An Interior Point Method with Column Aggregation and Selection

Jos F. Sturm

October 20, 2004

Abstract

We propose to apply the column and constraint generation idea within an interior point method. The technique can reduce computational times when the optimum number of basic variables and active constraints is relatively small. It can benefit from warm start information.

1 Introduction

Review of low-rank update strategies in IPMs. Analytic center cutting plane method (ACCPM). Paper Flippo-Janssen. Spectral Bundle. LP-DASA. LP-Simplex with partial rows+columns. Projection methods.

An essential element in our approach is the extended self-dual embedding of Ye, Todd and Mizuno. In this model, (a multiple of) the duality gap appears as a variable. Inactive ('non-basic') columns can be merged into the column of residuals, which has a duality-gap activity. In this way, the inactive columns can be revived instantaneously within the interior point method.

Wide region following method.

A key element in the polynomiality proof of the new active set IPM is a merit function, which was introduced recently in Sturm.

Consider the following self-dual linear program:

$$\begin{array}{lll} \min & b_0 y_0 & \\ \text{s.t.} & Mx + z = y_0 r & (1) \\ & r^{\mathrm{T}} x + s_0 = b_0 & (2) \\ & x \in \Re^m \times \Re^n_+,\ z \in \{0\}^m \times \Re^n_+,\ y_0, s_0 \in \Re_+, & (3) \end{array}$$

where $M = -M^{\mathrm{T}}$ is a given $(m+n) \times (m+n)$ skew symmetric matrix, $r \in \Re^{m+n}$ is a given vector, and b_0 is a scalar constant. The decision variables are x, y_0, s_0 and z.

The optimal value of (3) is zero. Our goal is to find a strictly complementary solution to (3).

Even though the $y_0 \geq 0$ constraint will be binding at optimality, it is a redundant constraint and its dual multiplier s_0 can always be set to zero. In the original Ye-Todd-Mizuno model, y_0

therefore appears as a free variable. In our setting, it is an advantage to have the flexibility of letting the slack variable $s_0 \geq 0$ float instead of forcing it to be zero.

2 Centering Step and Aggregation

Let (B, N) be a parition of $\{1, 2, \ldots, m+n\}$ such that $\{1, 2, \ldots, m\} \subseteq B$. The (B, N)-aggregated model at iterate (x^k, y^k, z^k) is

$$\begin{array}{lll} \min & b_0 y_0 & \\ \text{s.t.} & M_{BB} x_B + x_0 a_0 + z_B = y_0 r_B & (4) \\ & -a_0^{\mathrm{T}} x_B + z_0 = y_0 r_0 & (5) \\ & r_B^{\mathrm{T}} x_B + r_0 x_0 = b_0 & (6) \\ & x_B \in \Re^m \times \Re_+^{|B|-m},\ z_B \in \{0\}^m \times \Re_+^{|B|-m},\ x_0 \geq 0,\ z_0 \geq 0, y_0 \in \Re. & (7) \end{array}$$

The vector a_0 and the scalar r_0 in the (B, N, x_N)-aggregated model are defined as

$$a_0 = M_{BN} x_N^k,\ r_0 := r_N^{\mathrm{T}} x_N^k. \tag{8}$$

Based on the current solution for (3), we can construct a feasible solution to (7) by setting

$$x_0^k = 1,\ z_0^k = (x_N^k)^{\mathrm{T}} z_N^k.$$

Conversely, given a solution $(x_B^+, z_B^+, x_0^+, z_0^+, y_0^+)$ to (7), we let

$$x_N^+ = x_0^+ x_N,\ z_N^+ = y_0^+ r_N - M_{NB} x_B^+ - x_0^+ (M_{NN} x_N).$$

This solution is feasible to (3) if and only if $z_N^+ \geq 0$.

We let

$$w_i := x_i z_i \text{ for } i = 1, 2, \ldots, m+n, \tag{9}$$

so that in fact $w \in \{0\}^m \times \Re_{++}^n$. We define a target $\hat{w} \in \{0\}^m \times \mathrm{CR}(n)$. The centering direction is implicitly defined by

$$\begin{aligned} X_B \Delta z_B + Z_B \Delta x_B &= \hat{w}_B - w_B & (10) \\ x_0 \Delta z_0 + z_0 \Delta x_0 &= e^{\mathrm{T}} (\hat{w}_N - w_N), & (11) \end{aligned}$$

together with

$$\left\{ \begin{array}{l} M_{BB} \Delta x_B + a_0 \Delta x_0 + \Delta z_B = \Delta y_0 r_B \\ -a_0^{\mathrm{T}} \Delta x_B + \Delta z_0 = \Delta y_0 r_0 \\ r_B^{\mathrm{T}} \Delta x_B + r_0 \Delta x_0 = 0 \\ \Delta z_i = 0 \text{ for } i = 1, 2, \ldots, m. \end{array} \right. \tag{12}$$

We observe the following consequence of the Cauchy-Schwarz inequality:

$$X \succ 0 \Longrightarrow X^{-1} - \frac{e e^{\mathrm{T}}}{e^{\mathrm{T}} X e} \succeq 0. \tag{13}$$

This implies that

$$(\hat{w}_B - w_B)^{\mathrm{T}} W_B^{-1} (\hat{w}_B - w_B) + \frac{(e^{\mathrm{T}}(\hat{w}_N - w_N))^2}{e^{\mathrm{T}} w_N} \leq (\hat{w} - w)^{\mathrm{T}} W^{-1} (\hat{w} - w). \tag{14}$$

In other words, the proximity measure for the aggregated model is bounded above by the proximity measure in the complete model. We keep moving indices $j \in N$ to B until the w-distance of w to CR has been reduced sufficiently in the complete model.

3 Predictor Step and Selection

The (B, N, y_0, x_N, z_N)-selected model is

$$\begin{aligned}
\min \quad & \tilde{b}_0 y_0 \\
\text{s.t.} \quad & M_{BB} x_B + z_B = y_0 \tilde{r}_B && (15) \\
& \tilde{r}_B^{\mathrm{T}} x_B + s_0 = \tilde{b}_0 && (16) \\
& x_B \in \Re^m \times \Re_+^{|B|-m},\ z_B \in \{0\}^m \times \Re_+^{|B|-m}, y_0, s_0 \in \Re_+. && (17)
\end{aligned}$$

The right hand side vector in the selected model is defined as

$$\tilde{r}_B = r_B - M_{BN} x_N / y_0,\ \tilde{b}_0 := b_0 - \frac{x_N^{\mathrm{T}} z_N}{y_0}. \tag{18}$$

The construction is such that the current solution for (3), is also feasible to (17). Conversely, given a solution (x_B^+, z_B^+, y_0^+) to (7), we let

$$\begin{aligned}
x_N^+ &= (y_0^+ / y_0) x_N, \\
s_0^+ &= \max\{b_0 - r^{\mathrm{T}} x^+, \theta\}.
\end{aligned}$$

Now, let

$$\begin{aligned}
\Delta z_N &= (y_0^+ - y_0)\tilde{r}_N - M_{NB}(x_B^+ - x_B). \\
\Delta s_0 &= -r^{\mathrm{T}}(x^+ - x) = -x_N^{\mathrm{T}} \Delta z_N / y_0.
\end{aligned}$$

If $s_0 + \Delta s_0 \geq \theta$ then set $z_N^+ = z_N + \Delta z_N$ and leave r unchanged.

We observe that if $s_0 + \Delta s_0 < \theta$ then

$$x_N^{\mathrm{T}} \Delta z_N = -y_0 \Delta s_0 > y_0(s_0 - \theta).$$

In this case, we choose z_N^+ such that

$$\begin{cases}
z_i^+ - z_i = \Delta z_i \text{ if } \Delta z_i \leq 0,\ i \in N, \\
0 \leq z_i^+ - z_i < \Delta z_i \text{ if } \Delta z_i > 0,\ i \in N, \\
x_N^{\mathrm{T}} \Delta z_N = y_0(s_0 - \theta).
\end{cases}$$

We update r_N accordingly. Observe that $r^+ \leq r$. We get

$$s_0^+ = \theta.$$

Observe that these modifications do not influence the proximity to the central region. However, the progress is more than usual since not only is y_0 decreased, but also $\|[r]_+\}$.