

# Optimistic Noise-Aware Sequential Quadratic Programming for Equality Constrained Optimization with Rank-Deficient Jacobians

Albert S. Berahas<sup>\*†</sup>      Jiahao Shi<sup>‡</sup>      Baoyu Zhou<sup>§</sup>

April 2, 2025

## Abstract

We propose and analyze a sequential quadratic programming algorithm for minimizing a noisy nonlinear smooth function subject to noisy nonlinear smooth equality constraints. The algorithm uses a step decomposition strategy and, as a result, is robust to potential rank-deficiency in the constraints, allows for two different step size strategies, and has an early stopping mechanism. Under the linear independence constraint qualification, convergence is established to a neighborhood of a first-order stationary point, where the radius of the neighborhood is proportional to the noise levels in the objective function and constraints. Moreover, in the rank-deficient setting, the merit parameter may converge to zero, and convergence to a neighborhood of an infeasible stationary point is established. Numerical experiments demonstrate the efficiency and robustness of the proposed method.

## 1 Introduction

Constrained optimization problems with noisy objective functions and noisy equality constraints are ubiquitous and arise in a plethora of applications, e.g., machine learning [33], statistics [14], engineering design [34], and simulation optimization [1]. In this paper, we consider noisy equality-constrained optimization problems of the form

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad c(x) = 0, \quad (1.1)$$

---

<sup>\*</sup>Corresponding author.

<sup>†</sup>Dept. of Industrial and Operations Engineering, University of Michigan. ([albertberahas@gmail.com](mailto:albertberahas@gmail.com))

<sup>‡</sup>Dept. of Industrial and Operations Engineering, University of Michigan. ([jiahaos@umich.edu](mailto:jiahaos@umich.edu))

<sup>§</sup>School of Computing and Augmented Intelligence, Arizona State University. ([baoyu.zhou@asu.edu](mailto:baoyu.zhou@asu.edu))

where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$  are smooth, possibly nonconvex functions, and where access to the exact values and derivatives of both the objective and constraint functions is not possible. Instead, noisy approximations, denoted by over-bars, are used that satisfy

$$\|\bar{f}(x) - f(x)\| \leq \epsilon_f, \quad \|\bar{g}(x) - \nabla f(x)\| \leq \epsilon_g, \quad \|\bar{c}(x) - c(x)\| \leq \epsilon_c, \quad \text{and} \quad \|\bar{J}(x) - \nabla c(x)^T\| \leq \epsilon_J, \quad (1.2)$$

where  $(\bar{f}(x), \bar{g}(x), \bar{c}(x), \bar{J}(x))$  denote objective function and gradient and constraint function and Jacobian approximations, respectively, at  $x \in \mathbb{R}^n$ , and  $\{\epsilon_f, \epsilon_g, \epsilon_c, \epsilon_J\} \subset \mathbb{R}_{\geq 0}$  dictate the level of noise. Such problems pose significant challenges for optimization methods, particularly when noise impacts not only the objective function but also the constraints, potentially violating classical assumptions and making the notion of feasibility unclear.

Both the unconstrained and the constrained noisy settings have received much attention due to their wide applicability. In this paper, we focus on *noise-aware* algorithms for solving such problems, i.e., algorithms that exploit information about the noise and that are adaptive. In the unconstrained and bounded noise setting, several noise-aware algorithms that leverage noise-level dependent constants (e.g.,  $\epsilon_f$  and  $\epsilon_g$ ) to evaluate the acceptability of steps within line search [5, 6, 29, 48] or trust region [2, 12, 30, 44] methods have been proposed. A natural extension of these algorithms to the constrained setting assumes bounded noise in the objective function and associated derivatives, and possibly in the constraint functions. In this context, and when noise is present only in the objective function, [32] proposed a line search gradient projection method with a noise level estimation approach, and [35] proposed a line search sequential quadratic programming (SQP) method that uses a differentiable exact augmented Lagrangian as the merit function. Related to our problem setting, in the nonlinear equality-constrained setting with bounded noise in both the objective and constraint functions, [38] proposed a line search SQP method for the setting in which the linear independence constraint qualification (LICQ) is satisfied, and recently a study proposed a trust-region SQP method for the setting in which the LICQ is violated [45]. Under the same noise setting but for a different problem class, in [22] a line search interior-point method is proposed for solving bound-constrained problems with bounded noise in the objective function, and the methodology is extended in [17] to tackle general inequality-constrained problems with bounded noise in the objective and constraint functions under the LICQ.

Another related but distinct problem class that has received significant attention in recent years involves problems with deterministic constraints and a stochastic objective function. In this setting, it is often assumed that an unbiased gradient approximation with bounded variance is available; see e.g., [4, 7, 8, 9, 18, 20, 21, 24, 40]. This problem class presents a different set of challenges for algorithm design, analysis, and implementation, but shares some commonalities with our problem setting. Perhaps the most closely related works to ours are [7, 21]. In [7], the authors consider situations where the LICQ does not hold and propose an SQP method with a step decomposition strategy inspired by the Byrd-Omojokun approach [37]. In [21], an SQP method with step decomposition and

termination tests for the subproblems is introduced to enhance computational efficiency. Our approach builds upon these two key ideas with the necessary modifications to solve noisy equality-constrained problems.

## 1.1 Contributions

The goal of this paper is to design, analyze, and implement an SQP algorithm for solving noisy equality-constrained optimization problems with possibly rank-deficient constraint Jacobians. Specifically, we focus on problems where both the objective and constraint functions, as well as their associated derivatives, are contaminated with bounded noise. The challenge lies in dealing with this noise in conjunction with the nonlinear constraints and possible rank deficiency. Our algorithm is noise-aware and requires knowledge or estimates of the noise levels. The main contributions can be summarized as follows.

- We propose an SQP algorithm to solve noisy equality-constrained problems (1.1). Our algorithm is built upon the classical line search SQP framework [36, Chapter 18.4] and stochastic SQP methods [7, 8, 21], with the following adjustments to account for the noise, the possible rank deficiency, and for practicality.

**Optimistic feasibility:** Given the inherent noise in the constraint functions, achieving strict feasibility is not practical. Therefore, we consider feasibility to be attained if the constraint violation falls below a specified threshold related to the noise level in the constraint function ( $\epsilon_c$ ). We treat this as a feature of our algorithm, and build upon the notion of optimistic feasibility. Based on this, we design an early termination condition. This approach not only improves efficiency by potentially reducing the number of iterations but also retains the strong convergence guarantees associated with the SQP framework.

**Step decomposition and inexact solution:** Since rank deficiency in the Jacobian of the constraints (violation of the LICQ) is possible due to noise or other factors, we employ a Byrd-Omojukun step decomposition strategy to compute the search direction [37]. This strategy involves computing normal and tangential steps, which requires solving two subproblems. For practical purposes and to account for the noise, we allow for inexact solutions to these subproblems. We carefully design termination conditions for both subproblems that ensure that the inexact solutions are sufficiently accurate to achieve global convergence while enhancing computational efficiency.

**Flexible step size scheme:** We propose two efficient and theoretically sound schemes for selecting the step size parameter: an adaptive scheme and a line search scheme. The adaptive scheme is motivated by [7] and is objective-function-free, but requires estimates of the Lipschitz constants of both the objective and constraint functions. The line search scheme, motivated by [5, 36], requires function evaluations and knowledge of the noise levels (or estimates) in the objective and constraint functions.

- We prove that the iterates generated by the proposed SQP method, for both step size selection strategies, exhibit neighborhood convergence behavior, with the neighborhood size dictated by the noise in the objective and constraint functions. Specifically, in the full-rank setting, the iterates converge to a neighborhood of a first-order stationary point of (1.1). When the noisy constraint Jacobian is rank-deficient, the merit parameter may converge to zero, and the iterates converge to a neighborhood of an infeasible stationary point. These results are consistent with those for deterministic SQP variants [19], and when all noise parameters are zero, we recover the deterministic results. Additionally, we establish a complexity analysis for both feasibility (or infeasible stationarity) and stationarity measures, highlighting the efficiency of our approach even under noisy conditions. Under similar noisy conditions, in [45] a trust-region SQP method is proposed, and, in addition to the guarantees in the full-rank setting, convergence to a neighborhood of an infeasible stationary point is established in the rank-deficient setting. To establish the result in the full-rank setting, an assumption is made on the discrepancy between the null spaces of the noisy and the true Jacobians [45, Assumption 5], which can be significant even when the noise levels are small. Our analysis does not require this assumption. Furthermore, contrary to both [38, 45] which establish only asymptotic convergence of the iterates, we derive non-asymptotic bounds for the average stationarity and feasibility measures across iterations and iteration complexity results.
- We evaluate the performance of our proposed SQP method on a subset of equality-constrained problems from the CUTEst collection. By artificially adding different levels of noise, we conduct controlled experiments and test the robustness of the method, and consider both the full-rank and rank-deficient constraint Jacobian settings. We compare the robustness of the two step size selection strategies to the noise and also compare exact and inexact variants. Our experiments demonstrate that the method is robust to noise variations, with the early termination condition triggering appropriately, and that our method is competitive with existing noise-aware line search SQP methods [38].

## 1.2 Organization

We conclude this section by introducing the notation used throughout the paper. In Section 2, we discuss the main assumptions and present our proposed algorithm. The convergence guarantees of the algorithm are given in Section 3. Empirical results are presented in Section 4 and concluding remarks are provided in Section 5.

## 1.3 Notation

Let  $\mathbb{N} := \{0, 1, 2, \dots\}$  denote the set of natural numbers,  $\mathbb{R}$  denote the set of real numbers,  $\mathbb{R}^n$  denote the set of  $n$ -dimensional real vectors and  $\mathbb{R}^{m \times n}$  denote the set of  $m$ -by- $n$  di-

mensional real matrices. For any  $r \in \mathbb{R}$ ,  $\mathbb{R}_{\geq r}$  (resp.,  $\mathbb{R}_{> r}$ ) denotes the set of real numbers greater than or equal to  $r$  (resp., strictly greater than  $r$ ). Unless specified, norms  $\|\cdot\|$  are 2-norms. Let  $\text{Range}(A)$  and  $\text{Null}(A)$  denote the range space and the null space of  $A \in \mathbb{R}^{m \times n}$ , respectively. For any  $k \in \mathbb{N}$ , the subscript  $k$  denotes the quantity evaluated at iterate  $x_k$ , e.g.,  $f_k := f(x_k)$ . Any quantity with an overbar denotes a noisy estimate, e.g.,  $\bar{f}_k$  is a noisy estimate of  $f_k$ .

## 2 Noisy Sequential Quadratic Programming Method

In this section, we introduce our proposed noisy sequential quadratic programming (SQP) method. We begin with assumptions and preliminaries.

### 2.1 Assumptions and Preliminaries

We make the following assumption about problem (1.1) and the iterates generated by the algorithm.

**Assumption 2.1.** *We propose an iterative algorithm that generates a sequence of iterates  $\{x_k\} \subset \mathbb{R}^n$ . We assume that there exists an open convex set  $\mathcal{X} \subseteq \mathbb{R}^n$  containing all the iterates  $\{x_k\}$  and the trial points  $\{x_k + \bar{d}_k\}$  generated by our algorithm, where  $\{\bar{d}_k\}$  is the sequence of search directions. The objective function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is continuously differentiable and bounded below over  $\mathcal{X}$ , and its gradient  $\nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is Lipschitz continuous with constant  $L \in \mathbb{R}_{>0}$  and bounded over  $\mathcal{X}$ . The constraint function  $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$  and its Jacobian  $\nabla c^T : \mathbb{R}^n \rightarrow \mathbb{R}^{m \times n}$  are bounded over  $\mathcal{X}$ , and  $\nabla c^T$  is Lipschitz continuous with constant  $\Gamma \in \mathbb{R}_{\geq 0}$  over  $\mathcal{X}$ . Specifically, there exist constants  $(f_{inf}, f_{sup}, \kappa_g, \kappa_c, \kappa_J) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{>0} \times \mathbb{R}_{>0} \times \mathbb{R}_{>0}$  such that for all  $x \in \mathcal{X}$ ,*

$$f_{inf} \leq f(x) \leq f_{sup}, \quad \|\nabla f(x)\| \leq \kappa_g, \quad \|c(x)\| \leq \kappa_c, \quad \text{and} \quad \|\nabla c(x)\| \leq \kappa_J.$$

The next assumption pertains to the noisy evaluations of the objective and constraint functions, as well as their gradients.

**Assumption 2.2.** *There exist constants  $\{\epsilon_f, \epsilon_g, \epsilon_c, \epsilon_J\} \subset \mathbb{R}_{\geq 0}$  such that (1.2) holds for all  $x \in \mathcal{X}$ .*

**Remark 2.3.** *The smoothness and boundedness assumptions are standard in the deterministic continuous constrained optimization literature [11, 46], and the set  $\mathcal{X}$  is assumed to be open and convex but not necessarily bounded. Throughout the paper, we assume that the noise in the objective and constraint functions, as well as their associated gradients, is bounded (Assumption 2.2). Under this noise model, we show that the components of our proposed algorithm (e.g., step size and merit parameter) are well defined.*

Often, in nonlinear equality-constrained optimization, the linear independence constraint qualification (LICQ) is assumed to hold by imposing that the singular values of the Jacobian  $\nabla c(x)^T$  are bounded away from zero; see, e.g., [36, Assumption 18.1(a)] and [39]. We do not make such an assumption here; however, we consider this setting in our analysis (as a special case) in Section 3. Before proceeding, we make a few remarks about the singular values of the true and noisy Jacobians of the constraints, as they play a key role in equality-constrained optimization. The singular values of  $\nabla c(x)^T$  along with the upper bound of the noise in the constraint Jacobians  $\epsilon_J$  dictate the rank of  $\bar{J}(x)$ . Suppose that the singular values of  $\nabla c(x)^T$  are lower bounded by  $\kappa_\sigma \in \mathbb{R}_{>0}$  for all  $x \in \mathcal{X}$ , i.e., the true constraint Jacobian is of full rank. If  $\epsilon_J < \kappa_\sigma$  (as assumed in [38]), it then follows from Weyl's Theorem [47] that the singular values of  $\bar{J}(x)$  are lower bounded by  $\kappa_\sigma - \epsilon_J$  for all  $x \in \mathcal{X}$ . However, if  $\epsilon_J \geq \kappa_\sigma$ , then the full-rank property of  $\bar{J}(x)$  cannot be guaranteed. Now, suppose that  $\nabla c(x)^T$  is rank-deficient. In this setting, under Assumption 2.2 it is possible that  $\bar{J}(x)$  is also rank-deficient. That said, there are also examples for which  $\bar{J}(x)$  is of full rank with high probability; see [3, 13]. In this paper, we do not make explicit assumptions about the relationship between the singular values of  $\nabla c(x)$  and  $\epsilon_J$ , but consider the full rank setting as a special case in our analysis.

Let  $\mathcal{L}(x, y) = f(x) + c(x)^T y$  denote the Lagrangian corresponding to (1.1), where  $y \in \mathbb{R}^m$  represents a vector of Lagrange multipliers. Under Assumption 2.1 and the LICQ, a first-order stationary point  $x^*$  with respect to (1.1) satisfies

$$\nabla f(x^*) + \nabla c(x^*)y^* = 0 \quad \text{and} \quad c(x^*) = 0 \quad \text{for some } y^* \in \mathbb{R}^m. \quad (2.1)$$

Throughout this manuscript, we define the stationarity error at  $(x, y) \in (\mathbb{R}^n, \mathbb{R}^m)$  as  $\|\nabla f(x) + \nabla c(x)y\|$  and the feasibility error at  $x \in \mathbb{R}^n$  as  $\|c(x)\|$ . In the absence of constraint qualifications, it is possible that (1.1) is infeasible, or (1.1) can be degenerate, in which case (2.1) might not be a necessary condition. To address the difficulties that arise from the lack of constraint qualifications, it is essential to develop an algorithm that can transition its objective from finding a stationary point of (1.1) to identifying a stationary point relative to a measure of infeasibility. To this end, we define the infeasibility measure  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  as  $\varphi(x) = \frac{1}{2}\|c(x)\|^2$  and a point  $\tilde{x} \in \mathbb{R}^n$  is considered stationary with respect to  $\varphi$  if one of the following conditions is satisfied: (i)  $c(\tilde{x}) = 0$ , or (ii)  $c(\tilde{x}) \neq 0$  and  $0 = \nabla \varphi(\tilde{x}) = \nabla c(\tilde{x})c(\tilde{x})$ , where case (ii) is the definition of an infeasible stationary point. Henceforth, we define the infeasible stationarity error at  $x \in \mathbb{R}^n$  as  $\|\nabla c(x)c(x)\|$ .

## 2.2 A Noisy Sequential Quadratic Programming Method

In this section, we describe our proposed SQP method. The algorithm has foundations in the line search SQP paradigm [36, Chapter 18.4], with several modifications to account for the noise and possible violation of the LICQ.

To address the challenges posed by the discrepancy between the observed noisy constraint functions  $\bar{c}(x)$  and the true constraint functions  $c(x)$ , our algorithm has an early

termination feature. The concept, termed *optimistic feasibility*, is simple yet powerful. Due to the noise, even if  $\bar{c}(x) = 0$ , we cannot conclusively assert feasibility with respect to the true constraint functions. In fact, from a pessimistic standpoint, exact feasibility may never be guaranteed under our general noise model. To overcome this limitation, we adopt an optimistic perspective to feasibility. Specifically, if  $\|\bar{c}(x)\| \leq \epsilon_o$  for some  $\epsilon_o \in \mathbb{R}_{\geq 0}$ , we consider feasibility to be approximately achieved. Consequently, the search direction at an approximately feasible point focuses on improving the objective function while maintaining approximate feasibility. This approach prevents the algorithm from becoming sensitive to the noise, which could result in the unnecessary minimization of constraint violation. Moreover, our algorithm has an early termination scheme that is triggered when  $\|\bar{c}(x)\| \leq \epsilon_o$  and  $\Delta \bar{l}(x_k, \bar{r}_k, \bar{d}_k) \leq \epsilon_o$  and that returns an approximately first-order stationary iterate. In our algorithm,  $\epsilon_o \in [0, \epsilon_c]$  to ensure that if the algorithm terminates early, it does so at an iterate that is sufficiently feasible with respect to the noise in the constraint functions. Moreover,  $\epsilon_o = 0$  represents the pessimistic setting and  $\epsilon_o = \epsilon_c$  represents the most optimistic setting. In either setting, neighborhood convergence is established. In the optimistic setting, the algorithm may terminate early at an approximately first-order stationary point.

At every iteration  $k \in \mathbb{N}$  of our proposed algorithm, the search direction  $\bar{d}_k$  is computed using a step decomposition technique [37], i.e.,  $\bar{d}_k = \bar{v}_k + \bar{u}_k$  where  $\bar{v}_k \in \text{Range}(\bar{J}_k^T)$  is the *normal component* and  $\bar{u}_k \in \text{Null}(\bar{J}_k)$  is the *tangential component*. Each component is computed by solving an optimization subproblem. In particular, when  $\|\bar{J}_k^T \bar{c}_k\| \neq 0$ , the algorithm first computes the normal component  $\bar{v}_k$  as an approximate solution of the subproblem

$$\min_{\bar{v} \in \text{Range}(\bar{J}_k^T)} \frac{1}{2} \|\bar{c}_k + \bar{J}_k \bar{v}\|^2 \quad \text{s.t.} \quad \|\bar{v}\| \leq \sigma_{Jc} \|\bar{J}_k^T \bar{c}_k\|, \quad (2.2)$$

where  $\sigma_{Jc} \in \mathbb{R}_{>0}$ , that is required to satisfy the Cauchy-like decrease condition

$$\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\| \geq \bar{\gamma}_c \left( \|\bar{c}_k\| - \|\bar{c}_k + \bar{\alpha}_k^c \bar{J}_k \bar{v}_k^c\| \right), \quad (2.3)$$

where  $\bar{\gamma}_c \in (0, 1]$  is a parameter,  $\bar{v}_k^c := -\bar{J}_k^T \bar{c}_k$  is the negative gradient direction of the objective function of (2.2) at  $\bar{v} = 0$ , and  $\bar{\alpha}_k^c = \min \left\{ \sigma_{Jc}, \|\bar{J}_k^T \bar{c}_k\|^2 / \|\bar{J}_k \bar{J}_k^T \bar{c}_k\|^2 \right\}$  is the optimal step size along the direction of  $\bar{v}_k^c$  that minimizes  $\|\bar{c}_k + \alpha \bar{J}_k \bar{v}_k^c\|^2$  within the trust region constraint defined in (2.2). Subproblem (2.2) minimizes the reduced linearized constraint violation subject to a trust-region constraint. In our optimistic approach, we skip subproblem (2.2) when  $\|\bar{c}_k\| \leq \epsilon_o$  and set  $\bar{v}_k = 0$ , as reducing constraint violation is unnecessary. Efficient iterative techniques for solving (2.2) include the Steihaug-Toint conjugate gradient method (CG-Steihaug) [42] and GLTR [25].

After computing the normal component  $\bar{v}_k$ , the algorithm computes the tangential component  $\bar{u}_k$  as an approximate solution to the subproblem

$$\min_{\bar{u} \in \mathbb{R}^n} (\bar{g}_k + H_k \bar{v}_k)^T \bar{u} + \frac{1}{2} \bar{u}^T H_k \bar{u} \quad \text{s.t.} \quad \bar{J}_k \bar{u} = 0, \quad (2.4)$$

whose optimal solution  $\bar{u}_k^*$  and the associated dual multiplier  $\bar{y}_k^*$  (under Assumption 2.4) can be computed via the following linear system of equations

$$\begin{bmatrix} H_k & \bar{J}_k^T \\ \bar{J}_k & 0 \end{bmatrix} \begin{bmatrix} \bar{u}_k^* \\ \bar{y}_k^* \end{bmatrix} = - \begin{bmatrix} \bar{g}_k + H_k \bar{v}_k \\ 0 \end{bmatrix}. \quad (2.5)$$

Note that even when  $\bar{J}_k$  is rank-deficient, the solution component  $\bar{u}_k^*$  is unique while the dual multiplier component  $\bar{y}_k^*$  is not; see Lemma 3.21. Instead of computing an exact solution  $(\bar{u}_k^*, \bar{y}_k^*)$ , our algorithm relies on  $(\bar{u}_k, \bar{y}_k)$ , an inexact solution of (2.5), which can be computed via iterative Krylov subspace methods (e.g., MINRES [15]) that only require matrix-vector multiplications. We further define the residual of inexact solution  $(\bar{u}_k, \bar{y}_k)$  as

$$\begin{bmatrix} \bar{\rho}_k \\ \bar{r}_k \end{bmatrix} := \begin{bmatrix} H_k & \bar{J}_k^T \\ \bar{J}_k & 0 \end{bmatrix} \begin{bmatrix} \bar{u}_k \\ \bar{y}_k \end{bmatrix} + \begin{bmatrix} \bar{g}_k + H_k \bar{v}_k \\ 0 \end{bmatrix}. \quad (2.6)$$

We make the following standard assumptions about the matrix  $H_k$  in (2.4)–(2.6).

**Assumption 2.4.** *The sequence  $\{H_k\}$  is bounded, e.g., there exists a constant  $\kappa_H \in \mathbb{R}_{>0}$  such that  $\|H_k\| \leq \kappa_H$  for all  $k \in \mathbb{N}$ . In addition, for all  $k \in \mathbb{N}$ , the matrix  $H_k$  is sufficiently positive definite in the null space of both  $\nabla c_k^T$  and  $\bar{J}_k$ , e.g., there exists a constant  $\zeta \in \mathbb{R}_{>0}$  such that  $u^T H_k u \geq \zeta \|u\|^2$  for all  $u \in \text{Null}(\nabla c_k^T) \cup \text{Null}(\bar{J}_k)$ .*

**Remark 2.5.** *Classical SQP methods assume that  $H_k$  is sufficiently positive definite in the null space of  $\nabla c_k^T$  [36, Assumption 18.1(b)]. Although  $\nabla c_k$  is not computable in our setting, this assumption remains necessary for our analysis. Moreover, we require that  $H_k$  is sufficiently positive definite in the null space of the noisy constraint Jacobian  $\bar{J}_k$ . The identity matrix satisfies Assumption 2.4. In [38],  $H_k$  is a constant diagonal matrix that is strictly positive definite. We do not impose such structure on  $H_k$  in this paper.*

Next, our proposed algorithm relies on a merit function to balance the two possible competing goals of minimizing the objective function and minimizing the constraint violation. To this end, we use an exact  $\ell_2$ -merit function

$$\bar{\phi}(x, \bar{\tau}) := \bar{\tau} \bar{f}(x) + \|\bar{c}(x)\| \quad (2.7)$$

where  $\bar{\tau} \in \mathbb{R}_{>0}$  is a merit parameter. Several classical SQP methods (e.g., [27, 39]) use the  $\ell_1$  norm in the merit function. We use the  $\ell_2$  norm instead for ease of presentation and analysis, but note that the  $\ell_1$  norm or other norms could be used. Our algorithm makes use of a local first-order model of the merit function  $\bar{l} : \mathbb{R}^n \times \mathbb{R}_{>0} \times \mathbb{R}^n \rightarrow \mathbb{R}$ , and more specifically, the reduction in the model of the merit function  $\Delta \bar{l} : \mathbb{R}^n \times \mathbb{R}_{>0} \times \mathbb{R}^n \rightarrow \mathbb{R}$

$$\Delta \bar{l}(x, \bar{\tau}, \bar{d}) := \bar{l}(x, \bar{\tau}, 0) - \bar{l}(x, \bar{\tau}, \bar{d}) = -\bar{\tau} \bar{g}(x)^T \bar{d} + \|\bar{c}(x)\| - \|\bar{c}(x) + \bar{J}(x) \bar{d}\|. \quad (2.8)$$

The reduction in the model in the merit function is used in our early termination condition, i.e., when both  $\|\bar{c}_k\| \leq \epsilon_o$  and  $\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) \leq \epsilon_o$  the algorithm terminates.



The algorithm also uses the reduction in the model of the merit function in the termination tests and in the merit parameter update scheme. Two different termination tests are used depending on whether  $\|\bar{c}(x)\| \leq \epsilon_o$  or  $\|\bar{c}(x)\| > \epsilon_o$ . Given inexact normal components  $\bar{v}_k$  that satisfy (2.3), inexact tangential components  $\bar{u}_k$  that satisfy the following termination conditions are sufficient to guarantee convergence.

**Termination Test 1.** *This termination test is only considered when  $\|\bar{c}_k\| \leq \epsilon_o$ . Given  $\lambda_{\rho r} \in (0, 1), \kappa_{\rho r} \in \mathbb{R}_{>0}, \lambda_u \in (0, \zeta)$  where  $\zeta \in \mathbb{R}_{>0}$  is given in Assumption 2.4, and  $\sigma_u \in (0, 1), (\bar{u}_k, \bar{\rho}_k, \bar{r}_k)$  satisfy Termination Test 1 if*

$$\max\{\|\bar{\rho}_k\|, \|\bar{r}_k\|\} \leq \lambda_{\rho r} \min\{\max\{\|\bar{u}_k\|, \|\bar{J}_k^T \bar{c}_k\|\}, \kappa_{\rho r}\}, \quad (2.9)$$

$$\bar{u}_k^T H_k \bar{u}_k \geq \lambda_u \|\bar{u}_k\|^2 - \epsilon_o \quad (2.10)$$

$$\bar{g}_k^T \bar{u}_k + \frac{1}{2} \bar{u}_k^T H_k \bar{u}_k \leq \epsilon_o, \quad (2.11)$$

$$\text{and} \quad \Delta \bar{l}(x_k, \bar{\tau}_{k-1}, \bar{u}_k) \geq \bar{\tau}_{k-1} \sigma_u \max\{\bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2\} - \epsilon_o \quad (2.12)$$

hold. The algorithm then sets  $\bar{\tau}_k = \bar{\tau}_{k-1}$ .

**Termination Test 2.** *This termination test is only considered when  $\|\bar{c}_k\| > \epsilon_o$ . Given  $\lambda_{\rho r} \in (0, 1), \kappa_{\rho r} \in \mathbb{R}_{>0}, \lambda_{uv} \in \mathbb{R}_{>0}, \lambda_u \in (0, \zeta)$  where  $\zeta \in \mathbb{R}_{>0}$  is given in Assumption 2.4,  $\lambda_v \in \mathbb{R}_{>0}, \sigma_u \in (0, 1), \sigma_c \in (0, 1), \sigma_r \in (\sigma_c, 1)$ , and  $\bar{v}_k \in \text{Range}(\bar{J}_k^T)$  satisfying (2.3),  $(\bar{u}_k, \bar{\rho}_k, \bar{r}_k)$  satisfy Termination Test 2 if condition (2.9) holds and*

$$\|\bar{u}_k\| \leq \lambda_{uv} \|\bar{v}_k\| \text{ or } \left\{ \begin{array}{l} \bar{u}_k^T H_k \bar{u}_k \geq \lambda_u \|\bar{u}_k\|^2 \\ (\bar{g}_k + H_k \bar{v}_k)^T \bar{u}_k + \max\{\frac{1}{2}, 1 - \|\bar{J}_k^T \bar{c}_k\|\} \bar{u}_k^T H_k \bar{u}_k \leq \lambda_v \|\bar{v}_k\| \end{array} \right\} \quad (2.13)$$

together with at least one of the following two conditions,

$$\Delta \bar{l}(x_k, \bar{\tau}_{k-1}, \bar{d}_k) \geq \bar{\tau}_{k-1} \sigma_u \max\{\bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2\} + \sigma_c (\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|) \quad (2.14)$$

$$\text{or} \quad \|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k + \bar{r}_k\| \geq \sigma_r (\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|) > 0. \quad (2.15)$$

When (2.14) is satisfied, we set  $\bar{\tau}_k = \bar{\tau}_{k-1}$ . If (2.14) is not satisfied but (2.15) holds, given a user-defined parameter  $\sigma_\tau \in (0, 1)$ , the algorithm sets

$$\bar{\tau}_k \leftarrow \begin{cases} \bar{\tau}_{k-1} & \text{if } \bar{\tau}_{k-1} \leq (1 - \sigma_\tau) \bar{\tau}_k^{\text{trial}} \\ (1 - \sigma_\tau) \bar{\tau}_k^{\text{trial}} & \text{otherwise,} \end{cases} \quad (2.16)$$

where

$$\bar{\tau}_k^{\text{trial}} \leftarrow \begin{cases} \infty & \text{if } \bar{g}_k^T \bar{d}_k + \max\{\bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2\} \leq 0 \\ \frac{(1 - \frac{\sigma_c}{\sigma_r})(\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k + \bar{r}_k\|)}{\bar{g}_k^T \bar{d}_k + \max\{\bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2\}} & \text{otherwise.} \end{cases} \quad (2.17)$$

The conditions in Termination Test 2 are similar to those presented in [21]. The main motivation is to ensure that the reduction in the model of the merit function is sufficiently positive. This is achieved by either directly setting  $\bar{\tau}_k = \bar{\tau}_{k-1}$  when (2.14) is satisfied or via the merit parameter update rule (2.16)–(2.17). The conditions in Termination Test 1 are different, mainly due to the optimistic feasibility setting. In (2.10)–(2.12), we add a relaxation term ( $\epsilon_o$ ) to ensure that Termination Test 1 can be satisfied when the residual (i.e.,  $\max\{\|\bar{\rho}_k\|, \|\bar{r}_k\|\}$ ) is sufficiently small. Minimizing the constraint violation is not a priority in this optimistic setting ( $\|\bar{v}_k\| = 0$  since  $\|\bar{c}_k\| \leq \epsilon_o$ ), and we simply set  $\bar{\tau}_k = \bar{\tau}_{k-1}$  and consider the model reduction condition given in (2.12). Note that the relaxation parameters in (2.10)–(2.12) are only required to be proportional to  $\epsilon_o$ . We use  $\epsilon_o$  as the relaxation parameter as it simplifies the analysis and avoids the need for an additional parameter. The smaller the relaxation parameters, the more accurate  $\bar{u}_k$  is required to be. When  $\epsilon_o = 0$ , conditions (2.10)–(2.12) reduce to those in Termination Test 2. The  $\bar{\tau}_k$  update rules in both termination tests ensure that  $\{\bar{\tau}_k\}$  is monotonically non-increasing. It is important to note that (2.12) does not necessarily ensure that the reduction in the model of the merit function is non-negative, i.e.,  $\Delta\bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) = \Delta\bar{l}(x_k, \bar{\tau}_k, \bar{u}_k) \geq 0$ , which poses a significant challenge as this quantity is no longer a reliable measure of optimality. Fortunately, when  $\Delta\bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) < 0$ , one can show that both the stationarity and feasibility errors are sufficiently small and comparable in magnitude to the noise levels in the problem. In particular, Theorem 3.33 guarantees that if early termination is triggered due to  $\|\bar{c}_k\| \leq \epsilon_o$  and  $\Delta\bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) \leq \epsilon_o$  at some iteration, then both the stationarity and feasibility errors are sufficiently small. Note that the termination condition for  $\Delta\bar{l}(x_k, \bar{\tau}_k, \bar{d}_k)$  remains robust even if the constant on the right-hand side is replaced by different constant that is proportional to  $\epsilon_o$ . Another termination condition arises when  $\|\bar{c}_k\| > \epsilon_o$  and  $\|\bar{J}_k^T \bar{c}_k\| = 0$ , corresponding to iterates that are approximately infeasible stationary. This termination condition aligns with SQP methods developed under the rank-deficient constraint Jacobian setting (see, e.g., [19, 21]). Unlike prior works, due to the presence of noise in the constraints, we can only conclude that the terminated iterate is approximately an infeasible stationary point, with the associated error characterized in Theorem 3.34. On the other hand, if the algorithm does not terminate early, then an averaged measure of stationarity and feasibility error and/or infeasible stationarity error is guaranteed to be small (Theorems 3.36 and 3.38).

The final component of the algorithm is the mechanism for selecting the step sizes  $\{\bar{\alpha}_k\} \subset \mathbb{R}_{>0}$  used to update the iterates, i.e.,  $x_{k+1} \leftarrow x_k + \bar{\alpha}_k \bar{d}_k$ . Our algorithm, see Algorithm 1, allows for two step size schemes: an adaptive scheme inspired by [7] and a line search scheme inspired by [5, 36]. The goal of both step size procedures is to find a step size that ensures sufficient decrease as measured by the progress made on the merit function. A key difference between the two schemes is that the former is objective-function-free but requires knowledge (or estimates) of the Lipschitz constants of the objective and constraint gradients, while the latter requires function evaluations (that may be contaminated with noise) and knowledge (or estimate) of the noise level in the objective and constraint

functions. We discuss both approaches in detail below.

The adaptive step size scheme (**Option I**) is very similar to that proposed in [7]. For completeness and in order to introduce the notation used in the algorithm and analysis, we fully present the procedure. The key idea is to carefully select a step size based on the magnitude of the normal and tangential components, to safe-guard our algorithm from potential rank deficiency in the constraint Jacobians. The exact balance between these two components is problem-dependent and, to avoid the need for tuning, our algorithm generates two sequences  $\{\bar{\chi}_k\}$  and  $\{\bar{\zeta}_k\}$  to distinguish between tangentially and normally dominated steps. Specifically, at iteration  $k \in \mathbb{N}$ , after the two steps  $(\bar{v}_k, \bar{u}_k)$  are computed, the conditions

$$\|\bar{u}_k\|^2 \geq \bar{\chi}_{k-1} \|\bar{v}_k\|^2 \quad \text{and} \quad \frac{1}{2}(\bar{v}_k + \bar{u}_k)^T H_k(\bar{v}_k + \bar{u}_k) < \frac{1}{4}\bar{\zeta}_{k-1} \|\bar{u}_k\|^2, \quad (2.18)$$

are employed to update the sequences via

$$(\bar{\chi}_k, \bar{\zeta}_k) \leftarrow \begin{cases} ((1 + \sigma_\chi) \bar{\chi}_{k-1}, (1 - \sigma_\zeta) \bar{\zeta}_{k-1}) & \text{if (2.18) holds} \\ (\bar{\chi}_{k-1}, \bar{\zeta}_{k-1}) & \text{otherwise,} \end{cases} \quad (2.19)$$

where  $\sigma_\chi \in \mathbb{R}_{>0}$  and  $\sigma_\zeta \in (0, 1)$  are user-defined parameters. Under the conditions that  $\bar{\chi}_{-1} > 0$  and  $\bar{\zeta}_{-1} > 0$ , by (2.19), it follows that  $\{\bar{\chi}_k\}$  is monotonically non-decreasing and  $\{\bar{\zeta}_k\}$  is monotonically non-increasing. Moreover, our analysis will show that  $\{\bar{\chi}_k\}$  (resp.  $\{\bar{\zeta}_k\}$ ) is bounded above (resp. below) uniformly by a positive real number, i.e., these sequences are bounded deterministically and  $(\bar{\chi}_k, \bar{\zeta}_k) = (\bar{\chi}_{k-1}, \bar{\zeta}_{k-1})$  for all sufficiently large  $k \in \mathbb{N}$  in any run of the algorithm (see Lemma 3.10). With  $\bar{d}_k = \bar{u}_k + \bar{v}_k$  (note that  $\bar{d}_k \neq 0$  by Lemma 3.8), the algorithm then sets

$$\bar{\xi}_k := \begin{cases} \bar{\xi}_{k-1} & \text{if } \bar{\xi}_{k-1} \leq \bar{\xi}_k^{\text{trial}} \\ \min \{(1 - \sigma_\xi) \bar{\xi}_{k-1}, \bar{\xi}_k^{\text{trial}}\} & \text{otherwise,} \end{cases} \quad (2.20)$$

where

$$\bar{\xi}_k^{\text{trial}} := \begin{cases} \frac{\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k)}{\bar{\tau}_k \|\bar{d}_k\|^2} & \text{if } \|\bar{u}_k\|^2 \geq \bar{\chi}_k \|\bar{v}_k\|^2 \\ \frac{\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k)}{\|\bar{d}_k\|^2} & \text{otherwise,} \end{cases} \quad (2.21)$$

and  $\sigma_\xi \in (0, 1)$  is a user-defined parameter and  $\bar{\chi}_k$  is given in (2.19). By (2.20),  $\bar{\xi}_k \leq \bar{\xi}_k^{\text{trial}}$  for all  $k \in \mathbb{N}$ , and  $\{\bar{\xi}_k\}$  is monotonically non-increasing and bounded away from 0 (see Lemma 3.11). Then, the algorithm sets

$$\bar{\alpha}_k^{\text{suff}} := \min \left\{ \frac{2(1-\eta)\beta_k \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k)}{(\bar{\tau}_k L + \Gamma) \|\bar{d}_k\|^2}, 1 \right\}, \quad (2.22)$$

where  $\eta \in (0, 1)$  and  $\{\beta_k\}$  are user-defined parameters satisfying  $\frac{2(1-\eta)\beta_k \bar{\xi}_{-1} \max\{\bar{\tau}_{-1}, 1\}}{\bar{\tau}_{-1} L + \Gamma} \in (0, 1]$  for all  $k \in \mathbb{N}$  with  $(\bar{\tau}_{-1}, \bar{\xi}_{-1})$  representing the initial values of sequences  $\{\bar{\tau}_k\}$  and

$\{\bar{\xi}_k\}$ , respectively. We then employ a safeguarding scheme to ensure that the step size  $\bar{\alpha}_k$  is not too large or too small. To this end, since  $\bar{\alpha}_k^{\text{suff}}$  is based on stochastic quantities, we use the sequence  $\{\bar{\xi}_k\}$ , which is monotonically non-increasing and bounded from zero, as the proxy, and define

$$\bar{\alpha}_k^{\min} := \begin{cases} \frac{2(1-\eta)\beta_k\bar{\xi}_k\bar{\tau}_k}{\bar{\tau}_k L + \Gamma} & \text{if } \|\bar{u}_k\|^2 \geq \bar{\chi}_k \|\bar{v}_k\|^2 \\ \frac{2(1-\eta)\beta_k\bar{\xi}_k}{\bar{\tau}_k L + \Gamma} & \text{otherwise,} \end{cases} \quad (2.23)$$

and  $\bar{\alpha}_k^{\max} := \bar{\alpha}_k^{\min} + \theta\beta_k$ , where  $\theta \in \mathbb{R}_{>0}$  is an input parameter. The algorithm then sets the step size  $\bar{\alpha}_k$  by projecting  $\bar{\alpha}_k^{\text{suff}}$  onto the interval  $[\bar{\alpha}_k^{\min}, \bar{\alpha}_k^{\max}]$ , i.e.,  $\bar{\alpha}_k \leftarrow \text{Proj}_k(\bar{\alpha}_k^{\text{suff}})$ , where

$$\text{Proj}_k(\bar{\alpha}_k^{\text{suff}}) = \text{Proj}\left(\bar{\alpha}_k^{\text{suff}} \mid [\bar{\alpha}_k^{\min}, \bar{\alpha}_k^{\max}]\right) \quad (2.24)$$

and  $\text{Proj}(\bar{\alpha}_k^{\text{suff}} \mid \mathcal{I})$  denotes the projection of  $\bar{\alpha}_k^{\text{suff}}$  onto the interval  $\mathcal{I} \subset \mathbb{R}$ . Note that this projection step will never increase the step size (see Lemma 3.14). This is necessary especially when  $\bar{J}_k$  is rank-deficient and the merit parameter  $\bar{\tau}_k$  is close to 0, which makes  $\bar{\alpha}_k^{\min}$  close to 0 when  $\|\bar{u}_k\|$  dominates  $\|\bar{v}_k\|$ . In this case, there could be a large gap between  $\bar{\alpha}_k^{\text{suff}}$  and  $\bar{\alpha}_k^{\min}$ , and the projection scheme makes this gap controllable.

The other option for the step size selection is to use a line search mechanism **Option II**. Of course, noisy evaluations of the objective and constraint functions are required by this option (while **Option I** is objective- and constraint-function-free). This approach is inspired by the relaxed Armijo condition proposed in [5] for the unconstrained setting with bounded noise and classical line search SQP methods [36]. By Assumption 2.2 and  $\phi(x_k, \bar{\tau}_k) := \bar{\tau}_k f(x_k) + \|c(x_k)\|$ , it follows that

$$|\bar{\phi}(x_k, \bar{\tau}_k) - \phi(x_k, \bar{\tau}_k)| \leq \bar{\tau}_k |\bar{f}(x_k) - f(x_k)| + \|\bar{c}(x_k) - c(x_k)\| \leq \bar{\tau}_k \epsilon_f + \epsilon_c. \quad (2.25)$$

We consider a modified version of the sufficient decrease Armijo condition, applied to the merit function (2.7),

$$\bar{\phi}(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_k) \leq \bar{\phi}(x_k, \bar{\tau}_k) - \eta \bar{\alpha}_k \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) + \epsilon_{A_k}, \quad (2.26)$$

where  $\eta \in (0, 1)$  and  $\epsilon_{A_k} \in \mathbb{R}_{\geq 0}$  is a user-defined parameter that depends on  $(\epsilon_f, \epsilon_c, \epsilon_g, \epsilon_J)$  and accounts for the noise. The step size is selected via a standard backtracking mechanism, where the initial trial step size is a user-defined constant  $\alpha_u^L \in (0, 1]$ . If a given trial step size does not satisfy (2.26), then the trial step size is reduced by a constant factor. This process is repeated until a step size that satisfies (2.26) is found (see Lemma 3.15). The relaxation term  $\epsilon_{A_k}$  is crucial for both theory (by guaranteeing a well-defined line search procedure) and practice (by potentially accepting longer step sizes). The explicit expression of the relaxation term  $\epsilon_{A_k}$  is given in Lemma 3.15.

The complete algorithm (with both step size options) is given in Algorithm 1.

---

**Algorithm 1** Optimistic Noise-Aware SQP Algorithm
 

---

**Require:**  $x_0 \in \mathbb{R}^n$  (initial iterate);  $\bar{\tau}_{-1} \in \mathbb{R}_{>0}$  (initial merit parameter value);  $\epsilon_o \in [0, \epsilon_c]$  (optimistic feasibility parameter);

**Require (Option I: Adaptive):**  $\{\beta_k\} \subset (0, 1]$  (adaptive step size parameter),  $L \in \mathbb{R}_{>0}$ ,  $\Gamma \in \mathbb{R}_{>0}$  (Lipschitz constant estimates),  $\{\bar{\chi}_{-1}, \bar{\xi}_{-1}, \bar{\zeta}_{-1}\} \subset \mathbb{R}_{>0}$  (initial values for additional sequences)

**Require (Option II: Line Search):**  $\alpha_u^L \in (0, 1]$  (initial step size),  $\nu \in (0, 1)$  (search control parameters),  $\epsilon_{A_k}$  (line search relaxation parameter)

- 1: **for all**  $k \in \mathbb{N}$  **do**
  - 2:   Compute gradient  $\bar{g}_k$  and Jacobian  $\bar{J}_k$  estimates satisfying Assumption 2.2
  - 3:   **if**  $\|\bar{c}_k\| \leq \epsilon_o$  **then**
  - 4:     Set  $\bar{v}_k = 0$  and  $\bar{\tau}_k = \bar{\tau}_{k-1}$
  - 5:     Compute  $(\bar{u}_k, \bar{\rho}_k, \bar{r}_k)$  via (2.6) satisfying Termination Test 1 and set  $\bar{d}_k = \bar{u}_k$
  - 6:     **if**  $\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) \leq \epsilon_o$  **then**
  - 7:       **Terminate and return**  $x_k$  (*approximate first-order stationary point*)
  - 8:     **end if**
  - 9:   **else**
  - 10:    **if**  $\|\bar{J}_k^T \bar{c}_k\| = 0$  **then**
  - 11:     **Terminate and return**  $x_k$  (*approximate infeasible stationary point*)
  - 12:    **end if**
  - 13:    Compute  $\bar{v}_k$  via (2.2) satisfying (2.3)
  - 14:    Compute  $(\bar{u}_k, \bar{\rho}_k, \bar{r}_k)$  via (2.6) satisfying Termination Test 2 and set  $\bar{d}_k = \bar{v}_k + \bar{u}_k$
  - 15:    Update  $\bar{\tau}_k$  by (2.16) and (2.17)
  - 16:    **end if**
  - 17:    Choose step size  $\bar{\alpha}_k$  following one of the two options:
  - 18:     **Option I:** Update  $\bar{\alpha}_k$  via (2.18)–(2.24).
  - 19:     **Option II:** Set  $\bar{\alpha}_k = \alpha_u^L$  and update  $\bar{\alpha}_k = \nu \bar{\alpha}_k$  until (2.26) is satisfied
  - 20:    Set  $x_{k+1} \leftarrow x_k + \bar{\alpha}_k \bar{d}_k$
  - 21: **end for**
-

**Remark 2.6.** *We make a few remarks about Algorithm 1.*

- **Step size schemes:** *Our proposed algorithm has two options for selecting the step size. Algorithm 1 with **Option I** (adaptive) is similar to [7] and Algorithm 1 with **Option II** (line search) is similar to [38]. The main difference between these two schemes is that the adaptive scheme (**Option I**) requires estimates of the Lipschitz constants while the line search scheme (**Option II**) requires objective function and constraint evaluations and knowledge of  $(\epsilon_f, \epsilon_g, \epsilon_c, \epsilon_J)$  (or at least estimates of these quantities).*
- **Comparison to [7, 21]:** *The algorithms proposed in [7, 21] are developed for solving optimization problems with stochastic objective functions and deterministic equality constraints. We consider general bounded noise in both the objective and constraint functions. Moreover, another differentiating factor is our optimistic early termination scheme. All these algorithms employ a step decomposition strategy to compute the step, but the normal and tangential component computations are not exactly the same. Due to potential rank-deficiency in the constraint Jacobians, similar to [7] an inexact normal step is computed by solving a trust-region subproblem whose goal is to minimize linearized infeasibility. Then, similar to [21] an inexact tangential step is computed with appropriately modified termination conditions to account for optimistic feasibility.*
- **Comparison to [38]:** *The algorithm and analysis in [38] is restricted to the setting in which the LICQ holds. We consider the setting in which the constraint Jacobians may be rank-deficient. Other differences include: (i) the Hessian approximation in [38] is assumed to be a scalar multiple of the identity matrix, while we allow for more general Hessian approximations; (ii) we propose two step size selection schemes whereas [38] only proposes a line search strategy; (iii) exact solutions to subproblems are considered in [38] whereas we allow for inexact solutions; and, (iv) we adopt an optimistic approach to feasibility and develop early stopping mechanisms.*
- **Comparison to [45]:** *While our proposed algorithm addresses the same class of problems as [45], it differs in the following aspects: (i) [45] employs a trust-region approach whereas we use alternative step-size schemes; (ii) our approach allows for inexact solutions and includes an early stopping mechanism, whereas such features are not present in [45]; and, (iii) in the analysis presented in [45] a key assumption regarding the perturbation error of the orthogonal basis of the Jacobian matrix is made [45, Assumption 5], whereas our analysis does not impose such an assumption.*

### 3 Convergence Analysis

In this section, we provide convergence guarantees for Algorithm 1. Specifically, if the LICQ is satisfied, the algorithm generates a sequence of iterates whose averaged first-order

stationary and feasibility measures converge to a sufficiently small value dictated by the noise. When the LICQ is not satisfied, there exists a subsequence of the iterates generated by the algorithm whose infeasible stationary measure converges to a sufficiently small value dictated by the noise. Furthermore, we show that the averaged stationary error is sufficiently small if the merit parameter is bounded away from zero. The convergence results are consistent with the results that can be established for the deterministic SQP method [19], and when the noise levels are all zero we recover such results. To prove our main results (Section 3.4), we first establish the well-posedness of Algorithm 1 (Section 3.2). Next, we conduct perturbation (error) analysis for subproblems (2.2) and (2.6), and demonstrate that the solutions to these subproblems remain stable under noise perturbations (Section 3.3), and also show the deviation of the lower bound of merit parameter values is proportional to the noise. We begin with three assumptions in Section 3.1.

### 3.1 Assumptions

In this subsection, we present three key assumptions made throughout Section 3. Assumption 3.1 is related to the iterative solver used to compute an inexact tangential step, Assumption 3.2 is related to the iterates generated by Algorithm 1, and Assumption 3.4 is related to the early termination feature of Algorithm 1.

**Assumption 3.1.** For all  $k \in \mathbb{N}$ , the iterative linear system solver employed in (2.6) generates a sequence of inexact solutions  $\{(\bar{u}_{k,t}, \bar{y}_{k,t}, \bar{\rho}_{k,t}, \bar{r}_{k,t})\}_{t \in \mathbb{N}}$  satisfying

$$\begin{bmatrix} \bar{\rho}_{k,t} \\ \bar{r}_{k,t} \end{bmatrix} = \begin{bmatrix} H_k & \bar{J}_k^T \\ \bar{J}_k & 0 \end{bmatrix} \begin{bmatrix} \bar{u}_{k,t} \\ \bar{y}_{k,t} \end{bmatrix} + \begin{bmatrix} \bar{g}_k + H_k \bar{v}_k \\ 0 \end{bmatrix} \text{ for all } t \in \mathbb{N}$$

such that  $\lim_{t \rightarrow \infty} \|(\bar{u}_{k,t}, \bar{\rho}_{k,t}, \bar{r}_{k,t}) - (\bar{u}_k^*, 0, 0)\| = 0$ , where  $\bar{u}_k^*$  is the solution of (2.5).

**Assumption 3.2.** For all  $k \in \mathbb{N}$ ,  $\bar{J}_k^T \bar{c}_k \neq 0$  or  $\bar{g}_k \notin \text{Range}(\bar{J}_k^T)$ .

**Remark 3.3.** Assumption 3.1 is satisfied if one employs a linear system solver, such as MINRES [15], given that  $\bar{u}_k^*$  is unique (even though  $\bar{J}_k$  could be rank-deficient). Assumption 3.2 is necessary for the well-posedness of the proposed termination tests (see Lemma 3.7). Assumption 3.2 is violated only when  $\bar{J}_k^T \bar{c}_k = 0$  and  $\bar{g}_k$  lies precisely in  $\text{Range}(\bar{J}_k^T)$ . When the algorithm reaches an iterate for which  $\bar{J}_k^T \bar{c}_k = 0$ , it terminates if  $\|\bar{c}_k\| > \epsilon_o$  with an approximate infeasible stationary point.

**Assumption 3.4.** Algorithm 1 does not terminate finitely, i.e., Algorithm 1 does not terminate on Lines 7 or 11 for any  $k \in \mathbb{N}$ .

**Remark 3.5.** Theorems 3.36, 3.38 and 3.40 all rely on Assumption 3.4. However, in general, Assumption 3.4 is not guaranteed to be satisfied, i.e., the algorithm may terminate finitely. To handle the case where Assumption 3.4 is violated, we will show in Theorem 3.33 (resp., Theorem 3.34) that if Algorithm 1 terminates on Line 7 (resp., Line 11), a near stationary solution (resp., a near infeasible stationary solution) to (1.1) is attained.

### 3.2 Well-posedness of Algorithm

In this section, we show that Algorithm 1 is well-posed, i.e., every iteration of the algorithm is well-defined and is guaranteed to terminate finitely. We first show that the termination tests for tangential components  $\{\bar{u}_k\}$  are well-posed and that at least one of them is satisfied in finite time. Then, we show that the sequences  $\{\bar{\chi}_k\}$ ,  $\{\bar{\zeta}_k\}$ , and  $\{\bar{\xi}_k\}$  defined in the adaptive step size scheme (**Option I**, (2.18)–(2.21)) are bounded to ensure that the scheme is well-posed. Moreover, we show the boundedness of the merit parameter sequence  $\{\bar{\tau}_k\}$  in the presence of the LICQ. Finally, we show that both the adaptive step size (**Option I**) and the line search (**Option II**) schemes are well-defined in that the step sizes are guaranteed to be upper bounded on all iterates and lower bounded under the condition that the merit parameter is lower bounded or that both the infeasibility stationarity and the feasibility errors are sufficiently large.

We present a lemma that pertains to the normal component and that is common in the constrained optimization literature; see e.g., [19, Lemmas 3.5 & 3.6] and [7, Lemma 1].

**Lemma 3.6.** *Suppose that Assumptions 2.1 and 3.4 hold. If  $\|\bar{c}_k\| > \epsilon_o$ , there exist  $\sigma_v \in \mathbb{R}_{>0}$  and  $\underline{\sigma}_{Jc} \in \mathbb{R}_{>0}$  such that*

$$\|\bar{c}_k\| (\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|) \geq \sigma_v \|\bar{J}_k^T \bar{c}_k\|^2, \quad \text{and} \quad \underline{\sigma}_{Jc} \|\bar{J}_k^T \bar{c}_k\|^2 \leq \|\bar{v}_k\| \leq \sigma_{Jc} \|\bar{J}_k^T \bar{c}_k\|.$$

*Proof.* The proof follows the same logic [19, Lemmas 3.5 & 3.6] by replacing  $(c_k, J_k, v_k)$  with their noisy evaluations  $(\bar{c}_k, \bar{J}_k, \bar{v}_k)$ . The result only holds when  $\|\bar{c}_k\| > \epsilon_o$  since otherwise we do not solve subproblem (2.2) and directly set  $\bar{v}_k = 0$ .  $\square$

The following lemma guarantees the well-posedness of Termination Tests 1 and 2 when computing the tangential step, i.e., a solver will compute  $\bar{u}_k$  which satisfies our proposed termination tests within a finite number of iterations. Note that the MINRES method [15] guarantees an exact solution to the linear system within a uniformly bounded number of iterations.

**Lemma 3.7.** *Suppose that Assumptions 2.2, 2.4, 3.1, 3.2, and 3.4 hold. For all  $k \in \mathbb{N}$ , the iterative linear system solver computes  $(\bar{u}_k, \bar{\rho}_k, \bar{r}_k) \leftarrow (\bar{u}_{k,t}, \bar{\rho}_{k,t}, \bar{r}_{k,t})$  satisfying either Termination Test 1 or Termination Test 2 in a finite number of iterations  $t \in \mathbb{N}$ .*

*Proof.* By (2.2) it follows that  $\|\bar{v}_k\| \leq \sigma_{Jc} \|\bar{J}_k^T \bar{c}_k\|$  and by (2.5) it follows that  $H_k \bar{u}_k^* + \bar{J}_k^T \bar{y}_k^* + \bar{g}_k + H_k \bar{v}_k = 0$ . If  $\|\bar{u}_k^*\| = \|\bar{J}_k^T \bar{c}_k\| = 0$ , then  $\bar{v}_k = \bar{J}_k^T \bar{c}_k = 0$  and  $\bar{J}_k^T \bar{y}_k^* + \bar{g}_k = 0$ , which violates Assumption 3.2. Hence,  $\max\{\|\bar{u}_k^*\|, \|\bar{J}_k^T \bar{c}_k\|\} > 0$ . By Assumption 3.1 it follows that (2.9) is satisfied by  $(\bar{u}_k, \bar{\rho}_k, \bar{r}_k) = (\bar{u}_{k,t}, \bar{\rho}_{k,t}, \bar{r}_{k,t})$  for sufficiently large  $t \in \mathbb{N}$ . We will first show that if  $\|\bar{c}_k\| > \epsilon_o$  Termination Test 2 ((2.13) and (2.14) or (2.15)) will eventually be satisfied, and then if  $\|\bar{c}_k\| \leq \epsilon_o$  Termination Test 1 ((2.9)–(2.12)) will eventually be satisfied.

When  $\|\bar{c}_k\| > \epsilon_o$ , if  $\|\bar{J}_k^T \bar{c}_k\| = 0$ , Algorithm 1 terminates on Line 11. Hence, we consider  $\|\bar{J}_k^T \bar{c}_k\| \neq 0$  ( $\|\bar{v}_k\| > 0$  by (2.3)) when computing  $\bar{u}_k$ . To satisfy condition (2.13),



we consider two cases: (i) if  $\|\bar{u}_k^*\| = 0$ , then Assumption 3.1 implies that  $\{\|\bar{u}_{k,t}\|\}_{t \in \mathbb{N}} \rightarrow \|\bar{u}_k^*\| = 0$  so that the former condition in (2.13) holds with  $(\bar{u}_k, \bar{\rho}_k, \bar{r}_k) \leftarrow (\bar{u}_{k,t}, \bar{\rho}_{k,t}, \bar{r}_{k,t})$  for all sufficiently large  $t \in \mathbb{N}$ ; (ii) if  $\|\bar{u}_k^*\| > 0$ , it follows by Assumption 2.4 and (2.5) that

$$\begin{aligned} \bar{u}_k^{*T} (\bar{g}_k + H_k \bar{v}_k) + \max \left\{ \frac{1}{2}, 1 - \|\bar{J}_k^T \bar{c}_k\| \right\} \bar{u}_k^{*T} H_k \bar{u}_k^* &< \bar{u}_k^{*T} (\bar{g}_k + H_k \bar{v}_k + H_k \bar{u}_k^*) \\ &= -\bar{u}_k^{*T} \bar{J}_k^T \bar{y}_k^* = 0. \end{aligned}$$

Since  $\|\bar{v}_k\| > 0$ , the latter condition in (2.13) holds by Assumption 3.1  $((\bar{u}_k, \bar{\rho}_k, \bar{r}_k) \leftarrow (\bar{u}_{k,t}, \bar{\rho}_{k,t}, \bar{r}_{k,t})$  for all sufficiently large  $t \in \mathbb{N}$ ). Moreover, by  $\|\bar{v}_k\| > 0$  and (2.3) it follows that  $\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\| > 0$ , and by Assumption 3.1 and  $\sigma_r \in (\sigma_c, 1) \subset (0, 1)$ ,

$$\lim_{t \rightarrow \infty} (\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k + \bar{r}_{k,t}\|) = \|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\| > \sigma_r (\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|).$$

Condition (2.15) holds with  $(\bar{u}_k, \bar{\rho}_k, \bar{r}_k) \leftarrow (\bar{u}_{k,t}, \bar{\rho}_{k,t}, \bar{r}_{k,t})$  for all sufficiently large  $t \in \mathbb{N}$ . Thus, Termination Test 2 is guaranteed to be satisfied in a finite number of iterations.

When  $\|\bar{c}_k\| \leq \epsilon_o$ , then  $\|\bar{v}_k\| = 0$  (Line 4, Algorithm 1). We consider two cases,  $\epsilon_o > 0$  and  $\epsilon_o = 0$ . When  $\epsilon_o > 0$ , by Assumption 2.4 and (2.5), it follows that  $\bar{u}_k^{*T} H_k \bar{u}_k^* \geq \zeta \|\bar{u}_k^*\|^2 \geq \lambda_u \|\bar{u}_k^*\|^2 > \lambda_u \|\bar{u}_k^*\|^2 - \epsilon_o$  and  $\bar{g}_k^T \bar{u}_k^* + \frac{1}{2} \bar{u}_k^{*T} H_k \bar{u}_k^* = \bar{u}_k^{*T} (\bar{g}_k + H_k \bar{v}_k) + \frac{1}{2} \bar{u}_k^{*T} H_k \bar{u}_k^* = -\frac{1}{2} \bar{u}_k^{*T} H_k \bar{u}_k^* < \epsilon_o$ . By Assumption 2.4, (2.5), (2.8),  $\sigma_u \in (0, 1)$  and  $\bar{d}_k^* = \bar{u}_k^* + \bar{v}_k = \bar{u}_k^*$ ,

$$\begin{aligned} \Delta \bar{l}(x_k, \bar{r}_{k-1}, \bar{d}_k^*) &= \Delta \bar{l}(x_k, \bar{r}_{k-1}, \bar{u}_k^*) = -\bar{r}_{k-1} \bar{g}_k^T \bar{u}_k^* + \|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{u}_k^*\| \\ &= -\bar{r}_{k-1} \left( -H_k \bar{u}_k^* - \bar{J}_k^T \bar{y}_k^* \right)^T \bar{u}_k^* \\ &= \bar{r}_{k-1} \bar{u}_k^{*T} H_k \bar{u}_k^* \\ &> \bar{r}_{k-1} \sigma_u \max \left\{ \bar{u}_k^{*T} H_k \bar{u}_k^*, \lambda_u \|\bar{u}_k^*\|^2 \right\} - \epsilon_o. \end{aligned}$$

By Assumption 3.1, for sufficiently large  $t \in \mathbb{N}$ , the inexact solution  $\bar{u}_k \leftarrow \bar{u}_{k,t}$  satisfies (2.10)–(2.12) when  $\epsilon_o > 0$ , and Termination Test 1 is guaranteed to be satisfied in finite iterations.

On the other hand, when  $\|\bar{c}_k\| \leq \epsilon_o = 0$  by Assumption 3.2, it follows that  $\|\bar{c}_k\| = \|\bar{v}_k\| = 0$  and  $\|\bar{u}_k^*\| > 0$ . Therefore, by Assumption 2.4, (2.5) and  $\lambda_u \in (0, \zeta)$ , it follows that  $\bar{u}_k^{*T} H_k \bar{u}_k^* \geq \zeta \|\bar{u}_k^*\|^2 > \lambda_u \|\bar{u}_k^*\|^2$  and  $\bar{g}_k^T \bar{u}_k^* + \frac{1}{2} \bar{u}_k^{*T} H_k \bar{u}_k^* = \bar{u}_k^{*T} (\bar{g}_k + H_k \bar{v}_k) + \frac{1}{2} \bar{u}_k^{*T} H_k \bar{u}_k^* = -\frac{1}{2} \bar{u}_k^{*T} H_k \bar{u}_k^* < 0 = \epsilon_o$ . Additionally, by (2.8),  $\sigma_u \in (0, 1)$  and  $\bar{d}_k^* = \bar{u}_k^* + \bar{v}_k = \bar{u}_k^*$ , using similar logic as the former case, it follows that  $\Delta \bar{l}(x_k, \bar{r}_{k-1}, \bar{d}_k^*) > \bar{r}_{k-1} \sigma_u \max \left\{ \bar{u}_k^{*T} H_k \bar{u}_k^*, \lambda_u \|\bar{u}_k^*\|^2 \right\}$ , since  $\bar{u}_k^* \neq 0$ . Thus, by Assumption 3.1, for sufficiently large  $t \in \mathbb{N}$ , the inexact solution  $\bar{u}_k \leftarrow \bar{u}_{k,t}$  satisfies (2.10)–(2.12) when  $\epsilon_o = 0$ , and Termination Test 1 is guaranteed to be satisfied in finite iterations.  $\square$

Next, we prove that the search direction  $\bar{d}_k$  is non-zero at every iteration if the algorithm does not terminate early. This ensures that the step size schemes are well-posed.

**Lemma 3.8.** *Suppose that Assumptions 2.2, 2.4, 3.1, 3.2, and 3.4 hold. For all  $k \in \mathbb{N}$ ,  $\|\bar{d}_k\| > 0$ .*

*Proof.* We consider two cases  $\|\bar{c}_k\| \leq \epsilon_o$  and  $\|\bar{c}_k\| > \epsilon_o$  and prove the result in both cases using a contradiction argument.

Suppose that  $\|\bar{d}_k\| = 0$ . If  $\|\bar{c}_k\| \leq \epsilon_o$  and  $\epsilon_o = 0$ , then it follows that  $\|\bar{c}_k\| = \|\bar{v}_k\| = 0$ , which implies that  $\|\bar{u}_k\| = 0$ . By (2.9),  $\|\bar{\rho}_k\| = \|\bar{r}_k\| = 0$  and consequently  $\bar{u}_k^* = \bar{u}_k = 0$ , which violates Assumption 3.2. If  $\|\bar{c}_k\| \leq \epsilon_o$  and  $\epsilon_o > 0$ , then it follows by (2.8) that  $\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) = 0 < \epsilon_o$ . Algorithm 1 then terminates at line 7, which causes contradiction.

Suppose that  $\|\bar{d}_k\| = 0$ . If  $\|\bar{c}_k\| > \epsilon_o$ , by Assumption 3.4 the algorithm does not terminate on Line 11, and thus  $\|\bar{c}_k\| > \epsilon_o$ . This implies that  $\|\bar{v}_k\| > 0$  by (2.3), and so the right-hand-side of (2.14) is positive. Since  $\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) = 0$  by (2.8), (2.14) cannot be satisfied. On the other hand,  $\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k + \bar{r}_k\| = \|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{d}_k\| = 0$  and (2.15) cannot be satisfied. Thus, Termination Test 2 cannot be satisfied and we have reached a contradiction to Lemma 3.7.

In both cases,  $\|\bar{d}_k\| = 0$  leads to a contradiction, and thus the desired result follows.  $\square$

The following lemma characterizes the relationship between the reduction in the model of the merit function and the search direction under various conditions when early termination is not triggered, i.e., Algorithm 1 does not terminate on Lines 7 or 11. Similar results have been established in [8, Lemma 3.4] and [21, Lemma 8]. It is worth noting that in our setting, a small value of  $\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k)$  does not necessarily imply that the norm of the search direction is also small, which is different from [8, 21]. This distinction arises due to the potential rank deficiency in the noisy constraint Jacobians. We distinguish between two cases because  $\bar{v}_k = 0$  is set when  $\|\bar{c}_k\| \leq \epsilon_o$ , resulting in no reduction in constraint violation in this scenario. Moreover, if the singular values of  $\{\bar{J}_k\}$  are bounded below and  $\|\bar{c}_k\| > \epsilon_o$ , an alternative result can be derived by replacing the infeasible stationarity measure with the feasibility measure.

**Lemma 3.9.** *Suppose that Assumptions 2.1, 2.2, 2.4, 3.1, 3.2, and 3.4 hold. If  $\|\bar{c}_k\| \leq \epsilon_o$ ,*

$$\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) \geq \frac{\bar{\tau}_k \sigma_u \lambda_u}{2} \|\bar{u}_k\|^2 = \frac{\bar{\tau}_k \sigma_u \lambda_u}{2} \|\bar{d}_k\|^2. \quad (3.1)$$

*If  $\|\bar{c}_k\| > \epsilon_o$ , it follows that*

$$\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) \geq \bar{\tau}_k \sigma_u \lambda_u \|\bar{u}_k\|^2 + \frac{\sigma_c \sigma_v}{\kappa_c + \epsilon_c} \|\bar{J}_k^T \bar{c}_k\|^2 \geq \frac{\min\{\bar{\tau}_k \sigma_u \lambda_u, \sigma_c \sigma_v / (\kappa_c + \epsilon_c)\}}{\max\{2, 2\sigma_{J_c}^2\}} \|\bar{d}_k\|^2. \quad (3.2)$$

*Moreover, if  $\|\bar{c}_k\| > \epsilon_o$  and the singular values of  $\bar{J}_k$  are bounded below by  $\kappa_\sigma$  for all  $k \in \mathbb{N}$ ,*

$$\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) \geq \bar{\tau}_k \sigma_u \lambda_u \|\bar{u}_k\|^2 + \kappa_\sigma^2 \sigma_c \sigma_v \|\bar{c}_k\| \geq \frac{\min\{\bar{\tau}_k \sigma_u \lambda_u, \kappa_\sigma^2 \sigma_c \sigma_v\}}{\max\{2, 2\sigma_{J_c}^2 (\kappa_c + \epsilon_c) (\kappa_J + \epsilon_J)^2\}} \|\bar{d}_k\|^2. \quad (3.3)$$

*Proof.* If  $\|\bar{c}_k\| \leq \epsilon_o$ , then Termination Test 1 is satisfied, and

$$\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) \geq \bar{\tau}_k \sigma_u \max\left\{\bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2\right\} - \epsilon_o \quad (3.4)$$

by (2.12) and  $(\bar{d}_k, \bar{\tau}_k) = (\bar{u}_k, \bar{\tau}_{k-1})$ . Since  $\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) > \epsilon_o$  by Assumption 3.4, it follows that

$$\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) \geq \bar{\tau}_k \sigma_u \lambda_u \|\bar{u}_k\|^2 - \epsilon_o > \bar{\tau}_k \sigma_u \lambda_u \|\bar{u}_k\|^2 - \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k).$$

Re-arranging the above inequality and using the fact that  $\bar{v}_k = 0$  we conclude (3.1).

If  $\|\bar{c}_k\| > \epsilon_o$ , when (2.14) holds, we directly have  $\bar{\tau}_k = \bar{\tau}_{k-1}$  and

$$\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) \geq \bar{\tau}_k \sigma_u \max \left\{ \bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2 \right\} + \sigma_c (\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|).$$

Otherwise, when (2.14) does not hold, by Termination Test 2, (2.15) holds. By (2.16) and (2.17), it follows that  $\bar{\tau}_k \leq (1 - \sigma_\tau) \bar{\tau}_k^{\text{trial}}$  and

$$\bar{\tau}_k (\bar{g}_k^T \bar{d}_k + \max \{ \bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2 \}) \leq (1 - \sigma_\tau) (1 - \frac{\sigma_c}{\sigma_r}) (\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k + \bar{r}_k\|). \quad (3.5)$$

By (2.6), (2.8), (2.15) and (3.5), it follows that

$$\begin{aligned} \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) &= -\bar{\tau}_k \bar{g}_k^T \bar{d}_k + \|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{d}_k\| \\ &= -\bar{\tau}_k \bar{g}_k^T \bar{d}_k + \|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k + \bar{r}_k\| \\ &\geq \bar{\tau}_k \max \left\{ \bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2 \right\} + \frac{\sigma_c}{\sigma_r} (\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k + \bar{r}_k\|) \\ &\geq \bar{\tau}_k \max \left\{ \bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2 \right\} + \sigma_c (\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|). \end{aligned} \quad (3.6)$$

Combining with (2.14), we further know that when  $\|\bar{c}_k\| > \epsilon_o$ ,

$$\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) \geq \bar{\tau}_k \sigma_u \max \left\{ \bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2 \right\} + \sigma_c (\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|). \quad (3.7)$$

By the step update rule, Lemma 3.6, Assumptions 2.1 and 2.2, it follows that

$$\|\bar{d}_k\|^2 \leq 2 \left( \|\bar{u}_k\|^2 + \|\bar{v}_k\|^2 \right) \leq 2 \left( \|\bar{u}_k\|^2 + \sigma_{J_c}^2 \|\bar{J}_k^T \bar{c}_k\|^2 \right). \quad (3.8)$$

Furthermore, by (3.7), (3.8) and Lemma 3.6,

$$\begin{aligned} \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) &\geq \bar{\tau}_k \sigma_u \lambda_u \|\bar{u}_k\|^2 + \sigma_c \sigma_v \frac{\|\bar{J}_k^T \bar{c}_k\|^2}{\|\bar{c}_k\|} \\ &\geq \bar{\tau}_k \sigma_u \lambda_u \|\bar{u}_k\|^2 + \frac{\sigma_c \sigma_v}{\kappa_c + \epsilon_c} \|\bar{J}_k^T \bar{c}_k\|^2, \end{aligned}$$

we conclude (3.2).

If the singular values of  $\{\bar{J}_k\}$  are bounded below by  $\kappa_\sigma > 0$  for all  $k \in \mathbb{N}$ , it follows from the singular value decomposition of  $\bar{J}_k^T$  that

$$\|\bar{J}_k^T \bar{c}_k\| \geq \kappa_\sigma \|\bar{c}_k\|, \quad (3.9)$$

and by Lemma 3.6, Assumptions 2.1 and 2.2

$$\|\bar{d}_k\|^2 \leq 2 \left( \|\bar{u}_k\|^2 + \sigma_{J_c}^2 \|\bar{J}_k^T \bar{c}_k\|^2 \right) \leq 2 \left( \|\bar{u}_k\|^2 + \sigma_{J_c}^2 (\kappa_c + \epsilon_c) (\kappa_J + \epsilon_J)^2 \|\bar{c}_k\| \right). \quad (3.10)$$

If  $\|\bar{c}_k\| > \epsilon_o$ , by (3.7), (3.9), (3.10) and Lemma 3.6, it follows that

$$\begin{aligned} \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) &\geq \bar{\tau}_k \sigma_u \max \left\{ \bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2 \right\} + \sigma_c (\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|) \\ &\geq \bar{\tau}_k \sigma_u \lambda_u \|\bar{u}_k\|^2 + \frac{\sigma_c \sigma_v \|\bar{J}_k^T \bar{c}_k\|^2}{\|\bar{c}_k\|} \geq \bar{\tau}_k \sigma_u \lambda_u \|\bar{u}_k\|^2 + \kappa_\sigma^2 \sigma_c \sigma_v \|\bar{c}_k\|, \end{aligned}$$

from which we conclude (3.3).  $\square$

Next, we introduce two lemmas related to the properties of the sequences  $\{\bar{\chi}_k, \bar{\zeta}_k\}$  (Lemma 3.10) and  $\{\bar{\xi}_k\}$  (Lemma 3.11).

**Lemma 3.10.** *Suppose that Assumptions 2.4 and 3.4 hold. There exists  $(\chi_{\max}, \zeta_{\min}) \in \mathbb{R}_{>0} \times \mathbb{R}_{>0}$  such that for some  $K_{\chi, \zeta} \in \mathbb{N}$ ,  $(\bar{\chi}_k, \bar{\zeta}_k) = (\bar{\chi}_{K_{\chi, \zeta}}, \bar{\zeta}_{K_{\chi, \zeta}})$  for all  $k \in \mathbb{N}$  with  $k \geq K_{\chi, \zeta}$ , where  $(\bar{\chi}_{K_{\chi, \zeta}}, \bar{\zeta}_{K_{\chi, \zeta}}) \in (0, \chi_{\max}] \times [\zeta_{\min}, \infty)$ .*

*Proof.* The proof proceeds in the same manner as [7, Lemma 2].  $\square$

**Lemma 3.11.** *Suppose Assumptions 2.1, 2.2, 2.4, 3.1, 3.2, and 3.4 hold. The sequence  $\{\bar{\xi}_k\}$  is non-increasing. In addition, there exists  $\bar{\xi}_{\min} \in \mathbb{R}_{>0}$  such that  $\bar{\xi}_k \geq \bar{\xi}_{\min}$  for all  $k \in \mathbb{N}$ .*

*Proof.* By (2.20) either  $\bar{\xi}_k = \bar{\xi}_{k-1}$  or  $\bar{\xi}_k = \min\{(1 - \sigma_\xi) \bar{\xi}_{k-1}, \bar{\xi}_k^{\text{trial}}\} < \bar{\xi}_{k-1}$ . Therefore,  $\{\bar{\xi}_k\}$  is non-increasing.

If  $\|\bar{c}_k\| > \epsilon_o$ , by (3.2),

$$\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) \geq \max \left\{ \bar{\tau}_k \sigma_u \lambda_u \|\bar{u}_k\|^2, \sigma_c \sigma_v (\kappa_c + \epsilon_c)^{-1} \|\bar{J}_k^T \bar{c}_k\|^2 \right\}. \quad (3.11)$$

We consider two cases. When  $\|\bar{u}_k\|^2 \geq \bar{\chi}_k \|\bar{v}_k\|^2$ , by (2.20) and (2.21), either  $\bar{\xi}_k = \bar{\xi}_{k-1}$  or

$$\bar{\xi}_k \geq (1 - \sigma_\xi) \frac{\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k)}{\bar{\tau}_k \|\bar{d}_k\|^2} \geq (1 - \sigma_\xi) \frac{\bar{\tau}_k \sigma_u \lambda_u \|\bar{u}_k\|^2}{2 \bar{\tau}_k (1 + \bar{\chi}_k^{-1}) \|\bar{u}_k\|^2} = \frac{(1 - \sigma_\xi) \sigma_u \lambda_u}{2(1 + \bar{\chi}_k^{-1})} > 0.$$

On the other hand, when  $\|\bar{u}_k\|^2 < \bar{\chi}_k \|\bar{v}_k\|^2$ , by (2.2), (2.20), (2.21), (3.2), and Lemma 3.10 it follows that either  $\bar{\xi}_k = \bar{\xi}_{k-1}$  or

$$\bar{\xi}_k \geq (1 - \sigma_\xi) \frac{\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k)}{\|\bar{d}_k\|^2} \geq (1 - \sigma_\xi) \frac{\sigma_c \sigma_v \|\bar{J}_k^T \bar{c}_k\|^2}{2(\bar{\chi}_k + 1)(\kappa_c + \epsilon_c) \sigma_{J_c}^2 \|\bar{J}_k^T \bar{c}_k\|^2} \geq \frac{(1 - \sigma_\xi) \sigma_c \sigma_v}{2(\chi_{\max} + 1)(\kappa_c + \epsilon_c) \sigma_{J_c}^2} > 0. \quad (3.12)$$

If  $\|\bar{c}_k\| \leq \epsilon_o$ , it follows that  $\|\bar{u}_k\|^2 \geq \bar{\chi}_k \|\bar{v}_k\|^2$ , and by (2.20), (2.21), and (3.1)

$$\bar{\xi}_k \geq (1 - \sigma_\xi) \frac{\Delta \bar{l}(x_k, \bar{r}_k, \bar{d}_k)}{\bar{r}_k \|\bar{d}_k\|^2} \geq (1 - \sigma_\xi) \frac{\bar{r}_k \sigma_u \lambda_u \|\bar{d}_k\|^2}{2 \bar{r}_k \|\bar{d}_k\|^2} = \frac{(1 - \sigma_\xi) \sigma_u \lambda_u}{2} > 0.$$

Setting  $\bar{\xi}_{\min} = \min \left\{ \frac{(1 - \sigma_\xi) \sigma_u \lambda_u}{\max\{2(1 + \chi^{-1}), 2\}}, \frac{(1 - \sigma_\xi) \sigma_c \sigma_v}{2(\chi_{\max} + 1)(\kappa_c + \epsilon_c) \sigma_{j_c}^2} \right\}$  completes the proof.  $\square$

The next lemma pertains to the boundedness of the tangential and full search directions.

**Lemma 3.12.** *Suppose that Assumptions 2.1, 2.2, 2.4, 3.1, 3.2, and 3.4 hold. Suppose that the minimum non-zero singular values of  $\bar{J}_k$  are bounded below by  $\kappa_\sigma \in \mathbb{R}_{>0}$  for all  $k \in \mathbb{N}$ . Then, for all  $k \in \mathbb{N}$ , there exists  $\{\kappa_{\bar{u}}, \kappa_{\bar{d}}\} \subset \mathbb{R}_{>0}$  such that  $\|\bar{u}_k\| \leq \kappa_{\bar{u}}$  and  $\|\bar{d}_k\| \leq \kappa_{\bar{d}}$ .*

*Proof.* Suppose that  $\text{rank}(\bar{J}_k) = p$ . First, let us discuss the case when  $p = n \leq m$ . It follows by the rank-nullity theorem [28] that  $\text{nullity}(\bar{J}_k) = n - p = 0$ , i.e., the null space of  $\bar{J}_k$  is trivial (contains only the zero vector). Since  $\bar{J}_k$  has full column rank, the matrix  $\bar{J}_k^T \bar{J}_k$  is symmetric and positive definite, ensuring that it is invertible. It then follows by (2.6), (2.9), Assumptions 2.1, 2.2, and 2.4, and the condition of the minimum non-zero singular values of  $\bar{J}_k$  being bounded below by  $\kappa_\sigma \in \mathbb{R}_{>0}$  that  $\|\bar{u}_k\| = \|(\bar{J}_k^T \bar{J}_k)^{-1} \bar{J}_k^T \bar{r}_k\| \leq \|(\bar{J}_k^T \bar{J}_k)^{-1}\| \|\bar{J}_k\| \|\bar{r}_k\| \leq (\kappa_{J_k} + \epsilon_{J_k}) \kappa_\sigma^{-2} \lambda_{\rho r} \kappa_{\rho r}$ . Now, suppose that  $p = n \leq m$  does not hold. Since  $p \leq \min\{m, n\}$ , we consider the following cases: (i)  $p < n \leq m$  and (ii)  $p \leq m < n$ . In both cases, it follows that  $p < n$ . The null space of  $\bar{J}_k$  is therefore nontrivial. Suppose that  $\bar{\Sigma}_k$  is a diagonal matrix containing the non-zero singular values of  $\bar{J}_k$ , where the full SVD decomposition of  $\bar{J}_k$  is given as

$$\bar{J}_k^T = (\bar{U}_k \ \bar{Z}_k) \begin{pmatrix} \bar{\Sigma}_k & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \bar{V}_k \\ \bar{V}_k^\perp \end{pmatrix} = \bar{U}_k \bar{\Sigma}_k \bar{V}_k, \quad (3.13)$$

where  $\bar{U}_k \in \mathbb{R}^{n \times p}$ ,  $\bar{Z}_k \in \mathbb{R}^{n \times (n-p)}$ ,  $\bar{\Sigma}_k \in \mathbb{R}^{p \times p}$ ,  $\bar{V}_k \in \mathbb{R}^{p \times m}$ ,  $\bar{V}_k^\perp \in \mathbb{R}^{(m-p) \times m}$ . Let  $\tilde{J}_k^T = \bar{U}_k \bar{\Sigma}_k \in \mathbb{R}^{n \times p}$ . We have  $\text{Range}(\tilde{J}_k^T) = \text{Range}(\bar{J}_k^T)$ ,  $\|\tilde{J}_k \tilde{J}_k^T\| = \|\bar{J}_k \bar{J}_k^T\|$ . Since  $\bar{Z}_k$  represents the orthonormal basis of null space of  $\bar{J}_k$ , by the linear system (2.6)

$$\bar{u}_k = \tilde{J}_k^T \left( \tilde{J}_k \tilde{J}_k^T \right)^{-1} \bar{V}_k \bar{r}_k - \bar{Z}_k \left( \bar{Z}_k^T H_k \bar{Z}_k \right)^{-1} \bar{Z}_k^T \left( \bar{g}_k + H_k \bar{v}_k + H_k \tilde{J}_k^T \left( \tilde{J}_k \tilde{J}_k^T \right)^{-1} \bar{V}_k \bar{r}_k - \bar{\rho}_k \right). \quad (3.14)$$

Since  $\bar{U}_k$  and  $\bar{V}_k$  are both submatrices of orthogonal matrices,  $\|\bar{U}_k\| \leq 1$ ,  $\|\bar{V}_k\| \leq 1$ . By  $\|(\tilde{J}_k \tilde{J}_k^T)^{-1}\| = \|\bar{\Sigma}_k^{-2}\| \leq \kappa_\sigma^{-2}$ , Assumptions 2.1, 2.2, 2.4 and (2.9) it follows that

$$\begin{aligned} & \|\bar{u}_k\| \\ & \leq \left\| \tilde{J}_k^T \left( \tilde{J}_k \tilde{J}_k^T \right)^{-1} \bar{V}_k \bar{r}_k \right\| + \left\| \bar{Z}_k \left( \bar{Z}_k^T H_k \bar{Z}_k \right)^{-1} \bar{Z}_k^T \left( \bar{g}_k + H_k \bar{v}_k + H_k \tilde{J}_k^T \left( \tilde{J}_k \tilde{J}_k^T \right)^{-1} \bar{V}_k \bar{r}_k - \bar{\rho}_k \right) \right\| \end{aligned}$$

$$\begin{aligned}
&\leq (\kappa_J + \epsilon_J) \kappa_\sigma^{-2} \|\bar{r}_k\| + \zeta^{-1} (\kappa_g + \epsilon_g + \kappa_H \sigma_{Jc} (\kappa_J + \epsilon_J) (\kappa_c + \epsilon_c) + \kappa_H (\kappa_J + \epsilon_J) \kappa_\sigma^{-2} \|\bar{r}_k\| + \|\bar{\rho}_k\|) \\
&\leq ((1 + \zeta^{-1} \kappa_H) (\kappa_J + \epsilon_J) \kappa_\sigma^{-2} + \zeta^{-1}) \lambda_{\rho r} \kappa_{\rho r} + \zeta^{-1} (\kappa_g + \epsilon_g + \kappa_H \sigma_{Jc} (\kappa_J + \epsilon_J) (\kappa_c + \epsilon_c)) := \kappa_{\bar{u}}.
\end{aligned}$$

Since  $\kappa_{\bar{u}} > (\kappa_J + \epsilon_J) \kappa_\sigma^{-2} \lambda_{\rho r} \kappa_{\rho r}$ , the bound derived when  $\text{rank}(\bar{J}_k) = p = n$ , we have  $\|\bar{u}_k\| \leq \kappa_{\bar{u}}$  for all  $k \in \mathbb{N}$ . Given that  $\|\bar{v}_k\| \leq \sigma_{Jc} \|\bar{J}_k^T \bar{c}_k\| \leq \sigma_{Jc} (\kappa_J + \epsilon_J) (\kappa_c + \epsilon_c)$  and  $\|\bar{d}_k\| \leq \|\bar{v}_k\| + \|\bar{u}_k\|$ , there exists  $\kappa_{\bar{d}} \in \mathbb{R}_{>0}$  such that  $\|\bar{d}_k\| \leq \kappa_{\bar{d}}$ .  $\square$

Next, we prove that  $\{\bar{\tau}_k\}$  is bounded away from zero when the singular values of  $\bar{J}_k$  are bounded below by  $\kappa_\sigma \in \mathbb{R}_{>0}$  for all  $k \in \mathbb{N}$ . The step sizes for both schemes (**Option I** and **Option II**) are also lower bounded in this case (Lemmas 3.14 and 3.15).

**Lemma 3.13.** *Suppose that Assumptions 2.1, 2.2, 2.4, 3.1, 3.2, and 3.4 hold. The merit parameter sequence  $\{\bar{\tau}_k\}$  is non-increasing. If the singular values of  $\bar{J}_k$  are bounded below by  $\kappa_\sigma \in \mathbb{R}_{>0}$  for all  $k \in \mathbb{N}$ , then there exists  $\tilde{\tau}_{\min} \in \mathbb{R}_{>0}$  such that  $\bar{\tau}_k \geq \tilde{\tau}_{\min}$  for all  $k \in \mathbb{N}$ .*

*Proof.* When Termination Test 1 is triggered,  $\bar{\tau}_k = \bar{\tau}_{k-1}$ . When Termination Test 2 is triggered, either  $\bar{\tau}_k = \bar{\tau}_{k-1}$  or  $\bar{\tau}_k = (1 - \sigma_\tau) \bar{\tau}_k^{\text{trial}} < \bar{\tau}_{k-1}$ . Therefore,  $\{\bar{\tau}_k\}$  is a non-increasing sequence. By Algorithm 1, (2.16), and (2.17), it follows that  $\bar{\tau}_k < \bar{\tau}_{k-1}$  only when  $\|\bar{c}_k\| > \epsilon_o$  (Termination Test 2),  $\bar{g}_k^T \bar{d}_k + \max \left\{ \bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2 \right\} > 0$ , and

$$\bar{\tau}_{k-1} > \frac{(1 - \sigma_\tau) \left(1 - \frac{\sigma_c}{\sigma_r}\right) (\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k + \bar{r}_k\|)}{\bar{g}_k^T \bar{d}_k + \max \left\{ \bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2 \right\}}.$$

By Assumptions 2.1, 2.2, and 2.4, (2.2), and Lemma 3.12, considering the two cases in (2.13),

$$\begin{aligned}
\bar{g}_k^T \bar{d}_k + \max \left\{ \bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2 \right\} &= \bar{g}_k^T \bar{u}_k + \bar{g}_k^T \bar{v}_k + \max \left\{ \bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2 \right\} \\
&\leq (\kappa_g + \epsilon_g) (\lambda_{uv} + 1) \|\bar{v}_k\| + \max \{ \kappa_H, \lambda_u \} \|\bar{u}_k\|^2 \\
&\leq ((\kappa_g + \epsilon_g) (\lambda_{uv} + 1) + \kappa_{\bar{u}} \max \{ \kappa_H, \lambda_u \} \lambda_{uv}) \sigma_{Jc} \|\bar{J}_k^T \bar{c}_k\|,
\end{aligned}$$

or

$$\begin{aligned}
\bar{g}_k^T \bar{d}_k + \max \left\{ \bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2 \right\} &= \bar{g}_k^T \bar{u}_k + \bar{g}_k^T \bar{v}_k + \bar{u}_k^T H_k \bar{u}_k \\
&= (\bar{g}_k + H_k \bar{v}_k)^T \bar{u}_k + (1 - \|\bar{J}_k^T \bar{c}_k\|) \bar{u}_k^T H_k \bar{u}_k + \bar{g}_k^T \bar{v}_k \\
&\quad + \|\bar{J}_k^T \bar{c}_k\| \bar{u}_k^T H_k \bar{u}_k - \bar{v}_k^T H_k \bar{u}_k \\
&\leq \lambda_u \|\bar{v}_k\| + (\kappa_g + \epsilon_g) \|\bar{v}_k\| + \kappa_{\bar{u}}^2 \kappa_H \|\bar{J}_k^T \bar{c}_k\| + \kappa_{\bar{u}} \kappa_H \|\bar{v}_k\| \\
&\leq ((\lambda_u + \kappa_g + \epsilon_g + \kappa_{\bar{u}} \kappa_H) \sigma_{Jc} + \kappa_{\bar{u}}^2 \kappa_H) \|\bar{J}_k^T \bar{c}_k\|.
\end{aligned}$$

Thus, there exists a constant  $\kappa_D \in \mathbb{R}_{>0}$  such that  $\bar{g}_k^T \bar{d}_k + \max \left\{ \bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2 \right\} \leq \kappa_D \|\bar{J}_k^T \bar{c}_k\|$ . Hence, if the merit parameter  $\bar{\tau}_k$  is ever updated, it follows by (2.15), (3.9),

and Lemma 3.6 that

$$\begin{aligned}
\bar{\tau}_{k-1} &> \frac{(1-\sigma_\tau)\left(1-\frac{\sigma_c}{\sigma_r}\right)\left(\|\bar{c}_k\|-\|\bar{c}_k+\bar{J}_k\bar{v}_k+\bar{r}_k\|\right)}{\bar{g}_k^T\bar{d}_k+\max\{\bar{u}_k^T H_k \bar{u}_k, \lambda_u\|\bar{u}_k\|^2\}} \\
&\geq \frac{(1-\sigma_\tau)(\sigma_r-\sigma_c)\left(\|\bar{c}_k\|-\|\bar{c}_k+\bar{J}_k\bar{v}_k\|\right)}{\bar{g}_k^T\bar{d}_k+\max\{\bar{u}_k^T H_k \bar{u}_k, \lambda_u\|\bar{u}_k\|^2\}} \\
&\geq \frac{(1-\sigma_\tau)(\sigma_r-\sigma_c)\sigma_v\|J_k^T\bar{c}_k\|^2/\|\bar{c}_k\|}{\kappa_D\|J_k^T\bar{c}_k\|} \\
&\geq \frac{(1-\sigma_\tau)(\sigma_r-\sigma_c)\sigma_v\kappa_\sigma}{\kappa_D} := \tilde{\tau}_{\min} > 0.
\end{aligned} \tag{3.15}$$

The desired result is therefore satisfied with the defined  $\tilde{\tau}_{\min}$ .  $\square$

Lemma 3.13 shows that  $\bar{\tau}_k$  is uniformly bounded away from zero when  $\bar{J}_k$  is of full rank. In general, there are two possible outcomes for the sequence  $\{\bar{\tau}_k\}$ : (i) there exists  $\bar{\tau}_{\min} \in \mathbb{R}_{>0}$  such that  $\bar{\tau}_k \geq \bar{\tau}_{\min}$  for all  $k \in \mathbb{N}$ ; and (ii)  $\bar{\tau}_k \rightarrow 0$ . In Section 3.4, we derive convergence guarantees and complexity results for the different outcomes. The remainder of this subsection is about properties of step sizes.

Lemmas 3.14 and 3.15 provide lower and upper bounds for step sizes generated by both the adaptive and the line search schemes. The lower bounds of the step size are established when  $\bar{\tau}_k \geq \bar{\tau}_{\min}$ . However, when  $\{\bar{\tau}_k\}$  diminishes to 0, the step size may converge to zero, which poses challenges for the convergence of the algorithm. We prove that the step size does not vanish under certain conditions. For this purpose, we define

$$\mathcal{X}_\gamma := \{x \in \mathbb{R}^n : \|\bar{J}(x)^T \bar{c}(x)\| \geq \gamma, \|\bar{c}(x)\| > \epsilon_o\} \tag{3.16}$$

for some  $\gamma \in \mathbb{R}_{>0}$  and show that the step sizes are uniformly bounded away from zero when  $x_k \in \mathcal{X}_\gamma$  even when  $\{\bar{\tau}_k\}$  diminishes to zero. We first show that the conclusions above hold for the adaptive step size scheme (**Option I**).

**Lemma 3.14.** *Suppose that Assumptions 2.1, 2.2, 2.4, 3.1, 3.2, and 3.4 hold, and that **Option I** is used to select the step size in Algorithm 1. In addition, suppose that  $\beta_k = \beta \in (0, 1)$  satisfies  $\frac{2(1-\eta)\beta\bar{\xi}_{-1}\max\{\bar{\tau}_{-1}, 1\}}{\bar{\tau}_{-1}\Gamma+L} \in (0, 1]$  for all  $k \in \mathbb{N}$ . For any  $k \in \mathbb{N}$ , it holds that  $0 < \bar{\alpha}_k^{\min} \leq \bar{\alpha}_k^{\text{uff}}$ ,  $\bar{\alpha}_k \leq \bar{\alpha}_k^{\text{uff}} \leq 1$ , and*

$$0 < \bar{\alpha}_k^{\min} \leq \bar{\alpha}_k^{\max} \leq \left( \frac{2(1-\eta)\bar{\xi}_{-1}\max\{\bar{\tau}_{-1}, 1\}}{\Gamma} + \theta \right) \beta := \alpha_u^A. \tag{3.17}$$

- If  $\bar{\tau}_k \geq \bar{\tau}_{\min}$  for all  $k \in \mathbb{N}$ , it follows that  $\bar{\alpha}_k^{\min} \geq \frac{2(1-\eta)\beta\bar{\xi}_{\min}\min\{\bar{\tau}_{\min}, 1\}}{\bar{\tau}_{\min}L+\Gamma} := \alpha_{l\tau}^A$ .
- If  $\{\bar{\tau}_k\}$  diminishes to 0, for  $\gamma \in \mathbb{R}_{>0}$ , there exists  $\alpha_{i0}^A \in \mathbb{R}_{>0}$  such that if  $x_k \in \mathcal{X}_\gamma$ , then  $\bar{\alpha}_k \geq \alpha_{i0}^A$ .

*Proof.* Since  $\bar{\xi}_k \leq \bar{\xi}_k^{\text{trial}}$  by (2.20), by the update rule of  $\bar{\alpha}_k^{\min}$  (2.23) it follows that

$$\bar{\alpha}_k^{\min} \leq \frac{2(1-\eta)\beta_k \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k)}{(\bar{\tau}_k L + \Gamma) \|\bar{d}_k\|^2}.$$

In addition, by Lemmas 3.11 and 3.13, (2.23) and the prescribed condition for  $\{\beta_k\}$

$$\bar{\alpha}_k^{\min} \leq \frac{2(1-\eta)\beta_k \bar{\xi}_{-1} \max\{\bar{\tau}_{-1}, 1\}}{\bar{\tau}_{-1} L + \Gamma} \leq 1.$$

By (2.22),  $\bar{\alpha}_k^{\min} \leq \min \left\{ \frac{2(1-\eta)\beta_k \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k)}{(\bar{\tau}_k L + \Gamma) \|\bar{d}_k\|^2}, 1 \right\} = \bar{\alpha}_k^{\text{suff}}$ , and hence, the projection operator will never increase the step size, i.e.,  $\bar{\alpha}_k \leq \bar{\alpha}_k^{\text{suff}} \leq 1$ . Moreover, by  $\beta_k = \beta \in (0, 1)$  and the definitions of  $\bar{\alpha}_k^{\min}$  and  $\bar{\alpha}_k^{\max}$ , it follows that  $\bar{\alpha}_k^{\max} = \bar{\alpha}_k^{\min} + \theta\beta \leq \left( \frac{2(1-\eta)\bar{\xi}_{-1} \max\{\bar{\tau}_{-1}, 1\}}{\Gamma} + \theta \right) \beta$ .

If  $\bar{\tau}_k \geq \bar{\tau}_{\min}$ , it follows by (2.23) and Lemma 3.13 that

$$\frac{2(1-\eta)\beta \bar{\xi}_{\min} \min\{\bar{\tau}_{\min}, 1\}}{\bar{\tau}_{\min} L + \Gamma} \leq \frac{2(1-\eta)\beta \bar{\xi}_k \min\{\bar{\tau}_k, 1\}}{\bar{\tau}_k L + \Gamma} \leq \bar{\alpha}_k^{\min}.$$

The desired conclusion is therefore satisfied.

If  $\{\bar{\tau}_k\}$  vanishes to zero, define  $\mathcal{K}_\gamma := \{k \in \mathbb{N} : x_k \in \mathcal{X}_\gamma\}$ . It follows by Lemma 3.6 that  $\|\bar{v}_k\| \geq \underline{\sigma}_{Jc} \|\bar{J}_k^T \bar{c}_k\|^2 \geq \underline{\sigma}_{Jc} \gamma^2$  for all  $k \in \mathcal{K}_\gamma$ . Furthermore,  $\|\bar{u}_k\| \leq \frac{\kappa_{\bar{u}}}{\underline{\sigma}_{Jc} \gamma^2} \|\bar{v}_k\|$  for all  $k \in \mathcal{K}_\gamma$ , by Lemma 3.12. If  $\|\bar{u}_k\|^2 < \bar{\chi}_k \|\bar{v}_k\|^2$ , then  $\bar{\alpha}_k \geq \frac{2(1-\eta)\beta \bar{\xi}_{\min}}{(\bar{\tau}_{-1} L + \Gamma)}$  by (2.23) and Lemma 3.11. If  $\|\bar{u}_k\|^2 \geq \bar{\chi}_k \|\bar{v}_k\|^2$ , by (2.24), (3.2), Lemma 3.6, and  $\bar{\alpha}_k^{\min} \leq \bar{\alpha}_k^{\text{suff}}$ ,

$$\begin{aligned} \bar{\alpha}_k &= \min \left\{ \bar{\alpha}_k^{\text{suff}}, \bar{\alpha}_k^{\max} \right\} \\ &= \min \left\{ \frac{2(1-\eta)\beta \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k)}{(\bar{\tau}_k L + \Gamma) \|\bar{d}_k\|^2}, 1, \frac{2(1-\eta)\beta \bar{\xi}_k \bar{\tau}_k}{\bar{\tau}_k L + \Gamma} + \theta\beta \right\} \\ &\geq \min \left\{ \frac{(1-\eta)\beta \sigma_c \sigma_v \|\bar{J}_k^T \bar{c}_k\|^2}{(\bar{\tau}_k L + \Gamma)(\kappa_c + \epsilon_c)(\|\bar{u}_k\|^2 + \|\bar{v}_k\|^2)}, 1, \frac{2(1-\eta)\beta \bar{\xi}_k \bar{\tau}_k}{\bar{\tau}_k L + \Gamma} + \theta\beta \right\} \\ &\geq \min \left\{ \frac{(1-\eta)\beta \sigma_c \sigma_v \|\bar{J}_k^T \bar{c}_k\|^2}{(\bar{\tau}_k L + \Gamma)(\kappa_c + \epsilon_c)(\kappa_{\bar{u}}^2 / (\sigma_{Jc}^2 \gamma^4) + 1) \sigma_{Jc}^2 \|\bar{J}_k^T \bar{c}_k\|^2}, 1, \frac{2(1-\eta)\beta \bar{\xi}_k \bar{\tau}_k}{\bar{\tau}_k L + \Gamma} + \theta\beta \right\} \\ &\geq \min \left\{ \frac{(1-\eta)\beta \sigma_c \sigma_v \underline{\sigma}_{Jc}^2 \gamma^4}{(\bar{\tau}_{-1} L + \Gamma)(\kappa_c + \epsilon_c) \sigma_{Jc}^2 (\kappa_{\bar{u}}^2 + \sigma_{Jc}^2 \gamma^4)}, 1, \theta\beta \right\}. \end{aligned} \quad (3.18)$$

Let  $\alpha_{i0}^A := \min \left\{ \frac{(1-\eta)\beta \sigma_c \sigma_v \underline{\sigma}_{Jc}^2 \gamma^4}{(\bar{\tau}_{-1} L + \Gamma)(\kappa_c + \epsilon_c) \sigma_{Jc}^2 (\kappa_{\bar{u}}^2 + \sigma_{Jc}^2 \gamma^4)}, 1, \theta\beta, \frac{2(1-\eta)\beta \bar{\xi}_{\min}}{\bar{\tau}_{-1} L + \Gamma} \right\}$ , and the result holds.  $\square$

The next lemma proves that similar results apply to the line search scheme (**Option II**), albeit with different lower and upper bounds.

**Lemma 3.15.** *Suppose that Assumptions 2.1, 2.2, 2.4, 3.1, 3.2, and 3.4 hold, and that Option II is used to select the step size in Algorithm 1 with*

$$\epsilon_{A_k} = 2\bar{\tau}_k \epsilon_f + 4\epsilon_c + \bar{\tau}_k \alpha_u^L \epsilon_g \|\bar{d}_k\| + \alpha_u^L \epsilon_J \|\bar{d}_k\|. \quad (3.19)$$



- If  $\bar{\tau}_k \geq \bar{\tau}_{\min}$  for all  $k \in \mathbb{N}$ , there exists  $\kappa_{\Delta l} \in \mathbb{R}_{>0}$  (that depends on  $\bar{\tau}_{\min}$ ) such that  $\bar{\alpha}_k \geq \min \left\{ \alpha_u^L, \frac{\nu(1-\eta)\kappa_{\Delta l}}{\bar{\tau}_{-1}L+\Gamma} \right\} := \alpha_{l\tau}^L$ .
- If  $\{\bar{\tau}_k\}$  diminishes to 0, for  $\gamma \in \mathbb{R}_{>0}$ , there exists  $\alpha_{l0}^L \in \mathbb{R}_{>0}$  such that if  $x_k \in \mathcal{X}_\gamma$ , then  $\bar{\alpha}_k \geq \alpha_{l0}^L$ .

*Proof.* If  $\bar{\tau}_k \geq \bar{\tau}_{\min}$ , by Lemma 3.9 there exists  $\kappa_{\Delta l} \in \mathbb{R}_{>0}$  (that depends on  $\bar{\tau}_{\min}$ ) such that  $\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) \geq \kappa_{\Delta l} \|\bar{d}_k\|^2$ . By Assumptions 2.1 and 2.2, (2.7), (2.8), and  $\bar{\alpha}_k \leq \alpha_u^L \leq 1$

$$\begin{aligned}
& \phi(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_k) - \phi(x_k, \bar{\tau}_k) \\
&= \bar{\tau}_k f_{k+1} - \bar{\tau}_k f_k + \|c_{k+1}\| - \|c_k\| \\
&\leq \bar{\alpha}_k \bar{\tau}_k g_k^T \bar{d}_k + \|c_k + \bar{\alpha}_k J_k \bar{d}_k\| - \|c_k\| + \frac{1}{2} (\bar{\tau}_k L + \Gamma) \bar{\alpha}_k^2 \|\bar{d}_k\|^2 \\
&\leq \bar{\alpha}_k \bar{\tau}_k g_k^T \bar{d}_k + \|\bar{c}_k + \bar{\alpha}_k \bar{J}_k \bar{d}_k\| - \|\bar{c}_k\| + \frac{1}{2} (\bar{\tau}_k L + \Gamma) \bar{\alpha}_k^2 \|\bar{d}_k\|^2 + \|c_k - \bar{c}_k\| + \|c_k - \bar{c}_k + \bar{\alpha}_k (J_k - \bar{J}_k) \bar{d}_k\| \\
&\leq \bar{\alpha}_k \bar{\tau}_k \bar{g}_k^T \bar{d}_k + \bar{\alpha}_k \bar{\tau}_k (g_k - \bar{g}_k)^T \bar{d}_k - \bar{\alpha}_k \|\bar{c}_k\| + \bar{\alpha}_k \|\bar{c}_k + \bar{J}_k \bar{d}_k\| \\
&\quad + \|\bar{c}_k + \bar{\alpha}_k \bar{J}_k \bar{d}_k\| - \|\bar{c}_k\| + \bar{\alpha}_k \|\bar{c}_k\| - \bar{\alpha}_k \|\bar{c}_k + \bar{J}_k \bar{d}_k\| + \frac{1}{2} (\bar{\tau}_k L + \Gamma) \bar{\alpha}_k^2 \|\bar{d}_k\|^2 + 2\epsilon_c + \bar{\alpha}_k \epsilon_J \|\bar{d}_k\| \\
&\leq -\bar{\alpha}_k \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) + \bar{\alpha}_k \bar{\tau}_k (g_k - \bar{g}_k)^T \bar{d}_k \\
&\quad + \|\bar{c}_k + \bar{\alpha}_k \bar{J}_k \bar{d}_k\| - \|\bar{c}_k\| + \bar{\alpha}_k \|\bar{c}_k\| - \bar{\alpha}_k \|\bar{c}_k + \bar{J}_k \bar{d}_k\| + \frac{1}{2} (\bar{\tau}_k L + \Gamma) \bar{\alpha}_k^2 \|\bar{d}_k\|^2 + 2\epsilon_c + \bar{\alpha}_k \epsilon_J \|\bar{d}_k\| \\
&\leq -\bar{\alpha}_k \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) + \bar{\alpha}_k \bar{\tau}_k (g_k - \bar{g}_k)^T \bar{d}_k + (1 - \bar{\alpha}_k) \|\bar{c}_k\| + (\bar{\alpha}_k - 1) \|\bar{c}_k\| \\
&\quad + \frac{1}{2} (\bar{\tau}_k L + \Gamma) \bar{\alpha}_k^2 \|\bar{d}_k\|^2 + 2\epsilon_c + \bar{\alpha}_k \epsilon_J \|\bar{d}_k\| \\
&\leq -\bar{\alpha}_k \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) + \bar{\alpha}_k \bar{\tau}_k (g_k - \bar{g}_k)^T \bar{d}_k + \frac{\bar{\tau}_{-1}L+\Gamma}{2\kappa_{\Delta l}} \bar{\alpha}_k^2 \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) + 2\epsilon_c + \bar{\alpha}_k \epsilon_J \|\bar{d}_k\| \\
&\leq -\bar{\alpha}_k \left( 1 - \frac{\bar{\tau}_{-1}L+\Gamma}{2\kappa_{\Delta l}} \bar{\alpha}_k \right) \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) + \bar{\tau}_k \alpha_u^L \epsilon_g \|\bar{d}_k\| + 2\epsilon_c + \alpha_u^L \epsilon_J \|\bar{d}_k\|.
\end{aligned} \tag{3.20}$$

Moreover, (2.25) implies that when  $\bar{\alpha}_k \leq \frac{(1-\eta)\kappa_{\Delta l}}{\bar{\tau}_{-1}L+\Gamma}$ , (2.26) is satisfied by the definition of  $\epsilon_{A_k}$ . By setting  $\alpha_{l\tau}^L := \min \left\{ \alpha_u^L, \frac{\nu(1-\eta)\kappa_{\Delta l}}{\bar{\tau}_{-1}L+\Gamma} \right\}$ , the desired conclusion then follows.

If  $\{\bar{\tau}_k\}$  vanishes to zero, define  $\mathcal{K}_\gamma := \{k \in \mathbb{N} : x_k \in \mathcal{X}_\gamma\}$ . It follows by the same arguments as in the proof of Lemma 3.14 that  $\|\bar{u}_k\| \leq \frac{\kappa_{\bar{u}}}{\underline{\sigma}_{J_c} \gamma^2} \|\bar{v}_k\|$  for all  $k \in \mathcal{K}_\gamma$  and

$$\frac{\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k)}{\|\bar{d}_k\|^2} \geq \frac{\sigma_c \sigma_v \|\bar{J}_k^T \bar{c}_k\|^2}{2(\kappa_c + \epsilon_c)(\|\bar{u}_k\|^2 + \|\bar{v}_k\|^2)} \geq \frac{\sigma_c \sigma_v \|\bar{J}_k^T \bar{c}_k\|^2}{2(\kappa_c + \epsilon_c)(\kappa_{\bar{u}}^2 / (\underline{\sigma}_{J_c}^2 \gamma^4) + 1) \sigma_{J_c}^2 \|\bar{J}_k^T \bar{c}_k\|^2} \geq \frac{\sigma_c \sigma_v \underline{\sigma}_{J_c}^2 \gamma^4}{2(\kappa_c + \epsilon_c)(\kappa_{\bar{u}}^2 + \underline{\sigma}_{J_c}^2 \gamma^4) \sigma_{J_c}^2}.$$

The desired conclusion then follows by the definition of  $\kappa_{\Delta l}(\gamma)$ ,  $\alpha_{l0}^L$  and a similar rationale as in the previous case.  $\square$

**Remark 3.16.** Lemmas 3.14 and 3.15 provide useful lower and upper bounds of the step size for the adaptive and line search schemes, respectively. The results in Lemma 3.14 are similar to those in [7, Theorem 2 & Lemma 12]. The relaxation parameter  $\epsilon_{A_k} \in \mathbb{R}_{\geq 0}$

(defined in (3.19)) ensures that the step size that satisfies the line search condition (2.26) is bounded away from zero under reasonable conditions, ensuring that the backtracking line search procedure is well defined. Relaxation parameters are common in adaptive algorithms in the noisy setting. It is important to note that the relaxation parameters used in [5, 6, 10, 22, 29, 32] depend on  $\epsilon_f$  for problems with noisy objective functions, and the relaxation parameter in [38] depends on both  $\epsilon_f$  and  $\epsilon_c$  for problems with both noisy objective and constraint functions. Similar relaxation strategies have been developed in the context of trust-region methods; see e.g., [12, 30, 44] for problems with noisy objective functions and [45] for problems with both noisy objective and constraint functions. The fact that all relaxation parameters in previous works do not depend on  $\epsilon_g$  and  $\epsilon_J$  can be attributed to the following reasons: (i) stronger assumptions are made on the gradient approximations [5, 22] or (ii) the desired stationarity measure is assumed to be sufficiently large compared to  $\epsilon_g$  and/or  $\epsilon_J$  in the convergence analysis [6, 10, 29, 32, 38, 12, 30, 44]. The additional terms in the relaxation parameter (3.19),  $\bar{\tau}_k \alpha_u^L \epsilon_g \|\bar{d}_k\| + \alpha_u^L \epsilon_J \|\bar{d}_k\|$ , are necessary in order to establish non-asymptotic guarantees (Theorem 3.36) without the need for additional assumptions or restrictions. It is important to note that the additional terms can be replaced by quantities proportional to  $(\bar{\tau}_{-1} \epsilon_g + \epsilon_J)^2$ . Finally, when  $\{\bar{\tau}_k\}$  diminishes to zero, the noise levels in the constraint functions,  $\epsilon_c$  and  $\epsilon_J$ , dominate the relaxation parameter  $\epsilon_{A_k}$ .

Table 1 summarizes the results from Lemmas 3.14 and 3.15.

Table 1: Lower and upper bounds for  $\bar{\alpha}_k$  in Algorithm 1 (**Option I** and **Option II**).

	Upper Bound	Lower Bound ( $\bar{\tau}_k \geq \bar{\tau}_{\min}$ )	Lower Bound ( $\bar{\tau}_k \rightarrow 0$ and $x_k \in \mathcal{X}_\gamma$ )
<b>Option I</b>	$\left(\frac{2(1-\eta)\bar{\xi}_{-1} \max\{\bar{\tau}_{-1}, 1\}}{\Gamma} + \theta\right) \beta$	$\frac{2(1-\eta)\bar{\xi}_{\min} \min\{\bar{\tau}_{\min}, 1\}}{\bar{\tau}_{\min} L + \Gamma} \beta$	$\min \left\{ \frac{(1-\eta)\sigma_c \sigma_v g_{J_c}^2 \gamma^4}{(\bar{\tau}_{-1} L + \Gamma)(\kappa_c + \epsilon_c) \sigma_{J_c}^2 (\kappa_u^2 + \sigma_{J_c}^2 \gamma^4)}, \frac{1}{\beta}, \theta, \frac{2(1-\eta)\bar{\xi}_{\min}}{\bar{\tau}_{-1} L + \Gamma} \right\} \beta$
<b>Option II</b>	$\alpha_u^L \in (0, 1]$	$\min \left\{ \alpha_u^L, \frac{\nu(1-\eta)\kappa_{\Delta l}}{\bar{\tau}_{-1} L + \Gamma} \right\}$	$\min \left\{ \alpha_u^L, \frac{\nu(1-\eta)\kappa_{\Delta l}(\gamma)}{\bar{\tau}_{-1} L + \Gamma} \right\}$

### 3.3 Perturbation Analysis

In this section, we present theoretical guarantees that characterize the perturbations of the (possibly inexact) solutions to subproblems (2.2) and (2.4). Specifically, we establish how the solutions of the noisy subproblems deviate from those of the fictitious (never solved) true subproblems. Moreover, we derive useful upper bounds on the deviation of the lower bound of the merit parameter value as compared to the fictitious (never computed) true merit parameter value. The former results provide insight into the stationarity, feasibility, or infeasible stationarity measures of the true problem, and the latter results provide insights into the stability of the algorithm.

To simplify the notation, let  $g_k = \nabla f(x_k)$  and  $J_k = \nabla c(x_k)^T$ , and let  $v_k$ ,  $u_k^*$  and  $\tau_k$  denote the true normal and tangential components and true merit parameter (defined explicitly below in (3.23), (3.42), and (3.52)–(3.53)), respectively. As shown in [4], when

the constraint evaluations are exact (i.e.,  $\epsilon_c = \epsilon_J = 0$ ) and the singular values of  $J_k$  are bounded away from zero, the difference between the noisy and true search directions ( $\bar{d}_k$  and  $d_k$ ) is upper bounded by a quantity proportional to  $\epsilon_g$ . Deriving such a result in our noisy setting is not as straightforward. The main reason for this is that  $J_k$  or  $\bar{J}_k$  can be rank-deficient and the fact that our algorithm makes use of inexact subproblem solutions. As such, we do not directly prove that  $\|\bar{d}_k - d_k\|$  is sufficiently small. Instead, we establish the convergence of our algorithm by deriving upper bounds for  $\|\bar{v}_k - v_k\|$  when the LICQ holds and either exact solutions or Cauchy solutions are employed (see Lemmas 3.18 and 3.19) and  $\|\bar{u}_k - u_k^*\|$  regardless of whether  $J_k$  or  $\bar{J}_k$  are rank-deficient (see Lemma 3.25). When the LICQ is violated,  $\bar{v}_k$  and  $v_k$  can be non-unique and the norm difference  $\|\bar{v}_k - v_k\|$  can be large. The trust region constraint  $\|\bar{v}_k\| \leq \sigma_{J_c} \|\bar{J}_k^T \bar{c}_k\|$  is sufficient for algorithmic convergence, as we will demonstrate that  $\|\bar{J}_k^T \bar{c}_k\|$  eventually becomes small (see Theorem 3.36).

The first lemma bounds the difference between  $\|\bar{J}_k^T \bar{c}_k\|$  and  $\|J_k^T c_k\|$ .

**Lemma 3.17.** *Suppose that Assumptions 2.1 and 2.2 hold. There exists  $\mathcal{E}_{J_c}, \mathcal{E}_{J_c^2} \in \mathbb{R}_{\geq 0}$  such that, for all  $k \in \mathbb{N}$*

$$\|J_k^T c_k - \bar{J}_k^T \bar{c}_k\| \leq \mathcal{E}_{J_c} \quad \text{and} \quad \left| \|J_k^T c_k\|^2 - \|\bar{J}_k^T \bar{c}_k\|^2 \mathbb{1}\{\|\bar{c}_k\| > \epsilon_o\} \right| \leq \mathcal{E}_{J_c^2}.$$

*Proof.* By Assumptions 2.1 and 2.2 it follows that

$$\|J_k^T c_k - \bar{J}_k^T \bar{c}_k\| \leq \|\bar{J}_k^T (c_k - \bar{c}_k)\| + \|(J_k - \bar{J}_k)^T c_k\| \leq ((\kappa_J + \epsilon_J)\epsilon_c + \kappa_c \epsilon_J) := \mathcal{E}_{J_c}, \quad (3.21)$$

and

$$\begin{aligned} & \left| \|J_k^T c_k\|^2 - \|\bar{J}_k^T \bar{c}_k\|^2 \mathbb{1}\{\|\bar{c}_k\| > \epsilon_o\} \right| \\ & \leq \max \left\{ \left| \|\bar{J}_k^T \bar{c}_k\|^2 - \|J_k^T c_k\|^2 \right|, \|J_k^T c_k\|^2 \mathbb{1}\{\|\bar{c}_k\| \leq \epsilon_o\} \right\} \\ & \leq \max \left\{ \|J_k^T c_k - \bar{J}_k^T \bar{c}_k\| ((\kappa_J + \epsilon_J)(\kappa_c + \epsilon_c) + \kappa_J \kappa_c), \kappa_J^2 \epsilon_o^2 \right\} \\ & \leq \max \left\{ ((\kappa_J + \epsilon_J)\epsilon_c + \kappa_c \epsilon_J) ((\kappa_J + \epsilon_J)(\kappa_c + \epsilon_c) + \kappa_J \kappa_c), \kappa_J^2 \epsilon_c^2 \right\} := \mathcal{E}_{J_c^2}, \end{aligned} \quad (3.22)$$

which completes the proof.  $\square$

Lemma 3.17 is useful for establishing upper bounds of the perturbation errors of inexact solutions to normal (2.2) and tangential (2.6) subproblems, as shown in Lemmas 3.18 and 3.25, respectively. By (3.21) and (3.22), when  $\epsilon_c = \epsilon_J = 0$ , it follows  $\mathcal{E}_{J_c} = \mathcal{E}_{J_c^2} = 0$ .

Let  $v_k$  denote the inexact solution to the deterministic trust region subproblem

$$\min_{v \in \text{Range}(J_k^T)} \frac{1}{2} \|c_k + J_k v\|^2 \quad \text{s.t.} \quad \|v\| \leq \sigma_{J_c} \|J_k^T c_k\| \quad (3.23)$$

that is required to satisfy the Cauchy-like decrease condition

$$\|c_k\| - \|c_k + J_k v_k\| \geq \gamma_c (\|c_k\| - \|c_k + \alpha_k^c J_k v_k^c\|), \quad (3.24)$$

where  $\gamma_c \in (0, 1]$  is a user-defined parameter,  $v_k^c := -J_k^T c_k$  is the negative gradient direction for the objective of (3.23) at  $v = 0$ , and  $\alpha_k^c = \min \left\{ \sigma_{Jc}, \frac{\|J_k^T c_k\|^2}{\|J_k J_k^T c_k\|^2} \right\}$  is the optimal step length along the direction of  $v_k^c$  that minimizes  $\|c_k + \alpha J_k v_k^c\|^2$  under the trust region constraint defined in (3.23) over  $\alpha \in \mathbb{R}$ . The next lemma bounds the error between the true and noisy Cauchy steps. In the ideal setting, our goal is to show that  $\|\bar{v}_k - v_k\|$  is sufficiently small for any inexact solutions  $\bar{v}_k$  and  $v_k$  that satisfy Cauchy-like decrease. However, the fact that the difference between Cauchy step and exact step for either true or noisy problem cannot be bounded by  $\epsilon$  terms under all conditions, prohibits us from being able to establish such a result. As a compromise, we provide bounds on the error between the true and noisy normal components in two extreme cases, where either both steps are Cauchy steps or both steps are the exact steps.

**Lemma 3.18.** *Suppose that Assumptions 2.1, 2.2, and 3.4 hold. In addition, suppose that the singular values of both  $J_k$  and  $\bar{J}_k$  are bounded below by  $\kappa_\sigma \in \mathbb{R}_{>0}$  for all  $k \in \mathbb{N}$ . Furthermore, suppose that  $\bar{v}_k = \bar{\alpha}_k^c \bar{v}_k^c$  is the Cauchy solution to (2.2) and  $v_k = \alpha_k^c v_k^c$  is the Cauchy solution to (3.23). Then, there exists  $\mathcal{E}_v \in \mathbb{R}_{\geq 0}$  such that  $\|\bar{v}_k - v_k\| \leq \mathcal{E}_v$ .*

*Proof.* When  $\|\bar{c}_k\| > \epsilon_o$ , we have  $\bar{v}_k^c = -\bar{J}_k^T \bar{c}_k$ . Let  $\bar{\alpha}_k^* = \frac{\|\bar{J}_k^T \bar{c}_k\|^2}{\|\bar{J}_k \bar{J}_k^T \bar{c}_k\|^2}$  and  $\alpha_k^* = \frac{\|J_k^T c_k\|^2}{\|J_k J_k^T c_k\|^2}$ . It follows from Lemma 3.17 that

$$\begin{aligned} \|J_k J_k^T c_k\| - \|\bar{J}_k \bar{J}_k^T \bar{c}_k\| &\leq \|J_k J_k^T c_k - \bar{J}_k \bar{J}_k^T \bar{c}_k\| \\ &\leq \|J_k J_k^T c_k - \bar{J}_k J_k^T c_k\| + \|\bar{J}_k J_k^T c_k - \bar{J}_k \bar{J}_k^T \bar{c}_k\| \\ &\leq \kappa_J \kappa_c \epsilon_J + (\kappa_J + \epsilon_J)((\kappa_J + \epsilon_J)\epsilon_c + \kappa_c \epsilon_J) := \mathcal{E}_{JJc}. \end{aligned} \quad (3.25)$$

Let us discuss the following four cases:

- (a)  $\bar{\alpha}_k^* \leq \sigma_{Jc}$  and  $\alpha_k^* \leq \sigma_{Jc}$ :  $|\bar{\alpha}_k^c - \alpha_k^c| = |\bar{\alpha}_k^* - \alpha_k^*|$ ,
- (b)  $\bar{\alpha}_k^* \leq \sigma_{Jc}$  and  $\alpha_k^* > \sigma_{Jc}$ :  $|\bar{\alpha}_k^c - \alpha_k^c| = |\bar{\alpha}_k^* - \sigma_{Jc}| = \sigma_{Jc} - \bar{\alpha}_k^* \leq \alpha_k^* - \bar{\alpha}_k^* = |\alpha_k^* - \bar{\alpha}_k^*|$ ,
- (c)  $\bar{\alpha}_k^* > \sigma_{Jc}$  and  $\alpha_k^* \leq \sigma_{Jc}$ :  $|\bar{\alpha}_k^c - \alpha_k^c| = |\alpha_k^* - \sigma_{Jc}| = \sigma_{Jc} - \alpha_k^* \leq \bar{\alpha}_k^* - \alpha_k^* = |\alpha_k^* - \bar{\alpha}_k^*|$ ,
- (d)  $\bar{\alpha}_k^* > \sigma_{Jc}$  and  $\alpha_k^* > \sigma_{Jc}$ :  $|\bar{\alpha}_k^c - \alpha_k^c| = |\sigma_{Jc} - \sigma_{Jc}| = 0 \leq |\alpha_k^* - \bar{\alpha}_k^*|$ ,

where  $\sigma_{Jc}$  is the constant in the trust region constraint in (2.2) and (3.23). In all cases,  $|\bar{\alpha}_k^c - \alpha_k^c| \leq |\alpha_k^* - \bar{\alpha}_k^*|$ , from which it follows that

$$\begin{aligned} &\|\alpha_k^c v_k^c - \bar{\alpha}_k^c \bar{v}_k^c\| \\ &\leq \|\alpha_k^c v_k^c - \bar{\alpha}_k^c v_k^c\| + \|\bar{\alpha}_k^c v_k^c - \bar{\alpha}_k^c \bar{v}_k^c\| \\ &\leq |\alpha_k^* - \bar{\alpha}_k^*| \|v_k^c\| + \|\bar{\alpha}_k^c v_k^c - \bar{\alpha}_k^c \bar{v}_k^c\| \\ &\leq \left| \frac{\|J_k^T c_k\|^2}{\|J_k J_k^T c_k\|^2} - \frac{\|\bar{J}_k^T \bar{c}_k\|^2}{\|\bar{J}_k \bar{J}_k^T \bar{c}_k\|^2} \right| \|J_k^T c_k\| + \sigma_{Jc} \|J_k^T c_k - \bar{J}_k^T \bar{c}_k\| \end{aligned}$$

$$\begin{aligned}
&= \left| \frac{\|J_k^T c_k\|^2 \|\bar{J}_k \bar{J}_k^T \bar{c}_k\|^2 - \|\bar{J}_k^T \bar{c}_k\|^2 \|\bar{J}_k \bar{J}_k^T \bar{c}_k\|^2 + \|\bar{J}_k^T \bar{c}_k\|^2 \|\bar{J}_k \bar{J}_k^T \bar{c}_k\|^2 - \|\bar{J}_k^T \bar{c}_k\|^2 \|J_k J_k^T c_k\|^2}{\|J_k J_k^T c_k\|^2 \|\bar{J}_k \bar{J}_k^T \bar{c}_k\|^2} \right| \|J_k^T c_k\| \\
&\quad + \sigma_{Jc} \|J_k^T c_k - \bar{J}_k^T \bar{c}_k\| \\
&\leq \frac{\|J_k^T c_k\|}{\|J_k J_k^T c_k\|^2} (\|\bar{J}_k^T \bar{c}_k\| + \|J_k^T c_k\|) (\|\bar{J}_k^T \bar{c}_k\| - \|J_k^T c_k\|) \\
&\quad + \frac{\|J_k^T c_k\|}{\|J_k J_k^T c_k\|^2} \frac{\|\bar{J}_k^T \bar{c}_k\|^2}{\|\bar{J}_k \bar{J}_k^T \bar{c}_k\|^2} (\|J_k J_k^T c_k\| + \|\bar{J}_k \bar{J}_k^T \bar{c}_k\|) (\|J_k J_k^T c_k\| - \|\bar{J}_k \bar{J}_k^T \bar{c}_k\|) \\
&\quad + \sigma_{Jc} \|J_k^T c_k - \bar{J}_k^T \bar{c}_k\| \\
&\leq \frac{\|J_k^T c_k\|}{\|J_k J_k^T c_k\|^2} (\|\bar{J}_k^T \bar{c}_k\| - \|J_k^T c_k\| + 2\|J_k^T c_k\|) \|\bar{J}_k^T \bar{c}_k - J_k^T c_k\| \\
&\quad + \frac{\|J_k^T c_k\|}{\|J_k J_k^T c_k\|^2} \kappa_\sigma^{-2} (\|\bar{J}_k \bar{J}_k^T \bar{c}_k\| - \|J_k J_k^T c_k\| + 2\|J_k J_k^T c_k\|) \|J_k J_k^T c_k - \bar{J}_k \bar{J}_k^T \bar{c}_k\| \\
&\quad + \sigma_{Jc} \|J_k^T c_k - \bar{J}_k^T \bar{c}_k\| \\
&\leq \frac{2\|J_k^T c_k\|^2}{\|J_k J_k^T c_k\|^2} \mathcal{E}_{Jc} + \frac{\|J_k^T c_k\|}{\|J_k J_k^T c_k\|^2} \mathcal{E}_{Jc}^2 + \frac{2\|J_k^T c_k\|^2}{\|J_k J_k^T c_k\|^2} \kappa_J \kappa_\sigma^{-2} \mathcal{E}_{JJc} + \frac{\|J_k^T c_k\|}{\|J_k J_k^T c_k\|^2} \kappa_\sigma^{-2} \mathcal{E}_{JJc}^2 + \sigma_{Jc} \mathcal{E}_{Jc},
\end{aligned}$$

where the second to last inequality follows from the imposed assumption about singular values of  $\bar{J}_k$ , and the last inequality follows from Lemma 3.17 and (3.25). When  $\|J_k J_k^T c_k\| > \max\{\mathcal{E}_{Jc}, \mathcal{E}_{JJc}\}$ ,

$$\begin{aligned}
&\|\alpha_k^c v_k^c - \bar{\alpha}_k^c \bar{v}_k^c\| \\
&\leq \frac{2\|J_k^T c_k\|^2}{\|J_k J_k^T c_k\|^2} \mathcal{E}_{Jc} + \frac{\|J_k^T c_k\|}{\|J_k J_k^T c_k\|} \mathcal{E}_{Jc} + \frac{2\|J_k^T c_k\|^2}{\|J_k J_k^T c_k\|^2} \kappa_J \kappa_\sigma^{-2} \mathcal{E}_{JJc} + \frac{\|J_k^T c_k\|}{\|J_k J_k^T c_k\|} \kappa_\sigma^{-2} \mathcal{E}_{JJc} + \sigma_{Jc} \mathcal{E}_{Jc} \\
&\leq 2\kappa_\sigma^{-2} \mathcal{E}_{Jc} + \kappa_\sigma^{-1} \mathcal{E}_{Jc} + 2\kappa_\sigma^{-4} \kappa_J \mathcal{E}_{JJc} + \kappa_\sigma^{-3} \mathcal{E}_{JJc} + \sigma_{Jc} \mathcal{E}_{Jc}.
\end{aligned}$$

Otherwise,  $\|J_k^T c_k\| \leq \kappa_\sigma^{-1} \|J_k J_k^T c_k\| \leq \kappa_\sigma^{-1} \max\{\mathcal{E}_{Jc}, \mathcal{E}_{JJc}\}$ , and it follows by Lemma 3.17

$$\begin{aligned}
\|\alpha_k^c v_k^c - \bar{\alpha}_k^c \bar{v}_k^c\| &\leq \max\{\alpha_k^*, \bar{\alpha}_k^*\} \|v_k^c\| + \|\bar{\alpha}_k^c v_k^c - \bar{\alpha}_k^c \bar{v}_k^c\| \\
&\leq \sigma_{Jc} \kappa_\sigma^{-1} \max\{\mathcal{E}_{Jc}, \mathcal{E}_{JJc}\} + \sigma_{Jc} \mathcal{E}_{Jc}.
\end{aligned}$$

If  $\|\bar{c}_k\| \leq \epsilon_o$ , then  $\bar{v}_k^c = 0$  and  $\|c_k\| \leq \|\bar{c}_k\| + \|c_k - \bar{c}_k\| \leq 2\epsilon_c$ , from which it follows that

$$\|\alpha_k^c v_k^c - \bar{\alpha}_k^c \bar{v}_k^c\| = \|\alpha_k^c J_k^T c_k\| \leq 2\sigma_{Jc} \kappa_J \epsilon_c.$$

The desired conclusion follows by defining

$$\mathcal{E}_v := \max\{(2\kappa_\sigma^{-2} + \kappa_\sigma^{-1} + \sigma_{Jc})\mathcal{E}_{Jc} + (2\kappa_\sigma^{-4} \kappa_J + \kappa_\sigma^{-3})\mathcal{E}_{JJc}, \sigma_{Jc} \kappa_\sigma^{-1} \max\{\mathcal{E}_{Jc}, \mathcal{E}_{JJc}\} + \sigma_{Jc} \mathcal{E}_{Jc}, 2\sigma_{Jc} \kappa_J \epsilon_c\}.$$

□

The next lemma provides an upper bound of the perturbation errors of the exact (optimal) solutions to (2.2) and (3.23).

**Lemma 3.19.** *Suppose that Assumptions 2.1, 2.2, and 3.4 hold. In addition, suppose that the singular values of both  $J_k$  and  $\bar{J}_k$  are bounded below by  $\kappa_\sigma \in \mathbb{R}_{>0}$  for all  $k \in \mathbb{N}$ . Furthermore, suppose that  $\bar{v}_k^*$  and  $v_k^*$  are the exact solutions to (2.2) and (3.23), respectively. Then, there exists  $\mathcal{E}_{v^*} \in \mathbb{R}_{\geq 0}$  such that  $\|\bar{v}_k^* - v_k^*\| \leq \mathcal{E}_{v^*}$ .*

*Proof.* When  $\|\bar{c}_k\| \leq \epsilon_o \leq \epsilon_c$ , by (2.2) and Assumption 2.1, it follows that  $\|\bar{v}_k^* - v_k^*\| \leq \|v_k^*\| \leq \sigma_{J_c} \|J_k^T c_k\| \leq \sigma_{J_c} \kappa_{J_c} \epsilon_c$ .

For the remainder of the proof we consider the case where  $\|\bar{c}_k\| > \epsilon_o$ . Let  $\bar{v}_k = \bar{J}_k^T \bar{s}_k$  and  $v_k = J_k^T s_k$  with vectors  $\{\bar{s}_k, s_k\} \subset \mathbb{R}^m$ . We have  $\|\bar{J}_k^T \bar{c}_k\| \neq 0$  since otherwise Algorithm 1 terminates on Line 11 and violates Assumption 3.4. When  $\|J_k^T c_k\| = 0$ , we have  $\|v_k\| = 0$  and  $\|\bar{J}_k^T \bar{c}_k\| \leq \mathcal{E}_{J_c}$  by Lemma 3.17, which implies that  $\|\bar{v}_k - v_k\| \leq \|\bar{v}_k\| \leq \sigma_{J_c} \mathcal{E}_{J_c}$  and the desired conclusion holds. We therefore assume that  $\|J_k^T c_k\| \neq 0$  in the remainder of the analysis. Using the vectors  $\{\bar{s}_k, s_k\} \subset \mathbb{R}^m$  we reformulate problems (2.2) and (3.23)

$$\min_{\bar{s} \in \mathbb{R}^m} \frac{1}{2} \|\bar{c}_k + \bar{J}_k \bar{J}_k^T \bar{s}\|^2 \quad \text{s.t.} \quad \|\bar{J}_k^T \bar{s}\| \leq \sigma_{J_c} \|\bar{J}_k^T \bar{c}_k\|, \quad (3.26)$$

$$\text{and} \quad \min_{s \in \mathbb{R}^m} \frac{1}{2} \|c_k + J_k J_k^T s\|^2 \quad \text{s.t.} \quad \|J_k^T s\| \leq \sigma_{J_c} \|J_k^T c_k\|. \quad (3.27)$$

Let  $\mathcal{C} = \{s \in \mathbb{R}^m \mid \|J_k^T s\| \leq \sigma_{J_c} \|J_k^T c_k\|\}$ , where  $\mathcal{C}$  is a nonempty, closed, convex set, and

$$\hat{s} = \frac{\|\bar{J}_k^T \bar{c}_k\|}{\|J_k^T c_k\|} (\bar{J}_k \bar{J}_k^T)^{-1/2} (J_k J_k^T)^{1/2} s. \quad (3.28)$$

Note that  $\bar{J}_k \bar{J}_k^T$  and  $J_k J_k^T$  are both invertible since  $\bar{J}_k$  and  $J_k$  are of full rank, as well as  $\|\bar{J}_k^T \bar{c}_k\| \neq 0$  and  $\|J_k^T c_k\| \neq 0$ . Then  $\hat{s}$  in (3.28) satisfies

$$\|\bar{J}_k^T \hat{s}\|^2 = \frac{\|\bar{J}_k^T \bar{c}_k\|^2}{\|J_k^T c_k\|^2} s^T (J_k J_k^T)^{1/2} (\bar{J}_k \bar{J}_k^T)^{-1/2} \bar{J}_k \bar{J}_k^T (\bar{J}_k \bar{J}_k^T)^{-1/2} (J_k J_k^T)^{1/2} s = \frac{\|\bar{J}_k^T \bar{c}_k\|^2}{\|J_k^T c_k\|^2} \|J_k^T s\|^2.$$

Thus, we can reformulate (3.27) by changing the variable from  $s$  to  $\hat{s}$  as

$$\min_{\hat{s} \in \mathbb{R}^m} \frac{1}{2} \left\| c_k + \frac{\|J_k^T c_k\|}{\|\bar{J}_k^T \bar{c}_k\|} (J_k J_k^T)^{1/2} (\bar{J}_k \bar{J}_k^T)^{1/2} \hat{s} \right\|^2 \quad \text{s.t.} \quad \|\bar{J}_k^T \hat{s}\| \leq \sigma_{J_c} \|\bar{J}_k^T \bar{c}_k\|. \quad (3.29)$$

Let us further reformulate (3.26) and (3.29) to (3.30) and (3.31), respectively, by rewriting the objective functions

$$\min_{\bar{s} \in \mathbb{R}^m} \frac{1}{2} \left\| \bar{s} + (\bar{J}_k \bar{J}_k^T)^{-1} \bar{c}_k \right\|_{\bar{J}_k \bar{J}_k^T \bar{J}_k \bar{J}_k^T}^2 \quad \text{s.t.} \quad \|\bar{J}_k^T \bar{s}\| \leq \sigma_{J_c} \|\bar{J}_k^T \bar{c}_k\|, \quad \text{and} \quad (3.30)$$

$$\min_{\hat{s} \in \mathbb{R}^m} \frac{\|J_k^T c_k\|^2}{2 \|J_k^T \bar{c}_k\|^2} \left\| \hat{s} + \frac{\|\bar{J}_k^T \bar{c}_k\|}{\|J_k^T c_k\|} (\bar{J}_k \bar{J}_k^T)^{-1/2} (J_k J_k^T)^{-1/2} c_k \right\|_{(\bar{J}_k \bar{J}_k^T)^{1/2} J_k J_k^T (\bar{J}_k \bar{J}_k^T)^{1/2}}^2 \quad \text{s.t.} \quad \|\bar{J}_k^T \hat{s}\| \leq \sigma_{J_c} \|\bar{J}_k^T \bar{c}_k\|. \quad (3.31)$$

Note that (3.30) and (3.31) share the same constraint. Let  $\bar{s}^*$  and  $\hat{s}^*$  be the optimal solutions of (3.30) and (3.31), respectively. It then follows from the projection theorem [16, Theorem 4.3-1] that  $\bar{s}^*$  and  $\hat{s}^*$  are both unique solutions to (3.30) and (3.31), and

$$\forall t \in \mathcal{C}, \quad \left\langle -(\bar{J}_k \bar{J}_k^T)^{-1} \bar{c}_k - \bar{s}^*, t - \bar{s}^* \right\rangle_{\bar{J}_k \bar{J}_k^T \bar{J}_k \bar{J}_k^T} \leq 0, \quad (3.32)$$

$$\forall t \in \mathcal{C}, \quad \left\langle \hat{s}^* + \frac{\|\bar{J}_k^T \bar{c}_k\|}{\|\bar{J}_k^T c_k\|} (\bar{J}_k \bar{J}_k^T)^{-1/2} (J_k J_k^T)^{-1/2} c_k, \hat{s}^* - t \right\rangle_{(\bar{J}_k \bar{J}_k^T)^{1/2} J_k J_k^T (\bar{J}_k \bar{J}_k^T)^{1/2}} \leq 0, \quad (3.33)$$

where  $\langle x, y \rangle_A = x^T A y$ , and  $x, y \in \mathbb{R}^n$  and  $A \in \mathbb{R}^{n \times n}$  is a positive definite matrix. Plugging  $\hat{s}^*$  and  $\bar{s}^*$  into (3.32) and (3.33), respectively,

$$\left\langle -(\bar{J}_k \bar{J}_k^T)^{-1} \bar{c}_k - \bar{s}^*, \hat{s}^* - \bar{s}^* \right\rangle_{\bar{J}_k \bar{J}_k^T \bar{J}_k \bar{J}_k^T} \leq 0, \quad (3.34)$$

$$\text{and} \quad \left\langle \hat{s}^* + \frac{\|\bar{J}_k^T \bar{c}_k\|}{\|\bar{J}_k^T c_k\|} (\bar{J}_k \bar{J}_k^T)^{-1/2} (J_k J_k^T)^{-1/2} c_k, \hat{s}^* - \bar{s}^* \right\rangle_{(\bar{J}_k \bar{J}_k^T)^{1/2} J_k J_k^T (\bar{J}_k \bar{J}_k^T)^{1/2}} \leq 0. \quad (3.35)$$

Summing (3.34) and (3.35), we have

$$\begin{aligned} & \left\langle -(\bar{J}_k \bar{J}_k^T)^{-1} \bar{c}_k - \bar{s}^* + \hat{s}^* + \frac{\|\bar{J}_k^T \bar{c}_k\|}{\|\bar{J}_k^T c_k\|} (\bar{J}_k \bar{J}_k^T)^{-1/2} (J_k J_k^T)^{-1/2} c_k, \hat{s}^* - \bar{s}^* \right\rangle_{\bar{J}_k \bar{J}_k^T \bar{J}_k \bar{J}_k^T} \\ & + \left( \hat{s}^* + \frac{\|\bar{J}_k^T \bar{c}_k\|}{\|\bar{J}_k^T c_k\|} (\bar{J}_k \bar{J}_k^T)^{-1/2} (J_k J_k^T)^{-1/2} c_k \right)^T \left( (\bar{J}_k \bar{J}_k^T)^{1/2} J_k J_k^T (\bar{J}_k \bar{J}_k^T)^{1/2} - \bar{J}_k \bar{J}_k^T \bar{J}_k \bar{J}_k^T \right) (\hat{s}^* - \bar{s}^*) \leq 0. \end{aligned}$$

Since the minimum singular value of  $\bar{J}_k$  is lower bounded by  $\kappa_\sigma$ , it follows that the minimum eigenvalue of  $(\bar{J}_k \bar{J}_k^T)^2$  is bounded below by  $\kappa_\sigma^4$ . Thus,

$$\begin{aligned} & \|\hat{s}^* - \bar{s}^*\|^2 \\ & \leq \kappa_\sigma^{-4} \|\hat{s}^* - \bar{s}^*\|_{\bar{J}_k \bar{J}_k^T \bar{J}_k \bar{J}_k^T}^2 \\ & \leq \kappa_\sigma^{-4} \left\langle (\bar{J}_k \bar{J}_k^T)^{-1} \bar{c}_k - \frac{\|\bar{J}_k^T \bar{c}_k\|}{\|\bar{J}_k^T c_k\|} (\bar{J}_k \bar{J}_k^T)^{-1/2} (J_k J_k^T)^{-1/2} c_k, \hat{s}^* - \bar{s}^* \right\rangle_{\bar{J}_k \bar{J}_k^T \bar{J}_k \bar{J}_k^T} \\ & \quad + \kappa_\sigma^{-4} \left( \hat{s}^* + \frac{\|\bar{J}_k^T \bar{c}_k\|}{\|\bar{J}_k^T c_k\|} (\bar{J}_k \bar{J}_k^T)^{-1/2} (J_k J_k^T)^{-1/2} c_k \right)^T \left( \bar{J}_k \bar{J}_k^T \bar{J}_k \bar{J}_k^T - (\bar{J}_k \bar{J}_k^T)^{1/2} J_k J_k^T (\bar{J}_k \bar{J}_k^T)^{1/2} \right) (\hat{s}^* - \bar{s}^*). \end{aligned} \quad (3.36)$$

Since the minimum singular value of  $\bar{J}_k$  and  $J_k$  are both bounded below by  $\kappa_\sigma$ ,  $\|(\bar{J}_k \bar{J}_k^T)^{-1/2}\| \leq \kappa_\sigma^{-1}$  and  $\|(J_k J_k^T)^{-1/2}\| \leq \kappa_\sigma^{-1}$ . It follows by [41, Lemma 2.2] (perturbation bound for matrix square root) and [43, Theorem 2.5] (perturbation bound for matrix inverse) that

$$\begin{aligned} \left\| (\bar{J}_k \bar{J}_k^T)^{-1/2} - (J_k J_k^T)^{-1/2} \right\| & \leq \left\| (\bar{J}_k \bar{J}_k^T)^{-1/2} \right\| \left\| (J_k J_k^T)^{-1/2} \right\| \left\| (\bar{J}_k \bar{J}_k^T)^{1/2} - (J_k J_k^T)^{1/2} \right\| \\ & \leq \kappa_\sigma^{-2} \frac{1}{\kappa_\sigma + \kappa_\sigma} \|\bar{J}_k \bar{J}_k^T - J_k J_k^T\| \leq \frac{(2\kappa_J + \epsilon_J)\epsilon_J}{2\kappa_\sigma^3}. \end{aligned} \quad (3.37)$$

Moreover, by (3.37), Lemma 3.17, and Assumptions 2.1 and 2.2, it follows that

$$\begin{aligned} & \left\| (\bar{J}_k \bar{J}_k^T)^{-1} \bar{c}_k - \frac{\|\bar{J}_k^T \bar{c}_k\|}{\|\bar{J}_k^T c_k\|} (\bar{J}_k \bar{J}_k^T)^{-1/2} (J_k J_k^T)^{-1/2} c_k \right\| \\ & \leq \left\| (\bar{J}_k \bar{J}_k^T)^{-1} \bar{c}_k - (\bar{J}_k \bar{J}_k^T)^{-1/2} (J_k J_k^T)^{-1/2} \bar{c}_k \right\| + \left\| (\bar{J}_k \bar{J}_k^T)^{-1/2} (J_k J_k^T)^{-1/2} (\bar{c}_k - c_k) \right\| \end{aligned}$$

$$\begin{aligned}
& + \left\| \frac{\|J_k^T c_k\| - \|\bar{J}_k^T \bar{c}_k\|}{\|J_k^T c_k\|} (\bar{J}_k \bar{J}_k^T)^{-1/2} (J_k J_k^T)^{-1/2} c_k \right\| \\
& \leq \left\| (\bar{J}_k \bar{J}_k^T)^{-1/2} \left( (\bar{J}_k \bar{J}_k^T)^{-1/2} - (J_k J_k^T)^{-1/2} \right) \bar{c}_k \right\| + \left\| (\bar{J}_k \bar{J}_k^T)^{-1/2} (J_k J_k^T)^{-1/2} (\bar{c}_k - c_k) \right\| \\
& \quad + \frac{\|J_k^T c_k\| - \|\bar{J}_k^T \bar{c}_k\|}{\|J_k^T c_k\|} \left\| (\bar{J}_k \bar{J}_k^T)^{-1/2} (J_k J_k^T)^{-1/2} c_k \right\| \\
& \leq \left\| (\bar{J}_k \bar{J}_k^T)^{-1/2} \right\| \left\| (\bar{J}_k \bar{J}_k^T)^{-1/2} - (J_k J_k^T)^{-1/2} \right\| \|\bar{c}_k\| + \left\| (\bar{J}_k \bar{J}_k^T)^{-1/2} \right\| \left\| (J_k J_k^T)^{-1/2} \right\| \|\bar{c}_k - c_k\| \\
& \quad + \left\| J_k^T c_k - \bar{J}_k^T \bar{c}_k \right\| \frac{\|c_k\|}{\|J_k^T c_k\|} \left\| (\bar{J}_k \bar{J}_k^T)^{-1/2} \right\| \left\| (J_k J_k^T)^{-1/2} \right\| \\
& \leq \frac{((2\kappa_J + \epsilon_J)(\kappa_c + \epsilon_c) + 2\kappa_\sigma \kappa_c) \epsilon_J + (2\kappa_\sigma^2 + 2\kappa_\sigma(\kappa_J + \epsilon_J)) \epsilon_c}{2\kappa_\sigma^4}. \tag{3.38}
\end{aligned}$$

The first term of inequality (3.36) (excluding constant) can be bounded by

$$\begin{aligned}
& \left\langle (\bar{J}_k \bar{J}_k^T)^{-1} \bar{c}_k - \frac{\|\bar{J}_k^T \bar{c}_k\|}{\|J_k^T c_k\|} (\bar{J}_k \bar{J}_k^T)^{-1/2} (J_k J_k^T)^{-1/2} c_k, \hat{s}^* - \bar{s}^* \right\rangle_{\bar{J}_k \bar{J}_k^T \bar{J}_k \bar{J}_k^T} \\
& \leq \frac{((2\kappa_J + \epsilon_J)(\kappa_c + \epsilon_c) + 2\kappa_\sigma \kappa_c) \epsilon_J + (2\kappa_\sigma^2 + 2\kappa_\sigma(\kappa_J + \epsilon_J)) \epsilon_c}{2\kappa_\sigma^4} (\kappa_J + \epsilon_J)^4 \|\hat{s}^* - \bar{s}^*\|.
\end{aligned}$$

Additionally, it follows by Assumptions 2.1 and 2.2 that

$$\begin{aligned}
\| \bar{J}_k \bar{J}_k^T \bar{J}_k \bar{J}_k^T - (\bar{J}_k \bar{J}_k^T)^{1/2} J_k J_k^T (\bar{J}_k \bar{J}_k^T)^{1/2} \| & \leq \| (\bar{J}_k \bar{J}_k^T)^{1/2} (\bar{J}_k \bar{J}_k^T - J_k J_k^T) (\bar{J}_k \bar{J}_k^T)^{1/2} \| \\
& \leq (\kappa_J + \epsilon_J)^2 (2\kappa_J + \epsilon_J) \epsilon_J.
\end{aligned}$$

The second term of inequality (3.36) (excluding constant) can be bounded by

$$\begin{aligned}
& \left( \hat{s}^* + \frac{\|\bar{J}_k^T \bar{c}_k\|}{\|J_k^T c_k\|} (\bar{J}_k \bar{J}_k^T)^{-\frac{1}{2}} (J_k J_k^T)^{-\frac{1}{2}} c_k \right)^T \left( \bar{J}_k \bar{J}_k^T \bar{J}_k \bar{J}_k^T - (\bar{J}_k \bar{J}_k^T)^{\frac{1}{2}} J_k J_k^T (\bar{J}_k \bar{J}_k^T)^{\frac{1}{2}} \right) (\hat{s}^* - \bar{s}^*) \\
& \leq \left( \|\hat{s}^*\| + \frac{\|\bar{J}_k^T \bar{c}_k\|}{\|J_k^T c_k\|} \left\| (\bar{J}_k \bar{J}_k^T)^{-\frac{1}{2}} (J_k J_k^T)^{-\frac{1}{2}} c_k \right\| \right) (\kappa_J + \epsilon_J)^2 (2\kappa_J + \epsilon_J) \epsilon_J \|\hat{s}^* - \bar{s}^*\| \\
& \leq \left( \frac{\|\hat{s}^*\|}{\|\bar{J}_k^T \hat{s}^*\|} \|\bar{J}_k^T \hat{s}^*\| + \|\bar{J}_k^T \bar{c}_k\| \frac{\|c_k\|}{\|J_k^T c_k\|} \left\| (\bar{J}_k \bar{J}_k^T)^{-\frac{1}{2}} \right\| \left\| (J_k J_k^T)^{-\frac{1}{2}} \right\| \right) (\kappa_J + \epsilon_J)^2 (2\kappa_J + \epsilon_J) \epsilon_J \|\hat{s}^* - \bar{s}^*\| \\
& \leq (\kappa_\sigma^{-1} \sigma_{J_c} + \kappa_\sigma^{-3}) (\kappa_c + \epsilon_c) (\kappa_J + \epsilon_J)^3 (2\kappa_J + \epsilon_J) \epsilon_J \|\hat{s}^* - \bar{s}^*\|. \tag{3.39}
\end{aligned}$$

As a result, by plugging (3.38) and (3.39) into (3.36), it follows that

$$\begin{aligned}
\|\hat{s}^* - \bar{s}^*\| & \leq (\kappa_J + \epsilon_J)^4 \frac{((2\kappa_J + \epsilon_J)(\kappa_c + \epsilon_c) + 2\kappa_\sigma \kappa_c) \epsilon_J + (2\kappa_\sigma^2 + 2\kappa_\sigma(\kappa_J + \epsilon_J)) \epsilon_c}{2\kappa_\sigma^8} \\
& \quad + \kappa_\sigma^{-4} (\kappa_\sigma^{-1} \sigma_{J_c} + \kappa_\sigma^{-3}) (\kappa_c + \epsilon_c) (\kappa_J + \epsilon_J)^3 (2\kappa_J + \epsilon_J) \epsilon_J := \mathcal{E}_{\bar{s}}. \tag{3.40}
\end{aligned}$$

By Assumption 2.1, (3.27), (3.28), (3.37), and the fact that all singular values of  $\bar{J}_k$  are



bounded below by  $\kappa_\sigma$ ,

$$\begin{aligned}
\|\hat{s}^* - s^*\| &= \left\| \frac{\|\bar{J}_k^T \bar{c}_k\|}{\|J_k^T c_k\|} (\bar{J}_k \bar{J}_k^T)^{-1/2} (J_k J_k^T)^{1/2} s^* - s^* \right\| \\
&\leq \frac{\|J_k J_k^T\|^{1/2}}{\|J_k^T c_k\|} \left\| \|\bar{J}_k^T \bar{c}_k\| (\bar{J}_k \bar{J}_k^T)^{-1/2} - \|J_k^T c_k\| (J_k J_k^T)^{-1/2} \right\| \frac{\|s^*\|}{\|J_k^T s^*\|} \|J_k^T s^*\| \\
&\leq \kappa_J \sigma_{Jc} \kappa_\sigma^{-1} \left\| \|\bar{J}_k^T \bar{c}_k - J_k^T c_k\| (\bar{J}_k \bar{J}_k^T)^{-1/2} + \|J_k^T c_k\| \left( (\bar{J}_k \bar{J}_k^T)^{-1/2} - (J_k J_k^T)^{-1/2} \right) \right\| \\
&\leq \kappa_J \sigma_{Jc} \kappa_\sigma^{-1} \left( \|\bar{J}_k^T \bar{c}_k - J_k^T c_k\| \left\| (\bar{J}_k \bar{J}_k^T)^{-1/2} \right\| + \|J_k^T c_k\| \left\| (\bar{J}_k \bar{J}_k^T)^{-1/2} - (J_k J_k^T)^{-1/2} \right\| \right) \\
&\leq \kappa_J \sigma_{Jc} \kappa_\sigma^{-1} \left( \kappa_\sigma^{-1} (\kappa_c \epsilon_J + (\kappa_J + \epsilon_J) \epsilon_c) + \frac{\kappa_J \kappa_c (2\kappa_J + \epsilon_J) \epsilon_J}{2\kappa_\sigma^3} \right) := \mathcal{E}_s.
\end{aligned} \tag{3.41}$$

It then follows that  $\|\bar{s}^* - s^*\| \leq \|\hat{s}^* - s^*\| + \|\hat{s}^* - \bar{s}^*\| \leq \mathcal{E}_s + \mathcal{E}_{\bar{s}}$ . By Assumption 2.1, 2.2, (3.41), and the fact that  $\bar{v}_k^* = \bar{J}_k^T \bar{s}_k^*$  and  $v_k^* = J_k^T s_k^*$ ,

$$\|\bar{v}_k^* - v_k^*\| \leq \|\bar{J}_k^T \bar{s}^* - \bar{J}_k^T s^*\| + \|\bar{J}_k^T s^* - J_k^T s^*\| \leq (\kappa_J + \epsilon_J) (\mathcal{E}_s + \mathcal{E}_{\bar{s}}) + \sigma_{Jc} \kappa_\sigma^{-1} \kappa_J \kappa_c \epsilon_J.$$

Defining  $\mathcal{E}_{v^*} := \max \{ (\kappa_J + \epsilon_J) (\mathcal{E}_s + \mathcal{E}_{\bar{s}}) + \sigma_{Jc} \kappa_\sigma^{-1} \kappa_J \kappa_c \epsilon_J, \sigma_{Jc} \kappa_J \epsilon_c \}$  concludes the proof.  $\square$

**Remark 3.20.** Lemma 3.18 shows that when the LICQ holds for both true and noisy constraint Jacobians, Cauchy solutions to problems (2.2) and (3.23) are sufficiently close. Lemma 3.19 shows that a similar result holds for the optimal solutions of problems (2.2) and (3.23). It can be seen from the definitions of  $\mathcal{E}_v$  and  $\mathcal{E}_{v^*}$  that both perturbation errors depend on  $\epsilon_c$  and  $\epsilon_J$  linearly. Moreover, when  $\epsilon_c = \epsilon_J = 0$ , both errors are zero, i.e.,  $\mathcal{E}_v = \mathcal{E}_{v^*} = 0$ . The analysis in Lemmas 3.18 and 3.19 relies critically on the LICQ. Without it, one cannot establish bounds for the ratios  $\|\bar{J}_k^T \bar{c}_k\| / \|\bar{J}_k \bar{J}_k^T \bar{c}_k\|$  and  $\|J_k^T c_k\| / \|J_k J_k^T c_k\|$  (proof of Lemma 3.18), and the non-uniqueness of optimal solutions to (2.2) makes it challenging to derive the results in Lemma 3.19. We note that Lemmas 3.18 and 3.19 are not used to derive the main convergence guarantees (Section 3.4).

As we will see in Section 3.4, our main interest is the distance between the true and the noisy tangential components. Let  $u_k^*$  denote the exact solution to the deterministic version of the linear system (2.5), i.e.,

$$\begin{bmatrix} H_k & J_k^T \\ J_k & 0 \end{bmatrix} \begin{bmatrix} u_k^* \\ y_k^* \end{bmatrix} = - \begin{bmatrix} g_k + H_k v_k \\ 0 \end{bmatrix}, \tag{3.42}$$

where  $y_k^* \in \mathbb{R}^m$  is the vector of dual multipliers. We aim to derive an upper bound for  $\|u_k^* - \bar{u}_k\|$ , the difference between the exact solution to the deterministic linear system (3.42) and an inexact solution to the noisy linear system (2.5). The following results hold regardless of whether the LICQ holds for either the noisy or deterministic problems. First present a technical lemma that bounds the true and noisy tangential steps, the true search direction, and the dual multipliers.

**Lemma 3.21.** *Suppose that Assumptions 2.1, 2.2, 2.4, and 3.4 hold. In addition, suppose that the minimum non-zero singular values of both  $J_k$  and  $\bar{J}_k$  are bounded below by  $\kappa_\sigma \in \mathbb{R}_{>0}$  for all  $k \in \mathbb{N}$ . Then, for all  $k \in \mathbb{N}$ , there exist  $\{\kappa_{\bar{u}^*}, \kappa_{u^*}, \kappa_{d^*}\} \subset \mathbb{R}_{>0}$  such that  $\|\bar{u}_k^*\| \leq \kappa_{\bar{u}^*}$ ,  $\|u_k^*\| \leq \kappa_{u^*}$  and  $\|d_k^*\| \leq \kappa_{d^*}$ , where  $d_k^* = u_k^* + v_k$ , where  $v_k$  is the inexact solution to (3.23) satisfying (3.24). Moreover, for all  $k \in \mathbb{N}$ , there exist  $\{\kappa_{\bar{y}^*}, \kappa_{y^*}\} \subset \mathbb{R}_{>0}$  and solutions  $\{\bar{y}_k^*, y_k^*\} \subset \mathbb{R}^m$  from linear systems (2.6) and (3.42) such that  $\|\bar{y}_k^*\| \leq \kappa_{\bar{y}^*}$  and  $\|y_k^*\| \leq \kappa_{y^*}$ .*

*Proof.* Suppose that  $\text{rank}(\bar{J}_k) = p$  and  $\text{rank}(J_k) = q$ . First, let  $p = n \leq m$ , in which case the null space of  $\bar{J}_k$  is trivial (contains only the zero vector). By (2.5), it follows that  $\bar{u}_k^* = 0$  and  $\bar{y}_k^* = (\bar{J}_k^T \bar{J}_k)^{-1} \bar{J}_k^T (\bar{g}_k + \bar{H}_k \bar{v}_k)$ , where the latter is bounded by Assumptions 2.1 and 2.2. Similarly, when  $q = n \leq m$ , it follows that  $u_k^* = 0$  and  $y_k^* = (J_k^T J_k)^{-1} J_k^T (g_k + H_k v_k)$ , where the latter is bounded by Assumptions 2.1 and 2.2. The desired conclusion therefore holds for  $p = n \leq m$  or  $q = n \leq m$ .

Hence, for the remainder of the proof, we consider  $\text{rank}(\bar{J}_k) = p < n$  and  $\text{rank}(J_k) = q < n$ . Suppose that  $\Sigma_k$  are diagonal matrices containing the non-zero singular values of  $J_k$ . Similar to (3.13), it follows that

$$J_k^T = (U_k \ Z_k) \begin{pmatrix} \Sigma_k & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_k \\ V_k^\perp \end{pmatrix} = U_k \Sigma_k V_k,$$

where  $Z_k$  is the orthonormal basis of null space of  $J_k$ . By the linear systems (2.5) and (3.42),

$$\bar{u}_k^* = -\bar{Z}_k (\bar{Z}_k^T H_k \bar{Z}_k)^{-1} \bar{Z}_k^T (\bar{g}_k + H_k \bar{v}_k) \quad (3.43)$$

$$\text{and} \quad u_k^* = -Z_k (Z_k^T H_k Z_k)^{-1} Z_k^T (g_k + H_k v_k), \quad (3.44)$$

where  $\bar{Z}_k$  is defined in (3.13). In addition, it follows by Assumption 2.1 that

$$\begin{aligned} \|\bar{u}_k^*\| &\leq \zeta^{-1} (\kappa_g + \epsilon_g + \kappa_H \|\bar{v}_k\|) \leq \zeta^{-1} (\kappa_g + \epsilon_g + \kappa_H \sigma_{J_c} \|\bar{J}_k^T \bar{c}_k\|) \\ &\leq \zeta^{-1} (\kappa_g + \epsilon_g + \kappa_H \sigma_{J_c} (\kappa_J + \epsilon_J) (\kappa_c + \epsilon_c)) := \kappa_{\bar{u}^*} \end{aligned} \quad (3.45)$$

$$\text{and} \quad \|u_k^*\| \leq \zeta^{-1} (\kappa_g + \kappa_H \sigma_{J_c} \|J_k^T c_k\|) \leq \zeta^{-1} (\kappa_g + \kappa_H \sigma_{J_c} \kappa_J \kappa_c) := \kappa_{u^*}. \quad (3.46)$$

Given that  $\|v_k\| \leq \sigma_{J_c} \|J_k^T c_k\| \leq \sigma_{J_c} \kappa_J \kappa_c$  ((3.23) and Assumption 2.1) and (3.46), there exists  $\kappa_{d^*} \in \mathbb{R}_{>0}$  such that  $\|d_k^*\| \leq \kappa_{d^*}$ . Let  $\bar{V}_k^\dagger = \bar{V}_k^T (\bar{V}_k \bar{V}_k^T)^{-1}$  denote the Moore-Penrose pseudo-inverse of  $\bar{V}_k$  (defined in (3.13)) and  $V_k^\dagger = V_k^T (V_k V_k^T)^{-1}$ . Although the Lagrange multipliers are not unique for the linear systems (2.5) and (3.42),

$$\bar{y}_k^* = -\bar{V}_k^\dagger \left( \tilde{J}_k \tilde{J}_k^T \right)^{-1} \tilde{J}_k^T (\bar{g}_k + H_k (\bar{u}_k^* + \bar{v}_k)), \quad (3.47)$$

$$\text{and} \quad y_k^* = -V_k^\dagger \left( \tilde{J}_k \tilde{J}_k^T \right)^{-1} \tilde{J}_k^T (g_k + H_k (u_k^* + v_k)) \quad (3.48)$$

are one of the solutions to the linear systems (2.5) and (3.42), respectively. Suppose that  $\|\bar{V}_k^\dagger\| \leq \kappa_{\bar{V}^\dagger}$  and  $\|V_k^\dagger\| \leq \kappa_{V^\dagger}$ , it then follows by Assumptions 2.1 and 2.2, (3.45), (3.46), (3.47), and (3.48) that

$$\begin{aligned} \|\bar{y}_k^*\| &\leq \kappa_{\bar{V}^\dagger} \kappa_\sigma^{-2} (\kappa_J + \epsilon_J) (\kappa_g + \epsilon_g + \kappa_H (\kappa_{\bar{u}^*} + \sigma_{Jc} (\kappa_J + \epsilon_J) (\kappa_c + \epsilon_c))) \\ \text{and } \|y_k^*\| &\leq \kappa_{V^\dagger} \kappa_\sigma^{-2} \kappa_J (\kappa_g + \kappa_H (\kappa_{u^*} + \sigma_{Jc} \kappa_J \kappa_c)), \end{aligned}$$

which concludes the proof.  $\square$

Lemma 3.21 provides upper bounds for the tangential components, the search directions, and the dual multipliers required in the analysis. The next lemma (Lemma 3.25) shows that the distance between inexact noisy solution  $\bar{u}_k$  and optimal noisy solution  $\bar{u}_k^*$  depends on the norm of residuals ( $\|\bar{\rho}_k\|$  and  $\|\bar{r}_k\|$ ) and the errors in the approximations. Moreover, it provides an upper bound on the distance between the noisy tangential component  $\bar{u}_k$  in Algorithm 1 and the true exact tangential component  $u_k^*$ . Before we proceed we make an additional technical assumption about the null spaces of the true and noisy constraint Jacobians. Under a similar noise setting to that considered in this paper, in [45], the authors show that the true stationarity error measure is sufficiently small under the assumption that  $\|Z_k - \bar{Z}_k\|$  is sufficiently small. Instead, we directly analyze the error in the tangential components under the following assumption regarding the orthogonal basis of null spaces.

**Assumption 3.22.** *The null spaces of  $J_k$  and  $\bar{J}_k$  are nontrivial, i.e., both  $Z_k$  and  $\bar{Z}_k$  exist.*

**Remark 3.23.** *The null space of  $J_k$  (resp.,  $\bar{J}_k$ ) is trivial if and only if  $n \leq m$  and  $\text{rank}(J_k) = n$  (resp.,  $\text{rank}(\bar{J}_k) = n$ ). Assumption 3.22 is therefore equivalent to requiring that when  $n \leq m$ ,  $\text{rank}(\bar{J}_k) < n$  and  $\text{rank}(J_k) < n$ . This assumption is critical since our stationarity measures involve  $Z_k$  (resp.,  $\bar{Z}_k$ ), the null space of  $J_k$  (resp.,  $\bar{J}_k$ ). Assumption 3.22 is common in the literature; see, e.g., [19, 45]. Note that [45] additionally assumes that  $Z_k$  and  $\bar{Z}_k$  have the same dimension (otherwise  $\|Z_k - \bar{Z}_k\|$  is not computable), whereas we do not impose this restriction.*

When  $\text{rank}(\bar{J}_k) = \text{rank}(J_k) = n \leq m$ , the null spaces of  $J_k$  and  $\bar{J}_k$  are both trivial. In this case,  $u_k^* = 0$  and  $\bar{u}_k^* = 0$ . We use the following counterexample to show that when the null space of  $\bar{J}_k$  is trivial, the difference between  $u_k^*$  and  $\bar{u}_k^*$  can be significant.

**Example 3.24.** *Let  $n = m = 2$ , and set  $H_k = I$ . Consider the matrices*

$$J_k = \begin{bmatrix} 1.0 & 2.0 \\ 0.5 & 1.0 \end{bmatrix}, \quad \bar{J}_k = \begin{bmatrix} 1.0 & 2.0 \\ 0.5 & 1.0 + \epsilon_J \end{bmatrix}$$

*within the context of equations (2.5) and (3.42). By construction,  $J_k$  is rank-deficient, whereas  $\bar{J}_k$  has full rank, with the perturbation satisfying  $\|J_k - \bar{J}_k\| = \epsilon_J$ . Next, let  $g_k +$*

$H_k v_k = g_k + v_k = [z_1, z_2]^T$  for some  $(z_1, z_2) \in \mathbb{R}^2$ . By (3.44), it follows that

$$u_k^* = \begin{bmatrix} -0.8z_1 + 0.4z_2 \\ z_1 + 0.4z_2 \end{bmatrix}.$$

Conversely, since the null space of  $\bar{J}_k$  contains only the zero vector and is therefore independent of  $\bar{g}_k + H_k \bar{v}_k$ , we obtain  $\bar{u}_k^* = 0$ . Since  $u_k^*$  can attain arbitrarily large values, it is evident that the discrepancy between  $u_k^*$  and  $\bar{u}_k^*$  can be significant, and cannot be controlled by  $\epsilon_J$ .

**Lemma 3.25.** *Suppose that Assumptions 2.1, 2.2, 2.4, 3.1, 3.2, 3.4, and 3.22 hold. In addition, suppose that the minimum non-zero singular values of both  $J_k$  and  $\bar{J}_k$  are bounded below by  $\kappa_\sigma \in \mathbb{R}_{>0}$  for all  $k \in \mathbb{N}$ . Then, for all  $k \in \mathbb{N}$*

$$\begin{aligned} \|u_k^* - \bar{u}_k\|^2 &\leq 6\zeta^{-2} (\epsilon_g^2 + \kappa_{\bar{y}^*}^2 \epsilon_J^2 + \kappa_{y^*}^2 \epsilon_J^2) + 2\lambda_{pr}^2 (\zeta^{-1} + (\kappa_J + \epsilon_J) \kappa_\sigma^{-2})^2 \max\{\|\bar{u}_k\|^2, \|\bar{J}_k^T \bar{c}_k\|^2\} \\ &\quad + 12\zeta^{-2} \kappa_H^2 \sigma_{Jc}^2 (\|\bar{J}_k^T \bar{c}_k\|^2 \mathbb{1}\{\|\bar{c}_k\| > \epsilon_o\} + \|J_k^T c_k\|^2). \end{aligned}$$

*Proof.* By Assumption 3.22, both  $Z_k$  and  $\bar{Z}_k$  exist. It follows from [21, equation (37)] that  $\|I - \bar{Z}_k (\bar{Z}_k^T H_k \bar{Z}_k)^{-1} \bar{Z}_k^T H_k\| \leq 1$ . Since  $\bar{U}_k, \bar{V}_k$  (defined in (3.13)) are both submatrices of orthogonal matrices,  $\|\bar{U}_k\| \leq 1, \|\bar{V}_k\| \leq 1$ . Additionally, it follows by  $\|(\tilde{J}_k \tilde{J}_k^T)^{-1}\| = \|\bar{\Sigma}_k^{-2}\| \leq \kappa_\sigma^{-2}$ , Assumptions 2.2 and 2.4, and (3.14), (3.44) that

$$\begin{aligned} \|\bar{u}_k^* - \bar{u}_k\| &\leq \|\bar{Z}_k (\bar{Z}_k^T H_k \bar{Z}_k)^{-1} \bar{Z}_k^T \bar{\rho}_k\| + \|(I - \bar{Z}_k (\bar{Z}_k^T H_k \bar{Z}_k)^{-1} \bar{Z}_k^T H_k) \bar{U}_k \bar{\Sigma}_k (\tilde{J}_k \tilde{J}_k^T)^{-1} \bar{V}_k \bar{r}_k\| \\ &\leq \zeta^{-1} \|\bar{\rho}_k\| + (\kappa_J + \epsilon_J) \kappa_\sigma^{-2} \|\bar{r}_k\|. \end{aligned}$$

Let  $C_k$  denote the orthonormal basis of  $\text{Null}(\bar{J}_k) \cap \text{Null}(J_k)$ . Then, one can construct the orthonormal basis of  $\text{Null}(\bar{J}_k)$  and denote it as  $\bar{Z}_k = [C_k, \bar{Z}_k^c] \in \mathbb{R}^{n \times (n-p)}$ . Similarly, one can construct the orthonormal basis of  $\text{Null}(J_k)$  and denote it as  $Z_k = [Z_k^c, C_k] \in \mathbb{R}^{n \times (n-q)}$ . Suppose that  $\bar{u}_k^* = \bar{Z}_k \bar{w}_k$  and  $u_k^* = Z_k w_k$  where  $\bar{w}_k \in \mathbb{R}^{n-p}$  and  $w_k \in \mathbb{R}^{n-q}$ , respectively. By (2.5) and (3.42)

$$\begin{aligned} H_k [Z_k^c \ C_k \ \bar{Z}_k^c] \begin{bmatrix} 0 \\ \bar{w}_k \end{bmatrix} &= H_k [Z_k^c \ \bar{Z}_k] \begin{bmatrix} 0 \\ \bar{w}_k \end{bmatrix} = H_k \bar{Z}_k \bar{w}_k \\ &= H_k \bar{u}_k^* = -(\bar{g}_k + H_k \bar{v}_k + \bar{J}_k^T \bar{y}_k^*) \\ \text{and} \quad H_k [Z_k^c \ C_k \ \bar{Z}_k^c] \begin{bmatrix} w_k \\ 0 \end{bmatrix} &= H_k [Z_k \ \bar{Z}_k^c] \begin{bmatrix} w_k \\ 0 \end{bmatrix} = H_k Z_k w_k \\ &= H_k u_k^* = -(g_k + H_k v_k + J_k^T y_k^*). \end{aligned}$$

Multiplying  $[Z_k^c \ C_k \ \bar{Z}_k^c]^T$  to both sides, it follows from  $J_k Z_k = 0$  and  $\bar{J}_k \bar{Z}_k = 0$  that

$$[Z_k^c \ C_k \ \bar{Z}_k^c]^T H_k [Z_k^c \ C_k \ \bar{Z}_k^c] \begin{bmatrix} 0 \\ \bar{w}_k \end{bmatrix} = -[Z_k^c \ C_k \ \bar{Z}_k^c]^T (\bar{g}_k + H_k \bar{v}_k + \bar{J}_k^T \bar{y}_k^*)$$

$$= - \begin{bmatrix} Z_k^T (\bar{g}_k + H_k \bar{v}_k + (\bar{J}_k - J_k)^T \bar{y}_k^*) \\ \bar{Z}_k^{cT} (\bar{g}_k + H_k \bar{v}_k) \end{bmatrix}, \quad (3.49)$$

$$\begin{aligned} \text{and} \quad [Z_k^c \ C_k \ \bar{Z}_k^c]^T H_k [Z_k^c \ C_k \ \bar{Z}_k^c] \begin{bmatrix} w_k \\ 0 \end{bmatrix} &= -[Z_k^c \ C_k \ \bar{Z}_k^c]^T (g_k + H_k v_k + J_k^T y_k^*) \\ &= - \begin{bmatrix} Z_k^T (g_k + H_k v_k) \\ \bar{Z}_k^{cT} (g_k + H_k v_k + (\bar{J}_k - J_k)^T y_k^*) \end{bmatrix}. \end{aligned} \quad (3.50)$$

Since  $[Z_k^c \ C_k \ \bar{Z}_k^c]$  is an orthonormal basis of  $\text{Null}(J_k) \cup \text{Null}(\bar{J}_k)$ , it follows by Assumption 2.4 that  $\|([Z_k^c \ C_k \ \bar{Z}_k^c]^T H_k [Z_k^c \ C_k \ \bar{Z}_k^c])^{-1}\| \leq \zeta^{-1}$ . Thus,

$$\begin{aligned} \|u_k^* - \bar{u}_k^*\|^2 &= \|Z_k w_k - \bar{Z}_k \bar{w}_k\|^2 \\ &= \left\| [Z_k^c \ C_k \ \bar{Z}_k^c] \begin{bmatrix} w_k \\ 0 \end{bmatrix} - [Z_k^c \ C_k \ \bar{Z}_k^c] \begin{bmatrix} 0 \\ \bar{w}_k \end{bmatrix} \right\|^2 \\ &\leq \left\| [Z_k^c \ C_k \ \bar{Z}_k^c] ([Z_k^c \ C_k \ \bar{Z}_k^c]^T H_k [Z_k^c \ C_k \ \bar{Z}_k^c])^{-1} \begin{bmatrix} Z_k^T (\bar{g}_k + H_k \bar{v}_k + (\bar{J}_k - J_k)^T \bar{y}_k^*) \\ \bar{Z}_k^{cT} (\bar{g}_k + H_k \bar{v}_k) \end{bmatrix} \right. \\ &\quad \left. - [Z_k^c \ C_k \ \bar{Z}_k^c] ([Z_k^c \ C_k \ \bar{Z}_k^c]^T H_k [Z_k^c \ C_k \ \bar{Z}_k^c])^{-1} \begin{bmatrix} Z_k^T (g_k + H_k v_k) \\ \bar{Z}_k^{cT} (g_k + H_k v_k + (\bar{J}_k - J_k)^T y_k^*) \end{bmatrix} \right\|^2 \\ &= \left\| [Z_k^c \ C_k \ \bar{Z}_k^c] ([Z_k^c \ C_k \ \bar{Z}_k^c]^T H_k [Z_k^c \ C_k \ \bar{Z}_k^c])^{-1} \begin{bmatrix} Z_k^T \\ \bar{Z}_k^{cT} \end{bmatrix} ((\bar{g}_k - g_k) + H_k (\bar{v}_k - v_k)) \right. \\ &\quad \left. + [Z_k^c \ C_k \ \bar{Z}_k^c] ([Z_k^c \ C_k \ \bar{Z}_k^c]^T H_k [Z_k^c \ C_k \ \bar{Z}_k^c])^{-1} \begin{bmatrix} Z_k^T ((\bar{J}_k - J_k)^T \bar{y}_k^*) \\ -\bar{Z}_k^{cT} ((\bar{J}_k - J_k)^T y_k^*) \end{bmatrix} \right\|^2 \\ &\leq 3\zeta^{-2} \left( \|\bar{g}_k - g_k\|^2 + \|H_k (\bar{v}_k - v_k)\|^2 + \left\| \begin{bmatrix} Z_k^T ((\bar{J}_k - J_k)^T \bar{y}_k^*) \\ -\bar{Z}_k^{cT} ((\bar{J}_k - J_k)^T y_k^*) \end{bmatrix} \right\|^2 \right) \\ &\leq 3\zeta^{-2} (\epsilon_g^2 + \kappa_H^2 \|\bar{v}_k - v_k\|^2 + \kappa_{\bar{y}^*}^2 \epsilon_J^2 + \kappa_{y^*}^2 \epsilon_J^2) \\ &\leq 3\zeta^{-2} (\epsilon_g^2 + \kappa_{\bar{y}^*}^2 \epsilon_J^2 + \kappa_{y^*}^2 \epsilon_J^2) + 6\zeta^{-2} \kappa_H^2 \sigma_{J_c}^2 (\|\bar{J}_k^T \bar{c}_k\|^2 \mathbb{1}\{\|\bar{c}_k\| > \epsilon_o\} + \|J_k^T c_k\|^2), \end{aligned} \quad (3.51)$$

where the first inequality follows multiplying  $([Z_k^c \ C_k \ \bar{Z}_k^c]^T H_k [Z_k^c \ C_k \ \bar{Z}_k^c])^{-1}$  to both sides of (3.49) and (3.50), the second inequality follows by the orthonormality of  $[Z_k^c \ C_k \ \bar{Z}_k^c]$ , Assumption 2.4, and the fact that  $\|a_1 + a_2 + a_3\|^2 \leq 3(\|a_1\|^2 + \|a_2\|^2 + \|a_3\|^2)$ , the third inequality follows by Assumptions 2.2 and 2.4, Lemma 3.21, and the last inequality follows by (2.2) and (3.23). It then follows by Young's inequality that

$$\begin{aligned} \|u_k^* - \bar{u}_k\|^2 &\leq 2\|u_k^* - \bar{u}_k^*\|^2 + 2\|\bar{u}_k^* - \bar{u}_k\|^2 \\ &\leq 6\zeta^{-2} (\epsilon_g^2 + \kappa_{\bar{y}^*}^2 \epsilon_J^2 + \kappa_{y^*}^2 \epsilon_J^2) + 12\zeta^{-2} \kappa_H^2 \sigma_{J_c}^2 (\|\bar{J}_k^T \bar{c}_k\|^2 \mathbb{1}\{\|\bar{c}_k\| > \epsilon_o\} + \|J_k^T c_k\|^2) \\ &\quad + 2(\zeta^{-1} \|\bar{\rho}_k\| + (\kappa_J + \epsilon_J) \kappa_\sigma^{-2} \|\bar{r}_k\|)^2. \end{aligned}$$

The desired conclusion (second statement) then follows by (2.9).  $\square$

**Remark 3.26.** Lemma 3.25 shows that the upper bound of  $\|u_k^* - \bar{u}_k\|^2$  depends on noise levels  $\epsilon_g, \epsilon_J$ , residual norm of problem (2.6) (which is bounded by  $\max\{\|\bar{u}_k\|^2, \|\bar{J}_k^T \bar{c}_k\|^2\}$ ), and the infeasible stationarity errors  $\|\bar{J}_k^T \bar{c}_k\|^2$  and  $\|J_k^T c_k\|^2$ . Note that by (3.51) the dependence on  $\|\bar{J}_k^T \bar{c}_k\|^2 + \|J_k^T c_k\|^2$  can be replaced by  $\|\bar{v}_k - v_k\|^2$ . When  $\epsilon_c = \epsilon_J = \epsilon_g = 0$ ,  $\bar{v}_k$  and  $v_k$  are either both Cauchy or both exact solutions to subproblems (2.2) and (3.23), respectively, then  $\bar{u}_k = u_k^*$ . However, our proof does not require that  $\|u_k^* - \bar{u}_k\|^2$  is sufficiently small at each iteration. Instead,  $\|u_k^* - \bar{u}_k\|^2$  is sufficiently small once both the stationarity ( $\|\bar{u}_k\|$ ) and infeasible stationarity ( $\|\bar{J}_k^T \bar{c}_k\|$ ) errors are sufficiently small.

Given the fictitious (never computed in Algorithm 1) normal solution  $v_k$  and tangential solution  $u_k^*$  for the true SQP subproblems (3.23) and (3.42), respectively, we define a fictitious (never computed in Algorithm 1) merit parameter  $\tau_k$  update rule for the deterministic counterpart of our proposed algorithm,

$$\tau_k \leftarrow \begin{cases} \bar{\tau}_{k-1} & \text{if } \bar{\tau}_{k-1} \leq (1 - \sigma_\tau) \tau_k^{\text{trial}} \\ (1 - \sigma_\tau) \tau_k^{\text{trial}} & \text{otherwise,} \end{cases} \quad (3.52)$$

where

$$\tau_k^{\text{trial}} \leftarrow \begin{cases} \infty & \text{if } g_k^T d_k^* + \max\{u_k^{*T} H_k u_k^*, \lambda_u \|u_k^*\|^2\} \leq 0 \\ \frac{(1 - \frac{\sigma_c}{\sigma_\tau})(\|c_k\| - \|c_k + J_k v_k\|)}{g_k^T d_k^* + \max\{u_k^{*T} H_k u_k^*, \lambda_u \|u_k^*\|^2\}} & \text{otherwise.} \end{cases} \quad (3.53)$$

The next lemma characterizes the relationship between noisy and true merit parameters. Note that it requires  $\bar{u}_k$  to be the exact solution to subproblem (2.5) to maintain consistency with the exactness level of  $u_k^*$ , which restricts  $\lambda_{pr} = 0$  in (2.9).

**Lemma 3.27.** Suppose that Assumptions 2.1, 2.2, 2.4, 3.1, 3.2, 3.4, and 3.22 hold. In addition, suppose that  $\bar{v}_k$  and  $v_k$  are either both Cauchy solutions or both exact solutions to subproblems (2.2) and (3.23), respectively, and that  $\bar{u}_k$  and  $u_k^*$  are the exact solutions to subproblems (2.5) and (3.42), respectively, i.e.,  $\bar{u}_k = \bar{u}_k^*$ . Then, there exists  $\mathcal{E}_\tau \in \mathbb{R}_{\geq 0}$  such that for all  $k \in \mathbb{N}$ ,  $|(\bar{\tau}_k - \tau_k)(g_k^T d_k^* + u_k^{*T} H_k u_k^*)| \leq \mathcal{E}_\tau$ .

*Proof.* We first consider the case when  $\|\bar{c}_k\| \leq \epsilon_o$ , it follows by Assumptions 2.1, 2.2, and 2.4, Lemma 3.21, (3.23), and (3.42) that

$$\begin{aligned} |g_k^T d_k^* + u_k^{*T} H_k u_k^*| &= |g_k^T u_k^* + u_k^{*T} H_k u_k^* + g_k^T v_k| = |-u_k^{*T} H_k v_k + g_k^T v_k| \\ &\leq (\kappa_{u^*} \kappa_H + \kappa_g) \sigma_{Jc} \kappa_J \|c_k\| \leq 2(\kappa_{u^*} \kappa_H + \kappa_g) \sigma_{Jc} \kappa_{Jc} := \mathcal{E}_c, \end{aligned}$$

which implies that  $|(\bar{\tau}_k - \tau_k)(g_k^T d_k^* + u_k^{*T} H_k u_k^*)| \leq 2\bar{\tau}_{k-1} \mathcal{E}_c$  by Lemma 3.13.

Next, we consider the case when  $\|\bar{c}_k\| > \epsilon_o$ . Since we allow  $\bar{v}_k$  and  $v_k$  to be either both Cauchy or both exact solutions to subproblems (2.2) and (3.23), respectively, it follows that  $\|v_k - \bar{v}_k\| \leq \max\{\mathcal{E}_v, \mathcal{E}_{v^*}\}$ , where  $\mathcal{E}_v$  and  $\mathcal{E}_{v^*}$  are defined in Lemmas 3.18 and 3.19, respectively. The difference between the exact noisy tangential component  $\bar{u}_k = \bar{u}_k^*$  and the true tangential component  $u_k^*$  can be derived by (3.51),

$$\begin{aligned}
& \|u_k^* - \bar{u}_k\| \\
&= \|Z_k w_k - \bar{Z}_k \bar{w}_k\| \\
&\leq \left\| [Z_k^c \ C_k \ \bar{Z}_k^c] ([Z_k^c \ C_k \ \bar{Z}_k^c]^T H_k [Z_k^c \ C_k \ \bar{Z}_k^c])^{-1} \begin{bmatrix} Z_k^T \\ \bar{Z}_k^{cT} \end{bmatrix} ((\bar{g}_k - g_k) + H_k(\bar{v}_k - v_k)) \right. \\
&\quad \left. + [Z_k^c \ C_k \ \bar{Z}_k^c] ([Z_k^c \ C_k \ \bar{Z}_k^c]^T H_k [Z_k^c \ C_k \ \bar{Z}_k^c])^{-1} \begin{bmatrix} Z_k^T ((\bar{J}_k - J_k)^T \bar{y}_k^*) \\ -\bar{Z}_k^{cT} ((\bar{J}_k - J_k)^T y_k^*) \end{bmatrix} \right\| \\
&\leq \zeta^{-1} ((\kappa_{y^*} + \kappa_{\bar{y}^*})\epsilon_J + \epsilon_g + \kappa_H \max\{\mathcal{E}_v, \mathcal{E}_{v^*}\}) := \mathcal{E}_u.
\end{aligned}$$

We divide the proof into four cases that depend on whether  $(\tau_k, \bar{\tau}_k)$  is updated conditioned on  $\bar{\tau}_{k-1}$ . For conciseness, we define  $h_k = g_k^T d_k^* + \max\{u_k^{*T} H_k u_k^*, \lambda_u \|u_k^*\|^2\}$  and  $\bar{h}_k = \bar{g}_k^T \bar{d}_k + \max\{\bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2\}$ . We begin with a general bound that holds in all cases. By Assumptions 2.1, 2.2 and 2.4, and the fact that  $d_k^* = u_k^* + v_k$  and  $\bar{d}_k = \bar{u}_k + \bar{v}_k$ ,

$$\begin{aligned}
|h_k - \bar{h}_k| &= \left| g_k^T d_k^* + \max\{u_k^{*T} H_k u_k^*, \lambda_u \|u_k^*\|^2\} - \bar{g}_k^T \bar{d}_k - \max\{\bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2\} \right| \\
&\leq \left| g_k^T d_k^* + u_k^{*T} H_k u_k^* - \bar{g}_k^T \bar{d}_k - \bar{u}_k^T H_k \bar{u}_k \right| \\
&\leq \left| g_k^T d_k^* - g_k^T \bar{d}_k + g_k^T \bar{d}_k - \bar{g}_k^T \bar{d}_k + u_k^{*T} H_k u_k^* - \bar{u}_k^{*T} H_k \bar{u}_k + \bar{u}_k^{*T} H_k u_k^* - \bar{u}_k^T H_k \bar{u}_k \right| \\
&\leq \left| g_k^T d_k^* - g_k^T \bar{d}_k \right| + \left| g_k^T \bar{d}_k - \bar{g}_k^T \bar{d}_k \right| + \left| u_k^{*T} H_k u_k^* - \bar{u}_k^{*T} H_k \bar{u}_k \right| + \left| \bar{u}_k^{*T} H_k u_k^* - \bar{u}_k^T H_k \bar{u}_k \right| \\
&\leq \kappa_g (\mathcal{E}_v + \mathcal{E}_u) + \kappa_{\bar{d}} \epsilon_g + \kappa_H (\kappa_u + \kappa_{\bar{u}^*}) \mathcal{E}_u := \mathcal{E}_h.
\end{aligned}$$

Additionally, when  $\bar{v}_k$  and  $v_k$  are either both Cauchy solutions or both exact solutions to subproblems (2.2) and (3.23), respectively, it follows by Lemmas 3.6 and 3.18, and Assumption 2.2

$$\begin{aligned}
\left\| \bar{c}_k + \bar{J}_k \bar{v}_k \right\| - \|c_k + J_k v_k\| &\leq \left\| c_k + J_k v_k - \bar{c}_k - \bar{J}_k \bar{v}_k \right\| \\
&\leq \|c_k - \bar{c}_k\| + \|J_k v_k - \bar{J}_k \bar{v}_k\| + \|J_k \bar{v}_k - \bar{J}_k \bar{v}_k\| \\
&\leq \epsilon_c + \kappa_J \max\{\mathcal{E}_v, \mathcal{E}_{v^*}\} + \sigma_{J_c} (\kappa_J + \epsilon_J) (\kappa_c + \epsilon_c) \epsilon_J := \mathcal{E}_{cJv}.
\end{aligned} \tag{3.54}$$

Case (i): Neither  $\bar{\tau}_k$  nor  $\tau_k$  is updated, i.e.,  $\tau_k = \bar{\tau}_k = \bar{\tau}_{k-1}$ . The result holds in case (i).

Case (ii): Both  $\bar{\tau}_k$  and  $\tau_k$  are updated. By the update rule for  $\tau_k$  and  $\bar{\tau}_k$ , we have  $h_k > 0$ ,  $\bar{h}_k > 0$ , and  $(1 - \sigma_\tau)(1 - \sigma_c/\sigma_\tau) \|\bar{c}_k\| < \bar{\tau}_{k-1} \bar{h}_k \leq \bar{\tau}_{-1} \bar{h}_k$ . It then follows by

$\bar{u}_k = \bar{u}_k^*$ , (2.17), (3.53), and (3.54) that

$$\begin{aligned}
|\bar{\tau}_k - \tau_k| &= (1 - \sigma_\tau) \left| \bar{\tau}_k^{trial} - \tau_k^{trial} \right| \\
&= (1 - \sigma_\tau) \left| \frac{(1 - \sigma_c / \sigma_r)(\|c_k\| - \|c_k + J_k v_k\|)}{h_k} - \frac{(1 - \sigma_c / \sigma_r)(\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|)}{\bar{h}_k} \right| \\
&= (1 - \sigma_\tau) \left( 1 - \frac{\sigma_c}{\sigma_r} \right) \left| \frac{(\|c_k\| - \|c_k + J_k v_k\| - \|\bar{c}_k\| + \|\bar{c}_k + \bar{J}_k \bar{v}_k\|) \bar{h}_k + (\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|) (\bar{h}_k - h_k)}{h_k \bar{h}_k} \right| \\
&\leq (1 - \sigma_\tau) \left( 1 - \frac{\sigma_c}{\sigma_r} \right) \left| \frac{\|\bar{c}_k\| \mathcal{E}_h + \bar{h}_k (\epsilon_c + \mathcal{E}_{cJv})}{h_k \bar{h}_k} \right| \\
&\leq \frac{\bar{\tau}_{-1} \mathcal{E}_h}{h_k} + (1 - \sigma_\tau) \left( 1 - \frac{\sigma_c}{\sigma_r} \right) \frac{\epsilon_c + \mathcal{E}_{cJv}}{h_k}.
\end{aligned}$$

Thus, in case (ii),  $|(\bar{\tau}_k - \tau_k) h_k| \leq \bar{\tau}_{-1} \mathcal{E}_h + (1 - \sigma_\tau) \left( 1 - \frac{\sigma_c}{\sigma_r} \right) (\epsilon_c + \mathcal{E}_{cJv})$ .

Case (iii):  $\bar{\tau}_k$  is updated and  $\tau_k$  is not, i.e.,  $\bar{\tau}_k = \bar{\tau}_k^{trial} < \bar{\tau}_{k-1}$  and  $\tau_k = \bar{\tau}_{k-1} \leq (1 - \sigma_\tau) \bar{\tau}_k^{trial}$ . By the update rule of  $\bar{\tau}_k$ , we have  $\bar{h}_k > 0$ . Let us first discuss the case where  $\bar{h}_k > \mathcal{E}_h$ , which implies that  $h_k > 0$  and  $\tau_k^{trial} < \infty$ . It then follows by (3.53)

$$\begin{aligned}
|\bar{\tau}_k - \tau_k| &= \tau_k - (1 - \sigma_\tau) \left( 1 - \frac{\sigma_c}{\sigma_r} \right) \frac{\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k + \bar{\tau}_k\|}{\bar{h}_k} \\
&\leq (1 - \sigma_\tau) \left( 1 - \frac{\sigma_c}{\sigma_r} \right) \left( \frac{\|c_k\| - \|c_k + J_k v_k\|}{h_k} - \frac{\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|}{\bar{h}_k} \right),
\end{aligned}$$

from which we conclude that the result from case (ii) holds. When  $\bar{h}_k \leq \mathcal{E}_h$ , it follows that  $h_k \leq 2\mathcal{E}_h$  and therefore  $|(\bar{\tau}_k - \tau_k) h_k| \leq 4\bar{\tau}_{-1} \mathcal{E}_h$ .

Case (iv):  $\tau_k$  is updated and  $\bar{\tau}_k$  is not, i.e.,  $\tau_k \leq \bar{\tau}_k = \bar{\tau}_{k-1}$  and  $\bar{\tau}_k = \bar{\tau}_{k-1}$ . By the update rule of  $\tau_k$ , we have  $h_k > 0$ . Let us first discuss the case when  $h_k > \mathcal{E}_h$ , which implies that  $\bar{h}_k > 0$  and  $\bar{\tau}_k^{trial} < \infty$ . It then follows by (3.53)

$$\begin{aligned}
|\bar{\tau}_k - \tau_k| &= \bar{\tau}_{k-1} - (1 - \sigma_\tau) \left( 1 - \frac{\sigma_c}{\sigma_r} \right) \frac{\|c_k\| - \|c_k + J_k v_k\|}{h_k} \\
&\leq (1 - \sigma_\tau) \left( 1 - \frac{\sigma_c}{\sigma_r} \right) \left( \frac{\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|}{\bar{h}_k} - \frac{\|c_k\| - \|c_k + J_k v_k\|}{h_k} \right),
\end{aligned}$$

from which we conclude that the result from case (ii) holds. When  $h_k \leq \mathcal{E}_h$ , we have  $|(\bar{\tau}_k - \tau_k) h_k| \leq 2\bar{\tau}_{-1} \mathcal{E}_h$ .

Finally, defining  $\mathcal{E}_\tau := \max \left\{ \bar{\tau}_{-1} \mathcal{E}_h + (1 - \sigma_\tau) \left( 1 - \frac{\sigma_c}{\sigma_r} \right) (\epsilon_c + \mathcal{E}_{cJv}), 4\bar{\tau}_{-1} \mathcal{E}_h, 2\bar{\tau}_{-1} \mathcal{E}_c \right\}$  completes the proof.  $\square$

**Remark 3.28.** In Lemma 3.27, we bound  $|(\bar{\tau}_k - \tau_k)(g_k^T d_k^* + u_k^{*T} H_k u_k^*)|$  by quantities related to the noise in the objective and constraint functions. When  $\epsilon_g = \epsilon_c = \epsilon_J = 0$ , we have  $\mathcal{E}_v = \mathcal{E}_u = \mathcal{E}_h = \mathcal{E}_{cJv} = \mathcal{E}_c = 0$  and the upper bound is zero ( $\bar{\tau}_k = \tau_k$ ). When  $g_k^T d_k^* + u_k^{*T} H_k u_k^*$  is sufficiently small, i.e., the algorithm has reached an approximate first-order stationary point, the bound becomes useless. However, when  $g_k^T d_k^* + u_k^{*T} H_k u_k^*$  is sufficiently large, the



difference between  $\tau_k$  and  $\bar{\tau}_k$  is sufficiently small and proportional to the noise levels in the problem. Lemma 3.27 imposes conditions on the normal and tangential components, i.e., the fact that the exactness of both the noisy normal component and the noisy tangential component must be of the same level as the true components. It is important to note that we do not require such conditions in the main convergence analysis (Section 3.4).

Careful consideration of the behavior of the merit parameter  $\bar{\tau}_k$  is essential as it plays a pivotal role in both the formulation and analysis of the proposed noisy SQP method. It is shown in [19, Lemma 3.16] that when the problem has no noise and the LICQ holds, the merit parameter sequence is guaranteed to be bounded away from zero. Next, we discuss the deviation in the lower bound of the merit parameter sequence ( $\{\bar{\tau}_k\}$ ) from a fictitious lower bound of a true merit parameter sequence ( $\{\tau_k\}$ , never computed) under the LICQ.

**Lemma 3.29.** *Suppose that Assumptions 2.1, 2.2, 2.4, 3.1, 3.2, and 3.4 hold. In addition, suppose that the singular values of both  $\bar{J}_k$  and  $J_k$  are bounded below by  $\kappa_\sigma \in \mathbb{R}_{>0}$  for all  $k \in \mathbb{N}$ . Furthermore, suppose that  $\bar{v}_k$  and  $v_k$  are the exact solutions of (2.2) and (3.23), respectively, and that  $\bar{u}_k$  and  $u_k^*$  are exact solutions to subproblems (2.5) and (3.42), respectively, i.e.,  $\|\bar{c}_k + \bar{J}_k \bar{v}_k\| = 0$ ,  $\|\bar{r}_k\| = \|\bar{\rho}_k\| = 0$ , and  $\bar{u}_k = \bar{u}_k^*$ . Then, there exist lower bounds  $\{\tau_{\min}, \tilde{\tau}_{\min}\} \subset \mathbb{R}_{>0}$  such that  $\tau_k \geq \tau_{\min}$ ,  $\bar{\tau}_k \geq \tilde{\tau}_{\min}$ , and  $\mathcal{E}_{\tau, \min} \in \mathbb{R}_{\geq 0}$  such that  $|\tau_{\min} - \tilde{\tau}_{\min}| \leq \mathcal{E}_{\tau, \min}$ .*

*Proof.* Since  $\|\bar{r}_k\| = \|\bar{\rho}_k\| = 0$ , we have  $\bar{u}_k = \bar{u}_k^*$ . Note that  $\bar{\tau}_k < \bar{\tau}_{k-1}$  only when  $\|\bar{c}_k\| > \epsilon_o$ ,  $\bar{g}_k^T \bar{d}_k + \max\{\bar{u}_k^T H_k \bar{u}_k, \lambda_u \|\bar{u}_k\|^2\} > 0$ , and  $\bar{\tau}_{k-1} \geq \frac{(1-\sigma_\tau)(1-\sigma_c/\sigma_r)\|\bar{c}_k\|}{\bar{g}_k^T \bar{d}_k + \max\{\bar{u}_k^*{}^T H_k \bar{u}_k^*, \lambda_u \|\bar{u}_k^*\|^2\}}$ . Hence, if the merit parameter is ever updated, it follows by (3.15) that

$$\begin{aligned} \bar{\tau}_{k-1} &\geq \frac{(1-\sigma_\tau)(\sigma_r-\sigma_c)\sigma_v\|\bar{J}_k^T \bar{c}_k\|^2/\|\bar{c}_k\|}{\bar{g}_k^T \bar{u}_k^* + \bar{g}_k^T \bar{v}_k + \max\{\bar{u}_k^*{}^T H_k \bar{u}_k^*, \lambda_u \|\bar{u}_k^*\|^2\}} \\ &\geq \frac{(1-\sigma_\tau)(1-\sigma_c/\sigma_r)\sigma_v\|\bar{J}_k^T \bar{c}_k\|^2/\|\bar{c}_k\|}{-\bar{u}_k^*{}^T H_k \bar{v}_k + \bar{g}_k^T \bar{v}_k} \\ &\geq \frac{(1-\sigma_\tau)(1-\sigma_c/\sigma_r)\sigma_v\|\bar{J}_k^T \bar{c}_k\|^2/\|\bar{c}_k\|}{\kappa_{\bar{u}}^* \kappa_H \sigma_{Jc}(\kappa_J + \epsilon_J)\|\bar{c}_k\| + (\kappa_g + \epsilon_g)\sigma_{Jc}(\kappa_J + \epsilon_J)\|\bar{c}_k\|} \\ &= \frac{(1-\sigma_\tau)(1-\sigma_c/\sigma_r)\sigma_v \kappa_\sigma^2}{\kappa_{\bar{u}}^* \kappa_H \sigma_{Jc}(\kappa_J + \epsilon_J) + (\kappa_g + \epsilon_g)\sigma_{Jc}(\kappa_J + \epsilon_J)} := \tilde{\tau}_{\min}. \end{aligned}$$

As for deterministic SQP method (no noise), it follows from [19, Lemmas 3.5 & 3.6] that there exists  $\sigma_v \in \mathbb{R}_{>0}$  such that

$$\|c_k\| (\|c_k\| - \|c_k + J_k v_k\|) \geq \sigma_v \|J_k^T c_k\|^2. \quad (3.55)$$

Although  $\sigma_v$  is not necessarily the same constant as in Lemma 3.6, this discrepancy does not result in a loss of generality. Specifically,  $\sigma_v$  can be defined as the minimum of the two constants appearing in (3.55) and its noisy counterpart. Whenever  $\tau_k$  is updated, we have

$g_k^T d_k^* + \max \left\{ u_k^{*T} H_k u_k^*, \lambda_u \|u_k^*\|^2 \right\} > 0$ , from which it follows that

$$\begin{aligned} \tau_{k-1} &\geq \frac{(1-\sigma_\tau)(1-\sigma_c/\sigma_r)(\|c_k\| - \|c_k + J_k v_k\|)}{g_k^T u_k^* + g_k^T v_k + \max \left\{ u_k^{*T} H_k u_k^*, \lambda_u \|u_k^*\|^2 \right\}} \\ &\geq \frac{(1-\sigma_\tau)(1-\sigma_c/\sigma_r)\sigma_v \|J_k^T c_k\|^2 / \|c_k\|}{u_k^{*T} H_k v_k + g_k^T v_k} \\ &\geq \frac{(1-\sigma_\tau)(1-\sigma_c/\sigma_r)\sigma_v \kappa_\sigma^2 \|c_k\|}{\kappa_{u^*} \kappa_H \sigma_{J_c} \kappa_{J_c} \|c_k\| + \kappa_g \sigma_{J_c} \kappa_{J_c} \|c_k\|} \\ &= \frac{(1-\sigma_\tau)(1-\sigma_c/\sigma_r)\sigma_v \kappa_\sigma^2}{\kappa_{u^*} \kappa_H \sigma_{J_c} \kappa_{J_c} + \kappa_g \sigma_{J_c} \kappa_{J_c}} := \tau_{\min}. \end{aligned}$$

Given (3.45) and (3.46), it follows by Assumption 2.2 that

$$\kappa_{\bar{u}^*} - \kappa_{u^*} = \zeta^{-1} (\epsilon_g + \kappa_H \sigma_{J_c} (\kappa_{J_c} \epsilon_c + \kappa_c \epsilon_J + \epsilon_c \epsilon_J)) := \epsilon_{u^*}.$$

By Assumption 2.2, Lemma 3.21, the definitions of  $\tau_{\min}$  and  $\tilde{\tau}_{\min}$ , it follows that

$$\begin{aligned} |\tau_{\min} - \tilde{\tau}_{\min}| &\leq \left| \frac{(1-\sigma_\tau)(1-\sigma_c/\sigma_r)\sigma_v \kappa_\sigma^2}{\kappa_{u^*} \kappa_H \sigma_{J_c} \kappa_{J_c} + \kappa_g \sigma_{J_c} \kappa_{J_c}} - \frac{(1-\sigma_\tau)(1-\sigma_c/\sigma_r)\sigma_v \kappa_\sigma^2}{\kappa_{\bar{u}^*} \kappa_H \sigma_{J_c} (\kappa_{J_c} + \epsilon_J) + (\kappa_g + \epsilon_g) \sigma_{J_c} (\kappa_{J_c} + \epsilon_J)} \right| \\ &\leq (1 - \sigma_\tau) \left( 1 - \frac{\sigma_c}{\sigma_r} \right) \sigma_v \kappa_\sigma^2 \left| \frac{\kappa_{\bar{u}^*} \kappa_H \sigma_{J_c} (\kappa_{J_c} + \epsilon_J) + (\kappa_g + \epsilon_g) \sigma_{J_c} (\kappa_{J_c} + \epsilon_J) - \kappa_{u^*} \kappa_H \sigma_{J_c} \kappa_{J_c} + \kappa_g \sigma_{J_c} \kappa_{J_c}}{(\kappa_{u^*} \kappa_H \sigma_{J_c} \kappa_{J_c} + \kappa_g \sigma_{J_c} \kappa_{J_c})(\kappa_{\bar{u}^*} \kappa_H \sigma_{J_c} (\kappa_{J_c} + \epsilon_J) + (\kappa_g + \epsilon_g) \sigma_{J_c} (\kappa_{J_c} + \epsilon_J))} \right| \\ &\leq (1 - \sigma_\tau) \left( 1 - \frac{\sigma_c}{\sigma_r} \right) \sigma_v \kappa_\sigma^2 \left| \frac{\kappa_H \sigma_{J_c} \kappa_{J_c} \epsilon_{u^*} + \kappa_{\bar{u}^*} \kappa_H \sigma_{J_c} \epsilon_J + \kappa_g \sigma_{J_c} \epsilon_J + \kappa_{J_c} \sigma_{J_c} \epsilon_g + \epsilon_J \sigma_{J_c} \epsilon_g}{(\kappa_{u^*} \kappa_H \sigma_{J_c} \kappa_{J_c} + \kappa_g \sigma_{J_c} \kappa_{J_c})(\kappa_{\bar{u}^*} \kappa_H \sigma_{J_c} (\kappa_{J_c} + \epsilon_J) + (\kappa_g + \epsilon_g) \sigma_{J_c} (\kappa_{J_c} + \epsilon_J))} \right| \\ &:= \mathcal{E}_{\tau, \min}, \end{aligned}$$

which completes the proof.  $\square$

**Remark 3.30.** Lemma 3.29 bounds the deviation in the lower bound of the merit parameter sequence from a fictitious true (never computed) lower bound when the LICQ holds. By Lemma 3.13, this deviation is sufficiently small and proportional to the noise levels in the problem. When  $\epsilon_g = \epsilon_c = \epsilon_J = 0$ , we have  $\mathcal{E}_{\tau, \min} = 0$  and  $(\bar{\tau}_k, \tau_k)$  share the same positive lower bound.

### 3.4 Main Convergence Theorems

In this section, we establish the main convergence theorems for Algorithm 1 under both step size strategies. Before presenting the main results, we note that our results are analogous to results that can be established in the deterministic setting, see e.g., [19, Theorem 3.3], and are summarized in Table 2. In the most general case, convergence can only be established in terms of the *infeasible stationarity error*  $\|J_k^T c_k\|$ . When  $J_k$  is of full rank, we can prove a stronger convergence result in terms of the *feasibility error*  $\|c_k\|$ . Furthermore, as we will show later in this section, when  $\bar{J}_k$  is of full rank, we can also establish the convergence of  $\|c_k\|$ . Whether or not convergence can be established in terms of the *stationarity error*

$$\|Z_k^T (\nabla f(x_k) + H_k v_k)\| \tag{3.56}$$

$\sigma_{\min}(\bar{J}_k) \geq \kappa_\sigma, \bar{\tau}_k \geq \bar{\tau}_{\min}$		$\sigma_{\min}(\bar{J}_k) = 0, \bar{\tau}_k \geq \bar{\tau}_{\min}$		$\sigma_{\min}(\bar{J}_k) = 0, \bar{\tau}_k \rightarrow 0$	
		$\sigma_{\min}(J_k) \geq \kappa_\sigma$	$\sigma_{\min}(J_k) = 0$	$\sigma_{\min}(J_k) \geq \kappa_\sigma$	$\sigma_{\min}(J_k) = 0$
(in)fea.	$\ c_k\ $	$\ c_k\ ^2$	$\ J_k^T c_k\ ^2$	$\ c_k\ ^2$	$\ J_k^T c_k\ ^2$
sta.	$\ Z_k^T(\nabla f(x_k) + H_k v_k)\ ^2$	$\ Z_k^T(\nabla f(x_k) + H_k v_k)\ ^2$		–	

Table 2: Summary of feasibility or infeasible stationarity measure and stationarity measure used in different cases.

heavily depends on the behavior of the merit parameter sequence. When  $\{\bar{\tau}_k\}$  diminishes to zero, no meaningful conclusions can be drawn about the *stationarity error*. When  $\{\bar{\tau}_k\}$  is bounded below by some  $\bar{\tau}_{\min} \in \mathbb{R}_{>0}$ , we prove that the averaged *stationarity error* is sufficiently small. It is worth noting that the reduced gradient  $\|Z_k^T \nabla f(x_k)\| = \|Z_k^T(\nabla f(x_k) + J_k^T y_k^*)\|$  is used as a stationarity measure in [45]. We argue that (3.56) is a reasonable stationarity measure for problem (1.1) provided that  $\|J_k^T c_k\|$  vanishes or is sufficiently small. Since  $\|Z_k^T H_k v_k\| \leq \|H_k\| \|v_k\| \leq \kappa_H \sigma_{J_c} \|J_k^T c_k\|$  by Assumption 2.4 and (3.23), as long as  $\|J_k^T c_k\|$  is sufficiently small,  $\|Z_k^T(\nabla f(x_k) + H_k v_k)\|$  closely approximates the norm of the reduced gradient  $\|Z_k^T \nabla f(x_k)\|$  and is thus a reasonable stationarity measure. The stationarity error is related to the norm of tangential component.

We first derive upper bounds of the model reduction for both step size schemes when  $\bar{\tau}_k \geq \bar{\tau}_{\min}$ . However, when  $\{\bar{\tau}_k\}$  diminishes to zero, the step size may vanish, and a sufficient reduction in the merit function can no longer be ensured. In this case, we show that the algorithm converges to a nearly infeasible stationary solution. To address this scenario, we provide a result on the reduction of constraint violation, which supports the convergence analysis when the merit parameter vanishes. We emphasize the importance of demonstrating that the true merit function achieves sufficient reduction at each iteration under the proposed step size. To this end, we define

$$\phi(x, \tau) := \tau f(x) + \|c(x)\|. \quad (3.57)$$

**Lemma 3.31.** *Suppose that Assumptions 2.1, 2.2, 2.4, 3.1, 3.2, and 3.4 hold. Let*

$$\mathcal{E}^A := \alpha_u^A \bar{\tau}_{-1} \kappa_{\bar{d}} \epsilon_g + 2\epsilon_c + \alpha_u^A \kappa_{\bar{d}} \epsilon_J \quad (3.58)$$

$$\text{and } \mathcal{E}^L := 4\bar{\tau}_{-1} \epsilon_f + \bar{\tau}_{-1} \alpha_u^L \kappa_{\bar{d}} \epsilon_g + 6\epsilon_c + \alpha_u^L \kappa_{\bar{d}} \epsilon_J. \quad (3.59)$$

If  $\bar{\tau}_k \geq \bar{\tau}_{\min}$ , for all  $k \in \mathbb{N}$

$$\phi(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_k) - \phi(x_k, \bar{\tau}_k) \leq -\eta \alpha_{l^i}^i \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) + \mathcal{E}^i, \quad (3.60)$$

where  $i \in \{A, L\}$  denotes the adaptive (**Option I**) and line search (**Option II**) scheme, respectively, in Algorithm 1. If  $\{\bar{\tau}_k\}$  diminishes to 0 and  $x_k \in \mathcal{X}_\gamma$ , for all  $k \in \mathbb{N}$

$$\|c_{k+1}\| - \|c_k\| \leq -\eta \alpha_{l^i}^i \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) - \bar{\tau}_k (f_{k+1} - f_k) + \mathcal{E}^i. \quad (3.61)$$

Table 3: Table summarizing the notation (upper and lower bounds and merit function increase) from Lemma 3.31 for the two step size schemes  $i \in \{A, L\}$  and different scenarios for the merit parameter ( $\bar{\tau}_k \geq \bar{\tau}_{\min}$  and  $\bar{\tau}_k \rightarrow 0$ ,  $x_k \in \mathcal{X}_\gamma$ ).

	Upper Bound	Lower Bound ( $\bar{\tau}_k \geq \bar{\tau}_{\min}$ )	Lower Bound ( $\bar{\tau}_k \rightarrow 0$ and $x_k \in \mathcal{X}_\gamma$ )	Merit Function Increase
<b>Option I</b>	$\alpha_u^A$	$\alpha_{l\tau}^A$	$\alpha_{l0}^A$	$\mathcal{E}^A = \mathcal{O}(\epsilon_g + \epsilon_c + \epsilon_J)$
<b>Option II</b>	$\alpha_u^L$	$\alpha_{l\tau}^L$	$\alpha_{l0}^L$	$\mathcal{E}^L = \mathcal{O}(\epsilon_f + \epsilon_g + \epsilon_c + \epsilon_J)$

*Proof.* We first prove that (3.60) and (3.61) hold when  $i = A$ . By Lemma 3.14 it follows that  $\bar{\alpha}_k \leq 1$ . When  $\bar{\tau}_k \geq \bar{\tau}_{\min}$ , it follows that

$$\begin{aligned}
& \phi(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_k) - \phi(x_k, \bar{\tau}_k) \\
& \leq -\bar{\alpha}_k \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) + \bar{\alpha}_k \bar{\tau}_k (g_k - \bar{g}_k)^T \bar{d}_k + \frac{1}{2} (\bar{\tau}_k L + \Gamma) \bar{\alpha}_k^2 \|\bar{d}_k\|^2 + 2\epsilon_c + \bar{\alpha}_k \epsilon_J \|\bar{d}_k\| \\
& \leq -\bar{\alpha}_k \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) + \bar{\alpha}_k \bar{\tau}_k (g_k - \bar{g}_k)^T \bar{d}_k + (1 - \eta) \beta \bar{\alpha}_k \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) + 2\epsilon_c + \bar{\alpha}_k \epsilon_J \|\bar{d}_k\| \\
& \leq -\eta \bar{\alpha}_k \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) + \bar{\alpha}_k \bar{\tau}_k \kappa_{\bar{d}} \epsilon_g + 2\epsilon_c + \bar{\alpha}_k \kappa_{\bar{d}} \epsilon_J \tag{3.62} \\
& \leq -\eta \alpha_{l\tau}^A \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) + \alpha_u^A \bar{\tau}_{-1} \kappa_{\bar{d}} \epsilon_g + 2\epsilon_c + \alpha_u^A \kappa_{\bar{d}} \epsilon_J, \tag{3.63}
\end{aligned}$$

where the first inequality follows by the third last inequality in (3.20), the second inequality follows by the fact that  $\bar{\alpha}_k \leq \bar{\alpha}_k^{\text{suff}} \leq \frac{2(1-\eta)\beta_k \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k)}{(\bar{\tau}_k L + \Gamma) \|\bar{d}_k\|^2}$  (see Lemma 3.14 and (2.22)), the third inequality follows by Lemma 3.12, Assumption 2.2, and the fact that  $\beta \leq 1$ , and the last inequality follows by Lemmas 3.13 and 3.14. (3.60) is therefore satisfied. When  $\{\bar{\tau}_k\}$  diminishes to 0 and  $x_k \in \mathcal{X}_\gamma$ , it follows from Lemma 3.14 that  $\bar{\alpha}_k \leq \alpha_{l0}^A$ . Finally, (3.61) follows by (3.57) and (3.62).

Next, we prove that (3.60) and (3.61) hold when  $i = L$ . When  $\bar{\tau}_k \geq \bar{\tau}_{\min}$ , it follows by Assumption 2.2, (2.25), (2.26), Lemma 3.12 and Lemma 3.15 that

$$\begin{aligned}
\phi(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_k) - \phi(x_k, \bar{\tau}_k) & \leq \bar{\phi}(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_k) - \bar{\phi}(x_k, \bar{\tau}_k) + 2\bar{\tau}_k \epsilon_f + 2\epsilon_c \\
& \leq -\eta \bar{\alpha}_k \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) + \epsilon_{A_k} + 2\bar{\tau}_k \epsilon_f + 2\epsilon_c \tag{3.64}
\end{aligned}$$

$$\begin{aligned}
& \leq -\eta \alpha_{l\tau}^L \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) + 4\bar{\tau}_{-1} \epsilon_f + \bar{\tau}_{-1} \alpha_u^L \kappa_{\bar{d}} \epsilon_g + 6\epsilon_c + \alpha_u^L \kappa_{\bar{d}} \epsilon_J, \tag{3.65}
\end{aligned}$$

and (3.60) is therefore satisfied. When  $\{\bar{\tau}_k\}$  diminishes to 0 and  $x_k \in \mathcal{X}_\gamma$ , (3.61) follows by (3.57), (3.65), and Lemma 3.15.  $\square$

Table 3 summarizes the results of Lemma 3.31.

The following lemma shows a relationship between the true tangential step  $u_k^*$  and the stationarity error (3.56).

**Lemma 3.32.** *Suppose that Assumptions 2.4, 3.4, and 3.22 hold. For all  $k \in \mathbb{N}$ ,  $\|u_k^*\| \geq \kappa_H^{-1} \|Z_k^T (\nabla f(x_k) + H_k v_k)\|$ .*

*Proof.* The proof follows the same logic as [7, Lemma 9].  $\square$

The following lemma establishes bounds for feasibility and stationarity errors in the event that Algorithm 1 terminates finitely on Line 7.

**Theorem 3.33.** *Suppose that Assumptions 2.1, 2.2, 2.4, 3.1, 3.2, and 3.22 hold. If Algorithm 1 terminates finitely at iteration  $k = K_T \in \mathbb{N}$  on Line 7 (i.e.,  $\|\bar{c}_{K_T}\| \leq \epsilon_o$  and  $\Delta\bar{l}(x_{K_T}, \bar{\tau}_{K_T}, \bar{d}_{K_T}) \leq \epsilon_o$ ), where  $\epsilon_o \in [0, \epsilon_c]$ , then  $\|c_{K_T}\| \leq 2\epsilon_c$  and*

$$\begin{aligned} & \left\| Z_{K_T}^T (\nabla f(x_{K_T}) + H_{K_T} v_{K_T}) \right\|^2 \\ & \leq \frac{4\kappa_H^2 \epsilon_c}{\bar{\tau}_{K_T} \sigma_u \lambda_u} + 12\kappa_H^2 \zeta^{-2} (\epsilon_g^2 + \kappa_{\bar{y}^*}^2 \epsilon_J^2 + \kappa_{y^*}^2 \epsilon_J^2) + 96\kappa_H^4 \zeta^{-2} \sigma_{J_c}^2 \kappa_J^2 \epsilon_c^2 \\ & \quad + 4\kappa_H^2 \lambda_{pr}^2 (\zeta^{-1} + (\kappa_J + \epsilon_J) \kappa_\sigma^{-2})^2 \max \left\{ \frac{2\epsilon_c}{\bar{\tau}_{K_T} \sigma_u \lambda_u}, (\kappa_J + \epsilon_J)^2 \epsilon_c^2 \right\} \end{aligned} \quad (3.66)$$

*Proof.* Since Algorithm 1 terminates at line 7 only when  $\|\bar{c}_k\| \leq \epsilon_o$  and  $\Delta\bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) \leq \epsilon_o$ , it follows by Assumption 2.2 that  $\|c_k\| \leq \epsilon_c + \epsilon_o \leq 2\epsilon_c$ . Additionally, by (3.4)

$$\bar{\tau}_{K_T} \sigma_u \lambda_u \|\bar{u}_{K_T}\|^2 - \epsilon_o \leq \Delta\bar{l}(x_{K_T}, \bar{\tau}_{K_T}, \bar{d}_{K_T}) \leq \epsilon_o,$$

which implies that  $\|\bar{u}_{K_T}\|^2 \leq \frac{2\epsilon_o}{\bar{\tau}_{K_T} \sigma_u \lambda_u} \leq \frac{2\epsilon_c}{\bar{\tau}_{K_T} \sigma_u \lambda_u}$ . Moreover, by Assumptions 2.1 and 2.2 it follows that  $\|\bar{J}_{K_T}^T \bar{c}_{K_T}\| \leq (\kappa_J + \epsilon_J) \epsilon_o \leq (\kappa_J + \epsilon_J) \epsilon_c$  and  $\|J_{K_T}^T c_{K_T}\| \leq 2\kappa_J \epsilon_c$ . Thus, the stationarity result (3.66) holds by Lemmas 3.25 and 3.32, and the fact that  $\|u_{K_T}^*\|^2 \leq 2(\|\bar{u}_{K_T} - u_{K_T}^*\|^2 + \|\bar{u}_{K_T}\|^2)$ .  $\square$

Next, we show that if Algorithm 1 terminates finitely on Line 11, the *infeasible stationarity error* is sufficiently small.

**Theorem 3.34.** *Suppose that Assumptions 2.1, 2.2, 2.4, 3.1, and 3.2 hold. If Algorithm 1 terminates finitely at iteration  $k = K_T \in \mathbb{N}$  on Line 11 (i.e.,  $\|\bar{c}_{K_T}\| > \epsilon_o$  and  $\|\bar{J}_{K_T}^T \bar{c}_{K_T}\| = 0$ ), where  $\epsilon_o \in (0, \epsilon_c]$ , then  $\|J_{K_T}^T c_{K_T}\| \leq \kappa_c \epsilon_J + (\kappa_J + \epsilon_J) \epsilon_c$ .*

*Proof.* By Lemma 3.17,

$$\|J_{K_T}^T c_{K_T}\| \leq \|\bar{J}_{K_T}^T \bar{c}_{K_T}\| + \|J_{K_T}^T c_{K_T} - \bar{J}_{K_T}^T \bar{c}_{K_T}\| \leq \kappa_c \epsilon_J + (\kappa_J + \epsilon_J) \epsilon_c,$$

which concludes the statement.  $\square$

**Remark 3.35.** *By Theorem 3.33, we note that if Algorithm 1 ever terminates on Line 7, the terminated iterate is sufficiently feasible even if the constraint Jacobians are rank-deficient. Moreover, we provide an upper bound on the first-order stationarity error that depends on the noise level and merit parameter value at the terminated iterate ( $\bar{\tau}_{K_T}$ ). If the singular values of  $\bar{J}_k$  are bounded away from 0, then we have  $\bar{\tau}_{K_T} \geq \bar{\tau}_{\min}$ . However, if  $\bar{J}_k$  is rank-deficient, then  $\bar{\tau}_{K_T}$  can be arbitrarily close to zero, in which case the stationarity error can be large. By Theorem 3.34, we note that if Algorithm 1 ever terminates on Line 11, the infeasible stationarity error at terminated iterate is sufficiently small.*

We set up some general conditions necessary for the step size strategies required for Theorem 3.36, 3.38, and 3.40. If Algorithm 1 employs **Option I** for the step size scheme,  $\beta$  must satisfy the conditions outlined in Lemma 3.14. Alternatively, if Algorithm 1 employs **Option II**, the relaxation parameter  $\epsilon_{A_k}$  is defined in (3.19). Under these conditions, we show that when  $\bar{\tau}_k$  is bounded away from zero, the averaged stationarity and infeasible stationarity errors converge to a sufficiently small value, the size of which depends on the noise levels.

**Theorem 3.36.** *Suppose Assumptions 2.1, 2.2, 2.4, 3.1, 3.2, 3.4, and 3.22 hold. In addition, suppose that there exists  $\bar{\tau}_{\min} \in \mathbb{R}_{>0}$  such that  $\bar{\tau}_k \geq \bar{\tau}_{\min}$  for all  $k \in \mathbb{N}$ . Let  $\phi_{\min} = \min\{0, \bar{\tau}_{-1} f_{\text{inf}}\}$ , and let  $i \in \{A, L\}$  denote the adaptive (**Option I**) and line search (**Option II**) scheme, respectively, in Algorithm 1. Then, for all  $k \in \mathbb{N}$ ,*

$$\frac{1}{k+1} \sum_{j=0}^k \|J_j^T c_j\|^2 \leq \frac{(\bar{\tau}_{-1}|f_{\text{inf}}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min})(\kappa_c + \epsilon_c)}{(k+1)\sigma_c \sigma_v \eta \alpha_{l\tau}^i} + \frac{\kappa_c + \epsilon_c}{\sigma_c \sigma_v \eta \alpha_{l\tau}^i} \mathcal{E}^i + \mathcal{E}_{Jc^2}, \quad (3.67)$$

and

$$\begin{aligned} & \frac{1}{k+1} \sum_{j=0}^k \frac{\|Z_j^T (\nabla f(x_j) + H_j v_j)\|^2}{\kappa_H^2} \\ & \leq \left( \frac{4+2\kappa_1}{\bar{\tau}_{\min} \sigma_u \lambda_u \eta \alpha_{l\tau}^i} + \frac{(\kappa_1+2\kappa_2)(\kappa_c + \epsilon_c)}{\sigma_c \sigma_v \eta \alpha_{l\tau}^i} \right) \frac{\bar{\tau}_{-1}|f_{\text{inf}}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min}}{k+1} \\ & \quad + \left( \frac{4+2\kappa_1}{\bar{\tau}_{\min} \sigma_u \lambda_u \eta \alpha_{l\tau}^i} + \frac{(\kappa_1+2\kappa_2)(\kappa_c + \epsilon_c)}{\sigma_c \sigma_v \eta \alpha_{l\tau}^i} \right) \mathcal{E}^i \\ & \quad + 2\kappa_2 \mathcal{E}_{Jc^2} + \kappa_1 (\kappa_J + \epsilon_J)^2 \epsilon_c^2 + 12\zeta^{-2} (\epsilon_g^2 + \kappa_{\bar{y}^*}^2 \epsilon_J^2 + \kappa_{y^*}^2 \epsilon_J^2), \end{aligned} \quad (3.68)$$

where  $\kappa_1 = 4\lambda_{pr}^2 (\zeta^{-1} + (\kappa_J + \epsilon_J)\kappa_\sigma^{-2})^2$  and  $\kappa_2 = 24\zeta^{-2}\kappa_H^2\sigma_{Jc}^2$ .

If the singular values of  $J_k$  are bounded below by  $\kappa_\sigma \in \mathbb{R}_{>0}$  for all  $k \in \mathbb{N}$ ,

$$\frac{1}{k+1} \sum_{j=0}^k \|c_j\|^2 \leq \frac{(\bar{\tau}_{-1}|f_{\text{inf}}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min})(\kappa_c + \epsilon_c)}{(k+1)\sigma_c \sigma_v \eta \alpha_{l\tau}^i \kappa_\sigma^2} + \frac{\kappa_c + \epsilon_c}{\sigma_c \sigma_v \eta \alpha_{l\tau}^i \kappa_\sigma^2} \mathcal{E}^i + \frac{1}{\kappa_\sigma^2} \mathcal{E}_{Jc^2}. \quad (3.69)$$

If the singular values of  $\bar{J}_k$  are bounded below by  $\kappa_\sigma \in \mathbb{R}_{>0}$  for all  $k \in \mathbb{N}$ ,

$$\frac{1}{k+1} \sum_{j=0}^k \|c_j\| \leq \frac{\bar{\tau}_{-1}|f_{\text{inf}}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min}}{(k+1)\kappa_\sigma^2 \sigma_c \sigma_v \eta \alpha_{l\tau}^i} + \frac{1}{\kappa_\sigma^2 \sigma_c \sigma_v \eta \alpha_{l\tau}^i} \mathcal{E}^i + 2\epsilon_c. \quad (3.70)$$

*Proof.* It follows by Assumption 2.1 and Lemmas 3.9, 3.13 and 3.31 that

$$\begin{aligned} & \phi(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_{k+1}) - \phi(x_k, \bar{\tau}_k) \\ & = \phi(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_k) - \phi(x_k, \bar{\tau}_k) + \phi(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_{k+1}) - \phi(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_k) \end{aligned}$$

$$\begin{aligned}
&\leq -\eta\alpha_{l\tau}^i\Delta\bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) + \mathcal{E}^i + (\bar{\tau}_{k+1} - \bar{\tau}_k)f_{\text{inf}} \\
&\leq -\eta\alpha_{l\tau}^i\left(\frac{\bar{\tau}_k\sigma_u\lambda_u}{2}\|\bar{u}_k\|^2 + \frac{\sigma_c\sigma_v}{\kappa_c+\epsilon_c}\|\bar{J}_k^T\bar{c}_k\|^2\mathbb{1}\{\|\bar{c}_k\| > \epsilon_o\}\right) + \mathcal{E}^i + (\bar{\tau}_{k+1} - \bar{\tau}_k)f_{\text{inf}}. \quad (3.71)
\end{aligned}$$

When  $f_{\text{inf}} \geq 0$  it follows that  $\phi(x_k, \bar{\tau}_k) = \bar{\tau}_k f_k + \|c_k\| \geq 0$  and when  $f_{\text{inf}} < 0$  it follows that  $\phi(x_k, \bar{\tau}_k) \geq \bar{\tau}_{-1} f_{\text{inf}}$ . Hence, for all  $k \in \mathbb{N}$ , it follows that

$$\phi(x_k, \bar{\tau}_k) \geq \min\{0, \bar{\tau}_{-1} f_{\text{inf}}\} := \phi_{\min}. \quad (3.72)$$

Summing (3.71) for  $j \in \{0, 1, \dots, k\}$ , it follows by (3.72) that

$$\begin{aligned}
&\phi_{\min} - \phi(x_0, \bar{\tau}_0) \\
&\leq \phi(x_{k+1}, \bar{\tau}_{k+1}) - \phi(x_0, \bar{\tau}_0) \\
&\leq -\frac{\bar{\tau}_k\sigma_u\lambda_u\eta\alpha_{l\tau}^i}{2}\sum_{j=0}^k\|\bar{u}_j\|^2 - \frac{\sigma_c\sigma_v\eta\alpha_{l\tau}^i}{\kappa_c+\epsilon_c}\sum_{j=0}^k(\|\bar{J}_j^T\bar{c}_j\|^2\mathbb{1}\{\|\bar{c}_j\| > \epsilon_o\}) + \bar{\tau}_{-1}|f_{\text{inf}}| + (k+1)\mathcal{E}^i. \quad (3.73)
\end{aligned}$$

It then follows by (3.73) that

$$\frac{\sigma_c\sigma_v\eta\alpha_{l\tau}^i}{\kappa_c+\epsilon_c}\sum_{j=0}^k\|\bar{J}_j^T\bar{c}_j\|^2\mathbb{1}\{\|\bar{c}_j\| > \epsilon_o\} \leq (k+1)\mathcal{E}^i + \bar{\tau}_{-1}|f_{\text{inf}}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min}. \quad (3.74)$$

Since  $\left|\|\bar{J}_j^T\bar{c}_j\|^2\mathbb{1}\{\|\bar{c}_j\| > \epsilon_o\} - \|\bar{J}_j^T\bar{c}_j\|^2\right| \leq \|\bar{J}_j^T\bar{c}_j\|^2\mathbb{1}\{\|\bar{c}_j\| \leq \epsilon_o\} \leq (\kappa_J + \epsilon_J)^2\epsilon_c^2$ , we have

$$\begin{aligned}
\frac{\sigma_c\sigma_v\eta\alpha_{l\tau}^i}{\kappa_c+\epsilon_c}\sum_{j=0}^k\|\bar{J}_j^T\bar{c}_j\|^2 &\leq \frac{\sigma_c\sigma_v\eta\alpha_{l\tau}^i}{\kappa_c+\epsilon_c}\sum_{j=0}^k(\|\bar{J}_j^T\bar{c}_j\|^2\mathbb{1}\{\|\bar{c}_j\| > \epsilon_o\} + \|\bar{J}_j^T\bar{c}_j\|^2\mathbb{1}\{\|\bar{c}_j\| \leq \epsilon_o\}) \\
&\leq (k+1)\left(\mathcal{E}^i + \frac{\sigma_c\sigma_v\eta\alpha_{l\tau}^i}{\kappa_c+\epsilon_c}(\kappa_J + \epsilon_J)^2\epsilon_c^2\right) + \bar{\tau}_{-1}|f_{\text{inf}}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min}. \quad (3.75)
\end{aligned}$$

By (3.22) and (3.75), it follows that

$$\begin{aligned}
\frac{\sigma_c\sigma_v\eta\alpha_{l\tau}^i}{\kappa_c+\epsilon_c}\sum_{j=0}^k\|J_j^T c_j\|^2 &\leq \frac{\sigma_c\sigma_v\eta\alpha_{l\tau}^i}{\kappa_c+\epsilon_c}\sum_{j=0}^k(\|\bar{J}_j^T\bar{c}_j\|^2\mathbb{1}\{\|\bar{c}_j\| > \epsilon_o\} + \|\bar{J}_j^T\bar{c}_j\|^2\mathbb{1}\{\|\bar{c}_j\| > \epsilon_o\}) \\
&\leq (k+1)\left(\mathcal{E}^i + \frac{\sigma_c\sigma_v\eta\alpha_{l\tau}^i}{\kappa_c+\epsilon_c}\mathcal{E}_{Jc^2}\right) + \bar{\tau}_{-1}|f_{\text{inf}}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min}. \quad (3.76)
\end{aligned}$$

By re-arranging the above, it follows that (3.67) is satisfied.

Additionally, it follows by (3.73) that

$$\frac{\bar{\tau}_{\min}\sigma_u\lambda_u\eta\alpha_{l\tau}^i}{2}\sum_{j=0}^k\|\bar{u}_j\|^2 \leq \frac{\bar{\tau}_k\sigma_u\lambda_u\eta\alpha_{l\tau}^i}{2}\sum_{j=0}^k\|\bar{u}_j\|^2$$

$$\leq \bar{\tau}_{-1}|f_{\inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min} + (k+1)\mathcal{E}^i. \quad (3.77)$$

By Lemmas 3.25 and 3.32, it follows that

$$\begin{aligned}
& \sum_{j=0}^k \left\| Z_j^T (\nabla f(x_j) + H_j v_j) \right\|^2 \\
& \leq \sum_{j=0}^k \kappa_H^2 \|u_j^*\|^2 \\
& \leq \sum_{j=0}^k \kappa_H^2 \left( 2 \|\bar{u}_j\|^2 + 2 \|u_j^* - \bar{u}_j\|^2 \right) \\
& \leq 2\kappa_H^2 \sum_{j=0}^k \|\bar{u}_j\|^2 + 4\kappa_H^2 \lambda_{\rho r}^2 (\zeta^{-1} + (\kappa_J + \epsilon_J) \kappa_\sigma^{-2})^2 \sum_{j=0}^k (\|\bar{u}_j\|^2 + \|\bar{J}_j^T \bar{c}_j\|^2) \\
& \quad + 24\zeta^{-2} \kappa_H^4 \sigma_{Jc}^2 \sum_{j=0}^k (\|\bar{J}_j^T \bar{c}_j\|^2 \mathbb{1}\{\|\bar{c}_j\| > \epsilon_o\} + \|J_j^T c_j\|^2) + 12(k+1)\kappa_H^2 \zeta^{-2} (\epsilon_g^2 + \kappa_{\bar{y}^*}^2 \epsilon_J^2 + \kappa_{y^*}^2 \epsilon_J^2) \\
& \leq (k+1) \frac{4\kappa_H^2 + 8\kappa_H^2 \lambda_{\rho r}^2 (\zeta^{-1} + (\kappa_J + \epsilon_J) \kappa_\sigma^{-2})^2}{\bar{\tau}_{\min} \sigma_u \lambda_u \eta \alpha_{l\tau}^i} \mathcal{E}^i + \frac{4\kappa_H^2 + 8\kappa_H^2 \lambda_{\rho r}^2 (\zeta^{-1} + (\kappa_J + \epsilon_J) \kappa_\sigma^{-2})^2}{\bar{\tau}_{\min} \sigma_u \lambda_u \eta \alpha_{l\tau}^i} (\bar{\tau}_{-1}|f_{\inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min}) \\
& \quad + 4\kappa_H^2 \lambda_{\rho r}^2 (\zeta^{-1} + (\kappa_J + \epsilon_J) \kappa_\sigma^{-2})^2 \frac{\kappa_c + \epsilon_c}{\sigma_c \sigma_v \eta \alpha_{l\tau}^i} \left( (k+1) \left( \mathcal{E}^i + \frac{\sigma_c \sigma_v \eta \alpha_{l\tau}^i}{\kappa_c + \epsilon_c} (\kappa_J + \epsilon_J)^2 \epsilon_c^2 \right) + \bar{\tau}_{-1}|f_{\inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min} \right) \\
& \quad + 24\zeta^{-2} \kappa_H^4 \sigma_{Jc}^2 \frac{\kappa_c + \epsilon_c}{\sigma_c \sigma_v \eta \alpha_{l\tau}^i} \left( (k+1) \left( 2\mathcal{E}^i + \frac{\sigma_c \sigma_v \eta \alpha_{l\tau}^i}{\kappa_c + \epsilon_c} \mathcal{E}_{Jc^2} \right) + 2\bar{\tau}_{-1}|f_{\inf}| + 2\phi(x_0, \bar{\tau}_0) - 2\phi_{\min} \right) \\
& \quad + 12(k+1)\kappa_H^2 \zeta^{-2} (\epsilon_g^2 + \kappa_{\bar{y}^*}^2 \epsilon_J^2 + \kappa_{y^*}^2 \epsilon_J^2) \\
& \leq (k+1) \left( \frac{4\kappa_H^2 + 8\kappa_H^2 \lambda_{\rho r}^2 (\zeta^{-1} + (\kappa_J + \epsilon_J) \kappa_\sigma^{-2})^2}{\bar{\tau}_{\min} \sigma_u \lambda_u \eta \alpha_{l\tau}^i} + \left( 48\zeta^{-2} \kappa_H^4 \sigma_{Jc}^2 + 4\kappa_H^2 \lambda_{\rho r}^2 (\zeta^{-1} + (\kappa_J + \epsilon_J) \kappa_\sigma^{-2})^2 \right) \frac{\kappa_c + \epsilon_c}{\sigma_c \sigma_v \eta \alpha_{l\tau}^i} \right) \mathcal{E}^i \\
& \quad + 24(k+1)\zeta^{-2} \kappa_H^4 \sigma_{Jc}^2 \mathcal{E}_{Jc^2} + 4(k+1)\kappa_H^2 \lambda_{\rho r}^2 (\zeta^{-1} + (\kappa_J + \epsilon_J) \kappa_\sigma^{-2})^2 (\kappa_J + \epsilon_J)^2 \epsilon_c^2 \\
& \quad + 12(k+1)\kappa_H^2 \zeta^{-2} (\epsilon_g^2 + \kappa_{\bar{y}^*}^2 \epsilon_J^2 + \kappa_{y^*}^2 \epsilon_J^2) \\
& \quad + \frac{4\kappa_H^2 + 8\kappa_H^2 \lambda_{\rho r}^2 (\zeta^{-1} + (\kappa_J + \epsilon_J) \kappa_\sigma^{-2})^2}{\bar{\tau}_{\min} \sigma_u \lambda_u \eta \alpha_{l\tau}^i} (\bar{\tau}_{-1}|f_{\inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min}) \\
& \quad + \left( 48\zeta^{-2} \kappa_H^4 \sigma_{Jc}^2 + 4\kappa_H^2 \lambda_{\rho r}^2 (\zeta^{-1} + (\kappa_J + \epsilon_J) \kappa_\sigma^{-2})^2 \right) \frac{\kappa_c + \epsilon_c}{\sigma_c \sigma_v \eta \alpha_{l\tau}^i} (\bar{\tau}_{-1}|f_{\inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min}).
\end{aligned}$$

Thus, (3.68) is satisfied.

Starting with (3.76), if the singular values of  $J_k$  are bounded below by  $\kappa_\sigma$  for all  $k \in \mathbb{N}$ , (3.69) is then satisfied by (3.9).

Finally, if the singular values of  $\bar{J}_k$  are bounded below by  $\kappa_\sigma \in \mathbb{R}_{>0}$  for all  $k \in \mathbb{N}$ , it follows by Assumption 2.1, (2.7) and Lemmas 3.9 and 3.31 that

$$\phi(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_{k+1}) - \phi(x_k, \bar{\tau}_k)$$



$$\begin{aligned}
&\leq -\eta\alpha_{l\tau}^i\Delta\bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) + \mathcal{E}^i + (\bar{\tau}_{k+1} - \bar{\tau}_k)f_{\text{inf}} \\
&\leq -\eta\alpha_{l\tau}^i\left(\frac{\bar{\tau}_k\sigma_u\lambda_u}{2}\|\bar{u}_k\|^2 + \kappa_\sigma^2\sigma_c\sigma_v\|\bar{c}_k\|\mathbb{1}\{\|\bar{c}_j\| > \epsilon_o\}\right) + \mathcal{E}^i + (\bar{\tau}_{k+1} - \bar{\tau}_k)f_{\text{inf}}. \quad (3.78)
\end{aligned}$$

Summing (3.78) for  $j \in \{0, 1, \dots, k\}$  and re-arranging, it follows that

$$\eta\alpha_{l\tau}^i\kappa_\sigma^2\sigma_c\sigma_v\sum_{j=0}^k\|\bar{c}_j\|\mathbb{1}\{\|\bar{c}_j\| > \epsilon_o\} \leq (k+1)\mathcal{E}^i + \bar{\tau}_{-1}|f_{\text{inf}}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min}.$$

Since

$$\begin{aligned}
\|\|\bar{c}_k\|\mathbb{1}\{\|\bar{c}_k\| > \epsilon_o\} - \|c_k\|\| &\leq \max\{\|\|\bar{c}_k\| - \|c_k\|\|, \|\|c_k\|\mathbb{1}\{\|\bar{c}_k\| \leq \epsilon_o\}\|\| \\
&\leq \max\{\epsilon_c, \epsilon_c + \epsilon_o\} \leq 2\epsilon_c,
\end{aligned}$$

it follows that

$$\eta\alpha_{l\tau}^i\kappa_\sigma^2\sigma_c\sigma_v\sum_{j=0}^k\|c_j\| \leq (k+1)(\mathcal{E}^i + 2\eta\alpha_{l\tau}^i\kappa_\sigma^2\sigma_c\sigma_v\epsilon_c) + \bar{\tau}_{-1}|f_{\text{inf}}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min},$$

which completes the proof.  $\square$

**Remark 3.37.** *Theorem 3.36 provides the convergence guarantees for Algorithm 1 with either step size strategy when the merit parameter is bounded away from zero. In the most general setting, we can only prove that the averaged square of both the stationarity error and the infeasibility stationarity error converges to a neighborhood of zero as  $k \rightarrow \infty$ . By Lemmas 3.17 and 3.31, it follows that  $\mathcal{E}^A = \mathcal{O}(\epsilon_g + \epsilon_c + \epsilon_J)$ ,  $\mathcal{E}^L = \mathcal{O}(\epsilon_f + \epsilon_g + \epsilon_c + \epsilon_J)$ , and  $\mathcal{E}_{Jc^2} = \mathcal{O}(\epsilon_c + \epsilon_J)$  when  $\epsilon_c \leq 1$ . Thus, the neighborhood size is therefore  $\mathcal{O}(\epsilon_f + \epsilon_g + \epsilon_c + \epsilon_J)$ . If the singular values of  $J_k$  are bounded below by  $\kappa_\sigma$ , we prove that the averaged square of the feasibility error (3.69) converges to a neighborhood of zero as  $k \rightarrow \infty$ , the size of which is  $\mathcal{O}(\epsilon_f + \epsilon_g + \epsilon_c + \epsilon_J)$ . Furthermore, if the singular values of  $\bar{J}_k$  are bounded below by  $\kappa_\sigma$ , we can prove a stronger result; the averaged feasibility error (3.70) converges to a neighborhood of zero as  $k \rightarrow \infty$ , the size of which is still  $\mathcal{O}(\epsilon_f + \epsilon_g + \epsilon_c + \epsilon_J)$ . When  $\epsilon_f = \epsilon_g = \epsilon_c = \epsilon_J = 0$ , Theorem 3.36 recovers (3.9) and (3.11) in [19, Theorem 3.3] for a noise-free SQP method with rank-deficient Jacobians.*

**Theorem 3.38.** *Suppose Assumptions 2.1, 2.2, 2.4, 3.1, 3.2, and 3.4 hold. In addition, suppose that the merit parameter sequence  $\{\bar{\tau}_k\}$  diminishes to zero. Let  $\mathcal{E}_c^A := 2\epsilon_c + \alpha_u^A\kappa_{\bar{d}}\epsilon_J$  and  $\mathcal{E}_c^L := 6\epsilon_c + \alpha_u^L\kappa_{\bar{d}}\epsilon_J$ , respectively, and let  $i \in \{A, L\}$  denote the adaptive (**Option I**) and line search (**Option II**) scheme, respectively, in Algorithm 1. Then, for some  $\sigma_\gamma \in (0, 1)$ , for all  $k \in \mathbb{N}$ ,*

$$\liminf_{k \rightarrow \infty} \|J_k^T c_k\| \leq \max\left\{\epsilon_c(\epsilon_J + \kappa_J) + \kappa_c\epsilon_J + \sqrt{\frac{(1+\sigma_\gamma)(\kappa_c + \epsilon_c)\mathcal{E}_c^i}{\eta\alpha_{l_0}^i\sigma_c\sigma_v}}, 2\kappa_J\epsilon_c\right\}.$$

In addition, if the singular values of  $J_k$  are bounded below by  $\kappa_\sigma \in \mathbb{R}_{>0}$ , for all  $k \in \mathbb{N}$ ,

$$\liminf_{k \rightarrow \infty} \|c_k\| \leq \kappa_\sigma^{-1} \max \left\{ \epsilon_c(\epsilon_J + \kappa_J) + \kappa_c \epsilon_J + \sqrt{\frac{(1+\sigma_\gamma)(\kappa_c + \epsilon_c)\mathcal{E}_c^i}{\eta \alpha_{i_0}^i \sigma_c \sigma_v}}, 2\kappa_J \epsilon_c \right\}.$$

*Proof.* Our goal is to prove that for any  $\delta \in \mathbb{R}_{>0}$ , there exist infinitely many iterates  $x_k$  for which

$$\|\bar{J}_k^T \bar{c}_k\| < \sqrt{\frac{(1+\sigma_\gamma)(\kappa_c + \epsilon_c)\mathcal{E}_c^i}{\eta \alpha_{i_0}^i \sigma_c \sigma_v}} + \delta \text{ or } \|\bar{c}_k\| < \epsilon_o + \delta. \quad (3.79)$$

To derive a contradiction, we assume that there exists some  $\delta \in \mathbb{R}_{>0}$  and some  $k_\gamma \in \mathbb{N}$  such that for all  $k \geq k_\gamma$ ,

$$\|\bar{J}_k^T \bar{c}_k\| \geq \sqrt{\frac{(1+\sigma_\gamma)(\kappa_c + \epsilon_c)\mathcal{E}_c^i}{\eta \alpha_{i_0}^i \sigma_c \sigma_v}} + \delta \text{ and } \|\bar{c}_k\| \geq \epsilon_o + \delta.$$

Let  $\gamma := \sqrt{\frac{(1+\sigma_\gamma)(\kappa_c + \epsilon_c)\mathcal{E}_c^i}{\eta \alpha_{i_0}^i \sigma_c \sigma_v}}$  for some  $\sigma_\gamma \in (0, 1)$ . By the definition of  $\mathcal{X}_\gamma$  (see (3.16)), we have  $x_k \in \mathcal{X}_\gamma$  for all  $k \geq k_\gamma$ .

Let  $\mathcal{E}_f^A := \alpha_u^A \kappa_{\bar{d}} \epsilon_g$  and  $\mathcal{E}_f^L := 4\epsilon_f + \alpha_u^L \kappa_{\bar{d}} \epsilon_g$ , respectively. It then follows that for all  $k \geq k_\gamma$ ,

$$\begin{aligned} \|c_k\| - \|c_{k+1}\| &\geq \bar{\tau}_k (f_{k+1} - f_k) + \eta \alpha_{i_0}^i \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) - \mathcal{E}_f^i \bar{\tau}_k - \mathcal{E}_c^i \\ &\geq \eta \alpha_{i_0}^i \sigma_c \sigma_v (\kappa_c + \epsilon_c)^{-1} \|\bar{J}_k^T \bar{c}_k\|^2 + \bar{\tau}_k (f_{\inf} - f_{\sup}) - \mathcal{E}_f^i \bar{\tau}_k - \mathcal{E}_c^i \\ &\geq (1 + \sigma_\gamma) \mathcal{E}_c^i + \bar{\tau}_k (f_{\inf} - f_{\sup}) - \mathcal{E}_f^i \bar{\tau}_k - \mathcal{E}_c^i \\ &= \sigma_\gamma \mathcal{E}_c^i + \bar{\tau}_k (f_{\inf} - f_{\sup}) - \mathcal{E}_f^i \bar{\tau}_k, \end{aligned} \quad (3.80)$$

where the first inequality follows by (2.7) and Lemmas 3.14 and 3.15, (3.62), (3.64), the second inequality follows by Assumption 2.1 and Lemma 3.9, and the third inequality follows by the fact that  $x_k \in \mathcal{X}_\gamma$ . Since  $\bar{\tau}_k \rightarrow 0$ , there exists sufficiently large  $K_\gamma \geq k_\gamma$  such that when  $k \geq K_\gamma$ ,  $\bar{\tau}_k (f_{\sup} - f_{\inf}) + \mathcal{E}_f^i \bar{\tau}_k \leq \frac{1}{2} \sigma_\gamma \mathcal{E}_c^i$ , which implies that  $\|c_k\| - \|c_{k+1}\| \geq \frac{1}{2} \sigma_\gamma \mathcal{E}_c^i > 0$ .

As a result, for some iteration  $k_c$  it follows that  $\|c_{k_c}\| < \epsilon_o + \delta$ , which leads to a contradiction. Hence, there exists a subsequence  $\mathcal{K} \subset \{k_\gamma, k_\gamma + 1, \dots\}$  such that (3.79) is satisfied for all  $k \in \mathcal{K}$  and any  $\delta \in \mathbb{R}_{>0}$ . If the former case in (3.79) is satisfied, it follows by Lemma 3.17 that

$$\|J_k^T c_k\| \leq \|J_k^T c_k - \bar{J}_k^T \bar{c}_k\| + \|\bar{J}_k^T \bar{c}_k\| \leq (\kappa_J + \epsilon_J) \epsilon_c + \kappa_c \epsilon_J + \sqrt{\frac{(1+\sigma_\gamma)(\kappa_c + \epsilon_c)\mathcal{E}_c^i}{\eta \alpha_{i_0}^i \sigma_c \sigma_v}} + \delta$$

and if the latter case in (3.79) is satisfied it follows by Assumption 2.1 and 2.2 that

$$\|J_k^T c_k\| \leq \|J_k^T \bar{c}_k\| + \|J_k^T (\bar{c}_k - c_k)\| \leq \kappa_J (\epsilon_o + \delta + \epsilon_c) \leq 2\kappa_J \epsilon_c + \kappa_J \delta.$$

Hence, for any  $\delta \in \mathbb{R}_{>0}$  and all  $k \in \mathcal{K}$ , we have

$$\|J_k^T c_k\| < \max \left\{ \epsilon_c(\epsilon_J + \kappa_J) + \kappa_c \epsilon_J + \sqrt{\frac{(1+\sigma_\gamma)(\kappa_c + \epsilon_c)\mathcal{E}_c^i}{\eta \alpha_{i0}^2 \sigma_c \sigma_v}}, 2\kappa_J \epsilon_c \right\} + \max\{\kappa_J, 1\}\delta.$$

By choosing an arbitrarily small value for  $\delta > 0$ , the first conclusion is therefore satisfied.

If in addition we assume that the singular values of  $J_k$  are bounded below by  $\kappa_\sigma$  for all  $k \in \mathbb{N}$ , the second conclusion holds as  $\|J_k^T c_k\| \geq \kappa_\sigma \|c_k\|$ .  $\square$

**Remark 3.39.** *Theorem 3.38 provides convergence guarantees for Algorithm 1 with either step size strategy when the merit parameter sequence vanishes to zero. In the most general setting, when  $\{\epsilon_c, \epsilon_J\} \subset (0, 1]$ , we prove that there exists a subsequence of  $\|J_k^T c_k\|$  (infeasible stationarity measure) that converges to a neighborhood of zero whose size is  $\mathcal{O}((\epsilon_c + \epsilon_J)^{1/2})$ . Contrary to Theorem 3.36, we cannot prove convergence for the entire sequence of iterates since the step size is not necessarily lower bounded if  $x_k \notin \mathcal{X}_\gamma$ . This result is consistent with [45, Proposition 1], which shows that the iterates visit a critical region related to  $\|J_k^T c_k\|$  infinitely often. If we additionally assume that the singular values of  $J_k$  are bounded below by  $\kappa_\sigma$ , we can prove a similar lim inf-type result for  $\|c_k\|$ . Finally, under the condition that  $\epsilon_c = \epsilon_J = 0$ , Theorem 3.38 implies that  $\liminf_{k \rightarrow \infty} \|J_k^T c_k\| = 0$ . This result differs from equation (3.10) in [19, Theorem 3.3], which demonstrates the stronger conclusion that  $\lim_{k \rightarrow \infty} \|J_k^T c_k\| = 0$ . The discrepancy arises due to the step size scheme in the presence of noise, making our result the strongest possible under such conditions. Our lim inf-type result recovers the corresponding conclusion in [7, Theorem 3] when  $\epsilon_c = \epsilon_J = 0$ .*

We derive worst-case complexity bounds for the iterates generated by Algorithm 1 in different settings. We use  $K_{inf}$  and  $K_{fea}$  to denote the number of iterations required until

$$\|J_k^T c_k\| \leq \epsilon \tag{3.81}$$

$$\text{and} \quad \|c_k\| \leq \epsilon \tag{3.82}$$

are guaranteed to be satisfied, respectively. Moreover, we use  $K_{si}$  to denote the number of iterations required until

$$\|Z_k^T (\nabla f(x_k) + H_k v_k)\| \leq \epsilon \quad \text{and} \quad \|J_k^T c_k\| \leq \epsilon \tag{3.83}$$

are guaranteed to both be satisfied, we use  $K_{sf}$  to denote the number of iterations required until

$$\|Z_k^T (\nabla f(x_k) + H_k v_k)\| \leq \epsilon \quad \text{and} \quad \|c_k\| \leq \epsilon \tag{3.84}$$

are guaranteed to both be satisfied, and, finally, we use  $K_{\overline{sf}}$  to denote the number of iterations required until

$$\|Z_k^T (\nabla f(x_k) + H_k v_k)\| \leq \epsilon \quad \text{and} \quad \sqrt{\|c_k\|} \leq \epsilon \tag{3.85}$$

are both satisfied. Theorems 3.36 and 3.38 present neighborhood convergence guarantees for Algorithm 1 in different settings. As a result, the worst case complexity results presented below explicitly define this neighborhood, i.e., for the different convergence measures stated above the target accuracy levels  $\epsilon \in \mathbb{R}_{>0}$  are bounded below. Such conditions are analogous to those for noise-aware methods for noisy unconstrained optimization problems [6, 12, 29].

**Theorem 3.40.** *Suppose Assumptions 2.1, 2.2, 2.4, 3.1, 3.2, 3.4, and 3.22 hold. Let  $\phi_{\min} = \min\{0, \bar{\tau}_{-1} f_{\inf}\}$ , and let  $i \in \{A, L\}$  denote the adaptive (**Option I**) and line search (**Option II**) scheme, respectively, in Algorithm 1. In addition, let  $\sigma_\epsilon \in \mathbb{R}_{>1}$  and suppose that*

$$\epsilon^2 > \max \left\{ 4\mathcal{E}_{Jc}^2, \sigma_\epsilon \left( \mathcal{E}_{Jc^1} + \frac{\kappa_c + \epsilon_c}{\eta \alpha_{l_0}^i \sigma_c \sigma_v} \mathcal{E}^i \right), \sigma_\epsilon \left( \mathcal{E}_{Jc^2} + \frac{\kappa_c + \epsilon_c}{\eta \alpha_{l_r}^i \sigma_c \sigma_v} \mathcal{E}^i \right) \right\}, \quad (3.86)$$

where  $\mathcal{E}_{Jc^1} := ((\kappa_J + \epsilon_J)\epsilon_c + \kappa_c \epsilon_J) ((\kappa_J + \epsilon_J)(\kappa_c + \epsilon_c) + \kappa_J \kappa_c)$ . Then, (3.81) is guaranteed to be satisfied after

$$K_{inf} = \left\lceil \frac{(\kappa_c + \epsilon_c)(\bar{\tau}_{-1} |f_{\inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min})}{\eta \min\{\alpha_{l_0}^i, \alpha_{l_r}^i\} \sigma_c \sigma_v (1 - \sigma_\epsilon^{-1}) \epsilon^2} \right\rceil$$

iterations. Further, if the singular values of  $J_k$  are bounded below by  $\kappa_\sigma \in \mathbb{R}_{>0}$  for all  $k \in \mathbb{N}$  and

$$\epsilon^2 > \max \left\{ \frac{4\mathcal{E}_{Jc}^2}{\kappa_\sigma^2}, \sigma_\epsilon \left( \frac{1}{\kappa_\sigma^2} \mathcal{E}_{Jc^1} + \frac{\kappa_c + \epsilon_c}{\eta \min\{\alpha_{l_0}^i, \alpha_{l_r}^i\} \sigma_c \sigma_v \kappa_\sigma^2} \mathcal{E}^i \right), \sigma_\epsilon \left( \frac{1}{\kappa_\sigma^2} \mathcal{E}_{Jc^2} + \frac{\kappa_c + \epsilon_c}{\eta \alpha_{l_r}^i \sigma_c \sigma_v \kappa_\sigma^2} \mathcal{E}^i \right) \right\}, \quad (3.87)$$

then, (3.82) is guaranteed to be satisfied after

$$K_{fea} = \left\lceil \frac{(\kappa_c + \epsilon_c)(\bar{\tau}_{-1} |f_{\inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min})}{\eta \min\{\alpha_{l_0}^i, \alpha_{l_r}^i\} \sigma_c \sigma_v \kappa_\sigma^2 (1 - \sigma_\epsilon^{-1}) \epsilon^2} \right\rceil$$

iterations.

Moreover, suppose that there exists  $\bar{\tau}_{\min} \in \mathbb{R}_{>0}$  such that  $\bar{\tau}_k \geq \bar{\tau}_{\min}$  for all  $k \in \mathbb{N}$  and

$$\epsilon^2 > \sigma_\epsilon \left( \kappa_{si} \mathcal{E}^i + (2\kappa_H^2 \kappa_2 + 1) \mathcal{E}_{Jc^2} + \kappa_H^2 \left( \kappa_1 (\kappa_J + \epsilon_J)^2 \epsilon_c^2 + 12\zeta^{-2} (\epsilon_g^2 + \kappa_{y^*}^2 \epsilon_J^2 + \kappa_{y^*}^2 \epsilon_J^2) \right) \right), \quad (3.88)$$

then, (3.83) is guaranteed to be satisfied after

$$K_{si} = \left\lceil \frac{\kappa_{si}(\bar{\tau}_{-1} |f_{\inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min})}{(1 - \sigma_\epsilon^{-1}) \epsilon^2} \right\rceil$$

iterations. Further, if the singular values of  $J_k$  are bounded below by  $\kappa_\sigma \in \mathbb{R}_{>0}$  for all  $k \in \mathbb{N}$  and

$$\epsilon^2 > \sigma_\epsilon \left( \kappa_{sf} \mathcal{E}^i + (2\kappa_H^2 \kappa_2 + \kappa_\sigma^{-2}) \mathcal{E}_{Jc^2} + \kappa_H^2 \left( \kappa_1 (\kappa_J + \epsilon_J)^2 \epsilon_c^2 + 12\zeta^{-2} (\epsilon_g^2 + \kappa_{y^*}^2 \epsilon_J^2 + \kappa_{y^*}^2 \epsilon_J^2) \right) \right), \quad (3.89)$$

then, (3.84) is guaranteed to be satisfied after

$$K_{sf} = \left\lceil \frac{\kappa_{sf}(\bar{\tau}_{-1}|f_{\inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min})}{(1 - \sigma_\epsilon^{-1})\epsilon^2} \right\rceil$$

iterations. Finally, if the singular values of  $\bar{J}_k$  are bounded below by  $\kappa_\sigma \in \mathbb{R}_{>0}$  for all  $k \in \mathbb{N}$ , and

$$\epsilon^2 > \sigma_\epsilon \left( \kappa_{\bar{sf}} \mathcal{E}^i + 2\kappa_H^2 \kappa_2 \mathcal{E}_{Jc^2} + 2\epsilon_c + \kappa_H^2 (\kappa_1(\kappa_J + \epsilon_J)^2 \epsilon_c^2 + 12\zeta^{-2} (\epsilon_g^2 + \kappa_{\bar{y}^*}^2 \epsilon_J^2 + \kappa_{\bar{y}^*}^2 \epsilon_J^2)) \right), \quad (3.90)$$

then, (3.85) is guaranteed to be satisfied after

$$K_{\bar{sf}} = \left\lceil \frac{\kappa_{\bar{sf}}(\bar{\tau}_{-1}|f_{\inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min})}{(1 - \sigma_\epsilon^{-1})\epsilon^2} \right\rceil$$

iterations. Note that  $\kappa_{si}$ ,  $\kappa_{sf}$ , and  $\kappa_{\bar{sf}}$  are all problem-dependent constants and are explicitly defined in the proof.

*Proof.* We prove the first two general results with regard to  $K_{inf}$  and  $K_{fea}$  by discussing two cases: (i)  $\lim_{k \rightarrow \infty} \bar{\tau}_k = 0$  and (ii)  $\bar{\tau}_k \geq \bar{\tau}_{\min}$  for all  $k \in \mathbb{N}$ .

When  $\lim_{k \rightarrow \infty} \bar{\tau}_k = 0$ , we use  $K_{inf}$  to denote the number of iterations required until (3.81) is satisfied at an iteration. Since for all  $k < K_{inf}$ ,  $\|J_k^T c_k\| > \epsilon$ , it follows from Lemma 3.17 that  $\|\bar{J}_k^T \bar{c}_k\| \geq \|J_k^T c_k\| - \mathcal{E}_{Jc} > \epsilon - \mathcal{E}_{Jc} > \mathcal{E}_{Jc}$ . Additionally, it follows from (3.21) and Cauchy-Schwarz inequality that  $\|\bar{c}_k\| \geq \|c_k\| - \epsilon_c \geq \frac{\|J_k^T c_k\|}{\kappa_J} - \epsilon_c > \frac{2\mathcal{E}_{Jc}}{\kappa_J} - \epsilon_c \geq \epsilon_c \geq \epsilon_o$ . The lower bounds  $\alpha_{i0}^A$  and  $\alpha_{i0}^L$  provided in Lemmas 3.14 and 3.15 are therefore valid with  $\gamma = \mathcal{E}_{Jc}$ . It then follows from (3.2), (3.22), (3.57), and (3.61) that

$$\begin{aligned} \phi(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_k) - \phi(x_k, \bar{\tau}_k) &\leq -\eta \alpha_{i0}^i \Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) + \mathcal{E}^i \\ &\leq -\eta \alpha_{i0}^i \sigma_c \sigma_v (\kappa_c + \epsilon_c)^{-1} \|\bar{J}_k^T \bar{c}_k\|^2 + \mathcal{E}^i \\ &\leq -\eta \alpha_{i0}^i \sigma_c \sigma_v (\kappa_c + \epsilon_c)^{-1} \|J_k^T c_k\|^2 + \eta \alpha_{i0}^i \sigma_c \sigma_v (\kappa_c + \epsilon_c)^{-1} \mathcal{E}_{Jc^1} + \mathcal{E}^i, \end{aligned}$$

where  $\mathcal{E}_{Jc^1} := ((\kappa_J + \epsilon_J)\epsilon_c + \kappa_c \epsilon_J) ((\kappa_J + \epsilon_J)(\kappa_c + \epsilon_c) + \kappa_J \kappa_c)$ . The last inequality is satisfied since  $|\|J_k^T c_k\|^2 - \|\bar{J}_k^T \bar{c}_k\|^2| \leq \mathcal{E}_{Jc^1}$  is proved in (3.22). It then follows that

$$\begin{aligned} &\phi(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_{k+1}) - \phi(x_k, \bar{\tau}_k) \\ &\leq -\eta \alpha_{i0}^i \sigma_c \sigma_v (\kappa_c + \epsilon_c)^{-1} \|J_k^T c_k\|^2 + \eta \alpha_{i0}^i \sigma_c \sigma_v (\kappa_c + \epsilon_c)^{-1} \mathcal{E}_{Jc^1} + \mathcal{E}^i + (\bar{\tau}_{k+1} - \bar{\tau}_k) f_{\inf}. \end{aligned}$$

Summing the inequality above for all iterations within  $\{0, 1, \dots, K_{inf} - 1\}$  and doing rearrangement, we have from (3.86) and (3.72) that

$$\begin{aligned} &K_{inf} \eta \alpha_{i0}^i \sigma_c \sigma_v (\kappa_c + \epsilon_c)^{-1} \min_{j \in \{0, 1, \dots, K_{inf} - 1\}} \|J_j^T c_j\|^2 \leq \sum_{j=0}^{K_{inf}-1} \eta \alpha_{i0}^i \sigma_c \sigma_v (\kappa_c + \epsilon_c)^{-1} \|J_j^T c_j\|^2 \\ &\leq K_{inf} (\eta \alpha_{i0}^i \sigma_c \sigma_v (\kappa_c + \epsilon_c)^{-1} \mathcal{E}_{Jc^1} + \mathcal{E}^i) + \bar{\tau}_{-1} |f_{\inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min} \\ &\leq K_{inf} \eta \alpha_{i0}^i \sigma_c \sigma_v (\kappa_c + \epsilon_c)^{-1} \sigma_\epsilon^{-1} \epsilon^2 + \bar{\tau}_{-1} |f_{\inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min}. \end{aligned} \quad (3.91)$$

Since for all  $k < K_{inf}$ , we always have  $\|J_k^T c_k\| > \epsilon$ . It then follows by (3.91) that

$$\frac{\kappa_c + \epsilon_c}{\eta \alpha_{l_0}^i \sigma_c \sigma_v} > \frac{K_{inf}}{\bar{\tau}_{-1} |f_{inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{min}} (1 - \sigma_\epsilon^{-1}) \epsilon^2,$$

and we require at most  $K_{inf} = \left\lceil \frac{(\kappa_c + \epsilon_c)(\bar{\tau}_{-1} |f_{inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{min})}{\eta \alpha_{l_0}^i \sigma_c \sigma_v (1 - \sigma_\epsilon^{-1}) \epsilon^2} \right\rceil$  many iterations such that (3.81) is satisfied.

When  $\bar{\tau}_k \geq \bar{\tau}_{min}$  for all  $k \in \mathbb{N}$ , it follows from (3.67) that

$$\min_{j \in \{0, \dots, K_{inf} - 1\}} \|J_j^T c_j\|^2 \leq \frac{\kappa_c + \epsilon_c}{\sigma_c \sigma_v \eta \alpha_{l_\tau}^i} \mathcal{E}^i + \mathcal{E}_{Jc^2} + \frac{(\bar{\tau}_{-1} |f_{inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{min})(\kappa_c + \epsilon_c)}{K_{inf} \sigma_c \sigma_v \eta \alpha_{l_\tau}^i},$$

which implies by (3.86) that

$$\frac{(\bar{\tau}_{-1} |f_{inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{min})(\kappa_c + \epsilon_c)}{K_{inf} \sigma_c \sigma_v \eta \alpha_{l_\tau}^i} > \epsilon^2 - \frac{\kappa_c + \epsilon_c}{\sigma_c \sigma_v \eta \alpha_{l_\tau}^i} \mathcal{E}^i - \mathcal{E}_{Jc^2} > (1 - \sigma_\epsilon^{-1}) \epsilon^2,$$

and we require at most  $K_{inf} = \left\lceil \frac{(\kappa_c + \epsilon_c)(\bar{\tau}_{-1} |f_{inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{min})}{\eta \alpha_{l_\tau}^i \sigma_c \sigma_v (1 - \sigma_\epsilon^{-1}) \epsilon^2} \right\rceil$  iterations such that (3.81) is satisfied. The first result is therefore satisfied by combining the two cases.

If the singular values of  $J_k$  are bounded below by  $\kappa_\sigma$  for all  $k \in \mathbb{N}$ , the second desired conclusion (see  $K_{fea}$  and (3.82)) is directly satisfied by  $\|J_k^T c_k\| \geq \kappa_\sigma \|c_k\|$ .

When  $\bar{\tau}_k \geq \bar{\tau}_{min}$  for all  $k \in \mathbb{N}$ , combining (3.67) with a rearrangement of (3.68) yields

$$\begin{aligned} & \min_{j \in \{0, \dots, K_{si} - 1\}} \|Z_j^T (\nabla f(x_j) + H_j v_j)\|^2 + \|J_j^T c_j\|^2 \\ & \leq \left( \frac{\kappa_H^2 (4 + 2\kappa_1)}{\bar{\tau}_{min} \sigma_u \lambda_u \eta \alpha_{l_\tau}^i} + \frac{\kappa_H^2 (\kappa_1 + 2\kappa_2)(\kappa_c + \epsilon_c)}{\sigma_c \sigma_v \eta \alpha_{l_\tau}^i} + \frac{\kappa_c + \epsilon_c}{\sigma_c \sigma_v \eta \alpha_{l_\tau}^i} \right) \frac{\bar{\tau}_{-1} |f_{inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{min}}{K_{si}} \\ & \quad + \left( \frac{\kappa_H^2 (4 + 2\kappa_1)}{\bar{\tau}_{min} \sigma_u \lambda_u \eta \alpha_{l_\tau}^i} + \frac{\kappa_H^2 (\kappa_1 + 2\kappa_2)(\kappa_c + \epsilon_c)}{\sigma_c \sigma_v \eta \alpha_{l_\tau}^i} + \frac{\kappa_c + \epsilon_c}{\sigma_c \sigma_v \eta \alpha_{l_\tau}^i} \right) \mathcal{E}^i \\ & \quad + (2\kappa_H^2 \kappa_2 + 1) \mathcal{E}_{Jc^2} + \kappa_H^2 \kappa_1 (\kappa_J + \epsilon_J)^2 \epsilon_c^2 + 12\kappa_H^2 \zeta^{-2} (\epsilon_g^2 + \kappa_{\bar{y}^*}^2 \epsilon_J^2 + \kappa_{\bar{y}^*}^2 \epsilon_J^2). \end{aligned} \quad (3.92)$$

Let  $\kappa_{si} = \frac{\kappa_H^2 (4 + 2\kappa_1)}{\bar{\tau}_{min} \sigma_u \lambda_u \eta \alpha_{l_\tau}^i} + \frac{(\kappa_H^2 (\kappa_1 + 2\kappa_2) + 1)(\kappa_c + \epsilon_c)}{\sigma_c \sigma_v \eta \alpha_{l_\tau}^i}$ . We use  $K_{si}$  to denote the number of iterations required until (3.83) is satisfied at an iteration. Since for all  $k < K_{si}$ ,  $\|Z_k^T (\nabla f(x_k) + H_k v_k)\|^2 + \|J_k^T c_k\|^2 > \epsilon^2$ , it follows by (3.88) and (3.92) that

$$\kappa_{si} > \frac{K_{si}}{\bar{\tau}_{-1} |f_{inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{min}} (1 - \sigma_\epsilon^{-1}) \epsilon^2,$$

and the third desired conclusion (see  $K_{si}$  and (3.83)) is satisfied. If the singular values of  $J_k$  are bounded below by  $\kappa_\sigma$  for all  $k \in \mathbb{N}$ , combining (3.68) with a rearrangement of (3.69)

yields

$$\begin{aligned}
& \min_{j \in \{0, \dots, K_{sf}-1\}} \left\| Z_j^T (\nabla f(x_j) + H_j v_j) \right\|^2 + \|c_j\|^2 \\
& \leq \left( \frac{\kappa_H^2 (4+2\kappa_1)}{\bar{\tau}_{\min} \sigma_u \lambda_u \eta \alpha_{l\tau}^i} + \frac{\kappa_H^2 (\kappa_1+2\kappa_2)(\kappa_c+\epsilon_c)}{\sigma_c \sigma_v \eta \alpha_{l\tau}^i} + \frac{\kappa_c+\epsilon_c}{\sigma_c \sigma_v \eta \alpha_{l\tau}^i \kappa_\sigma^2} \right) \frac{\bar{\tau}_{-1} |f_{\inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min}}{K_{sf}} \\
& \quad + \left( \frac{\kappa_H^2 (4+2\kappa_1)}{\bar{\tau}_{\min} \sigma_u \lambda_u \eta \alpha_{l\tau}^i} + \frac{\kappa_H^2 (\kappa_1+2\kappa_2)(\kappa_c+\epsilon_c)}{\sigma_c \sigma_v \eta \alpha_{l\tau}^i} + \frac{\kappa_c+\epsilon_c}{\sigma_c \sigma_v \eta \alpha_{l\tau}^i \kappa_\sigma^2} \right) \mathcal{E}^i \\
& \quad + (2\kappa_H^2 \kappa_2 + \kappa_\sigma^{-2}) \mathcal{E}_{Jc^2} + \kappa_H^2 \kappa_1 (\kappa_J + \epsilon_J)^2 \epsilon_c^2 + 12\kappa_H^2 \zeta^{-2} (\epsilon_g^2 + \kappa_{\bar{y}^*}^2 \epsilon_J^2 + \kappa_{\bar{y}^*}^2 \epsilon_J^2).
\end{aligned} \tag{3.93}$$

Let  $\kappa_{sf} = \frac{\kappa_H^2 (4+2\kappa_1)}{\bar{\tau}_{\min} \sigma_u \lambda_u \eta \alpha_{l\tau}^i} + \frac{(\kappa_\sigma^2 \kappa_H^2 (\kappa_1+2\kappa_2)+1)(\kappa_c+\epsilon_c)}{\sigma_c \sigma_v \eta \alpha_{l\tau}^i \kappa_\sigma^2}$ . We use  $K_{sf}$  to denote the number of iterations required until (3.84) is satisfied at an iteration. Since for all  $k < K_{sf}$ ,  $\|Z_k^T (\nabla f(x_k) + H_k v_k)\|^2 + \|c_k\|^2 > \epsilon^2$ , it follows by (3.89) and (3.93) that

$$\kappa_{sf} > \frac{K_{sf}}{\bar{\tau}_{-1} |f_{\inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min}} (1 - \sigma_\epsilon^{-1}) \epsilon^2,$$

and the fourth desired conclusion (see  $K_{sf}$  and (3.84)) is satisfied. Finally, if the singular values of  $\bar{J}_k$  are bounded below by  $\kappa_\sigma$  for all  $k \in \mathbb{N}$ , by (3.68) and (3.70), we have

$$\begin{aligned}
& \min_{j \in \{0, \dots, K_{sf}^- - 1\}} \left\| Z_j^T (\nabla f(x_j) + H_j v_j) \right\|^2 + \|c_j\|^2 \\
& \leq \left( \frac{\kappa_H^2 (4+2\kappa_1)}{\bar{\tau}_{\min} \sigma_u \lambda_u \eta \alpha_{l\tau}^i} + \frac{\kappa_H^2 (\kappa_1+2\kappa_2)(\kappa_c+\epsilon_c)}{\sigma_c \sigma_v \eta \alpha_{l\tau}^i} + \frac{1}{\sigma_c \sigma_v \eta \alpha_{l\tau}^i \kappa_\sigma^2} \right) \frac{\bar{\tau}_{-1} |f_{\inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min}}{K_{sf}^-} \\
& \quad + \left( \frac{\kappa_H^2 (4+2\kappa_1)}{\bar{\tau}_{\min} \sigma_u \lambda_u \eta \alpha_{l\tau}^i} + \frac{\kappa_H^2 (\kappa_1+2\kappa_2)(\kappa_c+\epsilon_c)}{\sigma_c \sigma_v \eta \alpha_{l\tau}^i} + \frac{1}{\sigma_c \sigma_v \eta \alpha_{l\tau}^i \kappa_\sigma^2} \right) \mathcal{E}^i \\
& \quad + 2\kappa_H^2 \kappa_2 \mathcal{E}_{Jc^2} + 2\epsilon_c + \kappa_H^2 \kappa_1 (\kappa_J + \epsilon_J)^2 \epsilon_c^2 + 12\kappa_H^2 \zeta^{-2} (\epsilon_g^2 + \kappa_{\bar{y}^*}^2 \epsilon_J^2 + \kappa_{\bar{y}^*}^2 \epsilon_J^2).
\end{aligned} \tag{3.94}$$

Let  $\kappa_{sf}^- = \frac{\kappa_H^2 (4+2\kappa_1)}{\bar{\tau}_{\min} \sigma_u \lambda_u \eta \alpha_{l\tau}^i} + \frac{\kappa_\sigma^2 \kappa_H^2 (\kappa_1+2\kappa_2)(\kappa_c+\epsilon_c)+1}{\sigma_c \sigma_v \eta \alpha_{l\tau}^i \kappa_\sigma^2}$ . We use  $K_{sf}^-$  to denote the number of iterations required until (3.85) is satisfied at an iteration. Since for all  $k < K_{sf}^-$ , we always have  $\|Z_k^T (\nabla f(x_k) + H_k v_k)\|^2 + \|c_k\|^2 > \epsilon^2$ , it follows by (3.90) and (3.94) that

$$\kappa_{sf}^- > \frac{K_{sf}^-}{\bar{\tau}_{-1} |f_{\inf}| + \phi(x_0, \bar{\tau}_0) - \phi_{\min}} (1 - \sigma_\epsilon^{-1}) \epsilon^2,$$

and the final desired conclusion (see  $K_{sf}^-$  and (3.85)) is satisfied.  $\square$

**Remark 3.41.** *Theorem 3.40 provides worst-case iteration complexity guarantees for Algorithm 1 under the conditions that  $\epsilon$  is sufficiently large compared to  $\mathcal{O}((\epsilon_g + \epsilon_c + \epsilon_J)^{1/2})$  where  $\{\epsilon_g, \epsilon_c, \epsilon_J\} \subset (0, 1]$ . The results of the theorem can be summarized as follows: (i) In the most general case, Algorithm 1 requires  $\mathcal{O}(\epsilon^{-2})$  iterations to achieve the desired accuracy in terms of the infeasible stationarity error; (ii) If additionally  $\{J_k\}$  is with full rank, then Algorithm 1 can also achieve the desired feasibility error in at most  $\mathcal{O}(\epsilon^{-2})$  iterations;*

(iii) When  $\bar{\tau}_k \geq \bar{\tau}_{\min}$  for all  $k \in \mathbb{N}$ , Algorithm 1 requires  $\mathcal{O}(\epsilon^{-2})$  iterations to achieve both the desired accuracy in terms of stationarity error and infeasible stationarity error. (iv) When  $\bar{\tau}_k \geq \bar{\tau}_{\min}$  and  $\{J_k\}$  or  $\{\bar{J}_k\}$  are of full rank, then Algorithm 1 can also achieve the desired accuracy in terms of stationarity error and feasibility error in at most  $\mathcal{O}(\epsilon^{-2})$  iterations. The only difference is that when  $\{\bar{J}_k\}$  is of full rank, the desired accuracy in terms of feasibility error can be improved (see (3.85)). The target accuracies for each case are consistent with Table 2. In the setting with no noise, i.e.,  $\epsilon_f = \epsilon_g = \epsilon_c = \epsilon_J = 0$ , the conditions on  $\epsilon$  only require that these accuracy parameters are positive, which could be arbitrarily close to zero. Under the LICQ, the algorithm recovers the complexity results of the deterministic SQP method [20].

## 4 Numerical Experiments

In this section, we present numerical results that showcase the performance of the proposed algorithm applied to standard equality-constrained nonlinear optimization problems from the CUTEst collection [26], with different levels of manually introduced noise. We selected a subset of equality-constrained optimization problems from the CUTEst collection where the objective function is non-constant and the LICQ holds at all iterations for the deterministic line search SQP method [36]. This resulted in 88 problems. The performance of our algorithm is evaluated under two scenarios: (i) when the LICQ is satisfied and (ii) when the LICQ is violated. The latter scenario is carefully induced to ensure fair experiments.

To simulate noise, we add uniformly distributed random perturbations to the exact function and gradient evaluations of both the objective and constraints for each problem. This results in the noisy quantities  $(\bar{f}_k, \bar{g}_k, \bar{c}_k, \bar{J}_k)$ , which are used in the algorithms. Specifically, for  $e_f \sim \mathcal{U}(-\epsilon_f, \epsilon_f)$ ,  $e_g \sim \mathcal{U}\left(-\frac{\epsilon_g}{\sqrt{n}}\mathbf{1}_n, \frac{\epsilon_g}{\sqrt{n}}\mathbf{1}_n\right)$ ,  $e_c \sim \mathcal{U}\left(-\frac{\epsilon_c}{\sqrt{m}}\mathbf{1}_m, \frac{\epsilon_c}{\sqrt{m}}\mathbf{1}_m\right)$ , and  $e_J \sim \mathcal{U}\left(-\frac{\epsilon_J}{\sqrt{mn}}\mathbf{1}_{m \times n}, \frac{\epsilon_J}{\sqrt{mn}}\mathbf{1}_{m \times n}\right)$ , we set

$$\bar{f}(x) = f(x) + e_f, \quad \bar{c}(x) = c(x) + e_c, \quad \bar{g}(x) = g(x) + e_g, \quad \text{and} \quad \bar{J}(x) = J(x) + e_J,$$

where  $\{\epsilon_f, \epsilon_g, \epsilon_c, \epsilon_J\} \subset \mathbb{R}_{>0}$  dictate the level of the noise. We consider four different noise levels in the objective function and constraint functions,  $\epsilon_f \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$  and  $\epsilon_c \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$ , and all combinations, and set  $\epsilon_g = \sqrt{\epsilon_f}$  and  $\epsilon_J = \sqrt{\epsilon_c}$  to align the gradient noise levels with the upper bounds established for common DFO methods [31].

To implement the adaptive step size scheme (AdaSQP), we set  $\eta = 0.5$  and  $\beta = 1$  and estimate the Lipschitz constants of the objective gradient and the constraint Jacobian using gradient differences near the initial point and keep these estimates constant. As for the line search scheme (LSSQP), we set  $\alpha_u^L = 1$ ,  $\eta = 10^{-3}$ , and  $\nu = 0.5$ . The other parameters were set as  $\bar{\tau}_{-1} = 1$ ,  $\lambda_u = 5 \times 10^{-9}$ ,  $\sigma_{Jc} = 10^2$ ,  $\sigma_u = 0.99$ ,  $\sigma_c = 0.1$ ,  $\sigma_r = 0.9999$ ,  $\sigma_\tau = 10^{-2}$ ,  $\xi_{-1} = 1$ ,  $\chi_{-1} = 10^{-3}$ ,  $\zeta_{-1} = 10^3$ , and  $\theta = 10^4$ , to ensure that the parameters interfered as minimally as possible with the algorithm. We set  $\epsilon_o = 0$  for the pessimistic



algorithms (**AdaSQP-pes** and **LSSQP-pes**) and set  $\epsilon_o = \epsilon_c$  for the optimistic algorithms (**AdaSQP-opt** and **LSSQP-opt**). We employ the CG-Steihaug method [36] to solve (2.2) potentially inexactly; the method is terminated when (2.3) holds and the residual vectors, defined as  $\bar{R}_{k,t} = \bar{J}_k^T \bar{J}_k \bar{v}_{k,t} + \bar{J}_k^T \bar{c}_k$  where  $\bar{v}_{k,t}$  denotes the  $t$ th iteration of the CG-Steihaug algorithm, satisfy

$$\|\bar{R}_{k,t}\|_\infty \leq \kappa_v \min\{\epsilon_c, \epsilon_f\} \max\{1, \|\bar{J}_k^T \bar{c}_k\|_\infty\}. \quad (4.1)$$

We employ the MINRES method [15] to potentially solve (2.4) inexactly; the method is terminated when at least one of Termination Test 1 or 2 holds and the residual vectors  $\bar{\rho}_{k,t}$  and  $\bar{r}_{k,t}$  (defined in (2.6) where the second subscript defines the iterations of the MINRES procedure) satisfy

$$\left\| \begin{bmatrix} \bar{\rho}_{k,t} \\ \bar{r}_{k,t} \end{bmatrix} \right\|_\infty \leq \kappa_u \min\{\epsilon_c, \epsilon_f\} \max\left\{\min\left\{\max\left\{\|\bar{u}_k\|_\infty, \|\bar{J}_k^T \bar{c}_k\|_\infty\right\}, 10^2\right\}, 10^{-2}\right\}, \quad (4.2)$$

Compared to (2.9), we add a  $\max\{\cdot, 10^{-2}\}$  term to ensure that the right-hand-side is not too small. For the inexact algorithms (**AdaSQP-inexact** and **LSSQP-inexact**), we use  $\kappa_u = \kappa_v = 10^{-2}$  in (4.1) and (4.2), and for the exact variants (**AdaSQP-exact** and **LSSQP-exact**) we replace  $\kappa_v \min\{\epsilon_c, \epsilon_f\}$  in (4.1) and  $\kappa_u \min\{\epsilon_c, \epsilon_f\}$  in (4.2) with  $10^{-10}$ .

We first test the performance of **AdaSQP-opt** and **LSSQP-opt** under different noise levels and exactness levels. We start by investigating the numerical performance of the proposed algorithms on the original CUTEst problem instances where the LICQ is satisfied. The feasibility and stationarity errors, defined as  $\|c_k\|_\infty$  and  $\|\nabla f(x_k) + J_k^T y_k\|_\infty$ , respectively, where  $y_k$  is a least-squares multiplier obtained using the true gradient  $\nabla f(x_k)$  and the true Jacobian  $J_k$ , are reported based on the *best iterate* from a maximum of 1000 iterations. The best iterate was selected as follows. For each problem instance, we selected the iterate with the lowest stationarity error, from the subset of iterates that satisfied the feasibility condition  $\|c_k\|_\infty \leq 2 \max\{\epsilon_c, \epsilon_f\}$ . If no iterate satisfied the feasibility condition, we instead reported the iterate with the smallest feasibility error. If the algorithm terminates early because of  $\|\bar{c}_k\| \leq \epsilon_o$  and  $\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) \leq \epsilon_o$ , the *best iterate* is selected similarly, but only from the iterates up to the termination. It can be observed from Figure 1 that the performance of the inexact method is comparable to that of the exact method for both step size schemes. When  $\epsilon_c$  is fixed, the feasibility error remains at a consistent level, while smaller values of  $\epsilon_f$  lead to lower stationarity errors. Similarly, when  $\epsilon_f$  is fixed, the stationarity error remains stable and the feasibility error scales with the noise level  $\epsilon_c$ .

Next, we consider the same problems, noise levels and inexactness but in the setting for which the LICQ is violated. To do this in a controlled manner and without affecting the feasible region, we duplicated the last constraint (after adding noise) for each instance in the original set. For problems where the LICQ is violated, we adopt a similar approach to report the best iterate. In addition, we report the infeasible stationarity error  $\|J_k^T c_k\|_\infty$  at that iterate. Figure 2 indicates that our method remains effective even when the LICQ is not satisfied. Furthermore, the best infeasible stationarity error decreases as  $\epsilon_c$  is reduced.

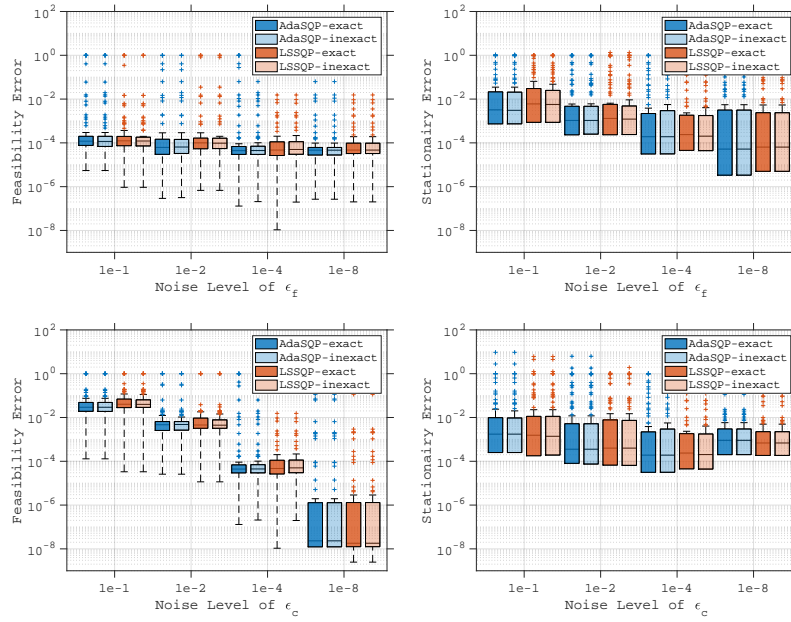


Figure 1: Box plots illustrating feasibility errors (left) and stationarity errors (right) for exact and inexact AdaSQP-opt and LSSQP-opt methods on CUTEst problems that satisfy the LICQ. First row:  $\epsilon_c = 10^{-4}$  and  $\epsilon_f \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$ . Second row:  $\epsilon_f = 10^{-4}$  and  $\epsilon_c \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$ .

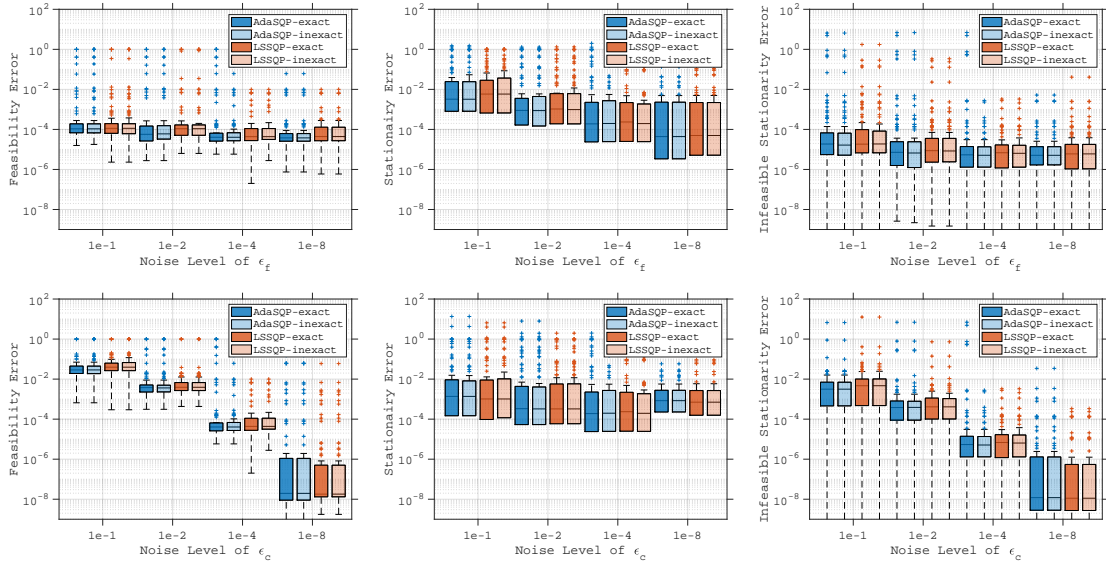


Figure 2: Box plots illustrating feasibility errors (left), stationarity errors (middle) and infeasible stationarity error (right) for exact and inexact AdaSQP-opt and LSSQP-opt methods on CUTEst problems that violate the LICQ. First row:  $\epsilon_c = 10^{-4}$  and  $\epsilon_f \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$ . Second row:  $\epsilon_f = 10^{-4}$  and  $\epsilon_c \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$ .

In the remainder of our experiments, all algorithms are run with a budget of 1,000 iterations and a total of 10,000 function evaluations, where the total number of function evaluations is defined as the sum of the number of function evaluations  $\bar{f}(\cdot)$  and two times the number of gradient evaluations  $\bar{g}(\cdot)$ . An algorithm is considered to have successfully solved a problem if an approximate stationary point is found, defined as

$$\|c(x_k)\|_\infty \leq 2 \max\{\epsilon_c, \epsilon_f\} \quad \text{and} \quad \|\nabla f(x_k) + \nabla c(x_k) y_k\|_\infty \leq 2(\epsilon_g + \|y_k\|_\infty \epsilon_J), \quad (4.3)$$

where  $y_k = \arg \min_{y \in \mathbb{R}^m} \|\nabla f(x_k) + \nabla c(x_k) y\|^2$ . If our proposed algorithm terminates early because of  $\|c(x_k)\|_\infty \leq \epsilon_o$  and  $\Delta \bar{l}(x_k, \bar{\tau}_k, \bar{d}_k) \leq \epsilon_o$ , it is deemed to have successfully solved the problem. For problems satisfying the LICQ, we consider both the pessimistic ( $\epsilon_o = 0$ ) and optimistic ( $\epsilon_o = \epsilon_c$ ) variants of **AdaSQP** and **LSSQP**, and compare our inexact methods with  $\kappa_v = \kappa_u = 10^{-2}$  against the Noise-Tolerant SQP (**NTSQP**) method proposed in [38], where in each iteration the following linear system

$$\begin{bmatrix} H_k & \bar{J}_k^T \\ \bar{J}_k & 0 \end{bmatrix} \begin{bmatrix} \bar{d}_k \\ \bar{y}_k \end{bmatrix} = - \begin{bmatrix} \bar{g}_k \\ \bar{c}_k \end{bmatrix}$$

is solved exactly using the MINRES procedure until

$$\left\| \begin{bmatrix} \bar{\rho}_{k,t} \\ \bar{r}_{k,t} \end{bmatrix} \right\|_\infty \leq 10^{-10} \max \left\{ \left\| \begin{bmatrix} \bar{g}_k \\ \bar{c}_k \end{bmatrix} \right\|_\infty, 10^{-2} \right\}, \quad (4.4)$$

to obtain the search direction. We present results in the form of Dolan-Moré performance profiles [23] in terms of function evaluations and MINRES iterations. While our proposed methods incur additional computational cost to solve problem (2.2), this cost can be mitigated by combining (2.2) and (2.4) into a single subproblem, similar to the **NTSQP** algorithm when LICQ is satisfied. Therefore, we focus on MINRES iterations to enable a fair comparison. In the first set of results, the LICQ is satisfied (Figures 3 and 4) and in the second set of results the LICQ is violated (Figures 5 and 6). We also showcase the sensitivity of our method to the inexactness parameters  $\kappa_u = \kappa_v \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$  (Figures 7 and 8).

Figures 3 and 4 illustrate the performance of the optimistic and pessimistic inexact variants of **AdaSQP** and **LSSQP**, and **NTSQP**, in terms of function evaluations and MINRES iterations, respectively, on problems where the LICQ holds and under varying noise levels. In general, across all noise levels and the two metrics of comparison, the optimistic variants consistently perform as well as or better than their pessimistic counterparts and **NTSQP**. In terms of function evaluations, **AdaSQP** is more efficient than **LSSQP**. This is due to the fact that **LSSQP** requires at least one additional function evaluation at every iteration to find a suitable step size. For certain noise instances (e.g.,  $\epsilon_f = \epsilon_c = 10^{-8}$ ), **AdaSQP** is able to solve fewer problem instances compared to **LSSQP**. This occurs because the step size for **AdaSQP** may be too conservative, which prevents the algorithm from achieving the desired

feasibility and stationarity within the budget. In terms of MINRES iterations, LSSQP is more efficient than AdaSQP. This is due to the fact that LSSQP ensures that the step size yields sufficient reduction in the merit function, whereas AdaSQP tends to adopt a more conservative step size. It is interesting to note that the optimistic variants solve a greater number of problem instances compared to the pessimistic methods in some noise settings (e.g.,  $\epsilon_f = 10^{-8}, \epsilon_c = 10^{-2}$ ). This can be attributed to the early termination condition and suggests that this condition can be utilized as a reliable termination condition for the algorithms. Although all methods solve fewer problem instances as the noise levels  $\epsilon_f$  and  $\epsilon_c$  decrease, they remain robust to changes in noise levels.

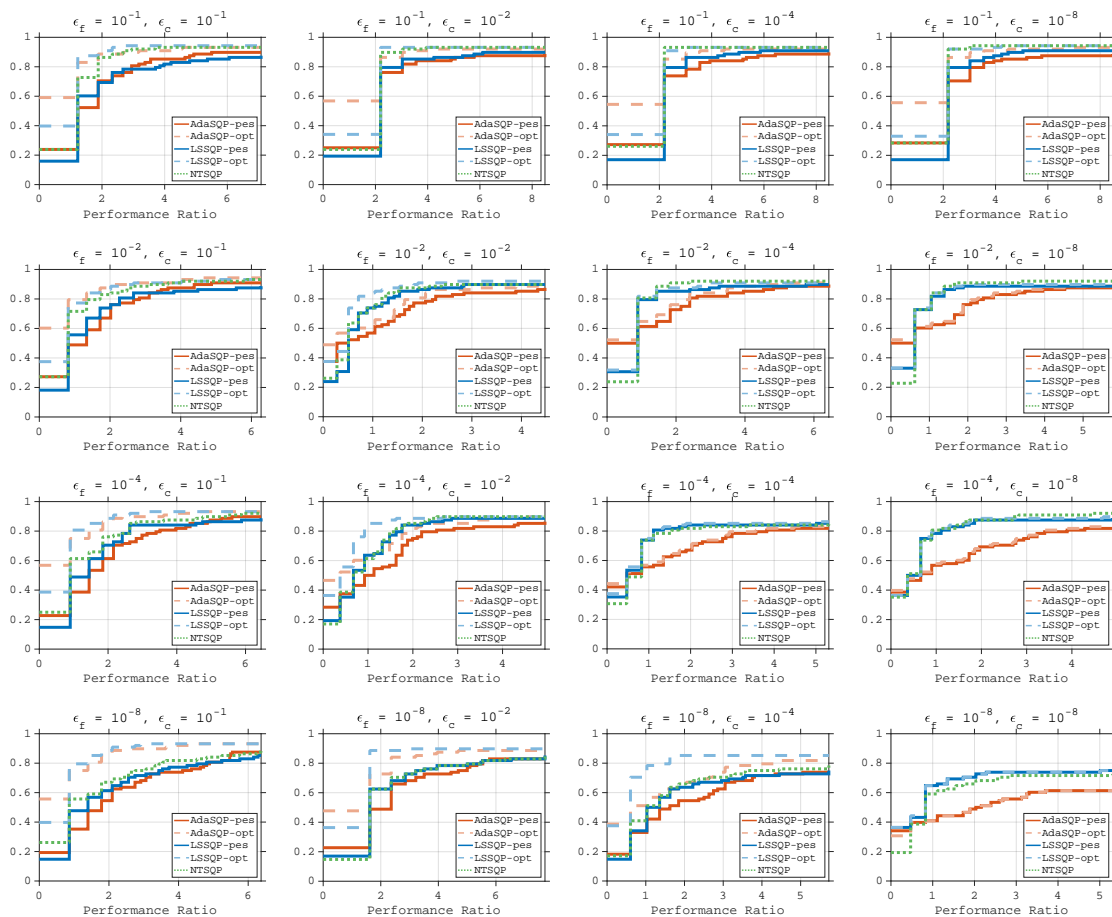


Figure 3: Dolan-Moré performance profiles comparing AdaSQP-pes, AdaSQP-opt, LSSQP-pes, LSSQP-opt and NTSQP on CUTEst collection of test problems that satisfy the LICQ in terms of number of **function evaluations** for  $\epsilon_c \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$  (from left to right) and  $\epsilon_f \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$  (from top to bottom).

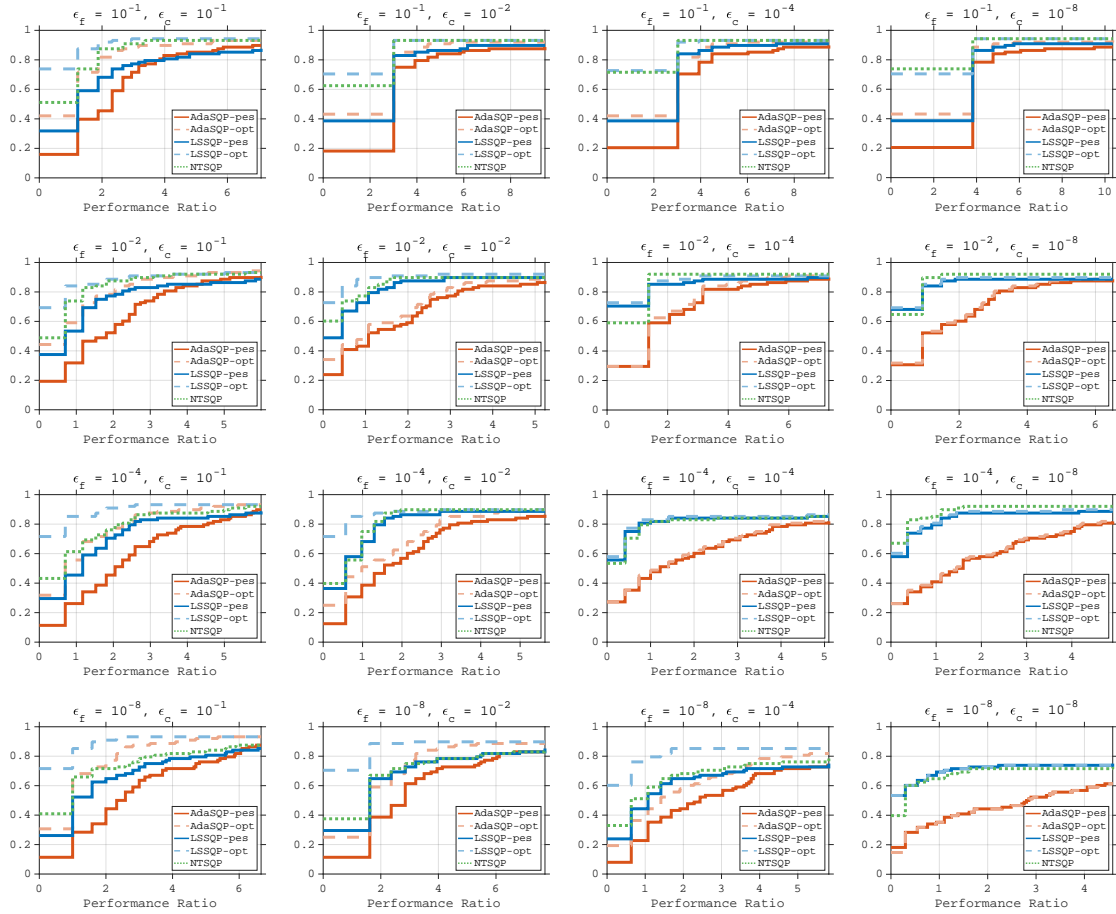


Figure 4: Dolan-Moré performance profiles comparing AdaSQP-pes, AdaSQP-opt, LSSQP-pes, LSSQP-opt and NTSQP on CUTEst collection of test problems that satisfy the LICQ in terms of **MINRES iterations** for  $\epsilon_c \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$  (from left to right) and  $\epsilon_f \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$  (from top to bottom).

To gain further insights into the pairwise comparison between the optimistic and pessimistic methods, we report the number of instances, out of 88 CUTEst problems, where the optimistic method terminates earlier than the pessimistic method, as well as the number of instances where the pessimistic method terminates earlier than the optimistic method. As shown in Table 4, for both **AdaSQP** and **LSSQP**, the optimistic variants terminated earlier for most of the problem instances and noise levels. The difference is more pronounced at high (objective function or constraint) noise levels. That said, there are a few problem instances where the optimistic method performs worse than the pessimistic method. This occurs because, even when  $\|\bar{c}_k\| \leq \epsilon_c$ , it may still be beneficial to continue to reduce the feasibility error in some cases since the simulated noise could be much smaller than the upper bound  $\epsilon_c$ .

$\epsilon_c$	$\epsilon_f$			
	$10^{-1}$	$10^{-2}$	$10^{-4}$	$10^{-8}$
$10^{-1}$	<b>(54/1, 45/0)</b>	<b>(47/1, 43/1)</b>	<b>(52/0, 47/0)</b>	<b>(57/0, 50/0)</b>
$10^{-2}$	<b>(44/0, 37/0)</b>	<b>(24/1, 27/0)</b>	<b>(36/2, 38/0)</b>	<b>(43/1, 44/0)</b>
$10^{-4}$	<b>(43/0, 40/0)</b>	(11/0, 7/0)	(8/7, 4/0)	<b>(37/3, 36/0)</b>
$10^{-8}$	<b>(46/0, 39/0)</b>	(4/0, 3/0)	(1/0, 2/0)	(2/9, 2/0)

Table 4: Number of problem instances for which the optimistic/pessimistic variants is better among the 88 CUTEst problems for both **AdaSQP** (first entry of the tuple) and **LSSQP** (second entry of the tuple).

Figures 5 and 6 illustrate the performance of the optimistic and pessimistic inexact variants of **AdaSQP** and **LSSQP** on the same 88 CUTEst problems (same as in Figures 3 and 4) but in the setting where the LICQ is violated in a controlled way (as described above). We do not compare to **NTSQP** since this method was designed for the setting in which LICQ holds. As shown in Figures 5 and 6, **AdaSQP** is more efficient than **LSSQP** when the evaluation metric is total function evaluations while **LSSQP** is more efficient when considering MINRES iterations. This observations are similar to those for the setting in which the LICQ is satisfied. Compared to the setting in which the LICQ is satisfied (Figures 3 and 4), the performance gap between the pessimistic and optimistic variants is smaller in this setting.

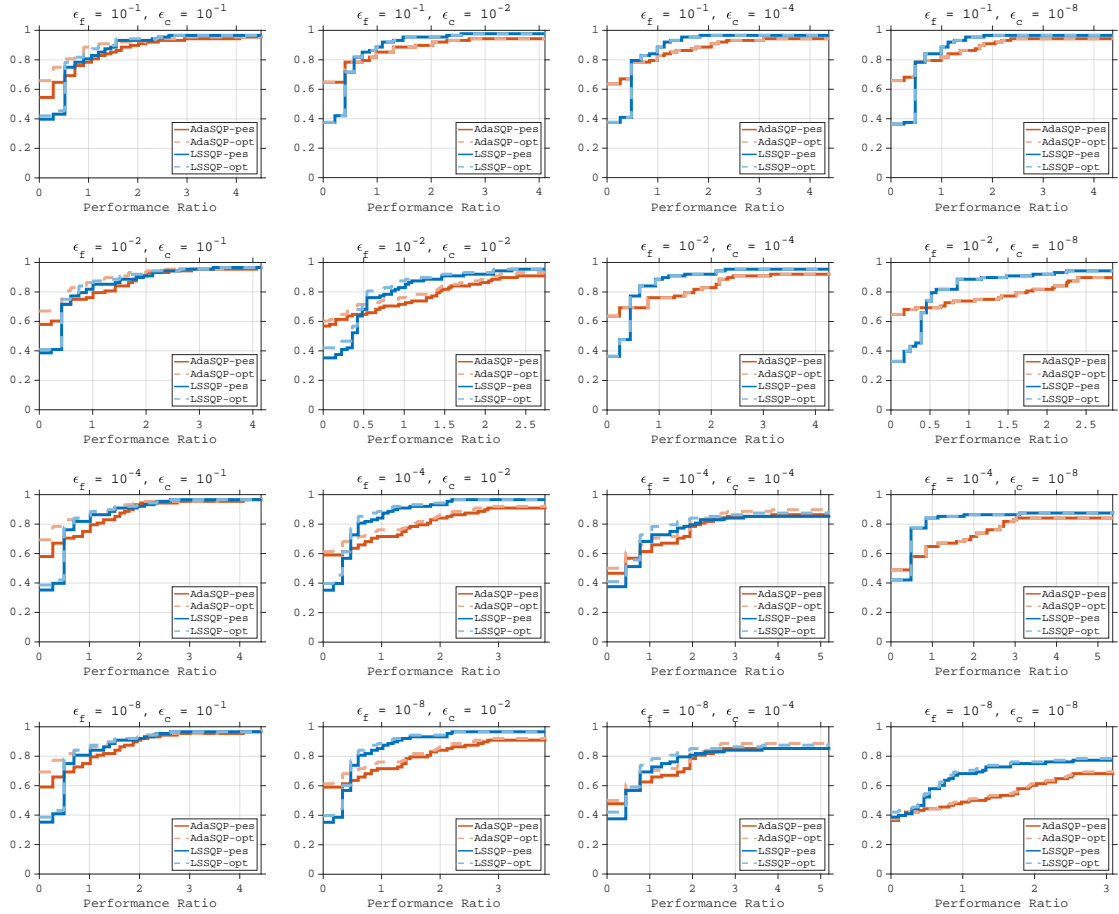


Figure 5: Dolan-Moré performance profiles comparing AdaSQP-pes, AdaSQP-opt, LSSQP-pes, and LSSQP-opt on CUTEst collection of test problems in the absence of the LICQ in terms of **function evaluations** for  $\epsilon_c \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$  (from left to right) and  $\epsilon_f \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$  (from top to bottom).



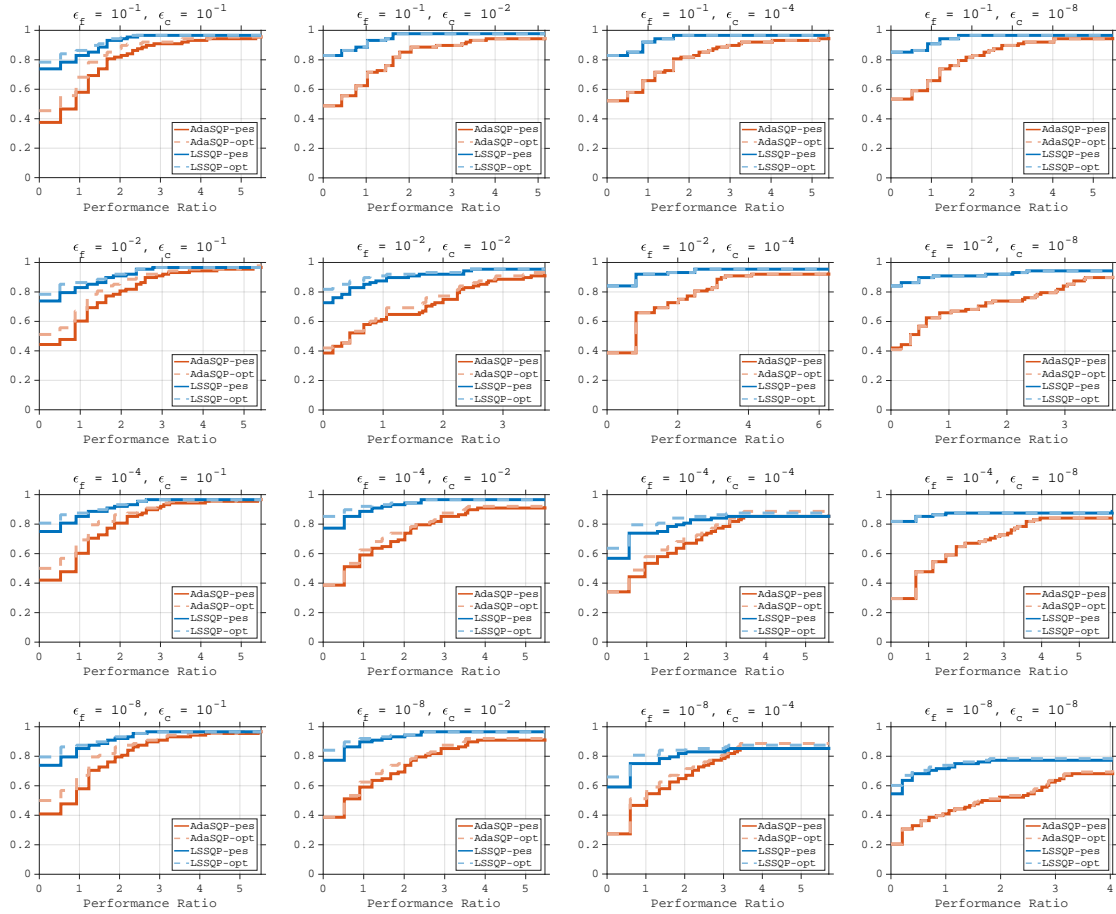


Figure 6: Dolan-Moré performance profiles comparing AdaSQP-pes, AdaSQP-opt, LSSQP-pes, and LSSQP-opt on CUTEst collection of test problems in the absence of the LICQ in terms of **MINRES iterations** for  $\epsilon_c \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$  (from left to right) and  $\epsilon_f \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$  (from top to bottom).

Finally, we compare the performance of the AdaSQP-opt and LSSQP-opt methods for different levels of inexactness in the subproblem solutions. Specifically, we set  $\kappa_u = \kappa_v \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$ . For conciseness, we present Dolan-Moré performance profiles in terms of both function evaluations and the total number of CG and MINRES iterations, focusing on the case where  $\epsilon_c = \epsilon_f \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$ . Figures 7 (AdaSQP-opt) and 8 (LSSQP-opt) demonstrate that our proposed algorithms are highly robust to changes in  $\kappa_u$  and  $\kappa_v$ . In particular, the most inexact method ( $\kappa_u = \kappa_v = 10^{-1}$ ) fails on only one additional problem instance compared to the other exactness levels when  $\epsilon_c = \epsilon_f = 10^{-4}$ . That said, in general, this inexactness level is more efficient than the other levels in terms of the total number of CG and MINRES iterations.

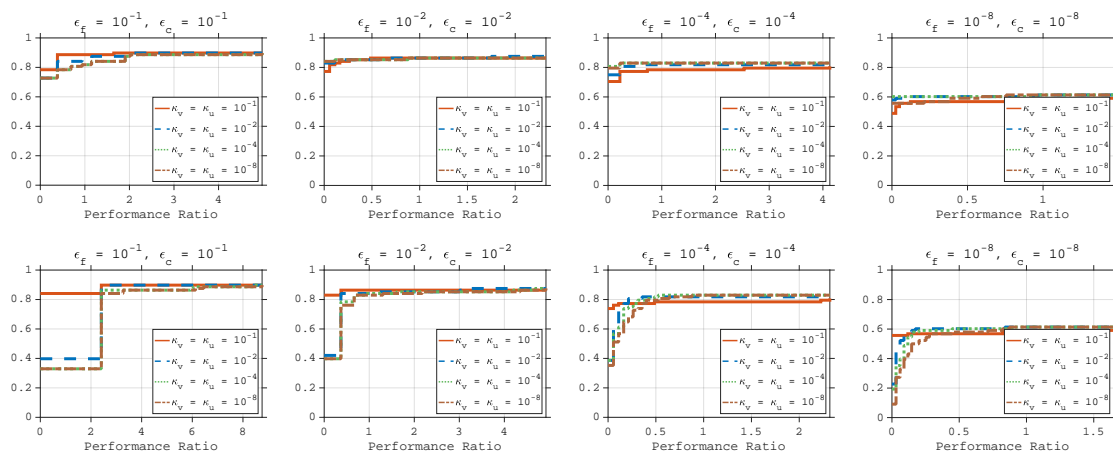


Figure 7: Dolan-Moré performance profiles comparing different inexactness levels for AdaSQP-opt on CUTEst collection of test problems in terms of **function evaluations** (top row) and **CG and MINRES iterations** (bottom row) for  $\epsilon_f = \epsilon_c \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$  (from left to right).

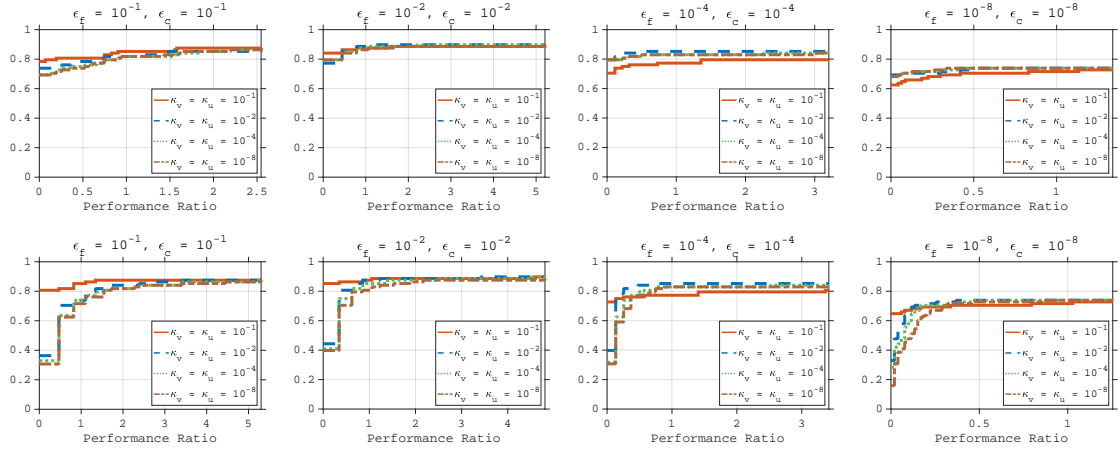


Figure 8: Dolan-Moré performance profiles comparing different inexactness levels for LSSQP-opt on CUTEst collection of test problems in terms of **function evaluations** (top row) and **CG and MINRES iterations** (bottom row) for  $\epsilon_f = \epsilon_c \in \{10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}\}$  (from left to right).

## 5 Final Remarks and Future Work

This paper introduces a sequential quadratic programming (SQP) algorithm to solve noisy nonlinear optimization problems with noisy nonlinear equality constraints, addressing the challenges posed by bounded noise in the objective and constraint functions as well as potential rank deficiency in the constraint Jacobians. Our approach incorporates a step decomposition strategy, inexact subproblem solutions, two flexible step size selection schemes, and an optimistic early termination scheme to enhance computational efficiency and robustness. We have established convergence to a neighborhood of a first-order stationary point in the presence of full-rank constraint Jacobians and convergence to a neighborhood of an infeasible stationary point in the rank-deficient case, with both neighborhood sizes dictated by the noise levels in the problem. Numerical experiments conducted on the CUTEst test set demonstrate the efficiency and robustness of the proposed method.

### Declaration

### Conflict of Interest

The authors declare no conflict of interest.

## References

- [1] Satyajith Amaran, Nikolaos V Sahinidis, Bikram Sharda, and Scott J Bury. Simulation optimization: a review of algorithms and applications. *Annals of Operations Research*, 240:351–380, 2016.
- [2] Stefania Bellavia, Gianmarco Gurioli, Benedetta Morini, and Philippe Louis Toint. The impact of noise on evaluation complexity: the deterministic trust-region case. *Journal of Optimization Theory and Applications*, 196(2):700–729, 2023.
- [3] Florent Benaych-Georges and Raj Rao Nadakuditi. The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 111:120–135, 2012.
- [4] Albert S Berahas, Raghu Bollapragada, and Jiahao Shi. Modified line search sequential quadratic methods for equality-constrained optimization with unified global and local convergence guarantees. *arXiv preprint arXiv:2406.11144*, 2024.
- [5] Albert S Berahas, Richard H Byrd, and Jorge Nocedal. Derivative-free optimization of noisy functions via quasi-newton methods. *SIAM Journal on Optimization*, 29(2):965–993, 2019.
- [6] Albert S Berahas, Liyuan Cao, and Katya Scheinberg. Global convergence rate analysis of a generic line search algorithm with noise. *SIAM Journal on Optimization*, 31(2):1489–1518, 2021.
- [7] Albert S Berahas, Frank E Curtis, Michael J O’Neill, and Daniel P Robinson. A stochastic sequential quadratic optimization algorithm for nonlinear-equality-constrained optimization with rank-deficient jacobians. *Mathematics of Operations Research*, 2023.
- [8] Albert S Berahas, Frank E Curtis, Daniel Robinson, and Baoyu Zhou. Sequential Quadratic Optimization for Nonlinear Equality Constrained Stochastic Optimization. *SIAM Journal on Optimization*, 31(2):1352–1379, 2021.
- [9] Albert S Berahas, Jiahao Shi, Zihong Yi, and Baoyu Zhou. Accelerating stochastic sequential quadratic programming for equality constrained optimization using predictive variance reduction. *Computational Optimization and Applications*, 86(1):79–116, 2023.
- [10] Albert S Berahas, Miaolan Xie, and Baoyu Zhou. A sequential quadratic programming method with high-probability complexity bounds for nonlinear equality-constrained stochastic optimization. *SIAM Journal on Optimization*, 35(1):240–269, 2025.
- [11] Richard H Byrd, Robert B Schnabel, and Gerald A Shultz. A trust region algorithm for nonlinearly constrained optimization. *SIAM Journal on Numerical Analysis*, 24(5):1152–1170, 1987.
- [12] Liyuan Cao, Albert S Berahas, and Katya Scheinberg. First-and second-order high probability complexity bounds for trust-region methods with noisy oracles. *Mathematical Programming*, pages 1–52, 2023.
- [13] Mireille Capitaine and Catherine Donati-Martin. Strong asymptotic freeness for wigner and wishart matrices. *Indiana University mathematics journal*, pages 767–803, 2007.

- [14] Nilanjan Chatterjee, Yi-Hau Chen, Paige Maas, and Raymond J Carroll. Constrained maximum likelihood estimation for model calibration using summary-level information from external big data sources. *Journal of the American Statistical Association*, 111(513):107–117, 2016.
- [15] Sou-Cheng T Choi, Christopher C Paige, and Michael A Saunders. MINRES-QLP: A Krylov subspace method for indefinite or singular symmetric systems. *SIAM Journal on Scientific Computing*, 33(4):1810–1836, 2011.
- [16] Philippe G Ciarlet. *Linear and nonlinear functional analysis with applications*. SIAM, 2013.
- [17] Frank E Curtis, Shima Dezfulian, and Andreas Waechter. An interior-point algorithm for continuous nonlinearly constrained optimization with noisy function and derivative evaluations. *arXiv preprint arXiv:2502.11302*, 2025.
- [18] Frank E Curtis, Vyacheslav Kungurtsev, Daniel P Robinson, and Qi Wang. A stochastic-gradient-based interior-point algorithm for solving smooth bound-constrained optimization problems. *arXiv preprint arXiv:2304.14907*, 2023.
- [19] Frank E Curtis, Jorge Nocedal, and Andreas Wächter. A matrix-free algorithm for equality constrained optimization problems with rank-deficient jacobians. *SIAM Journal on Optimization*, 20(3):1224–1249, 2010.
- [20] Frank E Curtis, Michael J O’Neill, and Daniel P Robinson. Worst-case complexity of an sqp method for nonlinear equality constrained stochastic optimization. *Mathematical Programming*, 205(1):431–483, 2024.
- [21] Frank E Curtis, Daniel P Robinson, and Baoyu Zhou. A stochastic inexact sequential quadratic optimization algorithm for nonlinear equality-constrained optimization. *INFORMS Journal on Optimization*, 2024.
- [22] Shima Dezfulian and Andreas Wächter. On the convergence of interior-point methods for bound-constrained nonlinear optimization problems with noise. *arXiv preprint arXiv:2405.11400*, 2024.
- [23] Elizabeth D Dolan and Jorge J Moré. Benchmarking optimization software with performance profiles. *Mathematical programming*, 91:201–213, 2002.
- [24] Yuchen Fang, Sen Na, Michael W Mahoney, and Mladen Kolar. Fully stochastic trust-region sequential quadratic programming for equality-constrained optimization problems. *SIAM Journal on Optimization*, 34(2):2007–2037, 2024.
- [25] Nicholas IM Gould, Stefano Lucidi, Massimo Roma, and Philippe L Toint. Solving the trust-region subproblem using the lanczos method. *SIAM Journal on Optimization*, 9(2):504–525, 1999.
- [26] Serge Gratton and Philippe L Toint. S2mpj and cutest optimization problems for matlab, python and julia. *arXiv preprint arXiv:2407.07812*, 2024.
- [27] Shih-Ping Han. A globally convergent method for nonlinear programming. *Journal of optimization theory and applications*, 22(3):297–309, 1977.
- [28] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

- [29] Billy Jin, Katya Scheinberg, and Miaolan Xie. High probability complexity bounds for adaptive step search based on stochastic oracles. *SIAM Journal on Optimization*, 34(3):2411–2439, 2024.
- [30] Jeffrey Larson, Matt Menickelly, and Jiahao Shi. A novel noise-aware classical optimizer for variational quantum algorithms. *arXiv preprint arXiv:2401.10121*, 2024.
- [31] Jeffrey Larson, Matt Menickelly, and Stefan M Wild. Derivative-free optimization methods. *Acta Numerica*, 28:287–404, 2019.
- [32] Yuchen Lou, Shigeng Sun, and Jorge Nocedal. Noise-tolerant optimization methods for the solution of a robust design problem. *arXiv preprint arXiv:2401.15007*, 2024.
- [33] Pablo Márquez-Neila, Mathieu Salzmann, and Pascal Fua. Imposing hard constraints on deep networks: Promises and limitations. *arXiv preprint arXiv:1706.02025*, 2017.
- [34] Joaquim RRA Martins and Andrew Ning. *Engineering design optimization*. Cambridge University Press, 2021.
- [35] Sen Na, Mihai Anitescu, and Mladen Kolar. An adaptive stochastic sequential quadratic programming with differentiable exact augmented lagrangians. *Mathematical Programming*, 199(1):721–791, 2023.
- [36] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Series in Operations Research and Financial Engineering. Springer-Verlag New York, 2006.
- [37] Emmanuel Omotayo Omojokun. *Trust region algorithms for optimization with nonlinear equality and inequality constraints*. University of Colorado at Boulder, 1989.
- [38] Figen Oztoprak, Richard Byrd, and Jorge Nocedal. Constrained optimization in the presence of noise. *SIAM Journal on Optimization*, 33(3):2118–2136, 2023.
- [39] Michael JD Powell. A fast algorithm for nonlinearly constrained optimization calculations. In *Numerical Analysis: Proceedings of the Biennial Conference Held at Dundee, June 28–July 1, 1977*, pages 144–157. Springer, 2006.
- [40] Sashank J Reddi, Suvrit Sra, Barnabás Póczos, and Alex Smola. Stochastic frank-wolfe methods for nonconvex optimization. In *2016 54th annual Allerton conference on communication, control, and computing (Allerton)*, pages 1244–1251. IEEE, 2016.
- [41] Bernhard A Schmitt. Perturbation bounds for matrix square roots and pythagorean sums. *Linear algebra and its applications*, 174:215–227, 1992.
- [42] Trond Steihaug. The conjugate gradient method and trust regions in large scale optimization. *SIAM Journal on Numerical Analysis*, 20(3):626–637, 1983.
- [43] Gilbert W Stewart. *Matrix perturbation theory*. 1990.
- [44] Shigeng Sun and Jorge Nocedal. A trust region method for noisy unconstrained optimization. *Mathematical Programming*, 202(1):445–472, 2023.
- [45] Shigeng Sun and Jorge Nocedal. A trust-region algorithm for noisy equality constrained optimization. *arXiv preprint arXiv:2411.02665*, 2024.
- [46] Andreas Wächter and Lorenz T Biegler. Line search filter methods for nonlinear programming: Motivation and global convergence. *SIAM Journal on Optimization*, 16(1):1–31, 2005.

- [47] Hermann Weyl. Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). *Mathematische Annalen*, 71(4):441–479, 1912.
- [48] Yuchen Xie, Richard H Byrd, and Jorge Nocedal. Analysis of the bfgs method with errors. *SIAM Journal on Optimization*, 30(1):182–209, 2020.