15.093 Optimization Methods

Lecture 21: The Affine Scaling Algorithm

1 Outline

Slide 1

- History
- Geometric intuition
- Algebraic development
- Affine Scaling
- Convergence
- Initialization
- Practical performance

2 History

Slide 2

- In 1984, Karmakar at AT&T "invented" interior point method
- In 1985, Affine scaling "invented" at IBM + AT&T seeking intuitive version of Karmarkar's algorithm
- In early computational tests, A.S. far outperformed simplex and Karmarkar's algorithm
- In 1989, it was realised Dikin invented A.S. in 1967

3 Geometric intuition

3.1 Notation

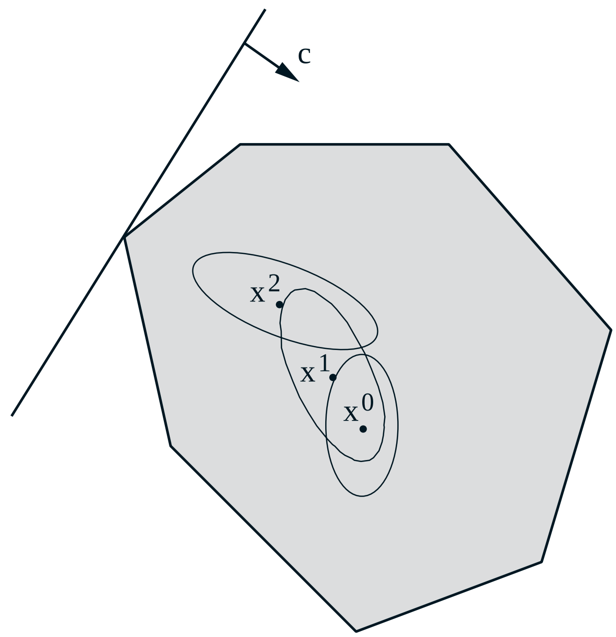
Slide 3

$$\begin{array}{rl} \min & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{array}$$

and its dual

$$\begin{array}{rl} \max & \boldsymbol{p}'\boldsymbol{b} \\ \text{s.t.} & \boldsymbol{p}'\boldsymbol{A} \leq \boldsymbol{c}' \end{array}$$

- $P = \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geq \boldsymbol{0}\}$
- $\{\boldsymbol{x} \in P \mid \boldsymbol{x} > \boldsymbol{0}\}$ the *interior* of P and its elements *interior points*



3.2 The idea

Slide 4

4 Algebraic development

4.1 Theorem

Slide 5

$\beta \in (0,1)$, $\boldsymbol{y} \in \Re^n$: $\boldsymbol{y} > \boldsymbol{0}$, and

$$S = \left\{ \boldsymbol{x} \in \Re^n \;\Big|\; \sum_{i=1}^{n} \frac{(x_i - y_i)^2}{y_i^2} \le \beta^2 \right\}.$$

Then, $\boldsymbol{x} > \boldsymbol{0}$ for every $\boldsymbol{x} \in S$
Proof

- $\boldsymbol{x} \in S$
- $(x_i - y_i)^2 \le \beta^2 y_i^2 < y_i^2$
- $|x_i - y_i| < y_i$; $-x_i + y_i < y_i$, and hence $x_i > 0$

Slide 6

$$\boldsymbol{x} \in S \text{ is equivalent to } \big|\big|\boldsymbol{Y}^{-1}(\boldsymbol{x} - \boldsymbol{y})\big|\big| \le \beta$$

Replace original LP:

$$\begin{array}{ll} \min & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \big|\big|\boldsymbol{Y}^{-1}(\boldsymbol{x} - \boldsymbol{y})\big|\big| \le \beta. \end{array}$$

$\boldsymbol{d} = \boldsymbol{x} - \boldsymbol{y}$

$$\begin{array}{ll} \min & \boldsymbol{c}'\boldsymbol{d} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{d} = \boldsymbol{0} \\ & ||\boldsymbol{Y}^{-1}\boldsymbol{d}|| \leq \beta \end{array}$$

4.2 Solution

SLIDE 7

If rows of $\boldsymbol{A}$ are linearly independent and $\boldsymbol{c}$ is not a linear combination of the rows of $\boldsymbol{A}$, then

- optimal solution $\boldsymbol{d}^*$:

$$\boldsymbol{d}^* = -\beta \frac{\boldsymbol{Y}^2(\boldsymbol{c} - \boldsymbol{A}'\boldsymbol{p})}{\big|\big|\boldsymbol{Y}(\boldsymbol{c} - \boldsymbol{A}'\boldsymbol{p})\big|\big|}, \quad \boldsymbol{p} = (\boldsymbol{A}\boldsymbol{Y}^2\boldsymbol{A}')^{-1}\boldsymbol{A}\boldsymbol{Y}^2\boldsymbol{c}.$$

- $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{d}^* \in P$
- $\boldsymbol{c}'\boldsymbol{x} = \boldsymbol{c}'\boldsymbol{y} - \beta\big|\big|\boldsymbol{Y}(\boldsymbol{c} - \boldsymbol{A}'\boldsymbol{p})\big|\big| < \boldsymbol{c}'\boldsymbol{y}$

4.2.1 Proof

SLIDE 8

- $\boldsymbol{A}\boldsymbol{Y}^2\boldsymbol{A}'$ is invertible;if not, there exists some $\boldsymbol{z} \neq \boldsymbol{0}$ such that $\boldsymbol{z}'\boldsymbol{A}\boldsymbol{Y}^2\boldsymbol{A}'\boldsymbol{z} = 0$
- $\boldsymbol{w} = \boldsymbol{Y}\boldsymbol{A}'\boldsymbol{z}$; $\boldsymbol{w}'\boldsymbol{w} = 0 \;\Rightarrow\; \boldsymbol{w} = \boldsymbol{0}$
- Hence $\boldsymbol{A}'\boldsymbol{z} = \boldsymbol{0}$ contradiction
- Since $\boldsymbol{c}$ is not a linear combination of the rows of $\boldsymbol{A}$, $\boldsymbol{c} - \boldsymbol{A}'\boldsymbol{p} \neq \boldsymbol{0}$ and $\boldsymbol{d}^*$ is well defined
- $\boldsymbol{d}^*$ feasible

$$\boldsymbol{Y}^{-1}\boldsymbol{d}^* = -\beta \frac{\boldsymbol{Y}(\boldsymbol{c} - \boldsymbol{A}'\boldsymbol{p})}{\big|\big|\boldsymbol{Y}(\boldsymbol{c} - \boldsymbol{A}'\boldsymbol{p})\big|\big|} \Rightarrow ||\boldsymbol{Y}^{-1}\boldsymbol{d}^*|| = \beta$$

  $\boldsymbol{A}\boldsymbol{d}^* = \boldsymbol{0}$, since $\boldsymbol{A}\boldsymbol{Y}^2(\boldsymbol{c} - \boldsymbol{A}'\boldsymbol{p}) = \boldsymbol{0}$

-

$$\begin{aligned} \boldsymbol{c}'\boldsymbol{d} &= (\boldsymbol{c}' - \boldsymbol{p}'\boldsymbol{A})\boldsymbol{d} \\ &= (\boldsymbol{c}' - \boldsymbol{p}'\boldsymbol{A})\boldsymbol{Y}\boldsymbol{Y}^{-1}\boldsymbol{d} \\ &\geq -\big|\big|\boldsymbol{Y}(\boldsymbol{c} - \boldsymbol{A}'\boldsymbol{p})\big|\big| \cdot ||\boldsymbol{Y}^{-1}\boldsymbol{d}|| \\ &\geq -\beta\big|\big|\boldsymbol{Y}(\boldsymbol{c} - \boldsymbol{A}'\boldsymbol{p})\big|\big|. \end{aligned}$$

SLIDE 9

-

$$\begin{aligned} \boldsymbol{c}'\boldsymbol{d}^* &= (\boldsymbol{c}' - \boldsymbol{p}'\boldsymbol{A})\boldsymbol{d}^* \\ &= -(\boldsymbol{c}' - \boldsymbol{p}'\boldsymbol{A})\beta \frac{\boldsymbol{Y}^2(\boldsymbol{c} - \boldsymbol{A}'\boldsymbol{p})}{\big|\big|\boldsymbol{Y}(\boldsymbol{c} - \boldsymbol{A}'\boldsymbol{p})\big|\big|} \\ &= -\beta \frac{\big(\boldsymbol{Y}(\boldsymbol{c} - \boldsymbol{A}'\boldsymbol{p})\big)'\big(\boldsymbol{Y}(\boldsymbol{c} - \boldsymbol{A}'\boldsymbol{p})\big)}{\big|\big|\boldsymbol{Y}(\boldsymbol{c} - \boldsymbol{A}'\boldsymbol{p})\big|\big|} \\ &= -\beta\big|\big|\boldsymbol{Y}(\boldsymbol{c} - \boldsymbol{A}'\boldsymbol{p})\big|\big|. \end{aligned}$$

- $\boldsymbol{c}'\boldsymbol{x} = \boldsymbol{c}'\boldsymbol{y} + \boldsymbol{c}'\boldsymbol{d}^* = \boldsymbol{c}'\boldsymbol{y} - \beta\big|\big|\boldsymbol{Y}(\boldsymbol{c} - \boldsymbol{A}'\boldsymbol{p})\big|\big|$

4.3 Interpretation

SLIDE 10

- $\boldsymbol{y}$ be a nondegenerate BFS with basis $\boldsymbol{B}$
- $\boldsymbol{A} = [\boldsymbol{B} \;\; \boldsymbol{N}]$
- $\boldsymbol{Y} = \text{diag}(y_1, \ldots, y_m, 0, \ldots, 0)$ and $\boldsymbol{Y}_0 = \text{diag}(y_1, \ldots, y_m)$, then $\boldsymbol{AY} = [\boldsymbol{BY}_0 \;\; \boldsymbol{0}]$

$$\begin{aligned} \boldsymbol{p} &= (\boldsymbol{AY}^2\boldsymbol{A}')^{-1}\boldsymbol{AY}^2\boldsymbol{c} \\ &= (\boldsymbol{B}')^{-1}\boldsymbol{Y}_0^{-2}\boldsymbol{B}^{-1}\boldsymbol{BY}_0^2\boldsymbol{c}_B \\ &= (\boldsymbol{B}')^{-1}\boldsymbol{c}_B \end{aligned}$$

- Vectors $\boldsymbol{p}$ dual estimates
- $\boldsymbol{r} = \boldsymbol{c} - \boldsymbol{A}'\boldsymbol{p}$ becomes reduced costs:

$$\boldsymbol{r} = \boldsymbol{c} - \boldsymbol{A}'(\boldsymbol{B}')^{-1}\boldsymbol{c}_B$$

- Under degeneracy?

4.4 Termination

SLIDE 11

$\boldsymbol{y}$ and $\boldsymbol{p}$ be primal and dual feasible solutions with

$$\boldsymbol{c}'\boldsymbol{y} - \boldsymbol{b}'\boldsymbol{p} < \epsilon$$

$\boldsymbol{y}^*$ and $\boldsymbol{p}^*$ be optimal primal and dual solutions. Then,

$$\begin{aligned} \boldsymbol{c}'\boldsymbol{y}^* &\le \boldsymbol{c}'\boldsymbol{y} < \boldsymbol{c}'\boldsymbol{y}^* + \epsilon, \\ \boldsymbol{b}'\boldsymbol{p}^* - \epsilon &< \boldsymbol{b}'\boldsymbol{p} \le \boldsymbol{b}'\boldsymbol{p}^* \end{aligned}$$

4.4.1 Proof

SLIDE 12

- $\boldsymbol{c}'\boldsymbol{y}^* \le \boldsymbol{c}'\boldsymbol{y}$
- By weak duality, $\boldsymbol{b}'\boldsymbol{p} \le \boldsymbol{c}'\boldsymbol{y}^*$
- Since $\boldsymbol{c}'\boldsymbol{y} - \boldsymbol{b}'\boldsymbol{p} < \epsilon$,

$$\boldsymbol{c}'\boldsymbol{y} < \boldsymbol{b}'\boldsymbol{p} + \epsilon \le \boldsymbol{c}'\boldsymbol{y}^* + \epsilon$$

$$\boldsymbol{b}'\boldsymbol{p}^* = \boldsymbol{c}'\boldsymbol{y}^* \le \boldsymbol{c}'\boldsymbol{y} < \boldsymbol{b}'\boldsymbol{p} + \epsilon$$

5 Affine Scaling

5.1 Inputs

SLIDE 13

- $(\boldsymbol{A},\ \boldsymbol{b},\ \boldsymbol{c})$;
- an initial primal feasible solution $\boldsymbol{x}^0 > \boldsymbol{0}$
- the optimality tolerance $\epsilon > 0$
- the parameter $\beta \in (0,1)$

5.2 The Algorithm

SLIDE 14

1. (Initialization) Start with some feasible $\boldsymbol{x}^0 > \boldsymbol{0}$; let $k = 0$.
2. (Computation of dual estimates and reduced costs) Given some feasible $\boldsymbol{x}^k > \boldsymbol{0}$, let

$$\begin{aligned}\boldsymbol{X}_k &= \operatorname{diag}(x_1^k, \ldots, x_n^k),\\ \boldsymbol{p}^k &= (\boldsymbol{A}\boldsymbol{X}_k^2\boldsymbol{A}')^{-1}\boldsymbol{A}\boldsymbol{X}_k^2\boldsymbol{c},\\ \boldsymbol{r}^k &= \boldsymbol{c} - \boldsymbol{A}'\boldsymbol{p}^k.\end{aligned}$$

3. (Optimality check) Let $\boldsymbol{e} = (1,1,\ldots,1)$. If $\boldsymbol{r}^k \geq \boldsymbol{0}$ and $\boldsymbol{e}'\boldsymbol{X}_k\boldsymbol{r}^k < \epsilon$, then stop; the current solution $\boldsymbol{x}^k$ is primal ϵ-optimal and $\boldsymbol{p}^k$ is dual ϵ-optimal.
4. (Unboundedness check) If $-\boldsymbol{X}_k^2\boldsymbol{r}^k \geq \boldsymbol{0}$ then stop; the optimal cost is $-\infty$.
5. (Update of primal solution) Let

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \beta\frac{\boldsymbol{X}_k^2\boldsymbol{r}^k}{||\boldsymbol{X}_k\boldsymbol{r}^k||}.$$

5.3 Variants

SLIDE 15

- $||\boldsymbol{u}||_\infty = \max_i |u_i|, \quad \gamma(\boldsymbol{u}) = \max\{u_i \mid u_i > 0\}$
- $\gamma(\boldsymbol{u}) \leq ||\boldsymbol{u}||_\infty \leq ||\boldsymbol{u}||$
- *Short-step* method.
- *Long-step* variants

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \beta\frac{\boldsymbol{X}_k^2\boldsymbol{r}^k}{||\boldsymbol{X}_k\boldsymbol{r}^k||_\infty}$$

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \beta\frac{\boldsymbol{X}_k^2\boldsymbol{r}^k}{\gamma(\boldsymbol{X}_k\boldsymbol{r}^k)}$$

6 Convergence

6.1 Assumptions

SLIDE 16

Assumptions A:

(a) The rows of the matrix $\boldsymbol{A}$ are linearly independent.

(b) The vector $\boldsymbol{c}$ is not a linear combination of the rows of $\boldsymbol{A}$.

(c) There exists an optimal solution.

(d) There exists a positive feasible solution.

Assumptions B:

(a) Every BFS to the primal problem is nondegenerate.

(b) At every BFS to the primal problem, the reduced cost of every nonbasic variable is nonzero.

6.2 Theorem

SLIDE 17

If we apply the long-step affine scaling algorithm with $\epsilon = 0$, the following hold:

(a) For the Long-step variant and under Assumptions A and B, and if $0 < \beta < 1$, $\boldsymbol{x}^k$ and $\boldsymbol{p}^k$ converge to the optimal primal and dual solutions

(b) For the second Long-step variant, and under Assumption A and if $0 < \beta < 2/3$, the sequences $\boldsymbol{x}^k$ and $\boldsymbol{p}^k$ converge to some primal and dual optimal solutions, respectively

7 Initialization

SLIDE 18

$$\begin{array}{lrcl}
\min & \boldsymbol{c}'\boldsymbol{x} & + & M x_{n+1} \\
\text{s.t.} & \boldsymbol{A}\boldsymbol{x} & + & (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{e})x_{n+1} = \boldsymbol{b} \\
& (\boldsymbol{x}, x_{n+1}) \geq \boldsymbol{0} & &
\end{array}$$

8 Example

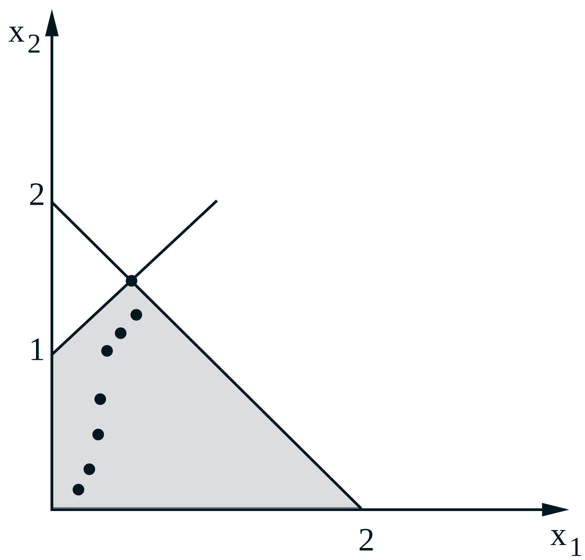
SLIDE 19

$$\begin{array}{lrcrcl}
\max & x_1 & + & 2x_2 & & \\
\text{s.t.} & x_1 & + & x_2 & \leq & 2 \\
& -x_1 & + & x_2 & \leq & 1 \\
& x_1, x_2 \geq 0 & & & &
\end{array}$$

9 Practical Performance

SLIDE 20

- Excellent practical performance, simple
- Major step: invert $\boldsymbol{A}\boldsymbol{X}_k^2\boldsymbol{A}'$
- Imitates the simplex method near the boundary


x2
2
1
2
x1