A new variant of the interior point method, for large scale and warm started linear programs

Jos F. Sturm
j.f.sturm@uvt.nl

http://center.uvt.nl/staff/sturm

CentER, Tilburg, The Netherlands

Overview

- IPM vs. Other Approaches
- Linear Programming
- IPM Centrality
- Predictor-Corrector
- Aggregate Model Corrector
- Submodel Predictor
- Future work

IPM vs. Other Approaches

- IPM iteration factors $n \times n$ matrix variable Z in each iteration. Slow for large n.

IPM vs. Other Approaches

- IPM iteration factors $n \times n$ matrix variable Z in each iteration. Slow for large n.
- Bundle and NLP approaches merely need p such that $p^T Z p < 0$. Uses Lanczos. Fast for large n.

IPM vs. Other Approaches

- IPM iteration factors $n \times n$ matrix variable Z in each iteration. Slow for large n.
- Bundle and NLP approaches merely need p such that $p^T Z p < 0$. Uses Lanczos. Fast for large n.
- IPM works on complete model (A, b, c).

IPM vs. Other Approaches

- IPM iteration factors $n \times n$ matrix variable Z in each iteration. Slow for large n.
- Bundle and NLP approaches merely need p such that $p^T Zp < 0$. Uses Lanczos. Fast for large n.
- IPM works on complete model (A, b, c).
- Simplex only on (A_B, b, c_B).

IPM vs. Other Approaches

- IPM iteration factors $n \times n$ matrix variable Z in each iteration. Slow for large n.
- Bundle and NLP approaches merely need p such that $p^T Zp < 0$. Uses Lanczos. Fast for large n.
- IPM works on complete model (A, b, c).
- Simplex only on (A_B, b, c_B).
- Simplex benefits from warm start.

IPM vs. Other Approaches

- IPM iteration factors $n \times n$ matrix variable Z in each iteration. Slow for large n.
- Bundle and NLP approaches merely need p such that $p^T Zp < 0$. Uses Lanczos. Fast for large n.
- IPM works on complete model (A, b, c).
- Simplex only on (A_B, b, c_B).
- Simplex benefits from warm start.
- Low rank updates.

IPM vs. Other Approaches

- IPM iteration factors $n \times n$ matrix variable Z in each iteration. Slow for large n.
- Bundle and NLP approaches merely need p such that $p^T Zp < 0$. Uses Lanczos. Fast for large n.
- IPM works on complete model (A, b, c).
- Simplex only on (A_B, b, c_B).
- Simplex benefits from warm start.
- Low rank updates.
- Exploiting structure.

Context

- Bundle: Helmberg, Rendl, Oustry, Madhu N. a.o.
- NLP approach of Burer, Monteiro, Zhang.
- Active set, simplex variants: Goldfarb, Y. Zhang.
- Projection methods: Lin-Han, Orsi-Rami-Moore.
- ACCPM: Goffin, Vial a.o.
- Low rank IPM: Karmarkar a.o.

Linear Programming

$$(P)\quad \begin{array}{ll} \min & c^{\mathrm{T}}x \\ \text{s.t.} & Ax - s = b \\ & x \in \Gamma,\ s \in \mathcal{K} \end{array}$$

$$(D)\quad \begin{array}{ll} \max & b^{\mathrm{T}}y \\ \text{s.t.} & A^{\mathrm{T}}y + z = c \\ & z \geq \Gamma^*, y \in \mathcal{K}^* \end{array}$$

Here $\mathcal{K} = \{0\}^m$ or $\mathcal{K} = \Re^m_+$.

Unify

$$x = \begin{bmatrix} y \\ `x' \end{bmatrix}, \; z = \begin{bmatrix} s \\ `z' \end{bmatrix},$$

$$\forall i \in \mathcal{I} : \begin{cases} x_i \geq 0 \\ z_i \geq 0. \end{cases}$$

$$\forall i \notin \mathcal{I} : \begin{cases} x_i \text{ is free} \\ z_i = 0. \end{cases}$$

All slacks (P,D) in z.

IPM vs. Active Set

- Active Set moves along the *boundary*

IPM vs. Active Set

- Active Set moves along the *boundary*
- Warm start initiates from *complementary* solution

IPM vs. Active Set

- Active Set moves along the *boundary*
- Warm start initiates from *complementary* solution

But IPM follows Central Path in *interior*!

IPM Centrality

Wide Region

- Let

$$w_i = x_i z_i, \ i \in \mathcal{I}.$$

Wide Region

- Let

$$w_i = x_i z_i,\ i \in \mathcal{I}.$$

- Central path

$$\{w \mid w_i = w_j > 0\, \forall i, j \in \mathcal{I}\}.$$

Wide Region

- Let

$$w_i = x_i z_i, \; i \in \mathcal{I}.$$

- Central path

$$\{w \mid w_i = w_j > 0 \, \forall i, j \in \mathcal{I}\}.$$

- Central region (CR)

$$\{w \mid |\mathcal{I}| \min_{i \in \mathcal{I}} w_i \geq \theta \sum_i w_i > 0\}.$$

Wide Region

- Let

$$w_i = x_i z_i,\ i \in \mathcal{I}.$$

- Central path

$$\{w \mid w_i = w_j > 0 \,\forall i, j \in \mathcal{I}\}.$$

- Central region (CR)

$$\{w \mid |\mathcal{I}| \min_{i \in \mathcal{I}} w_i \geq \theta \sum_i w_i > 0\}.$$

- $\theta = 1$ yields CP; choose $0 < \theta < 1$.

Wide Region(2)

- w^θ is projection of w onto CR

Wide Region(2)

- w^θ is projection of w onto CR
- Proximity measure (Roos et al.)

$$\delta_w^\theta = \frac{\|w^\theta - w\|_w}{\|w\|_w},$$

with $\|x\|_w^2 = \sum_{i\in\mathcal{I}} \frac{x_i^2}{w_i}.$.

Wide Region(2)

- w^θ is projection of w onto CR
- Proximity measure (Roos et al.)

$$\delta_w^\theta = \frac{\|w^\theta - w\|_w}{\|w\|_w},$$

with $\|x\|_w^2 = \sum_{i \in \mathcal{I}} \frac{x_i^2}{w_i}$..

- Neighborhood $\{w | \delta_w^\theta \leq \beta\}$.

Predictor-Corrector Method

- Corrector Step reduces $\delta_w^\theta \leq \beta$ to $O(\beta^2)$. (Newton step to $x_i z_i = w_i^\theta$.)
- Predictor Step reduces duality gap until $\delta_w^\theta \approx \beta$.

Corrector Basics

Linearize

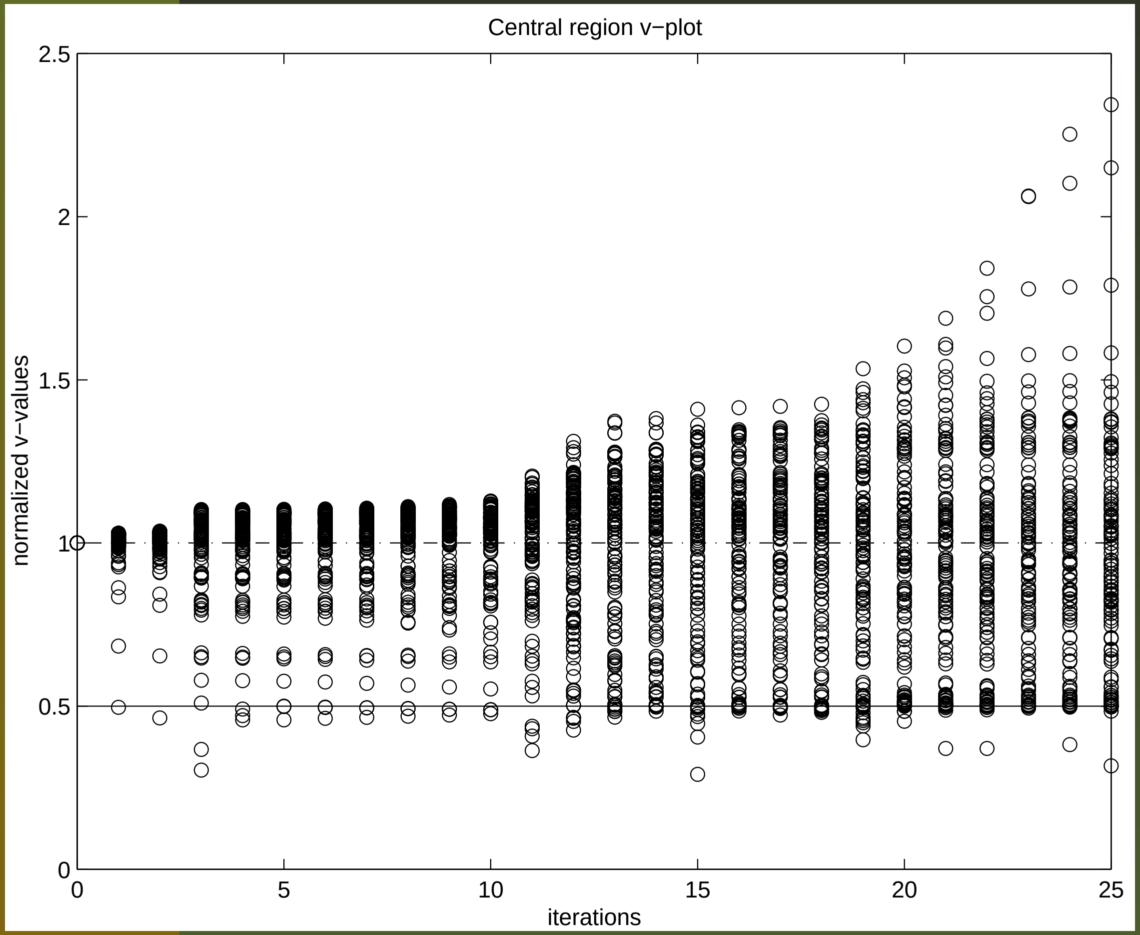
$$(x_i + \Delta x_i)(z_i + \Delta z_i) = w_i^\theta$$

yielding

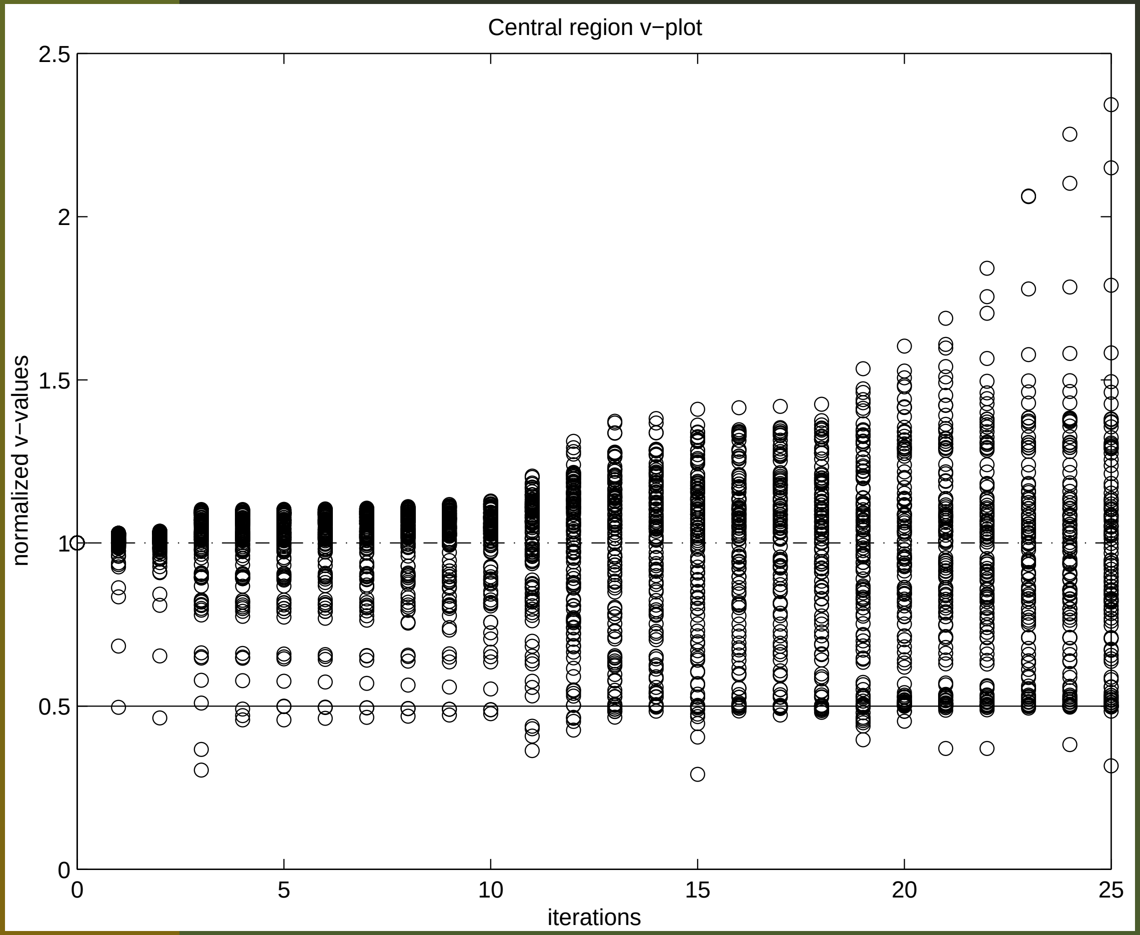
$$w_i + x_i \Delta z_i + z_i \Delta x_i = w_i^\theta.$$

$(\, i \in \mathcal{I} \,)$

Centrality Plot



Centrality Plot



Why not aggregate variables above threshold?

First some details...

Optimality Conditions

$$\begin{array}{lllll} & A^{\mathrm{T}}y & -c & +z & = 0 \\ -Ax & & +b & +s & = 0 \\ c^{\mathrm{T}}x & -b^{\mathrm{T}}y & & & \leq 0 \end{array}$$

with

$$x \geq 0,\ s \in \mathcal{K},$$

$$z \geq 0, y \in \mathcal{K}^*.$$

Skew-Symmetric Form

Data matrix: $M = -M^{\mathrm{T}}$.

$$\begin{array}{ll} \min & b_0 y_0 \\ \text{s.t.} & Mx + z = y_0 r \\ & r^{\mathrm{T}} x = b_0 \\ & x \in \Re^m \times \Re^n_+,\; z \in \{0\}^m \times \Re^n_+ \end{array}$$

Skew-Symmetric Form

Data matrix: $M = -M^{\mathrm{T}}$.

$$\begin{array}{ll} \min & b_0 y_0 \\ \text{s.t.} & Mx + z = y_0 r \\ & r^{\mathrm{T}} x = b_0 \\ & x \in \Re^m \times \Re^n_+, \; z \in \{0\}^m \times \Re^n_+ \end{array}$$

Observe: $x^{\mathrm{T}} z = y_0 r^{\mathrm{T}} x = b_0 y_0 \geq 0$. Optimum 0. r is residual vector (from initialization). Possible choice $b_0 = 1$.

Aggregate Model Corrector

For partition (B, N) aggregate at (x^k, z^k)

$$a_0 := M_{BN} x_N^k$$

$$r_0 := r_N^{\mathrm{T}} x_N^k$$

Aggregate model:

$$\begin{array}{ll} \min & b_0 y_0 \\ \text{s.t.} & M_{BB} x_B + x_0 a_0 + z_B = y_0 r_B \\ & -a_0^{\mathrm{T}} x_B + z_0 = y_0 r_0 \\ & r_B^{\mathrm{T}} x_B + r_0 x_0 = b_0 \end{array}$$

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

Aggregate Model Corrector (2)

- Initialize

$$B = \{i \,|\, |\mathcal{I}| w_i < \theta \sum_{\mathcal{I}} w_j\}$$

Aggregate Model Corrector (2)

- Initialize

$$B = \{i \,|\, |\mathcal{I}| w_i < \theta \sum_{\mathcal{I}} w_j\}$$

- As long as the inner centrality condition is not met in Master Problem, we disaggregate the most-violating variables.

Aggregate Model Corrector (2)

- Initialize

$$B = \{i \mid |\mathcal{I}| w_i < \theta \sum_{\mathcal{I}} w_j\}$$

- As long as the inner centrality condition is not met in Master Problem, we disaggregate the most-violating variables.
- Single disaggregation dominated by rank-1 update (quadratic time, not cubic).

Aggregate Model Corrector (2)

- Initialize

$$B = \{i | |\mathcal{I}| w_i < \theta \sum_{\mathcal{I}} w_j\}$$

- As long as the inner centrality condition is not met in Master Problem, we disaggregate the most-violating variables.
- Single disaggregation dominated by rank-1 update (quadratic time, not cubic).
- Efficiency determined by size of final B.

Aggregate Model Corrector (2)

- Initialize

$$B = \{i \| \mathcal{I}| w_i < \theta \sum_{\mathcal{I}} w_j\}$$

- As long as the inner centrality condition is not met in Master Problem, we disaggregate the most-violating variables.
- Single disaggregation dominated by rank-1 update (quadratic time, not cubic).
- Efficiency determined by size of final B.
- (SDP: achieve centrality only in subproblem)

Predictor Basics

Linearize

$$(x_i + \Delta x_i)(z_i + \Delta z_i) = 0$$

yielding

$$w_i + x_i \Delta z_i + z_i \Delta x_i = 0.$$

Submodel Predictor

Select only B-variables:

$$\begin{array}{ll} \min & \tilde{b}_0 y_0 \\ \text{s.t.} & M_{BB} x_B + z_B = y_0 \tilde{r}_B \\ & \tilde{r}_B^{\mathrm{T}} x_B = \tilde{b}_0 \end{array}$$

with residual $y_0^k \tilde{r}_B = M_{BB} x_B^k + z_B^k$,

$$y_0 (r_B - \tilde{r}_B) = M_{BN} x_N.$$

In Master Model:

$$x_N = \frac{y_0}{y_0^k} x_N^k.$$

Sub Model Predictor (2)

- Initialize B with estimate of binding dual constraints. First step: based on warm start. Subsequent steps: based on previous step.

Sub Model Predictor (2)

- Initialize B with estimate of binding dual constraints. First step: based on warm start. Subsequent steps: based on previous step.
- Step length based on $\|\Delta X_B \Delta z_B\|_w$.

Sub Model Predictor (2)

- Initialize B with estimate of binding dual constraints. First step: based on warm start. Subsequent steps: based on previous step.
- Step length based on $\|\Delta X_B \Delta z_B\|_w$.
- If not in Master Problem's neighborhood then introduce the most-violating variable $j \in N$, i.e.

$$w_j \leq 0 \text{ or } j \in \arg\max_i \frac{(w_i - w_i^\theta)^2}{w_i}.$$

Rank-1 update.

Sub Model Predictor (2)

- Initialize B with estimate of binding dual constraints. First step: based on warm start. Subsequent steps: based on previous step.
- Step length based on $\|\Delta X_B \Delta z_B\|_w$.
- If not in Master Problem's neighborhood then introduce the most-violating variable $j \in N$, i.e.

$$w_j \leq 0 \text{ or } j \in \arg\max_i \frac{(w_i - w_i^\theta)^2}{w_i}.$$

  Rank-1 update.
- Worst-case rate $1 - 1/O(\sqrt{|B|})$.

(SDP remarks)

- Update increments order of matrix variable, thus low (not 1) rank update
- MP: maintain feasibility
- SP: also centrality

Predictor Issues

- For Master-Problem $\Delta x^{\mathrm{T}} \Delta z \neq 0$ hence normalization b_0 changes. Merit Function.

Predictor Issues

- For Master-Problem $\Delta x^{\mathrm{T}} \Delta z \neq 0$ hence normalization b_0 changes. Merit Function.
- Impact on centrality, especially if $\Delta x^{\mathrm{T}} \Delta z > 0$.

Predictor Issues

- For Master-Problem $\Delta x^{\mathrm{T}} \Delta z \neq 0$ hence normalization b_0 changes. Merit Function.
- Impact on centrality, especially if $\Delta x^{\mathrm{T}} \Delta z > 0$.
- Alternative: keep b_0 constant, $\Delta x^{\mathrm{T}} \Delta z = 0$ by changing r_N. Maintain $-r_\infty \leq r_N \leq r_\infty$.

Implementational Issues

- Rank deficiency of reduced system $A_B D A_B^{\mathrm{T}}$. Standard rank-1 update formula not applicable.

Implementational Issues

- Rank deficiency of reduced system $A_B D A_B^{\mathrm{T}}$. Standard rank-1 update formula not applicable.
- Remedy: product-form-Cholesky

$$L_0 L_1 \dots L_k \Theta_k L_k^{\mathrm{T}} L_{k-1}^{\mathrm{T}} \dots L_0^{\mathrm{T}}.$$

Implementational Issues

- Rank deficiency of reduced system $A_B D A_B^{\mathrm{T}}$. Standard rank-1 update formula not applicable.
- Remedy: product-form-Cholesky

$$L_0 L_1 \ldots L_k \Theta_k L_k^{\mathrm{T}} L_{k-1}^{\mathrm{T}} \ldots L_0^{\mathrm{T}}.$$

- Not only columns, also rows may be added

Implementational Issues

- Rank deficiency of reduced system $A_B D A_B^{\mathrm{T}}$. Standard rank-1 update formula not applicable.
- Remedy: product-form-Cholesky

$$L_0 L_1 \dots L_k \Theta_k L_k^{\mathrm{T}} L_{k-1}^{\mathrm{T}} \dots L_0^{\mathrm{T}}.$$

- Not only columns, also rows may be added
- Exploiting sparsity

Implementational Issues

- Rank deficiency of reduced system $A_B D A_B^{\mathrm{T}}$. Standard rank-1 update formula not applicable.
- Remedy: product-form-Cholesky

$$L_0 L_1 \ldots L_k \Theta_k L_k^{\mathrm{T}} L_{k-1}^{\mathrm{T}} \ldots L_0^{\mathrm{T}}.$$

- Not only columns, also rows may be added
- Exploiting sparsity
- Pivot ordering in augmented self-dual system

Implementational Issues

- Rank deficiency of reduced system $A_B D A_B^{\mathrm{T}}$. Standard rank-1 update formula not applicable.
- Remedy: product-form-Cholesky

$$L_0 L_1 \dots L_k \Theta_k L_k^{\mathrm{T}} L_{k-1}^{\mathrm{T}} \dots L_0^{\mathrm{T}}.$$

- Not only columns, also rows may be added
- Exploiting sparsity
- Pivot ordering in augmented self-dual system
- Prepare for future extensions

Implementational Issues

- Rank deficiency of reduced system $A_B D A_B^{\mathrm{T}}$. Standard rank-1 update formula not applicable.
- Remedy: product-form-Cholesky

$$L_0 L_1 \ldots L_k \Theta_k L_k^{\mathrm{T}} L_{k-1}^{\mathrm{T}} \ldots L_0^{\mathrm{T}}.$$

- Not only columns, also rows may be added
- Exploiting sparsity
- Pivot ordering in augmented self-dual system
- Prepare for future extensions
- Submodel pre-solving?

Finally...

- Computational savings
- Iterates potentially more powerful
- Benefits from warm start
- Both rows and columns can be discarded
- Modifications for large SDP.
- More at ISMP2003, SIOPT2005, ISMP2006.