Lecture 20: The Affine Scaling Algorithm
1 Outline

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

2 History

- 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

\[
\begin{align*}
\min & \quad c'x \\
\text{s.t.} & \quad Ax = b \\
& \quad x \geq 0
\end{align*}
\]

and its dual

\[
\begin{align*}
\max & \quad p'b \\
\text{s.t.} & \quad p'A \leq c'
\end{align*}
\]

- \( P = \{ x \mid Ax = b, \ x \geq 0 \} \)
- \( \{ x \in P \mid x > 0 \} \) the interior of \( P \) and its elements interior points
3.2 The idea

4 Algebraic development

4.1 Theorem

\( \beta \in (0, 1), \ y \in \mathbb{R}^n: \ y > 0, \) and

\[
S = \left\{ x \in \mathbb{R}^n \mid \sum_{i=1}^{n} \frac{(x_i - y_i)^2}{y_i^2} \leq \beta^2 \right\}.
\]

Then, \( x > 0 \) for every \( x \in S \)

Proof

\begin{itemize}
  \item \( x \in S \)
  \item \( (x_i - y_i)^2 \leq \beta^2 y_i^2 < y_i^2 \)
  \item \( |x_i - y_i| < y_i; -x_i + y_i < y_i \), and hence \( x_i > 0 \)
\end{itemize}

\( x \in S \) is equivalent to \( \|Y^{-1}(x - y)\| \leq \beta \)

Replace original LP:

\[
\begin{align*}
\min & \quad c'x \\
\text{s.t.} & \quad Ax = b \\
& \quad \|Y^{-1}(x - y)\| \leq \beta.
\end{align*}
\]
\[ d = x - y \]
\[ \min \quad c'd \]
\[ \text{s.t.} \quad Ad = 0 \]
\[ ||Y^{-1}d|| \leq \beta \]

4.2 Solution

If rows of \( A \) are linearly independent and \( c \) is not a linear combination of the rows of \( A \), then

- optimal solution \( d^* \):

\[ d^* = -\beta \frac{Y^2(c - A'p)}{||Y(c - A'p)||}, \quad p = (AY^2A')^{-1}AY^2c. \]

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

4.2.1 Proof

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

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

\( Ad^* = 0 \), since \( AY^2(c - A'p) = 0 \)

\[ c'd = (c' - p'A)d \]
\[ = (c' - p'A)YY^{-1}d \]
\[ \geq -||Y(c - A'p)|| \cdot ||Y^{-1}d|| \]
\[ \geq -\beta ||Y(c - A'p)||. \]

\[ c'd^* = (c' - p'A)d^* \]
\[ = -(c' - p'A)\beta \frac{Y^2(c - A'p)}{||Y(c - A'p)||} \]
\[ = -\beta \frac{(Y(c - A'p))'(Y(c - A'p))}{||Y(c - A'p)||} \]
\[ = -\beta ||Y(c - A'p)||. \]

- \( c'x = c'y + c'd^* = c'y - \beta ||Y(c - A'p)|| \)
4.3 Interpretation

• $y$ be a nondegenerate BFS with basis $B$

• $A = [B \ N]$

• $Y = \text{diag}(y_1, \ldots, y_m, 0, \ldots, 0)$ and $Y_0 = \text{diag}(y_1, \ldots, y_m)$, then $AY = [BY_0 \ 0]$

$$p = (AY^2A')^{-1}AY^2c$$
$$= (B')^{-1}Y_0^{-2}B^{-1}BY_0^2c_B$$
$$= (B')^{-1}c_B$$

• Vectors $p$ dual estimates

• $r = c - A'p$ becomes reduced costs:

$$r = c - A'(B')^{-1}c_B$$

• Under degeneracy?

4.4 Termination

$y$ and $p$ be primal and dual feasible solutions with

$$c'y - b'p < \epsilon$$

$y^*$ and $p^*$ be optimal primal and dual solutions. Then,

$$c'y^* \leq c'y < c'y^* + \epsilon,$$
$$b'p^* - \epsilon < b'p \leq b'p^*$$

4.4.1 Proof

• $c'y^* \leq c'y$

• By weak duality, $b'p \leq c'y^*$

• Since $c'y - b'p < \epsilon$,

$$c'y < b'p + \epsilon \leq c'y^* + \epsilon$$
$$b'p^* = c'y^* \leq c'y < b'p + \epsilon$$
5 Affine Scaling

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

5.2 The Algorithm

1. (Initialization) Start with some feasible \(x^0 > 0\); let \(k = 0\).
2. (Computation of dual estimates and reduced costs) Given some feasible \(x^k > 0\), let
   \[
   X_k = \text{diag}(x_1^k, \ldots, x_n^k),
   \]
   \[
   p_k = (AX_k^2A')^{-1}AX_k^2c,
   \]
   \[
   r_k = c - A'p_k.
   \]
3. (Optimality check) Let \(e = (1, 1, \ldots, 1)\). If \(r_k \geq 0\) and \(e'X_kr_k < \epsilon\), then stop; the current solution \(x^k\) is primal \(\epsilon\)-optimal and \(p^k\) is dual \(\epsilon\)-optimal.
4. (Unboundedness check) If \(-X_k^2r_k \geq 0\) then stop; the optimal cost is \(-\infty\).
5. (Update of primal solution) Let
   \[
   x^{k+1} = x^k - \beta \frac{X_k^2r_k}{\|X_kr_k\|}.\]

5.3 Variants
- \(\|u\|_\infty = \max_i |u_i|, \quad \gamma(u) = \max\{u_i \mid u_i > 0\}\)
- \(\gamma(u) \leq \|u\|_\infty \leq \|u\|\)
- Short-step method.
- Long-step variants
  \[
  x^{k+1} = x^k - \beta \frac{X_k^2r_k}{\|X_kr_k\|_\infty}
  \]
  \[
  x^{k+1} = x^k - \beta \frac{X_k^2r_k}{\gamma(X_kr_k)}
  \]
6 Convergence

6.1 Assumptions

Assumptions A:
(a) The rows of the matrix $A$ are linearly independent.
(b) The vector $c$ is not a linear combination of the rows of $A$.
(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

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 $0 < \beta < 1$, $x^k$ and $p^k$ 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 $x^k$ and $p^k$ converge to some primal and dual optimal solutions, respectively.

7 Initialization

$$\min \quad c'x + Mx_{n+1}$$
$$\text{s.t.} \quad Ax + (b - Ae)x_{n+1} = b$$
$$(x, x_{n+1}) \geq 0$$

8 Example

$$\max \quad x_1 + 2x_2$$
$$\text{s.t.} \quad x_1 + x_2 \leq 2$$
$$-x_1 + x_2 \leq 1$$
$$x_1, x_2 \geq 0$$

9 Practical Performance

• Excellent practical performance, simple
• Major step: invert $AX^2A'$
• Imitates the simplex method near the boundary