15.081J/6.251J Introduction to Mathematical Programming

Lecture 22: Primal-dual Barrier Interior Point Algorithm

1 Outline

Slide 1

1. The Barrier Problem
2. Solving Equations
3. The Primal-Dual Barrier Algorithm
4. Insight on Behavior
5. Computational Aspects
6. Conclusions

2 The Barrier Problem

Slide 2

Barrier problem:

$$\begin{aligned} \min \quad & B_\mu(\boldsymbol{x}) = \boldsymbol{c}'\boldsymbol{x} - \mu \sum_{j=1}^{n} \log x_j \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

KKT:

$$\boldsymbol{c} - \mu \left(\frac{1}{x_1(\mu)}, \ldots, \frac{1}{x_n(\mu)} \right)' + \boldsymbol{A}'\boldsymbol{p}(\mu) = \boldsymbol{0}$$

$$\boldsymbol{A}\boldsymbol{x}(\mu) = \boldsymbol{b}, \qquad \boldsymbol{x}(\mu) \geq 0$$

2.1 Optimality Conditions

Slide 3

Set $s_j(\mu) = \dfrac{\mu}{x_j(\mu)}$

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x}(\mu) &= \boldsymbol{b} \\ \boldsymbol{x}(\mu) &\geq \boldsymbol{0} \\ \boldsymbol{A}'\boldsymbol{p}(\mu) + \boldsymbol{s}(\mu) &= \boldsymbol{c} \\ \boldsymbol{s}(\mu) &\geq \boldsymbol{0} \\ s_j(\mu)x_j(\mu) &= \mu \quad \text{or} \\ \boldsymbol{X}(\mu)\boldsymbol{S}(\mu)\boldsymbol{e} &= \boldsymbol{e}\mu \end{aligned}$$

$\boldsymbol{X}(\mu) = \text{diag}\big(x_1(\mu), \ldots, x_n(\mu)\big)$, $\boldsymbol{S}(\mu) = \text{diag}\big(s_1(\mu), \ldots, s_n(\mu)\big)$

3 Solving Equations

SLIDE 4

$$\boldsymbol{F}(\boldsymbol{z}) = \begin{bmatrix} \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \\ \boldsymbol{A}'\boldsymbol{p} + \boldsymbol{s} - \boldsymbol{c} \\ \boldsymbol{X}\boldsymbol{S}\boldsymbol{e} - \mu\boldsymbol{e} \end{bmatrix}$$

$\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{s})$, $r = 2n + m$
Solve

$$\boldsymbol{F}(\boldsymbol{z}^*) = \boldsymbol{0}$$

3.1 Newton's method

SLIDE 5

$$\boldsymbol{F}(\boldsymbol{z}^k + \boldsymbol{d}) \approx \boldsymbol{F}(\boldsymbol{z}^k) + \boldsymbol{J}(\boldsymbol{z}^k)\boldsymbol{d}$$

Here $\boldsymbol{J}(\boldsymbol{z}^k)$ is the $r \times r$ Jacobian matrix whose (i, j)th element is given by

$$\left.\frac{\partial F_i(\boldsymbol{z})}{\partial z_j}\right|_{\boldsymbol{z}=\boldsymbol{z}^k}$$

$$\boldsymbol{F}(\boldsymbol{z}^k) + \boldsymbol{J}(\boldsymbol{z}^k)\boldsymbol{d} = \boldsymbol{0}$$

SLIDE 6

Set $\boldsymbol{z}^{k+1} = \boldsymbol{z}^k + \boldsymbol{d}$ ($\boldsymbol{d}$ is the *Newton direction*)
$(\boldsymbol{x}^k, \boldsymbol{p}^k, \boldsymbol{s}^k)$ current primal and dual feasible solution
Newton direction $\boldsymbol{d} = (\boldsymbol{d}_x^k, \boldsymbol{d}_p^k, \boldsymbol{d}_s^k)$

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}' & \boldsymbol{I} \\ \boldsymbol{S}_k & \boldsymbol{0} & \boldsymbol{X}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{d}_x^k \\ \boldsymbol{d}_p^k \\ \boldsymbol{d}_s^k \end{bmatrix} = - \begin{bmatrix} \boldsymbol{A}\boldsymbol{x}^k - \boldsymbol{b} \\ \boldsymbol{A}'\boldsymbol{p}^k + \boldsymbol{s}^k - \boldsymbol{c} \\ \boldsymbol{X}_k\boldsymbol{S}_k\boldsymbol{e} - \mu^k\boldsymbol{e} \end{bmatrix}$$

3.2 Step lengths

SLIDE 7

$$\begin{aligned} \boldsymbol{x}^{k+1} &= \boldsymbol{x}^k + \beta_P^k \boldsymbol{d}_x^k \\ \boldsymbol{p}^{k+1} &= \boldsymbol{p}^k + \beta_D^k \boldsymbol{d}_p^k \\ \boldsymbol{s}^{k+1} &= \boldsymbol{s}^k + \beta_D^k \boldsymbol{d}_s^k \end{aligned}$$

To preserve nonnegativity, take

$$\begin{aligned} \beta_P^k &= \min\left\{1, \alpha \min_{\{i|(d_x^k)_i<0\}} \left(-\frac{x_i^k}{(d_x^k)_i}\right)\right\}, \\ \beta_D^k &= \min\left\{1, \alpha \min_{\{i|(d_s^k)_i<0\}} \left(-\frac{s_i^k}{(d_s^k)_i}\right)\right\}, \end{aligned}$$

$0 < \alpha < 1$

4 The Primal-Dual Barrier Algorithm

SLIDE 8

1. (Initialization) Start with $\boldsymbol{x}^0 > \boldsymbol{0},\ \boldsymbol{s}^0 > \boldsymbol{0},\ \boldsymbol{p}^0$, and set $k = 0$

2. (Optimality test) If $(\boldsymbol{s}^k)'\boldsymbol{x}^k < \epsilon$ stop; else go to Step 3.

3. (Computation of Newton directions)

$$\begin{aligned} \mu^k &= \frac{(\boldsymbol{s}^k)'\boldsymbol{x}^k}{n} \\ \boldsymbol{X}_k &= \mathrm{diag}(x_1^k, \ldots, x_n^k) \\ \boldsymbol{S}_k &= \mathrm{diag}(s_1^k, \ldots, s_n^k) \end{aligned}$$

Solve linear system

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}' & \boldsymbol{I} \\ \boldsymbol{S}_k & \boldsymbol{0} & \boldsymbol{X}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{d}_x^k \\ \boldsymbol{d}_p^k \\ \boldsymbol{d}_s^k \end{bmatrix} = - \begin{bmatrix} \boldsymbol{A}\boldsymbol{x}^k - \boldsymbol{b} \\ \boldsymbol{A}'\boldsymbol{p}^k + \boldsymbol{s}^k - \boldsymbol{c} \\ \boldsymbol{X}_k\boldsymbol{S}_k\boldsymbol{e} - \mu^k\boldsymbol{e} \end{bmatrix}$$

SLIDE 9

4. (Find step lengths)

$$\begin{aligned} \beta_P^k &= \min\left\{1, \alpha \min_{\{i \mid (d_x^k)_i < 0\}} \left(- \frac{x_i^k}{(d_x^k)_i} \right) \right\} \\ \beta_D^k &= \min\left\{1, \alpha \min_{\{i \mid (d_s^k)_i < 0\}} \left(- \frac{s_i^k}{(d_s^k)_i} \right) \right\} \end{aligned}$$

5. (Solution update)

$$\begin{aligned} \boldsymbol{x}^{k+1} &= \boldsymbol{x}^k + \beta_P^k \boldsymbol{d}_x^k \\ \boldsymbol{p}^{k+1} &= \boldsymbol{p}^k + \beta_D^k \boldsymbol{d}_p^k \\ \boldsymbol{s}^{k+1} &= \boldsymbol{s}^k + \beta_D^k \boldsymbol{d}_s^k \end{aligned}$$

6. Let $k := k + 1$ and go to Step 2

5 Insight on behavior

SLIDE 10

- Affine Scaling

$$\boldsymbol{d}_{\mathrm{affine}} = -\boldsymbol{X}^2\Big(\boldsymbol{I} - \boldsymbol{A}'(\boldsymbol{A}\boldsymbol{X}^2\boldsymbol{A}')^{-1}\boldsymbol{A}\boldsymbol{X}^2\Big)\boldsymbol{c}$$

- Primal barrier

$$\boldsymbol{d}_{\mathrm{primal-barrier}} = \Big(\boldsymbol{I} - \boldsymbol{X}^2\boldsymbol{A}'(\boldsymbol{A}\boldsymbol{X}^2\boldsymbol{A}')^{-1}\boldsymbol{A}\Big)\left(\boldsymbol{X}\boldsymbol{e} - \frac{1}{\mu}\boldsymbol{X}^2\boldsymbol{c}\right)$$

- For $\mu = \infty$

$$\boldsymbol{d}_{\text{centering}} = \Big(\boldsymbol{I} - \boldsymbol{X}^2\boldsymbol{A}'(\boldsymbol{A}\boldsymbol{X}^2\boldsymbol{A}')^{-1}\boldsymbol{A}\Big)\boldsymbol{X}\boldsymbol{e}$$

- Note that

$$\boldsymbol{d}_{\text{primal-barrier}} = \boldsymbol{d}_{\text{centering}} + \frac{1}{\mu}\boldsymbol{d}_{\text{affine}}$$

- When μ is large, then the centering direction dominates, i.e., in the beginning, the barrier algorithm takes steps towards the analytic center
- When μ is small, then the affine scaling direction dominates, i.e., towards the end, the barrier algorithm behaves like the affine scaling algorithm

6 Computational aspects of IPMs

SLIDE 11

Simplex vs. Interior point methods (IPMs)

- Simplex method tends to perform poorly on large, massively degenerate problems, whereas IP methods are much less affected.
- Key step in IPMs

$$\big(\boldsymbol{A}\boldsymbol{X}_k^2\boldsymbol{A}'\big)\boldsymbol{d} = \boldsymbol{f}$$

- In implementations of IPMs $\boldsymbol{A}\boldsymbol{X}_k^2\boldsymbol{A}'$ is usually written as

$$\boldsymbol{A}\boldsymbol{X}_k^2\boldsymbol{A}' = \boldsymbol{L}\boldsymbol{L}',$$

  where $\boldsymbol{L}$ is a square lower triangular matrix called the *Cholesky factor*

- Solve system

$$\big(\boldsymbol{A}\boldsymbol{X}_k^2\boldsymbol{A}'\big)\boldsymbol{d} = \boldsymbol{f}$$

  by solving the triangular systems

$$\boldsymbol{L}\boldsymbol{y} = \boldsymbol{f}, \qquad \boldsymbol{L}'\boldsymbol{d} = \boldsymbol{y}$$

- The construction of $\boldsymbol{L}$ requires $O(n^3)$ operations; but the actual computational effort is highly dependent on the sparsity (number of nonzero entries) of $\boldsymbol{L}$
- Large scale implementations employ heuristics (reorder rows and columns of $\boldsymbol{A}$) to improve sparsity of $\boldsymbol{L}$. If $\boldsymbol{L}$ is sparse, IPMs are stronger.

7 Conclusions

SLIDE 12

- IPMs represent the present and future of Optimization.
- Very successful in solving very large problems.
- Extend to general convex problems

6.251J / 15.081J Introduction to Mathematical Programming
Fall 2009