Maintaining a Basis Matrix in the Linear Programming Interior Point Method

Lukas Schork* Jacek Gondzio†

School of Mathematics, University of Edinburgh, Edinburgh EH9 3FD, Scotland, UK

Technical Report ERGO 17-009, September 14, 2017

Abstract

To precondition the normal equation system from the linear programming (LP) interior point method, basis preconditioners choose a basis matrix dependent on column scaling factors. Two criteria for choosing the basis matrix are compared which yield a maximum volume or maximum weight basis. Finding a maximum volume basis requires a combinatorial effort, but it gives a stronger bound on the condition number of the preconditioned normal matrix than the maximum weight basis. It is shown that neither of the two bases need to become an optimal LP basis as the interior point method approaches a solution. A crossover algorithm is presented to recover an optimal LP basis.

1 Introduction

The key ingredient in the linear programming interior point method is solving the normal equation system

$$AD^2A^T\Delta\boldsymbol{y} = \boldsymbol{r}, \tag{1}$$

where A is an $m \times n$ matrix of rank $m \leq n$ (often $m \ll n$) and D is a positive definite diagonal matrix that changes in every iteration. To precondition (1) for an iterative linear solver, basis preconditioners approximate AD^2A^T by $A_\mathcal{B}D_\mathcal{B}^2A_\mathcal{B}^T$, where $A_\mathcal{B}$ is a nonsingular matrix composed of m columns of A, denoted *basis matrix*, and $D_\mathcal{B}$ is the corresponding $m \times m$ submatrix of D. This approach is justified computationally because usually $A_\mathcal{B}$ can be factorized much faster than the normal matrix itself.

Building a concrete preconditioner of this form requires a criterion for choosing the basic columns. The obvious way is to select columns in decreasing order of the diagonal elements of D until m linearly independent columns are found. A basis selected by this criterion is denoted *maximum weight basis* and as a preconditioner is analysed in [4]. It is implemented in [5] in conjunction with the augmented system matrix from an interior point method.

A different criterion for choosing the basic columns is found in the literature which yields a *maximum volume basis* [3]. It is analysed and tested in [1] for preconditioning an iterative method for least squares problems. The maximum volume basis gives a better bound on the condition number of the preconditioned normal matrix than the maximum weight basis, but selecting the columns is a combinatorial problem.

*L.Schork@ed.ac.uk

†J.Gondzio@ed.ac.uk

This paper reviews the maximum volume basis as a preconditioner for the normal equation system and compares it to the maximum weight basis. It explains the difference between the bounds on the condition number of the preconditioned matrices. In conjunction with interior point methods it is shown that the maximum volume basis computed for a point on the central path is invariant to a row and column scaling of the LP, whereas the maximum weight basis is invariant to a row scaling only.

The paper also addresses the problem of recovering an LP basic solution from an interior solution. Maintaining a basis matrix in every interior point iteration raises the question if that basis becomes an optimal LP basis as the interior point method converges. It is shown that this need not be the case for the maximum volume and the maximum weight basis. However, they eventually attain a certain property that we call *correct degeneracy.* A simplex-type algorithm is presented to recover an optimal LP basis from a basis with correct degeneracy.

The following notations are used. $\boldsymbol{e}_j$ is the j-th unit vector in $\mathbb{R}^n$. When $\boldsymbol{x}$ is a vector, then $\text{diag}(\boldsymbol{x})$ is the diagonal matrix with entries x_i on the diagonal. When there is a vector $\boldsymbol{x}$ and a matrix X, then $X = \text{diag}(\boldsymbol{x})$. When $\mathcal{B}$ is an index set, then $x_{\mathcal{B}}$ is the vector with entries x_i for $i \in \mathcal{B}$ and $X_{\mathcal{B}} = \text{diag}(\boldsymbol{x}_{\mathcal{B}})$. When a matrix A does not have a corresponding lower case letter, then A_j is the j-th column of A and $A_{\mathcal{B}}$ is the matrix composed of columns A_j for $j \in \mathcal{B}$. Expressions like $|A|$, $|\boldsymbol{x}|$, $A \leq 1$ and $\boldsymbol{x} \leq 1$ are meant componentwise. $\|\cdot\|_2$, $\|\cdot\|_F$, $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the two-norm, the Frobenius norm, the smallest and the largest eigenvalue of a matrix.

2 Background

The linear program is stated in standard form of a primal-dual pair

$$\text{minimize } \boldsymbol{c}^T\boldsymbol{x} \quad \text{subject to } A\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geq 0, \tag{2a}$$

$$\text{maximize } \boldsymbol{b}^T\boldsymbol{y} \quad \text{subject to } A^T\boldsymbol{y} + \boldsymbol{z} = \boldsymbol{c},\ \boldsymbol{z} \geq 0, \tag{2b}$$

in which A is an $m \times n$ matrix of full row rank. It is assumed that a strictly feasible primal-dual point exists, i. e. a point $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ statisfying

$$A\boldsymbol{x} = \boldsymbol{b}, \quad A^T\boldsymbol{y} + \boldsymbol{z} = \boldsymbol{c}, \quad (\boldsymbol{x}, \boldsymbol{z}) > 0,$$

which is a sufficient condition for (2a) and (2b) to have optimal solutions.

The assumption of a strictly feasible primal-dual point implies the existence of the (feasible) central path $\mathcal{C}$, which is the set of points $\big(\boldsymbol{x}_\mu, \boldsymbol{y}_\mu, \boldsymbol{z}_\mu\big)$ parametrized by $\mu > 0$ that satisfy

$$\begin{aligned} A\boldsymbol{x}_\mu &= \boldsymbol{b}, \\ A^T\boldsymbol{y}_\mu + \boldsymbol{z}_\mu &= \boldsymbol{c}, \\ (\boldsymbol{x}_\mu, \boldsymbol{z}_\mu) &> 0, \\ (\boldsymbol{x}_\mu)_i(\boldsymbol{z}_\mu)_i &= \mu, \quad i = 1, \ldots, n. \end{aligned}$$

The sequence $\big(\boldsymbol{x}_\mu, \boldsymbol{y}_\mu, \boldsymbol{z}_\mu\big)$ has at least one limit point as $\mu \to 0$. Each limit point is a strictly complementary solution $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{z}^*)$ to (2), i. e. $\boldsymbol{x}^* + \boldsymbol{z}^* > 0$. The partition into $x_i^* > 0$ and $x_i^* = 0$ is identical for all strictly complementary solutions. See [8] for standard results about linear programming.

A primal-dual interior point method generates a sequence of points which lie in some vicinity of the central path. It computes the step directions from the normal equation system in which the diagonal matrix $D = \text{diag}(\sqrt{x_i/z_i})$ is defined by the current point. The results presented in this paper about preconditioners hold true for linear systems of the form (1) in which D is any positive definite diagonal matrix. For the results in conjunction with interior point methods it will be assumed that D is the scaling matrix corresponding to a point on the central path.

A basis $\mathcal{B}$ is a set of m column indices such that $A_{\mathcal{B}}$ is nonsingular. Associated with $\mathcal{B}$ is the nonbasic set $\mathcal{N} = \{1, \dots, n\} \setminus \mathcal{B}$ and the vertex $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$ satisfying primal and dual equations whilst $\hat{\boldsymbol{x}}_{\mathcal{N}} = 0$ and $\hat{\boldsymbol{z}}_{\mathcal{B}} = 0$. The vertex is primal feasible when $\hat{\boldsymbol{x}}_{\mathcal{B}} = A_{\mathcal{B}}^{-1}\boldsymbol{b} \geq 0$ and dual feasible when $\hat{\boldsymbol{z}}_{\mathcal{N}} = \boldsymbol{c}_{\mathcal{N}} - A_{\mathcal{N}}^T A_{\mathcal{B}}^{-T} \boldsymbol{c}_{\mathcal{B}} \geq 0$. The basis is optimal when the vertex is primal and dual feasible. The basis is primal degenerate when $\hat{x}_i = 0$ for some $i \in \mathcal{B}$ and dual degenerate when $\hat{z}_i = 0$ for some $i \in \mathcal{N}$.

For a basis $\mathcal{B}$ and a positive definite diagonal matrix D we denote $A_{\mathcal{B}}^{-1} A_{\mathcal{N}}$ *tableau matrix* and $D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} D_{\mathcal{N}}$ *scaled tableau matrix*. When $D = \operatorname{diag}(\sqrt{(\boldsymbol{x}_\mu)_i / (\boldsymbol{z}_\mu)_i})$ for a point $(\boldsymbol{x}_\mu, \boldsymbol{y}_\mu, \boldsymbol{z}_\mu) \in \mathcal{C}$, then the scaled tableau matrix can be expressed alternatively as

$$D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} D_{\mathcal{N}} = \frac{1}{\sqrt{\mu}} D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} D_{\mathcal{N}} \sqrt{\mu} = X_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} X_{\mathcal{N}},$$

$$D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} D_{\mathcal{N}} = \sqrt{\mu} D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} D_{\mathcal{N}} \frac{1}{\sqrt{\mu}} = Z_{\mathcal{B}} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} Z_{\mathcal{N}}^{-1}.$$

3 Maximum Volume Basis Preconditioner

When the interior point method converges, the diagonal elements of $D = \operatorname{diag}(\sqrt{x_i/z_i})$ tend to zero or to infinity. When there are less than m linearly independent columns of A whose scaling factors become large, then the condition number of the normal matrix becomes unbounded. In practice the normal equation system in an advanced interior point iteration almost always requires preconditioning to be tractable for an iterative linear solver.

Basis preconditioners choose a basis $\mathcal{B}$ and transform the normal equation system into

$$D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A D^2 A^T A_{\mathcal{B}}^{-T} D_{\mathcal{B}}^{-1} \Delta \boldsymbol{u} = D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} \boldsymbol{r}, \tag{4}$$

from which the solution to the original system is recovered by $\Delta \boldsymbol{y} = A_{\mathcal{B}}^{-T} D_{\mathcal{B}}^{-1} \Delta \boldsymbol{u}$. The challenge with basis preconditioning is to choose $\mathcal{B}$ to get a small condition number of the preconditioned system matrix. Writing this matrix as

$$C = I + D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} D_{\mathcal{N}}^2 A_{\mathcal{N}}^T A_{\mathcal{B}}^{-T} D_{\mathcal{B}}^{-1},$$

it is obvious that $\lambda_{\min}(C) \geq 1$ and that $\lambda_{\max}(C)$ can be bounded by bounding the entries in the scaled tableau matrix. It will be shown below that $|D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} D_{\mathcal{N}}| \leq \rho$ is equivalent to $A_{\mathcal{B}} D_{\mathcal{B}}$ having *local ρ-maximum volume* in AD. We use the definition of maximum volume given in [6] except that we restrict the submatrix to be square and nonsingular.

Definition 3.1. *For an $m \times m$ matrix B define $\operatorname{vol}(B) = |\det B|$. Let $A \in \mathbb{R}^{m \times n}$ have rank m and B be a square nonsingular matrix formed by m columns of A. Let $\rho \geq 1$. $\operatorname{vol}(B)$ is said to be a local ρ-maximum volume in A if*

$$\operatorname{vol}(B') \leq \rho \operatorname{vol}(B)$$

for any B' that is obtained by replacing one column of B by a column of A not in B.

In this paper maximum volume is always meant locally, i. e. we only consider increasing the volume by changing to a neighbouring basis. The next lemma proves the relation between a maximum volume submatrix and the magnitude of entries in the tableau matrix.

Lemma 3.2. *Let $A \in \mathbb{R}^{m \times n}$ have rank m and B be a nonsingular matrix formed by m columns of A. Let $\rho \geq 1$. Then B has local ρ-maximum volume in A if and only if $|B^{-1}A| \leq \rho$.*

Proof. Let B' be obtained by replacing column p of B by column A_q not in B. Denoting $\boldsymbol{v} = B^{-1}A_q$, then

$$(B')^{-1} = \begin{bmatrix} 1 & & -v_1/v_p & & \\ & \ddots & \vdots & & \\ & & v_p^{-1} & & \\ & & \vdots & \ddots & \\ & & -v_m/v_p & & 1 \end{bmatrix} B^{-1}.$$

It follows from the properties of determinants that $\mathrm{vol}(B'^{-1}) = v_p^{-1}\,\mathrm{vol}(B^{-1})$ and $\mathrm{vol}(B') = v_p\,\mathrm{vol}(B)$. Therefore

$$\begin{aligned} |B^{-1}A| \le \rho \iff & \quad |v_p| \le \rho \text{ for all possible choices of } p \text{ and } q \\ \iff & \quad \mathrm{vol}(B') \le \rho\,\mathrm{vol}(B) \text{ for all } B' \text{ obtained by replacing} \\ & \quad \text{one column of } B \text{ by one column of } A \text{ not in } B \\ \iff & \quad B \text{ has local } \rho\text{-maximum volume in } A. \end{aligned}$$

□

A similar result is proved in [3], Lemma 1, just that in their definition maximum volume (with $\rho = 1$) is meant globally and hence it implies that $|B^{-1}A| \le 1$ but not vice versa.

For a ρ-maximum volume basis matrix $A_{\mathcal{B}}D_{\mathcal{B}}$ in AD we obtain the following bound on the condition number of C. First

$$\left\| D_{\mathcal{B}}^{-1}A_{\mathcal{B}}^{-1}A_{\mathcal{N}}D_{\mathcal{N}} \right\|_2^2 \le \left\| D_{\mathcal{B}}^{-1}A_{\mathcal{B}}^{-1}A_{\mathcal{N}}D_{\mathcal{N}} \right\|_F^2 \le \rho^2 m(n-m).$$

Therefore

$$\lambda_{\max}(C) = 1 + \left\| D_{\mathcal{B}}^{-1}A_{\mathcal{B}}^{-1}A_{\mathcal{N}}D_{\mathcal{N}}^2A_{\mathcal{N}}^TA_{\mathcal{B}}^{-T}D_{\mathcal{B}}^{-1} \right\|_2 \le 1 + \rho^2 m(n-m). \tag{5}$$

Since $\lambda_{\min}(C) \ge 1$, $1 + \rho^2 m(n-m)$ also bounds the two-norm condition number of C.

The algorithm below gives a constructive proof that for any matrix AD of rank m there exists a 1-maximum volume basis. That means, there exists a basis which bounds the condition number of C by $1 + m(n-m)$ regardless of the numerical values of A and the scaling factors D.

[3] presents an algorithm to find a maximum volume basis. The algorithm repeatedly exchanges the nonbasic column with maximum tableau entry into the basis until all tableau entries are bounded by 1 in magnitude. Algorithm 1 is a simple modification which does not compute the entire tableau after each basis update. The algorithm terminates in a finite number of iterations because each pivot operation increases $\mathrm{vol}(A_{\mathcal{B}})$ so that a basis cannot repeat.

Finding a maximum volume basis by Algorithm 1 is not practical since it requires repeatedly computing the tableau matrix. The concept might be used in practice, however, by employing heuristics to choose candidate columns to enter the basis and applying lines 5–10 of Algorithm 1 only to those columns.

In linear programming, when D is the scaling matrix corresponding to a point on the central path, the volume of any basis matrix of AD is independent of the scaling of the LP. Assume that $R \in \mathbb{R}^{m\times m}$ and $S \in \mathbb{R}^{n\times n}$ are positive definite diagonal matrices and consider the LP with $\tilde{A} = RAS$, $\tilde{\boldsymbol{b}} = R\boldsymbol{b}$ and $\tilde{\boldsymbol{c}} = S\boldsymbol{c}$ instead of A, $\boldsymbol{b}$ and $\boldsymbol{c}$. Then the central path $\mathcal{C}$ is scaled to $\tilde{\mathcal{C}}$ by

$$(\boldsymbol{x}_\mu, \boldsymbol{y}_\mu, \boldsymbol{z}_\mu) \in \mathcal{C} \implies (S^{-1}\boldsymbol{x}_\mu, R^{-1}\boldsymbol{y}_\mu, S\boldsymbol{z}_\mu) \in \tilde{\mathcal{C}}$$

Algorithm 1 maxvolume

Require: $A \in \mathbb{R}^{m \times n}$, $\mathcal{B}$ such that $A_{\mathcal{B}}$ is nonsingular, $\rho \geq 1$

repeat
  finished = **true**
  for $j = 1$ **to** n **do**
    if $j \notin \mathcal{B}$ **then**
      $\boldsymbol{v} = A_{\mathcal{B}}^{-1} A_j$
      $p = \arg\max_i \{|v_i|\}$
      if $|v_p| > \rho$ **then**
        $\mathcal{B}_p = j$
        finished = **false**
      end if
    end if
  end for
until finished

and the scaling factors at a central point become $\tilde{D} = DS^{-1}$. Hence the scaled tableau

$$\tilde{D}_{\mathcal{B}}^{-1} \tilde{A}_{\mathcal{B}}^{-1} \tilde{A}_{\mathcal{N}} \tilde{D}_{\mathcal{N}} = S_{\mathcal{B}} D_{\mathcal{B}}^{-1} S_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} R^{-1} R A_{\mathcal{N}} S_{\mathcal{N}} D_{\mathcal{N}} S_{\mathcal{N}}^{-1} = D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} D_{\mathcal{N}}$$

remains unchanged and the volume of all basis matrices remains unchanged.

4 Maximum Weight Basis Preconditioner

A maximum weight basis is defined for a matrix $A \in \mathbb{R}^{m \times n}$ of rank m and column scaling factors d_j for $1 \leq j \leq n$. It treats A and $\boldsymbol{d}$ separately. Algorithm 2, cited from [4], determines a maximum weight basis.

Algorithm 2 maxweight

Require: $A \in \mathbb{R}^{m \times n}$ of rank m, $\boldsymbol{d} \in \mathbb{R}^n$, $\boldsymbol{d} > 0$

Order the elements of $\boldsymbol{d}$ such that $d_1 \geq \cdots \geq d_n$; order the columns of A accordingly
Let $\mathcal{B} = \emptyset$ and $l = 1$
while $\mathcal{B}$ contains less than m indices **do**
  If A_l is linearly independent of $A_{\mathcal{B}}$, set $\mathcal{B} = \mathcal{B} \cup \{l\}$.
  $l = l + 1$
end while

[4] analyses the maximum weight basis as a preconditioner for the normal equation system. It is seen from the proof of Lemma 2.1 in that paper that if $\mathcal{B}$ is a maximum weight basis, then every entry in the scaled tableau is bounded by its corresponding entry in the unscaled tableau. In particular this yields the bound

$$\left\| D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_j d_j \right\|_2 \leq \left\| A_{\mathcal{B}}^{-1} A_j \right\|_2 \quad \text{for } j \in \mathcal{N}. \tag{6}$$

It is concluded in [4] that

$$\lambda_{\max}(D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A D^2 A^T A_{\mathcal{B}}^{-T} D_{\mathcal{B}}^{-1}) \leq m \bar{\chi}_A^2, \tag{7}$$

where the quantity $\bar{\chi}_A$ can be defined by

$$\bar{\chi}_A = \max_{\mathcal{B} \text{ basis}} \{ \left\| A_{\mathcal{B}}^{-1} A \right\|_2 \}. \tag{8}$$

Because there are only a finite number of basis matrices of A, $\bar{\chi}_A$ is finite.

As in the previous section, the right-hand side in (7) also bounds the two-norm condition number of the preconditioned normal matrix. The bound is independent of D, but not independent of the numerical values of A. In general, it is weaker than that of the maximum volume basis because $\bar{\chi}_A$ is not bounded polynomially in the dimension of A. Using the well known example of Peters and Wilkinson [7] to construct

$$A_{\mathcal{B}} = \begin{bmatrix} 1 & -1 & & \cdots & -1 \\ & 1 & & & -1 \\ & & \ddots & & \vdots \\ & & & 1 & -1 \\ & & & & 1 \end{bmatrix}, \quad A_{\mathcal{N}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix},$$

it can be verified that $\left\|A_{\mathcal{B}}^{-1}A\right\|_2$ grows exponentially with m. We have observed similar effects in real-world linear programs when A has certain structures, originating for example from the discretization of PDEs.

The difference between the maximum weight and the maximum volume basis is that the first one does not take the values of $A_{\mathcal{B}}^{-1}A_{\mathcal{N}}$ into account. This simplification allows a maximum weight basis to be computed at reasonable cost, but the norm of the resulting tableau matrix can be as large as $\bar{\chi}_A$. When A is totally unimodular, then the nonzero entries in the tableau matrix are ± 1 and a maximum weight basis is also a maximum volume basis. For the special case that A is the node-arc incidence matrix of a graph, [4] already derived the bound from the maximum volume basis for the maximum weight basis.

In linear programming, a maximum weight basis depends on the column scaling of the LP. Assume that $S \in \mathbb{R}^{n\times n}$ is a positive definite diagonal matrix and consider the LP with $\tilde{A} = AS$ and $\tilde{\boldsymbol{c}} = S\boldsymbol{c}$ instead of A and $\boldsymbol{c}$. Then the scaling factors at a central point become $\tilde{D} = DS^{-1}$. Hence, given any basis $\mathcal{B}$ and $\mu > 0$ defining a central point, there exists a column scaling that makes $\mathcal{B}$ a maximum weight basis at that point, despite the scaled tableau matrix remaining unchanged. A row scaling of the LP obviously does not affect a maximum weight basis.

5 Recovering an Optimal Basis

The procedure that computes an LP basic solution from an (approximate) interior solution is called crossover. Conventionally, it constructs a starting basis guided by the ratio of primal to dual slack variables and updates the basis by simplex-type operations until it is optimal [2]. When the interior point method employs an iterative linear solver with basis preconditioning, then it is appealing to use the basis from the final interior point iteration as starting basis for the crossover. This section presents a specialized crossover procedure that starts from a maximum volume or a maximum weight basis.

Throughout this section $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ is a point on the central path with scaling matrix D, and $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$ is the vertex corresponding to the basis $\mathcal{B}$. Note that $(\boldsymbol{x}, \boldsymbol{z}) > 0$ and $\hat{\boldsymbol{x}}_{\mathcal{N}} = 0$, $\hat{\boldsymbol{z}}_{\mathcal{B}} = 0$.

To begin, the following lemma states relations between any central point and the vertex solution corresponding to any basis.

Lemma 5.1. *Let $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in \mathcal{C}$, $D = \operatorname{diag}(\sqrt{x_i/z_i})$. Let $\mathcal{B}$ be a basis with vertex $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$. Then*

(i) $D_{\mathcal{B}}^{-1}A_{\mathcal{B}}^{-1}A_{\mathcal{N}}D_{\mathcal{N}}\boldsymbol{e}_{\mathcal{N}} = X_{\mathcal{B}}^{-1}\hat{\boldsymbol{x}}_{\mathcal{B}} - \boldsymbol{e}_{\mathcal{B}}$,

(ii) $\boldsymbol{e}_{\mathcal{B}}^T D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} D_{\mathcal{N}} = \boldsymbol{e}_{\mathcal{N}}^T - \hat{\boldsymbol{z}}_{\mathcal{N}}^T Z_{\mathcal{N}}^{-1}$,

(iii) $\sum_{i\in\mathcal{B}} \hat{x}_i (x_i)^{-1} + \sum_{j\in\mathcal{N}} \hat{z}_j (z_j)^{-1} = n$.

Proof. Part (i) and (ii) are computed from

$$\begin{aligned}
D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} D_{\mathcal{N}} \boldsymbol{e}_{\mathcal{N}} &= X_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} X_{\mathcal{N}} \boldsymbol{e}_{\mathcal{N}} \\
&= X_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} \boldsymbol{x}_{\mathcal{N}} \\
&= X_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} (\boldsymbol{b} - A_{\mathcal{B}} \boldsymbol{x}_{\mathcal{B}}) \\
&= X_{\mathcal{B}}^{-1} \hat{\boldsymbol{x}}_{\mathcal{B}} - \boldsymbol{e}_{\mathcal{B}}, \\
\boldsymbol{e}_{\mathcal{B}}^T D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} D_{\mathcal{N}} &= \boldsymbol{e}_{\mathcal{B}}^T Z_{\mathcal{B}} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} Z_{\mathcal{N}}^{-1} \\
&= \boldsymbol{z}_{\mathcal{B}}^T A_{\mathcal{B}}^{-1} A_{\mathcal{N}} Z_{\mathcal{N}}^{-1} \\
&= (\boldsymbol{c}_{\mathcal{B}}^T A_{\mathcal{B}}^{-1} - \boldsymbol{y}^T) A_{\mathcal{N}} Z_{\mathcal{N}}^{-1} \\
&= (\boldsymbol{c}_{\mathcal{B}}^T A_{\mathcal{B}}^{-1} A_{\mathcal{N}} - \boldsymbol{c}_{\mathcal{N}}^T + \boldsymbol{z}_{\mathcal{N}}^T) Z_{\mathcal{N}}^{-1} \\
&= \boldsymbol{e}_{\mathcal{N}}^T - \hat{\boldsymbol{z}}_{\mathcal{N}}^T Z_{\mathcal{N}}^{-1}.
\end{aligned}$$

Part (iii) is obtained by setting equal the two expressions

$$\begin{aligned}
\boldsymbol{e}_{\mathcal{B}}^T D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} D_{\mathcal{N}} \boldsymbol{e}_{\mathcal{N}} &\overset{(i)}{=} \boldsymbol{e}_{\mathcal{B}}^T X_{\mathcal{B}}^{-1} \hat{\boldsymbol{x}}_{\mathcal{B}} - \boldsymbol{e}_{\mathcal{B}}^T \boldsymbol{e}_{\mathcal{B}}, \\
\boldsymbol{e}_{\mathcal{B}}^T D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} D_{\mathcal{N}} \boldsymbol{e}_{\mathcal{N}} &\overset{(ii)}{=} \boldsymbol{e}_{\mathcal{N}}^T \boldsymbol{e}_{\mathcal{N}} - \hat{\boldsymbol{z}}_{\mathcal{N}}^T Z_{\mathcal{N}}^{-1} \boldsymbol{e}_{\mathcal{N}}.
\end{aligned}$$

□

It is insightful to compare an optimal basis to a maximum volume basis in terms of the scaled tableau matrix. Because $\boldsymbol{x}_{\mathcal{B}}$ and $\boldsymbol{z}_{\mathcal{N}}$ are positive, a basis is primal feasible if $X_{\mathcal{B}}^{-1} \hat{\boldsymbol{x}}_{\mathcal{B}} \geq 0$ and dual feasible if $Z_{\mathcal{N}}^{-1} \hat{\boldsymbol{z}}_{\mathcal{N}} \geq 0$. By means of Lemma 5.1 (i) and (ii) this is equivalent to $D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} D_{\mathcal{N}} \boldsymbol{e}_{\mathcal{N}} \geq -\boldsymbol{e}_{\mathcal{B}}$ and $\boldsymbol{e}_{\mathcal{B}}^T D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} D_{\mathcal{N}} \leq \boldsymbol{e}_{\mathcal{N}}^T$; i.e. the row sums of the scaled tableau matrix are bounded from below by -1 and the column sums are bounded from above by 1. Rewriting these conditions as

$$X_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} \boldsymbol{x}_{\mathcal{N}} \geq -\boldsymbol{e}_{\mathcal{B}}, \quad -\boldsymbol{z}_{\mathcal{B}}^T A_{\mathcal{B}}^{-1} A_{\mathcal{N}} Z_{\mathcal{N}}^{-1} \geq -\boldsymbol{e}_{\mathcal{N}}^T,$$

their meaning becomes clear. $A_{\mathcal{B}}^{-1} A_{\mathcal{N}} \boldsymbol{x}_{\mathcal{N}}$ is the change in the primal basic variables when the primal nonbasic variables are set to zero whilst satisfying $A\boldsymbol{x} = \boldsymbol{b}$. When this change is bounded from below by $-\boldsymbol{x}_{\mathcal{B}}$ then the basic variables remain nonnegative. In the second condition $-\boldsymbol{z}_{\mathcal{B}}^T A_{\mathcal{B}}^{-1} A_{\mathcal{N}}$ is the change in the dual nonbasic variables when the dual basic variables are set to zero whilst satisfying $A^T \boldsymbol{y} + \boldsymbol{z} = \boldsymbol{c}$. When this change is bounded from below by $-\boldsymbol{z}_{\mathcal{N}}$ then the nonbasic variables remain nonnegative.

On the other hand a basis has maximum volume if $D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} D_{\mathcal{N}} (\boldsymbol{e}_j)_{\mathcal{N}} \geq -\boldsymbol{e}_{\mathcal{B}}$ for all $j \in \mathcal{N}$ and $(\boldsymbol{e}_i)_{\mathcal{B}}^T D_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} D_{\mathcal{N}} \leq \boldsymbol{e}_{\mathcal{N}}^T$ for all $i \in \mathcal{B}$; i.e. each entry in the scaled tableau matrix is bounded from below by -1 and bounded from above by 1. Rewriting these conditions as

$$\begin{aligned}
X_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_j x_j &\geq -\boldsymbol{e}_{\mathcal{B}} \quad \forall j \in \mathcal{N}, \\
-z_i (\boldsymbol{e}_i)_{\mathcal{B}}^T A_{\mathcal{B}}^{-1} A_{\mathcal{N}} Z_{\mathcal{N}}^{-1} &\geq -\boldsymbol{e}_{\mathcal{N}}^T \quad \forall i \in \mathcal{B},
\end{aligned}$$

it is seen that a maximum volume basis allows any primal nonbasic variable and any dual basic variable to be set to zero individually without destroying nonnegativity of the other variables.

It will be shown by example that a maximum volume basis need not be optimal and that an optimal basis need not have maximum volume. Consider the linear program

$$\min_{\boldsymbol{x}} 0 \qquad \text{s.t.} \begin{bmatrix} 1 & 0 & 1 & & & \\ 0 & 1 & & 1 & & \\ 1 & 1 & & & 1 & \\ -1 & -1 & & & & 1 \end{bmatrix} \boldsymbol{x} = \begin{pmatrix} \delta \\ \delta \\ 3/2 \\ -1/2 \end{pmatrix}, \quad \boldsymbol{x} \geq 0$$

for $\delta = \frac{4509}{3275} \approx 1.4$ whose feasible region is illustrated in Figure 1 by treating x_3, x_4, x_5 and x_6 as slack variables. Because the objective function is the zero vector, the primal components of the central path are constant. From the optimality conditions it can be computed analytically that $x_1 = x_2 = 0.54$, $x_3 = x_4 = \delta - 0.54$, $x_5 = 0.42$ and $x_6 = 0.58$. For any basis $\mathcal{B}$ the scaled tableau matrix is the same for all central points.

The basis $\mathcal{B} = \{2, 3, 5, 6\}$ defining the vertex $\hat{\boldsymbol{x}}$ is optimal. Computing

$$X_{\mathcal{B}}^{-1} A_{\mathcal{B}}^{-1} A_1 x_1 = \begin{pmatrix} 0 \\ x_1/x_3 \\ x_1/x_5 \\ -x_1/x_6 \end{pmatrix}$$

and using that $x_5 < x_1$ shows that $\mathcal{B}$ does not have maximum volume. On the other hand, the basis $\mathcal{B}' = \{2, 3, 4, 6\}$ defining the vertex $\hat{\boldsymbol{x}}'$ is primal infeasible. Computing

$$X_{\mathcal{B}'}^{-1} A_{\mathcal{B}'}^{-1} A_1 x_1 = \begin{pmatrix} x_1/x_2 \\ x_1/x_3 \\ -x_1/x_4 \\ 0 \end{pmatrix}, \quad X_{\mathcal{B}'}^{-1} A_{\mathcal{B}'}^{-1} A_5 x_5 = \begin{pmatrix} x_5/x_2 \\ 0 \\ -x_5/x_4 \\ x_5/x_6 \end{pmatrix}$$

and substituting the values for $\boldsymbol{x}$ verifies that $\mathcal{B}'$ has maximum volume.

The example also proves that a maximum weight basis need not be optimal, and that an optimal basis need not have maximum weight. It is readily verified that $\mathcal{B}'$ as defined above and $\mathcal{B}'' = \{1, 3, 4, 6\}$ are the only maximum weight bases for any point on the central path. Their vertices $\hat{\boldsymbol{x}}'$ and $\hat{\boldsymbol{x}}''$ are primal infeasible.

Although the maximum volume and the maximum weight basis may not yield an optimal basis as the central path approaches a solution, they share a certain property with an optimal basis that we call *correct degeneracy.*

Definition 5.2. *Let $\mathcal{B}$ be a basis with vertex $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$. $\mathcal{B}$ has* correct degeneracy *if for a strictly complementary solution $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{z}^*)$ it satisfies*

(i) $i \in \mathcal{B} \wedge x_i^* = 0 \quad \Longrightarrow \quad \hat{x}_i = 0,$

(ii) $j \in \mathcal{N} \wedge z_j^* = 0 \quad \Longrightarrow \quad \hat{z}_j = 0.$

Lemma 5.3. *The basis $\mathcal{B}$ has correct degeneracy if it satisfies any of the following conditions:*

(i) $\mathcal{B}$ is primal and dual feasible;

(ii) $\mathcal{B}$ is primal feasible and 5.2(ii) holds;

(iii) $\mathcal{B}$ is dual feasible and 5.2(i) holds;

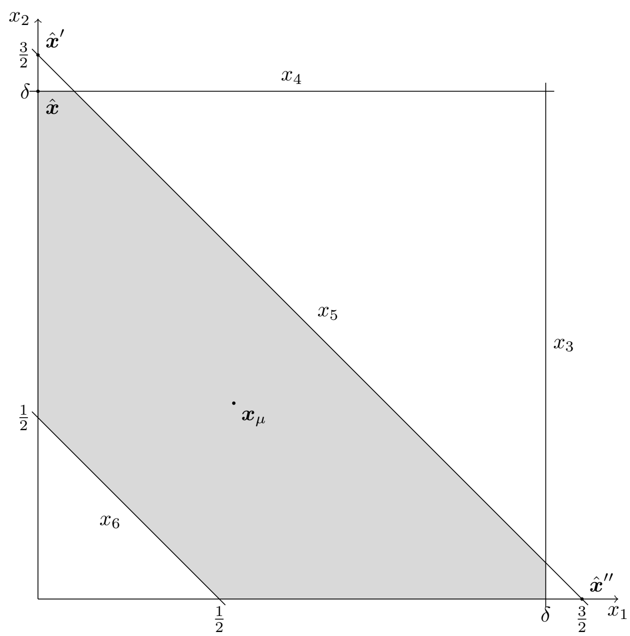

Figure 1: Feasible region of example LP. Constraints are labeled by their slack variable.

(iv) $\mathcal{B}$ is maximum volume basis for sufficiently small $\mu > 0$;

(v) $\mathcal{B}$ is maximum weight basis for sufficiently small $\mu > 0$;

Proof. (i) is immediate because if $x_i^* = 0$ or $z_j^* = 0$ for a strictly complementary solution, then this holds for any solution.

For (ii) and (iii) let $\left(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k\right) \in \mathcal{C}$ converge to a strictly complementary solution $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{z}^*)$. From Lemma 5.1(iii)

$$\sum_{i \in \mathcal{B}} \hat{x}_i (x_i^k)^{-1} + \sum_{j \in \mathcal{N}} \hat{z}_j (z_j^k)^{-1} = n.$$

When 5.2(ii) holds, the sum over $\mathcal{N}$ is bounded from below independently of k. When $\mathcal{B}$ is primal feasible, the summands in the sum over $\mathcal{B}$ are all nonnegative, so that the equality can be satisfied as $\boldsymbol{x}^k \to \boldsymbol{x}^*$ only if 5.2(i) holds. When 5.2(i) holds, the sum over $\mathcal{B}$ is bounded from below independently of k. When $\mathcal{B}$ is dual feasible, the summands in the sum over $\mathcal{N}$ are all nonnegative, so that the equality can be satisfied as $\boldsymbol{z}^k \to \boldsymbol{z}^*$ only if 5.2(ii) holds.

For (iv) and (v) consider the central path parametrized by $\mu > 0$ and from Lemma 5.1(iii)

$$\sum_{i \in \mathcal{B}} \hat{x}_i (\boldsymbol{x}_\mu)_i^{-1} + \sum_{j \in \mathcal{N}} \hat{z}_j (\boldsymbol{z}_\mu)_j^{-1} = n.$$

When $\mathcal{B}$ does not have correct degeneracy, then some of the summands become unbounded as μ tends to zero because any limit point of $\left(\boldsymbol{x}_\mu, \boldsymbol{y}_\mu, \boldsymbol{z}_\mu\right)$ is a strictly complementary solution. When a summand $i \in \mathcal{B}$ becomes unbounded, it follows from Lemma 5.1(i) that an entry in the corresponding row of the scaled tableau becomes unbounded. When a summand $j \in \mathcal{N}$ becomes unbounded, it follows from Lemma 5.1(ii) that an entry in the corresponding column of the scaled tableau becomes unbounded. In both cases $\mathcal{B}$ does not have maximum volume or maximum weight. $\square$

Algorithm 3 generic crossover

Require: basis $\mathcal{B}$ with correct degeneracy

(a) Update $\mathcal{B}$ by exchanging $i \in \mathcal{B}$ with $j \in \mathcal{N}^o$ until $\hat{\boldsymbol{x}}_\mathcal{B} \geq 0$.

(b) Update $\mathcal{B}$ by exchanging $j \in \mathcal{N}$ with $i \in \mathcal{B}^o$ until $\hat{\boldsymbol{z}}_\mathcal{N} \geq 0$.

Assuming that a basis with correct degeneracy is given, we now present an algorithm to recover an optimal basis. For a basis $\mathcal{B}$ with vertex $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$ let $\mathcal{B}^o = \{i \in \mathcal{B} : \hat{x}_i = 0\}$ and $\mathcal{N}^o = \{j \in \mathcal{N} : \hat{z}_j = 0\}$ denote the set of primal and dual degeneracies. Algorithm 3 states the crossover procedure. It is called generic because it does not specify a pivoting rule for choosing the columns to enter and leave the basis. It will be shown that with a suitable pivoting rule the algorithm terminates and that the final basis is optimal. Let $\mathcal{B}$ be an initial basis with correct degeneracy and $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{z}^*)$ be a strictly complementary solution. Then $x_j^* = 0$ for $j \notin \mathcal{B} \cup \mathcal{N}^o$. Since $A_{\mathcal{B}\cup\mathcal{N}^o}$ has rank m, there exists a primal feasible basis among $\mathcal{B} \cup \mathcal{N}^o$. Hence part (a) can terminate.

The important observation is that at the end of part (a) the basis has correct degeneracy. To see this, it is first verified that pivot operations in part (a) keep $\hat{\boldsymbol{z}}$ unchanged. Assume that $i = \mathcal{B}_p$ is exchanged by $j \in \mathcal{N}^o$ to form $\mathcal{B}'$. Denoting $\boldsymbol{v} = A_\mathcal{B}^{-1} A_j$, we have that

$$
\begin{aligned}
\frac{1}{v_p}\begin{bmatrix} -v_1 & \cdots & 1 & \cdots & -v_m \end{bmatrix} \boldsymbol{c}_{\mathcal{B}'} &= \frac{1}{v_p}(v_p c_i - \boldsymbol{v}^T \boldsymbol{c}_\mathcal{B} + c_j) \\
&= \frac{1}{v_p}(v_p c_i - \hat{\boldsymbol{y}}^T A_j + c_j) \\
&= \frac{1}{v_p}(v_p c_i + \hat{z}_j) = c_i.
\end{aligned}
$$

Using the well known formula for updating $A_\mathcal{B}^{-1}$ it follows that

$$
A_{\mathcal{B}'}^{-T} \boldsymbol{c}_{\mathcal{B}'} = A_\mathcal{B}^{-T} \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ -\frac{v_1}{v_p} & \cdots & \frac{1}{v_p} & \cdots & -\frac{v_m}{v_p} \\ & & & \ddots & \\ & & & & 1 \end{bmatrix} \boldsymbol{c}_{\mathcal{B}'} = A_\mathcal{B}^{-T} \boldsymbol{c}_\mathcal{B}
$$

and therefore

$$
\hat{\boldsymbol{z}}' = \boldsymbol{c} - A^T \hat{\boldsymbol{y}}' = \boldsymbol{c} - A^T A_{\mathcal{B}'}^{-T} \boldsymbol{c}_{\mathcal{B}'} = \boldsymbol{c} - A^T A_\mathcal{B}^{-T} \boldsymbol{c}_\mathcal{B} = \hat{\boldsymbol{z}}.
$$

Because $\hat{\boldsymbol{z}}$ is unchanged in part (a) and the initial basis had correct degeneracy, it follows from Lemma 5.3(ii) that the primal feasible basis at the end of part (a) has correct degeneracy.

Correct degeneracy at the beginning of part (b) implies that

$$
i \notin \mathcal{B}^o \cup \mathcal{N} \implies \hat{x}_i \neq 0 \implies x_i^* \neq 0 \implies z_i^* = 0,
$$

so that there exists a dual feasible basis $\mathcal{B}'$ with $\mathcal{N}' \subseteq \mathcal{B}^o \cup \mathcal{N}$. Hence part (b) can terminate. It can be verified analogously to above that pivot operations with $i \in \mathcal{B}^o$ keep $\hat{\boldsymbol{x}}$ unchanged. Therefore the final basis is also primal feasible and hence optimal.

Some comments are appropriate on the crossover method. The algorithm is solely based on a basis with correct degeneracy and does not require the interior solution itself. Part (a) is the phase I of the primal simplex method restricted to the columns $\mathcal{B} \cup \mathcal{N}^o$, and part (b) is the phase I of the dual simplex method restricted to the columns $\mathcal{B}^o \cup \mathcal{N}$. Parts

(a) and (b) can be swapped without affecting correctness of the algorithm. It follows from the correctness proof that a basis with correct degeneracy which is dual nondegenerate (i. e. $\mathcal{N}^o = \emptyset$) is primal feasible. Likewise, a basis with correct degeneracy which is primal nondegenerate (i. e. $\mathcal{B}^o = \emptyset$) is dual feasible.

6 Concluding Remarks

The obvious drawback of the maximum weight basis is to ignore the magnitude of entries in $A_{\mathcal{B}}^{-1} A_{\mathcal{N}}$. Since there is no reason to assume that these entries will always be small in magnitude, we do not see that the maximum weight basis preconditioner could yield a robust, general purpose LP solver. Indeed, [5] remarks that this preconditioner fails to achieve convergence on several problems. The maximum volume concept seems promising to us because it allows to bound directly the magnitude of entries in the scaled tableau matrix. Although the bound $1+\rho^2 m(n-m)$ is pessimistic for large scale problems, some preliminary numerical tests have shown that the preconditioner is effective in practice. Implementing the maximum volume concept will depend on finding (heuristic) methods to identify large entries in the scaled tableau matrix without computing it entirely.

The crossover algorithm that we presented is fundamentally different to existing algorithms in that it does not directly use the (approximate) interior solution. Whether the algorithm can be implemented in practice still remains an open question because the iterates from an interior point solver do not lie on the central path, they merely stay in its neighbourhood. Hence there is no guarantee that a maximum volume basis eventually attains correct degeneracy. In any case, a basis preconditioner provides an attractive starting basis for a crossover method.

References

[1] M. Arioli and I. S. Duff. Preconditioning linear least-squares problems by identifying a basis matrix. *SIAM J. Sci. Comput.*, 37(5):S544–S561, 2015.

[2] R. E. Bixby and M. J. Saltzman. Recovering an optimal LP basis from an interior point solution. *Oper. Res. Lett.*, 15(4):169–178, 1994.

[3] S. A. Goreinov, I. V. Oseledets, D. V. Savostyanov, E. E. Tyrtyshnikov, and N. L. Zamarashkin. How to find a good submatrix. In *Matrix methods: theory, algorithms and applications*, pages 247–256. World Sci. Publ., Hackensack, NJ, 2010.

[4] R. D. C. Monteiro, J. W. O'Neal, and T. Tsuchiya. Uniform boundedness of a preconditioned normal matrix used in interior-point methods. *SIAM J. Optim.*, 15(1):96–100, 2004.

[5] A. R. L. Oliveira and D. C. Sorensen. A new class of preconditioners for large-scale linear systems from interior point methods for linear programming. *Linear Algebra Appl.*, 394:1–24, 2005.

[6] C.-T. Pan. On the existence and computation of rank-revealing *LU* factorizations. *Linear Algebra Appl.*, 316(1-3):199–222, 2000.

[7] G. Peters and J. H. Wilkinson. The least squares problem and pseudo-inverses. *The Computer Journal*, 13(3):309, 1970.

[8] S. J. Wright. *Primal-dual interior-point methods*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.