Chapter 12: Interior Point Methods

(cp. Bertsimas & Tsitsiklis, Chapter 9)

Affine Scaling Algorithm: Basic Idea

Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ and consider primal-dual pair of LPs:

$$\begin{array}{llcll} \min & c^T \cdot x & \qquad & \max & p^T \cdot b \\ \text{s.t.} & A \cdot x = b & & \text{s.t.} & p^T \cdot A \leq c^T \\ & x \geq 0 & & & \end{array}$$

Definition 12.1.

Let $P := \{x \in \mathbb{R}^n \mid A \cdot x = b,\ x \geq 0\}$. A point $x \in P$ with $x > 0$ is an *interior point of P*. The set $\{x \in P \mid x > 0\}$ is the *interior of P*.

Main idea:

- Start with interior point x^0.
- Form ellipsoid S_0 centered at x^0 and contained in the interior of P.
- Optimize $c^T \cdot x$ over all $x \in S_0$ and find x^1.
- Form ellipsoid S_1 centered at x^1 etc.

Affine Scaling Algorithm: An Ellipsoid Contained in P

Lemma 12.2.

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

$$S := \left\{ x \in \mathbb{R}^n \;\middle|\; \sum_{i=1}^{n} \frac{(x_i - y_i)^2}{{y_i}^2} \le \beta^2 \right\} \;.$$

Then, $x > 0$ for every $x \in S$.

Proof: . . . □

- Fix some $y > 0$ with $A \cdot y = b$ and set $Y := \operatorname{diag}(y_1, \ldots, y_n) \in \mathbb{R}^{n \times n}$.
- Then,

$$x \in S \quad \Longleftrightarrow \quad \|Y^{-1} \cdot (x - y)\|_2 \le \beta \;,$$

- That is, S is an ellipsoid centered at y.
- Let $S_0 := \{x \in S \mid A \cdot x = b\}$ (section of ellipsoid S).
- S_0 is itself an ellipsoid contained in the interior of P.

Affine Scaling Algorithm: Optimizing over the Ellipsoid

We replace the original LP with $\min c^T \cdot x$ s.t. $x \in S_0$.

$$\min \quad c^T \cdot x \qquad \text{s.t.} \quad A \cdot x = b, \quad \|Y^{-1} \cdot (x - y)\|_2 \le \beta$$

Reformulate by setting $d := x - y$.

$$\min \quad c^T \cdot d \qquad \text{s.t.} \quad A \cdot d = 0, \quad \|Y^{-1} \cdot d\|_2 \le \beta \qquad (12.1)$$

Lemma 12.3.

Assume that rows of A are linearly independent and c is not a linear combination of rows of A. An optimal solution d^* to (12.1) is given by

$$d^* := -\beta \frac{Y^2 \cdot (c - A^T \cdot p)}{\|Y \cdot (c - A^T \cdot p)\|_2} \quad \text{with} \quad p := (A \cdot Y^2 \cdot A^T)^{-1} \cdot A \cdot Y^2 \cdot c \;.$$

Moreover, $x := y + d^* \in P$ and

$$c^T \cdot x = c^T \cdot y - \beta \, \|Y \cdot (c - A^T \cdot p)\|_2 < c^T \cdot y \;.$$

Proof: See Bertsimas & Tsitsiklis, proof of Lemma 9.2. □

Remark: If $d^* \ge 0$, the LP is unbounded since $c^T \cdot d^* < 0$, $A \cdot d^* = 0$, and $y + \alpha \cdot d^* > 0$ for all $\alpha \ge 0$.

Affine Scaling Algorithm: Interpretation of p

- To interpret p, assume for a moment that y is a non-degenerate basic feasible solution (contradicting $y > 0$!) with basis B and $A = [B, N]$.
- Let $Y := \text{diag}(y_1, \ldots, y_m, 0, \ldots, 0)$ and $Y_0 := \text{diag}(y_1, \ldots, y_m)$.
- Then, $A \cdot Y = [B \cdot Y_0, 0]$ and
$$\begin{aligned} p :=& (A \cdot Y^2 \cdot A^T)^{-1} \cdot A \cdot Y^2 \cdot c \\ =& (B^T)^{-1} \cdot Y_0^{-2} \cdot B^{-1} \cdot B \cdot Y_0{}^2 \cdot c_B = (B^T)^{-1} \cdot c_B \end{aligned}$$
is the corresponding dual basic solution.
- Thus, the vector p, corresponding to a primal solution y, is called dual estimate, even if y is not basic.
- The vector $r := c - A^T \cdot p$ becomes $r = c - A^T \cdot (B^T)^{-1} \cdot c_B$, i. e., the reduced cost vector.
- Notice that, if y is degenerate, the matrix $A \cdot Y^2 \cdot A^T$ is not invertible and this interpretation breaks down.

Affine Scaling Algorithm: Duality Gap

If $r = c - A^T \cdot p \geq 0$, then p is a dual feasible solution and

$$r^T \cdot y = (c - A^T \cdot p)^T \cdot y = c^T \cdot y - p^T \cdot A \cdot y = c^T \cdot y - p^T \cdot b$$

is the difference in objective function values between y and p, called the duality gap.

If $r^T \cdot y = 0$, complementary slackness cond. hold and y, p are optimal.

Lemma 12.4.

Let y and p be a primal and dual feasible solution, respectively, such that

$$r^T \cdot y = c^T \cdot y - p^T \cdot b < \varepsilon \ .$$

Let y^* and p^* be optimal primal and dual feasible solutions, respectively. Then,

$$\begin{aligned} c^T \cdot y^* \leq c^T \cdot y < c^T \cdot y^* + \varepsilon \ , \\ b^T \cdot p^* - \varepsilon < b^T \cdot p \leq b^T \cdot p^* \ . \end{aligned}$$

Proof:. . . □

Affine Scaling Algorithm

1. Start with feasible $x^0 > 0$; set $k := 0$; (initialization)

2. Let
$$X_k := \text{diag}(x_1^k, \dots, x_n^k)$$
$$p^k := (A \cdot {X_k}^2 \cdot A^T)^{-1} \cdot A \cdot {X_k}^2 \cdot c$$
$$r^k := c - A^T \cdot p^k$$
(computation of dual estimates and reduced costs)

3. If $r^k \geq 0$ and ${r^k}^T \cdot x^k < \varepsilon$, then stop;
(optimality check; x^k is primal ε-optimal and p^k is dual ε-optimal)

4. If $-{X_k}^2 \cdot r^k \geq 0$, then stop;
(unboundedness check; optimal cost is $-\infty$)

5. Let
$$x^{k+1} := x^k - \beta \frac{{X_k}^2 \cdot r^k}{\|X_k \cdot r^k\|_2} \;;$$
set $k := k + 1$ and go to 2. (update of primal solution)

Affine Scaling Algorithm: Convergence

Assumptions:

- rows of A are linearly independent
- c is not a linear combination of rows of A
- there exists an optimum solution
- there exists a positive feasible solution
- every basic feasible solution to the primal is non-degenerate
- at every basic feasible solution to the primal, the reduced costs of non-basic variables are all non-zero

Theorem 12.5.

Under these assumptions, for $\varepsilon = 0$, the algorithm converges to a pair of primal and dual optimal solutions. □

Affine Scaling Algorithm: Initialization and Performance

Initialization: Consider auxiliary problem:

$$
\begin{aligned}
\min\ \ & c^T \cdot x + M\, x_{n+1} \\
\text{s.t.}\ \ & A \cdot x + (b - A \cdot e)\, x_{n+1} = b \\
& (x, x_{n+1}) \geq 0
\end{aligned}
$$

Notice that $(e, 1)$ is a positive feasible solution.

Computational performance:

- Simple algorithm with excellent performance in practice.
- Computational bottleneck in each iteration: calculation of p^k.
- Computing matrix $A \cdot {X_k}^2 \cdot A^T$ takes $O(m^2 n)$ arithmetic operations.
- Solving system of linear equations involving matrix $A \cdot {X_k}^2 \cdot A^T$ takes $O(m^3)$ arithmetic operations.
- In total, $O(m^2 n)$ arithmetic operations per iteration.