Working Paper

Parallel Solution of Linear Programs Via Nash Equilibria

Markku Kallio
Andrzej Ruszczyński

WP-94-15
March 1994

International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 715210 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

Parallel Solution of Linear Programs Via Nash Equilibria

Markku Kallio
Andrzej Ruszczyński

WP-94-15
March 1994

International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 715210 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

Parallel solution of linear programs via Nash equilibria

Markku Kallio
Andrzej Ruszczyński

International Institute for Applied Systems Analysis
A-2361 Laxenburg, Austria

March 1994

Abstract

The linear programming problem is shown to be equivalent to a game in which primal players minimize the augmented Lagrangian function for the primal problem and dual players maximize the augmented Lagrangian function for the dual problem. Based on that, a parallel solution method is developed in which processors carry out under-relaxed Jacobi steps for the players. Strong convergence of the method is proved and the ratio of linear convergence estimated. Computational results are highly encouraging.

Keywords: Linear Programming, Augmented Lagrangians, Parallel Computing, Nash Equilibrium.

1. Introduction

The main objective of this paper is to develop a parallel procedure for solving the standard linear programming problem (see [5])

$$\begin{array}{l} \min c^T x \\ Ax = b, \\ x \geq 0, \end{array} \tag{1.1}$$

where $x \in R^n$ is the vector of decision variables, $c \in R^n$, $b \in R^m$ and A is an $m \times n$ matrix. As usual, we define the Lagrangian

$$L(x, \pi) = c^T x + \pi^T (b - Ax)$$

and the dual problem

$$\begin{array}{l} \max b^T \pi \\ A^T \pi \leq c. \end{array} \tag{1.2}$$

It is well-known that if (1.1) has a solution then (1.2) has a solution too, and that any pair $(\hat{x}, \hat{\pi})$ of such solutions is a saddle point of the Lagrangian:

$$\max_{\pi \in R^m} L(\hat{x}, \pi) = L(\hat{x}, \hat{\pi}) = \min_{x \geq 0} L(x, \hat{\pi}). \tag{1.3}$$

There were many attempts to solve linear programs by a saddle point seeking procedure for $L(x, \pi)$ (see, e.g. [1]). The simplest algorithm of this class may have the form

$$\begin{aligned} x^{k+1} &= \left[x^k - \tau_k \nabla_x L(x^k, \pi^k)\right]_+, \\ \pi^{k+1} &= \pi^k + \tau_k \nabla_\pi L(x^k, \pi^k), \quad k = 1, 2, \ldots. \end{aligned}$$

They are generally considered as inefficient, mainly because of the need to use very small stepsizes for primal and dual updates: $\tau_k \to 0$, $\sum_{k=0}^{\infty} \tau_k = \infty$.

A substantial improvement can be made by replacing $L(x, \pi)$ with the augmented Lagrangian:

$$\Lambda_P(x, \pi) = c^T x + \pi^T (b - Ax) + \frac{1}{2} \rho \|b - Ax\|^2,$$

where ρ is a positive penalty parameter. One of the main advantages of the augmented Lagrangian is the possibility of solving (1.1) and (1.2) by the following method of multipliers:

$$\pi^{k+1} = \pi^k + \rho(b - Ax^k), \; k = 1, 2, \ldots, \tag{1.4}$$

where

$$\Lambda(x^k, \pi^k) = \min_{x \geq 0} \Lambda(x, \pi^k). \tag{1.5}$$

It is well known (see, e.g., [11, 2]) that the method of multipliers stops after finitely many iterations at an optimal solution of (1.2). Any solution of (1.5) is then an optimal solution of (1.1) (see also [3, 6]).

In [8] a Jacobi-type method, based on our earlier works [7] and [13], was developed for the problem of minimizing the augmented Lagrangian in (1.5). The fundamental

advantage of the approach of [8] is that stepsizes in the iterative procedure need not converge to zero; in fact, they can be quite large and their safe values can be estimated by analyzing the sparsity pattern of the matrix A. Promising numerical results reported in [8] indicate that the new approach has the potential to solve very large problems. However, it is still a two-level method with the upper level responsible for updating the multipliers and the lower level solving subproblems (1.5) via Jacobi iterations.

In the next two sections we shall develop a new one-level method based on a more general idea: primal and dual steps which will be carried out in parallel for two different augmented Lagrangian functions defined for the primal and the dual problems. The original problem will be replaced by a game and a parallel method will be introduced for finding an equilibrium of this game (see [3] for an overview of parallel methods for variational inequalities).

2. The game

Let us define a non-cooperative game with two players: **P** and **D**. The objective of **P** is to minimize in the variables $x \in R^n$, $x \geq 0$, the augmented Lagrangian function for the primal problem (1.1):

$$\min_{x \geq 0} \left[\Lambda_P(x, \pi) = c^T x + \pi^T (b - Ax) + \frac{1}{2}\rho \|b - Ax\|^2 \right]. \tag{2.1}$$

The objective of **D** is to maximize in the variables $\pi \in R^m$ and $v \in R^n$, $v \geq 0$, the augmented Lagrangian function for the dual problem (1.2):

$$\max_{\pi,\ v \geq 0} \left[\Lambda_D(x, \pi, v) = b^T \pi + x^T (c - A^T \pi - v) - \frac{1}{2}\rho \|c - A^T \pi - v\|^2 \right]. \tag{2.2}$$

Let us note that we introduced slack variables $v \geq 0$ into the constraints of the dual problem to convert them into equations.

The following result is fundamental for our approach.

Theorem 1. *A point $\hat{x}$ is an optimal solution of (1.1) and a point $\hat{\pi}$ is an optimal solution of (1.2) if and only if $\hat{x}$, $\hat{\pi}$ and $\hat{v} = c - A^T \hat{\pi}$ constitute the Nash equilibrium of the game (2.1)-(2.2), i.e.,*

$$\Lambda_P(\hat{x}, \hat{\pi}) = \min_{x \geq 0} \Lambda_P(x, \hat{\pi}), \tag{2.3}$$

$$\Lambda_D(\hat{x}, \hat{\pi}, \hat{v}) = \max_{\pi,\ v \geq 0} \Lambda_D(\hat{x}, \pi, v). \tag{2.4}$$

Proof. Assume that $\hat{x}$ is a solution of (1.1) and $\hat{\pi}$ solves (1.2). Then $b - A\hat{x} = 0$, $c - A^T \hat{\pi} - \hat{v} = 0$ and $\hat{v}^T \hat{x} = 0$. This yields (2.3):

$$\Lambda_P(x, \hat{\pi}) = \hat{\pi}^T b + \hat{v}^T x + \frac{1}{2}\rho \|b - Ax\|^2 \geq \hat{\pi}^T b = c^T \hat{x} = \Lambda_P(\hat{x}, \hat{\pi})$$

and (2.4):

$$\begin{aligned}\Lambda_D(\hat{x},\pi,v) &= c^T\hat{x} + \pi^T(b - A\hat{x}) - v^T\hat{x} - \frac{1}{2}\rho\|c - A^T\pi - v\|^2 \\ &\leq c^T\hat{x} = \hat{\pi}^T b = \Lambda_D(\hat{x},\hat{\pi},\hat{v}).\end{aligned}$$

Let us now assume (2.3) and (2.4) and prove optimality of $\hat{x}$ and $\hat{\pi}$ for (1.1) and (1.2). Define

$$\hat{y} = b - A\hat{x}$$

$$\hat{z} = c - A^T\hat{\pi} - \hat{v}.$$

It is necessary and sufficient for (2.3) that

$$\nabla_x \Lambda_P(\hat{x},\hat{\pi}) = \hat{z} + \hat{v} - \rho A^T\hat{y} \geq 0, \tag{2.5}$$

$$\hat{x}^T\nabla_x\Lambda_P(\hat{x},\hat{\pi}) = \hat{x}^T\left(\hat{z} + \hat{v} - \rho A^T\hat{y}\right) = 0. \tag{2.6}$$

From (2.4) we obtain:

$$\nabla_\pi \Lambda_D(\hat{x},\hat{\pi},\hat{v}) = \hat{y} + \rho A\hat{z} = 0, \tag{2.7}$$

$$\nabla_v \Lambda_D(\hat{x},\hat{\pi},\hat{v}) = -\hat{x} + \rho\hat{z} \leq 0, \tag{2.8}$$

$$\hat{v}^T\nabla_v\Lambda_D(\hat{x},\hat{\pi},\hat{v}) = \hat{v}^T(-\hat{x} + \rho\hat{z}) = 0. \tag{2.9}$$

By (2.8), (2.5) and (2.6),

$$0 \leq (\hat{x} - \rho\hat{z})^T(\hat{z} + \hat{v} - \rho A^T\hat{y}) = -\rho\hat{z}^T(\hat{z} + \hat{v} - \rho A^T\hat{y}).$$

Using (2.7) and (2.9) in the last inequality we get

$$0 \leq -\hat{v}^T\hat{x} - \rho\|\hat{y}\|^2 - \rho\|\hat{z}\|^2.$$

Thus $\hat{v}^T\hat{x} = |\hat{y}\| = \|\hat{z}\| = 0$. The proof is complete.

Based on this result, we shall now decompose the game (2.1)-(2.2) into a game with $2n + m$ elementary players: primal players $\mathbf{P}_j$, $j = 1,\ldots,n$, and dual players $\mathbf{D}_i$, $i = 1,\ldots,m+n$. Their objectives are partial optimizations in (2.1) and (2.2), respectively. The objective of the player $\mathbf{P}_j$, $j = 1,\ldots,n$, is to solve the problem

$$\min_{x_j \geq 0} \Lambda_P(x,\pi). \tag{2.10}$$

Players $\mathbf{D}_i$, $i = 1,\ldots,m$, solve the problems

$$\max_{\pi_i} \Lambda_D(x,\pi,v). \tag{2.11}$$

Finally, players $\mathbf{D}_{m+j}$, $j = 1,\ldots,n$, aim at solving

$$\max_{v_j \geq 0} \Lambda_D(x,\pi,v); \tag{2.12}$$

The following theorem results immediately from Theorem 1.

Theorem 2. *A point $\hat{x}$ is an optimal solution of (1.1) and a point $\hat{\pi}$ is an optimal solution of (1.2) if and only if $\hat{x}$, $\hat{\pi}$ and $\hat{v} = c - A^T\hat{\pi}$ constitute the Nash equilibrium of the game (2.10)-(2.12).*

Proof. Nash equilibria of the game (2.1)-(2.2) and the game (2.10)-(2.12) are identical, because the functions Λ_P and Λ_D are continuously differentiable and the feasible sets in (2.1) and (2.2) are Cartesian products of feasible sets for single coordinates. The result follows then from Theorem 1.

In the next section we shall use this result to develop a parallel method for solving the problems (1.1)-(1.2) in which individual tasks (processors) will play the role of agents $\mathbf{P}_j$ and $\mathbf{D}_i$ in our game.

3. The method

Let us now describe in detail a decomposition method for solving the problems (1.1) and (1.2). It is, in fact, an under-relaxed Jacobi algorithm for solving the game (2.10)-(2.12). In the description below we use A_j to denote the j-th column of A, $j = 1, \ldots, n$, and A^i to denote the i-th row of A, $i = 1, \ldots, m$.

Initialization. Choose a starting point $x^0 \geq 0$, $\pi^0 \in R^m$ and $v^0 \geq 0$ and a parameter $\gamma \in (0, 2)$. Set $k = 0$.

Stopping Test. Determine the primal infeasibility

$$y^k = b - Ax^k, \tag{3.1}$$

the dual infeasibility

$$z^k = c - A^T\pi^k - v^k \tag{3.2}$$

and the error function

$$E_k = \langle x^k, v^k \rangle + \rho\|y^k\|^2 + \rho\|z^k\|^2. \tag{3.3}$$

If $E_k = 0$ then stop.

Prediction. Make the multiplier method updates

$$\tilde{\pi}^k = \pi^k + \rho y^k,$$

$$\tilde{x}^k = x^k - \rho z^k.$$

Stepsize Calculation. Define the sets

$$\begin{aligned} \mathcal{N}_k &= \left\{ j : v_j^k > 0 \text{ or } \tilde{x}_j^k < 0 \right\}, \\ \mathcal{B}_k &= \left\{ j : x_j^k > 0 \text{ or } c_j - A_j^T\tilde{\pi}^k < 0 \right\}. \end{aligned}$$

Calculate

$$D_k = \sum_{j \in \mathcal{B}_k} \frac{1}{\|A_j\|^2} \left(c_j - A_j^T \tilde{\pi}^k\right)^2 + \sum_{i=1}^{m} \frac{1}{\|A^i\|^2} \left(b_i - A^i \tilde{x}^k\right)^2 + \sum_{j \in \mathcal{N}_k} \left(\tilde{x}_j^k\right)^2$$

and the stepsize

$$\tau_k = \frac{\gamma \rho E_k}{D_k}. \tag{3.4}$$

Primal Optimization. For $j = 1, \ldots, n$ find the unconstrained solution of (2.10), make a step of length τ_k towards it and project onto the feasible set:

$$x_j^{k+1} = \left[x_j^k - \frac{\tau_k}{\rho \|A_j\|^2} \left(c_j - A_j^T \tilde{\pi}^k\right)\right]_+ .$$

Dual Optimization. For $i = 1, \ldots, m$ find the solution of (2.11) and make a step of length τ_k towards it:

$$\pi_i^{k+1} = \pi_i^k + \frac{\tau_k}{\rho \|A^i\|^2} \left(b_i - A^i \tilde{x}^k\right).$$

For $j = 1, \ldots, n$ find the unconstrained solution of (2.12) and make a step of length τ_k towards it:

$$v_j^{k+1} = \left[v_j^k - \frac{\tau_k}{\rho} \tilde{x}_j^k\right]_+ .$$

Increase k by one and go to the Stopping Test.

It is a matter of elementary calculations to show that the formulae used in the Primal Optimization step and in the Dual Optimization Step correspond to problems (2.10), (2.11) and (2.12).

An important property of the method described here is that it is highly parallelizable. Each step of the method contains a number of column or row operations that can be carried out simultaneously. It is also worth mentioning that the optimization steps are made in directions of the gradients of the usual Lagrangian $L(x, \pi)$, but calculated at the predicted point $(\tilde{x}, \tilde{\pi})$, analogously to the idea of the extragradient method of [9].

4. Convergence

The kth iteration of the algorithm introduced in the previous section can be compactly described as

$$x_j^{k+1} = \left[x_j^k - \frac{\tau_k \theta_{x_j}^k}{\rho \|A_j\|^2} \nabla_{x_j} \Lambda_P(x^k, \pi^k)\right]_+ , \quad j = 1, \ldots, n, \tag{4.1}$$

$$\pi_i^{k+1} = \pi_i^k + \frac{\tau_k}{\rho \|A^i\|^2} \nabla_{\pi_i} \Lambda_D(x^k, \pi^k, v^k), \quad i = 1, \ldots, m, \tag{4.2}$$

$$v_j^{k+1} = \left[v_j^k + \frac{\tau_k \theta_{v_j}^k}{\rho} \nabla_{v_j} \Lambda_D(x^k, \pi^k, v^k)\right]_+ , \quad j = 1, \ldots, n, \tag{4.3}$$

where

$$\theta^k_{x_j} = \begin{cases} 1 & \text{if } j \in \mathcal{B}_k \\ 0 & \text{otherwise,} \end{cases}$$

$$\theta^k_{v_j} = \begin{cases} 1 & \text{if } j \in \mathcal{N}_k \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 3. *Assume that (1.1) has a solution. Then the method generates a sequence* $\left\{(x^k, \pi^k, v^k)\right\}_{k=0}^{\infty}$ *convergent to a point* $(\hat{x}, \hat{\pi}, \hat{v})$ *such that* $\hat{x}$ *is an optimal solution of (1.1),* $\hat{\pi}$ *is an optimal solution of (1.2) and* $\hat{v} = c - A^T\hat{\pi}$.

Proof. Let (x^*, π^*, v^*) be an equilibrium of the game. Let S_x and S_π be diagonal matrices with diagonal elements $\|A_j\|^2$, $j = 1, \ldots, n$, and $\|A^i\|^2$, $i = 1, \ldots, m$, respectively, and let us denote $\|a\|_S^2 = \langle a, Sa\rangle$. We define

$$W_k = \|x^k - x^*\|^2_{S_x} + \|\pi^k - \pi^*\|^2_{S_\pi} + \|v^k - v^*\|^2. \tag{4.4}$$

We also define Θ^k_x and Θ^k_v as diagonal matrices with the diagonal entries $\theta^k_{x_j}$ and $\theta^k_{v_j}$, respectively.

Since the scaling matrices are diagonal, the projection on the set $\{x : x \geq 0\}$ is nonexpansive in the norm $\|\cdot\|_{S_x}$. Thus

$$\begin{aligned} \|x^{k+1} - x^*\|^2_{S_x} &\leq \|x^k - \frac{\tau_k}{\rho} S_x^{-1}\Theta^k_x \nabla_x \Lambda_P(x^k, \pi^k) - x^*\|^2_{S_x} \\ &= \|x^k - x^*\|^2_{S_x} - \frac{2\tau_k}{\rho}\langle \Theta^k_x \nabla_x \Lambda_P(x^k, \pi^k), x^k - x^*\rangle \\ &\quad + \frac{\tau_k^2}{\rho^2}\|\Theta^k_x \nabla_x \Lambda_P(x^k, \pi^k)\|^2_{S_x^{-1}}. \end{aligned} \tag{4.5}$$

In a similar way,

$$\begin{aligned} \|\pi^{k+1} - \pi^*\|^2_{S_\pi} &= \|\pi^k - \pi^*\|^2_{S_\pi} + \frac{2\tau_k}{\rho}\langle \nabla_\pi \Lambda_D(x^k, \pi^k, v^k), \pi^k - \pi^*\rangle \\ &\quad + \frac{\tau_k^2}{\rho^2}\|\nabla_\pi \Lambda_D(x^k, \pi^k, v^k)\|^2_{S_\pi^{-1}}, \end{aligned} \tag{4.6}$$

$$\begin{aligned} \|v^{k+1} - v^*\|^2 &\leq \|v^k - v^*\|^2 + \frac{2\tau_k}{\rho}\langle \Theta^k_v \nabla_v \Lambda_D(x^k, \pi^k, v^k), v^k - v^*\rangle \\ &\quad + \frac{\tau_k^2}{\rho^2}\|\Theta^k_v \nabla_v \Lambda_D(x^k, \pi^k, v^k)\|^2. \end{aligned} \tag{4.7}$$

Let us estimate the terms that stand at τ_k in (4.5)-(4.7). Since $\langle \nabla_{x_j} \Lambda_P(x^k, \pi^k), x^k_j - x^*_j\rangle \leq 0$ for $j \notin \mathcal{B}_k$,

$$\langle \Theta^k_x \nabla_x \Lambda_P(x^k, \pi^k), x^k - x^*\rangle \geq \langle \nabla_x \Lambda_P(x^k, \pi^k), x^k - x^*\rangle.$$

In a similar way, we obtain

$$\langle \Theta^k_v \nabla_v \Lambda_D(x^k, \pi^k, v^k), v^k - v^*\rangle \leq \langle \nabla_v \Lambda_D(x^k, \pi^k, v^k), v^k - v^*\rangle.$$

Using the last two inequalities, after lengthy but straightforward transformations, we can estimate the expression at $-2\tau_k/\rho$ in (4.5)-(4.7) as follows:

$$\langle \Theta_x^k \nabla_x \Lambda_P(x^k, \pi^k), x^k - x^* \rangle - \langle \nabla_\pi \Lambda_D(x^k, \pi^k, v^k), \pi^k - \pi^* \rangle$$
$$- \langle \Theta_v^k \nabla_v \Lambda_D(x^k, \pi^k, v^k), v^k - v^* \rangle \;\geq\; \langle x^k, v^k \rangle + \rho \|y^k\|^2 + \rho \|z^k\|^2 \;=\; E_k.$$

The terms at τ_k^2/ρ^2 sum up to

$$\|\Theta_x^k \nabla_x \Lambda_P(x^k, \pi^k)\|^2_{S_x^{-1}} + \|\nabla_\pi \Lambda_D(x^k, \pi^k, v^k)\|^2_{S_\pi^{-1}} + \|\Theta_v^k \nabla_v \Lambda_D(x^k, \pi^k, v^k)\|^2 = D_k.$$

Combining (4.5)-(4.7) with the last two equations we obtain

$$W_{k+1} \leq W_k - \frac{2\tau_k}{\rho} E_k + \frac{\tau_k^2}{\rho^2} D_k = W_k - \frac{\tau_k}{\rho^2}(2\rho E_k - \tau_k D_k).$$

By the stepsize rule (3.4),

$$W_{k+1} \leq W_k - \frac{(2-\gamma)\gamma E_k^2}{D_k}. \tag{4.8}$$

Thus, the sequence $\{W_k\}$ is non-increasing and

$$\lim_{k \to \infty} \frac{E_k^2}{D_k} = 0. \tag{4.9}$$

Since W_k is bounded, the sequence $\{(x^k, \pi^k, v^k)\}$ is bounded and has an accumulation point $(\hat{x}, \hat{\pi}, \hat{v})$. By (4.9) and the boundedness of $\{D_k\}$, $\lim_{k \to \infty} E_k = 0$. Thus,

$$\langle \hat{x}, \hat{v} \rangle + \rho \|b - A\hat{x}\|^2 + \rho \|c - A^T\hat{\pi} - \hat{v}\|^2 = 0,$$

i.e., $\hat{x}$ is an optimal solution of (1.1), $\hat{\pi}$ is an optimal solution of (1.2) and $\hat{v} = c - A^T\hat{\pi}$. By Theorem 2 we can use $(\hat{x}, \hat{\pi}, \hat{v})$ instead of (x^*, π^*, v^*) in (4.4). Then from (4.8) we see that the distance to $(\hat{x}, \hat{\pi}, \hat{v})$ is non-increasing. Consequently, $(\hat{x}, \hat{\pi}, \hat{v})$ is the only accumulation point of the sequence $\{(x^k, \pi^k, v^k)\}$. The proof is complete.

5. Rate of Convergence

We shall now analyze asymptotic properties of our method under the following assumption.

Strict Complementarity Condition. Problems (1.1) and (1.2) have unique solutions $\hat{x}$ and $\hat{\pi}$ such that for $\hat{v} = c - A^T\hat{\pi}$ one has $\hat{v}_j > 0$ if $\hat{x}_j = 0$ and $\hat{x}_j > 0$ if $\hat{v}_j = 0$.

As usual, we introduce the set of non-basic variables $\mathcal{N} = \{j : \hat{x}_j = 0\}$, the set of basic variables $\mathcal{B} = \{j : \hat{x}_j > 0\}$ and the basis matrix $B = (A_j)_{j \in \mathcal{B}}$. By the Strict Complementarity Condition the basis B is square and non-singular.

At first we show that the basis is identified in a finite number of iterations.

Lemma 1. *Assume that the Strict Complementarity Condition holds. Then, for any* $\gamma \in (1,2)$, *there is* k_0 *such that for all* $k \geq k_0$ *we have* $x_j^k = 0$ *for all* $j \in \mathcal{N}$ *and* $v_j^k = 0$ *for all* $j \in \mathcal{B}$.

Proof. Let us define the sets

$$\mathcal{M} = \{j \in \mathcal{N} : \ x_j^k > 0 \text{ for infinitely many } k\},$$

$$\mathcal{P} = \{j \in \mathcal{B} : \ v_j^k > 0 \text{ for infinitely many } k\}.$$

We shall prove that they are empty. Suppose the opposite, namely, that at least one of them is non-empty. By Theorem 1 and strict complementarity,

$$\lim_{k\to\infty} \nabla_{x_j} \Lambda_P(x^k, \pi^k) = \hat{v}_j > 0 \quad \text{for} \quad j \in \mathcal{N} \tag{5.1}$$

and

$$\lim_{k\to\infty} \nabla_{v_j} \Lambda_D(x^k, \pi^k, v^k) = -\hat{x}_j < 0 \quad \text{for} \quad j \in \mathcal{B}. \tag{5.2}$$

By (5.1) and (4.1), for all sufficiently large k,

$$x_j^k > x_j^{k+1} > \cdots \geq 0, \ j \in \mathcal{M}.$$

Similarly, (5.2) and (4.3) imply that for all sufficiently large k

$$v_j^k > v_j^{k+1} > \cdots \geq 0, \ j \in \mathcal{P}.$$

Let us denote $x_{\mathcal{M}} = (x_j)_{j\in\mathcal{M}}$ and, in a similar way, $x_{\mathcal{P}}$, $v_{\mathcal{M}}$, $v_{\mathcal{P}}$. From (4.1) and (4.3) we obtain:

$$x_{\mathcal{M}}^{k+1} = x_{\mathcal{M}}^k - \frac{\tau_k}{\rho} S_{x_{\mathcal{M}}}^{-1} \nabla_{x_{\mathcal{M}}} \Lambda_P(x^k, \pi^k), \tag{5.3}$$

$$v_{\mathcal{P}}^{k+1} = v_{\mathcal{P}}^k + \frac{\tau_k}{\rho} \nabla_{v_{\mathcal{P}}} \Lambda_D(x^k, \pi^k, v^k). \tag{5.4}$$

Let us multiply both sides of (5.3) by $v_{\mathcal{M}}^k$, both sides of (5.4) by $x_{\mathcal{P}}^k$ and add the results. Since the left sides are non-negative, we obtain

$$\begin{aligned} 0 &\leq \langle x_{\mathcal{M}}^k, v_{\mathcal{M}}^k \rangle + \langle x_{\mathcal{P}}^k, v_{\mathcal{P}}^k \rangle - \frac{\tau_k}{\rho} \left(\langle v_{\mathcal{M}}^k, S_{x_{\mathcal{M}}}^{-1} \nabla_{x_{\mathcal{M}}} \Lambda_P(x^k, \pi^k) \rangle - \langle x_{\mathcal{P}}^k, \nabla_{v_{\mathcal{P}}} \Lambda_D(x^k, \pi^k, v^k) \rangle \right) \\ &\leq E_k \left[1 - \frac{\gamma}{D_k} \left(\langle v_{\mathcal{M}}^k, S_{x_{\mathcal{M}}}^{-1} \nabla_{x_{\mathcal{M}}} \Lambda_P(x^k, \pi^k) \rangle - \langle x_{\mathcal{P}}^k, \nabla_{v_{\mathcal{P}}} \Lambda_D(x^k, \pi^k, v^k) \rangle \right) \right], \end{aligned} \tag{5.5}$$

where in the last inequality we used (3.4) and the estimate $\langle x^k, v^k \rangle \leq E_k$. By Theorem 1 and (5.1)-(5.2),

$$\begin{aligned} \lim_{k\to\infty} \left[\langle v_{\mathcal{M}}^k, S_{x_{\mathcal{M}}}^{-1} \nabla_{x_{\mathcal{M}}} \Lambda_P(x^k, \pi^k) \rangle - \langle x_{\mathcal{P}}^k, \nabla_{v_{\mathcal{P}}} \Lambda_D(x^k, \pi^k, v^k) \rangle \right] &= \langle \hat{v}_{\mathcal{M}}, S_{x_{\mathcal{M}}}^{-1} \hat{v}_{\mathcal{M}} \rangle + \|\hat{x}_{\mathcal{P}}\|^2 \\ &= \lim_{k\to\infty} D_k > 0, \end{aligned}$$

because in calculating $\lim_{k\to\infty} D_k$ directions with nonzero limits occur only for $x^k_{\mathcal{M}}$ and $v^k_{\mathcal{P}}$. Thus, with $1 < \gamma < 2$, the right side of (5.5) becomes negative for large k: a contradiction. The proof is complete.

Remark. Simple counter examples can be constructed to show that the assumption $\gamma > 1$ cannot be dropped.

We shall now introduce some measures of sparsity of the matrix A. Let M^i denote the number of nonzeros in the ith row of A, $i = 1, \ldots, m$. We define for each column A_j of A, $j = 1, \ldots, n$, the average row count

$$N_j = \frac{\sum_{i=1}^m a_{ij}^2 M^i}{\sum_{i=1}^m a_{ij}^2}.$$

Analogously, let M_j be the number of nonzeros in the jth column of A and let

$$N^i = \frac{\sum_{j=1}^n a_{ij}^2 M_j}{\sum_{j=1}^n a_{ij}^2}$$

denote the average column count associated with row i, $i = 1, \ldots, m$. Finally, let

$$N = \max\left(\max_{1\le j\le n} N_j, \max_{1\le i\le m} N^i\right).$$

Let us note that there is a simple upper bound on N: the maximum number of nonzeros in a row or column of A.

Next, let us define

$$\kappa_{\min} = \min\left(1, \min_{1\le j\le n} \|A_j\|^2, \min_{1\le i\le m} \|A^i\|^2\right),$$

and

$$\kappa_{\max} = \max\left(\max_{1\le j\le n} \|A_j\|^2, \max_{1\le i\le m} \|A^i\|^2\right) + 1.$$

We can estimate the limit properties of the stepsizes τ_k as follows.

Theorem 4. *Assume that the Strict Complementarity Condition holds. Then, for any $\gamma \in (1,2)$, there exists k_0 such that for all $k \ge k_0$*

$$\tau_k \ge \gamma\beta,$$

where

$$\beta = \frac{1}{(N+1)(\kappa_{\max}/\kappa_{\min})\max(1,\rho^{-2})}. \tag{5.6}$$

Proof. We shall estimate the ratio in (3.4) for a fixed (but sufficiently large) k. By Lemma 1,

$$E_k = \rho\|y^k\|^2 + \rho\|z^k\|^2,$$

if k is large enough. Let us estimate D_k (for brevity, we omit the subscripts and superscripts k). For large k, by Lemma 1, we have

$$\begin{aligned} D_k &= \|z_{\mathcal{B}} - \rho B^T y\|^2_{S_x^{-1}} + \|y + \rho A z\|^2_{S_\pi^{-1}} + \rho^2 \|z_{\mathcal{N}}\|^2 \\ &= \|z_{\mathcal{B}} - \rho B^T y\|^2_{S_x^{-1}} + \|y + \rho B z_{\mathcal{B}} + \rho A_{\mathcal{N}} z_{\mathcal{N}}\|^2_{S_\pi^{-1}} + \rho^2 \|z_{\mathcal{N}}\|^2 \\ &\leq \frac{1}{\kappa_{\min}} \left\| \begin{bmatrix} -\rho B^T & I & \\ I & \rho B & \rho A_{\mathcal{N}} \\ & & \rho I \end{bmatrix} \begin{bmatrix} y \\ z_{\mathcal{B}} \\ z_{\mathcal{N}} \end{bmatrix} \right\|^2, \end{aligned} \tag{5.7}$$

where B is the basis matrix. The right hand side of (5.7) can be estimated by Lemma A (see the Appendix):

$$D_k \leq (N+1)(\kappa_{\max}/\kappa_{\min}) \max(1, \rho^2) \left(\|y\|^2 + \|z_{\mathcal{B}}\|^2 + \|z_{\mathcal{N}}\|^2 \right),$$

because the row counts of the matrix mutiplied by $(y, z_{\mathcal{B}}, z_{\mathcal{N}})$ in (5.7) are bounded by $N+1$ and its squared column norms by $\kappa_{\max} \max(1, \rho^2)$. Consequently,

$$D_k \leq \rho(N+1)(\kappa_{\max}/\kappa_{\min}) \max(1, \rho^{-2}) E_k,$$

which proves our assertion.

We are now ready to estimate the rate of convergence of the method.

Theorem 5. *Assume that the Strict Complementarity Condition holds. Then, for any $\gamma \in (1, 2)$, the method is convergent at a linear rate, i.e. there is $q(\rho) \in (0, 1)$ such that, for all sufficiently large k, one has*

$$W_{k+1} \leq q(\rho) W_k,$$

where W_k is defined by (4.4). Moreover,

$$\limsup_{\rho \to \infty} q(\rho) < 1.$$

Proof. Let us rewrite inequality (4.8) as follows

$$W_{k+1} \leq W_k - (2-\gamma)\gamma \left(\frac{\rho E_k}{D_k} \right) \left(\frac{E_k}{\rho} \right). \tag{5.8}$$

By Theorem 4, for all sufficiently large k,

$$\frac{\rho E_k}{D_k} \geq \beta > 0. \tag{5.9}$$

Let us estimate from below E_k/ρ. By lemma 1, for all sufficiently large k,

$$\begin{aligned} E_k/\rho &= \|y^k\|^2 + \|z^k\|^2 \\ &= \|B(x^k_{\mathcal{B}} - x^*_{\mathcal{B}})\|^2 + \|A^T_{\mathcal{N}}(\pi^k - \pi^*) + (v^k_{\mathcal{N}} - v^*_{\mathcal{N}})\|^2 + \|B^T(\pi^k - \pi^*)\|^2 \end{aligned}$$

and

$$W_k = \|x_B^k - x_B^*\|_{S_x}^2 + \|\pi^k - \pi^*\|_{S_\pi}^2 + \|v_N^k - v_N^*\|^2.$$

Since the basis matrix B is non-singular, there is a number $\mu > 0$, independent of ρ, such that

$$E_k/\rho \geq \mu W_k. \tag{5.10}$$

Substituting (5.9) and (5.10) into (5.8) we obtain

$$W_{k+1} \leq (1 - (2 - \gamma)\gamma\beta\mu)\, W_k,$$

which proves the first assertion. By passing to the limit with $\rho \to \infty$ in (5.6) we obtain the second assertion.

6. Computational Illustration

An experimental computer code called *Nash* has been developed by means of revising our earlier *Jacobi* code for the multiplier method [8]. The procedure of section 3 was generalized in a straightforward manner to account explicitly for simple bounds on primal variables.

To begin the first iteration, we set all primal variables equal to zero and project onto the bounds. Similarly, we set all dual multipliers to zero initially. The iterations end when the error term (3.3) relative to the absolute value of the objective function value is smaller than a prespecified optimality tolerance ϕ, and when primal and dual infeasibilities are reasonably small (we used $\phi = 10^{-6}$). An iteration in *Nash* consists of a dual phase and a primal phase. It should be stressed that both phases can be run in parallel. Furthermore, all primal variables can be processed in parallel and the dual phase can be carried out simultaneously for all rows.

For illustration, *Nash* was tested on some problems from the Netlib library [10]. The set consists of the six largest problems used in our earlier study [8]. Table 1 shows the names and dimensions of these problems. The serial run times t_s on a HP9000/720 for the simplex code *Cplex 2.1* [4] are reported as reference times to be used for efficiency comparisons. For *Nash*, simulation runs were performed on the same computer to obtain estimates of the CPU work w. Assuming one processor for each column and row, we obtain a lower bound for the parallel run time

$$t_p = \frac{w}{n+m}. \tag{6.1}$$

This formula assumes that all processors are loaded with equal tasks, so that the execution time is the same for all. Alternatively, we might have defined the parallel run time based on the worst cases (the longest run times of primal and dual phases). However, the worst case results from an uneven distribution of nonzeros in rows and columns and the largest tasks can be further parallelized, for instance, by splitting dense columns or rows.

The simplex method does not suit parallel computation the way *Nash* does. Therefore for *Cplex*, we use the serial run time in the comparisons, and define the potential

speedup of *Nash* against *Cplex* as the serial run time of the latter, divided by the estimated parallel run time t_p of *Nash*. Obviously, there is some gain from parallel computation for the simplex methods as well. Besides, communication time is omitted for *Nash*. Therefore, our speed-ups should be regarded as rather optimistic.

Iteration counts, work (CPU seconds), estimated parallel run times t_p (seconds), potential speed-ups (t_s/t_p) against *Cplex* and the stepsize at the end are reported in Table 2.

As measures for precision, for each row i and column j, relative primal errors e_P^i and relative dual errors e_D^j are defined as follows:

$$\begin{aligned} e_P^i &= \frac{|y_i|}{1 + |b_i| + \sum_j |x_j a_{ij}|}, \\ e_D^j &= \frac{|z_j|}{1 + |c_j| + \sum_i |\pi_i a_{ij}|}, \end{aligned}$$

where y_i and z_j refer to primal and dual infeasibilities, respectively. Table 3 shows relative errors in the objective function (e_F), average relative primal and dual errors $\bar{e}_P$ and $\bar{e}_D$ and maximal relative primal and dual errors $\hat{e}_P$ and $\hat{e}_D$.

It is worth pointing out that in all cases there is a potential of speedup of orders of magnitude with respect to a sequential method and that the final accuracy is quite high. The results also show that our theoretical estimates concerning limit values of stepsizes are satisfied with a broad margin. As a whole, the new approach appears to be very promising.

Problem	Rows	Columns	Nonzeros	Time
stocfor3	16675	15695	74004	1023.2
80bau3b	2263	9799	29063	93.6
stocfor2	2158	2031	9492	14.9
degen3	1504	1818	26230	113.3
sctap3	1481	2480	10734	5.2
pilot	1442	3652	43220	432.5

Table 1: Problem statistics. Time is the reference time in seconds obtained by *Cplex*.

Problem	Iterations	Work	Parallel Time	Potential Speedup	Final Stepsize
stocfor3	4745	2955	0.091	11200	0.2
80bau3b	21064	3809	0.316	296	0.1
stocfor2	3408	274	0.066	227	0.4
degen3	9748	920	0.278	408	0.3
sctap3	2672	207	0.052	99	0.3
pilot	21022	2459	0.482	897	0.1

Table 2: Performance of the method.

Problem	e_F	$\bar{e}_P$	$\bar{e}_D$	$\hat{e}_P$	$\hat{e}_D$
stocfor3	3E-5	2E-5	5E-6	7E-4	6E-4
80bau3b	1E-5	4E-6	7E-7	7E-4	4E-4
stocfor2	2E-5	2E-5	2E-5	7E-4	6E-4
degen3	8E-6	5E-5	1E-5	6E-4	6E-4
sctap3	9E-5	3E-6	1E-6	7E-4	5E-4
pilot	3E-4	5E-5	6E-6	7E-4	2E-4

Table 3: Accuracy of the solution.

References

[1] K.J. Arrow, L. Hurwicz and H. Uzawa, *Studies in Linear and Nonlinear Programming*, (Stanford University Press, Stanford, 1958).

[2] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, (Academic Press,1982).

[3] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and Distributed Computation* (Prentice-Hall, Englewood Cliffs, 1989).

[4] R.E. Bixby, "Implementing the simplex method: the initial basis", *ORSA Journal on Computing* 4(1992) 267-284.

[5] G.B. Dantzig, *Linear Programming and Extensions* (Princeton University Press, Princeton, 1963).

[6] O. Güler, "Augmented Lagrangian algorithms for linear programming", *Journal of Optimization Theory and Applications* 75(1992) 445-470.

[7] M. Kallio and S. Salo, "Tatonnement procedures for linearly constrained convex optimization," Helsinki School of Economics, 1992 (accepted for publication in *Management Science*).

[8] M. Kallio, A. Ruszczyński and S. Salo, "A regularized Jacobi method for large-scale linear programming", working paper WP-93-61, IIASA, Laxenburg, 1993.

[9] G.M. Korpelevich, "The extragradient method for finding saddle points and other problems", *Matecon*, 12(1976) 747-756.

[10] Netlib, LP Test Problems, Bell Laboratories.

[11] B.T. Polyak and N.V. Tretyakov, "An iterative method for linear programming and its economic interpretation", *Matecon* 10(1974) 81-100.

[12] R.T. Rockafellar, "Augmented Lagrangians and applications of the proximal point algorithm in convex programming", *Mathematics of Operations Research* 1(1976) 97-116.

[13] A. Ruszczyński, "Augmented Lagrangian decomposition for sparse convex optimization", working paper WP-92-75, IIASA, Laxenburg, 1992 (accepted for publication in *Mathematics of Operations Research*).

Appendix

Lemma A (Kallio, Ruszczyński and Salo, 1993) *Let A be an $m \times n$ matrix and let*

$$N_j = \frac{\sum_{i=1}^{m} a_{ij}^2 M^i}{\sum_{i=1}^{m} a_{ij}^2}, \quad j = 1, \ldots, n,$$

where M^i is the number of nonzeros in the ith row of A. Then for every $d \in R^n$

$$\|Ad\|^2 \leq \sum_{j=1}^{n} N_j \|A_j\|^2 d_j^2.$$

Proof. Expansion into single entries yields

$$\|Ad\|^2 = \sum_{j=1}^{n} \sum_{s=1}^{n} A_j^T A_s d_j d_s = \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{s \in V(i,j)} (a_{ij} d_j)(a_{is} d_s),$$

where $V(i,j)$ is the set of such s that $a_{ij} a_{is} \neq 0$. Applying Schwarz inequality to the right side of the above equation and noting that $s \in V(i,j)$ if and only if $j \in V(i,s)$ we get

$$\|Ad\|^2 \leq \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{s \in V(i,j)} a_{ij}^2 d_j^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} M^i a_{ij}^2 d_j^2,$$

where in the last relation we used the fact that $V(i,j)$ has M^i elements. The required result follows now from the definition of the N_j's.