SIAM J. Optim.
Vol. 31, No. 2, pp. 1352–1379

SEQUENTIAL QUADRATIC OPTIMIZATION FOR NONLINEAR EQUALITY CONSTRAINED STOCHASTIC OPTIMIZATION[*]

ALBERT S. BERAHAS[†], FRANK E. CURTIS[‡], DANIEL ROBINSON[‡], AND BAOYU ZHOU[‡]

Abstract. Sequential quadratic optimization algorithms are proposed for solving smooth nonlinear optimization problems with equality constraints. The main focus is an algorithm proposed for the case when the constraint functions are deterministic, and constraint function and derivative values can be computed explicitly, but the objective function is stochastic. It is assumed in this setting that it is intractable to compute objective function and derivative values explicitly, although one can compute stochastic function and gradient estimates. As a starting point for this stochastic setting, an algorithm is proposed for the deterministic setting that is modeled after a state-of-the-art line-search SQP algorithm but uses a stepsize selection scheme based on Lipschitz constants (or adaptively estimated Lipschitz constants) in place of the line search. This sets the stage for the proposed algorithm for the stochastic setting, for which it is assumed that line searches would be intractable. Under reasonable assumptions, convergence (resp., convergence in expectation) from remote starting points is proved for the proposed deterministic (resp., stochastic) algorithm. The results of numerical experiments demonstrate the practical performance of our proposed techniques.

Key words. nonlinear optimization, stochastic optimization, sequential quadratic optimization

AMS subject classifications. 49M05, 49M10, 49M37, 65K05, 65K10, 90C15, 90C30, 90C55

DOI. 10.1137/20M1354556

1. Introduction.

We consider the design of algorithms for solving smooth nonlinear optimization problems with equality constraints. Such problems arise in various important applications throughout science and engineering, including optimal control, PDE-constrained optimization, network flow, and resource allocation [1, 2, 20, 30].

Numerous algorithms have been proposed for solving *deterministic* equality constrained optimization problems. Penalty methods [11, 14], including augmented Lagrangian methods [10, 18, 26], attempt to solve such problems by penalizing constraint violation through an objective term—weighted by a penalty parameter—and employing unconstrained optimization techniques for solving (approximately) a corresponding sequence of penalty subproblems. Such algorithms can behave poorly due to ill-conditioning and/or nonsmoothness of the penalty subproblems, depending on the type of penalty function employed. Their performance also often suffers due to sensitivity to the scheme for updating the penalty parameter.

Algorithms that consistently outperform penalty methods are those based on sequential quadratic optimization (commonly known as SQP), which in this setting of equality constrained optimization is intimately connected to the idea of applying Newton's method to stationarity conditions of the problem [35]. In particular, it is commonly accepted that one of the state-of-the-art algorithms for solving equality constrained optimization problems is such an SQP method that chooses stepsizes based on a line search applied to an exact penalty function [15, 16, 27]. In such a

[*] Received by the editors July 21, 2020; accepted for publication (in revised form) February 17, 2021; published electronically May 17, 2021.
https://doi.org/10.1137/20M1354556

Funding: This material is based upon work supported by the U.S. National Science Foundation, Division of Computing and Communication Foundations, under award CCF-1740796.

[†] Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109 USA (albertberahas@gmail.com).

[‡] Industrial and Systems Engineering, Lehigh University, Bethlehem, PA 18015 USA (frank.e.curtis@gmail.com, daniel.p.robinson@gmail.com, baz216@lehigh.edu).

method, the penalty function acts as a merit function only. It does not influence the computed search direction; it only influences the computed stepsize.

Significantly fewer algorithms have been proposed for solving equality constrained *stochastic* optimization problems. In particular, in this paper, we focus on such problems with constraint functions that are deterministic, but objective functions that are stochastic, in the sense that the objective is an expectation of a function defined with respect to a random variable with unknown distribution. (Various modeling paradigms have been proposed for solving problems involving stochastic constraints. These are out of our scope; we refer the reader to [31].) We assume that it is intractable to compute objective function and gradient values, although one is able to compute (unbiased) stochastic gradient estimates. A few algorithms have been proposed that may be employed in this setting [9, 19, 23, 28], but these are based on penalty methodologies, so do not benefit from advantages of SQP techniques. Let us also mention various proposed stochastic Frank–Wolfe algorithms [17, 21, 22, 29, 36] for (non)convex stochastic optimization with convex constraints. These are not applicable for our setting of having general nonlinear equality constraints.

1.1. Contributions.

In this paper, we propose two algorithms modeled after the aforementioned line-search SQP methodology. Our primary focus is an algorithm for the aforementioned setting of a problem with deterministic constraint functions, but a stochastic objective function. However, as a first step for considering this setting, we begin by proposing an algorithm for the deterministic setting that employs an adaptive stepsize selection scheme that makes use of Lipschitz constants (or adaptively updated Lipschitz constant estimates) rather than a line search. Based on this algorithm for the deterministic setting, we propose our algorithm for the stochastic setting that also uses Lipschitz constants (or, in practice, estimates of them) for stepsize selection.

We prove under common assumptions that our deterministic algorithm has convergence guarantees that match those of a state-of-the-art line-search SQP method. In addition, we prove under loose assumptions that our stochastic algorithm offers convergence guarantees that can match those of our deterministic algorithm *in expectation*. In particular, the results that we prove for our stochastic algorithm are of the type offered by stochastic gradient schemes for unconstrained optimization [4]. An additional challenge for constrained stochastic optimization is potentially poor behavior of an adaptive merit function parameter that balances emphasis between minimizing constraint violation and reducing the objective function. To address this, in addition to our aforementioned convergence analysis, which considers the behavior of the algorithm under good behavior of this adaptive parameter, we prove under a pragmatic assumption that a certain type of poor behavior cannot occur, and another type of poor behavior occurs with probability zero.

The results of numerical experiments show that our deterministic algorithm is as reliable as a state-of-the-art line-search SQP method, although, as should be expected, it is sometimes less efficient than such a method that performs line searches. Our experiments with our stochastic algorithm show that it consistently and significantly outperforms an approach that attempts to solve constrained problems by applying a stochastic (sub)gradient scheme to minimize an exact penalty function.

1.2. Notation.

Let $\mathbb{R}$ denote the set of real numbers (i.e., scalars), let $\mathbb{R}_{\geq r}$ (resp., $\mathbb{R}_{>r}$) denote the set of real numbers greater than or equal to (resp., greater than) $r \in \mathbb{R}$, let $\mathbb{R}^n$ denote the set of n-dimensional real vectors, let $\mathbb{R}^{m \times n}$ denote the set of m-by-n-dimensional real matrices, and let $\mathbb{S}^n$ denote the set of n-by-n-

dimensional symmetric matrices. The set of natural numbers is denoted as $\mathbb{N} := \{0, 1, 2, \dots\}$. For any $m \in \mathbb{N}$, let $[m]$ denote the set of integers $\{1, \dots, m\}$.

Each of our algorithms is iterative, generating a sequence of iterates $\{x_k\}$ with $x_k \in \mathbb{R}^n$ for all $k \in \mathbb{N}$. The iteration number is also appended as a subscript to other quantities corresponding to each iteration, e.g., $f_k := f(x_k)$ for all $k \in \mathbb{N}$.

1.3. Organization.

Our algorithm for the deterministic setting is proposed and analyzed in section 2. We present our analysis alongside that of a line-search SQP method for ease of comparison with this state-of-the-art strategy. Our algorithm for the stochastic setting is proposed and analyzed in section 3. The results of numerical experiments are provided in section 4 and concluding remarks are offered in section 5.

2. Deterministic setting.

Given an objective function $f : \mathbb{R}^n \to \mathbb{R}$ and a constraint function $c : \mathbb{R}^n \to \mathbb{R}^m$, consider the optimization problem

$$\min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } c(x) = 0. \tag{2.1}$$

We make the following assumption about the optimization problem (2.1) and the algorithms that we propose, each of which generates an iterate sequence $\{x_k\} \subset \mathbb{R}^n$, search direction sequence $\{d_k\} \subset \mathbb{R}^n$, and trial stepsize sequence $\{\alpha_{k,j}\} \subset \mathbb{R}_{>0}$.

Assumption 2.1. *Let $\mathcal{X} \subseteq \mathbb{R}^n$ be an open convex set containing the iterates $\{x_k\}$ and trial points $\{x_k + \alpha_{k,j} d_k\}$. The objective function $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and bounded below over $\mathcal{X}$, and its gradient $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous with constant L and bounded over $\mathcal{X}$. The constraint function $c : \mathbb{R}^n \to \mathbb{R}^m$ (where $m \le n$) and its Jacobian $\nabla c^T : \mathbb{R}^n \to \mathbb{R}^{m \times n}$ are bounded over $\mathcal{X}$, each gradient $\nabla c_i : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous with constant γ_i over $\mathcal{X}$ for all $i \in \{1, \dots, m\}$, and the singular values of ∇c^T are bounded away from zero over $\mathcal{X}$.*

Most of the statements in Assumption 2.1 are standard smoothness assumptions; see, e.g., [7, 33]. We do not assume that $\mathcal{X}$ is bounded. The assumption that the singular values of ∇c^T are bounded away from zero is equivalent to the linear independence constraint qualification (LICQ). This is a relatively strong assumption in the literature on algorithms for solving constrained optimization problems, but it holds for various real-world problems of interest [2], and in any case is reasonable in our context due to the significant challenges that arise in the stochastic setting in section 3.

Defining the Lagrangian $\ell : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ corresponding to (2.1) by $\ell(x, y) = f(x) + c(x)^T y$, first-order stationarity conditions for (2.1)—which are necessary due to the inclusion of the LICQ in Assumption 2.1—are given by

$$0 = \begin{bmatrix} \nabla_x \ell(x, y) \\ \nabla_y \ell(x, y) \end{bmatrix} = \begin{bmatrix} \nabla f(x) + \nabla c(x) y \\ c(x) \end{bmatrix}. \tag{2.2}$$

A consequence of Lipschitz continuity of the constraint functions is the following. Since this fact is well known and easily proved, we present it without proof.

Lemma 2.2. *Under Assumption 2.1, it follows for any $x \in \mathbb{R}^n$, $\alpha \in \mathbb{R}_{>0}$, and $d \in \mathbb{R}^n$ such that $(x, x + \alpha d) \in \mathcal{X} \times \mathcal{X}$ that*

$$|c_i(x + \alpha d)| \le |c_i(x) + \alpha \nabla c_i(x)^T d| + \tfrac{1}{2} \gamma_i \alpha^2 \|d\|_2^2 \quad \textit{for all} \quad i \in [m]$$
$$\textit{and} \quad \|c(x + \alpha d)\|_1 \le \|c(x) + \alpha \nabla c(x)^T d\|_1 + \tfrac{1}{2} \Gamma \alpha^2 \|d\|_2^2 \quad \textit{with} \quad \Gamma := \textstyle\sum_{i \in [m]} \gamma_i.$$

2.1. Merit function. As is common in SQP techniques, our algorithms use as a merit function the ℓ_1-norm penalty function $\phi : \mathbb{R}^n \times \mathbb{R}_{>0} \to \mathbb{R}$ defined by

$$\phi(x,\tau) = \tau f(x) + \|c(x)\|_1. \tag{2.3}$$

Here, $\tau \in \mathbb{R}_{>0}$ is a merit parameter, the value of which is chosen in the algorithm according to a positive sequence $\{\tau_k\}$ that is set adaptively. We make use of a local model of the merit function $q : \mathbb{R}^n \times \mathbb{R}_{>0} \times \mathbb{R}^n \times \mathbb{S}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by

$$q(x,\tau,\nabla f(x),H,d) = \tau(f(x) + \nabla f(x)^T d + \tfrac{1}{2}\max\{d^T H d, 0\}) + \|c(x) + \nabla c(x)^T d\|_1.$$

A critical quantity in our algorithms is the reduction in this model for a given $d \in \mathbb{R}^n$ with $c(x) + \nabla c(x)^T d = 0$, i.e., $\Delta q : \mathbb{R}^n \times \mathbb{R}_{>0} \times \mathbb{R}^n \times \mathbb{S}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by

$$\begin{aligned} \Delta q(x,\tau,\nabla f(x),H,d) &:= q(x,\tau,\nabla f(x),H,0) - q(x,\tau,\nabla f(x),H,d) \\ &= -\tau(\nabla f(x)^T d + \tfrac{1}{2}\max\{d^T H d, 0\}) + \|c(x)\|_1. \end{aligned} \tag{2.4}$$

The following lemma shows an important relationship between the directional derivative of the merit function and this model reduction function.

LEMMA 2.3. *Given $(x,\tau,H,d) \in \mathbb{R}^n \times \mathbb{R}_{>0} \times \mathbb{S}^n \times \mathbb{R}^n$ with $c(x) + \nabla c(x)^T d = 0$,*

$$\phi'(x,\tau,d) = \tau \nabla f(x)^T d - \|c(x)\|_1 \le -\Delta q(x,\tau,\nabla f(x),H,d), \tag{2.5}$$

where $\phi' : \mathbb{R}^n \times \mathbb{R}_{>0} \times \mathbb{R}^n \to \mathbb{R}$ is the directional derivative of ϕ at (x,τ) for d.

Proof. The first equation in (2.5) is well known; see, e.g., [25, Theorem 18.2]. On the other hand, from the definition (2.4) one finds that $\Delta q(x,\tau,\nabla f(x),H,d) = -\phi'(x,\tau,d) - \frac{1}{2}\tau\max\{d^T H d, 0\} \le -\phi'(x,\tau,d)$, which shows the inequality in (2.5). □

2.2. Algorithm preliminaries. The algorithms that we discuss for solving (2.1) are based on an SQP paradigm. Specifically, at x_k for all $k \in \mathbb{N}$, a search direction $d_k \in \mathbb{R}^n$ is computed by solving a quadratic optimization subproblem based on a local quadratic model of f and a local affine model of c about x_k. Letting $f_k := f(x_k)$, $g_k := \nabla f(x_k)$, $c_k := c(x_k)$, and $J_k := \nabla c(x_k)^T$ for all $k \in \mathbb{N}$ and given a sequence $\{H_k\}$ satisfying Assumption 2.4 below (a standard type of sufficiency condition for equality constrained optimization), this subproblem is given by

$$\min_{d \in \mathbb{R}^n} \; f_k + g_k^T d + \tfrac{1}{2} d^T H_k d \quad \text{s.t.} \quad c_k + J_k d = 0.$$

The optimal solution d_k of this subproblem, and an associated Langrange multiplier $y_k \in \mathbb{R}^m$, can be obtained by solving the linear system of equations

$$\begin{bmatrix} H_k & J_k^T \\ J_k & 0 \end{bmatrix} \begin{bmatrix} d_k \\ y_k \end{bmatrix} = - \begin{bmatrix} g_k \\ c_k \end{bmatrix}. \tag{2.6}$$

ASSUMPTION 2.4. *The sequence $\{H_k\}$ is bounded in norm by $\kappa_H \in \mathbb{R}_{>0}$. In addition, there exists a constant $\zeta \in \mathbb{R}_{>0}$ such that, for all $k \in \mathbb{N}$, the matrix H_k has the property that $u^T H_k u \ge \zeta \|u\|_2^2$ for all $u \in \mathbb{R}^n$ such that $J_k u = 0$.*

We stress that our algorithms and analysis do *not* assume that H_k is equal to the Hessian of the Lagrangian at x_k for some multiplier y_k, although choosing $\{H_k\}$ in this manner would be appropriate in order to ensure fast local convergence guarantees.

Since our focus is only on achieving convergence to stationarity from remote starting points, we merely assume that $\{H_k\}$ satisfies Assumption 2.4.

Under Assumptions 2.1 and 2.4, the following results are well known in the literature.

LEMMA 2.5. *For all $k \in \mathbb{N}$, the linear system* (2.6) *has a unique solution.*

LEMMA 2.6. *For any $k \in \mathbb{N}$, the solution (d_k, y_k) obtained by solving* (2.6) *has $d_k = 0$ if and only if the pair (x_k, y_k) satisfies* (2.2).

2.3. Algorithms. In this section, we present two algorithms for solving problem (2.1). The first algorithm chooses stepsizes based on a rule using Lipschitz constant estimates, which can be set adaptively. This algorithm is new to the literature and establishes a foundation upon which our method for the stochastic setting will be built. The second algorithm, by contrast, employs a standard type of backtracking line search. This algorithm is standard in the literature. We prove a convergence theory for it alongside that for our newly proposed algorithm for illustrative purposes.

In both algorithms, after d_k is computed, the merit parameter τ_k is set. This is done by first setting, for some $\sigma \in (0,1)$, a trial value $\tau_k^{trial} \in \mathbb{R}_{>0} \cup \{\infty\}$ by

$$\tau_k^{trial} \leftarrow \begin{cases} \infty & \text{if } g_k^T d_k + \max\{d_k^T H_k d_k, 0\} \leq 0, \\ \frac{(1-\sigma)\|c_k\|_1}{g_k^T d_k + \max\{d_k^T H_k d_k, 0\}} & \text{otherwise.} \end{cases} \tag{2.7}$$

(If $c_k = 0$, then it follows from (2.6) and Assumption 2.4 that $d_k^T H_k d_k \geq 0$ and $g_k^T d_k + d_k^T H_k d_k = 0$, meaning $\tau_k^{trial} \leftarrow \infty$. Hence, $\tau_k^{trial} < \infty$ requires $\|c_k\|_1 > 0$, in which case $\tau_k^{trial} > 0$.) Then, the merit parameter τ_k is set, for some $\epsilon \in (0,1)$, by

$$\tau_k \leftarrow \begin{cases} \tau_{k-1} & \text{if } \tau_{k-1} \leq \tau_k^{trial}, \\ (1-\epsilon)\tau_k^{trial} & \text{otherwise.} \end{cases} \tag{2.8}$$

This ensures that $\tau_k \leq \tau_k^{trial}$. Regardless of the case in (2.8), it follows that

$$\Delta q(x_k, \tau_k, g_k, H_k, d_k) \geq \tfrac{1}{2}\tau_k \max\{d_k^T H_k d_k, 0\} + \sigma\|c_k\|_1. \tag{2.9}$$

This inequality will be central in our analysis of both algorithms. In particular, it will be useful when combined with the fact that each algorithm ensures that, for all $k \in \mathbb{N}$, the stepsize $\alpha_k \in \mathbb{R}_{>0}$ is selected such that for $\eta \in (0,1)$ one finds

$$\phi(x_k + \alpha_k d_k, \tau_k) \leq \phi(x_k, \tau_k) - \eta\alpha_k \Delta q(x_k, \tau_k, g_k, H_k, d_k). \tag{2.10}$$

Remark 2.7. An alternative approach for setting the merit function parameter is to set it based on the computed Lagrange multiplier estimate y_k. For example, in the context of our ℓ_1-norm exact penalty function $\phi(x_k, \cdot)$, one can ensure that the computed search direction d_k is a direction of descent for $\phi(\cdot, \tau_k)$ from x_k if $\tau_k < \|y_k\|_\infty^{-1}$; see, e.g., [25]. However, it is often better in practice to set it based on ensuring sufficient reduction in a model of the merit function (see, e.g., [7, 8]).

LEMMA 2.8. *Under Assumption* 2.1*, the inner **for** loop in Algorithm* 2.1 *is well-posed in that for any $k \in \mathbb{N}$, it terminates finitely. In addition, for all $k \in \mathbb{N}$,*

$$\begin{aligned} L_k &\leq L_{\max} := \max\{L_{-1}, \rho L\} \\ \textit{and } \ \gamma_{k,i} &\leq \gamma_{\max,i} := \max\{\gamma_{-1,i}, \rho\gamma_i\} \quad \textit{for all } \ i \in [m]. \end{aligned} \tag{2.11}$$

Algorithm 2.1 SQP Algorithm with Adaptive Lipschitz Constant Estimates.

```
Require: $x_0 \in \mathbb{R}^n$; $\tau_{-1} \in \mathbb{R}_{>0}$; $\epsilon \in (0,1)$; $\sigma \in (0,1)$; $\eta \in (0,1)$; $\rho \in \mathbb{R}_{>1}$; $L_{-1} \in \mathbb{R}_{>0}$;
    $\gamma_{-1,i} \in \mathbb{R}_{>0}$ for all $i \in [m]$
for all $k \in \mathbb{N}$ do
    Compute $(d_k, y_k)$ as the solution of (2.6)
    if $(x_k, y_k)$ satisfies (2.2) then return $(x_k, y_k)$
    Set $\tau_k^{trial}$ by (2.7) and $\tau_k$ by (2.8)
    Initialize $L_{k,0} \in (0, L_{k-1}]$ and $\gamma_{k,i,0} \leftarrow (0, \gamma_{k-1,i}]$ for all $i \in [m]$
    for all $j \in \mathbb{N}$ do
        Set
        $$\left\{\begin{aligned} \widehat{\alpha}_{k,j} &\leftarrow \frac{2(1-\eta)\Delta q(x_k,\tau_k,g_k,H_k,d_k)}{(\tau_k L_{k,j}+\sum_{i\in[m]}\gamma_{k,i,j})\|d_k\|_2^2} \\ \widetilde{\alpha}_{k,j} &\leftarrow \widehat{\alpha}_{k,j} - \frac{4\|c_k\|_1}{(\tau_k L_{k,j}+\sum_{i\in[m]}\gamma_{k,i,j})\|d_k\|_2^2} \end{aligned}\right\}; \ \alpha_{k,j} \leftarrow \begin{cases} \widehat{\alpha}_{k,j} & \text{if } \widehat{\alpha}_{k,j} < 1 \\ 1 & \text{if } \widetilde{\alpha}_{k,j} \leq 1 \leq \widehat{\alpha}_{k,j} \\ \widetilde{\alpha}_{k,j} & \text{if } \widetilde{\alpha}_{k,j} > 1 \end{cases}$$
        if (2.10) or (2.12) holds then
            Set $L_k \leftarrow L_{k,j}$ and $\gamma_{k,i} \leftarrow \gamma_{k,i,j}$ for all $i \in [m]$
            Set $\alpha_k \leftarrow \alpha_{k,j}$ and $x_{k+1} \leftarrow x_k + \alpha_k d_k$ and break (loop over $j \in \mathbb{N}$)
        else
            if (2.12a) (resp., (2.12b) for some $i \in [m]$) is not satisfied
                Set $L_{k,j+1} \leftarrow \rho L_{k,j}$ (resp., $\gamma_{k,i,j+1} \leftarrow \rho\gamma_{k,i,j}$)
            else
                Set $L_{k,j+1} \leftarrow L_{k,j}$ (resp., $\gamma_{k,i,j+1} \leftarrow \gamma_{k,i,j}$)
```

Proof. To derive a contradiction, suppose that for some $k \in \mathbb{N}$ the inner **for** loop does not terminate. This means that for each iteration of the **for** loop at least one inequality in (2.12) does not hold. In such a case, the **for** loop sets $L_{k,j+1}$ (resp., $\gamma_{k,i,j+1}$ for some $i \in [m]$) as $\rho > 1$ times $L_{k,j}$ (resp., $\gamma_{k,i,j}$ for some $i \in [m]$). This leads to a contradiction to the fact that if $L_{k,j} \geq L$ and $\gamma_{k,i,j} \geq \gamma_i$ for all $i \in [m]$, then (2.12) holds. Finally, (2.11) follows from the initialization of the Lipschitz constant estimates; the fact that if any of these values is ever increased in the **for** loop, then this occurs by the value being multiplied by $\rho > 1$; and the fact that for all $k \in \mathbb{N}$ the algorithm initializes $L_{k,0} \in (0, L_{k-1}]$ and $\gamma_{k,i,0} \in (0, \gamma_{k-1,i}]$ for all $i \in [m]$. ◻

Our first algorithm is stated as Algorithm 2.1. A signifying feature of it is the manner in which it can adapt Lipschitz constant estimates, which are used in the stepsize selection scheme. For any $(k,j) \in \mathbb{N} \times \mathbb{N}$, if the estimates $L_{k,j}$ and $\{\gamma_{k,i,j}\}_{i=1}^m$ satisfy $L_{k,j} \geq L$ and $\gamma_{k,i,j} \geq \gamma_i$ for all $i \in [m]$, then it follows (see [24] and Lemma 2.2) that for $\alpha_{k,j} \in \mathbb{R}_{>0}$ yielding $x_k + \alpha_{k,j} d_k \in \mathcal{X}$ (recall Assumption 2.1) one has

$$f(x_k + \alpha_{k,j} d_k) \leq f_k + \alpha_{k,j} g_k^T d_k + \tfrac{1}{2} L_{k,j} \alpha_{k,j}^2 \|d_k\|_2^2 \tag{2.12a}$$

$$\text{and} \quad |c_i(x_k + \alpha_{k,j} d_k)| \leq |c_i(x_k) + \alpha_{k,j} \nabla c_i(x_k)^T d_k| + \tfrac{1}{2} \gamma_{k,i,j} \alpha_{k,j}^2 \|d_k\|_2^2 \tag{2.12b}$$

for all $i \in [m]$. If one knows Lipschitz constants for ∇f and $\{\nabla c_i\}_{i=1}^m$, then one could simply set $L_{k,0}$ and $\gamma_{k,i,0}$ for all $i \in [m]$ to these values for all $k \in \mathbb{N}$, in which case the inner **for** loop would terminate in iteration $j = 0$ for all $k \in \mathbb{N}$. However, if such Lipschitz constants are unknown, as is often the case, then the adaptive procedure in Algorithm 2.1 ensures that convergence can be guaranteed, as shown in the next subsection. For now, we simply prove the following lemma showing that the inner loop of the algorithm is well-posed. (One could choose a different increase factor $\rho \in \mathbb{R}_{>1}$ for each Lipschitz constant estimate; we use a common value of ρ for simplicity.)

The intuition behind the stepsize selection scheme in Algorithm 2.1 is that the stepsize is chosen to ensure a sufficient reduction in an upper bound on the change in the merit function. This upper bound is revealed in Lemma 2.13 later on; in particular, see (2.13). Due to the nonsmoothness of the merit function, which creates a *kink* at a unit stepsize, there are three cases: the minimizer may occur *before*, *at*, or *after* the kink. An illustration of these cases is shown in Figure 2.1. Certain situations that lead to each of the three cases is as follows. (There are additional situations that one may consider since the upper bounding function involves a combination of many terms, but the following are a few example situations to provide some intuition.) If the Lipschitz constant estimates are large enough, indicating high nonlinearity of the problem functions, then the minimizer may be at a stepsize less than 1. On the other hand, if the Lipschitz constant estimates are not too large and derivative information of the objective suggests that the merit function improves beyond a unit stepsize, then the minimizer is at a stepsize greater than 1. Otherwise, the minimizer occurs at a unit stepsize since at least this corresponds to a step toward linearized feasibility.

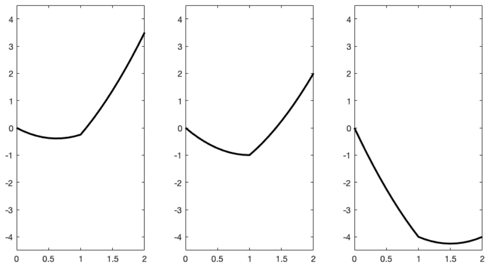

Fig. 2.1. *Illustration of three cases for an upper bounding function of the merit function (see Lemma* 2.13*) motivating the three cases in the stepsize selection scheme in Algorithm* 2.1*. Each graph shows the value of the upper bound on the change in the merit function as a function of α_k.*

The second algorithm is stated as Algorithm 2.2. In each iteration, it employs a traditional backtracking line search scheme until the reduction in the merit function is sufficiently large compared to the reduction in the model of the merit function. This is sufficient for showing a convergence result, as shown in the next subsection.

Algorithm 2.2 SQP Algorithm with Backtracking Line Search.

```
Require: $x_0 \in \mathbb{R}^n$; $\tau_{-1} \in \mathbb{R}_{>0}$; $\epsilon \in (0,1)$; $\sigma \in (0,1)$; $\eta \in (0,1)$; $\nu \in (0,1)$; $\alpha \in \mathbb{R}_{>0}$
for all $k \in \mathbb{N}$ do
    Compute $(d_k, y_k)$ as the solution of (2.6)
    if $(x_k, y_k)$ satisfies (2.2) then return $(x_k, y_k)$
    Set $\tau_k^{trial}$ by (2.7) and $\tau_k$ by (2.8)
    for all $j \in \mathbb{N}$ do
        Set $\alpha_{k,j} \leftarrow \nu^j \alpha$
        if (2.10) holds then
            Set $\alpha_k \leftarrow \alpha_{k,j}$ and $x_{k+1} \leftarrow x_k + \alpha_k d_k$ and break (loop over $j \in \mathbb{N}$)
```

2.4. Convergence analysis. We prove in this section that, from any initial iterate, each of Algorithm 2.1 and Algorithm 2.2 generates a sequence of iterates over which a first-order measure of primal-dual stationarity for (2.1) (recall (2.2)) vanishes. We assume throughout this section that both Assumptions 2.1 and 2.4 hold; for brevity, we do not remind the reader of this fact within the statement of each result. We also remark that if an algorithm terminates finitely, then it does so with (x_k, y_k) satisfying (2.2), meaning primal-dual stationarity has been achieved. Hence, we may assume without loss of generality in this section that neither algorithm terminates finitely, meaning that $\{x_k\}$ is infinite and $d_k \neq 0$ for all $k \in \mathbb{N}$ (recall Lemma 2.6).

In all of the results of this section, the statements are proved to hold with respect to *both* Algorithms 2.1 and 2.2. There are only a few differences in the results for the two algorithms; when a result differs, we say so explicitly. Much of our analysis, at least prior to Lemma 2.14, follows standard analysis for line-search SQP methods; see, e.g., [5, 6]. Nonetheless, we provide proofs of the results for completeness.

Our analysis uses the orthogonal decomposition of the search directions given by $d_k = u_k + v_k$, where $u_k \in \text{Null}(J_k)$ and $v_k \in \text{Range}(J_k^T)$ for all $k \in \mathbb{N}$. *We emphasize that the components u_k and v_k do not need to be computed explicitly for any $k \in \mathbb{N}$.* They are merely tools for our analysis. As is common in the literature, we refer to u_k as the tangential component and v_k as the normal component of d_k.

We first show an upper bound on the normal components of the search directions.

LEMMA 2.9. *There exists $\kappa_v \in \mathbb{R}_{>0}$ such that, for all $k \in \mathbb{N}$, the normal component v_k satisfies* $\max\{\|v_k\|_2, \|v_k\|_2^2\} \leq \kappa_v \|c_k\|_2$.

Proof. Let $k \in \mathbb{N}$ be arbitrary. From $J_k d_k = J_k(u_k + v_k) = -c_k$, $u_k \in \text{Null}(J_k)$, and $v_k \in \text{Range}(J_k^T)$, one has $v_k = -J_k^T(J_kJ_k^T)^{-1}c_k$; hence, by the definition of the matrix norm induced by the ℓ_2 vector norm, it follows that

$$\begin{aligned} & \|v_k\|_2 \leq \|J_k^T(J_kJ_k^T)^{-1}\|_2\|c_k\|_2 \\ \iff \quad & \|v_k\|_2^2 \leq (\|J_k^T(J_kJ_k^T)^{-1}\|_2\|c_k\|_2)^2 = (\|J_k^T(J_kJ_k^T)^{-1}\|_2^2\|c_k\|_2)\|c_k\|_2. \end{aligned}$$

Hence, the desired conclusion follows under Assumption 2.1. □

Our next result reveals that there exists a critical threshold between the norms of the tangential and normal components of the search directions, and in any iteration $k \in \mathbb{N}$ in which the search direction d_k is dominated by the tangential component u_k, the curvature of H_k along d_k has a useful lower bound defined with u_k.

LEMMA 2.10. *There exists $\kappa_{uv} \in \mathbb{R}_{>0}$ such that, for any $k \in \mathbb{N}$, if $\|u_k\|_2^2 \geq \kappa_{uv}\|v_k\|_2^2$, then $\frac{1}{2}d_k^TH_kd_k \geq \frac{1}{4}\zeta\|u_k\|_2^2$, where $\zeta \in \mathbb{R}_{>0}$ is defined in Assumption* 2.4.

Proof. Assumption 2.4 implies for any $\kappa_{uv} \in \mathbb{R}_{>0}$ that $\|u_k\|_2^2 \geq \kappa_{uv}\|v_k\|_2^2$ means

$$\begin{aligned} & \tfrac{1}{2}d_k^TH_kd_k = \tfrac{1}{2}u_k^TH_ku_k + u_k^TH_kv_k + \tfrac{1}{2}v_k^TH_kv_k \\ & \geq \tfrac{1}{2}\zeta\|u_k\|_2^2 - \|u_k\|_2\|H_k\|_2\|v_k\|_2 - \tfrac{1}{2}\|H_k\|_2\|v_k\|_2^2 \geq \left(\tfrac{\zeta}{2} - \tfrac{\kappa_H}{\sqrt{\kappa_{uv}}} - \tfrac{\kappa_H}{2\kappa_{uv}}\right)\|u_k\|_2^2. \end{aligned}$$

Thus, under Assumption 2.4, the result holds for $\kappa_{uv} \in \mathbb{R}_{>0}$ with $\frac{\kappa_H}{\sqrt{\kappa_{uv}}} + \frac{\kappa_H}{2\kappa_{uv}} \leq \frac{\zeta}{4}$. □

For the constant $\kappa_{uv} \in \mathbb{R}_{>0}$ defined in Lemma 2.10, let us define

$$\Psi_k := \begin{cases} \|u_k\|_2^2 + \|c_k\|_2 & \text{if } \|u_k\|_2^2 \geq \kappa_{uv}\|v_k\|_2^2, \\ \|c_k\|_2 & \text{otherwise,} \end{cases}$$

along with the corresponding index sets $\mathcal{K}_u := \{k \in \mathbb{N} : \|u_k\|_2^2 \geq \kappa_{uv}\|v_k\|_2^2\}$ and $\mathcal{K}_v := \{k \in \mathbb{N} : \|u_k\|_2^2 < \kappa_{uv}\|v_k\|_2^2\}$ (that form a partition of $\mathbb{N}$). Our next result shows that the squared norms of the search directions and the constraint violations are bounded above by this critical quantity in all iterations.

LEMMA 2.11. *There exists $\kappa_\Psi \in \mathbb{R}_{>0}$ such that, for all $k \in \mathbb{N}$, the search direction and constraint violation satisfy $\|d_k\|_2^2 \leq \kappa_\Psi \Psi_k$ and $\|d_k\|_2^2 + \|c_k\|_2 \leq (\kappa_\Psi + 1)\Psi_k$.*

Proof. For all $k \in \mathcal{K}_u$, it follows that

$$\|d_k\|_2^2 = \|u_k\|_2^2 + \|v_k\|_2^2 \leq (1 + \kappa_{uv}^{-1})\|u_k\|_2^2 \leq (1 + \kappa_{uv}^{-1})(\|u_k\|_2^2 + \|c_k\|_2).$$

For all $k \in \mathcal{K}_v$, one finds from Lemma 2.9 that

$$\|d_k\|_2^2 = \|u_k\|_2^2 + \|v_k\|_2^2 < (\kappa_{uv} + 1)\|v_k\|_2^2 \leq (\kappa_{uv} + 1)\kappa_v\|c_k\|_2.$$

Combining the results from the two cases implies the first desired result. To establish the second result, note that the definition of Ψ_k yields $\|c_k\|_2 \leq \Psi_k$ for all $k \in \mathbb{N}$. □

As revealed by our next lemma, the reduction in the model of the merit function is bounded below with respect to the same critical quantity.

LEMMA 2.12. *There exists $\kappa_q \in \mathbb{R}_{>0}$ such that, for all $k \in \mathbb{N}$, the reduction in the model of the merit function satisfies $\Delta q(x_k, \tau_k, g_k, H_k, d_k) \geq \kappa_q \tau_k \Psi_k$.*

Proof. Combining (2.9) and Lemma 2.10, it follows that $\Delta q(x_k, \tau_k, g_k, H_k, d_k) \geq \frac{1}{4}\tau_k\zeta\|u_k\|_2^2 + \sigma\|c_k\|_1$ for $k \in \mathcal{K}_u$. Similarly, (2.9) implies that $\Delta q(x_k, \tau_k, g_k, H_k, d_k) \geq \sigma\|c_k\|_1$ for all $k \in \mathcal{K}_v$. Combining the two cases, $\|\cdot\|_1 \geq \|\cdot\|_2$, and the fact that $\{\tau_k\}$ is monotonically nonincreasing, the result holds for $\kappa_q := \min\{\frac{1}{4}\zeta, \sigma/\tau_{-1}\} \in \mathbb{R}_{>0}$. □

Our next lemma shows an upper bound on the change in the merit function when the inner **for** loop of Algorithm 2.1 terminates with large Lipschitz constant estimates.

LEMMA 2.13. *For all $k \in \mathbb{N}$, if the inner **for** loop of Algorithm* 2.1 *terminates since* (2.12) *holds, then with $\Gamma_k := \sum_{i\in[m]} \gamma_{k,i} \in \mathbb{R}_{>0}$ it follows that*

$$\begin{aligned} &\phi(x_k + \alpha_k d_k, \tau_k) - \phi(x_k, \tau_k) \\ &\leq \alpha_k\tau_k g_k^T d_k + |1 - \alpha_k|\|c_k\|_1 - \|c_k\|_1 + \tfrac{1}{2}(\tau_k L_k + \Gamma_k)\alpha_k^2\|d_k\|_2^2. \end{aligned} \tag{2.13}$$

Proof. For such $k \in \mathbb{N}$, it follows from (2.12) and Lemma 2.2 that

$$\begin{aligned} &\phi(x_k + \alpha_k d_k, \tau_k) - \phi(x_k, \tau_k) \\ &\leq \alpha_k\tau_k g_k^T d_k + \|c_k + \alpha_k J_k d_k\|_1 - \|c_k\|_1 + \tfrac{1}{2}(\tau_k L_k + \Gamma_k)\alpha_k^2\|d_k\|_2^2 \\ &= \alpha_k\tau_k g_k^T d_k + |1 - \alpha_k|\|c_k\|_1 - \|c_k\|_1 + \tfrac{1}{2}(\tau_k L_k + \Gamma_k)\alpha_k^2\|d_k\|_2^2, \end{aligned}$$

as desired. □

Next, we show lower bounds for the reduction in the merit function in each iteration of each algorithm. For concision, let us define for all $k \in \mathbb{N}$ the values

$$\widehat{\mu}_k := \tfrac{2(1-\eta)\Delta q(x_k,\tau_k,g_k,H_k,d_k)}{(\tau_k L + \sum_{i\in[m]}\gamma_i)\|d_k\|_2^2} \quad \text{and} \quad \widetilde{\mu}_k := \widehat{\mu}_k - \tfrac{4\|c_k\|_1}{(\tau_k L + \sum_{i\in[m]}\gamma_i)\|d_k\|_2^2}.$$

For a given $k \in \mathbb{N}$, one should notice the similarity between these values and the pair $(\widehat{\alpha}_{k,j}, \widetilde{\alpha}_{k,j})$ defined for all $j \in \mathbb{N}$ in Algorithm 2.1, except that the pair $(\widehat{\mu}_k, \widetilde{\mu}_k)$ are defined with respect to L and γ_i for all $i \in [m]$ defined in Assumption 2.1.

LEMMA 2.14. *For all $k \in \mathbb{N}$, the inequality* (2.10) *holds, where in the case of Algorithm* 2.2 *this occurs with the stepsize satisfying $\alpha_k > \nu\min\{\widehat{\mu}_k, \max\{1, \widetilde{\mu}_k\}\} > 0$.*

Proof. Let $k \in \mathbb{N}$ be given. First, consider Algorithm 2.1. If the inner **for** loop terminates since the stepsize yields (2.10), then there is nothing left to prove. Hence, we may proceed by supposing that the loop terminates since (2.12) holds, which we shall now proceed to show means that (2.10) holds as well. Consider three cases, where as in Lemma 2.13 let us define $\Gamma_k := \sum_{i\in[m]} \gamma_{k,i} \in \mathbb{R}_{>0}$.

Case 1: Suppose that in the last iteration of the inner **for** loop one finds $\widehat{\alpha}_{k,j} < 1$, in which case the algorithm yields $\alpha_k = \frac{2(1-\eta)\Delta q(x_k,\tau_k,g_k,H_k,d_k)}{(\tau_k L_k+\Gamma_k)\|d_k\|_2^2} < 1$. Combining this fact with Lemmas 2.3 and 2.13, it follows that

$$\begin{aligned}
&\phi(x_k+\alpha_k d_k,\tau_k) - \phi(x_k,\tau_k)\\
\le{}& \alpha_k(\tau_k g_k^T d_k - \|c_k\|_1) + \tfrac{1}{2}(\tau_k L_k + \Gamma_k)\alpha_k^2\|d_k\|_2^2\\
\le{}& -\alpha_k\Delta q(x_k,\tau_k,g_k,H_k,d_k) + \tfrac{1}{2}(\tau_k L_k + \Gamma_k)\alpha_k^2\|d_k\|_2^2\\
={}& -\alpha_k\Delta q(x_k,\tau_k,g_k,H_k,d_k) + \tfrac{1}{2}\alpha_k(\tau_k L_k + \Gamma_k)\left(\tfrac{2(1-\eta)\Delta q(x_k,\tau_k,g_k,H_k,d_k)}{(\tau_k L_k+\Gamma_k)\|d_k\|_2^2}\right)\|d_k\|_2^2\\
={}& -\eta\alpha_k\Delta q(x_k,\tau_k,g_k,H_k,d_k).
\end{aligned}$$

Case 2: Suppose that in the last iteration of the inner **for** loop one finds $\widehat{\alpha}_{k,j} \ge 1$ and $\widetilde{\alpha}_{k,j} \le 1$, in which case the algorithm yields $\alpha_k = 1$. Combining this fact, the fact that $\widehat{\alpha}_{k,j} \ge 1$ in the last iteration of the loop, and Lemmas 2.3 and 2.13 yields the same string of relationships as in Case 1, except that since $\widehat{\alpha}_{k,j} \ge 1$ the first equation holds not as an equation, but as an "$\le$" inequality.

Case 3: Suppose that in the last iteration of the inner **for** loop one finds $\widetilde{\alpha}_{k,j} > 1$, in which case the algorithm yields $\alpha_k = \frac{2(1-\eta)\Delta q(x_k,\tau_k,g_k,H_k,d_k)-4\|c_k\|_1}{(\tau_k L_k+\Gamma_k)\|d_k\|_2^2} > 1$. Combining this fact with Lemmas 2.3 and 2.13, it follows that

$$\begin{aligned}
&\phi(x_k+\alpha_k d_k,\tau_k) - \phi(x_k,\tau_k)\\
\le{}& \alpha_k\tau_k g_k^T d_k + (\alpha_k-1)\|c_k\|_1 - \|c_k\|_1 + \tfrac{1}{2}(\tau_k L_k + \Gamma_k)\alpha_k^2\|d_k\|_2^2\\
={}& \alpha_k(\tau_k g_k^T d_k - \|c_k\|_1) + 2(\alpha_k-1)\|c_k\|_1 + \tfrac{1}{2}(\tau_k L_k + \Gamma_k)\alpha_k^2\|d_k\|_2^2\\
\le{}& -\alpha_k\Delta q(x_k,\tau_k,g_k,H_k,d_k) + 2\alpha_k\|c_k\|_1 + \tfrac{1}{2}(\tau_k L_k + \Gamma_k)\alpha_k^2\|d_k\|_2^2\\
={}& -\alpha_k\Delta q(x_k,\tau_k,g_k,H_k,d_k) + 2\alpha_k\|c_k\|_1\\
&\quad + \tfrac{1}{2}\alpha_k(\tau_k L_k + \Gamma_k)\left(\tfrac{2(1-\eta)\Delta q(x_k,\tau_k,g_k,H_k,d_k)-4\|c_k\|_1}{(\tau_k L_k+\Gamma_k)\|d_k\|_2^2}\right)\|d_k\|_2^2\\
={}& -\eta\alpha_k\Delta q(x_k,\tau_k,g_k,H_k,d_k).
\end{aligned}$$

Combining the three cases shows the desired result for Algorithm 2.1.

Now consider Algorithm 2.2. One finds that one of three cases occurs, which mimic those for Algorithm 2.1. In particular, if $\widehat{\mu}_k < 1$, then an analysis similar to that for Case 1 above shows that for $j \in \mathbb{N}$ with $\alpha_{k,j}/\nu > \widehat{\mu}_k$ and $\alpha_{k,j} \le \widehat{\mu}_k$, the backtracking line search will terminate by iteration $j \in \mathbb{N}$, from which it follows that $\alpha_k > \nu\widehat{\mu}_k$. If $\widehat{\mu}_k \ge 1$ and $\widetilde{\mu}_k \le 1$, or if $\widetilde{\mu}_k > 1$, then a similar argument combined with Case 2 or Case 3, respectively, completes the proof. □

Next, we show that the tangential components of the directions are bounded.

Lemma 2.15. *The tangential component sequence* $\{u_k\}$ *is bounded.*

Proof. The first block of (2.6) yields $u_k^T H_k(u_k+v_k) = -u_k^T g_k$. Hence, under Assumption 2.4, one finds that $\zeta\|u_k\|_2^2 \le u_k^T H_k u_k = -g_k^T u_k - v_k^T H_k u_k \le (\|g_k\|_2 + \kappa_H\|v_k\|_2)\|u_k\|_2$. Therefore, the result follows from Assumption 2.1 and Lemma 2.9. □

We now show that the merit parameter sequence is bounded and that it remains fixed at a value for all sufficiently large $k \in \mathbb{N}$.

LEMMA 2.16. *There exists $k_\tau \in \mathbb{N}$ and $\tau_{\min} \in \mathbb{R}_{>0}$ such that $\tau_k = \tau_{\min}$ for $k \geq k_\tau$.*

Proof. Recall that $\tau_k < \tau_{k-1}$ if and only if both $g_k^T d_k + \max\{d_k^T H_k d_k, 0\} > 0$ and

$$\tau_{k-1}(g_k^T d_k + \max\{d_k^T H_k d_k, 0\}) > (1-\sigma)\|c_k\|_1. \tag{2.14}$$

According to the first block equation of (2.6) (premultiplied by u_k^T) one has

$$g_k^T d_k + \max\{d_k^T H_k d_k, 0\} = \begin{cases} g_k^T v_k + v_k^T H_k u_k + v_k^T H_k v_k & \text{if } d_k^T H_k d_k \geq 0, \\ g_k^T v_k - v_k^T H_k u_k - u_k^T H_k u_k & \text{otherwise.} \end{cases}$$

The result follows from our ability to bound the left-hand side of this expression with respect to the constraint reduction. We consider two cases. First, if $d_k^T H_k d_k \geq 0$, then under Assumptions 2.1 and 2.4 it follows with Lemmas 2.9 and 2.15 and $\|\cdot\|_1 \geq \|\cdot\|_2$ that there exists a constant $\kappa_{\tau,1} \in \mathbb{R}_{>0}$ such that

$$g_k^T v_k + v_k^T H_k u_k + v_k^T H_k v_k \leq (\|g_k\|_2 + \kappa_H \|u_k\|_2)\|v_k\|_2 + \kappa_H \|v_k\|_2^2 \leq \kappa_{\tau,1}\|c_k\|_1.$$

Second, if $d_k^T H_k d_k < 0$, then under Assumptions 2.1 and 2.4 it follows from Lemmas 2.9 and 2.15 and $\|\cdot\|_1 \geq \|\cdot\|_2$ that there exists a constant $\kappa_{\tau,2} \in \mathbb{R}_{>0}$ such that

$$g_k^T v_k - v_k^T H_k u_k - u_k^T H_k u_k \leq (\|g_k\|_2 + \kappa_H \|u_k\|_2)\|v_k\|_2 \leq \kappa_{\tau,2}\|c_k\|_1.$$

Together, one has $g_k^T d_k + \max\{d_k^T H_k d_k, 0\} \leq \max\{\kappa_{\tau,1}, \kappa_{\tau,2}\}\|c_k\|_1$, meaning that to have $g_k^T d_k + \max\{d_k^T H_k d_k, 0\} > 0$ and (2.14) requires $\tau_{k-1} > (1-\sigma)/\max\{\kappa_{\tau,1}, \kappa_{\tau,2}\}$. Thus, if this inequality is not satisfied for $k = k_\tau$ for some $k_\tau \in \mathbb{N}$, then it remains unsatisfied for all $k \geq k_\tau$. This, with the fact that when Algorithm 2.1 or 2.2 decreases the merit parameter it does so by at least a constant factor, proves the result. □

We now prove that there is a positive lower bound for the stepsizes.

LEMMA 2.17. *There exists $\alpha_{\min} \in \mathbb{R}_{>0}$ such that $\alpha_k \geq \alpha_{\min}$ for all $k \in \mathbb{N}$.*

Proof. Let $k \in \mathbb{N}$ be given. With respect to Algorithm 2.1, one has that $\alpha_k \geq 1$ unless the inner **for** loop terminates in iteration $j \in \mathbb{N}$ with $\widehat{\alpha}_{k,j} < 1$. In such cases, it follows from the monotonicity of $\{\tau_k\}$ and Lemmas 2.8, 2.16, 2.11, and 2.12 that

$$\alpha_k = \tfrac{2(1-\eta)\Delta q(x_k,\tau_k,g_k,H_k,d_k)}{(\tau_k L_{k,j}+\sum_{i\in[m]}\gamma_{k,i,j})\|d_k\|_2^2} \geq \tfrac{2(1-\eta)\kappa_q\tau_{\min}}{(\tau_{-1}L_{\max}+\sum_{i\in[m]}\gamma_{\max,i})\kappa_\Psi} > 0.$$

Similarly, for Algorithm 2.2, Lemma 2.14 implies $\alpha_k \geq 1$ unless $\widehat{\mu}_k < 1$. In such cases, it follows from the monotonicity of $\{\tau_k\}$ and Lemmas 2.8, 2.16, 2.11, and 2.12 that

$$\alpha_k > \tfrac{2\nu(1-\eta)\Delta q(x_k,\tau_k,g_k,H_k,d_k)}{(\tau_k L+\sum_{i\in[m]}\gamma_i)\|d_k\|_2^2} \geq \tfrac{2\nu(1-\eta)\kappa_q\tau_{\min}}{(\tau_{-1}L+\sum_{i\in[m]}\gamma_i)\kappa_\Psi} > 0.$$

Overall, a positive lower bound has been proved for both algorithms. □

We now present our main convergence theorem for Algorithms 2.1 and 2.2.

THEOREM 2.18. *Algorithms* 2.1 *and* 2.2 *yield*

$$\lim_{k\to\infty}\|d_k\|_2 = 0, \quad \lim_{k\to\infty}\|c_k\|_2 = 0, \quad \textit{and} \quad \lim_{k\to\infty}\|g_k + J_k^T y_k\|_2 = 0.$$

Proof. For all $k \in \mathbb{N}$, it follows from Lemmas 2.12, 2.14, and 2.17 that

$$\phi(x_k, \tau_k) - \phi(x_{k+1}, \tau_k) \geq \eta \alpha_k \Delta q(x_k, \tau_k, g_k, H_k, d_k) \geq \eta \alpha_{\min} \kappa_q \tau_{\min} \Psi_k.$$

Combining this with Lemmas 2.11 and 2.16 shows for $k \in \mathbb{N}$ with $k > k_\tau$ that

$$\begin{aligned} &\phi(x_{k_\tau}, \tau_{\min}) - \phi(x_k, \tau_{\min}) \\ &= \sum_{j=k_\tau}^{k-1} (\phi(x_j, \tau_{\min}) - \phi(x_{j+1}, \tau_{\min})) \\ &\geq \eta \alpha_{\min} \kappa_q \tau_{\min} \sum_{j=k_\tau}^{k-1} \Psi_j \geq \tfrac{\eta \alpha_{\min} \kappa_q \tau_{\min}}{\kappa_\Psi + 1} \sum_{j=k_\tau}^{k-1} (\|d_j\|_2^2 + \|c_j\|_2). \end{aligned}$$

Since, under Assumption 2.1, $\phi(\cdot, \tau_{\min})$ is bounded below over the iterates, the above implies the first two desired limits. Note now that (2.6) implies

$$\|g_k + J_k^T y_k\|_2 = \|H_k d_k\|_2 \leq \|H_k\|_2 \|d_k\|_2 \leq \kappa_H \|d_k\|_2. \tag{2.15}$$

Hence, by Assumption 2.4 and $\{d_k\} \to 0$, the result follows. □

3. Stochastic setting.

Now consider the optimization problem

$$\min_{x \in \mathbb{R}^n} f(x) \ \text{ s.t. } \ c(x) = 0 \ \text{ with } \ f(x) = \mathbb{E}[F(x, \omega)], \tag{3.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $c : \mathbb{R}^n \to \mathbb{R}^m$, ω is a random variable with associated probability space $(\Omega, \mathcal{F}, P)$, $F : \mathbb{R}^n \times \Omega \to \mathbb{R}$, and $\mathbb{E}[\cdot]$ represents expectation taken with respect to P. We presume that one has access to values of the constraint function and its derivatives, but that it is intractable to evaluate the objective and/or its derivatives. That said, we presume that at a given iterate x_k, one can evaluate a stochastic gradient estimate $\bar{g}_k \in \mathbb{R}^n$ satisfying the following assumption.

ASSUMPTION 3.1. *For all $k \in \mathbb{N}$, the stochastic gradient estimate $\bar{g}_k \in \mathbb{R}^n$ is an unbiased estimator of the gradient of f at x_k, i.e., $\mathbb{E}_k[\bar{g}_k] = g_k$, where $\mathbb{E}_k[\cdot]$ denotes expectation taken with respect to the distribution of ω conditioned on the event that the algorithm has reached $x_k \in \mathbb{R}^n$ in iteration $k \in \mathbb{N}$. In addition, there exists a constant $M \in \mathbb{R}_{>0}$ such that, for all $k \in \mathbb{N}$, one has $\mathbb{E}_k[\|\bar{g}_k - g_k\|_2^2] \leq M$.*

3.1. Algorithm.

Similar to the deterministic setting, in order to solve (3.1), we consider a stochastic algorithm that computes a search direction $\bar{d}_k \in \mathbb{R}^n$ and Lagrange multiplier vector $\bar{y}_k \in \mathbb{R}^m$ in iteration $k \in \mathbb{N}$ by solving the linear system

$$\begin{bmatrix} H_k & J_k^T \\ J_k & 0 \end{bmatrix} \begin{bmatrix} \bar{d}_k \\ \bar{y}_k \end{bmatrix} = - \begin{bmatrix} \bar{g}_k \\ c_k \end{bmatrix}, \tag{3.2}$$

where $\{H_k\}$ satisfies Assumption 2.4. Generally, we use a "bar" over a quantity whose value in iteration $k \in \mathbb{N}$ depends on $\bar{g}_k$. Hence, as they are independent of $\bar{g}_k$ conditioned on the event that the algorithm reaches x_k as its kth iterate, we write the constraint value, constraint Jacobian, and $(1,1)$-block matrix as c_k, J_k, and H_k, respectively, but we write the solution of (3.2) as $(\bar{d}_k, \bar{y}_k)$ due to its dependence on $\bar{g}_k$.

The algorithm that we propose is stated as Algorithm 3.1. Paralleling Algorithm 2.1, the merit parameter is set based on the computation of a trial value

$$\bar{\tau}_k^{trial} \leftarrow \begin{cases} \infty & \text{if } \bar{g}_k^T \bar{d}_k + \max\{\bar{d}_k^T H_k \bar{d}_k, 0\} \leq 0, \\ \frac{(1-\sigma)\|c_k\|_1}{\bar{g}_k^T \bar{d}_k + \max\{\bar{d}_k^T H_k \bar{d}_k, 0\}} & \text{otherwise,} \end{cases} \tag{3.3}$$

Algorithm 3.1 Stochastic SQP Algorithm.

```
Require: $x_0 \in \mathbb{R}^n$; $\bar{\tau}_{-1} \in \mathbb{R}_{>0}$; $\epsilon \in (0,1)$; $\sigma \in (0,1)$; $\bar{\xi}_{-1} \in \mathbb{R}_{>0}$; $\{\beta_k\} \subset (0,1]$;
    $\theta \in \mathbb{R}_{\geq 0}$; $\{L_k\} \subset \mathbb{R}_{>0}$; $\{\Gamma_k\} \subset \mathbb{R}_{>0}$
for all $k \in \mathbb{N}$ do
    Compute $(\bar{d}_k, \bar{y}_k)$ as the solution of (3.2)
    if $\bar{d}_k = 0$ then
        Set $\bar{\tau}_k^{trial} \leftarrow \infty$, $\bar{\tau}_k \leftarrow \bar{\tau}_{k-1}$, $\bar{\xi}_k^{trial} \leftarrow \infty$, and $\bar{\xi}_k \leftarrow \bar{\xi}_{k-1}$
        Set $\bar{\hat{\alpha}}_{k,\mathrm{init}} \leftarrow 1$, $\bar{\tilde{\alpha}}_{k,\mathrm{init}} \leftarrow 1$, and $\bar{\alpha}_k \leftarrow 1$
    else (if $\bar{d}_k \neq 0$)
        Set $\bar{\tau}_k^{trial}$ by (3.3) and $\bar{\tau}_k$ by (3.4)
        Set $\bar{\xi}_k^{trial}$ and $\bar{\xi}_k$ by (3.6)
        Set

            $\bar{\hat{\alpha}}_{k,\mathrm{init}} \leftarrow \frac{\beta_k \Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k)}{(\bar{\tau}_k L_k + \Gamma_k)\|\bar{d}_k\|_2^2}$ and $\bar{\tilde{\alpha}}_{k,\mathrm{init}} \leftarrow \bar{\hat{\alpha}}_{k,\mathrm{init}} - \frac{4\|c_k\|_1}{(\bar{\tau}_k L_k + \Gamma_k)\|\bar{d}_k\|_2^2}$

        Set $\bar{\hat{\alpha}}_k \leftarrow \mathrm{Proj}_k(\bar{\hat{\alpha}}_{k,\mathrm{init}})$ and $\bar{\tilde{\alpha}}_k \leftarrow \mathrm{Proj}_k(\bar{\tilde{\alpha}}_{k,\mathrm{init}})$, then

            $\bar{\alpha}_k \leftarrow \begin{cases} \bar{\hat{\alpha}}_k & \text{if } \bar{\hat{\alpha}}_k < 1, \\ 1 & \text{if } \bar{\tilde{\alpha}}_k \leq 1 \leq \bar{\hat{\alpha}}_k, \\ \bar{\tilde{\alpha}}_k & \text{if } \bar{\tilde{\alpha}}_k > 1 \end{cases}$

    Set $x_{k+1} \leftarrow x_k + \bar{\alpha}_k \bar{d}_k$
```

followed by the rule

$$\bar{\tau}_k \leftarrow \begin{cases} \bar{\tau}_{k-1} & \text{if } \bar{\tau}_{k-1} \leq \bar{\tau}_k^{trial}, \\ (1-\epsilon)\bar{\tau}_k^{trial} & \text{otherwise}, \end{cases} \tag{3.4}$$

which ensures $\bar{\tau}_k \leq \bar{\tau}_k^{trial}$ and, similarly as for our deterministic algorithm (see (2.9)),

$$\Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k) \geq \tfrac{1}{2}\bar{\tau}_k \max\{\bar{d}_k^T H_k \bar{d}_k, 0\} + \sigma\|c_k\|_1. \tag{3.5}$$

A unique feature of our algorithm for this stochastic setting is that it adaptively estimates a lower bound for the ratio between the reduction in the model of the merit function and the merit parameter times the squared norm of a search direction. This is used to determine an interval into which the stepsize will be projected; control of this parameter is paramount to ensure convergence in expectation. We set

$$\bar{\xi}_k^{trial} \leftarrow \tfrac{\Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k)}{\bar{\tau}_k \|\bar{d}_k\|_2^2} \ \text{ then } \ \bar{\xi}_k \leftarrow \begin{cases} \bar{\xi}_{k-1} & \text{if } \bar{\xi}_{k-1} \leq \bar{\xi}_k^{trial}, \\ (1-\epsilon)\bar{\xi}_k^{trial} & \text{otherwise}, \end{cases} \tag{3.6}$$

which ensures $\bar{\xi}_k \leq \bar{\xi}_k^{trial}$ for all $k \in \mathbb{N}$. It will be shown in our analysis that $\{\bar{\xi}_k\}$ is bounded away from zero *deterministically*.

The sequences $\{\bar{\tau}_k\}$ and $\{\bar{\xi}_k\}$ are initialized with the input values $\bar{\tau}_{-1} \in \mathbb{R}_{>0}$ and $\bar{\xi}_{-1} \in \mathbb{R}_{>0}$, respectively. These values may be set deterministically, but we use a "bar" over these initial values for consistency with the remainders of the sequences.

For generality, Algorithm 3.1 is stated with Lipschitz constant estimates $\{L_k\}$ and $\{\Gamma_k\}$ given as inputs (with the idea that $\Gamma_k := \sum_{i \in [m]} \gamma_{k,i}$ for all $k \in \mathbb{N}$). Our analysis in the next subsection presumes that Lipschitz constants are known, although in practice these can be estimated using standard techniques (see, e.g., [12]) in an attempt to ensure that the same convergence results hold as for the case when the constants are known. The sequence $\{\beta_k\}$ is introduced to control the stepsizes. As in standard analysis for stochastic (sub)gradient-type methods, our analysis in the next subsection considers the case when $\{\beta_k\}$ is constant asymptotically and when it diminishes at an appropriate rate to ensure convergence in expectation. We define

$$\text{Proj}_k(\cdot) \equiv \text{Proj}\left(\cdot \;\middle|\; \left[\tfrac{\beta_k \bar{\xi}_k \bar{\tau}_k}{\bar{\tau}_k L_k + \Gamma_k}, \tfrac{\beta_k \bar{\xi}_k \bar{\tau}_k}{\bar{\tau}_k L_k + \Gamma_k} + \theta \beta_k^2\right]\right),$$

where $\text{Proj}(\cdot \mid \mathcal{I})$ represents the projection operator onto the interval $\mathcal{I} \subset \mathbb{R}$.

3.2. Convergence analysis.

In this section, we prove that Algorithm 3.1 has convergence properties that match those from the deterministic setting in expectation, with some caveats that we explain and justify. Our algorithm uses only the stochastic gradient estimates $\{\bar{g}_k\}$, computes $\{(\bar{d}_k, \bar{y}_k)\}$ by (3.2), sets merit parameter-related sequences $\{\bar{\tau}_k\}$ and $\{\bar{\tau}_k^{trial}\}$, and also sets steplength-related sequences $\{\bar{\xi}_k\}$ and $\{\bar{\xi}_k^{trial}\}$, but our analysis also references the gradients $\{g_k\}$ corresponding to $\{x_k\}$ as well as the corresponding sequence of solutions of (2.6), namely, $\{(d_k, y_k)\}$, and trial merit parameter values $\{\tau_k^{trial}\}$. In other words, for all $k \in \mathbb{N}$, conditioned on the event that the algorithm reaches x_k, we define (d_k, y_k) and τ_k^{trial} as they would be computed if the algorithm reached x_k as the kth iterate in Algorithm 2.1.

We assume throughout this section that L and $\Gamma := \sum_{i \in [m]} \gamma_i$ are known. In addition, we assume that Assumptions 2.1, 2.4, and 3.1 hold—where $\{H_k\}$ is a deterministic sequence chosen independently from $\{\bar{g}_k\}$—and for the sake of brevity we do not state this fact within each result. Explicitly, in addition to Assumption 3.1, we make the following assumption that subsumes Assumptions 2.1 and 2.4.

ASSUMPTION 3.2. *There exist universal quantities (including $\mathcal{X}$, L, $\{\gamma_i\}_{i \in [m]}$, κ_H, and ζ) such that Assumptions* 2.1 *and* 2.4 *hold for any realization of Algorithm* 3.1.

Remark 3.3. Assumption 3.2 subsumes Assumption 2.1, which means that it assumes that the iterates remain in an open convex set over which the objective and constraint function and derivative values are bounded. This is admittedly not ideal in a stochastic setting. For example, in the case of applying a stochastic gradient method (SG) in an unconstrained stochastic setting, it is not ideal to assume that the gradients at the iterates remain bounded in norm, since—as SG is not a descent method—it is unreasonable to assume that the iterates remain in a sublevel set of the objective function. However, we believe this assumption is more reasonable in a constrained setting, since the iterates are being driven to the *deterministic* feasible region. Further, we claim that Assumption 3.2 could be loosened if our algorithm were to choose a predetermined stepsize sequence, rather than one that mimicks the stepsize scheme from Algorithm 2.1. We discuss this issue further in section 5.

As in the deterministic setting, our analysis makes use of the orthogonal decomposition of the (stochastic) search directions given by $\bar{d}_k = \bar{u}_k + v_k$, where $\bar{u}_k \in \text{Null}(J_k)$ and $v_k \in \text{Range}(J_k^T)$ for all $k \in \mathbb{N}$. Let us emphasize that, conditioned on the event that the algorithm reaches x_k as its kth iterate, the normal component is *deterministic*, depending only on the constraint value c_k and Jacobian J_k; hence, we write v_k rather than $\bar{v}_k$ in the expression above. For all $k \in \mathbb{N}$, let Z_k be an orthogonal basis for the null space of J_k, which under Assumption 3.2 is a matrix in $\mathbb{R}^{n \times (n-m)}$. It follows that, for all $k \in \mathbb{N}$, $\bar{u}_k = Z_k \bar{w}_k$ and $u_k = Z_k w_k$ for some $(\bar{w}_k, w_k) \in \mathbb{R}^{n-m} \times \mathbb{R}^{n-m}$. Under Assumption 3.2, the reduced Hessian satisfies $Z_k^T H_k Z_k \succeq \zeta I$.

For our first lemma, we carry over properties of algorithmic quantities that hold in the same manner as in the deterministic case, conditioned on the event that the algorithm has reached x_k as the kth iterate. As in our analysis in the deterministic setting, for the constant $\kappa_{uv} \in \mathbb{R}_{>0}$ defined in the lemma, we define

$$\overline{\Psi}_k := \begin{cases} \|\bar{u}_k\|_2^2 + \|c_k\|_2 & \text{if } \|\bar{u}_k\|_2^2 \geq \kappa_{uv}\|v_k\|_2^2, \\ \|c_k\|_2 & \text{otherwise.} \end{cases}$$

Lemma 3.4. *For all $k \in \mathbb{N}$, (3.2) has a unique solution. In addition, due to the universality of the algorithmic conditions as described in Assumption 3.2, for the same constants $(\kappa_v, \kappa_{uv}, \kappa_\Psi, \kappa_q) \in \mathbb{R}_{>0} \times \mathbb{R}_{>0} \times \mathbb{R}_{>0} \times \mathbb{R}_{>0}$ that appear in Lemmas 2.9, 2.10, 2.11, and 2.12, the following statements hold true for all $k \in \mathbb{N}$.*

(a) *The normal component satisfies $\max\{\|v_k\|_2, \|v_k\|_2^2\} \leq \kappa_v \|c_k\|_2$.*

(b) *If $\|\bar{u}_k\|_2^2 \geq \kappa_{uv}\|v_k\|_2^2$, then $\frac{1}{2}\bar{d}_k^T H_k \bar{d}_k \geq \frac{1}{4}\zeta\|\bar{u}_k\|_2^2$.*

(c) *The search direction satisfies $\|\bar{d}_k\|_2^2 \leq \kappa_\Psi \overline{\Psi}_k$ and $\|\bar{d}_k\|_2^2 + \|c_k\|_2 \leq (\kappa_\Psi + 1)\overline{\Psi}_k$.*

(d) *The model reduction satisfies $\Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k) \geq \kappa_q \bar{\tau}_k \overline{\Psi}_k$.*

Finally, for all $k \in \mathbb{N}$, it follows that

$$\phi(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_k) - \phi(x_k, \bar{\tau}_k) \leq \bar{\alpha}_k \bar{\tau}_k g_k^T \bar{d}_k + |1 - \bar{\alpha}_k| \|c_k\|_1 - \|c_k\|_1 + \tfrac{1}{2}(\bar{\tau}_k L + \Gamma)\bar{\alpha}_k^2 \|\bar{d}_k\|_2^2.$$

Proof. That (3.2) has a unique solution for all $k \in \mathbb{N}$ follows for the same reason that Lemma 2.5 holds. The proofs of parts (a)–(d) follow in the same manner as the proofs of Lemmas 2.9, 2.10, 2.11, and 2.12, respectively, with the stochastic quantities $\{\bar{g}_k, \bar{d}_k, \bar{u}_k, \bar{\tau}_k\}$ in place of the deterministic quantities $\{g_k, d_k, u_k, \tau_k\}$, where it is important to recognize that the conclusions follow with the *same constants*, namely, $(\kappa_v, \kappa_{uv}, \kappa_\Psi, \kappa_q)$, as in the deterministic setting. The proof of the last conclusion follows in the same manner as that of Lemma 2.13. □

In the next lemma, we prove that the sequence $\{\bar{\xi}_k\}$ is bounded deterministically.

Lemma 3.5. *In any run of the algorithm, there exist $\bar{k}_\xi \in \mathbb{N}$ and $\bar{\xi}_{\min} \in \mathbb{R}_{>0}$ such that $\bar{\xi}_k = \bar{\xi}_{\min}$ for all $k \geq \bar{k}_\xi$, where $\bar{\xi}_{\min} \in [\xi_{\min}, \bar{\xi}_{-1}]$ with $\xi_{\min} := (1-\epsilon)\kappa_q/\kappa_\Psi$.*

Proof. If line 8 of the algorithm ever sets $\bar{\xi}_k < \bar{\xi}_{k-1}$, then it ensures that $\bar{\xi}_k \leq (1-\epsilon)\bar{\xi}_{k-1}$. This means that $\{\bar{\xi}_k\}$ is constant for sufficiently large k or it vanishes. On the other hand, by Lemma 3.4(c) and (d), it follows that $\frac{\Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k)}{\bar{\tau}_k \|\bar{d}_k\|_2^2} \geq \frac{\kappa_q \bar{\tau}_k \overline{\Psi}_k}{\kappa_\Psi \bar{\tau}_k \overline{\Psi}_k} = \frac{\kappa_q}{\kappa_\Psi}$, meaning that line 8 will never set $\bar{\xi}_k$ less than $(1-\epsilon)\kappa_q/\kappa_\Psi$ for any $k \in \mathbb{N}$. Therefore, $\{\bar{\xi}_k\}$ is constant for sufficiently large k in the manner stated. □

Next, we present the following obvious, but important, consequence of our stepsize selection scheme. In particular, the result shows that, even though the algorithm sets the stepsize adaptively, the difference between the largest and smallest possible stepsizes in a given iteration is $\mathcal{O}(\beta_k^2)$, so this difference is controlled by the algorithm.

Lemma 3.6. *For all $k \in \mathbb{N}$, $\bar{\alpha}_k \in [\bar{\alpha}_{k,\min}, \bar{\alpha}_{k,\max}] := \left[\frac{\beta_k \bar{\xi}_k \bar{\tau}_k}{\bar{\tau}_k L + \Gamma}, \frac{\beta_k \bar{\xi}_k \bar{\tau}_k}{\bar{\tau}_k L + \Gamma} + \theta \beta_k^2\right]$.*

Proof. The proof follows directly from the formula for $\bar{\alpha}_k$ in line 10. □

Our next result is a cornerstone of our analysis. It builds on the last conclusion in Lemma 3.4 to specify a useful upper bound for the merit function value after a step. Central to the proof is our specific stepsize selection strategy.

LEMMA 3.7. *Suppose that* $\{\beta_k\}$ *is chosen such that* $\beta_k\bar{\xi}_k\bar{\tau}_k/(\bar{\tau}_k L + \Gamma) \in (0,1]$ *for all* $k \in \mathbb{N}$. *Then, for all* $k \in \mathbb{N}$, *it follows that*

$$\begin{aligned}&\phi(x_k + \bar{\alpha}_k\bar{d}_k, \bar{\tau}_k) - \phi(x_k, \bar{\tau}_k)\\ \leq\ &-\bar{\alpha}_k\Delta q(x_k, \bar{\tau}_k, g_k, H_k, d_k) + \tfrac{1}{2}\bar{\alpha}_k\beta_k\Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k) + \bar{\alpha}_k\bar{\tau}_k g_k^T(\bar{d}_k - d_k),\end{aligned}$$

where $\bar{d}_k$ *and* d_k *are, respectively, defined by* (3.2) *and* (2.6).

Proof. Let $k \in \mathbb{N}$ be arbitrary. If $\bar{d}_k = 0$, then it follows from (3.2) that $c_k = 0$. Hence, along with (2.4), it follows that

$$-\bar{\alpha}_k\Delta q(x_k, \bar{\tau}_k, g_k, H_k, d_k) - \bar{\alpha}_k\bar{\tau}_k g_k^T d_k = \tfrac{1}{2}\bar{\alpha}_k\bar{\tau}_k \max\{d_k^T H_k d_k, 0\} \geq 0.$$

Along with the fact that $\bar{d}_k = 0$ implies that

$$\phi(x_k + \bar{\alpha}_k\bar{d}_k, \bar{\tau}_k) - \phi(x_k, \bar{\tau}_k) = 0 = \tfrac{1}{2}\bar{\alpha}_k\beta_k\Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k) + \bar{\alpha}_k\bar{\tau}_k g_k^T\bar{d}_k,$$

the desired conclusion follows. On the other hand, if $\bar{d}_k \neq 0$, we consider three cases, with a few subcases, depending on how the stepsize is set in line 10 of the algorithm.

Case 1: Suppose in line 10 that $\bar{\hat{\alpha}}_k < 1$, meaning that $\bar{\alpha}_k \leftarrow \bar{\hat{\alpha}}_k$. From Lemmas 3.4 and 2.3, it follows that

$$\begin{aligned}&\phi(x_k + \bar{\alpha}_k\bar{d}_k, \bar{\tau}_k) - \phi(x_k, \bar{\tau}_k)\\ &\leq \bar{\alpha}_k(\bar{\tau}_k g_k^T\bar{d}_k - \|c_k\|_1) + \tfrac{1}{2}(\bar{\tau}_k L + \Gamma)\bar{\alpha}_k^2\|\bar{d}_k\|_2^2\\ &= \bar{\alpha}_k(\bar{\tau}_k g_k^T d_k - \|c_k\|_1) + \tfrac{1}{2}(\bar{\tau}_k L + \Gamma)\bar{\alpha}_k^2\|\bar{d}_k\|_2^2 + \bar{\alpha}_k\bar{\tau}_k g_k^T(\bar{d}_k - d_k)\\ &\leq\ -\bar{\alpha}_k\Delta q(x_k, \bar{\tau}_k, g_k, H_k, d_k) + \tfrac{1}{2}(\bar{\tau}_k L + \Gamma)\bar{\alpha}_k^2\|\bar{d}_k\|_2^2 + \bar{\alpha}_k\bar{\tau}_k g_k^T(\bar{d}_k - d_k).\end{aligned}$$

Using this inequality, let us now consider two subcases. (For all $k \in \mathbb{N}$, since (3.6) ensures $\bar{\xi}_k \leq \bar{\xi}_k^{trial} = \frac{\Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k)}{\bar{\tau}_k\|\bar{d}_k\|_2^2}$, it follows that $\frac{\beta_k\Delta q(x_k, \bar{\tau}, \bar{g}_k, H_k, \bar{d}_k)}{(\bar{\tau}_k L + \Gamma)\|\bar{d}_k\|_2^2} \geq \frac{\beta_k\bar{\xi}_k\bar{\tau}_k}{\bar{\tau}_k L + \Gamma}$.)

Case 1a: If $\bar{\alpha}_k = \frac{\beta_k\Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k)}{(\bar{\tau}_k L + \Gamma)\|\bar{d}_k\|_2^2}$, then

$$\begin{aligned}&\phi(x_k + \bar{\alpha}_k\bar{d}_k, \bar{\tau}_k) - \phi(x_k, \bar{\tau}_k)\\ \leq\ &-\bar{\alpha}_k\Delta q(x_k, \bar{\tau}_k, g_k, H_k, d_k)\\ &\quad + \tfrac{1}{2}\bar{\alpha}_k(\bar{\tau}_k L + \Gamma)\left(\tfrac{\beta_k\Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k)}{(\bar{\tau}_k L + \Gamma)\|\bar{d}_k\|_2^2}\right)\|\bar{d}_k\|_2^2 + \bar{\alpha}_k\bar{\tau}_k g_k^T(\bar{d}_k - d_k)\\ =\ &-\bar{\alpha}_k\Delta q(x_k, \bar{\tau}_k, g_k, H_k, d_k) + \tfrac{1}{2}\bar{\alpha}_k\beta_k\Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k) + \bar{\alpha}_k\bar{\tau}_k g_k^T(\bar{d}_k - d_k).\end{aligned}$$

Case 1b: If $\bar{\alpha}_k = \frac{\beta_k\bar{\xi}_k\bar{\tau}_k}{\bar{\tau}_k L + \Gamma} + \theta\beta_k^2$, then the same argument applies as in Case 1a by plugging in for $\bar{\alpha}_k$ and using the fact that $\bar{\alpha}_k \leq \frac{\beta_k\Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k)}{(\bar{\tau}_k L + \Gamma)\|\bar{d}_k\|_2^2}$.

Case 2: Suppose in line 10 that $\bar{\tilde{\alpha}}_k \leq 1 \leq \bar{\hat{\alpha}}_k$, meaning that $\bar{\alpha}_k \leftarrow 1$. From Lemmas 3.4 and 2.3 and since $\frac{\beta_k\Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k)}{(\bar{\tau}_k L + \Gamma)\|\bar{d}_k\|_2^2} \geq 1 = \bar{\alpha}_k$, it follows that

$$\begin{aligned}&\phi(x_k + \bar{\alpha}_k\bar{d}_k, \bar{\tau}_k) - \phi(x_k, \bar{\tau}_k)\\ &\leq \bar{\alpha}_k(\bar{\tau}_k g_k^T\bar{d}_k - \|c_k\|_1) + \tfrac{1}{2}(\bar{\tau}_k L + \Gamma)\bar{\alpha}_k^2\|\bar{d}_k\|_2^2\\ &= \bar{\alpha}_k(\bar{\tau}_k g_k^T d_k - \|c_k\|_1) + \tfrac{1}{2}(\bar{\tau}_k L + \Gamma)\bar{\alpha}_k^2\|\bar{d}_k\|_2^2 + \bar{\alpha}_k\bar{\tau}_k g_k^T(\bar{d}_k - d_k)\\ &\leq\ -\bar{\alpha}_k\Delta q(x_k, \bar{\tau}_k, g_k, H_k, d_k) + \tfrac{1}{2}(\bar{\tau}_k L + \Gamma)\bar{\alpha}_k^2\|\bar{d}_k\|_2^2 + \bar{\alpha}_k\bar{\tau}_k g_k^T(\bar{d}_k - d_k)\\ &\leq\ -\bar{\alpha}_k\Delta q(x_k, \bar{\tau}_k, g_k, H_k, d_k) + \tfrac{1}{2}\bar{\alpha}_k\beta_k\Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k) + \bar{\alpha}_k\bar{\tau}_k g_k^T(\bar{d}_k - d_k).\end{aligned}$$

Case 3: Suppose in line 10 that $\bar{\bar{\alpha}}_k > 1$, meaning that $\bar{\alpha}_k \leftarrow \bar{\bar{\alpha}}_k$. From Lemmas 3.4 and 2.3, it follows that

$$\begin{aligned}
&\phi(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_k) - \phi(x_k, \bar{\tau}_k) \\
&\leq \bar{\alpha}_k \bar{\tau}_k g_k^T \bar{d}_k + (\bar{\alpha}_k - 1)\|c_k\|_1 - \|c_k\|_1 + \tfrac{1}{2}(\bar{\tau}_k L + \Gamma)\bar{\alpha}_k^2 \|\bar{d}_k\|_2^2 \\
&= \bar{\alpha}_k(\bar{\tau}_k g_k^T \bar{d}_k - \|c_k\|_1) + 2(\bar{\alpha}_k - 1)\|c_k\|_1 + \tfrac{1}{2}(\bar{\tau}_k L + \Gamma)\bar{\alpha}_k^2 \|\bar{d}_k\|_2^2 \\
&\leq \bar{\alpha}_k(\bar{\tau}_k g_k^T d_k - \|c_k\|_1) + 2\bar{\alpha}_k\|c_k\|_1 + \tfrac{1}{2}(\bar{\tau}_k L + \Gamma)\bar{\alpha}_k^2 \|\bar{d}_k\|_2^2 + \bar{\alpha}_k \bar{\tau}_k g_k^T(\bar{d}_k - d_k) \\
&\leq \ -\bar{\alpha}_k \Delta q(x_k, \bar{\tau}_k, g_k, H_k, d_k) + 2\bar{\alpha}_k\|c_k\|_1 + \tfrac{1}{2}(\bar{\tau}_k L + \Gamma)\bar{\alpha}_k^2 \|\bar{d}_k\|_2^2 + \bar{\alpha}_k \bar{\tau}_k g_k^T(\bar{d}_k - d_k).
\end{aligned}$$

Using this inequality, let us now consider two subcases. (Since the lemma requires $1 \geq \frac{\beta_k \bar{\xi}_k \bar{\tau}_k}{\bar{\tau}_k L + \Gamma}$ for all $k \in \mathbb{N}$, it is not possible that $\bar{\alpha}_k = \frac{\beta_k \bar{\xi}_k \bar{\tau}_k}{\bar{\tau}_k L + \Gamma}$ in this case.)

Case 3a: If $\bar{\alpha}_k = \frac{\beta_k \Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k) - 4\|c_k\|_1}{(\bar{\tau}_k L + \Gamma)\|\bar{d}_k\|_2^2}$, then

$$\begin{aligned}
&\phi(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_k) - \phi(x_k, \bar{\tau}_k) \\
&\leq \ -\bar{\alpha}_k \Delta q(x_k, \bar{\tau}_k, g_k, H_k, d_k) + 2\bar{\alpha}_k\|c_k\|_1 \\
&\qquad + \tfrac{1}{2}\bar{\alpha}_k(\bar{\tau}_k L + \Gamma)\left(\tfrac{\beta_k \Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k) - 4\|c_k\|_1}{(\bar{\tau}_k L + \Gamma)\|\bar{d}_k\|_2^2}\right)\|\bar{d}_k\|_2^2 + \bar{\alpha}_k \bar{\tau}_k g_k^T(\bar{d}_k - d_k) \\
&= \ -\bar{\alpha}_k \Delta q(x_k, \bar{\tau}_k, g_k, H_k, d_k) + \tfrac{1}{2}\bar{\alpha}_k \beta_k \Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k) + \bar{\alpha}_k \bar{\tau}_k g_k^T(\bar{d}_k - d_k).
\end{aligned}$$

Case 3b: If $\bar{\alpha}_k = \frac{\beta_k \bar{\xi}_k \bar{\tau}_k}{\bar{\tau}_k L + \Gamma} + \theta\beta_k^2$, then the same argument applies as in Case 3a by plugging in for $\bar{\alpha}_k$ and using the fact that $\bar{\alpha}_k \leq \frac{\beta_k \Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k) - 4\|c_k\|_1}{(\bar{\tau}_k L + \Gamma)\|\bar{d}_k\|_2^2}$.

The result follows by combining the conclusions of all cases and subcases. ∎

Our next two lemmas provide useful relationships between deterministic (i.e., dependent on g_k) and stochastic (i.e., dependent on $\bar{g}_k$) quantities.

LEMMA 3.8. *For all $k \in \mathbb{N}$, $\mathbb{E}_k[\bar{d}_k] = d_k$, $\mathbb{E}_k[\bar{u}_k] = u_k$, and $\mathbb{E}_k[\bar{y}_k] = y_k$. Moreover, there exists $\kappa_d \in \mathbb{R}_{>0}$, independent of k, such that $\mathbb{E}_k[\|\bar{d}_k - d_k\|_2] \leq \kappa_d\sqrt{M}$.*

Proof. The first statement follows from the fact that, conditioned on the kth iterate being x_k, the matrix on the left-hand side of (3.2) is deterministic and, under Assumption 3.2, it is invertible, along with the fact that expectation is a linear operator. For the second statement, notice that for any realization of $\bar{g}_k$, it follows that

$$\begin{bmatrix} \bar{d}_k - d_k \\ \bar{y}_k - y_k \end{bmatrix} = -\begin{bmatrix} H_k & J_k^T \\ J_k & 0 \end{bmatrix}^{-1} \begin{bmatrix} \bar{g}_k - g_k \\ 0 \end{bmatrix} \implies \|\bar{d}_k - d_k\|_2 \leq \kappa_d \|\bar{g}_k - g_k\|_2,$$

where (under Assumption 3.2) $\kappa_d \in \mathbb{R}_{>0}$ is an upper bound on the norm of the matrix shown above. It also follows from Jensen's inequality, concavity of the square root, and Assumption 3.1 that $\mathbb{E}_k[\|\bar{g}_k - g_k\|_2] \leq \sqrt{\mathbb{E}_k[\|\bar{g}_k - g_k\|_2^2]} \leq \sqrt{M}$. Combined with the displayed inequality above, the desired conclusion follows. ∎

Relationships between inner products involving deterministic and stochastic quantities are the subject of the next lemma.

LEMMA 3.9. *For all $k \in \mathbb{N}$, it follows that*

$$g_k^T d_k \geq \mathbb{E}_k[\bar{g}_k^T \bar{d}_k] \geq g_k^T d_k - \zeta^{-1}M \quad \textit{and} \quad d_k^T H_k d_k \leq \mathbb{E}_k[\bar{d}_k^T H_k \bar{d}_k].$$

Proof. From the first block equation in (3.2), it follows that

$$H_k(Z_k\overline{w}_k + v_k) + J_k^T\overline{y}_k = -\overline{g}_k \iff Z_k\overline{w}_k = -Z_k(Z_k^TH_kZ_k)^{-1}Z_k^T(\overline{g}_k + H_kv_k),$$

from which it follows that

$$\overline{g}_k^T\overline{u}_k = \overline{g}_k^TZ_k\overline{w}_k = -\overline{g}_k^TZ_k(Z_k^TH_kZ_k)^{-1}Z_k^T(\overline{g}_k + H_kv_k). \tag{3.7}$$

Following the same line of argument for (2.6), it follows that

$$g_k^Tu_k = -g_k^TZ_k(Z_k^TH_kZ_k)^{-1}Z_k^T(g_k + H_kv_k). \tag{3.8}$$

At the same time, under Assumptions 3.2 and 3.1, one finds that

$$\zeta^{-1}M \geq \mathbb{E}_k[\|Z_k^T(\overline{g}_k - g_k)\|^2_{(Z_k^TH_kZ_k)^{-1}}] \geq 0. \tag{3.9}$$

One finds that the middle term in this expression can be written as

$$\begin{aligned}&\mathbb{E}_k[\|Z_k^T(\overline{g}_k - g_k)\|^2_{(Z_k^TH_kZ_k)^{-1}}]\\ &= \mathbb{E}_k[\|Z_k^T\overline{g}_k\|^2_{(Z_k^TH_kZ_k)^{-1}}] - 2\mathbb{E}_k[\overline{g}_k^TZ_k(Z_k^TH_kZ_k)^{-1}Z_k^Tg_k] + \|Z_k^Tg_k\|^2_{(Z_k^TH_kZ_k)^{-1}}\\ &= \mathbb{E}_k[\|Z_k^T\overline{g}_k\|^2_{(Z_k^TH_kZ_k)^{-1}}] - \|Z_k^Tg_k\|^2_{(Z_k^TH_kZ_k)^{-1}}.\end{aligned}$$

Hence, combining (3.7), (3.8), (3.9), and the fact that $\mathbb{E}_k[\overline{g}_k] = g_k$ one finds

$$\begin{aligned}g_k^Tu_k - \mathbb{E}_k[\overline{g}_k^T\overline{u}_k] &= -g_k^TZ_k(Z_k^TH_kZ_k)^{-1}Z_k^T(g_k + H_kv_k)\\ &\quad + \mathbb{E}_k[\overline{g}_k^TZ_k(Z_k^TH_kZ_k)^{-1}Z_k^T(\overline{g}_k + H_kv_k)]\\ &= -\|Z_k^Tg_k\|^2_{(Z_k^TH_kZ_k)^{-1}} + \mathbb{E}_k[\|Z_k^T\overline{g}_k\|^2_{(Z_k^TH_kZ_k)^{-1}}] \in [0, \zeta^{-1}M].\end{aligned}$$

The first desired result follows from this fact, $\mathbb{E}_k[\overline{g}_k^Tv_k] = g_k^Tv_k$, and

$$g_k^Td_k - \mathbb{E}_k[\overline{g}_k^T\overline{d}_k] = g_k^Tu_k + g_k^Tv_k - \mathbb{E}_k[\overline{g}_k^T\overline{u}_k + \overline{g}_k^Tv_k] = g_k^Tu_k - \mathbb{E}_k[\overline{g}_k^T\overline{u}_k].$$

Now let us prove the second desired conclusion. From (3.2), it follows that

$$H_k(\overline{u}_k + v_k) = -\overline{g}_k - J_k^T\overline{y}_k \iff (\overline{u}_k + v_k)^TH_k(\overline{u}_k + v_k) = -\overline{g}_k^T(\overline{u}_k + v_k) + \overline{y}_k^Tc_k.$$

Following the same argument for (2.6), it follows that

$$(u_k + v_k)^TH_k(u_k + v_k) = -g_k^T(u_k + v_k) + y_k^Tc_k.$$

Combining these facts, it follows that

$$\overline{u}_k^TH_k\overline{u}_k + 2(\overline{u}_k - u_k)^TH_kv_k - u_k^TH_ku_k = -\overline{g}_k^T(\overline{u}_k + v_k) + g_k^T(u_k + v_k) + (\overline{y}_k - y_k)^Tc_k,$$

which after taking conditional expectation and using Lemma 3.8 yields

$$\mathbb{E}_k[\overline{u}_k^TH_k\overline{u}_k] - u_k^TH_ku_k = -\mathbb{E}_k[\overline{g}_k^T\overline{u}_k] + g_k^Tu_k.$$

The desired conclusion now follows since

$$\begin{aligned}\mathbb{E}_k[\overline{d}_k^TH_k\overline{d}_k] - d_k^TH_kd_k &= \mathbb{E}_k[(\overline{u}_k + v_k)^TH_k(\overline{u}_k + v_k)] - (u_k + v_k)^TH_k(u_k + v_k)\\ &= \mathbb{E}_k[\overline{u}_k^TH_k\overline{u}_k] - u_k^TH_ku_k,\end{aligned}$$

where again we have used the result of Lemma 3.8. ∎

In the remainder of our convergence analysis, we consider three cases depending on the behavior of the sequence $\{\bar{\tau}_k\}$ in a run of the algorithm. In the deterministic setting, it was proved that the merit parameter sequence eventually remains constant at a value that is sufficiently small to ensure that a primal-dual stationarity measure vanishes (see Lemma 2.16). However, under only Assumption 3.1, it is not possible to prove that such behavior is guaranteed for any possible run of Algorithm 3.1. Our analysis considers three mutually exclusive and exhaustive events: event $E_{\tau,\text{low}}$ that the merit parameter sequence eventually remains constant at a sufficiently small positive value; event $E_{\tau\searrow 0}$ that the merit parameter sequence vanishes; and event $E_{\tau\gg 0}$ that the merit parameter sequence eventually remains constant, but at a value that is not sufficiently small. Under modest assumptions, we prove that $E_{\tau\gg 0}$ occurs with probability zero, and under a stronger, yet pragmatic assumption, we prove that event $E_{\tau\searrow 0}$ does not occur. This leaves event $E_{\tau,\text{low}}$, which we consider first and show that, conditioned on this event, convergence comparable to the deterministic setting is achieved in expectation.

3.2.1. Constant, sufficiently small merit parameter.

For our purposes in this subsection, let us make the following assumption.

ASSUMPTION 3.10. *Event $E_{\tau,low}$ occurs in the sense that there exists an iteration number $\bar{k}_{\tau,\xi} \in \mathbb{N}$ and a merit parameter value $\bar{\tau}_{\min} \in \mathbb{R}_{>0}$ such that*

$$\bar{\tau}_k = \bar{\tau}_{\min} \leq \tau_k^{trial} \quad \textit{and} \quad \bar{\xi}_k = \bar{\xi}_{\min} \quad \textit{for all} \quad k \geq \bar{k}_{\tau,\xi}. \tag{3.10}$$

In addition, the stochastic gradient sequence $\{\bar{g}_k\}_{k\geq\bar{k}_{\tau,\xi}}$ satisfies

$$\mathbb{E}_{k,\tau,low}[\bar{g}_k] = g_k \quad \textit{and} \quad \mathbb{E}_{k,\tau,low}[\|\bar{g}_k - g_k\|_2^2] \leq M,$$

where $\mathbb{E}_{k,\tau,low}$ denotes expectation with respect to the distribution of ω conditioned on the event that $E_{\tau,low}$ occurs and the algorithm has reached x_k in iteration $k \in \mathbb{N}$.

The inequality $\bar{\tau}_k \leq \tau_k^{trial}$ in (3.10) is critical since it ensures that the model reduction value $\Delta q(x_k, \bar{\tau}_{\min}, g_k, H_k, d_k)$ satisfies the result of Lemma 2.12 for all $k \geq \bar{k}_{\tau,\xi}$ with $\bar{\tau}_{\min}$ in place of τ_k. In other words, it means that the merit parameter has become small enough such that, if one were to compute the *deterministic* search direction d_k using the *true* gradient g_k at x_k, then one would find that it is a direction of sufficient descent for the merit function $\phi(\cdot, \bar{\tau}_{\min})$ at x_k. The importance of this becomes clear in our final results at the end of this part of our analysis. The latter part of the assumption reaffirms the properties of the stochastic gradient estimates stated in Assumption 3.1, now conditioned on the occurrence of $E_{\tau,\text{low}}$. With this assumption, the results of Lemmas 3.8 and 3.9 continue to hold. For the sake of brevity, for the rest of this part of our analysis (section 3.2.1), let us redefine $\mathbb{E}_k[\,\cdot\,] \equiv \mathbb{E}_{k,\tau,low}[\,\cdot\,]$.

To derive our main result for this case, our goal is to prove upper bounds in expectation for the positive terms on the right-hand side of the conclusion of Lemma 3.7. Let us first consider the last term, which is addressed in our next lemma.

LEMMA 3.11. *Suppose that Assumption 3.10 holds. Let $\kappa_g \in \mathbb{R}_{>0}$ be an upper bound for $\{\|g_k\|_2\}$, the existence of which follows under Assumption 3.2. It follows, with $\kappa_d \in \mathbb{R}_{>0}$ from Lemma 3.8 and any $k \geq \bar{k}_{\tau,\xi}$, that $\mathbb{E}_k[\bar{\alpha}_k \bar{\tau}_k g_k^T(\bar{d}_k - d_k)] \leq \beta_k^2 \theta \bar{\tau}_{\min} \kappa_g \kappa_d \sqrt{M}$.*

Proof. For all $k \geq \bar{k}_{\tau,\xi}$, let E_k be the event that $g_k^T(\bar{d}_k - d_k) \geq 0$ and let E_k^c be the event that $g_k^T(\bar{d}_k - d_k) < 0$. Let $\mathbb{P}_k[\cdot]$ denote probability conditioned on the event that $E_{\tau,\text{low}}$ occurs and the algorithm has reached x_k in iteration k. By the law of total expectation, (3.10), and Lemma 3.6, it follows for $k \geq \bar{k}_{\tau,\xi}$ that

$$\begin{aligned}
&\mathbb{E}_k[\bar{\alpha}_k \bar{\tau}_k g_k^T(\bar{d}_k - d_k)] \\
&= \mathbb{E}_k[\bar{\alpha}_k \bar{\tau}_{\min} g_k^T(\bar{d}_k - d_k)|E_k]\mathbb{P}_k[E_k] + \mathbb{E}_k[\bar{\alpha}_k \bar{\tau}_{\min} g_k^T(\bar{d}_k - d_k)|E_k^c]\mathbb{P}_k[E_k^c] \\
&\leq \bar{\alpha}_{k,\max}\bar{\tau}_{\min}\mathbb{E}_k[g_k^T(\bar{d}_k - d_k)|E_k]\mathbb{P}_k[E_k] + \bar{\alpha}_{k,\min}\bar{\tau}_{\min}\mathbb{E}_k[g_k^T(\bar{d}_k - d_k)|E_k^c]\mathbb{P}_k[E_k^c].
\end{aligned}$$

Hence, since $\bar{d}_k$ is an unbiased estimator of d_k (by Lemma 3.8), it follows from the inequality above and the law of total expectation that

$$\begin{aligned}
&\mathbb{E}_k[\bar{\alpha}_k \bar{\tau}_k g_k^T(\bar{d}_k - d_k)] \\
&\leq \bar{\alpha}_{k,\min}\bar{\tau}_{\min}\mathbb{E}_k[g_k^T(\bar{d}_k - d_k)|E_k]\mathbb{P}_k[E_k] + \bar{\alpha}_{k,\min}\bar{\tau}_{\min}\mathbb{E}_k[g_k^T(\bar{d}_k - d_k)|E_k^c]\mathbb{P}_k[E_k^c] \\
&\quad + (\bar{\alpha}_{k,\max} - \bar{\alpha}_{k,\min})\bar{\tau}_{\min}\mathbb{E}_k[g_k^T(\bar{d}_k - d_k)|E_k]\mathbb{P}_k[E_k] \\
&= (\bar{\alpha}_{k,\max} - \bar{\alpha}_{k,\min})\bar{\tau}_{\min}\mathbb{E}_k[g_k^T(\bar{d}_k - d_k)|E_k]\mathbb{P}_k[E_k].
\end{aligned}$$

Now observe, by Cauchy–Schwarz and the law of total expectation, that

$$\begin{aligned}
\mathbb{E}_k[g_k^T(\bar{d}_k - d_k)|E_k]\mathbb{P}_k[E_k] &\leq \mathbb{E}_k[\|g_k\|_2\|\bar{d}_k - d_k\|_2|E_k]\mathbb{P}_k[E_k] \\
&= \mathbb{E}_k[\|g_k\|_2\|\bar{d}_k - d_k\|_2] - \mathbb{E}_k[\|g_k\|_2\|\bar{d}_k - d_k\|_2|E_k^c]\mathbb{P}_k[E_k^c] \\
&\leq \|g_k\|_2\mathbb{E}_k[\|\bar{d}_k - d_k\|_2].
\end{aligned}$$

Combining these results with Lemmas 3.6 and 3.8 yields the result. □

Our next result addresses the middle term on the right-hand side of Lemma 3.7.

LEMMA 3.12. *Suppose Assumption* 3.10 *holds. Then, for all $k \geq \bar{k}_{\tau,\xi}$, it follows that $\mathbb{E}_k[\Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k)] \leq \Delta q(x_k, \bar{\tau}_{\min}, g_k, H_k, d_k) + \bar{\tau}_{\min}\zeta^{-1}M$.*

Proof. Consider arbitrary $k \geq \bar{k}_{\tau,\xi}$. From (2.4), (3.10), Lemma 3.9, Jensen's inequality, and the convexity of $\max\{\cdot, 0\}$, it follows that

$$\begin{aligned}
\mathbb{E}_k[\Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k)] &= \mathbb{E}_k[-\bar{\tau}_{\min}(\bar{g}_k^T\bar{d}_k + \tfrac{1}{2}\max\{\bar{d}_k^T H_k \bar{d}_k, 0\}) + \|c_k\|_1] \\
&\leq -\bar{\tau}_{\min}(g_k^T d_k + \tfrac{1}{2}\max\{d_k^T H_k d_k, 0\}) + \bar{\tau}_{\min}\zeta^{-1}M + \|c_k\|_1 \\
&= \Delta q(x_k, \bar{\tau}_{\min}, g_k, H_k, d_k) + \bar{\tau}_{\min}\zeta^{-1}M,
\end{aligned}$$

as desired. □

We now prove our main theorem for this part of our analysis, where we define

$$\mathbb{E}_{\tau,\text{low}}[\ \cdot\ ] = \mathbb{E}[\ \cdot\ |\ \text{Assumption 3.10 holds}\ ].$$

THEOREM 3.13. *Suppose that Assumption* 3.10 *holds and the sequence $\{\beta_k\}$ is chosen such that $\beta_k\bar{\xi}_k\bar{\tau}_k/(\bar{\tau}_k L + \Gamma) \in (0, 1]$ for all $k \geq \bar{k}_{\tau,\xi}$. Define*

$$\overline{A} := \tfrac{\bar{\xi}_{\min}\bar{\tau}_{\min}}{\bar{\tau}_{\min}L+\Gamma} \quad \textit{and} \quad \overline{M} := \bar{\tau}_{\min}\big(\tfrac{1}{2}(\overline{A} + \theta)\zeta^{-1}M + \theta\kappa_g\kappa_d\sqrt{M}\big).$$

If $\beta_k = \beta \in (0, 2\overline{A}/(\overline{A} + \theta))$ for all $k \geq \bar{k}_{\tau,\xi}$, then

$$\begin{aligned}
&\mathbb{E}_{\tau,low}\left[\tfrac{1}{k+1}\sum_{j=\bar{k}_{\tau,\xi}}^{\bar{k}_{\tau,\xi}+k}\Delta q(x_j, \bar{\tau}_{\min}, g_j, H_j, d_j)\right] \\
&\leq \tfrac{\beta\overline{M}}{\overline{A}-\frac{1}{2}(\overline{A}+\theta)\beta} + \tfrac{\mathbb{E}_{\tau,low}[\phi(x_{\bar{k}_{\tau,\xi}},\bar{\tau}_{\min})]-\phi_{\min}}{(k+1)\beta(\overline{A}-\frac{1}{2}(\overline{A}+\theta)\beta)} \xrightarrow{k\to\infty} \tfrac{\beta\overline{M}}{\overline{A}-\frac{1}{2}(\overline{A}+\theta)\beta},
\end{aligned} \tag{3.11}$$

where $\phi_{\min} \in \mathbb{R}$ is a lower bound for $\phi(\cdot, \bar{\tau}_{\min})$ over $\mathcal{X}$, the existence of which follows by Assumption 3.2. *On the other hand, if $\sum_{k=\bar{k}_{\tau,\xi}}^{\infty} \beta_k = \infty$ and $\sum_{k=\bar{k}_{\tau,\xi}}^{\infty} \beta_k^2 < \infty$, then*

$$\lim_{k\to\infty} \mathbb{E}_{\tau,low}\left[\frac{1}{\left(\sum_{j=\bar{k}_{\tau,\xi}}^{\bar{k}_{\tau,\xi}+k} \beta_j\right)} \sum_{j=\bar{k}_{\tau,\xi}}^{\bar{k}_{\tau,\xi}+k} \beta_j \Delta q(x_j, \bar{\tau}_{\min}, g_j, H_j, d_j)\right] = 0. \tag{3.12}$$

Proof. Consider arbitrary $k \geq \bar{k}_{\tau,\xi}$. It follows from the definition of $\overline{A}$, Lemma 3.6, and the fact that $\beta_k \in (0,1]$ that $\overline{A}\beta_k \leq \bar{\alpha}_k \leq (\overline{A}+\theta)\beta_k$. Hence, it follows from Lemmas 3.4(d), 3.7, 3.11, and 3.12 that, under the conditions of the theorem,

$$\begin{aligned}
&\mathbb{E}_k[\phi(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_k)] - \mathbb{E}_k[\phi(x_k, \bar{\tau}_k)] \\
&\leq \mathbb{E}_k[-\bar{\alpha}_k \Delta q(x_k, \bar{\tau}_k, g_k, H_k, d_k) + \tfrac{1}{2}\bar{\alpha}_k \beta_k \Delta q(x_k, \bar{\tau}_k, \bar{g}_k, H_k, \bar{d}_k) + \bar{\alpha}_k \bar{\tau}_k g_k^T(\bar{d}_k - d_k)] \\
&\leq -\beta_k\big(\overline{A} - \tfrac{1}{2}(\overline{A}+\theta)\beta_k\big)\Delta q(x_k, \bar{\tau}_{\min}, g_k, H_k, d_k) + \beta_k^2 \overline{M}.
\end{aligned}$$

For the scenario of $\{\beta_k\}$ being a constant sequence for $k \geq \bar{k}_{\tau,\xi}$, one finds from above, taking total expectation conditioned on (3.10), that, for all $k \geq \bar{k}_{\tau,\xi}$,

$$\begin{aligned}
&\mathbb{E}_{\tau,\mathrm{low}}[\phi(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_{\min})] - \mathbb{E}_{\tau,\mathrm{low}}[\phi(x_k, \bar{\tau}_{\min})] \\
&\qquad\qquad \leq -\beta(\overline{A} - \tfrac{1}{2}(\overline{A}+\theta)\beta)\mathbb{E}_{\tau,\mathrm{low}}[\Delta q(x_k, \bar{\tau}_{\min}, g_k, H_k, d_k)] + \beta^2 \overline{M}.
\end{aligned}$$

Summing this inequality for $j \in \{\bar{k}_{\tau,\xi}, \dots, \bar{k}_{\tau,\xi}+k\}$, one finds by Assumption 3.2 that

$$\begin{aligned}
&\phi_{\min} - \mathbb{E}_{\tau,\mathrm{low}}[\phi(x_{\bar{k}_{\tau,\xi}}, \bar{\tau}_{\min})] \\
&\leq \mathbb{E}_{\tau,\mathrm{low}}[\phi(x_{\bar{k}_{\tau,\xi}+k+1}, \bar{\tau}_{\min})] - \mathbb{E}_{\tau,\mathrm{low}}[\phi(x_{\bar{k}_{\tau,\xi}}, \bar{\tau}_{\min})] \\
&\leq -\beta(\overline{A} - \tfrac{1}{2}(\overline{A}+\theta)\beta)\mathbb{E}_{\tau,\mathrm{low}}\left[\sum_{j=\bar{k}_{\tau,\xi}}^{\bar{k}_{\tau,\xi}+k} \Delta q(x_j, \bar{\tau}_{\min}, g_j, H_j, d_j)\right] + (k+1)\beta^2 \overline{M},
\end{aligned}$$

from which (3.11) follows. Now consider the scenario of $\{\beta_k\}$ diminishing as described. It follows that for sufficiently large $k \geq \bar{k}_{\tau,\xi}$ one finds $\beta_k \leq \overline{A}/(\overline{A}+\theta)$; hence, let us assume without loss of generality that, for all $k \geq \bar{k}_{\tau,\xi}$, one has $\beta_k \leq \overline{A}/(\overline{A}+\theta)$, which implies $\overline{A} - \frac{1}{2}(\overline{A}+\theta)\beta_k \geq \frac{1}{2}\overline{A}$. Similar to above, it follows for all $k \geq \bar{k}_{\tau,\xi}$ that

$$\begin{aligned}
&\mathbb{E}_{\tau,\mathrm{low}}[\phi(x_k + \bar{\alpha}_k \bar{d}_k, \bar{\tau}_{\min})] - \mathbb{E}_{\tau,\mathrm{low}}[\phi(x_k, \bar{\tau}_{\min})] \\
&\qquad\qquad \leq -\tfrac{1}{2}\overline{A}\beta_k \mathbb{E}_{\tau,\mathrm{low}}[\Delta q(x_k, \bar{\tau}_{\min}, g_k, H_k, d_k)] + \beta_k^2 \overline{M}.
\end{aligned}$$

Summing this inequality for $j \in \{\bar{k}_{\tau,\xi}, \dots, \bar{k}_{\tau,\xi}+k\}$, one finds by Assumption 3.2 that

$$\begin{aligned}
&\phi_{\min} - \mathbb{E}_{\tau,\mathrm{low}}[\phi(x_{\bar{k}_{\tau,\xi}}, \bar{\tau}_{\min})] \\
&\leq \mathbb{E}_{\tau,\mathrm{low}}[\phi(x_{\bar{k}_{\tau,\xi}+k+1}, \bar{\tau}_{\min})] - \mathbb{E}_{\tau,\mathrm{low}}[\phi(x_{\bar{k}_{\tau,\xi}}, \bar{\tau}_{\min})] \\
&\leq -\tfrac{1}{2}\overline{A}\mathbb{E}_{\tau,\mathrm{low}}\left[\sum_{j=\bar{k}_{\tau,\xi}}^{\bar{k}_{\tau,\xi}+k} \beta_j \Delta q(x_j, \bar{\tau}_{\min}, g_j, H_j, d_j)\right] + \overline{M}\sum_{j=\bar{k}_{\tau,\xi}}^{\bar{k}_{\tau,\xi}+k} \beta_j^2.
\end{aligned}$$

Rearranging this inequality yields

$$\mathbb{E}_{\tau,\mathrm{low}}\left[\sum_{j=\bar{k}_{\tau,\xi}}^{\bar{k}_{\tau,\xi}+k} \beta_j \Delta q(x_j, \bar{\tau}_{\min}, g_j, H_j, d_j)\right] \leq \frac{2(\mathbb{E}_{\tau,\mathrm{low}}[\phi(x_{\bar{k}_{\tau,\xi}}, \bar{\tau}_{\min})] - \phi_{\min})}{\overline{A}} + \frac{2\overline{M}}{\overline{A}}\sum_{j=\bar{k}_{\tau,\xi}}^{\bar{k}_{\tau,\xi}+k} \beta_j^2,$$

from which (3.12) follows. ∎

The following corollary carries the result of Theorem 3.13 to a statement about the expected properties of the sequences $\{\|g_k + J_k^T y_k\|_2\}$ and $\{\|c_k\|_2\}$, where the multiplier y_k is defined by (2.6) for all $k \in \mathbb{N}$. In particular, the corollary shows that the primal iterate sequence $\{x_k\}$ generated by the algorithm offers stationarity and feasibility in expectation. We close the corollary with the observation that at any iterate sufficiently close to a stationary point, the error of the Lagrange multiplier estimate is directly proportional to the error in the stochastic gradient estimate.

COROLLARY 3.14. *Under the conditions of Theorem* 3.13*, the following hold true.*
(a) *If $\beta_k = \beta \in (0, 2\overline{A}/(\overline{A} + \theta))$ for all $k \geq \bar{k}_{\tau,\xi}$, then*

$$\mathbb{E}_{\tau,low}\left[\tfrac{1}{k+1}\sum_{j=\bar{k}_{\tau,\xi}}^{\bar{k}_{\tau,\xi}+k}\left(\tfrac{\|g_j+J_j^T y_j\|_2^2}{\kappa_H^2} + \|c_j\|_2\right)\right] \xrightarrow{k\to\infty} \tfrac{(\kappa_\Psi+1)\beta\overline{M}}{\kappa_q\bar{\tau}_{\min}(\overline{A}-\frac{1}{2}(\overline{A}+\theta)\beta)}.$$

(b) *If $\sum_{k=\bar{k}_{\tau,\xi}}^{\infty}\beta_k = \infty$ and $\sum_{k=\bar{k}_{\tau,\xi}}^{\infty}\beta_k^2 < \infty$, then*

$$\lim_{k\to\infty}\mathbb{E}_{\tau,low}\left[\tfrac{1}{\left(\sum_{j=\bar{k}_{\tau,\xi}}^{\bar{k}_{\tau,\xi}+k}\beta_j\right)}\sum_{j=\bar{k}_{\tau,\xi}}^{\bar{k}_{\tau,\xi}+k}\beta_j\left(\tfrac{\|g_j+J_j^T y_j\|_2^2}{\kappa_H^2} + \|c_j\|_2\right)\right] = 0,$$

from which it follows that

$$\liminf_{k\to\infty}\ \mathbb{E}_{\tau,low}[\kappa_H^{-2}\|g_k + J_k^T y_k\|_2^2 + \|c_k\|_2] = 0.$$

In addition, in either case, there exists $\delta_x \in \mathbb{R}_{>0}$ such that if $\|x_k - x_*\|_2 \leq \delta_x$ for some stationary point $(x_*, y_*) \in \mathbb{R}^n \times \mathbb{R}^m$ for (3.1)*, then for any $\delta_g \in \mathbb{R}_{>0}$ one finds*

$$\left\|\begin{bmatrix}\bar{g}_k - \nabla f(x_*)\\ c_k\end{bmatrix}\right\|_2 \leq \delta_g \quad\Longrightarrow\quad \|\bar{y}_k - y_*\|_2 \leq 2\kappa_*\delta_g,$$

where, under Assumption 3.2*, $\kappa_* \in \mathbb{R}_{>0}$ is an upper bound for $\left\|\begin{bmatrix}H_k & J_*^T\\ J_* & 0\end{bmatrix}\right\|^{-1}$.*

Proof. Parts (a) and (b) follow by combining the results of Lemmas 2.11 and 2.12, the relation (2.15), and Theorem 3.13. The remainder follows with Lemma 2.6 since for x_k sufficiently close to x_*, one obtains with $g_* := \nabla f(x_*)$ and $c_* := c(x_*) = 0$ that

$$\begin{aligned}\|\bar{y}_k - y_*\|_2 &\leq \left\|\begin{bmatrix}H_k & J_k^T\\ J_k & 0\end{bmatrix}^{-1}\begin{bmatrix}\bar{g}_k\\ c_k\end{bmatrix} - \begin{bmatrix}H_k & J_*^T\\ J_* & 0\end{bmatrix}^{-1}\begin{bmatrix}g_*\\ c_*\end{bmatrix}\right\|_2\\ &= \left\|\begin{bmatrix}H_k & J_k^T\\ J_k & 0\end{bmatrix}^{-1}\begin{bmatrix}\bar{g}_k - g_*\\ c_k\end{bmatrix} + \left(\begin{bmatrix}H_k & J_k^T\\ J_k & 0\end{bmatrix}^{-1} - \begin{bmatrix}H_k & J_*^T\\ J_* & 0\end{bmatrix}^{-1}\right)\begin{bmatrix}g_*\\ 0\end{bmatrix}\right\|_2\\ &\leq \tfrac{3}{2}\kappa_*\left\|\begin{bmatrix}\bar{g}_k - g_*\\ c_k\end{bmatrix}\right\|_2 + \tfrac{1}{2}\kappa_*\delta_g,\end{aligned}$$

from which the desired conclusion follows. □

We close our analysis of this case with the following remark.

Remark 3.15. Consideration of the conclusion of Corollary 3.14(a) reveals the close relationship between our result and a conclusion that one reaches for a stochastic (sub)gradient method in an unconstrained setting. Notice that

$$\frac{2\kappa_\Psi\beta\overline{M}}{\kappa_q\bar{\tau}_{\min}(\overline{A}-\frac{1}{2}(\overline{A}+\theta)\beta)} = \frac{2\kappa_\Psi\beta\bar{\tau}_{\min}\left(\frac{1}{2}\left(\frac{\bar{\xi}_{\min}\bar{\tau}_{\min}}{\bar{\tau}_{\min}L+\Gamma}+\theta\right)\zeta^{-1}M+\theta\kappa_g\kappa_d\sqrt{M}\right)}{\kappa_q\bar{\tau}_{\min}\left(\frac{\bar{\xi}_{\min}\bar{\tau}_{\min}}{\bar{\tau}_{\min}L+\Gamma}-\frac{1}{2}\left(\frac{\bar{\xi}_{\min}\bar{\tau}_{\min}}{\bar{\tau}_{\min}L+\Gamma}+\theta\right)\beta\right)}.$$

Our first observation is a common one for the unconstrained setting: The value above is directly proportional to β. To reduce this value, one should choose smaller β, but the downside of choosing smaller β is that the algorithm takes shorter steps, meaning that it takes longer for this limiting value to be approached (recall (3.11)). On the other hand, while larger β means that the algorithm takes larger steps, this comes at the cost of a larger limiting value. A second observation, unique for our algorithm, is the influence of θ. The quantity above is directly proportional to θ, meaning that the optimal choice in terms of reducing this value is $\theta = 0$, in which case one obtains

$$\frac{2\kappa_\Psi\beta\overline{M}}{\kappa_q\bar{\tau}_{\min}(\overline{A}-\frac{1}{2}(\overline{A}+\theta)\beta)} \xrightarrow{\theta\to 0} \frac{\kappa_\Psi\beta\zeta^{-1}M}{\kappa_q(1-\frac{1}{2}\beta)}.$$

However, this results in a nonadaptive algorithm with $\bar{\alpha}_k = \beta_k\bar{\xi}_k\bar{\tau}_k/(\bar{\tau}_kL+\Gamma)$ for all $k\in\mathbb{N}$. This choice has some theoretical benefits (see also our discussion in section 5), but we have found this conservative choice to be detrimental in practice.

3.2.2. Poor merit parameter behavior. Theorem 3.13 and Corollary 3.14 show desirable convergence properties in expectation of Algorithm 3.1 in the event that the merit parameter sequence eventually remains constant at a value that is sufficiently small. This captures behavior similar to that of Algorithm 2.1 in the deterministic setting, in which the merit parameter is *guaranteed* to behave in this manner. However, for the stochastic Algorithm 3.1, one of two other events are possible, which we now define mathematically as follows:

- Event $E_{\tau\gg 0}$: there exists infinite $\overline{\mathcal{K}}_\tau\subseteq\mathbb{N}$ and $\bar{\tau}_{big}\in\mathbb{R}_{>0}$ such that $\bar{\tau}_k = \bar{\tau}_{big} > \tau_k^{trial}$ and $\bar{\xi}_k = \bar{\xi}_{\min}$ for all $k\in\overline{\mathcal{K}}_\tau$. Since $\bar{\tau}_k^{trial}\geq\bar{\tau}_k$ for all $k\in\mathbb{N}$, this means $\bar{\tau}_k^{trial} > \tau_k^{trial}$ for all $k\in\overline{\mathcal{K}}_\tau$.
- Event $E_{\tau\searrow 0}$: $\{\bar{\tau}_k\}\searrow 0$.

Our goal in this part of our analysis is to argue that these events, exhibiting what we refer to as poor behavior of the merit parameter sequence, are either impossible or only occur with probability zero. For these considerations, let us return to assuming that Assumptions 3.1 and 3.2 hold (and not Assumption 3.10).

Let us first consider event $E_{\tau\gg 0}$. As shown above in the definition of $E_{\tau\gg 0}$, the merit parameter remaining too large requires that the stochastic trial value $\bar{\tau}_k^{trial}$ *consistently* overestimates the deterministic trial value τ_k^{trial}. The following proposition shows that under a modest assumption about the behavior of the stochastic gradients and corresponding search directions, this behavior occurs with probability zero.

PROPOSITION 3.16. *If there exists $p\in(0,1]$ such that, for all $k\in\mathbb{N}$,*

$$\mathbb{P}_k[\bar{g}_k^T\bar{d}_k + \max\{\bar{d}_k^TH_k\bar{d}_k,0\}\geq g_k^Td_k+\max\{d_k^TH_kd_k,0\}]\geq p,$$

then $E_{\tau\gg 0}$ occurs with probability zero.

Proof. If, in any run of the algorithm, $g_k^Td_k+\max\{d_k^TH_kd_k,0\}\leq 0$ for all sufficiently large $k\in\mathbb{N}$, then $\tau_k^{trial}=\infty$ for all sufficiently large $k\in\mathbb{N}$ and event $E_{\tau\gg 0}$ does not occur. Hence, let us define $\mathcal{K}_{gd}\subseteq\mathbb{N}$ as the set of indices such that $k\in\mathcal{K}_{gd}$ if and only if $g_k^Td_k+\max\{d_k^TH_kd_k,0\}>0$, and let us restrict attention to runs in which $\mathcal{K}_{gd}$ is infinite. For any $k \in \mathcal{K}_{gd}$, it follows that the inequality $\bar{g}_k^T\bar{d}_k + \max\{\bar{d}_k^T H_k \bar{d}_k, 0\} \geq g_k^T d_k + \max\{d_k^T H_k d_k, 0\}$ holds if and only if

$$\bar{\tau}_k^{trial} = \frac{(1-\sigma)\|c_k\|_1}{\bar{g}_k^T\bar{d}_k + \max\{\bar{d}_k^T H_k \bar{d}_k, 0\}} \leq \frac{(1-\sigma)\|c_k\|_1}{g_k^T d_k + \max\{d_k^T H_k d_k, 0\}} = \tau_k^{trial}.$$

Hence, it follows from the conditions of the proposition, the fact that $\bar{\tau}_k \leq \bar{\tau}_k^{trial}$ for all $k \in \mathbb{N}$, and the fact that $\mathcal{K}_{gd}$ is infinite, that for any $k \in \mathbb{N}$ the probability is one that for a subsequent iteration number $\hat{k} \geq k$ one finds $\bar{\tau}_{\hat{k}} \leq \bar{\tau}_{\hat{k}}^{trial} \leq \tau_{\hat{k}}^{trial}$. This, the fact that Lemma 2.16 implies that $\{\tau_k^{trial}\}$ is bounded away from zero, and the fact that if the merit parameter is ever decreased, then it is done so by a constant factor, shows that one has $\bar{\tau}_k \leq \tau_k^{trial}$ for all sufficiently large $k \in \mathbb{N}$ with probability one. □

As a concrete example of a setting that offers the minimum probability required in Proposition 3.16, we offer the following. This is clearly only one of many example situations that one could consider to mimic real-world scenarios.

Example 3.17. If, for all $k \in \mathbb{N}$, one has $H_k \succ 0$ and $\bar{g}_k \sim \mathcal{N}(g_k, \Sigma_k)$ for some $\Sigma_k \in \mathbb{S}^n$ with $\Sigma_k \succ 0$, then the condition in Proposition 3.16 holds with $p = \frac{1}{2}$.

Proof. Let $k \in \mathbb{N}$ be arbitrary. The tangential component of the search direction is $\bar{u}_k = Z_k\bar{w}_k$, where, under Assumption 3.2 and the stated conditions, $\bar{w}_k = -(Z_k^T H_k Z_k)^{-1} Z_k^T(\bar{g}_k + H_k v_k)$. Plugging in this solution and simplifying yields

$$\bar{g}_k^T\bar{d}_k + \bar{d}_k^T H_k \bar{d}_k = v_k^T H_k^{1/2}(I - H_k^{1/2} Z_k (Z_k^T H_k Z_k)^{-1} Z_k^T H_k^{1/2})(H_k^{-1/2}\bar{g}_k + H_k^{1/2} v_k).$$

Since $\bar{g}_k$ is normally distributed with mean g_k, it follows that this value is normally distributed with a mean of the same form, but with g_k in place of $\bar{g}_k$ (see, e.g., [32]). Since a normally distributed random variable takes values greater than or equal to its expected value with probability $\frac{1}{2}$, the conclusion follows. □

Let us now consider the event $E_{\tau\searrow 0}$. One can learn from Lemmas 2.15 and 2.16 from the deterministic setting that the following holds true.

PROPOSITION 3.18. *Consider an arbitrary constant $g_{\max} \in \mathbb{R}_{>0}$. If, for a run of Algorithm* 3.1*, the stochastic gradient estimates satisfy $\|\bar{g}_k - g_k\|_2 \leq g_{\max}$ for all $k \in \mathbb{N}$, then the sequence of tangential step components $\{\bar{u}_k\}$ is bounded, and there exist $\bar{k}_\tau \in \mathbb{N}$ and $\bar{\tau}_{\min} \in \mathbb{R}_{>0}$ such that $\bar{\tau}_k = \bar{\tau}_{\min}$ for all $k \geq \bar{k}_\tau$.*

Proof. Boundedness in norm of the tangential step components follows in the same manner as in Lemma 2.15 with $(\bar{g}_k, \bar{u}_k)$ in place of (g_k, u_k). Further, the claimed behavior of the merit parameter sequence follows in the same manner as in the proof of Lemma 2.16 using $(\bar{g}_k, \bar{d}_k, \bar{u}_k)$ in place of (g_k, d_k, u_k), where in place of the constants $(\kappa_{\tau,1}, \kappa_{\tau,2})$ one derives constants $(\bar{\kappa}_{\tau,1}, \bar{\kappa}_{\tau,2})$ whose value depends on $g_{\max}$ as well as the upper bound on the sequence $\{\|g_k\|_2\}$ (under Assumption 3.2). □

By Proposition 3.18, if the differences $\{\bar{g}_k - g_k\}$ are bounded, then the merit parameter sequence will not vanish, i.e., event $E_{\tau\searrow 0}$ will not occur. This is guaranteed if the distributions defining the stochastic gradients $\{\bar{g}_k\}$ ensure uniform boundedness or, e.g., if $f(x) = \frac{1}{N}\sum_{i=1}^N f_i(x)$ and $\bar{g}_k := \nabla f_{i_k}(x_k)$ for all $k \in \mathbb{N}$, where the component functions $\{f_i\}$ have bounded derivatives over a set containing the iterates and in each iteration i_k is randomly sampled uniformly from $\{1, \ldots, N\}$.

4. Numerical results. In this section, we demonstrate the empirical performance of our proposed Algorithm 2.1 (for the deterministic setting) and Algorithm 3.1

(for the stochastic setting) using MATLAB implementations. We consider their performance on a subset of the equality constrained problems from the CUTE collection [3]. Specifically, of the 123 such problems in the set, we selected those for which (i) f is *not* a constant function, (ii) $n+m \leq 1000$, and (iii) the LICQ held at all iterates in all runs of all algorithms that we ran. This selection resulted in a total of 49 problems. Each problem comes with an initial point, which we used in our experiments.

4.1. Deterministic setting. Our goal in this setting is to demonstrate that, in practice, our proposed Algorithm 2.1 ("SQP Adaptive") is as reliable a method as the state-of-the-art Algorithm 2.2 ("SQP Backtracking"). We do not claim that SQP Adaptive is always as efficient as SQP Backtracking since, as has been verified by others in the literature, the line search scheme is typically very effective across a broad range of problems. That said, since our algorithm for the stochastic setting is based on SQP Adaptive, it is at least of interest to demonstrate that this approach is as reliable as SQP Backtracking. For these experiments, we chose each H_k to be the Hessian of the Lagrangian at (x_k, y_{k-1}). For both algorithms, for any k such that the inertia of the matrix in (2.6) is not correct with this choice, a multiple of the identity is added in an iterative manner until the correct inertia is attained. This is a common strategy in state-of-the-art constrained optimization software; see, e.g., [34].

For these experiments, the parameters were set as $\tau_{-1} = 1$, $\epsilon = 10^{-6}$, $\sigma = 1/2$, $\eta = 10^{-4}$, $\rho = 3$, $L_{-1} = 1$, $\gamma_{-1,i} = 1$, $\nu = 1/2$, and $\alpha = 1$. In line 5 of Algorithm 2.1, all Lipschitz constant estimates were set as 1/2 times the estimates from the previous iteration. A run terminated with a message of success if iteration $k \leq 10^4$ yielded

$$\|g_k + J_k^T y_k\|_\infty \leq 10^{-6} \max\{1, \|g_0 + J_0^T y_0\|_\infty\} \quad \text{and} \quad \|c_k\|_\infty \leq 10^{-6} \max\{1, \|c_0\|_\infty\};$$

otherwise, the run was considered a failure. Figure 4.1 provides Dolan–Moré performance profiles [13] for iterations and function evaluations required by the two methods. (The profiles are capped at $t = 20$.) As expected, the performance of SQP Backtracking was typically better than that of SQP Adaptive. That said, SQP Adaptive was as reliable as this state-of-the-art approach. Over all iterations of all runs of SQP Adaptive, the stepsize α_k was chosen less than one 40.9% of the time, equal to one 41.8% of the time, and greater than one 17.3% percent of the time.

We speculate that there may be situations in which SQP Adaptive can outperform SQP Backtracking in the deterministic regime. For example, we also ran the same set of experiments as above, but with $H_k = I$ for all $k \in \mathbb{N}$ for both algorithms, and

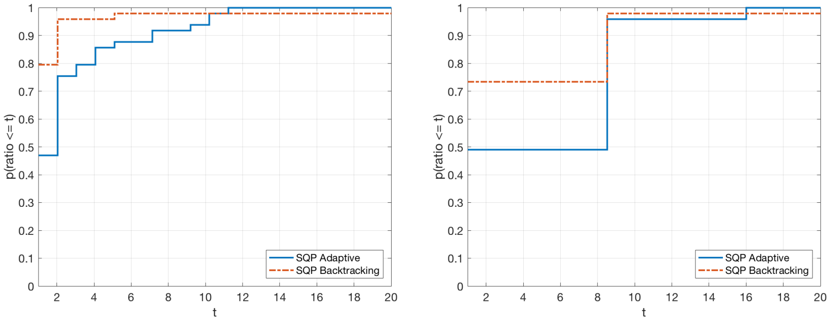

FIG. 4.1. *Performance profiles for SQP Adaptive and SQP Backtracking for problems from the CUTE test set in terms of iterations (left) and function evaluations (right).*

found that SQP Adaptive often outperforms SQP Backtracking, especially in terms of requiring fewer function evaluations. This appears to be due to SQP Backtracking requiring more function evaluations per iteration—during the line searches—when second-order derivatives are not employed.

4.2. Stochastic setting. Our goal in this setting is to compare the performance of our proposed Algorithm 3.1 ("Stochastic SQP") against that of a stochastic subgradient method ("Stochastic Subgradient") applied to minimize the exact penalty function (2.3) (which represents the current state-of-the-art for constrained stochastic optimization). For these experiments, we used our test set of 49 CUTE problems but considered multiple runs for different levels of noise. In particular, for a given run of an algorithm, we fixed $\epsilon_N \in \{10^{-8}, 10^{-4}, 10^{-2}, 10^{-1}\}$ and then for each iteration set the stochastic gradient estimate as $\bar{g}_k = \mathcal{N}(g_k, \epsilon_N I)$. For each problem and noise level, we ran 10 instances. This led to a total of 490 problem instances for each algorithm and noise level. Each run of Stochastic SQP was given a budget of 1000 iterations while each run of Stochastic Subgradient was given a budget of 10000 iterations. We tuned the value of τ individually for each problem instance for Stochastic Subgradient. In particular, for each problem instance, we ran the algorithm for the 11 values $\tau \in \{10^{-10}, 10^{-9}, \ldots, 10^{-1}, 10^{0}\}$ and selected the value for that instance that led to the best results in terms of feasibility and optimality errors (see below). Overall, this means that for each problem, Stochastic Subgradient was given 110 times the number of iterations that were allowed for Stochastic SQP. (This broad range of τ was needed by Stochastic Subgradient to obtain its best results. The selected τ values were roughly evenly distributed over the set from 10^{-10} to 10^{0}.)

For both methods, the Lipschitz constants L and $\Gamma = \sum_{i=1}^{m} \gamma_i$ were estimated using differences of gradients near the initial point and kept fixed for all subsequent iterations. (This process was done so that L and Γ were the same for both methods for each problem.) For Stochastic SQP, we set $H_k = I$ for all k for fairness of comparison with the (first-order) subgradient method. The other inputs for Stochastic SQP were set as $\bar{\tau}_{-1} = 1$, $\epsilon = 10^{-6}$, $\sigma = 1/2$, $\bar{\xi}_{-1} = 1$, $\theta = 10$, and $\beta_k = 1$ for all k. Stochastic Subgradient was run with a constant stepsize $\frac{\tau}{\tau L + \Gamma}$ for all k.

For each algorithm and each problem instance, we computed a resulting feasibility error and optimality error as follows. If a run produced an iterate that was sufficiently feasible in the sense that $\|c_k\|_\infty \leq 10^{-6} \max\{1, \|c_0\|_\infty\}$ for some k, then, with the largest k corresponding to such a feasible iterate, the feasibility error was reported as $\|c_k\|_\infty$ and the optimality error was reported as $\|g_k + J_k^T y_k\|_\infty$, where y_k was computed as a least-squares multiplier using the true gradient g_k and J_k. (In this manner, the optimality error is *not* based on a stochastic gradient; rather, it is a true measure of optimality corresponding to the iterate x_k.) On the other hand, if a run produced no sufficiently feasible iterate, then the feasibility error and optimality error were computed in this manner at the *least infeasible* iterate during the run. The results are reported in the form of box plots in Figure 4.2.

Finally, let us comment on the occurrence of the event (3.10). In all runs of Stochastic SQP, we found that $\bar{\tau}_k \leq \tau_k^{trial}$ held 100% of the time in the last 100 iterations. In fact, for the noise levels 10^{-8}, 10^{-4}, 10^{-2}, and 10^{-1}, this inequality held in 99.92%, 99.10%, 99.22%, and 99.65%, respectively, of *all* iterations. This provides evidence that the theory offered under the event $E_{\tau,\text{low}}$ is relevant in practice.

5. Conclusion. We have proposed SQP algorithms for solving smooth nonlinear optimization problems with equality constraints. Our first algorithm is based on a state-of-the-art line-search SQP method but employs a stepsize scheme based on

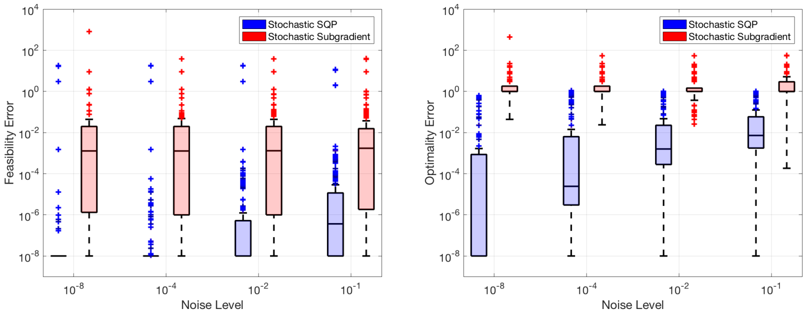

FIG. 4.2. *Box plots for feasibility errors (left) and optimality errors (right).*

(adaptively estimated) Lipschitz constants in place of the line search. We have shown that this method has convergence guarantees that match those of the state-of-the-art line-search SQP method, and our numerical experiments show that the algorithm is as reliable. Based on our first algorithm, our second algorithm is designed to solve problems involving deterministic constraint functions, but a stochastic objective. We have proved that under good behavior of the merit function parameter, the algorithm possesses convergence guarantees that match those of our deterministic algorithm in expectation. We have also argued that a certain type of poor behavior of this parameter only occurs with probability zero, and another type cannot occur under pragmatic assumptions. Our numerical experiments show that our algorithm for the stochastic setting consistently and significantly outperforms a (sub)gradient method employed to minimize a penalty function, which is an algorithm that represents the current state-of-the-art in the context of *stochastic* constrained optimization.

One assumption we have used is that the iterates remain in an open convex set over which the problem functions and their derivatives are bounded. One could loosen this assumption if one were to use our algorithm with $\theta = 0$. Indeed, notice that in our analysis in section 3.2.1, boundedness of $\{\|g_k\|_2\}$ is required in Lemma 3.11, but with $\theta = 0$ one has, for $k \geq \bar{k}_{\tau,\xi}$, $\mathbb{E}_k[\bar{\alpha}_k \bar{\tau}_k g_k^T(\bar{d}_k - d_k)] = \left(\frac{\beta_k \bar{\xi}_{\min} \bar{\tau}_{\min}}{\bar{\tau}_{\min} L_k + \Gamma_k}\right) \bar{\tau}_{\min} \mathbb{E}_k[g_k^T(\bar{d}_k - d_k)] = 0$. Hence, our assumption about the boundedness of $\{\|g_k\|_2\}$ is only needed when $\theta > 0$. We have proposed our algorithm for this setting since it is the context of $\theta > 0$ that allows the stepsize scheme in our algorithm to be adaptive, which has a significant benefit in terms of practical performance of the method.

Acknowledgments. The authors are grateful to the Associate Editor and two anonymous referees for their valuable comments and suggestions.

REFERENCES

[1] D. P. Bertsekas, *Network Optimization: Continuous and Discrete Models*, Athena Scientific Belmont, MA, 1998.

[2] J. T. Betts, *Practical Methods for Optimal Control and Estimation Using Nonlinear Programming*, SIAM, Philadelphia, 2010.

[3] I. Bongartz, A. R. Conn, N. Gould, and P. L. Toint, *Cute: Constrained and unconstrained testing environment*, ACM Trans. Math. Software, 21 (1995), pp. 123–160.

[4] L. Bottou, F. E. Curtis, and J. Nocedal, *Optimization methods for large-scale machine learning*, SIAM Rev., 60 (2018), pp. 223–311.

[5] R. H. Byrd, F. E. Curtis, and J. Nocedal, *An inexact SQP method for equality constrained optimization*, SIAM J. Optim., 19 (2008), pp. 351–369.

[6] R. H. Byrd, F. E. Curtis, and J. Nocedal, *An inexact Newton method for nonconvex equality constrained optimization*, Math. Program., 122 (2010), pp. 273–299.

[7] R. H. Byrd, J. C. Gilbert, and J. Nocedal, *A trust region method based on interior point techniques for nonlinear programming*, Math. Program., 89 (2000), pp. 149–185.

[8] R. H. Byrd, M. E. Hribar, and J. Nocedal, *An interior point algorithm for large-scale nonlinear programming*, SIAM J. Optim., 9 (1999), pp. 877–900.

[9] C. Chen, F. Tung, N. Vedula, and G. Mori, *Constraint-aware deep neural network compression*, in Proceedings of the ECCV, 2018, pp. 400–415.

[10] A. R. Conn, N. I. M. Gould, and P. L. Toint, *LANCELOT: A Fortran Package for Large-Scale Nonlinear Optimization*, Springer, New York, 1992.

[11] R. Courant, *Variational methods for the solution of problems of equilibrium and vibrations*, Bull. Amer. Math. Soc., 49 (1943), pp. 1–23.

[12] F. E. Curtis and D. P. Robinson, *Exploiting negative curvature in deterministic and stochastic optimization*, Math. Program. Ser. B, 176 (2019), pp. 69–94.

[13] E. D. Dolan and J. J. Moré, *Benchmarking optimization software with performance profiles*, Math. Program., 91 (2002), pp. 201–213.

[14] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, Chichester, UK, 1987.

[15] S. P. Han, *A globally convergent method for nonlinear programming*, J. Optim. Theory Appl., 22 (1977), pp. 297–309.

[16] S. P. Han and O. L. Mangasarian, *Exact penalty functions in nonlinear programming*, Math. Program, 17 (1979), pp. 251–269.

[17] E. Hazan and H. Luo, *Variance-reduced and projection-free stochastic optimization*, in Proceedings of the International Conference on Machine Learning, 2016, pp. 1263–1271.

[18] M. R. Hestenes, *Multiplier and Gradient Methods*, J. Optim. Theory Appl., 4 (1969), pp. 303–320.

[19] S. Kumar Roy, Z. Mhammedi, and M. Harandi, *Geometry aware constrained optimization techniques for deep learning*, in Proceedings of CVPR, 2018, pp. 4460–4469.

[20] F. Kupfer and E. W. Sachs, *Numerical solution of a nonlinear parabolic control problem by a reduced SQP method*, Comput. Optim. Appl., 1 (1992), pp. 113–135.

[21] F. Locatello, A. Yurtsever, O. Fercoq, and V. Cevher, *Stochastic Frank-Wolfe for composite convex minimization*, in Proceedings of NeurIPS, 2019, pp. 14269–14279.

[22] H. Lu and R. M. Freund, *Generalized stochastic Frank-Wolfe algorithm with stochastic "substitute" gradient for structured convex optimization*, Math. Program., 187 (2021), pp. 317–349.

[23] Y. Nandwani, A. Pathak, and P. Singla, *A primal-dual formulation for deep learning with constraints*, in Proceedings of NeurIPS, 2019, pp. 12157–12168.

[24] Y. Nesterov, *Introductory Lectures on Convex Optimization*, Appl. Optim., Springer, New York, 2004.

[25] J. Nocedal and S. Wright, *Numerical Optimization*, Springer Ser. Oper. Res. Financ. Eng., Springer, New York, 2006.

[26] M. J. D. Powell, *A Method for Nonlinear Constraints in Minimization Problems*, in Optimization, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298.

[27] M. J. D. Powell, *A fast algorithm for nonlinearly constrained optimization calculations*, in Numerical Analysis, Lecture Notes in Math., Springer, New York, 1978, pp. 144–157.

[28] S. N. Ravi, T. Dinh, V. S. Lokhande, and V. Singh, *Explicitly imposing constraints in deep networks via conditional gradients gives improved generalization and faster convergence*, in Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 2019, pp. 4772–4779.

[29] S. J. Reddi, S. Sra, B. Póczos, and A. Smola, *Stochastic Frank-Wolfe methods for nonconvex optimization*, in Proceedings of the 54th Annual Allerton Conference, IEEE, 2016, pp. 1244–1251.

[30] T. Rees, H. S. Dollar, and A. J. Wathen, *Optimal solvers for pde-constrained optimization*, SIAM J. Sci. Comput., 32 (2010), pp. 271–298.

[31] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, Philadelphia, 2009.

[32] Y. L. Tong, *The Multivariate Normal Distribution*, Springer, New York, 2012.

[33] A. Wächter and L. T. Biegler, *Line search filter methods for nonlinear programming: Motivation and global convergence*, SIAM J. Optim., 16 (2005), pp. 1–31.

[34] A. Waechter and L. T. Biegler, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, Math. Program., 106 (2006), pp. 25–57.

[35] R. B. Wilson, *A Simplicial Algorithm for Concave Programming*, Ph.D. thesis, Graduate School of Business Administration, Harvard University, Cambridge, MA, 1963.

[36] M. Zhang, Z. Shen, A. Mokhtari, H. Hassani, and A. Karbasi, *One sample stochastic Frank-Wolfe*, in Proceedings of AISTATS, 2020, pp. 4012–4023.