Finding a well-centred point for a set of polyhedral constraints

Nick Gould (RAL)
joint with
Coralia Cartis (OUCL)

\[ Ax = b, \; x \geq 0 \]

Motivation — feasible points

■ many optimization problems require \( x \in \mathbb{R}^n \):
  \[ Ax = b \quad \text{and} \quad x \geq 0 \]
  for given \( A, b \)
  ■ usually under determined \( \Longrightarrow \) select other important characteristics:
    - minimize \( c^T x \): linear programming
    - minimize \( \frac{1}{2} x^T H x + c^T x \): quadratic programming
    - minimize \( f(x) \) & maybe \( c(x) \geq 0 \): nonlinear programming

■ everything we say applies more generally to
  \[ Ax = b \quad \text{and} \quad l \leq x \leq u \]
  for given \( A, b, l, u \ldots \)
Motivation — interior points

- often desirable to find **interior-point** (IP) \( x \):

\[
Ax = b \quad \text{and} \quad x > 0
\]

Why?

- good starting point for **feasible-point** IP methods for convex problems
- low computational complexity of feasible-point IP methods
- reduces dimension of feasible region for more general non-convex problems
- subsequent iterates \( x + \Delta x \) satisfy

\[
A\Delta x = 0 \quad \text{and} \quad \Delta x \geq -x
\]

\( \implies \) all iterates lie in null-space of \( A \)
- often lessens influence of non-convexity

Motivation — best interior point

What is the “best” interior point?

- in the absence of any objective (bounded case):
  - **centroid**? — expensive
  - **analytic center**?

\[
\arg \max \prod_i x_i \equiv \arg \min -\sum_i \log x_i
\]

such that \( Ax = b \) and \( x > 0 \)
Best interior point II

If there is an objective, e.g. $c^T x$, **balance** analytic center with objective

$$\min_{x>0} c^T x - \mu \sum_i \log x_i$$

such that $Ax = b$

for some **fixed** target $\mu > 0$ $$\implies$$ **optimality conditions**

$$c - \mu \sum_i x_i^{-1} - A^T y = 0$$

and $Ax = b$ where $x > 0$

$$\implies$$ **central path**

$$Ax - b = 0$$

$$A^T y + s - c = 0$$

and $XSe = \mu e$

$$(x, s) > 0$$, parameter $\mu > 0$, $X = \text{diag } x$, $S = \text{diag } s$, $e = \text{vector of 1s}$

**Summary of the talk**

**Aim**: given $\mu > 0$, find $x > 0$ (along with $s > 0$ and $y$):

$$Ax - b = 0$$

$$A^T y + s - c = 0$$

and $XSe = \mu e$

- assumptions
- “obvious” Newton method with safeguards
- convergence behaviour
- numerical experience
- extensions and future work
Assumptions

\[ Ax - b = 0, \quad A^T y + s - c = 0 \quad \text{and} \quad XSe = \mu e \]

Assume (to start with)

- \( A \) is full rank
- \( \exists (x, s) > 0 : Ax = b \quad \text{and} \quad A^T y + s = c \)

\[ \implies \text{central path well defined} \]

Newton’s method

\[ Ax - b = 0, \quad A^T y + s - c = 0 \quad \text{and} \quad XSe = \mu e \]

Given \((x_k, s_k) > 0 \& y_k\), the primal-dual Newton correction satisfies

\[
\begin{pmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S_k & 0 & X_k \end{pmatrix} \begin{pmatrix} \dot{x}_k \\ \dot{y}_k \\ \dot{s}_k \end{pmatrix} = - \begin{pmatrix} Ax_k - b \\ A^T y_k + s_k - c \\ X_k S_k e - \mu e \end{pmatrix}
\]

Linesearch \((x_{k+1}, y_{k+1}, s_{k+1}) = v_k(\alpha_k)\):

\[
v_k(\alpha) \equiv \begin{pmatrix} x_k(\alpha) \\ y_k(\alpha) \\ s_k(\alpha) \end{pmatrix} = \begin{pmatrix} x_k \\ y_k \\ s_k \end{pmatrix} + \alpha \begin{pmatrix} \dot{x}_k \\ \dot{y}_k \\ \dot{s}_k \end{pmatrix}
\]

for some suitable \(\alpha_k \in (0, 1] \)
Safeguards

\[
(x_k(\alpha), y_k(\alpha), s_k(\alpha)) = (x_k + \alpha \dot{x}_k, y_k + \alpha \dot{y}_k, s_k + \alpha \dot{s}_k)
\]

Pick \( \alpha_k \in (0, 1] \) to

- ensure \( (x_k(\alpha_k), s_k(\alpha_k)) > 0 \)
- given \( \omega \in (0, 1) \), insist
  \[
  X_k(\alpha)s_k(\alpha) \geq \omega \mu e
  \]
  for all \( 0 \leq \alpha \leq \alpha_k \)

- (sufficiently) decrease some measure of optimality
- obvious merit function

\[
\Phi(x, y, s) = \|Xs - \mu e\| + r(x, y, s) : \quad r(x, y, s) = \|Ax - b\| + \|ATy + s - c\|
\]

Nonzero stepsizes

\[
\begin{pmatrix}
  A & 0 & 0 \\
  0 & AT & I \\
  S_k & 0 & X_k
\end{pmatrix}
\begin{pmatrix}
  \dot{x}_k \\
  \dot{y}_k \\
  \dot{s}_k
\end{pmatrix}
= -\begin{pmatrix}
  Ax_k - b \\
  ATy_k + s_k - c \\
  X_k s_k - \mu e
\end{pmatrix}
\]

Require

\[
X_k(\alpha)s_k(\alpha) \geq \omega \mu e
\]

for all \( 0 \leq \alpha \leq \alpha_k \) for given \( \omega \in (0, 1) \)

\[
X_k(\alpha) s_k(\alpha) - \omega \mu e
= X_k s_k + \alpha (S_k \dot{x}_k + X_k \dot{s}_k) + \alpha^2 \dot{X}_k \dot{s}_k - \omega \mu e
= (1 - \alpha)(X_k s_k - \omega \mu e) + \alpha (1 - \omega) \mu e + \alpha^2 \dot{X}_k \dot{s}_k
\]

At least one of the first two terms \( > 0 \) \( \implies \) nonzero stepsize
Decrease in merit functions

\[ \Phi(v) \equiv \Phi(x, y, s) = \|Xs - \mu e\| + r(x, y, s) : \]
\[ r(v) \equiv r(x, y, s) = \|Ax - b\| + \|ATy + s - c\| \]

- primal-dual residual \( r \) decreases linearly with \( \alpha \):
  \[ r(v_k(\alpha)) = (1 - \alpha)r(v_k) \]

- shifted complementarity behaves predictably:
  \[ X_k(\alpha)s_k(\alpha) - \mu e = (1 - \alpha)(X_k s_k - \mu e) + \alpha^2 \dot{X}_k \dot{s}_k \]
  \[ \Rightarrow \]
  \[ \Phi(v_k(\alpha)) \leq (1 - \alpha)\Phi(v_k) + \alpha^2 \|\dot{X}_k \dot{s}_k\| \quad \forall \alpha \in [0, 1] \]

Algorithm

Given target \( \mu \), initial point \( v_0 : (x_0, s_0) > 0 \), constant \( \omega \leq \min x_{i0}s_{i0}/\mu \), stopping tolerance \( \epsilon > 0 \) and \( k = 0 \)

- if \( \Phi(v_k) \leq \epsilon \), stop
- compute the Newton correction \( \dot{v}_k \)
- compute \( \alpha_k^L \in (0, 1] : X_k(\alpha)s_k(\alpha) \geq \omega \mu e \quad \forall \alpha \in [0, \alpha_k^L] \)
- Find \( \alpha_k \): global minimizer \( \Phi(v_k(\alpha)) \) in \( [0, \alpha_k^L] \)
- \( v_{k+1} = v_k + \alpha_k \dot{v}_k \)
- \( k \leftarrow k + 1 \)

Dominant cost / iteration: Newton correction
Convergence analysis

Require additionally that $(x_0, s_0) \geq (u_0, w_0)$ for some $(u_0, t_0, w_0)$: 
\[ Au_0 = b \text{ and } A^T t_0 + w_0 = c \] (Zhang)

Results:

$(x_k, s_k)$ bounded above and away from zero

$\implies$ (Newton system)

$\| \dot{x}_k \|, \| \dot{s}_k \| \leq \kappa_1 \Phi(v_k)$ for some constant $\kappa_1 > 0$

$\implies$ $(X_k(\alpha)s_k(\alpha) - \omega \mu e \geq \alpha(1 - \omega)\mu e - \alpha^2 \| \dot{X}_k \dot{s}_k \|)$

$\alpha_{L_k}^1 \geq \min \left\{ 1, \frac{\kappa_2}{\left[ \Phi(v_k) \right]^2} \right\}$ for some constant $\kappa_2 > 0$ …

Convergence analysis II

$\Phi(v_k(\alpha)) \leq (1 - \alpha)\Phi(v_k) + \alpha^2 \| \dot{X}_k \dot{s}_k \| \forall \alpha \in [0, 1]$

$\implies \Phi(v_{k+1}) \leq \min_{\alpha \in [0, \alpha_{L_k}^1]} (1 - \alpha)\Phi(v_k) + \alpha^2 \| \dot{X}_k \dot{s}_k \|$

$\implies \Phi(v_{k+1}) \leq \begin{cases} 
(1 - \frac{1}{2} \alpha_{L_k}^1)\Phi(v_k) & \text{if } \alpha_{L_k}^1 \leq \alpha_k \equiv \frac{1}{2} \Phi(v_k)/\| \dot{X}_k \dot{s}_k \| \\
(1 - \frac{1}{2} \alpha_k)\Phi(v_k) & \text{otherwise}
\end{cases}$

$\implies \Phi(v_{k+1}) \leq \kappa_3 \Phi(v_k)$ for some $\kappa_3 \in [0, 1)$

$\implies$ global linear convergence

Also $\alpha_{L_k}^1 \geq \min \left\{ 1, \frac{\kappa_2}{\left[ \Phi(v_k) \right]^2} \right\} \to 1$ and $\alpha_k \geq \frac{1}{2\kappa_1^2 \Phi(v_k)} \to \infty$

$\implies$ asymptotic superlinear convergence
Convergence analysis III

If \((x_0, s_0) > 0\) such that \(Ax_0 = b\) and \(A^Ty_0 + s_0 = c\) \(\implies\) at most

\[
\frac{1}{q} \log \left( \frac{\Phi(v_0)}{\epsilon} \right)
\]

iterations to find \(\Phi(v_k) \leq \epsilon\), where

\[
q = \min \left( \frac{\omega \mu}{2 \Phi(v_0)}, \frac{(1 - \omega) \omega \mu^2}{\Phi(v_0)^2} \right)
\]

depends only on \(v_0, \mu\) and \(\omega\)

\(\implies\) polynomial complexity

Good numerical experience ...

<table>
<thead>
<tr>
<th>Iter</th>
<th>p-feas</th>
<th>d-feas</th>
<th>com-slk</th>
<th>merit</th>
<th>step</th>
<th>mu</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3.9E+01</td>
<td>1.9E+01</td>
<td>1.6E+01</td>
<td>1.9E+03</td>
<td>-</td>
<td>1.0E+00</td>
<td>0.00</td>
</tr>
<tr>
<td>1r</td>
<td>3.5E+01</td>
<td>1.7E+01</td>
<td>3.1E+01</td>
<td>1.7E+03</td>
<td>1.0E-01</td>
<td>1.0E+00</td>
<td>0.00</td>
</tr>
<tr>
<td>2r</td>
<td>2.9E+01</td>
<td>1.4E+01</td>
<td>2.6E+01</td>
<td>1.4E+03</td>
<td>1.6E-01</td>
<td>1.0E+00</td>
<td>0.00</td>
</tr>
<tr>
<td>3r</td>
<td>2.1E+01</td>
<td>1.0E+01</td>
<td>1.8E+01</td>
<td>9.9E+02</td>
<td>3.0E-01</td>
<td>1.0E+00</td>
<td>0.00</td>
</tr>
<tr>
<td>4r</td>
<td>1.1E+01</td>
<td>5.2E+00</td>
<td>9.3E+00</td>
<td>5.2E+02</td>
<td>4.8E-01</td>
<td>1.0E+00</td>
<td>0.00</td>
</tr>
<tr>
<td>5r</td>
<td>1.4E-15</td>
<td>1.7E-14</td>
<td>9.8E-02</td>
<td>1.4E+01</td>
<td>1.0E+00</td>
<td>1.0E+00</td>
<td>0.00</td>
</tr>
<tr>
<td>6r</td>
<td>1.8E-15</td>
<td>6.8E-15</td>
<td>4.8E-04</td>
<td>3.5E+00</td>
<td>7.5E-01</td>
<td>1.0E+00</td>
<td>0.00</td>
</tr>
<tr>
<td>7r</td>
<td>1.8E-15</td>
<td>1.5E-15</td>
<td>2.1E-04</td>
<td>1.7E-01</td>
<td>1.0E-00</td>
<td>1.0E+00</td>
<td>0.00</td>
</tr>
<tr>
<td>8r</td>
<td>3.6E-15</td>
<td>0.0E+00</td>
<td>5.5E-07</td>
<td>7.6E-04</td>
<td>1.0E+00</td>
<td>1.0E+00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

====================== feasible point found =================

CERFACS Sparse Days meeting, 15th June 2006 – p. 15/22
...and bad numerical experience

<table>
<thead>
<tr>
<th>Iter</th>
<th>p-feas</th>
<th>d-feas</th>
<th>com-slk</th>
<th>merit</th>
<th>step</th>
<th>mu</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.1E+03</td>
<td>4.1E+01</td>
<td>1.2E+02</td>
<td>7.8E+03</td>
<td>-</td>
<td>1.0E+00</td>
<td>0.00</td>
</tr>
<tr>
<td>1r</td>
<td>2.1E+03</td>
<td>4.1E+01</td>
<td>1.9E+01</td>
<td>7.7E+03</td>
<td>1.0E-02</td>
<td>1.0E+00</td>
<td>0.00</td>
</tr>
<tr>
<td>377r</td>
<td>2.1E-06</td>
<td>4.1E-08</td>
<td>7.9E-05</td>
<td>1.1E-02</td>
<td>5.5E-03</td>
<td>1.0E+00</td>
<td>0.44</td>
</tr>
<tr>
<td>378r</td>
<td>2.1E-06</td>
<td>4.1E-08</td>
<td>7.9E-05</td>
<td>1.1E-02</td>
<td>5.4E-03</td>
<td>1.0E+00</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Why? No strict interior! $x_i \to 0$ and $s_i \approx \mu / x_i \to \infty \implies \alpha_k \not\to 1$

Controlled perturbations to cope with these difficulties

For $k = 1, 2, \ldots$,  
- pick $(\theta^x_k, \theta^s_k) \geq 0$:  
  
  \[
  \begin{align*}
  Ax - b &= 0 \\
  A^T y + s - c &= 0
  \end{align*}
  \]
  
  and  
  \[
  (X + \Theta^x_k)(S + \Theta^s_k)e = \mu e
  \]

  has an interior solution $(x_k, y_k, s_k)$
  - use previous algorithm to find $(x_k, y_k, s_k)$

- reduce $(\theta^x_k, \theta^s_k) \to (\theta^x_{k+1}, \theta^s_{k+1})$ via  
  $(\zeta \in (0, 1))$

  \[
  \begin{align*}
  (\theta^x_{k+1})_i &= \begin{cases} 
  0 & \text{if } (x_k)_i > 0 \\
  (1 - \zeta)(\theta^x_k)_i - \zeta(x_k)_i & \text{if } (x_k)_i \leq 0
  \end{cases} \\
  (\theta^s_{k+1})_i &= \begin{cases} 
  0 & \text{if } (s_k)_i > 0 \\
  (1 - \zeta)(\theta^s_k)_i - \zeta(s_k)_i & \text{if } (s_k)_i \leq 0
  \end{cases}
  \end{align*}
  \]

  $\implies (x_k + \theta^x_{k+1}, s_k + \theta^s_{k+1}) > 0$
Details

- equivalent to enlarging and then systematically contracting the primal-dual feasible region
- initial \((\theta^x_1, \theta^s_1)\)?
  - compute any \((x_0, y_0, s_0)\):
    \[
    Ax_0 = b \text{ and } A^Ty_0 + s_0 = c
    \]
  - compute sufficiently large \((\theta^x_1, \theta^s_1) \geq 0\):
    \[
    (x_0 + \theta^x_1, s_0 + \theta^s_1) > 0
    \]
  \[\implies\text{ all iterates satisfy}\]
  \[
  Ax_k = b \text{ and } A^Ty_k + s_k = c
  \]
  \[\implies\text{ polynomial complexity of each inner iteration}\]

Theoretical results

- if there is a strictly feasible point, there will be a finite number \(\sim 1/\mu\) of contractions before \((\theta^x_k, \theta^s_k) = 0\)
- if \(\lim_{k \to \infty} \max_i (\theta^x_k)_i > 0\), there is no primal feasible point
- if \(\lim_{k \to \infty} \max_i (\theta^s_k)_i > 0\), there is no dual feasible point
- if there is a primal feasible point, any index \(i\) for which
  \[
  \lim_{k \to \infty} (x_k)_i \to 0 \quad (\equiv \lim_{k \to \infty} (s_k)_i \to \infty)
  \]
  \(\iff\) \(x_i\) is always zero — implicit primal equality
  - remove variable from problem
- if there is a dual feasible point, any index \(i\) for which
  \[
  \lim_{k \to \infty} (s_k)_i \to 0 \quad (\equiv \lim_{k \to \infty} (x_k)_i \to \infty)
  \]
  \(\iff\) \(s_i\) is always zero — implicit dual equality
  - remove bound on variable
### Numerical experience — Netlib LP collection

<table>
<thead>
<tr>
<th>problem</th>
<th>n</th>
<th>m</th>
<th>x\text{imp}</th>
<th>s\text{imp}</th>
<th>c\text{imp}</th>
<th>y\text{imp}</th>
</tr>
</thead>
<tbody>
<tr>
<td>80BAU3B</td>
<td>9799</td>
<td>2262</td>
<td>10</td>
<td>6</td>
<td>38</td>
<td>4</td>
</tr>
<tr>
<td>ADLITTLE</td>
<td>97</td>
<td>56</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AGG2</td>
<td>302</td>
<td>516</td>
<td>2</td>
<td>58</td>
<td>1</td>
<td>44</td>
</tr>
<tr>
<td>BCDOUT</td>
<td>5940</td>
<td>5414</td>
<td>64</td>
<td>0</td>
<td>543</td>
<td>0</td>
</tr>
<tr>
<td>FIT2P</td>
<td>13525</td>
<td>3000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GREENBEA</td>
<td>5405</td>
<td>2392</td>
<td>699</td>
<td>84</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>SCRS8</td>
<td>1169</td>
<td>490</td>
<td>32</td>
<td>14</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>STOCFOR3</td>
<td>15695</td>
<td>16675</td>
<td>212</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WOODW</td>
<td>8405</td>
<td>1098</td>
<td>2511</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\(n = \# \text{variables}, \ m = \# \text{constraints}\)
\(x\text{imp} = \# \text{implied fixed variables}, \ s\text{imp} = \# \text{implied free variables}\)
\(c\text{imp} = \# \text{implied fixed constraints}, \ y\text{imp} = \# \text{implied free constraints}\)

### Conclusions

- Simple method for finding a well-centered feasible point of strictly feasible region
- Suitable for large-scale computation
- Globally linearly and locally superlinearly convergent
- Polynomial complexity
- Controlled perturbations allow us to identify less favourable outcomes
- What is a good target value \(\mu\)?
- Easily extensible for more general problems (convex QP, horizontal complementarity)
- Software package WCP coming as part of GALAHAD 2.0