An Inexact Trust-Region Method for Structured Nonsmooth Optimization with Application to Risk-Averse Stochastic Programming

Drew P. Kouri[1]

[1]Optimization and Uncertainty Quantification, Sandia National Laboratories, P.O. Box 5800, Albuquerque, 87185-1320, NM, USA.

Contributing authors: dpkouri@sandia.gov;

Abstract

We develop a trust-region method for efficiently minimizing the sum of a smooth function, a nonsmooth convex function, and the composition of a finite-valued support function with a smooth function. Optimization problems with this structure arise in numerous applications including risk-averse stochastic programming and subproblems for nonsmooth penalty nonlinear programming methods. Our method permits the use of inexact value and derivative information, enabling the solution of infinite-dimensional problems governed by, e.g., partial differential equations (PDEs). We prove global convergence of our method and under additional regularity assumptions, demonstrate that the sequence of iterates accumulates at a stationary point of our target problem. We demonstrate our method's efficiency on two PDE-constrained optimization examples, showing that its performance is invariant to the PDE discretization size.

Keywords: Trust Regions, Nonsmooth Optimization, Saddle Point Problems, Risk-Averse Optimization, Stochastic Optimization, Adaptivity

MSC Classification: 49M29 , 49M37 , 65K10 , 90C15 , 93E20

1 Introduction

We develop an efficient trust-region method for solving structured nonsmooth, nonconvex optimization problems with the form

$$\min_{x\in X}\,\{J(x) \coloneqq f(x) + \sup_{\theta\in\mathfrak{A}}\,(\theta, F(x))_Y + \phi(x)\},\tag{1}$$

where X and Y are real Hilbert spaces, $f : X \to \mathbb{R}$ and $F : X \to Y$ are smooth functions, $\phi : X \to (-\infty, +\infty]$ is proper, closed and convex, and $\mathfrak{A} \subset Y$ is nonempty, closed, convex and bounded. This class of problems arises in various applications including sparse estimation and learning [1], nonlinear programming with L^1-penalties [2], and simulating mechanical systems subject to Tresca friction [3].

In addition to these applications, an extremely important application of our method is risk-averse optimization, which has the form

$$\min_{x}\; f(x) + \mathcal{R}(F(x)) + \phi(x).\tag{2}$$

In the context of (2), f is a smooth deterministic function, F is a smooth stochastic function (i.e., $F(x)$ is a random variable for each x), ϕ is a nonsmooth convex function, and $\mathcal{R}$ is a coherent risk measure [4] (in which case, $\mathfrak{A}$ is the risk envelope associated with $\mathcal{R}$). Risk-averse optimization problems arise in numerous applications ranging from financial mathematics to reliability engineering [5, 6]. Although ubiquitous, these problems are often challenging to solve because of the nonsmoothness arising in $\mathcal{R}$ and ϕ. In fact, traditional nonsmooth optimization methods are often intractable when (2) is nonconvex and F is computationally expensive to evaluate, as is the case when F involves the solution of a system of partial differential equations (PDEs) with uncertain coefficients [7–11]. Recently, [12] introduced the specialized primal-dual risk minimization algorithm. At each iteration of the primal-dual risk minimization algorithm, one approximately solves a smoothed risk-averse optimization problem, which can be expensive. Moreover, it is unclear how to leverage inexact evaluations of f, F and its derivatives during the primal-dual iteration.

To address these shortcomings, we introduce a new trust-region method that can efficiently solve (1) even when f, F and their derivatives are computationally expensive to evaluate. In this setting, it is critical that *(i)* our algorithm exhibits rapid convergence and *(ii)* can leverage inexact evaluations of f, F and their derivatives while maintaining strong convergence guarantees. For example, in PDE-constrained optimization [13], evaluating F often requires the discretization and iterative numerical solution of a system of nonlinear PDEs—a cost that is expounded when the PDE has uncertain or random coefficients. To address the algorithmic requirements *(i)* and *(ii)* above, we extend the inexact proximal trust-region algorithm developed in [14]. The method in [14] is applicable for minimizing $f + \phi$ and is provably convergent even when f and its derivative are computed inexactly. Moreover, the method exhibits local superlinear—even quadratic—convergence rates under mild assumptions [15]. Roughly speaking, our algorithm applies the trust-region algorithm in [14] to (1) using a modified trust-region subproblem model to account for the supremum over $\mathfrak{A}$. By exploiting the structure of (1), we are able to extend the convergence theory from [14] to guarantee convergence of our algorithm.

The remainder of the paper is structured as follows. In Section 2, we discuss the problem formulation, assumptions and optimality conditions for (1). In Section 3, we introduce our inexact trust-region algorithm and describe efficient ways to solve the trust-region subproblem in Section 3.1. We prove convergence of our method in Section 4 and demonstrate its numerical performance in Section 5.

2 Preliminaries

Let X and Y be real Hilbert spaces. We denote the inner product and associated norm on X by $(\cdot,\cdot)_X$ and $\|\cdot\|_X$, respectively, and analogously for Y. We denote the Banach space of bounded linear operators mapping X into Y by $\mathcal{L}(X,Y)$ and the topological dual space of X (i.e., the space of continuous linear functionals on X) by $X^* \coloneqq \mathcal{L}(X,\mathbb{R})$. To simplify the presentation, we identify the dual space X^* with X via the Riesz representation theorem and we denote $\mathcal{L}(X) \coloneqq \mathcal{L}(X,X)$. Recall that an extended real-valued function $\psi : X \to [-\infty,+\infty]$ is proper if $\psi(x) > -\infty$ for all $x \in X$ and there exists at least one $x_0 \in X$ for which $\psi(x_0) < +\infty$. When ψ is convex, we denote its proximity operator by $\mathrm{prox}_\psi : X \to X$, i.e.,

$$\mathrm{prox}_\psi(x) \coloneqq \operatorname*{arg\,min}_{x' \in X}\{\tfrac{1}{2}\|x'-x\|_X^2 + \psi(x')\},$$

its subdifferential by $\partial\psi : X \rightrightarrows X$, i.e.,

$$\partial\psi(x) \coloneqq \{v \in X \mid \psi(x') \geq \psi(x) + (v, x'-x)_X \ \ \forall\, x' \in X\},$$

and its Fenchel conjugate by $\psi^* : X \to (-\infty,+\infty]$, i.e.,

$$\psi^*(v) \coloneqq \sup_{x\in X}\{(v,x)_X - \psi(x)\}.$$

If ψ is proper, closed and convex, then $\psi^{**} = \psi$ [16, Proposition 4.1]. We denote the effective domains of ψ and $\partial\psi$ by

$$\mathrm{dom}\,\psi \coloneqq \{x \in X \mid \psi(x) < +\infty\} \qquad \text{and} \qquad \mathrm{dom}\,\partial\psi \coloneqq \{x \in X \mid \partial\psi(x) \neq \emptyset\},$$

respectively. Finally, for a nonempty, closed and convex subset $C \subseteq X$, we denote the indicator and support functions of C by δ_C and σ_C, respectively, i.e.,

$$\delta_C(x) \coloneqq \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise} \end{cases} \qquad \text{and} \qquad \sigma_C(x) \coloneqq \sup_{v\in C}\,(v,x)_X.$$

Recall that $\sigma_C = \delta_C^* = \sigma_C^{**}$, the proximity operator of δ_C is the metric projection onto C, which we denote by proj_C, and that the subdifferential of δ_C is the normal cone

$$N_C(x) := \begin{cases} \{ v \in X \mid (v, x' - x)_X \le 0 \ \ \forall x' \in C\} & \text{if } x \in C \\ \emptyset & \text{otherwise.} \end{cases}$$

Notice that the second term in (1) is the support function $\sigma_{\mathfrak{A}}$.

We make the following assumptions on the problem data in (1).

Assumption 1 *The problem data in (1) satisfies the following conditions.*

1. *The function $\phi : X \to (-\infty, +\infty]$ is proper, closed and convex.*
2. *The function $f : X \to \mathbb{R}$ is M_f-smooth on $\operatorname{dom}\phi$, i.e., there exists an open set U containing $\operatorname{dom}\phi$ on which f is Fréchet differentiable and ∇f is Lipschitz continuous with modulus $M_f > 0$.*
3. *The set $\mathfrak{A} \subset Y$ is nonempty, closed, convex and bounded. We denote the bound on $\mathfrak{A}$ by $M_{\mathfrak{A}} := \sup_{\theta\in\mathfrak{A}} \|\theta\|_Y$.*
4. *The function $F : X \to Y$ is M_F-smooth on $\operatorname{dom}\phi$, i.e., there exists an open set V containing $\operatorname{dom}\phi$ on which F is Fréchet differentiable and F' is Lipschitz continuous with modulus $M_F > 0$.*
5. *The objective function $J = f + \sigma_{\mathfrak{A}} \circ F + \phi$ is bounded from below by $\kappa_{lb} \in \mathbb{R}$.*

Under Assumption 1, [17, Proposition 2.2.1] ensures that f and F are locally Lipschitz continuous. Consequently, this and the boundedness of $\mathfrak{A}$ ensure that $\sigma_{\mathfrak{A}} \circ F$ is also locally Lipschitz continuous. Therefore, [17, Proposition 2.3.3] and [17, Theorem 2.3.10] yield

$$\partial_C(f + \sigma_{\mathfrak{A}} \circ F)(x) = \nabla f(x) + F'(x)^* \partial\sigma_{\mathfrak{A}}(F(x))$$

for $x \in U \cap V$, where ∂_C denotes the Clarke subdifferential. It further follows from this, Corollary 1 of [17, Theorem 2.9.8] and [17, Theorem 2.9.9], that

$$\partial_C J(x) = \nabla f(x) + F'(x)^* \partial\sigma_{\mathfrak{A}}(F(x)) + \partial\phi(x) \tag{3}$$

for $x \in \operatorname{dom}\partial\phi$. Hence, [17, Proposition 2.4.11] provides a first-order necessary optimality condition for (1). In particular, if $\bar{x} \in X$ is a local minimizer of (1), then $\bar{x}$ satisfies

$$\exists\, \bar{\theta} \in \partial\sigma_{\mathfrak{A}}(F(\bar{x})) \quad \text{such that} \quad -(\nabla f(\bar{x}) + F'(\bar{x})^*\bar{\theta}) \in \partial\phi(\bar{x}), \tag{4}$$

which can be equivalently rewritten as

$$\exists\, \bar{\theta} \in \partial\sigma_{\mathfrak{A}}(F(\bar{x})) \quad \text{such that} \quad \bar{x} = \mathrm{prox}_{t\phi}(\bar{x} - t(\nabla f(\bar{x}) + F'(\bar{x})^*\bar{\theta})) \tag{5}$$

for arbitrary fixed $t > 0$. We will say that $\bar{x} \in X$ is a stationary point of (1) if it satisfies (5).

For the forthcoming analysis, it will be convenient to define the proper, closed and convex function $\psi_{\bar{x}} : X \to (-\infty, +\infty]$ for fixed $\bar{x} \in X$ defined by

$$\psi_{\bar{x}}(x) \coloneqq \sigma_{\mathfrak{A}}(F'(\bar{x})(x - \bar{x}) + F(\bar{x})) + \phi(x). \tag{6}$$

Using $\psi_{\bar{x}}$, we arrive at the following alternative characterization of stationary points for (1).

Theorem 2 *Let $\bar{x} \in X$ and consider $\psi_{\bar{x}}$ defined in (6). Then, $\bar{x}$ is a stationary point of (1) (i.e., (4) holds) if and only if*

$$\bar{x} = \mathrm{prox}_{t\psi_{\bar{x}}}(\bar{x} - t\nabla f(\bar{x})) \tag{7}$$

for arbitrary, fixed $t > 0$. In particular, if $\bar{x}$ is a stationary point of (1), then it is also a stationary point of the auxiliary optimization problem

$$\min_{x\in X} f(x) + \psi_{\bar{x}}(x). \tag{8}$$

Proof Leveraging the same arguments that produced (3), we obtain

$$\partial\psi_{\bar{x}}(x) = F'(\bar{x})^*\partial\sigma_{\mathfrak{A}}(F'(\bar{x})(x - \bar{x}) + F(\bar{x})) + \partial\phi(x).$$

From this follows the sequence of equivalences:

$$\begin{aligned}
(4) \quad &\iff \quad -\nabla f(\bar{x}) \in F'(\bar{x})^*\partial\sigma_{\mathfrak{A}}(F(\bar{x})) + \partial\phi(\bar{x}) = \partial\psi_{\bar{x}}(\bar{x}) \\
&\iff \quad \bar{x} - t\nabla f(\bar{x}) \in (\mathrm{Id} + t\partial\psi_{\bar{x}})(\bar{x}) \quad \forall\, t > 0 \\
&\iff \quad (7), \qquad \text{cf. [18, Example 23.3]}.
\end{aligned}$$

Here, $\mathrm{Id} \in \mathcal{L}(X)$ denotes the identity operator on X. To conclude, a stationary point $\bar{p} \in X$ for (8) satisfies

$$-\nabla f(\bar{p}) \in \partial\psi_{\bar{x}}(\bar{p}) = F'(\bar{x})^*\partial\sigma_{\mathfrak{A}}(F'(\bar{x})(\bar{p} - \bar{x}) + F(\bar{x})) + \partial\phi(\bar{p}). \tag{9}$$

Since $\bar{x}$ is a stationary point of (1), it satisfies (4) and therefore, substituting $\bar{x}$ for $\bar{p}$ in (9), we see that $\bar{x}$ is a stationary point of (8). □

Our trust-region algorithm leverages nonsmooth functions of the form

$$\psi_k(x) = \sigma_{\mathfrak{A}}(A_k(x - x_k) + b_k) + \phi(x),$$

where x_k is the k-th iterate, $A_k \in \mathcal{L}(X, Y)$ is an approximation of $F'(x_k)$ and $b_k \in Y$ is an approximation of $F(x_k)$ To prove convergence, we must relate the proximity operators of ψ_k with $\psi_{\bar{x}}$, which we can do leveraging the subsequent technical lemma.

Lemma 3 *Let $w_i : X \to Y$, $i = 1, 2$, be two affine maps defined by*

$$w_i(x) = D_i(x - u_i) + d_i, \quad i = 1, 2,$$

where $D_i \in \mathcal{L}(X, Y)$, $u_i \in X$ and $d_i \in Y$, $i = 1, 2$. Moreover, define $\Psi_i : X \to (-\infty, +\infty]$ by

$$\Psi_i(x) = \phi(x) + \sigma_{\mathfrak{A}}(w_i(x)), \quad i = 1, 2.$$

If $p_i = \mathrm{prox}_{t\Psi_i}(z)$, $i = 1, 2$, for fixed $t > 0$ and $z \in X$, then

$$\begin{aligned}
&\frac{\|p_1 - p_2\|_X^2}{\max\{1, \|p_1 - p_2\|_X\}} \\
&\le 2tM_{\mathfrak{A}}(\|D_1 - D_2\|_{\mathcal{L}(X,Y)}(1 + 2\|p_1 - u_1\|_X) + 2\|d_1 - d_2 - D_2(u_1 - u_2)\|_Y).
\end{aligned} \tag{10}$$

Proof Let $v_i : X \to (-\infty, +\infty]$ denote the objective function associated with the proximity operator of Ψ_i, i.e.,

$$v_i(x) = \tfrac{1}{2t}\|x - z\|_X^2 + \Psi_i(x), \quad i = 1, 2,$$

and notice that, for $x \in \operatorname{dom}\phi$,

$$v_1(x) - v_2(x) = \sigma_{\mathfrak{A}}(w_1(x)) - \sigma_{\mathfrak{A}}(w_2(x)).$$

Owing to the strong convexity of v_2 as well as the optimality of p_i, $i = 1, 2$, we have that

$$\begin{aligned}\tfrac{1}{2t}\|p_2 - p_1\|_X^2 &\le v_2(p_1) - v_2(p_2)\\ &\le [(v_1(p_2) - v_2(p_2)) - (v_1(p_1) - v_2(p_1))] + (v_1(p_1) - v_1(p_2))\\ &\le [(v_1(p_2) - v_2(p_2)) - (v_1(p_1) - v_2(p_1))]. \end{aligned} \tag{11}$$

It follows from the definition of the support function and the Fenchel-Young inequality that

$$(\theta_2(x), w_1(x) - w_2(x))_Y \le v_1(x) - v_2(x) \le (\theta_1(x), w_1(x) - w_2(x))_Y$$

for any $x \in X$ and $\theta_i(x) \in \partial\sigma_{\mathfrak{A}}(w_i(x))$, $i = 1, 2$. Consequently, we have that

$$[(v_1(x) - v_2(x)) - (v_1(y) - v_2(y))] \tag{12a}$$
$$\le (\theta_1(x), w_1(x) - w_2(x))_Y - (\theta_2(y), w_1(y) - w_2(y))_Y$$
$$= (\theta_1(x), (w_1(x) - w_2(x)) - (w_1(y) - w_2(y)))_Y \tag{12b}$$
$$+ (\theta_1(x) - \theta_2(y), w_1(y) - w_2(y))_Y \tag{12c}$$

for any x, $y \in \operatorname{dom}\phi$. To bound (12b), the boundedness of $\mathfrak{A}$ and the definitions of w_i yield

$$|(\theta_1(x), (w_1(x) - w_2(x)) - (w_1(y) - w_2(y)))_Y| \le M_{\mathfrak{A}}\|D_1 - D_2\|_{\mathcal{L}(X,Y)}\|x - y\|_X.$$

Now, to bound (12c), we have that

$$w_1(y) - w_2(y) = (D_1 - D_2)(y - u_1) + (d_1 - d_2 - D_2(u_1 - u_2)),$$

and so

$$\begin{aligned}&|(\theta_1(x) - \theta_2(y), w_1(y) - w_2(y))_Y|\\ &\qquad \le 2M_{\mathfrak{A}}(\|D_1 - D_2\|_{\mathcal{L}(X,Y)}\|y - u_1\|_X + \|d_1 - d_2 - D_2(u_1 - u_2)\|_Y).\end{aligned}$$

Using these estimates, we can bound (12a) by

$$\begin{aligned}&|(v_1(x) - v_2(x)) - (v_1(y) - v_2(y))|\\ &\le M_{\mathfrak{A}}(\|D_1 - D_2\|_{\mathcal{L}(X,Y)}(\|x - y\|_X + 2\|y - u_1\|_X) + 2\|d_1 - d_2 - D_2(u_1 - u_2)\|_Y).\end{aligned} \tag{13}$$

Combining (11) with (13), we achieve the bound

$$\tfrac{1}{2t}\|p_2 - p_1\|_X^2 \le M_{\mathfrak{A}}(\|D_1 - D_2\|_{\mathcal{L}(X,Y)}(\|p_2 - p_1\|_X + 2\|p_1 - u_1\|_X) + 2\|d_1 - d_2 - D_2(u_1 - u_2)\|_Y).$$

Applying Hölder's inequality to the right-hand side yields the bound (10). □

3 Algorithm

We now present an iterative method for solving (1). At each iteration of our method, we minimize a local model of the objective function J in (1) within a ball of radius $\Delta_k > 0$ called the *trust region*. To construct this subproblem, we employ a quadratic approximation of the Lagrangian functional

$$L(x, \theta) \coloneqq f(x) + (\theta, F(x))_Y.$$

Notice that L is continuously differentiable and that $x \mapsto L(x, \theta)$ has Lipschitz continuous gradient. We approximate L around the current iterate x_k by

$$L_k(x, \theta) \coloneqq q_k(x) + (\theta, \ell_k(x))_Y,$$

where $\theta_k \in \mathfrak{A}$ and

$$\begin{aligned} q_k(x) &\coloneqq \frac{1}{2}(B_k(x - x_k), x - x_k)_X + (g_k, x - x_k)_X \\ \ell_k(x) &\coloneqq A_k(x - x_k) + b_k. \end{aligned}$$

Here, $A_k \approx F'(x_k)$, $b_k \approx F(x_k)$, $g_k \approx \nabla f(x_k)$, and $B_k \in \mathcal{L}(X)$ encapsulates the curvature of $\ell(\cdot, \theta_k)$ at x_k. Using L_k, we approximate the objective function J by the local model

$$m_k(x) \coloneqq \sup_{\theta \in \mathfrak{A}} L_k(x, \theta) + \phi(x) = q_k(x) + \psi_k(x), \tag{14}$$

where

$$\psi_k(x) \coloneqq \sigma_{\mathfrak{A}}(\ell_k(x)) + \phi(x).$$

Note that the function ψ_k involves the sum of two potentially nonsmooth, yet convex, terms $\sigma_{\mathfrak{A}} \circ \ell_k$ and ϕ. To compute trial iterates x_k^+, we approximately solve the trust-region subproblems

$$\min_{x \in X} \{m_k(x) \coloneqq q_k(x) + \psi_k(x)\} \quad \text{subject to} \quad \|x - x_k\|_X \le \Delta_k, \tag{15}$$

where $\Delta_k > 0$ is the current trust-region radius.

To ensure convergence, we require that for each k the trial iterate x_k^+ satisfies

$$\|x_k^+ - x_k\|_X \le \kappa_{\text{rad}} \Delta_k \tag{16a}$$

$$m_k(x_k) - m_k(x_k^+) \ge \kappa_{\text{fcd}} h_k \min\left\{\frac{h_k}{1 + \|B_k\|_{\mathcal{L}(X)}}, \Delta_k\right\}, \tag{16b}$$

where $\kappa_{\text{rad}} > 0$ and $\kappa_{\text{fcd}} > 0$ are independent of k and h_k is the stationarity metric

$$h_k \coloneqq \frac{1}{t_k}\|\text{prox}_{t_k \psi_k}(x_k - t_k g_k) - x_k\|_X, \quad t_k > 0. \tag{17}$$

Note that (16b) ensures that $x_k^+ \in \operatorname{dom} \phi$ since otherwise the left-hand side would be $-\infty$. The positive parameter $t_k > 0$ in (17) is chosen as the step size for the Cauchy point discussed in Section 3.1. Once x_k^+ is computed, we decide whether or not to accept or reject x_k^+ based on the ratio of actual and predicted reduction:

$$\rho_k^* \coloneqq \frac{\text{ared}_k}{\text{pred}_k},$$

where

$$\text{ared}_k \coloneqq J(x_k) - J(x_k^+) \qquad \text{and} \qquad \text{pred}_k \coloneqq m_k(x_k) - m_k(x_k^+).$$

In many applications, the values and derivatives of f and F cannot be numerically computed exactly. For optimization problems governed by systems of PDEs, f and F may require the discretization and iterative solution of the underlying PDEs [13]. Similarly, in risk-averse stochastic programming, evaluating the risk measure requires approximation, which can be done in some cases using quadrature [7, 8, 19, 20]. To leverage this inexactness, we enforce the following conditions on g_k, b_k and A_k.

Condition 4 *There exist positive constants κ_{grad}, κ_{val} and κ_{jac}, independent of k, such that*

$$\|g_k - \nabla f(x_k)\|_X \le \kappa_{\text{grad}} \min\{h_k, \Delta_k\} \tag{18a}$$

$$\|b_k - F(x_k)\|_Y \le \kappa_{\text{val}} \min\{h_k, \Delta_k^2\} \tag{18b}$$

$$\|A_k - F'(x_k)\|_{\mathcal{L}(X,Y)} \le \kappa_{\text{jac}} \min\{h_k, \Delta_k\}. \tag{18c}$$

In these settings, the objective function value J, and thus ared_k, cannot be computed exactly. Instead, we replace ared_k with the *computed reduction* cred_k. We enforce the following condition on the approximation cred_k.

Condition 5 *There exists a positive constant κ_{obj}, independent of k, such that*

$$|\text{ared}_k - \text{cred}_k| \le \kappa_{\text{obj}} [\eta \min\{\text{pred}_k, \zeta_k\}]^\zeta, \tag{19}$$

where ζ, η, and ζ_k are (user-specified) positive real numbers that satisfy

$$\zeta > 1, \qquad 0 < \eta < \min\{\eta_1, 1 - \eta_2\}, \qquad \textit{and} \qquad \lim_{k \to \infty} \zeta_k = 0.$$

Using cred_k, we replace the ratio of actual and predicted reduction ρ_k^* with

$$\rho_k \coloneqq \frac{\text{cred}_k}{\text{pred}_k}.$$

As shown in [8, Lemma A.1], Condition 5 ensures that

$$\exists\, K_\eta \in \mathbb{N} \qquad \text{such that} \qquad |\rho_k^* - \rho_k| \le \eta \qquad \forall\, k \ge K_\eta. \tag{20}$$

We list the trust-region method in Algorithm 1 and in the forthcoming subsections, we discuss methods to efficiently compute the trial iterate x_k^+ in line 3 of Algorithm 1.

3.1 Cauchy Point

In traditional trust-region methods, the trial iterate x_k^+ is selected to produce at least a fraction of the model decrease achieved by a benchmark called the Cauchy point. For smooth unconstrained optimization, the Cauchy point is chosen as a point in the negative gradient direction that produces *sufficient* decrease of the (usually quadratic) model [21], while for convex-constrained trust-region methods, the Cauchy point is

Algorithm 1 TR-Risk Algorithm

Input: Initial guess $x_1 \in \operatorname{dom}\phi$, initial radius $\Delta_1 > 0$, $0 < \eta_1 < \eta_2 < 1$, and $0 < \gamma_1 \le \gamma_2 < 1 \le \gamma_3$

for $k = 1, 2, \dots$ **do**
  Model Selection: Choose A_k, b_k, and g_k that satisfy Condition 4
  Step Computation: Compute $x_k^+ \in X$ that satisfies (16)
  Step Acceptance and Radius Update: Compute the ratio of computed and predicted reduction ρ_k with cred_k satisfying Condition 5
  if $\rho_k < \eta_1$ **then**
    Set $x_{k+1} \leftarrow x_k$
    Choose $\Delta_{k+1} \in [\gamma_1\Delta_k, \gamma_2\Delta_k]$
  else
    Set $x_{k+1} \leftarrow x_k^+$
    if $\rho_k \in [\eta_1, \eta_2)$ **then**
      Choose $\Delta_{k+1} \in [\gamma_2\Delta_k, \Delta_k]$
    else
      Choose $\Delta_{k+1} \in [\Delta_k, \gamma_3\Delta_k]$
    end if
  end if
end for

a point along the projected gradient path [22]. For this work, we define the Cauchy point for (15) as

$$x_k^c := x_k + \alpha_k s_k(t_k) \qquad \text{with} \qquad s_k(t) := \operatorname{prox}_{t\psi_k}(x_k - t g_k) - x_k, \tag{21}$$

where $\alpha_k \in (0, 1]$ and $t_k > 0$ are selected to satisfy (16). When minimizing $f + \phi$, two basic approaches for selecting α_k and t_k are presented in [14, 23]. These approaches apply, without modification, to (15). The approach described in [14, Algorithm 2] sets $\alpha_k = 1$ and selects the step length t_k to satisfy certain sufficient decrease conditions using a bidirectional proximal search. In contrast, the method described in [23] first computes t_k as a safeguarded spectral (also called Barzilai-Borwein) step length and then selects α_k to minimize a quadratic upper bound of the model. In our numerical experiments, we employ the approach described in [23].

In order to compute x_k^c, we must be able to evaluate the proximity operator of ψ_k. However, ψ_k involves the sum of two convex functions, one of which involves the composition with an affine map. As a result, ψ_k is unlikely to admit an analytical proximity operator. If the proximity operator of ϕ is available and the projection onto $\mathfrak{A}$ is computable, then we can approximately compute the proximity operator of ψ_k iteratively by applying, e.g., projected gradient ascent to the dual problem. To this end, we recall that the dual problem associated with the optimization problem defining the proximity operator of ψ_k is given by

$$\max_{\theta\in\mathfrak{A}} \min_{x\in X} \{ \tfrac{1}{2r}\|x - z\|_X^2 + \phi(x) + (\theta, A_k(x - x_k) + b_k)_Y \}, \quad r > 0. \tag{22}$$

Notice that Sion's minimax theorem [24] ensures that there is no duality gap. Moreover, the inner minimization problem in (22) has the unique minimizer $\bar{p}(\theta)$, which satisfies the optimality conditions

$$\begin{aligned} 0 \in \tfrac{1}{r}(\bar{p}(\theta) - z) + A_k^*\theta + \partial\phi(\bar{p}(\theta)) \quad &\iff \quad (z - rA_k^*\theta) \in \bar{p}(\theta) + r\partial\phi(\bar{p}(\theta)) \\ &\iff \quad \bar{p}(\theta) = \operatorname{prox}_{r\phi}(z - rA_k^*\theta). \end{aligned}$$

Substituting $\bar{p}(\theta)$ into (22) yields the dual objective function

$$d(\theta) := \tfrac{1}{2r}\|\bar{p}(\theta) - z\|_X^2 + \phi(\bar{p}(\theta)) + (\theta, A_k(\bar{p}(\theta) - x_k) + b_k)_Y,$$

which is concave and differentiable with Lispchitz continuous gradient given by

$$\nabla d(\theta) = A_k(\operatorname{prox}_{r\phi}(z - rA_k^*\theta) - x_k) + b_k = \ell_k(\operatorname{prox}_{r\phi}(z - rA_k^*\theta))$$

and Lipschitz modulus $r\|A_k\|_{\mathcal{L}(X,Y)}^2$. One can maximize d using, e.g., the spectral projected gradient method [25], which we list in Algorithm 2. The following proposition

---

Algorithm 2 ψ_k Proximity Operator Computation

Input: Proximity operator arguments $z \in X$ and $r > 0$, initial guess $\theta^{(1)} \in \mathfrak{A}$, safeguards $0 < \gamma_{\min} < \gamma_{\max} < +\infty$, initial steplengths $\gamma^{(1)} \in [\gamma_{\min}, \gamma_{\max}]$ and $\lambda_0 \in (0, 1]$, and $0 < \sigma_1 < \sigma_2 < 1$, positive parameters $\alpha \in (0, 1)$ and $\beta \in [\sigma_1, \sigma_2]$, and nonmonotone memory limit $M \in \mathbb{N}$

for $n = 1, 2, \ldots$ **do**
    $p^{(n)} \leftarrow \operatorname{prox}_{r\phi}(z - rA_k^*\theta^{(n)})$
    $\mu^{(n)} \leftarrow \ell_k(p^{(n)})$
    if $n > 1$ **then**
        $\gamma^{(n)} \leftarrow \min\left\{\gamma_{\max}, \max\left\{\gamma_{\min}, \dfrac{\lambda^{(n-1)}\|s^{(n-1)}\|_Y^2}{(\mu^{(n-1)} - \mu^{(n)}, s^{(n-1)})_Y}\right\}\right\}$
    end if
    $s^{(n)} \leftarrow \operatorname{proj}_{\mathfrak{A}}(\theta^{(n)} + \gamma^{(n)}\mu^{(n)}) - \theta^{(n)}$
    $\lambda^{(n)} \leftarrow \lambda_0$
    $d_{\min} \leftarrow \min\{d(\theta^{(n-j)}) \mid 0 \le j \le \min\{n, M-1\}\}$
    while $d(\theta^{(n)} + \lambda^{(n)}s^{(n)}) < d_{\min} + \alpha\lambda^{(n)}(\mu^{(n)}, s^{(n)})_Y$ **do**
        $\delta \leftarrow -\dfrac{\frac{1}{2}(\lambda^{(n)})^2(\mu^{(n)}, s^{(n)})_Y}{d(\theta^{(n)} + \lambda^{(n)}s^{(n)}) - d(\theta^{(n)}) - \lambda^{(n)}(\mu^{(n)}, s^{(n)})_Y}$
        if $\delta \in [\sigma_1\lambda^{(n)}, \sigma_2\lambda^{(n)}]$ **then**
            $\lambda^{(n)} \leftarrow \delta$
        else
            $\lambda^{(n)} \leftarrow \beta\lambda^{(n)}$
        end if
    end while
    $\theta^{(n+1)} \leftarrow \theta^{(n)} + \lambda^{(n)}s^{(n)}$
end for

---

summarizes convergence results for Algorithm 2.

Proposition 6 *Let $\{(p^{(n)}, \theta^{(n)})\}$ denote the sequence of primal and dual iterates generated by Algorithm 2, then $\theta^{(n)} \rightharpoonup \bar{\theta}$, where $\bar{\theta} \in \mathfrak{A}$ is a solution to the dual problem (22). In addition, suppose one of the following conditions holds:*

(a) A_k is a compact operator;

(b) Y is finite dimensional;

(c) There exists $\varepsilon \in (0, \min\{1, (r\|A_k\|^2_{\mathcal{L}(X,Y)})^{-1}\})$ for which

$$\varepsilon \le \gamma^{(n)} \le \frac{2}{r\|A_k\|^2_{\mathcal{L}(X,Y)}} - \varepsilon \qquad \text{and} \qquad \varepsilon \le \lambda^{(n)} \le 1.$$

Then, $p^{(n)} \to \mathrm{prox}_{r\psi_k}(z) = \mathrm{prox}_{r\phi}(z - rA_k^*\bar{\theta})$.

Proof Weak convergence follows from [26, Theorem 3.2], where we note that the proofs of [26, Proposition 3.1 & Theorem 3.1] are easily modified to handle the more general nonmonotic line search in Algorithm 2. To prove the final claim, we first note that [18, Proposition 19.5] ensures that $\bar{p} = \mathrm{prox}_{r\psi_k}(z) = \mathrm{prox}_{r\phi}(z - rA_k^*\bar{\theta})$. Consequently, if (a) holds, then A_k^* is compact, which implies that $A_k^*\theta^{(n)} \to A_k^*\bar{\theta}$ and the result follows from the continuity of the proximity operator. Similarly, if (b) holds, then $\theta^{(n)} \to \bar{\theta}$ and again the result follows. Finally, if (c) holds, then by firm nonexpansivity [18, Proposition 4.4], we have that

$$\begin{aligned}
\|p^{(n)} - \bar{p}\|_X^2 &\le (-A_k^*\theta^{(n)} + A_k^*\bar{\theta}, \mathrm{prox}_{r\phi}(z - rA_k^*\theta^{(n)}) - \bar{p})_X \\
&= (\theta^{(n)} - \bar{\theta}, -A_k\mathrm{prox}_{r\phi}(z - rA_k^*\theta^{(n)}) + A_k\bar{p})_Y \\
&= (\theta^{(n)} - \bar{\theta}, \nabla d(\bar{\theta}) - \nabla d(\theta^{(n)}))_Y \\
&\le \|\theta^{(n)} - \bar{\theta}\|_Y \|\nabla d(\bar{\theta}) - \nabla d(\theta^{(n)})\|_Y.
\end{aligned}$$

Since $\{\theta^{(n)}\}$ is weakly convergent, $\{\|\theta^{(n)} - \bar{\theta}\|_Y\}$ is bounded. In addition, [18, Proposition 28.13] ensures that $\nabla d(\theta^{(n)}) \to \nabla d(\bar{\theta})$, proving the claim. □

Our subsequent convergence analysis assumes that the proximity operator of ψ_k is computed exactly. However, Algorithm 2 is only guaranteed to produce inexact evaluation, which motivates research on methods that can leverage inexact proximity operator evaluations [27, 28]. As we will see in the numerical results, Algorithm 2 typically produces sufficiently accurate approximations with only modest effort.

3.2 Trial Iterate Computation

To compute a trial step x_k^+ that improves upon the Cauchy point x_k^c and satisfies (16), we can employ any of the methods introduced in [23], including the dogleg, spectral proximal gradient, or truncated conjugate gradient algorithms. In our numerical results, we employ the truncated conjugate gradient method [23, Algorithm 4]. This method first computes the Cauchy point x_k^c and then improves upon it using modified nonlinear conjugate gradient iterations.

4 Convergence Theory

In this section, we prove convergence of Algorithm 1 assuming the proximity operator $\mathrm{prox}_{t\psi_k}$ is computed exactly. The technical results are partitioned into two classes. First, we prove global convergence of Algorithm 1 in the sense that $\{h_k\}$ accumulates at zero. These results leverage traditional trust-region proof techniques like those used to prove [14, Theorem 2]. We then postulate additional assumptions on the problem data that ensure the sequence of iterates $\{x_k\}$ generated by Algorithm 1 accumulate at a stationary point of (1). For these results, we employ the following notation to distinguish between quantities depending on approximations (i.e., b_k, A_k and g_k) and those that do not:

$$\begin{aligned}
\hat{\ell}_k(x) &\coloneqq F'(x_k)(x - x_k) + F(x_k) \\
\hat{\psi}_k(x) &\coloneqq \phi(x) + \sigma_{\mathfrak{A}}(\hat{\ell}_k(x)) \\
\hat{H}_k(t) &\coloneqq \tfrac{1}{t}\|x_k - \mathrm{prox}_{t\hat{\psi}_k}(x_k - t\nabla f(x_k))\|_X \\
H_k(t) &\coloneqq \tfrac{1}{t}\|x_k - \mathrm{prox}_{t\psi_k}(x_k - tg_k)\|_X.
\end{aligned}$$

Notice that $h_k = H_k(t_k)$ and we similarly denote $\hat{h}_k = \hat{H}_k(t_k)$.

4.1 Global Convergence

To prove global convergence of Algorithm 1, we extend the theory in [14]. In the subsequent lemma, we update [14, Lemma 8] to account for the more general nonsmooth term ψ_k.

Lemma 7 *Fix $k \geq K_\eta$ with K_η defined in (20) and let Assumption 1 hold. If $h_k > 0$ and*

$$(1 + \|B_k\|_{\mathcal{L}(X)})\Delta_k \leq \kappa_{\mathrm{vs}} h_k \tag{23}$$

with $\kappa_{\mathrm{vs}} \in (0, 1)$ defined by

$$\kappa_{\mathrm{vs}} \coloneqq \frac{\kappa_{\mathrm{fcd}}(1 - \eta_2 - \eta)}{\max\{\kappa_{\mathrm{fcd}}, \frac{1}{2}\kappa_{\mathrm{rad}}^2, \frac{M_f}{2}\kappa_{\mathrm{rad}}^2 + \kappa_{\mathrm{grad}}\kappa_{\mathrm{rad}} + M_{\mathfrak{A}}(\frac{M_F}{2}\kappa_{\mathrm{rad}}^2 + \kappa_{\mathrm{jac}}\kappa_{\mathrm{rad}} + 2\kappa_{\mathrm{val}})\}},$$

then $\rho_k \geq \eta_2$ and $\Delta_{k+1} \geq \Delta_k$.

Proof We first bound the difference between pred_k and ared_k. To this end, let $s_k = x_k^+ - x_k$ and recall that

$$\begin{aligned}
\mathrm{pred}_k - \mathrm{ared}_k &= (q_k(x_k) + \sigma_{\mathfrak{A}}(b_k) + \phi(x_k) - q_k(x_k^+) - \sigma_{\mathfrak{A}}(\ell_k(x_k^+)) - \phi(x_k^+)) \\
&\quad - (f(x_k) + \sigma_{\mathfrak{A}}(F(x_k)) + \phi(x_k) - f(x_k^+) - \sigma_{\mathfrak{A}}(F(x_k^+)) - \phi(x_k^+)) \\
&= (q_k(x_k) - q_k(x_k^+)) - (f(x_k) - f(x_k^+)) \\
&\quad + (\sigma_{\mathfrak{A}}(b_k) - \sigma_{\mathfrak{A}}(\ell_k(x_k^+)) - (\sigma_{\mathfrak{A}}(F(x_k)) - \sigma_{\mathfrak{A}}(F(x_k^+))).
\end{aligned}$$

We individually bound the smooth and nonsmooth terms. For the smooth term, we have that

$$(q_k(x_k) - q_k(x_k^+)) - (f(x_k) - f(x_k^+))$$

$$
\begin{aligned}
&= -\tfrac{1}{2}(B_k s_k, s_k)_X - (g_k, s_k)_X + \int_0^1 (\nabla f(x_k + t s_k), s_k)_X \, \mathrm{d}t \\
&\le \tfrac{1}{2}\|B_k\|_{\mathcal{L}(X)}\|s_k\|_X^2 + \tfrac{M_f}{2}\|s_k\|_X^2 + \kappa_{\mathrm{grad}}\|s_k\|_X \Delta_k \\
&\le (\tfrac{1}{2}\kappa_{\mathrm{rad}}^2\|B_k\|_{\mathcal{L}(X)} + \tfrac{M_f}{2}\kappa_{\mathrm{rad}}^2 + \kappa_{\mathrm{grad}}\kappa_{\mathrm{rad}})\Delta_k^2,
\end{aligned}
$$

where we used Assumption 1, Condition 4, and (16a). For the nonsmooth term, Condition 4 and (16a) ensure that

$$
\sigma_{\mathfrak{A}}(b_k) - \sigma_{\mathfrak{A}}(F(x_k)) \le M_{\mathfrak{A}}\|b_k - F(x_k)\|_Y \le M_{\mathfrak{A}}\kappa_{\mathrm{val}}\Delta_k^2
$$

and we have that

$$
\begin{aligned}
\sigma_{\mathfrak{A}}(F(x_k^+)) - \sigma_{\mathfrak{A}}(\ell_k(x_k^+)) = &(\sigma_{\mathfrak{A}}(F(x_k^+)) - \sigma_{\mathfrak{A}}(F(x_k) + F'(x_k)s_k)) \\
&+ (\sigma_{\mathfrak{A}}(F(x_k) + F'(x_k)s_k) - \sigma_{\mathfrak{A}}(\ell_k(x_k^+))).
\end{aligned} \tag{24}
$$

Again using Condition 4, we can bound the second term on the right-hand side of (24) as

$$
\begin{aligned}
&\sigma_{\mathfrak{A}}(F(x_k) + F'(x_k)s_k) - \sigma_{\mathfrak{A}}(\ell_k(x_k^+)) \\
&\qquad \le M_{\mathfrak{A}}(\|(F'(x_k) - A_k)s_k\|_Y + \|b_k - F(x_k)\|_Y) \\
&\qquad \le M_{\mathfrak{A}}(\kappa_{\mathrm{jac}}\|s_k\|_X\Delta_k + \kappa_{\mathrm{val}}\Delta_k^2) \\
&\qquad \le M_{\mathfrak{A}}(\kappa_{\mathrm{jac}}\kappa_{\mathrm{rad}} + \kappa_{\mathrm{val}})\Delta_k^2,
\end{aligned}
$$

Finally, to bound the first term on the right-hand side of (24), we first recall that support functions are subadditive[1] and so

$$
\sigma_{\mathfrak{A}}(F(x_k^+)) - \sigma_{\mathfrak{A}}(F(x_k) + F'(x_k)s_k)) \le \sigma_{\mathfrak{A}}(F(x_k^+) - F(x_k) - F'(x_k)s_k).
$$

To bound $\sigma_{\mathfrak{A}}(F(x_k^+) - F(x_k) - F'(x_k)s_k)$, Assumption 1 ensures that

$$
\begin{aligned}
\sigma_{\mathfrak{A}}(F(x_k^+) - F(x_k) - F'(x_k)s_k) &= \sup_{\theta\in\mathfrak{A}} \int_0^1 (\theta, (F'(x_k + t s_k) - F'(x_k))s_k)_Y \, \mathrm{d}t \\
&\le \tfrac{1}{2}M_{\mathfrak{A}}M_F\|s_k\|_X^2 \\
&\le \tfrac{1}{2}M_{\mathfrak{A}}M_F\kappa_{\mathrm{rad}}^2\Delta_k^2,
\end{aligned}
$$

where we have used the definition of $\sigma_{\mathfrak{A}}$ and the mean value theorem applied to the map $t \mapsto (\theta, F(x_k + t s_k))_Y$ for fixed $\theta \in \mathfrak{A}$ to arrive at the first equality. Combining these bounds yields

$$
|\mathrm{pred}_k - \mathrm{ared}_k| \le \frac{\kappa_{\mathrm{fcd}}(1 - \eta_2 - \eta)}{\kappa_{\mathrm{vs}}}(1 + \|B_k\|_{\mathcal{L}(X)})\Delta_k^2, \tag{25}
$$

where we have used Hölder inequality. Combining (25) and (16b), we obtain

$$
|\rho_k^* - 1| \le \frac{\kappa_{\mathrm{fcd}}(1 - \eta_2 - \eta)}{\kappa_{\mathrm{vs}}} \frac{(1 + \|B_k\|_{\mathcal{L}(X)})\Delta_k^2}{\kappa_{\mathrm{fcd}} h_k \min\left\{\frac{h_k}{1 + \|B_k\|_{\mathcal{L}(X)}}, \Delta_k\right\}} \le (1 - \eta_2 - \eta).
$$

Consequently, we have that $\rho_k^* \ge \eta_2 + \eta$ and $\rho_k \ge \eta_2$ as was to be shown. □

Replacing [14, Lemma 8] with Lemma 7 in the proof of [14, Theorem 3] enables us to prove the convergence of Algorithm 1.

---

[1] The subadditivity of $\sigma_{\mathfrak{A}}$ ensures that

$$
\sigma_{\mathfrak{A}}(y) = \sigma_{\mathfrak{A}}(y' + (y - y')) \le \sigma_{\mathfrak{A}}(y') + \sigma_{\mathfrak{A}}(y - y') \quad \forall\, y,\, y' \in Y.
$$

Theorem 8 *Let $\{x_k\}$ be the sequence of iterates generated by Algorithm 1. If Assumption 1 holds and if*

$$\sum_{k=1}^{\infty}(1+\max_{j=1,\ldots,k}\|B_j\|_{\mathcal{L}(X)})^{-1}=+\infty,$$

then

$$\liminf_{k\to\infty}\ h_k=0.$$

In addition, if there exists $t_{\max}>0$ satisfying $t_k\le t_{\max}$ for all k, then

$$\liminf_{k\to\infty}\ H_k(t)=0\quad\forall\, t>0.$$

Proof The proof of this result is nearly identical to the proof of [14, Theorem 3, Equation (40)] with [14, Lemma 8] replaced by Lemma 7. Finally, the monotonicity of $t\mapsto H_k(t)$ (cf. [14, Lemma 2]) and the upper bound on t_k ensure that $H_k(t_{\max})\le h_k$ proving the final result. □

In our first corollary to Theorem 8, we relate the convergence of h_k and $H_k(t)$ with $\hat{h}_k$ and $\hat{H}_k(t)$.

Corollary 9 *Let the assumptions of Theorem 8 hold. If there exist $0<t_{\min}\le t_{\max}$ such that $t_k\in[t_{\min},t_{\max}]$ for all k then*

$$\liminf_{k\to\infty}\ \hat{h}_k=0\qquad\textit{and}\qquad\liminf_{k\to\infty}\ \hat{H}_k(t)=0\quad\forall\, t>0.$$

Proof We first prove a bound between the proximity operators associated with ψ_k and $\hat{\psi}_k$. Let $p_k=\mathrm{prox}_{t_k\psi_k}(x_k-t_k\nabla f(x_k))$ and $\hat{p}_k=\mathrm{prox}_{t_k\hat{\psi}_k}(x_k-t_k\nabla f(x_k))$. Applying Lemma 3 with $\Psi_1=\psi_k$ and $\Psi_2=\hat{\psi}_k$ and noting that

$$\begin{cases}\|D_1-D_2\|_{\mathcal{L}(X,Y)}=\|A_k-F'(x_k)\|\le\kappa_{\mathrm{jac}}h_k\\ \|p_1-u_1\|_X=\|p_k-x_k\|_X\le\kappa_{\mathrm{grad}}t_kh_k+t_kh_k\le t_{\max}(\kappa_{\mathrm{grad}}+1)h_k\\ \|d_1-d_2-D_2(u_1-u_2)\|_Y=\|b_k-F(x_k)\|_Y\le\kappa_{\mathrm{val}}h_k,\end{cases}$$

we arrive at the bound

$$\frac{1}{t_k}\frac{\|p_k-\hat{p}_k\|_X^2}{\max\{1,\|p_k-\hat{p}_k\|_X\}}\le 2M_{\mathfrak{A}}(\kappa_{\mathrm{jac}}(1+2t_{\max}(\kappa_{\mathrm{grad}}+1)h_k)+2\kappa_{\mathrm{val}})h_k.$$

Notice that this and the bounds on t_k imply that

$$\liminf_{k\to\infty}\frac{1}{t_k}\|p_k-\hat{p}_k\|_X=0.$$

The nonexpansivity of the proximity operator and Condition 4 then ensure that

$$\begin{aligned}&\liminf_{k\to\infty}\tfrac{1}{t_k}\|x_k-\mathrm{prox}_{t_k\hat{\psi}_k}(x_k-t_k\nabla f(x_k))\|_X\\ &\quad\le\liminf_{k\to\infty}\tfrac{1}{t_k}\|x_k-\mathrm{prox}_{t_k\psi_k}(x_k-t_k\nabla f(x_k))\|_X+\liminf_{k\to\infty}\tfrac{1}{t_k}\|p_k-\hat{p}_k\|_X\\ &\quad\le\liminf_{k\to\infty}(\kappa_{\mathrm{grad}}+1)h_k=0.\end{aligned}$$

The final result follows from the monotonicity of $t\mapsto H_k(t)$ [14, Lemma 2]. In particular, $H_k(t)\le h_k$ for all $t\ge t_{\max}$ and for any $t\le t_{\max}$, $H_k(t)\le\frac{t_{\max}}{t}H_k(t_{\max})\le\frac{t_{\max}}{t}h_k$. □

Under additional assumptions, we can strengthen the lower limit in Theorem 8 to a limit.

Theorem 10 *Let the assumptions of Corollary 9 hold. If there exists* $\kappa_{\text{curv}} > 0$ *satisfying* $\|B_k\|_{\mathcal{L}(X)} \le \kappa_{\text{curv}}$ *for all* $k \in \mathbb{N}$*, then*

$$\lim_{k\to\infty} h_k = \lim_{k\to\infty} H_k(t) = \lim_{k\to\infty} \hat{H}_k(t) = 0 \quad \forall\, t > 0.$$

Proof The proof of this is similar to the proof of [15, Theorem 1] with modifications to account for the more general objective function. Let $\mathcal{S} \coloneqq \{k \in \mathbb{N} \,|\, \rho_k \ge \eta_1\}$ denote the set of indices corresponding to successful iterations. If $\mathcal{S}$ is finite, then $x_k = \bar{x}$ for $k \ge |\mathcal{S}|$ sufficiently large and $\Delta_k \to 0$. Define $p_k = \text{prox}_{t\psi_k}(\bar{x} - t\nabla f(\bar{x}))$ and $\bar{p}(t) = \text{prox}_{t\psi_{\bar{x}}}(\bar{x} - t\nabla f(\bar{x})) = \text{prox}_{t\hat{\psi}_k}(x_k - t\nabla f(x_k))$ for fixed $t > 0$ and $k \ge |\mathcal{S}|$. Applying Lemma 3 with $\Psi_1 = \psi_k$ and $\Psi_2 = \hat{\psi}_k = \psi_{\bar{x}}$ and noting that (via Condition 4)

$$\begin{cases} \|D_1 - D_2\|_{\mathcal{L}(X,Y)} = \|A_k - F'(\bar{x})\| \le \kappa_{\text{jac}}\Delta_k \\ \|p_1 - u_1\|_X = \|p_k - x_k\|_X \le \kappa_{\text{grad}} t\Delta_k + t\Delta_k \le t(\kappa_{\text{grad}} + 1)\Delta_k \\ \|d_1 - d_2 - D_2(u_1 - u_2)\|_Y = \|b_k - F(\bar{x})\|_Y \le \kappa_{\text{val}}\Delta_k^2, \end{cases}$$

we arrive at the bound

$$\frac{1}{t}\frac{\|p_k - \bar{p}(t)\|_X^2}{\max\{1, \|p_k - \bar{p}(t)\|_X\}} \le 2M_{\mathfrak{A}}(\kappa_{\text{jac}}(1 + 2t(\kappa_{\text{grad}} + 1)\Delta_k) + 2\kappa_{\text{val}}\Delta_k)\Delta_k,$$

which implies that

$$\lim_{k\to\infty} \frac{1}{t}\|p_k - \bar{p}(t)\|_X = 0.$$

Consequently, we have that

$$|H_k(t) - \tfrac{1}{t}\|\bar{p}(t) - \bar{x}\|_X| \le \kappa_{\text{grad}}\Delta_k + \tfrac{1}{t}\|p_k - \bar{p}(t)\|_X$$

and so $H_k(t) \to \frac{1}{t}\|\bar{p}(t) - \bar{x}\|_X$ and the desired result follows from monotonicity of $t \mapsto H_k(t)$ and the arguments in the proof of Corollary 9.

Now suppose that $\mathcal{S}$ is infinite and set $t = t_{\max}$. To arrive at a contradiction, we assume that there exists $\epsilon > 0$ and a subsequence $\mathcal{K} \subseteq \mathcal{S}$ satisfying

$$h_k \ge H_k(t) \ge 2\epsilon > 0 \quad \forall\, k \in \mathcal{K}. \tag{26}$$

By Theorem 8, for each $k \ge \mathcal{K}$ there exists $\ell > k$ for which $H_\ell(t) < \epsilon$. Let $k^+ > k$ denote the first such index and define

$$\mathcal{S}_0 \coloneqq \{j \in \mathcal{S} \,|\, k \le j < k^+ \;\; \forall\, k \in \mathcal{K}\}.$$

Note that for all $j \in \mathcal{S}_0$, we have that $H_j(t) \ge \epsilon$. By [14, Lemma 2], there exists $K_\eta \in \mathbb{N}$ such that $|\rho_j^* - \rho_j| \le \eta$ for all $j \ge K_\eta$. This, the boundedness of $\{\|B_k\|_{\mathcal{L}(X)}\}$ and (16b) ensure that

$$J(x_j) - J(x_{j+1}) \ge (\eta_1 - \eta)\kappa_{\text{fcd}}\epsilon \min\left\{\frac{\epsilon}{1 + \kappa_{\text{curv}}}, \Delta_j\right\}$$

for all $j \in \mathcal{S}_0$ with $j \ge K_\eta$. Since $\{J(x_k)\}$ is monotonically decreasing and bounded below, we have that the left-hand side converges to zero and hence so does Δ_j. Consequently, for sufficiently large $j \in \mathcal{S}_0$, we obtain

$$\Delta_j \le \frac{J(x_{j+1}) - J(x_j)}{\kappa_{\text{fcd}}\epsilon(\eta_1 - \eta)}.$$

For sufficiently large $k \in \mathcal{K}$, this and the triangle inequality ensure that

$$\|x_k - x_{k^+}\|_X \le \sum_{j=k}^{k^+-1} \|x_j - x_{j+1}\|_X \le \kappa_{\text{rad}} \sum_{j=k}^{k^+-1} \Delta_j \le \frac{\kappa_{\text{rad}}[J(x_k) - J(x_{k^+})]}{\kappa_{\text{fcd}}\epsilon(\eta_1 - \eta)}.$$

The right-hand side converges to zero and therefore so does the left-hand side as $\mathcal{K} \ni k \to \infty$. Lipschitz continuity then ensures that $\|\nabla f(x_k) - \nabla f(x_{k^+})\|_X$ and $\|F'(x_k) - F'(x_{k^+})\|$ tend to zero as $\mathcal{K} \ni k \to \infty$. Now, for $k \in \mathcal{K}$, the reverse triangle inequality yields

$$\begin{aligned}\epsilon &\le |H_k(t) - H_{k^+}(t)| \\ &\le \frac{1}{t}\|x_k - x_{k^+}\|_X + \frac{1}{t}\|\text{prox}_{t\psi_k}(x_k - tg_k) - \text{prox}_{t\psi_{k^+}}(x_{k^+} - tg_{k^+})\|_X.\end{aligned}$$

Nonexpansivity of the proximity operator then implies that

$$\begin{aligned}\|\text{prox}_{t\psi_k}&(x_k - tg_k) - \text{prox}_{t\psi_{k^+}}(x_{k^+} - tg_{k^+})\|_X \\ &\le \|x_k - x_{k^+}\|_X + t\|g_k - g_{k^+}\|_X \\ &\quad + \|\text{prox}_{t\psi_k}(x_k - tg_k) - \text{prox}_{t\psi_{k^+}}(x_k - tg_k)\|_X. \qquad (27)\end{aligned}$$

Owing to Condition 4 and Lipschitz continuity, we have that

$$\|g_k - g_{k^+}\|_X \le \kappa_{\text{grad}}(\Delta_k + \Delta_{k^+}) + M_f\|x_k - x_{k^+}\|_X.$$

It remains to bound the final term on the right-hand side of (27). For this, we employ Lemma 3 with $\Psi_1 = \psi_{k^+}$ and $\Psi_2 = \psi_k$. In this setting, we have that

$$\begin{cases}\|D_1 - D_2\|_{\mathcal{L}(X,Y)} = \|A_{k^+} - A_k\|_{\mathcal{L}(X,Y)} \\ \|p_1 - u_1\|_X = \|\text{prox}_{t\psi_{k^+}}(x_k - tg_k) - x_{k^+}\|_X \le \|x_k - x_{k^+}\|_X + t\|g_k - g_{k^+}\|_X + tH_{k^+}(t) \\ \|d_1 - d_2 - D_2(u_1 - u_2)\|_Y = \|b_{k^+} - b_k - A_k(x_{k^+} - x_k)\|_Y.\end{cases}$$

Note that $\|A_{k^+} - A_k\|_{\mathcal{L}(X,Y)}$ tends to zero because of Condition 4 and the fact that Δ_k and $\|x_k - x_{k^+}\|_X$ tend to zero. That is,

$$\begin{aligned}\|A_{k^+} - A_k\|_{\mathcal{L}(X,Y)} &\le \|A_k - F'(x_k)\|_{\mathcal{L}(X,Y)} + \|A_{k^+} - F'(x_{k^+})\|_{\mathcal{L}(X,Y)} \\ &\quad + \|F'(x_k) - F'(x_{k^+})\|_{\mathcal{L}(X,Y)} \\ &\le \kappa_{\text{jac}}(\Delta_k + \Delta_{k^+}) + M_F\|x_k - x_{k^+}\|_X.\end{aligned}$$

Using similar arguments combined with the mean value theorem, we can bound

$$\begin{aligned}\|b_{k^+} &- b_k - A_k(x_{k^+} - x_k)\|_Y \\ &\le \|b_k - F(x_k)\|_Y + \|b_{k^+} - F(x_{k^+})\|_Y + \|A_k - F'(x_k)\|_{\mathcal{L}(X,Y)}\|x_k - x_{k^+}\|_X \\ &\quad + \|F(x_{k^+}) - F(x_k) - F'(x_k)(x_{k^+} - x_k)\|_Y \\ &\le \kappa_{\text{val}}(\Delta_k^2 + \Delta_{k^+}^2) + \kappa_{\text{jac}}\Delta_k\|x_k - x_{k^+}\|_X + \tfrac{1}{2}M_F\|x_k - x_{k^+}\|_X^2.\end{aligned}$$

Finally, there exists $c > 0$, independent of k, such that $\|\text{prox}_{t\psi_{k^+}}(x_k - tg_k) - x_{k^+}\|_X \le c$ for all k since $\|x_k - x_{k^+}\|_X$ and $\|g_k - g_{k^+}\|_X$ converge to zero, and $H_{k^+}(t) \le h_{k^+} \le \varepsilon$. Defining

$$\begin{aligned}p_k &= \text{prox}_{t\psi_k}(x_k - tg_k), \qquad \tilde{p}_k = \text{prox}_{t\psi_{k^+}}(x_k - tg_k), \\ \mu_k &= M_{\mathfrak{A}}(\kappa_{\text{jac}}(\Delta_k + \Delta_{k^+}) + M_F\|x_k - x_{k^+}\|_X) \quad \text{and} \\ \nu_k &= M_{\mathfrak{A}}(\kappa_{\text{val}}(\Delta_k^2 + \Delta_{k^+}^2) + \kappa_{\text{jac}}\Delta_k\|x_k - x_{k^+}\|_X + \tfrac{1}{2}M_F\|x_k - x_{k^+}\|_X^2),\end{aligned}$$

we arrive at the bound

$$\frac{1}{t}\frac{\|p_k - \tilde{p}_k\|_X^2}{\max\{1, \|p_k - \tilde{p}_k\|_X\}} \le 2((1 + 2tc)\mu_k + \nu_k).$$

Notice that $\mu_k \xrightarrow{\mathcal{K}} 0$ and $\nu_k \xrightarrow{\mathcal{K}} 0$. Therefore, $\|p_k - \tilde{p}_k\|_X \xrightarrow{\mathcal{K}} 0$. As a result, $|H_k(t) - H_{k^+}(t)|$ converges to zero, arriving at a contradiction. Hence, no such subsequence exists, implying that $H_k(t) \to 0$ and the desired result follows as before. □

4.2 Convergence to Stationary Points

To prove convergence to a stationary point, we leverage results of the previous subsection. However, we require additional regularity of the problem data. The principle challenge is in demonstrating convergence of the proximity operators of ψ_k to the proximity operator of $\psi_{\bar{x}}$, where $\bar{x}$ is a stationary point of (1). The following proposition postulates conditions on F for which this property is valid.

Proposition 11 *Let $\{x_k\}$ be the sequence of iterates generated by Algorithm 1 and suppose Assumption 1 holds. Moreover, suppose there exists $\bar{x} \in X$ and an index set $\mathcal{K}$ such that $\{x_k\}_{\mathcal{K}}$ is bounded, $h_k \xrightarrow{\mathcal{K}} 0$, $F'(x_k) \xrightarrow{\mathcal{K}} F'(\bar{x})$, and*

$$F(x_k) - F(\bar{x}) - F'(\bar{x})(x_k - \bar{x}) \xrightarrow{\mathcal{K}} 0.$$

Then,

$$\mathrm{prox}_{t\psi_k}(x) \xrightarrow{\mathcal{K}} \mathrm{prox}_{t\psi_{\bar{x}}}(x) \quad \forall\, x \in X, \;\; t > 0. \tag{28}$$

Proof Fix $x \in X$ and $t > 0$. Lemma 3 with $\Psi_1 = \psi_{x_k}$ and $\Psi_2 = \psi_{\bar{x}}$ (i.e., $D_1 = A_k$, $d_1 = b_k$, $u_1 = x_k$, $D_2 = F'(\bar{x})$, $d_2 = F(\bar{x})$ and $u_2 = \bar{x}$), combined with Condition 4, yields

$$\begin{cases} \|D_1 - D_2\|_{\mathcal{L}(X,Y)} \le \kappa_{\mathrm{jac}} h_k + \|F'(x_k) - F'(\bar{x})\|_{\mathcal{L}(X,Y)} \\ \|p_1 - u_1\|_X = \|\mathrm{prox}_{t\psi_k}(x) - x_k\|_X \\ \|d_1 - d_2 - D_2(u_1 - u_2)\|_Y \le \kappa_{\mathrm{val}} h_k + \|F(x_k) - F(\bar{x}) - F'(\bar{x})(x_k - \bar{x})\|_Y. \end{cases}$$

If $\{\|\mathrm{prox}_{t\psi_k}(x) - x_k\|_X\}_{\mathcal{K}}$ is bounded, then our assumptions ensure that

$$\lim_{\mathcal{K} \ni k \to 0} \frac{\|\mathrm{prox}_{t\psi_k}(x) - \mathrm{prox}_{t\psi_{\bar{x}}}(x)\|_X^2}{\max\{1, \|\mathrm{prox}_{t\psi_k}(x) - \mathrm{prox}_{t\psi_{\bar{x}}}(x)\|_X\}} = 0,$$

and the desired result holds. As such, we will prove that $\{\|\mathrm{prox}_{t\psi_k}(x) - x_k\|_X\}_{\mathcal{K}}$ is bounded. To this end, we note that

$$\begin{aligned} |(\psi_k(z_1) - \psi_{\bar{x}}(z_1)) - (\psi_k(z_2) - \psi_{\bar{x}}(z_2))| &\le |(\sigma_{\mathfrak{A}}(\ell_k(z_1)) - \sigma_{\mathfrak{A}}(\ell_k(z_2))| \\ &\quad + |(\sigma_{\mathfrak{A}}(F(\bar{x}) + F'(\bar{x})(z_1 - \bar{x})) - \sigma_{\mathfrak{A}}(F(\bar{x}) + F'(\bar{x})(z_2 - \bar{x}))| \\ &\le M_{\mathfrak{A}}(\|A_k\|_{\mathcal{L}(X,Y)} + \|F'(\bar{x})\|_{\mathcal{L}(X,Y)})\|z_1 - z_2\|_X \end{aligned}$$

for all z_1, $z_2 \in X$ and therefore [29, Proposition 4.32] ensures that

$$\|\mathrm{prox}_{t\psi_k}(x) - \mathrm{prox}_{t\psi_{\bar{x}}}(x)\|_X \le t M_{\mathfrak{A}}(\|A_k\|_{\mathcal{L}(X,Y)} + \|F'(\bar{x})\|_{\mathcal{L}(X,Y)}).$$

Applying Condition 4 yields

$$\|A_k\|_{\mathcal{L}(X,Y)} \le \kappa_{\mathrm{jac}} h_k + \|F'(x_k) - F'(\bar{x})\|_{\mathcal{L}(X,Y)} + \|F'(\bar{x})\|_{\mathcal{L}(X,Y)},$$

which is bounded for $k \in \mathcal{K}$ since $F'(x_k) \xrightarrow{\mathcal{K}} F'(\bar{x})$ and $h_k \xrightarrow{\mathcal{K}} 0$. Therefore, we have that

$$\|\mathrm{prox}_{t\psi_k}(x)\|_X \le \|\mathrm{prox}_{t\psi_k}(x) - \mathrm{prox}_{t\psi_{\bar{x}}}(x)\|_X + \|\mathrm{prox}_{t\psi_{\bar{x}}}(x)\|_X$$

is bounded for $k \in \mathcal{K}$, proving the claim. □

In the subsequent corollaries, we provide conditions under which the assumptions of Proposition 11 hold. Our first corollary is particularly useful when X is finite dimensional as it assumes the existance of a strongly converging subsequence, while the second corollary is tailored to PDE-constrained optimization applications in which F and F' are completely continuous. Recall that completely continuous functions map weakly converging sequences into strongly converging sequences [30].

Corollary 12 *Let $\{x_k\}$ be the sequence of iterates generated by Algorithm 1 and suppose Assumption 1 holds. Moreover, suppose that $\bar{x} \in X$ and $\mathcal{K}$ are such that $h_k \xrightarrow{\mathcal{K}} 0$ and $x_k \xrightarrow{\mathcal{K}} \bar{x}$. Then (28) holds.*

Proof Assumption 1 and strong convergence ensure that $\{x_k\}_{\mathcal{K}}$ is bounded,

$$F(x_k) \xrightarrow{\mathcal{K}} F(\bar{x}) \quad \text{and} \quad F'(x_k) \xrightarrow{\mathcal{K}} F'(\bar{x}).$$

The result then follows from Proposition 11. □

Corollary 13 *Let $\{x_k\}$ be the sequence of iterates generated by Algorithm 1 and suppose Assumption 1 holds. Moreover, suppose that $\bar{x} \in X$ and $\mathcal{K}$ are such that $h_k \xrightarrow{\mathcal{K}} 0$ and $x_k \xrightarrow{\mathcal{K}} \bar{x}$ (weakly). Finally, assume that F and F' are completely continuous. Then (28) holds.*

Proof Weak convergence ensures that $\{x_k\}_{\mathcal{K}}$ is bounded and the assumed complete continuity ensures that

$$F(x_k) \xrightarrow{\mathcal{K}} F(\bar{x}) \quad \text{and} \quad F'(x_k) \xrightarrow{\mathcal{K}} F'(\bar{x}).$$

Moreover, $F'(\bar{x})$ is a completely continuous operator [30, Theorem 1.5.1] and therefore, $F'(\bar{x})(x_k - \bar{x}) \xrightarrow{\mathcal{K}} 0$. The result then follows from Proposition 11. □

Combining Corollaries 12 and 13 with Theorem 8 enables use to prove convergence to stationarity points of (1). Note that a consequence of Theorem 8 is that there exists a subsequence of the stationarity measures $\{h_{k_j}\}$ that converges to 0. Our first result demonstrates that strong accumulations points are stationary.

Theorem 14 *Let $\{x_k\}$ be the sequence of iterates generated by Algorithm 1 and let the assumptions of Theorem 8 hold. Moreover, suppose there exists $\bar{x} \in X$ and an index set $\mathcal{K} \subseteq \mathbb{N}$ such that $h_k \xrightarrow{\mathcal{K}} 0$ and $x_k \xrightarrow{\mathcal{K}} \bar{x}$. If there exists $t_{\max} > 0$ with $t_k \le t_{\max}$ for all k, then $\bar{x}$ is a stationary point of (1). In particular, (5) holds for all $t > 0$.*

Proof By Corollary 12, (28) holds. Using Condition 4, we have that

$$x_k - tg_k \xrightarrow{\mathcal{K}} \bar{x} - t\nabla f(\bar{x})$$

for fixed $t > 0$ and so the nonexspansivity of the proximity operator ensures that

$$\begin{aligned}\|\mathrm{prox}_{t\psi_k}(x_k - tg_k) - &\mathrm{prox}_{t\psi_{\bar{x}}}(\bar{x} - t\nabla f(\bar{x}))\|_X \\ &\leq \|(x_k - tg_k) - (\bar{x} - t\nabla f(\bar{x}))\|_X \\ &\quad + \|\mathrm{prox}_{t\psi_k}(\bar{x} - t\nabla f(\bar{x})) - \mathrm{prox}_{t\psi_{\bar{x}}}(\bar{x} - t\nabla f(\bar{x}))\|_X.\end{aligned}$$

In particular, we have that

$$\mathrm{prox}_{t\psi_k}(x_k - tg_k) \xrightarrow{\mathcal{K}} \mathrm{prox}_{t\psi_{\bar{x}}}(\bar{x} - t\nabla f(\bar{x})).$$

Now, Theorem 8 ensures that

$$H_k(t) = \|\mathrm{prox}_{t\psi_k}(x_k - tg_k) - x_k\|_X \xrightarrow{\mathcal{K}} 0$$

and therefore $\bar{x} = \mathrm{prox}_{t\psi_{\bar{x}}}(\bar{x} - t\nabla f(\bar{x}))$. The result then follows from Theorem 2. □

Theorem 14 is a useful result when X is finite dimensional. However, strong convergence is difficult to guarantee in infinite-dimensional Hilbert spaces where bounded sequences are only guaranteed to admit weak accumulation points. As demonstrated by the following theorem, this setting requires additional assumptions on f and F.

Theorem 15 *Let $\{x_k\}$ be the sequence of iterates generated by Algorithm 1 and let the assumptions of Corollary 9 hold. Moreover, suppose there exists $\bar{x} \in X$ and an index set $\mathcal{K} \subseteq \mathbb{N}$ such that $h_k \xrightarrow{\mathcal{K}} 0$ and $x_k \xrightharpoonup{\mathcal{K}} \bar{x}$. Finally, suppose that f is strongly convex and F and F' are completely continuous. If $\bar{x}$ is a stationary point for (1), then $x_k \xrightarrow{\mathcal{K}} \bar{x}$.*

Proof Let $\omega > 0$ denote the strong convexity modulus of f. To simplify the presentation, define

$$G_k(x,t) \coloneqq \tfrac{1}{t}(x - \mathrm{prox}_{t\psi_k}(x - t\nabla f(x))).$$

By [15, Lemma 2], we have that $G_k(\cdot, t)$ is a strongly monotone operator for all $t \in (0, 2\omega M_f^{-2})$, which yields

$$(\omega - \tfrac{1}{2}tM_f^{-2})\|x_k - \bar{x}\|_X^2 \leq (G_k(x_k,t) - G_k(\bar{x},t), x_k - \bar{x})_X.$$

Applying the Cauchy-Schwarz and triangle inequalities, we obtain

$$(\omega - \tfrac{1}{2}tM_f^{-2})\|x_k - \bar{x}\|_X \leq \|G_k(x_k,t)\|_X + \|G_k(\bar{x},t)\|_X.$$

By Corollary 9 and Condition 4, we have that

$$\lim_{\mathcal{K}\ni k\to\infty} \|G_k(x_k,t)\|_X \leq \lim_{\mathcal{K}\ni k\to\infty} (\kappa_{\mathrm{grad}} + 1)H_k(t) = 0$$

and by Corollary 13, in particular (28), we have that

$$\lim_{\mathcal{K}\ni k\to\infty} \|G_k(\bar{x},t)\|_X = \tfrac{1}{t}\|\bar{x} - \mathrm{prox}_{t\psi_{\bar{x}}}(\bar{x} - t\nabla f(\bar{x}))\|_X.$$

Since $\bar{x}$ is stationary, Theorem 2 ensures that $\bar{x} = \mathrm{prox}_{t\psi_{\bar{x}}}(\bar{x} - t\nabla f(\bar{x}))$ and therefore

$$\lim_{\mathcal{K}\ni k\to\infty} \|x_k - \bar{x}\|_X = 0,$$

proving the desired result. □

Theorem 15 demonstrates strong convergence when the weak limit of $\{x_k\}_{\mathcal{K}}$ is a stationary point. When f includes a Tikhonov regularization term, we can prove that this assumption is satisfied.

Theorem 16 *Let the assumptions of Corollary 9 hold and assume that*

$$f(x) = \frac{\tau}{2}\|x\|_X^2 + f_0(x),$$

where $\tau > 0$, $f_0 : X \to \mathbb{R}$ is Fréchet differentiable, and ∇f_0, F and F' are completely continuous. Moreover, suppose there exists $\bar{x} \in X$ and an index set $\mathcal{K}$ such that $h_k \xrightarrow{\mathcal{K}} 0$ and $x_k \xrightarrow{\mathcal{K}} \bar{x}$. Then, $\bar{x}$ is a stationary point of (1). In addition, if f_0 is τ_0-weakly convex with $\tau_0 \in (0, \tau)$, i.e.,

$$x \mapsto \frac{\tau_0}{2}\|x\|_X^2 + f_0(x)$$

is convex, then $x_k \xrightarrow{\mathcal{K}} \bar{x}$.

Proof Set $t = \frac{1}{\tau}$. Using Condition 4 and the assumed form of f, we have that

$$\begin{aligned}
\|(x_k - tg_k) - &(\bar{x} - t\nabla f(\bar{x}))\|_X \\
&\le t\|g_k - \nabla f(x_k)\|_X + \|(x_k - t\nabla f(x_k)) - (\bar{x} - t\nabla f(\bar{x})\|_X \\
&= t\|g_k - \nabla f(x_k)\|_X + t\|\nabla f_0(x_k) - \nabla f_0(\bar{x})\|_X \\
&\le t\kappa_{\text{grad}} h_k + t\|\nabla f_0(x_k) - \nabla f_0(\bar{x})\|_X
\end{aligned}$$

and hence

$$x_k - tg_k \xrightarrow{\mathcal{K}} \bar{x} - t\nabla f(\bar{x}).$$

Therefore, Corollary 13 and the nonexspansivity of the proximity operator yield

$$\text{prox}_{t\psi_k}(x_k - tg_k) \xrightarrow{\mathcal{K}} \text{prox}_{t\psi_{\bar{x}}}(\bar{x} - t\nabla f(\bar{x})).$$

Now, Theorem 8 ensures that

$$\|\text{prox}_{t\psi_k}(x_k - tg_k) - x_k\|_X \xrightarrow{\mathcal{K}} 0.$$

Therefore, the weak lower semicontinuity of the norm yields

$$0 = \liminf_{\mathcal{K}\ni k\to\infty} \|\text{prox}_{t\psi_k}(x_k - tg_k) - x_k\|_X \ge \|\text{prox}_{t\psi_{\bar{x}}}(\bar{x} - t\nabla f(\bar{x})) - \bar{x}\|_X$$

and so $\bar{x} = \text{prox}_{t\psi_{\bar{x}}}(\bar{x} - t\nabla f(\bar{x}))$, implying stationarity of $\bar{x}$. To conclude, if f_0 is τ_0-weakly convex with $\tau_0 < \tau$, then f is strongly convex and strong convergence follows from Theorem 15. □

Our final convergence result builds on Theorem 16, proving that when the Hessian sequence $\{B_k\}$ is bounded and J is convex, the entire sequence of iterates $\{x_k\}$ converges to a stationary point of (1).

Corollary 17 *Let the assumptions of Theorems 10 and 16 hold. Moreover, suppose J is convex and admits a unique minimizer $\bar{x} \in X$. If $\{x_k\}$ is bounded, then $x_k \to \bar{x}$.*

Proof Let $\tilde{x}$ be any weak accumulation point of $\{x_k\}$, which exists since $\{x_k\}$ is bounded, and let $\mathcal{K}$ be an index set on which $x_k \xrightarrow{\mathcal{K}} \tilde{x}$. By Theorem 10, $h_k \xrightarrow{\mathcal{K}} 0$ and by Theorem 15, we have that $x_k \xrightarrow{\mathcal{K}} \tilde{x}$ and $\tilde{x}$ is a stationary point of (1). Convexity of J ensures that the stationarity conditions (5) are necessary and sufficient for optimality, so $\tilde{x} = \bar{x}$. Furthermore, boundedness of $\{x_k\}$ ensures that every subsequence has a further weakly converging subsequence with limit $\bar{x}$. Consequently, $x_k \to \bar{x}$. □

5 Numerics

In this section, we demonstrate the numerical performance of Algorithm 1 on various examples from PDE-constrained optimization. Throughout, $D \subset \mathbb{R}^d$, $d = 1, 2$, is the physical domain with boundary ∂D and $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space with set of outcomes Ω, σ-algebra of events $\mathcal{F} \subseteq 2^\Omega$ and probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$. The optimization space in all examples is $X = L^2(D)$, while the space Y in the first example is $Y = L^2(\Omega, \mathcal{F}, \mathbb{P})$ and $Y = \mathbb{R}$ in the second. For each example, we employ the truncated conjugate gradients subproblem solver developed in [23, Algorithm 4] with the number of iterations capped at 15, lower and upper safeguards 10^{-12} and 10^{12}, respectively, descent parameter 10^{-4}, and relative and absolute tolerances 10^{-2} and 10^{-4}, respectively. In the notation of [23, Algorithm 4], these parameters are `maxit`, $t_{\min}$, $t_{\max}$, η, τ_k and $\bar{\tau}$, respectively. Note that this choice of subproblem solver ensures that the Cauchy point steplengths t_k satisfy $10^{-12} = t_{\min} \le t_k \le t_{\max} = 10^{12}$ for all k. In addition, we use Algorithm 2 to compute the subproblem proximity operator. For this, we set $\lambda_0 = 1$, $\gamma_{\min} = 10^{-6}$, $\gamma_{\max} = 10^6$, $\sigma_1 = 0.1$, $\sigma_2 = 0.9$, $\alpha = 10^{-4}$, $\beta = 0.5$ and $M = 1$ (i.e., monotonic line search). We terminate Algorithm 2 when

$$\frac{\|s^{(n)}\|_Y}{\gamma^{(n)}} \le 10^{-10}.$$

For Algorithm 1, we set $\Delta_1 = 10$, $\eta_1 = 10^{-4}$, $\eta_2 = 0.5$, $\gamma_1 = \gamma_2 = 0.25$, and $\gamma_3 = 10$. Finally, we terminate Algorithm 1 when

$$h_k \le 10^{-8}. \tag{29}$$

A consequence of Theorem 8 is that (29) is guaranteed to be satisfied after finitely many iterations.

5.1 Risk-Averse Control of Burger's Equation

The goal of this example is to compare our approach with the primal-dual risk minimization algorithm [12] and to demonstrate its ability to control inexact PDE solves via Conditions 4 and 5. Let $D = (0, 1)$ and $\tau > 0$, and consider the optimization problem

$$\min_{z \in L^2(D)} \; \mathcal{R}\left(\tfrac{1}{2}\|S(z) - 1\|^2_{L^2(D)}\right) + \tfrac{\tau}{2}\|z\|^2_{L^2(D)}, \tag{30}$$

where $u = S(z) : \Omega \to H^1(D)$ solves the weak form of Burger's equation

$$-\nu(\omega)\partial_{xx}u(\omega) + u(\omega)\partial_x u(\omega) = f(\omega) + z \qquad \text{in } D \text{ a.s.} \tag{31a}$$

$$[u(\omega)](0) = d_0(\omega), \quad [u(\omega)](1) = d_1(\omega) \qquad \text{a.s.} \tag{31b}$$

Here, ν, f, d_0 and d_1 are the random functions specified in [12]. For this example, $\sigma_{\mathfrak{A}} = \mathcal{R}$ is a convex combination of the expected value and the average value-at-risk[2]

$$\mathcal{R}(y) = (1 - \lambda)\mathbb{E}[y] + \lambda \, \mathrm{AVaR}_p[y]$$

with $\lambda = 0.75$ and $p = 0.9$, i.e., $\mathfrak{A} = \{\theta \in L^2(\Omega, \mathcal{F}, \mathbb{P}) \,|\, \mathbb{E}[\theta] = 1, \;\; 0.25 \leq \theta \leq 7.75 \text{ a.s.}\}$. We discretize the state u and the control z in space using continuous piecewise linear finite elements on a uniform mesh with 257 intervals and solve the resulting nonlinear system of equations using Newton's method globalized with a backtracking line search. Finally, we approximate the risk measure $\mathcal{R}$ using sample average approximation (SAA) with 10,000 Monte Carlo samples.

We summarize the performance of Algorithm 1 in Table 1 and include the performance of the primal-dual risk minimization algorithm (`PD-Risk`) [12] and the bundle method (`Bundle`) described in [31] for comparison. The results for `PD-Risk` and `Bundle` can also be found in [12, Tables 6 & 7]. In Table 1, `iter` is the number of iterations,

method	iter	nfval	ngrad	nhess
Alg. 1	7	8	8	119
PD-Risk	8	46	44	128
Bundle	69	182	182	---

Table 1: Comparison of Algorithm 1 with the primal-dual risk minimization algorithm `PD-Risk` and a bundle method `Bundle`: `iter` is the number of iterations, `nfval` is the number of function evaluations, `ngrad` is the number of gradient computations and `nhess` is the number of Hessian applications.

`nfval` is the number of evaluations of f and F, `ngrad` is the number of derivative evaluations for f and F, and `nhess` is the number of Hessian applications for f and F. Note that each gradient evaluation requires 10,000 deterministic adjoint (linear) PDE solves and each Hessian application requires 20,000 additional deterministic linearized PDE solves. Consequently, Algorithm 1 required 80,000 deterministic nonlinear PDEs to evaluate the objective function value and 2,460,000 deterministic linear PDE solves to evaluate and apply the derivatives. In contrast, `PD-Risk` required 460,000 deterministic nonlinear PDE solves and 3,000,000 deterministic linear PDE solves, while `Bundle` required 1,820,000 deterministic nonlinear and linear PDE solves. Although `Bundle` performed fewer linear PDE solves to compute derivatives, each nonlinear PDE solve requires multiple linear solves. When accounting for the linearized solves required to solve the state, `PD-Risk` required a total of 4,824,834 deterministic linear PDE solves, compared to 2,708,194 for Algorithm 1, and so Algorithm 1 reduced the

[2] Also called the conditional value-at-risk, expected shortfall and superquantile.

total number of deterministic linear PDE solves by a factor of 1.78. Algorithm 1 also provides additional benefits over `PD-Risk` since it can leverage inexact function and derivative evaluations.

Let $c_n(u, z) = 0$ denote the weak form of (31) and $u_k^{n,\ell}$ denote the ℓ-th iterate generated by Newton's method for solving c_n at the n-th sample and the k-iteration of Algorithm 1. Table 2 includes the iteration histories for two instances of Algorithm 1: The top half corresponds to the iteration history using highly accurate state PDE solves that we terminate based on the criterion

$$\|c_n(u_k^{n,\ell}, z_k)\| \le \tau \max\{1, \|c_k^{n,0}\|\},$$

where $\tau = 10^{-4}\sqrt{\epsilon_{\text{mach}}} \approx 1.49 \times 10^{-12}$ and ϵ_{mach} denotes machine precision, while the lower half is the iteration history where we have allowed Algorithm 1 to supply τ using Conditions 4 and 5 with $\kappa_{\text{grad}} = \kappa_{\text{val}} = \kappa_{\text{jac}} = 1$ and $\kappa_{\text{obj}} = 10^4$. We selected this value for κ_{obj} so that the initial tolerances generated by Conditions 4 and 5 were the same order of magnitude. Both the exact and inexact algorithms perform comparably, requiring the same number of iterations. However, the inexact variant leverages significantly relaxed PDE solver tolerances, reducing the total number of Newton iterations from 248,194 to 216,706.

k	$J(x_k)$	h_k	Δ_k	$\|x_k^+ - x_k\|$	`val tol`	`grad tol`	`itsp`
0	2.4289e-2	3.9059e-3	1.0000e+1	---	1.490e-12	1.490e-12	---
1	1.3441e-2	3.5555e-3	1.0000e+2	1.1950e+0	1.490e-12	1.490e-12	15
2	1.2453e-2	2.5784e-3	1.0000e+3	3.5392e-1	1.490e-12	1.490e-12	15
3	1.2151e-2	6.9818e-4	1.0000e+4	2.6859e-1	1.490e-12	1.490e-12	15
4	1.2129e-2	6.1039e-5	1.0000e+5	1.3944e-1	1.490e-12	1.490e-12	15
5	1.2128e-2	1.6595e-6	1.0000e+6	1.8164e-2	1.490e-12	1.490e-12	15
6	1.2128e-2	3.9913e-7	1.0000e+7	5.4855e-4	1.490e-12	1.490e-12	15
7	1.2128e-2	3.2924e-9	1.0000e+8	4.5861e-5	1.490e-12	1.490e-12	15
0	2.3866e-2	4.4689e-3	1.0000e+1	---	1.000e-2	4.469e-3	---
1	1.3256e-2	2.2653e-3	1.0000e+2	1.2599e+0	1.931e-3	2.265e-3	15
2	1.2478e-2	5.5037e-4	1.0000e+3	3.1316e-1	1.191e-4	5.504e-3	15
3	1.2132e-2	3.9939e-4	1.0000e+4	3.0440e-1	5.023e-5	3.994e-4	15
4	1.2128e-2	1.1822e-5	1.0000e+5	5.6905e-2	3.581e-7	1.182e-5	15
5	1.2128e-2	7.3445e-6	1.0000e+6	4.3233e-3	7.582e-10	7.344e-6	15
6	1.2128e-2	2.6303e-7	1.0000e+7	7.6969e-4	6.844e-11	2.630e-7	15
7	1.2128e-2	3.2007e-9	1.0000e+8	5.8116e-5	1.490e-12	3.201e-9	15

Table 2: Algorithm 1 iteration history using high-fidelity (rows 2 through 8) and inexact (rows 9 through 15) PDE solves. Columns 6 and 7 list the relative residual tolerance determined by the trust-region algorithm.

5.2 Sparse Optimal Control

The goal of this example is to demonstrate the mesh independence of Algorithm 1. Let $D = (0, 0.6) \times (0, 0.2)$. We consider a deterministic semilinear elliptic optimal control problem motivated by the one studied in [32]. Our control problem is given by

$$\min_{\substack{z \in L^2(D) \\ -10 \le z \le 10}} \max\left\{0, w - \frac{1}{|D_o|}\int_{D_o} S(z)\,\mathrm{d}x\right\} + \frac{\tau}{2}\|z\|^2_{L^2(D)} + \tau_1\|z\|_{L^1(D)}, \tag{32}$$

where $S(z) = u \in H^1(D)$ solves the weak form of the semilinear elliptic PDE

$$\begin{aligned} -\kappa\Delta u + \gamma u^3 &= 12\chi_{D_b} + z && \text{in } D \\ u &= 0 && \text{on } \Gamma = [0, 0.6] \times \{0\} \\ \kappa \nabla u \cdot n &= 0 && \text{on } \partial D \setminus \Gamma. \end{aligned} \tag{33}$$

Here, $D_b = (0, 0.1) \times (0.167, 0.2)$, $D_o = (0.5, 0.6) \times (0.167, 0.2)$, χ_{D_b} denotes the characteristic function of D_b, $\tau = 10^{-4}$, $\tau_1 = 10^{-2}$, $\kappa = 0.25$, $\gamma = 1.45$ and $w = 0.2$. For this problem, $\mathfrak{A} = [0, 1]$ and

$$F(z) = w - \frac{1}{|D_o|}\int_{D_o} S(z)\,\mathrm{d}x.$$

Moreover, ϕ is the sum of the L^1-penalty term and the indicator function associated with the bound constraints $-10 \le z \le 10$ a.e. We discretized u in (33) using P1 finite elements on a uniform triangular mesh and z using piecewise constants on the same mesh. As in the previous example, we solve the discretized nonlinear PDE using a linesearch globalized Newton's method. To investigate mesh independence, we performed a mesh refinement study, where we uniformly refined the initial mesh and solved the refined optimization problem from an initial guess of $z \equiv 0$. The initial mesh was generated by splitting the elements of a 60-by-20 uniform quadrilateral mesh into triangles. Subsequent meshes were generated by uniformly refining the initial quadrilateral mesh and then partitioning into triangles. Table 3 lists the results

mesh	iter	nfval	ngrad	nhess	npsi	nprox	aprox
60×20	2	3	3	34	1875	798	20.80
120×40	2	3	3	34	2084	772	20.06
240×80	2	3	3	34	2176	795	20.71
480×160	2	3	3	34	2192	788	20.51
960×320	2	3	3	34	2079	789	20.54
1920×640	2	3	3	34	2138	787	20.49

Table 3: Mesh refinement study of Algorithm 1 applied to the sparse control application, demonstrating mesh independence.

of this study. The columns of Table 3 include the total number of iterations (`iter`), the number of evaluations of f and F (`nfval`), the number of derivative computations for f and F (`ngrad`), the number of Hessian applications for f and F (`nhess`), the number of evaluations of ψ_k (`npsi`), the number of proximity operator evaluations for ψ_k (`nprox`) and the average number of iterations of Algorithm 2 (`aprox`). As seen in Table 3, the performance of Algorithm 1 is essentially constant, suggesting that Algorithm 1 is mesh independent. To achieve this, our implementation uses infinite-dimensionally consistent scaled inner products and norms for X and Y. In addition, we see that the average number of iterations of Algorithm 2 to compute the proximity operator of ψ_k is relatively modest, requiring roughly 20 iterations on average to achieve the requested tolerance of 10^{-10}.

6 Conclusions

We have introduced a provably convergent inexact trust-region algorithm, extended from [14], for the structured class of nonsmooth optimization problems represented by (1). This class encapsulates many important applications include risk-averse stochastic optimization and nonsmooth penalty methods for nonlinear programming. We verify through numerical examples that our method is quite efficient when solving discretized PDE-constrained optimization problems, demonstrating invariance to the discretization size and outperforming some existing nonsmooth methods. Although our method performs well in practice, the theoretical results rely heavily on the boundedness of $\mathfrak{A}$ and the exact computation of the proximity operator $\mathrm{prox}_{t\psi_k}$. Permitting unbounded $\mathfrak{A}$ would enable, e.g., the solution of nonlinearly constrained problems, while permitting inexact proximity operator evaluation may further reduce the computation burden. In future work, we hope to relax these requirements.

Acknowledgements. We would like to thank Thomas Surowiec for the many fruitful discussions about earlier versions of this work. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Declarations

- **Funding:** This research was sponsored, in part, by the U.S. Department of Energy Office of Science under the Early Career Research Program and the U.S. Air Force Office of Scientific Research.
- **Competing Interests:** The author has no competing interests to declare that are relevant to the content of this article.
- **Data Availability:** No data was generated or used for this article.

References

[1] Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology **58**(1), 267–288 (1996)

[2] Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer, Berlin, Heidelberg, New York (2006)

[3] Capatina, A.: Variational Inequalities and Frictional Contact Problems. Advances in Mechanics and Mathematics, vol. 31. Springer, Germany (2014)

[4] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D.: Coherent measures of risk. Math. Finance **9**(3), 203–228 (1999) https://doi.org/10.1111/1467-9965.00068

[5] Dentcheva, D., Ruszczyński, A.: Risk-Averse Optimization and Control: Theory and Methods. Springer Series in Operations Research and Financial Engineering Series. Springer, Germany (2024)

[6] Shapiro, A., Dentcheva, D., Ruszczynski, A.: Lectures on Stochastic Programming: Modeling and Theory, Second Edition. MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, Philadelphia (2014)

[7] Kouri, D.P., Heinkenschloss, M., Ridzal, D., Bloemen Waanders, B.G.: A trust-region algorithm with adaptive stochastic collocation for PDE optimization under uncertainty. SIAM Journal on Scientific Computing **35**(4), 1847–1879 (2013) https://doi.org/10.1137/120892362

[8] Kouri, D.P., Heinkenschloss, M., Ridzal, D., Bloemen Waanders, B.G.: Inexact objective function evaluations in a trust-region algorithm for PDE-constrained optimization under uncertainty. SIAM Journal on Scientific Computing **36**(6), 3011–3029 (2014) https://doi.org/10.1137/140955665

[9] Kouri, D.P., Shapiro, A.: In: Antil, H., Kouri, D.P., Lacasse, M.-D., Ridzal, D. (eds.) Optimization of PDEs with Uncertain Inputs, pp. 41–81. Springer, New York, NY (2018)

[10] Kouri, D.P., Surowiec, T.M.: Risk-averse PDE-constrained optimization using the conditional value-at-risk. SIAM Journal on Optimization **26**(1), 365–396 (2016) https://doi.org/10.1137/140954556

[11] Kouri, D.P., Surowiec, T.M.: Existence and optimality conditions for risk-averse PDE-constrained optimization. SIAM/ASA Journal on Uncertainty Quantification **6**(2), 787–815 (2018) https://doi.org/10.1137/16M1086613

[12] Kouri, D.P., Surowiec, T.M.: A primal–dual algorithm for risk minimization. Mathematical Programming **193**(1), 337–363 (2022)

[13] Hinze, M., Pinnau, R., Ulbrich, M., Ulbrich, S.: Optimization with PDE Constraints. Mathematical Modelling: Theory and Applications, vol. 23, p. 270. Springer, Germany (2009)

[14] Baraldi, R.J., Kouri, D.P.: A proximal trust-region method for nonsmooth optimization with inexact function and gradient evaluations. Mathematical Programming **201**(1), 1–40 (2022)

[15] Baraldi, R.J., Kouri, D.P.: Local convergence analysis of an inexact trust-region method for nonsmooth optimization. Optimization Letters **18**(3), 663–680 (2024)

[16] Ekeland, I., Temam, R.: Convex Analysis and Variational Problems. Classics in Applied Mathematics, Vol. 28. SIAM, Philadelphia (1999)

[17] Clarke, F.H.: Nonsmooth Analysis and Control Theory. Graduate Texts in Mathematics. Springer, Germany (1998)

[18] Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. CMS Books in Mathematics. Springer, Germany (2017)

[19] Chen, M., Mehrotra, S., Papp, D.: Scenario generation for stochastic optimization problems via the sparse grid method. Computational Optimization and applications **62**(3), 669–692 (2015)

[20] Heinkenschloss, M., Kouri, D.P.: Optimization problems governed by systems of pdes with uncertainties. Acta Numerica **34**, 491–577 (2025)

[21] Conn, A.R., Gould, N.I.M., Toint, P.L.: Trust–Region Methods. SIAM, Philadelphia (2000)

[22] Lin, C.-J., Moré, J.J.: Newton's method for large bound-constrained optimization problems. SIAM Journal on Optimization **9**(4), 1100–1127 (1999)

[23] Baraldi, R.J., Kouri, D.P.: Efficient subproblem solvers for a nonsmooth trust-region method. Computational Optimization and Applications, (2024)

[24] Sion, M.: On general minimax theorems. Pacific J. Math. **8**, 171–176 (1958)

[25] Birgin, E.G., Martínez, J.M., Raydan, M.: Nonmonotone spectral projected gradient methods on convex sets. SIAM Journal on Optimization **10**(4), 1196–1211 (2000)

[26] Bello Cruz, J.: On weak and strong convergence of the projected gradient method for convex optimization in real hilbert spaces. Numerical Functional Analysis and Optimization **37**(2), 129–144 (2016)

[27] Salzo, S., Villa, S., *et al.*: Inexact and accelerated proximal point algorithms. J. Convex Anal **19**(4), 1167–1192 (2012)

[28] Villa, S., Salzo, S., Baldassarre, L., Verri, A.: Accelerated and inexact forward-backward algorithms. SIAM Journal on Optimization **23**(3), 1607–1633 (2013) https://doi.org/10.1137/110844805

[29] Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer, Berlin, Heidelberg, New York (2000)

[30] Emelyanov, S.V., Korovin, S.K., Bobylev, N.A., Bulatov, A.V.: Homotopy of Extremal Problems: Theory and Applications. De Gruyter Series in Nonlinear Analysis and Applications. De Gruyter, Germany (2011)

[31] Schramm, H., Zowe, J.: A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, convergence analysis, numerical results. SIAM Journal on Optimization **2**(1), 121–152 (1992) https://doi.org/10.1137/0802008

[32] Kouri, D., Surowiec, T.: Risk-averse optimal control of semilinear elliptic pdes. ESAIM: Control, Optimisation and Calculus of Variations **26**, 53 (2020)