On Interior-Point Warmstarts for Linear and Combinatorial Optimization

Miguel F. Anjos

Department of Management Sciences
University of Waterloo, Ontario, Canada

Joint work with Alexander Engau (Waterloo & Colorado-Denver) and Anthony Vannelli (Guelph)

Research partially supported by **MITACS**

Aussois 2009

Warmstarting

The use of information from a *solved* problem
to accelerate the (re-)optimization of a *similar* problem

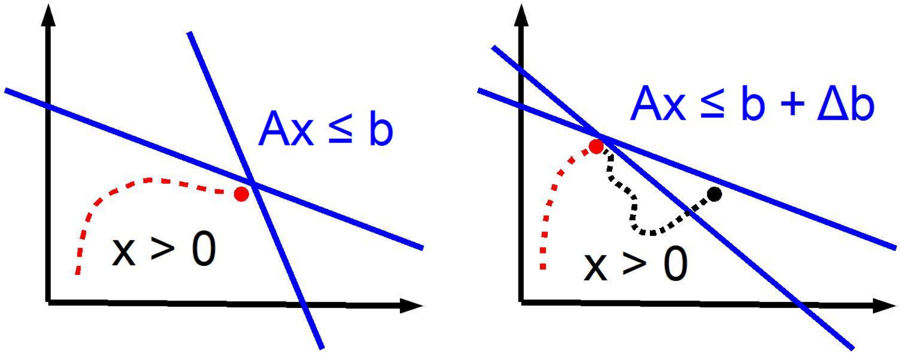
1. Same structure and size, similar data
   - changes of right-hand side
   - changes of objective function
   - changes of real-life data in financial and engineering applications (e.g., variation in market prices and/or product specifications)

2. Modified structure, more or less (of) the same data
   - series of LP/SDP relaxations for combinatorial optimization
   - removal of variables and/or constraints in branch-and-cut schemes

Results presented today are for LP, but an important motivation is to work towards improved warmstarting of interior-point methods (IPMs) for solving SDP relaxations of combinatorial optimization problems.

Illustration: Inherent Problem of IPM Warmstarts

Re-optimization after right-hand side perturbation



- The active set is unchanged $\Rightarrow$ Simplex warmstart in 1 iteration
- The interior has changed $\Rightarrow$ effective IPM warmstart not straightforward

Linear Programming in Primal-Dual Standard Form

(LP-P) $\min c^T x$ s.t. $Ax = b$, $x \geq 0$

(LP-D) $\max b^T y$ s.t. $A^T y + s = c$, $s \geq 0$

First-Order (Karush-Kuhn-Tucker) Optimality Conditions

$$Ax = b \quad (x \geq 0) \qquad \text{(Primal Feasibility)}$$

$$A^T y + s = c \quad (s \geq 0) \qquad \text{(Dual Feasibility)}$$

$$x_i s_i = \mu \quad \text{for all } i \qquad \text{(Complementarity)}$$

- System has a unique solution for every $\mu > 0$ ("central path")
- Central path converges to an optimal solution as $\mu \to 0$
- KKT system is nonlinear $\Rightarrow$ linearize & solve by Newton's method

Basic Idea of Interior-Point Methods: apply Newton's with decreasing μ

Infeasible Primal-Dual Path-Following IPM

1. Initialize $(x^0 > 0, y^0, s^0 > 0)$, let $\sigma \in [0,1)$, and set $k = 0$.
2. If $\|Ax^k - b\| \le \varepsilon_b$, $\|A^T y^k + s^k - c\| \le \varepsilon_c$, $(x^k)^T(s^k)/n \le \varepsilon_d$, stop.
3. Set $\mu^k = \sigma(x^k)^T(s^k)/n$ and solve for Newton direction

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = - \begin{bmatrix} Ax^k - b \\ A^T y^k + s^k - c \\ X^k S^k e - \mu^k e \end{bmatrix}$$

4. Compute step length α^k such that

$$x^k + \alpha^k \Delta x > 0 \text{ and } s^k + \alpha^k \Delta s > 0.$$

5. Take step $(x^{k+1}, y^{k+1}, s^{k+1}) = (x^k, y^k, s^k) + \alpha^k(\Delta x, \Delta y, \Delta s)$.
6. Increase k by 1 and go to Step 2.

Warmstarting in Linear Optimization

$$\begin{aligned} \min\ & (c^\circ + \Delta c)^T x \\ \text{s.t. } & (A^\circ + \Delta A)x = b^\circ + \Delta b \\ & x \geq 0 \end{aligned} \qquad \begin{aligned} \max\ & (b^\circ + \Delta b)^T y \\ \text{s.t. } & (A^\circ + \Delta A)^T y + s = c^\circ + \Delta c \\ & s \geq 0 \end{aligned}$$

Let $(x^\circ \geq 0, y^\circ, s^\circ \geq 0)$ be optimal for the initial problem

$$A^\circ x^\circ = b^\circ \qquad {A^\circ}^T y^\circ + s^\circ = c^\circ \qquad X^\circ S^\circ e = 0$$

then $(x^\circ, y^\circ, s^\circ)$ is typically infeasible for the perturbed problem

$$r_b^\circ = b - Ax^\circ = (b^\circ + \Delta b) - (A^\circ + \Delta A)x^\circ = \Delta b - \Delta A x^\circ$$
$$r_c^\circ = c - A^T y^\circ - s = (c^\circ + \Delta c) - (A^\circ + \Delta A)^T y^\circ - s^\circ = \Delta c - \Delta A^T y^\circ$$

- Good news: Feasibility can be handled by infeasible algorithm
- Bad news: Initial interiority must be achieved by alternative means

Warmstarting IPM (I): Adjustment of Initial Iterate

Newton adjustment (*Yildirim & Wright 2002, Gondzio & Grothey 2003*)

$$\begin{array}{ccccc} A\Delta x & & & & = r_b^\circ \\ & A^T\Delta y & + & \Delta s & = r_c^\circ \\ S^\circ\Delta x & & + & X^\circ\Delta s & = 0 \end{array}$$

(Weighted) least-squares adjustments (*Yildirim & Wright 2002*)

$$\min_{\Delta x} \|\Sigma\Delta x\| \text{ s.t. } A(x^\circ + \Delta x) = b, x^\circ + \Delta x \geq 0$$

$$\min_{\Delta y, \Delta s} \|\Lambda\Delta s\| \text{ s.t. } A^T(y^\circ + \Delta y) + (s^\circ + \Delta s) = c, s^\circ + \Delta s \geq 0$$

Warmstarting IPM (II): Shifted-Barrier Approaches

Shifted-barrier problem (*Freund 1991*)

$$\min\ c^T x - \varepsilon \sum \log(x_i + \varepsilon h_i) \text{ s.t. } Ax = b,\ x + \varepsilon h > 0$$

Infeasible-start shifted-barrier problem (*Freund 1996*)

$$\min\ (c + \varepsilon(A^T y^\circ + s^\circ - c))^T x - \mu\varepsilon \sum \log(x_i)$$
$$\text{s.t. } Ax = b + \varepsilon(Ax^\circ - b),\ x > 0$$

Exact primal-dual penalty-method approach (*Benson & Shanno 2007*)

$$\begin{aligned} \min\ & c^T x + d^T \xi \\ \text{s.t. }\ & Ax = b \\ & 0 \le x + \xi \le u \\ & \xi \ge 0 \end{aligned} \qquad\qquad \begin{aligned} \max\ & b^T y - u^T \psi \\ \text{s.t. }\ & A^T y + s = c \\ & -\psi \le s \le d \\ & \psi \ge 0 \end{aligned}$$

New: Primal-Dual Slack Approach

Key Idea: Introduce slack variables for the non-negativity constraints in both primal and dual:

$$\begin{aligned} \min\ & c^T x \\ \text{s.t. } & Ax = b \\ & x - \xi = 0 \\ & \xi \geq 0 \end{aligned} \qquad\qquad \begin{aligned} \max\ & b^T y \\ \text{s.t. } & A^T y + s = c \\ & s - \psi = 0 \\ & \psi \geq 0 \end{aligned}$$

- Good news: Penalty parameters are no longer needed
- Bad news: x (and s) are now free variables

Theoretical ("Nice") Complexity Results

Iteration complexity and worst-case iteration count

- Comparable computational cost per iteration as standard form LP
- Detailed analysis yields same worst-case iteration bound as IIPM

Initialization schemes for theoretical analysis

- Standard form: $(x^0, y^0, s^0) = (\zeta e, 0, \zeta e)$ where $\zeta \geq \|(x^*, s^*)\|_\infty$
- Slacked form: $(x^0, y^0, s^0, \xi^0, \psi^0) = (x^\circ, y^\circ, s^\circ, \zeta e, \zeta e)$

Iteration complexity of slacked-form IIPM (*Engau, A., Vannelli 2008*)

$$O(n) \log \left(\max \left\{ \frac{\|r_b^0\|}{\varepsilon_b}, \frac{\|r_c^0\|}{\varepsilon_c}, \frac{\|r_x^0\|}{\varepsilon_x}, \frac{\|r_s^0\|}{\varepsilon_s}, \frac{\zeta^2 n}{\varepsilon_d} \right\} \right)$$

- Easy to show that $\max\left\{\|r_x^0\|, \|r_s^0\|\right\} \leq \zeta^2 n$ for $\zeta \geq \|(x^\circ, s^\circ, 1)\|_\infty$

Practical ("Messy") Issue: Slack Initialization

- In theory: $(\xi^0, \psi^0) = (\zeta e, \zeta e)$ where $\zeta \geq \|(x^*, s^*)\|_\infty = \|(\xi^*, \psi^*)\|_\infty$
- In practice: $(\xi^0, \psi^0) = (x^\circ, s^\circ) + (r_x^0, r_s^0)$ such that

$$\Xi^0 \Psi^0 e = X^\circ S^\circ e + X^\circ R_s^0 e + R_x^0 S^\circ e + R_x^0 R_s^0 e$$

is well-balanced and similar in magnitude to $\rho = \left\|(r_b^\circ, r_c^\circ, 1)\right\|_\infty$.

Also: Add primal-dual indicator to detect variables that are sufficiently away from zero, and do not need to be "slacked".

Warmstarting for LP: Perturbation Scheme

(described by *Benson & Shanno 2007* and *Gondzio & Grothey 2008*)

- Test set: (perturbable) Netlib LP problems of size $m + n \leq 1000$
- "LP" solver: SDPT3 (freely available, easy supply of initial points) with supplemental "free-variable" code by *Anjos & Burer 2008*
- Perturbation: 10% or 20 (random) entries of A, b, c on average

$$\Delta c_i = \begin{cases} \epsilon\delta & \text{if } c_i^\circ = 0 \\ \epsilon\delta c_i^\circ & \text{otherwise} \end{cases} \quad \text{where } \epsilon \in [-1, 1], \delta \in \{0.1, 0.01, 0.001\}$$

(same scheme for b and A but preserving sparsity structure of A)

- Performance measures: perturbation levels, number of iterations

$$\text{WCR} = \text{Warm-to-Coldstart-Ratio} = \frac{\text{number of warmstart iterations}}{\text{number of coldstart iterations}}$$

Computational Results: LP Perturbations

(Feasible) Perturbations				WCR Statistics				
Δ	δ	Pert	(#)	avg	stdv	min	med	max
b	0.001	15.3	46	0.23	0.14	0.06	0.17	0.63
b	0.01	14.9	42	0.28	0.14	0.11	0.25	0.65
b	0.1	15.0	40	0.39	0.18	0.15	0.36	0.82
c	0.001	17.9	62	0.24	0.13	0.08	0.20	0.64
c	0.01	18.5	59	0.33	0.17	0.11	0.28	0.90
c	0.1	18.0	54	0.44	0.22	0.13	0.40	1.13
A	0.001	19.2	60	0.32	0.20	0.13	0.25	0.87
A	0.01	19.9	58	0.41	0.25	0.14	0.37	1.74
A	0.1	19.9	58	0.58	0.27	0.21	0.56	1.75
Abc	0.001	52.8	46	0.33	0.15	0.12	0.29	0.64
Abc	0.01	51.6	42	0.46	0.17	0.13	0.45	0.83
Abc	0.1	51.5	36	0.74	0.30	0.38	0.68	1.66

Warmstarting for Combinatorial Optimization: Max-cut

Let $x_{ij} = \begin{cases} 1 & \text{if vertices } i \text{ and } j \text{ lie on opposite sides of the cut} \\ 0 & \text{if vertices } i \text{ and } j \text{ lie on the same side of the cut} \end{cases}$

$$
\begin{array}{rll}
\max & \sum_{i<j} c_{ij} x_{ij} & \\
\text{subject to} & \left.\begin{array}{r} x_{ij} + x_{ik} + x_{jk} \leq 2 \\ x_{ij} - x_{ik} - x_{jk} \leq 0 \\ -x_{ij} + x_{ik} - x_{jk} \leq 0 \\ -x_{ij} - x_{ik} + x_{jk} \leq 0 \end{array}\right\} & \text{for } 1 \leq i < j < k \leq n \\
& x_{ij} \in \{0, 1\} & \text{for } 1 \leq i < j \leq n
\end{array}
$$

- Relax binary constraint to box constraint $0 \leq x_{ij} \leq 1$
- Drop triangle inequalities and sequentially add constraints as cuts
- Solve resulting relaxations from cold/warmstart to compute WCR

Computational Results: Max-cut on complete graphs

- Set of 100 randomly generated instances, up to 10 rounds of cuts

Problem size	# cuts	avg	sdev	min	med	max
30	1	0.62	0.04	0.54	0.63	0.76
30	10	0.77	0.06	0.62	0.76	0.87
30	100	1.03	0.04	0.96	1.03	1.15
60	1	0.60	0.05	0.54	0.60	0.81
60	10	0.76	0.08	0.59	0.78	0.92
60	100	0.92	0.06	0.75	0.93	1.14
100	1	0.59	0.04	0.52	0.56	0.72
100	10	0.75	0.09	0.59	0.76	0.90
100	100	0.79	0.06	0.68	0.80	0.85

Computational Results: Max-cut on toroidal square grid graphs

- Set of 100 randomly generated instances, up to 10 rounds of cuts

Problem size	# cuts	avg	sdev	min	med	max
36	1	0.73	0.10	0.44	0.75	0.86
36	10	0.71	0.09	0.50	0.75	0.86
36	100	0.69	0.06	0.44	0.67	0.75
64	1	0.73	0.05	0.50	0.75	0.86
64	10	0.73	0.04	0.50	0.75	0.86
64	100	0.68	0.06	0.38	0.67	0.75
100	1	0.70	0.07	0.44	0.75	0.75
100	10	0.70	0.07	0.40	0.67	0.75
100	100	0.66	0.06	0.40	0.67	0.75

Summary and Outlook

Take-home message: The proposed new warmstarting scheme

- is supported by theoretical and preliminary computational results
- is applicable for data perturbations and new variables and/or constraints
- consistently achieves a reduction of the iteration count by $\geq 30\%$ in comparison with coldstart; competitive with other approaches on perturbed Netlib problems.

Where we are going from here:

- improve benchmarking possibilities via collection of a meaningful set of LP test problems
- carefully compare, and explore combinations, with other warmstarting approaches
- extend from LP to NLP (and particularly to SOCP and SDP)

For Jack Edmonds (and everyone else):

- Paper is available on my webpage:
  `http://mfa.research.uwaterloo.ca`
- On the same page, there is also an online addendum with all the data files & detailed computational results.