ALESQP: AN AUGMENTED LAGRANGIAN EQUALITY-CONSTRAINED SQP METHOD FOR OPTIMIZATION WITH GENERAL CONSTRAINTS *

HARBIR ANTIL †, DREW P. KOURI ‡, AND DENIS RIDZAL §

Abstract. We present a new algorithm for infinite-dimensional optimization with general constraints, called ALESQP. In short, ALESQP is an augmented Lagrangian method that penalizes inequality constraints and solves equality-constrained nonlinear optimization subproblems at every iteration. The subproblems are solved using a matrix-free trust-region sequential quadratic programming (SQP) method that takes advantage of iterative, i.e., inexact linear solvers and is suitable for large-scale applications. A key feature of ALESQP is a constraint decomposition strategy that allows it to exploit problem-specific variable scalings and inner products.

We analyze convergence of ALESQP under different assumptions. We show that strong accumulation points are stationary. Consequently, in finite dimensions ALESQP converges to a stationary point. In infinite dimensions we establish that weak accumulation points are feasible in many practical situations. Under additional assumptions we show that weak accumulation points are stationary.

We present several infinite-dimensional examples where ALESQP shows remarkable discretization-independent performance in all of its iterative components, requiring a modest number of iterations to meet constraint tolerances at the level of machine precision. Also, we demonstrate a fully matrix-free solution of an infinite-dimensional problem with nonlinear inequality constraints.

Key words. ALESQP; augmented Lagrangian; composite step trust-region method; SQP; convergence analysis; constraint decomposition; nonlinear constraints

AMS subject classifications. 49M37, 90C30, 90C39, 93C20, 49K20, 49J20

1. Introduction.

In this paper, we develop a provably convergent algorithm for solving optimization problems of the form

$$\min_{x\in X} f(x) \quad \text{subject to} \quad g(x)=0, \quad Tx\in\bigcap_{i=1}^{m} C_i, \tag{1.1}$$

where X and Y are real Banach spaces, Z is a real Hilbert space, $f : X \to \mathbb{R}$, $g : X \to Y$, and $T : X \to Z$ is a linear operator. Moreover, the sets $C_1, \ldots, C_m$ are nonempty, closed and convex, where $C_i \subseteq Z$ for $i = 1, \ldots, m$. The optimization problem (1.1) encompasses many finite-dimensional and infinite-dimensional nonlinear optimization problems. Our proposed algorithm employs m individual augmented

*Submitted to the editors DATE.

Funding: HA was partially supported by NSF grants DMS-1818772, DMS-1913004, and the Air Force Office of Scientific Research under Award NO: FA9550-19-1-0036. DPK and DR were supported by the U.S. Air Force Office of Scientific Research, Optimization Program under Award NO: F4FGA09135G001. DR was also supported by Sandia National Laboratories' LDRD program.

†Department of Mathematical Sciences and Center for Mathematics and Artificial Intelligence (CMAI), George Mason University, Fairfax, VA 22030, USA. (hantil@gmu.edu, http://math.gmu.edu/~hantil/).

‡Optimization and Uncertainty Quantification, Sandia National Laboratories++, Albuquerque, NM 87185, USA. (dpkouri@sandia.gov, https://www.cs.sandia.gov/cr-dpkouri).

§Optimization and Uncertainty Quantification, Sandia National Laboratories++, Albuquerque, NM 87185, USA. (dridzal@sandia.gov, https://www.sandia.gov/~dridzal/).

++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 grant DE-NA-0003525. 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.

Lagrangian penalties to handle the constraints $Tx \in C_i$ for $i = 1, \ldots, m$. The resulting subproblems are *equality constrained* and can be solved efficiently using modern sequential quadratic programming (SQP) methods. By separately penalizing the constraints $Tx \in C_i$, for $i = 1, \ldots, m$, our algorithm increases the augmented Lagrangian penalty parameters according to the associated infeasibility of the current iterate, adapting the i-th penalty parameter to the specific scaling of the constraint $Tx \in C_i$. We call the proposed algorithm the *Augmented Lagrangian Equality-constrained SQP (ALESQP)* method.

To motivate (1.1) and ALESQP, we note that optimization problems constrained by partial differential equations (PDEs) can be written in the form (1.1), where $x = (u, z)$ is split into state u and control z variables and the equality constraint $g(x) = 0$ represents the governing PDE. For such problems, it is often difficult to prove regularity of multipliers, especially when constraints of the type $Tx \in C_i$ are enforced on the PDE solution [10, 22, 27]. As a result, nonlinear programming methods developed in finite dimensions often exhibit *mesh dependence* when applied to discretizations of such problems. More specifically, their algorithmic performance degrades with refinement of the PDE discretization—in other words, with increasing problem size. We tackle this particular challenge by deriving and analyzing ALESQP in an *infinite-dimensional* setting. A common approach to solving PDE-constrained optimization problems is to reformulate (1.1), with the aforementioned splitting $x = (u, z)$, by eliminating the PDE solution variable u. When the PDE is nonlinear, this approach requires a nonlinear solver, e.g., a Newton-type iteration, to solve the PDE and evaluate the objective function at every optimization iteration. In contrast, optimization formulation (1.1) allows us to maintain the PDE as an explicit constraint. In doing so, ALESQP does not require an accurate solution of the equality constraint $g(x) = 0$ until convergence, and, in fact, it balances the PDE solution accuracy with other feasibility and optimality metrics as the algorithm iterates. This further allows us to approximately solve equality-constrained subproblems using inexact matrix-free SQP methods that take advantage of iterative linear system solves [19, 20] and mesh adaptivity [41].

The augmented Lagrangian method, or the method of multipliers, was originally introduced in [21, 30] for finite-dimensional, equality-constrained optimization and subsequently extended and analyzed by numerous authors, see [6, 32, 33, 34]. Augmented Lagrangian also serves as the backbone of numerous successful numerical optimization software packages. For example, the MINOS solver uses an augmented Lagrangian penalty for linearized equality constraints and solves linearly constrained subproblems; the LANCELOT solver employs an augmented Lagrangian penalty for equality constraints and solves bound-constrained subproblems [14]; and the ALGENCAN solver has the ability to use augmented Lagrangian penalties to handle both equality and inequality constraints [8]. The ALESQP method is closely related to two existing augmented Lagrangian approaches: LANCELOT and sequential equality-constrained optimization (SECO) [7]. Our approach generalizes the problem formulation of SECO, and solves a sequence of penalized equality-constrained subproblems. An important addition to the SECO algorithm is in the use of *multiple augmented Lagrangian penalties* to handle disparate inequality constraint scalings. The general mechanics of the ALESQP algorithm are borrowed from the LANCELOT solver described in [12], including the penalty parameter and multiplier update procedures, with extensions to support multiple penalties. A principal difference between ALESQP and both SECO and LANCELOT is that we prove convergence for infinite-dimensional problems. This advance enables discretization-independent performance of ALESQP on such problems, including mesh-based discretizations. Notably, we observe mesh-independent performance in all iterative components of ALESQP, including the augmented Lagrangian iteration, its SQP subproblem solver and SQP's quadratic optimization solver.

In contrast to the extensive body of work on augmented Lagrangian methods and software for the solution of finite-dimensional optimization problems, there has been little work on solving general infinite-dimensional optimization problems using the augmented Lagrangian. For instance, the references [4, 5, 23, 24, 25] are limited to specific convex optimization problems, treat only finite-dimensional constraints, or require strong assumptions, and therefore do not support the solution of the general problem (1.1). Only recently Börgens et al. [9] introduced and analyzed a generally applicable infinite-dimensional augmented Lagrangian method. There are four major differences between ALESQP and the method presented in [9]. First, we consider a different problem formulation, with an emphasis on maintaining the explicit constraint $g(x) = 0$. In the context of PDE-constrained optimization, where $g(x) = 0$ encompasses the PDE constraint, this choice crucially enables an *inexact and therefore efficient* solution of the governing PDE, through rigorous use of iterative linear solvers [19] and mesh adaptivity [41]. Second, we treat all constraints of the type $Tx \in C_i$ in a unified fashion, through multiple penalties and the corresponding multiplier and penalty updates, and we solve equality-constrained subproblems. In contrast, due to strong regularity assumptions on the constraint function in [9] (complete continuity of the mapping G, [9, Assumption 5.1]), certain inequality constraints must be treated implicitly, as part of the subproblem, while others are penalized using the augmented Lagrangian. Third, we provide a complete algorithmic framework, with a discussion of methods that are chosen specifically for their suitability as ALESQP subproblem solvers. We demonstrate excellent performance on a variety of infinite-dimensional problems, with nearly constant iteration counts *in all algorithmic components* of the ALESQP framework, independent of problem size. Fourth, we do not employ a multiplier safeguard (also used in, e.g., [2, 8]). Rather, we use the multiplier update from LANCELOT, see [12].

The remainder of the paper is organized as follows. In Sections 2 and 3 we introduce the notation and describe the assumptions on (1.1), recalling the associated optimality conditions. In Sections 4 and 5 we introduce the augmented Lagrangian algorithm and prove asymptotic stationarity and asymptotic feasibility of the generated sequence of iterates. We build on these results and show that, under additional assumptions, weak accumulation points of the sequence of iterates are stationary points for (1.1). In Section 6 we extend the augmented Lagrangian formulation to handle nonlinear constraint operators T. In Section 7 we briefly discuss the remaining components of the ALESQP framework, including the SQP algorithm and its subroutines. We conclude with a variety of numerical results including statistical estimation and PDE-constrained optimization in Section 8.

2. Notation.

Given a Banach space $(X, \|\cdot\|_X)$, we denote the topological dual space of X by X^* and the associated dual pairing by $\langle \cdot, \cdot \rangle_{X^*,X}$. If X is a Hilbert space, we denote by $(\cdot, \cdot)_X$ the inner product on X and we assume that $\|\cdot\|_X$ is the usual norm on X. We denote by B_ρ^X for $\rho > 0$ the closed norm ball on X with radius ρ. For two Banach spaces X and Y, we denote the space of bounded linear operators that map X into Y by $\mathcal{L}(X, Y)$. For a closed, convex subset C of the Banach space X, we denote the projection of a point $x \in X$ onto C by $\mathbf{P}_C(x)$ and the distance from

x to C by $d_C(x)$. That is, $\mathbf{P}_C(x)$ and $d_C(x)$ satisfy

$$d_C(x) := \min_{y\in C} \|x-y\|_X = \|x - \mathbf{P}_C(x)\|_X.$$

In addition, we denote the normal cone to C at the point $x \in C$ by

$$N_C(x) := \{\lambda \in X^* \mid \langle \lambda, y - x\rangle_{X^*,X} \le 0 \ \ \forall y \in C\},$$

with $N_C(x) = \emptyset$ if $x \notin C$. Finally, we denote convergence with respect to the weak topology by $\rightharpoonup$, convergence with respect to the weak* topology by $\rightharpoonup^*$, and convergence with respect to the norm topology by $\to$.

3. Problem Formulation and Assumptions.

Let X and Y be real Banach spaces and let Z be a real Hilbert space. To simplify the presentation, we will associate Z^* with Z. Given the problem data $f : X \to \mathbb{R}$, $g : X \to Y$, $T \in \mathcal{L}(X, Z)$ and a nonempty, closed and convex set $C \subseteq Z$, we consider the optimization problem

$$\min_{x\in X} f(x) \quad \text{subject to} \quad g(x) = 0, \quad Tx \in C. \tag{3.1}$$

When f and g are Fréchet differentiable, we say that $\bar{x} \in X$ is a first-order stationary point of (3.1) if there exists $\bar{\zeta} \in Y$ such that

$$-\,(f'(\bar{x}) + g'(\bar{x})^*\bar{\zeta}) \in T^* N_C(T\bar{x}) \quad \text{and} \quad g(\bar{x}) = 0. \tag{3.2}$$

Note that this presumes $T\bar{x} \in C$ since the normal cone is empty otherwise.

Remark 3.1 (Banach Space Valued Constraints). As in [9], we could consider the case where Z is a real Banach space that is densely embedded in a real Hilbert space. However, this would complicate the presentation with little added benefit.

To prove convergence of our algorithm, we will require the following assumptions on the objective function f, the constraint operators g and T, and the constraint set C. In our subsequent analysis, we will explicitly state when each assumption is required. Assumptions (A0) (feasibility) and (A1) (differentiability) will be required throughout, whereas (A2), (A3) and (A4) will only be required to prove convergence.

Assumption 3.2 (Regularity of Problem Data).
- (A0) There exists $\bar{x} \in X$ such that $g(\bar{x}) = 0$ and $T\bar{x} \in C$.
- (A1) The functions f and g are continuously Fréchet differentiable.
- (A2) The adjoint operator T^* is injective.
- (A3) The functions f and $\|g(\cdot)\|_Y$ are weakly lower semicontinuous.
- (A4) There exist $C_i \subset Z$ for $i = 1, \ldots, m$ that are nonempty, closed and convex for which $C = C_1 \cap \cdots \cap C_m \neq \emptyset$ and $\{C_1, \ldots, C_m\}$ is boundedly regular in the sense that

$$\max_{i=1,\ldots,m} d_{C_i}(Tx_k) \to 0 \qquad \Longrightarrow \qquad d_C(Tx_k) \to 0$$

as $k \to \infty$ for every bounded sequence $\{x_k\} \subset X$.

Remark 3.3 (Assumption (A2)). Recall that the operator T is surjective if and only if T^* is injective and the range of T^* is norm-closed [35, Th. 4.15]. As a consequence, assumption (A2) is satisfied if T is surjective. In addition, recall that T^* is injective if and only if the kernel of T^* is trivial, i.e., $\ker T^* := \{z \in Z \mid T^*z = 0\} = \{0\}$.

4. Augmented Lagrangian with Explicit Equality Constraints. To develop the augmented Lagrangian portion of our algorithm, we recall that (3.1) is equivalent to the equality constrained problem

$$\min_{x\in X} \{f(x) + I_C(Tx)\} \quad \text{subject to} \quad g(x) = 0, \tag{4.1}$$

where $I_C(y) = 0$ if $y \in C$ and $I_C(y) = \infty$ if $y \notin C$. Here, $I_C(Tx)$ enforces the constraint $Tx \in C$. Given the constraint decomposition in assumption (A4), we can rewrite I_C as

$$I_C = \sum_{i=1}^{m} I_{C_i}.$$

As is typically done in augmented Lagrangian methods [34], we replace the indicator functions $I_{C_i}(Tx)$ with the relaxations $\Psi_i(x, \lambda, r)$, where $\Psi_i : X \times Z \times (0, \infty) \to \mathbb{R}$ is defined as

$$\Psi_i(x, \lambda, r) := \sup_{\mu\in Z}\{(\mu, Tx)_Z - I^*_{C_i}(\mu) - \tfrac{1}{2r}\|\mu - \lambda\|^2_Z\} \tag{4.2}$$

and $I^*_{C_i}$ is the Fenchel conjugate of I_{C_i}, i.e.,

$$I^*_{C_i}(\mu) := \sup_{z\in C_i} (\mu, z)_Z.$$

The augmented Lagrangian functional is given by

$$L(x, \lambda_1, \ldots, \lambda_m, r_1, \ldots, r_m) := f(x) + \sum_{i=1}^{m} \Psi_i(x, \lambda_i, r_i), \tag{4.3}$$

where $\lambda_i \in Z$ and $r_i > 0$ for $i = 1, \ldots, m$. The function being maximized in (4.2) is strongly concave and has the unique maximizer

$$\Lambda_i(x, \lambda, r) := r((r^{-1}\lambda + Tx) - \mathbf{P}_{C_i}(r^{-1}\lambda + Tx)). \tag{4.4}$$

Substituting $\Lambda_i(x, \lambda, r)$ into (4.2) and rearranging terms yields the usual augmented Lagrangian penalty function

$$\Psi_i(x, \lambda, r) = \tfrac{1}{2r}\|\Lambda_i(x, \lambda, r)\|^2_Z - \tfrac{1}{2r}\|\lambda\|^2_Z.$$

Motivated by [7, 12], we formulate Algorithm 4.1 using the augmented Lagrangian L.

Remark 4.1 (Penalty Parameter Update). The penalty parameter update in Algorithm 4.1 is completely decoupled for the first K_0 iterations. Here, K_0 can be taken arbitrarily large, but finite, e.g., $K_0 = 1000$. This allows each $r_i^{(k)}$ to be calibrated to the scaling associated with the i^{th} constraint. After K_0 iterations, Algorithm 4.1 switches schemes and updates all penalty parameters in unison. This penalty update scheme is a safeguard for the case in which the algorithm produces an infinite sequence of iterations, forcing the sequence to accumulate at a feasible point (under certain assumptions), and is typically never active in practice.

Algorithm 4.1 Multi-Penalty Equality-Constrained Augmented Lagrangian

Input: Initial multiplier estimates $\{\lambda_1^{(1)},\ldots,\lambda_m^{(1)}\} \in Z$, positive penalty parameters $\{r_1^{(1)},\ldots,r_m^{(1)}\}$, nonnegative null sequences $\{\delta^{(k)}\}$ and $\{\varepsilon^{(k)}\}$, $K_0 \in \mathbb{N}$, and positive constants $\{\nu_1,\ldots,\nu_m\}$, $\{\gamma_1,\ldots,\gamma_m\}$ with $\gamma_i < 1/2$, $\{\tau_1^{(0)},\ldots,\tau_m^{(0)}\}$, $\{\theta_1,\ldots,\theta_m\}$ with $\theta_i < 1$, τ_*, δ_*, ε_*, $\{\eta_1,\ldots,\eta_m\}$ with $\eta_i > 1$, $\overline{\eta} > 1$, $\{\alpha_1,\ldots,\alpha_m\}$ and $\{\beta_1,\ldots,\beta_m\}$. Set $\theta_i^{(1)} = \min\{1/r_i^{(1)},\theta_i\}$ and $\tau_i^{(1)} = \tau_i^{(0)}(\theta_i^{(1)})^{\alpha_i}$.

for $k = 1,2,3,\ldots$ **do**
  Compute $(x^{(k)},\zeta^{(k)}) \in X \times Y^*$ that satisfies
  $$\|g(x^{(k)})\|_Y \le \delta^{(k)} \quad \text{and} \quad \|f'(x^{(k)})+\textstyle\sum_i T^*\Lambda_i(x^{(k)},\lambda_i^{(k)},r_i^{(k)})+g'(x^{(k)})^*\zeta^{(k)}\|_{X^*} \le \varepsilon^{(k)}$$
  if $\|g(x^{(k)})\|_Y \le \delta_*,\quad \|f'(x^{(k)}) + \sum_i T^*\Lambda_i(x^{(k)},\lambda_i^{(k)},r_i^{(k)}) + g'(x^{(k)})^*\zeta^{(k)}\|_{X^*} \le \varepsilon_*$
    and $\max_i d_{C_i}(Tx^{(k)}) \le \tau_*$ **then**
    return $x^{(k)}$ as the approximate solution
  end if
  if $k = K_0 + 1$ **then**
    $\eta_i = \overline{\eta}$ for $i = 1,\ldots,m$
  end if
  `update` = `false`
  if $k > K_0$ **and** $\exists\, i$ such that $\|\Lambda_i(x^{(k)},\lambda_i^{(k)},r_i^{(k)}) - \lambda_i^{(k)}\|_Z > r_i^{(k)}\tau_i^{(k)}$ **then**
    `update` = `true`
  end if
  for $i = 1,\ldots,m$ **do**
    if $\|\Lambda_i(x^{(k)},\lambda_i^{(k)},r_i^{(k)}) - \lambda_i^{(k)}\|_Z > r_i^{(k)}\tau_i^{(k)}$ **or** `update` = `true` **then**
      $r_i^{(k+1)} = \eta_i r_i^{(k)}$
      $\theta_i^{(k+1)} = \min\{1/r_i^{(k+1)},\theta_i\}$
      $\tau_i^{(k+1)} = \tau_i^{(0)}(\theta_i^{(k+1)})^{\alpha_i}$
    else
      $r_i^{(k+1)} = r_i^{(k)}$
      $\theta_i^{(k+1)} = \min\{1/r_i^{(k+1)},\theta_i\}$
      $\tau_i^{(k+1)} = \tau_i^{(k)}(\theta_i^{(k+1)})^{\beta_i}$
    end if
    if $\|\Lambda_i(x^{(k)},\lambda_i^{(k)},r_i^{(k)})\|_Z \le \nu_i(r_i^{(k+1)})^{\gamma_i}$ **then**
      $\lambda_i^{(k+1)} = \Lambda_i(x^{(k)},\lambda_i^{(k)},r_i^{(k)})$
    else
      $\lambda_i^{(k+1)} = \lambda_i^{(k)}$
    end if
  end for
end for

Remark 4.2 (Subproblem Tolerance Sequences). The sequences $\{\varepsilon^{(k)}\}$ and $\{\delta^{(k)}\}$, are only required to be nonnegative and converge to zero. A basic choice is

$$\delta^{(k+1)} = \begin{cases} \delta_0\delta^{(k)} & \text{if } \exists\, i \text{ such that } r_i^{(k+1)} \ne r_i^{(k)} \\ \delta_1\delta^{(k)} & \text{otherwise} \end{cases} \qquad \text{and}$$
$$\varepsilon^{(k+1)} = \begin{cases} \varepsilon_0\varepsilon^{(k)} & \text{if } \exists\, i \text{ such that } r_i^{(k+1)} \ne r_i^{(k)} \\ \varepsilon_1\varepsilon^{(k)} & \text{otherwise} \end{cases}$$

for $0 < \delta_1 < \delta_0 < 1$ and $0 < \varepsilon_1 < \varepsilon_0 < 1$. More complicated updating strategies are possible. For example, adapting the tolerance sequence from [12] yields

$$\delta^{(k+1)} = \begin{cases} \delta^{(0)}(\theta^{(k+1)})^{\alpha_\delta} & \text{if } \exists\, i \text{ such that } r_i^{(k+1)} \neq r_i^{(k)} \\ \delta^{(k)}(\theta^{(k+1)})^{\beta_\delta} & \text{otherwise} \end{cases} \quad \text{and}$$

$$\varepsilon^{(k+1)} = \begin{cases} \varepsilon^{(0)}(\theta^{(k+1)})^{\alpha_\varepsilon} & \text{if } \exists\, i \text{ such that } r_i^{(k+1)} \neq r_i^{(k)} \\ \varepsilon^{(k)}(\theta^{(k+1)})^{\beta_\varepsilon} & \text{otherwise} \end{cases}$$

for positive constants $\delta^{(0)}$, $\varepsilon^{(0)}$, α_δ, α_ε, β_δ and β_ε, where, e.g., $\theta^{(k+1)} = (\max_i r_i^{(k+1)})^{-1}$, whereas adapting the sequences from [7] yields

$$\delta^{(k+1)} = \min\{\delta_0\delta^{(k)}, \delta_1\|g(x^{(k)}\|_Y\} \quad \text{and}$$
$$\varepsilon^{(k+1)} = \min\{\varepsilon_0\varepsilon^{(k)}, \varepsilon_1\|f'(x^{(k)}) + c'(x^{(k)})^*\zeta^{(k)}\|_{X^*}\}$$

for constants δ_0, δ_1, ε_0, $\varepsilon_1 \in (0,1)$.

4.1. Properties of the Augmented Lagrangian. We note that the first step in Algorithm 4.1 seeks an approximate stationary point of the equality constrained subproblem

$$\min_{x\in X}\ L_k(x) \qquad \text{subject to} \qquad g(x) = 0, \tag{4.5}$$

where $L_k(x) := L(x, \lambda_1^{(k)}, \dots, \lambda_m^{(k)}, r_1^{(k)}, \dots, r_m^{(k)})$. To facilitate the solution of (4.5), we first show that the penalty function $\Psi_i(\cdot,\lambda,r)$ is convex and continuously Fréchet differentiable with Lipschitz continuous gradient.

PROPOSITION 4.3. *For all $\lambda \in Z$ and $r > 0$, the penalty function $\Psi_i(\cdot,\lambda,r)$ is convex and continuous for $i = 1,\dots,m$. Additionally, $\Psi_i(\cdot,\lambda,r)$ is continuously Fréchet differentiable with gradient*

$$\nabla_x \Psi_i(x,\lambda,r) = T^*\Lambda_i(x,\lambda,r), \tag{4.6}$$

which is Lipschitz continuous with modulus $r\|T\|^2_{\mathcal{L}(X,Z)}$.

Proof. Notice that $\Psi_i(\cdot,\lambda,r)$ the Fenchel conjugate of $I_C^*(\cdot)+\frac{1}{2r}\|\cdot-\lambda\|_X^2$ composed with T. Therefore, $\Psi_i(\cdot,\lambda,r)$ is equal to the infimal convolution of the conjugates of $I_{C_i}^*$ and $\frac{1}{2r}\|\cdot-\lambda\|_X^2$ composed with T [3, Prop. 13.21(i)], i.e.,

$$\begin{aligned}\Psi_i(x,\lambda,r) &= \inf_{y\in Z}\{I_{C_i}(y) + (\lambda, Tx-y)_Z + \tfrac{r}{2}\|Tx-y\|_Z^2\} \\ &= \inf_{y\in C_i}\{(\lambda, Tx-y)_Z + \tfrac{r}{2}\|Tx-y\|_Z^2\}.\end{aligned}$$

Consequently, $\Psi_i(\cdot,\lambda,r)$ is continuous convex and Fréchet differentiable with Lipschitz continuous gradient (cf. Propositions 9.5 and 12.29 and Corollary 9.15 in [3]). ∎

COROLLARY 4.4. *For any fixed $\lambda_i \in Z$ and $r_i > 0$ for $i = 1,\dots,m$, the augmented Lagrangian $L(\cdot,\lambda_1,\dots,\lambda_m,r_1,\dots,r_m)$ is:* (i) *weakly lower semicontinuous if f is;* (ii) *convex if f is; and* (iii) *continuously Fréchet differentiable if f is. Moreover, in case* (iii)*, the derivative of $L(\cdot,\lambda_1,\dots,\lambda_m,r_1,\dots,r_m)$ is given by*

$$L_x(x,\lambda_1,\dots,\lambda_m,r_1,\dots,r_m) = f'(x) + \sum_{i=1}^m T^*\Lambda_i(x,\lambda_i,r_i) \tag{4.7}$$

and if f' is Lipschitz continuous, then so is $L_x(\cdot,\lambda_1,\dots,\lambda_m,r_1,\dots,r_m)$.

Proof. By Proposition 4.3, $\Psi_i(\cdot,\lambda_i,r_i)$ is convex and continuous, and therefore weakly lower semicontinuous. In addition, $\Psi_i(\cdot,\lambda_i,r_i)$ is continuously Fréchet differentiable with Lipschitz gradients. The desired results then follow from the assumed properties of f. ∎

5. Convergence Theory. In the subsequent results, we consider infinite sequences of iterates generated by Algorithm 4.1 ignoring the stopping conditions, i.e.,

$$\|g(x^{(k)})\|_Y \le \delta_* \tag{5.1a}$$

$$\|L_k'(x^{(k)}) + g'(x^{(k)})^*\zeta^{(k)}\|_{X^*} \le \varepsilon_* \tag{5.1b}$$

$$\max_{i=1,\ldots,m} d_{C_i}(Tx^{(k)}) \le \tau_* \tag{5.1c}$$

with $\delta_* = \varepsilon_* = \tau_* = 0$. We denote by $\mathbb{P}_i \subseteq \mathbb{N}$ the set of indices k that satisfy

$$\|\Lambda_i(x^{(k)},\lambda_i^{(k)},r_i^{(k)}) - \lambda_i^{(k)}\|_Z > r_i^{(k)}\tau_i^{(k)}. \tag{5.2}$$

We further denote by $\mathbb{M}_i \subseteq \mathbb{N}$ the subsets of indices for which

$$\lambda_i^{(k+1)} = \Lambda_i(x^{(k)},\lambda_i^{(k)},r_i^{(k)}) \tag{5.3}$$

holds. For any set $\mathbb{S} \subset \mathbb{N}$, we denote the complement of $\mathbb{S}$ by $\mathbb{S}^c := \mathbb{N}\setminus\mathbb{S}$ and the cardinality of $\mathbb{S}$ by $|\mathbb{S}|$. The set $\mathbb{P}_i$ encapsulates the iterations for which the penalty parameter for the i^{th} constraint is increased, while the set $\mathbb{M}_i$ encapsulates the iterations for which the multipliers for the i^{th} constraint are changed. We note that $r_i^{(k)} \to \infty$ if and only if there exists at least one $j = 1,\ldots,m$ such that $|\mathbb{P}_j| = \infty$; in this case, all penalty parameters are increased for any iteration $k \in \mathbb{P}_j$ with $k \ge K_0$ (see lines 9 through 22 in Algorithm 4.1). Our first result is technical and relates the satisfaction of the constraints in C_i to the multiplier updates in Algorithm 4.1.

LEMMA 5.1. *Let $x \in X$, $\lambda \in Z$ and $r > 0$ be arbitrary. Then,*

$$d_{C_i}(Tx) \le \frac{1}{r}\|\Lambda_i(x,\lambda,r) - \lambda\|_Z \le d_{C_i}(Tx) + \frac{1}{r}\|\lambda\|_Z \tag{5.4}$$

Proof. We first prove the lower bound. By definition of Λ_i, we have that

$$d_{C_i}(Tx) = \min_{y\in C_i}\|y - Tx\|_Z \le \|Tx - \mathbf{P}_{C_i}(r^{-1}\lambda + Tx)\|_Z = \frac{1}{r}\|\Lambda_i(x,\lambda,r) - \lambda\|_Z.$$

Similarly, the Lipschitz continuity (with unit modulus) of the projection [3, Prop. 4.8] ensures that

$$\begin{aligned}\frac{1}{r}\|\Lambda_i(x,\lambda,r) - \lambda\|_Z &\le \|Tx - \mathbf{P}_{C_i}(Tx)\|_Z + \|\mathbf{P}_{C_i}(Tx) - \mathbf{P}_{C_i}(r^{-1}\lambda + Tx)\|_Z \\ &\le d_{C_i}(Tx) + \frac{1}{r}\|\lambda\|_Z\end{aligned}$$

as desired. ∎

5.1. Dual Convergence. In this subsection, we analyze the sequence of dual variables $\{\lambda_i^{(k)}\}$ generated by Algorithm 4.1. Our first result is motivated by Lemma 4.2 in [12] and shows that the sequence $\{\|\lambda_i^{(k)}\|_Z\}$ does not grow too fast if $r_i^{(k)} \to \infty$. The second result relates the first to the asymptotic feasibility of the iterates $\{x^{(k)}\}$. Finally, we show that $\{\lambda_i^{(k)}\}$ converges strongly if $|\mathbb{P}_i| < \infty$.

LEMMA 5.2. *Let $\{\lambda_i^{(k)}\}$ be an infinite sequence of multipliers for the i^{th} constraint generated by Algorithm 4.1, ignoring the stopping conditions (5.1). If $r_i^{(k)} \to \infty$, then*

$$\lim_{k\to\infty} \frac{1}{(r_i^{(k)})^\alpha}\|\lambda_i^{(k)}\|_Z = 0 \quad \forall\, \alpha > \gamma_i,$$

where $\gamma_i < 1/2$ is defined in Algorithm 4.1.

Proof. We note that the proof of this fact is similar to the proof of [12, L. 4.2] for equality constrained augmented Lagrangian methods. If $\mathbb{M}_i = \{k_1, k_2, \ldots\}$ is finite, then the result clearly holds since $\lambda_i^{(k)}$ is fixed after finitely many iterations. Now suppose that $\mathbb{M}_i$ is infinite. For any $k_j < k \le k_{j+1}$, we have that $r_i^{(k)} \ge r_i^{(k_j+1)}$ and therefore,

$$\tfrac{1}{(r_i^{(k)})^\alpha}\|\lambda_i^{(k)}\|_Z \le \tfrac{1}{(r_i^{(k_j+1)})^\alpha}\|\lambda_i^{(k_j+1)}\|_Z \le \nu_i (r_i^{(k_j+1)})^{\gamma_i-\alpha}.$$

The upper bound follows from line 23 in Algorithm 4.1. Since $\alpha > \gamma_i$, the right hand side converges to zero and the desired result follows. ∎

Our next lemma builds on Lemma 5.2 and provides equivalent conditions that imply that the sequence $\{x^{(k)}\}$ is asymptotically feasible for the i^{th} constraint.

LEMMA 5.3. *Let $\{x^{(k)}\}$ be an infinite sequence of iterates generated by Algorithm 4.1, ignoring the stopping conditions (5.1), with the associated sequence of multipliers $\{\lambda_i^{(k)}\}$ for the i^{th} constraint. If $r_i^{(k)} \to \infty$, then the following three conditions are equivalent*

(a) $\liminf_{k\to\infty} d_{C_i}(Tx^{(k)}) = 0$

(b) $\liminf_{k\to\infty} \dfrac{1}{r_i^{(k)}}\|\Lambda_i(x^{(k)}, \lambda_i^{(k)}, r_i^{(k)}) - \lambda_i^{(k)}\|_Z = 0$

(c) $\liminf_{k\to\infty} \dfrac{1}{r_i^{(k)}}\|\Lambda_i(x^{(k)}, \lambda_i^{(k)}, r_i^{(k)})\|_Z = 0.$

These equivalences remain true if the limit inferiors are replaced by limits (or equivalently limit superiors).

Proof. By Lemma 5.1 with $y = x^{(k)}$, $\lambda = \lambda_i^{(k)}$ and $r = r_i^{(k)}$, we have that

$$d_{C_i}(Tx^{(k)}) \le \tfrac{1}{r_i^{(k)}}\|\Lambda_i(x^{(k)}, \lambda_i^{(k)}, r_i^{(k)}) - \lambda_i^{(k)}\|_Z \le d_{C_i}(Tx^{(k)}) + \tfrac{1}{r_i^{(k)}}\|\lambda_i^{(k)}\|_Z. \tag{5.5}$$

This and Lemma 5.2 yield (a) $\iff$ (b). The implication (b) $\iff$ (c) follows from Lemma 5.2 and the triangle and reverse triangle inequalities. ∎

Before proving our main finite-termination result, we provide situations for which the sequence of multiplier estimates converges strongly. Strong convergence will be useful for later results as it will allow us to demonstrate that weak accumulation points are stationary under certain assumptions.

THEOREM 5.4. *Let $\{x^{(k)}\}$ be an infinite sequence of iterates generated by Algorithm 4.1, ignoring the stopping conditions (5.1), with the associated sequence of multipliers $\{\lambda_i^{(k)}\}$ for the i^{th} constraint. If $|\mathbb{M}_i| < \infty$ or $|\mathbb{P}_j| < \infty$ for all $j = 1, \ldots, m$, then the sequence of multipliers $\{\lambda_i^{(k)}\}$ converges strongly to some $\bar{\lambda}_i \in Z$.*

Proof. First note that if $|\mathbb{M}_i| < \infty$, then $\lambda_i^{(k)} = \bar{\lambda}_i$ is constant for all $k > \max \mathbb{M}_i$ and the result follows. Now, consider the case when $|\mathbb{P}_j| < \infty$ for $j = 1, \ldots, m$. Let

$k' = \max_j \max \mathbb{P}_j + 1$. We will first show that $\{\lambda_i^{(k)}\}$ is Cauchy (and hence converges). Let $\epsilon > 0$ be arbitrary and choose $k_\epsilon \geq k'$ such that

$$\tau_i^{(k_\epsilon)} < \frac{1 - (\theta_i^{(k')})^{\beta_i}}{r_i^{(k')}} \epsilon.$$

Such a k_ϵ exists since $r_i^{(k)} = r_i^{(k')}$ and $\theta_i^{(k)} = \theta_i^{(k')}$ for all $k \geq k'$, and $\{\tau_i^{(k)}\}_{k \geq k'}$ is strictly decreasing to zero. For any $k \geq k_\epsilon$ and any $h \in \mathbb{N}$, we have that

$$\lambda_i^{(k+h)} - \lambda_i^{(k)} = \sum_{j=k}^{k+h-1} \lambda_i^{(j+1)} - \lambda_i^{(j)}.$$

Since $k \geq k_\epsilon$, we have that $\|\lambda_i^{(j+1)} - \lambda_i^{(j)}\|_Z$ is either equal to zero or is bounded above by $r_i^{(j)} \tau_i^{(j)} > 0$. Therefore, the triangle inequality ensures that

$$\|\lambda_i^{(k+h)} - \lambda_i^{(k)}\|_Z \leq \sum_{j=k}^{k+h-1} \|\lambda_i^{(j+1)} - \lambda_i^{(j)}\|_Z \leq \sum_{j=k}^{k+h-1} r_i^{(j)} \tau_i^{(j)} = r_i^{(k')} \tau_i^{(k)} \sum_{j=0}^{h} (\theta_i^{(k')})^{j\beta_i}.$$

Since $\theta_i^{(k')} < 1$, we have that the sum on the right hand side of the above inequality converges geometrically and thus

$$\|\lambda_i^{(k+h)} - \lambda_i^{(k)}\|_Z \leq \frac{\tau_i^{(k)} r_i^{(k')}}{1 - (\theta_i^{(k')})^{\beta_i}} < \frac{\tau_i^{(k_\epsilon)} r_i^{(k')}}{1 - (\theta_i^{(k')})^{\beta_i}} < \epsilon.$$

Consequently, $\{\lambda_i^{(k)}\}$ is Cauchy and hence converges strongly to some $\bar{\lambda}_i \in Z$. □

COROLLARY 5.5. *Consider the setting of Theorem 5.4. If* $|\mathbb{P}_j| < \infty$ *for* $j = 1, \ldots, m$, *then*

$$\Lambda_i(x^{(k)}, \lambda_i^{(k)}, r_i^{(k)}) \to \bar{\lambda}_i \qquad i = 1, \ldots, m.$$

Proof. The triangle inequality ensures that

$$\begin{aligned}\|\Lambda_i(x^{(k)}, \lambda_i^{(k)}, r_i^{(k)}) - \bar{\lambda}_i\|_Z &\leq \|\Lambda_i(x^{(k)}, \lambda_i^{(k)}, r_i^{(k)}) - \lambda_i^{(k)}\|_Z + \|\bar{\lambda}_i - \lambda_i^{(k)}\|_Z \\ &\leq r_i^{(k)} \tau_i^{(k)} + \|\bar{\lambda}_i - \lambda_i^{(k)}\|_Z \quad \forall\, k \geq \max_j \max \mathbb{P}_j + 1.\end{aligned}$$

Since $r_i^{(k)}$ is constant for all k sufficiently large and $\tau_i^{(k)} \to 0$, the result then follows from Theorem 5.4. □

5.2. Finite Termination. In this subsection, we investigate the scenarios for which Algorithm 4.1 terminates in a finite number of iterations. As the following result suggests, it is uncommon for the algorithm to produce infinitely many iterations without satisfying the stopping conditions (5.1). In particular, this result states that Algorithm 4.1 will satisfy the stopping conditions (5.1) after a finite number of iterations if there are infinitely many successful iterations (i.e., $|\mathbb{P}_i^c| = \infty$) or if the multiplier are updated infinitely often (i.e., $|\mathbb{M}_i| = \infty$). In fact, this result shows that the only case for which Algorithm 4.1 may not terminate in finitely many iterations is if $|\mathbb{P}_i^c| < \infty$ and the associated multiplier estimates are only updated a finite number of times, i.e., $|\mathbb{M}_i| < \infty$.

THEOREM 5.6. *Let $\{(x^{(k)}, \zeta^{(k)}, \lambda_i^{(k)}, r_i^{(k)})\}$ be an infinite sequence of iterates generated by Algorithm 4.1, ignoring the stopping conditions (5.1). Then, the sequence satisfies*

$$\lim_{k\to\infty} \|g(x^{(k)})\|_Y = 0 \quad \text{and} \quad \lim_{k\to\infty} \|L_k'(x^{(k)}) + g'(x^{(k)})^*\zeta^{(k)}\|_{X^*} = 0. \tag{5.6}$$

Consider arbitrary $i \in \{1,\ldots,m\}$. If $|\mathbb{P}_i^c \cup \mathbb{M}_i| = \infty$, then

$$\lim_{j\to\infty} d_{C_i}(Tx^{(k_j)}) = 0 \quad \text{where} \quad \mathbb{P}_i^c \cup \mathbb{M}_i = \{k_j\}_{j=1}^\infty. \tag{5.7}$$

In particular, if $|\mathbb{P}_i^c \cup \mathbb{M}_i| = \infty$, then

$$\liminf_{k\to\infty} d_{C_i}(Tx^{(k)}) = 0 \tag{5.8}$$

and if either $|\mathbb{P}_i| < \infty$ or $|\mathbb{M}_i^c| < \infty$, then $d_{C_i}(Tx^{(k)}) \to 0$ as $k \to \infty$.

Proof. We first note that the tolerance update rules in Algorithm 4.1 ensure that

$$\lim_{k\to\infty} \tau_i^{(k)} = 0, \quad \lim_{k\to\infty} \delta^{(k)} = 0, \quad \text{and} \quad \lim_{k\to\infty} \varepsilon^{(k)} = 0. \tag{5.9}$$

As a result (5.6) holds. Now, let $i \in \{1,\ldots,m\}$ be arbitrary. By Lemma 5.1, we have that $d_{C_i}(Tx^{(k)}) \le \tau_i^{(k)}$ for all $k \in \mathbb{P}_i^c$ and therefore $\{d_{C_i}(Tx^{(k)})\}_{k\in\mathbb{P}_i^c}$ converges to zero if $|\mathbb{P}_i^c| = \infty$. In particular, if $|\mathbb{P}_i| < \infty$, then we have that $d_{C_i}(Tx^{(k)}) \to 0$ as $k \to \infty$ since $d_{C_i}(Tx^{(k)}) \le \tau_i^{(k)}$ for all k sufficiently large. Now, suppose that $|\mathbb{P}_i| = \infty$. The multiplier update rule in Algorithm 4.1 ensures that

$$\frac{1}{r_i^{(k)}}\|\Lambda_i(x^{(k)}, \lambda_i^{(k)}, r_i^{(k)})\|_Z \le \frac{\nu_i}{r_i^{(k)}}(r_i^{(k+1)})^{\gamma_i} \le \nu_i\eta_i^{\gamma_i}(r_i^{(k)})^{\gamma_i-1} \quad \forall\, k \in \mathbb{M}_i. \tag{5.10}$$

Note that if $k > K_0$ then η_i is replaced by $\overline{\eta}$ in (5.10). Therefore, $\{d_{C_i}(Tx^{(k)})\}_{k\in\mathbb{M}_i}$ converges to zero by Lemma 5.3 if $|\mathbb{M}_i| = \infty$ since $\gamma_i < 1/2$. Consequently, if $|\mathbb{M}_i^c| < \infty$, then $d_{C_i}(Tx^{(k)}) \to 0$ as $k \to \infty$ since (5.10) holds for all k sufficiently large. Combining these results, we see that (5.7) holds if $|\mathbb{P}_i^c \cup \mathbb{M}_i| = \infty$. Finally, (5.8) is a direct consequence of (5.7). ∎

Theorem 5.6 provides conditions under which Algorithm 4.1 terminates in a finite number of iterations. However, it does not address the asymptotic satisfaction of the first order optimality conditions (3.2). Our next result demonstrates that the sequence of iterates generated by Algorithm 4.1 asymptotically satisfies (3.2) as long as (5.8) holds.

PROPOSITION 5.7. *Consider the setting of Theorem 5.6. Then the iterates satisfy*

$$\limsup_{k\to\infty} \langle T^*\Lambda_i(x^{(k)}, \lambda_i^{(k)}, r_i^{(k)}), y - x^{(k)}\rangle_{X^*,X} \le 0 \quad \forall\, y \in T^{-1}(C_i) \tag{5.11}$$

for $i = 1,\ldots,m$.

Proof. For the subsequent arguments, it will be convenient to define

$$s^{(k)} = \mathbf{P}_{C_i}((r^{(k)})^{-1}\lambda_i^{(k)} + Tx^{(k)}) \quad \text{and} \quad t^{(k)} = \Lambda_i(x^{(k)}, \lambda_i^{(k)}, r_i^{(k)}).$$

We have that $s^{(k)} \in C_i$ and $t^{(k)} \in N_{C_i}(s^{(k)})$ by [3, Prop. 6.46]. In addition, by [3, Th. 6.29], we can write

$$Tx^{(k)} = \frac{t^{(k)} - \lambda_i^{(k)}}{r_i^{(k)}} + s^{(k)}.$$

Consequently, for any $y \in T^{-1}(C_i)$, we have that

$$\langle T^* t^{(k)}, y - x^{(k)}\rangle_{X^*,X} = (t^{(k)}, Ty - (r_i^{(k)})^{-1}(t^{(k)} - \lambda_i^{(k)}) - s^{(k)})_Z$$

$$\leq \frac{1}{r_i^{(k)}}((t^{(k)}, \lambda_i^{(k)})_Z - \|t^{(k)}\|_Z^2), \tag{5.12}$$

where the upper bound follows from the fact that $t^{(k)} \in N_{C_i}(s^{(k)})$. If $|\mathbb{P}_j| < \infty$ for $j = 1, \ldots, m$, then Theorem 5.4 and Corollary 5.5 ensure that the upper bound in (5.12) converges to zero since $\lambda_i^{(k)} \to \bar{\lambda}_i$ and $t^{(k)} \to \bar{\lambda}_i$. Now, consider the case when $|\mathbb{P}_i| = \infty$. By maximizing the quadratic form on the right hand side of the above equation with respect to $t^{(k)}$, we see that

$$\langle T^* t^{(k)}, y - x^{(k)}\rangle_{X^*,X} \leq \frac{1}{4r_i^{(k)}}\|\lambda_i^{(k)}\|_Z^2.$$

After passing to the limit superior, the desired result follows as a consequence of Lemma 5.2 with $\alpha = 1/2 > \gamma_i$. ◻

Remark 5.8 (Relation to First-Order Optimality Conditions). By Theorem 5.6, we have that $g(x^{(k)}) \to 0$ and

$$-(f'(x^{(k)}) + g'(x^{(k)})^*\zeta^{(k)}) + e^{(k)} = \sum_{i=1}^m T^*\Lambda_i(x^{(k)}, \lambda_i^{(k)}, r_i^{(k)}), \tag{5.13}$$

where $e^{(k)} \to 0$. If $\langle e^{(k)}, x^{(k)}\rangle_{X^*,X} \to 0$, then Proposition 5.7 ensures that

$$\limsup_{k\to\infty} \langle -(f'(x^{(k)}) + g'(x^{(k)})^*\zeta^{(k)}), y - x^{(k)}\rangle_{X^*,X} \leq 0 \quad \forall\, y \in T^{-1}(C) \tag{5.14}$$

and therefore the sequence of iterates $\{(x^{(k)}, \zeta^{(k)})\}$ asymptotically satisfies the first-order optimality conditions (3.2), as long as (5.8) holds.

5.3. Convergence to Feasible Points and Asymptotic Stationarity. In this subsection, we show that weak accumulation points of the iterates $\{x^{(k)}\}$ generated by Algorithm 4.1 are nearly feasible. The assumptions required for this result follow from standard assumptions in the convergence theory for SQP and augmented Lagrangian methods [12, 15, 19, 20].

THEOREM 5.9. *Consider the setting of Theorem 5.6 and let assumption (A1)–(A4) hold. Suppose there exists a weakly converging subsequence $\{x^{(k_j)}\}$ with limit $\bar{x} \in X$ such that*

$$\theta^{(k_j)} := \Big(\sum_{i=1}^m r_i^{(k_j)}\Big)^{-1} \to 0 \quad \Longrightarrow \quad \theta^{(k_j)}(f'(x^{(k_j)}) + g'(x^{(k_j)})^*\zeta^{(k_j)}) \rightharpoonup^* 0. \tag{5.15}$$

If $|\mathbb{P}_i^c \cup \mathbb{M}_i| = \infty$ for all $i = 1, \ldots, m$, then $T\bar{x} \in C$. On the other hand, if there is at least one $i \in \{1, \ldots, m\}$ for which $|\mathbb{P}_i^c \cup \mathbb{M}_i| < \infty$, then there exists $\bar{t}_i \in (0,1)$ with $\bar{t}_1 + \ldots + \bar{t}_m = 1$ and $\bar{y}_i \in C_i$ such that

$$T\bar{x} = \sum_{i=1}^m \bar{t}_i \bar{y}_i. \tag{5.16}$$

In addition, the subsequence $\{(x^{(k_j)}, \zeta^{(k_j)})\}$ satisfies

$$\limsup_{k\to\infty} \langle -(f'(x^{(k_j)}) + g'(x^{(k_j)})^*\zeta^{(k_j)}), y - x^{(k_j)}\rangle_{X^*,X} \leq 0 \quad \forall\, y \in T^{-1}(C). \tag{5.17}$$

Proof. Assumption (A3) and Theorem 5.6 ensure that $g(\bar{x}) = 0$. If $|\mathbb{P}_i^c \cup \mathbb{M}_i| = \infty$, then Theorem 5.6 and the weak lower semicontinuity of $d_{C_i} \circ T$ [40, L. 1.5] imply that $T\bar{x} \in C_i$. As a result, if $|\mathbb{P}_i^c \cup \mathbb{M}_i| = \infty$ for all $i = 1, \ldots, m$, then $T\bar{x} \in C$. Now, assume that there exists at least one i for which $|\mathbb{P}_i^c \cup \mathbb{M}_i| < \infty$. For such i, $|\mathbb{P}_i| = \infty$, which implies that $r_\ell^{(k_j)} \to \infty$ for all $\ell = 1, \ldots, m$. Lemma 5.2 then ensure that $\theta^{(k_j)} \lambda_\ell^{(k_j)} \to 0$ for $\ell = 1, \ldots, m$. Now, by (5.13) and (5.15), we have

$$\begin{aligned} &\theta^{(k_j)} \sum_{i=1}^m T^* \Lambda_i(x^{(k_j)}, \lambda_i^{(k_j)}, r_i^{(k_j)}) \\ &\qquad = \theta^{(k_j)}(e^{(k_j)} - (f'(x^{(k_j)}) + g'(x^{(k_j)})^* \zeta^{(k_j)})) \rightharpoonup^* 0. \end{aligned} \tag{5.18}$$

Expanding the left hand side yields

$$\theta^{(k_j)} \sum_{i=1}^m \Lambda_i(x^{(k_j)}, \lambda_i^{(k_j)}, r_i^{(k_j)}) = Tx^{(k_j)} + \sum_{i=1}^m \{\theta^{(k_j)} \lambda_i^{(k_j)} - t_i^{(k_j)} \mathbf{P}_{C_i}(z_i^{(k_j)})\}, \tag{5.19}$$

where $t_i^{(k_j)} := \theta^{(k_j)} r_i^{(k_j)} > 0$, $t_1^{(k_j)} + \ldots + t_m^{(k_j)} = 1$, and $z_i^{(k_j)} := (r_i^{(k_j)})^{-1} \lambda_i^{(k_j)} + Tx^{(k_j)}$. Note that for all $k_j > K_0$, we have $t_i^{(k_j)} = t_i^{(K_0)} = \bar{t}_i$. By assumption (A2), we have that $\ker T^* = \{0\}$. Since $\{x^{(k_j)}\}$ converges weakly, it is bounded and hence $\{P_{C_i}(z_i^{(k_j)})\}$ is also bounded for $i = 1, \ldots, m$. Therefore, $\{P_C(z_i^{(k_j)})\}$ has a weakly converging subsequence (that we do not relabel) with limit $\bar{y}_i \in C_i$ since Z is a Hilbert space and C_i is closed and convex (hence, weakly closed) for $i = 1, \ldots, m$. Consequently, the sequence on the left hand side of (5.18) converges weakly* to $T^*(T\bar{x} - \sum_i \bar{t}_i \bar{y}_i)$. Owing to the uniqueness of weak* limits, we have that $T^*(T\bar{x} - \sum_i \bar{t}_i \bar{y}_i) = 0$ and hence (5.16) holds since $\ker T^* = \{0\}$. Moreover, (5.17) follows from (5.13) and Proposition 5.7 since $\{x^{(k_j)}\}$ is bounded. ∎

The next result is a simple consequence of Theorem 5.9 that employs common assumptions from the convergence theory for SQP (cf. [26] for more details) that ensure the results of Theorem 5.9 hold.

COROLLARY 5.10. *Suppose there exists a set $\Omega \subseteq X$ such that $x^{(k)} \in \Omega$ for all k and for which f' and g' are uniformly bounded on Ω. Moreover, assume that $\{\zeta^{(k)}\}$ is bounded. Then, any weak accumulation point of $\{x^{(k)}\}$ satisfies* (5.16) *and* (5.17) *holds for any bounded subsequence of $\{x^{(k)}\}$. In particular, Algorithm 4.1 either terminates in a finite number of iterations with an approximate stationary point or it produces an infinite sequence $\{x^{(k)}\}$ for which all bounded subsequence satisfy* (5.17) *and all weak accumulation points of $\{x^{(k)}\}$ satisfy* (5.16).

Remark 5.11 (Weak Limits and Feasibility). A consequence of Corollary 5.10 is that if $\{x^{(k)}\}$ has a weak accumulation point, then the feasible set for optimization problem (3.1) is nonempty. Notably, if X is reflexive and Ω is bounded, then $\{x^{(k)}\}$ has a weakly converging subsequence. The assumption that Ω is bounded is used to prove convergence of the augmented Lagrangian algorithm in [12, Assumption AS2].

Theorem 5.9 does not ensure that weak accumulation points $\bar{x}$ of $\{x^{(k)}\}$ are feasible. It only shows that $T\bar{x}$ is an element of the convex hull of $C_1 \cup \cdots \cup C_m$. We conclude this section with some common situations for which $T\bar{x}$ is guaranteed to be feasible. In these cases, Algorithm 4.1 either terminates in a finite number of iterations or it generates a sequence $\{x^{(k)}\}$ that satisfies the asymptotic stationarity condition (5.14) and whose weak accumulation points are feasible.

COROLLARY 5.12. *Let the assumptions of Theorem 5.9 hold and suppose one of the following conditions holds:*

(a) $m = 1$;
(b) $\mathbf{P}_{C_i}$ *is weakly continuous for* $i = 1, \ldots, m$;
(c) T *is completely continuous;*
(d) $x^{(k_j)}$ *converges strongly to* $\bar{x}$;
(e) There exists a Banach space X_0 *that is compactly embedded in* X *such that* $\{x^{(k_j)}\} \subset X_0$ *and* $x^{(k_j)} \rightharpoonup \bar{x}$ *in* X_0;
(f) X *is finite dimensional.*

Then, $\{x^{(k_j)}\}$ *satisfies the asymptotic stationarity condition* (5.17) *and* $T\bar{x} \in C$.

Proof. Case (a) is obvious. For cases (b)–(d), we have that $\bar{y}_i = \mathbf{P}_{C_i}(T\bar{x})$. Therefore, (5.16) shows that $T\bar{x}$ is a fixed point of the map $\sum_i \bar{t}_i \mathbf{P}_{C_i}(\cdot)$ and it follows from [3, Prop 4.34] that the fixed points of this map are exactly the set C. Moreover, if (e) holds, then the compact embedding of X_0 in X ensures that $x^{(k_j)} \to \bar{x}$ in X and the result follows from (d). Finally, if (f) holds, then (b), (c), (d) and (e) also hold. ∎

Remark 5.13 (Second-Order Optimality Conditions). It may be possible to prove strong convergence of $\{x^{(k_j)}\}$ under additional conditions such as second-order optimality conditions. See [9], where this is done for a related augmented Lagrangian algorithm in Banach space.

5.4. Convergence to Stationary Points. Theorem 5.6 gives sufficient conditions for Algorithm 4.1 to terminate in a finite number of iterations. However, it does not ensure that the sequence of iterates $\{x^{(k)}\}$ satisfies the first-order stationary conditions (3.2). In this subsection, we address this question. The next results demonstrate the limiting behavior of the sequence of iterates $\{x^{(k)}\}$ generated by Algorithm 4.1. In particular, if a weak accumulation point of $\{x^{(k)}\}$ exists, then that point must be a first-order stationary point. This result requires additional regularity assumptions on the problem data f and g. In particular, we assume the following.

Assumption 5.14 (Regularity of Derivatives).

(A5) The derivative of the equality constrained Lagrangian satisfies: If $x_k \rightharpoonup x$ in X, $\zeta_k \rightharpoonup^* \zeta$ in Y^* and

$$\limsup_{k\to\infty} \langle f'(x_k) + g'(x_k)^*\zeta_k, x_k - x\rangle_{X^*,X} \le 0,$$

then for all $y \in T^{-1}(C)$, the following holds

$$\langle f'(x) + g'(x)^*\zeta, x - y\rangle_{X^*,X} \le \liminf_{k\to\infty} \langle f'(x_k) + g'(x_k)^*\zeta_k, x_k - y\rangle_{X^*,X}.$$

A brief discussion of assumption (A5) is in order. If X is finite dimensional, then weak and strong convergence coincide, and therefore the continuity of $f' : X \to X^*$ and $g' : X \to \mathcal{L}(X, Y)$ (assumption (A1)) ensures that

$$f'(x_k) + g'(x_k)^*\zeta_k \to f'(x) + g'(x)^*\zeta \qquad \text{and} \qquad x_k \to x.$$

Consequently, (A5) is satisfied. In infinite dimensions, it may require quite strong assumptions to satisfy (A5). Following our next result, we provide assumptions that ensure that assumption (A5) holds.

THEOREM 5.15. *Consider the setting of Theorem 5.6 and let assumptions (A1)–(A5) hold. Let* $(\bar{x}, \bar{\zeta}) \in X \times Y^*$ *be a weak/weak* accumulation point of* $\{(x^{(k)}, \zeta^{(k)})\}$. *If* $T\bar{x} \in C$, *then* $\bar{x}$ *is a first-order stationary point of* (3.1). *That is,* $\bar{x}$ *satisfies* (3.2).

Proof. Let $\{(x^{(k_j)}, \zeta^{(k_j)})\}$ denote a subsequence such that $x^{(k_j)} \rightharpoonup \bar{x}$ and $\zeta^{(k_j)} \rightharpoonup^* \bar{\zeta}$ and suppose $T\bar{x} \in C$. Assumption (A3) ensures that $g(\bar{x}) = 0$. We now prove the first condition in (3.2). Proposition 5.7, the fact that $T\bar{x} \in C$, and the strong convergence of $\{e^{(k_j)}\}$ ensure that

$$\begin{aligned}\limsup_{k_j\to\infty} &\langle -(f'(x^{(k_j)}) + g'(x^{(k_j)})^*\zeta^{(k_j)}), \bar{x} - x^{(k_j)}\rangle_{X^*,X} \\ &= \limsup_{k_j\to\infty} \langle -(f'(x^{(k_j)}) + g'(x^{(k_j)})^*\zeta^{(k_j)}) + e^{(k_j)}, \bar{x} - x^{(k_j)}\rangle_{X^*,X} \\ &= \limsup_{k_j\to\infty} \sum_{i=1}^{m} (T^*\Lambda_i(x^{(k_j)}, \lambda_i^{(k_j)}, r_i^{(k_j)}), \bar{x} - x^{(k_j)})_Z \le 0.\end{aligned}$$

The result then follows from assumption (A5). ∎

The next result provides strong assumptions on the derivatives f' and g' that ensure assumption (A5) holds. In addition, the result employs a constraint qualification to ensure that $\{\zeta^{(k)}\}$ is bounded and hence has a weakly* converging subsequence. These are then used to show that the results of Theorem 5.9 hold. The assumptions used in this result were motivated by [9, § 5].

COROLLARY 5.16. *Consider the setting of Theorem 5.6, let assumptions (A1)–(A4) hold. Moreover, assume that f' and g' satisfy the following assumptions:*

(A6) The derivative $f' : X \to X^*$ is pseudomonotone, i.e.,

$$\begin{aligned} x_k \rightharpoonup x \quad \text{and} \quad &\limsup_{k\to\infty} \langle f'(x_k), x_k - x\rangle_{X^*,X} \le 0 \\ \Longrightarrow \quad &\langle f'(x), x - y\rangle_{X^*,X} \le \liminf_{k\to\infty} \langle f'(x_k), x_k - y\rangle_{X^*,X} \quad \forall\, y \in X;\end{aligned}$$

(A7) The Jacobian $g' : X \to \mathcal{L}(X, Y)$ is sequentially weak-to-strong continuous and $g'(x)^* \in \mathcal{L}(Y^*, X^*)$ is sequentially weak-to-strong continuous for all $x \in X$.*

Then, assumption (A5) holds. If, in addition,

(A8) For any bounded set $D \subset X$, the set $\{f'(x) \,|\, x \in D\} \subseteq X^*$ is bounded,

and $\bar{x}$ is a weak accumulation point of $\{x^{(k)}\}$ that satisfies the extended Robinson constraint qualification,

$$0 \in \operatorname{int} g'(\bar{x})(T^{-1}(C) - \bar{x}). \tag{5.20}$$

Then, $\bar{x}$ satisfies (3.2) as long as $T\bar{x} \in C$.

Proof. We first show that assumption (A5) holds. Let $x_k \rightharpoonup x$ in X and $\zeta_k \rightharpoonup^* \zeta$ in Y^*. By assumption (A7), we have that $g'(x_k) \to g'(x)$ and $g'(x)^*$ is sequentially weak*-to-strong continuous. As a result, we have that $g'(x)^*\zeta_k \to g'(x)^*\zeta$, which yields $g'(x_k)^*\zeta_k \to g'(x)^*\zeta$. Assumption (A5) then follows from assumption (A6).

Now, suppose (5.20) holds at $\bar{x}$ and let $\{x^{(k_j)}\}$ denote a subsequence of $\{x^{(k)}\}$ that weakly converges to $\bar{x}$ with associated multiplier subsequence $\{\zeta^{(k_j)}\}$. The generalized open mapping theorem [42, Th. 2.1] ensures that there exists $\rho > 0$ such that

$$B_\rho^Y \subset g'(\bar{x})[(T^{-1}(C) - \bar{x}) \cap B_1^X].$$

Let $y^{(k_j)} \in B_1^Y$ be a sequence of unit vectors satisfying $\langle \zeta^{(k_j)}, y^{(k_j)}\rangle_{Y^*,Y} \ge \frac{1}{2}\|\zeta^{(k_j)}\|_{Y^*}$. As a result of the above inclusion, we have that $-\rho y^{(k_j)} \in B_\rho^Y$ and there exists a bounded sequence $\{v^{(k_j)}\}$ in $T^{-1}(C)$ such that

$$-\rho y^{(k_j)} = g'(\bar{x})(v^{(k_j)} - \bar{x})$$

Assumption (A7) ensures that $-\rho y^{(k_j)} = g'(x^{(k_j)})(v^{(k_j)} - x^{(k_j)}) + \eta^{(k_j)}$, where $\eta^{(k_j)} \to 0$. Now, for k_j sufficiently large so that $\|\eta^{(k_j)}\|_Y \le \frac{\rho}{4}$, we have that

$$\begin{aligned}\frac{\rho}{2}\|\zeta^{(k_j)}\|_{Y^*} &\le \langle \zeta^{(k_j)}, \rho y^{(k_j)}\rangle_{Y^*,Y} \\ &\le \langle \zeta^{(k_j)}, -g'(x^{(k_j)})(v^{(k_j)} - x^{(k_j)})\rangle_{Y^*,Y} + \frac{\rho}{4}\|\zeta^{(k_j)}\|_{Y^*}.\end{aligned} \tag{5.21}$$

With $e^{(k)}$ as defined in (5.13), we can rewrite

$$-g'(x^{(k_j)})^*\zeta^{(k_j)} = f'(x^{(k_j)}) + \sum_{i=1}^m T^*\Lambda_i(x^{(k_j)}, \lambda_i^{(k_j)}, r_i^{(k_j)}) - e^{(k_j)}.$$

Substituting this expression into (5.21) and rearranging terms gives

$$\begin{aligned}\frac{\rho}{4}\|\zeta^{(k_j)}\|_{Y^*} &\le \langle f'(x^{(k_j)}), v^{(k_j)} - x^{(k_j)}\rangle_{X^*,X} \\ &\quad + \sum_{i=1}^m \langle T^*\Lambda_i(x^{(k_j)}, \lambda_i^{(k_j)}, r_i^{(k_j)}) - e^{(k_j)}, v^{(k_j)} - x^{(k_j)}\rangle_{X^*,X}.\end{aligned} \tag{5.22}$$

Notice that Proposition 5.7 ensures that there exists a sequence $\{n^{(k_j)}\} \in (0,\infty)$ such that $n^{(k_j)} \searrow 0$ and

$$\sum_{i=1}^m \langle T^*\Lambda_i(x^{(k_j)}, \lambda_i^{(k_j)}, r_i^{(k_j)}), v^{(k_j)} - x^{(k_j)}\rangle_{X^*,X} \le n^{(k_j)}$$

since $v^{k_j} \in T^{-1}(C)$. Therefore, the right hand side of (5.22), and hence $\{\zeta^{(k_j)}\}$, is bounded by (A8) and the fact that $e^{(k)} \to 0$. Since $\{\zeta^{(k_j)}\}$ is bounded, it has a weakly* convergent subsequence by the Banach-Alaoglu Theorem [16, Th. 5.18]. The desired result then follows from Theorem 5.15. ∎

6. Algorithmic Extensions.

We now present extensions to Algorithm 4.1. Our first extension allows us to separately penalize multiple individual constraints. The second extension allows us to handle nonlinear constraint operators T in (3.1).

6.1. Finitely Many Linear Constraints.

In this subsection, we consider the common setting in which there are finitely many constraints of the form $T_i x \in \widehat{C}_i$, where $T_i \in \mathcal{L}(X, Z_i)$ and $\widehat{C}_i \in Z_i$ is nonempty, closed and convex for $i = 1, \ldots, m$. Here, Z_i denotes a real Hilbert space for $i = 1, \ldots, m$. We first note that this setting can be stated in the more general setting of (3.1) by defining

$$Z := Z_1 \oplus \cdots \oplus Z_m, \qquad Tx := (T_1 x, \ldots, T_m x) \qquad \text{and} \qquad C := \widehat{C}_1 \times \cdots \times \widehat{C}_m. \tag{6.1}$$

Consequently, it is straightforward to apply Algorithm 4.1 with T and C defined above, either by treating $Tx \in C$ as a single constraint or by handling $T_i x \in \widehat{C}_i$ individually. In particular, define

$$C_i := \left(\prod_{j=1}^{i-1} Z_j\right) \times \widehat{C}_i \times \left(\prod_{j=i+1}^{m} Z_j\right), \quad i = 1, \ldots, m,$$

where $\prod$ denotes the Cartesian product and the first and last products are void if $i = 1$ and $i = m$, respectively. It is clear from the definition of C and C_i that assumption

(A4) holds. Moreover, using these definitions, we see that

$$I_C(Tx) = \sum_{i=1}^{m} I_{C_i}(Tx) = \sum_{i=1}^{m} I_{\widehat{C}_i}(T_i x)$$

and hence, Ψ_i and Λ_i only depend on $T_i x$ and $\widehat{C}_i$ for $i = 1, \ldots, m$. Unfortunately, T as defined above need not satisfy assumption (A2). In particular, the adjoint operator T^*, which is given by $T^* z = T_1^* z_1 + \ldots + T_m^* z_m$ for $z = (z_1, \ldots, z_m)$ and $z_i \in Z_i$ for $i = 1, \ldots, m$, need not be injective even if T_i^* is for $i = 1, \ldots, m$. Fortunately, there are practical situations where T^* is in fact injective. One such situation is when the optimization space X is a direct sum of Banach spaces. This is often the case for optimal control problems, in which case X is typically composed of the state and control spaces.

Suppose there exists real Banach spaces X_i and operators $\widehat{T}_i \in \mathcal{L}(X_i, Z_i)$ satisfying (A2), $i = 1, \ldots, m$, for which

$$X = X_1 \oplus \cdots \oplus X_m \qquad \text{and} \qquad T_i x = \widehat{T}_i x_i \quad \text{for} \quad i = 1, \ldots, m. \tag{6.2}$$

In this case, $T = \widehat{T}_1 \oplus \cdots \oplus \widehat{T}_m$ satisfies assumption (A2). Therefore, Theorem 5.9 applies directly to problems of this type. In fact, we can show that $T\bar{x} \in C$.

THEOREM 6.1. *Consider the setting of Theorem 5.6 and let assumption (A1)–(A4) hold. Let X, Z, T, and C be defined as in (6.1) and (6.2), and suppose there exists a weakly converging subsequence $\{x^{(k_j)}\}$ with limit $\bar{x} \in X$ such that*

$$r_i^{(k_j)} \to \infty \quad \Longrightarrow \quad \tfrac{1}{r_i^{(k_j)}} \big(f_{x_i}(x^{(k_j)}) + g_{x_i}(x^{(k_j)})^* \zeta^{(k_j)} \big) \rightharpoonup^* 0, \tag{6.3}$$

for $i = 1, \ldots, m$. Here, f_{x_i} and g_{x_i} denote the partial derivatives of f and g with respect to x_i, $i = 1, \ldots, m$, respectively. Then, $\bar{x}$ satisfies $T\bar{x} \in C$.

Proof. Clearly if $|\mathbb{P}_i^c \cup \mathbb{M}_i| = \infty$ then $\widehat{T}_i \bar{x}_i \in \widehat{C_i}$ for $i = 1, \ldots, m$. We assume that at least one $i = 1, \ldots, m$ satisfies $|\mathbb{P}_i| = \infty$. Using the product structure of the problem, (6.3) ensures that

$$\tfrac{1}{r_i^{(k_j)}} T_i^* \Lambda_i(x_i^{(k_j)}, \lambda_i^{(k_j)}, r_i^{(k_j)}) = \tfrac{1}{r_i^{(k_j)}} \Big(e_i^{(k_j)} - \big(f_{x_i}(x^{(k_j)}) + g_{x_i}(x^{(k_j)})^* \zeta^{(k_j)} \big) \Big) \rightharpoonup^* 0,$$

for which the left hand side can be expanded as

$$\tfrac{1}{r_i^{(k_j)}} T_i^* \Lambda_i(x_i^{(k_j)}, \lambda_i^{(k_j)}, r_i^{(k_j)}) = T_i^* (z_i^{(k_j)} - \mathbf{P}_{\widehat{C}_i}(z_i^{(k_j)})),$$

where $z_i^{(k_j)} = (r_i^{(k_j)})^{-1} \lambda_i^{(k_j)} + \widehat{T}_i x_i^{(k_j)}$ Since $(r_i^{(k_j)})^{-1} \lambda_i^{(k_j)} \to 0$ by Lemma 5.2 and $\widehat{T}_i x_i^{(k_j)} \rightharpoonup \widehat{T}_i \bar{x}_i$, the sequence $\{z_i^{(k_j)}\}$ is bounded and hence so is $\{\mathbf{P}_{\widehat{C}_i}(z_i^{(k_j)})\}$. Furthermore, since Z_i is a Hilbert space, $\{\mathbf{P}_{\widehat{C}_i}(z_i^{(k_j)})\}$ has a weakly converging subsequence (that we do not relabel) with limit $\bar{y}_i \in \widehat{C}_i$. The injectivity of $\widehat{T}_i^*$ and the uniqueness of weak* limits then ensure that $\widehat{T}_i \bar{x}_i = \bar{y}_i \in \widehat{C}_i$ for $i = 1, \ldots, m$. ∎

6.2. Nonlinear Constraints. We now consider the addition of the nonlinear constraint $T_0(x) \in C_0$ to (3.1). Here, $T_0 : X \to Z_0$, where Z_0 is a real Hilbert space and $C_0 \subseteq Z_0$ is nonempty, closed and convex. We define the penalty function Ψ_0 and the multiplier update functions Λ_0 analogously to Ψ_i and Λ_i for $i = 1, \ldots, m$.

Additionally, we define $r_0^{(k)}$, $\lambda_0^{(k)}$, $\mathbb{P}_0$ and $\mathbb{M}_0$ analogously. With these definitions, the results in Section 5.1 and Theorem 5.6 hold with no modifications. In the next theorem, we demonstrate how the result of Theorem 5.9 changes in this setting.

THEOREM 6.2. *Consider the setting of Theorem 5.9. Let $\bar{x} \in X$ be a weak accumulation point of $\{x^{(k)}\}$ with associated subsequence $\{x^{(k_j)}\}$ and suppose that there exists $\alpha \in (0,1)$ such that*

$$\theta^{(k_j)} := \Big(\sum_{i=0}^{m} r_i^{(k_j)}\Big)^{-1} \to 0 \quad \implies \quad (\theta^{(k_j)})^\alpha (f'(x^{(k_j)}) + g'(x^{(k_j)})^* \zeta^{(k_j)}) \rightharpoonup^* 0. \tag{6.4}$$

Moreover, assume that:

(A9) T_0 is completely continuous, continuously Fréchet differentiable, and the derivative T_0' satisfies

$$x_k \rightharpoonup x, \;\; y_k \rightharpoonup y \quad \text{in } X \qquad \implies \qquad T_0'(x_k) y_k \rightharpoonup T_0'(x) y.$$

If $|\mathbb{P}_i^c \cup \mathbb{M}_i| = \infty$ for $i = 0, \dots, m$, then $T\bar{x} \in C$ and $T_0(\bar{x}) \in C_0$. On the other hand, if there exists at least one $i = 0, \dots, m$ for which $|\mathbb{P}_i^c \cup \mathbb{M}_i| < \infty$, then $\bar{x}$ satisfies

$$\langle T_0'(\bar{x})^* (T_0(\bar{x}) - \mathbf{P}_{C_0}(T_0(\bar{x}))), y - \bar{x} \rangle_{X^*,X} \geq 0 \quad \forall\, y \in T^{-1}(C). \tag{6.5}$$

Finally, if $\bar{x}$ satisfies the extended Robinson constraint qualification

$$0 \in \operatorname{int}\{T_0(\bar{x}) + T_0'(\bar{x})(T^{-1}(C) - \bar{x}) - C_2\}, \tag{6.6}$$

then $T_0(\bar{x}) \in C_0$ and (5.16) holds.

Proof. If $|\mathbb{P}_i^c \cup \mathbb{M}_i| = \infty$ for $i = 0, \dots, m$, then the feasibility follows from the arguments in the proof of Theorem 5.9. To prove the remaining results, suppose that $|\mathbb{P}_i^c \cup \mathbb{M}_i| < \infty$ for at least one $i = 0, \dots, m$. In this case, (6.4) ensures that

$$\begin{aligned} &\theta^{(k_j)} \sum_{i=1}^{m} T^* \Lambda_i(x^{(k_j)}, \lambda_i^{(k_j)}, r_i^{(k_j)}) + \theta^{(k_j)} T_0'(x^{(k_j)})^* \Lambda_0(x^{(k_j)}, \lambda_0^{(k_j)}, r_0^{(k_j)}) \\ &\qquad = \theta^{(k_j)} \Big(e^{(k_j)} - \big(f'(x^{(k_j)}) + g'(x^{(k_j)})^* \zeta^{(k_j)} \big) \Big) \rightharpoonup^* 0 \end{aligned} \tag{6.7}$$

and assumption (A9) and Lemma 5.2 ensure that the left hand side of (6.7) weakly* converges to

$$\sum_{i=1}^{m} \bar{t}_i T^* (T\bar{x} - \bar{y}_i) + \bar{t}_0 T_0'(\bar{x})^* (T_0(\bar{x}) - \mathbf{P}_{C_0}(T_0(\bar{x}))) = 0 \tag{6.8}$$

for $\bar{y}_i \in C_i$, $i = 1, \dots, m$, as in Theorem 5.9. Assumption (A9) further ensures that

$$\begin{aligned} &\lim_{k_j \to \infty} \theta^{(k_j)} \langle T_0'(x^{(k_j)})^* \Lambda_0(x^{(k_j)}, \lambda_0^{(k_j)}, r_0^{(k_j)}), y - x^{(k_j)} \rangle_{X^*,X} \\ &\qquad = \bar{t}_0 \langle T_0'(\bar{x})^* (T_0(\bar{x}) - \mathbf{P}_{C_0}(T_0(\bar{x}))), y - \bar{x} \rangle_{X^*,X} \quad \forall\, y \in X \end{aligned}$$

and therefore Proposition 5.7 applied to the $i = 1, \dots, m$ constraints combined with (6.4) and (6.7) implies (6.5). In particular, $(\theta^{(k_j)})^{1-\alpha}(y - x^{(k_j)}) \to 0$ since $\{x^{(k_j)}\}$

converges weakly (and hence is bounded) and

$$
\begin{aligned}
&-\bar{t}_0\langle T_0'(\bar{x})^*(T_0(\bar{x})-\mathbf{P}_{C_0}(T_0(\bar{x})), y-\bar{x}\rangle_{X^*,X} \\
&= \lim_{k_j\to\infty} \theta^{(k_j)}\Big\{ -\langle T_0'(x^{(k_j)})^*\Lambda_0(x^{(k_j)},\lambda_0^{(k_j)},r_0^{(k_j)}), y-x^{(k_j)}\rangle_{X^*,X} \\
&\qquad + \langle e^{(k_j)} - (f'(x^{(k_j)}) + g'(x^{(k_j)})^*\zeta^{(k_j)}), y-x^{(k_j)}\rangle_{X^*,X}\Big\} \\
&\le \limsup_{k_j\to\infty} \theta^{(k_j)}\sum_{i=1}^m \langle T^*\Lambda_i(x^{(k_j)},\lambda_i^{(k_j)},r_i^{(k_j)}), y-x^{(k_j)}\rangle_{X^*,X} \le 0 \quad \forall\, y\in T^{-1}(C).
\end{aligned}
$$

To conclude, suppose that $\bar{x}$ satisfies (6.6), then there exists $\rho > 0$ such that $B_\rho^{Z_0} \subseteq T_0(\bar{x}) + T_0'(\bar{x})(T^{-1}(C)-\bar{x}) - C_0$. In particular, for any $z\in B_\rho^{Z_0}$, there exists $y\in T^{-1}(C)$ and $c\in C_0$ such that $z = T_0(\bar{x}) + T_0'(\bar{x})(y-\bar{x}) - c$. Therefore,

$$
\begin{aligned}
(T_0(\bar{x}) - \mathbf{P}_{C_0}(T_0(\bar{x})), z)_{Z_0} &= \langle T_0'(\bar{x})^*(T_0(\bar{x}) - \mathbf{P}_{C_0}(T_0(\bar{x}))), y-\bar{x}\rangle_{X^*,X} \\
&\quad + (T_0(\bar{x}) - \mathbf{P}_{C_0}(T_0(\bar{x})), T_0(\bar{x}) - c)_{Z_0}.
\end{aligned}
$$

The first term on the right hand side is nonnegative by the above arguments and the second term is nonnegative by [3, Th. 3.14]. Since this holds for all $z\in B_\rho^{Z_0}$, we have that $T_0(\bar{x}) - \mathbf{P}_{C_0}(T_0(\bar{x})) = 0$ and consequently $T_0(\bar{x})\in C_0$. Combining this with (6.8) and the injectivity of T^* shows that (5.16) holds. ∎

Remark 6.3 (Optimality of $\bar{x}$). Note that if $T\bar{x}\in C$, then the variational inequality (6.5) is the first-order optimality conditions for the optimization problem

$$\min_{Tx\in C}\; d_{C_0}(T_0(x))^2.$$

7. Solution of the subproblem. An important motivation for the ALESQP method is to enable iterative, and therefore inexact, solution of linear systems involving the discretizations of $g'(x)$. Therefore, a good choice for the solution of the augmented Lagrangian subproblem (4.5) is the inexact, matrix-free trust-region SQP method [19, 26]. To provide context for some modifications related to its use with the augmented Lagrangian, we give a short summary of the method. For this, we assume that X and Y are Hilbert spaces. We define the SQP Lagrangian $\mathscr{L} : X\times Y^*\to\mathbb{R}$ for (4.5), which includes the augmented Lagrangian,

$$\mathscr{L}(x,\zeta) := L(x) + \langle\zeta, g(x)\rangle_{Y^*,Y},$$

where $L(x) := L_k(x)$, and k denotes the k-th augmented Lagrangian iteration. The SQP method [19] extends the composite-step approach of [28] to rigorously handle inexact linear system solves. In the context of (4.5), the method comprises the following steps at its j^{th} iteration. We start with an iterate x_j, the corresponding Lagrange multiplier ζ_j, a trust-region radius Δ_j and a self-adjoint approximation of $\nabla_{xx}\mathscr{L}(x_j,\zeta_j)$, denoted by $H_j = H(x_j,\zeta_j)$, with $H\in\mathcal{L}(X,X)$. First, to reduce the linear infeasibility, $\|g'(x_j)n + g(x_j)\|_Y$, we approximately solve the *quasi-normal subproblem*,

$$\text{(7.1)}\qquad \min_{n\in X}\; \|g'(x_j)n + g(x_j)\|_Y^2 \qquad \text{subject to} \qquad \|n\|_X \le 0.8\Delta_j,$$

using Powell's dogleg method [31], where we compute the second-order (Newton) step by iteratively solving an *augmented system*, subject to the stopping condition provided in [26, Eqn. 34]. Second, given a solution n_j of (7.1), to improve optimality we solve the *tangential subproblem*,

$$\begin{aligned} &\min_{\widetilde{t}\in X} \ \tfrac{1}{2}(H_j\widetilde{t},\widetilde{t})_X + (\widetilde{W}_j(\nabla_x\mathscr{L}(x_j,\zeta_j)+H_jn_j),\widetilde{t})_X \\ &\text{subject to } \ \widetilde{t}\in \mathrm{Range}(\widetilde{W}_j)\,, \quad \|\widetilde{t}+n_j\|_X \le \Delta_j, \end{aligned} \tag{7.2}$$

using the projected conjugate gradient (CG) method with Steihaug-Toint termination criteria [36]. In (7.2) the linear operator $\widetilde{W}_j \in \mathcal{L}(X,X)$ represents an approximate projection onto the null-space of $g'(x_j)$, $\ker(g'(x_j))$. Its action on a vector is given by the solution of another augmented system, with the stopping conditions [26, Eqns. 37 and 39]. Third, to ensure that the trial step remains sufficiently close to $\ker(g'(x_j))$, i.e., maintains linearized feasibility, we additionally project the solution $\widetilde{t}_j$ of (7.2) onto $\ker(g'(x_j))$ to compute the *tangential step* t_j. To accomplish this, we solve an augmented system with the stopping condition [26, Eqn. 41]. This yields the trial step $s_j = n_j + t_j$. Fourth, we compute the Lagrange multipliers ζ_{j+1} by solving another augmented system, with the stopping conditions [26, Eqn. 43]. Finally, we apply trust-region acceptance and update criteria, see [26, Alg. 4, Steps 3 and 4].

The aforementioned augmented systems are optimality systems of the form

$$\begin{pmatrix} R_{X,X^*} & g'(x_j)^* \\ g'(x_j) & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}, \tag{7.3}$$

where $R_{X,X^*} \in \mathcal{L}(X,X^*)$ is the inverse Riesz map [18]. If the system (7.3) is solved directly, the residual $(e_1 \ \ e_2)^\top$ is ignored. If it is solved iteratively, the size of $(e_1 \ \ e_2)^\top$ can be controlled using stopping conditions of the form

$$\|e_1\|_{X^*} + \|e_2\|_Y \le \mathcal{T}(\|y_1\|_X, \|b_1\|_{X^*}, \|b_2\|_Y, \Delta_j, \tau_{\mathrm{nom}}),$$

where $\mathcal{T}:\mathbb{R}^5_+ \to \mathbb{R}_+$ is a tolerance function specific to each step, as referenced above, and $\tau_{\mathrm{nom}} > 0$ is a chosen nominal tolerance. The SQP method adjusts linear system tolerances based on its progress, in order to ensure global convergence under standard assumptions. A discussion of the theoretical assumptions, the nominal tolerance choice, the Riesz map, and some modifications to the augmented system (7.3) follows.

Function-space setting. The SQP algorithm [19] requires that X and Y be Hilbert spaces. In Section 8 all numerical examples satisfy these assumptions, justifying the application of the algorithm. However, our augmented Lagrangian algorithm and the corresponding convergence theory are developed in the more general setting of Banach spaces. In order to apply ALESQP in Banach space, extensions to the SQP algorithm are necessary. For instance, different notions of Cauchy points and Cauchy decrease conditions are needed, see [13, Sec. 8.3.2]; projections onto $\ker(g'(x))$ as discussed previously do not apply; the objective function in (7.2) must be modified; etc. The required extensions, while plausible, are beyond the scope of this paper.

Lipschitz continuous derivatives. In [19] it is assumed that the functions L and g are twice continuously differentiable. This is an appropriate assumption for all numerical examples in Section 8 *in the absence* of the constraints $Tx \in C$, i.e., when $L = f$. Once the constraints are included, the constraint penalty terms in L render L' Lipschitz continuous, see Corollary 4.4. The proof of Theorem 3.5 in [19] is easily extended to handle $\mathscr{L}$ with Lipschitz continuous derivatives. Specifically, the second-order Taylor expansion used on page 1537 of [19] can be replaced with the first-order expansion, followed by the use of Lipschitz continuity of $\mathscr{L}'$; see the assumption AW.1c for the composite-step algorithm analyzed in [13], and Theorem 3.1.4 in [13].

Nominal linear solver stopping tolerance. The theory in [19] permits an arbitrary choice of the nominal tolerance $\tau_{\text{nom}} > 0$. For good numerical performance, we choose $\tau_{\text{nom}} = \min\{\sqrt{\delta^{(k)}}, \sqrt{\epsilon^{(k)}}\}$. The same value is used for all augmented system solves, i.e., the nominal tolerances $\tau^{qn}, \tau^{pg}, \tau^{proj}, \tau^{tang}$ and τ^{lmh} from [19, 26].

Implementation of the Riesz maps. The SQP algorithm is posed in Hilbert space, and therefore naturally supports the use of Riesz maps, such as $R_{X^*,X}$. However, in large-scale applications the Riesz map may require an iterative solution of additional linear systems—nested within the iterative augmented system solve, or in other components of the SQP algorithm. Inexact or variable Riesz maps are not supported by [19]. To circumvent this challenge, in Section 8 we use diagonal Riesz map discretizations. This enables Riesz maps that are exact to within machine precision.

Preconditioning of the projected CG method. In certain applications we can accelerate the projected CG iteration for the solution of (7.2) by replacing the augmented system solve that yields the constraint null-space projection with the solution of a related linear system. Motivated by a comment about the "perfect preconditioner" for projected CG [17, p. 1381], we solve a system of the form

$$\begin{pmatrix} B(x_j) + T^*\Big(\sum_{i=1}^m r_i^{(k)}\big(R_{Z,Z^*} - D_{ij}\big)\Big)T & g'(x_j)^* \\ g'(x_j) & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ 0 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}, \tag{7.4}$$

where $B(x_j) \in \mathcal{L}(X, X^*)$ approximates $f''(x_j)$ and D_{ij} denotes the Newton derivative [11, Def. 2.1] of $\mathbf{P}_{C_i}((r_i^{(k)})^{-1}\lambda_i^{(k)} + Tx_j)$. Our assumptions for the use of this system are as follows: (i) f is twice continuously differentiable; (ii) X is a Hilbert space and B is nonnegative; and (iii) $(e_1 \;\; e_2)^\top = (0 \;\; 0)^\top$, i.e., only a direct solve is permitted. It is possible to relax the third assumption, and allow iterative solves of (7.4). However, this leads to two challenges, namely the derivation of stopping conditions for the iterative solve to replace conditions [26, Eqns. 37 and 39] and, more importantly, the question of efficient preconditioning of (7.4). Both are beyond the scope of this paper, which is why we enforce the third assumption whenever (7.4) is used in Section 8.

8. Applications. In this section, we demonstrate Algorithm 4.1 on three infinite dimensional optimization problems. Our first problem computes a probability density function (pdf) by maximizing the Rényi entropy. The second and third problems are optimization problems constrained by PDEs. Throughout, $\Omega \subset \mathbb{R}^d$ with $d = 1, 2$, and $L^p(\Omega)$, $p \in [1, \infty]$, denotes the usual p-order Lebesgue space. Moreover, $L^p_+(\Omega)$ will denote the subset of nonnegative $L^p(\Omega)$ functions. We denote by $\partial\Omega$ the boundary of Ω, and by $W^{1,s}(\Omega)$ and $H^1(\Omega) := W^{1,2}(\Omega)$ the Sobolev spaces of weakly differentiable functions [1]. Furthermore, we denote by $W_0^{1,s}(\Omega)$ the subspace of $W^{1,s}(\Omega)$-functions that are zero on the boundary in the trace sense, and $H_0^1(\Omega) := W_0^{1,2}(\Omega)$. All examples are discretized using continuous piecewise linear finite elements on regular simplicial meshes. We use diagonal Riesz map discretizations in all components of ALESQP, associated with the function spaces X, Y and Z. In particular, we use the lumped mass matrix for both $L^2(\Omega)$ and $H_0^1(\Omega)$.

We choose the following parameters for Algorithm 4.1:

(a) zero initial guesses throughout, i.e., $x^{(0)} = 0$, $\zeta^{(0)} = 0$ and $\lambda_i^{(1)} = 0$, for $i = 0, \dots, m$, with the exception of the initial guess $x^{(0)} = 1$ for the Rényi entropy example (due to the presence of the log function);

(b) initial SQP subproblem stopping tolerances

$$\delta^{(0)} = \epsilon^{(0)} = \max\{10^{-3}\|L_k'(x^{(0)}) + g'(x^{(0)})^*\zeta^{(0)}\|_{X^*}, 10^{-6}\};$$

(c) the basic tolerance update, see Remark 4.2, with reduction factors $\delta_0 = 0.25$, $\delta_1 = 0.9$, $\epsilon_0 = 0.25$ and $\epsilon_1 = 0.9$;
(d) the g-feasibility, optimality and T- feasibility stopping tolerances $\delta_* = \epsilon_* = \tau_* = 10^{-6}$, respectively;
(e) update factors $\eta_i = 5$, for $i = 0, 1, \dots, m$, for the augmented Lagrangian penalties; and
(f) $\overline{\eta} = 5$, $K_0 = 10^3$, $\theta_i = 0.1$, $\alpha_i = 0.1$, $\beta_i = 0.9$, $\tau_i^{(0)} = 1$, $\nu_i = 10^6$, $\gamma_i = 0.49$, for $i = 0, 1, \dots, m$.

Initial augmented Lagrangian penalty parameters. As in all augmented Lagrangian methods, the choice of the initial penalty parameters is important for good performance, and ALESQP is no exception. We use two general guidelines when choosing the initial parameters. First, they should be chosen as large as possible, without detriment to the convergence of the SQP subproblem solver. A conservative choice is $r_i^{(1)} = 10$, for $i = 0, 1, \dots, m$. This is the default choice in ALESQP. Second, they should be chosen so that all terms comprising the augmented Lagrangian functional are well balanced. In our first example, the inequality constraint scaling is such that the augmented Lagrangian terms are well balanced, and we can use the default penalty parameter choice. In the second and third examples, the problem structure—specifically the splitting of the variables into states and controls, combined with the PDE nature of the equality constraint linking the states and controls—dictates a more subtle choice, described in more detail in Section 8.2.

In the presented results, `AL` denotes the total number of augmented Lagrangian iterations, `SQP` the total number of SQP iterations, `CG` the total number of CG iterations, `normg` the equality constraint violation $\|g(\bar{x})\|_Y$, `grad-lag` the norm of the gradient of the subproblem Lagrangian $\|L_k'(\bar{x}) + g'(\bar{x})^*\bar{\zeta}\|_{X^*}$, and `feas` the constraint violation $\max_i d_{C_i}(T\bar{x})$. We implemented the entire ALESQP framework in Matlab (R2019a), and studied its performance using a single core of a 2.9GHz Intel Core i9 processor and 32GB of RAM. The problem instances studied here range in size from 4,225 to 524,801 optimization variables.

8.1. Maximum Entropy. The purpose of this example is to demonstrate mesh independent performance using direct and iterative linear system solves. Our maximum entropy problem seeks a pdf, x, that satisfies certain moment constraints. Let $\Omega = [0, 1]^2$, $X = L^p(\Omega)$, $Y = \mathbb{R}^3$, $Z = L^2(\Omega)$, $Z_0 = \mathbb{R}$, $C = L^2_+(\Omega)$, and $C_0 = [0, 1]$. We solve

$$\min_{X \in X} \left\{ f(x) := \frac{1}{p-1} \log\left(\int_\Omega x(\omega)^p \, \mathrm{d}\omega \right) \right\} \tag{8.1a}$$

$$\text{subject to} \quad Tx := x \geq 0 \ \text{ a.e.} \tag{8.1b}$$

$$g_1(x) := \int_\Omega x(\omega) \, \mathrm{d}\omega - 1 = 0 \tag{8.1c}$$

$$g_2(x) := \int_\Omega x(\omega)\omega \, \mathrm{d}\omega - \mu = 0 \tag{8.1d}$$

$$T_0(x) := \sigma^{-1} \det\left(\int_\Omega x(\omega)(\omega - \mu)(\omega - \mu)^\top \, \mathrm{d}\omega \right) \leq 1, \tag{8.1e}$$

where $\sigma > 0$ and $\mu \in \mathbb{R}^2$ are given, $g(x) = (g_1(x), g_2(x))$, and the objective function is the negative p-order Rényi entropy [38] with $p = 2.5$. Constraints (8.1b) and (8.1c) ensure that x is a pdf, (8.1d) ensures that the expected value associated with x is

μ, and (8.1e) ensures that the generalized variance [39] associated with x is smaller than σ. A straightforward computation shows that T_0 satisfies assumption (A9) and therefore Theorem 6.2 applies. We use the problem data

$$\mu = (0.45, 0.45) \qquad \text{and} \qquad \sigma = \frac{1}{2}\det\left(\int_\Omega (\omega-\mu)(\omega-\mu)^\top \,\mathrm{d}\omega\right) \approx 0.00368,$$

where the latter is chosen so the generalized variance associated with the optimal pdf is less than a half of the generalized variance associated with the uniform density.

For our numerical results, we use the default initial penalty parameters, $r_0^{(1)} = 10$ and $r_1^{(1)} = 10$, because the constraint values at the initial guess are well balanced. In particular, $\|Tx^{(0)}\|_Z = 1$ and $|T_0(x^{(0)})| = 2$. Table 1 documents ALESQP performance as the problem size grows, using direct solutions of the augmented systems (7.3). We observe nearly mesh-independent iteration numbers for the augmented Lagrangian loop and all its iterative components. We note that for this example the penalty parameters do not increase; e.g., for the 128×128 mesh the final values are $r_0^{(7)} = 10$ and $r_1^{(7)} = 10$. Table 2 documents ALESQP performance with iterative augmented system solves, where we have used unpreconditioned MINRES [29] to solve (7.3). Again, we observe nearly mesh-independent iteration numbers for the augmented Lagrangian loop and all its iterative components. Most notably, the total number of MINRES iterations is around 3,000, and it does not change significantly as the mesh is refined. In other words, we have demonstrated discretization-independent algorithmic performance of a fully matrix-free framework on an infinite-dimensional optimization problem with nonlinear inequality constraints. Finally, we note that the solution time for the matrix-free approach increases linearly with problem size, with the wallclock time of 5 seconds on the smallest mesh and 358 seconds on the largest mesh.

`Mesh`	`AL`	`SQP`	`CG`	`normg`	`grad-lag`	`feas`
64x64	8	44	249	1.36e-16	8.32e-08	2.06e-07
128x128	7	42	239	7.85e-17	1.61e-07	4.60e-07
256x256	7	47	246	6.23e-16	3.47e-07	8.98e-07
512x512	6	43	261	5.27e-16	9.35e-07	4.37e-07

TABLE 1

Maximum Entropy, direct solution of (7.3). *ALESQP performance for varying spatial discretization (*`Mesh`*). The* `AL`*,* `SQP`*, and* `CG` *iteration numbers are nearly mesh independent.*

`Mesh`	`AL`	`SQP`	`CG`	`normg`	`grad-lag`	`feas`	`tot.aug`	`avg.aug`
64x64	8	42	222	7.78e-14	2.00e-07	5.05e-07	2969	7.9
128x128	9	47	207	4.12e-15	3.01e-07	4.21e-08	2859	7.6
256x256	10	54	234	3.24e-16	8.72e-07	3.66e-08	3217	7.5
512x512	8	57	273	5.27e-16	9.57e-07	1.63e-07	3516	7.4

TABLE 2

Maximum Entropy, iterative solution of (7.3). *ALESQP performance for varying spatial discretization (*`Mesh`*). The* `AL`*,* `SQP`*, and* `CG` *iteration numbers are nearly mesh independent. The* `tot.aug` *column gives the total number of MINRES iterations in augmented system solves, for the entire run of ALESQP. The* `avg.aug` *column gives the average number of MINRES iterations per augmented system solve. The* `tot.aug` *and* `avg.aug` *iterations vary little as the mesh is refined.*

8.2. Semilinear Elliptic PDE with Control and State Constraints. The purpose of this example is to demonstrate nearly mesh-independent performance of

ALESQP on a PDE-constrained optimization problem with control and state constraints. From now on, we solve (7.4) to accelerate the projected CG method. As mentioned earlier, with (7.4) we only use direct linear system solves. Additionally, we demonstrate that ALESQP meets constraint tolerances at the level of machine precision with only marginally increased iteration counts. Let $\Omega = (0,1)^2$, $X = X_1 \oplus X_2$ with $X_1 = H_0^1(\Omega)$ and $X_2 = L^2(\Omega)$, $Y = H^{-1}(\Omega)$, and $Z_i = L^2(\Omega)$ for $i = 1,2,3$. We consider the problem

$$\min_{u\in X_1,\, z\in X_2} \left\{ f(u,z) := \frac{1}{2}\|u-u_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|z\|^2_{L^2(\Omega)} \right\} \tag{8.2a}$$

$$\text{subject to} \qquad u_a \le u \quad \text{a.e. in } \Omega \tag{8.2b}$$

$$z_a \le z \le z_b \quad \text{a.e. in } \Omega \tag{8.2c}$$

$$g(u,z) := \begin{cases} -\Delta u + u^3 - z = 0 & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases}, \tag{8.2d}$$

where $u_d \in L^2(\Omega)$, $z_a, z_b \in L^2(\Omega)$ with $z_a \le z_b$ a.e. in Ω, $u_a \in C(\overline{\Omega})$ with $u_a \le 0$ on $\partial\Omega$, and $\alpha > 0$ is the penalty parameter, are given. Moreover, $g(u,z) = 0$ is the weak form of (8.2d), $\widehat{T}_1$ is the compact embedding of $H_0^1(\Omega)$ into $L^2(\Omega)$, $\widehat{T}_2$ and $\widehat{T}_3$ are the identity operator on $L^2(\Omega)$, $\widehat{C}_1 := \{u \in Z_1 \,|\, u_a \le u\}$, $\widehat{C}_2 := \{z \in Z_2 \,|\, z_a \le z\}$, and $\widehat{C}_3 := \{z \in Z_3 \,|\, z \le z_b\}$. One can show that the solution to (8.2d), $u \in H_0^1(\Omega)$, in fact satisfies $u \in C_0(\Omega)$, where $C_0(\Omega)$ is the space of continuous functions on $\overline{\Omega}$ that vanish on the boundary $\partial\Omega$ [10]. In addition, from [27, Th. 2.14], we recall that the Lagrange multiplier ζ associated with the constraint $g(u,z) = 0$ satisfies $\zeta \in W_0^{1,s}(\Omega)$, with $s \in [1,2)$. Consequently, if z_a, $z_b \in W^{1,s}(\Omega)$, then we can conclude that the optimal control to (8.2) satisfies $z \in W^{1,s}(\Omega)$, where $W^{1,s}(\Omega)$ is compactly embedded in $L^2(\Omega)$. As a consequence of Theorem 6.1 and Corollary 5.12(e), any weak accumulation point in $H_0^1(\Omega) \oplus W^{1,s}(\Omega)$ of the sequence of iterates generated by Algorithm 4.1 is feasible, so long as (6.3) holds.

We investigate ALESQP performance on three scenarios: *(i)* only control constraints (i.e., $u_a = -\infty$); *(ii)* only state constraints (i.e., $z_a = -\infty$ and $z_b = \infty$); and *(iii)* both control and state constraints. For our numerical studies, we set $r_1^{(1)} = 10^3$, $r_2^{(1)} = \alpha$ and $r_3^{(1)} = \alpha$. The choice of initial penalty parameters is important to account for the differences in regularity and scaling of the associated multipliers. More precisely, we use α for the control-constraint parameters to balance the control penalty term $\frac{\alpha}{2}\|z\|^2_{L^2(\Omega)}$ in the objective function. Additionally, we note that the control-constraint multipliers are in $L^2(\Omega)$, and that we expect the corresponding penalty parameters to remain bounded. In contrast, the state-constraint multiplier is a measure, which suggests that the sequence $\{\|\lambda_1^{(k)}\|_Z\}_{k=1}^\infty$ is unbounded. Consequently, Lemma 5.2 suggests that sequence of penalty parameters $\{r_1^{(k)}\}_{k=1}^\infty$ is also unbounded. Therefore, it is appropriate to choose a very large initial parameter, here $r_1^{(1)} = 10^3$. In our studies, considerably larger values of $r_1^{(1)}$ had little impact on overall performance, including the solution of the SQP subproblems. Smaller values delayed the convergence of the outer augmented Lagrangian loop somewhat.

The problem data in (8.2) is motivated by [9]. In particular, we set $u_d \equiv -1, \alpha = 10^{-3}, z_a \equiv -10, z_b \equiv 10$ and

$$u_a(x) = -\tfrac{2}{3} + \tfrac{1}{2}\min\{x_1 + x_2, \min\{1 + x_1 - x_2, \min\{1 - x_1 + x_2, 2 - x_1 - x_2\}\}\}.$$

For scenario *(i)*, we replace $z_b = 10$ by $z_b = -1$, to ensure that the constraints are active. In Table 3 we observe nearly mesh-independent performance of ALESQP for all three scenarios. Additionally, we note that the state-constraint penalty parameter increases significantly and that the growth of the control-constraint penalty parameters is more moderate; e.g., for the 128×128 mesh in scenario *(iii)* the final values are $r_1^{(14)} = 3.12 \cdot 10^6$, $r_2^{(14)} = 1.25 \cdot 10^{-1}$ and $r_3^{(14)} = 6.25 \cdot 10^{-1}$. Moreover, if we tighten the outer stopping tolerances to 10^{-12}, as we do later in Table 4, the final penalty parameter values are $r_1^{(25)} = 3.91 \cdot 10^8$, $r_2^{(25)} = 6.25 \cdot 10^{-1}$ and $r_3^{(25)} = 6.25 \cdot 10^{-1}$. In other words, the state-constraint penalty continues to grow, while the control-constraint penalties stagnate. The discrepancy between $r_1^{(25)}$ and $r_2^{(25)}$ or $r_3^{(25)}$ strongly underlines the need for multiple penalties.

	Control			State			Control + State		
Mesh	AL	SQP	CG	AL	SQP	CG	AL	SQP	CG
64x64	9	20	41	11	28	57	12	35	71
128x128	9	20	40	12	33	66	14	44	90
256x256	9	22	45	14	40	80	16	50	102
512x512	9	22	47	16	47	95	18	60	123

TABLE 3

Semilinear. `Control`*: only control constraints;* `State`*: only state constraints;* `Control + State`*: control and state constraints. ALESQP performance for varying spatial discretization (*`Mesh`*). In all cases, we observe that the* `AL`*,* `SQP`*, and* `CG` *iterations are nearly mesh independent.*

In Table 4, we illustrate a remarkable feature of our algorithm. We consider a fixed mesh of size 128×128 and vary the outer stopping tolerances, including the T-feasibility tolerance τ_*. We observe that it is possible to achieve machine precision for all convergence measures with almost no increase in the total number of projected CG iterations.

tol	AL	SQP	CG	normg	grad-lag	feas
1e-6	14	44	90	9.51e-13	7.99e-12	1.35e-08
1e-8	17	46	94	9.56e-13	1.31e-13	7.43e-10
1e-10	21	48	98	9.58e-13	8.66e-14	2.77e-12
1e-12	25	50	102	9.67e-13	8.20e-14	2.49e-14

TABLE 4

Semilinear (Control + State). *ALESQP performance for varying stopping tolerances, $\varepsilon_* = \delta_* = \tau_* =$* `tol`*. We observe that it is possible to achieve machine precision for all convergence measures with a very mild increase in the iteration counts.*

8.3. Burgers' PDE with Control and State Constraints. This example showcases ALESQP in the context of *dynamic optimization.* Let $\Omega = (0,1)$, $Q := \Omega \times (0,T)$, $\Sigma := \partial\Omega \times (0,T)$, $X = X_1 \oplus X_2$ with $X_1 = L^2(0,T;H_0^1(\Omega))$ and $X_2 = L^2(Q)$, $Y = L^2(0,T;H^{-1}(\Omega))$, and $Z_1 = Z_2 = Z_3 = L^2(Q)$. We consider the problem

$$\min_{u \in X_1,\, z \in X_2} \left\{ f(u,z) := \frac{1}{2}\|u - u_d\|^2_{L^2(Q)} + \frac{\alpha}{2}\|z\|^2_{L^2(Q)} \right\} \tag{8.3a}$$

$$\text{(8.3b)} \qquad \text{subject to} \qquad u_a \le u \quad \text{a.e. in } Q$$

$$\text{(8.3c)} \qquad z_a \le z \le z_b \quad \text{a.e. in } Q$$

$$\text{(8.3d)} \qquad g(u,z) := \begin{cases} \partial_t u + u\partial_x u - \nu \partial_{xx}^2 u - z = 0 & \text{in } Q \\ u = 0 & \text{on } \Sigma \\ u(\cdot,0) - u_0 = 0 & \text{in } \Omega \end{cases},$$

with datum $u_d \in L^2(Q)$, $u_a \in C(\overline{Q})$ with $u_a(x,t) \le 0$ for all $(x,t) \in \partial\Omega \times [0,T]$, $z_a, z_b \in L^2(Q)$ with $z_a \le z_b$ a.e. in Q. Moreover, $g(u,z) = 0$ is the weak form of (8.3d), $\widehat{T}_1$ is the embedding of $L^2(0,T;H_0^1(\Omega))$ into $L^2(Q)$, $\widehat{T}_2$ and $\widehat{T}_3$ are the identity operators on $L^2(Q)$, and the constraint sets $\widehat{C}_i$, $i = 1,2,3$, are defined similarly to the ones in Section 8.2. It is unclear if the above problem fully satisfies our theory. In principle, one may be able to use regularity arguments similar to the ones from Section 8.2. However, such regularity results are not known for ζ associated with (8.3d), and the required study is beyond the scope of the paper. Nevertheless, we observe that our algorithm solves this problem efficiently. We refer to [37] for regularity results involving the control-constrained case, case (i) below.

Similar to the previous example, we test ALESQP in three different scenarios: *(i)* control constraints; *(ii)* state constraints; and *(iii)* mixed constraints. For our numerical results, we set $r_1^{(1)} = 10^3$, $r_2^{(1)} = \alpha$ and $r_3^{(1)} = \alpha$. The choice of the initial penalty parameters is justified by the problem structure, and closely follows the considerations given in Section 8.2. For $t \in (0,1)$, we set $u_d = 1$ for $x \in (0,1/2)$ and $u_d = 0$ otherwise, $\alpha = 5 \times 10^{-2}$, $z_a = -1$, $z_b = 2$ and $u_a = 0$. Table 5 shows ALESQP iteration counts. As in all previous examples, we observe nearly mesh-independent performance. Similar to Section 8.2 we note that the state-constraint penalty parameter increases significantly, while the control-constraint penalty parameters increase moderately; e.g., for the 128×128 case in scenario *(iii)* the final values are $r_1^{(15)} = 1.25 \cdot 10^5$, $r_2^{(15)} = 1.25 \cdot 10^0$ and $r_3^{(15)} = 2.5 \cdot 10^{-1}$.

	Control			State			Control + State		
Mesh	AL	SQP	CG	AL	SQP	CG	AL	SQP	CG
64x64	10	22	86	6	28	78	11	43	135
128x128	10	25	91	8	31	89	11	54	166
256x256	10	29	104	8	33	92	14	72	210
512x512	10	29	105	9	34	95	14	68	201

TABLE 5

Burger's Equation. `Control`*: only control constraints;* `State`*: only state constraints;* `Control + State`*: control and state constraints. ALESQP performance for varying spatial and temporal discretization (*`Mesh`*). In all cases, the* `AL`*,* `SQP`*, and* `CG` *iterations are nearly mesh independent.*

REFERENCES

[1] R. Adams and J. F. Fournier, *Sobolev spaces*, vol. 140 of Pure and Applied Mathematics (Amsterdam), Elsevier/Academic Press, Amsterdam, second ed., 2003.

[2] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt, *On augmented Lagrangian methods with general lower-level constraints*, SIAM J. Optim., 18 (2007), pp. 1286–1309, https://doi.org/10.1137/060654797.

[3] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer New York, 2011.

[4] M. Bergounioux, *Augmented Lagrangian method for distributed optimal control problems with state constraints*, J. Optim. Theory Appl., 78 (1993), pp. 493–521, https://doi.org/10.1007/BF00939879.

[5] M. Bergounioux and K. Kunisch, *Augmented Lagrangian techniques for elliptic state constrained optimal control problems*, SIAM J. Control Optim., 35 (1997), pp. 1524–1543, https://doi.org/10.1137/S036301299529330X.

[6] D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods*, Computer Science and Applied Mathematics, Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1982.

[7] E. G. Birgin, L. F. Bueno, and J. M. Martínez, *Sequential equality-constrained optimization for nonlinear programming*, Comput. Optim. Appl., 65 (2016), pp. 699–721, https://doi.org/10.1007/s10589-016-9849-6.

[8] E. G. Birgin and J. M. Martínez, *Practical augmented Lagrangian methods for constrained optimization*, vol. 10 of Fundamentals of Algorithms, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2014, https://doi.org/10.1137/1.9781611973365.

[9] E. Börgens, C. Kanzow, and D. Steck, *Local and global analysis of multiplier methods for constrained optimization in Banach spaces*, SIAM J. Control Optim., 57 (2019), pp. 3694–3722, https://doi.org/10.1137/19M1240186.

[10] E. Casas, *Control of an elliptic problem with pointwise state constraints*, SIAM J. Control Optim., 24 (1986), pp. 1309–1318, https://doi.org/10.1137/0324078.

[11] X. Chen, Z. Nashed, and L. Qi, *Smoothing methods and semismooth methods for nondifferentiable operator equations*, SIAM J. Numer. Anal., 38 (2000), pp. 1200–1216, https://doi.org/10.1137/S0036142999356719, https://doi.org/10.1137/S0036142999356719.

[12] A. R. Conn, N. I. M. Gould, and P. L. Toint, *A globally convergent augmented lagrangian algorithm for optimization with general constraints and simple bounds*, SIAM Journal on Numerical Analysis, 28 (1991), pp. 545–572, https://doi.org/10.1137/0728030.

[13] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Trust-Region Methods*, Society for Industrial and Applied Mathematics, USA, 2000.

[14] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Lancelot: A Fortran Package for Large-Scale Nonlinear Optimization (Release A)*, Springer Series in Computational Mathematics, Springer Berlin Heidelberg, 2013.

[15] J. E. Dennis, M. El-Alem, and M. C. Maciel, *A global convergence theory for general trust-region-based algorithms for equality constrained optimization*, SIAM Journal on Optimization, 7 (1997), pp. 177–207, https://doi.org/10.1137/S1052623492238881.

[16] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts, Wiley, 1999.

[17] N. I. M. Gould, M. E. Hribar, and N. J., *On the solution of equality constrained quadratic programming problems arising in optimization*, SIAM Journal of Scientific Computing, 23 (2002), pp. 1376–1395, https://doi.org/10.1137/S1064827598345667.

[18] A. Günnel, R. Herzog, and E. Sachs, *A note on preconditioners and scalar products in Krylov subspace methods for self-adjoint problems in hilbert space*, Electron. Trans. Numer. Anal., 41 (2014), pp. 13–20.

[19] M. Heinkenschloss and D. Ridzal, *A matrix-free trust-region SQP method for equality constrained optimization*, SIAM J. Optim., 24 (2014), pp. 1507–1541, https://doi.org/10.1137/130921738.

[20] M. Heinkenschloss and L. N. Vicente, *Analysis of inexact trust-region SQP algorithms*, SIAM J. Optim., 12 (2001/02), pp. 283–302, https://doi.org/10.1137/S1052623499361543.

[21] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appl., 4 (1969), pp. 303–320, https://doi.org/10.1007/BF00927673.

[22] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE constraints*, vol. 23 of Mathematical Modelling: Theory and Applications, Springer, New York, 2009.

[23] K. Ito and K. Kunisch, *The augmented Lagrangian method for equality and inequality constraints in Hilbert spaces*, Math. Programming, 46 (1990), pp. 341–360, https://doi.org/10.1007/BF01585750.

[24] K. Ito and K. Kunisch, *An augmented Lagrangian technique for variational inequalities*, Appl. Math. Optim., 21 (1990), pp. 223–241, https://doi.org/10.1007/BF01445164.

[25] K. Ito and K. Kunisch, *Lagrange multiplier approach to variational problems and applications*, vol. 15 of Advances in Design and Control, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008, https://doi.org/10.1137/1.9780898718614.

[26] D. P. Kouri and D. Ridzal, *Inexact trust-region methods for PDE-constrained optimization*, in Frontiers in PDE-constrained optimization, vol. 163 of IMA Vol. Math. Appl., Springer, New York, 2018, pp. 83–121.

[27] I. Neitzel, J. Pfefferer, and A. Rösch, *Finite element discretization of state-constrained elliptic optimal control problems with semilinear state equation*, SIAM J. Control Optim., 53 (2015), pp. 874–904, https://doi.org/10.1137/140960645.

[28] E. O. Omojokun, *Trust-region algorithms for optimization with nonlinear equality and inequality constraints*, PhD thesis, University of Colorado, Boulder, CO, USA, 1989.

[29] C. C. Paige and M. A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM Journal on Numerical Analysis, 12 (1975), pp. 617–629.

[30] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, in Optimization (Sympos., Univ. Keele, Keele, 1968), Academic Press, London, 1969, pp. 283–298.

[31] M. J. D. Powell, *A new algorithm for unconstrained optimization*, in Nonlinear Programming (Proc. Sympos., Univ. of Wisconsin, Madison, Wis., 1970), Academic Press, New York, 1970, pp. 31–65.

[32] R. T. Rockafellar, *The multiplier method of Hestenes and Powell applied to convex programming*, J. Optim. Theory Appl., 12 (1973), pp. 555–562, https://doi.org/10.1007/BF00934777.

[33] R. T. Rockafellar, *Augmented Lagrange multiplier functions and duality in nonconvex programming*, SIAM J. Control, 12 (1974), pp. 268–285. Collection of articles dedicated to the memory of Lucien W. Neustadt.

[34] R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res., 1 (1976), pp. 97–116, https://doi.org/10.1287/moor.1.2.97.

[35] W. Rudin, *Functional Analysis*, International series in pure and applied mathematics, McGraw-Hill, 1991.

[36] T. Steihaug, *The Conjugate Gradient Method and Trust Regions in Large Scale Optimization*, SIAM J. Numer. Anal., 20 (1983), pp. 626–637, https://doi.org/10.1137/0720042.

[37] F. Tröltzsch and S. Volkwein, *The SQP method for control constrained optimal control of the Burgers equation*, ESAIM Control Optim. Calc. Var., 6 (2001), pp. 649–674, https://doi.org/10.1051/cocv:2001127.

[38] T. Van Erven and P. Harremos, *Rényi divergence and kullback-leibler divergence*, IEEE Transactions on Information Theory, 60 (2014), pp. 3797–3820.

[39] S. S. Wilks, *Certain generalizations in the analysis of variance*, Biometrika, 24 (1932), pp. 471–494, https://doi.org/10.1093/biomet/24.3-4.471.

[40] E. H. Zarantonello, *Projections on convex sets in Hilbert space and spectral theory*, in Contributions to nonlinear functional analysis (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1971), 1971, pp. 237–424.

[41] J. Ziems and S. Ulbrich, *Adaptive multilevel inexact SQP methods for PDE-constrained optimization*, SIAM J. Optim., 21 (2011), pp. 1–40, https://doi.org/10.1137/080743160.

[42] J. Zowe and S. Kurcyusz, *Regularity and stability for the mathematical programming problem in Banach spaces*, Applied mathematics and Optimization, 5 (1979), pp. 49–62.