ISE

Active-Set Identification in Noisy and Stochastic Optimization

FRANK E. CURTIS, DANIEL P. ROBINSON, AND LARA ZEBIANE

Department of Industrial and Systems Engineering, Lehigh University

LEHIGH
UNIVERSITY®

COR@L
COMPUTATIONAL OPTIMIZATION
RESEARCH AT LEHIGH

Active-Set Identification in Noisy and Stochastic Optimization

Frank E. Curtis[*1], Daniel P. Robinson[†1], and Lara Zebiane[‡1]

[1]Department of Industrial and Systems Engineering, Lehigh University

August 30, 2025

Abstract

Identifying active constraints from a point near an optimal solution is important both theoretically and practically in constrained continuous optimization, as it can help identify optimal Lagrange multipliers and essentially reduces an inequality-constrained problem to an equality-constrained one. Traditional active-set identification guarantees have been proved under assumptions of smoothness and constraint qualifications, and assume exact function and derivative values. This work extends these results to settings when both objective and constraint function and derivative values have deterministic or stochastic noise. Two strategies are proposed that, under mild conditions, are proved to identify the active set of a local minimizer correctly when a point is close enough to the local minimizer and the noise is sufficiently small. Guarantees are also stated for the use of active-set identification strategies within a stochastic algorithm. We demonstrate our findings with two simple illustrative examples and a more realistic constrained neural-network training task.

1 Introduction

A critical feature of an algorithm for solving constrained continuous optimization problems is its ability to identify the inequality constraints that are active at a local minimizer. Suppose that, with respect to a function $c : \mathbb{R}^n \to \mathbb{R}^q$, a problem defined with respect to a decision variable $x \in \mathbb{R}^n$ involves inequality constraints of the form $c(x) \leq 0$. The aim of active-set identification is for an algorithm to be able to determine, with respect to a local minimizer $x_* \in \mathbb{R}^n$, the set $\{i \in \{1, \ldots, q\} : c_i(x_*) = 0\}$ of inequality constraints that hold at equality—i.e., are *active*—at x_*, even if the algorithm only has access to a solution estimate that is within a local neighborhood of x_*. Active-set identification is critical in multiple ways. One way is in the context of algorithms that, in each iteration, compute a search direction by (a) estimating the active set, and then (b) computing the search direction with respect to the estimated active set (while ignoring all of the constraints that are estimated to be *inactive*) [10, 9]. Another manner in which active-set identification is critical is in the computation of Lagrange multipliers, which in turn can be useful algorithmically and/or for sensitivity analysis of the local minimizer.

Various strong active-set identification guarantees have been proved for settings in which a problem's objective and constraint functions are continuously differentiable and an algorithm has access to exact function and derivative values; see, e.g., [7, 8, 18, 22, 24, 29, 34]. Typically, these guarantees require that a constraint qualification—such as the linear independence constraint qualification (LICQ) or Mangasarian-Fromovitz constraint qualification (MFCQ) [25]—and/or a second-order condition hold at the local minimizer x_*. Each such guarantee states that, if a point x is sufficiently close to x_*, then the active set at x_* can be identified correctly by solving a subproblem involving problem function and derivative values at x.

---

[*]E-mail: frank.e.curtis@lehigh.edu

[†]E-mail: daniel.p.robinson@lehigh.edu

[‡]E-mail: lara.zebiane@lehigh.edu

2

The contributions of this paper relate to showing that such guarantees can be extended to settings in which the problem function and derivative values contain errors. The design and analysis of optimization algorithms when function and derivative values are subject to errors are active areas of research, which motivates extensions of these techniques. Errors can arise in multiple settings. One setting of modern interest is that of *deterministic noise*, which we use to refer to situations in which the evaluation of a problem function or derivative always results in the same approximate value whenever the evaluation occurs multiple times at the same point. This may represent situations in which the problem function is defined by a numerical simulation that is subject to computational noise. We refer the reader to the recent articles [1, 11, 17, 31, 33, 36] on the design and analysis of algorithms for solving unconstrained or constrained optimization problems subject to deterministic noise. Another setting of modern interest is that of *stochastic noise*. This refers to situations in which a problem function value is defined for any point in its domain by, e.g., a large finite sum and/or an expectation over a random variable argument. The noise arises from the fact that, in practice, it is intractable to evaluate the value or derivative of such a function exactly, and instead one obtains an approximate value defined by a random sample of inputs. There are a large number of articles on the design and analysis of algorithms for solving unconstrained optimization problems when evaluations are subject to stochastic noise; see, e.g., the references in [5, 16]. For recent articles on solving constrained problems of this type, see [2, 3, 12, 13, 14, 15, 19, 26, 27].

Despite these efforts on algorithm design and analysis, little has been done to extend active-set identification guarantees to settings with errors. (An exception is [27] wherein an active-set algorithm is devised for solving problems when objective function and derivative values are subject to stochastic noise while constraint function and derivative values are exact. However, in that setting, the active-set estimation is tied to a specific algorithm and the results do not allow noise in the constraint function or derivative values.) In this paper, we consider strategies for active-set identification that may be employed irrespective of any particular algorithm, and when objective and/or constraint function and derivative values may be subject to noise. Following primarily [29], we show that when a point x is sufficiently close to a local minimizer x_* and the noise is sufficiently small, the active set at x_* can be identified correctly through solving two types of subproblems at x. We also state how our active-set identification guarantees can be employed for accurate active-set identification in the context of a stochastic algorithm for constrained optimization. Finally, we demonstrate our theoretical guarantees through a pair of illustrative numerical experiments, one involving a couple of small-dimensional examples where the results can be seen visually and another involving a constrained neural-network training problem.

1.1 Notation

We use $\mathbb{R}$ to denote the set of real numbers and, for any $r \in \mathbb{R}$, use $\mathbb{R}_{\geq r}$ to denote the set of real numbers greater than or equal to r. Real-valued n-vectors and m-by-n matrices are denoted by $\mathbb{R}^n$ and $\mathbb{R}^{m \times n}$, respectively. Given an integer $q \geq 1$, we define the corresponding set of integers $[q] := \{1, \ldots, q\}$. Given $\mathcal{A} \subseteq [q]$, we use $\mathcal{A}^c \subseteq [q]$ to denote the complement of $\mathcal{A}$ with respect to the set $[q]$, namely, $\mathcal{A}^c := [q] \setminus \mathcal{A}$. We use $\mathbb{1}$ to denote the vector with all components equal to 1; in each instance, the length of the vector is determined by the context in which it appears. Similarly, we use I to denote the identity matrix whose size is determined by its context. Given any vector or vector function v, its ith component is written as v_i. In addition, given such v, we define $[v]_+ := \max\{v, 0\}$ and $[v]_- := \max\{-v, 0\}$, where in each case the maximum is taken component-wise. Given a pair of vectors $(a, b) \in \mathbb{R}^n \times \mathbb{R}^n$, we use $a \perp b$ to indicate that $a_i b_i = 0$ for all $i \in [n]$. Given sets $\mathcal{A} \subseteq \mathbb{R}^n$ and $\mathcal{B} \subseteq \mathbb{R}^n$, we denote the ℓ_2-norm-induced distance between them as

$$\text{dist}(\mathcal{A}, \mathcal{B}) := \inf_{(a,b) \in \mathcal{A} \times \mathcal{B}} \|a - b\|_2. \tag{1}$$

1.2 Organization

In §2, we state the general form of the optimization problems that we consider throughout the paper, and provide background on optimality conditions and related concepts that are referenced throughout. In §3, we

discuss active-set identification guarantees in the context of having exact problem function and derivative values. This section includes one new contribution in the form of a modification of a result from [29] from the setting of employing a linear optimization problem (LP) with a trust-region constraint to one employing a quadratic optimization problem (QP). However, our main contributions are contained in §4, wherein we provide theoretical guarantees about active-set identification when function and derivative values can be subject to errors. We provide the results of illustrative numerical experiments in §5 and provide concluding remarks in §6.

2 Preliminaries

Given an objective function $f : \mathbb{R}^n \to \mathbb{R}$, equality-constraint function $e : \mathbb{R}^n \to \mathbb{R}^p$, and inequality-constraint function $c : \mathbb{R}^n \to \mathbb{R}^q$, where all of the problem functions (i.e., f, e, and c) are continuously differentiable, consider the constrained continuous optimization problem

$$\min_{x \in \mathbb{R}^n} \ f(x) \ \text{s.t.} \ e(x) = 0 \text{ and } c(x) \le 0. \tag{2}$$

Let the Lagrangian of (2) be denoted by $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ and defined by

$$\mathcal{L}(x, y, z) = f(x) + e(x)^T y + c(x)^T z,$$

where $y \in \mathbb{R}^p$ and $z \in \mathbb{R}^q$ are Lagrange multiplier variables. Under a constraint qualification, such as the LICQ or MFCQ, first-order necessary conditions for (2) are that at a local minimizer $x \in \mathbb{R}^n$ there exist $y \in \mathbb{R}^p$ and $z \in \mathbb{R}^q$ such that

$$\nabla f(x) + \nabla e(x) y + \nabla c(x) z = 0 \tag{3a}$$
$$e(x) = 0 \tag{3b}$$
$$0 \le z \perp c(x) \le 0, \tag{3c}$$

where (3a) and (3b) correspond to $\nabla_x \mathcal{L}(x, y, z) = 0$ and $\nabla_y \mathcal{L}(x, y, z) = 0$, respectively. We refer to any tuple $(x, y, z) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q$ satisfying (3) as a KKT (i.e., Karush-Kuhn-Tucker) point. Similarly, we refer to $x \in \mathbb{R}^n$ as a primal KKT point if and only if there exists $(y, z) \in \mathbb{R}^p \times \mathbb{R}^q$ such that (x, y, z) is a KKT point.

At a local minimizer $x \in \mathbb{R}^n$ of (2) at which a constraint qualification is presumed to hold, a Lagrange multiplier pair (y, z) such that (x, y, z) satisfies (3) is not necessarily unique. Therefore, for any $x \in \mathbb{R}^n$, let us define

$$\mathcal{D}(x) := \{(y, z) \in \mathbb{R}^p \times \mathbb{R}^q : (x, y, z) \text{ satisfies (3)}\}.$$

That is, for any $x \in \mathbb{R}^n$, the set $\mathcal{D}(x)$ (which may be empty) contains all Lagrange multiplier pairs that yield a KKT point along with x. It is well known that if the MFCQ holds at a local minimizer x of (2), then $\mathcal{D}(x)$ is nonempty and bounded, and if the LICQ holds at a local minimizer x of (2), then $\mathcal{D}(x)$ is a singleton. Strict complementarity is said to hold for a KKT point (x, y, z) for (2) if and only if $z_i > 0$ for all $i \in [q]$ such that $c_i(x) = 0$; thus, if the LICQ holds at a local minimizer x of (2), then strict complementarity holds if and only if the unique Lagrange multiplier pair (y, z) has $z_i > 0$ for all $i \in [q]$ such that $c_i(x) = 0$.

Let us indicate the set of active constraints at any primal point $x \in \mathbb{R}^n$ as the set of indices of the inequality constraints that are satisfied at equality at x, namely,

$$\mathcal{A}(x) := \{i \in [q] : c_i(x) = 0\}.$$

Let (x, y, z) be a KKT point. Of the constraints whose indices are contained in $\mathcal{A}(x)$, the set of *weakly* active inequality constraints at x are those such that the corresponding Lagrange multiplier is equal to zero for all elements of $\mathcal{D}(x)$, i.e.,

$$\mathcal{A}_0(x) := \{i \in \mathcal{A}(x) : z_i = 0 \text{ for all } (y, z) \in \mathcal{D}(x)\}.$$

Later in our analysis, we reference a particular second-order optimality condition. In particular, let us define the critical cone at a local minimizer $x \in \mathbb{R}^n$ as

$$\mathcal{C}(x) := \{d \in \mathbb{R}^n : \nabla e_i(x)^T d = 0 \text{ for all } i \in [p],$$
$$\nabla c_i(x)^T d \leq 0 \text{ for all } i \in \mathcal{A}_0(x), \text{ and}$$
$$\nabla c_i(x)^T d = 0 \text{ for all } i \in \mathcal{A}(x) \setminus \mathcal{A}_0(x)\}.$$

Then, the second-order condition that we reference requires that, at x, one has

$$d^T \nabla_{xx}^2 \mathcal{L}(x, y, z) d > 0 \text{ for all } d \in \mathcal{C}(x) \setminus \{0\} \text{ and all } (y, z) \in \mathcal{D}(x). \tag{4}$$

We close this section with an observation and a lemma that will be referenced later in our discussions and analysis. First, recall that by Farkas' theorem, one has that for any $g \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times p}$, and $B \in \mathbb{R}^{n \times \ell}$ where $\ell \in [q]$, one and only one of the following two systems of equations and inequalities has a solution:

$$g + Ay + B\hat{z} = 0 \text{ and } \hat{z} \geq 0 \text{ for some } (y, \hat{z}) \in \mathbb{R}^p \times \mathbb{R}^\ell; \text{ or}$$
$$A^T d = 0, \ B^T d \leq 0, \text{ and } g^T d < 0 \text{ for some } d \in \mathbb{R}^n.$$

Thus, at any $x \in \mathbb{R}^n$ that is feasible for (2) in the sense that $e(x) = 0$ and $c(x) \leq 0$, either $\mathcal{D}(x)$ is nonempty or for any $\theta \in (0, \infty)$ the QP given by

$$\min_{d \in \mathbb{R}^n} \ \nabla f(x)^T d + \tfrac{1}{2}\theta \|d\|_2^2$$
$$\text{s.t. } e(x) + \nabla e(x)^T d = 0 \text{ and } c(x) + \nabla c(x)^T d \leq 0 \tag{6}$$

has a negative optimal value. In addition, recall the following lemma pertaining to solutions of perturbed linear systems; see, e.g., [6, Theorem 7.29] or [30, Theorem 2.1.2].

Lemma 2.1. *Suppose $K \in \mathbb{R}^{n \times n}$ is nonsingular and $b \in \mathbb{R}^n \setminus \{0\}$. Let $v \in \mathbb{R}^n$ be the unique solution of $Kv = b$ (meaning $v \neq 0$ since $b \neq 0$) and let $\tilde{v}$ satisfy $\tilde{K}\tilde{v} = \tilde{b}$, where $\tilde{K} = K + \delta K$ and $\tilde{b} = b + \delta b$. Then, for any vector-induced norm $\|\cdot\|$ and with $\mathrm{cond}(K) := \|K\|\|K^{-1}\|$, one has that if $\|\delta K\|\|K^{-1}\| < 1$, then*

$$\frac{\|v - \tilde{v}\|}{\|v\|} \leq \frac{\mathrm{cond}(K)}{1 - \|\delta K\|\|K^{-1}\|} \left(\frac{\|\delta K\|}{\|K\|} + \frac{\|\delta b\|}{\|b\|} \right).$$

3 Exact Function and Derivative Values

The purpose of this section is to state two active-set identification results for the setting of having exact function and derivative values. These are the results upon which we build for our new theoretical guarantees for noisy settings, which are presented in the next section.

Let $x_* \in \mathbb{R}^n$ be a local minimizer of problem (2). Each active-set identification result that we present is based on solving a subproblem that is designed to determine an estimate of the optimal active set with respect to x_*, namely, $\mathcal{A}(x_*)$, at some distinct point $x \in \mathbb{R}^n$. In particular, the theoretical results that we present state that, under certain conditions, the estimate of the optimal active set is correct as long as x is sufficiently close to x_*. One subproblem that we introduce in this section, an LP defined with respect to problem function and derivative values at x, involves computing an estimate of an element of $\mathcal{D}(x_*)$ that can be employed to estimate $\mathcal{A}(x_*)$. The second subproblem, a QP also defined with respect to function and derivative values at x, involves computing a step in the primal space that in turn can be employed to estimate $\mathcal{A}(x_*)$. The former technique is based directly on one established in [29]. The latter technique is similar to one from [29], except that we consider a QP subproblem whereas the approach in [29] employs an LP subproblem with a trust-region constraint. Our motivation for employing a QP is that this allows active-set identification to be done through subproblems that may already be solved for computing search

directions within the context of recently designed algorithms for noisy settings. Extending the primal-step-based approach from [29, §3.1] to the noisy setting may also be possible, but we omit it from consideration here.

Let $x \in \mathbb{R}^n$ be given. A strategy for estimating the active set at a nearby local minimizer x_* involves first obtaining Lagrange multiplier values $(y, z) \in \mathbb{R}^p \times \mathbb{R}^q$, and then evaluating the KKT error given by the function $\psi : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q$ defined as

$$\psi(x, y, z) = \left\| \begin{bmatrix} \nabla f(x) + \nabla e(x)y + \nabla c(x)z \\ e(x) \\ \min\{z, -c(x)\} \end{bmatrix} \right\|_1, \tag{7}$$

where the minimum is taken component-wise; recall (3). The motivation for obtaining Lagrange multiplier values and evaluating the KKT error can be seen through the following theorem, also stated in [29], which is similar to results from [18, 20, 35].

Theorem 3.1. *[29, Theorem 2.1] Suppose that, with respect to problem* (2), *the MFCQ and the second-order condition* (4) *hold at a local minimizer* $x_* \in \mathbb{R}^n$. *Then, there exist* $\epsilon \in (0, 1]$ *and* $C \in \mathbb{R}_{>0}$ *such that, for all* $(x, y, z) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q$ *with* $z \geq 0$ *and* $\mathrm{dist}(\{(x, y, z)\}, \{x_*\} \times \mathcal{D}(x_*)) \leq \epsilon$, *one has that*

$$C^{-1}\psi(x, y, z) \leq \mathrm{dist}(\{(x, y, z)\}, \{x_*\} \times \mathcal{D}(x_*)) \leq C\psi(x, y, z).$$

Motivated by Theorem 3.1, a technique is proposed in [29] for estimating the active set from a primal point $x \in \mathbb{R}^n$. Specifically, at x, the idea is to compute Lagrange multiplier values by minimizing $\psi(x, y, z)$ with respect to (y, z) subject to $z \geq 0$, i.e., to evaluate the function $\omega : \mathbb{R}^n \to \mathbb{R}$ at x that is defined by

$$\omega(x) := \min_{(y,z) \in \mathbb{R}^p \times \mathbb{R}^q_{\geq 0}} \psi(x, y, z). \tag{8}$$

This can be done by solving a linear program with equilibrium constraints (LPEC), which in turn can be solved as a mixed-integer linear program (MIP). However, since it can be computationally expensive to solve such a problem in large-scale settings, a technique is subsequently proposed in [29] to solve an LP approximation to the LPEC, the solution of which can in turn be used to estimate the active set. Let us refer to this as the LP-LPEC approach. This is the approach that we describe in detail here and extend to the noisy setting in the next section.

To state the LP-LPEC approach, let us first introduce $\kappa : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$, $\rho : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q_{\geq 0} \to \mathbb{R}$, and $\bar{\rho} : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q_{\geq 0} \to \mathbb{R}$ defined respectively by

$$\kappa(x, y, z) := \|\nabla f(x) + \nabla e(x)y + \nabla c(x)z\|_1 + \|e(x)\|_1, \tag{9a}$$

$$\rho(x, y, z) := \kappa(x, y, z) + \sum_{\{i : c_i(x) < 0\}} -c_i(x)z_i + \sum_{\{i : c_i(x) \geq 0\}} c_i(x), \tag{9b}$$

$$\text{and} \quad \bar{\rho}(x, y, z) := \kappa(x, y, z) + \sum_{\{i : c_i(x) < 0\}} (-c_i(x)z_i)^{1/2} + \sum_{\{i : c_i(x) \geq 0\}} c_i(x). \tag{9c}$$

At $x \in \mathbb{R}^n$, an approximate solution of the aforementioned LPEC can then be obtained by solving the following subproblem for the given constant $M \in \mathbb{R}_{>0}$:

$$\min_{(y,z) \in \mathbb{R}^p \times \mathbb{R}^q} \rho(x, y, z) \quad \text{s.t.} \quad 0 \leq z \leq M\mathbb{1}$$

$$\text{where} \quad M := \max \left\{ \|c(x_*)\|_\infty, \max_{(y_*, z_*) \in \mathcal{D}(x_*)} \left\| \begin{bmatrix} y_* \\ z_* \end{bmatrix} \right\|_\infty \right\} + 1. \tag{10}$$

Observe that M is finite when the MFCQ is assumed to hold at the local minimizer x_*. (In practice, M can be estimated by a large constant, or one might consider the constraint simply to be $z \geq 0$ and, if the LP has

6

a solution, using such a solution instead.) A solution of this subproblem can be obtained by solving the LP

$$
\min_{(y,z,r,s)\in\mathbb{R}^p\times\mathbb{R}^q\times\mathbb{R}^n\times\mathbb{R}^n} \mathbb{1}^T r + \mathbb{1}^T s + \sum_{\{i:c_i(x)<0\}} -c_i(x)z_i
$$
$$
\text{s.t. } \nabla f(x) + \nabla e(x)y + \nabla c(x)z = r - s, \tag{11}
$$
$$
r \geq 0,\ s \geq 0,\ \text{and } M\mathbb{1} \geq z \geq 0.
$$

Given x, let (y_x, z_x) be an optimal solution of (10). An active-set estimate can then be obtained through evaluation of the function $\bar{\rho}$; in particular, let the estimate be

$$
\mathcal{A}_{\mathrm{LP}}(y_x, z_x; x) := \{i \in [q] : c_i(x) \geq -(\beta\bar{\rho}(x, y_x, z_x))^\sigma\}, \tag{12}
$$

where $\beta \in \mathbb{R}_{>0}$ and $\sigma \in (0, 1)$ are user-defined constants. The following theorem from [29] shows that this LP-LPEC approach can be employed under certain assumptions to provide an accurate active-set estimate from x near a local minimizer x_*.

Theorem 3.2. *[29, Theorem 3.7] Suppose that, with respect to problem* (2), *the MFCQ and the second-order condition* (4) *hold at a local minimizer* $x_* \in \mathbb{R}^n$. *Then, with* $\epsilon \in (0, 1]$ *defined as in Theorem 3.1 and any* $\beta \in \mathbb{R}_{>0}$ *and* $\sigma \in (0, 1)$, *there exists* $\hat{\epsilon} \in (0, \epsilon]$ *such that for all* $x \in \mathbb{R}^n$ *satisfying* $\|x - x_*\| \leq \hat{\epsilon}$, *the following hold.*

(a) *There exists a solution* (y_x, z_x) *of* (10) *such that* $\mathrm{dist}(\{(y_x, z_x)\}, \mathcal{D}(x_*)) \leq \frac{1}{2}\epsilon$.

(b) *Any* (y_x, z_x) *solving* (10) *has* $M^{-1}\rho(x, y_x, z_x) \leq \omega(x) \leq M\bar{\rho}(x, y_x, z_x)$.

(c) *Any* (y_x, z_x) *solving* (10) *yields* $\mathcal{A}_{\mathrm{LP}}(y_x, z_x; x) = \mathcal{A}(x_*)$.

This completes our discussion of the LP-LPEC approach that we extend to the setting of noisy function and derivative values in the next section.

Let us now introduce the QP-based approach that we extend to the noisy setting as well. Note that if a local minimizer x_* is a primal KKT point and $x = x_*$, then (6) is feasible and by Farkas' theorem it has the unique solution $d_* = 0$. However, if x is not a local minimizer, then (6) might be infeasible, and even if it is feasible it might have a negative optimal value, indicating that x is not a primal KKT point. To handle each of these possibilities, and in particular to compute a primal search direction that can also be used to determine an optimal active-set estimate at x, one can choose $(\theta, \nu) \in \mathbb{R}_{>0} \times \mathbb{R}_{>0}$ and solve the nonsmooth subproblem

$$
\min_{d\in\mathbb{R}^n} \nabla f(x)^T d + \tfrac{1}{2}\theta\|d\|_2^2 + \nu\|e(x) + \nabla e(x)^T d\|_1 + \nu\|[c(x) + \nabla c(x)^T d]_+\|_1. \tag{13}
$$

Alternatively, one can solve the equivalent constrained QP subproblem

$$
\min_{(d,r,s,t)\in\mathbb{R}^n\times\mathbb{R}^p\times\mathbb{R}^p\times\mathbb{R}^q} \nabla f(x)^T d + \nu(\mathbb{1}^T r + \mathbb{1}^T s + \mathbb{1}^T t) + \tfrac{1}{2}\theta\|d\|_2^2
$$
$$
\text{s.t. } e(x) + \nabla e(x)^T d = r - s,
$$
$$
c(x) + \nabla c(x)^T d \leq t, \tag{14}
$$
$$
r \geq 0,\ s \geq 0,\ \text{and } t \geq 0.
$$

This subproblem is feasible for any (x, θ, ν); indeed, $d = 0$, $r = [e(x)]_+$, $s = [e(x)]_-$, and $t = [c(x)]_+$ form a feasible point. The dual of (14) can be written as the QP

$$
\max_{(\alpha,\beta)\in\mathbb{R}^p\times\mathbb{R}^q} e(x)^T\alpha + c(x)^T\beta - \tfrac{1}{2}\theta^{-1}\|\nabla f(x) + \nabla e(x)\alpha + \nabla c(x)\beta\|_2^2
$$
$$
\text{s.t. } -\nu\mathbb{1} \leq \alpha \leq \nu\mathbb{1} \text{ and } 0 \leq \beta \leq \nu\mathbb{1}. \tag{15}
$$

This dual QP subproblem is also feasible for any (x, θ, ν); indeed, $(\alpha, \beta) = (0, 0)$ is a feasible point. Consequently, by weak duality, the optimal values of (14) and (15) are both finite, and one has that these optimal values are attained by some (d, r, s, t) and (α, β), respectively; see, e.g., [4, Proposition 1.4.12].

Given any (x, θ, ν) and a corresponding solution $(d_{x,\theta,\nu}, r_{x,\theta,\nu}, s_{x,\theta,\nu}, t_{x,\theta,\nu})$ of subproblem (14), an estimate of the optimal active set is given by

$$\mathcal{A}_{\mathrm{QP}}(d_{x,\theta,\nu}; x) := \{i \in [q] : c_i(x) + \nabla c_i(x)^T d_{x,\theta,\nu} \geq 0\}. \tag{16}$$

Theorem 3.3 below shows that this approach is successful when x is sufficiently close to a local minimizer of (2) at which the LICQ and strict complementarity hold. Our proof follows closely that of Lemma 3.1, Theorem 3.2, and Corollary 3.3 in [29], although various details are different due to our use of a QP subproblem, whereas [29] considers an LP subproblem with a trust-region constraint.

Theorem 3.3. *Suppose that, with respect to problem (2), the MFCQ holds at a local minimizer $x_* \in \mathbb{R}^n$. Then, the following hold.*

(a) *There exists $(\epsilon, \underline{\theta}) \in \mathbb{R}_{>0} \times \mathbb{R}_{>0}$ such that, for any $(x, \theta, \nu) \in \mathbb{R}^n \times \mathbb{R}_{>\underline{\theta}} \times \mathbb{R}_{>0}$ with $\|x - x_*\| \leq \epsilon$, the solution $d_{x,\theta,\nu}$ of (13) yields $\mathcal{A}_{\mathrm{QP}}(d_{x,\theta,\nu}; x) \subseteq \mathcal{A}(x_*)$.*

(b) *With respect to any $\zeta \in (0,1)$, there exists $(\epsilon, \underline{\theta}, \underline{\nu}) \in \mathbb{R}_{>0}$ such that for any $(x, \theta, \nu) \in \mathbb{R}^n \times \mathbb{R}_{>\underline{\theta}} \times \mathbb{R}_{>\underline{\nu}}$ with $\|x - x_*\|_\infty \leq \epsilon$ and $\theta \leq \|x - x_*\|^{-\zeta}$, the solution subproblem $d_{x,\theta,\nu}$ of (13) yields*

$$-\nabla f(x_*) \in \mathrm{range}[\nabla e(x_*)] + \mathrm{pos}[\{\nabla c_i(x_*) : i \in \mathcal{A}_{\mathrm{QP}}(d_{x,\theta,\nu}; x)\}], \tag{17}$$

where $\mathrm{range}[\cdot]$ denotes the range space of the matrix input and $\mathrm{pos}[\cdot]$ denotes the set of nonnegative combinations of the input set of vectors.

(c) *If the LICQ holds at x_*, then for any $(\zeta, \epsilon, \underline{\theta}, \underline{\nu}, x, \theta, \nu)$ satisfying the conditions of part (b), the solution $d_{x,\theta,\nu}$ of (13) yields*

$$\mathcal{A}(x_*) \setminus \mathcal{A}_0(x_*) \subseteq \mathcal{A}_{\mathrm{QP}}(d_{x,\theta,\nu}; x) \subseteq \mathcal{A}(x_*).$$

If strict complementarity also holds at x_*, then $\mathcal{A}_{\mathrm{QP}}(d_{x,\theta,\nu}; x) = \mathcal{A}(x_*)$.

Proof. Consider part (a). First, by continuity of c, it follows that one can choose $\epsilon \in \mathbb{R}_{>0}$ small enough such that for any $x \in \mathbb{R}^n$ with $\|x - x_*\|_\infty \leq \epsilon$ and any $i \notin \mathcal{A}(x_*)$ one has that $c_i(x) \leq \frac{1}{2}c_i(x_*) < 0$. Second, given such an ϵ and for any such x and positive ν, it follows from (13) that $\|d_{x,\theta,\nu}\| \to 0$ as $\theta \nearrow \infty$. Thus, one can choose $\underline{\theta}$ sufficiently large enough such that $i \notin \mathcal{A}(x_*)$ implies $c_i(x) + \nabla c_i(x)^T d_{x,\theta,\nu} < 0$. Thus, the claim in part (a) follows from the definition of $\mathcal{A}_{\mathrm{QP}}(d_{x,\theta,\nu}; x)$ in (16).

Next, consider part (b). Let us first suppose that $(\epsilon, \underline{\theta})$ satisfies the conditions of part (a). Furthermore, to prove the claim, let us introduce

$$\underline{\nu} := \max \left\{ \left\| \begin{bmatrix} y_* \\ z_* \end{bmatrix} \right\|_\infty + 1 : (y_*, z_*) \in \mathcal{D}(x_*) \right\}. \tag{18}$$

The value $\underline{\nu}$ is well-defined and positive since x_* is a primal KKT point and the MFCQ assumption ensures that $\mathcal{D}(x_*)$ is bounded. For any pair $(y_*, z_*) \in \mathcal{D}(x_*)$, the dual subproblem (15) at x_* has an objective value of 0, as previously stated. Otherwise, for the dual subproblem at $x \neq x_*$ with $\|x - x_*\| \leq \epsilon$ and $\nu \geq \bar{\nu}$, a feasible point is given by $(\alpha, \beta) = (y_*, z_*)$, for which the corresponding dual objective value satisfies

$$e(x)^T y_* + c(x)^T z_* - \tfrac{1}{2}\theta^{-1}\|\nabla f(x) + \nabla e(x)y_* + \nabla c(x)z_*\|_2^2$$
$$= (e(x) - e(x_*))^T y_* + (c(x) - c(x_*))^T z_*$$
$$\quad - \tfrac{1}{2}\theta^{-1}\|\nabla f(x) - \nabla f(x_*) + (\nabla e(x) - \nabla e(x^*))y_* + (\nabla c(x) - \nabla c(x_*))z_*\|_2^2$$
$$\geq -L(\|x - x_*\|^2 + \|x - x_*\|)$$

for a sufficiently large constant $L \in \mathbb{R}_{>0}$. This latter inequality follows due to the continuous differentiability of the problem functions and $\|x - x_*\| \leq \epsilon$.

Now let us assume, for the sake of deriving a contradiction, that no matter how small one chooses $\epsilon \in \mathbb{R}_{>0}$, there exists a point $x \in \mathbb{R}^n$ satisfying $\|x - x_*\|_\infty \leq \epsilon$ such that the resulting active-set estimate $\mathcal{A}_{\mathrm{QP}}(d_{x,\theta,\nu}; x)$ yields

$$-\nabla f(x_*) \notin \mathrm{range}[\nabla e(x_*)] + \mathrm{pos}[\{\nabla c_i(x_*) : i \in \mathcal{A}_{\mathrm{QP}}(d_{x,\theta,\nu}; x)\}].$$

Since there are only finitely many active set estimates (i.e., subsets of $[q]$), it follows that there exists a sequence of points converging to x_* such that along this sequence the active-set estimate is the same set of indices, call it $\widehat{\mathcal{A}}$. At the same time, observe that by Rockafellar [32, Theorem 19.1] the set $\mathrm{range}[\nabla e(x_*)] + \mathrm{pos}[\{\nabla c_i(x_*) : i \in \widehat{\mathcal{A}}\}]$ is finitely generated, meaning that it is a closed set. Now, by the definition of the distance function in equation (1), one can define the quantity τ as

$$\tau := \tfrac{1}{2} \mathrm{dist}(-\nabla f(x_*), \mathrm{range}[\nabla e(x_*)] + \mathrm{pos}[\{\nabla c_i(x_*) : i \in \widehat{\mathcal{A}}\}]) > 0. \tag{19}$$

Reducing ϵ from above, if necessary, one has that

$$\mathrm{dist}(-\nabla f(x), \mathrm{range}[\nabla e(x)] + \mathrm{pos}[\{\nabla c_i(x) : i \in \widehat{\mathcal{A}}\}]) \geq \tau$$

for the given $\widehat{\mathcal{A}}$ and all x satisfying $\|x - x_*\|_\infty \leq \epsilon$. The justification for this claim follows from standard continuity arguments; see [29, Appendix 1].

Now consider any such x such that $\mathcal{A}_{\mathrm{QP}}(d_{x,\theta,\nu}; x) = \widehat{\mathcal{A}}$. Let (d_x, r_x, s_x, t_x) and (α_x, β_x) denote the solutions to (14) and (15), respectively. For any index $i \notin \widehat{\mathcal{A}}$, the KKT conditions for (14)–(15) and equation (16) together imply that

$$c_i(x) + \nabla c_i(x)^T d_x < 0, \quad [r_x]_i = 0, \quad \text{and} \quad [\beta_x]_i = 0. \tag{20}$$

One can now analyze the objective of the dual subproblem (15) by separating it into two parts. First, using property (20), one can observe that

$$
\begin{aligned}
& \frac{1}{2\theta} \|\nabla f(x) + \nabla e(x)\alpha_x + \nabla c(x)\beta_x\|_2^2 \\
\geq{} & \frac{1}{2\theta} \min_{(\alpha,\beta) \in \mathbb{R}^p \times \mathbb{R}^q_{\geq 0}} \|\nabla f(x) + \nabla e(x)\alpha + \nabla c(x)\beta\|_2^2 \\
={} & \frac{1}{2\theta} \min_{(\alpha,\beta) \in \mathbb{R}^p \times \mathbb{R}^q_{\geq 0}} \left\| \nabla f(x) + \nabla e(x)\alpha + \sum_{i \in \widehat{\mathcal{A}}} \nabla c_i(x)\beta_i \right\|_2^2 \\
={} & \frac{1}{2\theta} \mathrm{dist}(-\nabla f(x), \mathrm{range}[\nabla e(x)] + \mathrm{pos}[\{\nabla c_i(x) : i \in \widehat{\mathcal{A}}\}])^2 \geq \frac{1}{2\theta}\tau^2.
\end{aligned}
$$

From $\beta_x \geq 0$ and (15), one also that

$$e(x)^T \alpha_x + c(x)^T \beta_x \leq e(x)^T \alpha_x + (c(x))_+^T \beta_x \leq \nu\|e(x)\|_1 + \nu\|(c(x))_+\|_1.$$

Combining these inequalities into the dual objective (15), one has

$$e(x)^T \alpha_x + c(x)^T \beta_x - \frac{1}{2\theta}\|\nabla f(x) + \nabla e(x)\alpha_x + \nabla c(x)\beta_x\|_2^2$$

$$\leq \nu\|e(x)\|_1 + \nu\|(c(x))_+\|_1 - \frac{1}{2\theta}\tau^2. \tag{21}$$

Finally, one can decrease ϵ further, if necessary, to have that

$$\nu\|e(x)\|_1 + \nu\|(c(x))_+\|_1 - \frac{1}{2\theta}\tau^2 < -\frac{\|x - x_*\|^\zeta}{4}\tau^2. \tag{22}$$

9

This choice is possible since $e(x)$ and $(c(x))_+$ are both $O(\|x - x_*\|)$ and under the conditions of the lemma one has that $\theta \leq \|x - x_*\|^{-\zeta}$. Since $\zeta \in (0,1)$, the bounds (21)–(22) stand in direct contradiction to our earlier result, which established that the optimal value of the dual subproblem satisfies

$$-L(\|x - x_*\|^2 + \|x - x_*\|) \leq e(x)^T \alpha_x + c(x)^T \beta_x - \frac{1}{2\theta}\|\nabla f(x) + \nabla e(x)\alpha_x + \nabla c(x)\beta_x\|_2^2.$$

This inconsistency forces the conclusion that τ defined in equation (19) cannot be positive, and in fact must be 0. Thus, the inclusion in (17) must hold, as desired.

Finally, consider part (c). Since the LICQ holds, one has that $[z_*]_i > 0$ for all $i \in \mathcal{A}(x_*) \setminus \mathcal{A}_0(x_*)$. Thus, for the inclusion (17) to hold, any index i for which $[z_*]_i > 0$ must satisfy $i \in \mathcal{A}_{\mathrm{QP}}(d_{x,\theta,\nu}; x)$; otherwise, (17) would be violated. Therefore, one can conclude the left-hand side inclusion $\mathcal{A}(x_*) \setminus \mathcal{A}_0(x_*) \subseteq \mathcal{A}_{\mathrm{QP}}(d_{x,\theta,\nu}; x)$. The final conclusion follows directly from the fact that strict complementarity implies that $\mathcal{A}_0(x_*) = \emptyset$, meaning there are no active inequality constraints at the solution for which the corresponding multiplier is zero. The proof is complete. $\qquad\square$

The proof of Theorem 3.3 reveals the influence of the parameters of the QP approach. Intuitively, the lower threshold $\underline{\theta}$ ensures that the component $d_{x,\theta,\nu}$ is sufficiently small in norm such that the step in (16) does not "go too far" to include indices of constraints that are inactive at x_*. The lower threshold $\underline{\nu}$, on the other hand, is a typical type of bound for exact penalization; it ensures that, when possible, the solution of (14) will have the appropriate auxiliary variable values equal to zero.

4   Noisy Function and Derivative Values

In this section, we continue to consider techniques for active-set identification with respect to problem (2). However, we now assume that, at any $x \in \mathbb{R}^n$, one only has access to approximations of the values of f, e, and c and their first-order derivatives. More precisely, throughout this section, we assume that there exist positive error bounds $(\epsilon_f, \epsilon_e, \epsilon_c, \epsilon_{\nabla f}, \epsilon_{\nabla e}, \epsilon_{\nabla c})$ such that at all $x \in \mathbb{R}^n$ one can only compute $\tilde{f}(x) \in \mathbb{R}$, $\tilde{e}(x) \in \mathbb{R}^p$, $\tilde{c}(x) \in \mathbb{R}^q$, $\widetilde{\nabla}f(x) \in \mathbb{R}^n$, $\widetilde{\nabla}e(x) \in \mathbb{R}^{n \times p}$, and $\widetilde{\nabla}c(x) \in \mathbb{R}^{n \times q}$ satisfying

$$\|f(x) - \tilde{f}(x)\| \leq \epsilon_f, \quad \|e(x) - \tilde{e}(x)\| \leq \epsilon_e, \quad \|c(x) - \tilde{c}(x)\| \leq \epsilon_c,$$
$$\|\nabla f(x) - \widetilde{\nabla}f(x)\| \leq \epsilon_{\nabla f}, \quad \|\nabla e(x) - \widetilde{\nabla}e(x)\| \leq \epsilon_{\nabla e}, \quad \|\nabla c(x) - \widetilde{\nabla}c(x)\| \leq \epsilon_{\nabla c}. \tag{23}$$

Our theoretical guarantees are designed to hold regardless of whether these approximate function and derivative values are computed subject to deterministic or stochastic noise. For convenience, let us define a vector of these error bounds as

$$\epsilon_{\mathrm{err}} := [\epsilon_f, \epsilon_e, \epsilon_c, \epsilon_{\nabla f}, \epsilon_{\nabla e}, \epsilon_{\nabla c}]^T. \tag{24}$$

Our theoretical guarantees will show that, under reasonable assumptions, accurate active-set identification will occur when the components of ϵ_{err} are sufficiently small.

4.1   Identification through Lagrange Multipliers

We first describe a technique for identifying the active set by estimating Lagrange multipliers at a primal point $x \in \mathbb{R}^n$. This approach extends the LP-LPEC method described in §3, adapting it to settings with noisy function and derivative evaluations.

With the goal of again constructing a two-sided bound on $\omega(x)$—the minimization of $\psi(x, \cdot, \cdot)$ with respect to (y, z) as defined in (7) and (8)—we follow the structure of the functions in (9) to define their noisy counterparts $\tilde{\kappa} : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$, $\tilde{\rho} : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q_{\geq 0} \to \mathbb{R}$, and $\tilde{\bar{\rho}} : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q_{\geq 0} \to \mathbb{R}$ by

$$\tilde{\kappa}(x,y,z) := \|\widetilde{\nabla}f(x) + \widetilde{\nabla}e(x)y + \widetilde{\nabla}c(x)z\|_1 + \|\tilde{e}(x)\|_1, \tag{25a}$$

$$\tilde{\rho}(x,y,z) := \tilde{\kappa}(x,y,z) + \sum_{\{i:\tilde{c}_i(x)<0\}} -\tilde{c}_i(x)z_i + \sum_{\{i:\tilde{c}_i(x)\geq 0\}} \tilde{c}_i(x), \tag{25b}$$

10

$$\text{and} \quad \tilde{\bar{\rho}}(x,y,z) := \tilde{\kappa}(x,y,z) + \sum_{\{i:\tilde{c}_i(x)<0\}} (-\tilde{c}_i(x)z_i)^{1/2} + \sum_{\{i:\tilde{c}_i(x)\geq 0\}} \tilde{c}_i(x). \tag{25c}$$

The following relationship can be established for $\tilde{\rho}$ and $\tilde{\bar{\rho}}$, drawing on [29, Lemma 3.5] with respect to ρ and $\bar{\rho}$. We provide a proof for the sake of completeness.

Lemma 4.1. *For any* $(x,y,z) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}_{\geq 0}^q$, *it follows that*

$$\tilde{\bar{\rho}}(x,y,z) \leq \tilde{\rho}(x,y,z) + q^{1/2} \tilde{\rho}(x,y,z)^{1/2}.$$

Proof. Let $[(-\tilde{c}_i(x)z_i)^{1/2}]_{\tilde{c}_i(x)<0}$ denote a vector with elements $(-\tilde{c}_i(x)z_i)^{1/2}$ for all $i \in [q]$ such that $\tilde{c}_i(x) < 0$. Using the definitions of $\tilde{\rho}$ and $\tilde{\bar{\rho}}$ and basic norm properties, it follows that for any $(x,y,z) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q$ with $z \geq 0$ one finds

$$\begin{aligned}
\tilde{\bar{\rho}}(x,y,z) &= \tilde{\kappa}(x,y,z) + \left\| [(-\tilde{c}_i(x)z_i)^{1/2}]_{\tilde{c}_i(x)<0} \right\|_1 + \sum_{\{i:\tilde{c}_i(x)\geq 0\}} \tilde{c}_i(x) \\
&\leq \tilde{\kappa}(x,y,z) + q^{1/2} \left\| [(-\tilde{c}_i(x)z_i)^{1/2}]_{\tilde{c}_i(x)<0} \right\|_2 + \sum_{\{i:\tilde{c}_i(x)\geq 0\}} \tilde{c}_i(x) \\
&= \tilde{\kappa}(x,y,z) + q^{1/2} \left(\sum_{\{i:\tilde{c}_i(x)<0\}} (-\tilde{c}_i(x)z_i) \right)^{1/2} + \sum_{\{i:\tilde{c}_i\geq 0\}} \tilde{c}_i(x) \\
&\leq \tilde{\rho}(x,y,z) + q^{1/2} \tilde{\rho}(x,y,z)^{1/2},
\end{aligned}$$

which completes the proof. $\qquad \square$

A relationship between $\tilde{\rho}$, $\tilde{\bar{\rho}}$, and ψ (recall (7)) is given in our next result.

Lemma 4.2. *Consider arbitrary* $\overline{M} \geq 1$. *Then, there exist positive error bounds* ϵ_ρ *and* $\epsilon_{\bar{\rho}}$ *that tend to zero as* ϵ_{err} *tends to zero such that for all* $(x,y,z) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q$ *with* $z \geq 0$, $\|c\|_\infty \leq \overline{M}$, $\|y\|_\infty \leq \overline{M}$, *and* $\|z\|_\infty \leq \overline{M}$ *one has that*

$$\overline{M}^{-1} \tilde{\rho}(x,y,z) - \epsilon_\rho \leq \psi(x,y,z) \leq \tilde{\bar{\rho}}(x,y,z) + \epsilon_{\bar{\rho}}. \tag{26}$$

Furthermore, if such ϵ_ρ *and* (x,y,z) *also satisfy*

$$\rho(x,y,z) > 0 \quad \text{and} \quad \epsilon_\rho < \frac{\overline{M}^{-1} \rho(x,y,z)}{\overline{M}^{-1} + 1}, \tag{27}$$

then the left-hand side of (26) is strictly positive.

Proof. Consider arbitrary (x,y,z) satisfying the conditions of the lemma. Let us begin by proving the right-hand side of (26) by analyzing the following difference:

$$\begin{aligned}
& |\bar{\rho}(x,y,z) - \tilde{\bar{\rho}}(x,y,z)| \\
= & \left| \sum_{\{i:c_i(x)<0\}} (-c_i(x)z_i)^{1/2} - \sum_{\{i:\tilde{c}_i(x)<0\}} (-\tilde{c}_i(x)z_i)^{1/2} \right. \\
& + \sum_{\{i:c_i(x)\geq 0\}} c_i(x) - \sum_{\{i:\tilde{c}_i(x)\geq 0\}} \tilde{c}_i(x) \\
& + \|e(x)\|_1 - \|\tilde{e}(x)\|_1 \\
& \left. + \|\nabla f(x) + \nabla e(x)y + \nabla c(x)z\|_1 - \|\widetilde{\nabla f}(x) + \widetilde{\nabla e}(x)y + \widetilde{\nabla c}(x)z\|_1 \right| \\
\leq & \left| \sum_{\{i:c_i(x)<0\}} (-c_i(x)z_i)^{1/2} - \sum_{\{i:\tilde{c}_i(x)<0\}} (-\tilde{c}_i(x)z_i)^{1/2} \right.
\end{aligned}$$

11

$$+ \left. \sum_{\{i : c_i(x) \geq 0\}} c_i(x) - \sum_{\{i : \tilde{c}_i(x) \geq 0\}} \tilde{c}_i(x) \right|$$
$$+ \big| \|e(x)\|_1 - \|\tilde{e}(x)\|_1 \big|$$
$$+ \| \nabla f(x) + \nabla e(x) y + \nabla c(x) z \|_1 - \| \widetilde{\nabla} f(x) + \widetilde{\nabla} e(x) y + \widetilde{\nabla} c(x) z \|_1 \big|.$$

We proceed by bounding each of the two terms separately. Since it can be bounded more simply, let us start with the second term and observe that by (23) one finds

$$\big| \|e(x)\|_1 - \|\tilde{e}(x)\|_1$$
$$+ \| \nabla f(x) + \nabla e(x) y + \nabla c(x) z \|_1 - \| \widetilde{\nabla} f(x) + \widetilde{\nabla} e(x) y + \widetilde{\nabla} c(x) z \|_1 \big|$$
$$\leq \| e(x) - \tilde{e}(x) \|_1 + \| (\nabla f(x) - \widetilde{\nabla} f(x)) + (\nabla e(x) - \widetilde{\nabla} e(x)) y + (\nabla c(x) - \widetilde{\nabla} c(x)) z \|_1$$
$$\leq p^{1/2} \epsilon_e + n^{1/2} \epsilon_{\nabla f} + n^{1/2} \overline{M} (\epsilon_{\nabla e} + \epsilon_{\nabla c}).$$

Let us now return to bound the first term from above, namely,

$$\left| \sum_{\{i : c_i(x) < 0\}} (-c_i(x) z_i)^{1/2} - \sum_{\{i : \tilde{c}_i(x) < 0\}} (-\tilde{c}_i(x) z_i)^{1/2} \right. \tag{28}$$
$$\left. + \sum_{\{i : c_i(x) \geq 0\}} c_i(x) - \sum_{\{i : \tilde{c}_i(x) \geq 0\}} \tilde{c}_i(x) \right|.$$

For bookkeeping purposes, let us define the following sets of indices based on the sign of the constraint function values and their approximations. Specifically, let us define

$$\mathcal{PP}(x) := \{i : c_i(x) \geq 0 \text{ and } \tilde{c}_i(x) \geq 0\},$$
$$\mathcal{PN}(x) := \{i : c_i(x) \geq 0 \text{ and } \tilde{c}_i(x) < 0\},$$
$$\mathcal{NP}(x) := \{i : c_i(x) < 0 \text{ and } \tilde{c}_i(x) \geq 0\},$$
$$\text{and } \mathcal{NN}(x) := \{i : c_i(x) < 0 \text{ and } \tilde{c}_i(x) < 0\}.$$

The difference (28) can be expressed through sums over these sets:

$$\left| \sum_{i \in \mathcal{NP}(x)} (-c_i(x) z_i)^{1/2} + \sum_{i \in \mathcal{NN}(x)} (-c_i(x) z_i)^{1/2} + \sum_{i \in \mathcal{PP}(x)} c_i(x) + \sum_{i \in \mathcal{PN}(x)} c_i(x) \right.$$
$$\left. - \sum_{i \in \mathcal{PN}(x)} (-\tilde{c}_i(x) z_i)^{1/2} - \sum_{i \in \mathcal{NN}(x)} (-\tilde{c}_i(x) z_i)^{1/2} - \sum_{i \in \mathcal{PP}(x)} \tilde{c}_i(x) - \sum_{i \in \mathcal{NP}(x)} \tilde{c}_i(x) \right|.$$

Note that for all of the terms with indices corresponding to values of opposite signs—specifically for $i \in \mathcal{NP}(x)$ and $i \in \mathcal{PN}(x)$—it is straightforward to verify graphically that both $|c_i(x)|$ and $|\tilde{c}_i(x)|$ are bounded above by ϵ_c. Combined with $\|z\|_\infty \leq \overline{M}$, this allow us to bound each of these terms by expressions involving ϵ_c, $\epsilon_c^{1/2}$, and $\overline{M}^{1/2}$. Also, for the pair of terms with indices both corresponding to positive signs, one has

$$\left| \sum_{i \in \mathcal{PP}(x)} c_i(x) - \sum_{i \in \mathcal{PP}(x)} \tilde{c}_i(x) \right| \leq q \epsilon_c.$$

Now for the terms with indices both corresponding to negative signs, one finds that

$$\left| \sum_{i \in \mathcal{NN}(x)} \left((-c_i(x) z_i)^{1/2} - (-\tilde{c}_i(x) z_i)^{1/2} \right) \right|$$

$$= \left| \sum_{i \in \mathcal{NN}(x) \cap \{i: \tilde{c}_i(x) \le c_i(x)\}} ((-c_i(x)z_i)^{1/2} - (-\tilde{c}_i(x)z_i)^{1/2}) \right.$$

$$+ \left. \sum_{i \in \mathcal{NN}(x) \cap \{i: \tilde{c}_i(x) > c_i(x)\}} ((-c_i(x)z_i)^{1/2} - (-\tilde{c}_i(x)z_i)^{1/2}) \right|,$$

which one can assume without loss of generality implies that

$$\left| \sum_{i \in \mathcal{NN}(x)} ((-c_i(x)z_i)^{1/2} - (-\tilde{c}_i(x)z_i)^{1/2}) \right|$$

$$\le \left| \sum_{i \in \mathcal{NN}(x) \cap \{i: \tilde{c}_i(x) > c_i(x)\}} ((-c_i(x)z_i)^{1/2} - (-\tilde{c}_i(x)z_i)^{1/2}) \right|.$$

(Our analysis would be similar if the sum over $i \in \mathcal{NN}(x) \cap \{i : \tilde{c}_i(x) \le c_i(x)\}$, which is a negative value, were to dominate the sum within the absolute value.) In order to further bound this term on the right-hand side, let us consider two subcases. First, if for some i in the sum one has $-c_i(x) \le \epsilon_c$, then it follows (since $i \in \mathcal{NN}(x)$) that

$$|(-c_i(x)z_i)^{1/2} - (-\tilde{c}_i(x)z_i)^{1/2}| \le (-c_i(x)z_i)^{1/2} \le \epsilon_c^{1/2} \overline{M}^{1/2}.$$

Otherwise, if for some i in the sum one has that $-c_i(x) > \epsilon_c$, then by the Mean Value Theorem there exists $\zeta \in [-\tilde{c}_i(x), -c_i(x)]$ with

$$(-c_i(x))^{1/2} - (-\tilde{c}_i(x))^{1/2} = s'(\zeta)((-c_i(x)) - (-\tilde{c}_i(x))),$$

where $s(\cdot)$ is the square root function. Since $-c_i(x) > \epsilon_c$, the largest value that $s'(\zeta)$ can take is bounded by the slope of the line segment from $(0,0)$ to the point $(\epsilon_c, \epsilon_c^{1/2})$, which is $\frac{\epsilon_c^{1/2}}{\epsilon_c}$. Consequently, one finds by $i \in \mathcal{NN}(x) \cap \{i : \tilde{c}_i(x) > c_i(x)\}$ that

$$|(-c_i(x)z_i)^{1/2} - (-\tilde{c}_i(x)z_i)^{1/2}|$$

$$= ((-c_i(x))^{1/2} - (-\tilde{c}_i(x))^{1/2})z_i^{1/2} \le \frac{\epsilon_c^{1/2}}{\epsilon_c}((-c_i(x)) - (-\tilde{c}_i(x)))z_i^{1/2} \le \epsilon_c^{1/2} \overline{M}^{1/2}.$$

Putting everything together, one finds that

$$\begin{aligned} &|\bar{\rho}(x,y,z) - \tilde{\bar{\rho}}(x,y,z)| \\ &\le 3q^{1/2}(\epsilon_c + \epsilon_c^{1/2}\overline{M}^{1/2}) + p^{1/2}\epsilon_e + n^{1/2}\epsilon_{\nabla f} + n^{1/2}\overline{M}(\epsilon_{\nabla e} + \epsilon_{\nabla c}) =: \epsilon_{\bar{\rho}}, \end{aligned} \tag{29}$$

where, as claimed, one has that $\epsilon_{\bar{\rho}} \to 0$ as $\epsilon_{\text{err}} \to 0$. Hence, with [29, Lemma 3.6],

$$\begin{aligned} \psi(x,y,z) &\le \bar{\rho}(x,y,z) \\ &= \tilde{\bar{\rho}}(x,y,z) + \bar{\rho}(x,y,z) - \tilde{\bar{\rho}}(x,y,z) \\ &\le \tilde{\bar{\rho}}(x,y,z) + \epsilon_{\bar{\rho}}, \end{aligned}$$

which establishes the right-hand side of (26).

Applying the same procedure, now to $|\tilde{\rho}(x,y,z) - \rho(x,y,z)|$, one finds that this difference can be bounded by ϵ_{ρ}, where $\epsilon_{\rho} \to 0$ as $\epsilon_{\text{err}} \to 0$; the proof is omitted since it follows by similar arguments to those above. Thus, with [29, Lemma 3.6],

$$\begin{aligned} \overline{M}\psi(x,y,z) &\ge \rho(x,y,z) \\ &= \rho(x,y,z) - \tilde{\rho}(x,y,z) + \tilde{\rho}(x,y,z) \end{aligned}$$

13

$$\geq \tilde{\rho}(x, y, z) - \epsilon_{\rho},$$

thus verifying the left-hand side of (26).

Finally, since $|\tilde{\rho}(x, y, z) - \rho(x, y, z)| \leq \epsilon_{\rho}$, one has $\tilde{\rho}(x, y, z) \geq \rho(x, y, z) - \epsilon_{\rho}$, so

$$\overline{M}^{-1}\tilde{\rho}(x, y, z) - \epsilon_{\rho} \geq \overline{M}^{-1}(\rho(x, y, z) - \epsilon_{\rho}) - \epsilon_{\rho} = \overline{M}^{-1}\rho(x, y, z) - \epsilon_{\rho}(\overline{M}^{-1} + 1),$$

which is strictly positive provided that (27) holds, as claimed. □

It should be emphasized that the result of Lemma 4.2 shows that, in order for the left-hand side of (26) to be strictly positive, the error bound ϵ_{ρ} must be small in proportion to $\rho(x, y, z) > 0$, a quantity that can be smaller as x approaches x_*. We will provide further comments about the consequence of this in §4.3.

The LP-LPEC-based active-set identification strategy can now be established as follows. First, one computes a solution (y_x, z_x) of the optimization problem

$$\min_{(y,z) \in \mathbb{R}^p \times \mathbb{R}^q} \tilde{\rho}(x, y, z) \text{ s.t. } 0 \leq z \leq M\mathbb{1}, \tag{30}$$

where M is defined in (10). (See our comments after (10) about how this constant can be chosen in practice.) The solution of (30) can be obtained by solving the LP

$$\min_{(y,z,r,s) \in \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^n \times \mathbb{R}^n} \mathbb{1}^T r + \mathbb{1}^T s + \sum_{\{i : \tilde{c}_i(x) < 0\}} -\tilde{c}_i(x) z_i$$

$$\text{s.t. } \widetilde{\nabla f}(x) + \widetilde{\nabla e}(x) y + \widetilde{\nabla c}(x) z = r - s, \tag{31}$$

$$r \geq 0, \ s \geq 0, \ M\mathbb{1} \geq z \geq 0.$$

The active-set estimate is then defined, similarly to (12), by

$$\widetilde{\mathcal{A}}_{\text{LP}}(y_x, z_x; x) := \{i \in [q] : \tilde{c}_i(x) \geq -(\beta\tilde{\rho}(x, y_x, z_x))^{\sigma}\} \tag{32}$$

where $\beta \in \mathbb{R}_{>0}$ and $\sigma \in (0, 1)$ are user-defined constants.

In the following theorem, we demonstrate that this methodology yields a two-sided bound on the value of $\omega(x)$ and can accurately identify the optimal active set at a point near a KKT point under moderate assumptions.

Theorem 4.1. *Suppose that, with respect to* (2), *the MFCQ and the second-order condition* (4) *hold at a local minimizer* $x_* \in \mathbb{R}^n$. *Let* ϵ *be as defined in Theorem 3.1, and for arbitrary* $\overline{M} \geq 1$ *let* ϵ_{ρ} *and* $\epsilon_{\tilde{\rho}}$ *be defined as in Lemma 4.2. Then, there exists* $\hat{\epsilon} \in (0, \epsilon/2]$ *such that for each* x *with* $\|x - x_*\| \leq \hat{\epsilon}$ *and* $\bar{\rho}(x, y_x, z_x) > 0$, *there exist* $\tilde{\epsilon}_x \in \mathbb{R}_{>0}$, $\epsilon_e \in \mathbb{R}_{>0}$, $\epsilon_c \in \mathbb{R}_{>0}$, $\epsilon_{\nabla f} \in \mathbb{R}_{>0}$, $\epsilon_{\nabla e} \in \mathbb{R}_{>0}$, *and* $\epsilon_{\nabla c} \in \mathbb{R}_{>0}$ *such that if* ϵ_{err} *defined by* (24) *satisfies* $\|\epsilon_{err}\|_{\infty} < \tilde{\epsilon}_x$, *then*

(a) $0 < M^{-1}\tilde{\rho}(x, y_x, z_x) - \epsilon_{\rho} \leq \omega(x) \leq \tilde{\rho}(x, y_x, z_x) + \epsilon_{\tilde{\rho}}$ *and*

(b) $\widetilde{\mathcal{A}}_{\text{LP}}(y_x, z_x; x) = \mathcal{A}(x_*).$

Proof. Consider part (a). First, let $\hat{\epsilon}$ satisfy $0 < \hat{\epsilon} \leq \bar{\epsilon} \leq \epsilon/2$, where $\bar{\epsilon}$ is defined in [29, Theorem 3.4]. Then, reduce $\hat{\epsilon}$ to a smaller positive number, if necessary, to ensure that $\|c(x)\|_{\infty} \leq M$ for all x with $\|x - x_*\| \leq \hat{\epsilon}$. Next, observe that according to [29, Theorem 3.4(i)], the minimizer of the problem in (8) that defines $\omega(x)$ satisfies $\text{dist}((y, z), \mathcal{D}(x_*)) \leq \epsilon/2$. Hence, for any Lagrange multiplier pair $(y_*, z_*) \in \mathcal{D}(x_*)$, it follows that $\|y\|_{\infty} \leq \|y_*\| + 1 \leq M$ and $\|z\|_{\infty} \leq \|z_*\| + 1 \leq M$.

Now using Lemma 4.2, and by replacing $\overline{M}$ by M, one has

$$\tilde{\rho}(x, y, z) + \epsilon_{\tilde{\rho}} \geq \psi(x, y_x, z_x)$$
$$\geq \min_{y, 0 \leq z \leq M\mathbb{1}} \psi(x, y, z)$$

14

$$\geq \min_{y, 0 \leq z \leq M\mathbb{1}} \left(M^{-1}\tilde{\rho}(x, y, z) - \epsilon_\rho \right)$$

$$= M^{-1}\tilde{\rho}(x, y_x, z_x) - \epsilon_\rho > 0.$$

However, since $M \geq 1$ and $\epsilon < 1$ and as was shown in [29, Theorem 3.4], the minimizer of $\psi(x, y, z)$ over (y, z) with $z \geq 0$ is attained at values of (y, z) that satisfy the restriction $\|z\|_\infty \leq M$. Thus one can write, by (8), that

$$0 < M^{-1}\tilde{\rho}(x, y_x, z_x) - \epsilon_\rho \leq \min_{y, 0 \leq z} \psi(x, y, z) \leq \tilde{\tilde{\rho}}(x, y_x, z_x) + \epsilon_{\bar{\rho}}.$$

Now consider part (b). One has from Theorem 3.2 that $\bar{\rho}(x, y_x, z_x) \to 0$ as $x \to x_*$. Thus, for any $\hat{\beta} > \beta$, and using continuity of c and (29), it follows for $i \notin \mathcal{A}(x_*)$ that one can decrease $\hat{\epsilon}$ if necessary to ensure that for $\|x - x_*\| \leq \hat{\epsilon}$ one finds

$$\tilde{c}_i(x) = \tilde{c}_i(x) - c_i(x) + c_i(x) \leq \epsilon_c + c_i(x)$$

$$< \epsilon_c + \frac{1}{2}c_i(x_*)$$

$$\leq \epsilon_c - (\hat{\beta}\bar{\rho}(x, y_x, z_x))^\sigma$$

$$\leq \epsilon_c - (\hat{\beta}(\tilde{\rho}(x, y_x, z_x) + \epsilon_{\bar{\rho}}))^\sigma.$$

For any such x there exists $\tilde{\epsilon}_x > 0$ sufficiently small such that $\|\epsilon_{\mathrm{err}}\|_\infty < \tilde{\epsilon}_x$ implies

$$\epsilon_c - (\hat{\beta}(\tilde{\rho}(x, y_x, z_x) + \epsilon_{\bar{\rho}}))^\sigma \leq -(\beta\tilde{\tilde{\rho}}(x, y_x, z_x))^\sigma.$$

Hence, $i \notin \widetilde{\mathcal{A}}_{\mathrm{LP}}(y_x, z_x; x)$. Now for $i \in \mathcal{A}(x_*)$ and sufficiently small $\hat{\epsilon} > 0$, one has, using [29, Theorem 3.4(ii)] and [29, Theorem 3.7(ii)], that for any $\hat{\beta} > \beta$

$$|\tilde{c}_i(x)| = |\tilde{c}_i(x) - c_i(x) + c_i(x)|$$

$$\leq |\tilde{c}_i(x) - c_i(x)| + |c_i(x)|$$

$$\leq \epsilon_c + |c_i(x)|$$

$$\leq \epsilon_c + (\hat{\beta}\bar{\rho}(x, y_x, z_x))^\sigma,$$

For any such x there exists $\tilde{\epsilon}_x > 0$ sufficiently small such that $\|\epsilon_{\mathrm{err}}\|_\infty < \tilde{\epsilon}_x$ implies

$$\epsilon_c + (\hat{\beta}\bar{\rho}(x, y_x, z_x))^\sigma \leq \epsilon_c + (\hat{\beta}(\tilde{\rho}(x, y_x, z_x) + \epsilon_\rho))^\sigma \leq (\beta\tilde{\tilde{\rho}}(x, y_x, z_x))^\sigma.$$

Hence, one finds $i \in \widetilde{\mathcal{A}}_{\mathrm{LP}}(y_x, z_x; x)$. The proof is complete. □

4.2 Identification through a Primal Step

We now present a technique for active-set identification in noisy settings that is based on computing a primal step through solving a QP. The approach is an extension of the one that offers Theorem 3.3 for the exact setting. Our primary tool for extending the approach to noisy settings is perturbation theory for the solution of linear systems, in particular, Lemma 2.1.

Analogously to (14), consider the QP subproblem

$$\min_{(d,r,s,t)\in\mathbb{R}^n\times\mathbb{R}^p\times\mathbb{R}^p\times\mathbb{R}^q} \widetilde{\nabla f}(x)^T d + \nu(\mathbb{1}^T r + \mathbb{1}^T s + \mathbb{1}^T t) + \frac{1}{2}\theta\|d\|_2^2$$

$$\text{s.t. } \tilde{e}(x) + \widetilde{\nabla e}(x)^T d = r - s,$$

$$\tilde{c}(x) + \widetilde{\nabla c}(x)^T d \leq t, \tag{33}$$

$$r \geq 0, \ s \geq 0, \text{ and } t \geq 0.$$

Like (14), this QP is always feasible and has a finite optimal value. Later in our analysis we will recall the well-known fact from exact penalty function theory that under a constraint qualification and if the parameter ν is sufficiently large, the (unique) solution component $d_{x,\theta,\nu}$ of (33) also solves the QP given by

$$
\min_{d \in \mathbb{R}^n} \; \widetilde{\nabla}f(x)^T d + \tfrac{1}{2}\theta\|d\|_2^2
$$
$$
\text{s.t. } \tilde{e}(x) + \widetilde{\nabla}e(x)^T d = 0 \text{ and } \tilde{c}(x) + \widetilde{\nabla}c(x)^T d \le 0; \tag{34}
$$

i.e., under these conditions, the solution of (33) has $(r_{x,\theta,\nu}, s_{x,\theta,\nu}, t_{x,\theta,\nu}) = (0,0,0)$.

Our first lemma is based on the relatively straightforward observation that if the LICQ holds at a local minimizer x_* of problem (2), a point x is sufficiently close to x_*, and the noise in certain constraint derivative values at x is sufficiently small, then the corresponding constraint derivative matrix has full rank.

Lemma 4.3. *Suppose that, with respect to problem (2), the LICQ holds at a local minimizer $x_* \in \mathbb{R}^n$. Let $\mathcal{A}_* := \mathcal{A}(x_*)$ and $\ell := |\mathcal{A}_*|$. Then, there exist $\epsilon_x \in \mathbb{R}_{>0}$, $\epsilon_{\nabla e} \in \mathbb{R}_{>0}$, and $\epsilon_{\nabla c} \in \mathbb{R}_{>0}$ so any $(x, \widetilde{\nabla}e(x), \widetilde{\nabla}c_{\mathcal{A}_*}(x)) \in \mathbb{R}^n \times \mathbb{R}^{n\times p} \times \mathbb{R}^{n\times\ell}$ with*

$$
\|x - x_*\| \le \epsilon_x, \; \|\widetilde{\nabla}e(x) - \nabla e(x)\| \le \epsilon_{\nabla e}, \text{ and } \|\widetilde{\nabla}c_{\mathcal{A}_*}(x) - \nabla c_{\mathcal{A}_*}(x)\| \le \epsilon_{\nabla c} \tag{35}
$$

yields that the matrix $\begin{bmatrix} \widetilde{\nabla}e(x) & \widetilde{\nabla}c_{\mathcal{A}_*}(x) \end{bmatrix}$ has full column rank.

Proof. Under the conditions of the lemma, specifically that the LICQ with respect to problem (2) holds at x_*, the matrix $\begin{bmatrix} \nabla e(x_*) & \nabla c_{\mathcal{A}_*}(x_*) \end{bmatrix}$ has full column rank. Then, since the functions $e_i(x)$ and $c_i(x)$ are continuously differentiable, it follows that $\begin{bmatrix} \nabla e(x) & \nabla c_{\mathcal{A}_*}(x) \end{bmatrix}$ is continuous as a function of x. Since the set of full-column-rank matrices is open, there exists $\epsilon_x \in \mathbb{R}_{>0}$ such that for all $x \in \mathbb{R}^n$ with $\|x - x_*\| \le \epsilon_x$, the matrix $\begin{bmatrix} \nabla e(x) & \nabla c_{\mathcal{A}_*}(x) \end{bmatrix}$ has full column rank. Consequently, for any such ϵ_x, there exist sufficiently small $\epsilon_{\nabla e} \in \mathbb{R}_{>0}$ and $\epsilon_{\nabla c} \in \mathbb{R}_{>0}$ such that $\begin{bmatrix} \widetilde{\nabla}e(x) & \widetilde{\nabla}c_{\mathcal{A}_*}(x) \end{bmatrix}$ has full column rank as well, which completes the proof. $\square$

Our next lemma shows that if x is sufficiently close to a local minimizer x_* and the noise in the function and gradient evaluation is all sufficiently small, then the QP subproblem (34) is feasible and its primal-dual solution is given by the solution of a certain linear system defined with respect to the active set at x_*.

Lemma 4.4. *Suppose that, with respect to problem (2), the LICQ and strict complementarity hold at a local minimizer $x_* \in \mathbb{R}^n$. Let $\mathcal{A}_* := \mathcal{A}(x_*)$ and $\ell := |\mathcal{A}_*|$. Then, for any $\theta \in \mathbb{R}_{>0}$, there exist $\epsilon_x \in \mathbb{R}_{>0}$, $\epsilon_e \in \mathbb{R}_{>0}$, $\epsilon_c \in \mathbb{R}_{>0}$, $\epsilon_{\nabla f} \in \mathbb{R}_{>0}$, $\epsilon_{\nabla e} \in \mathbb{R}_{>0}$, and $\epsilon_{\nabla c} \in \mathbb{R}_{>0}$ such that for all $(x, \tilde{e}(x), \tilde{c}(x), \widetilde{\nabla}f(x), \widetilde{\nabla}e(x), \widetilde{\nabla}c(x))$ satisfying $\|x - x_*\| \le \epsilon_x$ and the error bounds (23) it follows that the linear system*

$$
\begin{bmatrix} \theta I & \widetilde{\nabla}e(x) & \widetilde{\nabla}c_{\mathcal{A}_*}(x) \\ \widetilde{\nabla}e(x)^T & 0 & 0 \\ \widetilde{\nabla}c_{\mathcal{A}_*}(x)^T & 0 & 0 \end{bmatrix} \begin{bmatrix} d \\ \alpha \\ \beta_{\mathcal{A}_*} \end{bmatrix} = - \begin{bmatrix} \widetilde{\nabla}f(x) \\ \tilde{e}(x) \\ \tilde{c}_{\mathcal{A}_*}(x) \end{bmatrix} \tag{36}
$$

has a unique solution $(d, \alpha, \beta_{\mathcal{A}_*}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^\ell$ which, along with $\beta_{\mathcal{A}_*^c} = 0$, satisfies the KKT conditions of subproblem (34); in particular, one has

$$
\widetilde{\nabla}f(x) + \theta d + \widetilde{\nabla}e(x)\alpha + \widetilde{\nabla}c(x)^T\beta = 0, \tag{37a}
$$

$$
\tilde{e}(x) + \widetilde{\nabla}e(x)^T d = 0, \quad \tilde{c}_{\mathcal{A}_*}(x) + \widetilde{\nabla}c_{\mathcal{A}_*}(x)^T d = 0, \tag{37b}
$$

$$
\tilde{c}_{\mathcal{A}_*^c}(x) + \widetilde{\nabla}c_{\mathcal{A}_*^c}(x)^T d < 0, \tag{37c}
$$

$$
\beta_{\mathcal{A}_*} > 0, \quad \beta_{\mathcal{A}_*^c} = 0, \tag{37d}
$$

$$
\text{and } \tilde{c}(x) + \widetilde{\nabla}c(x)^T d \perp \beta. \tag{37e}
$$

Proof. Under the conditions of the lemma, the conditions of Lemma 4.3 hold, which in turn implies that there exist $\epsilon_x \in \mathbb{R}_{>0}$, $\epsilon_{\nabla e} \in \mathbb{R}_{>0}$, and $\epsilon_{\nabla c} \in \mathbb{R}_{>0}$ such that the matrix in (36), call it $\widetilde{K}$, is invertible and the linear system in (36) has a unique solution. Let such $(\epsilon_x, \epsilon_{\nabla e}, \epsilon_{\nabla c})$ be given. Now we note that, by construction of the linear system (36) and with $\beta_{\mathcal{A}_*^c} = 0$, the conditions (37a), (37b), and (37e) are all satisfied. All that remains is to show that, when x is sufficiently close to x_* and all of the noise in the function and derivative values is sufficiently small, the conditions (37c) and (37d) hold as well. Toward this end, let us define

$$K := \begin{bmatrix} \theta I & \nabla e(x_*) & \nabla c_{\mathcal{A}_*}(x_*) \\ \nabla e(x_*)^T & 0 & 0 \\ \nabla c_{\mathcal{A}_*}(x_*)^T & 0 & 0 \end{bmatrix} \text{ and } \begin{bmatrix} d_* \\ \alpha_* \\ [\beta_{\mathcal{A}_*}]_* \end{bmatrix} = -K^{-1} \begin{bmatrix} \nabla f(x_*) \\ e(x_*) \\ c_{\mathcal{A}_*}(x_*) \end{bmatrix}.$$

Under the conditions of the lemma, it follows by Farkas' theorem that $d_* = 0$ and, along with $[\beta_{\mathcal{A}_*^c}]_* = 0$, the pair $(\alpha_*, \beta_*) =: (y_*, z_*)$ is the unique Lagrange multiplier such that (x_*, y_*, z_*) is a KKT point for problem (2). Indeed, by the strict complementarity assumption in the lemma, it follows that $[\beta_{\mathcal{A}_*}]_* > 0$. Now note that the matrix K is invertible (e.g., by setting $x = x_*$ in Lemma 4.3), and that one has $\widetilde{K} = K + \delta K$, where— since f, e, and c are continuously differentiable—one finds that $\delta K \to 0$ as $x \to x_*$ and $\epsilon_{\text{err}} \to 0$, where ϵ_{err} is defined in (24). Consequently, $\|\delta K\|\|K^{-1}\| \to 0$ as $x \to x_*$ and $\epsilon_{\text{err}} \to 0$, which means that, for x sufficiently close to x_* and ϵ_{err} sufficiently small, one has from Lemma 2.1 that (36) yields

$$\frac{\left\| \begin{bmatrix} d_* - d \\ \alpha_* - \alpha \\ [\beta_{\mathcal{A}_*}]_* - \beta_{\mathcal{A}_*} \end{bmatrix} \right\|}{\left\| \begin{bmatrix} d_* \\ \alpha_* \\ [\beta_{\mathcal{A}_*}]_* \end{bmatrix} \right\|} \leq \frac{\text{cond}(K)}{1 - \|\delta K\|\|K^{-1}\|} \left(\frac{\|\delta K\|}{\|K\|} + \frac{\left\| \begin{bmatrix} \widetilde{\nabla} f(x) \\ \tilde{e}(x) \\ \tilde{c}_{\mathcal{A}_*}(x) \end{bmatrix} - \begin{bmatrix} \nabla f(x_*) \\ e(x_*) \\ c_{\mathcal{A}_*}(x_*) \end{bmatrix} \right\|}{\left\| \begin{bmatrix} \nabla f(x_*) \\ e(x_*) \\ c_{\mathcal{A}_*}(x_*) \end{bmatrix} \right\|} \right).$$

Overall, since $d_* = 0$ and $[\beta_{\mathcal{A}_*}]_* > 0$, it follows that for $\epsilon_x \in \mathbb{R}_{>0}$, $\epsilon_e \in \mathbb{R}_{>0}$, $\epsilon_c \in \mathbb{R}_{>0}$, $\epsilon_{\nabla f} \in \mathbb{R}_{>0}$, $\epsilon_{\nabla e} \in \mathbb{R}_{>0}$, and $\epsilon_{\nabla c} \in \mathbb{R}_{>0}$ all sufficiently small, any $(x, \tilde{e}(x), \tilde{c}(x), \widetilde{\nabla} f(x), \widetilde{\nabla} e(x), \widetilde{\nabla} c(x))$ satisfying $\|x - x_*\| \leq \epsilon_x$ and the error bounds (23) yield that all conditions in (37) hold, as desired. $\square$

We are now prepared to prove our main theorem of this subsection, for which we refer to the active-set estimate defined similarly to (16) as

$$\widetilde{\mathcal{A}}_{\text{QP}}(d; x) := \{i \in [q] : \tilde{c}_i(x) + \widetilde{\nabla} c_i(x)^T d \geq 0\}. \tag{38}$$

Theorem 4.2. *Suppose that, with respect to (2), the LICQ and strict complementarity hold at a strict local minimizer $x_* \in \mathbb{R}^n$. Let $\mathcal{A}_* := \mathcal{A}(x_*)$ and $\ell := |\mathcal{A}_*|$. Then, for any $\theta \in \mathbb{R}_{>0}$, there exist $\epsilon_x \in \mathbb{R}_{>0}$, $\epsilon_e \in \mathbb{R}_{>0}$, $\epsilon_c \in \mathbb{R}_{>0}$, $\epsilon_{\nabla f} \in \mathbb{R}_{>0}$, $\epsilon_{\nabla e} \in \mathbb{R}_{>0}$, $\epsilon_{\nabla c} \in \mathbb{R}_{>0}$, and $\underline{\nu} \in \mathbb{R}_{>0}$ such that for all $(x, \tilde{e}(x), \tilde{c}(x), \widetilde{\nabla} f(x), \widetilde{\nabla} e(x), \widetilde{\nabla} c(x), \nu)$ with $\|x - x_*\| \leq \epsilon_x$, the error bounds (23), and $\nu \geq \underline{\nu}$, the solution (d, r, s, t) of (33) is unique and the active-set estimate defined by (38) yields $\widetilde{\mathcal{A}}_{\text{QP}}(d; x) = \mathcal{A}_*$.*

Proof. Under the stated conditions, it follows by Lemma 4.4 that for any $\theta \in \mathbb{R}_{>0}$ one has for x sufficiently close to x_* and sufficiently small errors that (34) is feasible and has a unique minimizer at which the LICQ holds. Consequently, by [21, Theorem 4.4], it follows that for any $\theta \in \mathbb{R}_{>0}$ there exist $(\epsilon_x, \epsilon_e, \epsilon_c, \epsilon_{\nabla f}, \epsilon_{\nabla e}, \epsilon_{\nabla c})$ as stated such that with any $\nu \geq \underline{\nu} := M$ (where M is defined as in (10)) the solution (d, r, s, t) of (33) has $(r, s, t) = 0$, which in turn means by Lemma 4.4 that d is the unique solution of (34). All that remains is to observe that by Lemma 4.4 one can conclude that for x sufficiently close to x_*, sufficiently small noise, and ν sufficiently large, one also has the the active-set estimate given by (38) yields $\mathcal{A}_*$. In particular, this fact follows by the latter equation in (37b) along with (37c). $\square$

4.3 Active-Set Identification within a Stochastic Algorithm

Let us now consider the use of active-set identification procedures within the context of a stochastic algorithm for solving (2). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space that captures all outcomes of a run of the algorithm.

That is, each possible run of the algorithm is associated with an outcome $\omega \in \Omega$, an infinite-dimensional tuple whose kth element determines the outcome of any stochastic quantity in iteration $k \in \mathbb{N}$. In this manner, the algorithm defines a stochastic process $\{X_k\}$, where for all $k \in \mathbb{N}$ the random variable X_k is the primal iterate generated by the algorithm. Given, for simplicity, the initial condition that $X_1(\omega) = x_1$ for some $x_1 \in \mathbb{R}^n$ for all $\omega \in \Omega$, let $\mathcal{F}_k$ denote the sub-σ-algebra of $\mathcal{F}$ corresponding to the initial condition and all of the stochastic quantities prior to iteration k. In this manner $\{\mathcal{F}_k\}$ is a filtration.

Our first setting, for identification through Lagrange multipliers presented in §4.1, requires that the MFCQ and the second-order condition (4) hold at a local minimizer x_*. For all $k \in \mathbb{N}$, the LP (30) is solved with $x = X_k$ to yield a Lagrange multiplier pair (Y_k, Z_k). The following theorem presents a consequence of Theorem 4.1, namely, that if the primal iterate sequence converges almost surely to x_*, and if at each iterate the conditions of Theorem 4.1 hold with probability at least p—which include that $\bar{\rho}(X_k, Y_k, Z_k) > 0$ and the noise in the function and derivative values is sufficiently small relative to $\bar{\rho}(X_k, Y_k, Z_k)$—then with probability one there exists an iteration threshold such that in all iterations beyond the threshold the optimal active set will be identified correctly with probability at least p. It should be emphasized that, as mentioned in §4.1, the conditions of Theorem 4.1 may require that the noise is smaller near x_*; in particular, the conditions may require that the bound on the noise is in proportion to the distance from the primal iterate to x_*. Hence, practically speaking, to ensure the probability lower bound required in the theorem, one may require noise reduction (if possible) in the function and derivative estimates.

Theorem 4.3. *Suppose that, with respect to* (2), *the conditions of Theorem 4.1 hold. In addition, with* $\hat{\epsilon} \in \mathbb{R}_{>0}$ *as defined in Theorem 4.1, suppose that*

$$\mathbb{P}[(23) \text{ holds with } x = X_k$$
$$\wedge \, \bar{\rho}(X_k, Y_k, Z_k) > 0 \text{ where } (Y_k, Z_k) \text{ is computed to solve } (30) \text{ with } x = X_k$$
$$\wedge \, \tilde{\epsilon}_x \text{ is as small as required in Theorem 4.1}$$
$$\mid \mathcal{F}_k \wedge \|X_k - x_*\| \le \hat{\epsilon}] \ge p \text{ for all } k \in \mathbb{N}$$

for some universal constant $p \in (0, 1]$. *Then, for any parameters* $\beta \in \mathbb{R}_{>0}$ *and* $\sigma \in (0, 1)$, *if* $\{X_k\}$ *converges almost surely to* x_*, *it holds with probability one that there exists an index* $K \in \mathbb{N}$ *such that for all* $k \in \mathbb{N}$ *with* $k \ge K$ *it holds with probability at least* p *that the computed solution* (Y_k, Z_k) *of* (30) *with* $x = X_k$ *yields an active-set estimate defined by* (32) *satisfying* $\widetilde{\mathcal{A}}_{\mathrm{LP}}(Y_k, Z_k; X_k) = \mathcal{A}_*$. *Thus, it holds with probability one that* $\mathcal{A}_*$ *will be identified infinitely often by the algorithm.*

Proof. Under the stated conditions, it holds with probability one that there exists $K \in \mathbb{N}$ such that for all $k \in \mathbb{N}$ with $k \ge K$ the primal iterate has $\|X_k - x_*\| \le \epsilon_x$. Consequently, the desired conclusion follows along with the result of Theorem 4.1. $\qquad\square$

Our second setting, for identification through a primal step as presented in §4.2, requires that the LICQ and strict complementarity hold at a strict local minimizer x_*. These conditions may be seen as more restrictive than only the MFCQ and second-order condition, as in the previous setting. However, this setting does not necessarily require noise reduction in the function and derivative value estimates. In particular, once an iterate sequence enters into a neighborhood of the local minimizer, accurate active-set identification is achieved whenever the noise in the function and derivative values is sufficiently small. This is captured in the following theorem, where we recall that for all $k \in \mathbb{N}$ the active-set estimate is obtained by solving the QP (33), which yields the subproblem solution we denote as (D_k, R_k, S_k, T_k).

Theorem 4.4. *Suppose that, with respect to* (2), *the conditions of Theorem 4.2 hold. In addition, with* $(\epsilon_x, \epsilon_e, \epsilon_c, \epsilon_{\nabla f}, \epsilon_{\nabla c}, \underline{\nu})$ *as defined in Theorem 4.2, suppose*

$$\mathbb{P}[(23) \text{ holds with } x = X_k \mid \mathcal{F}_k \wedge \|X_k - x_*\| \le \epsilon_x] \ge p \text{ for all } k \in \mathbb{N}$$

for some universal constant $p \in (0, 1]$. *Then, for any parameter* $\theta \in \mathbb{R}_{>0}$ *and* $\nu \ge \underline{\nu}$, *if* $\{X_k\}$ *converges almost surely to* x_*, *it holds with probability one that there exists an index* $K \in \mathbb{N}$ *such that for all* $k \in \mathbb{N}$

with $k \geq K$ it holds with probability at least p that the solution (D_k, R_k, S_k, T_k) of (33) with $x = X_k$ is unique and the active-set estimate defined by (38) yields $\widetilde{\mathcal{A}}_{\mathrm{QP}}(D_k; X_k) = \mathcal{A}_*$. Thus, it holds with probability one that $\mathcal{A}_*$ will be identified infinitely often by the algorithm.

Proof. Under the stated conditions, it holds with probability one that there exists $K \in \mathbb{N}$ such that for all $k \in \mathbb{N}$ with $k \geq K$ the primal iterate has $\|X_k - x_*\| \leq \epsilon_x$. Consequently, the desired conclusion follows along with the result of Theorem 4.2. $\qquad\square$

5   Numerical Results

The purpose of our numerical experiments is to demonstrate the behavior of the active-set identification techniques analyzed in §4 in the context of a few test problems. We present the results of two sets of experiments. Our first experiment focuses on two related two-dimensional test problems that allow us to provide clear visualizations of active-set identification behavior in both deterministic and noisy settings. Our second experiment focuses on behavior in the context of a larger-dimensional neural-network training problem. We emphasize that any number of problems could be used to demonstrate the behavior of the active-set identification methods from §4, each potentially leading to interesting conclusions about the two techniques in general, the parameter choices for them, and so on. We chose our first experiment since two-dimensional problems allow for nice visualizations, and we chose our second experiment since it involves the problem of training of a neural network, which is a problem type of immense interest in modern optimization.

5.1   Visualization with Two-Dimensional Problems

Let us consider a pair of two-dimensional problems of the form

$$\min_{x \in \mathbb{R}^2} \ f(x) \ \text{ s.t. } \ c_1(x) := x_1^2 - x_2 \leq 0 \text{ and } c_2(x) := x_1^2 + x_2 - \tfrac{1}{2} \leq 0, \tag{39}$$

where c_i and x_i for each $i \in \{1, 2\}$ denote the ith component of the constraint function $c : \mathbb{R}^2 \to \mathbb{R}^2$ and the ith component of the vector $x \in \mathbb{R}^2$, respectively. (Since there are no equality constraints in (39), we denote the constraint function simply as c.) The feasible region is thus the region in $\mathbb{R}^2$ contained between two parabolas, one upward-facing and one downward-facing. Let us also consider two objective functions $f_1 : \mathbb{R}^2 \to \mathbb{R}$ and $f_2 : \mathbb{R}^2 \to \mathbb{R}$ that are defined respectively as

$$f_1(x) = (x_1 + \tfrac{1}{2})^2 + 4(x_2 - \tfrac{1}{2})^2$$
$$\text{and } \ f_2(x) = 4(x_1 + \tfrac{3}{5})^2 + (x_2 - \tfrac{1}{4})^2.$$

The problems to minimize f_1 and minimize f_2, each subject to $c(x) \leq 0$, have convex feasible regions and strongly convex objectives. Therefore, each problem has a unique solution, call it x_* in each case. It is straightforward to confirm that the problem to minimize f_1 subject to $c(x) \leq 0$ has one active constraint at x_* such that $\mathcal{A}(x_*) = \{2\}$, and at x_* the LICQ and strict complementarity hold. On the other hand, for the problem to minimize f_2 subject to $c(x) \leq 0$, both constraints are active at x_*, i.e., $\mathcal{A}(x_*) = \{1, 2\}$, and the LICQ and strict complementarity hold at x_*.

For each problem, we generated points in a grid around x_*. For each point x in the grid, we solved various instances of the LP-LPEC subproblem (30) (with $M \leftarrow 10^8$, $\beta \leftarrow 0.7071$,[1] and $\sigma \leftarrow 0.7$) and the QP subproblem (33) (with $\nu \leftarrow 100$ and $\theta \leftarrow 5$) in order to determine active-set estimates, as described in §4.1 and §4.2, respectively. For each problem, we first considered the deterministic setting, denoted as $\epsilon = 0$ (for zero noise). For each point in the grid, we determined whether the active-set estimate exactly matched $\mathcal{A}(x_*)$ or not, yielding a value of 1 or 0, respectively. Then, for two noise levels, denoted by $\epsilon \in \{10^{-2}, 10^{-1}\}$,

---

[1]This value of β was determined by choosing a point $\bar{x}$ randomly within the grid and evaluating $\beta \leftarrow 1/\max\{\|\nabla f(\bar{x})\|_\infty, \|c(\bar{x})\|_\infty, \|\nabla c(\bar{x})\|_\infty\}$ in order to account for the scaling of the problem functions. We conjecture that this may be an effective choice in practice.

respectively, we solved 8 instances of each subproblem with noise added to the function and derivative values, then determined the fraction of times that the active-set estimate exactly matched $\mathcal{A}(x_*)$. This yielded a fractional value between 0 and 1 for each point, where a value of 1 indicates that the active set was always identified correctly. With respect to the noise, for each function value or component of a function or derivative value, we added a number generated from a uniform distribution over the interval $[-\epsilon, \epsilon]$.

The results associated with objective functions f_1 and f_2 are presented in Figures 1 and 2, respectively. In each case, one finds that for noise level 10^{-2}, the active-set identification at points near x_* is relatively robust to noise in the function and derivative values, as can be guaranteed by our theoretical results. On the other hand, for noise level 10^{-1}, the active-set identification is less reliable, failing to identify $\mathcal{A}(x_*)$ in some cases. This is also consistent with our theoretical guarantees, since our theory requires that the noise is sufficiently small in order to guarantee accurate active-set identification. Generally speaking, at least for these test problems, the LP-LPEC approach identifies the active set correctly less often than the QP approach when only one constraint is active at x_*, and identifies the active set correctly more often when both constraints are active, at least with our choices of the parameters β, σ, ν, and θ. Naturally, the behavior of the methods can vary with these parameter choices, so these comparisons do not necessarily hold in general.

A notable observation about the heatmaps in Figures 1 and 2 is that, when the noise level is nonzero, one should only assume that the errors in the function and derivative values are sufficiently small with some probability such that active-set identification can be performed accurately, even within a small neighborhood of x_*. This is an empirical demonstration of the comments that we made in §4.3.

5.2 Neural Network Training Problem

Now let us consider a constrained neural network training problem. The problem that we consider is one of supervised learning of a prediction function that aims to predict, based on feature data for a given patient, whether or not the patient is at high risk for developing heart disease. The constraints are designed to ensure that the prediction function yields a higher probability of heart disease when the patient has a known high-risk factor. We emphasize that the constraints that we impose are for demonstration purposes only, and may not be ones that lead to better predictions in reality for this type of problem. That being said, the problem that we consider in this section might represent the type of constrained supervised learning problem that could potentially lead to better and more explainable prediction functions in the context of medical diagnoses, as opposed to prediction functions obtained purely through empirical loss minimization.

Let $\{(a_i, b_i)\}_{i=1}^N$ be a set of training data, where for each $i \in [N]$ the value $a_i \in \mathbb{R}^{n_f}$ is an n_f-vector of feature data for a given patient (including age, sex, and cholesterol level) and the value $b_i \in \{0, 1\}$ indicates whether the patient eventually developed heart disease (indicated by $b_i = 1$) or not (indicated by $b_i = 0$). In particular, let us suppose that for each $i \in [N]$ the jth component of a_i, namely, $a_{i,j}$, indicates the patient's cholesterol level. For the purposes of devising a tractable training problem, let us further suppose that the N data points are separated into K groups, where the indices in $\mathcal{N}_1$ correspond to group 1, the indices in $\mathcal{N}_2$ correspond to group 2, etc. (In this manner, one has that $\{\mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_K\}$ are disjoint sets and $|\mathcal{N}_1| + |\mathcal{N}_2| + \cdots |\mathcal{N}_K| = N$. In practice, the groups might be determined randomly or according to certain features, such as age and/or sex.) Given a prediction function p parameterized by a vector $x \in \mathbb{R}^n$ of trainable parameters—in our setting, defined by a neural network—and a loss function ℓ, our problem of interest can be written as

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^N \ell(p(x, a_i), b_i) \text{ s.t. } \max_{i \in \mathcal{N}_k} \left\{ -\frac{\partial p}{\partial a_{\cdot, j}}(x, a_i) \right\} \leq 0 \text{ for all } k \in [K], \tag{40}$$

where $\partial p / \partial a_{\cdot, j}$ indicates the partial derivative of p with respect to the jth element of the input of the neural network, which in our setting corresponds to cholesterol. Here, the objective function is a typical one for empirical loss minimization that aims to maximize the accuracy of predictions on unseen data. The constraints, on the other hand, aim to ensure that—all else being equal—an increase in a patient's cholesterol would increase (or at least not decrease) the prediction that the patient will develop heart disease. Rather
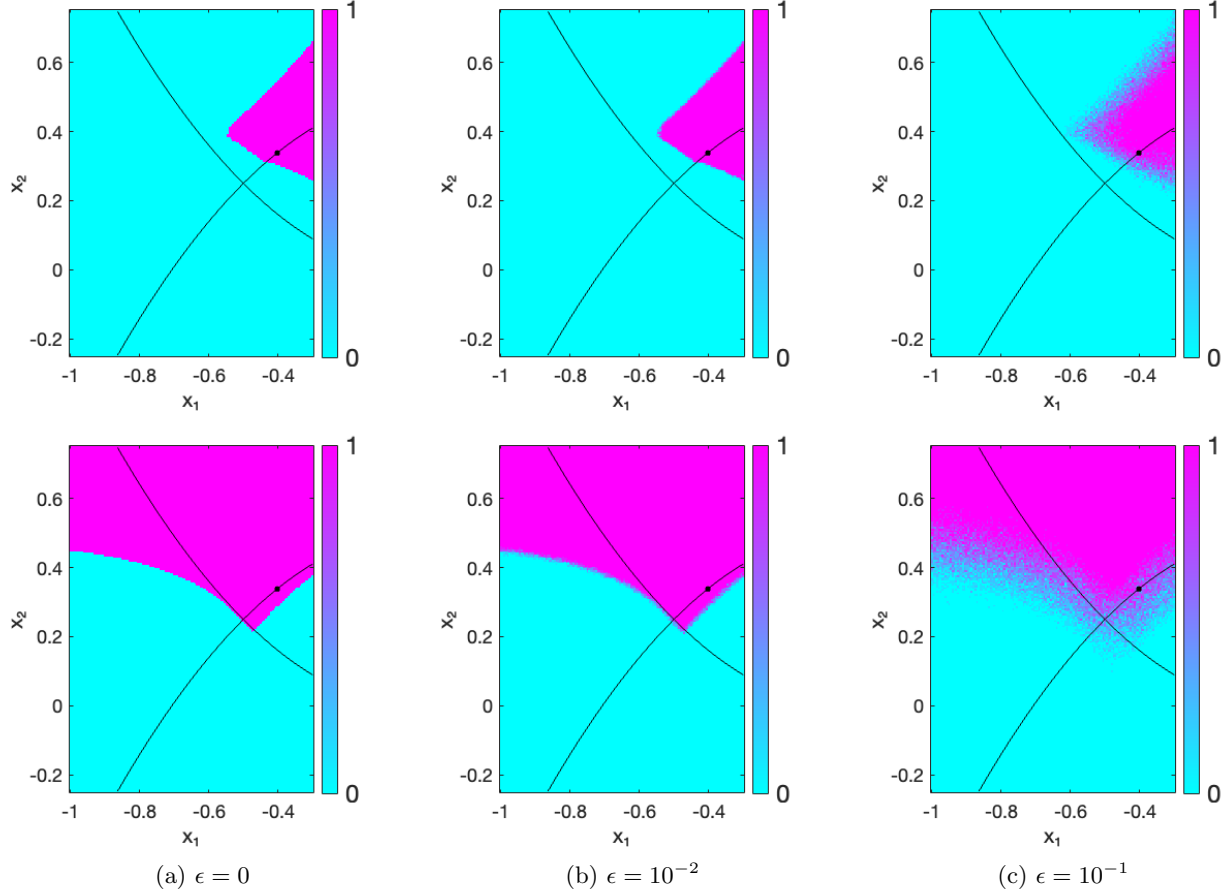
Figure 1: Heatmaps for the problem to minimize f_1 subject to $c(x) \leq 0$ for noise levels $\epsilon = 0$ (deterministic), 10^{-2}, and 10^{-1}, respectively. The top row corresponds to the LP-LPEC approach based on computing Lagrange multiplier estimates; the bottom row corresponds to the QP approach based on computing a primal step. The value of the heatmap at each point indicates the fraction of times that the active set $\mathcal{A}(x_*)$ was identified correctly by solving a subproblem at the point. (The unique optimal solution x_* is indicated by a black dot.) In the case of positive noise levels, the value was determined as a fraction over 8 randomly generated subproblem instances.
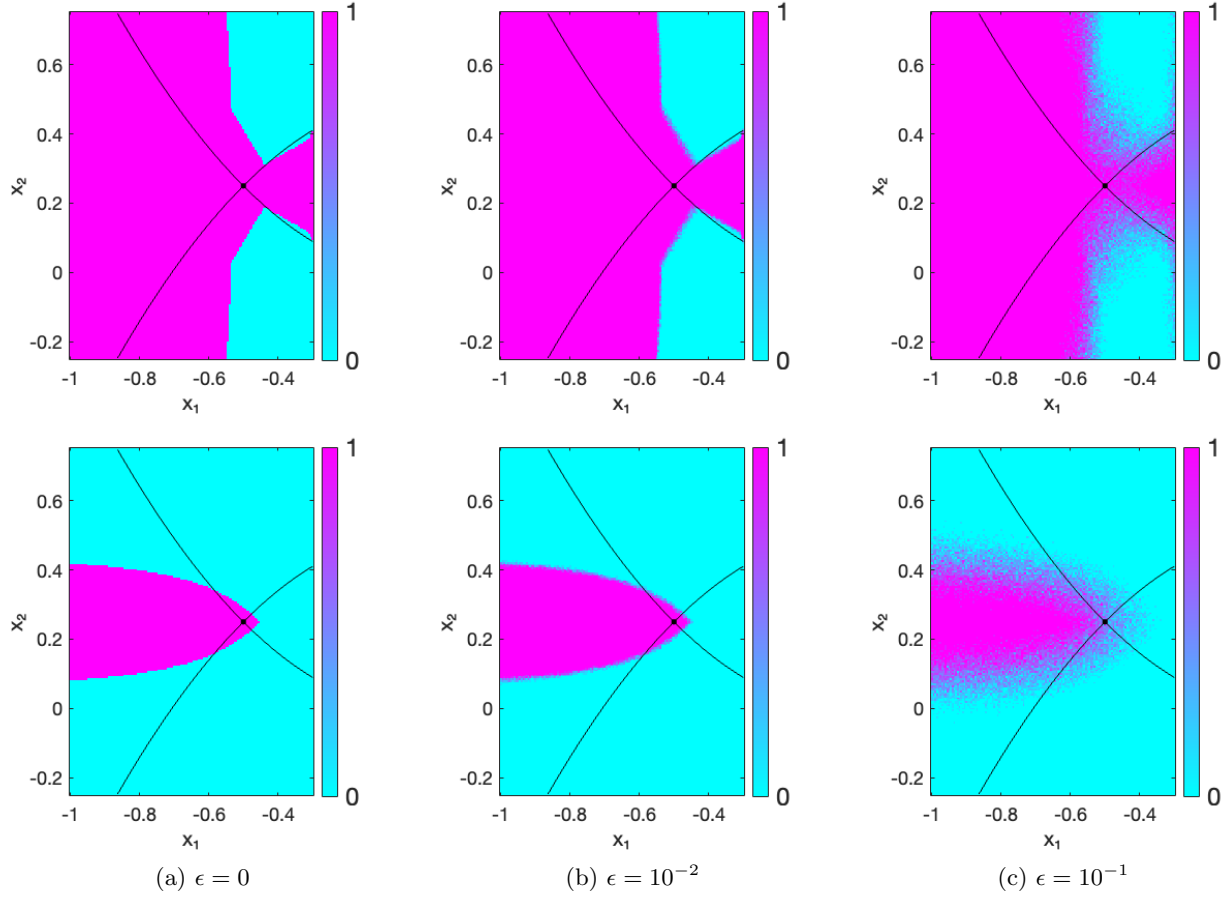
Figure 2: Heatmaps for the problem to minimize f_2 subject to $c(x) \leq 0$. The details are the same as stated in the caption for Figure 1.

than impose a constraint for each data point, which may lead to an intractable problem in large-scale settings, a constraint is imposed for each group in the data set. It should be noted that the maximum function is nonsmooth, and in practice one could impose a smooth constraint through a smoothed approximation of the maximum function. However, for our experiments we did not smooth the functions since (a) we were able to solve the problem to sufficient accuracy despite nonsmoothness in the constraint functions and (b) it is often the case in modern practice that algorithms employed for neural network training ignore nonsmoothness due to maximum functions, as with rectified linear unit (ReLU) activation.

We acquired a data set of 1025 data points with each point having $n_f = 13$ features [23]. For the prediction function p, we used a neural network with one hidden layer with 6 nodes and tanh activation. Including both neural network edge weights and bias terms as trainable parameters, this led to a problem with 91 variables. For the loss function we used the log-loss function. The data was split initially into training and testing data according to cholesterol level; in particular, we sorted the data points by cholesterol and used the top 20% according to cholesterol as the testing data. In this manner, training was conducted on those with lower cholesterol and testing was conducted on those with higher cholesterol. As an interesting side-note to our study of the behavior of active-set identification strategies, the idea here was to test the effect of training without the highest-cholesterol individuals in order to see if the presence of the constraints in (40) was able to improve the predictions on the testing data.

The training data was split into 10 groups in order to define a set of 10 constraints. We began our experimentation by employing an implementation of an L-BFGS method with a backtracking Armijo line search in order to minimize the objective of (40), without any constraints. This resulted in an approximate stationary point—we used a threshold of 10^{-4} for the norm of the gradient—with a maximum constraint function value of `1.1e+03`; i.e., it resulted in a point that was infeasible for the constraints in (40). Corresponding to this point, the training accuracy was determined to be 92.4% while the testing accuracy was determined to be 78.5%.

We then employed the same L-BFGS method to minimize a quadratic penalty function corresponding to (40) with a penalty parameter of `1e+02`; see [28] for further details on penalty methods. With a tolerance of 10^{-4} for the norm of the gradient, this resulted in a feasible point for (40). Corresponding to this point, the training accuracy was determined to be 89.0% while the testing accuracy was determined to be 83.9%. This demonstrates a case in which the presence of the constraints during the training process was able to improve testing accuracy by over 5%.

In terms of active-set identification, we considered the final iterate of the algorithm to minimize the quadratic penalty function as a solution estimate. At this point, in order to determine the estimated optimal active set, we employed the active-set identification techniques from §4 to solve both an LP-LPEC subproblem and a QP subproblem at x_*. Here and throughout the remainder of these experiments, we employed the parameters $M \leftarrow 10^8$, $\beta \leftarrow 10^{-12}$, $\sigma \leftarrow 0.7$, $\nu \leftarrow 10^2$, and $\theta \leftarrow 5$.) These both led to the same active set $\mathcal{A}(x_*) = [10]$, i.e., all constraints being active.

We studied the behavior of the active-set identification techniques from §4 by following the sequence of iterates in the L-BFGS method for minimizing a quadratic penalty function, as described previously. At each iterate, Figure 3 shows the number of correctly identified active constraints by the LP-LPEC and QP approaches when there *is no noise* in the function and derivative values used to define the subproblems for each approach. One finds that the LP-LPEC approach identifies the correct active set even in early iterations. We found this to be due to the fact that the iterates are consistently points where all constraints are violated. Combined with the fact that the right-hand side of the inequality in (32) is always negative, this demonstrates a situation in which the LP-LPEC approach is lucky always to identify violated constraints as being active at the optimal solution. The QP approach, on the other hand, does not identify the optimal active set until the iterates are closer to x_*. That said, it should be emphasized that both the LP-LPEC and QP approaches correctly identify the optimal active set in all later iterations.

To observe the effect of noise in the function and derivative values on active-set identification, we ran the algorithm with exact function and derivative values, but at each iterate we generated 10 different sets of noisy objective and constraint function and derivative values for active-set identification by sampling the data defining the objective and constraints. In particular, at each iterate, we used a random batch of data
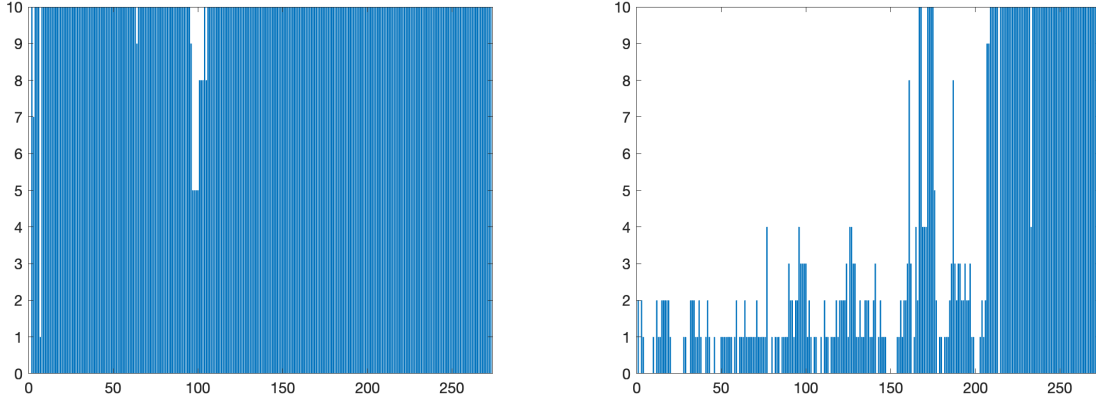
Figure 3: Over the iterates of an algorithm to solve (40), the number of correctly identified elements of the optimal active set by the LP-LPEC approach (left) and the QP approach (right) when there *is no noise* in the function and derivative values.

points with batch size equal to 32 in order to evaluate the function and derivative values employed in the subproblems for active-set identification. (We ran the code with other similar batch sizes and the results were qualitatively similar to those presented here.) The results are shown in Figure 4, where for each iteration index we plot the average number of correctly identified active constraints over the 10 different batchings that were generated. One finds that the LP-LPEC approach does a poorer job (compared to the results in Figure 3) of identifying the optimal active set in earlier iterations, but it still identifies the optimal active set correctly in later iterations. The QP approach, on the other hand, happens to do a better job (compared to the results in Figure 3) of identifying the optimal active set in later iterations. We found this to be due to the fact that the noisy function and derivative values more often caused the QP solution to have linearized constraint values to be positive (see (38)), which caused more constraints to be predicted to be active.
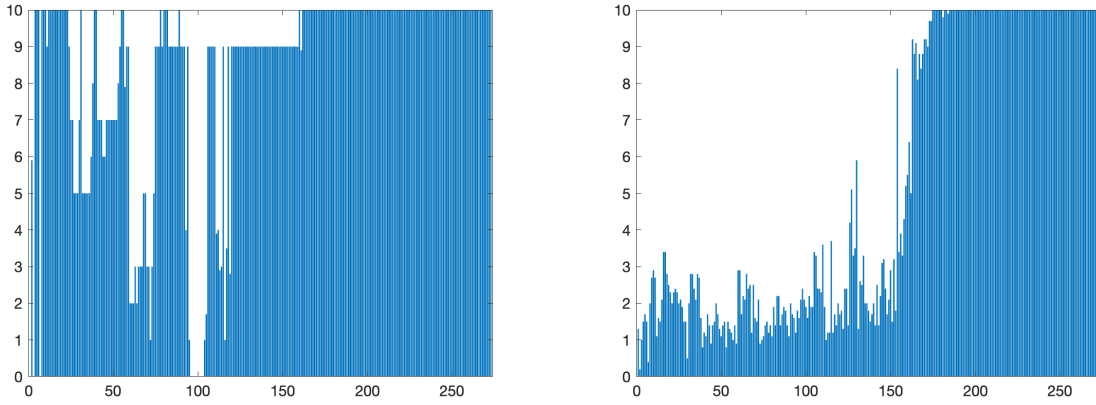

Figure 4: Over the iterates of an algorithm to solve (40), the number of correctly identified elements of the optimal active set by the LP-LPEC approach (left) and the QP approach (right) when there *is noise* in the function and derivative values.

As for the experiments in §5.1, the behavior of the methods can vary with different problems and parameter choices, so these specific comparisons do not necessarily hold in general. That said, the important

24

thing is that we observe accurate active-set identification in later iterations for both the LP-LPEC and QP approaches.

6 Conclusion

We have shown that state-of-the-art techniques for active-set identification when solving constrained continuous optimization problems can also be employed in contexts when function and derivative values are corrupted by noise. In particular, our theoretical guarantees show that when a point x is sufficiently close to a local minimizer x_* and the noise in the function and derivative values is sufficiently small, one can predict the active set at x_* correctly by solving two different types of subproblems defined at the point x. We have also shown how such techniques can be employed to offer active-set identification guarantees in the context of a stochastic algorithm for solving constrained continuous optimization problems. We have also provided the results of numerical experiments that demonstrate the behaviors of these techniques on a pair of illustrative problems and a neural-network training problem.

Interesting follow-up investigations to our work in this paper include studying how active-set identification techniques can be employed within noisy and stochastic algorithms for solving constrained continuous optimization problems both for computing certain types of primal search directions and for more accurate Lagrange multiplier estimation. For example, one may consider whether, despite noise in problem function and derivative values, one can design algorithms with convergence guarantees that operate by (a) estimating the active set, then (b) computing each search direction with respect to the estimated active set, as has been done in the deterministic setting [10, 9]. It would also be interesting to investigate how, e.g., the theoretical guarantees for Lagrange multiplier estimation presented in [12] (for the case of equality constraints only) can be extended to the setting of an algorithm for solving inequality-constrained problems. This may involve employing the active-set identification techniques and theoretical guarantees from this paper in order to provide guarantees for optimal Lagrange multiplier estimation.

References

[1] Albert S. Berahas, Richard H. Byrd, and Jorge Nocedal. Derivative-free optimization of noisy functions via quasi-Newton methods. *SIAM Journal on Optimization*, 29(2):965–993, 2019.

[2] Albert S. Berahas, Frank E. Curtis, Michael J. O'Neill, and Daniel P. Robinson. A stochastic sequential quadratic optimization algorithm for nonlinear equality constrained optimization with rank-deficient jacobians. *Mathematics of Operations Research*, 2023.

[3] Albert S. Berahas, Frank E. Curtis, Daniel P. Robinson, and Baoyu Zhou. Sequential quadratic optimization for nonlinear equality constrained stochastic optimization. *SIAM Journal on Optimization*, 31(2):1352–1379, 2021.

[4] Dimitri P. Bertsekas. *Convex Optimization Theory*. Athena Scientific, Belmont, Mass, 2009.

[5] Léon Bottou, Frank E. Curtis, and Jorge Nocedal. Optimization methods for large-scale machine learning. *SIAM Review*, 60(2):223–311, 2018.

[6] Richard L. Burden and John Douglas Faires. *Numerical Analysis*. Thomson Brooks/Cole, Belmont, CA, 8th edition, 2006.

[7] James V. Burke. On the identification of active constraints II: The nonconvex case. *SIAM Journal on Numerical Analysis*, 27(4):1081–1102, 1990.

[8] James V. Burke and Jorge J. Moré. On the identification of active constraints. *SIAM Journal on Numerical Analysis*, 25(5):1197–1211, 1988.

[9] Richard H. Byrd, Nicholas I. M. Gould, Jorge Nocedal, and Richard A. Waltz. On the convergence of successive linear-quadratic programming algorithms. *SIAM Journal on Optimization*, 16(2):471–489, 2005.

[10] Richard H. Byrd, Nicholas I.M. Gould, Jorge Nocedal, and Richard A. Waltz. An algorithm for nonlinear optimization using linear programming and equality constrained subproblems. *Mathematical Programming*, 100(1):27–48, 2003.

[11] Frank E. Curtis, Shima Dezfulian, and Andreas Wächter. An interior-point algorithm for continuous nonlinearly constrained optimization with noisy function and derivative evaluations. *arXiv preprint arXiv:2502.11302*, 2025.

[12] Frank E. Curtis, Xin Jiang, and Qi Wang. Almost-sure convergence of iterates and multipliers in stochastic sequential quadratic optimization. *Journal of Optimization Theory and Applications*, 2024.

[13] Frank E. Curtis, Michael J. O'Neill, and Daniel P. Robinson. Worst-case complexity of an SQP method for nonlinear equality constrained stochastic optimization. *Mathematical Programming*, 205:431–483, 2024.

[14] Frank E. Curtis, Daniel P. Robinson, and Baoyu Zhou. Sequential quadratic optimization for stochastic optimization with deterministic nonlinear inequality and equality constraints. *SIAM Journal on Optimization*, 34(4):3592–3622, 2024.

[15] Frank E. Curtis, Daniel P. Robinson, and Baoyu Zhou. A stochastic inexact sequential quadratic optimization algorithm for nonlinear equality-constrained optimization. *INFORMS Journal on Optimization*, 2024.

[16] Frank E. Curtis and Katya Scheinberg. Optimization methods for supervised machine learning: From linear models to deep learning. In *INFORMS Tutorials in Operations Research*, chapter 5, pages 89–114. Institute for Operations Research and the Management Sciences (INFORMS), 2017.

[17] Shima Dezfulian and Andreas Wächter. On the convergence of interior-point methods for bound-constrained nonlinear optimization problems with noise. *arXiv preprint arXiv:2405.11400*, 2024.

[18] Francisco Facchinei, Andreas Fischer, and Christian Kanzow. On the accurate identification of active constraints. *SIAM Journal on Optimization*, 9(1):14–32, 1998.

[19] Yuchen Fang, Sen Na, Michael W. Mahoney, and Mladen Kolar. Fully stochastic trust-region sequential quadratic programming for equality-constrained optimization problems. *SIAM Journal on Optimization*, 34(2):2007–2037, 2024.

[20] William W. Hager and M. Seetharama Gowda. Stability in the presence of degeneracy and error estimation. *Mathematical Programming*, 85(1):181–192, May 1999.

[21] S.-P. Han and Olvi L. Mangasarian. Exact penalty functions in nonlinear programming. *Mathematical Programming*, 17(1):251–269, December 1979.

[22] Warren Hare and Adrian S. Lewis. Identifying active constraints via partial smoothness and prox-regularity. *Journal of Convex Analysis*, 11:251–266, 2004.

[23] Andras Janosi, William Steinbrunn, Matthias Pfisterer, and Robert Detrano. Heart Disease. UCI Machine Learning Repository, 1989. DOI: https://doi.org/10.24432/C52P4X.

[24] Adrian S. Lewis. Active sets, nonsmoothness, and sensitivity. *SIAM Journal on Optimization*, 13(3):702–725, 2002.

[25] Olvi L. Mangasarian and Stan Fromovitz. The Fritz John Necessary Optimality Conditions in the Presence of Equality and Inequality Constraints. *Journal of Mathematical Analysis and Applications*, 17(1):37–47, 1967.

[26] Sen Na, Mihai Anitescu, and Mladen Kolar. An adaptive stochastic sequential quadratic programming with differentiable exact augmented lagrangians. *Mathematical Programming*, pages 1–71, jun 2022.

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

[28] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer New York, 2006.

[29] Christina Oberlin and Stephen J. Wright. Active Set Identification in Nonlinear Programming. *SIAM Journal on Optimization*, 17(2):577, 2006.

[30] James M. Ortega and Werner Rheinboldt. *Numerical Analysis: A Second Course*. Elsevier Science, Burlington, 2014.

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

[32] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

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

[34] Stephen J. Wright. Identifiable surfaces in constrained optimization. *SIAM Journal on Control and Optimization*, 31(4):1063–1079, 1993.

[35] Stephen J. Wright. Modifying SQP for degenerate problems. *SIAM Journal on Optimization*, 13(2):470–497, 2002.

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