# Constrained Optimization in the Presence of Noise

Figen Oztoprak* Richard Byrd † Jorge Nocedal ‡

October 5, 2021

**Abstract**

The problem of interest is the minimization of a nonlinear function subject to nonlinear equality constraints using a sequential quadratic programming (SQP) method. The minimization must be performed while observing only noisy evaluations of the objective and constraint functions. In order to obtain stability, the classical SQP method is modified by relaxing the standard Armijo line search based on the noise level in the functions, which is assumed to be known. Convergence theory is presented giving conditions under which the iterates converge to a neighborhood of the solution characterized by the noise level and the problem conditioning. The analysis assumes that the SQP algorithm does not require regularization or trust regions. Numerical experiments indicate that the relaxed line search improves the practical performance of the method on problems involving uniformly distributed noise. One important application of this work is in the field of derivative-free optimization, when finite differences are employed to estimate gradients.

## 1 Introduction

Let us consider the equality constrained nonlinear optimization problem

$$\min_x f(x) \qquad \text{s.t.} \;\; c(x) = 0, \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c(x) : \mathbb{R}^n \to \mathbb{R}^m$ are smooth functions. We assume that the minimization must be performed while observing approximate evaluations $\tilde{f}(x), \tilde{c}(x)$ of the functions $f, c$ and their derivatives.

We consider the application of a sequential quadratic programming (SQP) algorithm that employs an $\ell_1$ merit function to control the stepsize. The goal of the paper is to study the effect of noise on the behavior of the SQP algorithm, particularly the achievable accuracy

*Artelys Corporation

†Computer Science Department, University of Colorado, Boulder, USA

‡Department of Industrial Engineering and Management Sciences, Northwestern University, USA. This author was supported by National Science Foundation grant DMS-2011494, AFOSR grant FA95502110084, and ONR grant N00014-21-1-2675.

in the solution, and to highlight the aspects of the algorithm that are most susceptible to errors (or noise)—and redesign them. This work was motivated by applications in which the derivatives of $f$ and $c$ are approximated by finite differences [15], and thus contain errors, but the algorithm and analysis apply to the more general setting when stochastic or deterministic noise are present in both the function and derivative evaluations.

Let us define

$$g(x) = \nabla f(x), \qquad J(x) = \nabla c(x) \in \mathbb{R}^{m\times n}, \quad m < n, \tag{1.2}$$

and let $\tilde{g}(x), \tilde{J}(x)$ be the corresponding noisy evaluations. The iteration of the SQP algorithm is given by

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.3}$$

where $d_k$ is the solution of the quadratic subproblem

$$\min_{d\in\mathbb{R}^n} \ \tfrac{1}{2} d^T H_k d + \tilde{g}_k^T d \tag{1.4}$$

$$\text{s.t. } \tilde{c}_k + \tilde{J}_k d = 0, \tag{1.5}$$

and the steplength $\alpha_k > 0$ is chosen so as to ensure sufficient decrease in the merit function

$$\tilde{\phi}(x) = \tilde{f}(x) + \pi \|\tilde{c}(x)\|_1 \tag{1.6}$$

when the iterates are far away from a solution. Here $\pi > 0$ is a penalty parameter that is adjusted during the course of the optimization. The symmetric matrix $H_k$ is generally chosen as an approximation to the Hessian of the Lagrangian. However, in this paper we assume that $H_k$ is a multiple of the identity matrix,

$$H_k = \beta_k I \qquad \beta_k > 0, \tag{1.7}$$

because allowing more general choices introduces more constants in the analysis without contributing to the main goals of this investigation.

As in the noiseless case, the control of the penalty parameter in (1.6) is of critical importance in the SQP algorithm. $\pi$ should be chosen so that the SQP direction $d_k$ is a descent direction for $\tilde{\phi}$ at $x_k$, and it should provide adequate control on the size of $\alpha_k$. The proposed algorithm has the general form of a classical SQP method [7], specialized to the case when $H_k$ is a multiple of the identity matrix, and introduces a modification in the line search designed to handle noise.

We assume throughout that the noise in the function and gradient evaluations is bounded by some constants $\epsilon_f$ and $\epsilon_c$. This is not always the case in practice (e.g. when noise is Gaussian) but it covers many important practical settings, including computational noise [11]. Furthermore, we assume that $\epsilon_f, \epsilon_c$ are known, or can be estimated, and that the algorithm has access to them.

This study was motivated by some practical computations performed by the authors using the KNITRO software package [5]. They selected a few challenging nonlinear optimization

problems involving equality and inequality constraints, injected noise in the objective and constraints, and computed derivatives using noise-aware finite difference formula; see e.g., Moré and Wild [11] and Berahas et al. [1]. They observed that, for low levels of noise, KNITRO returned acceptable answers, even though one might suspect the default algorithm to be brittle in this setting. As the noise level was increased, the quality of the solution deteriorated markedly, suggesting that classical optimization methods should be redesigned to handle noise. To guide this investigation, it is essential to develop a convergence theory. In this paper, we focus on the case when noise cannot be diminished, and characterize the accuracy of a noise tolerant optimization algorithm.

As a first step in this investigation, we find it convenient to consider equality constrained optimization, and study the performance of a sequential quadratic optimization method, which is a simple method in this setting and must yet confront some important challenges raised by the presence of noise.

### 1.1 Contributions of this work

The main contribution of this paper is the development of a convergence theory for a classical sequential quadratic programming (SQP) algorithm for equality constrained optimization in the presence of noise. It is shown that, by introducing a relaxation in the line search procedure while keeping all other components of the SQP method unchanged, the iterates of the algorithm reach an acceptable neighborhood $C_1$ of the solution defined by a stationarity measure for the problem. Furthermore, once the iterates enter $C_1$ they cannot escape a larger neighborhood $C_2$ and must revisit $C_1$ an infinite number of times. The analysis gives a detailed characterization of these neighborhoods in terms of the noise level and problem characteristics. Numerical experiments show that the relaxed line search is, in fact, beneficial in practice.

Our convergence results assume that errors in function and gradients are bounded, and the analysis is deterministic, yielding somewhat pessimistic bounds. We believe, however, that the results can be useful in the design of robust constrained optimization methods. Specifically, our analysis suggests that only slight modifications are needed so that a classical SQP method is able to handle bounded noise.

### 1.2 Literature Review

Early work on constrained optimization in the presence of noise is reviewed by Poljak (a.k.a. Polyak) [13]. His study includes penalty, Lagrange, or extended Lagrange functions, and establishes probabilistic convergence theorems provided the steplength is chosen small enough from the start. Hintermueller [10] studies a penalty SQP method in which equality constraints are replaced with upper and lower bounding surrogates. Assuming that the noise level in the function is known, it is shown that in the limit the bounds contain a solution. Schittkowski [14] uses a non-monotone line search to handle errors due to approximate function and derivative evaluations. His algorithm was implemented in the NLPQLP software, which is reported to be successful in practice, but no convergence theory were presented.

The work that is most closely related to this study is [2, 3, 6]. In [3], an SQP method for equality constrained optimization is presented to handle the case when the objective function is stochastic and the constraints are deterministic. The stepsize is obtained by adaptively estimating Lipschitz constants in place of a line search. Conditions for convergence in expectation are established. [2] considers the case when Jacobians can be rank deficient, proposes a step decomposition approach, and presents compelling numerical results. [6] studies an SQP algorithm with an inexact step computation for the same problem setting. These three papers give careful attention to the behavior of the penalty parameter. For example, in [3] the penalty parameter is chosen in a way that provides sufficient descent in the quadratic model of the merit function in the deterministic setting. In the stochastic setting, they employ the stochastic gradient of the objective in the same formulae for updating the penalty parameter, but they can no longer guarantee that the resulting penalty parameter will be large enough and bounded. They prove their convergence results assuming that the penalty parameter is well behaved. Then, they discuss the probability of having small penalty values, and note that the boundedness issue is resolved by making the same assumption as in this paper, namely that noise is always bounded.

*Notation.* We let $\|\cdot\|$ denote the $\ell_2$ norm, unless otherwise stated. As is the convention, $f_k$ stands for $f(x_k)$ and similarly for other functions. The terms *error* and *noise* in the functions is used interchangeably. Since we assume absolute bounds on these quantities, the distinction between them is not important in this study.

## 2 The Algorithm

Before presenting the algorithm, we introduce some notation. We model the first-order change in the merit function $\phi$ at an iterate $x_k$ as

$$\tilde{\ell}(x_k; d_k) = \tilde{g}_k^T d_k + \pi_k \|\tilde{c}_k + \tilde{J}_k d_k\|_1 - \pi_k \|\tilde{c}_k\|_1. \tag{2.1}$$

We also define

$$\hat{\lambda}_k = (\tilde{J}_k \tilde{J}_k^T)^{-1} \tilde{J}_k \tilde{g}_k, \tag{2.2}$$

which is the standard least squares multiplier estimate [12, eqn(18.21)], accounting for noisy function evaluations. We assume that $\tilde{J}_k$ is full rank for all $k$, hence $\hat{\lambda}_k$ is well defined.

The penalty parameter will be updated using the following classical formula [12, eqn(18.32)]. Given a (fixed) parameter $\tau \in (0, 1)$, we set at every iteration

$$\pi_k = \begin{cases} \pi_{k-1} & \text{if } \pi_{k-1} \geq \frac{1}{1-\tau} \|\hat{\lambda}_k\|_\infty \\ \frac{2}{1-\tau} \|\hat{\lambda}_k\|_\infty & \text{otherwise.} \end{cases} \tag{2.3}$$

The factor 2 in the second line of (2.3) is introduced so that when $\pi_k$ is increased, it is increased substantially. We will see that this rule ensures that $\pi_k$ is eventually fixed. (In general, SQP methods do not set $H_k = \beta_k I$. In that case, using the least squares multiplier estimate in (2.3) will not lead to a convergent method.)

The algorithm for solving problem (1.1), when only noisy evaluations of the functions $\tilde{f}, \tilde{c}, \tilde{g}, \tilde{J}$ are available, is as follows.

**Algorithm 1** Noise Tolerant SQP Algorithm

**Input:** Initial iterate $x_0$, initial merit parameter $\pi_{-1} > 0$, bounds $\epsilon_f, \epsilon_c$ on the noise (3.1), and constants $\tau, \nu \in (0, 1)$.
Set $k \leftarrow 0$
**Repeat** until a termination test is satisfied:

1: Compute $\beta_k > 0$ and set $H_k = \beta_k I$ in (1.4)
2: Compute $d_k$ by solving (1.4)-(1.5)
3: Compute $\hat{\lambda}_k$ via (2.2)
4: Update penalty parameter $\pi_k$ by (2.3)
5: Compute $\tilde{\ell}(x_k; d_k)$ as in (2.1)
6: Set $\epsilon_R = 2(\epsilon_f + \pi_k \epsilon_c)$
7: Choose steplength $\alpha_k > 0$ such that

$$\tilde{\phi}(x_k + \alpha_k d_k) \leq \tilde{\phi}(x_k) + \nu \alpha_k \tilde{\ell}(x_k; d_k) + \epsilon_R, \tag{2.4}$$

8: Compute new iterate: $x_{k+1} = x_k + \alpha_k d_k$
9: Set $k \leftarrow k + 1$

The steplength $\alpha_k$ is computed in Step 7 using a backtracking line search. We refer to (2.4) as the *relaxed Armijo condition*. The term $\epsilon_R$ introduces a margin that facilitates the convergence analysis in the presence of noise, and as discussed in Section 4, is also useful in practice. Note that the line search cannot fail since (2.4) is satisfied for sufficiently small $\alpha_k$, by definition of $\epsilon_R$. In this paper, we assume that the quadratic subproblem (1.4)-(1.5) has a unique solution at every iteration—admittedly a strong assumption, but one that helps us focus on the effect of noise without the complicating effects of regularization parameters or trust regions. The study of a practical algorithm that employs those globalization strategies will be the subject of future work.

# 3 Global Convergence

In this section we show that the iterates generated by Algorithm 1 converge to a neighborhood of the solution determined by the noise level and certain characteristics of the problem. We also show that once the iterates reach this neighborhood they cannot stray away from it (under normal circumstances). We start by stating the assumptions upon which our analysis is built.

**Assumptions 3.1.** *The function $f$ has a Lipschitz continuous gradient with constant $L_f$. The functions $\nabla c_i$ are Lipschitz continuous for $i = 1, \ldots, m$ with the corresponding constants held in the vector $L_c$.*

We also assume that the error (or noise) in the evaluation of the functions is bounded.

**Assumptions 3.2.** *There exist positive constants $\epsilon_f, \epsilon_c, \epsilon_g, \epsilon_J$ such that for all $x \in \mathbb{R}^n$,*

$$|\tilde{f}(x) - f(x)| \leq \epsilon_f, \qquad \|\tilde{c}(x) - c(x)\|_1 \leq \epsilon_c, \tag{3.1}$$

$$\|\tilde{g}(x) - g(x)\| \leq \epsilon_g, \qquad \|\tilde{J}(x) - J(x)\|_{1,2} \leq \epsilon_J. \tag{3.2}$$

Here, $\|\cdot\|$ denotes the Euclidean norm and $\|\cdot\|_{1,2}$ denotes the matrix norm induced by the $\ell_1$ norm on $\mathbb{R}^m$ and the Euclidean norm on $\mathbb{R}^n$.

As already mentioned, we assume that, for all $k$, the matrices $\tilde{J}_k$ have full rank so that the quadratic problem (1.4)-(1.5) has a unique solution. To state this precisely, we let $\sigma_{min}(A)$ denote the smallest singular value of a matrix $A$.

**Assumptions 3.3.** *For all $k$, the scalar $\beta_k$ in (1.7) satisfies*

$$0 < b_l \leq \beta_k \leq b_u, \tag{3.3}$$

*for some constants $b_l, b_u$, and there is a constant $\gamma > 0$ such that*

$$\sigma_{min}(J_k) \geq \gamma, \quad \textit{with} \quad \gamma > \epsilon_J, \quad \forall k. \tag{3.4}$$

*Furthermore, the sequences $\{f_k\}, \{\|c_k\|\}$, $\{\|g_k\|\}$, $\{\|J_k\|\}$ generated by the algorithm are bounded.*

By the matrix inversion lemma [8] and (3.2), if $J_k$ has full rank and $\gamma > \epsilon_J$, then $\tilde{J}_k$ is also full rank and

$$\|\tilde{J}_k^T(\tilde{J}_k\tilde{J}_k^T)^{-1}\| \leq \frac{1}{\gamma - \epsilon_J} \equiv \delta, \quad \forall k. \tag{3.5}$$

The assumption that the sequences $\{f_k\}, \{\|c_k\|\}$, $\{\|g_k\|\}$, $\{\|J_k\|\}$ generated by the algorithm are bounded is fairly standard in the literature and is designed to avoid pathological situations. For example, the merit function $\phi$ may be unbounded below away from the solution if $\pi$ is not large enough. Although there are strategies to avoid these situations (see e.g. [12, §18.5], we do not include them in our algorithm, for simplicity.

Given these three sets of assumptions, we are ready to study the convergence properties of Algorithm 1. Let us apply the well known descent lemma (see e.g.[4]) to the true (noiseless) merit function

$$\phi(x) = f(x) + \pi\|c(x)\|_1. \tag{3.6}$$

We have that for any $(x, d)$

$$\phi(x + \alpha d) \leq \phi(x) + \alpha g(x)^T d + \pi[\|c(x) + \alpha J(x)d\|_1 - \|c(x)\|_1] + \tfrac{1}{2}(L_f + \pi\|L_c\|_1)\alpha^2\|d\|^2. \tag{3.7}$$

Thus, we can write

$$\phi(x + \alpha d) - \phi(x) \leq \ell(x; \alpha d) + \tfrac{1}{2}(L_f + \pi\|L_c\|_1)\alpha^2\|d\|^2, \tag{3.8}$$

where

$$\ell(x;s) = g(x)^T s + \pi\|c(x) + J(x)s\|_1 - \pi\|c(x)\|_1. \tag{3.9}$$

When function and derivatives are exact, it is easy to show that for $\pi$ sufficiently large and $\alpha$ sufficiently small we can guarantee a reduction in $\phi$; see [12]. We must establish that this is also the case in the noisy setting—before the iterates approach the region around the solution where noise dominates. We begin by establishing bounds on the step $d_k$.

### 3.1 Preliminary results

The optimality conditions of the quadratic problem (1.4)-(1.5) are given by

$$\begin{pmatrix} H_k & \tilde{J}_k^T \\ \tilde{J}_k & 0 \end{pmatrix} \begin{pmatrix} d_k \\ d_y \end{pmatrix} = - \begin{pmatrix} \tilde{g}_k + \tilde{J}_k^T y \\ \tilde{c}_k \end{pmatrix}, \tag{3.10}$$

for some Lagrange multiplier $y \in \mathbb{R}^m$. The step $d_k$ can be written as the sum of two orthogonal components,

$$d_k = v_k + u_k, \tag{3.11}$$

where $v_k$ is in the range space of $\tilde{J}_k^T$ and $u_k$ is in the null space of $J_k$. A simple computation from (3.10) shows that

$$v_k = -\tilde{J}_k^T(\tilde{J}_k\tilde{J}_k^T)^{-1}\tilde{c}_k, \qquad u_k = -\frac{1}{\beta_k}\tilde{P}_k\tilde{g}_k, \tag{3.12}$$

where

$$\tilde{P}_k = I - \tilde{J}_k^T\left(\tilde{J}_k\tilde{J}_k^T\right)^{-1}\tilde{J}_k \tag{3.13}$$

is an orthogonal projection matrix onto the tangent space of the constraints. We now establish bounds on $u_k, v_k$. In what follows, we let $J^\dagger$ denote the Moore-Penrose generalized inverse of a matrix $J$, and define $P_k = I - J_k^T\left(J_kJ_k^T\right)^{-1}J_k$. Since $\tilde{P}_k$ and $P_k$ are orthogonal projections, we have that $\|\tilde{P}_k\| = \|P_k\| = 1$.

**Lemma 3.4.** *Under Assumptions 3.1 and 3.2 we have both*

$$\|v_k\|_1 \le \delta\|\tilde{c}_k\|_1 \le \delta(\|c_k\|_1 + \epsilon_c) \tag{3.14}$$

$$\|u_k\| \le \frac{1}{\beta_k}(\|P_kg_k\| + \|g_k\|\eta\epsilon_J + \epsilon_g), \tag{3.15}$$

*where $\delta$ is defined in* (3.5) *and*

$$\eta = 1/\gamma. \tag{3.16}$$

*Therefore,*

$$\|d_k\| \le \delta(\|c_k\|_1 + \epsilon_c) + \frac{1}{\beta_k}(\|P_kg_k\| + \|g_k\|\eta\epsilon_J + \epsilon_g). \tag{3.17}$$

*Proof.* The bounds (3.14) follow directly from (3.12), (3.5), and (3.1). By (3.2), we can bound the norm of the tangential component as follows

$$\begin{aligned}\|u_k\| &= \frac{1}{\beta_k}\|\tilde{P}_k\tilde{g}_k\| \\ &\leq \frac{1}{\beta_k}\left(\|P_k g_k\| + \|(\tilde{P}_k - P_k)g_k\| + \|\tilde{P}_k\|\|g_k - \tilde{g}_k\|\right) \\ &\leq \frac{1}{\beta_k}\left(\|P_k g_k\| + \|(\tilde{P}_k - P_k)\|\|g_k\| + \epsilon_g\right). \end{aligned} \tag{3.18}$$

Moreover, by the bounds on perturbed projection matrices [16, Theorems 2.3 and 2.4] we have that

$$\|\tilde{P}_k - P_k\| \leq \frac{\epsilon_J}{\gamma} \equiv \eta\epsilon_J. \tag{3.19}$$

This yields (3.15). □

### 3.2 Penalty Parameter and Model Decrease

We note from (3.8) that in order to obtain a decrease in the true merit function $\phi$, we must ensure that $\ell(x_k; \alpha_k d_k)$ is negative. We will see that this can be achieved for $\alpha_k = 1$ by choosing a sufficiently large penalty parameter $\pi$, and provided noise does not dominate.

**Lemma 3.5.** *If at every iteration $k$ the penalty parameter satisfies*

$$\pi_k \geq \frac{1}{1-\tau}\|(\tilde{J}_k\tilde{J}_k^T)^{-1}\tilde{J}_k\tilde{g}_k\|_\infty, \quad \tau \in (0,1), \tag{3.20}$$

*then*

$$\begin{aligned}\ell(x_k; d_k) \leq &-\frac{1}{\beta_k}g_k^T P_k g_k + \frac{1}{\beta_k}(\|g_k\|^2\eta\epsilon_J + \epsilon_g\|g_k\|) - \tau\pi_k\|c_k\|_1 + \epsilon_g\delta(\|c_k\|_1 + \epsilon_c) \\ &+ \pi_k\left((2-\tau)\epsilon_c + \epsilon_J\left(\delta(\|c_k\|_1 + \epsilon_c) + \frac{1}{\beta_k}(\|P_k g_k\| + \|g_k\|\eta\epsilon_J + \epsilon_g)\right)\right).\end{aligned} \tag{3.21}$$

*Proof.* Since $d_k = -\frac{1}{\beta_k}\tilde{P}_k\tilde{g}_k - \tilde{J}_k^T(\tilde{J}_k\tilde{J}_k^T)^{-1}\tilde{c}_k$, we have from (3.9), (1.5), (3.5), (3.1), (3.2), and the definition of the $\|\cdot\|_{1,2}$ norm in (3.2), that

$$\begin{aligned}\ell(x_k; d_k) &= g_k^T d_k + \pi_k\|c_k + J_k d_k\|_1 - \pi_k\|c_k\|_1 \\ &\leq -\frac{1}{\beta_k}g_k^T\tilde{P}_k\tilde{g}_k - g_k^T\tilde{J}_k^T(\tilde{J}_k\tilde{J}_k^T)^{-1}\tilde{c}_k + \pi_k\|c_k + J_k d_k\|_1 - \pi_k\|c_k\|_1 \\ &\leq -\frac{1}{\beta_k}g_k^T\tilde{P}_k\tilde{g}_k - g_k^T\tilde{J}_k^T(\tilde{J}_k\tilde{J}_k^T)^{-1}\tilde{c}_k + \pi_k\|(c_k - \tilde{c}_k) + (J_k - \tilde{J}_k)d_k)\|_1 - \pi_k\|c_k\|_1 \\ &\leq -\frac{1}{\beta_k}g_k^T\tilde{P}_k\tilde{g}_k - \tilde{g}_k^T\tilde{J}_k^T(\tilde{J}_k\tilde{J}_k^T)^{-1}\tilde{c}_k + \epsilon_g\delta\|\tilde{c}_k\|_1 + \pi_k(\epsilon_c + \epsilon_J\|d_k\|) - \pi_k\|c_k\|_1 \\ &\leq -\frac{1}{\beta_k}g_k^T\tilde{P}_k\tilde{g}_k - \tilde{g}_k^T\tilde{J}_k^T(\tilde{J}_k\tilde{J}_k^T)^{-1}\tilde{c}_k + \epsilon_g\delta(\|c_k\|_1 + \epsilon_c) \\ &\quad + \pi_k\left[\epsilon_c + \epsilon_J\left(\delta(\|c_k\|_1 + \epsilon_c) + \frac{1}{\beta_k}(\|P_k g_k\| + \|g_k\|\eta\epsilon_J + \epsilon_g)\right)\right] - \pi_k\|c_k\|_1,\end{aligned} \tag{3.22}$$

the last line following by (3.17). Next, since $\|\tilde{P}_k\| = 1$ and recalling (3.19), we obtain

$$\begin{aligned}
-g_k^T \tilde{P}_k \tilde{g}_k &\leq -g_k^T P_k g_k + \|g_k\| \|P_k g_k - \tilde{P}_k \tilde{g}_k\| \\
&\leq -g_k^T P_k g_k + \|g_k\| \|P_k g_k - \tilde{P}_k g_k + \tilde{P}_k g_k - \tilde{P}_k \tilde{g}_k\| \\
&\leq -g_k^T P_k g_k + \|g_k\|^2 \|P_k - \tilde{P}_k\| + \|g_k\| \|g_k - \tilde{g}_k\| \\
&\leq -g_k^T P_k g_k + \|g_k\|^2 \eta \epsilon_J + \|g_k\| \epsilon_g.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\ell(x_k; d_k) \leq & -\frac{1}{\beta_k} g_k^T P_k g_k + \frac{1}{\beta_k} (\|g_k\|^2 \eta \epsilon_J + \epsilon_g \|g_k\|) - \tilde{g}_k^T \tilde{J}_k^T (\tilde{J}_k \tilde{J}_k^T)^{-1} \tilde{c}_k + \epsilon_g \delta (\|c_k\|_1 + \epsilon_c) \\
& + \pi_k \left[ \epsilon_c + \epsilon_J \left( \delta (\|c_k\|_1 + \epsilon_c) + \frac{1}{\beta_k} (\|P_k g_k\| + \|g_k\| \eta \epsilon_J + \epsilon_g) \right) \right] - \pi_k \|c_k\|_1.
\end{aligned}$$

Now suppose that we choose the parameter $\pi_k$ so that (3.20) holds. Then

$$-\tilde{g}_k^T \tilde{J}_k^T (\tilde{J}_k \tilde{J}_k^T)^{-1} \tilde{c}_k \leq \|\tilde{g}_k^T \tilde{J}_k^T (\tilde{J}_k \tilde{J}_k^T)^{-1}\|_\infty \|\tilde{c}_k\|_1 \leq (1 - \tau) \pi_k (\|c_k\|_1 + \epsilon_c),$$

and it follows that

$$\begin{aligned}
\ell(x_k; d_k) \leq & -\frac{1}{\beta_k} g_k^T P_k g_k + \frac{1}{\beta_k} (\|g_k\|^2 \eta \epsilon_J + \epsilon_g \|g_k\|) - \tau \pi_k \|c_k\|_1 + \epsilon_g \delta (\|c_k\|_1 + \epsilon_c) \\
& + \pi_k \left[ (2 - \tau) \epsilon_c + \epsilon_J \left( \delta (\|c_k\|_1 + \epsilon_c) + \frac{1}{\beta_k} (\|P_k g_k\| + \|g_k\| \eta \epsilon_J + \epsilon_g) \right) \right].
\end{aligned}$$

□

Lemma 3.5 implies that for any $x_k$ such that the right hand side of (3.21) is negative, we have $\ell(x_k; d_k) < 0$. We now provide conditions under which the decrease in $\ell$ is proportional to the optimality conditions of the nonlinear problem (1.1). Specifically, since $g_k^T P_k g_k = \|P_k g_k\|^2$ is the norm squared of the projected gradient, a combination of $g_k^T P_k g_k$ and $\|c_k\|_1$ can be regarded as a measure of stationarity of the constrained optimization problem. The following result assumes that the optimality measure is not small compared to the errors (or noise).

**Corollary 3.6.** *Choose any $\theta_1 \in [0, 1)$. For any $x_k$ sufficiently far from the solution such that*

$$(1 - \theta_1) \Big( \frac{1}{\beta_k} g_k^T P_k g_k + \tau \pi_k \|c_k\|_1 \Big) \geq E(x_k, \beta_k, \pi_k), \tag{3.23}$$

*where*

$$\begin{aligned}
E(x, \beta, \pi) = & \frac{1}{\beta} (\|g(x)\|^2 \eta \epsilon_J + \epsilon_g \|g(x)\|) + \epsilon_g \delta (\|c(x)\|_1 + \epsilon_c) \\
& + \pi \left[ (2 - \tau) \epsilon_c + \epsilon_J \left( \delta (\|c(x)\|_1 + \epsilon_c) + \frac{1}{\beta} (\|P(x) g(x)\| + \|g(x)\| \eta \epsilon_J + \epsilon_g) \right) \right],
\end{aligned} \tag{3.24}$$

*we have*

$$\ell(x_k; d_k) \leq -\theta_1 \left( \frac{1}{\beta_k} g_k^T P_k g_k + \tau \pi_k \|c_k\|_1 \right). \tag{3.25}$$

*Proof.* For any $\theta_1 \in [0,1)$, we can rewrite (3.21) as

$$\begin{aligned}\ell(x_k;d_k) \leq & -\theta_1(\frac{1}{\beta_k}g_k^T P_k g_k + \tau\pi_k\|c_k\|_1) - (1-\theta_1)(\frac{1}{\beta_k}g_k^T P_k g_k + \tau\pi_k\|c_k\|_1) \\ & + \frac{1}{\beta_k}(\|g_k\|^2\eta\epsilon_J + \epsilon_g\|g_k\|) + \epsilon_g\delta\|(\|c_k\|_1 + \epsilon_c) \\ & + \pi_k\left((2-\tau)\epsilon_c + \epsilon_J\left(\delta(\|c_k\|_1+\epsilon_c) + \frac{1}{\beta_k}(\|P_k g_k\| + \|g_k\|\eta\epsilon_J + \epsilon_g)\right)\right),\end{aligned}$$

from which (3.25) follows by condition (3.23). □

In order to make this result, and similar results to be proved later, more understandable and more convenient to use, we recall that $g(x)^T P(x) g(x) = \|P(x)g(x)\|^2$, and define the function

$$\psi_\pi(x) = \frac{1}{b_u}\|P(x)g(x)\|^2 + \pi\tau\|c(x)\|_1, \tag{3.26}$$

where $b_u$ is given in (3.3). Clearly, $\psi_\pi$ may be viewed as a measure of non-stationarity since $\psi_\pi(x^*) = 0$ when $x^*$ is a stationary point of the problem (1.1). Given this notation we can restate a slightly weaker version of Corollary 3.6.

**Corollary 3.7.** *Choose any $\theta_1 \in [0,1)$. For any $x_k$ sufficiently far from the solution such that*

$$\psi_{\pi_k}(x_k) \geq E(x_k, \beta_k, \pi_k)/(1-\theta_1), \tag{3.27}$$

*we have*

$$\ell(x_k;d_k) \leq -\theta_1\left(\frac{1}{\beta_k}g_k^T P_k g_k + \tau\pi_k\|c_k\|_1\right) \leq -\theta_1\psi_{\pi_k}(x_k). \tag{3.28}$$

*Proof.* The result follows from the fact that

$$\psi_{\pi_k}(x_k) \leq \big(\frac{1}{\beta_k}g_k^T P_k g_k + \tau\pi_k\|c_k\|_1\big).$$

□

### 3.3 Line search

Since $\pi_k$ is defined by (2.3) and (2.2), and by Assumptions 3.3, we have that $\{\|\hat{\lambda}_k\|\}$ is bounded. Moreover, since $\{\pi_k\}$ is monotone and since $\pi_k - \pi_{k-1}$ is either zero or greater than $\pi_{k-1}$, there exists values $k_0$ and $\bar{\pi}$ such that:

$$\pi_k = \bar{\pi}, \quad \forall k \geq k_0, \tag{3.29}$$

and (3.20) is satisfied. The rest of the analysis assumes that the penalty parameter has attained that fixed value $\bar{\pi}$. Thus, for the rest of the section

$$\tilde{\phi}(x_k) \equiv \tilde{f}(x_k) + \bar{\pi}\|\tilde{c}(x_k)\|_1, \qquad \phi(x_k) \equiv f(x_k) + \bar{\pi}\|c(x_k)\|_1, \ \forall k \geq k_0. \tag{3.30}$$

Algorithm 1 sets $x_{k+1} = x_k + \alpha_k d_k$, where $\alpha_k$ is chosen by repeated halving until the *relaxed Armijo condition* is satisfied:

$$\tilde{\phi}(x_k + \alpha_k d_k) \leq \tilde{\phi}(x_k) + \nu \alpha_k \tilde{\ell}(x_k; d_k) + \epsilon_R,$$

for some constants $\nu \in (0, 1)$ and $\epsilon_R \geq 2(\epsilon_f + \bar{\pi}\epsilon_c)$, where $\tilde{\ell}(x_k; d_k)$ is defined in (2.1). In other words, we require that the decrease in the noisy merit function be a fraction of the decrease of the noisy first-order model $\tilde{\ell}$, plus a relaxation term.

To ensure that the line search yields significant progress toward a solution, we need to show that $\alpha_k$ is bounded away from zero and that $\tilde{\ell}(x_k; d_k)$ is sufficiently negative. To do so, we recall that we have established in (3.28) that the noiseless first-order model $\ell(x_k; d)_k$ is sufficiently negative when condition (3.27) is satisfied. To relate $\ell(x_k; d_k)$ to $\tilde{\ell}(x_k; d_k)$, we recall (3.9) and (2.1), and measure the difference between these two quantities. By (3.17)

$$\begin{aligned}
|\tilde{\ell}(x_k; d_k) - \ell(x_k; d_k)| &\leq \epsilon_g \|d_k\| + 2\bar{\pi}\epsilon_c + \bar{\pi}\epsilon_J \|d_k\| \\
&\leq (\epsilon_g + \bar{\pi}\epsilon_J) \left( \delta(\|c_k\|_1 + \epsilon_c) + \frac{1}{\beta_k}(\|P_k g_k\| + \|g_k\|\eta\epsilon_J + \epsilon_g) \right) + 2\bar{\pi}\epsilon_c \qquad (3.31) \\
&\leq (\epsilon_g + \bar{\pi}\epsilon_J) \left( \delta(C_c + \epsilon_c) + \frac{1}{b_l}(C_g + C_g\eta\epsilon_J + \epsilon_g) \right) + 2\bar{\pi}\epsilon_c \\
&\equiv \epsilon_\ell, \qquad (3.32)
\end{aligned}$$

where $C_g, C_c$ are constants such that

$$\|g(x_k)\| \leq C_g, \qquad \|c(x_k)\|_1 \leq C_c \quad \forall k > k_0. \qquad (3.33)$$

We know that these constants exist because of Assumption 3.3. We now describe conditions under which one can characterize the size of the steplength $\alpha_k$. Let

$$L = L_f + \bar{\pi}\|L_c\|_1, \qquad (3.34)$$

where $L_f, L_c$ are defined in Assumptions 3.1.

**Theorem 3.8.** *Let $\theta_1$ be defined as in Corollary 3.6, choose constants $\theta_2 < \theta_1$, $\nu \in (0, 1)$ and*

$$\epsilon_R \geq 2(\epsilon_f + \bar{\pi}\epsilon_c) \equiv 2\epsilon_\phi. \qquad (3.35)$$

*Then, for all iterates $x_k$ with $k \geq k_0$ that satisfy both (3.27) and*

$$(1 - \nu)(\theta_1 - \theta_2) \left( \frac{1}{\beta_k}\|P_k g_k\|^2 + \bar{\pi}\tau\|c_k\|_1 \right) > 2\nu\epsilon_\ell, \qquad (3.36)$$

*if the steplength satisfies*

$$\alpha_k < \frac{(1 - \nu)\theta_2 \left( \frac{1}{\beta_k}\|P_k g_k\|^2 + \bar{\pi}\tau\|c_k\|_1 \right)}{\frac{L}{2}[\delta^2(\|c_k\|_1 + \epsilon_c)^2 + \frac{1}{\beta_k^2}(\|P_k g_k\| + \|g_k\|\eta\epsilon_J + \epsilon_g)^2]} \equiv \hat{\alpha}_k, \qquad (3.37)$$

*then*

$$\tilde{\phi}(x_k + \alpha_k d_k) \leq \tilde{\phi}(x_k) + \nu\alpha_k \tilde{\ell}(x_k; d_k) + \epsilon_R. \qquad (3.38)$$

*Proof.* By the definition (3.35) of $\epsilon_\phi$, (3.8), (3.34), the convexity of $\ell(x_k;\cdot)$, (3.32), (3.28), the fact that $P_k^2 = P_k$, and (3.17), we get

$$\begin{aligned}
\tilde{\phi}(x_k+\alpha d_k)-\tilde{\phi}(x_k) &\le \phi(x_k+\alpha d_k)-\phi(x_k)+2\epsilon_\phi \\
&\le \ell(x_k;\alpha d_k)+2\epsilon_\phi+\tfrac{L}{2}\alpha^2\|d_k\|^2 \\
&\le \alpha\ell(x_k;d_k)+2\epsilon_\phi+\tfrac{L}{2}\alpha^2\|d_k\|^2 \\
&= \nu\alpha\ell(x_k;d_k)+2\epsilon_\phi+(1-\nu)\alpha\ell(x_k;d_k)+\tfrac{L}{2}\alpha^2\|d_k\|^2 \\
&\le \nu\alpha\tilde{\ell}(x_k;d_k)+\nu\alpha\epsilon_\ell+2\epsilon_\phi+(1-\nu)\alpha\ell(x_k;d_k)+\tfrac{L}{2}\alpha^2\|d_k\|^2 \\
&\le \nu\alpha\tilde{\ell}(x_k;d_k)+2\epsilon_\phi+2\nu\alpha\epsilon_\ell-(1-\nu)\theta_1\alpha\Big(\frac{1}{\beta_k}g_k^TP_kg_k+\bar{\pi}\tau\|c_k\|_1\Big) \\
&\quad+\tfrac{L}{2}\alpha^2\|d_k\|^2 \\
&\le \nu\alpha\tilde{\ell}(x_k;d_k)+2\epsilon_\phi+2\nu\alpha\epsilon_\ell-(1-\nu)\theta_1\alpha\Big(\frac{1}{\beta_k}\|P_kg_k\|^2+\bar{\pi}\tau\|c_k\|_1\Big) \\
&\quad+\tfrac{L}{2}\alpha^2\Big[\delta^2(\|c_k\|_1+\epsilon_c)^2+\frac{1}{\beta_k^2}\big(\|P_kg_k\|+\|g_k\|\eta\epsilon_J+\epsilon_g\big)^2\Big],
\end{aligned}$$

the last line following from the orthogonality of the components (3.11) of $d_k$.

Now we choose a constant $\theta_2 < \theta_1$, and consider iterates $x_k$ such that (3.36) holds. For such iterates we have,

$$\begin{aligned}
\tilde{\phi}(x_k+\alpha d_k)-\tilde{\phi}(x_k) \le{}& \nu\alpha\tilde{\ell}(x_k;d_k)+2\epsilon_\phi-(1-\nu)\theta_2\alpha\Big(\frac{1}{\beta_k}\|P_kg_k\|^2+\bar{\pi}\tau\|c_k\|_1\Big) \\
&+\tfrac{L}{2}\alpha^2\left[\delta^2(\|c_k\|_1+\epsilon_c)^2+\frac{1}{\beta_k^2}\Big(\|P_kg_k\|+\|g_k\|\eta\epsilon_J+\epsilon_g\Big)^2\right].
\end{aligned}$$

Then, for any steplength satisfying (3.37) where $x_k$ satisfies the (3.23) and (3.36), we have

$$\tilde{\phi}(x_k+\alpha d_k)-\tilde{\phi}(x_k) \le \nu\alpha\tilde{l}(x_k;d_k)+2\epsilon_\phi,$$

and thus (3.38) holds since $\epsilon_R \ge 2\epsilon_\phi$. □

Note that condition (3.36) is implied by the slightly weaker inequality

$$(1-\nu)(\theta_1-\theta_2)\psi_{\bar{\pi}}(x_k) > 2\nu\epsilon_\ell. \tag{3.39}$$

Since the numerator in (3.37) is bounded away from zero by (3.36), and the denominator is bounded above given the assumed global upper bounds on $c_k, g_k$, and lower bound on $\beta_k$ stated in Assumptions 3.3, it follows that there is a constant $\bar{\alpha}$ such that $\hat{\alpha}_k > 2\bar{\alpha}$ for all $k \ge k_0$. The algorithm employs a backtracking line search that halves each trial step, hence we can conclude that

$$\bar{\alpha} \le \alpha_k, \qquad \text{for } k \ge k_0. \tag{3.40}$$

This will allow us to show that, when the conditions in Theorem 3.8 are satisfied, the algorithm will make non-negligible progress.

### 3.4 The Main Convergence Result

Now we show that Algorithm 1 will eventually generate iterates close to a stationary point of the problem, as measured by the function $\psi_{\bar{\pi}}(x)$ defined in (3.26). To do so, we note that condition (3.27) implies that the linear model decrease $\ell$ is sufficiently negative in the sense of (3.28), and we have established a bound in (3.32) for the distance between $\ell$ and $\tilde{\ell}$. Furthermore, we have shown that condition (3.39) ensures that the relaxed Armijo condition (3.38) is satisfied for steplengths $\alpha_k$ that are bounded away from zero. Those two conditions—(3.27), (3.39)—are necessary to ensure that the algorithm makes significant progress, but they are not sufficient. To control the effect of noise in testing (2.4) as well as the effect of the relaxation factor, we impose one additional condition,

$$\psi_{\bar{\pi}}(x_k) \geq \frac{2\epsilon_R + 4\epsilon_\phi}{\nu\bar{\alpha}\theta_2}, \tag{3.41}$$

to help define the region where Algorithm 1 progresses toward stationarity.

One more refinement is needed. The definition of the term $E(x_k, \beta_k, \pi_k)$ defined in (3.24) involves $c(x_k)$ and $g(x_k)$, which makes the region defined by (3.27) difficult to interpret. Therefore, we compute an upper bound for $E$. If we define

$$\begin{aligned}\mathcal{E} =& \frac{1}{b_l}(C_g^2\eta\epsilon_J + \epsilon_g C_g) + \epsilon_g\delta(C_c + \epsilon_c) \\ &+\bar{\pi}\left[(2-\tau)\epsilon_c + \epsilon_J\left(\delta(C_c+\epsilon_c) + \frac{1}{b_l}(C_g + C_g\eta\epsilon_J + \epsilon_g)\right)\right],\end{aligned} \tag{3.42}$$

where $C_g, C_c$ are given in (3.33), then we have that $E(x_k, \beta_k, \bar{\pi}) \leq \mathcal{E}$ for all $k \geq k_0$. We can thus state a condition that implies (3.27):

$$\psi_{\bar{\pi}}(x_k) \geq \frac{\mathcal{E}}{(1-\theta_1)}, \qquad \forall k \geq k_0. \tag{3.43}$$

In summary, the analysis presented above holds if conditions (3.43), (3.39) are satisfied and we also impose condition (3.41). This allows us to characterize a region, which we denote by $C_1$, where errors dominate and improvement in the merit function $\phi$ cannot be guaranteed. In other words, $C_1$ is the region where at least one of the three conditions—(3.43), (3.39), (3.41)—is not satisfied.

**Definition 3.9.** *The critical region $C_1$ is defined as the set of $x \in \mathbb{R}^n$ satisfying*

$$\psi_{\bar{\pi}}(x) \leq \max\left\{\frac{\mathcal{E}}{(1-\theta_1)}, \frac{2\nu\epsilon_\ell}{(1-\nu)(\theta_1-\theta_2)}, \frac{2\epsilon_R + 4\epsilon_\phi}{\nu\bar{\alpha}\theta_2}\right\}, \tag{3.44}$$

*where $\mathcal{E}$ and $\epsilon_\ell$ are defined by* (3.42) *and* (3.32)*, respectively, and $\theta_1, \theta_2$ are constants such that $0 < \theta_2 < \theta_1 < 1$.*

We also define the following set.

**Definition 3.10.** *Let $w = \sup\{\phi(x) : x \in C_1\}$, and define the level set*

$$C_2 = \{x : \phi(x) \leq w + 2\epsilon_\phi + \epsilon_R\}.$$

Note that by construction $C_1 \subseteq C_2$. We are now ready to state the main convergence result.

**Theorem 3.11.** *Suppose that Algorithm 1 generates a sequence $\{x_k\}$ from $x_0$ satisfying Assumptions 3.1-3.3. There is an iteration $k_1$ at which $\{x_k\}$ enters the critical region $C_1$, and for all $k > k_1$ the iterates remain in the critical level set $C_2$. The iterates may leave $C_1$, but there must be infinitely many iterates in $C_1$.*

*Proof.* Recall that the index $k_0$ is defined in (3.29). If $k \notin C_1$ and $k \geq k_0$, then the assumptions of Theorem 3.8 are satisfied and (3.38) holds. Therefore, by (3.40), (3.32), (3.25), (3.28), (3.26), (3.36)

$$\begin{aligned}
\phi(x_k + \alpha_k d_k) - \phi(x_k) &\leq \tilde{\phi}(x_k + \alpha_k d_k) - \tilde{\phi}(x_k) + 2\epsilon_\phi && (3.45)\\
&\leq \nu\bar{\alpha}\tilde{\ell}_\phi(x_k; d_k) + 2\epsilon_\phi + \epsilon_R && (3.46)\\
&\leq \nu\bar{\alpha}\ell_\phi(x_k; d_k) + \nu\bar{\alpha}\epsilon_\ell + 2\epsilon_\phi + \epsilon_R \\
&\leq -\nu\bar{\alpha}\theta_1\left(\frac{1}{\beta_k} g_k^T P_k g_k + \tau\bar{\pi}\|c_k\|_1\right) + \nu\bar{\alpha}\epsilon_\ell + 2\epsilon_\phi + \epsilon_R \\
&\leq -\nu\bar{\alpha}\theta_1\psi_{\bar{\pi}}(x_k) + \nu\bar{\alpha}\epsilon_\ell + 2\epsilon_\phi + \epsilon_R \\
&= -[\nu\bar{\alpha}\theta_2 + \hat{\alpha}\nu(\theta_1 - \theta_2)]\psi_{\bar{\pi}}(x_k) + \nu\bar{\alpha}\epsilon_\ell + 2\epsilon_\phi + \epsilon_R \\
&\leq -\nu\bar{\alpha}\theta_2\psi_{\bar{\pi}}(x_k) + 2\epsilon_\phi + \epsilon_R. && (3.47)
\end{aligned}$$

Combining this bound with (3.41), we have that if $x_k \notin C_1$ then

$$\phi(x_{k+1}) - \phi(x_k) \leq -\frac{\nu\bar{\alpha}\theta_2}{2}\psi_{\bar{\pi}}(x_k). \tag{3.48}$$

Since the sequence $\{\phi(x_k)\}$ is bounded below by Assumptions 3.3, $\psi_{\bar{\pi}}(x_k)$ converges to zero and thus it follows that Algorithm 1 eventually generates an iterate in $C_1$.

Now if $x_k \in C_1$, then by Step 6 in Algorithm 1, $\phi(x_{k+1}) \leq \phi(x_k) + 2\epsilon_\phi + \epsilon_R \leq w + 2\epsilon_\phi + \epsilon_R$, so that $x_{k+1} \in C_2$.

On the other hand, if $x_k \in C_2 \setminus C_1$, then by (3.48)

$$\phi(x_{k+1}) - \phi(x_k) \leq 0,$$

which implies $x_{k+1} \in C_2$. Thus the rest of the sequence lies in $C_2$, with infinitely many iterates in $C_1$. □

We should note that since we are not assuming that the objective function is strongly convex or satisfies a quadratic growth condition, it is possible that the supremum in Definition 3.10 is $w = \infty$. This is, however, an unlikely scenario.

### 3.5 Discussion

Let us take a closer look the main result of this paper, Theorem 3.11, since the critical region $C_1$ defined in (3.44) is complex.

By the definitions (3.42) and (3.32), we have that $\mathcal{E}$ and $\epsilon_\ell$ are both of order $O(\epsilon_c, \epsilon_g, \epsilon_J)$, and so is the right hand side in (3.44). This is as desired. The constants in these orders of magnitude matter, so we must characterize them.

First note that the critical region $C_1$, the set $C_2$ and $\bar{\pi}$ depend on the starting point $x_0$. It is then possible that $\bar{\pi}$ could be very large in some cases, although in practice this does not seem to be a major concern. The constants $C_g, C_c$, which also enter in the definition of $\mathcal{E}$ and $\epsilon_\ell$ could be quite large. One can, however, give a tighter definition of $C_1$ by not introducing these constants. In this case, we would define $\epsilon_\ell$ by (3.31) and employ (3.27), rather than (3.43). This makes the main theorem more precise, albeit more difficult to interpret.

Returning to the constants in (3.42) and (3.32), we have that

$$\epsilon_\ell, \mathcal{E} \sim \left[\delta, \frac{1}{b_l}, \frac{\eta}{b_l}\right],$$

and from (3.4), (3.5), (3.16) we observe that

$$\sigma_{min}(\tilde{J}_k) \geq \gamma, \qquad \delta = \frac{1}{\gamma - \epsilon_J} \geq \frac{1}{\sigma_{min}(\tilde{J}_k) - \epsilon_J}, \qquad \text{and } \eta = \frac{1}{\gamma} = \frac{1}{\sigma_{min}(\tilde{J}_k)}.$$

The effect of a near rank-deficient Jacobian and Hessian approximations $\beta_k I$ are now apparent.

It is interesting to compare $C_1$ with the region obtained by Berahas et al. [1] for unconstrained strongly convex optimization. When constraints are not present, i.e., $m = 0$, conditions (3.23) and (3.36) defining $C_1$ reduce to requirements of form $\|g_k\| \geq c_1 \epsilon_g$ and $\|g_k\|^2 \geq c_2(\epsilon_g \|g_k\| + \epsilon_g^2)$ for some constants $c_1$ and $c_2$, respectively. That corresponds to *Case 1* in the analysis of [1], in which case $\epsilon_g$ is small as compared to $\|g_k\|$ by some factor $\beta \in (0, 1)$, so that the line search ensures an improvement in the exact objective function – $f(x)$ in our notation. Similar to the setting in this paper, [1] employs a relaxed line search which does not fail even in the critical region; that is, when $\|g_k\| \leq \beta \epsilon_g$. Their analysis then provides a level set that the iterates cannot leave, which depends on the relaxation term $\epsilon_R$ as well as $\epsilon_\phi$ (i.e. $\epsilon_f$ in the unconstrained case) as in the definition of $C_2$ in our analysis. Since strong convexity is assumed in [1], they can define this level set in terms of a strong convexity parameter rather than a bound such as $w$ in Definition 3.10.

## 4 Numerical Experiments

We implemented Algorithm 1 in Python. We set $\nu = 0.1$, $\tau = 0.9$, and $\beta_k = 50$, for all $k$. The purpose of the numerical experiments is to supplement the theoretical results, which are stated in terms of the merit function $\phi$, by reporting the distance to the solution $\|x_k - x^*\|$ as the iteration progresses. In order to gain an idea of this behavior, it suffices to test only

| problem | classification | objective | constraints |
|---|---|---|---|
| HS7 | OOR2-AN-2-1 | $\ln(1+x_1^2) - x_2$ | $(1+x_1^2)^2 + x_2^2 = 4$ |
| BT11 | OOR2-AN-4-3 | $-x_1x_2x_3x_4$ | $x_1^3 + x_2^2 = 1$<br>$x_1^2x_4 - x_3 = 0$<br>$x_4^2 - x_2 = 0$ |
| HS40 | OOR2-AY-5-3 | $(x_1-1)^2 + (x_1-x_2)^2 + (x_2-x_3)^2$<br>$+(x_3-x_4)^4 + (x_4-x_5)^4$ | $x_1 + x_2^2 + x_3^3 = -2+\sqrt{18}$<br>$x_2 + x_4 + x_3^2 = -2+\sqrt{8}$<br>$x_1 - x_5 = 2$ |

a few examples. We selected the following three small-scale equality-constrained problems from the CUTEst set [9].

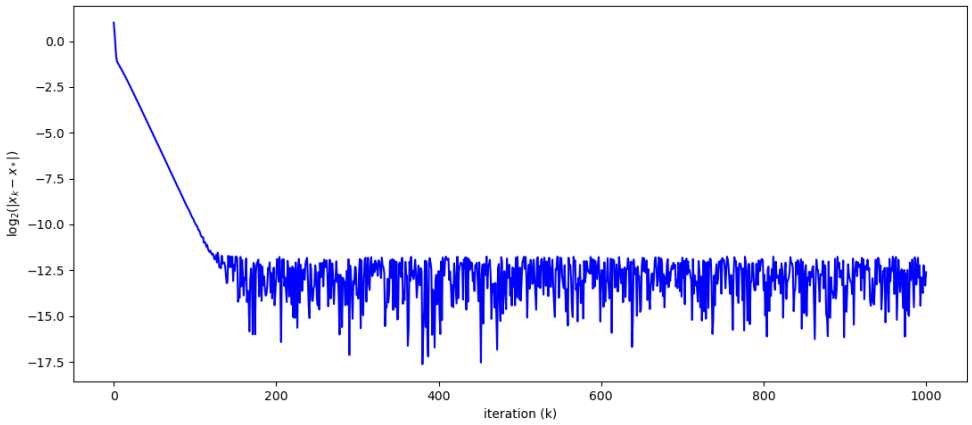
We add uniformly distributed random noise to the exact function values and to each component of the exact gradients; i.e., for $\xi_i \sim \mathcal{U}(-\epsilon_1, \epsilon_1)$, and $\psi_{ij} \sim \mathcal{U}(-\epsilon_2, \epsilon_2)$ we set

$$\tilde{f}(x) = f(x) + \xi_0, \qquad \tilde{c}_i(x) = c_i(x) + \xi_i$$
$$\tilde{g}_i(x) = g_i(x) + \psi_{0j}, \qquad \tilde{J}_{ij}(x) = J_{ij}(x) + \psi_{ij}.$$

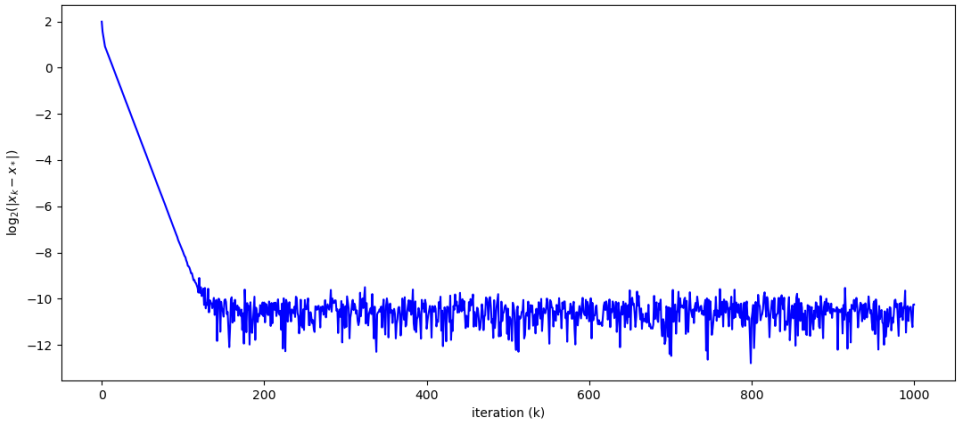
In our tests, we vary $\epsilon_1, \epsilon_2$, and report $\|x_k - x^*\|$, where $x^*$ is a locally optimal solution obtained by using exact gradients in the algorithm. For each of these problems, $x^*$ is a a nondegenerate stationary point.

**Asymptotic Behavior.** In Figure 4.1, we plot $\|x_k - x^*\|$ for 1000 iterations, for $\epsilon_1 = \epsilon_2 = 10^{-3}$ in the definitions of $\xi_i$, and $\psi_{ij}$ We also display the values of $\epsilon_f, \epsilon_c, \epsilon_g, \epsilon_J$ defined in (3.1)-(3.2) We should note that in each of the runs the penalty parameter $\pi_k$ became fixed within the first 15 iterations. We observe that $\{\|x_k - x^*\|\}$ is contained in a band whose upper bound is frequently visited by the algorithm, whereas the lower bound is defined by large irregular spikes. These results suggest that if one desires the highest accuracy in the solution, the algorithm should continue beyond the point where oscillations in the merit function occur, since there is little risk that the iterates will stray away from the neighborhood of the solution, and there is a chance that significantly higher accuracy is achieved at some iterates.

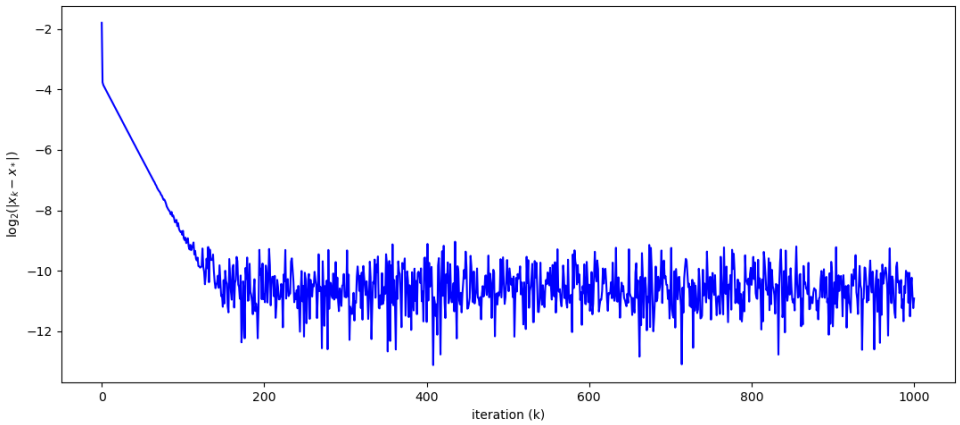
Figure 4.1: Distance to optimality ($\log_2(\|x_k - x^*\|)$) vs iteration number for $\epsilon_1 = \epsilon_2 = 10^{-3}$




(a) HS7. $\epsilon_f = 10^{-3}, \epsilon_c = 10^{-3}, \epsilon_g = 1.41 \times 10^{-3}, \epsilon_J = 1.41 \times 10^{-3}$




(b) BT11. $\epsilon_f = 10^{-3}, \epsilon_c = 3 \times 10^{-3}, \epsilon_g = 2.24 \times 10^{-3}, \epsilon_J = 6.71 \times 10^{-3}$




(c) HS40. $\epsilon_f = 10^{-3}, \epsilon_c = 3 \times 10^{-3}, \epsilon_g = 2 \times 10^{-3}, \epsilon_J = 6 \times 10^{-3}$

**Benefits of the relaxed line search.** The only unconventional part of Algorithm 1 is the relaxed line search (2.4). To observe the effect of the relaxation, we solved the test problems with and without it; the results are reported in Tables 4.1–4.3. We observe that when the relaxation is disabled, the line search often fails in a neighborhood of $x^*$ (we terminate the algorithm as soon as there is a line search failure). When the relaxation is enabled, the line search is always successful. In this case, we let the algorithm run for 100, 500, and 1000 iterations. It is apparent that the relaxed line search allows the algorithm to continue iterating past the point where the traditional line search would fail, yielding much better accuracy in the solution.

Table 4.1: $\min_k\{\|x_k - x^*\|\}$ when $\epsilon_1 = \epsilon_2 = 10^{-5}$

| | relaxation disabled | | relaxation enabled | | |
|---|---|---|---|---|---|
| problem | iter. of failure | $\min_k\{\|x_k - x^*\|\}$ | $k_{\max} = 100$ | $k_{\max} = 500$ | $k_{\max} = 1000$ |
| HS7 | 77 | 7.8260E-3 | 1.0234E-3 | 4.9413E-8 | 4.9413E-8 |
| BT11 | 64 | 4.8346E-2 | 3.9258E-3 | 1.9791E-6 | 1.4133E-6 |
| HS40 | 26 | 3.4728E-2 | 2.1251E-3 | 1.09888E-6 | 1.0988E-6 |

Table 4.2: $\min_k\{\|x_k - x^*\|\}$ when $\epsilon_1 = \epsilon_2 = 10^{-3}$

| | relaxation disabled | | relaxation enabled | | |
|---|---|---|---|---|---|
| problem | iter. of failure | $\min_k\{\|x_k - x^*\|\}$ | $k_{\max} = 100$ | $k_{\max} = 500$ | $k_{\max} = 1000$ |
| HS7 | 42 | 8.0390E-2 | 1.0401E-3 | 4.9328E-6 | 4.9328E-6 |
| BT11 | 18 | 9.3324E-1 | 4.0003E-3 | 1.9804E-4 | 1.4060E-4 |
| HS40 | 6 | 6.4144E-2 | 2.2293E-3 | 1.1183E-4 | 4.9328E-6 |

Table 4.3: $\min_k\{\|x_k - x^*\|\}$ when $\epsilon_1 = \epsilon_2 = 10^{-1}$

| | relaxation disabled | | relaxation enabled | | |
|---|---|---|---|---|---|
| problem | iter. of failure | $\min_k\{\|x_k - x^*\|\}$ | $k_{\max} = 100$ | $k_{\max} = 500$ | $k_{\max} = 1000$ |
| HS7 | 10 | 3.7404E-1 | 1.3113E-3 | 4.5607E-4 | 2.5422E-4 |
| BT11 | 8 | 1.7108 | 2.0598E-2 | 2.0598E-2 | 1.9451E-2 |
| HS40 | 2 | 1.1817E-1 | 5.8202E-2 | 3.8673E-2 | 3.8673E-2 |

**Effect of incorrect noise level estimations.** In Algorithm 1, estimations of $\epsilon_f$ and $\epsilon_c$ are needed to set the relaxation bound $\epsilon_R$ in (2.4). It is clear that underestimating the noise level can cause failure of the relaxed line search, which never fails when the true level (or an overestimation) is provided. On the other hand, overestimation can lead to large oscillations. The precise behavior of the algorithm will depend on the stop test, and there is no universally adopted stopping criterion in the noisy setting, to our knowledge.

Nevertheless, we performed the following experiments using a stop test that that could be considered as a naive modification of termination tests in standard packages. We simply

terminate the algorithm when the observed (noisy) feasibility and optimality errors are smaller than the (estimated) noise provided for these quantities, i.e.,

$$\|\tilde{c}(x_k)\|_1 \leq \epsilon_c^{est} \quad \text{and} \quad \|\tilde{g}(x_k) + \tilde{J}(x_k)^T \lambda_k\| \leq \epsilon_g^{est} + \|\lambda_k\|_\infty \epsilon_J^{est}. \tag{4.1}$$

Figures 4.6–4.4 report the quantity $\min_k\{\|x_k - x^*\|\}$ when the algorithm employs estimated noise levels $\epsilon_1^{est}$ and $\epsilon_2^{est}$ that are 10, 100 and 1000 times larger or smaller than the correct values. We perform this experiment for $\epsilon_i = 10^{-1}, 10^{-3}, 10^{-5}$. A termination due to the satisfaction of the condition (4.1) is marked with *(opt)*, and a line search failure is marked with *(ls)*.

Table 4.4: $\min_k\{\|x_k - x^*\|\}$ when true $\epsilon_i = 10^{-5}$; $i = 1, 2$

| | $\epsilon_i^{est} = \epsilon_i$ | | $\epsilon_i^{est} = 0.001\epsilon_i$ | | $\epsilon_i^{est} = 1000\epsilon_i$ | |
|---|---|---|---|---|---|---|
| problem | iter. | $\min_k\{\|x_k - x^*\|\}$ | iter. | $\min_k\{\|x_k - x^*\|\}$ | iter. | $\min_k\{\|x_k - x^*\|\}$ |
| HS7 | 188 (opt) | 3.2017E-6 | 69 (ls) | 8.8000E-3 | 74 (opt) | 5.5704E-3 |
| BT11 | 233 (opt) | 2.4010E-6 | 64 (ls) | 4.8346E-2 | 64 (opt) | 4.0112E-2 |
| HS40 | 2703 (opt) | 8.0766E-7 | 26 (ls) | 3.4728E-2 | 27 (opt) | 2.8305E-2 |

Table 4.5: $\min_k\{\|x_k - x^*\|\}$ when true $\epsilon_i = 10^{-3}$; $i = 1, 2$

| | $\epsilon_i^{est} = \epsilon_i$ | | $\epsilon_i^{est} = 0.01\epsilon_i$ | | $\epsilon_i^{est} = 100\epsilon_i$ | |
|---|---|---|---|---|---|---|
| problem | iter. | $\min_k\{\|x_k - x^*\|\}$ | iter. | $\min_k\{\|x_k - x^*\|\}$ | iter. | $\min_k\{\|x_k - x^*\|\}$ |
| HS7 | 117 (opt) | 3.5750E-4 | 42 (ls) | 8.0390E-2 | 39 (opt) | 5.4414E-2 |
| BT11 | 149 (opt) | 2.7466E-4 | 29 (ls) | 5.3925E-1 | 22 (opt) | 5.9597E-1 |
| HS40 | 154 (opt) | 4.2653E-4 | 7 (ls) | 6.4142E-2 | 2 (opt) | 6.9002E-2 |

Table 4.6: $\min_k\{\|x_k - x^*\|\}$ when true $\epsilon_i = 10^{-1}$; $i = 1, 2$

| | $\epsilon_i^{est} = \epsilon_i$ | | $\epsilon_i^{est} = 0.1\epsilon_i$ | | $\epsilon_i^{est} = 10\epsilon_i$ | |
|---|---|---|---|---|---|---|
| problem | iter. | $\min_k\{\|x_k - x^*\|\}$ | iter. | $\min_k\{\|x_k - x^*\|\}$ | iter. | $\min_k\{\|x_k - x^*\|\}$ |
| HS7 | 51 (opt) | 2.6752E-2 | 556 (ls) | 2.5682E-4 | 5 (opt) | 4.2796E-1 |
| BT11 | 20 (opt) | 6.6650E-1 | 3233 (ls) | 6.8738E-3 | 2 (opt) | 2.4045 |
| HS40 | 210 (opt) | 5.82E-2 | 982 (ls) | 1.4785E-2 | 0 (opt) | 2.8877E-1 |

As expected, underestimations cause line search failures while overestimations cause (4.1) to be triggered at earlier iterations. Another consequence of underestimating $\epsilon_2$ is that the algorithm might never be able to satisfy (4.1), even if a line search failure occurs sufficiently late in the run; see for example the entry corresponding to $\epsilon_i = 10^{-1}, \epsilon_i^{est} = 0.1\epsilon_i$. In summary, over-and underestimation of the noise levels can be harmful in ways that are dependent on the implementation.

We must point out that an optimization algorithm may provide an indication that the noise estimates must be re-computed. For example, the recovery procedure described by

Berahas et al. [1] uses information from the line search to request a better estimate (e.g. through sampling or finite difference tables), and can take precautions to avoid harmful iterations. Robust implementations of methods for constrained optimization in the presence of noise should include such features.

# 5 Final Remarks

Two questions guided this research. What is the best behavior one can expect of a constrained optimization method when functions and constraints contain a moderate amount of bounded noise that cannot be diminished at will? What are the minimal modifications of a classical optimization algorithm that allow it to tolerate noise, when the noise level can be estimated?

In this paper, we focused on a classical sequential quadratic programming method applied to equality constrained problems. We showed that a modification (relaxation) of the line search allows the iterates to approach a region around the solution where noise dominates—and that the iterates remain in a vicinity of this region, under normal circumstances. The analysis is presented under benign assumptions, for example that the Jacobian of the constraints is never close to singular, which facilitates the choice of the penalty parameter. Nevertheless, we believe that the essence of the analysis captures some of the main challenges to be confronted when functions and derivatives contain noise. The accuracy bounds presented in this paper will be sharpened in a forthcoming paper that studies the local behavior of the method near a well behaved minimizer.

The thorny issue of how to design a proper stop test that reflects the desires of the users has not been addressed in this paper and is worthy of research. The treatment of singularity and the use of a nondiagonal Hessian also requires attention, as well as the very important question of how to handle noisy inequality constraints.

*Acknowledgement.* We thank Shigeng Sun for his careful reading of the paper and useful suggestions.

## 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, Frank E Curtis, Michael J O'Neill, and Daniel P Robinson. A stochastic sequential quadratic optimization algorithm for nonlinear equality constrained optimization with rank-deficient jacobians. *arXiv preprint arXiv:2106.13015*, 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] Dimitri P Bertsekas. *Convex Optimization Algorithms*. Athena Scientific, 2015.

[5] R. H. Byrd, J. Nocedal, and R.A. Waltz. KNITRO: An integrated package for nonlinear optimization. In G. di Pillo and M. Roma, editors, *Large-Scale Nonlinear Optimization*, pages 35–59. Springer, 2006.

[6] 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.

[7] R. Fletcher. *Practical Methods of Optimization*. Wiley, second edition, 1987.

[8] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, second edition, 1989.

[9] Nicholas IM Gould, Dominique Orban, and Philippe L Toint. CUTEst: a constrained and unconstrained testing environment with safe threads for mathematical optimization. *Computational Optimization and Applications*, 60(3):545–557, 2015.

[10] M Hintermüller. Solving nonlinear programming problems with noisy function values and noisy gradients. *Journal of optimization theory and applications*, 114(1):133–169, 2002.

[11] Jorge J Moré and Stefan M Wild. Estimating derivatives of noisy simulations. *ACM Transactions on Mathematical Software (TOMS)*, 38(3):19, 2012.

[12] Jorge Nocedal and Stephen Wright. *Numerical Optimization*. Springer New York, 2 edition, 1999.

[13] BT Poljak. Nonlinear programming methods in the presence of noise. *Mathematical programming*, 14(1):87–97, 1978.

[14] K Schittkowski. Nlpqlp-nonlinear programming with non-monotone and distributed line search, 2014.

[15] Hao-Jun Michael Shi, Melody Qiming Xuan, Figen Oztoprak, and Jorge Nocedal. On the numerical performance of derivative-free optimization methods based on finite-difference approximations. *arXiv preprint arXiv:2102.09762*, 2021.

[16] Gilbert W Stewart. On the perturbation of pseudo-inverses, projections and linear least squares problems. *SIAM review*, 19(4):634–662, 1977.