INDUSTRIAL AND SYSTEMS ENGINEERING

ISE

An Interior-Point Algorithm for Continuous Nonlinearly Constrained Optimization with Noisy Function and Derivative Evaluations

FRANK E. CURTIS

Department of Industrial and Systems Engineering, Lehigh University

SHIMA DEZFULIAN AND ANDREAS WAECHTER

Department of Industrial Engineering and Management Sciences, Northwestern University

LEHIGH
UNIVERSITY®

An Interior-Point Algorithm for Continuous Nonlinearly Constrained Optimization with Noisy Function and Derivative Evaluations

Frank E. Curtis[*1], Shima Dezfulian[†2], and Andreas Waechter[‡2]

[1]Department of Industrial and Systems Engineering, Lehigh University
[2]Department of Industrial Engineering and Management Sciences, Northwestern University

February 16, 2025

Abstract

An algorithm based on the interior-point methodology for solving continuous nonlinearly constrained optimization problems is proposed, analyzed, and tested. The distinguishing feature of the algorithm is that it presumes that only noisy values of the objective and constraint functions and their first-order derivatives are available. The algorithm is based on a combination of a previously proposed interior-point algorithm that allows inexact subproblem solutions and recently proposed algorithms for solving bound- and equality-constrained optimization problems with only noisy function and derivative values. It is shown that the new interior-point algorithm drives a stationarity measure below a threshold that depends on bounds on the noise in the function and derivative values. The results of numerical experiments show that the algorithm is effective across a wide range of problems.

1 Introduction

Interior-point methods have been studied extensively and have proved to be successful in practice for solving continuous nonlinearly constrained optimization problems. Two prominent categories of interior-point methods are those based on line-search [15, 30, 31, 33] and trust-region [6, 7, 34] mechanisms. Many state-of-the-art nonlinear optimization solvers, such as IPOPT [31] and KNITRO [8], are based on interior-point methods.

An important feature of these contemporary interior-point methods is that they rely on exact values of the objective and constraint functions and their first-order derivatives. On the other hand, there has been little work done on developing and analyzing such algorithms for cases when function evaluations are affected by noise. In noisy settings, only estimates of the function and derivative values are available. Generally speaking, noise can be stochastic or deterministic. By stochastic noise, we refer to situations in which it is reasonable to model a function value through

a probability distribution, in which case any realization of a value is a realization of a random variable. By deterministic noise, on the other hand, we are referring to situations in which any request for a function or derivative value with respect to an input value always results in the same *noisy* value that approximates some original *true* value. The original function of interest may be presumed to be, e.g., continuously differentiable, but the corresponding noisy function may be nonsmooth or even discontinuous. Computational noise—say, from a numerical simulation—is an example of deterministic noise [26], where it might only be assumed that the noise is bounded over the domain of a given function.

Recently, a number of algorithms have been proposed and analyzed for solving unconstrained, bound-constrained, or equality-constrained optimization problems in the presence of stochastic or deterministic noise. For further information, we refer the reader to [1, 2, 3, 12, 13, 14, 21, 23, 24, 25, 28, 32]. There has also been some recent work on the design of stochastic-gradient-based interior-point methods [9, 10]. However, to our knowledge, there have not yet been extensions of recent algorithms for the bound- or equality-constrained setting with deterministic noise to the *nonlinear-inequality-constrained setting with deterministic noise*. In this paper, we propose, analyze, and test such an algorithm. Due to its practical success in noiseless settings, our algorithm is based on the interior-point methodology. As a result, we expect our approach to yield good performance in practice, and expect that our proposed techniques can be incorporated readily into state-of-the-art software packages for solving continuous nonlinearly constrained optimization problems.

1.1 Contributions

This paper builds primarily upon the interior-point algorithm proposed in [16] for solving bound-constrained optimization problems in the presence of deterministic noise by extending the methodology and theoretical results to settings involving nonlinear inequality constraints. The method proposed in [16] is a line-search interior-point algorithm. Proposing such a method for solving nonlinearly constrained problems comes with significant additional challenges. For example, as is well known even in settings when exact function and derivative values are available, line-search interior-point methods that compute the search directions through Newton-based techniques may fail to converge due to inconsistency between the step computation and enforcement of nonnegativity constraints on variables [29]. In addition, if the Jacobian of the constraint function at a given point can be rank deficient, then a Newton-based approach may result in an ill-posed subproblem. To address these issues, we follow the step-decomposition approach proposed for the interior-point method in [15].

Another challenge that we face as we transition from the bound-constrained case to the nonlinear-inequality-constrained setting (with deterministic noise) is assessing the progress of the algorithm in terms of minimizing the objective function while aiming to satisfy the constraints. For solving bound-constrained problems with the interior-point method in [16], the barrier subproblem is unconstrained and satisfaction of the bounds is enforced through a fraction-to-the-boundary rule. Thus, the barrier objective function measures the progress of the algorithm and the fraction-to-the-boundary rule ensures satisfaction of the constraints. However, in the nonlinearly constrained case, it is typically inefficient to enforce feasibility of the constraints at every algorithm iterate, meaning that it is preferred to employ a so-called infeasible algorithm. In the context of such an algorithm, one needs a measure of progress that balances the improvement in the objective function and constraint satisfaction. In noiseless settings, this is typically accomplished by defining a merit function that is a weighted sum of the objective function and a measure of constraint violation.

The weight assigned to the objective function relative to the constraint satisfaction [22] is called the merit parameter. Typically, the merit parameter is updated in an adaptive manner in such a way that it guarantees that the original optimization problem is being solved. Extending such an approach and its analysis to the noisy setting is the main contribution of this work.

We mention upfront that our proposed algorithm and analysis leave certain natural questions unanswered. In particular, we focus our attention in this paper on an algorithm for solving a single barrier subproblem that arises in an interior-point framework. This subproblem is equality-constrained, but unlike for an approach for the equality-constrained setting with deterministic noise such as that proposed in [23], significant obstacles arise in our setting due to the fact that our method needs to enforce nonnegativity of slack variables that are introduced for our problem formulation. Along with our concluding remarks, we suggest a heuristic strategy for decreasing the barrier parameter so that, in practice, a sequence of barrier subproblems may be solved, as is typical for an interior-point method. However, our analysis does not cover convergence guarantees for solving the original constrained optimization problem as the barrier parameter vanishes. On this note, one should recognize that such an analysis would be questionable in any case. After all, even for a method for the equality-constrained setting such as that in [23], the presence of deterministic noise means that one cannot guarantee convergence to a solution of the original problem of interest. Such is the case for our method for solving each barrier subproblem as well, meaning that it may be impossible to guarantee that a point is reached that is sufficiently close to optimality (or even stationarity) of the noiseless barrier subproblem to warrant a decrease in the barrier parameter. That said, in practice, a decrease of the barrier parameter may yield good behavior, so we suggest a heuristic in our concluding remarks.

1.2 Notation

Our optimization problem of interest is stated in terms of a decision variable $x \in \mathbb{R}^n$; see (1) in Section 2. For any $k \in \mathbb{N}$, the value of x in the kth iteration of our algorithm is denoted as x_k. Such subscripts are used for other variables and adaptive parameters that are employed. For any $i \in [n] := \{1, \ldots, n\}$, the ith component of a vector $x \in \mathbb{R}^n$ is denoted $x^{(i)}$. Generally, vectors are expressed using lowercase letters. With respect to any such a vector, the corresponding uppercase letter denotes a diagonal matrix with the components of the vector on the diagonal, e.g., with respect to $s \in \mathbb{R}^q$ we denote $S := \text{diag}(s) \in \mathbb{R}^{q \times q}$.

Given a matrix $M \in \mathbb{R}^{m \times n}$, its range space is denoted as $\text{range}(M)$, its null space is denoted as $\text{null}(M)$, its minimum singular value is denoted as $\sigma_{\min}(M)$, its maximum singular value is denoted as $\sigma_{\max}(M)$, and its condition number is denoted as $\kappa(M) := \sigma_{\max}(M)/\sigma_{\min}(M)$. Given a square matrix $M \in \mathbb{R}^{n \times n}$, a square root of M is any matrix $M^{1/2}$ such that $M^{1/2} M^{1/2} = M$. Given a pair of vectors $(u, v) \in \mathbb{R}^n \times \mathbb{R}^n$, their inner product is denoted as $\langle u, v \rangle = u^T v$. We use the operator notation $\| \cdot \| := \| \cdot \|_2 = \sqrt{\langle \cdot, \cdot \rangle}$ to denote the 2-norm of a vector and similarly use $\| \cdot \|$ to denote the induced 2-norm for a matrix. For a symmetric and positive definite matrix $M \in \mathbb{R}^{n \times n}$ and a pair of vectors $(u, v) \in \mathbb{R}^n \times \mathbb{R}^n$, we write $\langle u, v \rangle_M = u^T M v$ and denote $\| \cdot \|_M = \sqrt{\langle \cdot, \cdot \rangle_M}$.

Given two sequences $\{u_k\}$ and $\{v_k\}$ with $u_k \in \mathbb{R}$ and $v_k \in [0, \infty)$ for all $k \in \mathbb{N}$, the "big-O" notation, namely, $u_k = O(v_k)$, indicates that there exists a real number $c \in (0, \infty)$—defined independently from k—such that for all sufficiently large $k \in \mathbb{N}$ one has $|u_k| \leq c v_k$.

1.3 Outline

Our problem of interest and related optimization problem formulations for which our proposed algorithm is designed are stated in Section 2. Our proposed algorithm for solving a so-called barrier subproblem that arises in our discussions is presented in Section 3. Our analysis of the convergence properties of our proposed algorithm from Section 3 is provided in Section 4. In Section 5, we present the results of numerical experiments that show the effectiveness of the proposed algorithm. A conclusion is provided in Section 6.

2 Problem Description

Our main algorithm, stated as Algorithm 1 on page 11, generates an iterate sequence $\{x_k\}$ with $x_k \in \mathbb{R}^n$ for all generated $k \in \mathbb{N}$. Its aim is to solve (if only approximately) the continuous optimization problem

$$\min_{x \in \mathbb{R}^n} \ \bar{f}_0(x) \ \text{ s.t. } \ \bar{c}_I(x) \leq 0, \tag{1}$$

where $\bar{f}_0 : \mathbb{R}^n \to \mathbb{R}$ and $\bar{c}_I : \mathbb{R}^n \to \mathbb{R}^q$. (Throughout the paper, a "bar" above a quantity is used to indicate a noiseless quantity, as opposed to noisy quantities that have no "bar" above them.) The algorithm can be extended to settings in which equality constraints are also present, but for simplicity and since the main algorithmic components and corresponding analysis are nearly identical for that situation and this one, we restrict attention to the setting with only inequality constraints. The problem functions $\bar{f}_0$ and $\bar{c}_I$ are assumed to satisfy the following loose assumption pertaining to the generated iterates. The assumption includes that the functions are continuously differentiable so that the objective gradient function $\bar{g}_0 := \nabla \bar{f}_0 : \mathbb{R}^n \to \mathbb{R}^n$ and constraint Jacobian function $\bar{J}_I := \nabla \bar{c}_I^T : \mathbb{R}^n \to \mathbb{R}^{q \times n}$ are well defined. As previously mentioned, the main challenge of our setting is that the algorithm only has access to noisy values of the objective function $\bar{f}_0$, its gradient function $\bar{g}_0$, the constraint function $\bar{c}_I$, and its Jacobian function $\bar{J}_I$. Our assumption about the noisy evaluations of these functions is included in the following assumption.

Assumption 2.1. *The iterate sequence $\{x_k\}$ generated by Algorithm 1 is contained in an open convex set $\mathcal{X} \subseteq \mathbb{R}^n$ over which $\bar{f}_0$ and $\bar{c}_I$ are continuously differentiable, $\bar{f}_0$ is bounded below, $\bar{c}_I$ is bounded in norm, and both functions $\bar{g}_0 := \nabla \bar{f}_0$ and $\bar{J}_I := \nabla \bar{c}_I^T$ are Lipschitz continuous with values that are bounded uniformly in norm. Moreover, for any $x \in \mathcal{X}$, a call for $\bar{f}_0(x)$, $\bar{g}_0(x)$, $\bar{c}_I(x)$, or $\bar{J}_I(x)$ results in the approximate value $f_0(x)$, $g_0(x)$, $c_I(x)$, or $J_I(x)$, respectively, where for some known constants $(\epsilon_f, \epsilon_g, \epsilon_c, \epsilon_J) \in (0, \infty)^4$ one has*

$$|\bar{f}_0(x) - f_0(x)| \leq \epsilon_f, \quad \|\bar{g}_0(x) - g_0(x)\| \leq \epsilon_g,$$
$$\|\bar{c}_I(x) - c_I(x)\| \leq \epsilon_c, \quad \text{and} \quad \|\bar{J}_I(x) - J_I(x)\| \leq \epsilon_J. \tag{2}$$

Under Assumption 2.1, there exist constants $f_{\inf} \in \mathbb{R}$ and $(c_{I,\sup}, g_{0,\sup}, J_{I,\sup}) \in (0, \infty)^3$ such that for all $x \in \mathcal{X}$ one has

$$\bar{f}_0(x) \geq f_{\inf}, \ \|\bar{c}_I(x)\| \leq c_{I,\sup}, \ \|\bar{g}_0(x)\| \leq g_{0,\sup}, \ \text{and} \ \|\bar{J}_I(x)\| \leq J_{I,\sup},$$

and there exist $(L_g, L_J) \in (0, \infty)^2$ such that for all $(x, \bar{x}) \in \mathcal{X} \times \mathcal{X}$ one has

$$\|\bar{g}_0(x) - \bar{g}_0(\bar{x})\| \leq L_g \|x - \bar{x}\| \text{ and } \|\bar{J}_I(x) - \bar{J}_I(\bar{x})\| \leq L_J \|x - \bar{x}\|.$$

5

Under Assumption 2.1—specifically the assumption that the objective and constraint functions are continuously differentiable—and a constraint qualification, e.g., the Mangasarian-Fromovitz constraint qualification (MFCQ) [19], it follows that at a local minimizer $x \in \mathbb{R}^n$ of (1) there exists $\bar{y}_I \in \mathbb{R}^q$ with

$$\bar{g}_0(x) + \bar{J}_I(x)^T \bar{y}_I = 0, \quad \bar{c}_I(x) \le 0, \quad \bar{y}_I \ge 0, \quad \text{and} \quad \bar{c}_I(x)^T \bar{y}_I = 0.$$

However, more generally, it is possible that the constraints of problem (1) are infeasible (at least locally near a point), or that at a local minimizer of (1) the constraints are degenerate in the sense that a constraint qualification does not hold. With respect to these situations, the aim of the algorithm is at least to solve the infeasibility-minimization problem

$$\min_{x \in \mathbb{R}^n} \tfrac{1}{2} \| \max\{\bar{c}_I(x), 0\} \|^2, \tag{3}$$

where the max is defined component-wise. The objective of this problem is continuously differentiable and first-order conditions for optimality for it are

$$\bar{J}_I(x)^T \max\{\bar{c}_I(x), 0\} = 0. \tag{4}$$

We refer to any x yielding (4) with $\bar{c}_I(x) \not\le 0$ as an infeasible stationary point.

Following the design of the algorithm in [15], our algorithm aims to solve (perhaps only approximately) the optimization problem (1) or at least the infeasibility-minimization problem (3) by working with related formulations involving slack variables. In particular, observe that (1) is equivalent to

$$\min_{x \in \mathbb{R}^n} \bar{f}_0(x) \quad \text{s.t.} \quad \bar{c}_I(x) + s = 0 \quad \text{and} \quad s \ge 0, \tag{5}$$

and that if $s \ge 0$ and $\bar{c}_I(x) + s \ge 0$, then (4) is equivalent to

$$\bar{J}_I(x)^T (\bar{c}_I(x) + s) = 0 \quad \text{and} \quad S(\bar{c}_I(x) + s) = 0. \tag{6}$$

The conditions $s \ge 0$ and $\bar{c}_I(x) + s \ge 0$ motivate a *slack reset* that is incorporated into the algorithm to ensure that these inequalities always hold at least with respect to the noisy constraint function evaluation; see Section 3.

Our algorithmic strategy falls within the interior-point methodology, which means that it does not aim to solve (5) or (6) directly; rather, it employs a log-barrier function of the slack variables to replace the inequality constraints. The general idea of an interior-point method would be to solve to some accuracy a subproblem involving the barrier function on the slack variables for a given value of a barrier parameter $\mu \in (0, \infty)$, then decrease this parameter and solve the next subproblem in an iterative manner such that the overall algorithm converges to a solution of (5) or at least (6). The goal of our main algorithm and our corresponding analysis of it is to solve such a barrier subproblem approximately using only noisy function and derivative evaluations. The barrier subproblem that we employ (with $\log(\cdot)$ denoting the natural logarithm) is

$$\min_{(x,s) \in \mathbb{R}^n \times \mathbb{R}^q} \bar{f}_0(x) - \mu \sum_{i=1}^{q} \log(s^{(i)}) \quad \text{s.t.} \quad \bar{c}_I(x) + s = 0 \quad \text{and} \quad s > 0. \tag{7}$$

Observe that, at any point (x, s) with $s > 0$, the linear independence constraint qualification (LICQ) with respect to (7) holds. Hence, at any local minimizer (x, s) of (7), there exists a Lagrange multiplier $\bar{y}_{I,\mu} \in \mathbb{R}^q$ such that

$$\bar{g}_0(x) + \bar{J}_I(x)^T \bar{y}_{I,\mu} = 0, \quad -\mu S^{-1} + \bar{y}_{I,\mu} = 0, \quad \bar{c}_I(x) + s = 0, \quad \text{and} \quad s > 0. \tag{8}$$

6

We remark that in a setting in which equality constraints are also present, the LICQ would not be guaranteed to hold, but a similar set of conditions would be guaranteed under a suitable constraint qualification, such as the MFCQ.

For notational convenience in our presentation and analysis of our proposed algorithm, let us introduce the combined primal iterate vector $z \in \mathbb{R}^{n+q}$ as $z := [x^T \ s^T]^T$. For a given barrier parameter μ, the objective and constraint functions of (7) can be expressed as $\bar{f} : \mathbb{R}^{n+q} \to \mathbb{R}$ and $\bar{c} : \mathbb{R}^{n+q} \to \mathbb{R}^q$ where

$$\bar{f}(z) = \bar{f}_0(x) - \mu \sum_{i=1}^{q} \log(s^{(i)}) \text{ and } \bar{c}(z) = \bar{c}_I(x) + s \text{ for all } z \in \mathbb{R}^{n+q}.$$

Following [15], our algorithm employs *scaled* first-order derivatives of these functions. In particular, we define the scaled gradient function $\bar{g} : \mathbb{R}^{n+q} \to \mathbb{R}$ and scaled Jacobian function $\bar{J} : \mathbb{R}^{n+q} \to \mathbb{R}^{q \times (n+q)}$ such that, for all $z \in \mathbb{R}^{n+q}$,

$$\bar{g}(z) = \begin{bmatrix} \bar{g}_0(x) \\ -\mu e \end{bmatrix}, \text{ and } \bar{J}(z) = \begin{bmatrix} \bar{J}_I(x) & S \end{bmatrix},$$

where $e \in \mathbb{R}^q$ is a vector with all entries equal to one. (The imposed scaling is that the second block of the gradient of the objective and the second block of the constraint Jacobian are both multiplied by the diagonal matrix S.) Observe that in terms of this scaled Jacobian, the conditions in (6) are equivalent to

$$\bar{J}(z)^T \bar{c}(z) = 0. \tag{9}$$

Letting f, g, c, and J denote noisy approximations of $\bar{f}$, $\bar{g}$, $\bar{c}$, and $\bar{J}$, respectively, Assumption 2.1 implies that for all $z = [x^T \ s^T]^T$ with $x \in \mathcal{X}$ one has

$$|\bar{f}(z) - f(z)| \leq \epsilon_f, \quad \|\bar{g}(z) - g(z)\| \leq \epsilon_g,$$
$$\|\bar{c}(z) - c(z)\| \leq \epsilon_c, \quad \text{and} \quad \|\bar{J}(z) - J(z)\| \leq \epsilon_J. \tag{10}$$

3 Algorithm Description, Fixed Barrier Parameter

Let us now present our main algorithm, which aims to produce at least an approximate solution of (1) by solving the barrier subproblem (7) for a fixed barrier parameter $\mu \in (0, \infty)$, or at least to yield convergence to satisfying the stationarity conditions (9). To describe the steps of our main algorithm, suppose that it has reached iteration $k \in \mathbb{N}$, in which case the kth iteration proceeds as described in this section with the noisy function and derivative values $f_k := f(z_k) := f_0(x_k) - \mu \sum_{i=1}^{q} \log(s_k^{(i)})$, $c_k := c(z_k) := c_I(x_k) + s_k$, $g_k^T := g(z_k)^T := [g_0(x_k)^T \ -\mu e^T]^T$, and $J_k := J(z_k) := [J_I(x_k) \ S_k]$. The corresponding noiseless values $\bar{f}_k$, $\bar{c}_k$, $\bar{g}_k$, and $\bar{J}_k$ are defined similarly. Our proposed algorithm is stated formally as Algorithm 1 on page 11. We remark that, as previously mentioned, the initial conditions and structure of the algorithm will ensure that $s_k \geq 0$ and $c(z_k) = c_I(x_k) + s_k \geq 0$ for all $k \in \mathbb{N}$. Let us also define, for all $k \in \mathbb{N}$, the error values $\varepsilon_{f,k} \in \mathbb{R}$, $\varepsilon_{g,k} \in \mathbb{R}^n$, $\varepsilon_{c,k} \in \mathbb{R}^q$, and $\varepsilon_{J,k} \in \mathbb{R}^{q \times (n+q)}$ according to the equations

$$f_k = \bar{f}_k + \varepsilon_{f,k}, \ g_k = \bar{g}_k + \varepsilon_{g,k}, \ c_k = \bar{c}_k + \varepsilon_{c,k}, \ \text{and} \ J_k = \bar{J}_k + \varepsilon_{J,k}, \tag{11}$$

where by (10) one has $|\varepsilon_{f,k}| \leq \epsilon_f$, $\|\varepsilon_{g,k}\| \leq \epsilon_g$, $\|\varepsilon_{c,k}\| \leq \epsilon_c$, and $\|\varepsilon_{J,k}\| \leq \epsilon_J$.

The following occurs for each iteration index $k \in \mathbb{N}$. First, if $J_k^T c_k = 0$ and $c_I(x_k) \not\leq 0$, then the algorithm has reached an infeasible stationary point (based on noisy evaluations) and it terminates; recall that the algorithm ensures $s_k \geq 0$ and $c_k = c_I(x_k) + s_k \geq 0$, so $J_k^T c_k = 0$ (i.e., (9)) corresponds to (6). Otherwise, the algorithm computes a normal step toward the satisfaction of a linear approximation of the constraint functions within a trust region. Specifically, for a prescribed constant $\omega \in (0, \infty)$, the normal step v_k is computed by solving the trust-region subproblem

$$\min_{v \in \text{range}(J_k^T)} \frac{1}{2}\|c_k + J_k v\|^2 \quad \text{s.t.} \quad \|v\| \leq \omega \|J_k^T c_k\|. \tag{12}$$

For later reference, let us define $\bar{v}_k$ as the solution of (12) (with respect to x_k) if the true values $(\bar{c}_k, \bar{J}_k)$ are used in place of the noisy values (c_k, J_k). Second, the algorithm computes a tangential step to minimize a local model of the objective function subject to maintaining the progress toward linearized feasibility that was attained by the normal step. This computation employs symmetric

$$W_k \leftarrow \begin{bmatrix} H_k & 0 \\ 0 & \Sigma_k \end{bmatrix}, \tag{13}$$

where $H_k \in \mathbb{R}^{n \times n}$ and $\Sigma_k \in \mathbb{R}^{q \times q}$. (See below for further discussion on the requirements of W_k for each $k \in \mathbb{N}$.) The tangential step is defined as $u_k := d_k - v_k$, where the full step d_k (i.e., normal step plus tangential step) along with a new Lagrange multiplier estimate y_{k+1} is computed by solving

$$\begin{bmatrix} W_k & J_k^T \\ J_k & 0 \end{bmatrix} \begin{bmatrix} d_k \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} -g_k \\ J_k v_k \end{bmatrix}. \tag{14}$$

For later reference, let us define $(\bar{u}_k, \bar{d}_k, \bar{y}_{k+1})$ as the values that would have been computed (with respect to the same symmetric matrix W_k) if the true values $(\bar{J}_k, \bar{g}_k, \bar{v}_k)$ are used in place of the noisy values (J_k, g_k, v_k).

A few comments on the choice of $\{W_k\}$ are in order. First, we remark that the traditional choice in an interior-point method in the noiseless setting is for H_k to be chosen as the Hessian of a Lagrangian at the current iterate, or an approximation to it. However, for our noisy setting wherein we do not guarantee a fast rate of local convergence, we require H_k to be a symmetric matrix, but do not require it to be a matrix of second-order derivatives. (It can be constructed using noisy second-order derivative estimates, if they are available, but this is not necessary for our analysis.) As for Σ_k, if it is chosen as $\Sigma_k = \mu I$, then this is said to be the primal approach, whereas if it is chosen as $\Sigma_k = S_k \hat{Y}_k$ for some multiplier vector $\hat{y}_k \in \mathbb{R}^q$, then this is said to be the primal-dual approach. That said, we do not restrict our analysis to any such choice of Σ_k, as long as it is diagonal and positive definite. In addition to each element of the sequence $\{W_k\}$ defined by (13) being symmetric, we also require that (a) the linear system (14) has a unique solution for all $k \in \mathbb{N}$, (b) W_k is sufficiently positive definite in the null space of J_k for all $k \in \mathbb{N}$, and (c) $\{\|W_k\|\}$ is bounded. We formalize these assumptions in Assumption 4.1 on page 13, where in fact (for simplicity) we go a bit further and assume that W_k satisfies such properties also with respect to true function and derivative values. To ensure these conditions with respect to noisy quantities in practice, during the kth iteration for any $k \in \mathbb{N}$, the algorithm can start with initial symmetric H_k and Σ_k such that W_k in (13) satisfies a prescribed bound on its norm. Then, the algorithm can determine if (14) has a unique solution and W_k is positive definite over $\text{null}(J_k)$, say, by checking the inertia of the matrix in (14); see, e.g., [22]. If not, then H_k and/or Σ_k can be replaced with a convex combination of its current value and a prescribed positive definite matrix, or it can be

modified by adding a multiple of the identity matrix; see, e.g., [31]. This process can be repeated iteratively until the desired conditions are satisfied. Observe that modifying W_k as discussed above guarantees that the system (14) has a unique solution because J_k has full row rank.

Upon the computation of the normal step v_k, tangential step u_k, combined search direction d_k, and updated multiplier estimate y_{k+1}, the algorithm checks another termination condition. If $J_k^T c_k = 0$ (implying $c_I(x_k) \leq 0$ since the algorithm did not terminate previously) and $g_k + J_k^T y_{k+1} = 0$, then the algorithm has reached a point at which no further progress can be made (with respect to noisy values). This situation implies that the algorithm has reached an approximate stationary point for the barrier subproblem (7). We discuss the meaning of this situation further in our analysis in Section 4.

If the algorithm proceeds in the kth iteration after checking this second termination condition and has not terminated, then it needs to determine the amount to move along the search direction, i.e., it needs to determine a positive step size. This is done using a merit function—which balances the objective function and a constraint violation measure—whose aim is to measure progress of the algorithm toward a minimizer. Before determining the step size, the algorithm first updates the merit parameter that weighs the contributions of the merit function terms. For the merit function, our proposed algorithm employs $\phi : \mathbb{R}^{n+q} \times (0, \infty) \to \mathbb{R}$ defined by

$$\phi(z, \tau) = \tau f(z) + \|c(z)\|,$$

where $\tau \in (0, \infty)$ is the merit parameter that is updated dynamically. For later reference, let us also introduce the corresponding true merit function $\bar{\phi} : \mathbb{R}^{n+q} \times (0, \infty) \to \mathbb{R}$ with $\bar{\phi}(z, \tau) = \tau \bar{f}(z) + \|\bar{c}(z)\|$. To determine whether and by how much the merit parameter should be decreased, the algorithm uses a (noisy) model of ϕ at x_k, namely, $m_k : \mathbb{R}^{n+q} \times (0, \infty) \to \mathbb{R}$ with

$$m_k(d, \tau) = \tau(f_k + g_k^T d) + \|c_k + J_k d\|.$$

The change in this model corresponding to $(d_k, \tau_k) \in \mathbb{R}^{n+q} \times (0, \infty)$ is measured by the model reduction function $\Delta m_k : \mathbb{R}^{n+q} \times (0, \infty) \to \mathbb{R}$ defined by

$$\begin{aligned}
\Delta m_k(d_k, \tau_k) &= m_k(0, \tau_k) - m_k(d_k, \tau_k) \\
&= -\tau_k g_k^T d_k + \|c_k\| - \|c_k + J_k d_k\| \\
&= -\tau_k g_k^T d_k + \|c_k\| - \|c_k + J_k v_k\|,
\end{aligned} \tag{15}$$

where the last equation follows by (14). For reference in our analysis, let us also note that the reduction in a true (i.e., noiseless) model of the true merit function $\bar{\phi}$ at x_k can be defined as $\Delta \bar{m}_k : \mathbb{R}^{n+q} \times (0, \infty) \to \mathbb{R}$ defined similarly. This true model reduction is notable since having $\{(d_k, \tau_k)\}$ with $\tau_k \geq \tau$ for some $\tau \in (0, \infty)$ and $\{\Delta \bar{m}_k(d_k, \tau_k)\} \to 0$ is central to analyses of an algorithm of our type in the noiseless setting. We discuss this further in Section 4.

The goal of our update strategy of the merit parameter is to ensure that the chosen value $\tau_k \in (0, \tau_{k-1}]$ (where $\tau_{k-1} \in (0, \infty)$ is the value from the prior iteration) yields, for some prescribed $\sigma \in (0, 1)$, the condition

$$\Delta m_k(d_k, \tau_k) \geq \tfrac{1}{2} \tau_k u_k^T W_k u_k + \sigma(\|c_k\| - \|c_k + J_k v_k\|). \tag{16}$$

(This is the same kind of approach as has been employed in the noiseless setting; see, e.g., [15].) From (12) and positive-definiteness of W_k with respect to null(J_k), the right-hand side of this

9

expression is always nonnegative, and in fact in our analysis we show that this quantity is strictly positive when the algorithm does not terminate in iteration k. To determine such a value for the merit parameter τ_k, the algorithm uses the trial merit parameter value

$$\tau_k^{\text{trial}} \leftarrow \begin{cases} \infty & \text{if } g_k^T d_k + \frac{1}{2} u_k^T W_k u_k \leq 0 \\ \frac{(1-\sigma)(\|c_k\| - \|c_k + J_k v_k\|)}{g_k^T d_k + \frac{1}{2} u_k^T W_k u_k} & \text{otherwise,} \end{cases} \tag{17}$$

then, with prescribed $\delta_\tau \in (0, 1)$, sets the merit parameter as

$$\tau_k = \begin{cases} \tau_{k-1} & \text{if } \tau_{k-1} \leq \tau_k^{\text{trial}} \\ \min\{(1-\delta_\tau)\tau_{k-1}, \tau_k^{\text{trial}}\} & \text{otherwise.} \end{cases} \tag{18}$$

(In Lemma 4.11, we show that this ensures that (16) holds.) For future reference, let us also define the sequence $\{\bar{\tau}_k\}$ as the one that would have been generated in this manner if the algorithm generated the same iterate sequence $\{x_k\}$, but employed noiseless values for updating the merit parameter value for all $k \in \mathbb{N}$.

Now that the value of the merit parameter has been determined, the algorithm proceeds to compute a step size to take along the primal search direction. At this stage, it is necessary to characterize the primal search direction in terms of two components—corresponding to the variables x and s, respectively—and the search direction corresponding to the unscaled derivatives as well. The scaled and unscaled search directions, respectively, are

$$d_k =: \begin{bmatrix} d_k^x \\ d_k^s \end{bmatrix} \quad \text{and} \quad \hat{d}_k \leftarrow \begin{bmatrix} d_k^x \\ S_k d_k^s \end{bmatrix}, \quad \text{where} \quad (d_k^x, d_k^s) \in \mathbb{R}^n \times \mathbb{R}^q.$$

One rule for the step size is that it must ensure that the slack variables remain sufficiently positive. This is done through a fraction-to-the-boundary rule. The unscaled direction in the slack variables is the latter component of the direction $\hat{d}_k$, namely, $S_k d_k^s$. Therefore, the fraction-to-the-boundary rule involves computing the largest value of $\alpha_k^{\max}$ in $(0, 1]$ such that

$$s_k + \alpha_k^{\max} S_k d_k^s \geq (1 - \eta_s) s_k, \tag{19}$$

where $\eta_s \in (0, 1)$ is a prescribed parameter. Once this maximum step size has been determined, the algorithm backtracks, if necessary, until a relaxed Armijo condition is satisfied [1, 23]. Defining $\varepsilon_k \in (0, \infty)$ for all $k \in \mathbb{N}$ by $\varepsilon_k := \tau_k \epsilon_f + \epsilon_c$, and with $\eta_\phi \in (0, 1)$ and $\zeta \in (0, \infty)$, the condition is

$$\phi(z_k + \alpha_k \hat{d}_k, \tau_k) \leq \phi(z_k, \tau_k) - \eta_\phi \alpha_k \Delta m_k(d_k, \tau_k) + (2 + \zeta)\varepsilon_k. \tag{20}$$

(In practice, we expect one to set ζ relatively small, at least less than 1. For example, in our numerical experiments, $\zeta \leftarrow 0.1$.) Observe that (20) is well defined since it is always satisfied for sufficiently small $\alpha_k \in (0, 1]$ in the presence of positive ϵ_f and/or ϵ_c. Given $\alpha_k \in (0, 1]$ obtained by backtracking (starting from $\alpha_{\max}$) that yields (20), the algorithm sets $x_{k+1} \leftarrow x_k + \alpha_k d_k^x$ and $s_{k+1} \leftarrow \max\{s_k + \alpha_k S_k d_k^s, -c_I(x_{k+1})\} \geq 0$, for which one finds that

$$\begin{aligned} c(z_{k+1}) &= c_I(x_{k+1}) + s_{k+1} \\ &= c_I(x_{k+1}) + \max\{s_k + \alpha_k S_k d_k^s, -c_I(x_{k+1})\} \geq 0. \end{aligned} \tag{21}$$

It is important to observe that, after employing this update for the slack variables, which as in the literature we refer to as a slack reset, one finds that $\phi(z_{k+1}, \tau_k) \leq \phi(z_k + \alpha_k \hat{d}_k, \tau_k)$, meaning that (20) holds with the left-hand side replaced by $\phi(z_{k+1}, \tau_k)$. This is clarified in the following remark.

10

Remark 3.1. *Line 11 of Algorithm 1 guarantees $\phi(z_{k+1}, \tau_k) \leq \phi(z_k + \alpha_k \hat{d}_k, \tau_k)$. To see this, consider arbitrary $i \in \{1, \dots, q\}$. If $[c_I(x_{k+1})]^{(i)} \geq -[s_k + \alpha_k S_k d_k^s]^{(i)}$, then it yields $s_{k+1}^{(i)} \leftarrow [s_k + \alpha_k S_k d_k^s]^{(i)}$; otherwise, $s_{k+1}^{(i)} \leftarrow [-c_I(x_{k+1})]^{(i)}$, in which case $[c(z_{k+1})]^{(i)} = 0$. In any case, one finds $s_{k+1} \geq s_k + \alpha_k S_k d_k^s$, $f(z_{k+1}) \leq f\left(\begin{bmatrix} x_{k+1} \\ s_k + \alpha_k S_k d_k^s \end{bmatrix}\right)$, and $\|c(z_{k+1})\| \leq \left\| c\left(\begin{bmatrix} x_{k+1} \\ s_k + \alpha_k S_k d_k^s \end{bmatrix}\right)\right\|$, so $\phi(z_{k+1}, \tau_k) \leq \phi(z_k + \alpha_k \hat{d}_k, \tau_k)$, as claimed.*

Algorithm 1 : Algorithm for Solving (7) with Noisy Function Evaluations

---

Require: $(x_0, s_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^q \times \mathbb{R}^q$ with $s_0 \geq 0$ and $c(z_0) = c_I(x_0) + s_0 \geq 0$; $(\epsilon_f, \epsilon_c) \in (0, \infty)^2$ from Assumption 2.1; $\tau_{-1} \in (0, \infty)$; $\omega \in (0, \infty)$; $\sigma \in (0, 1)$; $\delta_\tau \in (0, 1)$; $\eta_s \in (0, 1)$; $\eta_\phi \in (0, 1)$; and $\zeta \in (0, \infty)$

1: **for** $k = 0, 1, \dots$ **do**
2:      **if** $J_k^T c_k = 0$ and $c_I(x_k) \not\leq 0$ **then return** (x_k, y_{k+1})
3:      compute v_k as the solution of (12)
4:      choose W_k satisfying (upcoming) Assumption 4.1
5:      compute (d_k, y_{k+1}), where $d_k \equiv v_k + u_k \equiv \begin{bmatrix} d_k^x \\ d_k^s \end{bmatrix}$, by solving (14)
6:      **if** $J_k^T c_k = 0$ (so $c_I(x_k) \leq 0$) and $g_k + J_k^T y_{k+1} = 0$ **then return** (x_k, y_{k+1})
7:      set τ_k by (18)
8:      set $\alpha_k^{\max}$ as the largest value in $(0, 1]$ such that (19) holds
9:      set $\alpha_k \leftarrow (\frac{1}{2})^j \alpha_k^{\max}$, where $j = \min\left\{ j \in \mathbb{N} : (20) \text{ holds with } \hat{d}_k \leftarrow \begin{bmatrix} d_k^x \\ S_k d_k^s \end{bmatrix}\right\}$
10:     set $x_{k+1} \leftarrow x_k + \alpha_k d_k^x$
11:     set $s_{k+1} \leftarrow \max\{s_k + \alpha_k S_k d_k^s, -c_I(x_{k+1})\}$
12: **end for**

---

4   Analysis of Algorithm 1

In this section, we analyze the behavior of Algorithm 1 for solving (7). Before starting our analysis, we discuss its main goals by remarking on the convergence guarantees that one aims to show in a noiseless setting for an algorithm such as ours. Then, we discuss our main assumptions and show that the algorithm is well posed before proceeding to our main convergence analysis.

4.1   Discussion of Aims of Convergence Analysis

Given the fact that Algorithm 1 aims to solve (7), but only has access to noisy function and derivative values, it is important to establish at the outset what one can hope to prove about the behavior of the algorithm. First, we note that if for some $k \in \mathbb{N}$ the algorithm terminates in line 2, then at least with respect to noisy quantities it has reached an infeasible stationary point. The point is not necessarily an infeasible stationary point with respect to noiseless quantities, but it is still reasonable for the algorithm to terminate since, by (11) and upcoming Lemma 4.2 in our analysis, one respectively has

$$\|\bar{c}_I(x_k) - c_I(x_k)\| \leq \epsilon_c \quad \text{and} \quad \|\bar{J}_k^T \bar{c}_k - J_k^T c_k\| = O(\epsilon_c + \epsilon_J).$$

Similarly, if for some $k \in \mathbb{N}$ the algorithm terminates in line 6, then at least with respect to the noisy quantities it has reached a point that appears to be stationary for (7). The point is not necessarily stationary for (7) with respect to noiseless quantities, but it is still reasonable for the algorithm to terminate since, under the assumptions of our upcoming analysis, one can show that

$$\|\bar{g}_k + \bar{J}_k^T \bar{y}_{k+1} - (g_k + J_k^T y_{k+1})\| = O(\epsilon_c + \epsilon_J + \epsilon_g).$$

For the sake of brevity and since finite termination is unlikely in practice, we omit the details of this bound, but claim that it would follow under the assumptions and results in Section 4.3 up through and including Corollary 4.1.

Assuming that Algorithm 1 does not terminate finitely, the aims of our analysis are to prove results that are modeled on the noiseless setting. In particular, we note the main convergence result in [11] and the main convergence result for a fixed barrier parameter in [15]. In both of these cases, the main aim is to show that, when the sequence of constraint Jacobians has singular values that are bounded away from zero, one has that the merit parameter remains bounded below by a positive real number and in the limit one finds

$$\{\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k)\} \to 0, \text{ which in turn yields } \left\{ \left\| \begin{bmatrix} \bar{g}_k + \bar{J}_k^T \bar{y}_{k+1} \\ \bar{c}_k \end{bmatrix} \right\| \right\} \to 0. \tag{22}$$

(See, e.g., Lemma 3.18 and its proof in [11].) Otherwise (when the constraint Jacobians tend toward rank deficiency), one at least has that $\{\|\bar{J}_k^T \bar{c}_k\|\} \to 0$. In our noisy setting, the case of potential rank deficiency of the constraint Jacobians presents a practical hurdle, especially when the noisy Jacobians tend toward rank deficiency while the true Jacobians do not. Our analysis thus focuses on the stronger results that can be proved when the constraint Jacobians do not tend toward rank deficiency and do not differ too much from their true values (see Assumptions 4.3 and 4.4 in Section 4.3). Overall, the main aim of our convergence analysis is to show that, under reasonable assumptions, Algorithm 1 generates merit parameter values that remain bounded below by a positive number and generates iterates such that, at least for some iteration index $k \in \mathbb{N}$, the value $\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k)$ (see (22)) is below a threshold that depends on the magnitude of the noise in the function and derivative values.

We also show under looser assumptions that the algorithm generates iterates such that, at least for some $k \in \mathbb{N}$, the value $\|\bar{J}_k^T \bar{c}_k\|^2$ is below a threshold that depends on the magnitude of the noise. This result, which is arguably of less interest in practice, is relegated to Appendix A.

4.2 Assumptions and Well-Posedness

Before proceeding to our analysis, let us introduce the assumptions that we make beyond Assumption 2.1. We also make an additional assumption on the bounds on the noisy function and derivative evaluations; this is presented after Lemma 4.8 since it relies on quantities introduced in that lemma.

First, we formalize our assumption pertaining to the sequence $\{W_k\}$ as follows. Recall that this assumption with respect to the noisy quantities has already been justified in the paragraph following (14). We strengthen it slightly to refer to noiseless values as well, since later on this will allow us to quantify, e.g., the difference between d_k and $\bar{d}_k$. (Alternatively, we could introduce a "true" matrix $\bar{W}_k$ that satisfies similar properties with respect to noiseless quantities as W_k does with respect to noisy ones. However, this would add extra complication without adding significant value to our analysis.)

Assumption 4.1. *The symmetric matrices $\{W_k\}$ defined by (13) is chosen such that: (a) for all $k \in \mathbb{N}$, (14) has a unique solution when either the noisy quantities (J_k, g_k, v_k) or noiseless quantities $(\bar{J}_k, \bar{g}_k, \bar{v}_k)$ are used; (b) for some $\sigma_W \in (0, \infty)$ and all $k \in \mathbb{N}$, one has $p^T W_k p \geq \sigma_W \|p\|^2$ for all $p \in \text{null}(J_k) \cup \text{null}(\bar{J}_k)$; and (c) for some $\kappa_W \in (0, \infty)$ and all $k \in \mathbb{N}$, one has $\|W_k\| \leq \kappa_W$.*

Second, while most of our analysis merely requires that the generated slack iterates have positive components, our later results require that the slack variables remain bounded. Thus, we introduce Assumption 4.2, below. In the noiseless setting, boundedness of the slack variables can be proved by the behavior of the algorithm and boundedness of the function and derivative values, as in Assumption 2.1; see [6, 15]. However, in our noisy setting, it happens that a similar conclusion cannot be drawn in the same way since the merit function values are not monotonically decreasing. In any case, we contend that the following Assumption 4.2 is weak under Assumption 2.1.

Assumption 4.2. *There exists $s_{\text{sup}} \in (0, \infty)$ such that $\|s_k\| \leq s_{\text{sup}}$ for all $k \in \mathbb{N}$. Thus, along with Assumption 2.1, $\{\bar{\phi}(z_k, \tau_k)\}$ is bounded below.*

Under Assumptions 2.1 and 4.2, there exists $(g_{\text{sup}}, J_{\text{sup}}) \in (0, \infty)^2$ such that

$$\|\bar{g}_k\| \leq g_{\text{sup}}, \quad \text{and} \quad \|\bar{J}_k\| \leq J_{\text{sup}} \quad \text{for all} \quad k \in \mathbb{N}. \tag{23}$$

Third, many parts of our analysis merely require that $\bar{J}_k$ has full row rank for all $k \in \mathbb{N}$, which in our setting is guaranteed since s_k has positive components for all $k \in \mathbb{N}$. However, some of our results require that, at least eventually, the Jacobians have singular values that are bounded above the noise level for the Jacobian values. Hence, let us introduce the following assumption.

Assumption 4.3. *There exist $k_J \in \mathbb{N}$ and $\gamma \in (\epsilon_J, \infty)$ such that, for all $k \in \mathbb{N}$ with $k \geq k_J$, one has that $\sigma_{\min}(J_k) \geq \gamma$.*

Under Assumptions 2.1 and 4.3, it follows that for all $k \in \mathbb{N}$ with $k \geq k_J$ one has $\sigma_{\min}(\bar{J}_k) \geq \gamma - \epsilon_J > 0$. Thus, there is $(\bar{\sigma}_J, \sigma_J) \in (0, \infty)^2$ such that, for such k,

$$\|(\bar{J}_k \bar{J}_k^T)^{-1}\| \leq \bar{\sigma}_J \quad \text{and} \quad \|(J_k J_k^T)^{-1}\| \leq \sigma_J. \tag{24}$$

It is worth emphasizing that both Assumption 4.3 and upcoming Assumption 4.4 impose implicit bounds on the noise level for the constraint Jacobian matrices.

Let us close this subsection by showing that, under our stated assumptions, the algorithm is well posed. For this purpose, it is worthwhile to note that the normal step v_k, which is the exact solution of (12), satisfies the Cauchy decrease condition [22], i.e., for some prescribed $\delta_v \in (0, 1]$, one has

$$\|c_k\| - \|c_k + J_k v_k\| \geq \delta_v(\|c_k\| - \|c_k + \hat{\alpha}_k J_k \hat{v}_k\|) \geq 0, \tag{25}$$

where $\hat{v}_k := -J_k^T c_k$ and $\hat{\alpha}_k$ is a solution of the one-dimensional problem

$$\min_{\hat{\alpha}} \tfrac{1}{2}\|c_k + \hat{\alpha} J_k \hat{v}_k\|^2 \quad \text{s.t.} \quad \hat{\alpha} \in [0, \omega]. \tag{26}$$

A Cauchy decrease condition also holds for the noiseless normal step $\bar{v}_k$ with $(\bar{c}_k, \bar{J}_k)$ in place of (c_k, J_k) and with $\hat{\alpha}_k$ redefined accordingly.

The lemma below summarizes prior remarks to show that, by its construction, Algorithm 1 is well defined in the sense that if it reaches iteration $k \in \mathbb{N}$, then each step involves finite computation and the algorithm will either terminate in iteration k or proceed to iteration $k + 1$. The lemma also shows that, as claimed, one finds $s_k \geq 0$ and $c(z_k) \geq 0$ for all generated $k \in \mathbb{N}$.

13

Lemma 4.1. *Suppose that Assumptions 2.1 and 4.1 hold. Then, if Algorithm 1 reaches iteration $k \in \mathbb{N}$ such that $s_k \geq 0$ and $c(z_k) \geq 0$, then it either terminates finitely or proceeds to iteration $k+1$ such that $s_{k+1} \geq 0$ and $c(z_{k+1}) \geq 0$. Thus, the algorithm either terminates finitely or generates an infinite sequence of iterates, where in either case $s_k \geq 0$ and $c(z_k) \geq 0$ for all generated $k \in \mathbb{N}$.*

Proof. Suppose that Algorithm 1 reaches iteration $k \in \mathbb{N}$ such that $s_k \geq 0$ and $c(z_k) \geq 0$. First, if it terminates in line 2, then there is nothing left to prove. Hence, we may proceed under the assumption that it continues, in which case one finds that (12) is feasible and there exists $v_k \in \text{range}(J_k^T)$ that is feasible for (12) and satisfies (25), meaning that the computation of v_k is well defined. It then follows under Assumption 4.1 that the computation of u_k and (d_k, y_{k+1}) is well defined. Next, either the algorithm terminates in line 6 or it reaches line 7, in which case it is guaranteed to reach line 8. Since $\eta_s \in (0, 1)$, line 8 is well defined. Similarly, since $\varepsilon_k \in (0, \infty)$, it follows that (20) holds for sufficiently small α_k, meaning that line 9 is well defined. It follows that line 11 is reached, which guarantees that $s_{k+1} \geq 0$ and $c(z_{k+1}) \geq 0$. $\qquad\square$

4.3 Convergence Analysis

Let us now proceed to the main part of our analysis. Since we have already discussed the consequences of finite terminate of the algorithm in Section 4.1, our analysis in this section presumes that the algorithm does not terminate.

Let us define for all $k \in \mathbb{N}$ the errors

$$\varepsilon_{v,k} := v_k - \bar{v}_k, \quad \varepsilon_{u,k} := u_k - \bar{u}_k, \quad \varepsilon_{d,k} := d_k - \bar{d}_k, \quad \text{and} \quad \varepsilon_{\tau,k} := \tau_k - \bar{\tau}_k. \tag{27}$$

Our preliminary analysis, prior to our main results, focuses on providing bounds on these errors in terms of the bounds on the errors from (11). Observe that, since $\{\tau_k\}$ and $\{\bar{\tau}_k\}$ are monotonically nonincreasing and bounded below by zero, it follows trivially that $\varepsilon_{\tau,k} \leq \tau_0$ for all $k \in \mathbb{N}$.

Our next few lemmas focus on properties of the constraint function value, constraint Jacobian value, and normal step. The lemmas involve properties of these quantities as well as differences between noisy and noiseless values.

Lemma 4.2. *Suppose Assumptions 2.1 and 4.1 hold. Then, for all $k \in \mathbb{N}$,*

$$\|J_k^T J_k - \bar{J}_k^T \bar{J}_k\| \leq \epsilon_J(\epsilon_J + 2\|\bar{J}_k\|), \tag{28a}$$

$$\|J_k J_k^T - \bar{J}_k \bar{J}_k^T\| \leq \epsilon_J(\epsilon_J + 2\|\bar{J}_k\|), \tag{28b}$$

$$\|(J_k J_k^T)^{1/2} - (\bar{J}_k \bar{J}_k^T)^{1/2}\| \leq (\epsilon_J(\epsilon_J + 2\|\bar{J}_k\|))^{1/2}, \tag{28c}$$

$$and \quad \|\bar{J}_k^T \bar{c}_k - J_k^T c_k\| \leq \|\bar{J}_k\|\epsilon_c + (\|\bar{c}_k\| + \epsilon_c)\epsilon_J. \tag{28d}$$

Proof. Consider arbitrary $k \in \mathbb{N}$. By Assumption 2.1, the triangle inequality, and submultiplicity of the matrix 2-norm, one finds that

$$\|J_k^T J_k - \bar{J}_k^T \bar{J}_k\| = \|J_k^T J_k - J_k^T \bar{J}_k + J_k^T \bar{J}_k - \bar{J}_k^T \bar{J}_k\|$$
$$\leq \|J_k^T\|\epsilon_J + \|\bar{J}_k\|\epsilon_J \leq (\epsilon_J + \|\bar{J}_k^T\|)\epsilon_J + \|\bar{J}_k\|\epsilon_J = \epsilon_J(\epsilon_J + 2\|\bar{J}_k\|),$$

which gives (28a). The proof of inequality (28b) is similar to that of (28a). Next, by $\frac{1}{2}$-Hölder continuity of the square root operator for symmetric positive definite matrices and the fact that

14

the matrix 2-norm is unitarily invariant, inequality (28c) follows from inequality (28a). Lastly, by (11), one finds

$$\|\bar{J}_k^T \bar{c}_k - J_k^T c_k\| \leq \|\bar{J}_k^T \bar{c}_k - \bar{J}_k^T c_k + \bar{J}_k^T c_k - J_k^T c_k\| \leq \|\bar{J}_k\|\epsilon_c + (\|\bar{c}_k\| + \epsilon_c)\epsilon_J,$$

which gives (28d). $\qquad\square$

The following useful lemma is well known. Its proof can be found in [6].

Lemma 4.3. *Given $a \in \mathbb{R}$, $b \in [0,\infty)$, and $\omega \in (0,\infty)$, the optimal value Φ^* of*

$$\min_{z \in [0,\omega)} \ \Phi(z) \equiv \tfrac{1}{2}z^2 a - zb \ \ \text{satisfies} \ \ \Phi^* \leq -\tfrac{b}{2}\min\{\tfrac{b}{|a|},\omega\}.$$

Lemma 4.3 can be used to prove our next lemma. A proof of a similar result, with less detail on the form of the lower bound that is proved, can be found in [11]. We provide a proof here to show the precise form of the lower bound.

Lemma 4.4. *Suppose Assumptions 2.1 and 4.1 hold. Then, for all $k \in \mathbb{N}$,*

$$\|\bar{c}_k\|(\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k\bar{v}_k\|) \geq \xi_{1,k}\|\bar{J}_k^T\bar{c}_k\|^2, \ \text{where} \ \xi_{1,k} = \tfrac{1}{2}\delta_v \min\left\{\frac{1}{\|\bar{J}_k^T\bar{J}_k\|}, \omega\right\}.$$

Proof. Consider arbitrary $k \in \mathbb{N}$ and recall that $\bar{v}_k$ is the solution of (12) when (c_k, J_k) is replaced by $(\bar{c}_k, \bar{J}_k)$. If $\|\bar{J}_k^T\bar{c}_k\| = 0$, then $\bar{v}_k = 0$, so the desired conclusion follows trivially. Hence, we may proceed under the assumption that $\|\bar{J}_k^T\bar{c}_k\| > 0$. By (26) with respect to noiseless quantities, for which the steepest-descent direction is $\hat{v}_k = -\bar{J}_k^T\bar{c}_k$, Cauchy decrease is attained with

$$\hat{\alpha}_k \in \arg\min_{\hat{\alpha} \in [0,\omega]} \ \tfrac{1}{2}\hat{\alpha}^2\|\bar{J}_k\bar{J}_k^T\bar{c}_k\|^2 - \hat{\alpha}\|\bar{J}_k^T\bar{c}_k\|^2.$$

Hence, by Lemma 4.3 and $\|\bar{J}_k^T\bar{c}_k\|^2\|\bar{J}_k^T\bar{J}_k\| \geq \|\bar{J}_k\bar{J}_k^T\bar{c}_k\|^2$, it follows that

$$\begin{aligned}
\tfrac{1}{2}(\|\bar{c}_k + \hat{\alpha}_k\bar{J}_k\hat{v}_k\|^2 - \|\bar{c}_k\|^2) &= \tfrac{1}{2}\hat{\alpha}_k^2\|\bar{J}_k\bar{J}_k^T\bar{c}_k\|^2 - \hat{\alpha}_k\|\bar{J}_k^T\bar{c}_k\|^2 \\
&\leq -\tfrac{1}{2}\|\bar{J}_k^T\bar{c}_k\|^2 \min\left\{\frac{\|\bar{J}_k^T\bar{c}_k\|^2}{\|\bar{J}_k\bar{J}_k^T\bar{c}_k\|^2}, \omega\right\} \\
&\leq -\tfrac{1}{2}\|\bar{J}_k^T\bar{c}_k\|^2 \min\left\{\frac{1}{\|\bar{J}_k^T\bar{J}_k\|}, \omega\right\}. \qquad (29)
\end{aligned}$$

Since $\bar{v}_k$ satisfies Cauchy decrease (see (25)) with respect to noiseless quantities and since $2a(a-b) \geq a^2 - b^2$ for arbitrary $(a,b) \in \mathbb{R} \times \mathbb{R}$, one finds that

$$\begin{aligned}
\|\bar{c}_k\|(\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k\bar{v}_k\|) &\geq \delta_v\|\bar{c}_k\|(\|\bar{c}_k\| - \|\bar{c}_k + \hat{\alpha}_k\bar{J}_k\hat{v}_k\|) \\
&\geq \tfrac{1}{2}\delta_v(\|\bar{c}_k\|^2 - \|\bar{c}_k + \hat{\alpha}_k\bar{J}_k\hat{v}_k\|^2) \\
&\geq \tfrac{1}{2}\delta_v\|\bar{J}_k^T\bar{c}_k\|^2 \min\left\{\frac{1}{\|\bar{J}_k^T\bar{J}_k\|}, \omega\right\},
\end{aligned}$$

which is the desired conclusion. $\qquad\square$

We now show that, under all of our aforementioned assumptions, the difference between the noisy and noiseless normal step is bounded uniformly for all sufficiently large $k \in \mathbb{N}$. The proof follows a similar argument as in [4].

Lemma 4.5. *Suppose Assumptions 2.1, 4.1, 4.2, and 4.3 hold. Then, there exists $\epsilon_v \in [0, \infty)$ such that for all $k \in \mathbb{N}$ with $k \geq k_J$ one has*

$$\|\bar{v}_k - v_k\| \leq \epsilon_v. \tag{30}$$

In addition, (30) holds with $\epsilon_v \to 0$ as $(\epsilon_c, \epsilon_J) \searrow (0,0)$.

Proof. Consider arbitrary $k \in \mathbb{N}$ with $k \geq k_J$. Since (12) requires $v \in \text{range}(J_k^T)$, one can express the solution of (12) as $v_k = J_k^T t_k$, where t_k solves

$$\min_{t \in \mathbb{R}^q} \quad \tfrac{1}{2}\|c_k + J_k J_k^T t\|^2 \quad \text{s.t.} \quad \|J_k^T t\| \leq \omega\|J_k^T c_k\|. \tag{31}$$

Let us now define, as allowed under Assumption 4.3, the change of variables

$$\hat{t} := \frac{\|J_k^T c_k\|}{\|\bar{J}_k^T \bar{c}_k\|}(J_k J_k^T)^{-1/2}(\bar{J}_k \bar{J}_k^T)^{1/2}\bar{t}. \tag{32}$$

Since $\|(J_k J_k^T)^{1/2}\hat{t}\| = \|J_k^T \hat{t}\|$ and $\|(\bar{J}_k \bar{J}_k^T)^{1/2}\bar{t}\| = \|\bar{J}_k^T \bar{t}\|$, it follows that $\|J_k^T \hat{t}\| = \frac{\|J_k^T c_k\|}{\|\bar{J}_k^T \bar{c}_k\|}\|\bar{J}_k^T \bar{t}\|$. Now considering (31) with respect to the noiseless quantities, but replacing the variable $\bar{t}$ with $\frac{\|\bar{J}_k^T \bar{c}_k\|}{\|J_k^T c_k\|}(\bar{J}_k \bar{J}_k^T)^{-1/2}(J_k J_k^T)^{1/2}\hat{t}$, one obtains

$$\min_{\hat{t} \in \mathbb{R}^q} \quad \tfrac{1}{2}\|\bar{c}_k + \frac{\|\bar{J}_k^T \bar{c}_k\|}{\|J_k^T c_k\|}(\bar{J}_k \bar{J}_k^T)^{1/2}(J_k J_k^T)^{1/2}\hat{t}\|^2 \quad \text{s.t.} \quad \|J_k^T \hat{t}\| \leq \omega\|J_k^T c_k\|. \tag{33}$$

The feasible regions of (31) and (33) are identical. Thus, let us denote the feasible region of these problems as the nonempty, closed, and convex set $\Pi := \{t \in \mathbb{R}^q : \|J_k^T t\| \leq \omega\|J_k^T c_k\|\}$. Problems (31) and (33) can now be written respectively as

$$\min_{t \in \Pi} \quad \tfrac{1}{2}\|(J_k J_k^T)^{-1} c_k + t\|^2_{(J_k J_k^T)^2}, \quad \text{and} \tag{34}$$

$$\min_{\hat{t} \in \Pi} \quad \tfrac{1}{2}\left\|\frac{\|J_k^T c_k\|}{\|\bar{J}_k^T \bar{c}_k\|}(J_k J_k^T)^{-1/2}(\bar{J}_k \bar{J}_k^T)^{-1/2}\bar{c}_k + \hat{t}\right\|^2_{(J_k J_k^T)^{1/2}(\bar{J}_k \bar{J}_k^T)(J_k J_k^T)^{1/2}}. \tag{35}$$

Let $(t^*, \hat{t}^*) \in \Pi \times \Pi$ be the optimal solutions of (34) and (35), respectively. (Since their objective functions are strongly convex, these solutions are unique.) By first-order optimality, respectively, for all $t \in \Pi$ one has

$$\langle -(J_k J_k^T)^{-1} c_k - t^*, t - t^* \rangle_{(J_k J_k^T)^2} \leq 0$$

$$\langle \frac{\|J_k^T c_k\|}{\|\bar{J}_k^T \bar{c}_k\|}(J_k J_k^T)^{-1/2}(\bar{J}_k \bar{J}_k^T)^{-1/2}\bar{c}_k + \hat{t}^*, \hat{t}^* - t \rangle_{(J_k J_k^T)^{1/2}\bar{J}_k \bar{J}_k^T (J_k J_k^T)^{1/2}} \leq 0.$$

Substituting $\hat{t}^*$ for t in the first inequality, substituting t^* for t in the second inequality, and summing the results, one finds that $I \leq 0$, where

$$I := -((J_k J_k^T)^{-1} c_k + t^*)^T (J_k J_k^T)^2 (\hat{t}^* - t^*)$$
$$+ (\frac{\|J_k^T c_k\|}{\|\bar{J}_k^T \bar{c}_k\|}(J_k J_k^T)^{-1/2}(\bar{J}_k \bar{J}_k^T)^{-1/2}\bar{c}_k + \hat{t}^*)^T (J_k J_k^T)^{1/2}\bar{J}_k \bar{J}_k^T (J_k J_k^T)^{1/2}(\hat{t}^* - t^*).$$

By expanding the inner products in I, one may write $I = I_1 + I_2$, where

$$I_1 := -(t^*)^T (J_k J_k^T)^2 (\hat{t}^* - t^*) + (\hat{t}^*)^T (J_k J_k^T)^{1/2}\bar{J}_k \bar{J}_k^T (J_k J_k^T)^{1/2}(\hat{t}^* - t^*)$$

16

and $I_2 := \frac{\|J_k^T c_k\|}{\|\bar{J}_k^T \bar{c}_k\|} \bar{c}_k^T (\bar{J}_k \bar{J}_k^T)^{1/2} (J_k J_k^T)^{1/2} (\hat{t}^* - t^*) - c_k^T J_k J_k^T (\hat{t}^* - t^*).$

By adding and subtracting $(\hat{t}^*)^T (J_k J_k^T)^2 (\hat{t}^* - t^*)$, I_1 can be written as

$$I_1 = (\hat{t}^* - t^*)^T (J_k J_k^T)^2 (\hat{t}^* - t^*)$$
$$+ (\hat{t}^*)^T (J_k J_k^T)^{1/2} \left(\bar{J}_k \bar{J}_k^T - J_k J_k^T \right) (J_k J_k^T)^{1/2} (\hat{t}^* - t^*)$$
$$= \|\hat{t}^* - t^*\|^2_{(J_k J_k^T)^2} + (\hat{t}^*)^T (J_k J_k^T)^{1/2} \left(\bar{J}_k \bar{J}_k^T - J_k J_k^T \right) (J_k J_k^T)^{1/2} (\hat{t}^* - t^*),$$

In addition, by adding and subtracting both $\bar{c}_k^T (\bar{J}_k \bar{J}_k^T)^{1/2} (J_k J_k^T)^{1/2} (\hat{t}^* - t^*)$ and $c_k^T (\bar{J}_k \bar{J}_k^T)^{1/2} (J_k J_k^T)^{1/2} (\hat{t}^* - t^*)$, I_2 can be written as

$$I_2 = (-(1 - \frac{\|J_k^T c_k\|}{\|\bar{J}_k^T \bar{c}_k\|}) \bar{c}_k^T (\bar{J}_k \bar{J}_k^T)^{1/2}$$
$$+ (\bar{c}_k - c_k)^T (\bar{J}_k \bar{J}_k^T)^{1/2} + c_k^T ((\bar{J}_k \bar{J}_k^T)^{1/2} - (J_k J_k^T)^{1/2}))(J_k J_k^T)^{1/2} (\hat{t}^* - t^*).$$

Thus, $I = I_1 + I_2 \leq 0$ along with these inequalities yields

$$\|\hat{t}^* - t^*\|^2_{(J_k J_k^T)^2}$$
$$\leq ((\hat{t}^*)^T (J_k J_k^T)^{1/2} (J_k J_k^T - \bar{J}_k \bar{J}_k^T) + (1 - \frac{\|J_k^T c_k\|}{\|\bar{J}_k^T \bar{c}_k\|}) \bar{c}_k^T (\bar{J}_k \bar{J}_k^T)^{1/2}$$
$$+ (c_k - \bar{c}_k)^T (\bar{J}_k \bar{J}_k^T)^{1/2} + c_k^T ((J_k J_k^T)^{1/2} - (\bar{J}_k \bar{J}_k^T)^{1/2}))(J_k J_k^T)^{1/2} (\hat{t}^* - t^*).$$

Thus, by submultiplicity of the matrix 2-norm, the facts that $\|(J_k J_k^T)^{1/2} \hat{t}^*\| = \|J_k^T \hat{t}^*\|$ and $\|(\bar{J}_k \bar{J}_k^T)^{1/2} \bar{c}_k\| = \|\bar{J}_k^T \bar{c}_k\|$, the definition of the feasible set Π, and the fact that for any pair of vectors $(a, b) \in \mathbb{R}^n \times \mathbb{R}^n$ one has $\|\|a\| - \|b\|\| \leq \|a - b\|$, it follows that

$$\|\hat{t}^* - t^*\|^2_{(J_k J_k^T)^2}$$
$$\leq (\|J_k^T \hat{t}^*\| \|J_k J_k^T - \bar{J}_k \bar{J}_k^T\| + |1 - \frac{\|J_k^T c_k\|}{\|\bar{J}_k^T \bar{c}_k\|}| \|\bar{J}_k^T \bar{c}_k\| + \|c_k - \bar{c}_k\| \|(\bar{J}_k \bar{J}_k^T)^{1/2}\|$$
$$+ \|c_k\| \|(J_k J_k^T)^{1/2} - (\bar{J}_k \bar{J}_k^T)^{1/2}\|)\|(J_k J_k^T)^{1/2} (\hat{t}^* - t^*)\|$$
$$\leq (\omega \|J_k^T c_k\| \|J_k J_k^T - \bar{J}_k \bar{J}_k^T\| + \|\bar{J}_k^T \bar{c}_k - J_k^T c_k\| + \|c_k - \bar{c}_k\| \|\bar{J}_k\|$$
$$+ \|c_k\| \|(J_k J_k^T)^{1/2} - (\bar{J}_k \bar{J}_k^T)^{1/2}\|)\|(J_k J_k^T)^{1/2}\| \|\hat{t}^* - t^*\|.$$

This bound, submultiplicity of the matrix 2-norm, and Lemma 4.2 yield

$$\|\hat{t}^* - t^*\|^2_{(J_k J_k^T)^2}$$
$$\leq (\omega(\|\bar{J}_k\| + \epsilon_J)(\|\bar{c}_k\| + \epsilon_c)\epsilon_J(\epsilon_J + 2\|\bar{J}_k\|) + 2\|\bar{J}_k\|\epsilon_c + (\|\bar{c}_k\| + \epsilon_c)\epsilon_J$$
$$+ (\|\bar{c}_k\| + \epsilon_c)(\epsilon_J(\epsilon_J + 2\|\bar{J}_k\|))^{1/2})(\|\bar{J}_k\| + \epsilon_J)\|\hat{t}^* - t^*\|.$$

Hence, along with $\|\hat{t}^* - t^*\|^2 \leq \|(J_k J_k^T)^{-2}\| \|\hat{t}^* - t^*\|^2_{(J_k J_k^T)^2}$ and (24) one has

$$\|\hat{t}^* - t^*\|$$
$$\leq \sigma_J^2 (\|\bar{J}_k\| + \epsilon_J)(\omega(\|\bar{J}_k\| + \epsilon_J)(\|\bar{c}_k\| + \epsilon_c)\epsilon_J(\epsilon_J + 2\|\bar{J}_k\|)$$
$$+ 2\|\bar{J}_k\|\epsilon_c + (\|\bar{c}_k\| + \epsilon_c)\epsilon_J + (\|\bar{c}_k\| + \epsilon_c)(\epsilon_J(\epsilon_J + 2\|\bar{J}_k\|))^{1/2})$$

$$=: \epsilon_{t_1,k}. \tag{36}$$

Now let $\bar{t}^* \in \Pi$ be the unique optimal solution of (31) with respect to the noiseless quantities. Our next aim is to bound $\|\bar{t}^* - t^*\|$. That being said, since $\|\bar{t}^* - t^*\| \leq \|\bar{t}^* - \hat{t}^*\| + \|\hat{t}^* - t^*\|$, where $\|\hat{t}^* - t^*\|$ has the bound given in (36), we only need to bound $\|\hat{t}^* - \bar{t}^*\|$. By (24), (32), and Lemma 4.2,

$$\|\bar{t}^* - \hat{t}^*\|$$
$$= \|(I - \tfrac{\|J_k^T c_k\|}{\|\bar{J}_k^T \bar{c}_k\|}(J_k J_k^T)^{-1/2}(\bar{J}_k \bar{J}_k^T)^{1/2})\bar{t}^*\|$$
$$= \|(\tfrac{\|J_k^T c_k\|}{\|\bar{J}_k^T \bar{c}_k\|}I - \tfrac{\|J_k^T c_k\|}{\|\bar{J}_k^T \bar{c}_k\|}I + I - \tfrac{\|J_k^T c_k\|}{\|\bar{J}_k^T \bar{c}_k\|}(J_k J_k^T)^{-1/2}(\bar{J}_k \bar{J}_k^T)^{1/2})\bar{t}^*\|$$
$$\leq \tfrac{\|J_k^T c_k\|}{\|\bar{J}_k^T \bar{c}_k\|}(|\tfrac{\|\bar{J}_k^T \bar{c}_k\|}{\|J_k^T c_k\|} - 1| + \|I - (J_k J_k^T)^{-1/2}(\bar{J}_k \bar{J}_k^T)^{1/2}\|)$$
$$\quad \cdot \|(\bar{J}_k \bar{J}_k^T)^{-1/2}\|\|(\bar{J}_k \bar{J}_k^T)^{1/2}\bar{t}^*\|$$
$$\leq \omega\|J_k^T c_k\|\|(\bar{J}_k \bar{J}_k^T)^{-1/2}\|$$
$$\quad \cdot (|1 - \tfrac{\|\bar{J}_k^T \bar{c}_k\|}{\|J_k^T c_k\|}| + \|(J_k J_k^T)^{-1/2}\|\|(J_k J_k^T)^{1/2} - (\bar{J}_k \bar{J}_k^T)^{1/2}\|)$$
$$= \omega\|(\bar{J}_k \bar{J}_k^T)^{-1/2}\|(|\|J_k^T c_k\| - \|\bar{J}_k^T \bar{c}_k\||$$
$$\quad + \|J_k^T c_k\|\|(J_k J_k^T)^{-1/2}\|\|(J_k J_k^T)^{1/2} - (\bar{J}_k \bar{J}_k^T)^{1/2}\|)$$
$$\leq \omega\sqrt{\bar{\sigma}_J}(\|\bar{J}_k\|\epsilon_c + (\|\bar{c}_k\| + \epsilon_c)\epsilon_J$$
$$\quad + (\|\bar{J}_k^T \bar{c}_k\| + \|\bar{J}_k\|\epsilon_c + \epsilon_J(\|\bar{c}_k\| + \epsilon_c))\sqrt{\sigma_J}(\epsilon_J(\epsilon_J + 2\|\bar{J}_k\|))^{1/2}) =: \epsilon_{t_2,k}.$$

Thus, with (24) and (36), one has $\|\hat{t}^* - t^*\| \leq \epsilon_{t_1,k} + \epsilon_{t_2,k} =: \epsilon_{t,k}$, so

$$\|\bar{v}_k - v_k\| = \|\bar{J}_k^T \hat{t}^* - J_k^T \hat{t}^* + J_k^T \hat{t}^* - J_k^T t^*\|$$
$$\leq \|\bar{J}_k - J_k\|\|(\bar{J}_k \bar{J}_k^T)^{-1/2}\|\|(\bar{J}_k \bar{J}_k^T)^{1/2}\hat{t}^*\| + \|J_k\|\|\hat{t}^* - t^*\|$$
$$\leq \epsilon_J \omega \sqrt{\bar{\sigma}_J}\|\bar{J}_k^T \bar{c}_k\| + (\|\bar{J}_k\| + \epsilon_J)\epsilon_{t,k}.$$

Thus, under Assumptions 2.1 and 4.2, the fact that $\epsilon_{t,k} := \epsilon_{t_1,k} + \epsilon_{t_2,k}$, and the definitions of $\epsilon_{t_1,k}$ and $\epsilon_{t_2,k}$, the desired conclusions follow. $\qquad\square$

We now present a few lemmas that focus on properties of the tangential steps. Our goal in these lemmas is to prove a result similar to Lemma 4.5, except in terms of the tangential steps. Our derivation of this result relies on expressing each tangential step as the result of an oblique projection onto the null space of a constraint Jacobian. Hence, our first result pertains to oblique projection matrices. The result is known, but we present it for the sake of completeness. For a matrix $A \in \mathbb{R}^{n \times n}$, the M-adjoint of A is a matrix $A^* \in \mathbb{R}^{n \times n}$ such that $\langle u, Av \rangle_M = \langle A^*u, v \rangle_M$ for all $(u, v) \in \mathbb{R}^n \times \mathbb{R}^n$. For a symmetric positive definite matrix M, the M-adjoint of A equals $A^* = M^{-1}A^T M$. In addition, a matrix P is called an M-orthogonal projection matrix if and only if it is idempotent ($P^2 = P$) and M-self-adjoint ($P = P^*$).

Lemma 4.6. *Suppose that the matrix $J \in \mathbb{R}^{q \times (n+q)}$ has full row rank, the columns of the matrix $Z \in \mathbb{R}^{(n+q) \times n}$ form an orthonormal basis for $\mathrm{null}(J)$, and $\Gamma \in \mathbb{R}^{(n+q) \times (n+q)}$ is symmetric and positive definite. Then,*

$$P := Z(Z^T \Gamma Z)^{-1} Z^T \Gamma = I - \Gamma^{-1} J^T (J \Gamma^{-1} J^T)^{-1} J \in \mathbb{R}^{(n+q) \times (n+q)}$$

is a Γ-orthogonal projection matrix that projects $u \in \mathbb{R}^{n+q}$ onto $\mathrm{null}(J)$ *and*

$$Q := \Gamma^{-1} J^T (J\Gamma^{-1} J^T)^{-1} J \in \mathbb{R}^{(n+q) \times (n+q)}$$

is a Γ-orthogonal projection matrix that projects $v \in \mathbb{R}^{n+q}$ onto $\mathrm{range}(\Gamma^{-1} J^T)$, *i.e., the space that is Γ-orthogonal to* $\mathrm{null}(J)$.

Proof. It is straightforward to verify that the stated matrices for P are equal and that both P and Q are idempotent and Γ-self-adjoint. Furthermore, given any $u \in \mathbb{R}^{n+q}$, one finds that $JPu = 0$, which shows that Pu lies in $\mathrm{null}(J)$, and given any $v \in \mathbb{R}^{n+q}$, one finds that $Qv \in \mathrm{range}(\Gamma^{-1} J^T)$, as claimed. $\qquad\square$

Let us now introduce the following result, which shows that the noisy and noiseless tangential steps can be expressed in terms of solutions of linear systems involving certain positive definite matrices that, under our assumptions, have eigenvalues that are uniformly contained within a positive interval.

Lemma 4.7. *Suppose that Assumptions 2.1, 4.1, 4.2, and 4.3 hold. Then, there exists $(\lambda, \sigma_\Gamma) \in [0, \infty) \times (0, \infty)$ such that, for all $k \in \mathbb{N}$, one can express*

$$
\begin{bmatrix} W_k + \lambda J_k^T J_k & J_k^T \\ J_k & 0 \end{bmatrix} \begin{bmatrix} u_k \\ y_{k+1} \end{bmatrix} = - \begin{bmatrix} g_k + W v_k \\ 0 \end{bmatrix}
$$

$$
and \quad \begin{bmatrix} W_k + \lambda \bar{J}_k^T \bar{J}_k & \bar{J}_k^T \\ \bar{J}_k & 0 \end{bmatrix} \begin{bmatrix} \bar{u}_k \\ \bar{y}_{k+1} \end{bmatrix} = - \begin{bmatrix} \bar{g}_k + W \bar{v}_k \\ 0 \end{bmatrix}, \tag{37}
$$

where $\Gamma_k := W_k + \lambda J_k^T J_k$ and $\bar{\Gamma}_k := W_k + \lambda \bar{J}_k^T \bar{J}_k$ satisfy

$$\min\{\sigma_{\min}(\Gamma_k), \sigma_{\min}(\bar{\Gamma}_k)\} \geq \sigma_\Gamma.$$

Proof. Consider arbitrary $k \in \mathbb{N}$. By the second block equation in (14), one finds that $u_k \in \mathrm{null}(J_k)$ and $\bar{u}_k \in \mathrm{null}(\bar{J}_k)$. Therefore, by (14), the pairs (u_k, y_{k+1}) and $(\bar{u}_k, \bar{y}_{k+1})$ satisfy (37) for any $\lambda \in [0, \infty)$. The remainder of the proof follows by known results; see, e.g., Section 5.4.2 of [18]. $\quad\square$

Lemma 4.8. *Suppose that Assumptions 2.1, 4.1, 4.2, and 4.3 hold. Then, for all $k \in \mathbb{N}$ with Γ_k and $\bar{\Gamma}_k$ defined as in Lemma 4.7, one has that*

$$\|\Gamma_k^{-1} - \bar{\Gamma}_k^{-1}\| \leq \lambda \sigma_\Gamma^{-2} \epsilon_J (\epsilon_J + 2\|\bar{J}_k\|). \tag{38}$$

Proof. Consider arbitrary $k \in \mathbb{N}$. By Lemma 4.7, submultiplicity of the matrix 2-norm, and Lemma 4.2, it follows that

$$
\begin{aligned}
\|\Gamma_k^{-1} - \bar{\Gamma}_k^{-1}\| &= \|\Gamma_k^{-1}(\bar{\Gamma}_k - \Gamma_k)\bar{\Gamma}_k^{-1}\| \\
&\leq \|\Gamma_k^{-1}\| \|\bar{\Gamma}_k^{-1}\| \|\Gamma_k - \bar{\Gamma}_k\| \\
&\leq \lambda \sigma_\Gamma^{-2} \|J_k J_k^T - \bar{J}_k \bar{J}_k^T\| \leq \lambda \sigma_\Gamma^{-2} \epsilon_J (\epsilon_J + 2\|\bar{J}_k\|),
\end{aligned}
$$

which gives the desired conclusion. $\qquad\square$

We are now prepared to prove a uniform bound on the norm of the differences between the noisy and noiseless tangential steps. The bound that we prove, stated in Lemma 4.9 below, requires an additional assumption pertaining to the noise in the constraint Jacobians, which we state next. That the assumption pertains to the noise in the constraint Jacobians specifically can be seen by observing the manner in which the matrices $\{\Gamma_k\}$ and $\{\bar{\Gamma}_k\}$ are defined through Lemma 4.7 (i.e., with the same matrix W_k for both the noisy and noiseless linear systems in (37)). Whereas Assumption 4.3 requires that, for large k, the noise in each constraint Jacobian estimate is small relative to singular values, the following assumption essentially requires that the noise in the constraint Jacobian estimates is small enough such that the null space defined by the noisy and noiseless values does not differ too much.

Assumption 4.4. *Let $k_J \in \mathbb{N}$ be defined as in Assumption 4.3, and for all $k \in \mathbb{N}$ let Γ_k and $\bar{\Gamma}_k$ be defined as in Lemma 4.7. Then, for all $k \in \mathbb{N}$ with $k \geq k_J$,*

$$\|(\bar{J}_k \bar{\Gamma}_k^{-1} \bar{J}_k)^{-1}\|(\|\bar{J}_k\|\|\Gamma_k^{-1} J_k^T - \bar{\Gamma}_k^{-1} \bar{J}_k^T\| + \|\bar{\Gamma}_k^{-1} \bar{J}_k^T\|\epsilon_J + \|\Gamma_k^{-1} J_k^T - \bar{\Gamma}_k^{-1} \bar{J}_k^T\|\epsilon_J) \quad < \quad 1.$$

Observe that if $\epsilon_J = 0$, then $J_k = \bar{J}_k$ and $\Gamma_k = \bar{\Gamma}_k$, in which case the expression on the left-hand side of the inequality in Assumption 4.4 is 0, so the assumption holds. It follows that the assumption also holds for ϵ_J sufficiently small relative to the magnitudes of the elements of $\bar{J}_k$ and W_k for all $k \in \mathbb{N}$.

Lemma 4.9. *Suppose that Assumptions 2.1, 4.1, 4.2, 4.3, and 4.4 hold. Then, there exists $\epsilon_u \in [0, \infty)$ such that for all $k \in \mathbb{N}$ with $k \geq k_J$ one has*

$$\|u_k - \bar{u}_k\| \leq \epsilon_u. \tag{39}$$

In addition, (39) holds with $\epsilon_u \to 0$ as $(\epsilon_c, \epsilon_J, \epsilon_g) \searrow (0, 0, 0)$.

Proof. Consider arbitrary $k \in \mathbb{N}$ with $k \geq k_J$. Let Z_k and $\bar{Z}_k$ have columns that form orthonormal bases for $\text{null}(J_k)$ and $\text{null}(\bar{J}_k)$, respectively. From (37),

$$u_k = -Z_k (Z_k^T \Gamma_k Z_k)^{-1} Z_k^T \Gamma_k (\Gamma_k^{-1} g_k + \Gamma_k^{-1} W_k v_k),$$
$$\text{and} \quad \bar{u}_k = -\bar{Z}_k (\bar{Z}_k^T \bar{\Gamma}_k \bar{Z}_k)^{-1} \bar{Z}_k^T \bar{\Gamma}_k (\bar{\Gamma}_k^{-1} \bar{g}_k + \bar{\Gamma}_k^{-1} W_k \bar{v}_k),$$

where Γ_k and $\bar{\Gamma}_k$ are defined as in Lemma 4.7. By Lemma 4.6, the matrices

$$P_k := Z_k (Z_k^T \Gamma_k Z_k)^{-1} Z_k^T \Gamma_k \quad \text{and} \quad \bar{P}_k := \bar{Z}_k (\bar{Z}_k^T \bar{\Gamma}_k \bar{Z}_k)^{-1} \bar{Z}_k^T \bar{\Gamma}_k$$

are a Γ_k-orthogonal projection matrix on $\text{null}(J_k)$ and a $\bar{\Gamma}_k$-orthogonal projection matrix on $\text{null}(\bar{J}_k)$, respectively. One finds that

$$\begin{aligned}
\|u_k - \bar{u}_k\| &= \|P_k (\Gamma_k^{-1} g_k + \Gamma_k^{-1} W_k v_k) - \bar{P}_k (\Gamma_k^{-1} g_k + \Gamma_k^{-1} W_k v_k) \\
&\quad + \bar{P}_k (\Gamma_k^{-1} g_k + \Gamma_k^{-1} W_k v_k) - \bar{P}_k (\bar{\Gamma}_k^{-1} \bar{g}_k + \bar{\Gamma}_k^{-1} W_k \bar{v}_k)\| \\
&\leq \|P_k - \bar{P}_k\|\|\Gamma_k^{-1} g_k + \Gamma_k^{-1} W_k v_k\| \\
&\quad + \|\bar{P}_k\|\|\Gamma_k^{-1}(g_k + W_k v_k) - \bar{\Gamma}_k^{-1}(\bar{g}_k + W_k \bar{v}_k)\|.
\end{aligned} \tag{40}$$

Our aim now is to establish upper bounds for the terms on the right-hand side of (40). To bound $\|P_k - \bar{P}_k\|$, we define $Q_k = \Gamma_k^{-1} J_k^T (J \Gamma_k^{-1} J_k^T)^{-1} J_k$ and $\bar{Q}_k = \bar{\Gamma}_k^{-1} \bar{J}_k^T (\bar{J}_k \bar{\Gamma}_k^{-1} \bar{J}_k^T)^{-1} \bar{J}_k$. By Lemma 4.6,

Theorem 3.4 in [27] regarding the properties of the oblique projection matrices (under Assumption 4.4), one finds

$$\|P_k - \bar{P}_k\| = \|Q_k - \bar{Q}_k\|$$

$$\leq \|\bar{Q}_k\|(1 + \|\bar{Q}_k\|)\min\{\kappa(\bar{\Gamma}_k^{-1}\bar{J}_k^T), \kappa((\bar{J}_k\bar{\Gamma}_k^{-1}\bar{J}_k^T)^{-1}\bar{J}_k)\}\frac{\|\Gamma_k^{-1}J_k^T - \bar{\Gamma}_k^{-1}\bar{J}_k^T\|}{\|\bar{\Gamma}_k^{-1}\bar{J}_k^T\|}$$

$$+ \|\bar{Q}_k\|(1 + \|\bar{Q}_k\|)\min\{\kappa(\bar{J}_k^T), \kappa(\bar{\Gamma}_k^{-1}\bar{J}_k^T(\bar{J}_k\bar{\Gamma}_k^{-1}\bar{J}_k^T)^{-1})\}\frac{\epsilon_J}{\|\bar{J}_k^T\|}$$

$$+ O(\|\bar{Q}_k\|\max\{\|\Gamma_k^{-1}J_k^T - \bar{\Gamma}_k^{-1}\bar{J}_k^T\|, \epsilon_J\}^2). \tag{41}$$

By Lemmas 4.2, 4.7, and 4.8 along with Assumption 2.1, one has

$$\|\Gamma_k^{-1}J_k^T - \bar{\Gamma}_k^{-1}\bar{J}_k^T\| = \|\Gamma_k^{-1}J_k^T - \Gamma_k^{-1}\bar{J}_k^T + \Gamma_k^{-1}\bar{J}_k^T - \bar{\Gamma}_k^{-1}\bar{J}_k^T\|$$

$$\leq \|\Gamma_k^{-1}\|\epsilon_J + \|\Gamma_k^{-1} - \bar{\Gamma}_k^{-1}\|\|\bar{J}_k\|$$

$$\leq \sigma_\Gamma^{-1}(\epsilon_J + \lambda\sigma_\Gamma^{-1}\epsilon_J(\epsilon_J + 2\|\bar{J}_k\|)\|\bar{J}_k\|).$$

Similarly, under the assumptions of the lemma, one can establish uniform upper bounds on the remaining terms on the right-hand side of (41) in terms of ϵ_J, λ, σ_Γ, $J_{I,\sup}$, and $s_{\sup}$ in such a manner that they all vanish as $\epsilon_J \searrow 0$.

Let us now turn to the next term on the right-hand side in (40). By Lemma 4.7, (12), and the assumptions of the lemma, one finds that

$$\|\Gamma_k^{-1}g_k + \Gamma_k^{-1}W_k v_k\| \leq \sigma_\Gamma^{-1}(\|\bar{g}_k\| + \epsilon_g + \kappa_W\omega(\|\bar{J}_k\| + \epsilon_J)(\|\bar{c}_k\| + \epsilon_c)).$$

Consequently, one has a uniform upper bound on this expression in terms of the error bounds $(\epsilon_c, \epsilon_J, \epsilon_g)$ that vanishes as $(\epsilon_c, \epsilon_J, \epsilon_g) \searrow (0, 0, 0)$.

Next, with respect to $\|\bar{P}_k\|$, one finds that

$$\|\bar{P}_k\| = \|I - \bar{\Gamma}_k^{-1}\bar{J}_k^T(\bar{J}_k\bar{\Gamma}_k^{-1}\bar{J}_k^T)^{-1}\bar{J}_k\| \leq 1 + \|\bar{\Gamma}_k^{-1}\bar{J}_k^T\|\|(\bar{J}_k\bar{\Gamma}_k^{-1}\bar{J}_k^T)^{-1}\bar{J}_k\|.$$

Following a similar approach as for $\|P_k - \bar{P}_k\|$, one can derive a uniform upper bound for this expression that vanishes as $\epsilon_J \searrow 0$.

Finally, with respect to the last term on the right-hand side in (40), one has along with (12) (i.e., $\|v_k\| \leq \omega\|J_k^T c_k\|$), Lemma 4.7, and Lemma 4.8 that

$$\|\Gamma_k^{-1}(g_k + W_k v_k) - \bar{\Gamma}_k^{-1}(\bar{g}_k + W_k\bar{v}_k)\|$$

$$= \|\Gamma_k^{-1}(g_k + W_k v_k) - \bar{\Gamma}_k^{-1}(g_k + W_k v_k)$$

$$\quad + \bar{\Gamma}_k^{-1}(g_k + W_k v_k) - \bar{\Gamma}_k^{-1}(\bar{g}_k + W_k\bar{v}_k)\|$$

$$\leq \|\Gamma_k^{-1} - \bar{\Gamma}_k^{-1}\|\|g_k + W_k v_k\| + \|\bar{\Gamma}_k^{-1}\|\|(g_k - \bar{g}_k) + W_k(v_k - \bar{v}_k)\|$$

$$\leq \sigma_\Gamma^{-1}(\lambda\sigma_\Gamma^{-1}\epsilon_J(\epsilon_J + 2\|\bar{J}_k\|)(\|\bar{g}_k\| + \epsilon_g + \kappa_W\omega(\|\bar{J}_k\| + \epsilon_J)(\|\bar{c}_k\| + \epsilon_c))$$

$$\quad + \epsilon_g + \kappa_W\epsilon_v).$$

Consequently, under the assumptions of the lemma, one can establish a uniform upper bound for this term in terms of ϵ_J, λ, σ_J, $J_{I,\sup}$, $s_{\sup}$, $g_{0,\sup}$, μ, ϵ_g, κ_W, ω, $c_{I,\sup}$, ϵ_c, and ϵ_v in such a manner that the upper bound vanishes as $(\epsilon_c, \epsilon_J, \epsilon_g) \searrow (0, 0, 0)$. Combined with prior results, the proof is complete. $\square$

The following corollary now follows with $d_k = u_k + v_k$ and $\bar{d}_k = \bar{u}_k + \bar{v}_k$.

Corollary 4.1. *Suppose that Assumptions 2.1, 4.1, 4.2, 4.3, and 4.4 hold. Then, there exists $\epsilon_d \in [0, \infty)$ such that for all $k \in \mathbb{N}$ with $k \geq k_J$ one has*

$$\|d_k - \bar{d}_k\| \leq \epsilon_d. \tag{42}$$

In addition, (42) holds with $\epsilon_d \to 0$ as $(\epsilon_c, \epsilon_J, \epsilon_g) \searrow (0, 0, 0)$.

Proof. The proof follows from Lemmas 4.5 and 4.9 with $\epsilon_d := \epsilon_v + \epsilon_u$. □

Observe that we have shown upper bounds for the quantities in (27), as desired. For the next phase of our analysis, we focus on properties of the computed search directions, where in particular we are interested in showing that they yield positive reductions in the model of the merit function, which in turn are reductions that (generally) vanish as the algorithm progresses. As a brief aside, we present the following lemma, which shows that in terms of noiseless quantities the model reduction is a valid stationarity measure. This lemma provides further detail to our claim in (22) that reducing the noiseless model reduction corresponds to improvement toward stationarity of the barrier subproblem (namely, (7)) that the algorithm aims to solve.

Lemma 4.10. *Suppose that Assumptions 2.1 and 4.1 hold. Then, for all $k \in \mathbb{N}$ such that $\bar{\tau}_{k-1} > 0$, one has that either*

(a) $\bar{J}_k^T \bar{c}_k = 0$ and $\bar{c}_I(x_k) \nleq 0$;

(b) $\bar{J}_k^T \bar{c}_k = 0$, $\bar{c}_I(x_k) \leq 0$, and $\bar{g}_k + \bar{J}_k^T \bar{y}_{k+1} = 0$; or

(c) $\bar{\tau}_k > 0$ and $\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k) \geq \frac{1}{2} \bar{\tau}_k \bar{u}_k^T W_k \bar{u}_k + \sigma(\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|) > 0$.

Thus, if x_k is not an infeasible stationary point, then $\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k) = 0$ if and only if (x_k, y_{k+1}) satisfies the first-order conditions for (7), i.e., (8).

Proof. Consider arbitrary $k \in \mathbb{N}$ such that $\bar{\tau}_{k-1} > 0$. If (a) or (b) holds, then there is nothing left to prove. Thus, one may proceed under the assumption that $\bar{J}_k^T \bar{c}_k \neq 0$ and/or $\bar{g}_k + \bar{J}_k^T \bar{y}_{k+1} \neq 0$. If $\bar{J}_k^T \bar{c}_k \neq 0$, then it follows from well-known properites of (12) (e.g., see [22]) that $\bar{v}_k \neq 0$ and $\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\| > 0$, which in turn shows by (15), (16), and (17) that (c) holds. On the other hand, if $\bar{J}_k^T \bar{c}_k = 0$, but $\bar{g}_k + \bar{J}_k^T \bar{y}_{k+1} \neq 0$, then from (12) and (14) one has $\bar{v}_k = 0$ and

$$W_k \bar{u}_k + \bar{J}_k^T \bar{y}_{k+1} = -\bar{g}_k \implies \bar{u}_k \neq 0 \text{ and } 0 < \bar{u}_k^T W_k \bar{u}_k = -\bar{g}_k^T \bar{u}_k.$$

Thus, $\bar{\tau}_k \leftarrow \bar{\tau}_{k-1} > 0$ and from (15)–(16) the conditions of part (c) hold. □

Following a similar line of argument as in the proof of Lemma 4.10, but with respect to the noisy quantities, the following result can be established due to our previous assumption that the algorithm does not terminate finitely.

Lemma 4.11. *Suppose that Assumptions 2.1 and 4.1 hold. Then, for all $k \in \mathbb{N}$, one has that $\tau_k > 0$ and (d_k, τ_k) together satisfy (16).*

Proof. Since the algorithm does not terminate, for all $k \in \mathbb{N}$ either $J_k^T c_k \neq 0$ or $g_k + J_k^T y_{k+1} \neq 0$. Thus, the proof follows like that of Lemma 4.10. □

The aim of our next three results is to establish a bound on the difference between $\Delta m_k(d_k, \tau_k)$ and $\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k)$ for all $k \in \mathbb{N}$. Ultimately, the purpose of this bound is to show that, as the algorithm is designed to drive $\Delta m_k(d_k, \tau_k)$ to small values, this in turn means that $\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k)$ is driven to be small as well. We establish the bound in two steps, first by proving a bound on the difference between Δm_k and $\Delta \bar{m}_k$ with respect to (d_k, τ_k), then by proving a bound on the difference between $\Delta \bar{m}_k$ with respect to (d_k, τ_k) and $(\bar{d}_k, \bar{\tau}_k)$.

Lemma 4.12. *Suppose that Assumptions 2.1 and 4.1 hold. Then, for all $k \in \mathbb{N}$, one has that $\Delta m_k(d_k, \tau_k) = \Delta \bar{m}_k(d_k, \tau_k) + E_k^{m,1}$, where*

$$
\begin{aligned}
E_k^{m,1} = \\
& - (\bar{\tau}_k + \varepsilon_{\tau,k}) \varepsilon_{g,k}^T (\bar{d}_k + \varepsilon_{d,k}) + \|\bar{c}_k + \varepsilon_{c,k}\| - \|\bar{c}_k\| \\
& - \left(\|\bar{c}_k + \bar{J}_k \bar{d}_k + \bar{J}_k \varepsilon_{d,k} + \varepsilon_{c,k} + \varepsilon_{J,k} \bar{d}_k + \varepsilon_{J,k} \varepsilon_{d,k}\| - \|\bar{c}_k + \bar{J}_k \bar{d}_k + \bar{J}_k \varepsilon_{d,k}\| \right).
\end{aligned}
$$

Moreover, supposing that Assumptions 4.2, 4.3, and 4.4 also hold, then for all $k \in \mathbb{N}$ with $k \geq k_J$ and $\epsilon_d \in [0, \infty)$ defined in Corollary 4.1, one has

$$
|E_k^{m,1}| \leq (\bar{\tau}_k + \varepsilon_{\tau,k}) \epsilon_g \epsilon_d + 2\epsilon_c + \epsilon_J \epsilon_d + ((\bar{\tau}_k + \varepsilon_{\tau,k}) \epsilon_g + \epsilon_J) \|\bar{d}_k\| := \mathcal{E}_k^{m,1}.
$$

Proof. For any $k \in \mathbb{N}$, one finds from (11), (15), and (27) that

$$
\begin{aligned}
& \Delta m_k(d_k, \tau_k) - \Delta \bar{m}_k(d_k, \tau_k) \\
= \; & - \tau_k \varepsilon_{g,k}^T d_k + \|c_k\| - \|\bar{c}_k\| - (\|c_k + J_k d_k\| - \|\bar{c}_k + \bar{J}_k d_k\|) \\
= \; & - (\bar{\tau}_k + \varepsilon_{\tau,k}) \varepsilon_{g,k}^T (\bar{d}_k + \varepsilon_{d,k}) + \|\bar{c}_k + \varepsilon_{c,k}\| - \|\bar{c}_k\| \\
& - (\|\bar{c}_k + \varepsilon_{c,k} + \bar{J}_k \bar{d}_k + \bar{J}_k \varepsilon_{d,k} + \varepsilon_{J,k} \bar{d}_k + \varepsilon_{J,k} \varepsilon_{d,k}\| - \|\bar{c}_k + \bar{J}_k \bar{d}_k + \bar{J}_k \varepsilon_{d,k}\|),
\end{aligned}
$$

as claimed. Now by the additional assumptions of the lemma, Corollary 4.1, and the triangle inequality (applied twice), one finds that

$$
\begin{aligned}
|E_k^{m,1}| \leq \; & (\bar{\tau}_k + \varepsilon_{\tau,k}) |\varepsilon_{g,k}^T \varepsilon_{d,k}| + \|\varepsilon_{c,k}\| \\
& + \|\varepsilon_{c,k} + \varepsilon_{J,k} \varepsilon_{d,k}\| + ((\bar{\tau}_k + \varepsilon_{\tau,k}) \epsilon_g + \epsilon_J) \|\bar{d}_k\| \\
\leq \; & (\bar{\tau}_k + \varepsilon_{\tau,k}) \epsilon_g \epsilon_d + 2\epsilon_c + \epsilon_J \epsilon_d + ((\bar{\tau}_k + \varepsilon_{\tau,k}) \epsilon_g + \epsilon_J) \|\bar{d}_k\|,
\end{aligned}
$$

which completes the proof. $\qquad \square$

Our next lemma provides the aforementioned complementary result.

Lemma 4.13. *Suppose that Assumptions 2.1 and 4.1 hold. Then, for all $k \in \mathbb{N}$, one has that $\Delta \bar{m}_k(d_k, \tau_k) = \Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k) + E_k^{m,2}$, where*

$$
E_k^{m,2} = -\bar{\tau}_k \bar{g}_k^T \varepsilon_{d,k} - \varepsilon_{\tau,k} \bar{g}_k^T \bar{d}_k - \varepsilon_{\tau,k} \bar{g}_k^T \varepsilon_{d,k} + \|\bar{c}_k + \bar{J}_k \bar{d}_k\| - \|\bar{c}_k + \bar{J}_k \bar{d}_k + \bar{J}_k \varepsilon_{d,k}\|.
$$

Moreover, supposing that Assumptions 4.2, 4.3, and 4.4 also hold, then for all $k \in \mathbb{N}$ with $k \geq k_J$ and $\epsilon_d \in [0, \infty)$ defined in Corollary 4.1, one has

$$
|E_k^{m,2}| \leq (\bar{\tau}_k + \varepsilon_{\tau,k}) \|\bar{g}_k\| \epsilon_d + |\varepsilon_{\tau,k}| \|\bar{g}_k\| \|\bar{d}_k\| + \|\bar{J}_k\| \epsilon_d =: \mathcal{E}_k^{m,2}.
$$

Proof. For any $k \in \mathbb{N}$, one finds from (11), (15), and (27) that

$$\Delta \bar{m}_k(d_k, \tau_k)$$
$$= -\tau_k \bar{g}_k^T d_k + \|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k d_k\|$$
$$= -\bar{\tau}_k \bar{g}_k^T \bar{d}_k + \|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{d}_k\|$$
$$\quad - \bar{\tau}_k \bar{g}_k^T \varepsilon_{d,k} - \varepsilon_{\tau,k} \bar{g}_k^T \bar{d}_k - \varepsilon_{\tau,k} \bar{g}_k^T \varepsilon_{d,k} + \|\bar{c}_k + \bar{J}_k \bar{d}_k\| - \|\bar{c}_k + \bar{J}_k \bar{d}_k + \bar{J}_k \varepsilon_{d,k}\|$$
$$= \Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k) + E_k^{m,2},$$

as claimed. Now by the additional assumptions of the lemma, Corollary 4.1, and the triangle inequality, one finds that

$$|E_k^{m,2}| \leq (\bar{\tau}_k + \varepsilon_{\tau,k})|\bar{g}_k^T \varepsilon_{d,k}| + |\varepsilon_{\tau,k}|\|\bar{g}_k\|\|\bar{d}_k\| + \|\bar{J}_k \varepsilon_{d,k}\|$$
$$\leq (\bar{\tau}_k + \varepsilon_{\tau,k})\|\bar{g}_k\|\epsilon_d + |\varepsilon_{\tau,k}|\|\bar{g}_k\|\|\bar{d}_k\| + \|\bar{J}_k\|\epsilon_d,$$

which completes the proof. $\qquad \square$

Remark 4.1. *If, for some $k \in \mathbb{N}$, one finds $\varepsilon_{f,k} = \|\varepsilon_{c,k}\| = \|\varepsilon_{g,k}\| = \|\varepsilon_{J,k}\| = 0$, then $E_k^{m,1} = 0$. However, these errors being zero does not imply that $\varepsilon_{\tau,k}$ is zero since one might find $\varepsilon_{\tau,k} \neq 0$ due to updates of the merit parameter in prior iterations. Thus, one finds that noise in any iteration can cause a nonzero difference between $\Delta \bar{m}_k(d_k, \tau_k)$ and $\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k)$ for any subsequent index k.*

The following corollary follows easily from the previous two lemmas. Thus, we state the corollary without a proof, since it is straightforward.

Corollary 4.2. *Suppose that Assumptions 2.1 and 4.1 hold. Then, for all $k \in \mathbb{N}$, one has that $\Delta m_k(d_k, \tau_k) = \Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k) + E_k^m$, where $E_k^m := E_k^{m,1} + E_k^{m,2}$ with $E_k^{m,1}$ and $E_k^{m,2}$ defined in Lemma 4.12 and Lemma 4.13, respectively. Moreover, supposing that Assumptions 4.2, 4.3, and 4.4 also hold, then for all $k \in \mathbb{N}$ with $k \geq k_J$ and $\epsilon_d \in [0, \infty)$ defined in Corollary 4.1, one has*

$$|E_k^m| \leq \mathcal{E}_k^m := \mathcal{E}_k^{m,1} + \mathcal{E}_k^{m,2} \tag{43}$$

where $\mathcal{E}_k^{m,1}$ and $\mathcal{E}_k^{m,2}$ are defined in Lemmas 4.12 and 4.13, respectively.

We have by now established bounds on differences between the noisy and noiseless model reductions. For our next two lemmas, we explore the effect of steps on changes in the *noiseless* merit function values. By construction of the algorithm, it follows that each step yields an upper bound on the change in the *noisy* merit function value given by (20). However, to establish that the algorithm also ensures certain progress with respect to the noiseless quantities, the next two lemmas provide bounds with respect to the noiseless merit function $\bar{\phi}$. The proof of our next lemma is similar to that of [15, Lemma 3.5].

Lemma 4.14. *Suppose that Assumptions 2.1, 4.1, and 4.2 hold. Then, for all $k \in \mathbb{N}$, one finds for all $\alpha \in (0, 1]$ with $\alpha d_k^s \geq -\eta_s$ that*

$$\bar{\phi}(z_k + \alpha \hat{d}_k, \tau_k) - \bar{\phi}(z_k, \tau_k) \leq -\alpha \Delta \bar{m}_k(d_k, \tau_k) + \xi_2(\tau_k)\alpha^2\|d_k\|^2,$$

where $\xi_2(\tau_k) := \max\{\frac{1}{2}(\tau_k L_g + L_J), \frac{\tau_k \mu}{1 - \eta_s}\}$.

Proof. For any real numbers ξ and ξ' with $\xi > 0$ and $\xi' \geq -\eta_s\xi$, one has

$$\left|\log(\xi + \xi') - \log\xi - \frac{\xi'}{\xi}\right| \leq \sup_{\xi'' \in [\xi,\xi+\xi']}\left|\frac{\xi'}{\xi''} - \frac{\xi'}{\xi}\right| \leq \frac{1}{1-\eta_s}\left(\frac{\xi'}{\xi}\right)^2.$$

Thus, by Taylor's theorem and Assumptions 2.1, 4.1, and 4.2, for any $k \in \mathbb{N}$

$$\begin{aligned}
&\bar{\phi}(z_k + \alpha\hat{d}_k, \tau_k) - \bar{\phi}(z_k, \tau_k) \\
&= \tau_k\left(\bar{f}_0(x_k + \alpha d_k^x) - \bar{f}_0(x_k) - \mu\sum_{i=1}^{q}\log((s_k + \alpha S_k d_k^s)^{(i)}) + \mu\sum_{i=1}^{q}\log(s_k^{(i)})\right) \\
&\quad + \|\bar{c}_I(x_k + \alpha d_k^x) + (s_k + \alpha S_k d_k^s)\| - \|\bar{c}_I(x_k) + s_k\| \\
&\leq \tau_k\alpha\bar{g}_0(x_k)^T d_k^x + \tfrac{1}{2}\tau_k\alpha^2 L_g\|d_k^x\|^2 - \tau_k\alpha\mu e^T d_k^s + \frac{\tau_k\mu}{1-\eta_s}\alpha^2\|d_k^s\|^2 \\
&\quad + \|\bar{c}_I(x_k) + \alpha\nabla\bar{c}_I(x_k)^T d_k^x + (s_k + \alpha S_k d_k^s)\| - \|\bar{c}_I(x_k) + s_k\| + \tfrac{1}{2}\alpha^2 L_J\|d_k^x\|^2 \\
&= \tau_k\alpha\bar{g}_k^T d_k + \tfrac{1}{2}\alpha^2(\tau_k L_g + L_J)\|d_k^x\|^2 + \frac{\tau_k\mu}{1-\eta_s}\alpha^2\|d_k^s\|^2 \\
&\quad + \|\bar{c}(z_k) + \alpha\bar{J}_k d_k\| - \|\bar{c}(z_k)\| \\
&\leq -\alpha(-\tau_k\bar{g}_k^T d_k + \|\bar{c}(z_k)\| - \|\bar{c}(z_k) + \bar{J}_k d_k\|) + \xi_2(\tau_k)\alpha^2\|d_k\|^2 \\
&= -\alpha\Delta\bar{m}_k(d_k, \tau_k) + \xi_2(\tau_k)\alpha^2\|d_k\|^2,
\end{aligned}$$

as desired. $\qquad\square$

Our next lemma extends the result of Lemma 4.14 by expressing the right-hand side of the inequality in terms of the stationarity measure $\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k)$.

Lemma 4.15. *Suppose that Assumptions 2.1, 4.1, and 4.2 hold. Then, for all $k \in \mathbb{N}$, one finds for all $\alpha \in (0,1]$ with $\alpha d_k^s \geq -\eta_s$ that*

$$\bar{\phi}(z_k + \alpha\hat{d}_k, \tau_k) - \bar{\phi}(z_k, \tau_k) \leq -\alpha\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) - \alpha E_k^{m,2} + \xi_2(\tau_k)\alpha^2\|\bar{d}_k + \varepsilon_{d,k}\|^2,$$

where $E_k^{m,2}$ and $\xi_2(\tau_k)$ are defined as in Lemmas 4.13 and 4.14, respectively.

Proof. The result follows directly from (27) along with Lemmas 4.13 and 4.14. $\qquad\square$

Following the result of Lemma 4.15, our next aim is to establish, for all $k \in \mathbb{N}$, a relationship between the noiseless model reduction $\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k)$ and the norm of the noiseless search direction $\bar{d}_k$. This relationship is established by tying both quantities to a combination of $\|\bar{u}_k\|^2$ and $\|\bar{J}_k^T \bar{c}_k\|^2$.

Lemma 4.16. *Suppose that Assumptions 2.1, 4.1, and 4.2 hold. Then, there exist $\xi_3 \in (0,\infty)$ and $\xi_4 \in (0,\infty)$ such that, for all $k \in \mathbb{N}$, one has*

$$\begin{aligned}
\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) &\geq \xi_3(\bar{\tau}_k\|\bar{u}_k\|^2 + \|\bar{J}_k^T \bar{c}_k\|^2) \\
and \quad \|\bar{d}_k\|^2 &\leq \xi_4(\|\bar{u}_k\|^2 + \|\bar{J}_k^T \bar{c}_k\|^2).
\end{aligned}$$

Proof. Consider arbitrary $k \in \mathbb{N}$. Following a similar argument as in Lemma 4.10, one has $\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) \geq \tfrac{1}{2}\bar{\tau}_k\bar{u}_k^T W_k\bar{u}_k + \sigma(\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k\bar{v}_k\|)$, regardless of whether x_k is an infeasible

stationary point or a first-order stationary point for (7). Hence, along with Lemma 4.4, one has under the given conditions that

$$\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k) \geq \tfrac{1}{2}\bar{\tau}_k \bar{u}_k^T W_k \bar{u}_k + \sigma(\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|)$$

$$\geq \tfrac{1}{2}\bar{\tau}_k \sigma_W \|\bar{u}_k\|^2 + \sigma \xi_{1,k} \frac{\|\bar{J}_k^T \bar{c}_k\|^2}{\|\bar{c}_k\|}.$$

The existence of ξ_3, as claimed, now follows since $\|\bar{c}_k\|$ is bounded under the conditions of the lemma. On the other hand, by the triangle inequality and the definition of the normal subproblem in (12), one finds that

$$\|\bar{d}_k\|^2 \leq 2(\|\bar{u}_k\|^2 + \|\bar{v}_k\|^2) \leq 2(\|\bar{u}_k\|^2 + \omega^2 \|\bar{J}_k^T \bar{c}_k\|^2),$$

from which the desired conclusion holds with $\xi_4 := 2\max\{1, \omega^2\}$. $\qquad\square$

Following Lemma 4.16, we are able to prove a stronger relationship between $\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k)$ and $\|\bar{d}_k\|^2$ if there exist a positive lower bound for the elements of the noiseless merit parameter sequence $\{\bar{\tau}_k\}$. This may occur more generally, but let us at least establish that this is guaranteed to occur under Assumptions 2.1, 4.1, 4.2, and 4.3. We do this through our next few lemmas.

Lemma 4.17. *Suppose that Assumptions 2.1, 4.1, and 4.2 hold. Then, for all $k \in \mathbb{N}$, one has $\|\bar{v}_k\| \leq \omega J_{\sup}(c_{I,\sup} + s_{\sup})$. Moreover, supposing Assumption 4.3 also holds, there exists $\xi_5 \in (0, \infty)$ such that for all $k \in \mathbb{N}$ with $k \geq k_J$ one finds*

$$\|\bar{v}_k\| \leq \xi_5(\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|).$$

Proof. Under the initially stated conditions, the first desired conclusion follows from (12) and (23). Now suppose Assumption 4.3 also holds and consider arbitrary $k \in \mathbb{N}$ with $k \geq k_J$. Then, $\|\bar{J}_k^T \bar{c}_k\| \geq \gamma \|\bar{c}_k\|$. Hence, by Lemma 4.4,

$$\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\| \geq \xi_{1,k} \frac{\|\bar{J}_k^T \bar{c}_k\|^2}{\|\bar{c}_k\|} \geq \xi_{1,k} \gamma^2 \|\bar{c}_k\|.$$

Thus, by the conditions of the lemma and (12), there exists $\xi_5 \in (0, \infty)$ with

$$\|\bar{v}_k\| \leq \omega \|\bar{J}_k^T\| \|\bar{c}_k\| \leq \xi_{1,k}^{-1} \gamma^{-2} \omega \|\bar{J}_k^T\| (\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|)$$

$$\leq \xi_5(\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|),$$

which completes the proof. $\qquad\square$

Lemma 4.18. *Suppose that Assumptions 2.1, 4.1, and 4.2 hold. Then, for all $k \in \mathbb{N}$, one has that $\|\bar{u}_k\| \leq \sigma_W^{-1}(g_{\sup} + \kappa_W \omega J_{\sup}(c_{I,\sup} + s_{\sup}))$.*

Proof. Consider arbitrary $k \in \mathbb{N}$. Let $\bar{Z}_k \in \mathbb{R}^{(n+q)\times n}$ be a matrix whose columns form an orthonormal basis for null$(\bar{J}_k)$. As in the proof of Lemma 4.9, one has $\bar{u}_k = -\bar{Z}_k(\bar{Z}_k^T \bar{\Gamma}_k \bar{Z}_k)^{-1} \bar{Z}_k^T (\bar{g}_k + W_k \bar{v}_k)$. Thus, under the conditions of the lemma, one finds with $\|\bar{Z}_k\| \leq 1$ and Lemmas 4.7 and 4.17 that

$$\|\bar{u}_k\| \leq \|\bar{Z}_k(\bar{Z}_k^T \Gamma_k \bar{Z}_k)^{-1} \bar{Z}_k^T\| \|\bar{g}_k + W_k \bar{v}_k\|$$

$$= \|\bar{Z}_k(\bar{Z}_k^T W_k \bar{Z}_k)^{-1} \bar{Z}_k^T\| \|\bar{g}_k + W_k \bar{v}_k\|$$

$$\leq \sigma_W^{-1}(g_{\sup} + \kappa_W \omega J_{\sup}(c_{I,\sup} + s_{\sup})),$$

which completes the proof. $\qquad\square$

We now show the aforementioned fact that the noiseless merit parameter values can be guaranteed to be bounded below by a positive real number. The proof of the following lemma is similar to that of Lemma 3.16 in [11].

Lemma 4.19. *Suppose that Assumptions 2.1, 4.1, 4.2, and 4.3 hold. Then, there exist $k_{\bar{\tau}} \in \mathbb{N}$ and $\bar{\tau} \in (0, \infty)$ such that $\bar{\tau}_k = \bar{\tau}$ for all $k \in \mathbb{N}$ with $k \geq k_{\bar{\tau}}$.*

Proof. By construction, $\{\bar{\tau}_k\}$ is monotonically nonincreasing. Thus, by (18), one can reach the desired conclusion by showing that the corresponding sequence of trial merit parameter values is bounded below. Consider arbitrary $k \in \mathbb{N}$ with $k \geq k_J$. By (14), it follows that $W_k \bar{d}_k + \bar{g}_k + \bar{J}_k^T \bar{y}_{k+1} = 0$. This, along with $\bar{u}_k \in \text{null}(\bar{J}_k)$, yields $\bar{u}_k^T W_k \bar{u}_k + \bar{u}_k^T W_k \bar{v}_k + \bar{u}_k^T \bar{g}_k = 0$. Thus, along with Lemmas 4.17 and 4.18, there exists $\xi_6 \in (0, \infty)$ such that

$$
\begin{aligned}
\bar{g}_k^T \bar{d}_k + \tfrac{1}{2} \bar{u}_k^T W_k \bar{u}_k = \bar{g}_k^T \bar{u}_k + \bar{g}_k^T \bar{v}_k + \tfrac{1}{2} \bar{u}_k^T W_k \bar{u}_k & \\
&\leq -\bar{u}_k^T W_k \bar{v}_k + \bar{g}_k^T \bar{v}_k - \tfrac{1}{2} \bar{u}_k^T W_k \bar{u}_k \\
&\leq -\tfrac{1}{2} \sigma_W \|\bar{u}_k\|^2 + \|\bar{g}_k\| \|\bar{v}_k\| - \bar{u}_k^T W_k \bar{v}_k \\
&\leq \|\bar{g}_k\| \|\bar{v}_k\| + \|\bar{u}_k\| \|W_k\| \|\bar{v}_k\| \\
&\leq \xi_6 \|\bar{v}_k\| \\
&\leq \xi_6 \xi_5 (\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|).
\end{aligned}
$$

Hence, $\frac{(1-\sigma)(\|\bar{c}_k\| - \|\bar{c}_k + \bar{J}_k \bar{v}_k\|)}{\bar{g}_k^T \bar{d}_k + \tfrac{1}{2} \bar{u}_k^T W_k \bar{u}_k} \geq \frac{1-\sigma}{\xi_6 \xi_5}$, which, as stated, completes the proof. $\qquad\square$

Our next lemma now builds upon the result of Lemma 4.16.

Lemma 4.20. *Suppose that Assumptions 2.1, 4.1, 4.2, and 4.3 hold. Then, there exists $\xi_d \in (0, \infty)$ such that, for all $k \in \mathbb{N}$, one finds*

$$
\xi_d \|\bar{d}_k\|^2 \leq \Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k).
$$

Proof. Consider arbitrary $k \in \mathbb{N}$. By Lemmas 4.16 and 4.19, one finds that

$$
\begin{aligned}
\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k) &\geq \xi_3 (\bar{\tau} \|\bar{u}_k\|^2 + \|\bar{J}_k^T \bar{c}_k\|^2) \\
&\geq \xi_3 \min\{\bar{\tau}, 1\} (\|\bar{u}_k\|^2 + \|\bar{J}_k^T \bar{c}_k\|^2) \geq \xi_d \|\bar{d}_k\|^2,
\end{aligned}
$$

where $\xi_d := \xi_3 \xi_4^{-1} \min\{\bar{\tau}, 1\}$, which completes the proof. $\qquad\square$

We are almost prepared to present our main theorem about the behavior of our proposed algorithm. The theorem considers two situations: one when only Assumptions 2.1, 4.1, and 4.2 hold, and another when Assumptions 4.3 and 4.4 also hold. For the latter situation, we provide the following corollary.

Corollary 4.3. *Suppose Assumptions 2.1, 4.1, 4.2, 4.3, and 4.4 hold. Then, for all $k \in \mathbb{N}$,*

$$
\mathcal{E}_k^{m,1} \leq e_1^{m,1} + e_2^{m,1} \|\bar{d}_k\| \quad \text{and} \quad \mathcal{E}_k^{m,2} \leq c_1^{m,2} + c_2^{m,2} \|\bar{d}_k\|, \tag{44}
$$

where $\mathcal{E}_k^{m,1}$ and $\mathcal{E}_k^{m,2}$ are defined in Lemmas 4.12 and 4.13, respectively, and

$$
\begin{aligned}
e_1^{m,1} &:= (\tau_0 \epsilon_g + \epsilon_J) \epsilon_d + 2\epsilon_c, \quad e_2^{m,1} := \tau_0 \epsilon_g + \epsilon_J, \\
c_1^{m,2} &:= (\tau_0 g_{\text{sup}} + J_{\text{sup}}) \epsilon_d, \quad \text{and} \quad c_2^{m,2} := \tau_0 g_{\text{sup}},
\end{aligned}
$$

with g_{sup} and J_{sup} defined in (23).

Proof. Under the stated assumptions, the desired conclusions follow from Lemmas 4.12 and 4.13, (23), and the facts that $\bar{\tau}_k + \varepsilon_{\tau,k} = \tau_k \leq \tau_0$ and $|\varepsilon_{\tau,k}| \leq \tau_0$. □

Our last lemma prior to our main theorem, stated next, shows situations in which the relaxed line search condition is guaranteed to hold.

Lemma 4.21. *Suppose that Assumptions 2.1, 4.1, and 4.2 hold, and for all $k \in \mathbb{N}$ with $\alpha_k^{\max}$ computed to satisfy the fraction-to-the-boundary rule, $\varepsilon_{d,k}$ defined in (27), and $\xi_2(\tau_k)$ defined in Lemma 4.14, let*

$$\hat{\alpha}_k = \min\left\{ \frac{(1-\eta_\phi)\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k)}{2\xi_2(\tau_k)\|\bar{d}_k + \varepsilon_{d,k}\|^2}, \alpha_k^{\max} \right\}. \tag{45}$$

Then, with $(E_k^{m,1}, E_k^{m,2})$ from Lemmas 4.12–4.13, one has for any $k \in \mathbb{N}$ that

$$\eta_\phi E_k^{m,1} - (1-\eta_\phi)E_k^{m,2} \leq \tfrac{1}{2}(1-\eta_\phi)\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) \tag{46}$$

implies that (20) holds for all $\alpha \in [0, \hat{\alpha}_k]$. Now suppose that, in addition, Assumptions 4.3 and 4.4 hold, and let $k_{\bar{\tau}}$ be defined in Lemma 4.19, ξ_d be defined in Lemma 4.20, $(e_1^{m,1}, e_2^{m,1}, c_1^{m,2}, c_2^{m,2})$ be defined in Corollary 4.3, and

$$\bar{\alpha}_k := \min\left\{ \frac{(1-\eta_\phi)\xi_d\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k)}{4\xi_2(\tau_k)(\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) + \xi_d\epsilon_d^2)}, \alpha_k^{\max} \right\} \quad \text{for all} \ \ k \geq k_{\bar{\tau}}. \tag{47}$$

Then, one has for any $k \geq k_{\bar{\tau}}$ that

$$(\eta_\phi e_1^{m,1} + (1-\eta_\phi)c_1^{m,2}) + \xi_d^{-1/2}(\eta_\phi e_2^{m,1} + (1-\eta_\phi)c_2^{m,2})\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k)^{1/2}$$
$$\leq \tfrac{1}{2}(1-\eta_\phi)\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) \tag{48}$$

implies that (20) holds for all $\alpha \in [0, \bar{\alpha}_k]$.

Proof. Consider arbitrary $k \in \mathbb{N}$. Our first aim is to show that if (46) holds, then (20) holds for all $\alpha \in [0, \hat{\alpha}_k]$. Toward this end, observe that by (10) and Lemma 4.15, one finds for all $\alpha \in (0, \alpha_k^{\max}]$ that

$$
\begin{aligned}
&\phi(z_k + \alpha\hat{d}_k, \tau_k) - \phi(z_k, \tau_k) \\
&\leq \bar{\phi}(z_k + \alpha\hat{d}_k, \tau_k) - \bar{\phi}(z_k, \tau_k) + 2\varepsilon_k \\
&\leq -\alpha\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) - \alpha E_k^{m,2} + \alpha^2\xi_2(\tau_k)\|\bar{d}_k + \varepsilon_{d,k}\|^2 + 2\varepsilon_k \\
&= -\alpha\eta_\phi\Delta m_k(d_k, \tau_k) + 2\varepsilon_k - \alpha(1-\eta_\phi)\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) \\
&\quad + \alpha\eta_\phi E_k^{m,1} - \alpha(1-\eta_\phi)E_k^{m,2} + \alpha^2\xi_2(\tau_k)\|\bar{d}_k + \varepsilon_{d,k}\|^2.
\end{aligned} \tag{49}
$$

Hence, along with (45) and (46), one finds that

$$
\begin{aligned}
&\phi(z_k + \alpha\hat{d}_k, \tau_k) - \phi(z_k, \tau_k) \\
&\leq -\alpha\eta_\phi\Delta m_k(d_k, \tau_k) + 2\epsilon_k - \tfrac{1}{2}\alpha(1-\eta_\phi)\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) + \alpha^2\xi_2(\tau_k)\|\bar{d}_k + \varepsilon_{d,k}\|^2 \\
&\leq -\alpha\eta_\phi\Delta m_k(d_k, \tau_k) + 2\varepsilon_k,
\end{aligned}
$$

which implies that (20) is satisfied, as desired.

Now let us prove the second part of the lemma with the addition of Assumptions 4.3 and 4.4. Consider arbitrary $k \in \mathbb{N}$ with $k \geq k_{\bar{\tau}}$. By (49), Lemmas 4.12 and 4.13, and Corollary 4.3, one finds for all $\alpha \in (0, \alpha_k^{\max}]$ that

$$
\begin{aligned}
& \phi(z_k + \alpha \hat{d}_k, \tau_k) - \phi(z_k, \tau_k) \\
\leq\ & -\alpha \eta_\phi \Delta m_k(d_k, \tau_k) + 2\varepsilon_k - \alpha(1 - \eta_\phi)\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k) \\
& + \alpha \eta_\phi \mathcal{E}_k^{m,1} + \alpha(1 - \eta_\phi)\mathcal{E}_k^{m,2} + 2\alpha^2 \xi_2(\tau_k)(\|\bar{d}_k\|^2 + \epsilon_d^2) \\
\leq\ & -\alpha \eta_\phi \Delta m_k(d_k, \tau_k) + 2\varepsilon_k - \alpha(1 - \eta_\phi)\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k) \\
& + \alpha \eta_\phi(e_1^{m,1} + e_2^{m,1}\|\bar{d}_k\|) \\
& + \alpha(1 - \eta_\phi)(c_1^{m,2} + c_2^{m,2}\|\bar{d}_k\|) + 2\alpha^2 \xi_2(\tau_k)(\|\bar{d}_k\|^2 + \epsilon_d^2).
\end{aligned}
$$

Now by Lemmas 4.19 and Lemma 4.20, one has

$$
\begin{aligned}
& \phi(z_k + \alpha \hat{d}_k, \tau_k) - \phi(z_k, \tau_k) \\
\leq\ & -\alpha \eta_\phi \Delta m_k(d_k, \tau_k) + 2\varepsilon_k - \alpha(1 - \eta_\phi)\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k) \\
& + \alpha(\eta_\phi e_1^{m,1} + (1 - \eta_\phi)c_1^{m,2}) \\
& + \alpha \xi_d^{-1/2}(\eta_\phi e_2^{m,1} + (1 - \eta_\phi)c_2^{m,2})\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k)^{1/2} \\
& + 2\alpha^2 \xi_2(\tau_k)\xi_d^{-1}\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k) + 2\alpha^2 \xi_2(\tau_k)\epsilon_d^2.
\end{aligned}
$$

Hence, by (47) and (48), one finds that

$$
\begin{aligned}
& \phi(z_k + \alpha \hat{d}_k, \tau_k) - \phi(z_k, \tau_k) \\
\leq\ & -\alpha \eta_\phi \Delta m_k(d_k, \tau_k) + 2\varepsilon_k - \tfrac{1}{2}\alpha(1 - \eta_\phi)\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k) \\
& + 2\alpha^2 \xi_2(\tau_k)\xi_d^{-1}\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k) + 2\alpha^2 \xi_2(\tau_k)\epsilon_d^2 \\
\leq\ & -\alpha \eta_\phi \Delta m_k(d_k, \tau_k) + 2\varepsilon_k,
\end{aligned}
$$

which implies that (20) is satisfied, as desired. □

We are now ready to prove our main theorem. We following the proof of the theorem with commentary on the theorem's consequences.

Theorem 1. *Suppose Assumptions 2.1, 4.1, and 4.2 hold. Then, with $E_k^{m,1}$ defined in Lemma 4.12, $E_k^{m,2}$ defined in Lemma 4.13, E_k^m defined in Corollary 4.2, and $\hat{\alpha}_k$ defined in Lemma 4.21, there exists $k \in \mathbb{N}$ such that*

$$
\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}_k) \leq \max\left\{ \frac{2(\eta_\phi E_k^{m,1} - (1 - \eta_\phi)E_k^{m,2})}{(1 - \eta_\phi)}, -2E_k^m, \frac{4(4 + 2\zeta)\varepsilon_k}{\hat{\alpha}_k \eta_\phi} \right\}. \tag{50}
$$

Now suppose that, in addition, Assumptions 4.3 and 4.4 hold, and let $k_{\bar{\tau}}$ be defined in Lemma 4.19, ξ_d be defined in Lemma 4.20, and $(e_1^{m,1}, e_2^{m,1}, c_1^{m,2}, c_2^{m,2})$ be defined in Corollary 4.3. Then, there exists $k \in \mathbb{N}$ with $k \geq k_{\bar{\tau}}$ such that

$$
\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}) \leq \max\{\rho_{k,1}, \rho_{k,2}, \rho_{k,3}\}, \tag{51}
$$

where

$$\rho_{k,1} = 2\left(\tfrac{\eta_\phi}{1-\eta_\phi}e_1^{m,1} + c_1^{m,2} + \xi_d^{-1/2}(\tfrac{\eta_\phi}{1-\eta_\phi}e_2^{m,1} + c_2^{m,2})\Delta\bar{m}_k(\bar{d}_k,\bar{\tau})^{1/2}\right),$$

$$\rho_{k,2} = 2\left(e_1^{m,1} + c_1^{m,2} + \xi_d^{-1/2}(e_2^{m,1} + c_2^{m,2})\Delta\bar{m}_k(\bar{d}_k,\bar{\tau})^{1/2}\right),$$

$$and \ \ \rho_{k,3} = \frac{4(4+2\zeta)\varepsilon_k}{\bar{\alpha}_k\eta_\phi} = \frac{4(4+2\zeta)\varepsilon_k}{\eta_\phi}\max\left\{\frac{4\xi_2(\bar{\tau})(\Delta\bar{m}_k(\bar{d}_k,\bar{\tau})+\xi_d\epsilon_d^2)}{(1-\eta_\phi)\xi_d\Delta\bar{m}_k(\bar{d}_k,\bar{\tau})}, \frac{1}{\alpha_k^{\max}}\right\}.$$

Proof. To derive a contradiction to the first desired conclusion, suppose (50) does not hold for all $k \in \mathbb{N}$. Under this supposition, consider arbitrary $k \in \mathbb{N}$. Since (50) does not hold, it follows that (46) holds, which in turn means by Lemma 4.21 that (20) holds for some step size $\alpha \geq \hat{\alpha}_k/2$. Thus, with Corollary 4.2,

$$\begin{aligned}
&\bar{\phi}(z_k + \alpha\hat{d}_k, \tau_k) - \bar{\phi}(z_k, \tau_k) \\
&\leq \phi(z_k + \alpha\hat{d}_k, \tau_k) - \phi(z_k, \tau_k) + 2\varepsilon_k \\
&\leq -\tfrac{1}{2}\hat{\alpha}_k\eta_\phi\Delta m_k(d_k, \tau_k) + 2\varepsilon_k + (2+\zeta)\varepsilon_k \\
&= -\tfrac{1}{2}\hat{\alpha}_k\eta_\phi\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) - \tfrac{1}{2}\hat{\alpha}_k\eta_\phi E_k^m + (4+\zeta)\varepsilon_k.
\end{aligned} \tag{52}$$

This, again with the fact that (50) does not hold, implies

$$\bar{\phi}(z_k + \alpha\hat{d}_k, \tau_k) - \bar{\phi}(z_k, \tau_k) \leq -\tfrac{1}{4}\hat{\alpha}_k\eta_\phi\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) + (4+\zeta)\varepsilon_k \leq -\zeta\varepsilon_k.$$

However, with Remark 3.1 this implies that $\bar{\phi}(z_{k+1}, \tau_k) - \bar{\phi}(z_k, \tau_k) \leq -\zeta\varepsilon_k$, which contradicts Assumption 4.2, specifically the fact that $\{\bar{\phi}(z_k, \tau_k)\}$ is bounded below. Consequently, (50) must hold for some $k \in \mathbb{N}$, as desired.

Now consider the second desired conclusion under the addition of Assumptions 4.3 and 4.4. To derive a contradiction, suppose that (51) does not hold for all $k \in \mathbb{N}$ with $k \geq k_{\bar{\tau}}$. Under this supposition, consider arbitrary such k. Since (51) does not hold, it follows that (48) holds, which in turn means by Lemma 4.21 that (20) holds for some step size $\alpha \geq \bar{\alpha}_k/2$. Thus, by a similar derivation that led to (52), Corollary 4.2, Corollary 4.3, and Lemma 4.20,

$$\begin{aligned}
&\bar{\phi}(z_k + \alpha\hat{d}_k, \tau_k) - \bar{\phi}(z_k, \tau_k) \\
&\leq -\tfrac{1}{2}\bar{\alpha}_k\eta_\phi\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) + \tfrac{1}{2}\bar{\alpha}_k\eta_\phi\mathcal{E}_k^m + (4+\zeta)\varepsilon_k \\
&\leq -\tfrac{1}{2}\bar{\alpha}_k\eta_\phi\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) + (4+\zeta)\varepsilon_k \\
&\quad + \tfrac{1}{2}\bar{\alpha}_k\eta_\phi(e_1^{m,1} + c_1^{m,2} + (e_2^{m,1} + c_2^{m,2})\|\bar{d}_k\|) \\
&\leq -\tfrac{1}{2}\bar{\alpha}_k\eta_\phi\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) + (4+\zeta)\varepsilon_k \\
&\quad + \tfrac{1}{2}\bar{\alpha}_k\eta_\phi(e_1^{m,1} + c_1^{m,2} + \xi_d^{-1/2}(e_2^{m,1} + c_2^{m,2})\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k)^{1/2}).
\end{aligned}$$

This, again with the fact that (51) does not hold, implies

$$\bar{\phi}(z_k + \alpha\hat{d}_k, \tau_k) - \bar{\phi}(z_k, \tau_k) \leq -\tfrac{1}{4}\bar{\alpha}_k\eta_\phi\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) + (4+\zeta)\varepsilon_k \leq -\zeta\varepsilon_k$$

However, with Remark 3.1 this implies that $\bar{\phi}(z_{k+1}, \tau_k) - \bar{\phi}(z_k, \tau_k) < -\zeta\varepsilon_k$, which contradicts Assumption 4.2, specifically the fact that $\{\bar{\phi}(z_k, \tau_k)\}$ is bounded below. Consequently, (51) must hold for some $k \in \mathbb{N}$, as desired. $\qquad\square$

The following observations explain the strong consequences of Theorem 1.

- The first conclusion of the theorem, expressed in (50), shows a preliminary upper bound on the stationarity measure $\Delta \bar{m}(\bar{d}_k, \bar{\tau}_k)$ in terms of error quantities in the given iteration $k \in \mathbb{N}$. If $E_k^{m,1}$, $E_k^{m,2}$, and E_k^m are each relatively small and the step size threshold $\hat{\alpha}_k$ (recall (45)) is not exceedingly small, then (50) describes that the algorithm has reached an iteration in which $\Delta \bar{m}(\bar{d}_k, \bar{\tau}_k)$ is small, as is desirable; recall §4.1. One should of course note that $\hat{\alpha}_k$ itself depends on $\Delta \bar{m}(\bar{d}_k, \bar{\tau}_k)$, $\|\bar{d}_k + \varepsilon_{d,k}\|^2$, and $\alpha_k^{\max}$, and in particular $\hat{\alpha}_k$ not being exceedingly small means that $\|\bar{d}_k + \varepsilon_{d,k}\|^2$ is not exceedingly large relative to $\Delta \bar{m}(\bar{d}_k, \bar{\tau}_k)$ and that $\alpha_k^{\max}$ is not exceedingly small. It is not straightforward to guarantee that such a situation occurs without considering the overall behavior of the algorithm over the entire sequence of iterations, rather than in the particular iteration $k \in \mathbb{N}$ in which (50) holds. Therefore, let us turn to the second desired conclusion wherein we have effectively shown an iteration-independent bound on the threshold reached by the stationary measure $\Delta \bar{m}(\bar{d}_k, \bar{\tau}_k)$.

- In the second conclusion of the theorem under the additional Assumptions 4.3 and 4.4, the bound $\Delta \bar{m}(\bar{d}_k, \bar{\tau}_k) \leq \max\{\rho_{k,1}, \rho_{k,2}\}$ would imply that

$$\Delta \bar{m}(\bar{d}_k, \bar{\tau}_k) \leq c_1 + c_2 \Delta \bar{m}(\bar{d}_k, \bar{\tau}_k)^{1/2}$$

for some iteration-independent constants $c_1 \in (0, \infty)$ and $c_2 \in (0, \infty)$, which in turn would bound $\Delta \bar{m}(\bar{d}_k, \bar{\tau}_k)^{1/2}$ in terms of these constants. Recalling Corollary 4.3, one finds that these constants would all vanish if the noise bounds were to vanish, except for the term derived from $c_2^{m,2} := \tau_0 g_{\sup}$, which depends on the initial merit parameter value. This directs us back to Remark 4.1, where we first observed that poor behavior can be exhibited by the algorithm if the noisy and noiseless merit parameter sequences diverge substantially. Specifically, the worst case is when the noisy merit parameter values does not decrease sufficiently, in which case the algorithm does not put enough weight on reducing constraint violation. One way in which this poor behavior can be mitigated is by choosing τ_0 to a conservative, small value, although this might not always yield the best performance in real-world practice.

- In the second conclusion, the bound $\Delta \bar{m}(\bar{d}_k, \bar{\tau}_k) \leq \rho_{k,3}$ would imply that

$$\Delta \bar{m}(\bar{d}_k, \bar{\tau}_k) \leq \frac{4(4 + 2\zeta)\varepsilon_k}{\eta_\phi} \max\left\{ \frac{4\xi_2(\bar{\tau})(\Delta \bar{m}_k(\bar{d}_k, \bar{\tau}) + \xi_d \epsilon_d^2)}{(1 - \eta_\phi)\xi_d \Delta \bar{m}_k(\bar{d}_k, \bar{\tau})}, \frac{1}{\alpha_k^{\max}} \right\}.$$

Recall that by Corollary 4.1 it follows that $\|d_k - \bar{d}_k\| \leq \epsilon_d$. One can also show under our assumptions that $\{d_k\}$ is bounded in norm; see, e.g., Lemma 3.7 in [15]. Consequently, following Lemma 3.11 in [15], one can show that there exists a constant $d_{\sup}^s$ such that $(\alpha_k^{\max})^{-1} \leq \eta_s^{-1}\|d_k^s\| \leq \eta_s^{-1} d_{\sup}^s$ for all $k \in \mathbb{N}$. In fact, this is quite a pessimistic bound, since in fact one should expect $\|d_k^s\|$ to be small at points at which $\Delta \bar{m}(\bar{d}_k, \bar{\tau}_k)$ is small. In any case, overall one can essentially conclude that $\Delta \bar{m}(\bar{d}_k, \bar{\tau}_k) \leq \rho_{k,3}$ implies that $\Delta \bar{m}(\bar{d}_k, \bar{\tau}_k)$ is bounded by constants that vanish with the noise.

5 Numerical Experiments

The purpose of our numerical experiments is to demonstrate the practical performance of our proposed algorithm on a large and diverse set of test problems. Our aim is to show that, in practice, our algorithm is able to drive stationarity measures below levels that are proportional

to the noise in the function and derivative values. Similarly as for our theoretical results in the preceding section, our experiments consider runs with a fixed barrier parameter. We discuss a proposed heuristic for reducing the barrier parameter in Section 6.

We implemented our algorithm in Matlab. We chose, in the following manner, test problems from CUTE (Constrained and Unconstrained Testing Environment) [5] for which AMPL models [17] were available.[1] First, of all available CUTE/AMPL problems, we restricted attention to those involving at least one (potentially nonlinear) constraint given by CUTE/AMPL in their standard form `l <= c(x) <= u`. (Problems with only bound constraints on an individual variable or variables were ignored.) Second, of these problems, we restricted attention to those with `l < u`, componentwise. Third, we removed from consideration all problems with more than 100 primal or dual variables. This resulted in a set of 137 test problems for our numerical experiments.

The following initialization strategy and algorithmic parameters were fixed for all runs. (Values not mentioned here were chosen differently for different experiments, as discussed later on.) For the initial primal iterate x_0, we took the initial point offered by CUTE/AMPL and projected it inside any present bound constraints using the strategy proposed in Section 3.6 of [31]. The initial slack variable s_0 was set to a vector of all ones, then a slack reset was performed; recall (21). The initial merit parameter was set to $\tau_{-1} \leftarrow 10^{-1}$, the normal-step trust-region-radius parameter was set to $\omega \leftarrow 10^3$, the merit parameter update parameters were set to $\sigma \leftarrow 10^{-1}$ and $\delta_\tau \leftarrow 10^{-4}$, and the relaxed Armijo line-search parameters were set to $\eta_\phi \leftarrow 10^{-8}$ and $\zeta \leftarrow 10^{-1}$.

Important implementation details were as follows. First, as is typical for software for solving continuous optimization problems, each problem was scaled before our algorithm was applied. Specifically, if the noiseless gradient of the objective function evaluated at the initial point had an ℓ_∞ norm greater than 10, then the objective function was scaled by $10/\|\bar{g}_0\|_\infty$. Similarly, if the noiseless gradient of any constraint function evaluated at the initial point had an ℓ_∞ norm greater than 10, then the constraint function was scaled similarly. All of our subsequent numerical results refer to solving the resulting scaled problems. Second, during the run of our proposed algorithm to compute the normal step v_k for all $k \in \mathbb{N}$, our code employs an implementation of the well-known Moré-Sorensen method [20] for solving the trust-region subproblem (12). Third, for all $k \in \mathbb{N}$, to compute the tangential step u_k, we initialize W_k as in (13) with a noisy Hessian value (see next paragraph) and $\Sigma_k \leftarrow S_k Y_k$. Then, a Hessian modification subroutine is performed to ensure that it is sufficiently positive definite over null(J_k). This is done by using Matlab's built-in `ldl` method to compute a block LDL factorization of the matrix in (14), then checking its inertia [22]. If the block diagonal matrix in this factorization has n positive eigenvalues with value at least 10^{-10}, then no modification is performed. Otherwise, sequentially larger multiples of the identity matrix are added to W_k until the block diagonal matrix in its LDL factorization has n positive eigenvalues with value at least 10^{-10}. Once the Hessian modification strategy is completed, the tangential step u_k and multipliers y_{k+1} are computed using Matlab's built-in `mldivide` routine for solving linear systems of equations.

We present the results of experiments with two barrier parameter values and two noise levels. For the barrier parameter, we considered $\mu \leftarrow 10^{-1}$ and $\mu \leftarrow 10^{-4}$. In each case, the initial dual variable y_0 was set to $\mu S_0^{-1} e$ and the fraction-to-the-boundary parameter was set to $\eta_s \leftarrow \max\{0.99, 1 - \mu\}$. For the noise levels, we considered $\epsilon_f = \epsilon_c \leftarrow 10^{-2}$ and $\epsilon_f = \epsilon_c \leftarrow 10^{-6}$. In each case, we chose $\epsilon_g = \epsilon_J = \epsilon_H \leftarrow \sqrt{\epsilon_f} = \sqrt{\epsilon_c}$. Given these constants, noise was added to each CUTE/AMPL function and derivative value for each $k \in \mathbb{N}$ as follows. For the objective and constraint functions,

---

[1]https://vanderbei.princeton.edu/ampl/nlmodels/index.html

we set $f_0(x_k) \leftarrow \bar{f}_0(x_k) + \varepsilon_{f_0,k}$ and $c_I(x_k) \leftarrow \bar{c}_I(x_k) + \varepsilon_{c_I,k}$, where $\varepsilon_{f_0,k}$ was drawn from a uniform distribution over $[-\epsilon_f, \epsilon_f]$ and $\varepsilon_{c_I,k}$ was drawn from a uniform distribution over an ℓ_2-norm ball with radius ϵ_c. For the objective and constraint first-order derivatives, we set $g_0(x_k) \leftarrow \bar{g}_0(x_k) + \varepsilon_{g_0,k}$ and $J_I(x_k) \leftarrow \bar{J}(x_k) + \varepsilon_{J_I,k}$, where $\varepsilon_{g_0,k}$ was drawn from a uniform distribution over an ℓ_2-norm ball with radius ϵ_g, $\varepsilon_{J_I,k}$ had the same sparsity pattern as $\bar{J}(x_k)$, and the nonzero components of each row of $\varepsilon_{J_I,k}$ were drawn from a uniform distribution over an ℓ_2-norm ball with radius $\epsilon_J/\sqrt{q}$. Finally, with $\bar{H}_k$ defined as the Hessian of the Lagrangian evaluated at (x_k, y_k), we set $H_k \leftarrow \bar{H}_k + \varepsilon_{H,k}$, where $\varepsilon_{H,k}$ is a diagonal matrix with each diagonal element drawn from a uniform distribution over $[-\epsilon_H, \epsilon_H]$.

For each test problem, barrier parameter, and noise level, we ran our code until 2000 iterations were conducted or a time limit of 1 hour was reached. In actual practice, it would be more reasonable to run Algorithm 1 with termination conditions corresponding to those stated in the algorithm itself, i.e., for prescribed thresholds, the algorithm may terminate when—according to noisy quantities—an approximate infeasible stationary point or approximate first-order stationary point has been reached. However, for our purposes here, we chose to disable these termination conditions in order to be confident that the algorithm has reached the highest quality solution that it can in each run.

In Figure 1, we show the quality of the solutions found with $\mu = 10^{-1}$ in terms of *noisy* stationarity measures, both with respect to $\epsilon_f = 10^{-2}$ and $\epsilon_f = 10^{-6}$. (Recall from above that the value for ϵ_f determines all of the values for ϵ_c, ϵ_g, ϵ_J, and ϵ_H as well.) Specifically, for varying thresholds on the stationarity measure for solving the barrier subproblem (left) and a stationarity measure for minimizing infeasibility (right), we show the percentages of problems for which the geometric average of the last 10 iterates satisfied the given threshold for each measure. Figure 2 provides the results for these same runs of the algorithm (using noisy quantities), but considers values of the measures with respect to *noiseless* function and derivative values. These are included to show that the algorithm is reaching points that can be considered nearly stationary for noiseless values, even though it only has access to noisy ones.

(a) Stationarity w.r.t. barrier subproblem

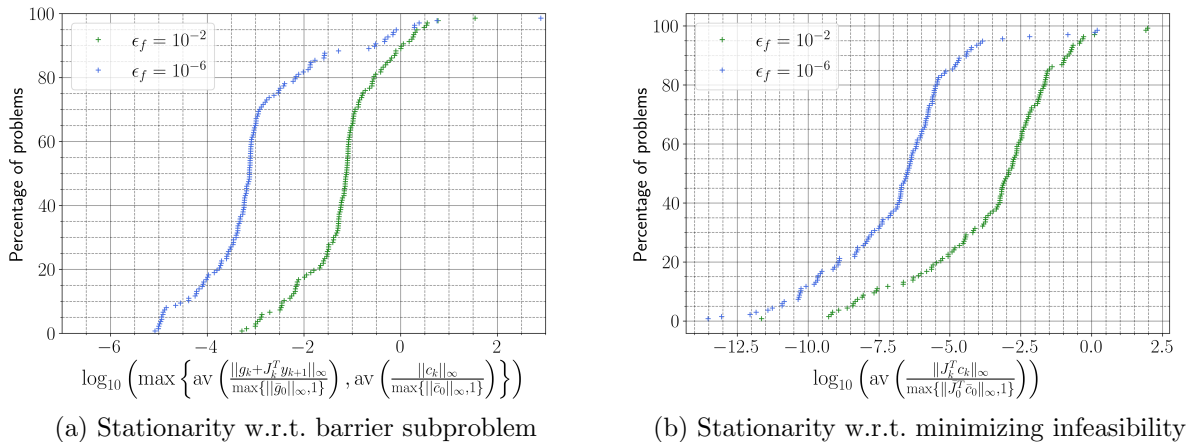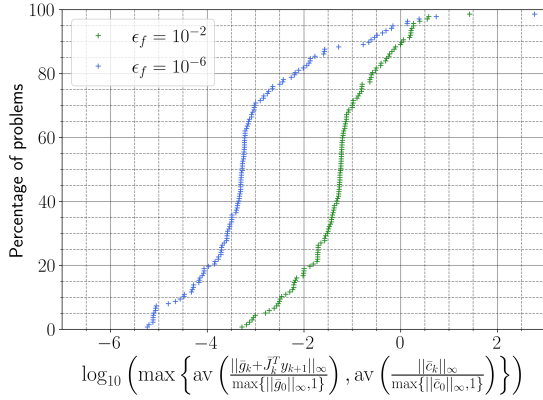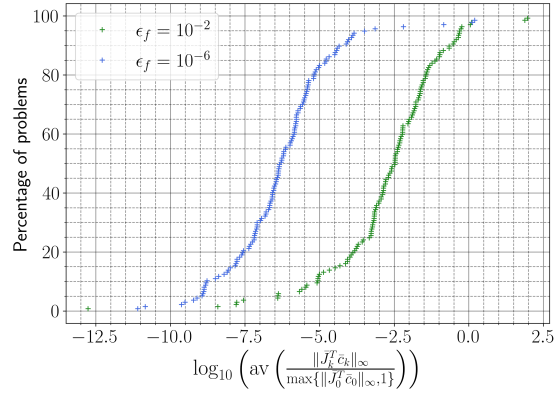(b) Stationarity w.r.t. minimizing infeasibility

Figure 1: $\mu = 10^{-1}$, measured with noisy function and derivative values

Figures 3 and 4, respectively, show results in terms of *noisy* and *noiseless* stationarity measures, this time when the barrier parameter is $\mu = 10^{-4}$.
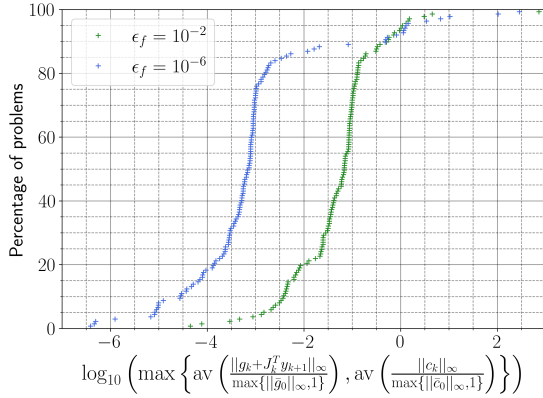
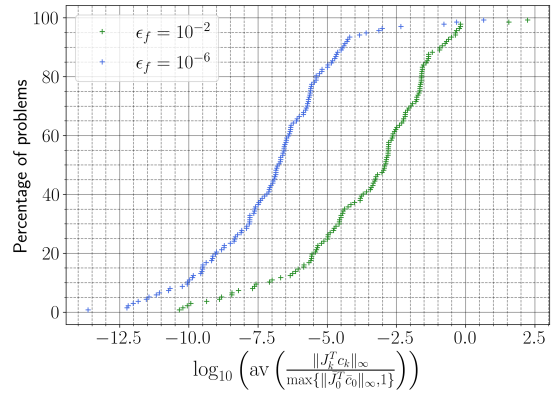(a) Stationarity w.r.t. barrier subproblem   (b) Stationarity w.r.t. minimizing infeasibility

Figure 2: $\mu = 10^{-1}$, measured with noiseless function and derivative values

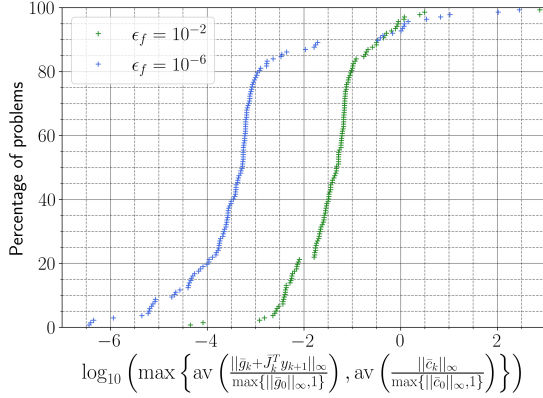(a) Stationarity w.r.t. barrier subproblem   (b) Stationarity w.r.t. minimizing infeasibility

Figure 3: $\mu = 10^{-4}$, measured with noisy function and derivative values

The results for $\mu = 10^{-4}$ are comparable to those with $\mu = 10^{-1}$. On one hand, these results show that the performance of our algorithm is robust with respect to the barrier parameter value, which is good news. At the same time, the results show that with a smaller barrier parameter, one should not necessarily expect to be able to obtain significantly better final solution estimates. This can be attributed to the fact, which we mentioned at the end of §1.1, that regardless of the barrier parameter value, one may be inhibited from obtaining high-accuracy solutions due to the presence of noise.

6   Conclusion and Future Work

We have presented, analyzed, and tested the practical performance of an interior-point algorithm for solving continuous inequality-constrained optimization problems when function and derivative

(a) Stationarity w.r.t. barrier subproblem
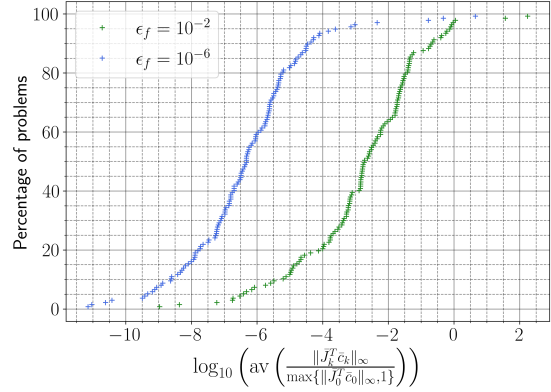
(b) Stationarity w.r.t. minimizing infeasibility

Figure 4: $\mu = 10^{-4}$, measured with noiseless function and derivative values

values are corrupted by noise. Our algorithm focuses on solving a single barrier subproblem. We have shown that the algorithm either terminates finitely due to one of two reasonable termination criteria, or it generates an infinite sequence of iterates, and in this latter case it is guaranteed under certain assumptions to reach a point at which a stationarity measure is below a threshold that depends on the noise level.

As previously mentioned in the paper and shown in our numerical experiments, in the noisy setting there is not necessarily a benefit of reducing the barrier parameter too small, since in any case the noise in the function and derivative values may inhibit the ability to acquire a highly accurate solution (to the underlying noiseless problem). That being said, it is very reasonable to expect that, in some cases, one may obtain better solutions by solving a sequence of barrier subproblems for a diminishing sequence of barrier parameters. We propose that this may be done by incorporating a practical stopping condition for a given barrier subproblem that builds off of the practical condition proposed in Theorem 3.9 in [16]. This would essentially involve employing an iteration-dependent variant of (51) where the unknown quantities are estimated. Since $(\epsilon_f, \epsilon_g, \epsilon_c, \epsilon_J)$ are presumed to be known, and both $g_{\sup}$ and $J_{\sup}$ are easily estimated during a run of the algorithm using (23), the main quantity that requires some attention is ξ_d. By its definition in Lemma 4.20, one finds that this value is affected by the choice of ω in (12). If ω is chosen to be a relatively large value, then ξ_d becomes relatively small, which in turn can cause $\{\rho_{k,1}, \rho_{k,2}, \rho_{k,3}\}$ to become relatively large. In some sense, this makes (51) easier to satisfy, but at the expense of only having a looser bound on $\Delta \bar{m}_k(\bar{d}_k, \bar{\tau})$. By contrast, one may obtain a tighter, computable bound on $\Delta \bar{m}_k(\bar{d}_k, \bar{\tau})$ by replacing ω in the definition of ϵ_d with $\|J_k^T c_k\|/\|v_k\|$. Passing this value through the definition of ϵ_d, one can use the resulting right-hand side of (51) as a threshold for terminating the solve for a given barrier subproblem. (Specifically, the solve may terminate if/when $\Delta m(d_k, \tau_k)$ is below this threshold.) If/when this bound holds, the barrier parameter can be reduced using a typical update rule; see, e.g., [31].

References

[1] 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.

[2] 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.

[3] 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.

[4] Albert S. Berahas, Jiahao Shi, and Baoyu Zhou. Optimistic noise-aware sequential quadratic programming for equality constrained optimization with rank-deficient jacobians. *Working paper*, 2025.

[5] Ingrid Bongartz, Andrew R Conn, Nick Gould, and Ph L Toint. Cute: Constrained and unconstrained testing environment. *ACM Transactions on Mathematical Software (TOMS)*, 21(1):123–160, 1995.

[6] Richard H Byrd, Jean Charles Gilbert, and Jorge Nocedal. A trust region method based on interior point techniques for nonlinear programming. *Mathematical programming*, 89(1):149–185, 2000.

[7] Richard H Byrd, Mary E Hribar, and Jorge Nocedal. An interior point algorithm for large-scale nonlinear programming. *SIAM Journal on Optimization*, 9(4):877–900, 1999.

[8] Richard H Byrd, Jorge Nocedal, and Richard A Waltz. K nitro: An integrated package for nonlinear optimization. *Large-scale nonlinear optimization*, pages 35–59, 2006.

[9] Frank E. Curtis, Xin Jiang, and Qi Wang. Single-Loop Deterministic and Stochastic Interior-Point Algorithms for Nonlinearly Constrained Optimization. arXiv 2408.16186, 2024.

[10] 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 2304.14907, 2023.

[11] 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.

[12] Frank E Curtis, Daniel P Robinson, and Baoyu Zhou. Inexact sequential quadratic optimization for minimizing a stochastic objective function subject to deterministic nonlinear equality constraints. *arXiv preprint arXiv:2107.03512*, 2021.

[13] Frank E Curtis, Daniel P Robinson, and Baoyu Zhou. Sequential quadratic optimization for stochastic optimization with deterministic nonlinear inequality and equality constraints. *arXiv preprint arXiv:2302.14790*, 2023.

[14] Frank E Curtis, Katya Scheinberg, and Rui Shi. A stochastic trust region algorithm based on careful step normalization. *Informs Journal on Optimization*, 1(3):200–220, 2019.

[15] Frank E. Curtis, Olaf Schenk, and Andreas Waechter. An Interior-Point Algorithm for Large-Scale Nonlinear Optimization with Inexact Step Computations. *SIAM Journal on Scientific Computing*, 32(6):3447–3475, 2010.

[16] 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.

[17] David M Gay. Hooking your solver to ampl. Technical report, Citeseer, 1997.

[18] Philip E Gill, Walter Murray, and Margaret H Wright. *Practical optimization*. SIAM, 2019.

[19] O. L. Mangasarian and S. Fromovitz. The Fritz John Necessary Optimality Conditions in the Presence of Equality and Inequality Constraints. *Journal of Mathematical Analysis and Applications*, 17:37–47, 1967.

[20] Jorge J Moré and Danny C Sorensen. Computing a trust region step. *SIAM Journal on scientific and statistical computing*, 4(3):553–572, 1983.

[21] Sen Na, Mihai Anitescu, and Mladen Kolar. Inequality constrained stochastic nonlinear optimization via active-set sequential quadratic programming. *Mathematical Programming*, pages 1–75, 2023.

[22] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 1999.

[23] Figen Oztoprak, Richard Byrd, and Jorge Nocedal. Constrained optimization in the presence of noise. *SIAM Journal on Optimization*, 33(3):2118–2136, 2023.

[24] Courtney Paquette and Katya Scheinberg. A stochastic line search method with expected complexity analysis. *SIAM Journal on Optimization*, 30(1):349–376, 2020.

[25] Songqiang Qiu and Vyacheslav Kungurtsev. A sequential quadratic programming method for optimization with stochastic objective functions, deterministic inequality constraints and robust subproblems. *arXiv preprint arXiv:2302.07947*, 2023.

[26] Hao-Jun Michael Shi. *Methods for Stochastic, Noisy, and Derivative-Free Optimization*. PhD thesis, Northwestern University, 2021.

[27] GW Stewart. On the numerical analysis of oblique projectors. *SIAM Journal on Matrix Analysis and Applications*, 32(1):309–348, 2011.

[28] Shigeng Sun and Jorge Nocedal. A trust region method for noisy unconstrained optimization. *Mathematical Programming*, pages 1–28, 2023.

[29] Andreas Wächter and Lorenz T Biegler. Failure of global convergence for a class of interior point methods for nonlinear programming. *Mathematical Programming*, 88:565–574, 2000.

[30] 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.

[31] Andreas Wächter and Lorenz T Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106:25–57, 2006.

[32] Yuchen Xie, Richard H Byrd, and Jorge Nocedal. Analysis of the bfgs method with errors. *SIAM Journal on Optimization*, 30(1):182–209, 2020.

[33] Hiroshi Yamashita. A globally convergent primal-dual interior point method for constrained optimization. *Optimization Methods and Software*, 10(2):443–469, 1998.

[34] Hiroshi Yamashita, Hiroshi Yabe, and Takahito Tanabe. A globally and superlinearly convergent primal-dual interior point trust region method for large scale constrained optimization. *Mathematical Programming*, 102:111–151, 2005.

A   Convergence of $\|\bar{J}_k^T \bar{c}_k\|^2$

Under looser assumptions than for the first part of Theorem 1—specifically, without Assumptions 4.3 and 4.4—we prove in this appendix that Algorithm 1 is guaranteed to generate a sequence of iterates such that, for some $k \in \mathbb{N}$, one finds that $\|\bar{J}_k^T \bar{c}_k\|^2$ is below a threshold that depends on noisy quantities. The statement is similar to the first part of Theorem 1. As in Section 4.3, the result presumes that the algorithm does not terminate finitely.

Theorem 2. *Suppose that Assumptions 2.1, 4.1, and 4.2 hold, and for all $k \in \mathbb{N}$, with $\xi_2(\tau_k)$ defined in Lemma 4.14, (ξ_3, ξ_4) defined in Lemma 4.16, $\varepsilon_{d,k}$ defined in (27), and $\alpha_k^{\max}$ computed to satisfy the fraction-to-the-boundary rule, let*

$$\check{\alpha}_k := \min\left\{ \frac{1}{4\xi_2(\tau_k)} \left(\frac{(1-\eta_\phi)\xi_3\|\bar{J}_k^T \bar{c}_k\|^2}{\xi_4(\|\bar{u}_k\|^2 + \|\bar{J}_k^T \bar{c}_k\|^2) + \|\varepsilon_{d,k}\|^2} \right), \alpha_k^{\max} \right\}. \tag{53}$$

Then, with $E_k^{m,1}$ defined in Lemma 4.12, $E_k^{m,2}$ defined in Lemma 4.13, and E_k^m defined in Corollary 4.2, there exists $k \in \mathbb{N}$ such that

$$\|\bar{J}_k^T \bar{c}_k\|^2 \leq \max\left\{ \frac{2(\eta_\phi E_k^{m,1} - (1-\eta_\phi)E_k^{m,2})}{(1-\eta_\phi)\xi_3}, -\frac{2E_k^m}{\xi_3}, \frac{8(2+\zeta)\varepsilon_k}{\check{\alpha}_k \eta_\phi \xi_3} \right\}. \tag{54}$$

Proof. Consider arbitrary $k \in \mathbb{N}$. Our first aim is to show that if

$$\eta_\phi E_k^{m,1} - (1-\eta_\phi)E_k^{m,2} \leq \tfrac{1}{2}(1-\eta_\phi)\xi_3\|\bar{J}_k^T \bar{c}_k\|^2, \tag{55}$$

then the relaxed Armijo condition (20) is satisfied for all $\alpha \leq \check{\alpha}_k$. Toward this end, first observe that by (10), Lemma 4.13, and 4.14, one finds for all $\alpha \in (0, \alpha_k^{\max}]$ that

$$\begin{aligned}
\phi(z_k + \alpha \hat{d}_k, \tau_k) - \phi(z_k, \tau_k) &\leq \bar{\phi}(z_k + \alpha \hat{d}_k, \tau_k) - \bar{\phi}(z_k, \tau_k) + 2\varepsilon_k \\
&\leq -\alpha\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) - \alpha E_k^{m,2} + \xi_2(\tau_k)\alpha^2\|d_k\|^2 + 2\varepsilon_k \\
&= -\alpha\eta_\phi\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) - \alpha(1-\eta_\phi)\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) \\
&\quad - \alpha E_k^{m,2} + \xi_2(\tau_k)\alpha^2\|d_k\|^2 + 2\varepsilon_k.
\end{aligned}$$

Consequently, by Corollary 4.2 and $\|d_k\|^2 \leq 2(\|\bar{d}_k\|^2 + \|\varepsilon_{d,k}\|^2)$, one finds

$$\phi(z_k + \alpha\hat{d}_k, \tau_k) - \phi(z_k, \tau_k)$$
$$\leq -\alpha\eta_\phi\Delta m_k(d_k, \tau_k) + \alpha\eta_\phi E_k^m + 2\varepsilon_k$$
$$\quad -\alpha(1-\eta_\phi)\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) - \alpha E_k^{m,2} + 2\xi_2(\tau_k)\alpha^2(\|\bar{d}_k\|^2 + \|\varepsilon_{d,k}\|^2)$$
$$= -\alpha\eta_\phi\Delta m_k(d_k, \tau_k) + 2\varepsilon_k$$
$$\quad -\alpha(1-\eta_\phi)\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) + 2\xi_2(\tau_k)\alpha^2(\|\bar{d}_k\|^2 + \|\varepsilon_{d,k}\|^2)$$
$$\quad + \alpha\eta_\phi E_k^{m,1} - \alpha(1-\eta_\phi)E_k^{m,2}.$$

Thus, under (55), one finds from Lemma 4.16 that

$$\phi(z_k + \alpha\hat{d}_k, \tau_k) - \phi(z_k, \tau_k) \leq -\alpha\eta_\phi\Delta m_k(d_k, \tau_k) + 2\varepsilon_k$$
$$\quad -\alpha(1-\eta_\phi)\xi_3(\bar{\tau}_k\|\bar{u}_k\|^2 + \|\bar{J}_k^T\bar{c}_k\|^2)$$
$$\quad + 2\xi_2(\tau_k)\xi_4\alpha^2(\|\bar{u}_k\|^2 + \|\bar{J}_k^T\bar{c}_k\|^2) + 2\xi_2(\tau_k)\alpha^2\|\varepsilon_{d,k}\|^2$$
$$\quad + \tfrac{1}{2}\alpha(1-\eta_\phi)\xi_3\|\bar{J}_k^T\bar{c}_k\|^2$$
$$\leq -\alpha\eta_\phi\Delta m_k(d_k, \tau_k) + 2\varepsilon_k$$
$$\quad -\tfrac{1}{2}\alpha(1-\eta_\phi)\xi_3\|\bar{J}_k^T\bar{c}_k\|^2$$
$$\quad + 2\xi_2(\tau_k)\xi_4\alpha^2(\|\bar{u}_k\|^2 + \|\bar{J}_k^T\bar{c}_k\|^2) + 2\xi_2(\tau_k)\alpha^2\|\varepsilon_{d,k}\|^2,$$

which along with (53) implies that, for all $\alpha \leq \check{\alpha}_k$, (20) holds, as claimed.

Now, to derive a contradiction to the desired conclusion, suppose that (54) does not hold for all $k \in \mathbb{N}$. Under this supposition, consider arbitrary $k \in \mathbb{N}$. Since (54) does not hold, it follows that (55) holds, which in turn means that (20) holds for some step size $\alpha \geq \check{\alpha}_k/2$. Thus, with Corollary 4.2, one has

$$\bar{\phi}(z_k + \alpha\hat{d}_k, \tau_k) - \bar{\phi}(z_k, \tau_k) \leq \phi(z_k + \alpha\hat{d}_k, \tau_k) - \phi(z_k, \tau_k) + 2\varepsilon_k$$
$$\leq -\tfrac{1}{2}\check{\alpha}_k\eta_\phi\Delta m_k(d_k, \tau_k) + 2\varepsilon_k + (2+\zeta)\varepsilon_k$$
$$= -\tfrac{1}{2}\check{\alpha}_k\eta_\phi\Delta\bar{m}_k(\bar{d}_k, \bar{\tau}_k) - \tfrac{1}{2}\check{\alpha}_k\eta_\phi E_k^m + 2\varepsilon_k + (2+\zeta)\varepsilon_k.$$

Now by Lemma 4.16, $\|\bar{J}_k^T\bar{c}_k\|^2 > -\frac{2E_k^m}{\xi_3}$, and $\|\bar{J}_k^T\bar{c}_k\|^2 > \frac{8(2+\zeta)\varepsilon_k}{\check{\alpha}_k\eta_\phi\xi_3}$, one finds

$$\bar{\phi}(z_k + \alpha\hat{d}_k, \tau_k) - \bar{\phi}(z_k, \tau_k) \leq -\tfrac{1}{2}\check{\alpha}_k\eta_\phi\xi_3\|\bar{J}_k^T\bar{c}_k\|^2 - \tfrac{1}{2}\check{\alpha}_k\eta_\phi E_k^m + 2\varepsilon_k + (2+\zeta)\varepsilon_k$$
$$< -\tfrac{1}{4}\check{\alpha}_k\eta_\phi\xi_3\|\bar{J}_k^T\bar{c}_k\|^2 + 2\varepsilon_k + (2+\zeta)\varepsilon_k < -\zeta\varepsilon_k.$$

However, with Remark 3.1 this implies that $\bar{\phi}(z_{k+1}, \tau_k) - \bar{\phi}(z_k, \tau_k) < -\zeta\varepsilon_k$ for all $k \in \mathbb{N}$, which contradicts Assumption 4.2, specifically the fact that $\{\bar{\phi}(z_k, \tau_k)\}$ is bounded below. Consequently, (54) must hold for some $k \in \mathbb{N}$, as desired. $\qquad\square\qquad\qquad\qquad\square$