

# Optimal Acceleration for Proximal Minimization of the Sum of Convex and Strongly Convex Functions

Govind M. Chari · Uijeong Jang · Ernest K. Ryu · Behçet Açıkmüşe

Received: date / Accepted: date

**Abstract** When minimizing the sum of a convex and a strongly convex function, or when finding the zero of the sum of a monotone operator and a strongly monotone operator, Chambolle and Pock (2010) and Davis and Yin (2015) proposed accelerated mechanisms that achieve an  $\mathcal{O}(1/N^2)$  convergence rate for the squared distance to the solution, but the optimality of this rate was not established. In this work, we present Fast Douglas–Rachford Splitting (FDR), an accelerated method that improves the constants established in the prior works, and provide a complexity lower bound establishing that both the  $\mathcal{O}(1/N^2)$  convergence rate and the leading-order constant of FDR’s rate are optimal.

## 1 Introduction

We consider the composite minimization problem

$$\underset{x \in \mathbb{R}^d}{\text{minimize}} \quad f(x) + g(x) \tag{1}$$

where  $f: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$  is closed, convex, and proper and  $g: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$  is closed, proper, and  $\mu$ -strongly convex. When  $f$  and  $g$  are *proximable*, i.e., their proximal operators can be evaluated efficiently, Douglas–Rachford splitting (DRS) [7, 16]

$$\begin{aligned} x_{k+1/2} &= \text{PROX}_{\alpha g}(z_k) \\ x_{k+1} &= \text{PROX}_{\alpha f}(2x_{k+1/2} - z_k) \\ z_{k+1} &= z_k + x_{k+1} - x_{k+1/2}, \end{aligned}$$

---

Govind M. Chari  
University of Washington, Seattle  
E-mail: gchari@uw.edu

Uijeong Jang  
University of California, Los Angeles  
E-mail: uijeongjang@math.ucla.edu

Ernest K. Ryu  
University of California, Los Angeles  
E-mail: eryl@math.ucla.edu

Behçet Açıkmüşe  
University of Washington, Seattle  
E-mail: behcet@uw.edu

with stepsize  $\alpha > 0$ , is a widely used solution method.

While convergence of DRS only requires  $g$  to be convex, strong convexity of  $g$  produces the rate

$$\|x_N - x_\star\|^2 = \mathcal{O}(1/N).$$

The problem setup can be further generalized to the monotone inclusion problem

$$\underset{x \in \mathbb{R}^d}{\text{find}} \quad 0 \in (\mathbf{A} + \mathbf{B})x \tag{2}$$

where  $\mathbf{A}$  is maximal monotone and  $\mathbf{B}$  is maximal monotone and  $\mu$ -strongly monotone.

For a maximal monotone operator  $\mathbf{A}$ , let

$$\mathbf{J}_{\alpha\mathbf{A}} = (\mathbf{I} + \alpha\mathbf{A})^{-1}$$

denote its resolvent. Douglas–Rachford splitting applied to (2), with stepsize  $\alpha > 0$ , is

$$\begin{aligned} x_{k+1/2} &= \mathbf{J}_{\alpha\mathbf{B}}(z_k) \\ x_{k+1} &= \mathbf{J}_{\alpha\mathbf{A}}(2x_{k+1/2} - z_k) \\ z_{k+1} &= z_k + x_{k+1} - x_{k+1/2}, \end{aligned}$$

and attains the same  $\mathcal{O}(1/N)$  rate under  $\mu$ -strong monotonicity of  $\mathbf{B}$ .

In optimization, an *acceleration* is a modification of a base algorithm that improves the worst-case convergence rate of a given performance measure. While the classical Nesterov acceleration [18] achieves the optimal accelerated rate for function-value suboptimality, the modern theory of accelerations considers a broader range of performance measures, including distance to the solution and the squared fixed-point residual. In particular, the Chambolle–Pock [4] and Davis–Yin splitting [6] methods admit accelerated variants—based on mechanisms distinct from Nesterov’s acceleration—that can be applied to Problem (1) to obtain an accelerated rate of  $\mathcal{O}(1/N^2)$  for the squared distance to the optimal solution. However, it was unknown whether this  $\mathcal{O}(1/N^2)$  rate is optimal for this problem class, or whether the leading constants achieved by these accelerated methods are tight.

*Contributions.* This work presents two main contributions. First, we present an improved accelerated algorithm for Problems (1) and (2) whose leading constant improves upon those of the accelerated Chambolle–Pock and Davis–Yin methods by a factor of 4. Specifically, the iterates satisfy

$$\|x_N - x_\star\|^2 \leq \frac{\|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2}{1 + 4N^2\mu^2} \sim \frac{\|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2}{4N^2\mu^2},$$

where  $x_\star$  denotes an optimal solution,  $u_\star$  is a corresponding dual solution,  $x_0$  and  $u_0$  are the primal and dual initializations respectively,  $\mu$  is the strong convexity parameter, and  $N$  is the iteration count. Second, we establish a matching complexity lower bound showing that both the  $\mathcal{O}(1/N^2)$  convergence rate and the leading-order term  $(\|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2)/4N^2\mu^2$  are optimal for this problem class.

## 1.1 Notation and preliminaries

Throughout,  $\mathbb{N}$  denotes the set of natural numbers,  $\mathbb{R}^d$  the  $d$ -dimensional Euclidean space,  $\langle \cdot, \cdot \rangle$  the standard inner product, and  $\|\cdot\|$  the induced norm.

We recall standard definitions from convex analysis and monotone operator theory [23, 1, 24].

A function  $f: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$  is *closed* if its epigraph, defined as

$$\text{epi } f = \{(x, t) \in \mathbb{R}^d \times \mathbb{R} \mid f(x) \leq t\},$$

is a closed set. It is *convex* if

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y) \quad \forall x, y \in \mathbb{R}^d \text{ and } \theta \in [0, 1],$$

and *proper* if it never takes the value  $-\infty$  and has finite value for some  $x_0 \in \mathbb{R}^d$ .

The *subdifferential* of  $f$  at  $x$  is

$$\partial f(x) = \{v \in \mathbb{R}^d \mid f(y) \geq f(x) + \langle v, y - x \rangle, \forall y \in \mathbb{R}^d\}.$$

We write  $f'(x) \in \partial f(x)$  to denote a subgradient.

A function  $g: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$  is  $\mu$ -strongly convex if  $g(x) - \frac{\mu}{2}\|x\|^2$  is convex, or equivalently, if for all  $x, y \in \mathbb{R}^d$  and all  $v \in \partial g(x)$ ,

$$g(y) \geq g(x) + \langle v, y - x \rangle + \frac{\mu}{2}\|y - x\|^2.$$

If  $f$  is closed, convex, and proper, then its subdifferential  $\partial f$  is a *monotone operator*, which means that

$$\langle v - w, x - y \rangle \geq 0 \quad \forall x, y \in \mathbb{R}^d, \forall v \in \partial f(x), \forall w \in \partial f(y).$$

Similarly, if  $g$  is closed, proper, and  $\mu$ -strongly convex, then  $\partial g$  is a  $\mu$ -strongly monotone operator, meaning that

$$\langle v - w, x - y \rangle \geq \mu\|x - y\|^2 \quad \forall x, y \in \mathbb{R}^d, \forall v \in \partial g(x), \forall w \in \partial g(y).$$

For a function  $f: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$ , define

$$\operatorname{argmin}_{x \in \mathbb{R}^d} f(x) = \{z \in \mathbb{R}^d \mid f(z) \leq f(x) \quad \forall x \in \mathbb{R}^d\}.$$

For a closed, convex, and proper function  $f$  and  $\gamma > 0$ , the proximal operator  $\operatorname{prox}_{\gamma f}: \mathbb{R}^d \rightarrow \mathbb{R}^d$  is

$$\operatorname{prox}_{\gamma f}(x) = \operatorname{argmin}_{z \in \mathbb{R}^d} \left\{ f(z) + \frac{1}{2\gamma}\|z - x\|^2 \right\}$$

From the optimality conditions of this minimization problem, if  $y = \operatorname{prox}_{\gamma f}(x)$ , then

$$0 \in \partial f(y) + \frac{1}{\gamma}(y - x).$$

Equivalently,

$$y = x - \gamma f'(y).$$

Consider the primal-dual problem pair

$$\underset{x \in \mathbb{R}^d}{\text{minimize}} \quad f(x) + g(x) \tag{\mathcal{P}}$$

$$\underset{u \in \mathbb{R}^d}{\text{maximize}} \quad -f^*(-u) - g^*(u) \tag{\mathcal{D}}$$

where  $f: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$  is closed, convex, and proper, and  $g: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$  is closed, proper, and  $\mu$ -strongly convex. Assume that the primal problem ( $\mathcal{P}$ ) admits a minimizer  $x_*$ , which is unique by strong convexity of  $g$ , and total duality holds. Let  $u_* \in \partial g(x_*)$  denote a corresponding dual solution to ( $\mathcal{D}$ ). Then necessarily  $-u_* \in \partial f(x_*)$  [24].

**Lemma 1** *If  $(x_*, u_*)$  is a primal-dual solution (i.e.  $u_* \in \partial g(x_*)$  and  $-u_* \in \partial f(x_*)$ ), then  $x_* = \text{prox}_{\gamma g}(x_* + \gamma u_*)$  and  $x_* = \text{prox}_{\gamma f}(x_* - \gamma u_*)$  for  $\gamma > 0$ .*

*Proof* We can verify the equalities by writing optimality conditions for the proximal operator

$$\begin{aligned} 0 &\in \partial g(x_*) + \frac{1}{\gamma}(x_* - (x_* + \gamma u_*)) \\ &= \partial g(x_*) - u_*. \end{aligned}$$

Since  $u_* \in \partial g(x_*)$ , the inclusion holds.

$$\begin{aligned} 0 &\in \partial f(x_*) + \frac{1}{\gamma}(x_* - (x_* - \gamma u_*)) \\ &= \partial f(x_*) + u_*. \end{aligned}$$

Since  $-u_* \in \partial f(x_*)$ , the inclusion holds.

## 1.2 Prior accelerations

We quickly review other acceleration mechanisms that have been published in prior works. Note that each of these acceleration mechanisms is distinct in the sense that the metric they accelerate differs.

*Nesterov-type Acceleration.* The classical Nesterov accelerated gradient method (NAG) reduces the function-value suboptimality at an accelerated  $\mathcal{O}(1/N^2)$  rate [18]. In the modern literature, the Optimized Gradient Method (OGM) [10, 14] improves the leading constant of NAG by a factor of 2 and attains the exact optimal worst-case rate for this setup with an exact matching complexity lower bound [8].

In the composite minimization setup where one function is smooth and the other is proximable, NAG can be extended to FISTA [2] and OGM to OptISTA [12]. Analogously, OptISTA is exactly optimal for the smooth composite convex minimization setup and improves the leading constant of FISTA by a factor of 2.

In Problem (1), however, neither function is assumed to be smooth, so Nesterov-type acceleration does not apply directly. To clarify the distinction, suppose instead that  $g$  were  $L$ -smooth rather than  $\mu$ -strongly convex. Then FISTA applied to Problem (1) is

$$\begin{aligned} x_{k+1} &= \text{prox}_{f/L} \left( y_k - \frac{1}{L} \nabla g(y_k) \right) \\ y_{k+1} &= x_{k+1} + \left( \frac{t_k - 1}{t_{k+1}} \right) (x_{k+1} - x_k) \end{aligned} \tag{FISTA}$$

where

$$y_1 = x_1 \in \mathbb{R}^d, \quad t_1 = 1, \quad t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}.$$

**Fact 1** *If  $\{x_k\}$  is generated by FISTA, then for any  $N \geq 1$ , the following bound is satisfied for any minimizer  $x_*$  [2, Theorem 4.4]*

$$f(x_N) + g(x_N) - f(x_*) - g(x_*) \leq \frac{2L\|x_1 - x_*\|^2}{N^2}$$

Nesterov acceleration can be applied to the corresponding dual of (1), since the dual involves the sum of a convex and a smooth, convex function. However, the quantity that is reduced at an accelerated rate in this case is the dual-function-value suboptimality, and this is different from the distance to the solution.

*Halpern Acceleration.* The Halpern acceleration [11], also referred to as the anchor acceleration in the modern literature, reduces the squared fixed-point residual at an accelerated  $\mathcal{O}(1/N^2)$  rate [13,15]. Specifically, suppose  $\mathbf{T}: \mathbb{R}^d \rightarrow \mathbb{R}^d$  is a non-expansive operator. Then, the optimized Halpern method (OHM)

$$x_{k+1} = \frac{1}{k+2}x_0 + \frac{k+1}{k+2}\mathbf{T}x_k$$

exhibits the rate

$$\|\mathbf{T}x_{N-1} - x_{N-1}\|^2 \leq \frac{4\|x_0 - x_\star\|^2}{N^2},$$

where  $x_\star \in \text{Fix } \mathbf{T}$  (i.e.  $x_\star = \mathbf{T}x_\star$ ).

OHM can be applied directly to Problem (1) by taking  $\mathbf{T}$  to be the Peaceman–Rachford splitting (PRS) operator. This algorithm is

$$\begin{aligned} y_{k+1} &= 2\text{prox}_{\alpha f}(x_k) - x_k \\ x_{k+1} &= \frac{1}{k+2}x_0 + \frac{k+1}{k+2}(2\text{prox}_{\alpha g}(y_{k+1}) - y_{k+1}), \end{aligned} \tag{OHM}$$

where  $\alpha > 0$ . This yields the rate

$$\|\mathbf{T}_{\text{PRS}}x_{N-1} - x_{N-1}\|^2 = \mathcal{O}(1/N^2),$$

where  $\mathbf{T}_{\text{PRS}} = (2\text{prox}_{\alpha g} - \mathbf{I}) \circ (2\text{prox}_{\alpha f} - \mathbf{I})$  is the PRS operator.

However, the quantity that is reduced at an accelerated rate in this case is the fixed-point residual of the PRS operator, and this is different from the distance to the solution.

*Chambolle–Pock–Davis–Yin Acceleration.* The accelerated Chambolle and Pock method [4, Alg. 2] applied to (1) is

$$\begin{aligned} u_{k+1} &= u_k - \sigma_k z_k + \sigma_k \text{prox}_{f/\sigma_k}(z_k - (1/\sigma_k)u_k) \\ x_{k+1} &= \text{prox}_{\tau_k g}(x_k + \tau_k u_{k+1}) \\ z_{k+1} &= x_{k+1} + \theta_k(x_{k+1} - x_k), \end{aligned} \tag{CP}$$

where

$$\tau_{k+1} = \theta_k \tau_k, \quad \sigma_{k+1} = \sigma_k / \theta_k, \quad \theta_k = \frac{1}{\sqrt{1 + 2\mu\tau_k}},$$

and  $\tau_0, \sigma_0 > 0$  with  $\tau_0\sigma_0 \leq 1$ , and  $z_0 = x_0$ . Similarly, the accelerated Davis–Yin method [6, Alg. 2] applied to (1) is

$$\begin{aligned} x_0 &= \text{prox}_{\gamma_0 g}(y_0) \\ u_0 &= \frac{1}{\gamma_0}(y_0 - x_0) \\ x_k &= \text{prox}_{\gamma_{k-1} g}(y_{k-1} + \gamma_{k-1}u_{k-1}) \\ u_k &= \frac{1}{\gamma_{k-1}}(y_{k-1} + \gamma_{k-1}u_{k-1} - x_k) \\ y_k &= \text{prox}_{\gamma_k f}(x_k - \gamma_k u_k), \end{aligned} \tag{DYS}$$

where  $y_0 \in \mathbb{R}^d$  is an initialization, and the proximal stepsizes are given by

$$\gamma_{k+1} = \frac{\gamma_k}{\sqrt{1 + 2\gamma_k\mu}} \text{ for } k = 0, \dots, N-1,$$

where  $\gamma_0 \in (0, \infty)$ .

**Fact 2** *If  $\{x_k\}$  is generated by CP with  $\tau_0\sigma_0 = 1$ , then for any  $\epsilon > 0$ , there exists  $N_0$  such that for all  $N \geq N_0$  [4, Theorem 2],*

$$\|x_N - x_\star\|^2 \leq \frac{1 + \epsilon}{N^2} \left( \frac{\|x_0 - x_\star\|^2}{\mu^2\tau_0^2} + \frac{\|u_0 - u_\star\|^2}{\mu^2} \right),$$

where  $x_\star$  is the minimizer of Problem (1) and  $u_\star \in \partial g(x_\star)$ , equivalently  $-u_\star \in \partial f(x_\star)$ , is a corresponding dual solution.

**Fact 3** *If  $\{x_k\}$  is generated by DYS and  $x_\star$  is the minimizer of Problem (1), then  $\|x_N - x_\star\|^2 = \mathcal{O}(1/N^2)$  [6, Theorem 1.2].*

Conceptually, our method in Section 2 can be viewed as a Chambolle–Pock–Davis–Yin-type acceleration, but further refined to achieve the optimal constant in the leading order term.

## 2 Fast Douglas–Rachford Splitting

We now present Fast Douglas–Rachford Splitting (FDR):

$$\begin{aligned} w_0 &= x_0 - \eta_0 u_0 \\ y_{k+1} &= \text{prox}_{\eta_k g}(2x_k - w_k) \\ w_{k+1} &= \left(1 + \frac{\eta_{k+1}}{\eta_k}\right) y_{k+1} - \left(\frac{\eta_{k+1}}{\eta_k}\right) (2x_k - w_k) \\ x_{k+1} &= \text{prox}_{\eta_{k+1} f}(w_{k+1}) \end{aligned} \tag{FDR}$$

for  $k = 0, \dots, N-1$ , where  $N \in \mathbb{N}$  is a pre-specified total iteration count,  $x_0, u_0 \in \mathbb{R}^d$  are respectively primal and dual starting points, and the proximal stepsizes are given by

$$\eta_k = \frac{2N\mu}{1 + 4kN\mu^2} \quad \text{for } k = 0, \dots, N.$$

In Section 2.1, we establish the following convergence guarantee for FDR.

**Theorem 1** *Let  $N \geq 1$  and let  $f: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$  be closed, convex, and proper, and  $g: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$  be closed, proper, and  $\mu$ -strongly convex with  $\mu > 0$ . Assume that a primal solution  $x_\star \in \text{argmin}(f + g)$  and a dual solution  $u_\star$  satisfying  $u_\star \in \partial g(x_\star)$ ,  $-u_\star \in \partial f(x_\star)$  exist. Then, the iterates generated by FDR satisfy the rate*

$$\|x_N - x_\star\|^2 \leq \frac{\|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2}{1 + 4N^2\mu^2}.$$

*Discussion.* For FDR, the total number of iterations  $N$  must be specified *a priori*, as the proximal step-size sequence  $\{\eta_k\}_{k=0}^N$  explicitly depends on  $N$ . Additionally, Theorem 1 provides a guarantee on the final iterate,  $x_N$ . Both properties are shared by exact optimal methods such as OGM and OptISTA.

We observe that if the step-sizes are constant ( $\eta_0 = \eta_1 = \dots = \eta_N$ ), our algorithm reduces to Peaceman–Rachford splitting up to our choice of initialization [22].

Moreover, as  $N, k \rightarrow \infty$  the ratio  $\eta_{k+1}/\eta_k \rightarrow 1$ , so the later FDR iterations resemble Peaceman–Rachford splitting with varying, decreasing proximal stepsizes.

Theorem 1 shows that FDR improves the leading asymptotic constant of the accelerated Chambolle–Pock and Davis–Yin methods by a factor of four. Details are provided in Appendix A.

## 2.1 Proof of Theorem 1

We establish the convergence rate of FDR with a Lyapunov analysis, which is a widely used proof template in the analysis of first-order methods [12, 14, 20, 21]. For notational simplicity, define

$$R^2 = \|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2.$$

Specifically, define

$$\begin{aligned} \mathcal{V}_{-1} &= \nu R^2 - \nu^2 \| -2N\mu(x_0 - x_\star) + (u_0 - u_\star) \|^2 \\ \mathcal{V}_0 &= \nu^2 \|x_\star + 2N\mu u_\star - x_0 - 2N\mu u_0\|^2 \\ \mathcal{V}_k &= \nu^2 \|(1 + 4kN\mu^2)x_\star + 2N\mu u_\star - (1 + 4kN\mu^2)(2x_k - w_k)\|^2 \quad \forall k = 1, \dots, N-1 \\ \mathcal{V}_N &= \|x_N - x_\star\|^2 + 4N^2\mu^2\nu^2 \|u_\star + f'(x_N)\|^2, \end{aligned}$$

where  $x_k$  and  $w_k$  are generated by (FDR),  $f'(x_N) = (w_N - x_N)/\eta_N$ , and

$$\nu = \frac{1}{1 + 4N^2\mu^2}.$$

Once we establish that the Lyapunov sequence is dissipative, i.e., that

$$\mathcal{V}_N \leq \mathcal{V}_{N-1} \leq \dots \leq \mathcal{V}_1 \leq \mathcal{V}_0 \leq \mathcal{V}_{-1},$$

it follows immediately that

$$\|x_N - x_\star\|^2 \leq \mathcal{V}_N \leq \dots \leq \mathcal{V}_{-1} \leq \nu R^2 = \frac{\|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2}{1 + 4N^2\mu^2},$$

which implies the stated rate

$$\|x_N - x_\star\|^2 \leq \frac{\|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2}{1 + 4N^2\mu^2}.$$

It remains to show that the Lyapunov sequence is non-increasing.

Case 1:  $\mathcal{V}_0 \leq \mathcal{V}_{-1}$

$$\begin{aligned} \mathcal{V}_0 - \mathcal{V}_{-1} &= \nu^2 \|x_\star + 2N\mu u_\star - x_0 - 2N\mu u_0\|^2 - \nu R^2 + \nu^2 \| -2N\mu(x_0 - x_\star) + (u_0 - u_\star) \|^2 \\ &= \nu^2 \|x_\star - x_0 + 2N\mu(u_\star - u_0)\|^2 + \nu^2 \|2N\mu(x_\star - x_0) + (u_0 - u_\star)\|^2 - \nu R^2 \\ &= \nu^2 [(1 + 4N^2\mu^2)\|x_0 - x_\star\|^2 + (1 + 4N^2\mu^2)\|u_0 - u_\star\|^2] - \nu R^2 \\ &= \nu(\|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2 - R^2) \\ &= 0 \end{aligned}$$

Case 2:  $\mathcal{V}_1 \leq \mathcal{V}_0$  for  $N \geq 2$

First, we will give some formulas and definitions.

$$\|a\|^2 - \|b\|^2 = \langle a + b, a - b \rangle \quad (3a)$$

$$\Omega = 1 + 4N\mu^2 \quad (3b)$$

$$\Omega\eta_1 = \eta_0 \quad (3c)$$

$$g'(y_1) = \frac{x_0 + \eta_0 u_0 - y_1}{\eta_0} \quad (3d)$$

$$f'(x_1) = \frac{w_1 - x_1}{\eta_1} \quad (3e)$$

$$f'(x_1) + g'(y_1) = \frac{(y_1 - x_1)\Omega}{\eta_0} \quad (3f)$$

$$2\eta_0(1 + \Omega)\mu = \Omega^2 - 1 \quad (3g)$$

$$\begin{aligned} \mathcal{V}_1 - \mathcal{V}_0 &= \nu^2[\|\Omega x_\star + 2N\mu u_\star - \Omega(2x_1 - w_1)\|^2 - \|x_\star + 2N\mu u_\star - x_0 - 2N\mu u_0\|^2] \\ &= \nu^2[\|\Omega x_\star + \eta_0 u_\star - \Omega(x_1 - \eta_1 f'(x_1))\|^2 - \\ &\quad \|x_\star + \eta_0 u_\star - (y_1 + \eta_0 g'(y_1))\|^2] \quad \triangleright \text{using (3d) and (3e)} \\ &= \nu^2[\langle \Omega x_\star + \eta_0 u_\star - \Omega x_1 + \Omega \eta_1 f'(x_1) + x_\star + \eta_0 u_\star - y_1 - \eta_0 g'(y_1), \\ &\quad \Omega x_\star + \eta_0 u_\star - \Omega x_1 + \Omega \eta_1 f'(x_1) - x_\star - \eta_0 u_\star + y_1 + \eta_0 g'(y_1) \rangle] \quad \triangleright \text{using (3a)} \\ &= \nu^2[\langle (1 + \Omega)x_\star - y_1 - \Omega x_1 + 2\eta_0 u_\star + \Omega \eta_1 f'(x_1) - \eta_0 g'(y_1), \\ &\quad (-1 + \Omega)x_\star + y_1 - \Omega x_1 + \Omega \eta_1 f'(x_1) + \eta_0 g'(y_1) \rangle] \\ &= \nu^2[\langle (1 + \Omega)x_\star - y_1 - \Omega x_1 + 2\eta_0 u_\star + \eta_0(f'(x_1) - g'(y_1)), \\ &\quad (-1 + \Omega)x_\star + y_1 - \Omega x_1 + (y_1 - x_1)\Omega \rangle] \quad \triangleright \text{using (3c) and (3f)} \\ &= \nu^2[\langle (1 + \Omega)x_\star - y_1 - \Omega x_1 + 2\eta_0 u_\star + \eta_0(f'(x_1) - g'(y_1)), (1 + \Omega)(y_1 - x_\star) + 2\Omega(x_\star - x_1) \rangle] \\ &= \nu^2[\langle (1 + \Omega)(y_1 - x_\star), (1 + \Omega)x_\star - y_1 - \Omega x_1 + 2\eta_0 u_\star + \eta_0(f'(x_1) - g'(y_1)) \rangle + \\ &\quad \langle 2\Omega(x_\star - x_1), (1 + \Omega)x_\star - y_1 - \Omega x_1 + 2\eta_0 u_\star + \eta_0(f'(x_1) - g'(y_1)) \rangle] \\ &= \nu^2[\langle (1 + \Omega)(y_1 - x_\star), (1 + \Omega)x_\star - y_1 - \Omega x_1 + 2\eta_0 u_\star + (y_1 - x_1)\Omega - 2\eta_0 g'(y_1) \rangle + \\ &\quad \langle 2\Omega(x_\star - x_1), (1 + \Omega)x_\star - y_1 - \Omega x_1 + 2\eta_0 u_\star + 2\eta_0 f'(x_1) - (y_1 - x_1)\Omega \rangle] \quad \triangleright \text{using (3f)} \\ &= \nu^2[\langle y_1 - x_\star, (1 + \Omega)^2 x_\star + (\Omega^2 - 1)y_1 - 2\Omega(1 + \Omega)x_1 + 2\eta_0(1 + \Omega)(u_\star - g'(y_1)) \rangle + \\ &\quad \langle x_\star - x_1, 2\Omega(1 + \Omega)x_\star - 2\Omega(1 + \Omega)y_1 + 4\eta_0\Omega(u_\star + f'(x_1)) \rangle] \\ &= \nu^2[\langle y_1 - x_\star, (1 + \Omega)^2 x_\star + (\Omega^2 - 1)y_1 - 2\Omega(1 + \Omega)x_\star + 2\eta_0(1 + \Omega)(u_\star - g'(y_1)) \rangle + \\ &\quad \langle x_\star - x_1, 4\eta_0\Omega(u_\star + f'(x_1)) \rangle] \\ &= \nu^2[\langle y_1 - x_\star, (\Omega^2 - 1)(y_1 - x_\star) + 2\eta_0(1 + \Omega)(u_\star - g'(y_1)) \rangle - 4\eta_0\Omega\langle x_\star - x_1, -u_\star - f'(x_1) \rangle] \\ &= \nu^2\{2\eta_0(1 + \Omega)[-\langle y_1 - x_\star, g'(y_1) - u_\star \rangle + \mu\|y_1 - x_\star\|^2] - 4\eta_0\Omega\langle x_\star - x_1, -u_\star - f'(x_1) \rangle\} \\ &\leq 0, \end{aligned}$$

where the final inequality holds by monotonicity of  $\partial f$ ,  $\mu$ -strong monotonicity of  $\partial g$ , since  $-u_\star \in \partial f(x_\star)$ , and because  $2\eta_0(1 + \Omega) \geq 0$  and  $4\eta_0\Omega \geq 0$ .

Case 3:  $\mathcal{V}_{k+1} \leq \mathcal{V}_k \quad \forall k = 1, \dots, N - 2$

First we will give some important formulas and definitions.

$$g'(y_{k+1}) = \frac{2x_k - w_k - y_{k+1}}{\eta_k} \quad (4a)$$

$$f'(x_{k+1}) = \frac{w_{k+1} - x_{k+1}}{\eta_{k+1}} \quad (4b)$$

$$\Delta_k = 1 + 4kN\mu^2 \quad (4c)$$

$$\Delta_{k+1} = 1 + 4(k+1)N\mu^2 \quad (4d)$$

$$\eta_k \Delta_k = \eta_{k+1} \Delta_{k+1} = 2N\mu \quad (4e)$$

$$f'(x_{k+1}) + g'(y_{k+1}) = \frac{(y_{k+1} - x_{k+1})\Delta_{k+1}}{2N\mu} \quad (4f)$$

$$\begin{aligned} \mathcal{V}_{k+1} - \mathcal{V}_k &= \nu^2 [\|\Delta_{k+1}x_\star + 2N\mu u_\star - \Delta_{k+1}(2x_{k+1} - w_{k+1})\|^2 \\ &\quad - \|\Delta_k x_\star + 2N\mu u_\star - \Delta_k(2x_k - w_k)\|^2] \\ &= \nu^2 [\|\Delta_{k+1}x_\star + 2N\mu u_\star - \Delta_{k+1}(x_{k+1} - \eta_{k+1}f'(x_{k+1}))\|^2 \\ &\quad - \|\Delta_k x_\star + 2N\mu u_\star - \Delta_k(y_{k+1} + \eta_k g'(y_{k+1}))\|^2] \quad \triangleright \text{from (4a) and (4b)} \\ &= \nu^2 [\langle (\Delta_{k+1} + \Delta_k)x_\star + 4N\mu u_\star - \Delta_{k+1}(x_{k+1} - \eta_{k+1}f'(x_{k+1})) - \Delta_k(y_{k+1} + \eta_k g'(y_{k+1})), \\ &\quad (\Delta_{k+1} - \Delta_k)x_\star - \Delta_{k+1}(x_{k+1} - \eta_{k+1}f'(x_{k+1})) + \Delta_k(y_{k+1} + \eta_k g'(y_{k+1})) \rangle] \quad \triangleright \text{from (3a)} \\ &= \nu^2 [\langle \Delta_{k+1}(x_\star - x_{k+1}) + \Delta_k(x_\star - y_{k+1}) + 4N\mu u_\star + \Delta_{k+1}\eta_{k+1}f'(x_{k+1}) - \Delta_k\eta_k g'(y_{k+1}), \\ &\quad \Delta_{k+1}(x_\star - x_{k+1}) - \Delta_k(x_\star - y_{k+1}) + \Delta_{k+1}\eta_{k+1}f'(x_{k+1}) + \Delta_k\eta_k g'(y_{k+1}) \rangle] \\ &= \nu^2 [\Delta_{k+1}^2 \|x_\star - x_{k+1}\|^2 - \Delta_k^2 \|x_\star - y_{k+1}\|^2 + 2\Delta_{k+1}^2 \langle x_\star - x_{k+1}, \eta_{k+1}f'(x_{k+1}) \rangle \\ &\quad + 2\Delta_k^2 \langle x_\star - y_{k+1}, \eta_k g'(y_{k+1}) \rangle + \Delta_{k+1}^2 \eta_{k+1}^2 \|f'(x_{k+1})\|^2 - \Delta_k^2 \eta_k^2 \|g'(y_{k+1})\|^2 \\ &\quad + \langle 4N\mu u_\star, \Delta_{k+1}(x_\star - x_{k+1}) - \Delta_k(x_\star - y_{k+1}) + \Delta_{k+1}\eta_{k+1}f'(x_{k+1}) + \Delta_k\eta_k g'(y_{k+1}) \rangle] \\ &= \nu^2 [\langle x_\star - x_{k+1}, 2\Delta_{k+1}^2 \eta_{k+1}f'(x_{k+1}) + 4\Delta_{k+1}N\mu u_\star + \Delta_{k+1}^2(x_\star - x_{k+1}) \rangle \\ &\quad + \langle x_\star - y_{k+1}, 2\Delta_k^2 \eta_k g'(y_{k+1}) - 4\Delta_k N\mu u_\star - \Delta_k^2(x_\star - y_{k+1}) \rangle \\ &\quad + 4N^2\mu^2 \langle 2u_\star + f'(x_{k+1}) - g'(y_{k+1}), f'(x_{k+1}) + g'(y_{k+1}) \rangle] \quad \triangleright \text{from (4e)} \\ &= \nu^2 [\langle x_\star - x_{k+1}, 2\Delta_{k+1}^2 \eta_{k+1}f'(x_{k+1}) + 4\Delta_{k+1}N\mu u_\star + \Delta_{k+1}^2(x_\star - x_{k+1}) \rangle \\ &\quad + \langle x_\star - y_{k+1}, 2\Delta_k^2 \eta_k g'(y_{k+1}) - 4\Delta_k N\mu u_\star - \Delta_k^2(x_\star - y_{k+1}) \rangle \\ &\quad + 2N\mu \Delta_{k+1} \langle 2u_\star + f'(x_{k+1}) - g'(y_{k+1}), y_{k+1} - x_{k+1} \rangle] \quad \triangleright \text{from (4f)} \\ &= \nu^2 [\langle x_\star - x_{k+1}, 6N\Delta_{k+1}\mu f'(x_{k+1}) + 8N\Delta_{k+1}\mu u_\star - 2N\Delta_{k+1}\mu g'(y_{k+1}) + \\ &\quad \Delta_{k+1}^2(x_\star - x_{k+1}) + \langle x_\star - y_{k+1}, 2N\mu(2\Delta_k + \Delta_{k+1})g'(y_{k+1}) - 4N\mu(\Delta_k + \Delta_{k+1})u_\star \\ &\quad - 2N\mu\Delta_{k+1}f'(x_{k+1}) - \Delta_k^2(x_\star - y_{k+1}) \rangle] \quad \triangleright \text{from (4e)} \\ &= \nu^2 [\langle x_\star - x_{k+1}, 8N\Delta_{k+1}\mu(f'(x_{k+1}) + u_\star) + \Delta_{k+1}^2(x_\star - y_{k+1}) \rangle \\ &\quad + \langle x_\star - y_{k+1}, 4N\mu(\Delta_k + \Delta_{k+1})g'(y_{k+1}) - 4N\mu(\Delta_k + \Delta_{k+1})u_\star + (x_{k+1} - y_{k+1})\Delta_{k+1}^2 \\ &\quad - \Delta_k^2(x_\star - y_{k+1}) \rangle] \quad \triangleright \text{from (4f)} \\ &= \nu^2 [-8N\Delta_{k+1}\mu \langle x_{k+1} - x_\star, f'(x_{k+1}) + u_\star \rangle \\ &\quad + \langle x_\star - y_{k+1}, 4N\mu(\Delta_k + \Delta_{k+1})(g'(y_{k+1}) - u_\star) + (\Delta_{k+1}^2 - \Delta_k^2)(x_\star - y_{k+1}) \rangle] \\ &= \nu^2 \{-8N\Delta_{k+1}\mu \langle x_{k+1} - x_\star, f'(x_{k+1}) + u_\star \rangle \\ &\quad + 4N\mu(\Delta_k + \Delta_{k+1})[-\langle x_\star - y_{k+1}, u_\star - g'(y_{k+1}) \rangle + \mu \|x_\star - y_{k+1}\|^2]\} \end{aligned}$$

$$\leq 0$$

where the final inequality holds by monotonicity of  $\partial f$ ,  $\mu$ -strong monotonicity of  $\partial g$ , since  $-u_\star \in \partial f(x_\star)$ , and because the coefficients  $8N\Delta_{k+1}\mu$  and  $4N\mu(\Delta_k + \Delta_{k+1})$  are non-negative.

Case 4:  $\mathcal{V}_N \leq \mathcal{V}_{N-1}$  (including the case  $\mathcal{V}_1 \leq \mathcal{V}_0$  when  $N = 1$ )

First, we will give some formulas and definitions.

$$g'(y_N) = \frac{2x_{N-1} - w_{N-1} - y_N}{\eta_{N-1}} \quad (5a)$$

$$f'(x_N) = \frac{w_N - x_N}{\eta_N} \quad (5b)$$

$$\Delta_{N-1} = 1 + 4(N-1)N\mu^2 \quad (5c)$$

$$\Delta_N = 1 + 4N^2\mu^2 \quad (5d)$$

$$\Delta_{N-1}\eta_{N-1} = \Delta_N\eta_N = 2N\mu \quad (5e)$$

$$f'(x_N) + g'(y_N) = \frac{(y_N - x_N)\Delta_N}{2N\mu} \quad (5f)$$

$$\begin{aligned} \mathcal{V}_N - \mathcal{V}_{N-1} &= \|x_N - x_\star\|^2 + 4N^2\mu^2\nu^2\|u_\star + f'(x_N)\|^2 - \nu^2\|\Delta_{N-1}x_\star + 2N\mu u_\star - \Delta_{N-1}(2x_{N-1} - w_{N-1})\|^2 \\ &= \|x_N - x_\star\|^2 + 4N^2\mu^2\nu^2\|u_\star + f'(x_N)\|^2 - \nu^2\|\Delta_{N-1}x_\star + 2N\mu u_\star - \Delta_{N-1}(y_N + \eta_{N-1}g'(y_N))\|^2 \\ &= \|x_N - x_\star\|^2 + 4N^2\mu^2\nu^2\|u_\star + f'(x_N)\|^2 - \nu^2\|\Delta_{N-1}(x_\star - y_N) + 2N\mu(u_\star - g'(y_N))\|^2 \quad \triangleright \text{from (5e)} \\ &= \|x_N - x_\star\|^2 + \nu^2\langle 2N\mu(u_\star + f'(x_N)) + \Delta_{N-1}(x_\star - y_N) + 2N\mu(u_\star - g'(y_N)), \\ &\quad 2N\mu(u_\star + f'(x_N)) - \Delta_{N-1}(x_\star - y_N) - 2N\mu(u_\star - g'(y_N)) \rangle \quad \triangleright \text{from (3a)} \\ &= \|x_N - x_\star\|^2 + \nu^2\langle 2N\mu(u_\star + f'(x_N)) + \Delta_{N-1}(x_\star - y_N) + 2N\mu(u_\star - g'(y_N)), \\ &\quad (y_N - x_N)\Delta_N - \Delta_{N-1}(x_\star - y_N) \rangle \quad \triangleright \text{from (5f)} \\ &= \|x_N - x_\star\|^2 + \nu^2\langle 2N\mu(u_\star + f'(x_N)) + \Delta_{N-1}(x_\star - y_N) + 2N\mu(u_\star - g'(y_N)), \\ &\quad (\Delta_{N-1} + \Delta_N)(y_N - x_\star) - \Delta_N(x_N - x_\star) \rangle \quad \triangleright \text{add and subtract } \Delta_N x_\star \\ &= \nu^2\langle 2N\mu(u_\star + f'(x_N)) - \Delta_{N-1}(y_N - x_\star) + 2N\mu(u_\star - g'(y_N)), (\Delta_{N-1} + \Delta_N)(y_N - x_\star) \rangle \\ &\quad + \nu^2\langle -\Delta_N(x_N - x_\star) + 2N\mu(u_\star + f'(x_N)) - \Delta_{N-1}(y_N - x_\star) + 2N\mu(u_\star - g'(y_N)), -\Delta_N(x_N - x_\star) \rangle \\ &= \nu^2\langle 2N\mu u_\star + (y_N - x_N)\Delta_N - 2N\mu g'(y_N) - \Delta_{N-1}(y_N - x_\star) + 2N\mu(u_\star - g'(y_N)), \\ &\quad (\Delta_{N-1} + \Delta_N)(y_N - x_\star) \rangle + \nu^2\langle -\Delta_N(x_N - x_\star) + 2N\mu(u_\star + f'(x_N)) - \Delta_{N-1}(y_N - x_\star) + \\ &\quad 2N\mu u_\star + 2N\mu f'(x_N) + (x_N - y_N)\Delta_N, -\Delta_N(x_N - x_\star) \rangle \quad \triangleright \text{from (5f)} \\ &= \nu^2\langle 4N\mu(u_\star - g'(y_N)) + (\Delta_N - \Delta_{N-1})(y_N - x_\star) + (x_\star - x_N)\Delta_N, (\Delta_{N-1} + \Delta_N)(y_N - x_\star) \rangle \\ &\quad + \nu^2\langle 4N\mu(u_\star + f'(x_N)) - (\Delta_N + \Delta_{N-1})(y_N - x_\star), -\Delta_N(x_N - x_\star) \rangle \\ &= \nu^2\langle 4N\mu(u_\star - g'(y_N)) + (\Delta_N - \Delta_{N-1})(y_N - x_\star), (\Delta_{N-1} + \Delta_N)(y_N - x_\star) \rangle + \\ &\quad \nu^2\langle 4N\mu(u_\star + f'(x_N)), -\Delta_N(x_N - x_\star) \rangle \\ &= 8N\mu(1 + 2N(2N-1)\mu^2)\nu^2(-\langle g'(y_N) - u_\star, y_N - x_\star \rangle + \mu\|y_N - x_\star\|^2) + \\ &\quad 4N\mu\nu(-\langle f'(x_N) + u_\star, x_N - x_\star \rangle) \\ &\leq 0 \end{aligned}$$

where the final inequality holds by monotonicity of  $\partial f$ ,  $\mu$ -strong monotonicity of  $\partial g$ , since  $-u_\star \in \partial f(x_\star)$ , and because  $8N\mu(1 + 2N(2N-1)\mu^2)\nu^2 \geq 0$  and  $4N\mu\nu \geq 0$ .

□

### 3 Matching lower bound

In this section, we present a complexity lower bound that matches the  $\mathcal{O}(1/N^2)$  rate and its leading constant of Theorem 1.

We follow a standard approach: we first construct a pair of worst-case functions for the class of methods satisfying the *proximal span condition*, and then apply a resisting oracle technique to extend the lower bound to apply to all deterministic algorithms.

#### 3.1 Proximal span condition

Intuitively, for first-order methods, the span condition is defined by requiring that each iterate belongs to the span of the previous iterates and the associated first-order information. This notion has been formalized in various forms in prior works [17, 3, 8, 19, 9, 21, 12].

In our particular setting, we consider composite problems accessed through proximal oracles. Before giving the formal definition in Section 3.3, we use the term *deterministic  $N$ -step prox-prox method* informally to refer to any deterministic algorithm that makes a total of  $2N$  proximal oracle calls, each call being to either  $f$  or  $g$ , and then outputs a point  $x_N$ . Since the method has access to two starting points, namely  $x_0$  and  $u_0$ , the corresponding *proximal span condition* can be stated as below for fixed iteration count  $N > 0$ . Specifically, for fixed  $(x_0, u_0)$ , we say a *trajectory* or a *sequence*

$$\{(z_i, \delta_i, \gamma_i)\}_{i=0}^{2N-1}, x_N\}$$

satisfies the *proximal span condition* if

$$\begin{aligned} \delta_i &\in \{0, 1\} && \text{for } i = 0, \dots, 2N - 1, \\ z_0 &\in x_0 + \text{span}\{u_0\} \\ d_i &= \begin{cases} z_i - \text{prox}_{\gamma_i f}(z_i) \text{ for some } \gamma_i > 0, & \text{if } \delta_i = 0, \\ z_i - \text{prox}_{\gamma_i g}(z_i) \text{ for some } \gamma_i > 0, & \text{if } \delta_i = 1, \end{cases} && \text{for } i = 0, \dots, 2N - 1, \\ z_i &\in x_0 + \text{span}\{u_0, d_0, \dots, d_{i-1}\} && \text{for } i = 1, \dots, 2N - 1, \\ x_N &\in x_0 + \text{span}\{u_0, d_0, \dots, d_{2N-1}\} \end{aligned}$$

and we say a *method* satisfies the *proximal span condition* if the trajectory  $\{(z_i, \delta_i, \gamma_i)\}_{i=0}^{2N-1}, x_N\}$  produced by the first-order deterministic method satisfies the proximal span condition for any starting point  $(x_0, u_0)$ .

To clarify, we make no assumption on either the order of the proximal oracle evaluations or which function is queried at each step. We only assume that the total number of proximal oracle evaluations is  $2N$ .

Under the proximal span condition, we establish the following lower bound.

**Theorem 2** *Fix a positive integer  $N$  and let  $\mu > 0$ . For any initial pair  $(x_0, u_0) \in \mathbb{R}^{2N+2} \times \mathbb{R}^{2N+2}$ , there exist a closed, proper, and convex function  $f: \mathbb{R}^{2N+2} \rightarrow \mathbb{R} \cup \{\infty\}$ , a closed, proper, and  $\mu$ -strongly convex function  $g: \mathbb{R}^{2N+2} \rightarrow \mathbb{R} \cup \{\infty\}$ ,  $x_\star \in \text{argmin}(f + g)$ , and  $u_\star \in \partial g(x_\star)$  with  $-u_\star \in \partial f(x_\star)$  such that the following holds:*

*For every deterministic  $N$ -step prox-prox method satisfying the proximal span condition, if  $x_N$  denotes its final output when run from  $(x_0, u_0)$  on the pair  $(f, g)$ , then*

$$\|x_N - x_\star\|^2 \geq \frac{\|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2}{1 + 4N^2\mu^2 + 4N\mu}.$$

### 3.2 Proof of Theorem 2

We now construct a worst-case instance for first-order deterministic prox-prox methods satisfying the proximal span condition. Throughout this subsection, the starting pair  $(x_0, u_0)$  is given and fixed.

Fix  $N \in \mathbb{N}$  and  $\mu > 0$ . Choose an orthonormal set of vectors

$$e_{-1}, e_0, e_1, \dots, e_{2N}$$

of  $\mathbb{R}^{2N+2}$  such that  $u_0 \in \text{span}\{e_{-1}\}$ . Let  $t \in \mathbb{R}^d$  such that

$$t := \sum_{i=0}^{2N} t_i e_i \in \mathbb{R}^d.$$

where  $t_i > 0$  for  $0 \leq i \leq 2N$ . We define two closed convex sets by

$$C_t := \left\{ y \in \mathbb{R}^{2N+2} : y_{-1} = 0, y_0 = t_0, (y_{2k-1}, y_{2k}) \in [(0, 0), (t_{2k-1}, t_{2k})], k = 1, \dots, N \right\},$$

and

$$D_t := \left\{ y \in \mathbb{R}^{2N+2} : y_{-1} = 0, (y_{2k}, y_{2k+1}) \in [(0, 0), (t_{2k}, t_{2k+1})], k = 0, \dots, N-1 \right\},$$

where  $y_i$  is shorthand for  $\langle y, e_i \rangle$  for  $-1 \leq i \leq 2N$  and  $[(0, 0), (a, b)]$  is the line segment with endpoints  $(0, 0)$  and  $(a, b)$ . We then define closed, proper, and  $\mu$ -strongly convex  $g$

$$g(x) := \frac{\mu}{2} \|x - x_0\|^2 + \langle u_0, x - x_0 \rangle + \delta_{x_0 + C_t}(x),$$

and closed, proper, and convex (possibly non-smooth)  $f$  as

$$f(x) := -\langle u_0, x - x_0 \rangle + \delta_{x_0 + D_t}(x).$$

Then minimizing  $f + g$  is equivalent to minimizing

$$\frac{\mu}{2} \|x - x_0\|^2$$

over  $(x_0 + C_t) \cap (x_0 + D_t)$ . Because

$$C_t \cap D_t = \{y \in \mathbb{R}^{2N+2} : y_{-1} = 0, y_i = t_i \text{ for } 0 \leq i \leq 2N\},$$

we have

$$(x_0 + C_t) \cap (x_0 + D_t) = \{x_0 + y \in \mathbb{R}^{2N+2} : y_{-1} = 0, y_i = t_i \text{ for } 0 \leq i \leq 2N\}.$$

Hence  $x_* := x_0 + t$  is the unique minimizer of  $f + g$ .

Moreover, the proximal mappings admit the explicit representations

$$\text{prox}_{\gamma f}(x_0 + y) = x_0 + \Pi_{D_t}(y + \gamma u_0),$$

and

$$\text{prox}_{\gamma g}(x_0 + y) = x_0 + \Pi_{C_t}\left(\frac{y - \gamma u_0}{1 + \gamma \mu}\right),$$

where  $\Pi_{D_t}$  and  $\Pi_{C_t}$  denote the Euclidean projections onto  $D_t$  and  $C_t$ , respectively.

For  $k = -1, 0, \dots, 2N$ , define

$$S_k := \text{span}\{e_{-1}, e_0, \dots, e_k\}, \quad S_{-1} := \text{span}\{e_{-1}\}.$$

The sets  $C_t$  and  $D_t$  are arranged so that the two proximal operators reveal coordinates one at a time. We will refer to this as the *proximal zero-chain property*, and the next lemma formalizes this construction.

**Lemma 2** For every  $\gamma > 0$  and every  $k = -1, 0, \dots, 2N - 1$ ,

$$y \in S_k \implies \text{prox}_{\gamma f}(x_0 + y) = x_0 + \Pi_{D_t}(y + \gamma u_0) \in x_0 + S_{k+1},$$

and

$$y \in S_k \implies \text{prox}_{\gamma g}(x_0 + y) = x_0 + \Pi_{C_t}\left(\frac{y - \gamma u_0}{1 + \gamma \mu}\right) \in x_0 + S_{k+1}.$$

*Proof* Both  $C_t$  and  $D_t$  are Cartesian products of one-dimensional and two-dimensional blocks, and hence their Euclidean projections act blockwise. Since  $u_0 \in \text{span}\{e_{-1}\}$  and both  $C_t$  and  $D_t$  impose the constraint  $z_{-1} = 0$ , the shift by  $\pm\gamma u_0$  only affects the  $e_{-1}$ -component and does not reveal any new coordinate. Therefore, if  $y \in S_k$ , then after projecting onto either  $C_t$  or  $D_t$ , at most the next coordinate  $e_{k+1}$  can become nonzero. This proves the claim.  $\square$

This lemma shows that if  $y \in S_k$ , then

$$(x_0 + y) - \text{prox}_{\gamma g}(x_0 + y) \in S_{k+1}$$

because each proximal evaluation reveals at most one new coordinate. Consequently, for any prox-prox method satisfying the proximal span condition, the final output  $x_N$  generated after at most  $2N$  proximal evaluations must belong to

$$x_0 + S_{2N-1}.$$

Since

$$x_\star - x_0 = t = \sum_{i=0}^{2N} t_i e_i$$

has a nonzero  $e_{2N}$ -component, we obtain

$$\inf_{y \in x_0 + S_{2N-1}} \|y - x_\star\|^2 = t_{2N}^2 > 0.$$

Thus, after at most  $2N$  proximal oracle calls, no method satisfying the proximal span condition can guarantee an error smaller than  $t_{2N}^2$  on this instance. The next lemma rewrites the optimality conditions at the endpoint  $x_\star = x_0 + t$  in blockwise form.

**Lemma 3** Let

$$x_\star = x_0 + t = x_0 + \sum_{i=0}^{2N} t_i e_i,$$

and let

$$g_\star = \sum_{i=0}^{2N} g_i e_i \in \mathbb{R}^d$$

where  $g_i \in \mathbb{R}$  for  $0 \leq i \leq 2N$ . Assume that

$$g_\star \in \mu t + N_{C_t}(t) \quad \text{and} \quad -g_\star \in N_{D_t}(t).$$

Then, with

$$u_\star := u_0 + g_\star,$$

we have

$$u_\star \in \partial g(x_\star), \quad -u_\star \in \partial f(x_\star).$$

Moreover, the above inclusions on  $g_\star$  are equivalent to the blockwise inequalities

$$\begin{aligned} (g_{2k+1} - \mu t_{2k+1}, g_{2k+2} - \mu t_{2k+2}) \cdot (t_{2k+1}, t_{2k+2}) &\geq 0, & k = 0, \dots, N-1, \\ (g_{2k}, g_{2k+1}) \cdot (t_{2k}, t_{2k+1}) &\leq 0, & k = 0, \dots, N-1, \end{aligned}$$

and  $g_{2N} = 0$ .

*Proof* Recall that

$$g(x) = \frac{\mu}{2} \|x - x_0\|^2 + \langle u_0, x - x_0 \rangle + \delta_{x_0 + C_t}(x),$$

and

$$f(x) = -\langle u_0, x - x_0 \rangle + \delta_{x_0 + D_t}(x).$$

Since  $x_\star = x_0 + t$ , we have

$$\partial g(x_\star) = \mu(x_\star - x_0) + u_0 + N_{x_0 + C_t}(x_\star) = \mu t + u_0 + N_{C_t}(t),$$

where we used the translation rule

$$N_{x_0 + C_t}(x_0 + t) = N_{C_t}(t).$$

Hence, if

$$g_\star \in \mu t + N_{C_t}(t),$$

then

$$u_\star := u_0 + g_\star \in \partial g(x_\star).$$

Similarly,

$$\partial f(x_\star) = -u_0 + N_{x_0 + D_t}(x_\star) = -u_0 + N_{D_t}(t),$$

and therefore

$$-g_\star \in N_{D_t}(t) \implies -u_\star = -(u_0 + g_\star) \in \partial f(x_\star).$$

It remains to characterize the normal cone conditions explicitly. Since  $C_t$  is the Cartesian product of the singleton constraint  $z_0 = t_0$ , the constraint  $z_{-1} = 0$ , and the line segments

$$[(0, 0), (t_{2k-1}, t_{2k})], \quad k = 1, \dots, N,$$

its normal cone at  $t$  is the product of the corresponding blockwise normal cones. The equality constraints fixing the coordinates  $-1$  and  $0$  contribute the full normal spaces in those coordinates, and hence impose no restrictions on the corresponding components of a normal vector. Thus, only the two-dimensional segment blocks need to be considered. For a segment

$$[(0, 0), (a, b)],$$

the normal cone at the endpoint  $(a, b)$  is  $\{w \in \mathbb{R}^2 : w \cdot (a, b) \geq 0\}$ . Applying this with

$$w = (g_{2k+1} - \mu t_{2k+1}, g_{2k+2} - \mu t_{2k+2})$$

gives

$$(g_{2k+1} - \mu t_{2k+1}, g_{2k+2} - \mu t_{2k+2}) \cdot (t_{2k+1}, t_{2k+2}) \geq 0, \quad k = 0, \dots, N-1.$$

Likewise,  $D_t$  is the product of the segment blocks

$$[(0, 0), (t_{2k}, t_{2k+1})], \quad k = 0, \dots, N-1,$$

together with the constraint  $z_{-1} = 0$ . Therefore,

$$-g_\star \in N_{D_t}(t)$$

is equivalent to

$$(g_{2k}, g_{2k+1}) \cdot (t_{2k}, t_{2k+1}) \leq 0, \quad k = 0, \dots, N-1.$$

Finally,  $D_t$  leaves the coordinate  $2N$  free and hence its normal component there must vanish. So,  $g_{2N} = 0$  and this proves the result.  $\square$

We now choose the coefficients of  $t$  and  $g_*$  so that every blockwise inequality is satisfied with equality under the normalization

$$\|x_0 - x_*\|^2 + \|u_0 - u_*\|^2 = 1.$$

**Lemma 4** *Define*

$$\begin{aligned} t_{2k}^* &= \sqrt{\frac{\mu}{(1+2N\mu)(1+2k\mu)(1+(2k+1)\mu)}}, & k = 0, \dots, N-1, \\ t_{2k+1}^* &= \sqrt{\frac{\mu}{(1+2N\mu)(1+(2k+1)\mu)(1+(2k+2)\mu)}}, & k = 0, \dots, N-1, \\ t_{2N}^* &= \frac{1}{1+2N\mu}, \end{aligned}$$

and

$$g_{2k}^* = -(1+2k\mu)t_{2k}^*, \quad g_{2k+1}^* = (1+(2k+2)\mu)t_{2k+1}^*, \quad k = 0, \dots, N-1,$$

with

$$g_{2N}^* = 0.$$

Then the pair

$$t = \sum_{i=0}^{2N} t_i^* e_i, \quad g_* = \sum_{i=0}^{2N} g_i^* e_i$$

with

$$x_* = x_0 + t, \quad u_* = u_0 + g_*$$

satisfies

$$\begin{aligned} \|x_0 - x_*\|^2 + \|u_0 - u_*\|^2 &= \sum_{i=0}^{2N} (t_i^*)^2 + \sum_{i=0}^{2N} (g_i^*)^2 = 1, \\ (g_{2k+1}^* - \mu t_{2k+1}^*, g_{2k+2}^* - \mu t_{2k+2}^*) \cdot (t_{2k+1}^*, t_{2k+2}^*) &= 0, \quad k = 0, \dots, N-1, \\ (g_{2k}^*, g_{2k+1}^*) \cdot (t_{2k}^*, t_{2k+1}^*) &= 0, \quad k = 0, \dots, N-1, \end{aligned}$$

so that  $u_* \in \partial g(x_*)$ ,  $-u_* \in \partial f(x_*)$ , and therefore  $x_* \in \operatorname{argmin}(f + g)$ .

*Proof* We first verify the inequalities.

For each  $k = 0, \dots, N-2$ , we have

$$g_{2k+1}^* - \mu t_{2k+1}^* = (1+(2k+1)\mu)t_{2k+1}^*,$$

and

$$g_{2k+2}^* - \mu t_{2k+2}^* = -(1+(2k+3)\mu)t_{2k+2}^*.$$

Therefore,

$$\begin{aligned} &(g_{2k+1}^* - \mu t_{2k+1}^*, g_{2k+2}^* - \mu t_{2k+2}^*) \cdot (t_{2k+1}^*, t_{2k+2}^*) \\ &= (1+(2k+1)\mu)(t_{2k+1}^*)^2 - (1+(2k+3)\mu)(t_{2k+2}^*)^2. \end{aligned}$$

By the definition of  $t_{2k+1}^*$  and  $t_{2k+2}^*$ ,

$$(1+(2k+1)\mu)(t_{2k+1}^*)^2 = \frac{\mu}{(1+2N\mu)(1+(2k+2)\mu)},$$

and

$$(1 + (2k + 3)\mu)(t_{2k+2}^*)^2 = \frac{\mu}{(1 + 2N\mu)(1 + (2k + 2)\mu)}.$$

Hence

$$(g_{2k+1}^* - \mu t_{2k+1}^*, g_{2k+2}^* - \mu t_{2k+2}^*) \cdot (t_{2k+1}^*, t_{2k+2}^*) = 0.$$

For  $k = N - 1$ ,

$$g_{2N-1}^* - \mu t_{2N-1}^* = (1 + (2N - 1)\mu)t_{2N-1}^*, \quad g_{2N}^* - \mu t_{2N}^* = -\mu t_{2N}^*.$$

Therefore,

$$\begin{aligned} & (g_{2N-1}^* - \mu t_{2N-1}^*, g_{2N}^* - \mu t_{2N}^*) \cdot (t_{2N-1}^*, t_{2N}^*) \\ &= (1 + (2N - 1)\mu)(t_{2N-1}^*)^2 - \mu(t_{2N}^*)^2. \end{aligned}$$

Now

$$(t_{2N-1}^*)^2 = \frac{\mu}{(1 + 2N\mu)(1 + (2N - 1)\mu)(1 + 2N\mu)} = \frac{\mu}{(1 + (2N - 1)\mu)(1 + 2N\mu)^2},$$

so

$$(1 + (2N - 1)\mu)(t_{2N-1}^*)^2 = \frac{\mu}{(1 + 2N\mu)^2}.$$

Also,

$$(t_{2N}^*)^2 = \frac{1}{(1 + 2N\mu)^2}.$$

Hence

$$(1 + (2N - 1)\mu)(t_{2N-1}^*)^2 = \mu(t_{2N}^*)^2,$$

and therefore

$$(g_{2N-1}^* - \mu t_{2N-1}^*, g_{2N}^* - \mu t_{2N}^*) \cdot (t_{2N-1}^*, t_{2N}^*) = 0.$$

Thus,

$$(g_{2k+1}^* - \mu t_{2k+1}^*, g_{2k+2}^* - \mu t_{2k+2}^*) \cdot (t_{2k+1}^*, t_{2k+2}^*) = 0, \quad k = 0, \dots, N - 1.$$

Similarly, for each  $k = 0, \dots, N - 1$ ,

$$(g_{2k}^*, g_{2k+1}^*) \cdot (t_{2k}^*, t_{2k+1}^*) = -(1 + 2k\mu)(t_{2k}^*)^2 + (1 + (2k + 2)\mu)(t_{2k+1}^*)^2.$$

Again, by the definitions,

$$(1 + 2k\mu)(t_{2k}^*)^2 = \frac{\mu}{(1 + 2N\mu)(1 + (2k + 1)\mu)},$$

and

$$(1 + (2k + 2)\mu)(t_{2k+1}^*)^2 = \frac{\mu}{(1 + 2N\mu)(1 + (2k + 1)\mu)}.$$

Therefore,

$$(g_{2k}^*, g_{2k+1}^*) \cdot (t_{2k}^*, t_{2k+1}^*) = 0, \quad k = 0, \dots, N - 1.$$

Thus, all block inequalities are satisfied with equality. It remains to verify the normalization constraint.

For each  $j = 0, \dots, 2N - 1$ ,

$$(t_j^*)^2 = \frac{\mu}{(1 + 2N\mu)(1 + j\mu)(1 + (j + 1)\mu)} = \frac{1}{1 + 2N\mu} \left( \frac{1}{1 + j\mu} - \frac{1}{1 + (j + 1)\mu} \right).$$

Hence

$$\sum_{j=0}^{2N-1} (t_j^*)^2 = \frac{1}{1+2N\mu} \sum_{j=0}^{2N-1} \left( \frac{1}{1+j\mu} - \frac{1}{1+(j+1)\mu} \right) = \frac{1}{1+2N\mu} \left( 1 - \frac{1}{1+2N\mu} \right).$$

Adding  $(t_{2N}^*)^2 = \frac{1}{(1+2N\mu)^2}$  yields

$$\sum_{j=0}^{2N} (t_j^*)^2 = \frac{1}{1+2N\mu}.$$

Next, for each  $k = 0, \dots, N-1$ ,

$$(g_{2k}^*)^2 = (1+2k\mu)^2 (t_{2k}^*)^2 = \frac{\mu(1+2k\mu)}{(1+2N\mu)(1+(2k+1)\mu)},$$

and

$$(g_{2k+1}^*)^2 = (1+(2k+2)\mu)^2 (t_{2k+1}^*)^2 = \frac{\mu(1+(2k+2)\mu)}{(1+2N\mu)(1+(2k+1)\mu)}.$$

Therefore,

$$\begin{aligned} (g_{2k}^*)^2 + (g_{2k+1}^*)^2 &= \frac{\mu}{1+2N\mu} \left( \frac{1+2k\mu}{1+(2k+1)\mu} + \frac{1+(2k+2)\mu}{1+(2k+1)\mu} \right) \\ &= \frac{\mu}{1+2N\mu} \cdot \frac{2+(4k+2)\mu}{1+(2k+1)\mu} = \frac{2\mu}{1+2N\mu}. \end{aligned}$$

Summing over  $k = 0, \dots, N-1$  and using  $g_{2N}^* = 0$ , we obtain

$$\sum_{j=0}^{2N} (g_j^*)^2 = \sum_{k=0}^{N-1} ((g_{2k}^*)^2 + (g_{2k+1}^*)^2) = \frac{2N\mu}{1+2N\mu}.$$

Combining the two identities gives

$$\sum_{j=0}^{2N} (t_j^*)^2 + \sum_{j=0}^{2N} (g_j^*)^2 = \frac{1}{1+2N\mu} + \frac{2N\mu}{1+2N\mu} = 1.$$

□

Together with the fact that

$$\|x_N - x_\star\|^2 \geq \inf_{y \in x_0 + S_{2N-1}} \|y - x_\star\|^2 = (t_{2N}^*)^2$$

and

$$(t_{2N}^*)^2 = \frac{1}{(1+2N\mu)^2} = \frac{\|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2}{1+4N^2\mu^2+4N\mu},$$

we obtain

$$\|x_N - x_\star\|^2 \geq \frac{\|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2}{1+4N^2\mu^2+4N\mu}.$$

This proves Theorem 2.

### 3.3 Extension using the resisting oracle technique

Next, using the so-called *resisting oracle technique*, we extend the result of Theorem 2 to all deterministic  $N$ -step prox-prox methods. First, we make a formal definition.

Let  $N > 0$  be a positive integer. A *deterministic  $N$ -step prox-prox method  $\mathbf{A}$*  is a mapping from an initial pair  $(x_0, u_0)$  and a tuple of functions  $(f, g)$  to a sequence of outputs

$$\mathbf{A}((x_0, u_0), (f, g)) = \left\{ \left\{ (z_i, \delta_i, \gamma_i) \right\}_{i=0}^{2N-1}, x_N \right\},$$

where  $\delta_i \in \{0, 1\}$  and  $\gamma_i > 0$  for  $i = 0, \dots, 2N - 1$ . For clarity, we call  $\{z_i\}_{0 \leq i \leq 2N-1}$  the *query points* and  $x_N$  the *final output*. Moreover,  $\mathbf{A}$  requires that the output sequence depends on  $(f, g)$  only through  $(x_0, u_0)$ , previous query points, and previous proximal oracle calls. More precisely, letting

$$y_i := \begin{cases} \text{prox}_{\gamma_i f}(z_i), & \delta_i = 0, \\ \text{prox}_{\gamma_i g}(z_i), & \delta_i = 1, \end{cases} \quad i = 0, \dots, 2N - 1,$$

there exist deterministic maps  $\mathbf{A}_0, \dots, \mathbf{A}_{2N}$  such that

$$(z_0, \delta_0, \gamma_0) = \mathbf{A}_0(x_0, u_0),$$

$$(z_i, \delta_i, \gamma_i) = \mathbf{A}_i\left((x_0, u_0), \{(z_j, \delta_j, \gamma_j, y_j)\}_{j=0}^{i-1}\right), \quad i = 1, \dots, 2N - 1,$$

and

$$x_N = \mathbf{A}_{2N}\left((x_0, u_0), \{(z_j, \delta_j, \gamma_j, y_j)\}_{j=0}^{2N-1}\right).$$

With this definition, we are now ready to formally state the more general lower bound.

**Theorem 3** *Fix a positive integer  $N$  and let  $\mu > 0$ . Let  $d \geq 4N + 4$ . For every deterministic  $N$ -step prox-prox method  $\mathbf{A}$  on  $\mathbb{R}^d$  and every initial pair  $(x_0, u_0) \in \mathbb{R}^d \times \mathbb{R}^d$ , there exist a closed, proper, and convex function  $f: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$  and a closed, proper, and  $\mu$ -strongly convex function  $g: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$ , together with  $x_* \in \text{argmin}(f + g)$  and  $u_* \in \partial g(x_*)$  with  $-u_* \in \partial f(x_*)$ , such that, if  $x_N$  denotes the final output of  $\mathbf{A}$  when run from  $(x_0, u_0)$  on  $(f, g)$ , then*

$$\|x_N - x_*\|^2 \geq \frac{\|x_0 - x_*\|^2 + \|u_0 - u_*\|^2}{1 + 4N^2\mu^2 + 4N\mu}.$$

Theorem 3 extends Theorem 2 from methods satisfying the proximal span condition to all deterministic  $N$ -step prox-prox methods, using the resisting-oracle technique. To do this, we introduce a *prox-prox zero-respecting condition*. It is closely related to the proximal span condition and may in fact be somewhat redundant here, but we include it for clarity of exposition.

Let  $f, g: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$ , and let

$$\left\{ \left\{ (\tilde{z}_i, \delta_i, \gamma_i) \right\}_{i=0}^{2N-1}, \tilde{x}_N \right\}$$

be a sequence with  $\delta_i \in \{0, 1\}$  and  $\gamma_i > 0$  for  $i = 0, \dots, 2N - 1$ . Define, for each  $i = 0, \dots, 2N - 1$ ,

$$\tilde{d}_i := \begin{cases} \tilde{z}_i - \text{prox}_{\gamma_i f}(\tilde{z}_i), & \delta_i = 0, \\ \tilde{z}_i - \text{prox}_{\gamma_i g}(\tilde{z}_i), & \delta_i = 1. \end{cases}$$

We say that the trajectory

$$\left\{ \left\{ (\tilde{z}_i, \delta_i, \gamma_i) \right\}_{i=0}^{2N-1}, \tilde{x}_N \right\}$$

is *prox-prox zero-respecting with respect to*  $(f, g)$  if

$$\text{supp}(\tilde{z}_i) \subseteq \bigcup_{k=0}^{i-1} \text{supp}(\tilde{d}_k), \quad i = 0, \dots, 2N - 1$$

and

$$\text{supp}(\tilde{x}_N) \subseteq \bigcup_{k=0}^{2N-1} \text{supp}(\tilde{d}_k).$$

where the support is taken with respect to the fixed orthonormal basis with the convention  $\text{supp}() = \emptyset$ .

The next lemma formalizes that, for functions  $(f, g)$  constructed in Theorem 2, the zero-respecting condition is enough to recover the proximal span condition with  $(x_0, u_0) = (0, 0)$ . The key idea is that proximal residuals can reveal new coordinates only consecutively, one at a time. All supports below are taken with respect to the orthonormal coordinates  $e_{-1}, e_0, \dots, e_{2N}$  used in the construction of  $(f, g)$  in Section 3.2.

**Lemma 5** *Let  $f, g : \mathbb{R}^{2N+2} \rightarrow \mathbb{R} \cup \{\infty\}$  be the functions defined in Theorem 2. Let*

$$\{(\tilde{z}_i, \delta_i, \gamma_i)\}_{i=0}^{2N-1}, \tilde{x}_N\}$$

*be prox-prox zero-respecting with respect to  $(f, g)$ , and define*

$$\tilde{d}_i := \begin{cases} \tilde{z}_i - \text{prox}_{\gamma_i f}(\tilde{z}_i), & \delta_i = 0, \\ \tilde{z}_i - \text{prox}_{\gamma_i g}(\tilde{z}_i), & \delta_i = 1, \end{cases} \quad i = 0, \dots, 2N - 1.$$

*Then  $\{(\tilde{z}_i, \delta_i, \gamma_i)\}_{i=0}^{2N-1}, \tilde{x}_N\}$  satisfies the proximal span condition with initial pair  $(x_0, u_0) = (0, 0)$ , namely,*

$$\begin{aligned} \tilde{z}_i &\in \text{span}\{\tilde{d}_0, \dots, \tilde{d}_{i-1}\}, \quad i = 0, \dots, 2N - 1, \\ \tilde{x}_N &\in \text{span}\{\tilde{d}_0, \dots, \tilde{d}_{2N-1}\}. \end{aligned}$$

*Consequently, the proof of Theorem 2 applies to every prox-prox zero-respecting trajectory with  $(x_0, u_0) = (0, 0)$ .*

*Proof* For notational simplicity, define

$$E_{-1} := \{0\}, \quad E_m := \text{span}\{e_0, \dots, e_m\}, \quad m = 0, \dots, 2N - 1.$$

We prove by induction on  $k = 0, \dots, 2N - 1$  that there exists  $m_k \in \{-1, 0, \dots, k\}$  such that

$$\text{span}\{\tilde{d}_0, \dots, \tilde{d}_k\} = E_{m_k}.$$

For  $k = 0$ , the zero-respecting property gives

$$\text{supp}(\tilde{z}_0) \subseteq \emptyset,$$

so  $\tilde{z}_0 = 0$ . Hence, by the prox-zero-chain property,

$$\tilde{d}_0 \in \text{span}\{e_0\}.$$

Therefore  $\text{span}\{\tilde{d}_0\}$  is either  $\{0\}$  or  $\text{span}\{e_0\}$ , so the claim holds with  $m_0 = -1$  or  $m_0 = 0$ . Now suppose the claim holds for  $k - 1$ , namely,

$$\text{span}\{\tilde{d}_0, \dots, \tilde{d}_{k-1}\} = E_{m_{k-1}}.$$

Then

$$\bigcup_{i=0}^{k-1} \text{supp}(\tilde{d}_i) \subseteq \{0, \dots, m_{k-1}\}.$$

Here, we use the convention  $\{0, \dots, -1\} = \emptyset$  when  $m_{k-1} = -1$ . Since the trajectory is prox-prox zero-respecting,

$$\text{supp}(\tilde{z}_k) \subseteq \bigcup_{i=0}^{k-1} \text{supp}(\tilde{d}_i),$$

and thus

$$\tilde{z}_k \in \text{span}\{e_0, \dots, e_{m_{k-1}}\} = E_{m_{k-1}}.$$

Applying the prox-zero-chain property, we obtain

$$\tilde{d}_k \in \text{span}\{e_0, \dots, e_{m_{k-1}}, e_{m_{k-1}+1}\}.$$

Therefore

$$\text{span}\{\tilde{d}_0, \dots, \tilde{d}_k\}$$

is either

$$\text{span}\{e_0, \dots, e_{m_{k-1}}\} = E_{m_{k-1}} \quad \text{or} \quad \text{span}\{e_0, \dots, e_{m_{k-1}+1}\} = E_{m_{k-1}+1}.$$

Thus by induction, for each  $k = 0, \dots, 2N - 1$ ,

$$\bigcup_{i=0}^{k-1} \text{supp}(\tilde{d}_i) \subseteq \{0, \dots, m_{k-1}\},$$

so the zero-respecting property implies

$$\tilde{z}_k \in \text{span}\{e_0, \dots, e_{m_{k-1}}\} = \text{span}\{\tilde{d}_0, \dots, \tilde{d}_{k-1}\}.$$

Likewise,

$$\text{supp}(\tilde{x}_N) \subseteq \bigcup_{i=0}^{2N-1} \text{supp}(\tilde{d}_i) \subseteq \{0, \dots, m_{2N-1}\},$$

and therefore

$$\tilde{x}_N \in \text{span}\{e_0, \dots, e_{m_{2N-1}}\} = \text{span}\{\tilde{d}_0, \dots, \tilde{d}_{2N-1}\}.$$

This proves that the trajectory satisfies the proximal span condition with initial pair  $(x_0, u_0) = (0, 0)$ .  $\square$

We now turn to the lifting construction used in the resisting-oracle argument.

**Lemma 6** *Let  $U \in \mathbb{R}^{d' \times d}$  have orthonormal columns with  $d' \geq d$ . For any vectors  $x_0, u_0 \in \mathbb{R}^{d'}$ , if  $f: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$  is a closed, proper, and convex function, then  $f_U: \mathbb{R}^{d'} \rightarrow \mathbb{R} \cup \{\infty\}$  defined by*

$$f_U(x) := f(U^\top(x - x_0)) - \langle u_0, x - x_0 \rangle$$

*is a convex, closed, and proper function.*

*Proof* Since the mapping  $x \mapsto U^\top(x - x_0)$  is affine and  $f$  is convex, closed, and proper, the composition

$$x \mapsto f(U^\top(x - x_0))$$

is convex, closed, and proper. Moreover, the function

$$x \mapsto -\langle u_0, x - x_0 \rangle$$

is affine and finite-valued on  $\mathbb{R}^{d'}$ . Therefore,

$$f_U(x) = f(U^\top(x - x_0)) - \langle u_0, x - x_0 \rangle$$

is the sum of a proper, closed, convex function and an affine function. Hence  $f_U$  is proper, closed, and convex.  $\square$

**Lemma 7** *Let  $U \in \mathbb{R}^{d' \times d}$  have orthonormal columns with  $d' \geq d$ . For any vectors  $x_0, u_0 \in \mathbb{R}^{d'}$ , if  $g: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$  is closed, proper, and  $\mu$ -strongly convex, then the function  $g_U: \mathbb{R}^{d'} \rightarrow \mathbb{R} \cup \{\infty\}$  defined by*

$$g_U(x) := g(U^\top(x - x_0)) + \langle u_0, x - x_0 \rangle + \frac{\mu}{2} \|P(x - x_0)\|^2, \quad P := I - UU^\top,$$

*is  $\mu$ -strongly convex.*

*Proof* Fix  $x, y \in \mathbb{R}^{d'}$  and let  $v \in \partial g_U(x)$  be arbitrary and write it uniquely as

$$v = Us + u_0 + \mu P(x - x_0)$$

where  $s \in \partial g(U^\top(x - x_0))$ . Since  $g$  is  $\mu$ -strongly convex, we have

$$g(U^\top(y - x_0)) \geq g(U^\top(x - x_0)) + \langle s, U^\top(y - x) \rangle + \frac{\mu}{2} \|U^\top(y - x)\|^2.$$

Also,

$$\langle u_0, y - x_0 \rangle = \langle u_0, x - x_0 \rangle + \langle u_0, y - x \rangle,$$

and

$$\frac{\mu}{2} \|P(y - x_0)\|^2 = \frac{\mu}{2} \|P(x - x_0)\|^2 + \mu \langle P(x - x_0), y - x \rangle + \frac{\mu}{2} \|P(y - x)\|^2.$$

Adding these identities and inequalities yields

$$\begin{aligned} g_U(y) &\geq g_U(x) + \langle Us + u_0 + \mu P(x - x_0), y - x \rangle \\ &\quad + \frac{\mu}{2} \|U^\top(y - x)\|^2 + \frac{\mu}{2} \|P(y - x)\|^2. \end{aligned}$$

Since

$$\|U^\top(y - x)\|^2 + \|P(y - x)\|^2 = \|y - x\|^2,$$

we obtain

$$\begin{aligned} g_U(y) &\geq g_U(x) + \langle Us + u_0 + \mu P(x - x_0), y - x \rangle + \frac{\mu}{2} \|y - x\|^2 \\ &= g_U(x) + \langle v, y - x \rangle + \frac{\mu}{2} \|y - x\|^2. \end{aligned}$$

Therefore  $g_U$  is  $\mu$ -strongly convex.  $\square$

From now on,  $U \in \mathbb{R}^{d' \times d}$  has orthonormal columns,  $x_0, u_0$  are fixed, and  $f_U$  and  $g_U$  are defined as in the previous two lemmas. Next, we derive explicit formulas for the proximal mappings of the lifted functions.

**Lemma 8** For every  $z \in \mathbb{R}^d$  and every  $\gamma > 0$ , assume that  $U^\top u_0 = 0$ . Write

$$\tilde{z} := U^\top(z - x_0), \quad p_z := P(z - x_0),$$

where  $P := I - UU^\top$ . Then

$$\text{prox}_{\gamma f_U}(z) = x_0 + U \text{prox}_{\gamma f}(\tilde{z}) + p_z + \gamma u_0,$$

and therefore

$$z - \text{prox}_{\gamma f_U}(z) = U(\tilde{z} - \text{prox}_{\gamma f}(\tilde{z})) - \gamma u_0.$$

In particular,

$$U^\top(z - \text{prox}_{\gamma f_U}(z)) = \tilde{z} - \text{prox}_{\gamma f}(\tilde{z}).$$

*Proof* Since  $U^\top U = I_d$  and  $P = I - UU^\top$ , every  $x \in \mathbb{R}^d$  admits a unique decomposition

$$x = x_0 + U\tilde{x} + p, \quad \tilde{x} \in \mathbb{R}^d, \quad p \in \text{range}(P),$$

and we write

$$z = x_0 + U\tilde{z} + p_z, \quad \tilde{z} := U^\top(z - x_0), \quad p_z := P(z - x_0).$$

Since  $U^\top u_0 = 0$ , we have  $u_0 \in \text{range}(P)$ , and hence

$$\langle u_0, x - x_0 \rangle = \langle u_0, U\tilde{x} + p \rangle = \langle u_0, p \rangle.$$

Using  $\text{range}(U) \perp \text{range}(P)$ , we obtain

$$\|x - z\|^2 = \|\tilde{x} - \tilde{z}\|^2 + \|p - p_z\|^2,$$

and therefore

$$f_U(x) + \frac{1}{2\gamma}\|x - z\|^2 = f(\tilde{x}) - \langle u_0, p \rangle + \frac{1}{2\gamma}\|\tilde{x} - \tilde{z}\|^2 + \frac{1}{2\gamma}\|p - p_z\|^2.$$

Hence the minimization separates. Minimizing with respect to  $\tilde{x}$ , we get

$$\tilde{x} = \text{prox}_{\gamma f}(\tilde{z}),$$

while the minimizing  $p$  satisfies

$$\frac{1}{\gamma}(p - p_z) - u_0 = 0,$$

so

$$p = p_z + \gamma u_0.$$

This proves that

$$\text{prox}_{\gamma f_U}(z) = x_0 + U \text{prox}_{\gamma f}(\tilde{z}) + p_z + \gamma u_0.$$

Subtracting from  $z = x_0 + U\tilde{z} + p_z$  gives

$$z - \text{prox}_{\gamma f_U}(z) = U(\tilde{z} - \text{prox}_{\gamma f}(\tilde{z})) - \gamma u_0.$$

Finally, since  $U^\top u_0 = 0$ , applying  $U^\top$  yields

$$U^\top(z - \text{prox}_{\gamma f_U}(z)) = \tilde{z} - \text{prox}_{\gamma f}(\tilde{z}).$$

□

**Lemma 9** For every  $z \in \mathbb{R}^d$  and every  $\gamma > 0$ , assume that  $U^\top u_0 = 0$ . Write

$$\tilde{z} := U^\top(z - x_0), \quad p_z := P(z - x_0),$$

where  $P := I - UU^\top$ . Then

$$\text{prox}_{\gamma g_U}(z) = x_0 + U \text{prox}_{\gamma g}(\tilde{z}) + \frac{1}{1 + \gamma\mu}(p_z - \gamma u_0),$$

and

$$z - \text{prox}_{\gamma g_U}(z) = U(\tilde{z} - \text{prox}_{\gamma g}(\tilde{z})) + \frac{\gamma\mu}{1 + \gamma\mu}p_z + \frac{\gamma}{1 + \gamma\mu}u_0.$$

In particular,

$$U^\top(z - \text{prox}_{\gamma g_U}(z)) = \tilde{z} - \text{prox}_{\gamma g}(\tilde{z}).$$

*Proof* Again write

$$x = x_0 + U\tilde{x} + p, \quad z = x_0 + U\tilde{z} + p_z, \quad p, p_z \in \text{range}(P).$$

Since  $U^\top u_0 = 0$ , we have  $u_0 \in \text{range}(P)$ , and therefore

$$\langle u_0, x - x_0 \rangle = \langle u_0, U\tilde{x} + p \rangle = \langle u_0, p \rangle.$$

Using  $\text{range}(U) \perp \text{range}(P)$ , we obtain

$$\|x - z\|^2 = \|\tilde{x} - \tilde{z}\|^2 + \|p - p_z\|^2.$$

Hence

$$g_U(x) + \frac{1}{2\gamma}\|x - z\|^2 = g(\tilde{x}) + \langle u_0, p \rangle + \frac{\mu}{2}\|p\|^2 + \frac{1}{2\gamma}\|\tilde{x} - \tilde{z}\|^2 + \frac{1}{2\gamma}\|p - p_z\|^2.$$

The minimization separates. Minimizing with respect to  $\tilde{x}$ , we get

$$\tilde{x} = \text{prox}_{\gamma g}(\tilde{z}),$$

while the minimizing  $p$  satisfies

$$u_0 + \mu p + \frac{1}{\gamma}(p - p_z) = 0.$$

Equivalently,

$$(1 + \gamma\mu)p = p_z - \gamma u_0,$$

so

$$p = \frac{1}{1 + \gamma\mu}(p_z - \gamma u_0).$$

This proves that

$$\text{prox}_{\gamma g_U}(z) = x_0 + U \text{prox}_{\gamma g}(\tilde{z}) + \frac{1}{1 + \gamma\mu}(p_z - \gamma u_0).$$

Subtracting from

$$z = x_0 + U\tilde{z} + p_z$$

gives

$$\begin{aligned} z - \text{prox}_{\gamma g_U}(z) &= U(\tilde{z} - \text{prox}_{\gamma g}(\tilde{z})) + p_z - \frac{1}{1 + \gamma\mu}(p_z - \gamma u_0) \\ &= U(\tilde{z} - \text{prox}_{\gamma g}(\tilde{z})) + \frac{\gamma\mu}{1 + \gamma\mu}p_z + \frac{\gamma}{1 + \gamma\mu}u_0. \end{aligned}$$

Finally, since  $U^\top u_0 = 0$  and  $U^\top p_z = 0$ , applying  $U^\top$  yields

$$U^\top(z - \text{prox}_{\gamma g_U}(z)) = \tilde{z} - \text{prox}_{\gamma g}(\tilde{z}).$$

□

These identities allow us to match proximal evaluations in the lifted space with projected proximal evaluations in the original space. We are now ready to carry out the resisting-oracle construction.

**Lemma 10** *Let  $N > 0$  be a positive integer. Assume  $\mu > 0$ ,  $d' \geq d + 2N + 2$ , and let  $\mathbf{A}$  be any deterministic  $N$ -step prox-prox method with initial pair  $(x_0, u_0) \in \mathbb{R}^{d'} \times \mathbb{R}^{d'}$ . Let*

$$f: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\} \quad \text{and} \quad g: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$$

*be such that  $f$  is closed, proper, and convex and  $g$  is closed, proper, and  $\mu$ -strongly convex. Then there exists a matrix*

$$U \in \mathbb{R}^{d' \times d}$$

*with orthonormal columns such that  $U^\top u_0 = 0$ , and if*

$$\mathbf{A}((x_0, u_0), (f_U, g_U)) = \left\{ \left\{ (z_i, \delta_i, \gamma_i) \right\}_{i=0}^{2N-1}, x_N \right\},$$

*and we define the projected iterates*

$$\tilde{z}_i := U^\top (z_i - x_0), \quad \tilde{x}_N := U^\top (x_N - x_0),$$

*then the sequence*

$$\left\{ \left\{ (\tilde{z}_i, \delta_i, \gamma_i) \right\}_{i=0}^{2N-1}, \tilde{x}_N \right\}$$

*is prox-prox zero-respecting with respect to  $(f, g)$ .*

*Proof* Let  $e_0, \dots, e_{d-1}$  be the standard basis of  $\mathbb{R}^d$ . First, we recursively construct

$$I_0 \subseteq I_1 \subseteq \dots \subseteq I_{2N} \subseteq \{0, \dots, d-1\},$$

an orthonormal family  $\{v_j\}_{j \in I_{2N}} \subseteq \mathbb{R}^{d'}$ , and a method trajectory

$$(z_0, \delta_0, \gamma_0, y_0), \dots, (z_{2N-1}, \delta_{2N-1}, \gamma_{2N-1}, y_{2N-1}),$$

where  $y_i \in \mathbb{R}^{d'}$  is intended to be the output of the  $i$ -th prox call.

We begin with

$$I_0 := \emptyset.$$

Next, fix  $i \in \{0, \dots, 2N-1\}$ , and suppose that  $I_i$ , the vectors  $\{v_j\}_{j \in I_i}$ , and the previous trajectory

$$(z_0, \delta_0, \gamma_0, y_0), \dots, (z_{i-1}, \delta_{i-1}, \gamma_{i-1}, y_{i-1})$$

have already been constructed. Since  $\mathbf{A}$  is deterministic, the trajectory determines the next query triple  $(z_i, \delta_i, \gamma_i)$ . Define

$$\bar{z}_i := \sum_{j \in I_i} \langle v_j, z_i - x_0 \rangle e_j \in \mathbb{R}^d,$$

so that  $\text{supp}(\bar{z}_i) \subseteq I_i$ . Next define

$$\bar{y}_i := \begin{cases} \text{prox}_{\gamma_i f}(\bar{z}_i), & \delta_i = 0, \\ \text{prox}_{\gamma_i g}(\bar{z}_i), & \delta_i = 1, \end{cases}$$

and the corresponding residual

$$\bar{d}_i := \bar{z}_i - \bar{y}_i.$$

Set

$$I_{i+1} := I_i \cup \text{supp}(\bar{d}_i).$$

Since  $\bar{y}_i = \bar{z}_i - \bar{d}_i$ , we also have

$$\text{supp}(\bar{y}_i) \subseteq I_{i+1}.$$

We now choose the new vectors

$$\{v_j\}_{j \in I_{i+1} \setminus I_i}$$

as an orthonormal family in  $S_i^\perp$ , where

$$S_i := \text{span}\left(\{u_0, z_0 - x_0, \dots, z_i - x_0\} \cup \{v_j\}_{j \in I_i}\right).$$

This is possible because

$$\dim S_i \leq (i + 2) + |I_i|,$$

hence

$$\dim S_i^\perp \geq d' - (i + 2) - |I_i| \geq d - |I_i| \geq |I_{i+1} \setminus I_i|.$$

Now define

$$q_i := (z_i - x_0) - \sum_{j \in I_i} \langle v_j, z_i - x_0 \rangle v_j,$$

and  $y_i \in \mathbb{R}^{d'}$  by

$$y_i := \begin{cases} x_0 + \sum_{j \in I_{i+1}} (\bar{y}_i)_j v_j + q_i + \gamma_i u_0, & \delta_i = 0, \\ x_0 + \sum_{j \in I_{i+1}} (\bar{y}_i)_j v_j + \frac{1}{1 + \gamma_i \mu} (q_i - \gamma_i u_0), & \delta_i = 1. \end{cases}$$

This completes the recursive construction for  $i = 0, \dots, 2N - 1$ . Since  $\mathbf{A}$  is deterministic, this transcript uniquely determines the final output  $x_N \in \mathbb{R}^{d'}$ . Then choose the orthonormal vectors indexed by  $j \notin I_{2N}$  in the orthonormal complement of

$$S_{2N+1} := \text{span}\left(\{u_0, z_0 - x_0, \dots, z_{2N-1} - x_0, x_N - x_0\} \cup \{v_j\}_{j \in I_{2N}}\right).$$

This is possible because

$$\dim S_{2N+1} \leq (2N + 2) + |I_{2N}|,$$

so

$$\dim S_{2N+1}^\perp \geq d' - (2N + 2) - |I_{2N}| \geq d - |I_{2N}|.$$

Now set

$$U := [v_0 | \dots | v_{d-1}] \in \mathbb{R}^{d' \times d}, \quad P := I - UU^\top.$$

By construction, each column of  $U$  is orthogonal to  $u_0$ , and therefore  $U^\top u_0 = 0$ .

For the next step, we claim that if  $\mathbf{A}$  is run on  $(f_U, g_U)$ , then after  $i$  oracle calls its trajectory is exactly

$$(z_0, \delta_0, \gamma_0, y_0), \dots, (z_{i-1}, \delta_{i-1}, \gamma_{i-1}, y_{i-1}), \quad i = 0, \dots, 2N,$$

and consequently, its final output is  $x_N$ . We prove this by induction on  $i$ . For  $i = 0$ , we only have the initial state  $(x_0, u_0)$  and empty transcript, so the claim is immediate. Now assume the claim holds for some  $i \in \{0, \dots, 2N - 1\}$ . Then  $\mathbf{A}$  run on  $(f_U, g_U)$  has seen exactly the same trajectory

$$(z_0, \delta_0, \gamma_0, y_0), \dots, (z_{i-1}, \delta_{i-1}, \gamma_{i-1}, y_{i-1})$$

as in the recursive construction. Hence, its next query is exactly  $(z_i, \delta_i, \gamma_i)$ .

Now for every  $j \notin I_i$ , the vector  $v_j$  is orthogonal to  $z_i - x_0$ , so  $\langle v_j, z_i - x_0 \rangle = 0$ . Therefore,

$$U^\top (z_i - x_0) = \bar{z}_i, \quad P(z_i - x_0) = q_i.$$

Now we apply Lemma 8 and Lemma 9. If  $\delta_i = 0$ , then

$$\begin{aligned}\text{prox}_{\gamma_i f_U}(z_i) &= x_0 + U \text{prox}_{\gamma_i f}(U^\top(z_i - x_0)) + P(z_i - x_0) + \gamma_i u_0 \\ &= x_0 + U \text{prox}_{\gamma_i f}(\bar{z}_i) + q_i + \gamma_i u_0.\end{aligned}$$

Since  $\bar{y}_i = \text{prox}_{\gamma_i f}(\bar{z}_i)$  and  $\text{supp}(\bar{y}_i) \subseteq I_{i+1}$ , this gives

$$\text{prox}_{\gamma_i f_U}(z_i) = x_0 + \sum_{j \in I_{i+1}} (\bar{y}_i)_j v_j + q_i + \gamma_i u_0 = y_i.$$

If  $\delta_i = 1$ , then

$$\begin{aligned}\text{prox}_{\gamma_i g_U}(z_i) &= x_0 + U \text{prox}_{\gamma_i g}(U^\top(z_i - x_0)) + \frac{1}{1 + \gamma_i \mu} (P(z_i - x_0) - \gamma_i u_0) \\ &= x_0 + U \text{prox}_{\gamma_i g}(\bar{z}_i) + \frac{1}{1 + \gamma_i \mu} (q_i - \gamma_i u_0).\end{aligned}$$

Since  $\bar{y}_i = \text{prox}_{\gamma_i g}(\bar{z}_i)$  and  $\text{supp}(\bar{y}_i) \subseteq I_{i+1}$ , this gives

$$\text{prox}_{\gamma_i g_U}(z_i) = x_0 + \sum_{j \in I_{i+1}} (\bar{y}_i)_j v_j + \frac{1}{1 + \gamma_i \mu} (q_i - \gamma_i u_0) = y_i.$$

Thus the true  $i$ -th oracle output is exactly  $y_i$ , which proves the induction step. Hence the actual run of  $\mathbf{A}$  on  $(f_U, g_U)$  has exactly the recursively constructed  $2N$ -step transcript. Since the final output depends only on this trajectory, the final output is exactly  $x_N$ .

We now verify the zero-respecting property. For each  $i = 0, \dots, 2N - 1$ , define

$$\tilde{d}_i := \begin{cases} \tilde{z}_i - \text{prox}_{\gamma_i f}(\tilde{z}_i), & \delta_i = 0, \\ \tilde{z}_i - \text{prox}_{\gamma_i g}(\tilde{z}_i), & \delta_i = 1. \end{cases}$$

Since  $y_i = \text{prox}_{\gamma_i f_U}(z_i)$  when  $\delta_i = 0$  and  $y_i = \text{prox}_{\gamma_i g_U}(z_i)$  when  $\delta_i = 1$ , Lemma 8 and Lemma 9 give

$$\tilde{d}_i = U^\top(z_i - y_i).$$

Next, we show that

$$U^\top(y_i - x_0) = \bar{y}_i.$$

If  $\delta_i = 0$ , then by construction,

$$y_i = x_0 + \sum_{j \in I_{i+1}} (\bar{y}_i)_j v_j + q_i + \gamma_i u_0.$$

Multiplying by  $U^\top$ , and using  $U^\top q_i = 0$  and  $U^\top u_0 = 0$ , we obtain

$$U^\top(y_i - x_0) = \bar{y}_i.$$

If  $\delta_i = 1$ , then

$$y_i = x_0 + \sum_{j \in I_{i+1}} (\bar{y}_i)_j v_j + \frac{1}{1 + \gamma_i \mu} (q_i - \gamma_i u_0),$$

and again  $U^\top q_i = 0$  and  $U^\top u_0 = 0$  imply

$$U^\top(y_i - x_0) = \bar{y}_i.$$

Since also

$$U^\top(z_i - x_0) = \bar{z}_i,$$

we conclude that

$$\tilde{d}_i = U^\top(z_i - y_i) = U^\top(z_i - x_0) - U^\top(y_i - x_0) = \bar{z}_i - \bar{y}_i = \bar{d}_i.$$

Because  $I_0 = \emptyset$  and

$$I_{i+1} = I_i \cup \text{supp}(\bar{d}_i) = I_i \cup \text{supp}(\tilde{d}_i),$$

it follows by induction that

$$I_i = \bigcup_{k=0}^{i-1} \text{supp}(\tilde{d}_k), \quad i = 0, \dots, 2N.$$

Moreover,

$$\text{supp}(\tilde{z}_i) = \text{supp}(\bar{z}_i) \subseteq I_i = \bigcup_{k=0}^{i-1} \text{supp}(\tilde{d}_k), \quad i = 0, \dots, 2N-1.$$

Finally, for every  $j \notin I_{2N}$  we have  $v_j \in S_{2N+1}^\perp$  and  $x_N - x_0 \in S_{2N+1}$ , so

$$\langle v_j, x_N - x_0 \rangle = 0.$$

Therefore

$$\text{supp}(\tilde{x}_N) = \text{supp}(U^\top(x_N - x_0)) \subseteq I_{2N} = \bigcup_{k=0}^{2N-1} \text{supp}(\tilde{d}_k).$$

Hence

$$\text{supp}(\tilde{x}_N) \subseteq \bigcup_{k=0}^{2N-1} \text{supp}(\tilde{d}_k),$$

which is exactly the prox-prox zero-respecting property.  $\square$

Combining the resisting-oracle lemma with the zero-respecting property, we obtain the desired lower bound for arbitrary deterministic prox-prox methods.

*Proof* By Theorem 2 with initial pair  $(0, 0)$ , there exist a closed, proper, and convex function  $\tilde{f}: \mathbb{R}^{2N+2} \rightarrow \mathbb{R} \cup \{\infty\}$  and a closed, proper, and  $\mu$ -strongly convex function  $\tilde{g}: \mathbb{R}^{2N+2} \rightarrow \mathbb{R} \cup \{\infty\}$  with

$$\tilde{x}_\star \in \text{argmin}(\tilde{f} + \tilde{g}), \quad \tilde{u}_\star \in \partial\tilde{g}(\tilde{x}_\star), \quad -\tilde{u}_\star \in \partial\tilde{f}(\tilde{x}_\star).$$

Since  $d \geq 4N + 4 = (2N + 2) + (2N + 2)$ , by Lemma 10 there exists a matrix

$$U \in \mathbb{R}^{d \times (2N+2)}$$

with orthonormal columns such that  $U^\top u_0 = 0$ , and if

$$\mathbf{A}((x_0, u_0), (\tilde{f}_U, \tilde{g}_U)) = \left\{ \left\{ (z_i, \delta_i, \gamma_i) \right\}_{i=0}^{2N-1}, x_N \right\},$$

and we define

$$\tilde{z}_i := U^\top(z_i - x_0), \quad \tilde{x}_N := U^\top(x_N - x_0),$$

then the projected sequence

$$\left\{ \left\{ (\tilde{z}_i, \delta_i, \gamma_i) \right\}_{i=0}^{2N-1}, \tilde{x}_N \right\}$$

is prox-prox zero-respecting with respect to  $(\tilde{f}, \tilde{g})$ . Therefore, by Theorem 2 and Lemma 5,

$$\|\tilde{x}_N - \tilde{x}_\star\|^2 \geq \frac{\|\tilde{x}_\star\|^2 + \|\tilde{u}_\star\|^2}{1 + 4N^2\mu^2 + 4N\mu}. \quad (6)$$

Now define

$$f := \tilde{f}_U, \quad g := \tilde{g}_U,$$

and

$$x_\star := x_0 + U\tilde{x}_\star, \quad u_\star := u_0 + U\tilde{u}_\star.$$

We first show that  $x_\star \in \operatorname{argmin}(f + g)$ . Every  $z \in \mathbb{R}^d$  can be written uniquely as

$$z = x_0 + Uy + p, \quad y \in \mathbb{R}^{2N+2}, \quad p \in \operatorname{range}(P),$$

where  $P := I - UU^\top$ . By definition of  $f_U$  and  $g_U$ , we have

$$f(z) = \tilde{f}(y) - \langle u_0, Uy + p \rangle,$$

and

$$g(z) = \tilde{g}(y) + \langle u_0, Uy + p \rangle + \frac{\mu}{2}\|p\|^2.$$

Hence

$$(f + g)(z) = \tilde{f}(y) + \tilde{g}(y) + \frac{\mu}{2}\|p\|^2.$$

This is minimized exactly when

$$y \in \operatorname{argmin}(\tilde{f} + \tilde{g}) \quad \text{and} \quad p = 0.$$

Thus

$$x_\star = x_0 + U\tilde{x}_\star \in \operatorname{argmin}(f + g).$$

Next we show that  $u_\star \in \partial g(x_\star)$  and  $-u_\star \in \partial f(x_\star)$ . By definition,

$$g(x) = \tilde{g}(U^\top(x - x_0)) + \langle u_0, x - x_0 \rangle + \frac{\mu}{2}\|P(x - x_0)\|^2.$$

Hence

$$\partial g(x) = U \partial \tilde{g}(U^\top(x - x_0)) + u_0 + \mu P(x - x_0).$$

Since

$$x_\star = x_0 + U\tilde{x}_\star,$$

we have

$$U^\top(x_\star - x_0) = \tilde{x}_\star, \quad P(x_\star - x_0) = 0.$$

Therefore

$$\partial g(x_\star) = U \partial \tilde{g}(\tilde{x}_\star) + u_0.$$

Since  $\tilde{u}_\star \in \partial \tilde{g}(\tilde{x}_\star)$ , we conclude that

$$u_\star = u_0 + U\tilde{u}_\star \in \partial g(x_\star).$$

Similarly,

$$f(x) = \tilde{f}(U^\top(x - x_0)) - \langle u_0, x - x_0 \rangle.$$

So,

$$\partial f(x) = U \partial \tilde{f}(U^\top(x - x_0)) - u_0,$$

and

$$\partial f(x_\star) = U\partial\tilde{f}(\tilde{x}_\star) - u_0.$$

Thus, we conclude that  $-u_\star = -u_0 - U\tilde{u}_\star \in \partial f(x_\star)$ . Putting altogether, we have

$$\|x_0 - x_\star\|^2 = \|U\tilde{x}_\star\|^2 = \|\tilde{x}_\star\|^2,$$

and

$$\|u_0 - u_\star\|^2 = \|U\tilde{u}_\star\|^2 = \|\tilde{u}_\star\|^2.$$

Thus

$$\|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2 = \|\tilde{x}_\star\|^2 + \|\tilde{u}_\star\|^2.$$

Finally,

$$x_N - x_\star = x_N - x_0 - U\tilde{x}_\star = U(\tilde{x}_N - \tilde{x}_\star) + P(x_N - x_0),$$

and the two terms on the right-hand side are orthogonal. Hence

$$\|x_N - x_\star\|^2 = \|\tilde{x}_N - \tilde{x}_\star\|^2 + \|P(x_N - x_0)\|^2 \geq \|\tilde{x}_N - \tilde{x}_\star\|^2.$$

Combining with (6) yields

$$\|x_N - x_\star\|^2 \geq \frac{\|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2}{1 + 4N^2\mu^2 + 4N\mu}.$$

□

## 4 Numerical Results

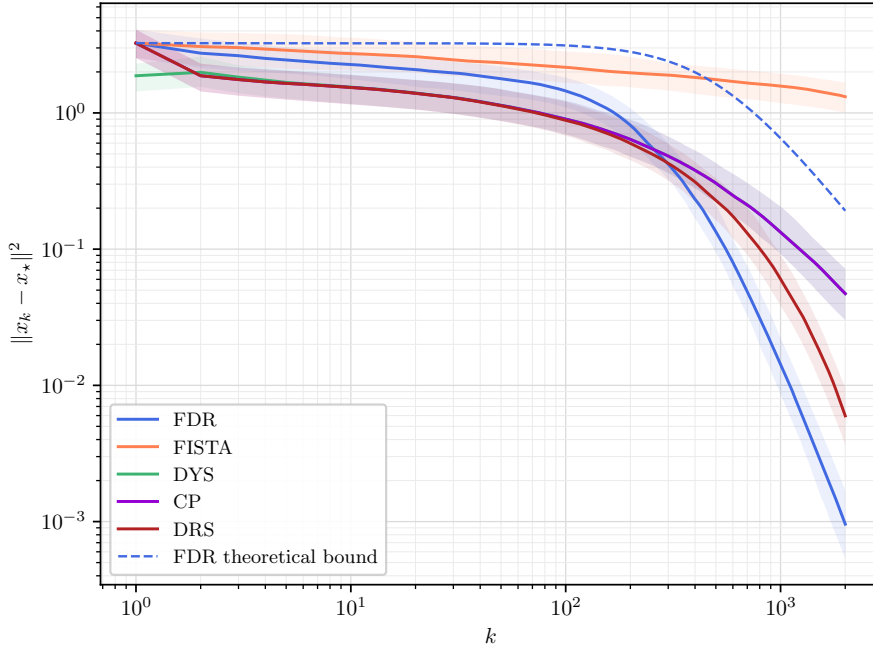
We compare FDR with FISTA, Douglas–Rachford splitting (DRS), accelerated Chambolle–Pock (CP), and accelerated Davis–Yin splitting (DYS) on a family of elastic-net regression problems. Each instance has the form

$$\underset{x}{\text{minimize}} \quad \underbrace{\|Ax - b\|_2^2 + \frac{\mu}{2}\|x\|^2}_{=g(x)} + \underbrace{\lambda\|x\|_1}_{=f(x)}.$$

We randomly generated 100 instances with  $A \in \mathbb{R}^{40 \times 100}$  and  $b \in \mathbb{R}^{40}$  being computed as  $Ax_{\text{true}} + \epsilon$ , where  $\epsilon$  is drawn from a Gaussian with standard deviation 0.01, and  $x_{\text{true}}$  is generated randomly with only 10% of elements being nonzero. We use  $\mu = \lambda = 10^{-3}$ .

Since the guarantee on FDR's squared distance to the solution is on the final iterate,  $x_N$ , the FDR curve in Figure 1 reports the terminal error  $\|x_k - x_\star\|^2$  obtained by running FDR for  $N = k$  iterations. The solid curves are the median of each algorithm's performance, and the shaded region shows the interquartile range for each on the 100 instances.

Figure 1 shows that FDR eventually achieves the smallest error among the tested methods. The observed FDR errors also remain below the theoretical bound, which is valid, but we see these errors are much smaller than the theoretical upper bound on these random elastic-net instances.



**Fig. 1** Median squared distance to the solution over 100 elastic-net instances. Shaded bands show the interquartile range.

## 5 Conclusion

In this work, we presented Fast Douglas–Rachford Splitting (FDR), an accelerated Douglas–Rachford-type method for minimizing the sum of a convex and a  $\mu$ -strongly convex function. We established the guarantee

$$\|x_N - x_\star\|^2 \leq \frac{\|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2}{1 + 4N^2\mu^2},$$

which improves upon the leading asymptotic constant of the accelerated methods of Chambolle–Pock and Davis–Yin by a factor of 4. We further established a complexity lower bound showing that both the  $\mathcal{O}(1/N^2)$  convergence rate and the leading-order constant achieved by FDR are optimal.

It is worth noting that the acceleration mechanism and the lower-bound construction are fundamentally different than those of Nesterov acceleration. In this sense, the accelerated methods of Chambolle–Pock, Davis–Yin, and FDR should be viewed as belonging to a class of acceleration mechanism distinct from the Nesterov-type acceleration. Developing a more principled understanding of the different classes of acceleration mechanisms is an interesting direction for future work.

Finally, while our lower bound asymptotically matches the upper bound in the leading-order term, a gap remains in the higher-order terms. Since FDR was discovered through BnB-PEP [5], a framework for discovering optimal first-order algorithms, we conjecture that the upper bound achieved by FDR is in fact exactly optimal, implying that the current lower bound is likely loose. We leave the task of fully tightening the lower bound to future work.

**Acknowledgements** GMC and BA were supported by the Office of Naval Research grants N000142512231 and N00014-25-1-2319. UJ and EKR were supported by the Air Force Office of Scientific Research under award number FA95502510183.

## Appendix A Leading constants for accelerated splitting methods

In this appendix we provide the asymptotic leading constants of the accelerated Davis–Yin splitting algorithm and the accelerated Chambolle–Pock algorithm when specialized to  $(\mathcal{P})$ .

### A.1 Accelerated Davis–Yin Splitting

From Proposition A.1 in [6], the iterates of Algorithm (DYS) satisfy

$$(1 + 2\gamma_k\mu)\|x_{k+1} - x_\star\|^2 + \gamma_k^2\|u_{k+1} - u_\star\|^2 \leq \|x_k - x_\star\|^2 + \gamma_k^2\|u_k - u_\star\|^2.$$

Furthermore, in the proof of Theorem 1.2 in Appendix A of [6], we see that the proximal step sizes satisfy

$$\frac{1}{\gamma_{k+1}^2} = \frac{1 + 2\gamma_k\mu}{\gamma_k^2},$$

and

$$\lim_{k \rightarrow \infty} (k+1)\gamma_k = \frac{1}{\mu}.$$

After  $N$  proximal evaluations of  $g$  we obtain

$$\begin{aligned} \|x_{N-1} - x_\star\|^2 &\leq \gamma_{N-1}^2 \left( \frac{1}{\gamma_0^2} \|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2 \right) \\ &= \frac{\gamma_{N-1}^2}{\gamma_0^2} (\|x_0 - x_\star\|^2 + \gamma_0^2 \|u_0 - u_\star\|^2). \end{aligned}$$

Setting  $\gamma_0^2 = 1$  to match our assumed initial condition yields the asymptotic rate

$$\|x_{N-1} - x_\star\|^2 \sim \frac{\|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2}{N^2\mu^2}.$$

### A.2 Accelerated Chambolle–Pock

The convergence rate of the accelerated Chambolle–Pock algorithm is given in [4, Theorem 2], assuming  $\sigma_0 = 1/\tau_0$ . For any  $\epsilon > 0$ , there exists  $N_0$  such that for all  $N \geq N_0$ ,

$$\begin{aligned} \|x_N - x_\star\|^2 &\leq \frac{1 + \epsilon}{N^2} \left( \frac{\|x_0 - x_\star\|^2}{\mu^2\tau_0^2} + \frac{\|u_0 - u_\star\|^2}{\mu^2} \right) \\ &= \frac{1 + \epsilon}{N^2\mu^2\tau_0^2} (\|x_0 - x_\star\|^2 + \tau_0^2\|u_0 - u_\star\|^2). \end{aligned}$$

Setting  $\tau_0^2 = 1$  to match our assumed initial condition and taking  $\epsilon \rightarrow 0$  yields

$$\|x_N - x_\star\|^2 \sim \frac{\|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2}{N^2\mu^2}.$$

### A.3 Comparison with FDR

Recall that FDR satisfies

$$\|x_N - x_\star\|^2 \leq \frac{\|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2}{1 + 4N^2\mu^2} \sim \frac{\|x_0 - x_\star\|^2 + \|u_0 - u_\star\|^2}{4N^2\mu^2}.$$

Thus FDR improves the asymptotic constant of both accelerated Davis–Yin splitting and accelerated Chambolle–Pock by a factor of 4.

### References

1. Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer New York (2011). DOI 10.1007/978-1-4419-9467-7. URL <http://dx.doi.org/10.1007/978-1-4419-9467-7>
2. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences* **2**(1), 183–202 (2009)
3. Carmon, Y., Duchi, J.C., Hinder, O., Sidford, A.: Lower bounds for finding stationary points i. *Mathematical Programming* **184**(1), 71–120 (2020)
4. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision* **40**(1), 120–145 (2010). DOI 10.1007/s10851-010-0251-1. URL <http://dx.doi.org/10.1007/s10851-010-0251-1>
5. Das Gupta, S., Van Parys, B.P., Ryu, E.K.: Branch-and-bound performance estimation programming: A unified methodology for constructing optimal optimization methods. *Mathematical Programming* **204**(1), 567–639 (2024)
6. Davis, D., Yin, W.: A three-operator splitting scheme and its optimization applications. *Set-valued and variational analysis* **25**, 829–858 (2017)
7. Douglas, J., Rachford, H.H.: On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American mathematical Society* **82**(2), 421–439 (1956)
8. Drori, Y.: The exact information-based complexity of smooth convex minimization. *Journal of Complexity* **39**, 1–16 (2017)
9. Drori, Y., Taylor, A.: On the oracle complexity of smooth strongly convex minimization. *Journal of Complexity* **68**, 101590 (2022)
10. Drori, Y., Teboulle, M.: Performance of first-order methods for smooth convex minimization: a novel approach. *Mathematical Programming* **145**(1), 451–482 (2014)
11. Halpern, B.: Fixed points of nonexpanding maps. *Bulletin of the American Mathematical Society* **73**(6), 957–961 (1967). DOI 10.1090/S0002-9904-1967-11864-0. URL <http://dx.doi.org/10.1090/S0002-9904-1967-11864-0>
12. Jang, U., Gupta, S.D., Ryu, E.K.: Computer-assisted design of accelerated composite optimization methods: Optista. *Mathematical Programming* pp. 1–109 (2025)
13. Kim, D.: Accelerated proximal point method for maximally monotone operators. *Mathematical Programming* **190**(1), 57–87 (2021)
14. Kim, D., Fessler, J.A.: Optimized first-order methods for smooth convex minimization. *Mathematical programming* **159**(1), 81–107 (2016)
15. Lieder, F.: On the convergence rate of the Halpern-iteration. *Optimization letters* **15**(2), 405–418 (2021)
16. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis* **16**(6), 964–979 (1979)
17. Nemirovsky, A.S.: On optimality of krylov’s information when solving linear operator equations. *Journal of Complexity* **7**(2), 121–130 (1991)
18. Nesterov, Y.: A method for solving the convex programming problem with convergence rate  $\mathcal{O}(1/k^2)$ . In: *Dokl akad nauk Sssr*, vol. 269, pp. 543–547 (1983)
19. Ouyang, Y., Xu, Y.: Lower complexity bounds of first-order methods for convex-concave bilinear saddle-point problems. *Mathematical Programming* **185**(1), 1–35 (2021)
20. Park, C., Park, J., Ryu, E.K.: Factor- $\sqrt{2}$  acceleration of accelerated gradient methods. *Applied Mathematics & Optimization* **88**(3), 77 (2023)
21. Park, J., Ryu, E.K.: Exact optimal accelerated complexity for fixed-point iterations. In: *International Conference on Machine Learning*, pp. 17420–17457. PMLR (2022)
22. Peaceman, D.W., Rachford Jr., H.H.: The numerical solution of parabolic and elliptic differential equations. *Journal of the Society for Industrial and Applied Mathematics* **3**(1), 28–41 (1955). DOI 10.1137/0103003
23. Rockafellar, R.T.: *Convex Analysis*. Princeton University Press (1970). DOI 10.1515/9781400873173. URL <http://dx.doi.org/10.1515/9781400873173>
24. Ryu, E.K., Yin, W.: *Large-Scale Convex Optimization via Monotone Operators*. Cambridge University Press (2022)