Science & Technology Facilities Council

Preprint

RAL-P-2016-007

The iterative solution of linear systems arising in the primal-dual interior point algorithm for linear programming

T Rees

July 2016

Submitted for publication in SIAM Journal on Scientific Computing

RAL Library
STFC Rutherford Appleton Laboratory
R61
Harwell Oxford
Didcot
OX11 0QX

Tel: +44(0)1235 445384
Fax: +44(0)1235 446403
email: libraryral@stfc.ac.uk

ISSN 1361- 4762

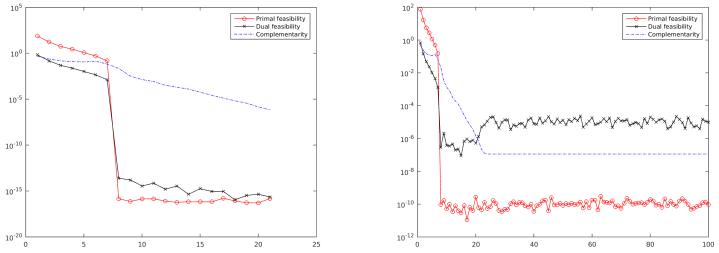
THE ITERATIVE SOLUTION OF LINEAR SYSTEMS ARISING IN THE PRIMAL-DUAL INTERIOR POINT ALGORITHM FOR LINEAR PROGRAMMING

TYRONE REES*

Abstract. A critical component of interior point algorithms is a linear system solve, and we develop conditions on how accurate this solve must be in order to preserve the convergence of the interior point method. We show that the error, measured in a certain norm, must be reduced proportional to the square-root of the duality measure, $\mu_k^{1/2}$, in comparison to other approaches that suggest a reduction proportional to μ_k is needed. We also show that this norm is strongly related to the natural norm in which many of the traditional iterative methods converge. We give numerical results that suggest that this framework is descriptive of the convergence behaviour of the interior point method in the presence of inexactness.

1. Introduction. Interior point methods have proved to be one of the most efficient ways to solve linear programming problems. Such methods iterate towards the optimal solution, solving a linear system at each step in order to calculate the next search direction. The linear system is traditionally solved using direct methods. However, the dimension of the systems that we need to solve has grown rapidly as the volume and velocity of data has increased in recent years. Therefore we increasingly need to replace the direct solve in interior point codes by an iterative solve; see, for example, the work towards a matrix-free interior point method [16].

Iterative methods for linear systems, by their nature, only solve the system inexactly, and this can cause difficulties. To illustrate this, consider the `adlittle` problem from the Netlib test set [13], which we solve in MATLAB using a modified version of Michael Saunders' PDCO code [40]. We changed the routine to solve the augmented system using Matlab's `backslash`, and in addition introduced an error by adding a small perturbation (specifically, `1e-12` $*$ `randn(n,1)`) to the solution. Figure 1.1 shows the results of this test, and it's clear that—even with a perturbation this small—the error is significant enough to cause the interior point method to fail to converge. In particular, the problem loses dual feasibility and never recovers.



(a) Augmented system solved with `backslash`

(b) Augmented system solved with `backslash` then perturbed

Fig. 1.1: Convergence measures, direct method and perturbed solution

*Rutherford Appleton Laboratory, Chilton, Didcot, UK, tyrone.rees@stfc.ac.uk. This author was supported by Engineering and Physical Sciences Research Council grant EP/I013067/1.

This example is a simple—and representative—illustration of the fact that we need to take some care when solving for the Newton correction inexactly. In particular, it would appear that having a small error is insufficient, and instead we need to focus on the size of the residual.

An obvious question arises: how accurate must the solution to the linear system be in order to retain the convergence of the exact algorithm? Researchers have been trying to quantify this since at least the mid 1990's [24, 4, 12], and the case where the inexact solution is still primal feasible (for example, by solving with a constraint preconditioner) has been particularly well studied [24, 35, 36, 2]. We describe the prior work in more detail in Section 5, but in general these studies predict we should solve the linear system to a tolerance proportional to the duality measure, μ_k. It can be observed experimentally that we get convergence at significantly looser tolerances [3, 37, 28].

The main contributions of this work are the following:

- We describe the inexact infeasible interior point method in Section 2. We state three assumptions on the accuracy of the solution, including requirements that we solve the linear system to an accuracy proportional to $\mu_k^{1/2}$ in a certain norm, and that the residual is reduced. Theorem 2.7 shows that, provided these assumptions hold, this algorithm converges at the same rate as predicted by the theory for the equivalent interior point algorithm with exact solves.
- We argue in Section 2.3 that we can compute a value for δ_k such that, if we solve the linear system to a tolerance of $\delta_k \mu_k^{1/2}$ (measured in the 'natural' norm), then the residual reduction requirement is satisfied. The value of δ_k relies on the largest entry of the primal and dual solutions, $\mathbf{x}_k$ and $\mathbf{s}_k$, and the largest singular value of the constraint matrix, B.
- We claim that the norms predicted by the theory above are simply related to the norms in which four commonly used iterative methods naturally converge. In particular, we consider preconditioned conjugate gradients, conjugate gradients applied to the normal equations with a constraint preconditioner, and MINRES with Schur-complement and augmented Lagrangian preconditioners. We justify these claims in Sections 3.1 and 3.2.

In addition, we present numerical validation of the theoretical results above in Sections 2.5 and 4. The code used to generate these results is available to download from https://github.com/tyronerees/IterativeInnerLP.

2. Inexactness in the interior point algorithm. We begin by developing the framework that we will use throughout. We state a version of the inexact infeasible interior point algorithm (IIIP), and prove the main result of the paper, giving conditions on the inexactness such that IIIP converges at the same rate that we can show the (exact) infeasible interior point method converges.

2.1. Background. Consider the linear program

$$\begin{aligned} \min \ & \mathbf{f}^T \mathbf{x} \\ \text{s.t. } & B\mathbf{x} = \mathbf{g}, \\ & \mathbf{x} \geq \mathbf{0}, \end{aligned} \tag{2.1}$$

where we assume that $B \in \mathbb{R}^{m\times n}$, $m < n$, is of full rank. The dual problem to (2.1) is

$$\begin{aligned} \max\ & \mathbf{g}^T\mathbf{y} \\ \text{s.t. } & B^T\mathbf{y} + \mathbf{s} = \mathbf{f} \\ & \mathbf{s} \geq \mathbf{0}. \end{aligned} \tag{2.2}$$

The primal-dual solutions of (2.1)–(2.2) are given by the Karush-Kuhn-Tucker (KKT) conditions, the application of which gives us the non-linear system

$$\mathbf{s} + B^T\mathbf{y} = \mathbf{f} \tag{2.3}$$

$$B\mathbf{x} = \mathbf{g} \tag{2.4}$$

$$XS\mathbf{e} = 0 \tag{2.5}$$

$$\mathbf{x}_i \geq 0, \quad \mathbf{s}_i \geq 0, \tag{2.6}$$

where $X = \text{diag}(\mathbf{x})$, $S = \text{diag}(\mathbf{s})$ and $\mathbf{e}$ denotes the vector of ones. A method for solving these equations that has proved to be particularly effective is the primal-dual interior point method – see, e.g., Wright [45] and the references within for an overview of the area.

The interior point method is an adaptation of the standard Newton algorithm that biases the search direction towards the interior of the region $\mathbf{x}_k,\ \mathbf{s}_k > 0$, and also to keep the components of $\mathbf{x}_k,\ \mathbf{s}_k$ from getting too close to 0. In practice we do this by relaxing the complementarity condition so that we allow $(\mathbf{x}_k)_i(\mathbf{s}_k)_i \neq 0$; it is usual to require $(\mathbf{x}_k)_i(\mathbf{s}_k)_i = \sigma_k\mu_k$, where $\sigma_k \in [0, 1]$ is a centering parameter (which may change with each iteration) and $\mu_k = \mathbf{x}_k^T\mathbf{s}_k/n$ is the duality measure. It can be shown that, with such a choice, we obtain strict complementarity at convergence.

It may not be practical to find a feasible starting point, that is a point $(\mathbf{x}_0, \mathbf{y}_0, \mathbf{s}_0)$ such that (2.3) and (2.4) hold exactly. For such cases the *infeasible interior-point* applies, and the Newton-like iteration at each step is given by solving a system of the form

$$\begin{bmatrix} 0 & B^T & I \\ B & 0 & 0 \\ S_k & 0 & X_k \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x}_k^* \\ \Delta\mathbf{y}_k^* \\ \Delta\mathbf{s}_k^* \end{bmatrix} = \begin{bmatrix} -\zeta_k \\ -\xi_k \\ -X_kS_k\mathbf{e} + \sigma_k\mu_k\mathbf{e}, \end{bmatrix} \tag{2.7}$$

where $\zeta_k = \mathbf{s}_k + B^T\mathbf{y}_k - \mathbf{f}$ and $\xi_k = B\mathbf{x}_k - \mathbf{g}$. Once the algorithm computes the search directions, it takes a step in that direction

$$\begin{bmatrix} \mathbf{x}_{k+1} \\ \mathbf{y}_{k+1} \\ \mathbf{s}_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_k + \alpha_k\Delta\mathbf{x}_k^* \\ \mathbf{y}_k + \alpha_k\Delta\mathbf{y}_k^* \\ \mathbf{s}_k + \alpha_k\Delta\mathbf{s}_k^* \end{bmatrix},$$

where $\alpha_k \in [0, 1]$ is a step length, with the specific value prescribed by the flavour of interior point method used.

Solving the linear system (2.7) is the main computational effort at each iteration of the interior point method. We consider two ways of re-writing (2.7) to make its solution easier. First, consider the block 2×2 formulation, known as the *augmented system*:

$$\begin{bmatrix} X_k^{-1}S_k & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x}_k^* \\ -\Delta\mathbf{y}_k^* \end{bmatrix} = \begin{bmatrix} X_k^{-1}\tau_k - \zeta_k \\ \xi_k \end{bmatrix}, \tag{2.8}$$

where $\tau_k = -X_k S_k \mathbf{e} + \sigma_k \mu_k \mathbf{e}$.

The alternative formulation is to solve (2.8) using the range-space method, further reducing the coefficient matrix to its primal Schur complement. This gives the equations

$$B(S_k^{-1}X_k)B^T \Delta \mathbf{y}_k^* = \xi_k + BS_k^{-1}(\tau_k - X_k\zeta_k) \tag{2.9}$$

(in this context called the *normal equations*), which are subsequently solved using a Choleksy factorization of $B(S_k^{-1}X_k)B^T$. This formulation is usually denser than (2.8), and has worse stability properties, but is often used in interior point codes; see the recent paper of Greif, Moulding and Orban [22] for a discussion of the relative merits of these approaches.

When solving large scale optimziation problems there is a point after which it is no longer feasible to use direct methods, and an iterative method is the only viable approach. In response to this there has been a lot of interest recently in the so-called 'matrix-free' interior point method—see Gondzio [16] and the references therein—in which the solution algorithm is only allowed access to the matrix in terms of a matrix-vector product.

2.2. An inexact infeasible interior point algorithm. In this section we give conditions on the accuracy with which we should solve (2.7) (or, in practice, one of the reduced versions) so that we can prove convergence of IIIP.

First, we define a neighbourhood of the central path,

$$\mathcal{N}_{-\infty}(\gamma,\beta) = \left\{ \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{s} \end{bmatrix} \mid \left\| \begin{bmatrix} \zeta \\ \xi \end{bmatrix} \right\| \leq \frac{\beta\mu}{\mu_0} \left\| \begin{bmatrix} \zeta_0 \\ \xi_0 \end{bmatrix} \right\|, \ \begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix} > 0, \quad x_i s_i \geq \gamma\mu \ \forall i \right\}, \tag{2.10}$$

where $\gamma \in (0,1)$ and $\beta \geq 1$ are parameters, and ζ_0, ξ_0 and μ_0 are the primal residual, dual residual, and duality measure evaluated at the initial point. Furthermore, ζ, ξ and μ the equivalent quantities evaluated at the point $[\mathbf{x}^T, \mathbf{y}^T, \mathbf{s}^T]^T$. We will require the iterates to lie within this set.

The algorithm we consider here is given below as Algorithm 1. This is an extension of that of Kojima, Megiddo and Mizuno [26], our addition being that we allow the inexact solution of the linear system. In Section 2.2.1 we outline the proof of the usual algorithm, with exact solves, as presented in Wright [46]. We then extend this proof in Section 2.2.1 to cover the case where the linear system is only approximated.

2.2.1. An outline of convergence of the exact version. In the case where we solve the linear system in Algorithm 1 using a direct method, a very nice description of the proof is given by Wright [46, Chapter 6]. We will use this proof as the basis of the proof of the extended algorithm in Section 2.2.2 and, for the reader's convenience, we repeat the main steps here.

LEMMA 2.1. *There is a positive constant C_1^* such that*

$$\nu_k \|(\mathbf{x}_k, \mathbf{s}_k)\|_1 \leq C_1^* \mu_k \qquad \textit{for all } k \geq 0.$$

Proof. See proof in Wright [46, Lemma 6.3]. Note that

$$C_1^* = \left(\beta n + n + \frac{\beta}{\mu_0} \left\| \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{s}_0 \end{bmatrix} \right\|_\infty \left\| \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix} \right\|_1 \right) / \left(\min_{i=1,2,\ldots,n} \min \{ (\mathbf{x}_0)_i, (\mathbf{s}_0)_i \} \right).$$

Algorithm 1 Inexact Infeasible Interior Point method

Given $\gamma \in (0,1)$, $\beta \geq 1$, $0 < \sigma_{\min} < \sigma_{\max} \leq 0.5$, $\epsilon_{\max} < \sigma_{\max}$
Choose $(\mathbf{x}_0, \mathbf{y}_0, \mathbf{s}_0)$ with $(\mathbf{x}_0, \mathbf{s}_0) > 0$
for $k = 0, 1, 2, \ldots$ **do**
  Choose $\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$
  Choose $\epsilon_k < \epsilon_{\max}$
  Solve inexactly

$$\begin{bmatrix} 0 & B^T & I \\ B & 0 & 0 \\ S_k & 0 & X_k \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x}_k^* \\ \Delta\mathbf{y}_k^* \\ \Delta\mathbf{s}_k^* \end{bmatrix} = \begin{bmatrix} -\zeta_k \\ -\xi_k \\ -X_k S_k \mathbf{e} + \sigma_k \mu_k \mathbf{e}, \end{bmatrix}.$$

  Choose α_k as the largest value $\alpha \in [0,1]$ such that

$$(\mathbf{x}_k + \alpha\Delta\mathbf{x}_k, \mathbf{y}_k + \alpha\Delta\mathbf{y}_k, \mathbf{s}_k + \alpha\Delta\mathbf{s}_k) \in \mathcal{N}_{-\infty}(\gamma, \beta) \tag{2.11}$$

  and the following Armijo condition holds:

$$\mu_k(\alpha) := \frac{(\mathbf{x}_k + \alpha\Delta\mathbf{x}_k)^T(\mathbf{s}_k + \alpha\Delta\mathbf{x}_k)}{n} \leq (1 - 0.01\alpha)\mu_k \tag{2.12}$$

  Set

$$(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}, \mathbf{s}_{k+1}) = (\mathbf{x}_k + \alpha\Delta\mathbf{x}_k, \mathbf{y}_k + \alpha\Delta\mathbf{y}_k, \mathbf{s}_k + \alpha\Delta\mathbf{s}_k)$$

end for

□

The result of Lemma 2.1 is used to prove the following result, in which we bound the size of the direction of change in the primal and dual variables in a certain norm:

LEMMA 2.2. *There is a positive constant C_2^* such that*

$$\|\Delta\mathbf{x}_k\|_{X_k^{-1}S_k} \leq C_2^* \mu_k^{1/2}, \ \|\Delta\mathbf{s}_k\|_{S_k^{-1}X_k} \leq C_2^* \mu_k^{1/2},$$

for all $k \geq 0$.

Proof. See proof in Wright [46, Lemma 6.5]. The proof relies on the result of Lemma 2.1. Note that

$$C_2^* = 2\frac{C_1^*}{\gamma^{1/2}} \max(\|\mathbf{x}_0 - \mathbf{x}_*\|, \|\mathbf{s}_0 - \mathbf{s}_*\|) + \frac{n}{\gamma^{1/2}},$$

□

Lemma 2.2 is, in turn, used to prove the final technical lemma, which guarantees that significant progress is made at each step of Algorithm 1.

LEMMA 2.3. *There is a value $\bar{\alpha} \in (0,1)$ such that the following three conditions are satisfied for all $\alpha \in [0, \bar{\alpha}]$ and all $k \geq 0$:*

$$(\mathbf{x}_k + \alpha\Delta\mathbf{x}_k)^T(\mathbf{s}_k + \alpha\Delta\mathbf{s}_k) \geq (1 - \alpha)\mathbf{x}_k^T\mathbf{s}_k,$$
$$((\mathbf{x}_k)_i + \alpha(\Delta\mathbf{x}_k)_i)^T((\mathbf{s}_k)_i + \alpha(\Delta\mathbf{s}_k)_i) \geq (\gamma/n)(\mathbf{x}_k + \alpha\Delta\mathbf{x}_k)^T(\mathbf{s}_k + \alpha\Delta\mathbf{s}_k),$$
$$(\mathbf{x}_k + \alpha\Delta\mathbf{x}_k)^T(\mathbf{s}_k + \alpha\Delta\mathbf{s}_k) \leq (1 - .01\alpha)\mathbf{x}_k^T\mathbf{s}_k.$$

Therefore, conditions (2.11) and (2.12) are satisfied for all $\alpha \in [0, \bar{\alpha}]$ and all $k \geq 0$.

Proof. See proof in Wright [46, Lemma 6.7]. ∎

Once we have the proof of Lemma 2.3, the proof of the following result is straightforward.

THEOREM 2.4. *The sequence $\{\mu_k\}$ generated by Algorithm 1 with exact solves converges Q-linearly to zero, and the sequence of residual norms $\left\{\|(\mathbf{r}_k^f, \mathbf{r}_k^g)\|\right\}$ converges R-linearly to zero.*

2.2.2. The proof of convergence of the inexact version. We now turn our attention to the case where we allow an inexact solution of the linear system.

We are going to concentrate not on solving the block 3×3 system directly, but instead solving the reduced augmented system (2.8) or the normal equations (2.9). It's easy to see that—as X_k is diagonal—either of these approaches result in an approximate solution $(\Delta\mathbf{x}_k, \Delta\mathbf{y}_k, \Delta\mathbf{s}_k)$ where the final equation in (2.7) is satisfied exactly. We formally state our assumption below:

ASSUMPTION 1: *The last row of (2.7) is satisfied exactly. That is,*

$$S_k\Delta\mathbf{s}_k + X_k\Delta\mathbf{s}_k = (\sigma_k\mu_k - X_kS_k)\mathbf{e}.$$

If Assumption 1 is satisfied, the inexact solve is equivalent to solving *exactly* a perturbed problem of the form

$$\begin{bmatrix} 0 & B^T & I \\ B & 0 & 0 \\ S_k & 0 & X_k \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x}_k \\ \Delta\mathbf{y}_k \\ \Delta\mathbf{s}_k \end{bmatrix} = - \begin{bmatrix} \zeta_k \\ \xi_k \\ X_kS_k\mathbf{e} - \sigma_k\mu_k\mathbf{e} \end{bmatrix} + \begin{bmatrix} \mathbf{r}_k^f \\ \mathbf{r}_k^g \\ \mathbf{0} \end{bmatrix}, \tag{2.13}$$

where the residuals are given by $\mathbf{r}_k^f = B^T\Delta\mathbf{x}_k + \Delta\mathbf{s}_k + \zeta_k$ and $\mathbf{r}_k^g = B\Delta\mathbf{x}_k + \xi_k$.

We follow the method of the proof outlined in Section 2.2.1, making adaptations to suit our needs as necessary. We first prove a lemma that extends Lemma 2.2 to the inexact case.

LEMMA 2.5. *Suppose the linear system (2.7) is solved inexactly in such a way that Assumption 1 holds and, furthermore, the errors satisfy*

$$\|\Delta\mathbf{x}_k - \Delta\mathbf{x}_k^*\|_{X_k^{-1}S_k} \leq \mu_k^{1/2}\epsilon_k, \tag{2.14}$$

$$\|\Delta\mathbf{y}_k - \Delta\mathbf{y}_k^*\|_{B(X_k^{-1}S_k)^{-1}B^T} \leq \mu_k^{1/2}\epsilon_k. \tag{2.15}$$

Then there is a positive constant C_2 such that

$$\|\Delta\mathbf{x}_k\|_{X_k^{-1}S_k} \leq C_2\mu_k^{1/2} \qquad \|\Delta\mathbf{s}_k\|_{S_k^{-1}X_k} \leq C_2\mu_k^{1/2}$$

for all $k \geq 0$.

Proof. We apply Lemma 2.2 to get a bound on the required quantities by applying the triangle inequality:

$$\begin{aligned} \|\Delta\mathbf{x}_k\|_{X_k^{-1}S_k} &\leq \|\Delta\mathbf{x}_k - \Delta\mathbf{x}_k^*\|_{X_k^{-1}S_k} + \|\Delta\mathbf{x}_k^*\|_{X_k^{-1}S_k} \\ &\leq \|\Delta\mathbf{x}_k - \Delta\mathbf{x}_k^*\|_{X_k^{-1}S_k} + C_2^*\mu_k^{1/2} \\ &\leq (C_2^* + \epsilon_{\max})\mu_k^{1/2}. \end{aligned}$$

Applying a similar argument, and using the fact that the third equation in (2.13) has zero residual, gives that

$$\begin{aligned}\|\Delta\mathbf{s}_k\|_{S_k^{-1}X_k} &\leq \|\Delta\mathbf{s}_k - \Delta\mathbf{s}_k^*\|_{S_k^{-1}X_k} + C_2^*\mu_k^{1/2} \\ &= \left\|X_k^{-1}S_k(\Delta\mathbf{x}_k - \Delta\mathbf{x}_k^*)\right\|_{S_k^{-1}X_k} + C_2^*\mu_k^{1/2} \\ &= \|\Delta\mathbf{x}_k - \Delta\mathbf{x}_k^*\|_{X_k^{-1}S_k} + C_2^*\mu_k^{1/2} \\ &\leq (C_2^* + \epsilon_{\max})\mu_k^{1/2}.\end{aligned}$$

The result therefore holds with $C_2 = C_2^* + \epsilon_{\max}$. □

We highlight the fact that, in the statement of this result, we have made an additional assumption about the quality of our approximate solution:

ASSUMPTION 2: *The linear system is solved with an accuracy such that (2.14) and (2.15) hold for some given $\epsilon_k > 0$.*

The reason for this assumption is clear from the proof, but it is somewhat unsatisfactory as it depends on the size of the *error*, which, by its very nature, is not explicitly available. Despite this limitation, we claim in Section 2.4 that this is, in fact, a natural requirement. Furthermore, in Section 3 we show how this quantity is related to the Krylov subspace methods typically used in this context.

We will now prove the final lemma, the inexact analogue to Lemma 2.3, which shows that sufficient progress is made at each step of the algorithm:

LEMMA 2.6. *Suppose that Assumptions 1 and 2 hold, and additionally that the residuals $\mathbf{r}_k^f = B^T\Delta\mathbf{y}_k + \Delta\mathbf{s}_k + \zeta_k$ and $\mathbf{r}_k^g = B\Delta\mathbf{x}_k + \xi_k$ satisfy*

$$\left\|\begin{bmatrix}\mathbf{r}_k^f \\ \mathbf{r}_k^g\end{bmatrix}\right\|_2 \leq \epsilon_k \left\|\begin{bmatrix}\zeta_k \\ \xi_k\end{bmatrix}\right\|_2 \tag{2.16}$$

for some $\epsilon_k \in [0, 1)$. Then there is a value $\bar{\alpha} \in (0, 1)$ such that the following three conditions are satisfied for all $\alpha \in [0, \bar{\alpha}]$ and for all $k \geq 0$:

$$(\mathbf{x}_k + \alpha\Delta\mathbf{x}_k)^T(\mathbf{s}_k + \alpha\Delta\mathbf{s}_k) \geq (1 - (1 - \epsilon_{\max})\alpha)\mathbf{x}_k^T\mathbf{s}_k \tag{2.17}$$

$$((\mathbf{x}_k)_i + \alpha(\Delta\mathbf{x}_k)_i)^T((\mathbf{s}_k)_i + \alpha(\Delta\mathbf{s}_k)_i) \geq (\gamma/n)(1 - \alpha)\mathbf{x}_k^T\mathbf{s}_k \tag{2.18}$$

$$(\mathbf{x}_k + \alpha\Delta\mathbf{x}_k)^T(\mathbf{s}_k + \alpha\Delta\mathbf{s}_k) \leq (1 - 0.01\alpha)\mathbf{x}_k^T\mathbf{s}_k. \tag{2.19}$$

Therefore the conditions (2.11) and (2.12) are satisfied for all $\alpha \in [0, \bar{\alpha}]$ and for all $k \geq 0$.

Proof. First, we note the relationships

$$\Delta\mathbf{x}_k^T\Delta\mathbf{s}_k \leq C_2^2\mu_k, \qquad \mathbf{s}_k^T\Delta\mathbf{x}_k + \mathbf{x}_k^T\Delta\mathbf{s}_k = (\sigma_k - 1)\mathbf{x}_k^T\mathbf{s}_k,$$

which are easy to prove using Lemma 2.5 and the fact that the third equation of (2.7) holds exactly (see the proof of Lemma 6.7 in [46]). Using these we get that

$$\begin{aligned}&(\mathbf{x}_k + \alpha\Delta\mathbf{x}_k)^T(\mathbf{s}_k + \alpha\Delta\mathbf{s}_k) \\ &\quad= \mathbf{x}_k^T\mathbf{s}_k + \alpha(\sigma_k - 1)\mathbf{x}_k^T\mathbf{s}_k + \alpha^2\Delta\mathbf{x}_k^T\Delta\mathbf{s}_k \\ &\quad\geq (1 - \alpha)\mathbf{x}_k^T\mathbf{s}_k + \alpha\sigma_k\mathbf{x}_k^T\mathbf{s}_k - \alpha^2C_2^2\mu_k \\ &\quad= (1 - (1 - \epsilon_{\max})\alpha)\mathbf{x}_k^T\mathbf{s}_k + (\sigma_k - \epsilon_{\max})\alpha\mathbf{x}_k^T\mathbf{s}_k - \alpha^2C_2^2\mu_k \\ &\quad\geq (1 - (1 - \epsilon_{\max})\alpha)\mathbf{x}_k^T\mathbf{s}_k + \left((\sigma_k - \epsilon_{\max})\alpha - \frac{\alpha^2C_2^2}{n}\right)\mathbf{x}_k^T\mathbf{s}_k,\end{aligned}$$

where $\epsilon_{\max} = \max \epsilon_k$. Hence (2.17) holds provided the final term above is non-negative, which is true when

$$\alpha \leq \frac{(\sigma_{\min} - \epsilon_{\max})n}{C_2^2}.$$

To show that (2.17) implies (2.11), we first note that, using the definition of ζ_{k+1}, ξ_{k+1}, and (2.13), it is easy to see that

$$\begin{bmatrix} \zeta_{k+1} \\ \xi_{k+1} \end{bmatrix} = (1-\alpha) \begin{bmatrix} \zeta_k \\ \xi_k \end{bmatrix} + \alpha \begin{bmatrix} \mathbf{r}_k^f \\ \mathbf{r}_k^g \end{bmatrix}. \tag{2.20}$$

We can therefore, using condition (2.16), bound the residuals of the interior point method by

$$\left\| \begin{bmatrix} \zeta_{k+1} \\ \xi_{k+1} \end{bmatrix} \right\| \leq (1-\alpha) \left\| \begin{bmatrix} \zeta_k \\ \xi_k \end{bmatrix} \right\| + \alpha\epsilon_k \left\| \begin{bmatrix} \zeta_k \\ \xi_k \end{bmatrix} \right\|$$

and hence we obtain

$$\left\| \begin{bmatrix} \zeta_{k+1} \\ \xi_{k+1} \end{bmatrix} \right\| \leq (1-(1-\epsilon_k)\alpha) \left\| \begin{bmatrix} \zeta_k \\ \xi_k \end{bmatrix} \right\|. \tag{2.21}$$

Then

$$\begin{aligned}
\frac{1}{\mu_{k+1}} \left\| \begin{bmatrix} \zeta_{k+1} \\ \xi_{k+1} \end{bmatrix} \right\| &\leq \frac{1-(1-\epsilon_{\max})\alpha}{\mu_{k+1}} \left\| \begin{bmatrix} \zeta_k \\ \xi_k \end{bmatrix} \right\| \qquad \text{[by (2.21)]} \\
&= \frac{n(1-(1-\epsilon_{\max})\alpha)}{(\mathbf{x}_k + \alpha\Delta\mathbf{x}_k)^T(\mathbf{s}_k + \alpha\Delta\mathbf{x}_k)} \left\| \begin{bmatrix} \zeta_k \\ \xi_k \end{bmatrix} \right\| \\
&\leq \frac{n(1-(1-\epsilon_{\max})\alpha)}{(1-(1-\epsilon_{\max})\alpha)\mathbf{x}_k^T\mathbf{s}_k} \left\| \begin{bmatrix} \zeta_k \\ \xi_k \end{bmatrix} \right\| \qquad \text{[by (2.17)]} \\
&= \frac{1}{\mu_k} \left\| \begin{bmatrix} \zeta_k \\ \xi_k \end{bmatrix} \right\| \\
&\leq \frac{\beta}{\mu_0} \left\| \begin{bmatrix} \zeta_0 \\ \xi_0 \end{bmatrix} \right\|,
\end{aligned}$$

the final step holding because we are within the neighbourhood of the central path (2.10).

The remainder of the proof is identical to that of [46, Lemma 6.7], and all the conditions (2.17– 2.19) hold if $\alpha \in [0, \bar{\alpha}]$, where

$$\bar{\alpha} = \min\left(\frac{n(\sigma_{\min} - \epsilon_{\max})}{C_2^2}, \frac{\sigma_{\min}(1-\gamma)}{C_2^2}, \frac{0.49n}{C_2^2}, 1 \right).$$

□

In proving Lemma 2.6 we made an additional assumption on the quality of the solution, which we will now formalize.

ASSUMPTION 3: *The residuals* $\mathbf{r}_k^f = B^T\Delta\mathbf{y}_k + \Delta\mathbf{s}_k + \zeta_k$ *and* $\mathbf{r}_k^g = B\Delta\mathbf{x}_k + \xi_k$ *satisfy (2.16) for some* $\epsilon_k \in [0, 1)$.

We will give conditions on the tolerances required so that Assumptions 1 and 2 imply Assumption 3 in Section 2.3. First, we state and prove our main result.

THEOREM 2.7. *Suppose that (2.7) is solved in such a way that Assumptions 1—3 hold. Then the sequence* $\{\mu_k\}$ *generated by Algorithm 1 converges Q-linearly to zero, and the sequence of residual norms* $\left\{ \left\| \begin{bmatrix} \zeta_k \\ \xi_k \end{bmatrix} \right\| \right\}$ *converges R-linearly to zero.*

Proof. From Lemma 2.6, there is a constant $\bar{\alpha} > 0$ such that $\alpha_k > \bar{\alpha}$ for all k. From the Armijo condition we have that

$$\mu_{k+1} \leq (1 - 0.01\alpha_k)\mu_k \leq (1 - 0.01\bar{\alpha})\mu_k$$

for all k, and hence the sequence of duality gaps converges Q-linearly to zero.

Next, since

$$\left\| \begin{bmatrix} \zeta_k \\ \xi_k \end{bmatrix} \right\| \leq \mu_k \cdot \frac{\beta}{\mu_0} \left\| \begin{bmatrix} \zeta_0 \\ \xi_0 \end{bmatrix} \right\|,$$

the sequence $\left\{ \left\| \begin{bmatrix} \zeta_k \\ \xi_k \end{bmatrix} \right\| \right\}$ is bounded above by another sequence that converges Q-linearly, so the sequence of residual norms is R-linearly convergent. ∎

2.3. Removing the requirement of residual reduction. Recall Theorem 2.7 holds provided the inexact solve satisfies three requirements: Assumption 1 (the last row of (2.7) is satisfied exactly), Assumption 2 (which bounds the error in a norm that changes every iteration), and Assumption 3 (which bounds the residual in a global norm). We now give a condition that enables us to select a value for ϵ_k in (2.14) and (2.15) such that Assumptions 1 and 2 imply Assumption 3.

PROPOSITION 2.8. *Let $\epsilon \in (0, 1)$ be given. There exists $\delta \in (0, 1)$ such that if the iterates satisfy*

$$\|\Delta \mathbf{x}_k - \Delta \mathbf{x}_k^*\|_{X_k^{-1} S_k} \leq \mu_k^{1/2} \delta, \qquad \|\Delta \mathbf{y}_k - \Delta \mathbf{y}_k^*\|_{B(X_k^{-1} S_k)^{-1} B^T} \leq \mu_k^{1/2} \delta$$

then the residuals satisfy

$$\left\| \begin{bmatrix} \mathbf{r}_k^f \\ \mathbf{r}_k^g \end{bmatrix} \right\|_2 \leq \epsilon \left\| \begin{bmatrix} \zeta_k \\ \xi_k \end{bmatrix} \right\|_2. \tag{2.22}$$

Proof. Let us define $H = X_k^{-1} S_k$, $\mathbf{e}_k^x = \Delta \mathbf{x}_k^* - \Delta \mathbf{x}_k$, $\mathbf{e}_k^y = \Delta \mathbf{y}_k^* - \Delta \mathbf{y}_k$, and $\mathbf{e}_k^s = \Delta \mathbf{s}_k^* - \Delta \mathbf{s}_k$, and note that

$$\begin{aligned} \left\| \begin{bmatrix} \mathbf{r}_k^f \\ \mathbf{r}_k^g \end{bmatrix} \right\|_2 &= \left(\|\mathbf{r}_k^f\|_2^2 + \|\mathbf{r}_k^g\|_2^2 \right)^{1/2} \\ &\leq \|\mathbf{r}_k^f\|_2 + \|\mathbf{r}_k^g\|_2. \end{aligned} \tag{2.23}$$

We take the two components on the right hand side of (2.23) separately. First, note that, since all norms in finite dimensions are equivalent, there is a constant C such

$$\|\mathbf{r}_k^f\|_2 \leq C \|\mathbf{r}_k^f\|_{H^{-1}}.$$

We focus on the alternative norm, for which

$$\|\mathbf{r}_k^f\|_{H^{-1}} = \|B^T \mathbf{e}_k^y + \mathbf{e}_k^s\|_{H^{-1}} \leq (\|B^T \mathbf{e}_k^y\|_{H^{-1}} + \|\mathbf{e}_k^s\|_{H^{-1}}) = \|\mathbf{e}_k^y\|_{BH^{-1}B^T} + \|H^{-1/2} \mathbf{e}_k^s\|_2$$

Using the fact that $\mathbf{e}_k^s = H \mathbf{e}_k^x$ (since the last equation of (2.7) is satisfied exactly), we therefore can bound $\|\mathbf{r}_k^f\|_H^{-1}$ by

$$\|\mathbf{r}_k^f\|_{H^{-1}} \leq \|\mathbf{e}_k^y\|_{BH^{-1}B^T} + \|H^{1/2} \mathbf{e}_k^x\|_2 = \|\mathbf{e}_k^y\|_{BH^{-1}B^T} + \|\mathbf{e}_k^x\|_H$$

Therefore, using the error conditions in the statement of the proposition we obtain

$$\|\mathbf{r}_k^f\|_{H^{-1}} \leq \delta\|\zeta_k\|_2 + \delta\|\xi_k\|_2 \leq \sqrt{2}\delta\mu^{1/2} \left\| \begin{bmatrix} \zeta_k \\ \xi_k \end{bmatrix} \right\|_2.$$

We now need to obtain a value for C. Since H^{-1} is symmetric positive definite, C can be taken to be the square root of the maximal eigenvalue of H. As $H = X_k^{-1}S_k$ is diagonal, the maximal eigenvalue is the largest entry. Thus

$$\max \lambda(H) = \max_i x_i^{-1}s_i = \max_i (x_i s_i)^{-1} s_i^2 \leq (\gamma\mu_k)^{-1} \max_i s_i^2.$$

Therefore

$$\|\mathbf{r}_k^f\|_2 \leq (2/\gamma)^{1/2} \|\mathbf{s}_k\|_1 \delta \left\| \begin{bmatrix} \zeta_k \\ \xi_k \end{bmatrix} \right\|_2. \tag{2.24}$$

Taking the second term of (2.23) we see

$$\|\mathbf{r}_k^g\|_2 = \|B\mathbf{e}_k^x\|_2 = \|\mathbf{e}_k^x\|_{B^TB}.$$

There exists a constant D such that $\|\mathbf{x}\|_{B^TB} \leq D\|\mathbf{x}\|_H$. Therefore

$$\|\mathbf{r}_k^g\|_2 \leq D\|\mathbf{e}_k^x\|_H \leq D\delta\mu_k^{1/2} \left\| \begin{bmatrix} \zeta_k \\ \xi_k \end{bmatrix} \right\|_2.$$

Now we just have to find the constant of equivalence D. Note that

$$D^2 = \max_{\mathbf{x}} \frac{\mathbf{x}^T B^T B\mathbf{x}}{\mathbf{x}^T H\mathbf{x}} = \max_{\mathbf{x}} \frac{\mathbf{x}^T B^T B\mathbf{x}}{\mathbf{x}^T\mathbf{x}} \cdot \frac{\mathbf{x}^T\mathbf{x}}{\mathbf{x}^T H\mathbf{x}} \leq \max_{\mathbf{x}} \frac{\mathbf{x}^T B^T B\mathbf{x}}{\mathbf{x}^T\mathbf{x}} \cdot \max_{\mathbf{z}} \frac{\mathbf{z}^T H^{-1}\mathbf{z}}{\mathbf{z}^T H\mathbf{z}}.$$

The first term on the right hand side is $\Sigma_{\max}^2$, the square of largest singular value of B. The second term is the maximum eigenvalue of H^{-1}, which is bounded above by $(\gamma\mu_k)^{-1}\max_i x_i^2$. We therefore have that

$$\|\mathbf{r}_k^g\|_2 \leq (\gamma)^{-1/2}\|\mathbf{x}_k\|_1 \Sigma_{\max}\delta \left\| \begin{bmatrix} \zeta_k \\ \xi_k \end{bmatrix} \right\|_2. \tag{2.25}$$

Combining (2.24) and (2.25) we have that

$$\left\| \begin{bmatrix} \mathbf{r}_k^f \\ \mathbf{r}_k^g \end{bmatrix} \right\|_2 \leq \frac{\sqrt{2}\|\mathbf{s}_k\|_1 + \Sigma_{\max}\|\mathbf{x}_k\|_1}{(\gamma)^{1/2}} \delta \left\| \begin{bmatrix} \zeta_k \\ \xi_k \end{bmatrix} \right\|_2.$$

Therefore taking any

$$\delta < \frac{(\gamma)^{1/2}\epsilon}{\sqrt{2}\|\mathbf{s}_k\|_1 + \Sigma_{\max}\|\mathbf{x}_k\|_1}, \tag{2.26}$$

will ensure contraction in the residual. □

Proposition 2.8 tells us that, given a desired ϵ_k, we can prescribe an accuracy with which we should solve the linear systems (in a norm that changes with the outer iteration) to a reduction of the residual (in a constant norm) at each step.

We caution that, despite the fact that only an estimate of $\Sigma_{\max}$ is needed in practice, it may be the case that computing this is not possible. In practice, iterating until the condition (2.16) is satisfied may be preferable, although this generally requires additional work per iteration.

2.4. What is special about these norms?. In Section 2.3 we saw that if we can solve (2.7) so that Assumption 1 holds and the errors (2.14) and (2.15) are small enough, then Assumption 3 also holds. Before turn our attention to the linear solvers in Section 3, it is instructive to discuss the norms that appear in Assumption 2. In particular, we contend that far from being unnatural and technical, these are the norms in which we *should* measure convergence of the inner iteration. We highlight the fact that although all norms are equivalent in finite dimensions, the constants involved may depend on parameters that are present in the problem (either physically or introduced by the mathematical modelling), and we want to converge in a norm independent of such things.

2.4.1. The normal equations. We know that the solution to the normal equations formulation (2.9) is given by $\Delta\mathbf{y}_k^*$, but the question is, given a perturbed solution $\Delta\mathbf{y}_k = \Delta\mathbf{y}_k^* + \mathbf{e}_k^y$, what norm $\|\cdot\|_\sharp$ shall we impose on $\mathbb{R}^m$ so that we can ensure that if $\|\mathbf{e}_k^x\|_\sharp$ is small, then $\Delta\mathbf{y}_k$ is 'close' to the solution we want?

One common way of answering this question is to use the fact that solving the normal equations (2.9) is equivalent to finding $\mathbf{v} \in \mathbb{R}^m$ that minimizes the quadratic

$$\phi(\mathbf{v}) := \frac{1}{2}\mathbf{v}^T\widehat{S}\mathbf{v} - \mathbf{v}^T\mathbf{f}, \tag{2.27}$$

where $\widehat{S} = B(S_k^{-1}X_k)B^T$ and $\mathbf{f} = \xi_k + BS_k^{-1}(\tau_k - X_k\zeta_k)$. This is a consequence of the fact that $\widehat{S}$ is symmetric positive definite.

The quadratic ϕ forms a 'bowl' here, and so the solution of (2.9) is at the bottom of the bowl. It is therefore appropriate to think of $\Delta\mathbf{y}_k$ as being 'close to' the exact solution $\Delta\mathbf{y}_k^*$ if $\phi(\Delta\mathbf{y}_k)$ also lies near the bottom of the bowl. Formally, we would like to identify a norm $\|\cdot\|_\sharp$ such that, given $\epsilon > 0$, we can find $\delta > 0$ such that

$$\|\Delta\mathbf{y}_k - \Delta\mathbf{y}_k^*\|_\sharp < \delta \Rightarrow |\phi(\Delta\mathbf{y}_k) - \phi(\Delta\mathbf{y}_k^*)| < \epsilon.$$

It is straightforward to show that $\phi(\Delta\mathbf{y}_k) - \phi(\Delta\mathbf{y}_k^*) = \mathbf{e}_k^y\widehat{S}\mathbf{e}_k^y$ (see, e.g., Nocedal and Wright [30, Section 5.1]), and so choosing $\|\cdot\|_\sharp = \|\cdot\|_{\widehat{S}}$ satisfies this requirement trivially. This argument is commonly used to explain why the norm in which the conjugate gradient method converges in is reasonable; for more on this, see Section 3.2.

In our context, this is the norm that is associated with $\mathbb{R}^m$ in Assumption 2, and so we argue that, far from being technical, it is the natural norm that we should expect to see in a convergence proof.

Once we obtain a suitable vector $\Delta\mathbf{y}_k$, we typically obtain $\Delta\mathbf{x}_k$ by solving the first row of (2.8), obtaining

$$\Delta\mathbf{x}_k = S_k^{-1}X_k(\mathbf{b} + B^T\Delta\mathbf{y}_k), \tag{2.28}$$

where $\mathbf{b} = X_k^{-1}\tau_k - \zeta_k$. Since S_k is diagonal, we can solve this exactly. Note that (2.28) also holds for the exact solutions, so we get that the errors are related by

$$(X_k^{-1}S_k)^{1/2}(\Delta\mathbf{x}_k - \Delta\mathbf{x}_k^*) = (X_k^{-1}S_k)^{-1/2}B^T(\Delta\mathbf{y}_k - \Delta\mathbf{y}_k^*).$$

Therefore we have that, if we solve for $\Delta\mathbf{y}_k$ first,

$$\|\Delta\mathbf{x}_k - \Delta\mathbf{x}_k^*\|_{X_k^{-1}S_k} = \|\Delta\mathbf{y}_k - \Delta\mathbf{y}_k^*\|_{B(X_k^{-1}S_k)^{-1}B^T}, \tag{2.29}$$

and so the condition on the size of $\mathbf{e}_k^x$ in Assumption 2 is also natural, and holds automatically.

2.4.2. The augmented system. We argue similarly for the augmented system. First, note that solving (2.8) is equivalent to solving a quadratic program, namely

$$\min_{\mathbf{u}} \frac{1}{2}\mathbf{u}^T H\mathbf{u} - \mathbf{u}^T\mathbf{b} \tag{2.30}$$

$$\text{s.t. } B\mathbf{u} = \xi_k, \tag{2.31}$$

with $H = X_k^{-1}S_k$, $\mathbf{b} = X_k^{-1}\tau_k - \zeta_k$, and $\mathbf{b} = X_k^{-1}\tau_k - \zeta_k$. It is well-known that minimizing (2.30) subject to (2.31) is equivalent to finding the stationary points of the Lagrangian

$$\mathcal{L}(\mathbf{u}, \mathbf{p}) = \frac{1}{2}\mathbf{u}^T H\mathbf{u} - \mathbf{u}^T\mathbf{b} - \mathbf{p}^T(B\mathbf{u} - \xi_k), \tag{2.32}$$

which recovers the linear system (2.8). The solution to (2.30—2.31) is therefore $\mathbf{u} = \Delta\mathbf{x}_k^*$, which has the associated Lagrange multiplier $\mathbf{p} = -\Delta\mathbf{y}_k^*$.

When we solve the linear system (2.8) (or, equivalently, the quadratic program (2.30—2.31)) only approximately, we obtain $\Delta\mathbf{x}_k \approx \Delta\mathbf{x}_k^*$ and $\Delta\mathbf{y}_k \approx \Delta\mathbf{y}_k^*$. We are interested in finding $\Delta\mathbf{x}_k$ and $\Delta\mathbf{y}_k$ not independently, but as solutions to the coupled linear system (2.8). Subsequently, we need to select norms $\|\cdot\|_\sharp$ on $\mathbb{R}^n$ and $\|\cdot\|_\flat$ on $\mathbb{R}^m$ that reflect this.

Again, we want to ensure that we remain sympathetic to the underlying problem by considering not the matrix formulation (2.8), but the equivalent optimization problem (2.30-2.31), and in particular the Lagrangian (2.32). As in Section 2.4.1, we want to use norms such that, given $\epsilon > 0$, there exists $\delta_x, \delta_y > 0$ such that

$$\|\Delta\mathbf{x}_k - \Delta\mathbf{x}_k^*\|_\sharp < \delta_x, \|\Delta\mathbf{y}_k - \Delta\mathbf{y}_k^*\|_\flat < \delta_y \Rightarrow |\mathcal{L}(\Delta\mathbf{x}_k^*, \Delta\mathbf{y}_k^*) \; - \; \mathcal{L}(\Delta\mathbf{x}_k, \Delta\mathbf{y}_k)| < \epsilon,$$

where δ_x, δ_y are independent of any parameters. This will ensure we remain near to the turning point, and hence solve the underlying (coupled) problem.

Let $\Delta\mathbf{x}_k = \Delta\mathbf{x}_k^* + \mathbf{e}_k^x$ and $\Delta\mathbf{y}_k = \Delta\mathbf{y}_k^* + \mathbf{e}_k^y$. Then

$$\begin{aligned}
\mathcal{L}(\Delta\mathbf{x}_k, \Delta\mathbf{y}_k) =& \frac{1}{2}(\Delta\mathbf{x}_k^* + \mathbf{e}_k^x)^T H(\Delta\mathbf{x}_k^* + \mathbf{e}_k^x) - (\Delta\mathbf{x}_k^* + \mathbf{e}_k^x)^T\mathbf{b} \\
& - (\Delta\mathbf{y}_k^* + \mathbf{e}_k^y)^T(B(\Delta\mathbf{x}_k^* + \mathbf{e}_k^x) - \xi_k) \\
=& \frac{1}{2}(\Delta\mathbf{x}_k^*)^T H\Delta\mathbf{x}_k - (\Delta\mathbf{x}_k^*)^T\mathbf{b} - (\Delta\mathbf{y}_k^*)^T(B\Delta\mathbf{x}_k^* - \xi_k) \\
& + (\Delta\mathbf{x}_k^*)^T H\mathbf{e}_k^x - (\mathbf{e}_k^x)^T\mathbf{b} - (\Delta\mathbf{y}_k^*)^T B\mathbf{e}_k^x - (\mathbf{e}_k^y)^T(B\Delta\mathbf{x}_k^* - \xi_k) \\
& + \frac{1}{2}(\mathbf{e}_k^x)^T H\mathbf{e}_k^x - (\mathbf{e}_k^y)^T B\mathbf{e}_k^x \\
=& \mathcal{L}(\Delta\mathbf{x}_k^*, \Delta\mathbf{x}_k^*) + (\mathbf{e}_k^x)^T(H\Delta\mathbf{x}_k^* - B^T\Delta\mathbf{y}_k^* - \mathbf{b}) - (\mathbf{e}_k^y)^T(B\Delta\mathbf{x}_k^* - \xi_k) \\
& + \frac{1}{2}(\mathbf{e}_k^x)^T H\mathbf{e}_k^x - (\mathbf{e}_k^y)^T B\mathbf{e}_k^x \\
=& \mathcal{L}(\Delta\mathbf{x}_k^*, \Delta\mathbf{x}_k^*) + \frac{1}{2}(\mathbf{e}_k^x)^T H\mathbf{e}_k^x - (\mathbf{e}_k^y)^T B\mathbf{e}_k^x,
\end{aligned}$$

where we have used the fact that the stationary points $\Delta\mathbf{x}_k^*$, $\Delta\mathbf{y}_k^*$ satisfy (2.8). We

can therefore bound the difference between the Lagrangian at the two points by:

$$\begin{aligned}|\mathcal{L}(\Delta\mathbf{x}_k, \Delta\mathbf{y}_k) - \mathcal{L}(\Delta\mathbf{x}_k^*, \Delta\mathbf{y}_k^*)| &\leq \frac{1}{2}\|\mathbf{e}_k^x\|_H^2 + \left|(\mathbf{e}_k^y)^T B\mathbf{e}_k^x\right| \\ &= \frac{1}{2}\|\mathbf{e}_k^x\|_H^2 + \left|(\mathbf{e}_k^y)^T BH^{-1/2}H^{1/2}\mathbf{e}_k^x\right| \\ &\leq \frac{1}{2}\|\mathbf{e}_k^x\|_H^2 + \|H^{-1/2}B^T\mathbf{e}_k^y\|_2\|H^{1/2}\mathbf{e}_k^x\|_2 \\ &= \frac{1}{2}\|\Delta\mathbf{x}_k^* - \Delta\mathbf{x}_k\|_H^2 + \|\Delta\mathbf{y}_k^* - \Delta\mathbf{y}_k\|_{BH^{-1}B^T}\|\Delta\mathbf{x}_k^* - \Delta\mathbf{x}_k^*\|_H.\end{aligned}$$

Therefore, if we have a method where the primal variables converge in the $H-$norm and the dual variables converge in the $\widehat{S}-$norm, where $\widehat{S} = BH^{-1}B^T$, then we have a method which converges to the minimum independently of any parameters. For the interior point method applied to LPs, where $H = X_k^{-1}S_k$, these again are exactly the conditions (2.14) and (2.15) of Assumption 2.

2.5. Numerical validation. In this section we perform some numerical tests to demonstrate empirically that the convergence of the interior point method relies on the inner linear system being solved to an accuracy proportional to $\mu_k^{1/2}$.

We run the NETLIB problem `lp_adlittle` and solve it using a fork of Michael Saunders' PDCO package [40]. We ask for the interior point method to consider the problem solved when the primal feasibility, dual feasibility, and the complimentarity measure are all less than 10^{-6}. The only thing that changes between the tests is the solver for the linear system (2.7). We use Matlab's `backslash` as the direct solver, and compare with the performance of a preconditioned Krylov method.

We only give results for one test problem (which is representative of typical behaviour) as an illustration here; more extensive numerical tests are reported in Section 4, where use Theorem 2.7 and knowledge of specific solvers to predict the accuracy that we need in order to guarantee convergence.

Our aim in this test is to show that the convergence of IIIP depends on the errors being smaller than $\mu_k^{1/2}$ in the norms given in Assumption 2. To this end, in these examples we use fixed tolerances for all iterations, some of which are too loose to allow the interior point method to converge within 50 iterations. We must use standard Krylov methods here, but we delay a detailed discussion of their properties until Section 3.

EXAMPLE 2.9. *We compare the behaviour the interior point method when solving for the Newton direction (2.7) using the normal equation formulation (2.9). We compare the behaviour of* `backslash` *with that of CG solved to a tolerance of* 10^{-1}, 10^{-3}, *and* 10^{-5}. *We use the incomplete Cholesky routine* `HSL_MI28` *[1, 42] as a preconditioner.*

When the linear system is solved using backslash, the convergence curves of the interior point iterations are given in Figure 2.1. Plots showing the progress of the interior point method when the normal equations are solved inexactly, together with the norm of the error and $\mu_k^{1/2}$, are given in Figure 2.2. The figures on the left show measures of convergence of the interior point iteration, while the figures on the right show the norm of the error of the final iterate of CG, where the norm used is that of Assumption 2. To get the error, we solve the equation using a direct method (`backslash`) and calculate (an approximation to) $\|\mathbf{e}_k\|_{(H,\widehat{S})}$ from there

In Figure 2.2b the error lies above the value of $\mu_k^{1/2}$, and this translates to the lack of rapid convergence in the interior point method, as shown in Figure 2.2a. On

the other hand, in Figure 2.2f the error is always well below the value of $\mu_k^{1/2}$, and correspondingly the interior point method convergence in Figure 2.2e follows roughly the same pattern as that of the direct method in Figure 2.1. This behaviour is as we might expect given Theorem 2.7, and suggests that the conditions there are descriptive of actual convergence behaviour.

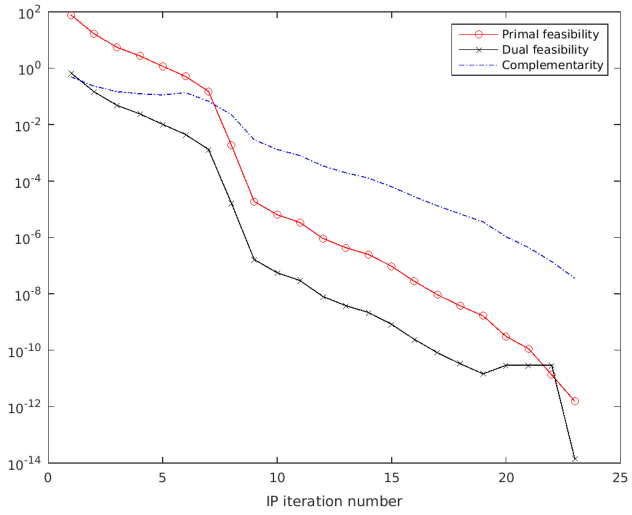


Fig. 2.1: `lp_adlittle`, Normal equations solved with backslash

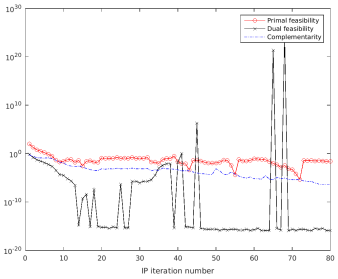
EXAMPLE 2.10. *We compare the behaviour the interior point method when solving for the Newton direction (2.7) using the augmented system formulation (2.8). We compare the behaviour of* `backslash` *with that of MINRES with a block-diagonal Schur complement preconditioner solved to a tolerance of* 10^{-4}, 10^{-6}, 10^{-8} *, and* 10^{-10}. *We use* `HSL_MI28`

The ideal case, where we solve the linear system using a direct method, looks very similar to Figure 2.1 (with only a very slight variation at the last few iterations), so we do not reproduce it here.

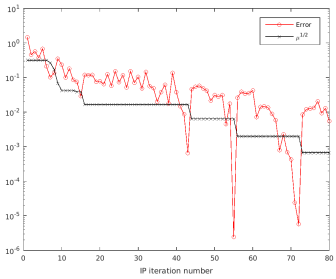
Figure 2.3 shows the results of this test. As in Example 2.9, the figures on the left show measures of convergence of the interior point iteration, while the figures on the right show the norm of the error of the final iterate of MINRES.

It is clear from Figure 2.3 that for the two cases where the interior point method converges in less than 50 iterations the error $\|\mathbf{e}_k\|_{(H,\widehat{S})}$ never strays too far above the $\mu_k^{1/2}$ line, which is what was predicted by the theory above. In the cases where the interior point algorithm fails to converge, we can see the convergence stops being superlinear in the figures on the left by seeing the iteration number at which the error goes above $\mu_k^{1/2}$ in the figures on the right, again in agreement with the theory derived above.

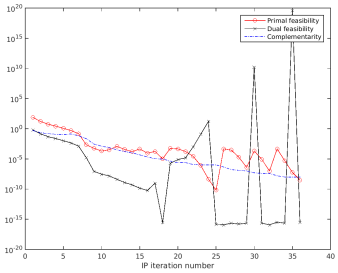
It's interesting to note that, although we solve the linear system until the relative residual (measured in the preconditioned norm) is smaller than a fixed tolerance, after a certain point the error measured in the (H, A) norm 'tracks' $\mu_k^{1/2}$. This curious behaviour is a result of the relationship between the natural norms for the interior



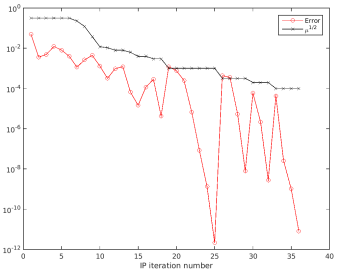
(a) Interior point convergence, tolerance of 10^{-1}



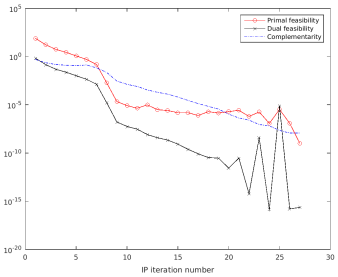
(b) $\mu_k^{1/2}$ and $\|\mathbf{e}_k\|_{H,\widehat{S}}$ for the final iterate, tolerance of 10^{-1}



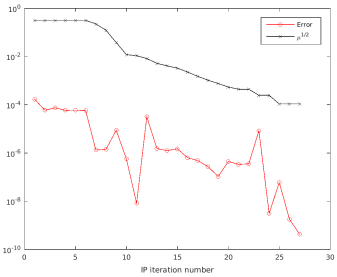
(c) Interior point convergence, tolerance of 10^{-3}



(d) $\mu_k^{1/2}$ and $\|\mathbf{e}_k\|_{H,\widehat{S}}$ for the final iterate, tolerance of 10^{-3}



(e) Interior point convergence, tolerance of 10^{-5}

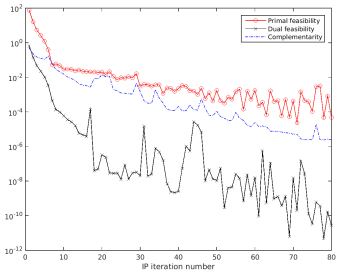


(f) $\mu_k^{1/2}$ and $\|\mathbf{e}_k\|_{H,\widehat{S}}$ for the final iterate, tolerance of 10^{-5}

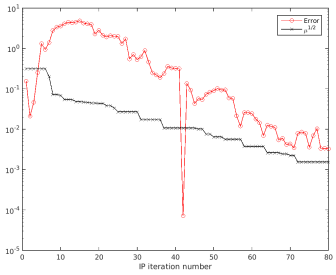
Fig. 2.2: Solving the problem `adlittle` with a range of iterative solvers

point method and that of MINRES; see Section 3.2.

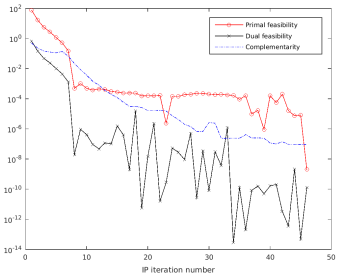
3. The iterative solution of the linear system. In the preceeding sections we derived conditions on the accuracy with which we must solve the linear system to ensure convergence of the interior point method. However, these conditions rely on measuring the error in a non-standard norm. To be practical, we need to ensure that we have available complementary iterative methods for solving the linear system.



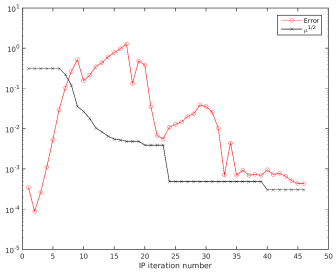
(a) Interior point convergence, tolerance of 10^{-4}



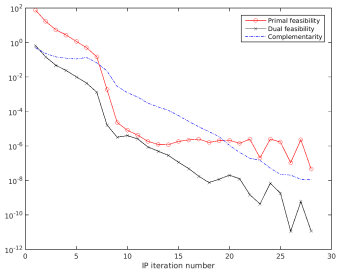
(b) $\mu_k^{1/2}$ and $\|\mathbf{e}_k\|_{H,\widehat{S}}$ for the final iterate, tolerance of 10^{-4}



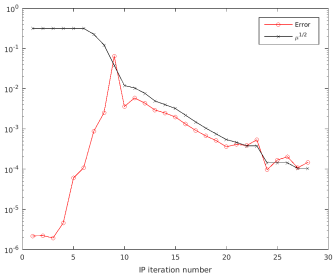
(c) Interior point convergence, tolerance of 10^{-6}



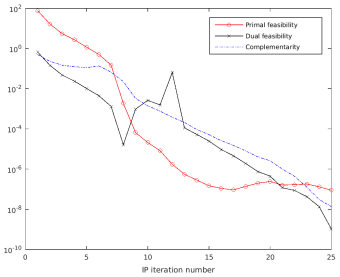
(d) $\mu_k^{1/2}$ and $\|\mathbf{e}_k\|_{H,\widehat{S}}$ for the final iterate, tolerance of 10^{-6}



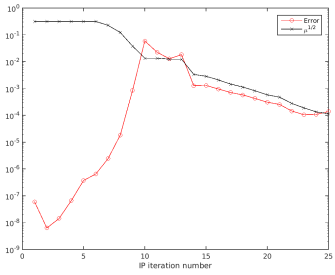
(e) Interior point convergence, tolerance of 10^{-8}



(f) $\mu_k^{1/2}$ and $\|\mathbf{e}_k\|_{H,\widehat{S}}$ for the final iterate, tolerance of 10^{-8}



(g) Interior point convergence, tolerance of 10^{-10}



(h) $\mu_k^{1/2}$ and $\|\mathbf{e}_k\|_{H,\widehat{S}}$ for the final iterate, tolerance of 10^{-10}

Fig. 2.3: Solving the problem `adlittle` with a range of iterative solvers

3.1. Solving the normal equations. Recall that a popular way of solving the linear system (2.7) is to form the normal equations (2.9) to obtain $\Delta\mathbf{y}_k^*$ and use back substitution to calculate the other unknowns. Note that, in this case, the only error in the residual will come from the $\Delta\mathbf{y}_k^*$ term, i.e., $\mathbf{r}_k^f = \mathbf{0}$.

Since the system (2.9) is symmetric positive definite, the Krylov subspace method of choice is typically preconditioned conjugate gradients (PCG), applied with a symmetric positive definite preconditioner, $\hat{S}$. Note that in this case the PCD algorithm finds the vector in the (preconditioned) Krylov subspace that minimizes

$$\|\Delta\mathbf{y}_k - \Delta\mathbf{y}_k^*\|_{B(S_k^{-1}X_k)B^T},$$

independently of the choice of preconditioner. Recall also from (2.29) that if $\|\Delta\mathbf{y}_k - \Delta\mathbf{y}_k^*\|_{B(S_k^{-1}X_k)B^T}$ is sufficiently small, then the condition on the error in the primal variable is also satisfied.

It is therefore straightforward to see that, for any choice of preconditioner, solving the system (2.9) using PCG will give convergence in a norm that satisfies the requirements of Theorem 2.7. In particular, picking a convergence tolerance by applying Proposition 2.8 will result in convergence of the interior point method.

Finally, we note that, even though PCG finds the vector that minimizes the error in the 'right' norm, it is not straightforward to calculate, and hence test, this quantity. However, an easy to calculate lower bound on this quantity is described by Strakoš and Tichý [44].

3.2. Solving the augmented system. Another choice is to solve the augmented system (2.8). A number of iterative methods are commonly used for such systems, and knowing which method is best is not always straightforward. Here we consider two Krylov methods, projected conjugate gradients and MINRES.

3.2.1. Constraint preconditioners and PPCG. Constraint precondtioning [25, 10] has proved particularly attractive within the optimization community. Such a preconditioner takes the form

$$\mathcal{P}_{con} = \begin{bmatrix} G & B^T \\ B & 0 \end{bmatrix},$$

and gets its name from the fact that the constraint block is left untouched. Such preconditioners can be applied with a projected preconditioned conjugate gradient algorithm (PPCG) [27, 18, 39], which itself—provided a primal-feasible starting vector is chosen—is equivalent to using standard PCG with a constraint preconditioner [39, 19].

This combination was shown by Rozložník and Simoncini [39] to (in exact arithmetic) always generate iterates that satisfy the constraints, which is attractive from an optimization perspective. Analyses of convergence that exploit this fact have appeared in the literature; see Section 5.2.

This is the approach used to solve the linear system iteratively in KNITRO [7] and GALAHAD [20]. HOPDM [15] also uses a variant of projected conjugate gradients with a constraint preconditioner to solve the linear systems. For a description of constraint preconditioners (both exact and inexact) for optimization problems, see, e.g., the survey by D'Apusso, De Simone, and di Serafino [8] and the references therein.

PPCG is intimately related to another method for solving (2.8), the null-space method [5, Section 6], [38]. Given any primal-feasible vector $\mathbf{w}$ and a matrix, Z, that

spans the nullspace of B, we can decompose the unknown $\Delta\mathbf{x}_k^* = Z\mathbf{z}^* + \mathbf{w}$. Thus, if we can find the unknown $\mathbf{z}^*$, which is the solution of the reduced linear system

$$Z^T(X_k^{-1}S_k)Z\mathbf{z}^* = Z^T(X_k^{-1}(\tau_k - S_k\mathbf{w}) - \zeta_k), \tag{3.1}$$

then we can reconstruct $\Delta\mathbf{x}_k^*$. Gould, Hribar and Nocedal [18] showed that PPCG applied to (2.8) with a constraint preconditioner $\mathcal{P}_{con}$ is equivalent to solving (3.1) using PCG with a preconditioner Z^TGZ. Therefore the vector $\mathbf{z} \approx \mathbf{z}^*$ is found that minimizes $\|\mathbf{z} - \mathbf{z}^*\|_{Z^TX_k^{-1}S_kZ}$ over the Krylov subspace. This, in turn, gives us an approximation to $\Delta\mathbf{x}_k^*$ by taking $\Delta\mathbf{x}_k^* = Z\mathbf{z} + \mathbf{w}$. Simple manipulation then gives that

$$\begin{aligned}\|\mathbf{z} - \mathbf{z}^*\|_{Z^TX_k^{-1}S_kZ} &= \|Z\mathbf{z} - Z\mathbf{z}^*\|_{X_k^{-1}S_k} \\ &= \|Z\mathbf{z} + \mathbf{w} - (Z\mathbf{z}^* + \mathbf{w})\|_{X_k^{-1}S_k} \\ &= \|\Delta\mathbf{x}_k - \Delta\mathbf{x}_k^*\|_{X_k^{-1}S_k}.\end{aligned}$$

This shows that constraint preconditioners applied with (projected) conjugate gradients will give a solution to the inner iteration that converges in a norm sympathetic with the outer (interior point) iteration. Furthermore, as was the case with the normal equations, this result is independent of the choice of preconditioner.

3.2.2. MINRES with $\mathcal{P}_{BD}$. The minimal residual method (MINRES) was developed by Paige and Saunders [34] for solving symmetric, indefinite linear systems $\mathcal{A}\mathbf{z} = \mathbf{b}$ using an underlying Lanczos process. Given a symmetric positive definite preconditioner, $\mathcal{P}$, the preconditioned MINRES method finds the vector that minimizes $\|\mathbf{r}_k\|_{\mathcal{P}^{-1}}$ over the preconditioned Krylov subspace. As with PCG, we aim for a preconditioner $\mathcal{P}$ that clusters the eigenvalues of $\mathcal{P}^{-1}\mathcal{A}$.

For saddle point (or KKT) systems of the form

$$\begin{bmatrix} H & B^T \\ B & 0 \end{bmatrix}\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{g} \end{bmatrix}, \tag{3.2}$$

where $H \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{m\times n}$, $m < n$, a popular paradigm is to select a preconditioner that approximates the 'ideal' block-diagonal preconditioner

$$\mathcal{P}_{\text{BD}} = \begin{bmatrix} H_0 & 0 \\ 0 & \widehat{S}_0 \end{bmatrix} \approx \begin{bmatrix} H & 0 \\ 0 & BH^{-1}B^T \end{bmatrix}.$$

Murphy, Golub and Wathen [29] showed that when $\mathcal{P}_{\text{BD}}$ is applied exactly MINRES will converge in at most three iterations.

Note that, as pointed out by Olivera and Sorensen [32], the primal Schur complement $BH^{-1}B^T$ and the normal equations (2.9) are identical. Therefore any preconditioner for the normal equations translates to a corresponding preconditioner for the augmented system. Gill et al [14] suggest a few block-diagonal preconditioners specifically for interior point methods that fit into this framework.

Recall that Lemma 2.8 requires us to have a method that minimizes $\|\mathbf{e}_k\|_{\mathcal{E}}$, where $\mathbf{e}_k$ is the error and $\mathcal{E} = \texttt{blkdiag}(H, BH^{-1}B^T)$. Now

$$\|\mathbf{r}_k\|_{\mathcal{P}^{-1}} = \|\mathcal{A}\mathbf{e}_k\|_{\mathcal{P}^{-1}} = \|\mathbf{e}_k\|_{\mathcal{A}\mathcal{P}^{-1}\mathcal{A}}.$$

Since, in finite dimensions, all vector norms are equivalent, there exist constants c, C such that

$$c\|\mathbf{e}_k\|_{\mathcal{E}} \le \|\mathbf{e}_k\|_{\mathcal{A}\mathcal{P}^{-1}\mathcal{A}} \le C\|\mathbf{e}_k\|_{\mathcal{E}}. \tag{3.3}$$

We just need to show that these constants are 'useful', in the sense that they are independent of the properties of the matrices $\mathcal{A}$ and $\mathcal{P}$.

For a general vector $\mathbf{x}$, the condition (3.3) is equivalent to finding constants $c,\ C$ such that

$$c^2\mathbf{x}^T\mathcal{E}\mathbf{x} \leq \mathbf{x}^T\mathcal{A}\mathcal{P}^{-1}\mathcal{A}\mathbf{x} \leq C^2\mathbf{x}^T\mathcal{E}\mathbf{x},$$

or, equivalently,

$$c^2\mathbf{x}^T\mathbf{x} \leq \mathbf{x}^T\mathcal{E}^{-1/2}\mathcal{A}\mathcal{P}^{-1}\mathcal{A}\mathcal{E}^{-1/2}\mathbf{x}^T \leq C^2\mathbf{x}^T\mathbf{x}.$$

The constants c and C are therefore the upper and lower bounds of the Rayleigh quotient

$$\frac{\mathbf{x}^T\mathcal{E}^{-1/2}\mathcal{A}\mathcal{P}^{-1}\mathcal{A}\mathcal{E}^{-1/2}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}. \tag{3.4}$$

We consider first the idealized case, $\mathcal{P} = \mathcal{E}$. In this case the Rayleigh quotient (3.4) is bounded above and below by the largest and smallest eigenvalues respectively of

$$\mathcal{E}^{-1}\mathcal{A}\mathcal{E}^{-1}\mathcal{A} = (\mathcal{E}^{-1}\mathcal{A})^2.$$

The eigenvalues of $\mathcal{E}^{-1}\mathcal{A}$ are 1 or $\frac{1\pm\sqrt{5}}{2}$ [29]. Hence

$$\left(\frac{1-\sqrt{5}}{2}\right)^2 \leq \frac{\mathbf{x}^T\mathcal{E}^{-1/2}\mathcal{A}\mathcal{E}^{-1}\mathcal{A}\mathcal{E}^{-1/2}\mathbf{x}^T}{\mathbf{x}^T\mathbf{x}} \leq \left(\frac{1+\sqrt{5}}{2}\right)^2.$$

Therefore in the ideal case where $\mathcal{P} = \mathcal{E}$, we will have constants

$$c = \left|\frac{1-\sqrt{5}}{2}\right| = \frac{\sqrt{5}-1}{2},\ C = \frac{1+\sqrt{5}}{2}.$$

Of course, for the vast majority of cases $\mathcal{E}$ will not be a practical preconditioner. However, note that we can write

$$\frac{\mathbf{x}^T\mathcal{E}^{-1/2}\mathcal{A}\mathcal{P}^{-1}\mathcal{A}\mathcal{E}^{-1/2}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \frac{\mathbf{y}^T\mathcal{E}^{1/2}\mathcal{P}^{-1}\mathcal{E}^{1/2}\mathbf{y}}{\mathbf{y}^T\mathbf{y}} \cdot \frac{\mathbf{x}^T\mathcal{E}^{-1/2}\mathcal{A}\mathcal{E}^{-1}\mathcal{A}\mathcal{E}^{-1/2}\mathbf{x}}{\mathbf{x}^T\mathbf{x}},$$

where $\mathbf{y} = \mathcal{E}^{-1/2}\mathcal{A}\mathcal{E}^{-1/2}\mathbf{x}$. Therefore, in general, if $\mathcal{P}$ is a matrix that is spectrally equivalent to $\mathcal{E}$, in the sense that the eigenvalues of $\mathcal{P}^{-1}\mathcal{E}$ are contained in an interval $[\psi, \Psi]$, say, where ψ, Ψ are constants close to unity, then for $\mathbf{x} \neq \mathbf{0}$ we get

$$\frac{\sqrt{\psi}(\sqrt{5}-1)}{2} \leq \frac{\|\mathbf{x}\|_{\mathcal{A}\mathcal{P}^{-1}\mathcal{A}}}{\|\mathbf{x}\|_{\mathcal{E}}} \leq \frac{\sqrt{\Psi}(1+\sqrt{5})}{2}. \tag{3.5}$$

Therefore in this case MINRES with $\mathcal{P}_{\mathrm{BD}}$ converges in the 'right' norm.

3.2.3. MINRES with $\mathcal{P}_{\rm AL}$. Another paradigm suggested for solving linear systems of the form (3.2), particularly where the (1,1) block H is singular or ill-conditioned, is to use MINRES with the augmented Lagrangian preconditioner:

$$\mathcal{P}_{\rm AL} = \begin{bmatrix} H_{\rm AL} & 0 \\ 0 & W \end{bmatrix} \approx \begin{bmatrix} H + B^T W^{-1} B & 0 \\ 0 & W \end{bmatrix},$$

where W is some easy to invert matrix, usually a scaled identity. Preconditioners of the form $\mathcal{P}_{\rm AL}$ have been studied in, e.g., [23], and in the specific context we consider here in [28, 37, 43]. Note that we can think of $\mathcal{P}_{\rm BD}$ as a preconditioner with a primal Schur complement, and $\mathcal{P}_{\rm AL}$ as a dual Schur complement preconditioner.

As in the previous section, we wish to show that also converges in a norm equivalent to $\|\cdot\|_{{\rm diag}(H,BH^{-1}B^T)}$. We would like to find constants c, C such that, for all $\mathbf{z} \in \mathbb{R}^{n+m}, \mathbf{z} \neq \mathbf{0}$,

$$c\|\mathbf{z}\|_{\mathcal{E}} \leq \|\mathbf{z}\|_{\mathcal{A}\mathcal{P}_{\rm AL}^{-1}\mathcal{A}} \leq C\|\mathbf{z}\|_{\mathcal{E}}.$$

This is equivalent to finding bounds for the generalized Rayleigh quotient

$$\frac{\mathbf{z}^T \mathcal{A}\mathcal{P}_{AL}^{-1}\mathcal{A}\mathbf{z}}{\mathbf{z}^T \mathcal{E}\mathbf{z}},$$

or equivalently bounding the eigenvalues of $\mathcal{E}^{-1}\mathcal{A}\mathcal{P}_{AL}^{-1}\mathcal{A}$, or, indeed, the similar matrix $\mathcal{P}_{AL}^{-1}\mathcal{A}\mathcal{E}^{-1}\mathcal{A}$. We will use this final form to derive the bounds.

Let $\widehat{S} := BH^{-1}B^T$. Then

$$\begin{bmatrix} H & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} H & 0 \\ 0 & \widehat{S} \end{bmatrix}^{-1} \begin{bmatrix} H & B^T \\ B & 0 \end{bmatrix} = \begin{bmatrix} H + B^T \widehat{S}^{-1} B & B^T \\ B & \widehat{S} \end{bmatrix},$$

and so we can just consider the generalized eigenvalue problem

$$\begin{bmatrix} H + B^T \widehat{S}^{-1} B & B^T \\ B & \widehat{S} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \lambda \begin{bmatrix} H + B^T W^{-1} B & 0 \\ 0 & W \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}.$$

Consider the first row of equations. First, suppose that $B^T\mathbf{y} = \mathbf{0}$, so that

$$(H + B^T \widehat{S}^{-1} B)\mathbf{x} = \lambda(H + B^T W^{-1} B)\mathbf{x}.$$

Then

$$(1 - \lambda) = \lambda \frac{\mathbf{x}^T H^{-1} B^T W^{-1} B\mathbf{x}}{\mathbf{x}^T\mathbf{x}} - \frac{\mathbf{x}^T H^{-1} B^T \widehat{S}^{-1} B\mathbf{x}}{\mathbf{x}^T\mathbf{x}}.$$

Note that the last term on the right hand side is equal to 0 or 1, since the matrix is a projection, and so we have either

$$(1 - \lambda) = \lambda\nu \Rightarrow \lambda = \frac{1}{1 + \nu},$$
$$\text{or} \qquad (1 - \lambda) = \lambda\nu - 1 \Rightarrow \lambda = \frac{2}{1 + \nu},$$

where $\nu := \nu(\mathbf{x}) = \frac{\mathbf{x}^T H^{-1} B^T W^{-1} B\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$.

Now, excluding these cases, we have that

$$(H + B^T\widehat{S}^{-1}B\mathbf{x}) + B^T\mathbf{y} = \lambda(H + B^TW^{-1}B)\mathbf{x}.$$

Since we've excluded the case where $B^T\mathbf{y} = \mathbf{0}$, we can assume that the coefficient of $\mathbf{x}$ is invertible and get

$$\mathbf{x} = -((1-\lambda)H + B^T\widehat{S}^{-1}B - \lambda B^TW^{-1}B)^{-1}B^T\mathbf{y},$$

and so substituting this into the second row we get

$$-B\left((1-\lambda)H + B^T\widehat{S}^{-1}B - \lambda B^TW^{-1}B\right)^{-1}B^T\mathbf{y} + \widehat{S}\mathbf{y} = \lambda W\mathbf{y}.$$

Note that $(B(A + B^TC^{-1}B)B^T)^{-1} = (BA^{-1}B^T)^{-1} + C^{-1}$ (see, e.g., [11, Problem 12.12]), and applying this result gives

$$\begin{aligned}\mathbf{y} &= ((2-\lambda)\widehat{S}^{-1} - \lambda W^{-1})(\widehat{S}\mathbf{y} - \lambda W\mathbf{y}) \\ &= -(2\lambda - \lambda^2)\widehat{S}^{-1}W\mathbf{y} + \lambda^2\mathbf{y} + (2-\lambda)\mathbf{y} - \lambda W^{-1}\widehat{S}\mathbf{y}.\end{aligned}$$

Multiplying through on the right by $\mathbf{y}^T$, dividing by $\mathbf{y}^T\mathbf{y}$, and setting $\eta := \eta(\mathbf{y}) = \frac{\mathbf{y}^T\widehat{S}^{-1}W\mathbf{y}}{\mathbf{y}^T\mathbf{y}}$ we get

$$\begin{aligned}(\lambda^2 - 2\lambda)\eta + \lambda^2 + (2-\lambda) - \frac{\lambda}{\eta} &= 1 \\ \text{or} \quad 2\lambda^2 - (2\eta + \frac{1}{\eta} + 1)\lambda + 1 &= 0\end{aligned}$$

and therefore

$$\lambda = \frac{(2\eta + \frac{1}{\eta} + 1) \pm \sqrt{(2\eta + \frac{1}{\eta} + 1)^2 - 8}}{4}.$$

Furthermore, we can link the two cases by noting that the spectrum of $H^{-1}B^TW^{-1}B$ satisfies $\lambda(H^{-1}B^TW^{-1}B) = \{0\} \cup \lambda(W^{-1}\widehat{S})$. We've proved the following theorem.

THEOREM 3.1. *Let* $\mathcal{P}_{\text{AL}} = \begin{bmatrix} H + B^TW^{-1}B & 0 \\ 0 & W \end{bmatrix}$ *for some* W. *Then the norm in which MINRES preconditioned with* $\mathcal{P}_{\text{AL}}$ *minimizes the error,* $\|\cdot\|_{\mathcal{A}\mathcal{P}_{\text{AL}}^{-1}\mathcal{A}}$ *is equivalent to* $\|\cdot\|_{\mathcal{E}}$, *i.e.*

$$c\|\mathbf{z}\|_{\mathcal{E}} \leq \|\mathbf{z}\|_{\mathcal{A}\mathcal{P}_{\text{AL}}^{-1}\mathcal{A}} \leq C\|\mathbf{z}\|_{\mathcal{E}}, \tag{3.6}$$

where c, C *are positive constants such that*

$$c^2 = \min\left(\frac{\eta_{\min}}{\eta_{\min}+1}, \min_{\eta\in\lambda(\widehat{S}^{-1}W)}\left(\frac{(2\eta + \frac{1}{\eta} + 1) \pm \sqrt{(2\eta + \frac{1}{\eta} + 1)^2 - 8}}{4}\right)\right)$$

$$\text{and } C^2 = \max\left(2, \max_{\eta\in\lambda(\widehat{S}^{-1}W)}\left(\frac{(2\eta + \frac{1}{\eta} + 1) \pm \sqrt{(2\eta + \frac{1}{\eta} + 1)^2 - 8}}{4}\right)\right),$$

where $\eta_{\min}$ is the smallest eigenvalue of $\widehat{S}^{-1}W$.

REMARK 3.2. *Note that the constants in Theorem 3.1 only depend on the quality of the approximation of W to $BH^{-1}B^T$. In the 'ideal' case where $W = BH^{-1}B^T$ we have the bounds*

$$\sqrt{1 - \frac{1}{\sqrt{2}}}\|\mathbf{z}\|_{\mathcal{E}} \leq \|\mathbf{z}\|_{\mathcal{A}\mathcal{P}_{\mathrm{AL}}^{-1}\mathcal{A}} \leq \sqrt{2}\|\mathbf{z}\|_{\mathcal{E}}.$$

Provided we pick a matrix W which is close enough to the 'ideal' choice of the Schur complement, then MINRES applied with the augmented Lagrangian preconditioner will also converge in a norm sympathetic to the underlying optimization problem. However, the constant of equivalence is a little harder to enumerate in this case, and in general it may be more beneficial to over-solve, rather than calculate the specific values of c and C in any given case.

Finally, we note that Morini, Simoncini, and Tani [28] recently developed an augmented Lagrangian preconditioner for (2.8) and the equivalent preconditioner for (2.7). They found that, in certain cases, the interior point method solved with the augmented system formulation (2.8) converged when an inexact solve of the 3x3 system (2.7) did not, a result that can be explained by the theory in Section 2 (in particular, the requirement that the third row of (2.7) be solved exactly).

3.3. Other preconditioners. In Sections 3.1 and 3.2 we considered four of the most popular methods. Similar arguments could be made for solving the augmented system with, e.g., GMRES with a block triangular preconditioner (such as the preconditioner proposed by Gill et al.[14]). We now highlight a few other methods for solving the linear system that are not covered by the theory in Section 2 above.

We reiterate that it is a requirement of the theory that the final equation of (2.7) be satisfied exactly, meaning the results of Section 3 will not be applicable to any method that solves the block 3×3 system (2.7).

Some preconditioners have to proposed to solve systems of the form (2.6) with MINRES or GMRES that are as yet outside the scope of the theory of Section 3. In particular, an approach that has attracted interest recently is to use an incomplete factorization on the whole indefinite system [6, 21, 41]. Orban has recently applied this approach to interior point methods [33]. The question of how such preconditioners fit into the framework of Section 3 (if at all) is future work.

3.4. Summary. We have looked at CG on the normal equations, projected CG on the augmented system, and MINRES with a Schur complement or an augmented Lagrangian preconditioner on the augmented system. For each of these common solution methods we derived the tolerance to which the iterative method should be solved to guarantee convergence of the interior point method at the same rate as when using a direct method. For clarity, we collect these tolerances in Table 3.1.

We highlight that in all cases the tolerance is inversely proportional to both $\|\mathbf{s}_k\|_1$ and $\|\mathbf{x}_k\|_1$. Very large entries in these terms will therefore subsequently require a very accurate solve; see Section 4. Enforcing an additional acceptance test on the approximate solution may safeguard against this, but we do not pursue this further here.

While the accuracy of the solve is independent of the quality of the preconditioner in the CG cases, for MINRES the less accurate the preconditioner, the more accurate the solve must be. This may be one of the reasons why the CG-based methods are

Method	Preconditioner	tolerance	best-case	reference
PCG	any	$\delta_k \mu_k^{1/2}$	$\delta_k \mu_k^{1/2}$	Sec. 3.1
PPCG	$\begin{bmatrix} G & B^T \\ B & 0 \end{bmatrix}$	$\delta_k \mu_k^{1/2}$	$\delta_k \mu_k^{1/2}$	Sec. 3.2.1
MINRES	$\begin{bmatrix} X_k^{-1} S_k & 0 \\ 0 & \widehat{S}_0 \end{bmatrix}$	$\frac{\sqrt{5}-1}{2}\sqrt{\phi}\delta_k \mu_k^{1/2}$	$\frac{\sqrt{5}-1}{2}\delta_k \mu_k^{1/2}$	Sec. 3.2.2
MINRES	$\begin{bmatrix} H_{\mathrm{AL}} & 0 \\ 0 & W \end{bmatrix}$	$\sigma_k \delta_k \mu_k^{1/2}$	$\sqrt{1-\frac{1}{\sqrt{2}}}\delta_k \mu_k^{1/2}$	Sec. 3.2.3

Table 3.1: Recommended tolerances. Here μ_k is the duality measure, $\delta_k = \frac{(\gamma)^{1/2}\epsilon}{\sqrt{2}\|\mathbf{s}_k\|_1 + \Sigma_{\max}\|\mathbf{x}_k\|_1}$ where $\epsilon \leq 1$ and $\gamma \in (0,1)$ as defined in (2.10). $\widehat{S}_0 \approx B S_k^{-1} X_k B^T$ and $H_{\mathrm{AL}} \approx X_k^{-1} S_k + B^T W^{-1} B$. $\phi = \lambda_{\min}\left(\widehat{S}_0^{-1}(B S_k^{-1} X_k B^T)\right)$, and if $\eta_{\min} = \lambda_{\min}((B S_k^{-1} X_k B^T)^{-1} W)$, then $\sigma_k^2 = \min\left(\frac{\eta_{\min}}{\eta_{\min}+1}, \min_{\eta \in \lambda((B S_k^{-1} X_k B^T)^{-1} W)}\left(\frac{(2\eta+\frac{1}{\eta}+1) \pm \sqrt{(2\eta+\frac{1}{\eta}+1)^2-8}}{4}\right)\right)$.

so popular in the optimization community. On the other hand, understanding the ideal properties of, say, a Schur complement preconditioner—namely that ϕ is known and close to one—is the first step in developing fast iterative solvers based on this paradigm in the future.

4. Numerical validation. We use a fork of Michael Saunders' PDCO interior point solver [40] as the basis of our numerical tests. In particular, we modified the code to allow the use of a custom linear solver, we removed the automatic reduction of the convergence tolerance as the iteration progresses, and we set the regularization parameters to zero.

We consider two of the cases above, the normal equations with PCG and the augmented system with MINRES and a block diagonal preconditioner, and test the tolerances predicted in Section 3.

4.1. Normal equations. First, we solve problems from the Netlib test set using preconditioned conjugate gradients. We only report results for problems where the interior point method with a direct solver converged. We used a incomplete Cholesky preconditioner with PCG, namely the HSL routine `HSL_MI28` (applied via the Matlab interface) [42, 1]. We give the results in Table 4.1. For each problem we run three tests: solving with backslash, and solving with PCG at tolerances of $10^2 \mu_k^{1/2}\delta_k$ and $\mu_k^{1/2}\delta_k$. Here, as usual, μ_k is the duality measure and $\delta_k = 1/(\sqrt{2}\|\mathbf{s}_k\|_1 + \Sigma_{\max}\|\mathbf{x}_k\|_1)$ – i.e., ϵ and γ in (2.26) are both taken to be one. We report the interior point iterations required for backslash, and the number of *extra* interior point iterations required to find the optimal solution for both the PCG tests, along with the upper and lower bounds on the tolerance required.

The numerical results in Table 4.1 are consistent with the theory in the preceding sections: of the 42 problems, solving to a tolerance two orders of magnitude higher than that predicted above resulted in failures for over 20% of the problems, whereas solving to the predicted tolerance only resulted in one failure, which in turn was due to the fact that the minimum tolerance required was unachievable, being well below machine precision. Furthermore, for the problems that were successful, the average number of extra iterations was 4.64 in the case of the looser tolerance (3.5 if we exclude

Problem	b'slash	N.E., tol $= 10^2\mu^{1/2}\delta$			N.E., tol $= \mu^{1/2}\delta$		
	# its	extra its	max(tol)	min(tol)	extra its	max(tol)	min(tol)
lp_adlittle	23	9	1.2e-01	1.7e-12	5	1.2e-03	6.9e-07
lp_afiro	13	8	4.0e-01	2.2e-04	2	4.0e-03	2.2e-06
lp_agg2	29	*	4.6e-03	3.9e-08	0	4.6e-05	5.3e-08
lp_agg3	29	*	4.6e-03	1.0e-08	0	4.6e-05	4.4e-08
lp_agg	31	*	6.0e-03	5.8e-12	0	6.0e-05	3.5e-08
lp_bandm	32	1	4.6e-04	4.1e-07	0	4.6e-06	3.4e-09
lp_beaconfd	26	*	4.7e-04	2.6e-27	0	4.7e-06	2.3e-13
lp_blend	20	7	1.1e-02	9.2e-06	0	1.1e-04	9.2e-08
lp_bnl2	57	*	7.6e-05	1.5e-23	0	7.6e-07	1.0e-13
lp_czprob	75	0	5.3e-04	2.1e-07	0	5.3e-06	2.1e-09
lp_d2q06c	59	0	9.1e-06	2.6e-11	0	9.1e-08	2.6e-13
lp_e226	37	0	1.7e-04	2.6e-11	0	1.7e-06	2.2e-13
lp_finnis	38	*	4.0e-04	5.4e-23	0	4.0e-06	9.1e-11
lp_fit1d	30	0	5.6e-06	1.8e-10	0	5.6e-08	9.7e-13
lp_fit1p	30	0	3.9e-06	2.0e-08	0	3.9e-08	2.0e-10
lp_ganges	25	41	3.2e-02	4.9e-68	2	3.2e-04	3.8e-07
lp_gfrd_pnc	30	0	8.8e-06	1.4e-09	0	8.8e-08	1.4e-11
lp_grow15	40	4	5.4e+00	6.5e-11	-1	5.4e-02	6.5e-13
lp_grow22	41	4	3.7e+00	4.4e-11	1	3.7e-02	4.4e-13
lp_grow7	34	6	1.2e+01	2.6e-10	3	1.2e-01	2.3e-12
lp_israel	42	3	9.5e-04	8.8e-07	0	9.5e-06	8.4e-09
lp_kb2	29	1	2.3e-03	6.3e-10	0	2.3e-05	6.3e-12
lp_lotfi	36	*	2.1e-04	2.0e-25	*	2.1e-06	2.1e-29
lp_sc105	17	9	5.9e-02	1.0e-07	1	5.9e-04	5.8e-07
lp_sc205	17	8	2.7e-02	9.2e-06	1	2.7e-04	8.7e-08
lp_sc50a	16	5	1.8e-01	1.9e-04	1	1.8e-03	1.3e-06
lp_sc50b	12	14	1.6e-01	5.3e-08	2	1.6e-03	2.0e-06
lp_scagr25	27	5	1.5e-02	1.1e-05	1	1.5e-04	8.1e-08
lp_scagr7	23	5	5.0e-02	8.0e-05	0	5.0e-04	6.1e-07
lp_scrs8	48	0	3.8e-05	4.6e-13	0	3.8e-07	4.6e-15
lp_scsd1	13	4	6.1e-03	1.9e-05	0	6.1e-05	4.7e-07
lp_scsd6	17	4	3.7e-03	9.0e-06	0	3.7e-05	8.7e-08
lp_scsd8	16	3	1.7e-03	4.0e-06	0	1.7e-05	3.6e-08
lp_sctap1	31	4	4.6e-04	3.1e-06	0	4.6e-06	2.4e-08
lp_sctap2	30	3	1.1e-04	2.0e-06	0	1.1e-06	1.2e-08
lp_sctap3	34	1	8.5e-05	1.7e-06	0	8.5e-07	1.3e-08
lp_share1b	41	0	1.4e-04	6.6e-09	0	1.4e-06	6.6e-11
lp_share2b	21	4	6.2e-04	6.6e-07	1	6.2e-06	3.5e-09
lp_standata	47	*	8.6e-04	1.4e-05	0	8.6e-06	5.9e-08
lp_standmps	54	*	6.7e-03	5.0e-06	0	6.7e-05	5.1e-08
lp_stocfor1	22	0	5.1e-04	1.6e-08	0	5.1e-06	1.6e-10
lp_stocfor2	39	0	1.9e-05	1.7e-09	0	1.9e-07	1.7e-11

Table 4.1: Convergence, conjugate gradients on the normal equations. Runs that reach the maximum number of iterations (80) are labelled with a '*'.

the outlier `lp_ganges`), compared to an average of 0.46 extra iterations if we use the tolerance predicted in the theory above. It's interesting to note that, since δ_k directly depends on the path taken to the optimal solution, in many cases solving to a looser tolerance initially results in requiring solves to a tighter tolerance eventually. This is notably the case for the failures `lp_bnl2, lp_finnis,` and `lp_lofti`, and probably the cause of the delayed convergence in `lp_ganges`.

4.2. Augmented system. Here we repeat the tests of Section 4.1, but solving the augmented system (2.8). We use MINRES with a block diagonal preconditioner. The Schur complement here is the same as the normal equations (a fact first pointed

out in this context by Oliveira and Sorensen [31]), and so, as in Section 4.1, we again use `HSL_MI28` [1, 42] as the Schur complement approximation. In practice, forming the Schur complement—as required by most incomplete Cholesky routines—is not advisable, as there is little (if any) benefit of doing this over solving the normal equations. However, our focus here is not on comparing the merits of particular solvers, but analysing the behaviour of the interior point method in the presence of inexact solves; to that end, we want a robust preconditioner, and `HSL_MI28` gives us that. The results are reported in Table 4.2.

There is a fundamental difference between MINRES-based methods and the CG-based methods: the norm equivalence constants of (3.5) and (3.6), for example, depend on the quality of the preconditioner. This is reflected by the results in Table 4.2. This difference is predicted by the theory in Section 3. In order to obtain the correct bound, we would need to calculate the smallest eigenvalue of the preconditioned system, which is not feasible in general. In order to take into account this extra requirement we run tests with three tolerances – $\{10^2, 1, 10^{-2}\} \times \mu_k^{1/2}\delta_k$. Note that, for convenience, we absorb the constant $(\sqrt{5}-1)/2$ required in (3.5) to δ_k. The tightest bound corresponds to a minimum eigenvalue of 10^{-4}, which seems to be sufficient for most of these problems.

With an inexact solve at the tightest tolerance, 17 of the 55 problems fail. Most of these can be explained by the fact that the theory requires a convergence tolerance that is at or around machine precision, and is therefore below the maximal attainable accuracy of MINRES. This is also true for the outliers `lp_adlittle` and `lp_gfrd_pn`, for which the interior point iteration was able to recover convergence. However, this is not the case for some of the problems; to see what's happening here we look in detail at `lp_standata` that, having a minimum tolerance of 10^{-9}, should be attainable by MINRES.

Figure 4.1 shows the eigenvalues of the Schur complement, $BS_k^{-1}X_kB^T$, and the preconditioned Schur complement for iteration 47 (the point at which we have convergence with exact solves). The eigenvalues of the Schur complement are particularly nastily distributed, ranging continuously from 10^{-20} to 10^8. The preconditioner does a good job at improving the conditioning, but the smallest eigenvalue of the preconditioned system is still around 10^{-10}, meaning the smallest scaling value used in 4.2 of 10^{-2} is too large by three orders of magnitude.

Figure 4.2 shows the convergence curves for this problem, where we vary the stopping tolerances (Figure 4.2a and Figure 4.2b show tolerances of $10^{-2}\delta_k\mu_k^{1/2}$ and $10^{-5}\delta_k\mu_k^{1/2}$ respectively). As we can see, by tightening the tolerance to that predicted by the theory, we can achieve convergence at the same rate as the exact algorithm. We remark that the tolerance of $10^{-5}\delta_k\mu_k^{1/2}$ (being a worst case bound) is in fact too tight here, and solving to a tolerance of $10^{-3}\delta_k\mu_k^{1/2}$ obtains convergence curves with the same behaviour as in Figure 4.2b.

5. Prior work. In this section we give a brief overview of the current understanding of the required accuracy of the inner solve, as given in the literature.

5.1. Approaches independent of the solution method. In the general interior point context, Freund and Jarre [12, Section 3.3], propose stopping when the norm of the residual is smaller than

$$\eta(1-\alpha_k)\max(\|\zeta_k\|, \mu_k.)$$

Problem	b'slash	N.E., tol $= 10^2\mu^{1/2}\delta$		N.E., tol $= \mu^{1/2}\delta$		N.E., tol $= 10^{-2}\mu^{1/2}\delta$	
	# its	extra its	min(tol)	extra its	min(tol)	extra its	min(tol)
lp_adlittle	23	38	3.8e-70	26	4.5e-52	18	2.3e-31
lp_afiro	13	12	1.3e-04	2	1.3e-06	0	1.3e-08
lp_agg2	29	*	1.2e-08	*	1.9e-11	0	3.3e-10
lp_agg3	29	*	1.1e-14	*	8.8e-13	2	2.7e-10
lp_agg	31	*	1.3e-13	*	9.7e-15	0	2.2e-10
lp_bandm	32	*	1.9e-09	*	2.6e-13	0	2.1e-11
lp_beaconfd	26	*	1.0e-15	*	2.0e-16	*	2.9e-20
lp_blend	20	5	5.7e-06	0	5.7e-08	0	5.7e-10
lp_bnl2	57	*	4.5e-12	*	2.0e-14	*	1.1e-16
lp_capri	44	2	3.7e-07	0	5.5e-09	0	4.4e-11
lp_czprob	75	*	1.5e-09	*	6.8e-11	0	1.3e-11
lp_d2q06c	59	*	5.0e-11	*	3.4e-13	*	3.3e-15
lp_e226	37	*	1.0e-12	*	3.2e-14	*	3.3e-17
lp_etamacro	30	*	2.0e-78	45	2.0e-73	*	2.6e-72
lp_finnis	38	1	2.1e-09	0	5.6e-11	0	5.6e-13
lp_fit1d	30	0	6.6e-11	0	6.0e-13	0	6.0e-15
lp_fit1p	30	0	1.2e-08	0	1.2e-10	0	1.2e-12
lp_ganges	25	48	6.2e-76	37	4.0e-70	0	2.0e-09
lp_gfrd_pnc	30	*	4.3e-67	*	2.2e-69	48	1.9e-71
lp_grow15	37	*	8.9e-10	*	2.8e-12	*	1.9e-14
lp_grow22	39	*	1.0e-09	24	8.3e-13	*	5.3e-14
lp_grow7	34	*	2.0e-10	*	1.3e-11	*	1.0e-13
lp_israel	42	2	5.7e-07	0	5.2e-09	0	5.2e-11
lp_kb2	29	2	4.3e-10	0	3.9e-12	0	3.9e-14
lp_lotfi	36	11	6.3e-13	5	3.4e-13	7	1.7e-15
lp_maros	48	*	4.0e-13	*	4.4e-15	*	5.2e-17
lp_perold	66	*	3.3e-10	*	1.8e-09	*	8.5e-12
lp_pilot4	59	*	7.8e-11	*	1.8e-10	0	4.3e-13
lp_pilot_we	59	*	7.0e-08	*	3.0e-10	0	3.0e-13
lp_recipe	24	*	9.2e-12	*	1.1e-13	*	1.1e-15
lp_sc105	17	30	4.9e-57	19	1.0e-31	1	3.6e-09
lp_sc205	17	29	5.1e-48	29	2.8e-50	1	6.9e-10
lp_sc50a	16	34	7.2e-68	19	2.9e-35	4	1.2e-08
lp_sc50b	12	33	9.7e-60	23	3.2e-38	8	6.0e-11
lp_scagr25	27	7	9.7e-06	0	1.0e-07	0	8.4e-10
lp_scagr7	23	5	8.6e-05	0	3.8e-07	0	3.8e-09
lp_scfxm1	34	*	4.6e-13	*	1.5e-14	*	7.2e-18
lp_scfxm2	39	*	1.2e-11	0	3.7e-13	0	3.7e-15
lp_scfxm3	40	*	9.8e-12	*	9.1e-14	*	5.3e-16
lp_scrs8	48	*	5.4e-13	*	5.3e-14	*	1.1e-16
lp_scsd1	13	3	2.7e-05	0	2.9e-07	0	2.9e-09
lp_scsd6	17	3	5.9e-06	0	5.4e-08	0	5.4e-10
lp_scsd8	16	5	3.5e-06	0	3.8e-08	0	2.2e-10
lp_sctap1	31	4	1.1e-06	0	1.5e-08	0	1.5e-10
lp_sctap2	30	1	1.5e-06	0	7.3e-09	0	7.3e-11
lp_sctap3	34	1	1.0e-06	0	7.7e-09	0	7.7e-11
lp_share1b	41	1	5.1e-09	0	4.1e-11	0	4.1e-13
lp_share2b	21	5	3.8e-07	1	6.4e-09	0	5.4e-11
lp_stair	26	6	9.9e-10	0	1.5e-08	0	1.5e-10
lp_standata	47	*	6.0e-08	*	3.0e-09	*	1.2e-09
lp_standmps	54	*	3.3e-09	*	7.1e-11	*	5.2e-11
lp_stocfor1	22	1	2.3e-08	0	9.9e-11	0	9.9e-13
lp_stocfor2	39	0	1.0e-09	0	1.0e-11	0	1.0e-13
lp_vtp_base	31	*	3.6e-12	*	1.9e-13	*	2.3e-14
lp_wood1p	65	*	3.0e-11	*	3.3e-09	1	1.0e-11

Table 4.2: Convergence, MINRES with a Schur complement preconditioner on the augmented system. Runs that reach the maximum number of iterations (80) are labelled with a '*'.

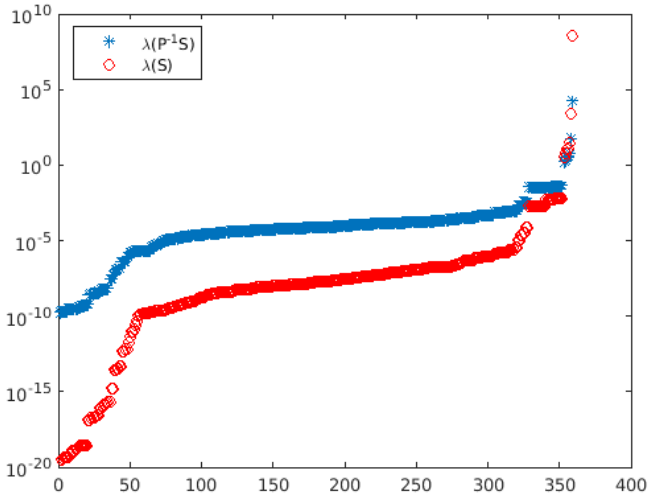
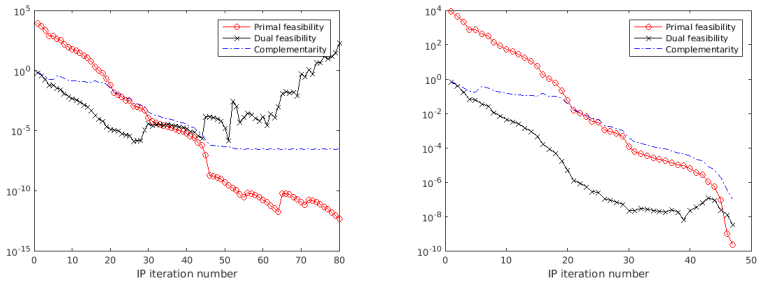


Fig. 4.1: Eigenvalues of the Schur complement and preconditioned Schur complement, `standata_eigs`



(a) `tol` $= 10^{-2}\delta_k\mu_k^{1/2}$ (b) `tol` $= 10^{-5}\delta_k\mu_k^{1/2}$

Fig. 4.2: Convergence curves for `standata`, varying the tolerance of the inner solve

For this choice of stopping criterion a reduction in the residual can be proved, and problems with convergence are guarded against, but a naive application will only guarantee linear convergence.

An approach that has been successful for getting sharper convergence estimates of convergence behaviour is to cast the problem in the framework of an inexact Newton method [9]. Note that we can rewrite (2.13) as

$$\underbrace{\begin{bmatrix} 0 & B^T & I \\ B & 0 & 0 \\ S_k & 0 & X_k \end{bmatrix}}_{F'(\mathbf{x}_k,\mathbf{y}_k,\mathbf{s}_k)} \begin{bmatrix} \Delta\mathbf{x}_k \\ \Delta\mathbf{y}_k \\ \Delta\mathbf{s}_k \end{bmatrix} = -\underbrace{\begin{bmatrix} \zeta_k \\ \xi_k \\ X_kS_k\mathbf{e} \end{bmatrix}}_{F(\mathbf{x}_k,\mathbf{y}_k,\mathbf{s}_k)} + \underbrace{\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ -\sigma_k\mu_k\mathbf{e} \end{bmatrix}}_{\mathbf{r}_k^\mu} + \underbrace{\begin{bmatrix} \mathbf{r}_k^f \\ \mathbf{r}_k^g \\ \mathbf{r}_k^h \end{bmatrix}}_{\mathbf{r}_k}, \tag{5.1}$$

where we have relaxed the assumptions of this paper by allowing an error in the third equation. If we had a true Newton method, where $\mathbf{r}_k^\mu$, then we could directly apply the theory of inexact Newton methods. In this cased we know that the method will converge (locally) if we solve the system to an accuracy such that $\mathbf{r}_k$ satisfies

$$\frac{\|\mathbf{r}_k\|}{\|F(\mathbf{x}_k, \mathbf{y}_k, \mathbf{s}_k)\|} \leq \eta_k$$

for some forcing sequence η_k, where necessarily $\eta_k < 1$. If the forcing sequence $\eta_k \to 0$ sufficiently fast, then we get superlinear convergence [9, Corollary 3.5].

However, since the interior point method is based on a perturbed Newton iteration we have to be more careful. In particular, since we only enforce complimentarity at convergence, the $\sigma_k \mu_k$ term must be treated carefully.

There are two competing properties here. First, the interior point method must converge, and so we need $\|\mathbf{r}_k^\mu + \mathbf{r}_k\|/\|F(\mathbf{x}_k, \mathbf{y}_k, \mathbf{s}_k)\| \leq \eta_k$ for some forcing sequence η_k. Note that, if the linear system is solved by a direct method, then this converges by virtue of the fact that $\mu_k \to 0$.

However, we only have explicit control over the size of the μ_k-less residual $\mathbf{r}_k$, and so we can adjust the stopping criterion to make $\|\mathbf{r}_k\|/\|F(\mathbf{x}_k, \mathbf{y}_k, \mathbf{s}_k) + \mathbf{r}_k^\mu\| \leq \epsilon_k$ for any suitable ϵ_k, but need to choose this latter tolerance to be sympathetic with the outer iteration.

Bellavia [4] showed that, since $\|F(\mathbf{x}_k, \mathbf{y}_k, \mathbf{s}_k)\|_2 \leq \frac{\mathbf{x}_k^T \mathbf{s}_k}{\sqrt{n}}$, (where n is the length of $\mathbf{x}_k$), we have

$$\|\mathbf{r}_k + \mathbf{r}_k^\mu\| \leq \|\mathbf{r}_k\| + \mu_k \sqrt{n}.$$

Therefore, assuming that μ_k satisfies $\mu_k = \beta_1 \frac{\mathbf{x}_k^T \mathbf{s}_k}{n}$, and that we use the stopping rule $\|\mathbf{r}_k\|_2 \leq \eta_k \mathbf{x}_k^T \mathbf{s}_k$, Bellavia was able to show that

$$\|\mathbf{r}_k^\mu + \mathbf{r}_k\| = (\beta_1 + \sqrt{n}\eta_k)\|F(\mathbf{x}_k, \mathbf{y}_k, \mathbf{s}_k)\|_2,$$

and so $\beta_1 + \sqrt{n}\eta_k$ can be taken to be the forcing sequence.

Since $\mathbf{x}_k$ and $\mathbf{s}_k$ approach complimentarity as the interior point converges, the stopping criterion $\mathbf{x}_k^T \mathbf{s}_k$ can be very stringent, and the additional requirements on η_k exasperate this problem. In particular, the analysis requires that $\eta_k < (1 - \beta_1)/\sqrt{n}$.

Baryamureeba and Steihaug [3] note that the choice of norm is unimportant, and if we instead consider the 1-norm then, since $\|F(\mathbf{x}_k, \mathbf{y}_k, \mathbf{s}_k)\|_1 \geq \mathbf{x}_k^T \mathbf{s}_k$, simply requiring that $\|\mathbf{r}_k\|_1 \leq \eta_k \mathbf{x}_k^T \mathbf{s}_k$ is enough to show

$$\|\mathbf{r}_k + \mathbf{r}_k^\mu\|_1 \leq (\beta_1 + \eta_k) \|F(\mathbf{x}_k, \mathbf{y}_k, \mathbf{s}_k)\|_1 .$$

We therefore remove the factor of $1/\sqrt{n}$ from the tolerance (but are measuring convergence in the unnatural 1-norm).

These approaches are nice, in they do not prescribe any specific iterative method, but they do not explain why the interior point method converges at looser convergence tolerances, as shown in the numerical experiments above.

5.2. Approaches that require primal feasibility. In our analysis above we assumed that the third equation of (2.7) must be satisfied exactly. A number of people have considered the case where, instead, the second equation is solved without error at certain points of the iteration. This is known in the optimization context as primal feasibility.

Ito, Kelley and Sachs [24] and Portugal et al [35, 36] describe methods which need to remain primal feasible once the approximate solution lies on the constraint manifold, as the primal residual lies on the right hand side of the stopping condition.

Similar to our proof of Theorem 2.7, Al-Jeiroudi and Gondzio [2] adapt the proof of [46, Theorem 6.1] to show convergence of an inexact interior point method for linear programming under the assumption that the primal residual is zero at each iteration. The theory here recommends using the stopping condition

$$\|\mathbf{r}_k^h\|_\infty \leq \eta_k \mu_k,$$

where η_k is a forcing term that is sufficiently small (see [2, Lemma 4.3 p. 243]). Gondzio [17] later extended this result to show convergence of a feasible interior point method for quadratic programs under the condition that the primal residual remains zero.

6. Conclusion. We have described conditions that the solution given by an iterative solver should satisfy in order to guarantee convergence of the interior point method. In particular, if the third equation of (2.7) is satisfied exactly, and need the error in the approximation to $[\Delta\mathbf{x}_k^*, \Delta\mathbf{y}_k^*]$ in bounded by $\delta\mu_k^{1/2}$ (for some known constant δ) in the $(X_k^{-1}S_k, BS_k^{-1}X_kB^T)$-norm, then we can guarantee convergence. This differs from what's known in the literature, since it measures the error in an iteration-dependent norm, and predicts a looser stopping tolerance.

We also show that, despite the norm being iteration-dependent, it is closely related to the natural norm in which certain Krylov subspace methods that satisfy an optimality property converge in. We show this explicitly for CG (applied to both the normal equations and the augmented system) and MINRES with block-diagonal preconditioners. We also present numerical experiments that appear to show that the theory developed is descriptive of the true convergence behaviour of the interior point method under inexact solves.

We have presented the theory exclusively for linear programming, but numerical experiments show that a similar result holds true for quadratic programs. We hope that, as has proved to be the case for solving linear systems arising from discretizations of systems of PDEs, an understanding of the norm in which the algebraic error should be naturally measured will lead to better preconditioners for this important class of problems.

REFERENCES

[1] HSL (2013)., *A collection of Fortran codes for large scale scientific computation.* http://www.hsl.rl.ac.uk.

[2] G. Al-Jeiroudi and J. Gondzio, *Convergence analysis of the inexact infeasible interior-point method for linear optimization*, Journal of optimization theory and applications **141** (2009), no. 2, 231–247.

[3] Venansius Baryamureeba and Trond Steihaug, *On the convergence of an inexact primal-dual interior point method for linear programming*, Large-Scale Scientific Computing, Springer, 2006, pp. 629–637.

[4] Stefania Bellavia, *Inexact interior-point method*, Journal of Optimization Theory and Applications **96** (1998), no. 1, 109–121.

[5] Michele Benzi, Gene H. Golub, and Jörg Liesen, *Numerical solution of saddle point problems*, Acta Numer. **14** (2005), 1–137.

[6] R. Bridson, *An ordering method for the direct solution of saddle-point matrices*, Unpublished preprint available from http://www.cs.ubc.ca/~rbridson/kktdirect/, 2007.

[7] Richard H Byrd, Jorge Nocedal, and Richard A Waltz, *KNITRO: An integrated package for nonlinear optimization*, Large-scale nonlinear optimization, Springer, 2006, pp. 35–59.

[8] Marco D'Apuzzo, Valentina De Simone, and Daniela di Serafino, *On mutual impact of numerical linear algebra and large-scale optimization with focus on interior point methods*, Computational Optimization and Applications **45** (2010), no. 2, 283–310.
[9] Ron S Dembo, Stanley C Eisenstat, and Trond Steihaug, *Inexact newton methods*, SIAM Journal on Numerical analysis **19** (1982), no. 2, 400–408.
[10] H. S. Dollar, *Iterative linear algebra for constrained optimization*, Ph.D. thesis, University of Oxford, 2005.
[11] Roger Fletcher, *Practical methods of optimization*, John Wiley & Sons, 2013.
[12] Roland W Freund and Florian Jarre, *A qmr-based interior-point algorithm for solving linear programs*, Mathematical Programming **76** (1997), no. 1, 183–210.
[13] David M Gay, *Electronic mail distribution of linear programming test problems*, Mathematical Programming Society COAL Newsletter **13** (1985), 10–12.
[14] Philip E Gill, Walter Murray, Dulce B Ponceleón, and Michael A Saunders, *Preconditioners for indefinite systems arising in optimization*, SIAM Journal on Matrix Analysis and Applications **13** (1992), no. 1, 292–311.
[15] Jacek Gondzio, *HOPDM (version 2.12) – A fast LP solver based on a primal-dual interior point method*, European Journal of Operational Research **85** (1995), no. 1, 221–225.
[16] ———, *Matrix-free interior point method*, Computational Optimization and Applications **51** (2012), no. 2, 457–480.
[17] ———, *Convergence analysis of an inexact feasible interior point method for convex quadratic programming*, SIAM Journal on Optimization **23** (2013), no. 3, 1510–1527.
[18] Nicholas I. M. Gould, Mary E. Hribar, and Jorge Nocedal, *On the solution of equality constrained quadratic programming problems arising in optimization*, SIAM J. Sci. Comput. **23** (2001), no. 4, 1376–1395.
[19] Nicholas I.M. Gould, Dominique Orban, and Tyrone Rees, *Projected Krylov methods for saddle-point systems*, SIAM J. Matrix Anal. & Appl. **35** (2014), 1329–1343.
[20] Nicholas IM Gould, Dominique Orban, and Philippe L Toint, *GALAHAD, a library of thread-safe Fortran 90 packages for large-scale nonlinear optimization*, ACM Transactions on Mathematical Software (TOMS) **29** (2003), no. 4, 353–372.
[21] C. Greif, S. He, and P. Liu, *SYM-ILDL: C++ package for incomplete factorizations of symmetric indefinite matrices*, https://github.com/inutard/matrix-factor, 2013.
[22] Chen Greif, Erin Moulding, and Dominique Orban, *Bounds on eigenvalues of matrices arising from interior-point methods*, SIAM Journal on Optimization **24** (2014), no. 1, 49–83.
[23] Chen Greif and D. Schötzau, *Preconditioners for saddle point linear systems with highly singular (1, 1) blocks*, ETNA, Special Volume on Saddle Point Problems **22** (2006), 114–121.
[24] S. Ito, C.T. Kelley, and E.W. Sachs, *Inexact primal-dual interior point iteration for linear programs in function spaces*, Computational Optimization and Applications **4** (1995), no. 3, 189–201.
[25] Carsten Keller, Nicholas I. M. Gould, and Andrew J. Wathen, *Constraint preconditioning for indefinite linear systems*, SIAM J. Matrix Anal. Appl. **21** (2000), no. 4, 1300–1317.
[26] Masakazu Kojima, Nimrod Megiddo, and Shinji Mizuno, *A primaldual infeasible-interior-point algorithm for linear programming*, Mathematical programming **61** (1993), no. 1-3, 263–280.
[27] L. Lukšan and J. Vlček, *Indefinitely preconditioned inexact newton method for large sparse equality constrained nonlinear programming problems*, Numerical linear algebra with applications **5** (1998), no. 3, 219–247.
[28] Benedetta Morini, Valeria Simoncini, and Mattia Tani, *Unreduced symmetric KKT systems arising from interior point methods. Part II: preconditioning*, Tech. report, http://www.optimization-online.org/DB_HTML/2014/07/4418.html, 2014.
[29] Malcolm F. Murphy, Gene H. Golub, and Andrew J. Wathen, *A note on preconditioning for indefinite linear systems*, SIAM J. Sci. Comput. **21** (2000), no. 6, 1969–1972.
[30] Jorge Nocedal and Stephen J. Wright, *Numerical optimization*, Springer, 1999.
[31] A.R.L. Oliveira and D.C. Sorensen, *Computational experience with a preconditioner for interior point methods for linear programming*, Tech. Report CRPC-TR97772, Center for Research on Parallel Computation, Rice University, 1997.
[32] Aurelio R.L. Oliveira and Danny C. Sorensen, *A new class of preconditioners for large-scale linear systems from interior point methods for linear programming*, Linear Algebra and its applications **394** (2005), 1–24.
[33] Dominique Orban, *Limited-memory LDL^T factorization of symmetric quasi-definite matrices with application to constrained optimization*, Numerical Algorithms (2014), 1–33.
[34] C. C. Paige and M. A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal. **12** (1975), no. 4, 617–629.
[35] Luis Portugal, Luis Fernandes, and Joaquim Judice, *A truncated Newton interior-point algo-*

rithm for the solution of a multicommodity spatial equilibrium model, Tech. report, Department of Mathematics, University of Coimbra, 1995.

[36] Luis F. Portugal, Mauricio G.C. Resende, Geraldo Veiga, and J.J. Júdice, *A truncated primal-infeasible dual-infeasible network interior point method*, Networks **35** (2000), no. 2, 91–108.

[37] T. Rees and C. Greif, *A preconditioner for linear systems arising from interior point optimization methods*, SIAM Journal on Scientific Computing **29** (2010), no. 5, 1992.

[38] Tyrone Rees and Jennifer Scott, *The null-space method and its relationship with matrix factorizations for sparse saddle point systems*, Tech. Report RAL-TR-2014-016, STFC Rutherford Appleton Laboratory, 2014.

[39] M. Rozložník and V. Simoncini, *Krylov subspace methods for saddle point problems with indefinite preconditioning*, SIAM journal on matrix analysis and applications **24** (2002), no. 2, 368–391.

[40] M. Saunders et al., *PDCO: A primal-dual barrier method for convex objectives*, DOI: http://dx.doi.org/10.5281/zenodo.31159.

[41] J.A. Scott and M. Tůma, *On signed incomplete cholesky factorization preconditioners for saddle-point systems*, SIAM Journal on Scientific Computing **36** (2014), no. 6, A2984–A3010.

[42] Jennifer Scott and Miroslav Tůma, *On positive semidefinite modification schemes for incomplete Cholesky factorization*, SIAM Journal on Scientific Computing **36** (2014), no. 2, A609–A633.

[43] Amir Shahzad, Eric C. Kerrigan, and George A. Constantinides, *Preconditioners for inexact interior point methods for predictive control*, American Control Conference (ACC), 2010, IEEE, 2010, pp. 5714–5719.

[44] Zdeněk Strakoš and Petr Tichý, *On estimation of the A-norm of the error in CG and PCG*, PAMM **3** (2003), no. 1, 553–554.

[45] Margaret H. Wright, *Ill-conditioning and computational error in interior methods for nonlinear programming*, SIAM Journal on Optimization **9** (1998), no. 1, 84–111.

[46] Stephen J Wright, *Primal-dual interior-point methods*, vol. 54, SIAM, 1997.