SIAM J. OPTIM.
Vol. 12, No. 3, pp. 782–810

WARM-START STRATEGIES IN INTERIOR-POINT METHODS FOR LINEAR PROGRAMMING*

E. ALPER YILDIRIM† AND STEPHEN J. WRIGHT‡

Abstract. We study the situation in which, having solved a linear program with an interior-point method, we are presented with a new problem instance whose data is slightly perturbed from the original. We describe strategies for recovering a "warm-start" point for the perturbed problem instance from the iterates of the original problem instance. We obtain worst-case estimates of the number of iterations required to converge to a solution of the perturbed instance from the warm-start points, showing that these estimates depend on the size of the perturbation and on the conditioning and other properties of the problem instances.

Key words. warm-start, reoptimization, interior-point methods, linear programming

AMS subject classifications. 90C51, 90C05

PII. S1052623400369235

1. Introduction.

This paper describes and analyzes warm-start strategies for interior-point methods applied to linear programming (LP) problems. We consider the situation in which one linear program, the "original instance," has been solved by an interior-point method, and we are then presented with a new problem of the same dimensions, the "perturbed instance," in which the data is slightly different. Interior-point iterates for the original instance are used to obtain warm-start points for the perturbed instance, so that when an interior-point method is started from this point, it finds the solution in fewer iterations than if no prior information were available. Although our results are theoretical, the strategies proposed here can be applied to practical situations, an aspect that is the subject of ongoing study.

The situation we have outlined arises, for instance, when linearization methods are used to solve nonlinear problems, as in the sequential LP algorithm. (One extension of this work that we plan to investigate is the extension to convex quadratic programs, which would be relevant to the solution of subproblems in sequential quadratic programming algorithms.) Our situation is different from the one considered by Gondzio [5], who deals with the case in which the number of variables or constraints in the problem is increased and the dimensions of the problem data objects are correspondingly expanded. The latter situation arises in solving subproblems arising from cutting-plane or column-generation algorithms, for example. The reader is also referred to Mitchell and Borchers [8] and Gondzio and Vial [6] for consideration of warm-start strategies in a cutting-plane scheme.

Freund [4] develops and analyzes a potential reduction algorithm from an infeasible warm-start, in which the iterate satisfies the equality constraints but is allowed to violate nonnegativity.

*Received by the editors March 7, 2000; accepted for publication (in revised form) August 18, 2001; published electronically February 27, 2002. This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing Research, U.S. Department of Energy, under contract W-31-109-Eng-38, and by the National Science Foundation under grants 0082065 and 0086559.

http://www.siam.org/journals/siopt/12-3/36923.html

†Department of Applied Mathematics and Statistics, State University of New York at Stony Brook, 1-107 Math Tower, Stony Brook, NY 11794–3600 (yildirim@ams.sunysb.edu).

‡Computer Sciences Department, University of Wisconsin-Madison, 1210 West Dayton Street, Madison, WI 53706 (swright@cs.wisc.edu).

For our analysis, we use the tools developed by Nunez and Freund [10], which in turn are based on the work of Renegar [11, 12, 13, 14] on the conditioning of linear programs and the complexity of algorithms for solving them. We also use standard complexity analysis techniques from the interior-point literature for estimating the number of iterations required to solve a linear program to given accuracy.

We start in section 2 with an outline of notation and a restatement and slight generalization of the main result from Nunez and Freund [10]. Section 3 outlines the warm-start strategies that we analyze in the paper and describes how they can be used to obtain reduced complexity estimates for interior-point methods. In section 4 we consider a warm-start technique in which a least-squares change is applied to a feasible interior-point iterate for the original instance to make it satisfy the constraints for the perturbed instance. We analyze this technique for central path neighborhoods based on both the Euclidean norm and the ∞ norm, deriving in each case a worst-case estimate for the number of iterations required by an interior-point method to converge to an approximate solution of the perturbed instance. In section 5 we study the technique of applying one iteration of Newton's method to a system of equations that is used to recover a strictly feasible point for the perturbed instance from a feasible iterate for the original instance. Section 6 discusses the relationship between the two warm-start strategies and the weighted versions of least-squares corrections. A small example is used to illustrate the behavior of different correction strategies. Finally, we conclude the paper with some discussions in section 7.

2. Preliminaries: Conditioning of LPs, central path neighborhoods, bounds on feasible points.

We consider the LP in the following standard form:

$$\text{(P)} \qquad \min_x \ c^T x \quad \text{subject to (s.t.)} \ \ Ax = b, \ x \geq 0,$$

where $A \in R^{m \times n}$, $b \in R^m$, and $c \in R^n$ are given and $x \in R^n$. The associated dual LP is given by the following:

$$\text{(D)} \qquad \max_{y,s} \ b^T y \quad \text{s.t.} \ \ A^T y + s = c, \ s \geq 0,$$

where $y \in R^m$ and $s \in R^n$. We borrow the notation of Nunez and Freund [10], denoting by d the data triplet (A, b, c) that defines the problems (P) and (D). We define the norm of d (differently from Nunez and Freund) as the maximum of the Euclidean norms of the three data components:

$$\|d\| \stackrel{\text{def}}{=} \max(\|A\|_2, \|b\|_2, \|c\|_2). \tag{2.1}$$

We use the norm notation $\| \cdot \|$ on a vector or matrix to denote the Euclidean norm and the operator norm it induces, respectively, unless explicitly indicated otherwise.

We use $\mathcal{F}$ to denote the space of strictly feasible data instances, that is,

$$\mathcal{F} = \{(A, b, c) : \exists \ x, y, s \text{ with } (x, s) > 0 \text{ such that } Ax = b, \ A^T y + s = c\}.$$

The complement of $\mathcal{F}$, denoted by $\mathcal{F}^C$, consists of data instances d for which either (P) or (D) does not have any strictly feasible solutions. The (shared) boundary of $\mathcal{F}$ and $\mathcal{F}^C$ is given by

$$\mathcal{B} = \text{cl}(\mathcal{F}) \cap \text{cl}(\mathcal{F}^C),$$

where $\text{cl}(\cdot)$ denotes the closure of a set. Since $(0, 0, 0) \in \mathcal{B}$, we have that $\mathcal{B} \neq \emptyset$. The data instances $d \in \mathcal{B}$ will be called *ill-posed data instances*, since arbitrarily small perturbations in the data d can result in data instances in either $\mathcal{F}$ or $\mathcal{F}^C$. The *distance to ill-posedness* is defined as

$$\rho(d) = \inf\{\|\Delta d\| \, : \, d + \Delta d \in \mathcal{B}\}, \tag{2.2}$$

where we use the norm (2.1) to define $\|\Delta d\|$. The condition number of a feasible problem instance d is defined as

$$\mathcal{C}(d) \stackrel{\text{def}}{=} \frac{\|d\|}{\rho(d)} \quad (\text{with } \mathcal{C}(d) \stackrel{\text{def}}{=} \infty \text{ when } \rho(d) = 0). \tag{2.3}$$

Since the perturbation $\Delta d = -d$ certainly has $d + \Delta d = 0 \in \mathcal{B}$, we have that $\rho(d) \leq \|d\|$ and therefore $\mathcal{C}(d) \geq 1$. Note, too, that $\mathcal{C}(d)$ is invariant under a nonnegative multiplicative scaling of the data d; that is, $\mathcal{C}(\beta d) = \mathcal{C}(d)$ for all $\beta > 0$.

Robinson [15] and Ashmanov [1] showed that a data instance $d \in \mathcal{F}$ satisfies $\rho(d) > 0$ (that is, d lies in the interior of $\mathcal{F}$) if and only if A has full row rank. For such d, another useful bound on $\rho(d)$ is provided by the minimum singular value of A. If we write the singular value decomposition of A as

$$A = USV^T = \sum_{i=1}^{m} \sigma_i(A) u_i v_i^T,$$

where U and V are orthogonal and $S = \operatorname{diag}(\sigma_1(A), \sigma_2(A), \ldots, \sigma_m(A))$, with $\sigma_1(A) \geq \sigma_2 \geq \cdots \geq \sigma_m(A) > 0$ denoting the singular values of A, then the perturbation

$$\Delta A = -\sigma_m(A) u_m v_m^T$$

is such that $A + \Delta A$ is singular, and moreover $\|\Delta A\| = \sigma_m(A)$ due to the fact that the Euclidean norm of a rank-one matrix satisfies the property

$$\|\beta u v^T\|_2 = |\beta| \, \|u\|_2 \, \|v\|_2. \tag{2.4}$$

We conclude that

$$\rho(d) \leq \sigma_m(A). \tag{2.5}$$

It is well known that for such $d \in \operatorname{int}(\mathcal{F})$ the system given by

$$Ax = b, \tag{2.6a}$$
$$A^T y + s = c, \tag{2.6b}$$
$$XSe = \mu e, \tag{2.6c}$$
$$(x, s) > 0 \tag{2.6d}$$

has a unique solution for every $\mu > 0$, where e denotes the vector of ones in the appropriate dimension, and X and S are the diagonal matrices formed from the components of x and s, respectively. We denote the solutions of (2.6) by $(x(\mu), y(\mu), s(\mu))$ and use $\mathcal{P}$ to denote the *central path* traced out by these solutions for $\mu > 0$, that is,

$$\mathcal{P} \stackrel{\text{def}}{=} \{(x(\mu), y(\mu), s(\mu)) \, : \, \mu > 0\}\,. \tag{2.7}$$

Throughout this paper, we assume that the original data instance d lies in $\mathcal{F}$ and that $\rho(d) > 0$. In sections 4 and 5, we assume further that the original data instance d has been solved by a feasible path-following interior-point method. Such a method generates a sequence of iterates (x^k, y^k, s^k) that satisfy the relations (2.6a), (2.6b), and (2.6d) and for which the pairwise products $x_i^k s_i^k$, $i = 1, 2, \ldots, n$, are not too different from one another, in the sense of remaining within some well-defined "neighborhood" of the central path. The duality measure $(x^k)^T s^k$ is driven toward zero as $k \to \infty$, and search directions are obtained by applying a modified Newton's method to the nonlinear system formed by (2.6a), (2.6b), and (2.6c).

We now give some notation for feasible sets and central path neighborhoods associated with the particular problem instance $d = (A, b, c)$. Let $\mathcal{S}$ and $\mathcal{S}^0$ denote the set of feasible and strictly feasible primal-dual points, respectively; that is,

$$\begin{aligned} \mathcal{S} &= \{(x, y, s) \; : \; Ax = b, \; A^T y + s = c, \; (x, s) \geq 0\}, \\ \mathcal{S}^0 &= \{(x, y, s) \in \mathcal{S} \; : \; (x, s) > 0\}. \end{aligned}$$

(Note that $d \in \mathcal{F}$ if and only if $\mathcal{S}^0 \neq \emptyset$.) We refer to the central path neighborhoods most commonly used in interior-point methods as the *narrow* and *wide* neighborhoods. The narrow neighborhood denoted by $\mathcal{N}_2(\theta)$ is defined as

$$\mathcal{N}_2(\theta) = \{(x, y, s) \in \mathcal{S}^0 \; : \; \|XSe - (x^T s/n)e\|_2 \leq \theta(x^T s/n)\} \tag{2.8}$$

for $\theta \in [0, 1)$. The wide neighborhood, which is denoted by $\mathcal{N}_{-\infty}(\gamma)$, is given by

$$\mathcal{N}_{-\infty}(\gamma) = \{(x, y, s) \in \mathcal{S}^0 \; : \; x_i s_i \geq \gamma(x^T s/n) \; \forall \; i = 1, 2, \ldots, n\}, \tag{2.9}$$

where u_i denotes the ith component of the vector u, and the parameter γ lies in $(0, 1]$.

We typically use a bar to denote the corresponding quantities for the perturbed problem instance $d + \Delta d$. That is, we have

$$\begin{aligned} \bar{\mathcal{S}} &= \{(x, y, s) \; : \; (A + \Delta A)x = (b + \Delta b), \; (A + \Delta A)^T y + s = (c + \Delta c), \; (x, s) \geq 0\}, \\ \bar{\mathcal{S}}^o &= \{(x, y, s) \in \bar{\mathcal{S}} \; : \; (x, s) > 0\}, \end{aligned}$$

whereas

$$\bar{\mathcal{N}}_2(\theta) = \{(x, y, s) \in \bar{\mathcal{S}}^o \; : \; \|XSe - (x^T s/n)e\|_2 \leq \theta(x^T s/n)\}, \tag{2.10a}$$

$$\bar{\mathcal{N}}_{-\infty}(\gamma) = \{(x, y, s) \in \bar{\mathcal{S}}^o \; : \; x_i s_i \geq \gamma(x^T s/n) \; \forall \; i = 1, 2, \ldots, n\}. \tag{2.10b}$$

We associate a value of μ with each iterate $(x, y, s) \in \mathcal{S}$ (or $\bar{\mathcal{S}}$) by setting

$$\mu = x^T s/n. \tag{2.11}$$

We call this μ the *duality measure* of the point (x, y, s). When (x, y, s) is feasible, it is easy to show that the duality gap $c^T x - b^T y$ is equal to $n\mu$.

Finally, we state a modified version of Theorem 3.1 from Nunez and Freund [10], which uses our definition (2.1) of the norm of the data instance and takes note of the fact that the proof in [10] continues to hold when we consider strictly feasible points that do not lie exactly on the central path $\mathcal{P}$.

THEOREM 2.1. *If $d = (A, b, c) \in \mathcal{F}$ and $\rho(d) > 0$, then for any point (x, y, s) satisfying the conditions*

$$Ax = b, \quad A^T y + s = c, \quad (x, s) > 0, \tag{2.12}$$

the following bounds are satisfied:

$$\|x\| \le \mathcal{C}(d)(\mathcal{C}(d) + \mu n/\|d\|), \tag{2.13a}$$

$$\|y\| \le \mathcal{C}(d)(\mathcal{C}(d) + \mu n/\|d\|), \tag{2.13b}$$

$$\|s\| \le 2\|d\|\mathcal{C}(d)(\mathcal{C}(d) + \mu n/\|d\|), \tag{2.13c}$$

where we have defined μ *as in* (2.11).

The proof exactly follows the logic of the proof in [10, Theorem 3.1], but differs in many details because of our use of Euclidean norms on the matrices and vectors. For instance, where the original proof defines a perturbation $\Delta A = -be^T/\|x\|_1$, to obtain an infeasible data instance, we instead use $\Delta A = -bx^T/\|x\|_2^2$. We also use the observation (2.4) repeatedly.

3. Warm starts and reduced complexity. Before describing specific strategies for warm starts, we preview the nature of our later results and show how they can be used to obtain improved estimates of the complexity of interior-point methods that use these warm starts.

We start by recalling some elements of the complexity analysis of interior-point methods. These methods typically produce iterates (x^k, y^k, s^k) that lie within a neighborhood such as (2.8) or (2.9) and for which the duality measure μ_k (defined as in (2.11) by $\mu_k = (x^k)^T s^k/n$) decreases monotonically with k, according to a bound of the following form:

$$\mu_{k+1} \le \left(1 - \frac{\delta}{n^\tau}\right)\mu_k, \tag{3.1}$$

where δ and τ are positive constants that depend on the algorithm. Typically, τ is 0.5, 1, or 2, while δ depends on the parameters θ or γ that define the neighborhood (see (2.8) and (2.9)) and on various other algorithmic parameters. Given a starting point (x^0, y^0, s^0) with duality measure μ_0, the number of iterations required to satisfy the stopping criterion

$$\mu \le \epsilon\|d\| \tag{3.2}$$

(for some small positive ϵ) is bounded by

$$\frac{\log(\epsilon\|d\|) - \log\mu_0}{\log\left(1 - \delta/n^\tau\right)} = \mathcal{O}\left(n^\tau \log\frac{\mu_0}{\|d\|\epsilon}\right). \tag{3.3}$$

It follows from this bound that, provided we have

$$\frac{\mu_0}{\|d\|} = \mathcal{O}(1/\epsilon^\eta)$$

for some fixed $\eta > 0$—which can be guaranteed for small ϵ when we apply a cold-start procedure—the number of iterations required to achieve (3.2) is

$$\mathcal{O}(n^\tau|\log\epsilon|). \tag{3.4}$$

Our warm-start strategies aim to find a starting point for the *perturbed* instance that lies inside one of the neighborhoods (2.10) and for which the initial duality measure $\bar{\mu}_0$ is not too large. By applying (3.3) to the perturbed instance, we see that if $\bar{\mu}_0/\|d+\Delta d\|$ is less than 1, then the formal complexity of the method will be better than the general estimate (3.4).

Both warm-start strategies that we describe in subsequent sections proceed by taking a point (x, y, s) from a neighborhood such as (2.8), (2.9) for the original instance and calculating an adjustment $(\Delta x, \Delta y, \Delta s)$ based on the perturbation Δd to obtain a starting point for the perturbed instance. The strategies are simple; their computational cost is no greater than the cost of one interior-point iteration. They do not succeed in producing a valid starting point unless the point (x, y, s) from the original problem has a large enough value of $\mu = x^T s/n$. That is, we must retreat to a prior iterate for the original instance until the adjustment $(\Delta x, \Delta y, \Delta s)$, when added to this iterate, does not cause some components of x or s to become negative. (Indeed, we require a stronger condition to hold: that the adjusted point $(x + \Delta x, y + \Delta y, s + \Delta s)$ belong to a neighborhood such as those of (2.10).) Since larger perturbations Δd generally lead to larger adjustments $(\Delta x, \Delta y, \Delta s)$, the prior iterate to which we must retreat is further back in the iteration sequence when Δd is larger. Most of the results in the following sections quantify this observation. They give a lower bound on $\mu/\|d\|$—expressed in terms of the size of the components of Δd, the conditioning $\mathcal{C}(d)$ of the original problem, and other quantities—such that the warm-start strategy, applied from a point (x, y, s) satisfying $\mu = x^T s/n$ and a neighborhood condition, yields a valid starting point for the perturbed problem.

Our strategy contrasts with that of Gondzio, who uses the solution of the original problem as a starting point in the computation of a central path point for the new problem, which has additional columns in the matrix A. Our strategies instead rely on a single correction to an interior-point iterate for the original problem to obtain a loosely centered starting point for the modified problem. We focus just on correcting the infeasibility of the linear equality conditions in (P) and (D), relying on the loose centrality of the original iterate to provide us with sufficient centrality of the adjusted starting point.

These results can be applied in a practical way when an interior-point approach is used to solve the original instance. Let $\{(x^k, y^k, s^k),\ k = 0, \ldots, K\}$ denote the iterates generated while solving the original problem. One can then store a subset of the iterates $\{(x^{k_i}, y^{k_i}, s^{k_i}),\, i = 0, 1, \ldots, L\}$ with $k_0 = 0$, which is the shortest sequence satisfying the property

$$\mu_{k_{i+1}} \geq \nu \mu_{k_i} \quad \forall i = 0, 1, 2, \ldots \tag{3.5}$$

for some ν with $0 < \nu \ll 1$. Suppose that we denote the lower bound discussed in the preceding paragraph by $\mu^*/\|d\|$. Then the best available starting point from the saved subsequence is the one with index k_ℓ, where ℓ is the largest index for which

$$\mu_{k_\ell} \geq \mu^*.$$

Because of (3.5) and the choice of ℓ, we have in fact that

$$\mu^* \leq \mu_{k_\ell} \leq (1/\nu)\mu^*. \tag{3.6}$$

The warm-start point is then

$$(\bar{x}^0, \bar{y}^0, \bar{s}^0) = (x^{k_\ell}, y^{k_\ell}, s^{k_\ell}) + (\Delta x, \Delta y, \Delta s), \tag{3.7}$$

where $(\Delta x, \Delta y, \Delta s)$ is the adjustment computed from one of our warm-start strategies. The duality measure corresponding to this point is

$$\bar{\mu}_0 = (\bar{x}^0)^T \bar{s}^0/n = \mu_{k_\ell} + (x^{k_\ell})^T \Delta s + (s^{k_\ell})^T \Delta x + \Delta x^T \Delta s.$$

By using the bounds on the components of $(\Delta x, \Delta y, \Delta s)$ that are obtained during the proofs of each major result, in conjunction with the bounds (2.13), we find that $\bar{\mu}_0$ can be bounded above by some multiple of $\mu^* + \mu_{k_\ell}$. Because of (3.6), we can deduce in each case that

$$\bar{\mu}_0 \leq \kappa \mu^* \tag{3.8}$$

for some κ independent of the problem instance d and the perturbation Δd. We conclude, by applying (3.3) to the perturbed instance, that the number of iterations required to satisfy the stopping criterion

$$\mu \leq \epsilon \|d + \Delta d\|, \tag{3.9}$$

starting from $(\bar{x}^0, \bar{y}^0, \bar{s}^0)$, is bounded by

$$\mathcal{O}\left(n^\tau \log \frac{\mu^*}{\|d + \Delta d\| \epsilon}\right). \tag{3.10}$$

Since our assumptions on $\|\Delta d\|$ usually ensure that

$$\|\Delta d\| \leq 0.5 \|d\|, \tag{3.11}$$

we have that

$$\frac{1}{\|d + \Delta d\|} \leq \frac{1}{\|d\| - \|\Delta d\|} \leq \frac{2}{\|d\|},$$

so that (3.10) can be expressed more conveniently as

$$\mathcal{O}\left(n^\tau \log \frac{\mu^*}{\|d\| \epsilon}\right). \tag{3.12}$$

After some of the results in subsequent sections, we will substitute for τ and μ^* in (3.12), to express the bound on the number of iterations in terms of the conditioning $\mathcal{C}(d)$ of the original instance and the size of the perturbation Δd.

Our first warm-start strategy, a least-squares correction, is described in section 4. The second strategy, a "Newton step correction," is based on a recent paper by Yıldırım and Todd [18] and is described in section 5.

4. Least-squares correction. For much of this section, we restrict our analysis to the changes in b and c only; that is, we assume

$$\Delta d = (0, \Delta b, \Delta c). \tag{4.1}$$

Perturbations to A will be considered in section 4.3.

Given any primal-dual feasible point (x, y, s) for the instance d, the least-squares correction for the perturbation (4.1) is the vector $(\Delta x, \Delta y, \Delta s)$ obtained from the solutions of the following subproblems:

$$\min \|\Delta x\| \quad \text{s.t. } A(x + \Delta x) = b + \Delta b,$$
$$\min \|\Delta s\| \quad \text{s.t. } A^T(y + \Delta y) + (s + \Delta s) = c + \Delta c.$$

Because $Ax = b$ and $A^T y + s = c$, we can restate these problems as

$$\min \|\Delta x\| \quad \text{s.t. } A\Delta x = \Delta b, \tag{4.2a}$$
$$\min \|\Delta s\| \quad \text{s.t. } A^T \Delta y + \Delta s = \Delta c, \tag{4.2b}$$

which are independent of (x, y, s). Given the following QR factorization of A^T,

$$A^T = \begin{bmatrix} Y & Z \end{bmatrix} \begin{bmatrix} R \\ 0 \end{bmatrix} = YR, \tag{4.3}$$

where $[Y \quad Z]$ is orthogonal and R is upper triangular, we find by simple manipulation of the optimality conditions that the solutions can be written explicitly as

$$\Delta x = YR^{-T}\Delta b, \tag{4.4a}$$
$$\Delta y = R^{-1}Y^T\Delta c, \tag{4.4b}$$
$$\Delta s = (I - YY^T)\Delta c. \tag{4.4c}$$

Contrary to a usual feasible interior-point step, Δx is in the range space of A^T, and Δs is in the null space of A. Consequently,

$$\Delta x^T \Delta s = 0. \tag{4.5}$$

Our strategy is as follows: We calculate the correction (4.4) just once, then choose an iterate (x^k, y^k, s^k) for the original problem such that $(x^k + \Delta x, s^k + \Delta s) > 0$, $(x^k + \Delta x, y^k + \Delta y, s^k + \Delta s)$ lies within either $\bar{\mathcal{N}}_2(\theta)$ or $\bar{\mathcal{N}}_{-\infty}(\gamma)$, and k is the largest index for which these properties hold. We hope to be able to satisfy these requirements for some index k for which the parameter μ_k is not too large. In this manner, we hope to obtain a starting point for the perturbed problem for which the initial value of μ is not large, so that we can solve the problem using a smaller number of interior-point iterations than if we had started without the benefit of the iterates from the original problem.

Some bounds that we use throughout our analysis follow immediately from (4.4):

$$\|\Delta s\| \leq \|\Delta c\|, \qquad \|\Delta x\| \leq \frac{\|\Delta b\|}{\sigma_m(A)} \leq \frac{\|\Delta b\|}{\rho(d)}, \tag{4.6}$$

where, as in (2.5), $\sigma_m(A)$ is the minimum singular value of A. These bounds follow from the fact that $I - YY^T$ is an orthogonal projection matrix onto the null space of A and from the observation that R has the same singular values as A. By defining

$$\delta_b = \frac{\|\Delta b\|}{\|d\|}, \qquad \delta_c = \frac{\|\Delta c\|}{\|d\|}, \tag{4.7}$$

we can rewrite (4.6) as

$$\|\Delta s\| \leq \|d\|\delta_c, \qquad \|\Delta x\| \leq \mathcal{C}(d)\delta_b. \tag{4.8}$$

We also define the following quantity, which occurs frequently in the analysis:

$$\delta_{bc} = \delta_c + 2\mathcal{C}(d)\delta_b. \tag{4.9}$$

In the remainder of the paper, we make the mild assumption that

$$\delta_b < 1, \qquad \delta_c < 1. \tag{4.10}$$

4.1. Small neighborhood. Suppose that we have iterates for the original problem that satisfy the following property, for some $\theta_0 \in (0, 1)$:

$$\|XSe - \mu e\|_2 \leq \theta_0 \mu, \quad \text{where } \mu = x^T s/n. \tag{4.11}$$

That is, $(x, y, s) \in \mathcal{N}_2(\theta_0)$. Iterates of a short-step path-following algorithm typically satisfy a condition of this kind. Since (x, y, s) is a strictly feasible point, its components satisfy the bounds (2.13). Note, too, that we have

$$\|XSe - \mu e\| \leq \theta_0 \mu \;\; \Rightarrow \;\; (1 - \theta_0)\mu \leq x_i s_i \leq (1 + \theta_0)\mu. \tag{4.12}$$

Our first proposition gives conditions on δ_{bc} and μ that ensure that the least-squares correction yields a point in the neighborhood $\bar{\mathcal{N}}_{-\infty}(\gamma)$.

Proposition 4.1. *Let $\gamma \in (0, 1 - \theta_0)$ be given, and let $\xi \in (0, 1 - \gamma - \theta_0)$. Assume that Δd satisfies*

$$\delta_{bc} \leq \frac{1 - \theta_0 - \gamma - \xi}{(n + 1)\mathcal{C}(d)}. \tag{4.13}$$

Let $(x, y, s) \in \mathcal{N}_2(\theta_0)$, and suppose that $(\Delta x, \Delta y, \Delta s)$ is the least-squares correction (4.4)*. Then $(x + \Delta x, y + \Delta y, s + \Delta s)$ lies in $\bar{\mathcal{N}}_{-\infty}(\gamma)$ if*

$$\mu \geq \frac{\|d\|}{\xi} 3\mathcal{C}(d)^2 \delta_{bc} \stackrel{\text{def}}{=} \mu_1^*. \tag{4.14}$$

Proof. By using (4.12), (2.13), (4.8), and (4.9), we obtain a lower bound on $(x_i + \Delta x_i)(s_i + \Delta s_i)$ as follows:

$$\begin{aligned}
&(x_i + \Delta x_i)(s_i + \Delta s_i) \\
&= x_i s_i + x_i \Delta s_i + \Delta x_i s_i + \Delta x_i \Delta s_i \\
&\geq (1 - \theta_0)\mu - \|x\|\|\Delta s\| - \|\Delta x\|\|s\| - \|\Delta x\|\|\Delta s\| \\
&\geq (1 - \theta_0)\mu - \mathcal{C}(d)\left(\mathcal{C}(d) + \mu n/\|d\|\right)\|d\|\delta_c \\
&\qquad - 2\|d\|\mathcal{C}(d)^2(\mathcal{C}(d) + \mu n/\|d\|)\delta_b - \|d\|\mathcal{C}(d)\delta_b\delta_c \\
&\geq \mu\left(1 - \theta_0 - n\mathcal{C}(d)\delta_{bc}\right) - \mathcal{C}(d)^2\|d\|\delta_{bc} - \mathcal{C}(d)\|d\|\delta_b\delta_c \\
&\geq \mu\left(1 - \theta_0 - n\mathcal{C}(d)\delta_{bc}\right) - 2\mathcal{C}(d)^2\|d\|\delta_{bc}.
\end{aligned} \tag{4.15}$$

Because of our assumption (4.13), the coefficient of μ in (4.15) is positive, and thus (4.15) represents a positive lower bound on $(x_i + \Delta x_i)(s_i + \Delta s_i)$ for all μ sufficiently large.

For an upper bound on $(x + \Delta x)^T(s + \Delta s)/n$, we have from (2.13), (4.8), and the relation (4.5) that

$$\begin{aligned}
&(x + \Delta x)^T(s + \Delta s)/n \\
&\leq \mu + \|\Delta x\|\|s\|/n + \|x\|\|\Delta s\|/n \\
&\leq \mu + 2\mathcal{C}(d)^2\|d\|\delta_b(\mathcal{C}(d) + \mu n/\|d\|)/n + \mathcal{C}(d)\|d\|\delta_c(\mathcal{C}(d) + \mu n/\|d\|)/n \\
&\leq \mu(1 + \mathcal{C}(d)\delta_{bc}) + \mathcal{C}(d)^2\|d\|\delta_{bc}/n.
\end{aligned} \tag{4.16}$$

It follows from this bound and (4.15) that a sufficient condition for the conclusion of the proposition to hold is that

$$\mu(1 - \theta_0 - n\mathcal{C}(d)\delta_{bc}) - 2\mathcal{C}(d)^2\|d\|\delta_{bc} \geq \gamma\mu(1 + \mathcal{C}(d)\delta_{bc}) + \gamma\mathcal{C}(d)^2\|d\|\delta_{bc}/n,$$

which is equivalent to

$$\mu \geq \frac{\|d\|\mathcal{C}(d)^2\delta_{bc}(2+\gamma/n)}{1-\theta_0-\gamma-\mathcal{C}(d)\delta_{bc}(n+\gamma)}, \tag{4.17}$$

provided that the denominator is positive. Because of condition (4.13), and using $\gamma \in (0,1)$ and $n \geq 1$, the denominator is in fact bounded below by the positive quantity ξ, and thus the condition (4.17) is implied by (4.14).

Finally, we show that our bounds ensure the positivity of $x + \Delta x$ and $s + \Delta s$. It is easy to show that the right-hand side of (4.15) is also a lower bound on $(x_i + \alpha\Delta x_i)(s_i + \alpha\Delta s_i)$ for all $\alpha \in [0,1]$ and all $i = 1, 2, \ldots, n$. Because μ satisfies (4.17), we have $(x_i + \alpha\Delta x_i)(s_i + \alpha\Delta s_i) > 0$ for all $\alpha \in [0,1]$. Since we know that $(x, s) > 0$, we conclude that $x_i + \Delta x_i > 0$ and $s_i + \Delta s_i > 0$ for all i as well, completing the proof. ∎

Next, we seek conditions on δ_{bc} and μ that ensure that the corrected iterate lies in a narrow central path neighborhood for the perturbed problem.

PROPOSITION 4.2. *Let $\theta > \theta_0$ be given, and let $\xi \in (0, \theta - \theta_0)$. Assume that the perturbation Δd satisfies*

$$\delta_{bc} \leq \frac{\theta - \theta_0 - \xi}{(2n+1)\mathcal{C}(d)}. \tag{4.18}$$

Suppose that $(x, y, s) \in \mathcal{N}_2(\theta_0)$ for the original problem and that $(\Delta x, \Delta y, \Delta s)$ is the least-squares correction (4.4). *Then, $(x + \Delta x, y + \Delta y, s + \Delta s)$ will lie in $\bar{\mathcal{N}}_2(\theta)$ if*

$$\mu \geq \frac{\|d\|}{\xi} 4\mathcal{C}(d)^2\delta_{bc} \stackrel{\text{def}}{=} \mu_2^*. \tag{4.19}$$

Proof. We start by finding a bound on the norm of the vector

$$[(x_i + \Delta x_i)(s_i + \Delta s_i)]_{i=1,2,\ldots,n} - \left[(x+\Delta x)^T(s+\Delta s)/n\right] e. \tag{4.20}$$

Given two vectors y and z in R^n, we have that

$$\left\| [y_i z_i]_{i=1,2,\ldots,n} \right\| \leq \|y\|\,\|z\|, \qquad \left| y^T z \right| \leq \|y\|\,\|z\|. \tag{4.21}$$

By using these elementary inequalities together with (4.5), (4.8), (4.9), and (2.13), we have that the norm of (4.20) is bounded by

$$\begin{aligned}
&\left\| [x_i s_i]_{i=1,2,\ldots,n} - \mu e \right\| + 2\,[\,\|\Delta x\|\,\|s\| + \|x\|\,\|\Delta s\|\,] + \|\Delta x\|\,\|\Delta s\| \\
&\leq \theta_0\mu + 2\mathcal{C}(d)\|d\|\delta_{bc}\,(\mathcal{C}(d) + n\mu/\|d\|) + \mathcal{C}(d)\|d\|\delta_b\delta_c \\
&\leq [\theta_0 + 2n\mathcal{C}(d)\delta_{bc}]\,\mu + 3\|d\|\mathcal{C}(d)^2\delta_{bc}.
\end{aligned}$$

Meanwhile, we obtain a lower bound on the duality measure after the correction by using the same set of relations:

$$\begin{aligned}
(x+\Delta x)^T(s+\Delta s)/n &\geq \mu - [\,\|\Delta x\|\,\|s\| + \|x\|\,\|\Delta s\|\,]/n \\
&\geq \mu - \mathcal{C}(d)\|d\|\delta_{bc}(\mathcal{C}(d) + n\mu/\|d\|)/n \\
&\geq \mu\,[1 - \mathcal{C}(d)\delta_{bc}] - \mathcal{C}(d)^2\|d\|\delta_{bc}/n.
\end{aligned} \tag{4.22}$$

Therefore, a sufficient condition for

$$(x + \Delta x, y + \Delta y, s + \Delta s) \in \bar{\mathcal{N}}_2(\theta)$$

is that

$$[\theta_0 + 2n\mathcal{C}(d)\delta_{bc}]\,\mu + 3\|d\|\mathcal{C}(d)^2\delta_{bc} \le \theta\mu\,[1 - \mathcal{C}(d)\delta_{bc}] - \theta\mathcal{C}(d)^2\|d\|\delta_{bc}/n,$$

which after rearrangement becomes

$$\mu\,[\theta - \theta_0 - 2n\mathcal{C}(d)\delta_{bc} - \theta\mathcal{C}(d)\delta_{bc}] \ge 3\|d\|\mathcal{C}(d)^2\delta_{bc} + \theta\|d\|\mathcal{C}(d)^2\delta_{bc}/n. \tag{4.23}$$

We have from (4.18) that the coefficient of μ on the left-hand side of this expression is bounded below by ξ. By dividing both sides of (4.23) by this expression and using $\theta \in (0, 1)$ and $n \ge 1$, we find that (4.19) is a sufficient condition for (4.23). A similar argument as in the proof of Proposition 4.1, together with the fact that $\mu_2^* > \mu_1^*$, ensures positivity of $(x + \Delta x, s + \Delta s)$. □

We now specialize the discussion of section 3 to show that Propositions 4.1 and 4.2 can be used to obtain lower complexity estimates for the interior-point warm-start strategy.

Considering first the case of Proposition 4.1, we have from the standard analysis of a long-step path-following algorithm that constrains its iterates to lie in $\bar{\mathcal{N}}_{-\infty}(\gamma)$ (see, for example, Wright [16, Chapter 5]) that the reduction in duality measure at each iteration satisfies (3.1) with

$$\tau = 1, \quad \delta = 2^{\frac{3}{2}}\gamma\frac{1-\gamma}{1+\gamma}\min\{\sigma_{\min}(1-\sigma_{\min}), \sigma_{\max}(1-\sigma_{\max})\},$$

where $0 < \sigma_{\min} < \sigma_{\max} < 1$ are the lower and upper bounds on the centering parameter σ at each iteration. Choosing one of the iterates of this algorithm (x^ℓ, y^ℓ, s^ℓ) in the manner of section 3 and defining the starting point as in (3.7), we have from (4.16), (4.13), (4.14), and the conditions $0 < \xi < 1$ and $n \ge 1$ that

$$\begin{aligned}\bar{\mu}_0 &= (\bar{x}^0)^T\bar{s}^0/n \\ &\le \mu_\ell(1 + \mathcal{C}(d)\delta_{bc}) + \mathcal{C}(d)^2\|d\|\delta_{bc}/n \le \mu_\ell(1 + 1/n) + \mu_1^*(\xi/n) \le 2\mu_\ell + \mu_1^*.\end{aligned}$$

Now from the property (3.6), it follows that

$$\bar{\mu}_0 \le (1 + 2/\nu)\mu_1^*.$$

It is easy to verify that (4.13) implies that $\|\Delta d\| \le \|d\|/2$, so that we can use the expression (3.12) to estimate the number of iterations. By substituting $\tau = 1$ and $\mu^* = \mu_1^*$ into (3.12), we obtain

$$\mathcal{O}\left(n\log\left(\frac{1}{\epsilon}\mathcal{C}(d)^2\delta_{bc}\right)\right) \quad \text{iterations.} \tag{4.24}$$

We conclude that if δ_{bc} is small in the sense that $\delta_{bc} \ll \mathcal{C}(d)^{-2}$, then the estimate (4.24) is an improvement on the cold-start complexity estimate (3.4), and thus it is advantageous to use the warm-start strategy.

Taking now the case of a starting point in the smaller neighborhood of Proposition 4.2, we set $\theta = 0.4$, and the centering parameter σ to the constant value $1 - 0.4/n^{1/2}$. The standard analysis of the short-step path-following algorithm (see, for example, [16, Chapter 4]) then shows that (3.1) holds with

$$\tau = 0.5, \qquad \delta = 0.4.$$

By using the procedure outlined in section 3 to derive the warm-start point, the argument of the preceding paragraph can be applied to obtain the following on the number of iterations:

$$\mathcal{O}\left(n^{1/2}\log\left(\frac{1}{\epsilon}\mathcal{C}(d)^2\delta_{bc}\right)\right). \tag{4.25}$$

We conclude as before that improved complexity over a cold start is available, provided that $\delta_{bc} \ll \mathcal{C}(d)^{-2}$.

4.2. Wide neighborhood. We now consider the case in which the iterates for the original problem lie in a wide neighborhood of the central path. To be specific, we suppose that they satisfy $x_i s_i \geq \gamma_0\mu$ for some $\gamma_0 \in (0,1)$, that is, $(x,y,s) \in \mathcal{N}_{-\infty}(\gamma_0)$. Note that, in this case, we have the following bounds on the pairwise products:

$$\gamma_0\mu \leq x_i s_i \leq (n-(n-1)\gamma_0)\mu. \tag{4.26}$$

Similarly to the upper bounds (2.13) on $\|x\|$ and $\|s\|$, we can derive lower bounds on x_i and s_i by combining (2.13) with (4.26) and using $x_i \leq \|x\|$ and $s_i \leq \|s\|$:

$$x_i \geq \frac{\gamma_0\mu}{2\|d\|\mathcal{C}(d)(\mathcal{C}(d)+n\mu/\|d\|)}, \tag{4.27a}$$

$$s_i \geq \frac{\gamma_0\mu}{\mathcal{C}(d)(\mathcal{C}(d)+n\mu/\|d\|)}. \tag{4.27b}$$

These lower bounds will be useful in the later analysis. The following proposition gives a sufficient condition for the least-squares corrected point to be a member of the wide neighborhood for the perturbed problem. The proof uses an argument identical to the proof of Proposition 4.1, with γ_0 replacing $(1-\theta_0)$.

PROPOSITION 4.3. *Given γ and γ_0 such that $0 < \gamma < \gamma_0 < 1$, suppose that ξ is a parameter satisfying $\xi \in (0, \gamma_0 - \gamma)$. Assume that Δd satisfies*

$$\delta_{bc} \leq \frac{\gamma_0 - \gamma - \xi}{(n+1)\mathcal{C}(d)}. \tag{4.28}$$

Suppose also that $(x,y,s) \in \mathcal{N}_{-\infty}(\gamma_0)$, and denote by $(\Delta x, \Delta y, \Delta s)$ the least-squares correction (4.4)*. Then a sufficient condition for*

$$(x+\Delta x, y+\Delta y, s+\Delta s) \in \bar{\mathcal{N}}_{-\infty}(\gamma) \tag{4.29}$$

is that

$$\mu \geq \frac{\|d\|}{\xi}3\mathcal{C}(d)^2\delta_{bc} \stackrel{\text{def}}{=} \mu_3^*. \tag{4.30}$$

An argument like the one leading to (4.24) can now be used to show that a long-step path-following method requires

$$\mathcal{O}\left(n\log\left(\frac{1}{\epsilon}\mathcal{C}(d)^2\delta_{bc}\right)\right) \quad \text{iterations} \tag{4.31}$$

to converge from the warm-start point to a point that satisfies (3.9).

4.3. Perturbations in *A*. We now allow for perturbations in A as well as in b and c. By doing so, we introduce some complications in the analysis that can be circumvented by imposing an a priori upper bound on the values of μ that we are willing to consider. This upper bound is large enough to encompass all values of μ of interest from the viewpoint of complexity, in the sense that when μ exceeds this bound, the warm-start strategy does not lead to an appreciably improved complexity estimate over the cold-start approach.

For some constant $\zeta > 1$, we assume that μ satisfies the bound

$$\mu \leq \frac{\zeta - 1}{n} \|d\| \mathcal{C}(d) \stackrel{\text{def}}{=} \mu_{\text{up}}, \tag{4.32}$$

so that, for a subexpression that recurs often in the preceding sections, we have

$$\mathcal{C}(d) + n\mu/\|d\| \leq \zeta \mathcal{C}(d).$$

For $\mu \in [0, \mu_{\text{up}}]$, we can simplify a number of estimates in the preceding sections, to remove their explicit dependence on μ. In particular, the bounds (2.13) on the strictly feasible point (x, y, s) with $\mu = x^T s/n$ become

$$\|x\| \leq \zeta \mathcal{C}(d)^2, \quad \|y\| \leq \zeta \mathcal{C}(d)^2, \quad \|s\| \leq 2\zeta \|d\| \mathcal{C}(d)^2. \tag{4.33}$$

Given a perturbation $\Delta d = (\Delta A, \Delta b, \Delta c)$ with $\|\Delta d\| < \rho(d)$, we know that $A + \Delta A$ has full rank. In particular, for the smallest singular value, we have

$$\sigma_m(A + \Delta A) \geq \sigma_m(A) - \|\Delta A\|. \tag{4.34}$$

To complement the definitions (4.7), we introduce

$$\delta_A = \frac{\|\Delta A\|}{\|d\|}. \tag{4.35}$$

As before, we consider a warm-start strategy obtained by applying least-squares corrections to a given point (x, y, s) that is strictly feasible for the unperturbed problem. The correction Δx is the solution of the following subproblem:

$$\min \|\Delta x\| \;\text{ s.t. }\; (A + \Delta A)(x + \Delta x) = b + \Delta b, \tag{4.36}$$

which is given explicitly by

$$\Delta x = (A + \Delta A)^T \left[(A + \Delta A)(A + \Delta A)^T\right]^{-1} (\Delta b - \Delta A x), \tag{4.37}$$

where we have used $Ax = b$. By using the QR factorization of $(A + \Delta A)^T$ as in (4.3) and (4.4), we find the following bound on $\|\Delta x\|$:

$$\|\Delta x\| \leq \frac{\|\Delta b\| + \|\Delta A\| \|x\|}{\sigma_m(A + \Delta A)}. \tag{4.38}$$

By using (4.34), (2.5), and the definitions (4.7), (4.35), and (2.3), we have

$$\|\Delta x\| \leq \frac{\|\Delta b\| + \|\Delta A\| \|x\|}{\sigma_m(A) - \|\Delta A\|} \leq \frac{\|\Delta b\| + \|\Delta A\| \|x\|}{\rho(d) - \|\Delta A\|} = \frac{\delta_b + \delta_A \|x\|}{1/\mathcal{C}(d) - \delta_A}.$$

In particular, when x is strictly feasible for the original problem, we have from (4.33) that

$$\|\Delta x\| \leq \mathcal{C}(d)\frac{\delta_b + \zeta\mathcal{C}(d)^2\delta_A}{1-\delta_A\mathcal{C}(d)},$$

while if we make the additional simple assumption that

$$\delta_A \leq \frac{1}{2\mathcal{C}(d)}, \tag{4.39}$$

then we have immediately that

$$\|\Delta x\| \leq 2\mathcal{C}(d)\delta_b + 2\zeta\mathcal{C}(d)^3\delta_A. \tag{4.40}$$

By using (4.39) again, together with (4.10) and the known bounds $\mathcal{C}(d) \geq 1$ and $\zeta > 1$, we obtain

$$\|\Delta x\| \leq 2\mathcal{C}(d)\delta_b + 2\zeta\mathcal{C}(d)^3\delta_A \leq 2\mathcal{C}(d) + \zeta\mathcal{C}(d)^2 \leq 3\zeta\mathcal{C}(d)^2. \tag{4.41}$$

The dual perturbation is the solution of the problem

$$\min \|\Delta s\| \ \text{ s.t. } \ (A+\Delta A)^T(y+\Delta y) + (s+\Delta s) = c + \Delta c. \tag{4.42}$$

Once again, the minimum norm solution is unique and given by

$$\Delta s = \left[I - (A+\Delta A)^T\left((A+\Delta A)(A+\Delta A)^T\right)^{-1}(A+\Delta A)\right](\Delta c - \Delta A^T y). \tag{4.43}$$

Therefore, we have the following upper bound:

$$\|\Delta s\| \leq \|\Delta c\| + \|\Delta A\|\|y\|. \tag{4.44}$$

Using (4.33), we have for (x,y,s) strictly feasible for the original problem that

$$\|\Delta s\| \leq \|\Delta c\| + \|\Delta A\|\zeta\mathcal{C}(d)^2 \leq \|d\|\delta_c + \zeta\|d\|\mathcal{C}(d)^2\delta_A. \tag{4.45}$$

By using these inequalities, we can prove a result similar to Proposition 4.3.

PROPOSITION 4.4. *Suppose we are given γ and γ_0 such that $0 < \gamma < \gamma_0 < 1$, and a feasible primal-dual point $(x,y,s) \in \mathcal{N}_{-\infty}(\gamma_0)$. Assume further that $\mu = x^Ts/n$ satisfies* (4.32) *and that the perturbation component ΔA satisfies* (4.39). *For the perturbation Δd, suppose that $(\Delta x, \Delta y, \Delta s)$ is the least-squares correction obtained from* (4.36) *and* (4.42). *We then have*

$$(x+\Delta x, y+\Delta y, s+\Delta s) \in \bar{\mathcal{N}}_{-\infty}(\gamma), \tag{4.46}$$

provided that μ satisfies the following lower bound:

$$\mu \geq 19\zeta\mathcal{C}(d)^2\frac{\|d\|}{\gamma_0-\gamma}\max\left(\delta_{bc}, \zeta\mathcal{C}(d)^3\delta_A\right) \stackrel{\text{def}}{=} \mu_4^*. \tag{4.47}$$

Proof. By using the upper bounds (4.40) and (4.41) on $\|\Delta x\|$, (4.45) on $\|\Delta s\|$, and (4.33) on $\|x\|$ and $\|s\|$, we have

$$\begin{aligned}
&(x_i+\Delta x_i)(s_i+\Delta s_i)\\
&\geq \gamma_0\mu - (\|x\| + \|\Delta x\|)\|\Delta s\| - \|s\|\|\Delta x\|\\
&\geq \gamma_0\mu - [4\zeta\mathcal{C}(d)^2][\|d\|\delta_c + \zeta\|d\|\mathcal{C}(d)^2\delta_A]\\
&\qquad - [2\|d\|\zeta\mathcal{C}(d)^2][2\mathcal{C}(d)\delta_b + 2\zeta\mathcal{C}(d)^3\delta_A]\\
&\geq \gamma_0\mu - 4\|d\|\zeta\mathcal{C}(d)^3\delta_b - 4\|d\|\zeta\mathcal{C}(d)^2\delta_c - 8\|d\|\zeta^2\mathcal{C}(d)^5\delta_A\\
&\geq \gamma_0\mu - 4\|d\|\zeta\mathcal{C}(d)^2\delta_{bc} - 8\|d\|\zeta^2\mathcal{C}(d)^5\delta_A,
\end{aligned}$$

where for the last inequality we have used the definition (4.9). By similar logic, and using (4.5), we have for the updated duality measure that

$$\begin{aligned}
&(x+\Delta x)^T(s+\Delta s)/n \\
&\le \mu + \|\Delta x\|\,\|s\|/n + \|x\|\,\|\Delta s\|/n \\
&\le \mu + [2\mathcal{C}(d)\delta_b + 2\zeta\mathcal{C}(d)^3\delta_A]2\zeta\|d\|\mathcal{C}(d)^2/n + \zeta\mathcal{C}(d)^2[\|d\|\delta_c + \zeta\|d\|\mathcal{C}(d)^2\delta_A]/n \\
&= \mu + 4\zeta\mathcal{C}(d)^3\|d\|\delta_b/n + \zeta\mathcal{C}(d)^2\|d\|\delta_c/n + 5\zeta^2\mathcal{C}(d)^5\|d\|\delta_A/n \\
&\le \mu + 2\zeta\mathcal{C}(d)^2\|d\|\delta_{bc}/n + 5\zeta^2\mathcal{C}(d)^5\|d\|\delta_A/n.
\end{aligned}$$

By comparing these two inequalities in the usual way and using $\gamma \in (0,1)$ and $n \ge 1$, we have that a sufficient condition for the conclusion (4.46) to hold is that

$$(\gamma_0 - \gamma)\mu \ge 6\|d\|\zeta\mathcal{C}(d)^2\delta_{bc} + 13\|d\|\zeta^2\mathcal{C}(d)^5\delta_A. \tag{4.48}$$

Since from (4.47), we have

$$\begin{aligned}
\frac{6}{19}(\gamma_0 - \gamma)\mu &\ge 6\|d\|\zeta\mathcal{C}(d)^2\delta_{bc}, \\
\frac{13}{19}(\gamma_0 - \gamma)\mu &\ge 13\|d\|\zeta^2\mathcal{C}(d)^5\delta_A,
\end{aligned}$$

then (4.48) holds. Finally, the positivity of $x + \Delta x$ and that of $s + \Delta s$ can be shown in a way similar to the proof of Proposition 4.1. Once again, the lower bound for $(x_i + \Delta x_i)(s_i + \Delta s_i)$ also holds for $(x_i + \alpha\Delta x_i)(s_i + \alpha\Delta s_i)$ for any $\alpha \in [0,1]$. Using the simple inequality $a + b \le 2\max(a,b)$, we obtain

$$(x_i + \alpha\Delta x_i)(s_i + \alpha\Delta s_i) \ge \gamma_0\mu - 8\zeta\mathcal{C}(d)^2\|d\|\max\left(\delta_{bc}, 2\zeta\mathcal{C}(d)^3\delta_A\right),$$

which yields a positive lower bound by (4.47), and the proof is complete. ◻

By using an argument like the ones leading to (4.24) and (4.31), we deduce that a long-step path-following algorithm that uses the warm start prescribed in Proposition 4.4 requires

$$\mathcal{O}\left(n\left[\log\left(\frac{1}{\epsilon}\mathcal{C}(d)^2\delta_{bc}\right) + \log\left(\frac{1}{\epsilon}\mathcal{C}(d)^5\delta_A\right)\right]\right) \quad \text{iterations} \tag{4.49}$$

to converge to a point that satisfies (3.9).

5. Newton step correction. In a recent study, Yıldırım and Todd [18] analyzed the perturbations in b and c in linear and semidefinite programming using interior-point methods. For such perturbations they stated a sufficient condition on the norm of the perturbation, which depends on the current iterate, so that an adjustment to the current point based on applying an iteration of Newton's method to the system (2.6a), (2.6b), (2.6c) yields a feasible iterate for the perturbed problem with a lower duality gap than that of the original iterate. In this section, we augment some of the analysis of [18] with other results, like those of section 4, to find conditions on the duality gap $\mu = x^T s/n$ and the perturbation size under which the Newton step yields a warm-start point that gives significantly better complexity than a cold start.

Each iteration of a primal-dual interior-point method involves solving a Newton-like system of linear equations whose coefficient matrix is the Jacobian of the system (2.6a), (2.6b), (2.6c). The general form of these equations is

$$\begin{array}{rcccl}
 & A\Delta x & & = & r_p, \\
A^T\Delta y & & + \;\; \Delta s & = & r_d, \\
 & S\Delta x & + \;\; X\Delta s & = & r_{xs},
\end{array} \tag{5.1}$$

where typically $r_p = b - Ax$ and $r_d = c - A^T y - s$. The choice of r_{xs} typically depends on the particular method being applied, but usually represents a Newton or higher-order step toward some "target point" (x', y', s'), which often lies on the central path $\mathcal{P}$ defined in (2.7).

In the approach used in Yıldırım and Todd [18] and in this section, this Newton-like system is used to correct for perturbations in the data (A, b, c) rather than to advance to a new primal-dual iterate. The right-hand side quantities are chosen so that the adjustment $(\Delta x, \Delta y, \Delta s)$ yields a point that is strictly feasible for the perturbed problem and whose duality gap is no larger than that of the current point (x, y, s).

In section 5.1, we consider the case of perturbations in b and c but not in A. In section 5.2 we allow perturbations in A as well.

5.1. Perturbations in *b* and *c*. In our strategy, we assume that

- the current point (x, y, s) is strictly primal-dual feasible for the original problem;
- the target point (x', y', s') used to define r_{xs} is a point that is strictly feasible for the perturbed problem for which $x_i' s_i' = x_i s_i$ for all $i = 1, 2, \ldots, n$;
- the step is a pure Newton step toward (x', y', s'); that is, $r_p = \Delta b$, $r_d = \Delta c$, and $r_{xs} = X'S'e - XSe = 0$.

Note that, in general, the second assumption is not satisfied for an arbitrary current point (x, y, s) because such a feasible point for the perturbed problem need not exist. However, Newton's method is still well defined with the above choices of r_p, r_d, and r_{xs}, and that assumption is merely stated for the sake of a complete description of our strategy.

Since A has full row rank by our assumption of $\rho(d) > 0$, we have, by substituting our right-hand side in (5.1) and performing block elimination, that the solution is given explicitly by

$$\Delta y = (AD^2A^T)^{-1}(\Delta b + AD^2\Delta c), \tag{5.2a}$$

$$\Delta s = \Delta c - A^T \Delta y, \tag{5.2b}$$

$$\Delta x = -S^{-1}X\Delta s, \tag{5.2c}$$

where

$$D^2 \stackrel{\text{def}}{=} S^{-1}X. \tag{5.3}$$

Since A has full row rank and D is positive diagonal, AD^2A^T is invertible.

The following is an extension of the results in Yıldırım and Todd [18] to the case of simultaneous perturbations in b and c. Note in particular that the Newton step yields a decrease in the duality gap $x^T s$.

PROPOSITION 5.1. *Assume that (x, y, s) is a strictly feasible point for d. Let $\Delta d = (0, \Delta b, \Delta c)$. Consider a Newton step $(\Delta x, \Delta y, \Delta s)$ taken from (x, y, s) targeting the point (x', y', s') that is strictly feasible for the perturbed problem and satisfies $X'S'e = XSe$, and let*

$$(\tilde{x}, \tilde{y}, \tilde{s}) \stackrel{\text{def}}{=} (x, y, s) + (\Delta x, \Delta y, \Delta s). \tag{5.4}$$

Then if

$$\left\| \begin{bmatrix} \Delta c \\ \Delta b \end{bmatrix} \right\|_\infty \leq \left\| \left[S^{-1}\left(I - A^T(AD^2A^T)^{-1}AD^2\right) \quad - S^{-1}A^T(AD^2A^T)^{-1} \right] \right\|_\infty^{-1}, \tag{5.5}$$

$(\tilde{x}, \tilde{y}, \tilde{s})$ *is feasible for the perturbed problem and satisfies*

$$\tilde{x}^T \tilde{s} \leq x^T s. \tag{5.6}$$

Proof. By rearranging (5.2c) and writing it componentwise, we have

$$s_i \Delta x_i + x_i \Delta s_i = 0 \Longleftrightarrow \frac{\Delta x_i}{x_i} + \frac{\Delta s_i}{s_i} = 0, \qquad i = 1, 2, \ldots, n. \tag{5.7}$$

The next iterate will be feasible if and only if

$$\frac{\Delta x_i}{x_i} \geq -1, \quad \frac{\Delta s_i}{s_i} \geq -1, \quad i = 1, 2, \ldots, n.$$

By combining these inequalities with (5.7), we find that feasibility requires

$$\left| \frac{\Delta x_i}{x_i} \right| \leq 1, \quad \left| \frac{\Delta s_i}{s_i} \right| \leq 1, \quad i = 1, 2, \ldots, n,$$

or, equivalently,

$$\left\| S^{-1} \Delta s \right\|_\infty = \left\| X^{-1} \Delta x \right\|_\infty \leq 1. \tag{5.8}$$

By using (5.2a) and (5.2c), we have

$$\begin{aligned} &\|S^{-1} \Delta s\|_\infty \\ &\quad = \left\| S^{-1} [\Delta c - A^T \Delta y] \right\|_\infty \\ &\quad = \left\| S^{-1} \left[\Delta c - A^T (AD^2 A^T)^{-1} AD^2 \Delta c - A^T (AD^2 A^T)^{-1} \Delta b \right] \right\|_\infty \\ &\quad \leq \left\| \left[S^{-1} \left(I - A^T (AD^2 A^T)^{-1} AD^2 \right) \quad - S^{-1} A^T (AD^2 A^T)^{-1} \right] \right\|_\infty \left\| \begin{bmatrix} \Delta c \\ \Delta b \end{bmatrix} \right\|_\infty . \end{aligned} \tag{5.9}$$

Hence, (5.5) is sufficient to ensure that $\|S^{-1} \Delta s\|_\infty \leq 1$.

By summing (5.7) over $i = 1, 2, \ldots, n$, we obtain $x^T \Delta s + s^T \Delta x = 0$. It is also clear from (5.7) that Δx_i and Δs_i have opposite signs for each $i = 1, 2, \ldots, n$, and thus $\Delta x^T \Delta s \leq 0$. Therefore, we have

$$(x + \Delta x)^T (s + \Delta s) = x^T s + x^T \Delta s + s^T \Delta x + \Delta x^T \Delta s = x^T s + \Delta x^T \Delta s \leq x^T s,$$

proving (5.6). ∎

Proposition 5.1 does not provide any insight about the behavior of the expression on the right-hand side of (5.5) as a function of μ. To justify our strategy of retreating to successively earlier iterates of the original problem, we need to show that the expression in question increases as μ corresponding to (x, y, s) increases, so that we can handle larger perturbations by considering iterates with larger values of μ. In the next theorem, we will show that there exists an increasing function $f(\mu)$ with $f(0) = 0$ that is a lower bound to the corresponding expression in (5.5) for all values of μ. The key to our result is the following bound:

$$\chi(H) \stackrel{\text{def}}{=} \sup_{\Sigma \in \mathcal{D}_+} \left\| \Sigma H^T (H \Sigma H^T)^{-1} \right\|_\infty < \infty, \tag{5.10}$$

where $\mathcal{D}_+$ denotes the set of diagonal matrices in $R^{n \times n}$ with strictly positive diagonal elements (i.e., positive definite diagonal matrices) and $\| \cdot \|_\infty$ is the ℓ_∞ matrix norm defined as the maximum of the sums of the absolute values of the entries in each row. This result, by now well known, was apparently first proved by Dikin [2]. For a survey of the background and applications of this and related results, see Forsgren [3].

THEOREM 5.2. *Consider points (x, y, s) in the neighborhood $\mathcal{N}_{-\infty}(\gamma_0)$ for the original problem, with $\gamma_0 \in (0, 1)$ and $\mu = x^T s/n$ as defined in (2.11). Then there exists an increasing function $f(\mu)$ with $f(0) = 0$ such that the expression on the right-hand side of (5.5) is bounded below by $f(\mu)$ for all (x, y, s) in this neighborhood.*

Proof. Let (x, y, s) be a strictly feasible pair of points for the original problem, which lies in $\mathcal{N}_{-\infty}(\gamma_0)$ for some $\gamma_0 \in (0, 1)$. From (4.27) and (5.10), we have

$$\begin{aligned} \left\|S^{-1}A^T(AD^2A^T)^{-1}\right\|_\infty &= \left\|S^{-1}D^{-2}D^2A^T(AD^2A^T)^{-1}\right\|_\infty \\ &\leq \left\|X^{-1}\right\|_\infty \left\|D^2A^T(AD^2A^T)^{-1}\right\|_\infty \\ &\leq \left(\frac{1}{\mu}\right)\frac{2\|d\|\mathcal{C}(d)}{\gamma_0}\left(\mathcal{C}(d) + n\mu/\|d\|\right)\chi(A). \end{aligned} \tag{5.11}$$

The first inequality is simply the matrix norm inequality. Since $D^2 = XS^{-1}$, and x and s are strictly feasible, D^2 is a positive definite diagonal matrix, and thus the bound in (5.10) applies.

Similarly, consider the following:

$$\begin{aligned} &\left\|S^{-1}\left(I - A^T(AD^2A^T)^{-1}AD^2\right)\right\|_\infty \\ &= \left\|S^{-1}D^{-1}\left(I - DA^T(AD^2A^T)^{-1}AD\right)D\right\|_\infty. \end{aligned} \tag{5.12}$$

Note that $(I - DA^T(AD^2A^T)^{-1}AD)$ is a projection matrix onto the null space of AD; therefore, its ℓ_2-norm is bounded by 1. Using the elementary matrix norm inequality $\|P\|_\infty \leq n^{1/2}\|P\|_2$ for any $P \in R^{n\times n}$, we obtain the following sequence of inequalities:

$$\begin{aligned} &\left\|S^{-1}\left(I - A^T(AD^2A^T)^{-1}AD^2\right)\right\|_\infty \\ &= \left\|S^{-1}D^{-1}(I - DA^T(AD^2A^T)^{-1}AD)D\right\|_\infty \\ &\leq \left\|X^{-1/2}S^{-1/2}\right\|_\infty \left\|I - DA^T(AD^2A^T)^{-1}AD\right\|_\infty \left\|X^{1/2}S^{-1/2}\right\|_\infty \\ &\leq \max_{i=1,2,\dots,n}\frac{1}{\sqrt{x_is_i}}\, n^{1/2} \max_{i=1,2,\dots,n}\sqrt{\frac{x_i}{s_i}} \\ &\leq n^{1/2}\frac{1}{\sqrt{\gamma_0\mu}}\max_{i=1,2,\dots,n}\frac{x_i}{\sqrt{x_is_i}} \\ &\leq \left(\frac{1}{\mu}\right)\frac{n^{1/2}\mathcal{C}(d)}{\gamma_0}\left(\mathcal{C}(d) + n\mu/\|d\|\right), \end{aligned} \tag{5.13}$$

where we used $D^2 = XS^{-1}$, $x_is_i \geq \gamma_0\mu$, and (2.13).

If we consider the reciprocal of the right-hand side of expression (5.5), we obtain

$$\begin{aligned} &\left\|\left[S^{-1}\left(I - A^T(AD^2A^T)^{-1}AD^2\right) \quad - S^{-1}A^T(AD^2A^T)^{-1}\right]\right\|_\infty \\ &\leq \left\|S^{-1}\left(I - A^T(AD^2A^T)^{-1}AD^2\right)\right\|_\infty + \left\|S^{-1}A^T(AD^2A^T)^{-1}\right\|_\infty \\ &\leq \left(\frac{1}{\mu}\right)\frac{n^{1/2}\mathcal{C}(d)}{\gamma_0}\left(\mathcal{C}(d) + \{n\mu\}\|d\|\right) + \left(\frac{1}{\mu}\right)\frac{2\|d\|\mathcal{C}(d)}{\gamma_0}\left(\mathcal{C}(d) + n\mu/\|d\|\right)\chi(A), \end{aligned} \tag{5.14}$$

which follows from (5.11) and (5.13). Therefore, (5.14) implies

$$\frac{1}{\left\| [S^{-1} \left(I - A^T (AD^2 A^T)^{-1} AD^2 \right) \quad - S^{-1} A^T (AD^2 A^T)^{-1}] \right\|_\infty} \geq f(\mu) \stackrel{\text{def}}{=} \frac{\gamma_0 \mu}{\mathcal{C}(d) \left(n^{1/2} + 2\|d\| \chi(A) \right) [\mathcal{C}(d) + n\mu/\|d\|]}. \tag{5.15}$$

It is easy to verify our claims both that f is monotone increasing in μ and that $f(0) = 0$. □

Note that Proposition 5.1 guarantees only that the point $(\tilde{x}, \tilde{y}, \tilde{s})$ is feasible for the perturbed problem. To initiate a feasible path-following interior-point method, we need to impose additional conditions to obtain a strictly feasible point for the perturbed problem that lies in some neighborhood of the central path. For example, in the proof, we imposed only the condition $(\tilde{x}, \tilde{s}) \geq 0$. Strict positivity of $\tilde{x}$ and $\tilde{s}$ could be ensured by imposing the following condition, for some $\epsilon \in (0, 1)$:

$$x_i + \Delta x_i \geq \epsilon x_i, \qquad s_i + \Delta s_i \geq \epsilon s_i \quad \forall\ i = 1, 2, \ldots, n. \tag{5.16}$$

Equivalently, we can replace the necessary and sufficient condition $\|S^{-1} \Delta s\|_\infty \leq 1$ in (5.8) by the condition $(\epsilon - 1)e \leq S^{-1} \Delta s \leq (1 - \epsilon)e$, that is,

$$\|S^{-1} \Delta s\|_\infty \leq 1 - \epsilon,$$

in the proof of Proposition 5.1. With this requirement, we obtain the following bounds:

$$\epsilon x_i \leq \tilde{x}_i \leq (2 - \epsilon) x_i, \qquad \epsilon s_i \leq \tilde{s}_i \leq (2 - \epsilon) s_i. \tag{5.17}$$

Note that if $(\Delta x, \Delta y, \Delta s)$ is the Newton step given by (5.2), then $\Delta x_i \Delta s_i \leq 0$ for all $i = 1, 2, \ldots, n$. First, consider the case $\Delta x_i \geq 0$, which implies $\tilde{x}_i \geq x_i$. We have from (5.17) that

$$\tilde{x}_i \tilde{s}_i \geq x_i \tilde{s}_i \geq \epsilon x_i s_i. \tag{5.18}$$

A similar set of inequalities holds for the case $\Delta s_i \geq 0$. Thus, if we define $\tilde{\mu} = \tilde{x}^T \tilde{s}/n$, we obtain

$$\tilde{\mu} \geq \epsilon \mu. \tag{5.19}$$

Note that by (5.6), we already have $\tilde{\mu} \leq \mu$. With this observation, we can relate the neighborhood in which the original iterate (x, y, s) lies to the one in which the adjusted point $(\tilde{x}, \tilde{y}, \tilde{s})$ lies.

PROPOSITION 5.3. *Let (x, y, s) be a strictly feasible point for d, and suppose that $\Delta d = (0, \Delta b, \Delta c)$ and $\epsilon \in (0, 1)$ are given. Consider the Newton step of Proposition 5.1 and the adjusted point $(\tilde{x}, \tilde{y}, \tilde{s})$ of (5.4). If*

$$\left\| \begin{bmatrix} \Delta c \\ \Delta b \end{bmatrix} \right\|_\infty \leq \frac{1 - \epsilon}{\left\| [S^{-1} (I - A^T (AD^2 A^T)^{-1} AD^2) \quad - S^{-1} A^T (AD^2 A^T)^{-1}] \right\|_\infty}, \tag{5.20}$$

with D defined in (5.3), then $(\tilde{x}, \tilde{y}, \tilde{s})$ is strictly feasible for $d + \Delta d$ with $\tilde{\mu} \leq \mu$. Moreover, if $(x, y, s) \in \mathcal{N}_{-\infty}(\gamma_0)$ for the original problem with $\gamma_0 \in (0, 1)$, then $(\tilde{x}, \tilde{y}, \tilde{s})$ satisfies $(\tilde{x}, \tilde{y}, \tilde{s}) \in \bar{\mathcal{N}}_{-\infty}(\epsilon \gamma_0)$.

Proof. It suffices to prove the final statement of the theorem. If we assume that $(x, y, s) \in \mathcal{N}_{-\infty}(\gamma_0)$, then, using (5.18) and (5.6), we have

$$\tilde{x}_i\tilde{s}_i \geq \epsilon x_i s_i \geq \epsilon\gamma_0\mu \geq \epsilon\gamma_0\tilde{\mu}, \tag{5.21}$$

which implies that $(\tilde{x}, \tilde{y}, \tilde{s}) \in \bar{\mathcal{N}}_{-\infty}(\epsilon\gamma_0)$, as required. □

We now have all the tools to be able to prove results like those of section 4. Suppose that the iterates of the original problem lie in a wide neighborhood with parameter γ_0. For convenience we define

$$\|\Delta d\|_\infty \stackrel{\text{def}}{=} \left\| \begin{bmatrix} \Delta b \\ \Delta c \end{bmatrix} \right\|_\infty = \max(\|\Delta b\|_\infty, \|\Delta c\|_\infty). \tag{5.22}$$

We also define the relative perturbation measure δ_d as follows:

$$\delta_d \stackrel{\text{def}}{=} \frac{\|\Delta d\|_\infty}{\|d\|}. \tag{5.23}$$

Note from (4.7) and (4.9) that

$$\delta_d = \max\left(\frac{\|\Delta b\|_\infty}{\|d\|}, \frac{\|\Delta c\|_\infty}{\|d\|}\right) \leq \max(\delta_b, \delta_c) \leq \delta_{bc}.$$

Hence, it is easy to compare results such as Proposition 5.4 below, which obtain a lower bound on μ in terms of δ_d, to similar results in preceding sections.

Note that Theorem 5.2 provides a lower bound $f(\mu)$ on the term on the right-hand side of (5.5). Therefore, combining this result with Proposition 5.3, we conclude that a sufficient condition for the perturbation Δd to satisfy (5.20) is that $\|\Delta d\|_\infty$ is bounded above by the lower bound (5.15) multiplied by $(1 - \epsilon)$, that is,

$$\|\Delta d\|_\infty \leq \frac{(1-\epsilon)\gamma_0\mu}{\mathcal{C}(d)\left(n^{1/2} + 2\|d\|\chi(A)\right)\left(\mathcal{C}(d) + n\mu/\|d\|\right)},$$

which by rearrangement yields

$$\mu \geq \frac{\mathcal{C}(d)^2\|\Delta d\|_\infty(n^{1/2} + 2\|d\|\chi(A))}{(1-\epsilon)\gamma_0 - n\mathcal{C}(d)\|\Delta d\|_\infty(n^{1/2} + 2\|d\|\chi(A))/\|d\|}, \tag{5.24}$$

provided that the denominator of this expression is positive. To ensure the latter condition, we impose the following bound on δ_d:

$$\delta_d = \frac{\|\Delta d\|_\infty}{\|d\|} < \frac{(1-\epsilon)\gamma_0}{n\mathcal{C}(d)(n^{1/2} + 2\|d\|\chi(A))}. \tag{5.25}$$

Indeed, when this bound is not satisfied, the perturbation may be so large that the adjusted point $(\tilde{x}, \tilde{y}, \tilde{s})$ may not be feasible for $d + \Delta d$ no matter how large we choose μ for the original iterate (x, y, s).

We now state and prove a result like Proposition 4.3 that gives a condition on $\|\Delta d\|_\infty$ and μ sufficient to ensure that the adjusted point $(\tilde{x}, \tilde{y}, \tilde{s})$ lies within a wide neighborhood of the central path for the perturbed problem.

PROPOSITION 5.4. *Let γ and γ_0 be given with $0 < \gamma < \gamma_0 < 1$, and suppose that ξ satisfies $\xi \in (0, \gamma_0 - \gamma)$. Assume that δ_d satisfies*

$$\delta_d \leq \frac{\gamma_0 - \gamma - \xi}{n\mathcal{C}(d)(n^{1/2} + 2\|d\|\chi(A))}. \tag{5.26}$$

Suppose that $(x, y, s) \in \mathcal{N}_{-\infty}(\gamma_0)$ *for the original problem, and let* $(\tilde{x}, \tilde{y}, \tilde{s})$ *be as defined in* (5.4). *Then if*

$$\mu \geq \frac{\|d\|}{\xi} \mathcal{C}(d)^2 \delta_d \left(n^{1/2} + 2\|d\|\chi(A) \right) \stackrel{\text{def}}{=} \mu_5^*, \tag{5.27}$$

we have $(\tilde{x}, \tilde{y}, \tilde{s}) \in \bar{\mathcal{N}}_{-\infty}(\gamma)$.

Proof. Setting $\epsilon = \gamma/\gamma_0$, we note that (5.26) satisfies condition (5.25), and so the Newton step adjustment yields a strictly feasible point for the perturbed problem. By the argument preceding the proposition, (5.24) gives a condition sufficient for the resulting iterate to lie in $\bar{\mathcal{N}}_{-\infty}(\gamma)$ by Proposition 5.3 since $\gamma = \epsilon\gamma_0$ by the hypothesis. However, (5.26) implies that the denominator of (5.24) is bounded below by ξ; hence, (5.24) is implied by (5.27), as required. ∎

The usual argument can now be used to show that a long-step path-following method requires

$$\mathcal{O}\left(n \log \left(\frac{1}{\epsilon} \mathcal{C}(d)^2 \delta_d \left(n^{1/2} + \|d\|\chi(A) \right) \right) \right) \quad \text{iterations} \tag{5.28}$$

to converge from the warm-start point to a point that satisfies (3.9).

5.2. Perturbations in *A*. In this section, we also allow perturbations in A (i.e., we let $\Delta d = (\Delta A, \Delta b, \Delta c)$) and propose a Newton step correction strategy to recover warm-start points for the perturbed problem from the iterates of the original problem.

The underlying idea is the same as that in section 5.1. Given a strictly feasible iterate $(x, y, s) \in \mathcal{N}_{-\infty}(\gamma_0)$ for the original problem, we apply Newton's method to recover a feasible point for the perturbed problem by keeping the pairwise products $x_i s_i$ fixed. As in section 4.3, we impose an upper bound on μ that excludes values of μ that are not likely to yield an adjusted starting point with significantly better complexity than a cold-start strategy. In particular, we assume that μ satisfies (4.32) for some $\zeta > 1$. Let

$$\bar{A} \stackrel{\text{def}}{=} A + \Delta A. \tag{5.29}$$

Given a feasible iterate (x, y, s) for the original problem, the Newton step correction is then given by the solution to

$$\begin{array}{rcccl} & \bar{A}\Delta x & & = & \Delta b - \Delta A x, \\ \bar{A}^T \Delta y & & + \quad \Delta s & = & \Delta c - \Delta A^T y, \\ & S\Delta x & + \quad X\Delta s & = & 0. \end{array} \tag{5.30}$$

Under the assumption that $\bar{A}$ has full row rank, the solution to (5.30) is then given by

$$\Delta y = (\bar{A}D^2\bar{A}^T)^{-1}(\bar{A}D^2(\Delta c - \Delta A^T y) + \Delta b - \Delta A x), \tag{5.31a}$$

$$\Delta s = \Delta c - \Delta A^T y - \bar{A}^T \Delta y, \tag{5.31b}$$

$$\Delta x = -S^{-1} X \Delta s, \tag{5.31c}$$

where $D^2 = S^{-1}X$ as in (5.3).

By a similar argument, a necessary and sufficient condition to have strictly feasible iterates for the perturbed problem is

$$\|S^{-1}\Delta s\|_\infty \leq 1 - \epsilon \quad \text{for some } \epsilon \in (0, 1). \tag{5.32}$$

By Proposition 5.3, the duality gap of the resulting iterate will also be smaller than that of the original iterate. We will modify the analysis in section 5 to incorporate the perturbation in A and will refer to the previous analysis without repeating the propositions.

Using (5.31), we get

$$S^{-1}\Delta s = S^{-1}(I - \bar{A}^T(\bar{A}D^2\bar{A}^T)^{-1}\bar{A}D^2)(\Delta c - \Delta A^T y) \\ - S^{-1}\bar{A}^T(\bar{A}D^2\bar{A}^T)^{-1}(\Delta b - \Delta A x).$$

Therefore, $\|S^{-1}\Delta s\|_\infty$ is bounded above by

$$\left\| [S^{-1}(I - \bar{A}^T(\bar{A}D^2\bar{A}^T)^{-1}\bar{A}D^2) \quad - S^{-1}\bar{A}^T(\bar{A}D^2\bar{A}^T)^{-1}] \right\|_\infty \left\| \begin{bmatrix} \Delta c - \Delta A^T y \\ \Delta b - \Delta A x \end{bmatrix} \right\|_\infty .$$

By Theorem 5.2, the first term in this expression is bounded above by $1/\bar{f}(\mu)$, where $\bar{f}(\mu)$ is obtained from $f(\mu)$ in (5.15) by replacing $\chi(A)$ by $\chi(\bar{A})$. For the second term, we extend the definition in (5.22) to account for the perturbations in A as follows,

$$\|\Delta d\|_\infty \stackrel{\text{def}}{=} \max(\|\Delta b\|_\infty, \|\Delta c\|_\infty, \|\Delta A\|_\infty, \|\Delta A^T\|_\infty), \tag{5.33}$$

and continue to define δ_d as in (5.23). We obtain that

$$\begin{aligned} &\left\| \begin{bmatrix} \Delta c - \Delta A^T y \\ \Delta b - \Delta A x \end{bmatrix} \right\|_\infty \\ &\le \max\{\|\Delta c\|_\infty + \|\Delta A^T\|_\infty \|y\|_\infty, \|\Delta b\|_\infty + \|\Delta A\|_\infty \|x\|_\infty\} \\ &\le \max\{\|\Delta d\|_\infty(1 + \|y\|_\infty), \|\Delta d\|_\infty(1 + \|x\|_\infty)\} \\ &\le \|\Delta d\|_\infty(1 + \zeta\mathcal{C}(d)^2) \\ &\le 2\|\Delta d\|_\infty \zeta \mathcal{C}(d)^2, \qquad (5.34) \end{aligned}$$

where we used (5.33), (4.33), $\zeta > 1$, and $\mathcal{C}(d) \ge 1$ to derive the inequalities. By combining the two upper bounds we obtain

$$\|S^{-1}\Delta s\|_\infty \le \left(\frac{1}{\mu}\right) \frac{1}{\gamma_0} 2\zeta\mathcal{C}(d)^3 \left(n^{1/2} + 2\|d\|\chi(\bar{A})\right) \left(\mathcal{C}(d) + n\mu/\|d\|\right) \|\Delta d\|_\infty. \tag{5.35}$$

Therefore, a sufficient condition to ensure (5.32) is obtained by requiring the upper bound in (5.35) to be less than $1 - \epsilon$. Rearranging the resulting inequality yields a lower bound on μ,

$$\mu \ge \frac{2\zeta\mathcal{C}(d)^4 \left(n^{1/2} + 2\|d\|\chi(\bar{A})\right) \|\Delta d\|_\infty}{\gamma_0(1 - \epsilon) - 2\zeta n\mathcal{C}(d)^3 \left(n^{1/2} + 2\|d\|\chi(\bar{A})\right) \|\Delta d\|_\infty / \|d\|}, \tag{5.36}$$

provided that the denominator is positive, which is ensured by the condition

$$\delta_d = \frac{\|\Delta d\|_\infty}{\|d\|} < \frac{\gamma_0(1 - \epsilon)}{2\zeta n\mathcal{C}(d)^3 \left(n^{1/2} + 2\|d\|\chi(\bar{A})\right)}. \tag{5.37}$$

The proof of the following result is similar to that of Proposition 5.4.

Proposition 5.5. *Let γ and γ_0 be given with $0 < \gamma < \gamma_0 < 1$, and suppose that ξ satisfies $\xi \in (0, \gamma_0 - \gamma)$. Assume that Δd satisfies*

$$\delta_d \le \frac{\gamma_0 - \gamma - \xi}{2\zeta n\mathcal{C}(d)^3 \left(n^{1/2} + 2\|d\|\chi(\bar{A})\right)}. \tag{5.38}$$

Suppose that $(x, y, s) \in \mathcal{N}_{-\infty}(\gamma_0)$ *and that* $(\tilde{x}, \tilde{y}, \tilde{s})$ *is the adjusted point defined in* (5.4). *Then we have* $(\tilde{x}, \tilde{y}, \tilde{s}) \in \bar{\mathcal{N}}_{-\infty}(\gamma)$, *provided that*

$$\mu \geq \frac{\|d\|}{\xi} 2\zeta \mathcal{C}(d)^4 \delta_d \left(n^{1/2} + 2\|d\|\chi(\bar{A})\right) \stackrel{\text{def}}{=} \mu_6^*. \tag{5.39}$$

The usual argument can be used again to show that a long-step path-following method requires

$$\mathcal{O}\left(n \log\left(\frac{1}{\epsilon}\mathcal{C}(d)^4 \delta_d \left(n^{1/2} + \|d\|\chi(\bar{A})\right)\right)\right) \quad \text{iterations} \tag{5.40}$$

to converge from the warm-start point to a point that satisfies (3.9).

6. Comparison of the strategies. Here we comment on the relationship between the Newton step correction strategy of section 5, the least-squares correction strategy of section 4, and a weighted least-squares approach described below. In particular, we discuss the effects of weighting in different situations and show that the strategies of sections 4 and 5 jointly retain all the benefits of the weighted least-squares strategy. The weighted least-squares strategy is discussed in section 6.1, relationships between the strategies in various circumstances are discussed in section 6.2, and some numerical results are presented in section 6.3.

6.1. Weighted least-squares strategy. When the data perturbations are confined to the vectors b and c, we can define $n \times n$ positive diagonal matrices Σ and Λ and solve the following variants on the subproblems (4.2):

$$\min \|\Sigma \Delta x\| \quad \text{s.t.} \quad A\Delta x = \Delta b, \tag{6.1a}$$

$$\min \|\Lambda \Delta s\| \quad \text{s.t.} \quad A^T \Delta y + \Delta s = \Delta c. \tag{6.1b}$$

The solutions are as follows:

$$\Delta x_\Sigma = \Sigma^{-2} A^T (A \Sigma^{-2} A^T)^{-1} \Delta b, \tag{6.2a}$$

$$\Delta y_\Lambda = (A \Lambda^2 A^T)^{-1} A \Lambda^2 \Delta c, \tag{6.2b}$$

$$\Delta s_\Lambda = (I - A^T (A \Lambda^2 A^T)^{-1} A \Lambda^2) \Delta c. \tag{6.2c}$$

When $\Sigma = \Lambda = I$, we recover the least-squares solutions (4.4). Alternative scalings include the following:

$$\Sigma = X^{-1}, \quad \Lambda = S^{-1} \tag{6.3}$$

and

$$\Sigma = D^{-1} = X^{-1/2} S^{1/2}, \quad \Lambda = D = X^{1/2} S^{-1/2}. \tag{6.4}$$

The second scaling is of particular interest, as a comparison of (6.2) with the substitutions (6.4) yields corrections quite similar to (5.2). The difference arises from the fact that the Newton step contains an additional condition that couples Δx and Δs; namely, $X\Delta s + S\Delta x = 0$. If the perturbation is confined to b (that is, $\Delta c = 0$), then the correction in x given by the Newton step scheme (5.2) is the same as the one obtained from (6.2a) with Σ as defined in (6.4). In this case, the Newton step correction reduces to a weighted least-squares correction.

In fact, Mitchell and Todd [9] use a similar weighted least-squares strategy in a column-generation framework. Assuming that $\tilde{x}$ is feasible for problem (P), a new column a is introduced so that $(\tilde{x}, 0)$ is feasible for the new problem. Then, in order to obtain a strictly feasible point to restart the interior-point algorithm, a step is taken in the direction given by $(d, 1)$, where d solves (6.1a) with $\Sigma = X^{-1}$ and $\Delta b = -a$. The reader is also referred to [7] and the references therein for consideration of other directions.

The weighted least-squares approach suffers some disadvantages relative to the approaches of sections 4 and 5. When the weighting matrices Σ and Λ depend on x and s, the solutions (6.2) must be computed afresh for each candidate initial point, whereas for unweighted least-squares, a single solution suffices (4.4). Unlike the Newton step, the weighted least-squares approach does not guarantee a smaller value of μ than at the initial point.

6.2. Relating the strategies. We now focus on the primal correction Δx obtained from the least-squares, weighted least-squares, and Newton step strategies. In deciding how to choose Δx, we need to recover primal feasibility while ensuring that our choice of Δx does not unnecessarily compromise the positivity of x. The strategy of section 4 minimizes the norm $\|\Delta x\|$, while in section 5 our analysis of the Newton step strategy used the quantity $\|X^{-1}\Delta x\|_\infty$ (and its dual counterpart) to bound the size of the perturbations that could be corrected by this strategy. The weighted least-squares approach in which we aim to minimize $\|X^{-1}\Delta x\|$ explicitly is a natural alternative to both these strategies. We now discuss this strategy in the case in which the perturbation is confined to b, that is, $\Delta c = 0$ and $\Delta A = 0$.

Suppose we partition A as $[B \quad N]$, where B represents the "basic" columns for the original problem—the columns i such that $x_i^* > 0$ for some solution x^* of problem (P). Consider first the case in which Δb does not lie in the range of B. Since A is assumed to have full row rank, there exist v_B and $v_N \neq 0$ such that

$$Bv_B + Nv_N = \Delta b. \tag{6.5}$$

From (6.2a), we have

$$\Delta x = \Sigma^{-2} A^T w, \quad \text{where} \quad A\Sigma^{-2}A^T w = \Delta b = Bv_B + Nv_N.$$

It follows that

$$B\Sigma_B^{-2}B^T w + N\Sigma_N^{-2}N^T w = Bv_B + Nv_N,$$

where Σ_B and Σ_N are the appropriate partitions of Σ. Since Δb does not lie in the range space of B, we have $\|\Delta x_N\| = \|\Sigma_N^{-2}N^T w\| \geq \alpha$ for some $\alpha > 0$. If Σ is defined as in (6.3), we have from $\|x_N\| = O(\mu)$ that

$$\|X_N^{-1}\Delta x_N\| \geq \frac{\alpha}{\max(x_N)_i} \geq \frac{\bar{\alpha}}{\mu}$$

for some constant $\bar{\alpha} > 0$. A similar result applies when we choose the scaling as in (6.4), and thus the Newton step strategy will also yield a large value of $\|X_N^{-1}\Delta x_N\|$ under these circumstances. Clearly the unweighted least-squares strategy ($\Sigma = I$) also will yield a correction Δx_N with $\|\Delta x_N\| \geq \alpha$, so in this case too we may have to back off to a much earlier iterate of the interior-point method for the original problem before the correction strategy yields a feasible starting point. The point here is that, even though we are minimizing $\|X^{-1}\Delta x\|$ explicitly in the weighted strategy, we cannot expect Δx_N to be appreciably smaller than in the other strategies when Δb does not lie in the range of B.

If Δb *does* lie in the range space of B, then there are two cases. First, consider the case in which B has full column rank. The analysis of Yıldırım and Todd [17] can be modified to show that the weighted least-squares correction using $\Sigma = X^{-1}$ converges to $(v^T, 0)^T$ as μ tends to 0, where v is the (unique) vector such that $Bv = \Delta b$. In [17], it is shown that the strategy of section 5 also converges to $(v^T, 0)^T$ as $\mu \downarrow 0$, so that the Newton step strategy gives asymptotically the same results as the weighted least-squares strategy in this case. Second, if B does not have full column rank, then it is shown in [17] that the Newton step strategy yields a correction Δx for which $\|X^{-1}\Delta x\|$ is well-behaved, in the sense that it remains bounded asymptotically as μ tends to 0. The analysis in [17] is based on a technical lemma (Lemma 5.1) which holds for the Newton step but does not necessarily hold for the weighted least-squares correction with $\Sigma = X^{-1}$. Hence, when Δb lies in the range space of B, it appears that the Newton step correction behaves at least as well as the weighted least-squares strategy that uses the scaling $\Sigma = X^{-1}$, at least asymptotically as $\mu \downarrow 0$.

Let us examine further the case in which B has full column rank and Δb is in the range space of B, but the perturbation is *not* necessarily small. In the notation of (6.5), we have that there exists a unique vector v_B such that

$$Bv_B = \Delta b. \tag{6.6}$$

If an interior-point method has been used to solve the original problem, and if we are close to the solution obtained with such a method, then the basic components (those contained in the subvector x_B) will be bounded away from zero, while the components of x_N will have size $O(\mu)$. By setting $v = (v_B^T, 0)^T$, we note that $Av = Bv_B = \Delta b$, so that v is feasible in both (4.2a) and (6.1a). In fact, we would expect v to be near-optimal in (6.1a) because it would yield an objective of size $O(\|\Delta b\|)$, and both weighting schemes (6.3) and (6.4) discourage solutions in which Δx_N is appreciably different from zero. As discussed above, we would also expect v to be near-optimal for the Newton step strategy. The plain least-squares strategy, on the other hand, does not discriminate between B and N components and may give a solution in which Δx_N is not especially small.

If it happens that $x+v \geq 0$, then $x+v$ will lie very close to a primal solution of the perturbed problem whenever x lies near a solution of the original problem. Hence, we would expect the weighted least-squares strategy to perform very well from an initial point that is an advanced iterate for the original problem, provided that the perturbation Δb can be accommodated without a change of basis. The plain least-squares method will usually perform less well, because it will be necessary to choose an initial point from the iterates for the original problem with an appreciably larger value of μ, to ensure that $x_N + \Delta x_N$ is nonnegative. In general, we would expect to need a μ that is bounded below by a multiple of $\|\Delta b\|$ in the plain least-squares strategy, whereas much smaller values of μ may be permissible in the other two strategies.

In the more interesting case in which the perturbation is large enough to force a change of basis, it is not at all clear that the weighted least-squares and Newton step strategies retain their advantage over plain least-squares. To be specific, if we do *not* have $x_B + \Delta x_B \geq 0$, it will be necessary to back up to an initial point in which the components of x_N are large enough to allow some of the perturbation to be absorbed by the Δx_N components

The need for backing up sufficiently far along the central path can also be motivated with reference to the dual problem (D) and to the geometry of the central path. When the perturbation is large enough to change the basis, the dual solution will usually change to a different vertex of its feasible polytope. Consequently, the central paths $\mathcal{P}$ (see (2.7)) for the original and perturbed problems (and therefore the neighborhoods $\mathcal{N}_2$ and $\bar{\mathcal{N}}_2$, and $\mathcal{N}_{-\infty}$ and $\bar{\mathcal{N}}_{-\infty}$) diverge significantly as $\mu \downarrow 0$. For large μ, however, the paths and neighborhood are quite similar for the original and perturbed problems. We need to choose μ sufficiently large that the neighborhoods are broad enough, and have a wide enough overlap, to ensure that the adjusted point $(x + \Delta x, y + \Delta y, s + \Delta s)$ lies inside the appropriate neighborhood for the perturbed problem.

We conclude that in the case of a perturbation to b, the strategies of sections 4 and 5 capture the potential advantages of using a weighted least-squares correction. Similar arguments can be made for the dual-only scaling, due to the symmetry between the primal and the dual problems.

6.3. Numerical results. We illustrate the remarks of the previous subsection—particularly the remarks about the relative performance of the strategies when the primal perturbation is and is not large enough to force a change of basis—with the following simple problem in R^2:

$$\min x_1 + x_2 \ \text{ s.t. } x_1 - x_2 = \epsilon, \ \ x \geq 0, \tag{6.7}$$

where $\epsilon > 0$ is a constant. We set $\epsilon = 10^{-2}$ throughout this section. This problem is well-conditioned (a large perturbation to the data is needed to make it infeasible) and has solution $x = (\epsilon, 0)^T$. Its dual is

$$\max \epsilon y \ \text{ s.t. } \begin{bmatrix} 1 \\ -1 \end{bmatrix} y + s = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \ \ s \geq 0.$$

Since $\epsilon > 0$, the dual solution is $s^* = (0, 2)^T$, $y^* = 1$. It is easy to show that the central path defined by (2.6) is as follows:

$$x(\mu) = \left(\frac{\mu + \epsilon}{2} + \frac{1}{2}\sqrt{\epsilon^2 + \mu^2}, \frac{\mu - \epsilon}{2} + \frac{1}{2}\sqrt{\epsilon^2 + \mu^2} \right)^T, \tag{6.8a}$$

$$s_1(\mu) = \frac{\mu}{x_1(\mu)}, \qquad s_2(\mu) = \frac{\mu}{x_2(\mu)}, \qquad y(\mu) = 1 - s_1(\mu). \tag{6.8b}$$

Note that for $\mu \gg \epsilon$ we have that

$$x(\mu) \approx (\mu, \mu)^T, \qquad s(\mu) = (1, 1)^T, \qquad y(\mu) \approx 0. \tag{6.9}$$

We plot $(x(\mu), y(\mu), s(\mu))$ for various values of μ in Table 6.1.

We consider perturbations to the right-hand side ϵ of the equality constraint in (6.7); namely, $\Delta b = \beta$. For $\beta > -\epsilon$, the solution of the perturbed primal can be attained without a change of basis. The solution of the perturbed primal becomes $x = (\epsilon + \beta, 0)^T$, while the solution of the dual remains unchanged. For $\beta < -\epsilon$, however, the solution of the perturbed primal becomes $x = (0, -(\beta + \epsilon))^T$, while the solution of the dual becomes $s^* = (2, 0)^T$, $y^* = 1$. Even for the latter case, we still have for $\mu \gg -(\beta + \epsilon)$ that the limits (6.9) hold, indicating that central paths for the original and perturbed problems are quite similar for larger values of μ.

TABLE 6.1
Central path points for (6.7), *with* $\epsilon = 10^{-2}$.

μ	x	y	s
1.e-5	(1.0e-2,5.0e-6)	1.0e0	(1.0e-3, 2.0e0)
1.e-4	(1.0e-2, 5.0e-5)	9.9e-1	(1.0e-2, 2.0e0)
1.e-3	(1.1e-2, 5.2e-4)	9.0e-1	(9.5e-2, 1.9e0)
5.e-3	(1.3e-2, 3.1e-3)	6.2e-1	(3.8e-1, 1.6e0)
1.e-2	(1.7e-2, 7.1e-3)	4.1e-1	(5.9e-1, 1.4e0)
2.e-2	(2.6e-2, 1.6e-2)	2.4e-1	(7.6e-1, 1.2e0)
1.e-1	(1.1e-1, 9.5e-2)	5.0e-2	(9.5e-1, 1.0e0)
5.e-1	(5.1e-1, 5.0e-1)	1.0e-2	(9.9e-1, 1.0e0)

By solving (4.4), we find that the plain least-squares adjustment is

$$\Delta x_{LS} = (\beta/2, -\beta/2), \qquad \Delta y_{LS} = 0, \qquad \Delta s_{LS} = 0$$

(independently of x). By substituting (6.3) into (6.2), we obtain the following weighted least-squares adjustments:

$$\Delta x_{WLS} = \frac{\beta}{x_1^2 + x_2^2} \begin{bmatrix} x_1^2 \\ -x_2^2 \end{bmatrix}, \qquad \Delta y_{WLS} = 0, \qquad \Delta s_{WLS} = 0.$$

Finally, we obtain the Newton step adjustment by substituting (6.4) into (6.2):

$$\Delta x_{NS} = \frac{\beta}{(\frac{x_1}{s_1} + \frac{x_2}{s_2})} \begin{bmatrix} x_1/s_1 \\ -x_2/s_2 \end{bmatrix}, \quad \Delta s_{NS} = -X^{-1} S \Delta x_{NS}, \quad A^T \Delta y_{NS} + \Delta s_{NS} = 0.$$

The primal corrections Δx_{WLS} and Δx_{NS} coincide when (x, y, s) is on the central path, confirming our previous observation regarding the asymptotic coincidence. The dual correction is of course different for the two strategies.

In Tables 6.2, 6.3, and 6.4 we indicate the effects of the plain least-squares, weighted least-squares, and Newton step adjustments for perturbations $\beta = -0.1\epsilon$, $\beta = -\epsilon$, and $\beta = -10\epsilon$, respectively. We tabulate the following quantities against a selection of values of μ:

- The values of μ obtained after each of the adjustment strategies, that is, $\mu(x + \Delta x, s + \Delta s) = (x + \Delta x)^T (s + \Delta s)/2$, provided that all components of $(x + \Delta x, s + \Delta s)$ are positive. If not, we enter "-".
- The centrality indicators $(x + \Delta x)_i (s + \Delta s)_i / \mu(x + \Delta x, s + \Delta s)$, $i = 1, 2$. If any components of $(x + \Delta x, s + \Delta s)$ are nonpositive, we enter "-".

A good starting point for the perturbed problem is one for which the centrality indicators are not too far from 1 in all components (all greater than 10^{-1}, say), while the value of $\mu(x + \Delta x, s + \Delta s)$ is as small as possible.

In Table 6.2, the perturbation is small enough that a basis change is not needed and, as expected, the weighted least-squares and Newton adjustments perform well. Even when the central path point with $\mu = 10^{-5}$ is used as the basis for adjustment, well-centered starting points with small duality gaps are obtained from both strategies. The plain least-squares approach does not give particularly good adjusted points when applied at the central path points with $\mu = 10^{-5}$ or $\mu = 10^{-4}$, but becomes comparable for higher values of μ (that is, when the initial point is taken to be slightly further back along the central path).

In Table 6.3, where the perturbation is large enough to make the problem degenerate, the performances of the plain and weighted least-squares adjustment strategies

TABLE 6.2
Centrality of adjusted points for various μ, with $\epsilon = 10^{-2}$ and $\beta = -10^{-3}$.

μ	μ_{LS}	Centrality	μ_{WLS}	Centrality	μ_{NS}	Centrality
1.e-5	5.1e-4	(1.9e-2, 2.0e0)	9.5e-6	(9.5e-1, 1.1e0)	1.0e-5	(9.9e-1, 1.0e0)
1.e-4	6.0e-4	(1.6e-1, 1.8e0)	9.5e-5	(9.5e-1, 1.1e0)	1.0e-4	(1.0e0, 1.0e0)
1.e-3	1.5e-3	(6.6e-1, 1.3e0)	9.5e-4	(9.5e-1, 1.1e0)	1.0e-3	(1.0e0, 1.0e0)
5.e-3	5.3e-3	(9.1e-1, 1.1e0)	4.9e-3	(9.5e-1, 1.0e0)	5.0e-3	(1.0e0, 1.0e0)
1.e-2	1.0e-2	(9.5e-1, 1.0e0)	9.9e-3	(9.6e-1, 1.0e0)	1.0e-2	(1.0e0, 1.0e0)
2.e-2	2.0e-2	(9.8e-1, 1.0e0)	2.0e-2	(9.8e-1, 1.0e0)	2.0e-2	(1.0e0, 1.0e0)
1.e-1	1.0e-1	(1.0e0, 1.0e0)	1.0e-1	(1.0e0, 1.0e0)	1.0e-1	(1.0e0, 1.0e0)
5.e-1	5.0e-1	(1.0e0, 1.0e0)	5.0e-1	(1.0e0, 1.0e0)	5.0e-1	(1.0e0, 1.0e0)

TABLE 6.3
Centrality of adjusted points for various μ, with $\epsilon = 10^{-2}$ and $\beta = -10^{-2}$.

μ	μ_{LS}	Centrality	μ_{WLS}	Centrality	μ_{N}	Centrality
1.e-5	5.0e-3	(1.0e-3, 2.0e0)	5.0e-6	(1.0e-3, 2.0e0)	5.0e-6	(2.0e-3, 2.0e0)
1.e-4	5.1e-3	(1.0e-2, 2.0e0)	5.0e-5	(1.0e-2, 2.0e0)	5.0e-5	(2.0e-2, 2.0e0)
1.e-3	5.5e-3	(9.5e-2, 1.9e0)	5.5e-4	(9.5e-2, 1.9e0)	5.5e-4	(1.9e-1, 1.8e0)
5.e-3	8.1e-3	(3.8e-1, 1.6e0)	3.6e-3	(3.8e-1, 1.6e0)	3.6e-3	(6.6e-1, 1.3e0)
1.e-2	1.2e-2	(5.9e-1, 1.4e0)	8.5e-3	(5.9e-1, 1.4e0)	8.5e-3	(8.8e-1, 1.1e0)
2.e-2	2.1e-2	(7.6e-1, 1.2e0)	1.9e-2	(7.6e-1, 1.2e0)	1.9e-2	(9.8e-1, 1.0e0)
1.e-1	1.0e-1	(9.5e-1, 1.0e0)	1.0e-1	(9.5e-1, 1.0e0)	9.8e-2	(1.0e0, 1.0e0)
5.e-1	5.0e-1	(9.9e-1, 1.0e0)	5.0e-1	(9.9e-1, 1.0e0)	5.0e-1	(1.0e0, 1.0e0)

TABLE 6.4
Centrality of adjusted points for various μ, with $\epsilon = 10^{-2}$ and $\beta = -10^{-1}$.

μ	μ_{LS}	Centrality	μ_{WLS}	Centrality	μ_{N}	Centrality
1.e-5	-	-	-	-	-	-
1.e-4	-	-	-	-	-	-
1.e-3	-	-	-	-	-	-
5.e-3	-	-	-	-	-	-
1.e-2	-	-	-	-	-	-
2.e-2	-	-	-	-	-	-
1.e-1	1.0e-1	(5.1e-1, 1.5e0)	9.8e-2	(4.9e-1, 1.5e0)	7.5e-2	(9.7e-1, 1.0e0)
5.e-1	5.0e-1	(9.0e-1, 1.1e0)	5.0e-1	(9.0e-1, 1.1e0)	5.0e-1	(1.0e0, 1.0e0)

are similar. For both strategies, we need to adjust from a central path point with value around $\mu = 10^{-3}$ or $\mu = 5 \times 10^{-3}$ to obtain a well-centered starting point for the perturbed problem. The Newton step correction strategy yields adjusted points that are better centered, but again we need to use the central path point with $\mu = 10^{-3}$ to obtain a reasonably adjusted point.

In Table 6.4, the perturbation is large enough to force a change of basis, and we see that all approaches behave in a similar fashion. To obtain a starting point that is well centered, we need to choose a central path point from the original problem with a duality gap of $\mu = 10^{-1}$.

7. Conclusions. We have described two schemes by which the iterates of an interior-point method applied to an LP instance can be adjusted to obtain starting points for a perturbed instance. We have derived worst-case estimates for the number of iterations required to obtain convergence from these warm starting points. These estimates depend chiefly on the size of the perturbation, on the conditioning of the original problem instance, and on a key property of the constraint matrix.

In future work, we plan to extend the techniques to infeasible interior-point methods and perform computational experiments to determine the practical usefulness of these techniques. We will also investigate extensions to wider classes of problems, such as convex quadratic programs and linear complementarity problems.

Acknowledgments. The first author would like to thank James Renegar for many helpful discussions and Mike Todd for pointing out the correct references to Dikin's result.

We are grateful to the referees for their careful reading of the first version of the paper and their perceptive comments, which prompted the addition of section 6.

REFERENCES

[1] S. A. Ashmanov, *Stability conditions for linear programming problems*, Comput. Math. Math. Phys., 21 (1981), pp. 40–49.

[2] I. I. Dikin, *On the speed of an iterative process*, Upravlyaemye Sistemi, 12 (1974), pp. 54–60.

[3] A. Forsgren, *On linear least-squares problems with diagonally dominant weight matrices*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 763–788.

[4] R. M. Freund, *A potential–function reduction algorithm for solving a linear program directly from an infeasible "warm start,"* Math. Programming, 52 (1991), pp. 441–466.

[5] J. Gondzio, *Warm start of the primal-dual method applied in the cutting-plane scheme*, Math. Programming, 83 (1998), pp. 125–143.

[6] J. Gondzio and J.-P. Vial, *Warm start and epsilon-subgradients in the cutting plane scheme for block-angular linear programs*, Comput. Optim. Appl., 14 (1999), pp. 17–36.

[7] J. E. Mitchell, *Karmarkar's Algorithm and Combinatorial Optimization Problems*, Ph.D. thesis, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, 1988.

[8] J. E. Mitchell and B. Borchers, *Solving real-world linear ordering problems using a primal–dual interior point cutting plane method*, Ann. Oper. Res., 62 (1996), pp. 253–276.

[9] J. E. Mitchell and M. J. Todd, *Solving combinatorial optimization problems using Karmarkar's algorithm*, Math. Programming, 56 (1992), pp. 245–284.

[10] M. A. Nunez and R. M. Freund, *Condition measures and properties of the central trajectory of a linear program*, Math. Programming, 83 (1998), pp. 1–28.

[11] J. Renegar, *Some perturbation theory for linear programming*, Math. Programming, 65 (1994), pp. 73–92.

[12] J. Renegar, *Incorporating condition measures into the complexity theory of linear programming*, SIAM J. Optim., 5 (1995), pp. 506–524.

[13] J. Renegar, *Linear programming, complexity theory, and elementary functional analysis*, Math. Programming, 70 (1995), pp. 279–351.

[14] J. Renegar, *Condition numbers, the barrier method, and the conjugate-gradient method*, SIAM J. Optim., 6 (1996), pp. 879–912.

[15] S. M. Robinson, *A characterization of stability in linear programming*, Oper. Res., 25 (1977), pp. 435–447.

[16] S. J. Wright, *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, 1997.

[17] E. A. Yildirim and M. J. Todd, *An interior-point approach to sensitivity analysis in degenerate linear programs*, SIAM J. Optim., to appear.

[18] E. A. Yildirim and M. J. Todd, *Sensitivity analysis in linear programming and semidefinite programming using interior-point methods*, Math. Program., 90 (2001), pp. 229–261.