcal{X} \to \mathcal{X}$, point $x_* \in \operatorname{Fix} T$ is said to solve the variational-inequality problem $\operatorname{VIP}(\nabla f + \partial g, \operatorname{Fix} T)$ if there exists $\xi_* \in \partial g(x_*)$ s.t. $\forall y \in \operatorname{Fix} T$, $\langle y - x_* | \nabla f(x_*) + \xi_* \rangle \geq 0$.

Fact 2.14 ([3, Prop. 26.5(vi), p. 383]): Consider a mapping $T \in \mathfrak{T}_{\mathcal{A}}$ (recall $\operatorname{Fix} T = \mathcal{A}$), and assume that one of the following holds:

(1) $0 \in \operatorname{sri}(\mathcal{A} - \operatorname{domain}(f + g))$ ([*cf.*]Prop. 6.19, p. 95] [3] for special cases);
(2) $\mathcal{X}$ is Euclidean and $\mathcal{A} \cap \operatorname{ri}[\operatorname{dom}(f + g)] \neq \emptyset$.

Then, point x_* solves $\operatorname{VIP}(\nabla f + \partial g, \operatorname{Fix} T)$ iff $x_* \in \arg\min_{x \in \operatorname{Fix} T}[f(x) + g(x)]$.

Proposition 2.15: *Given the closed affine set $\mathcal{A} \subset \mathcal{X}$, consider any $T \in \mathfrak{T}_{\mathcal{A}}$ (cf. Proposition 2.12). If U stands for the square root of the linear operator $\operatorname{Id} - Q$ in the description of T (cf. Definition 2.7), let $\overline{\operatorname{ran}}\, U$ denote the closure (in the strong topology) of the range of U. Then,*

$$\begin{aligned} &x_* \text{ solves } \operatorname{VIP}(\nabla f + \partial g, \operatorname{Fix} T) \\ &\quad \Leftrightarrow x_* \in \mathcal{A}_* := \{x \in \operatorname{Fix} T \mid \big[\nabla f(x) + \partial g(x)\big] \cap \overline{\operatorname{range}} U \neq \emptyset\}. \end{aligned} \tag{9a}$$

Moreover, for an arbitrarily fixed $\lambda \in \mathbb{R} \setminus \{0\}$, define the subset

$$\Upsilon_*^{(\lambda)} := \{(x, v) \in \operatorname{Fix} T \times \mathcal{X} \mid -\tfrac{1}{\lambda} U v \in \nabla f(x) + \partial g(x)\}. \tag{9b}$$

Then,

$$(x_*, v_*) \in \Upsilon_*^{(\lambda)} \Rightarrow x_* \text{ solves } \operatorname{VIP}(\nabla f + \partial g, \operatorname{Fix} T). \tag{9c}$$

Furthermore, in the case where $\mathcal{X}$ is finite dimensional,

$$\mathbf{x}_* \text{ solves } \operatorname{VIP}(\nabla f + \partial g, \operatorname{Fix} T) \Leftrightarrow \exists \mathbf{v}_* \in \mathcal{X} \text{ s.t. } (\mathbf{x}_*, \mathbf{v}_*) \in \Upsilon_*^{(\lambda)}. \tag{9d}$$

Proof: First, recall that $(\ker U)^{\perp} = \overline{\operatorname{ran}}\, U^* = \overline{\operatorname{ran}}\, U$ [3, Fact 2.18(iii), p. 32]. According to Definition 2.13,

$$\begin{aligned} &x_* \text{ solves } \operatorname{VIP}(\nabla f + \partial g, \operatorname{Fix} T) \\ &\Leftrightarrow \; x_* \in \operatorname{Fix} T \text{ and } \exists \xi_* \in \partial g(x_*) \text{ s.t. } \forall y \in \operatorname{Fix} T,\ \langle y - x_* | \nabla f(x_*) + \xi_* \rangle \geq 0 \\ &\Leftrightarrow \; x_* \in \operatorname{Fix} T \text{ and } \exists \xi_* \in \partial g(x_*) \text{ s.t. } \forall z \in \ker U,\ \langle z | \nabla f(x_*) + \xi_* \rangle \geq 0 \quad (10a) \\ &\Leftrightarrow \; x_* \in \operatorname{Fix} T \text{ and } \exists \xi_* \in \partial g(x_*) \text{ s.t. } \forall z \in \ker U,\ \langle z | \nabla f(x_*) + \xi_* \rangle \leq 0 \quad (10b) \end{aligned}$$

$$\Leftrightarrow\ x_* \in \operatorname{Fix} T \text{ and } \exists \xi_* \in \partial g(x_*) \text{ s.t. } \forall z \in \ker U,\ \langle z | \nabla f(x_*) + \xi_* \rangle = 0$$
$$\Leftrightarrow\ x_* \in \operatorname{Fix} T \text{ and } \exists \xi_* \in \partial g(x_*) \text{ s.t. } \nabla f(x_*) + \xi_* \in (\ker U)^{\perp} = \overline{\operatorname{ran}}\, U$$
$$\Leftrightarrow\ x_* \in \operatorname{Fix} T \text{ and } \left[\nabla f(x_*) + \partial g(x_*)\right] \cap \overline{\operatorname{ran}}\, U \neq \emptyset$$
$$\Leftrightarrow\ x_* \in \mathcal{A}_*, \tag{10c}$$

which establishes (9a). Notice that Proposition 2.12 is used in (10a) and $z \in \ker U \Leftrightarrow -z \in \ker U$ in (10b).

Moreover,

$$(x_*, v_*) \in \Upsilon_*^{(\lambda)}$$
$$\Leftrightarrow\ x_* \in \operatorname{Fix} T \text{ and } U\left(-\tfrac{v_*}{\lambda}\right) \in \nabla f(x_*) + \partial g(x_*)$$
$$\Leftrightarrow\ x_* \in \operatorname{Fix} T \text{ and } \exists v'_* \in \mathcal{X} \text{ s.t. } Uv'_* \in \left[\nabla f(x_*) + \partial g(x_*)\right] \cap \operatorname{ran} U \quad \left(v'_* = -\tfrac{v_*}{\lambda}\right)$$
$$\Leftrightarrow\ x_* \in \operatorname{Fix} T \text{ and } \left[\nabla f(x_*) + \partial g(x_*)\right] \cap \operatorname{ran} U \neq \emptyset$$
$$\Rightarrow\ x_* \in \operatorname{Fix} T \text{ and } \left[\nabla f(x_*) + \partial g(x_*)\right] \cap \overline{\operatorname{ran}}\, U \neq \emptyset \tag{11a}$$

$$\Leftrightarrow\ x_* \in \mathcal{A}_*,$$

which establishes (9c) via (9a).

In the case where $\mathcal{X}$ is Euclidean, (9d) is established by the well-known fact $\overline{\operatorname{ran}}\, U = \operatorname{ran} U$ [4, Thm. 2.4-3, p. 74], which turns '$\Rightarrow$' into '$\Longleftrightarrow$' in (11a). ■

3. Algorithm and convergence analysis

For any $T \in \mathfrak{T}_{\mathcal{A}}$ and any $\alpha \in (0, 1)$, define the α-averaged mapping

$$T_\alpha x := [\alpha T + (1 - \alpha)\mathrm{Id}\,]x = Q_\alpha x + \alpha\pi, \tag{12}$$

where $Q_\alpha := \alpha Q + (1 - \alpha)\mathrm{Id}$.

Theorem 3.1: *Consider $f, g \in \Gamma_0(\mathcal{X})$, with L being the Lipschitz-continuity constant of ∇f. Moreover, given the closed affine set $\mathcal{A}$, consider any $T \in \mathfrak{T}_{\mathcal{A}}$. For $\lambda \in \mathbb{R}_{>0}$, an arbitrarily fixed $x_0 \in \mathcal{X}$, and for all $n \in \mathbb{Z}_{\geq 0}$, the FM-HSDM is stated as follows:*

$$x_{1/2} := T_\alpha x_0 - \lambda \nabla f(x_0), \tag{13a}$$

$$x_1 := \mathrm{Prox}_{\lambda g}(x_{1/2}), \tag{13b}$$

$$x_{n+3/2} := x_{n+1/2} - \left[T_\alpha x_n - \lambda \nabla f(x_n)\right] + \left[T x_{n+1} - \lambda \nabla f(x_{n+1})\right], \tag{13c}$$

$$x_{n+2} := \mathrm{Prox}_{\lambda g}(x_{n+3/2})\,. \tag{13d}$$

Consider also $\alpha \in [0.5, 1)$ and $\lambda \in (0, 2(1 - \alpha)/L)$. Then, the following hold true.

(i) *There exist a sequence* $(v_n)_{n\in\mathbb{Z}_{\geq 0}} \subset \mathcal{X}$ *and a strongly positive operator* $\Theta : \mathcal{X}^2 \to \mathcal{X}^2$ *s.t. sequence* $(y_n := (x_n, v_n))_{n\in\mathbb{Z}_{>0}\setminus\{1\}}$ *is Fejér monotone [3, Def. 5.1, p. 75] w.r.t.* $\Upsilon_*^{(\lambda)}$ *of Proposition* 2.15 *in the Hilbert space* $(\mathcal{X}^2, \langle\cdot|\cdot\rangle_\Theta)$, *i.e.*

$$\|(x_{n+1}, v_{n+1}) - (x_*, v_*)\|_\Theta \leq \|(x_n, v_n) - (x_*, v_*)\|_\Theta, \quad \forall (x_*, v_*) \in \Upsilon_*^{(\lambda)} .$$

(ii) *Sequence* $(x_n)_{n\in\mathbb{Z}_{\geq 0}}$ *of* (13) *converges weakly to a point that solves* $\mathrm{VIP}(\nabla f + \partial g, \mathrm{Fix}\, T)$.

Proof: **(i)** By (13c),

$$x_{n+3/2} - x_{n+1/2} = Tx_{n+1} - T_\alpha x_n - \lambda \left[\nabla f(x_{n+1}) - \nabla f(x_n)\right] . \tag{14}$$

Since $z = \mathrm{Prox}_{\lambda g}(y) \Leftrightarrow (\exists \xi \in \partial g(z) \text{ s.t. } z + \lambda\xi = y)$, then

$$\exists \xi_{n+2} \in \partial g(x_{n+2}) \tag{15}$$

s.t. $x_{n+3/2} = x_{n+2} + \lambda\xi_{n+2}$ and thus $\exists \xi_{n+1} \in \partial g(x_{n+1})$ s.t. $x_{n+1/2} = x_{n+1} + \lambda\xi_{n+1}$. Incorporating the previous equations in (14) yields that $\forall n \in \mathbb{Z}_{\geq 0}$,

$$\begin{aligned} x_1 &= T_\alpha x_0 - \lambda \left[\nabla f(x_0) + \xi_1\right], \\ x_{n+2} - x_{n+1} &= Tx_{n+1} - T_\alpha x_n - \lambda \left[\nabla f(x_{n+1}) + \xi_{n+2}\right] + \lambda \left[\nabla f(x_n) + \xi_{n+1}\right] . \end{aligned} \tag{16}$$

Moreover, adding consecutive equations of (16) results into the following fact:

$$\begin{aligned} x_{n+1} &= Tx_n - \sum_{\nu=1}^{n-1} (T_\alpha - T)x_\nu - \lambda \left[\nabla f(x_n) + \xi_{n+1}\right] \\ &= Tx_n - \sum_{\nu=1}^{n+1} (T_\alpha - T)x_\nu + (T_\alpha - T)x_n + (T_\alpha - T)x_{n+1} \\ &\quad - \lambda \left[\nabla f(x_n) + \xi_{n+1}\right] \\ &= 2T_\alpha x_{n+1} - Tx_{n+1} + (T_\alpha x_n - T_\alpha x_{n+1}) - \sum_{\nu=1}^{n+1} (T_\alpha - T)x_\nu \\ &\quad - \lambda \left[\nabla f(x_n) + \xi_{n+1}\right], \end{aligned}$$

where the last equality holds true $\forall n \in \mathbb{Z}_{\geq 0}$. Consequently,

$$\begin{aligned} &(\mathrm{Id} + T - 2T_\alpha)x_{n+1} + (T_\alpha x_{n+1} - T_\alpha x_n) \\ &\quad = (1 - 2\alpha)(T - \mathrm{Id}\,)x_{n+1} + Q_\alpha(x_{n+1} - x_n) \\ &\quad = -\sum_{\nu=1}^{n+1} (T_\alpha - T)x_\nu - \lambda \left[\nabla f(x_n) + \xi_{n+1}\right], \end{aligned} \tag{17}$$

where the first equation is due to (12).

Choose arbitrarily a $w_* \in \operatorname{Fix} T$, i.e. $(\mathrm{Id} - T)w_* = 0$. Then,

$$\begin{aligned}(T_\alpha - T)x_\nu &= (1-\alpha)(\mathrm{Id} - T)x_\nu \\ &= (1-\alpha)\left[(\mathrm{Id} - T)x_\nu - (\mathrm{Id} - T)w_*\right] \\ &= (1-\alpha)(\mathrm{Id} - Q)(x_\nu - w_*)\,.\end{aligned}$$

Define also

$$v_{n+1} := (1-\alpha)\sum_{\nu=1}^{n+1} U(x_\nu - w_*)\,.$$

Point v_{n+1} does not depend on the choice of the fixed point w_*. Indeed, by Proposition 2.12, it can be verified that for any $w_\# \in \operatorname{Fix} T$, $w_\# - w_* \in \ker U$, and that

$$\begin{aligned}v_{n+1} &= (1-\alpha)\sum_{\nu=1}^{n+1} U(x_\nu - w_\# + w_\# - w_*) \\ &= (1-\alpha)\sum_{\nu=1}^{n+1} \left[U(x_\nu - w_\#) + U(w_\# - w_*)\right] \\ &= (1-\alpha)\sum_{\nu=1}^{n+1} U(x_\nu - w_\#)\,. \end{aligned} \tag{18}$$

Moreover,

$$\begin{aligned}v_{n+1} - v_n &= (1-\alpha)\sum_{\nu=1}^{n+1} U(x_\nu - w_*) - (1-\alpha)\sum_{\nu=1}^{n} U(x_\nu - w_*) \\ &= (1-\alpha)U(x_{n+1} - w_*), \quad \forall w_* \in \operatorname{Fix} T,\end{aligned} \tag{19}$$

and

$$\begin{aligned}-\sum_{\nu=1}^{n+1}(T_\alpha - T)x_\nu &= -(1-\alpha)\sum_{\nu=1}^{n+1}(\mathrm{Id} - Q)(x_\nu - w_*) \\ &= -U(1-\alpha)\sum_{\nu=1}^{n+1} U(x_\nu - w_*) \\ &= -Uv_{n+1}\,.\end{aligned} \tag{20}$$

Under the previous considerations, (17) becomes

$$(1-2\alpha)(T - \mathrm{Id}\,)x_{n+1} + Q_\alpha(x_{n+1} - x_n) + Uv_{n+1} = -\lambda\left[\nabla f(x_n) + \xi_{n+1}\right]. \tag{21}$$

Recall now Proposition 2.15, and consider *any* $(x_*, v_*) \in \Upsilon_*^{(\lambda)}$. By the definition of $\Upsilon_*^{(\lambda)}$, $(\mathrm{Id} \; - T)x_* = 0$ and there exists $\xi_* \in \partial g(x_*)$ s.t. $Uv_* +$

$\lambda[\nabla f(x_*) + \xi_*] = 0$. These arguments, (21) and $(T - \mathrm{Id})x_{n+1} - (T - \mathrm{Id})x_* = (Q - \mathrm{Id})(x_{n+1} - x_*)$ yield

$$\begin{aligned} &\lambda[\nabla f(x_n) - \nabla f(x_*)] + \lambda(\xi_{n+1} - \xi_*) \\ &\quad = -(1-2\alpha)(Q - \mathrm{Id})(x_{n+1} - x_*) - Q_\alpha(x_{n+1} - x_n) - U(v_{n+1} - v_*)\,. \end{aligned} \tag{22}$$

The Baillon–Haddad theorem [29], [3, Cor. 18.16, p. 270] states that the L-Lipschitz continuous ∇f is $(1/L)$-inverse strongly monotone, i.e. $\forall (x, x') \in \mathcal{X}^2$, $\langle x - x' | \nabla f(x) - \nabla f(x')\rangle \geq (1/L)\|\nabla f(x) - \nabla f(x')\|^2$. This property, the fact that ∂g is monotone [3, Example 20.3, p. 294], i.e. $\forall x, x', \xi, \xi'$ s.t. $\xi \in \partial g(x)$ and $\xi' \in \partial g(x')$, $\langle x - x' | \xi - \xi'\rangle \geq 0$, and the fact that U is self-adjoint imply

$$\begin{aligned} &\tfrac{2\lambda}{L}\|\nabla f(x_n) - \nabla f(x_*)\|^2 \\ &\quad \leq 2\lambda\langle x_n - x_* | \nabla f(x_n) - \nabla f(x_*)\rangle \\ &\quad \leq 2\lambda\langle x_{n+1} - x_* | \nabla f(x_n) - \nabla f(x_*)\rangle + 2\lambda\langle x_n - x_{n+1} | \nabla f(x_n) - \nabla f(x_*)\rangle \\ &\qquad + 2\lambda\langle x_{n+1} - x_* | \xi_{n+1} - \xi_*\rangle \\ &\quad = 2\langle x_{n+1} - x_* | \lambda[\nabla f(x_n) - \nabla f(x_*)] + \lambda(\xi_{n+1} - \xi_*)\rangle \\ &\qquad + 2\lambda\langle x_n - x_{n+1} | \nabla f(x_n) - \nabla f(x_*)\rangle \\ &\quad = -2(1-2\alpha)\langle x_{n+1} - x_* | (Q - \mathrm{Id})(x_{n+1} - x_*)\rangle \\ &\qquad - 2\langle x_{n+1} - x_* | Q_\alpha(x_{n+1} - x_n)\rangle \\ &\qquad - 2\langle x_{n+1} - x_* | U(v_{n+1} - v_*)\rangle + 2\lambda\langle x_n - x_{n+1} | \nabla f(x_n) - \nabla f(x_*)\rangle \qquad \text{(23a)} \\ &\qquad = -2(1-2\alpha)\langle x_{n+1} - x_* | (Q - \mathrm{Id})(x_{n+1} - x_*)\rangle \\ &\qquad\quad - 2\langle x_{n+1} - x_* | Q_\alpha(x_{n+1} - x_n)\rangle \\ &\qquad\quad - 2\langle U(x_{n+1} - x_*) | v_{n+1} - v_*\rangle + 2\lambda\langle x_n - x_{n+1} | \nabla f(x_n) - \nabla f(x_*)\rangle \\ &\qquad \leq -2(1-2\alpha)\langle x_{n+1} - x_* | (Q - \mathrm{Id})(x_{n+1} - x_*)\rangle \\ &\qquad\quad - 2\langle x_{n+1} - x_* | Q_\alpha(x_{n+1} - x_n)\rangle - \tfrac{2}{1-\alpha}\langle v_{n+1} - v_n | v_{n+1} - v_*\rangle \\ &\qquad\quad + \tfrac{\lambda L}{2}\|x_n - x_{n+1}\|^2 + \tfrac{2\lambda}{L}\|\nabla f(x_n) - \nabla f(x_*)\|^2, \qquad \text{(23b)} \end{aligned}$$

where (22) is used in (23a), and (19) as well as

$$2\langle \tfrac{a}{\sqrt{\eta}} | \sqrt{\eta}\, b\rangle_\Pi \leq \tfrac{1}{\eta}\|a\|_\Pi^2 + \eta\|b\|_\Pi^2, \quad \begin{cases} \forall (a, b) \in \mathcal{X}^2, \forall \eta \in \mathbb{R}_{>0}, \\ \forall \text{ strongly positive } \Pi \in \mathfrak{B}(\mathcal{X}), \end{cases} \tag{24}$$

with $\eta := 2/L$, $a := x_n - x_{n+1}$, $b := \nabla f(x_n) - \nabla f(x_*)$, and $\Pi := \mathrm{Id}$, were used in (23b).

Recall (12) to verify that the positivity of Q implies that for any $x \in \mathcal{X}$,

$$\langle Q_\alpha x | x\rangle = \alpha\langle Qx | x\rangle + (1-\alpha)\|x\|^2 \geq (1-\alpha)\|x\|^2, \tag{25}$$

i.e. Q_α is strongly positive. Hence, upon defining the linear mapping $\Theta : \mathcal{X}^2 \to \mathcal{X}^2 : (x, v) \mapsto (Q_\alpha x, v/(1-\alpha))$, it can be easily seen that Θ is strongly positive, under the standard inner product $\langle (x,v) | (x',v') \rangle := \langle x|x'\rangle + \langle v|v'\rangle$, $\forall (x,v),(x',v') \in \mathcal{X}^2$, due to the fact that both Q_α and $\mathrm{Id}/(1-\alpha)$ are strongly positive. Consequently, $(\mathcal{X}^2, \langle\cdot|\cdot\rangle_\Theta)$ can be considered to be a Hilbert space equipped with the inner product $\langle\cdot|\cdot\rangle_\Theta$.

Notation $y := (x,v)$, $\alpha \geq 1/2$ as well as the positivity of $\mathrm{Id} - Q$ in (23) yield

$$\begin{aligned}
0 \leq\ & 2\langle (x_{n+1} - x_n, v_{n+1} - v_n) | \Theta(x_* - x_{n+1}, v_* - v_{n+1})\rangle \\
& - 2(2\alpha - 1)\langle x_{n+1} - x_* | (\mathrm{Id} - Q)(x_{n+1} - x_*)\rangle \\
& + \tfrac{\lambda L}{2}\|x_n - x_{n+1}\|^2 \\
=\ & 2\langle y_{n+1} - y_n | \Theta(y_* - y_{n+1})\rangle - 2(2\alpha - 1)\langle x_{n+1} - x_* | (\mathrm{Id} - Q)(x_{n+1} - x_*)\rangle \\
& + \tfrac{\lambda L}{2}\|x_n - x_{n+1}\|^2 \\
\leq\ & 2\langle y_{n+1} - y_n | y_* - y_{n+1}\rangle_\Theta + \tfrac{\lambda L}{2}\|x_n - x_{n+1}\|^2 \\
=\ & \|y_n - y_*\|_\Theta^2 - \|y_{n+1} - y_*\|_\Theta^2 - \|y_{n+1} - y_n\|_\Theta^2 + \tfrac{\lambda L}{2}\|x_n - x_{n+1}\|^2 .
\end{aligned}$$

Hence,

$$\|y_n - y_*\|_\Theta^2 - \|y_{n+1} - y_*\|_\Theta^2 \geq \|y_{n+1} - y_n\|_\Theta^2 - \tfrac{\lambda L}{2}\|x_n - x_{n+1}\|^2 . \tag{26}$$

Since $\lambda < 2(1-\alpha)/L$, choose any $\zeta \in (\lambda L/[2(1-\alpha)], 1)$. Then, by (25), $\forall y := (x,v)$,

$$\tfrac{\lambda L}{2}\|x\|^2 < \zeta(1-\alpha)\|x\|^2 \leq \zeta\langle x|Q_\alpha x\rangle \leq \zeta\langle x|Q_\alpha x\rangle + \zeta\tfrac{1}{1-\alpha}\|v\|^2 = \zeta\|y\|_\Theta^2,$$

and by (26),

$$\begin{aligned}
\|y_n - y_*\|_\Theta^2 - \|y_{n+1} - y_*\|_\Theta^2 &\geq \|y_{n+1} - y_n\|_\Theta^2 - \tfrac{\lambda L}{2}\|x_n - x_{n+1}\|^2 \\
&\geq \|y_{n+1} - y_n\|_\Theta^2 - \zeta\|y_{n+1} - y_n\|_\Theta^2 \\
&= (1-\zeta)\|y_{n+1} - y_n\|_\Theta^2,
\end{aligned} \tag{27}$$

i.e. sequence $(y_n)_{n\in\mathbb{Z}_{\geq 0}} \subset (\mathcal{X}^2, \langle\cdot|\cdot\rangle_\Theta)$ is Fejér monotone w.r.t. $\Upsilon_*^{(\lambda)}$ of Proposition 2.15.

(ii) Due to Fejér monotonicity, sequence $(y_n)_n$ is bounded [as well as $(x_n)_n$ and $(v_n)_n$] [3, Prop. 5.4(i), p. 76] and possesses a non-empty set of weakly sequential cluster points $\mathfrak{W}[(y_n)_n]$ [3, Lem. 2.37, p. 36]. Moreover, it can be verified by (27) that $\forall n \in \mathbb{Z}_{\geq 0}$,

$$(1-\zeta)\sum_{\nu=2}^{n}\|y_{\nu+1} - y_\nu\|_\Theta^2 \leq \|y_2 - y_*\|_\Theta^2 - \|y_{n+1} - y_*\|_\Theta^2 \leq \|y_2 - y_*\|_\Theta^2,$$

and hence there exist $C', C \in \mathbb{R}_{>0}$ s.t. for any n,

$$\sum_{\nu=0}^{n}\|y_{\nu+1} - y_\nu\|_\Theta^2 \leq \tfrac{C'}{1-\zeta} =: C, \tag{28}$$

which leads to $\lim_{n\to\infty}\|y_{n+1}-y_n\|_\Theta = 0$, and which further implies that

$$\lim_{n\to\infty}(x_{n+1}-x_n) = 0, \quad \lim_{n\to\infty}(v_{n+1}-v_n) = 0\,. \tag{29}$$

Adding the following equations, which result from (21),

$$\begin{aligned} -\tfrac{1}{\lambda}(1-2\alpha)(T-\mathrm{Id}\,)x_{n+1} - \tfrac{1}{\lambda}Q_\alpha(x_{n+1}-x_n) - \tfrac{1}{\lambda}Uv_{n+1} - \nabla f(x_n) &= \xi_{n+1} \\ \tfrac{1}{\lambda}(1-2\alpha)(T-\mathrm{Id}\,)x_n + \tfrac{1}{\lambda}Q_\alpha(x_n-x_{n-1}) + \tfrac{1}{\lambda}Uv_n + \nabla f(x_{n-1}) &= -\xi_n \end{aligned} \tag{30}$$

yields

$$\begin{aligned} \xi_{n+1}-\xi_n = \tfrac{1-2\alpha}{\lambda}(T-\mathrm{Id}\,)(x_n-x_{n+1}) + \tfrac{1}{\lambda}Q_\alpha(x_n-x_{n-1}) - \tfrac{1}{\lambda}Q_\alpha(x_{n+1}-x_n) \\ + \tfrac{1}{\lambda}U(v_n-v_{n+1}) + [\nabla f(x_{n-1}) - \nabla f(x_n)]\,. \end{aligned} \tag{31}$$

By applying $\lim_{n\to\infty}$ to the previous equality, and by using the Lipschitz continuity of ∇f, i.e. $\|\nabla f(x_n) - \nabla f(x_{n-1})\| \leq L\|x_n - x_{n-1}\|$, (29), as well as the continuity of $\mathrm{Id}\, - T$, Q_α and U, it can be verified that

$$\lim_{n\to\infty}(\xi_{n+1}-\xi_n) = 0\,. \tag{32}$$

Now, by (16),

$$\begin{aligned} &x_{n+2}-x_{n+1} \\ &\quad = Tx_{n+1} - T_\alpha x_{n+1} + T_\alpha x_{n+1} - T_\alpha x_n - \lambda[\nabla f(x_{n+1}) \\ &\qquad - \nabla f(x_n)] - \lambda[\xi_{n+2}-\xi_{n+1}] \\ &\quad = (T-T_\alpha)x_{n+1} + Q_\alpha(x_{n+1}-x_n) - \lambda[\nabla f(x_{n+1}) \\ &\qquad - \nabla f(x_n)] - \lambda[\xi_{n+2}-\xi_{n+1}], \end{aligned}$$

which leads to

$$\begin{aligned} (1-\alpha)(\mathrm{Id}-T)x_n = (x_n - x_{n+1}) + Q_\alpha(x_n-x_{n-1}) \\ - \lambda[\nabla f(x_n) - \nabla f(x_{n-1})] - \lambda[\xi_{n+1}-\xi_n]\,. \end{aligned} \tag{33}$$

Choose any $\bar{y} := (\bar{x}, \bar{v}) \in \mathfrak{W}[(y_n)_{n\in\mathbb{Z}_{\geq 0}}] \neq \emptyset$, i.e. there exists a subsequence $(y_{n_k} := (x_{n_k}, v_{n_k}))_k$ s.t. $x_{n_k} \rightharpoonup_{k\to\infty} \bar{x}$ and $v_{n_k} \rightharpoonup_{k\to\infty} \bar{v}$. Furthermore, by (29), (32), (33), and the Lipschitz continuity of ∇f,

$$\begin{aligned}
\limsup_{n\to\infty}\|(\mathrm{Id}-T)x_n\| &\le \tfrac{1}{1-\alpha}\lim_{k\to\infty}\|x_n-x_{n+1}\| + \lim_{k\to\infty}\tfrac{1}{1-\alpha}\|Q_\alpha(x_n-x_{n-1})\| \\
&\quad + \tfrac{\lambda}{1-\alpha}\lim_{k\to\infty}\|\nabla f(x_n)-\nabla f(x_{n-1})\| \\
&\quad + \tfrac{\lambda}{1-\alpha}\lim_{k\to\infty}\|\xi_{n+1}-\xi_n\| \\
&\le \tfrac{1}{1-\alpha}\lim_{k\to\infty}\|x_n-x_{n+1}\| + \lim_{k\to\infty}\tfrac{\|Q_\alpha\|}{1-\alpha}\|x_n-x_{n-1}\| \\
&\quad + \tfrac{\lambda L}{1-\alpha}\lim_{k\to\infty}\|x_n-x_{n-1}\| + \tfrac{\lambda}{1-\alpha}\lim_{k\to\infty}\|\xi_{n+1}-\xi_n\| \\
&= 0\,. \qquad (34)
\end{aligned}$$

Hence, due to $x_{n_k} \rightharpoonup_{k\to\infty} \bar{x}$, $\lim_{k\to\infty}(\mathrm{Id} - T)x_{n_k} = 0$, and the demiclosedness property of the nonexpansive mapping T [3, Thm. 4.17, p. 63], it follows that

$$\bar{x} \in \operatorname{Fix} T\,. \qquad (35)$$

Fix arbitrarily an $x_\# \in \mathcal{X}$. Since $(x_n)_n$ is bounded, there exist $C'', C_{\nabla f} \in \mathbb{R}_{>0}$ s.t. for any n,

$$\begin{aligned}
\|\nabla f(x_n)\| &\le \|\nabla f(x_n)-\nabla f(x_\#)\| + \|\nabla f(x_\#)\| \\
&\le L\|x_n - x_\#\| + \|\nabla f(x_\#)\| \\
&\le L(\|x_n\| + \|x_\#\|) + \|\nabla f(x_\#)\| \\
&\le L(C'' + \|x_\#\|) + \|\nabla f(x_\#)\| \le C_{\nabla f}\,. \qquad (36)
\end{aligned}$$

Now, according to the Baillon–Haddad theorem [29], [3, Cor. 18.16, p. 270],

$$\begin{aligned}
&\tfrac{2\lambda}{L}\|\nabla f(x_{n_k})-\nabla f(\bar{x})\|^2 \\
&\quad \le 2\lambda\langle x_{n_k}-\bar{x} | \nabla f(x_{n_k})-\nabla f(\bar{x})\rangle \\
&\quad = 2\lambda\langle x_{n_k+1}-\bar{x} | \nabla f(x_{n_k})\rangle \\
&\qquad - 2\lambda\langle x_{n_k+1}-\bar{x} | \nabla f(\bar{x})\rangle + 2\lambda\langle x_{n_k}-x_{n_k+1} | \nabla f(x_{n_k})-\nabla f(\bar{x})\rangle \\
&\quad = -2\lambda\langle x_{n_k+1}-\bar{x} | \xi_{n_k+1}\rangle - 2\langle x_{n_k+1}-\bar{x} | Uv_{n_k+1}\rangle \\
&\qquad - 2\langle x_{n_k+1}-\bar{x} | Q_\alpha(x_{n_k+1}-x_{n_k})\rangle - (1-2\alpha)\langle x_{n_k+1}-\bar{x} | (T-\mathrm{Id}\,)x_{n_k+1}\rangle \\
&\qquad - 2\lambda\langle x_{n_k+1}-\bar{x} | \nabla f(\bar{x})\rangle + 2\lambda\langle x_{n_k}-x_{n_k+1} | \nabla f(x_{n_k})-\nabla f(\bar{x})\rangle \qquad (37a) \\
&\quad \le 2\lambda\left[g(\bar{x})-g(x_{n_k+1})\right] - 2\langle U(x_{n_k+1}-\bar{x}) | v_{n_k+1}\rangle \\
&\qquad - 2\langle x_{n_k+1}-\bar{x} | Q_\alpha(x_{n_k+1}-x_{n_k})\rangle - (1-2\alpha)\langle x_{n_k+1}-\bar{x} | (T-\mathrm{Id}\,)x_{n_k+1}\rangle \\
&\qquad - 2\lambda\langle x_{n_k+1}-\bar{x} | \nabla f(\bar{x})\rangle + 2\lambda\langle x_{n_k}-x_{n_k+1} | \nabla f(x_{n_k})-\nabla f(\bar{x})\rangle \qquad (37b) \\
&\quad \le 2\lambda\left[g(\bar{x})-g(x_{n_k+1})\right] - \tfrac{2}{1-\alpha}\langle v_{n_k+1}-v_{n_k} | v_{n_k+1}\rangle \\
&\qquad - 2\langle x_{n_k+1}-\bar{x} | Q_\alpha(x_{n_k+1}-x_{n_k})\rangle - (1-2\alpha)\langle x_{n_k+1}-\bar{x} | (T-\mathrm{Id}\,)x_{n_k+1}\rangle \\
&\qquad - 2\lambda\langle x_{n_k+1}-\bar{x} | \nabla f(\bar{x})\rangle + 2\lambda\left(C_{\nabla f} + \|\nabla f(\bar{x})\|\right)\|x_{n_k}-x_{n_k+1}\|, \qquad (37c)
\end{aligned}$$

where (21) is used in (47a), the convexity of g, (15) and the self-adjointness of U in (37b), and finally (19) and (36) in (37c). Since $\lim_{k\to\infty}(x_{n_k} - x_{n_k+1}) = 0$ by (29), the continuity of Q_α implies $\lim_{k\to\infty} Q_\alpha(x_{n_k+1} - x_{n_k}) = 0$, and (34) yields $\lim_{k\to\infty}(T - \mathrm{Id}\ )x_{n_k+1} = 0$. Notice again by (29) that $\lim_{k\to\infty}(v_{n_k+1} - v_{n_k}) = 0$. Furthermore, (29), together with $(x_{n_k} - \bar{x}) \rightharpoonup_{k\to\infty} 0$, yields $(x_{n_k+1} - \bar{x}) \rightharpoonup_{k\to\infty} 0$. Similarly, $(v_{n_k+1} - \bar{v}) \rightharpoonup_{k\to\infty} 0$ can be deduced from (29) and $(v_{n_k} - \bar{v}) \rightharpoonup_{k\to\infty} 0$. Due to [3, Lem. 2.41(iii), p. 37], all of the previous arguments result in $\lim_{k\to\infty}\langle v_{n_k+1} - v_{n_k} | v_{n_k+1}\rangle = 0$, $\lim_{k\to\infty}\langle x_{n_k+1} - \bar{x} | Q_\alpha(x_{n_k+1} - x_{n_k})\rangle = 0$, $\lim_{k\to\infty}\langle x_{n_k+1} - \bar{x} | (T - \mathrm{Id}\ )x_{n_k+1}\rangle = 0$, $\lim_{k\to\infty} \langle x_{n_k+1} - \bar{x} | \nabla f(\bar{x})\rangle = 0$ and $\lim_{k\to\infty}\|x_{n_k} - x_{n_k+1}\| = 0$. Hence, the application of $\limsup_{k\to\infty}$ onto both sides of (37c) yields

$$\limsup_{k\to\infty} \|\nabla f(x_{n_k}) - \nabla f(\bar{x})\|^2 \leq \limsup_{k\to\infty} L\left[g(\bar{x}) - g(x_{n_k+1})\right] = L\left[g(\bar{x}) - \liminf_{k\to\infty} g(x_{n_k+1})\right] \leq 0,$$

where the last inequality is deduced from the fact that $g \in \Gamma_0(\mathcal{X})$ turns out to be also weakly sequentially lower semicontinuous [3, Thm. 9.1, p. 129]. In other words,

$$\lim_{k\to\infty} \nabla f(x_{n_k}) = \nabla f(\bar{x}) \,. \tag{38}$$

Since $v_{n_k+1} \rightharpoonup_k \bar{v}$, i.e. $\forall z \in \mathcal{X}$, $\lim_{k\to\infty}\langle z | v_{n_k+1}\rangle = \langle z | \bar{v}\rangle$, it can be easily seen that $\forall z \in \mathcal{X}$, $\lim_{k\to\infty}\langle z | U v_{n_k+1}\rangle = \lim_{k\to\infty}\langle Uz | v_{n_k+1}\rangle = \langle Uz | \bar{v}\rangle = \langle z | U\bar{v}\rangle$, i.e. $Uv_{n_k+1} \rightharpoonup_k U\bar{v}$. Hence, having this result and (38) plugged into (30) yields that

$$\xi_{n_k+1} \rightharpoonup_{k\to\infty} \bar{\xi} := -\tfrac{1}{\lambda} U\bar{v} - \nabla f(\bar{x}) \,. \tag{39}$$

Using (21) once again,

$$\begin{aligned}
\langle x_{n_k+1} - \bar{x} | \xi_{n_k+1}\rangle &= -\langle x_{n_k+1} - \bar{x} | \nabla f(x_{n_k})\rangle - \tfrac{1}{\lambda}\langle x_{n_k+1} - \bar{x} | U v_{n_k+1}\rangle \\
&\quad - \tfrac{1}{\lambda}\langle x_{n_k+1} - \bar{x} | Q_\alpha(x_{n_k+1} - x_{n_k})\rangle \\
&\quad - \tfrac{1}{\lambda}(1 - 2\alpha)\langle x_{n_k+1} - \bar{x} | (T - \mathrm{Id}\ )x_{n_k+1}\rangle \\
&= -\langle x_{n_k+1} - \bar{x} | \nabla f(x_{n_k})\rangle - \tfrac{1}{\lambda}\langle U(x_{n_k+1} - \bar{x}) | v_{n_k+1}\rangle \\
&\quad - \tfrac{1}{\lambda}\langle x_{n_k+1} - \bar{x} | Q_\alpha(x_{n_k+1} - x_{n_k})\rangle \\
&\quad - \tfrac{1}{\lambda}(1 - 2\alpha)\langle x_{n_k+1} - \bar{x} | (T - \mathrm{Id}\ )x_{n_k+1}\rangle \\
&= -\langle x_{n_k+1} - \bar{x} | \nabla f(x_{n_k})\rangle - \tfrac{1}{\lambda(1-\alpha)}\langle v_{n_k+1} - v_{n_k} | v_{n_k+1}\rangle \\
&\quad - \tfrac{1}{\lambda}\langle x_{n_k+1} - \bar{x} | Q_\alpha(x_{n_k+1} - x_{n_k})\rangle \\
&\quad - \tfrac{1}{\lambda}(1 - 2\alpha)\langle x_{n_k+1} - \bar{x} | (T - \mathrm{Id}\ )x_{n_k+1}\rangle ,
\end{aligned} \tag{40}$$

where (19) is used in (40). Since $(x_{n_k+1} - \bar{x}) \rightharpoonup_k 0$ and $v_{n_k+1} \rightharpoonup_k \bar{v}$, and due to (29), (34) and (38), as well as the continuity of the linear mapping Q_α, it turns out by [3, Lem. 2.41(iii), p. 37] and (40) that $\lim_{k\to\infty}\langle x_{n_k+1} - \bar{x} | \xi_{n_k+1}\rangle = 0$. In other words,

$$\begin{aligned}\lim_{k\to\infty}\langle x_{n_k+1}|\xi_{n_k+1}\rangle &= \lim_{k\to\infty}\big(\langle x_{n_k+1} - \bar{x}|\xi_{n_k+1}\rangle + \langle \bar{x}|\xi_{n_k+1}\rangle\big)\\ &= \lim_{k\to\infty}\langle x_{n_k+1} - \bar{x}|\xi_{n_k+1}\rangle + \lim_{k\to\infty}\langle \bar{x}|\xi_{n_k+1}\rangle\\ &= \lim_{k\to\infty}\langle \bar{x}|\xi_{n_k+1}\rangle = \langle \bar{x}|\bar{\xi}\rangle\,. \end{aligned} \tag{41}$$

Now, by $(x_{n_k+1}, \xi_{n_k+1}) \in \operatorname{gra}\partial g$, the maximal monotonicity of ∂g [3, Thm. 20.40, p. 304] and the property manifested in (41), [3, Cor. 20.49(ii), p. 306] suggests that $(\bar{x}, \bar{\xi}) \in \operatorname{gra}\partial g \Leftrightarrow \bar{\xi} \in \partial g(\bar{x})$. Hence, according also to (39), $-U(\bar{v}/\lambda) \in \nabla f(\bar{x}) + \partial g(\bar{x})$, which together with (35) imply $(\bar{x}, \bar{v}) \in \Upsilon_*^{(\lambda)}$. Since $(\bar{x}, \bar{v})$ was arbitrarily chosen within $\mathfrak{W}[(y_n)_n]$, it follows that $\mathfrak{W}[(y_n)_n] \subset \Upsilon_*^{(\lambda)}$. Adding also to that the Fejér monotonicity property (27) of $(y_n)_{n\in\mathbb{Z}_{\geq 0}}$ w.r.t. $\Upsilon_*^{(\lambda)}$ yields that $(y_n)_n$ converges weakly to a point in $\Upsilon_*^{(\lambda)}$ [3, Thm. 5.5, p. 76]. According to (9c), the weak limit of $(x_n)_n$ solves $\operatorname{VIP}(\nabla f + \partial g, \operatorname{Fix} T)$. ■

Definition 3.2 ([3, (10.2), p. 144]): A proper convex function $h : \mathcal{X} \to (-\infty, +\infty]$ is called *uniformly convex* on a non-empty subset $\mathcal{S}$ of $\operatorname{dom} h$, if there exists an increasing function $\varphi_{\mathcal{S}} : [0, +\infty] \to [0, +\infty]$, which vanishes only at 0, s.t. $\forall x, x' \in \mathcal{S}$ and $\forall \mu \in (0, 1)$,

$$h(\mu x + (1 - \mu)x') + \mu(1 - \mu)\varphi_{\mathcal{S}}(\|x - x'\|) \leq \mu h(x) + (1 - \mu)h(x')\,.$$

In the case where $\mathcal{S} := \operatorname{dom} h$ and $\varphi_{\mathcal{S}} := (\beta_{\mathcal{S}}/2)(\cdot)^2$, for some $\beta_{\mathcal{S}} \in \mathbb{R}_{>0}$, then h is called *strongly convex* with constant $\beta_{\mathcal{S}}$. Moreover, 'strong convexity' $\Rightarrow$ 'uniform convexity' $\Rightarrow$ 'strict convexity'.

Assumption 3.3: (i) Function f is uniformly convex on every non-empty bounded subset of $\mathcal{X}$.
(ii) Function g is uniformly convex on every non-empty bounded subset of $\operatorname{dom}\partial g$.

Lemma 3.4: *In addition to the setting of Theorem* 3.1, *if either Assumption* 3.3*(i) or Assumption* 3.3*(ii) holds true, then sequence* $(x_n)_{n\in\mathbb{Z}_{\geq 0}}$ *of* (13) *converges strongly to a point that solves* $\operatorname{VIP}(\nabla f + \partial g, \operatorname{Fix} T)$.

Proof: As part **(ii)** of the proof of Theorem 3.1 has demonstrated, sequences $(x_n)_n$ and $(Uv_n)_n$ converge weakly to $\bar{x}$ and $U\bar{v}$, respectively. Consequently, (29), the continuity of Q_α, (30), (34), (38) and (39) suggest that $(\xi_n)_n$ converges weakly to $\bar{\xi}$.

Let Assumption 3.3(i) holds true. Then, according to [3, Ex. 22.3(iii), p. 324], given a bounded set $\mathcal{B} \subset \mathcal{X}$, there exists an increasing function $\varphi_{\mathcal{B}} : [0, +\infty) \to [0, +\infty]$, which vanishes only at 0, s.t. $\forall x, x' \in \mathcal{B}$,

$$\langle x - x' | \nabla f(x) - \nabla f(x') \rangle \geq 2\varphi_{\mathcal{B}} \left(\|x - x'\| \right) . \tag{42}$$

Define $\mathcal{B} := (x_n)_n \cup \{\bar{x}\}$ (recall that $(x_n)_n$ is bounded). Set $x := x_n$ and $x' := \bar{x}$ in (42) to obtain

$$\langle x_n - \bar{x} | \nabla f(x_n) - \nabla f(\bar{x}) \rangle \geq 2\varphi_{\mathcal{B}} \left(\|x_n - \bar{x}\| \right) , \quad \forall n . \tag{43}$$

Since $x_n \rightharpoonup_{n\to\infty} \bar{x}$ and $\lim_{n\to\infty} \nabla f(x_n) = \nabla f(\bar{x})$ by (38), the application of $\lim_{n\to\infty}$ to (43) and [3, Lem. 2.41(iii), p. 37] suggest that $\lim_{n\to\infty} \varphi_{\mathcal{B}}(\|x_n - \bar{x}\|) = 0$, and thus $\lim_{n\to\infty} \|x_n - \bar{x}\| = 0$, due to the properties of $\varphi_{\mathcal{B}}$.

Let now Assumption 3.3(ii) holds true. Then, according to [3, Ex. 22.3(iii), p. 324], given a bounded set $\mathcal{B} \subset \operatorname{dom} \partial g$, there exists an increasing function $\varphi_{\mathcal{B}} : [0, +\infty) \to [0, +\infty]$, which vanishes only at 0, s.t. $\forall x, x' \in \mathcal{B}$, and $\forall \xi \in \partial g(x)$, $\forall \xi' \in \partial g(x')$,

$$\langle x - x' | \xi - \xi' \rangle \geq 2\varphi_{\mathcal{B}} \left(\|x - x'\| \right) . \tag{44}$$

According to (15), $x_n \in \operatorname{dom} \partial g$, $\forall n$. Moreover, as the discussion after (41) demonstrated, $\bar{x} \in \operatorname{dom} \partial g$. Define thus the bounded set $\mathcal{B} := (x_n)_n \cup \{\bar{x}\} \subset \operatorname{dom} \partial g$, and set $x := x_n$, $x' := \bar{x}$, $\xi := \xi_n$ and $\xi' := \bar{\xi}$ in (44) to obtain

$$\langle x_n - \bar{x} | \xi_n - \bar{\xi} \rangle \geq 2\varphi_{\mathcal{B}} \left(\|x_n - \bar{x}\| \right) , \quad \forall n . \tag{45}$$

Similarly to (41), it can be verified that $\lim_{n\to\infty} \langle x_n | \xi_n \rangle = \langle \bar{x} | \bar{\xi} \rangle$. Thus

$$\begin{aligned} \lim_{n\to\infty} \langle x_n - \bar{x} | \xi_n - \bar{\xi} \rangle &= \lim_{n\to\infty} \langle x_n | \xi_n \rangle - \lim_{n\to\infty} \langle x_n | \bar{\xi} \rangle - \lim_{n\to\infty} \langle \bar{x} | \xi_n \rangle + \langle \bar{x} | \bar{\xi} \rangle \\ &= \langle \bar{x} | \bar{\xi} \rangle - \langle \bar{x} | \bar{\xi} \rangle - \langle \bar{x} | \bar{\xi} \rangle + \langle \bar{x} | \bar{\xi} \rangle = 0 . \end{aligned}$$

Hence, the application of $\lim_{n\to\infty}$ to (45) yields $\lim_{n\to\infty} \varphi_{\mathcal{B}}(\|x_n - \bar{x}\|) = 0$, and thus $\lim_{n\to\infty} \|x_n - \bar{x}\| = 0$. ■

Corollary 3.5: *Consider again the setting of Theorem* 3.1. *In the case where the non-smooth part of the composite loss becomes zero, i.e. $g:=0$, then* (13) *takes the special form*

$$x_{1/2} := T_\alpha x_0 - \lambda \nabla f(x_0), \tag{46a}$$

$$x_1 := x_{1/2}, \tag{46b}$$

$$x_{n+3/2} := x_{n+1/2} - \left[T_\alpha x_n - \lambda \nabla f(x_n)\right] + \left[T x_{n+1} - \lambda \nabla f(x_{n+1})\right], \tag{46c}$$

$$x_{n+2} := x_{n+3/2}\,. \tag{46d}$$

Consider $\alpha \in [0.5, 1)$ and $\lambda \in (0, 2(1-\alpha)/L)$. Then the following hold true.

(i) *For sequence $(x_n)_{n\in\mathbb{Z}_{\geq 0}}$ of* (46), *there exist a sequence $(v_n)_{n\in\mathbb{Z}_{\geq 0}} \subset \mathcal{X}$ and a strongly positive operator $\Theta : \mathcal{X}^2 \to \mathcal{X}^2$ s.t. sequence $(y_n := (x_n, v_n))_{n\in\mathbb{Z}_{>0}\setminus\{1\}}$ is Fejér monotone [3, Def. 5.1, p. 75] w.r.t. $\Upsilon_*^{(\lambda)}$ of Proposition* 2.15 *(under $g=0$) in the Hilbert space $(\mathcal{X}^2, \langle\cdot|\cdot\rangle_\Theta)$.*

(ii) Sequence $(x_n)_{n\in\mathbb{Z}_{\geq 0}}$ of (46) converges weakly to a point that solves $\mathrm{VIP}(\nabla f, \mathrm{Fix}\, T)$.

In the case where $f:=0$, the FM-HSDM recursions take the form

$$x_{1/2} := T_\alpha x_0, \tag{47a}$$

$$x_1 := \mathrm{Prox}_{\lambda g}(x_{1/2}), \tag{47b}$$

$$x_{n+3/2} := x_{n+1/2} - T_\alpha x_n + T x_{n+1}, \tag{47c}$$

$$x_{n+2} := \mathrm{Prox}_{\lambda g}(x_{n+3/2})\,. \tag{47d}$$

Consider $\alpha \in [0.5, 1)$ and $\lambda \in \mathbb{R}_{>0}$. Then the following hold true.

(i) *For sequence $(x_n)_{n\in\mathbb{Z}_{\geq 0}}$ of* (47), *there exist a sequence $(v_n)_{n\in\mathbb{Z}_{\geq 0}} \subset \mathcal{X}$ and a strongly positive operator $\Theta : \mathcal{X}^2 \to \mathcal{X}^2$ s.t. sequence $(y_n := (x_n, v_n))_{n\in\mathbb{Z}_{>0}\setminus\{1\}}$ is Fejér monotone [3, Def. 5.1, p. 75] w.r.t. $\Upsilon_*^{(\lambda)}$ of Proposition* 2.15 *(under $f=0$) in the Hilbert space $(\mathcal{X}^2, \langle\cdot|\cdot\rangle_\Theta)$.*

(ii) *Sequence $(x_n)_{n\in\mathbb{Z}_{\geq 0}}$ of* (47) *converges weakly to a point that solves* $\mathrm{VIP}(\partial g, \mathrm{Fix}\, T)$.

Proof: The proof becomes a special case of the one of Theorem 3.1, after setting $f:=0$ or $g:=0$. With regards to the reason behind the relaxation of λ offered by (47), notice that any $\lambda \in \mathbb{R}_{>0}$ can serve as the Lipschitz constant of $\nabla f = 0$. ■

The following theorem draws even stronger links with the original form of HSDM.

Theorem 3.6: *Consider $f \in \Gamma_0(\mathcal{X})$, with L being the Lipschitz-continuity constant of ∇f. Moreover, given the closed affine set $\mathcal{A}$, consider any $T \in \mathfrak{T}_{\mathcal{A}}$, and for $\lambda \in \mathbb{R}_{>0}$, an arbitrarily fixed $x_0 \in \mathcal{X}$, and for all $n \in \mathbb{Z}_{\geq 0}$ form the iterations:*

$$x_{1/2} := T_\alpha x_0 - \lambda \nabla f(T_\alpha x_0), \tag{48a}$$

$$x_1 := x_{1/2}, \tag{48b}$$

$$x_{n+3/2} := x_{n+1/2} - \left[T_\alpha x_n - \lambda \nabla f(T_\alpha x_n)\right] + \left[T x_{n+1} - \lambda \nabla f(T_\alpha x_{n+1})\right], \tag{48c}$$

$$x_{n+2} := x_{n+3/2}, \tag{48d}$$

where T_α is defined in (12). Consider also $\alpha \in [0.5, 1)$ and $\lambda \in (0, 2(1-\alpha)^2/L)$. Then, the following hold true.

- (i) *There exist a sequence $(v_n)_{n \in \mathbb{Z}_{\geq 0}} \subset \mathcal{X}$ and a strongly positive operator $\Upsilon : \mathcal{X}^2 \to \mathcal{X}^2$ s.t. sequence $(y_n := (x_n, v_n))_{n \in \mathbb{Z}_{>0} \setminus \{1\}}$ is Fejér monotone [3, Def. 5.1, p. 75] w.r.t. $\Upsilon_*^{(\lambda)}$ of Proposition 2.15 (under $g = 0$) in the Hilbert space $(\mathcal{X}^2, \langle \cdot | \cdot \rangle_\Upsilon)$.*
- (ii) *Sequence $(x_n)_n$ of (48) converges weakly to a point that solves $\mathrm{VIP}(\nabla f, \mathrm{Fix}\, T)$.*
- (iii) *If Assumption 3.3(i) also holds true, then $(x_n)_n$ of (48) converges strongly to a point that solves $\mathrm{VIP}(\nabla f, \mathrm{Fix}\, T)$.*

Proof: **(i)** Proposition 2.15 takes the following special form in the present context: if $\exists v_* \in \mathcal{X}$ s.t.

$$(x_*, v_*) \in \Upsilon_*^{(\lambda)} := \{(x, v) \in \mathrm{Fix}\, T \times \mathcal{X} \mid -\tfrac{1}{\lambda} U v = \nabla f(x)\}, \tag{49}$$

then x_* solves $\mathrm{VIP}(\nabla f, \mathrm{Fix}\, T)$.

By following the same steps which start from the beginning of the proof of Theorem 3.1 till (20), it can be verified that

$$-(1-2\alpha)(T - \mathrm{Id}\,)x_{n+1} - Q_\alpha(x_{n+1} - x_n) - U v_{n+1} = \lambda \nabla f(T_\alpha x_n), \tag{50}$$

and by considering any $(x_*, v_*) \in \Upsilon_*^{(\lambda)}$,

$$\begin{aligned}&\lambda[\nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)] \\ &\quad = -(1-2\alpha)(Q - \mathrm{Id}\,)(x_{n+1} - x_*) - Q_\alpha(x_{n+1} - x_n) - U(v_{n+1} - v_*)\,.\end{aligned} \tag{51}$$

As in the proof of Theorem 3.1, the Baillon–Haddad theorem [29], [3, Cor. 18.16, p. 270] suggests that

$$
\begin{aligned}
&\tfrac{2\lambda}{L}\|\nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)\|^2 \\
&\quad \leq 2\lambda\langle T_\alpha x_n - T_\alpha x_* | \nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)\rangle \\
&\quad = 2\lambda\langle Q_\alpha(x_n - x_*) | \nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)\rangle \\
&\quad = 2\lambda\langle x_n - x_* | Q_\alpha[\nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)]\rangle \\
&\quad = 2\lambda\langle x_{n+1} - x_* | Q_\alpha[\nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)]\rangle \\
&\qquad + 2\lambda\langle x_n - x_{n+1} | Q_\alpha[\nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)]\rangle \\
&\quad = -2(1-2\alpha)\langle x_{n+1} - x_* | Q_\alpha(Q - \mathrm{Id}\,)(x_{n+1} - x_*)\rangle \\
&\qquad - 2\langle x_{n+1} - x_* | Q_\alpha^2(x_{n+1} - x_n)\rangle - 2\langle x_{n+1} - x_* | Q_\alpha U(v_{n+1} - v_*)\rangle \\
&\qquad + 2\lambda\langle x_n - x_{n+1} | Q_\alpha[\nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)]\rangle \\
&\quad = -2(1-2\alpha)\langle x_{n+1} - x_* | Q_\alpha(Q - \mathrm{Id}\,)(x_{n+1} - x_*)\rangle \\
&\qquad - 2\langle x_{n+1} - x_* | Q_\alpha^2(x_{n+1} - x_n)\rangle - 2\langle U(x_{n+1} - x_*) | Q_\alpha(v_{n+1} - v_*)\rangle \\
&\qquad + 2\lambda\langle x_n - x_{n+1} | Q_\alpha[\nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)]\rangle \\
&\quad \leq -2(1-2\alpha)\langle x_{n+1} - x_* | Q_\alpha(Q - \mathrm{Id}\,)(x_{n+1} - x_*)\rangle \\
&\qquad - 2\langle x_{n+1} - x_* | Q_\alpha^2(x_{n+1} - x_n)\rangle - \tfrac{2}{1-\alpha}\langle v_{n+1} - v_n | Q_\alpha(v_{n+1} - v_*)\rangle \\
&\qquad + \tfrac{\lambda L}{2}\|x_n - x_{n+1}\|^2 + \tfrac{2\lambda}{L}\|Q_\alpha[\nabla f(x_n) - \nabla f(x_*)]\|^2 \\
&\quad \leq -2(2\alpha-1)\langle x_{n+1} - x_* | Q_\alpha(\mathrm{Id} - Q)(x_{n+1} - x_*)\rangle \\
&\qquad - 2\langle x_{n+1} - x_* | Q_\alpha^2(x_{n+1} - x_n)\rangle - \tfrac{2}{1-\alpha}\langle v_{n+1} - v_n | Q_\alpha(v_{n+1} - v_*)\rangle \\
&\qquad + \tfrac{\lambda L}{2}\|x_n - x_{n+1}\|^2 + \tfrac{2\lambda}{L}\|\nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)\|^2 \\
&\quad \leq 2\langle x_* - x_{n+1} | Q_\alpha^2(x_{n+1} - x_n)\rangle + \tfrac{2}{1-\alpha}\langle v_{n+1} - v_n | Q_\alpha(v_* - v_{n+1})\rangle \\
&\qquad + \tfrac{\lambda L}{2}\|x_n - x_{n+1}\|^2 + \tfrac{2\lambda}{L}\|\nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)\|^2 .
\end{aligned} \tag{52}
$$

Mapping Q_α^2 is strongly positive: indeed, if U_α denotes the square root of the strongly positive Q_α [*cf.* (29)], then $\forall x \in \mathcal{X}$, $\langle Q_\alpha^2 x | x\rangle = \langle U_\alpha Q_\alpha U_\alpha x | x\rangle = \langle Q_\alpha U_\alpha x | U_\alpha x\rangle \geq (1-\alpha)\langle U_\alpha x | U_\alpha x\rangle = (1-\alpha)\langle Q_\alpha x | x\rangle \geq (1-\alpha)^2\|x\|^2$. Define now the mapping $\Upsilon : \mathcal{X}^2 \to \mathcal{X}^2 : (x, v) \mapsto (Q_\alpha^2 x, [1/(1-\alpha)]Q_\alpha v)$. Mapping Υ turns out to be strongly positive, w.r.t. the standard inner product of $\mathcal{X}^2$: $\langle (x, v) | (x, v')\rangle := \langle x | x'\rangle + \langle v | v'\rangle$, $\forall (x, v), (x', v') \in \mathcal{X}^2$, due to the strong positivity of Q_α^2 and $[1/(1-\alpha)]Q_\alpha$. Consequently, one can consider $(\mathcal{X}^2, \langle\cdot|\cdot\rangle_\Upsilon)$ as a Hilbert space equipped with the inner product $\langle (x, v) | (x, v')\rangle_\Upsilon := \langle x | Q_\alpha^2 x'\rangle + [1/(1-\alpha)]\langle v | Q_\alpha v'\rangle$, $\forall (x, v), (x', v') \in \mathcal{X}^2$. As such, (52) becomes

$$0 \leq 2\langle y_{n+1} - y_n | \Upsilon(y_* - y_{n+1})\rangle + \tfrac{\lambda L}{2}\|x_n - x_{n+1}\|^2$$

$$= 2\langle y_{n+1} - y_n | y_* - y_{n+1}\rangle_\Upsilon + \tfrac{\lambda L}{2}\|x_n - x_{n+1}\|^2$$
$$= \|y_n - y_*\|_\Upsilon^2 - \|y_{n+1} - y_*\|_\Upsilon^2 - \|y_{n+1} - y_n\|_\Upsilon^2 + \tfrac{\lambda L}{2}\|x_n - x_{n+1}\|^2 . \quad (53)$$

Choose, now, any ζ' with $\lambda L/[2(1-\alpha)^2] < \zeta' < 1$. Then, for any $y = (x, v) \in \mathcal{X}^2$,

$$\tfrac{\lambda L}{2}\|x\|^2 < \zeta'(1-\alpha)^2\|x\|^2 \leq \zeta'\langle x|Q_\alpha^2 x\rangle$$
$$\leq \zeta'\langle x|Q_\alpha^2 x\rangle + \zeta'\tfrac{1}{1-\alpha}\langle v|Q_\alpha v\rangle = \zeta'\|y\|_\Upsilon^2 .$$

This argument together with (53) yield

$$\|y_n - y_*\|_\Upsilon^2 - \|y_{n+1} - y_*\|_\Upsilon^2 \geq \|y_{n+1} - y_n\|_\Upsilon^2 - \tfrac{\lambda L}{2}\|x_n - x_{n+1}\|^2$$
$$\geq \|y_{n+1} - y_n\|_\Upsilon^2 - \zeta'\|y_{n+1} - y_n\|_\Upsilon^2$$
$$= (1-\zeta')\|y_{n+1} - y_n\|_\Upsilon^2, \quad (54)$$

i.e. sequence $(y_n)_{n\in\mathbb{Z}_{\geq 0}} \subset (\mathcal{X}^2, \langle\cdot|\cdot\rangle_\Upsilon)$ is Fejér monotone w.r.t. $\Upsilon_*^{(\lambda)}$ of (49).

(ii) Due to Fejér monotonicity, (y_n) is bounded [3, Prop. 5.4(i), p. 76] and possesses a non-empty set of weakly sequential cluster points $\mathfrak{W}[(y_n)_n]$ [3, Lem. 2.37, p. 36]. Moreover, it can be readily verified, as in (29), that $\lim_{n\to\infty}(y_{n+1} - y_n) = 0$, $\lim_{n\to\infty}(x_{n+1} - x_n) = 0$ and $\lim_{n\to\infty}(v_{n+1} - v_n) = 0$. The rest of the proof follows steps similar to those after (29) in the proof of Theorem 3.1, but with the following twist: $\nabla f(x_{n_k})$ is replaced by $\nabla f(T_\alpha x_{n_k})$, where all the asymptotic results of the proof of Theorem 3.1 continue to hold due to the Lipschitz continuity of ∇f and the nonexpansiveness of T_α, e.g. $\forall x, x' \in \mathcal{X}$,

$$\|\nabla f(T_\alpha x) - \nabla f(T_\alpha x')\| \leq L\|T_\alpha x - T_\alpha x'\| \leq L\|x - x'\| .$$

(iii) Part **(ii)** of this proof has demonstrated that sequences $(x_n)_n$ and $(Uv_n)_n$ converge weakly to $\bar{x}$ and $U\bar{v}$, respectively. Consequently, in a way similar to part **(ii)** of the proof of Theorem 3.1, it can be shown also here that $(\xi_n)_n$ converges weakly to $\bar{\xi}$.

Let Assumption 3.3(i) holds true. Then, according to [3, Ex. 22.3(iii), p. 324], given a bounded set $\mathcal{B} \subset \mathcal{X}$, there exists an increasing function $\varphi_\mathcal{B} : [0, +\infty) \to [0, +\infty]$, which vanishes only at 0, s.t. $x, x' \in \mathcal{B}$,

$$\langle x - x'|\nabla f(x) - \nabla f(x')\rangle \geq 2\varphi_\mathcal{B}\left(\|x - x'\|\right) . \quad (55)$$

Due to the nonexpansiveness of T_α and the boundedness of $(x_n)_n$, by part **(i)** of the proof, it turns out that $(T_\alpha x_n)_n$ is also bounded: $\|T_\alpha x_n\| \leq \|T_\alpha x_n - T_\alpha\bar{x}\| + \|T_\alpha\bar{x}\| \leq \|x_n - \bar{x}\| + \|\bar{x}\| \leq \|x_n\| + 2\|\bar{x}\| \leq C'' + 2\|\bar{x}\|$, for some $C'' \in \mathbb{R}_{>0}$ (recall that $\bar{x} \in \text{Fix } T_\alpha = \text{Fix } T$). Define, thus, the bounded set $\mathcal{B} :=$

$(T_\alpha x_n)_n \cup \{\bar{x}\}$. As such, (55) yields

$$\begin{aligned}&\langle (T_\alpha - \mathrm{Id}\,)x_n | \nabla f(T_\alpha x_n) - \nabla f(\bar{x})\rangle + \langle x_n - \bar{x} | \nabla f(T_\alpha x_n) - \nabla f(\bar{x})\rangle \\ &\quad = \langle T_\alpha x_n - \bar{x} | \nabla f(T_\alpha x_n) - \nabla f(\bar{x})\rangle \geq 2\varphi_{\mathcal{B}}\left(\|T_\alpha x_n - \bar{x}\|\right), \quad \forall n. \qquad (56)\end{aligned}$$

Part **(i)** of this proof has already showed that $\lim_{n\to\infty}(T - \mathrm{Id}\,)x_n = 0$. As such, $\lim_{n\to\infty}(T_\alpha - \mathrm{Id}\,)x_n = \alpha \lim_{n\to\infty}(T - \mathrm{Id}\,)x_n = 0$. Moreover, note that $x_n \rightharpoonup_{n\to\infty} \bar{x}$ and $\lim_{n\to\infty} \nabla f(T_\alpha x_n) = \nabla f(\bar{x})$. Hence, due also to [3, Lem. 2.41(iii), p. 37], an application of $\lim_{n\to\infty}$ to both sides of (56) results in $\lim_{n\to\infty} \varphi_{\mathcal{B}}(\|T_\alpha x_n - \bar{x}\|) = 0$, and thus $\lim_{n\to\infty} T_\alpha x_n = \bar{x}$. Using $\lim_{n\to\infty}(T_\alpha - \mathrm{Id}\,)x_n = 0$, one can easily verify that $\lim_{n\to\infty} x_n = \lim_{n\to\infty}(\mathrm{Id}\, - T_\alpha)x_n + \lim_{n\to\infty} T_\alpha x_n = \bar{x}$, which establishes part (iii) of Theorem 3.6. ■

The following theorems present convergence rates on the sequence of FM-HSDM estimates.

Theorem 3.7: *For sequence $(x_n)_{n\in\mathbb{Z}_{\geq 0}}$ of* (13)*, there exists $\xi_n \in \partial g(x_n), \forall n$, s.t. for any $x_* \in \mathrm{Fix}\, T$,*

$$\tfrac{1}{n+1}\sum_{\nu=0}^{n}\langle x_{\nu+1} - x_* | (\mathrm{Id} - Q)(x_{\nu+1} - x_*)\rangle = O(\tfrac{1}{n+1}), \qquad (57a)$$

$$\tfrac{1}{n+1}\sum_{\nu=0}^{n}\| Uv_{\nu+1} + \lambda[\nabla f(x_\nu) + \xi_{\nu+1}]\|^2 = O(\tfrac{1}{n+1}), \qquad (57b)$$

$$\tfrac{1}{n+1}\sum_{\nu=0}^{n}\|(\mathrm{Id}\, - T)x_{\nu+1}\|^2 = O(\tfrac{1}{n+1}), \qquad (57c)$$

where the big-oh notation $a_n = O(b_n)$, $b_n > 0$, means $\limsup_{n\to\infty} |a_n|/b_n < +\infty$. Regarding sequence $(x_n)_{n\in\mathbb{Z}_{\geq 0}}$ of (46), (57a)–(57c) *still hold true, but $\xi_{\nu+1}$ is set equal to 0 in* (57b)*. Similarly, for sequence $(x_n)_{n\in\mathbb{Z}_{\geq 0}}$ of* (48), (57a), (57c) *as well as*

$$\tfrac{1}{n+1}\sum_{\nu=0}^{n}\| Uv_{\nu+1} + \lambda\nabla f(T_\alpha x_\nu)\|^2 = O(\tfrac{1}{n+1})$$

hold true.

Proof: First, notice by (25), Proposition A.5 and $\|Q_\alpha\| \leq 1$ that Q_α^{-1} exists and it is strongly positive with

$$\|Q_\alpha^{-1}\| \leq \tfrac{1}{1-\alpha}; \quad (1-\alpha)\|x\|^2 \leq \tfrac{(1-\alpha)}{\|Q_\alpha\|^2}\|x\|^2 \leq \langle Q_\alpha^{-1}x | x\rangle, \quad \forall x \in \mathcal{X}. \qquad (58)$$

Then, going back to the discussion following (25),

$\|y_{n+1} - y_n\|_\Theta^2$

$$= \|x_{n+1} - x_n\|^2_{Q_\alpha} + \tfrac{1}{1-\alpha}\|v_{n+1} - v_n\|^2 \tag{59a}$$

$$= \|Q_\alpha(x_{n+1} - x_n)\|^2_{Q_\alpha^{-1}} + \tfrac{1}{1-\alpha}\|(1-\alpha)U(x_{n+1} - x_*)\|^2 \tag{59b}$$

$$\begin{aligned}
&= \|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}] - (1-2\alpha)(\mathrm{Id} - T)x_{n+1}\|^2_{Q_\alpha^{-1}} \\
&\quad + \tfrac{1}{1-\alpha}\|(1-\alpha)U(x_{n+1} - x_*)\|^2
\end{aligned} \tag{59c}$$

$$\begin{aligned}
&= \|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|^2_{Q_\alpha^{-1}} + (1-2\alpha)^2\|(\mathrm{Id} - T)x_{n+1}\|^2_{Q_\alpha^{-1}} \\
&\quad - 2\langle Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}] \mid (1-2\alpha)(\mathrm{Id} - T)x_{n+1}\rangle_{Q_\alpha^{-1}} \\
&\quad + \tfrac{1}{1-\alpha}\|(1-\alpha)U(x_{n+1} - x_*)\|^2
\end{aligned}$$

$$\begin{aligned}
&\geq \tfrac{1}{\rho}\|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|^2_{Q_\alpha^{-1}} - \tfrac{(1-2\alpha)^2}{\rho-1}\|(\mathrm{Id} - T)x_{n+1}\|^2_{Q_\alpha^{-1}} \\
&\quad + \tfrac{1}{1-\alpha}\|(1-\alpha)U(x_{n+1} - x_*)\|^2
\end{aligned} \tag{59d}$$

$$\begin{aligned}
&= \tfrac{1}{\rho}\|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|^2_{Q_\alpha^{-1}} - \tfrac{(1-2\alpha)^2}{\rho-1}\|(\mathrm{Id} - T)x_{n+1} \\
&\quad - (\mathrm{Id} - T)x_*\|^2_{Q_\alpha^{-1}} + (1-\alpha)\langle x_{n+1} - x_* \mid (\mathrm{Id} - Q)(x_{n+1} - x_*)\rangle
\end{aligned}$$

$$\begin{aligned}
&= \tfrac{1}{\rho}\|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|^2_{Q_\alpha^{-1}} - \tfrac{(1-2\alpha)^2}{\rho-1}\|(\mathrm{Id} - Q)(x_{n+1} - x_*)\|^2_{Q_\alpha^{-1}} \\
&\quad + (1-\alpha)\langle x_{n+1} - x_* \mid (\mathrm{Id} - Q)(x_{n+1} - x_*)\rangle
\end{aligned}$$

$$\begin{aligned}
&= \tfrac{1}{\rho}\|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|^2_{Q_\alpha^{-1}} \\
&\quad - \tfrac{(1-2\alpha)^2}{\rho-1}\langle x_{n+1} - x_* \mid (\mathrm{Id} - Q)Q_\alpha^{-1}(\mathrm{Id} - Q)(x_{n+1} - x_*)\rangle \\
&\quad + (1-\alpha)\langle x_{n+1} - x_* \mid (\mathrm{Id} - Q)(x_{n+1} - x_*)\rangle
\end{aligned}$$

$$\geq \tfrac{1}{\rho}\|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|^2_{Q_\alpha^{-1}} \tag{59e}$$

$$\begin{aligned}
&\quad - \tfrac{(2\alpha-1)^2}{(\rho-1)(1-\alpha)}\langle x_{n+1} - x_* \mid (\mathrm{Id} - Q)(x_{n+1} - x_*)\rangle \\
&\quad + (1-\alpha)\langle x_{n+1} - x_* \mid (\mathrm{Id} - Q)(x_{n+1} - x_*)\rangle
\end{aligned} \tag{59f}$$

$$= \tfrac{1}{\rho}\|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|^2_{Q_\alpha^{-1}} + \theta\langle x_{n+1} - x_* \mid (\mathrm{Id} - Q)(x_{n+1} - x_*)\rangle \tag{59g}$$

$$\geq \tfrac{(1-\alpha)}{\rho}\|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|^2 + \theta\langle x_{n+1} - x_* \mid (\mathrm{Id} - Q)(x_{n+1} - x_*)\rangle, \tag{59h}$$

$$\geq \tfrac{(1-\alpha)}{\rho}\|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|^2 + \theta(1-\alpha)\|(\mathrm{Id} - Q)(x_{n+1} - x_*)\|^2_{Q_\alpha^{-1}} \tag{59i}$$

$$= \tfrac{(1-\alpha)}{\rho}\|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|^2 + \theta(1-\alpha)\|(\mathrm{Id} - T)x_{n+1}\|^2_{Q_\alpha^{-1}} \tag{59j}$$

$$\geq \tfrac{(1-\alpha)}{\rho}\|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|^2 + \theta(1-\alpha)^2\|(\mathrm{Id} - T)x_{n+1}\|^2, \tag{59k}$$

where the definition of Υ, given after (52), is used in (59a), (19) in (59b), (21) in (59c), (24) with $\eta := \rho/(\rho-1)$, $a := Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]$, $b := (1-2\alpha)(\mathrm{Id} - T)x_{n+1}$ and $\Pi := Q_\alpha^{-1}$, as well as $\rho > 1$ in (59d), and

$$\begin{aligned}
&\langle x_{n+1} - x_* | (\mathrm{Id} - Q) Q_\alpha^{-1} (\mathrm{Id} - Q)(x_{n+1} - x_*) \rangle \\
&\quad = \langle x_{n+1} - x_* | U^2 Q_\alpha^{-1} U^2 (x_{n+1} - x_*) \rangle \\
&\quad = \langle U(x_{n+1} - x_*) | \left(U Q_\alpha^{-1} U \right) U(x_{n+1} - x_*) \rangle \\
&\quad \leq \| U Q_\alpha^{-1} U \| \, \langle U(x_{n+1} - x_*) | U(x_{n+1} - x_*) \rangle && \text{(60a)} \\
&\quad = \| U Q_\alpha^{-1} U \| \, \langle x_{n+1} - x_* | (\mathrm{Id} - Q)(x_{n+1} - x_*) \rangle \\
&\quad \leq \| U \|^2 \| Q_\alpha^{-1} \| \, \langle x_{n+1} - x_* | (\mathrm{Id} - Q)(x_{n+1} - x_*) \rangle \\
&\quad = \| \mathrm{Id} - Q \| \| Q_\alpha^{-1} \| \, \langle x_{n+1} - x_* | (\mathrm{Id} - Q)(x_{n+1} - x_*) \rangle \\
&\quad \leq \tfrac{1}{1-\alpha} \langle x_{n+1} - x_* | (\mathrm{Id} - Q)(x_{n+1} - x_*) \rangle && \text{(60b)}
\end{aligned}$$

with (58) and $\|\mathrm{Id} - Q\| \leq 1$ in (59f). Note that [4, Thm. 9.2-2, p. 466] is used in (60a). Moreover, $\theta := (1-\alpha) - (2\alpha-1)^2/[(1-\alpha)(\rho-1)]$ becomes positive for any $\rho > 1 + (2\alpha-1)^2/(1-\alpha)^2$ in (59g), (58) in (59h), (60b) in (59i), the fact $(\mathrm{Id} - Q)(x_{n+1} - x_*) = (\mathrm{Id} - T)x_{n+1} - (\mathrm{Id} - T)x_* = (\mathrm{Id} - T)x_{n+1}$ in (59j), and (58) in (59k).

Due to (28), the previous considerations suggest that there exists $C \in \mathbb{R}_{>0}$ s.t. $\forall n$,

$$\begin{aligned}
\tfrac{C}{n+1} &\geq \tfrac{1}{n+1} \sum_{\nu=0}^{n} \| y_{\nu+1} - y_\nu \|_\Theta^2 \\
&\geq \tfrac{1}{\rho(n+1)} \sum_{\nu=0}^{n} \| U v_{\nu+1} + \lambda[\nabla f(x_\nu) + \xi_{\nu+1}] \|^2 \\
&\quad + \tfrac{\theta}{n+1} \sum_{\nu=0}^{n} \langle x_{\nu+1} - x_* | (\mathrm{Id} - Q)(x_{\nu+1} - x_*) \rangle \\
&\geq \tfrac{1}{\rho(n+1)} \sum_{\nu=0}^{n} \| U v_{\nu+1} + \lambda[\nabla f(x_\nu) + \xi_{\nu+1}] \|^2 + \tfrac{\theta(1-\alpha)^2}{n+1} \sum_{\nu=0}^{n} \| (\mathrm{Id} - T) x_{\nu+1} \|^2,
\end{aligned}$$

which establishes the claim of Theorem 3.7 regarding the sequence of (13). The proof of the claim with regards to the sequence of (47) follows the same steps as the previous one, but with the twist of replacing $\nabla f(x_n)$ by $\nabla f(T_\alpha x_n)$ and $g = 0$. ■

Theorem 3.8: *For the sequence* $(x_n)_{n\in\mathbb{N}}$ *of* (47), *there exists* $\xi_n \in \partial g(x_n)$, $\forall n$, *s.t. for any* $x_* \in \operatorname{Fix} T$,

$$\begin{aligned}
&\langle x_{n+1} - x_* | (\mathrm{Id} - Q)(x_{n+1} - x_*)\rangle = O(\tfrac{1}{n+1}),\\
&\|Uv_{n+1} + \lambda\xi_{n+1}\|^2 = O(\tfrac{1}{n+1}),\\
&\|(\mathrm{Id} \ - T)x_{n+1}\|^2 = O(\tfrac{1}{n+1})\,.
\end{aligned}$$

Proof: Define here $\Delta x_n := x_{n-1} - x_n$, $\Delta v_n := v_{n-1} - v_n$, $\Delta y_n := (\Delta x_n, \Delta v_n)$ and $\Delta\xi_n := \xi_{n-1} - \xi_n$, $\forall n$. Under these definitions and in the case of $f = 0$, (31) yields

$$\begin{aligned}
&(1-2\alpha)(Q - \mathrm{Id}\,)(x_n - x_{n+1}) + Q_\alpha\left[(x_n - x_{n+1}) - (x_{n-1} - x_n)\right]\\
&\quad = -U(v_n - v_{n+1}) - \lambda(\xi_n - \xi_{n+1})\\
&\quad \Leftrightarrow (1-2\alpha)(Q - \mathrm{Id}\,)\Delta x_{n+1} + Q_\alpha(\Delta x_{n+1} - \Delta x_n) = -U\Delta v_{n+1} - \lambda\Delta\xi_{n+1}\\
&\quad \Leftrightarrow \lambda\Delta\xi_{n+1} = -U\Delta v_{n+1} - Q_\alpha(\Delta x_{n+1} - \Delta x_n) - (1-2\alpha)(Q - \mathrm{Id}\,)\Delta x_{n+1}\,.
\end{aligned} \tag{61}$$

Moreover, (19) suggests that $-\Delta v_{n+1} = (1-\alpha)U(x_{n+1} - x_*)$, and thus

$$\tfrac{1}{1-\alpha}(\Delta v_{n+1} - \Delta v_n) = U\Delta x_{n+1}\,. \tag{62}$$

The monotonicity of $\partial g(\cdot)$, (61), (62), and the definition of Θ, introduced after (25), imply that

$$\begin{aligned}
&0 \le \langle \Delta x_{n+1} | \lambda\Delta\xi_{n+1}\rangle\\
&\quad \Leftrightarrow 0 \le \langle \Delta x_{n+1} | -U\Delta v_{n+1} - Q_\alpha(\Delta x_{n+1} - \Delta x_n) - (2\alpha - 1)(\mathrm{Id} - Q)\Delta x_{n+1}\rangle\\
&\quad \Leftrightarrow (2\alpha - 1)\langle \Delta x_{n+1} | (\mathrm{Id} - Q)\Delta x_{n+1}\rangle\\
&\quad \le -\langle U\Delta x_{n+1} | \Delta v_{n+1}\rangle - \langle \Delta x_{n+1} | Q_\alpha(\Delta x_{n+1} - \Delta x_n)\rangle\\
&\quad \Leftrightarrow (2\alpha - 1)\langle \Delta x_{n+1} | (\mathrm{Id} - Q)\Delta x_{n+1}\rangle\\
&\quad \le -\tfrac{1}{1-\alpha}\langle \Delta v_{n+1} - \Delta v_n | \Delta v_{n+1}\rangle - \langle \Delta x_{n+1} | Q_\alpha(\Delta x_{n+1} - \Delta x_n)\rangle\\
&\quad \Leftrightarrow (2\alpha - 1)\langle \Delta x_{n+1} | (\mathrm{Id} - Q)\Delta x_{n+1}\rangle \le \langle \Delta y_{n+1} | \Delta y_n - \Delta y_{n+1}\rangle_\Theta\\
&\quad \Leftrightarrow (2\alpha - 1)\langle \Delta x_{n+1} | (\mathrm{Id} - Q)\Delta x_{n+1}\rangle \le \tfrac{1}{2}\big(\|\Delta y_n\|_\Theta^2 - \|\Delta y_{n+1}\|_\Theta^2\\
&\qquad - \|\Delta y_n - \Delta y_{n+1}\|_\Theta^2\big)\\
&\quad \Leftrightarrow 2(2\alpha - 1)\langle \Delta x_{n+1} | (\mathrm{Id} - Q)\Delta x_{n+1}\rangle + \|\Delta y_n - \Delta y_{n+1}\|_\Theta^2\\
&\quad \le \|\Delta y_n\|_\Theta^2 - \|\Delta y_{n+1}\|_\Theta^2,
\end{aligned} \tag{63}$$

and due to $\alpha \ge 1/2$ as well as the positive-definiteness of Id $- Q$, (63) yields

$$\|y_{n+1} - y_n\|_\Theta^2 \le \|y_n - y_{n-1}\|_\Theta^2, \quad \forall n\,. \tag{64}$$

Now, (28) and (64) imply that there exists $C > 0$ s.t. for any n,

$$(n+1)\|y_{n+1} - y_n\|_\Theta^2 \leq \sum_{\nu=0}^{n} \|y_{\nu+1} - y_\nu\|_\Theta^2 \leq C,$$

and thus $\|y_{n+1} - y_n\|_\Theta^2 \leq C/(n+1)$. This result applied to (59h) and (59k) establishes the claim of Theorem 3.8. ■

4. Numerical tests

To validate the previous theoretical findings, tests are conducted on a simple scenario which is motivated by [13, Prob. 4.1]. More elaborate tests, involving noisy real data, are deferred to an upcoming publication where FM-HSDM is extended to a stochastic setting.

Given dimension $d \in \mathbb{Z}_{>0}$, the real Euclidean space $\mathscr{X}_0 := \mathbb{R}^d$ is considered. Upon defining the closed ball $\mathcal{B}[\mathbf{u}_c, r] := \{\mathbf{u} \in \mathscr{X}_0 \mid \|\mathbf{u} - \mathbf{u}_c\|_2 \leq r\}$, for centre $\mathbf{u}_c \in \mathscr{X}_0$ and radius $r \in \mathbb{R}_{>0}$, let $\mathcal{B}_1 := \mathcal{B}[\mathbf{u}_{c1}, r_1] := \mathcal{B}[2\mathbf{e}_1, 1]$ and $\mathcal{B}_2 := \mathcal{B}[\mathbf{u}_{c2}, r_2] := \mathcal{B}[\mathbf{0}, 2]$, where $\mathbf{e}_1$ stands for the first column of the $d \times d$ identity matrix $\mathbf{I}_d$. In all tests, $d := 10{,}000$. Let also $\mathbf{P}$ denote a $d \times d$ diagonal positive-definite matrix, whose *unique* smallest entry $[\mathbf{P}]_{11} \leq 1$ is fixed at position $(1, 1)$, and its largest entry, placed at position (d, d), is set to be equal to 10. This setting is fixed across all experiments. Each experiment in the sequel randomly draws numbers from the interval $([\mathbf{P}]_{11}, 10)$, under the uniform distribution, and places them in the remaining $d-2$ entries of the diagonal of $\mathbf{P}$. Moreover, in all scenarios, parameter α of FM-HSDM is set equal to 0.5, since this value produced the best performance among all theoretically supported values taken from $[0.5, 1)$.

Along the lines of [13, Prob. 4.1], the following constrained quadratic minimization task is considered:

$$\min_{\mathbf{u} \in \mathcal{B}_1 \cap \mathcal{B}_2} \mathbf{u}^\top \mathbf{P} \mathbf{u} = \min_{\mathbf{x} := (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}) \in \mathscr{X}_0^3 =: \mathcal{X}} \tfrac{1}{2} \mathbf{x}^{(1)\top} \mathbf{P} \mathbf{x}^{(1)} + \iota_{\mathcal{B}_1}(\mathbf{x}^{(2)}) + \iota_{\mathcal{B}_2}(\mathbf{x}^{(3)})$$
$$\text{s.to } \mathbf{x}^{(1)} = \mathbf{x}^{(2)} = \mathbf{x}^{(3)}, \tag{65}$$

where $\mathbf{x} := (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}) := [\mathbf{x}^{(1)\top}, \mathbf{x}^{(2)\top}, \mathbf{x}^{(3)\top}]^\top \in \mathscr{X}_0^3$, and $\mathcal{X} := \mathscr{X}_0^3$ with inner product defined as the standard Euclidean dot-vector product. The definition of the indicator functions $\iota_{\mathcal{B}_1}, \iota_{\mathcal{B}_2}$ can be found in Section 1.2. Since $\mathbf{P} \succ \mathbf{0}$ and the smallest entry of $\mathbf{P}$ is located at the $(1, 1)$ position, the unique solution to (65) is $\mathbf{x}_* := (\mathbf{e}_1, \mathbf{e}_1, \mathbf{e}_1)$. There are several ways of viewing (65) as a special case of (1). For example, $f(\mathbf{x}) := (1/2)\mathbf{x}^{(1)\top}\mathbf{P}\mathbf{x}^{(1)}$ and $g(\mathbf{x}) := \iota_{\mathcal{B}_1}(\mathbf{x}^{(2)}) + \iota_{\mathcal{B}_2}(\mathbf{x}^{(3)})$, for any $\mathbf{x} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)})$. The Lipschitz coefficient of ∇f is the largest entry of $\mathbf{P}$, i.e. $L = 10$, and $\operatorname{Prox}_{\lambda g}(\mathbf{x}) = (\mathbf{x}^{(1)}, P_{\mathcal{B}_1}(\mathbf{x}^{(2)}), P_{\mathcal{B}_2}(\mathbf{x}^{(3)}))$. For any $\lambda \in \mathbb{R}_{>0}$, the proximal mapping of $\iota_{\mathcal{B}_i}$ becomes $\operatorname{Prox}_{\lambda \iota_{\mathcal{B}_i}} = P_{\mathcal{B}_i}$, where $P_{\mathcal{B}_i}$ denotes the metric projection mapping onto the ball $\mathcal{B}_i$, given by $P_{\mathcal{B}_i}(\mathbf{u}) = \mathbf{u}_{ci} + (\mathbf{u} - \mathbf{u}_{ci}) r_i / \max\{\|\mathbf{u} - \mathbf{u}_{ci}\|, r_i\}$, for any $\mathbf{u} \in \mathscr{X}_0$. Furthermore, $\mathcal{A} := \{\mathbf{x} =$

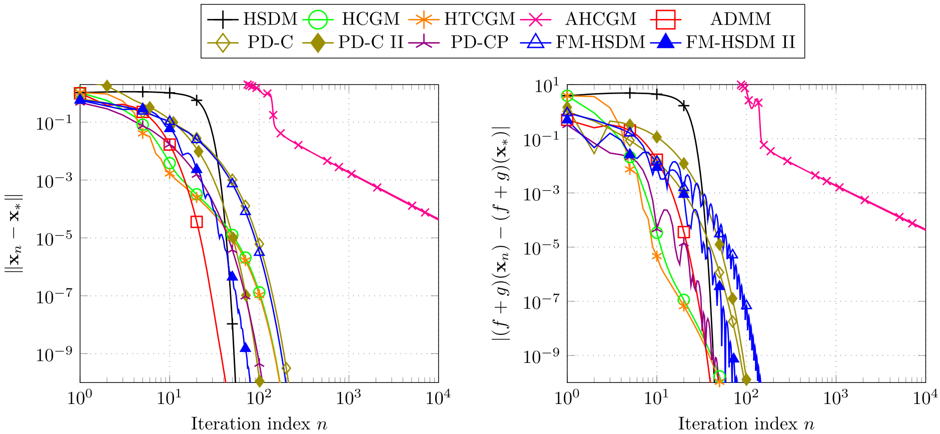

Figure 1. Deviation of the estimate $\mathbf{x}_n$ from the unique minimizer $\mathbf{x}_*$ of (65) and deviation of the loss-function value $(f+g)(\mathbf{x}_n)$ from the optimal $(f+g)(\mathbf{x}_*)$ vs iteration index n, in the case where $[\mathbf{P}]_{11} := 1$ and thus the condition number of $\mathbf{P}$ equals 10.

$(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}) \in \mathcal{X} \mid \mathbf{x}^{(1)} = \mathbf{x}^{(2)} = \mathbf{x}^{(3)}\}$ is a closed linear subspace and thus an affine set. According to Example A.1, a nonexpansive mapping T with $T \in \mathfrak{T}_{\mathcal{A}}$ is the metric projection mapping $P_{\mathcal{A}}(\mathbf{x}) = (1/3)(\sum_{i=1}^{3} \mathbf{x}^{(i)}, \sum_{i=1}^{3} \mathbf{x}^{(i)}, \sum_{i=1}^{3} \mathbf{x}^{(i)})$, $\forall \mathbf{x} := (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}) \in \mathcal{X}$.

Under the previous view of (65) as a special case of (1), FM-HSDM is compared with other HSDM family members such as the original HSDM [5], the hybrid conjugate gradient method (HCGM) [11], the hybrid three-term conjugate gradient method (HTCGM) [12] and the accelerated hybrid conjugate gradient method (AHCGM) [13]. Other competing methods include ADMM [21,22,24,25] in the standard 'scaled form' [23, §3.1.1], and the PD methods of 'CP-C' [15] and 'PD-CP' [14]. Due to the strongly convex nature of $\mathbf{x}^{(1)\top}\mathbf{P}\mathbf{x}^{(1)}$, the accelerated Alg. 2 of [14] with adaptive step sizes is used in 'PD-CP.'

To test (47) and address also the case where $[\mathbf{P}]_{11} \in \mathbb{R}_{>0}$ is close to zero (*cf.* Figure 2), i.e. $\mathbf{P}$ is 'nearly' singular, f and g can be considered in a different way than the previous setting: $f := 0$ and $g(\mathbf{x}) := (1/2)\mathbf{x}^{(1)\top}\mathbf{P}\mathbf{x}^{(1)} + \iota_{\mathcal{B}_1}(\mathbf{x}^{(2)}) + \iota_{\mathcal{B}_2}(\mathbf{x}^{(3)})$. Results that associate with this take on (65) as a special case of (1) and with FM-HSDM are shown in the subsequent figures under the tag 'FM-HSDM II'. The PD method of [15] is also adjusted to accommodate this view of (65), and the associated results are shown in Figures 1 and 2 under the tag of 'PD-C II'. It is worth stressing here that for this specific g, the proximal mapping $\mathrm{Prox}_{\lambda g}(\mathbf{x}) = ((\mathbf{I}_d + \lambda\mathbf{P})^{-1}\mathbf{x}^{(1)}, P_{\mathcal{B}_1}(\mathbf{x}^{(2)}), P_{\mathcal{B}_2}(\mathbf{x}^{(3)}))$. In other words, both PD-C II and FM-HSDM II use the resolvent $(\mathbf{I}_d + \gamma\mathbf{P})^{-1}$, for some adequate $\gamma \in \mathbb{R}_{>0}$, similarly to the case of ADMM and PD-CP.

Parameters in all methods were tuned to yield best performance. In all tests, methods start from the same initial point, randomly drawn from a unit-norm

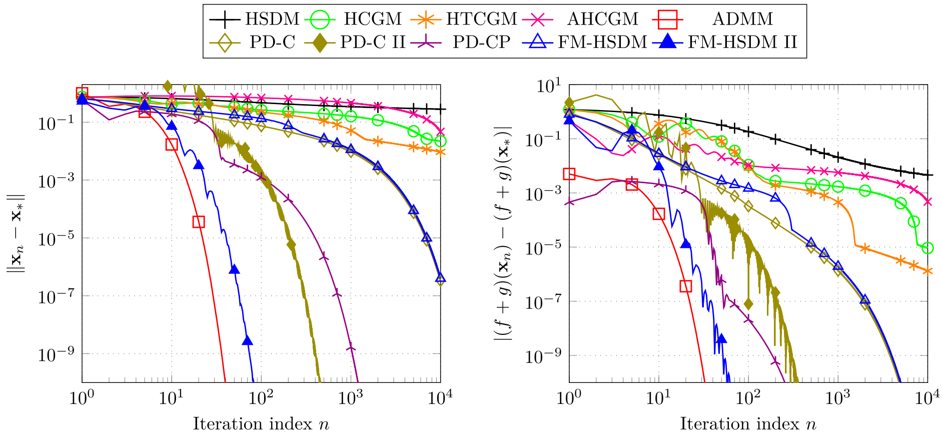

Figure 2. This setting follows that of Figure 1, but with $[\mathbf{P}]_{11} := 10^{-2}$, which results in a condition number $10/10^{-2} = 10^3$ for $\mathbf{P}$.

sphere and centred at the unique minimizer of (65). Each curve in Figures 1 and 2 is the uniform average of the curves obtained from 100 Monte-Carlo runs.

Figure 1 considers $[\mathbf{P}]_{11} := 1$, and since the largest entry of $\mathbf{P}$ is 10, the condition number of $\mathbf{P}$ is $10/1 = 10$. According to the developed theory, parameter λ of FM-HSDM is set equal to $\lambda := 0.99 \cdot 2(1-\alpha)/L$. Figure 1 shows that all methods, apart from AHCGM, perform similarly. All HSDM-family members, excluding FM-HSDM II, as well as PD-C score similar complexities since they use ∇f once per iteration. On the contrary, ADMM, PD-CP, PD-C II and FM-HSDM II do not utilize ∇f but build around the resolvent $(\mathbf{I}_d + \gamma\mathbf{P})^{-1}$ [3], for appropriate $\gamma \in \mathbb{R}_{>0}$.

The next set of tests follows that of Figure 1, but with $[\mathbf{P}]_{11} := 10^{-2}$, which yields the condition number $10/10^{-2} = 10^3$ for $\mathbf{P}$. As in the previous setting, parameter λ of FM-HSDM is set equal to $\lambda := 0.99 \cdot 2(1-\alpha)/L$. Notice that since the theory which associates with HSDM, HCGM, HTCGM and AHCGM offers guarantees of convergence in cases where f is strongly convex, i.e. $\mathbf{P}$ is positive definite, Figure 2 shows that the performance of the aforementioned algorithms degrades due to the fact that $\mathbf{P}$ was purposefully chosen to be 'nearly singular'. Figure 2 suggests also that FM-HSDM II pays the price, by using $(\mathbf{I}_d + \gamma\mathbf{P})^{-1}$, to achieve a performance similar to ADMM. The 'simpler' FM-HSDM and PD-C, where no matrix inversion is required, face difficulties in following the ADMM, FM-HSDM II, PD-C II and PD-CP curves for such an ill-conditioned minimization task. In theory, any $\lambda \in \mathbb{R}_{>0}$ can serve FM-HSDM due to the fact that $f := 0$. In practice, tuning is necessary, and the value of $\lambda = 100$ is used. Figure 2 underlines the flexibility of FM-HSDM, where mappings and computational complexity can be tuned to suit the minimization task at hand.

To compare (46) with (48), tests are performed on the following task:

$$\min_{\mathbf{x} \in \mathscr{X}_0} \ \mathbf{x}^\top \mathbf{P}\mathbf{x} \quad \text{s.t.}\ \mathbf{x} \in \mathcal{V} := \{\mathbf{u} \in \mathscr{X}_0 \mid \mathbf{e}_1^\top \mathbf{u} = 1\}, \tag{66}$$

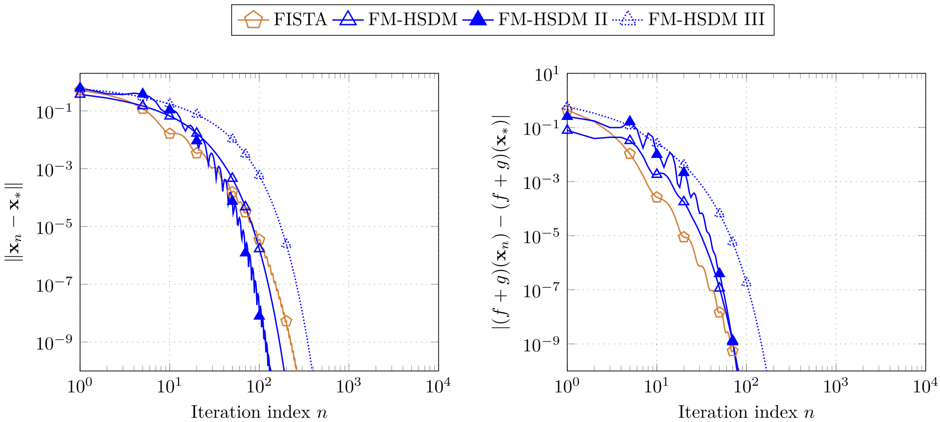

Figure 3. Deviation of the estimate $\mathbf{x}_n$ from the unique minimizer $\mathbf{x}_*$ of (66) and deviation of the loss-function value $(f+g)(\mathbf{x}_n)$ from the optimal $(f+g)(\mathbf{x}_*)$ vs iteration index n, in the case where $[\mathbf{P}]_{11} := 1$ and thus the condition number of $\mathbf{P}$ equals 10.

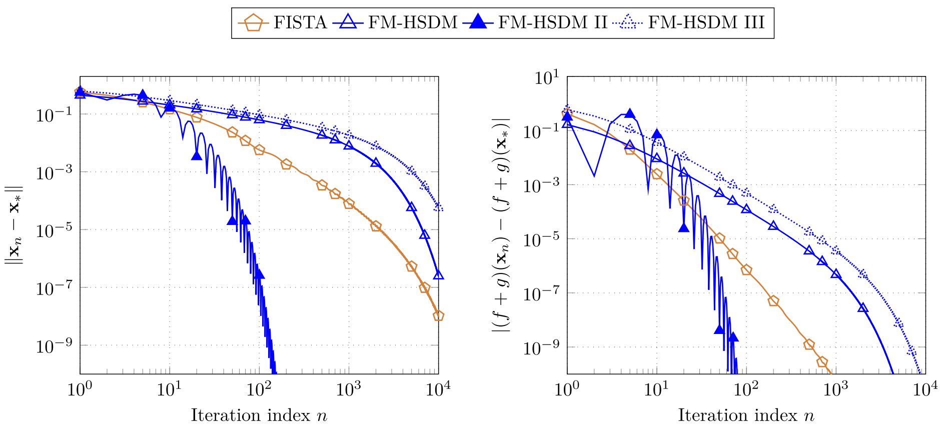

Figure 4. This setting follows that of Figure 3, but with $[\mathbf{P}]_{11} := 10^{-2}$, which results in a condition number $10/10^{-2} = 10^3$ for $\mathbf{P}$.

where $\mathscr{X}_0$, $\mathbf{P}$ and $\mathbf{e}_1$ were defined earlier in this section, and $\mathcal{V}$ is a hyperplane; hence, an affine set. Due to the construction of $\mathbf{P}$, it can be verified that the minimizer of (66) is $\mathbf{x}_* = \mathbf{e}_1$. Both (46) and (48) are employed with $T := P_{\mathcal{V}}$, where $P_{\mathcal{V}}$ stands for the metric projection mapping onto $\mathcal{V}$ (*cf.* Example A.2). The results of the application of (46) and (48) are illustrated in Figures 3 and 4 as 'FM-HSDM' and 'FM-HSDM III', respectively.

The state-of-the-art FISTA method [30, (4.1)–(4.3)] is also employed here after recasting (66) as $\min_{\mathbf{x}\in\mathscr{X}_0}(1/2)\mathbf{x}^\top\mathbf{P}\mathbf{x} + \iota_{\mathcal{V}}(\mathbf{x})$, where $\iota_{\mathcal{V}}$ stands for the indicator function of $\mathcal{V}$. This take on (66) opens also the door for (47), under $g(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) := g_1(\mathbf{x}^{(1)}) + g_2(\mathbf{x}^{(2)})$, $\forall(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) \in \mathscr{X}_0^2$, with $g_1(\mathbf{x}^{(1)}) := (1/2)\mathbf{x}^{(1)\top}\mathbf{P}\mathbf{x}^{(1)}, \forall\mathbf{x}^{(1)}$, $g_2 := \iota_{\mathcal{V}}$, and $\mathcal{A} := \{(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) \in \mathscr{X}_0^2 \mid \mathbf{x}^{(1)} = \mathbf{x}^{(2)}\}$, similarly to the application of FM-HSDM II to (65). Tag 'FM-HSDM II' is used also in Figures 3 and 4 to indicate the performance of (47). It is worth noticing that (48) can be applied to (66), but not to (65), due to the limitation of $g = 0$ in (48). Moreover, FISTA cannot be applied 'innocently' to (65), since its proximal-mapping step [30, (4.1)] amounts to identifying the metric projection of a point onto the intersection $\mathcal{B}_1 \cap \mathcal{B}_2$, which is itself the outcome of an iterative procedure, such as the projections-onto-convex-sets (POCS) algorithm [3, Cor. 5.23, p. 84]. Such computational issues would have been surmounted, had FISTA the ability to employ the convenient tool of 'splitting of variables', which is embedded in ADMM and PD methods, as well as in FM-HSDM via the affine constraint $\mathcal{A}$ [*cf.* (65)].

The way to construct $\mathbf{P}$ is identical to that in the case of (65). Parameters α (:= 0.5) and λ for FM-HSDM and FM-HSDM II are identical to those of the (65) scenario. The step size λ' of FM-HSDM III is defined as $\lambda' := 0.99 \cdot 2(1-\alpha)^2/L$, according to the specifications dictated by Theorem 3.6. In all tests, methods start from the same initial point, randomly drawn from a unit-norm sphere and centred at the unique minimizer of (66). Results are depicted in Figures 3 and 4, where each curve is the uniform average of the curves obtained from 100 Monte-Carlo runs. FM-HSDM III demonstrates slower convergence speed than that of the rest of the methods. Note that FISTA guarantees *optimal* convergence rate $|(f+g)(\mathbf{x}_n) - (f+g)(\mathbf{x}_*)| = \mathcal{O}[1/(n+1)^2]$ [30, Thm. 4.4]. The fast convergence speed of FM-HSDM II becomes prominent in the case of Figure 4, where $\mathbf{P}$ suffers a large condition number.

5. Conclusion

This paper introduced the FM-HSDM for solving affinely constrained composite minimization tasks in real Hilbert spaces. Only differential and proximal mappings are used to provide low-computational complexity recursions with enhanced flexibility towards the accommodation of affine constraints. The advocated scheme enjoys Fejér monotonicity, a constant step-size parameter across iterations, and minimal presuppositions on the smooth and non-smooth loss functions to establish weak, and under certain hypotheses, strong convergence to an optimal point. Results on the rate of convergence of the FM-HSDM's sequence of estimates were also presented. Numerical tests on synthetic data were also demonstrated to validate the theoretical findings. Thorough tests on noisy real data, which showcase the flexibility of the family of mappings $\mathfrak{T}_{\mathcal{A}}$ [*cf.* (7)] in a stochastic setting, are deferred to an upcoming publication.

Acknowledgments

The authors are grateful to a Referee for an insightful comment that led to Proposition 2.9.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was partially supported by the Division of Computing and Communication Foundations (NSF) awards 1514056, 1525194 and 1718796.

ORCID

Konstantinos Slavakis http://orcid.org/0000-0002-3370-3154

References

[1] Slavakis K, Yamada I. Accelerated hybrid steepest descent method for solving affinely constrained convex composite optimization problems. In: Intl. Conf. Continuous Optim. (ICCOPT); Aug.; Tokyo: Japan; 2016.

[2] Slavakis K, Yamada I, Ono S. Accelerating the hybrid steepest descent method for affinely constrained convex composite minimization tasks. In: Proc. ICASSP; Mar.; New Orleans; 2017

[3] Bauschke HH, Combettes PL. Convex Analysis and Monotone Operator Theory in Hilbert Spaces. New York: Springer; 2011.

[4] Kreyszig E. Introductory Functional Analysis and Applications. New York: Wiley; 1989.Wiley Classics Library ed.

[5] Yamada I. The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings. In: Butnariu D, Censor Y, Reich S, editors. Inherently parallel algorithms for feasibility and optimization and their applications. Elsevier; 2001. p. 473–504.

[6] Yamada I, Ogura N. Hybrid steepest descent method for variational inequality problem over the fixed point set of certain quasi-nonexpansive mappings. Numer Funct Anal Optim. 2004;25:619–655.

[7] Yamada I, Yukawa M, Yamagishi M. Minimizing the Moreau envelope of nonsmooth convex functions over the fixed point set of certain quasi-nonexpansive mappings. In: Bauschke HH, Burachik R, Combettes PL, et al., editors. Fixed-point algorithms for inverse problems in science and engineering. Springer; 2011. p. 345–390.

[8] Yamagishi M, Yamada I. Nonexpansiveness of a linearized augmented Lagrangian operator for hierarchical convex optimization. Inverse Probl. 2017;33:044003.

[9] Ogura N, Yamada I. Non-strictly convex minimization over the fixed point set of the asymptotically shrinking nonexpansive mapping. Numer Funct Anal Optim. 2002;23:113–137.

[10] Bauschke HH, Borwein JM. On projection algorithms for solving convex feasibility problems. SIAM Rev. 1996;38:367–426.

[11] Iiduka H, Yamada I. A use of conjugate gradient direction for the convex optimization problem over the fixed point set of a nonexpansive mapping. SIAM J Optim. 2009;19(4):1881–1893.

[12] Iiduka H. Three-term conjugate gradient method for the convex optimization problem over the fixed point set of a nonexpansive mapping. Appl Math Comput. 2011;217:6315–6327.

[13] Iiduka H. Acceleration method for convex optimization over the fixed point set of a nonexpansive mapping. Math Program. 2015;149:131–165.

[14] Chambolle A, Pock T. A first-order primal–dual algorithm for convex problems with applications to imaging. J Math Imaging Vision. 2011;40(1):120–145.
[15] Condat L. A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. J Optim Theory Appl. 2013;158(2):460–479.
[16] Combettes PL, Pesquet JC. Primal–dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. Set-Valued Var Anal. 2012;20(2):307–330.
[17] Vũ BC. A splitting algorithm for dual monotone inclusions involving cocoercive operators. Adv Comput Math. 2013;38(3):667–681.
[18] Davis D, Yin W. A three-operator splitting scheme and its optimization applications. 2015.
[19] Cevher V, Vũ CB, Yurtsever A. Stochastic forward-Douglas–Rachford splitting for monotone inclusions. EPFL; 2016. 215759.
[20] Necoara I, Patrascu A. A random coordinate descent algorithm for optimization problems with composite objective function and linear coupled constraints. Comput Optim Appl. 2014;57(2):307–337.
[21] Glowinski R, Marrocco A. Sur l'approximation par éléments finis et la résolution par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. Rev Francaise d'Aut Inf Rech Oper. 1975;9(2):41–76.
[22] Gabay D, Mercier B. A dual algorithm for the solution of nonlinear variational problems via finite-element approximations. Comp Math Appl. 1976;2:17–40.
[23] Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found Trends Mach Learn. 2011;3(1):1–122.
[24] Bredies K, Sun H. Preconditioned Douglas–Rachford splitting methods for convex-concave saddle-point problems. SIAM J Numer Anal. 2015;53(1):421–444.
[25] Sun H. Analysis of fully preconditioned ADMM with relaxation in Hilbert spaces. 2016.
[26] Shi W, Ling Q, Wu G, et al. EXTRA: An exact first-order algorithm for decentralized consensus optimization. SIAM J Optim. 2014;25(2):944–966.
[27] Shi W, Ling Q, Wu G, et al. A proximal gradient algorithm for decentralized composite optimization. IEEE Trans Signal Process. 2015;63(22):6013–6023.
[28] Combettes PL, Yamada I. Compositions and convex combinations of averaged nonexpansive operators. J Math Anal Appl. 2015;425:55–70.
[29] Baillon JB, Haddad G. Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones. Israel J Math. 1977;26(2):137–150.
[30] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J Imaging Sci. 2009;2(1):183–202.
[31] Rockafellar RT. Convex Analysis. Princeton, NJ: Princeton University Press; 1970.
[32] Ben-Israel A, Greville TNE. Generalized Inverses: Theory and Applications. 2nd ed. New York: Springer-Verlag; 2003.

Appendix

Several special cases of $\mathcal{A}$, of large interest in optimization tasks, together with members of the family of mappings $\mathfrak{T}_{\mathcal{A}}$ follow.

Example A.1: Given a Hilbert space $\mathscr{X}_0$ and $I \in \mathbb{Z}_{>0}$, consider the Hilbert space $\mathcal{X} := \mathscr{X}_0 \times \mathscr{X}_0 \times \cdots \times \mathscr{X}_0 = \{x := (x^{(1)}, x^{(2)}, \ldots, x^{(I)}) \mid x^{(i)} \in \mathscr{X}_0, \forall i \in \{1, \ldots, I\}\}$, equipped with the inner product $\langle x|x'\rangle_{\mathcal{X}} := \sum_{i=1}^{I} \langle x^{(i)}|x'^{(i)}\rangle$. Then, upon defining the (closed) linear subspace

$\mathcal{S} := \{x \in \mathcal{X} \mid x^{(1)} = x^{(2)} = \cdots = x^{(I)}\}$, the metric projection mapping onto $\mathcal{S}$ satisfies

$$P_{\mathcal{S}}(x) = \left(\frac{1}{I}\sum_{i=1}^{I} x^{(i)}, \frac{1}{I}\sum_{i=1}^{I} x^{(i)}, \ldots, \frac{1}{I}\sum_{i=1}^{I} x^{(i)}\right), \quad \forall x \in \mathcal{X}, \tag{A1}$$

and $P_{\mathcal{S}} \in \mathfrak{T}_{\mathcal{S}}$.

Proof: Formula (A1) can be easily derived by applying Example 2.6(i) to the special cases of $\mathcal{X}$ and $\mathcal{S}$: $\|x - P_{\mathcal{S}}x\|_{\mathcal{X}}^2 = \min_{z\in\mathscr{X}_0} \sum_{i=1}^{I} \|x^{(i)} - z\|^2$. Then, claim $P_{\mathcal{S}} \in \mathfrak{T}_{\mathcal{S}}$ is established by noticing that $\mathcal{S}$ is a closed affine set and by Proposition 2.11. ■

Example A.2 (Metric projection mapping onto a hyperplane): For a non-zero $a \in \mathcal{X}$ and a real number b, consider the metric projection mapping onto the hyperplane $\mathcal{V} := \{x \in \mathcal{X} \mid \langle a|x\rangle = b\}$ [3, (3.11), p. 49]

$$P_{\mathcal{V}} = \mathrm{Id} - \frac{\langle a|\mathrm{Id}\,\rangle}{\|a\|^2} a + \frac{b}{\|a\|^2} a\,. \tag{A2}$$

Then, $P_{\mathcal{V}} \in \mathfrak{T}_{\mathcal{V}}$.

Proof: The claim follows by the observations that $\mathcal{V}$ is a closed affine set, $(b/\|a\|^2)a \in \mathcal{V}$, and by introducing $\mathcal{V} = \{x \in \mathcal{X} \mid \langle a|x\rangle = 0\}$, with $P_{\mathcal{V}} = \mathrm{Id} - \frac{\langle a|\mathrm{Id}\,\rangle}{\|a\|^2} a$ and $P_{\mathcal{V}}[(b/\|a\|^2)a] = 0$, in Proposition 2.11. ■

As the following fact states, affine sets obtain a specific form in Euclidean spaces.

Fact A.3 ([31, Thm. 1.4, p. 5]): Given $\mathbf{b} \in \mathbb{R}^M$ ($M \in \mathbb{Z}_{>0}$) and $\mathbf{A} \in \mathbb{R}^{M\times D}$ ($D \in \mathbb{Z}_{>0}$) the set $\{\mathbf{x} \in \mathbb{R}^D \mid \mathbf{Ax} = \mathbf{b}\}$, if non-empty, is an affine set. Moreover, every affine set in $\mathcal{X} := \mathbb{R}^D$ can be represented in this way.

Motivated by the previous fact and aiming at an algorithmic scheme with wide applicability in Euclidean spaces, where most of the minimization problems reside, the following example and proposition offer a view of affine sets via *least-squares* (LS) tasks and nonexpansive mappings.

Example A.4 (Affinely constrained LS in Euclidean spaces): For vector $\mathbf{b}$ and matrix $\mathbf{A}$ of Fact A.3, consider the following LS solution set [3, Prop. 3.25, p. 50]:

$$\mathcal{A} := \mathrm{Argmin}_{\mathbf{x}\in\mathbb{R}^D} \tfrac{1}{2}\|\mathbf{Ax} - \mathbf{b}\|^2 = \{\mathbf{x} \in \mathbb{R}^D \mid \mathbf{A}^\top\mathbf{Ax} = \mathbf{A}^\top\mathbf{b}\}. \tag{A3}$$

Now, considering the $D \times 1$ vectors $\{\boldsymbol{\alpha}_m\}_{m=1}^M$, defined by the rows of $\mathbf{A}$, i.e. $[\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_M] := \mathbf{A}^\top$, as well as the $D \times 1$ vectors $\{\mathbf{g}_d\}_{d=1}^D$ defined via $[\mathbf{g}_1, \ldots, \mathbf{g}_D] := \mathbf{G}$, where $\mathbf{G} := \mathbf{A}^\top\mathbf{A}$ and $\mathbf{c} := [c_1, c_2, \ldots, c_D]^\top := \mathbf{A}^\top\mathbf{b}$, let the hyperplanes $\mathcal{A}_m := \{\mathbf{x} \in \mathbb{R}^D \mid \langle\boldsymbol{\alpha}_m|\mathbf{x}\rangle = b_m\}$, $(m = 1, \ldots, M)$, as well as $\mathcal{G}_d := \{\mathbf{x} \in \mathbb{R}^D \mid \langle\mathbf{g}_d|\mathbf{x}\rangle = c_d\}$, $(d = 1, \ldots, D)$, with associated metric projection mappings $P_{\mathcal{A}_m}$ and $P_{\mathcal{G}_d}$, respectively [*cf.* (A2)]. Then, any of the following mappings, with † denoting the Moore–Penrose pseudoinverse operation [32],

$$T = \begin{cases} \left(\mathbf{I} - \frac{\mu}{\varrho}\mathbf{A}^\top\mathbf{A}\right)\mathrm{Id} + \frac{\mu}{\varrho}\mathbf{A}^\top\mathbf{b}\,, & \varrho \geq \|\mathbf{A}\|^2, \mu \in (0,1]\,, \quad \text{(A4a)} \\ (\mathbf{I} - \mathbf{A}^\top\mathbf{A}^{\dagger\top})\mathrm{Id} + \mathbf{A}^\dagger\mathbf{b}\,, & \text{(A4b)} \\ (\mathbf{I} - \mathbf{G}\mathbf{G}^\dagger)\mathrm{Id} + \mathbf{G}^\dagger\mathbf{A}^\top\mathbf{b}\,, & \text{(A4c)} \\ (\mathbf{I} + \gamma\mathbf{A}^\top\mathbf{A})^{-1}\mathrm{Id} + \gamma(\mathbf{I} + \gamma\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{b}\,, & \gamma \in \mathbb{R}_{>0}\,, \quad \text{(A4d)} \\ (1-\beta)\mathrm{Id} + \beta\sum_{m=1}^{M}\frac{\|\boldsymbol{\alpha}_m\|^2}{\|\mathbf{A}\|_\mathrm{F}^2}P_{\mathcal{A}_m}\,, & \beta \in (0,1]\,, \quad \text{(A4e)} \\ (1-\theta)\mathrm{Id} + \theta\sum_{d=1}^{D}\omega_d P_{\mathcal{G}_d}\,, & \begin{cases}\theta \in (0,1]\,, \omega_d \in (0,1)\,, \\ \sum_{d=1}^{D}\omega_d = 1\,,\end{cases} \quad \text{(A4f)} \end{cases}$$

satisfies $T \in \mathfrak{T}_{\mathcal{A}}$.

Furthermore, given also the $M_0 \times 1$ ($M_0 \in \mathbb{Z}_{>0}$) vector $\mathbf{b}_0$, the $M_0 \times D$ matrix $\mathbf{A}_0$, let the non-empty affine constraint set $\mathcal{K} := \{\mathbf{x} \in \mathbb{R}^D \mid \mathbf{A}_0\mathbf{x} = \mathbf{b}_0\}$, with metric projection mapping $P_{\mathcal{K}} = (\mathbf{I} - \mathbf{A}_0^\top\mathbf{A}_0^{\dagger\top})\mathrm{Id} + \mathbf{A}_0^\dagger\mathbf{b}_0$ [3, Prop. 3.17, p. 47]. Then, according to [32, Ex. 34, p. 120],

$$\begin{aligned} &\mathbf{x} \in \mathcal{A}_{\mathcal{K}} := \mathrm{Argmin}_{\mathbf{z}\in\mathcal{K}}\, \tfrac{1}{2}\|\mathbf{A}\mathbf{z} - \mathbf{b}\|^2 \\ &\Leftrightarrow \exists\boldsymbol{\mu} \in \mathbb{R}^{M_0} \text{ s.t. } (\mathbf{x}, \boldsymbol{\mu}) \in \overline{\mathcal{A}} := \left\{ (\mathbf{x}', \boldsymbol{\mu}') \in \mathbb{R}^D \times \mathbb{R}^{M_0} \mid \overbrace{\begin{bmatrix} \mathbf{A}^\top\mathbf{A} & \mathbf{A}_0^\top \\ \mathbf{A}_0 & \mathbf{0} \end{bmatrix}}^{\mathbf{L}:=} \begin{bmatrix} \mathbf{x}' \\ \boldsymbol{\mu}' \end{bmatrix} = \overbrace{\begin{bmatrix} \mathbf{A}^\top\mathbf{b} \\ \mathbf{b}_0 \end{bmatrix}}^{\mathbf{e}:=} \right\}, \end{aligned} \tag{A5}$$

or, in other words, $\mathcal{A}_{\mathcal{K}} = \Pi_{\mathbb{R}^D}\overline{\mathcal{A}}$, where $\Pi_{\mathbb{R}^D}$ denotes the mapping $\Pi_{\mathbb{R}^D} : \mathbb{R}^D \times \mathbb{R}^{M_0} \to \mathbb{R}^D : (\mathbf{x}, \boldsymbol{\mu}) \mapsto \mathbf{x}$. Define also the $(D + M_0) \times 1$ vectors $[\mathbf{l}_1, \ldots, \mathbf{l}_{D+M_0}] := \mathbf{L}$, as well as the hyperplanes $\mathcal{L}_d := \{(\mathbf{x}', \boldsymbol{\mu}') \in \mathbb{R}^D \times \mathbb{R}^{M_0} \mid \langle \mathbf{l}_d | (\mathbf{x}', \boldsymbol{\mu}') \rangle = e_d\}$, with $P_{\mathcal{L}_d}$ denoting the associated metric projection mapping [*cf.* (A2)]. Then, any of the following mappings $\bar{T} : \mathbb{R}^{D+M_0} \to \mathbb{R}^{D+M_0}$:

$$\bar{T} = \begin{cases} \left(\mathbf{I} - \frac{\overline{\mu}}{\overline{\varrho}}\mathbf{L}^\top\mathbf{L}\right)\mathrm{Id} + \frac{\overline{\mu}}{\overline{\varrho}}\mathbf{L}^\top\mathbf{b}, & \overline{\varrho} \geq \|\mathbf{L}\|^2, \overline{\mu} \in (0,1], \quad \text{(A6a)} \\ \left(\mathbf{I} - \mathbf{L}^\top\mathbf{L}^{\dagger\top}\right)\mathrm{Id} + \mathbf{L}^\dagger\mathbf{e}, & \text{(A6b)} \\ \left(\mathbf{I} + \overline{\gamma}\mathbf{L}^\top\mathbf{L}\right)^{-1}\mathrm{Id} + \overline{\gamma}\left(\mathbf{I} + \overline{\gamma}\mathbf{L}^\top\mathbf{L}\right)^{-1}\mathbf{L}^\top\mathbf{e}, & \overline{\gamma} \in \mathbb{R}_{>0}, \quad \text{(A6c)} \\ (1-\overline{\theta})\mathrm{Id} + \overline{\theta}\sum_{d=1}^{D+M_0}\overline{w}_d P_{\mathcal{L}_d}, & \begin{cases}\overline{\theta} \in (0,1], \overline{\omega}_d \in (0,1), \\ \sum_{d=1}^{D+M_0}\overline{\omega}_d = 1,\end{cases} \quad \text{(A6d)} \end{cases}$$

satisfies $\bar{T} \in \mathfrak{T}_{\overline{\mathcal{A}}}$. Moreover, the mapping $T : \mathbb{R}^D \to \mathbb{R}^D$, defined by

$$T := (1 - \overline{\beta})P_{\mathcal{K}} + \overline{\beta}P_{\mathcal{K}}\sum_{m=1}^{M}\frac{\|\boldsymbol{\alpha}_m\|^2}{\|\mathbf{A}\|_\mathrm{F}^2}P_{\mathcal{A}_m}P_{\mathcal{K}}, \quad \overline{\beta} \in (0,1], \tag{A6e}$$

satisfies $T \in \mathfrak{T}_{\mathcal{A}_{\mathcal{K}}}$.

Proof: For $\delta \in \mathbb{R}_{>0}$, define

$$\varphi_\delta(\mathbf{x}) := \tfrac{1}{2\delta}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2, \quad \forall \mathbf{x} \in \mathbb{R}^D, \tag{A7}$$

and verify that $\nabla\varphi_\delta = (1/\delta)\mathbf{A}^\top\mathbf{A}\mathrm{Id} - (1/\delta)\mathbf{A}^\top\mathbf{b}$. According to (A3), all points $\mathbf{x} \in \mathbb{R}^D$ s.t. $\nabla\varphi_\delta(\mathbf{x}) = \mathbf{0}$ constitute $\mathcal{A}$. Moreover, for any $\varrho \geq \|\mathbf{A}\|^2/\delta$, $\|\nabla\varphi_\delta(\mathbf{x}) - \nabla\varphi_\delta(\mathbf{x}')\| \leq (\|\mathbf{A}\|^2/\delta)\|\mathbf{x}-\mathbf{x}'\| \leq \varrho\|\mathbf{x}-\mathbf{x}'\|$, $\forall \mathbf{x},\mathbf{x}' \in \mathbb{R}^D$, since $\|\mathbf{A}^\top\mathbf{A}\| = \|\mathbf{A}\|^2$. In other words, $\nabla\varphi_\delta$ is ϱ-Lipschitz continuous, which, according to the Baillon–Haddad theorem [29], [3, Cor. 18.16, p. 270], is equivalent to that $(1/\varrho)\nabla\varphi_\delta$ is firmly nonexpansive iff $\mathrm{Id} - (1/\varrho)\nabla\varphi_\delta$ is firmly nonexpansive [*cf.* Example 2.6(iii)] with fixed-point set equal to $\mathcal{A}$. By utilizing once again Example 2.6(iii), $R := 2[\mathrm{Id} - (1/\varrho)\nabla\varphi_\delta] - \mathrm{Id}$ is nonexpansive, and for any $\zeta \in (0,1]$, $R' := \zeta R + (1-\zeta)\mathrm{Id} = \mathrm{Id} - (2\zeta/\varrho)\nabla\varphi_\delta = \{\mathbf{I} - [2\zeta/(\varrho\delta)]\mathbf{A}^\top\mathbf{A}\}\mathrm{Id} + [2\zeta/(\varrho\delta)]\mathbf{A}^\top\mathbf{b}$ is nonexpansive with $\mathrm{Fix}\,(R') = \mathcal{A}$. Due to the nonexpansiveness of R', $\|\mathbf{I} - [2\zeta/(\varrho\delta)]\mathbf{A}^\top\mathbf{A}\| \leq 1$ (*cf.* Fact 2.8). Constraining $\zeta \in (0,1/2]$ guarantees that $\mathbf{I} - [2\zeta/(\varrho\delta)]\mathbf{A}^\top\mathbf{A} \succeq \mathbf{0}$. By defining $\mu := 2\zeta$ and $\delta := 1$, the claim regarding (A4a) is established.

The metric projection mapping $P_{\ker\mathbf{A}}$ onto $\ker\mathbf{A}$ is $P_{\ker\mathbf{A}} = (\mathbf{I} - \mathbf{A}^\top\mathbf{A}^{\top\dagger})\mathrm{Id}$ [3, Prop. 3.28(iii), p. 51]. Since $\mathcal{A} = \ker\mathbf{A} + \mathbf{A}^\dagger\mathbf{b}$ [3, Prop. 3.28(i), p. 51], [3, Prop. 3.17, p. 47] suggests that the metric projection mapping $P_{\mathcal{A}}$ onto $\mathcal{A}$ becomes $P_{\mathcal{A}} = P_{\ker\mathbf{A}} + \mathbf{A}^\dagger\mathbf{b} - P_{\ker\mathbf{A}}(\mathbf{A}^\dagger\mathbf{b}) = P_{\ker\mathbf{A}} + \mathbf{A}^\dagger\mathbf{b}$, due to $P_{\ker\mathbf{A}}(\mathbf{A}^\dagger\mathbf{b}) = \mathbf{0}$ [3, Prop. 3.28(i), p. 51]. Hence, (A4b) is an immediate consequence of Proposition 2.11. By [32, Ex. 18(d), p. 49], $\mathbf{A}^\top\mathbf{A}^{\top\dagger} = \mathbf{A}^\top\mathbf{A}(\mathbf{A}^\top\mathbf{A})^\dagger = \mathbf{G}\mathbf{G}^\dagger$ and $\mathbf{A}^\dagger\mathbf{b} = (\mathbf{A}^\top\mathbf{A})^\dagger\mathbf{A}^\top\mathbf{b} = \mathbf{G}^\dagger\mathbf{A}^\top\mathbf{b}$. Hence, (A4c) follows easily from (A4b).

Now, for any $\gamma' \in \mathbb{R}_{>0}$, $\mathrm{Prox}_{\gamma'\varphi_\delta} = (\mathbf{I} + (\gamma'/\delta)\mathbf{A}^\top\mathbf{A})^{-1}\mathrm{Id} + (\gamma'/\delta)(\mathbf{I} + (\gamma'/\delta)\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{b}$. Setting $\gamma := \gamma'/\delta$, the nonexpansiveness of $\mathrm{Prox}_{\gamma\delta\varphi_\delta}$, stated by Example 2.6(ii), suggests that $\|(\mathbf{I} + \gamma\mathbf{A}^\top\mathbf{A})^{-1}\| \leq 1$ (*cf.* Fact 2.8), and that $\mathrm{Fix}\,(\mathrm{Prox}_{\gamma\delta\varphi_\delta}) = \mathcal{A}$. Due also to the fact that $(\mathbf{I} + \gamma\mathbf{A}^\top\mathbf{A})^{-1}$ is positive, the claim regarding (A4d) is established.

Let $\delta := \|\mathbf{A}\|_{\mathrm{F}}^2$ in (A10), so that

$$\begin{aligned}\varphi_{\|\mathbf{A}\|_{\mathrm{F}}^2}(\mathbf{x}) &= \tfrac{1}{2\|\mathbf{A}\|_{\mathrm{F}}^2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2 = \tfrac{1}{2\|\mathbf{A}\|_{\mathrm{F}}^2}\sum_{m=1}^{M}(\langle\boldsymbol{\alpha}_m|\mathbf{x}\rangle - b_m)^2 \\ &= \tfrac{1}{2}\sum_{m=1}^{M}\tfrac{\|\boldsymbol{\alpha}_m\|^2}{\|\mathbf{A}\|_{\mathrm{F}}^2}\|\mathbf{x} - P_{\mathcal{A}_m}(\mathbf{x})\|^2 = \tfrac{1}{2}\sum_{m=1}^{M} w_m\|\mathbf{x} - P_{\mathcal{A}_m}(\mathbf{x})\|^2,\end{aligned}$$

where the explicit expression of $P_{\mathcal{A}_m}$ is given in (A2), and the non-negative weights $\{w_m := \|\boldsymbol{\alpha}_m\|^2/\|\mathbf{A}\|_{\mathrm{F}}^2\}_{m=1}^M$ satisfy $\sum_{m=1}^M w_m = 1$. It can be also verified by the Fréchet-gradient definition [3, Def. 2.45, p. 38] that $\nabla\|(\mathrm{Id} - P_{\mathcal{A}_m})\mathbf{x}\|^2 = 2(\mathrm{Id} - P_{\mathcal{A}_m})\mathbf{x}$, which yields

$$\nabla\varphi_{\|\mathbf{A}\|_{\mathrm{F}}^2} = \sum_{m=1}^{M} w_m(\mathrm{Id} - P_{\mathcal{A}_m}) = \mathrm{Id} - \sum_{m=1}^{M} w_m P_{\mathcal{A}_m}.$$

Hence, all minimizers of $\varphi_{\|\mathbf{A}\|_{\mathrm{F}}^2}$, i.e. $\mathcal{A}$, constitute the fixed-point set of $\sum_m w_m P_{\mathcal{A}_m}$, which is equal to the fixed-point set of the mapping in (A4e). Hence, by utilizing the trivial fact $\mathrm{Id} \in \mathfrak{T}$ and by applying also Proposition 2.10(i) to $(1-\beta)\mathrm{Id} + \beta\sum_m w_m P_{\mathcal{A}_m}$, the claim of (A4e) is established.

Regarding (A6e), notice first that $\mathcal{A} = \cap_{d=1}^D \mathcal{G}_d$. According to Example 2.6(iv), $\mathcal{A} = \mathrm{Fix}\,(\sum_d \omega_d P_{\mathcal{G}_d})$. Since $P_{\mathcal{G}_d} \in \mathfrak{T}$ (*cf.* Example A.2), Proposition 2.10(i) yields $\sum_d \omega_d P_{\mathcal{G}_d} \in \mathfrak{T}$. As a result, fact $\mathrm{Id} \in \mathfrak{T}$ and Proposition 2.10(i) yield $(1-\theta)\mathrm{Id} + \theta\sum_d \omega_d P_{\mathcal{G}_d} \in \mathfrak{T}$, which establishes the claim of (A4f). Due to $\overline{\mathcal{A}} = \arg\min_{(\mathbf{x},\boldsymbol{\mu})}\|\mathbf{L}[\mathbf{x}^\top, \boldsymbol{\mu}^\top]^\top - \mathbf{e}\|^2$, arguments similar to those developed for (A4a), (A4b) and (A4d) yield (A6a), (A6b) and (A6c), respectively. Furthermore, notice that since $\overline{\mathcal{A}} = \cap_{d=1}^{D+M_0}\mathcal{L}_d$, (A6d) is deduced in a way similar to the derivation of (A4f) from (A3).

Regarding (A9a), notice that $\mathcal{A}_{\mathcal{K}} = \operatorname{Fix} T_{\mathcal{A}_{\mathcal{K}}}$ [5, Prop. 4.2(a)], where

$$T_{\mathcal{A}_{\mathcal{K}}} := (1-\overline{\beta})\mathrm{Id} + \overline{\beta} P_{\mathcal{K}} \sum_{m=1}^{M} \frac{\|\boldsymbol{\alpha}_m\|^2}{\|\mathbf{A}\|_{\mathrm{F}}^2} P_{\mathcal{A}_m}$$

is nonexpansive for $\overline{\beta} \in (0, 3/2]$. Since $\mathcal{A}_{\mathcal{K}} = \operatorname{Fix} T_{\mathcal{A}_{\mathcal{K}}} = \operatorname{Fix} T_{\mathcal{A}_{\mathcal{K}}} \cap \mathcal{K} = \operatorname{Fix} T_{\mathcal{A}_{\mathcal{K}}} \cap \operatorname{Fix} P_{\mathcal{K}}$, Example 2.6(v) suggests that $\mathcal{A}_{\mathcal{K}}$ can be seen also as the fixed-point set of the nonexpansive mapping $T_{\mathcal{A}_{\mathcal{K}}} P_{\mathcal{K}}$, which is nothing but the mapping appearing at (A9a). Now, due to Proposition 2.10(i) and Example A.2, $\sum_m w_m P_{\mathcal{A}_m} \in \mathfrak{T}$, with $w_m := \|\boldsymbol{\alpha}_m\|^2/\|\mathbf{A}\|_{\mathrm{F}}^2$. Hence, Proposition 2.10(ii) suggests also that $P_{\mathcal{K}}(\sum_m w_m P_{\mathcal{A}_m}) P_{\mathcal{K}} \in \mathfrak{T}$. Once again, since $P_{\mathcal{K}} \in \mathfrak{T}$ (*cf.* Proposition 2.11), Proposition 2.10(i) guarantees $(1-\overline{\beta})P_{\mathcal{K}} + \overline{\beta} P_{\mathcal{K}} \sum_m w_m P_{\mathcal{A}_m} P_{\mathcal{K}} \in \mathfrak{T}$, for $\overline{\beta} \in (0, 1]$, which establishes the claim of (A9a). ■

An auxiliary proposition, used in Theorem 3.7, follows.

Proposition A.5: *Given the surjective and strongly positive mapping* $\Pi \in \mathfrak{B}(\mathcal{X})$, *i.e. there exists* $\delta \in \mathbb{R}_{>0}$ *s.t.* $\langle \Pi x | x \rangle \geq \delta \|x\|^2$, $\forall x \in \mathcal{X}$, *the inverse* Π^{-1} *exists and* $\Pi^{-1} \in \mathfrak{B}(\mathcal{X})$ *with* $\|\Pi^{-1}\| \leq 1/\delta$. *Moreover,* Π^{-1} *is strongly positive and* $(\delta/\|\Pi\|^2)\|x\|^2 \leq \langle \Pi^{-1} x | x \rangle \leq (1/\delta)\|x\|^2$, $\forall x \in \mathcal{X}$.

Proof: [4, Sec. 2.7, Prob. 7, p. 101] guarantees the existence of Π^{-1} and $\Pi^{-1} \in \mathfrak{B}(\mathcal{X})$. By the strong positivity of Π, $\forall x \in \mathcal{X} \setminus (\{0\} = \ker \Pi^{-1})$, $\|\Pi^{-1}x\|^2 \leq (1/\delta)\langle \Pi^{-1}x | \Pi(\Pi^{-1}x)\rangle = (1/\delta)\langle \Pi^{-1}x|x\rangle \leq (1/\delta)\|\Pi^{-1}x\|\|x\| \Rightarrow \|\Pi^{-1}x\| \leq (1/\delta)\|x\| \Rightarrow \|\Pi^{-1}\| \leq (1/\delta)$. By [4, Thm. 9.4-2, p. 476] and the previous result, $\forall x \in \mathcal{X}$, $\langle \Pi^{-1}x|x\rangle \leq \|\Pi^{-1}\|\|x\|^2 \leq (1/\delta)\|x\|^2$. Moreover, $\forall x' \in \mathcal{X}$, $\langle \Pi x' | \Pi^{-1}\Pi x'\rangle = \langle \Pi x'|x'\rangle \geq \delta\|x'\|^2 \geq (\delta/\|\Pi\|^2)\|\Pi x'\|^2$, which yields, under $x := \Pi x'$, that $\forall x \in \mathcal{X}$, $(\delta/\|\Pi\|^2)\|x\|^2 \leq \langle \Pi^{-1}x|x\rangle$. ■