COMPUTATIONAL METHODS
IN OPTIMIZATION

October 2, 2023

Outline:

Primal-dual IPM (continued)

Today + Tomorrow (ft. exercises)

Next week: Barrier IPMs on Monday Oct. 9

No class on Tuesday Oct 10

PRIMAL-DUAL INTERIOR-POINT IMPLEMENTATION

Goal: Solve the linear program

$$(P) \quad \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad c^T x \quad \text{s.t.} \quad Ax = b, \quad x \geq 0$$

$A \in \mathbb{R}^{m \times n}$

] short for $x_i \geq 0 \ \forall i = 1..n$

All solvers consider (explicitly or implicitly) the dual program

$$(D) \quad \underset{\substack{y \in \mathbb{R}^m \\ s \in \mathbb{R}^n}}{\text{maximize}} \quad b^T y \quad \text{s.t.} \quad A^T y + s = c, \quad s \geq 0$$

↳ If both problems are feasible, the set of primal-dual solutions (x^*, y^*, s^*) such that x^* solves (P) and (y^*, s^*) solve (D) is given by the solutions of the KKT equations

$$\begin{cases} Ax^* - b = 0 \\ A^T y^* + s^* - c = 0 \\ x^* \geq 0 \\ s^* \geq 0 \\ x_i^* s_i^* = 0 \quad \forall i = 1..n \end{cases} \iff X^* S^* e = 0$$

$$X^* = \begin{bmatrix} x_1^* & & 0 \\ & \ddots & \\ 0 & & x_n^* \end{bmatrix} \quad S^* = \begin{bmatrix} s_1^* & & 0 \\ & \ddots & \\ 0 & & s_n^* \end{bmatrix}$$

$$e = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^n$$

In interior-point methods, the last 3 conditions are replaced by

$$x^k > 0$$

$$s^k > 0$$

$$X^k S^k e = \mu^k e \qquad \text{where } \mu^k = \frac{(x^k)^T (s^k)}{n} > 0$$

Primal-dual IPMs compute (x^k, y^k, s^k) that converge to (x^*, y^*, s^*) that satisfies KKT but $\forall k$, (x^k, y^k, s^k) never satisfies $X^k S^k e = 0$ $\left(\|X^k S^k e\| \xrightarrow[k\to\infty]{} 0 \right)$

$\begin{pmatrix} x^k > 0 \\ s^k > 0 \end{pmatrix}$

Algorithm (A, b, c)

Initialization: Compute (x^0, y^0, s^0) that is primal-dual strictly feasible

$\{ Ax^0 - b = 0,\ A^T y^0 + s^0 - c = 0,\ x^0 > 0,\ s^0 > 0 \}$

Set $\mu^0 = \frac{(x^0)^T (s^0)}{n}$ and $k = 0$. Choose $\varepsilon > 0$.

While $(\mu^k \geq \varepsilon)$

- Solve the linear system

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I_m \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{bmatrix} = - \begin{bmatrix} 0 \\ 0 \\ X^k S^k e \end{bmatrix}$$

to obtain a direction $\begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{bmatrix} \in \mathbb{R}^{n+m+n}$

- Compute a stepsize $\alpha^k > 0$ such that $\begin{bmatrix} x^k \\ y^k \\ s^k \end{bmatrix} + \alpha^k \begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{bmatrix}$ is primal-dual strictly feasible.
- Set $\begin{bmatrix} x^{k+1} \\ y^{k+1} \\ s^{k+1} \end{bmatrix} = \begin{bmatrix} x^k \\ y^k \\ s^k \end{bmatrix} + \alpha^k \begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{bmatrix}$

and $k = k+1$

↳ Implementation principles

(1) Add a budget criterion to the main loop

(typically stop the method after $K \in \mathbb{N}$ iterations)

Motivation: If the method converges, it should satisfy $\mu_k \leq \varepsilon$ in at most $O\left(n^{\nu} \log(1/\varepsilon)\right)$ iterations

$O(A)$: "big-O" of A $= C \times A$ — C is a positive constant that does not depend on A

Complexity bound for IPMs

with $\nu \geq \frac{1}{2}$ depends on the class of IPMs that is used

(primal-dual, barrier, infeasible, ...)

NB: Other possible budget criteria include CPU time, number of arithmetic operations, number of linear system solves, ...

(2) Solve a "perturbed" linear system

Idea behind primal-dual IPM — At (x^k, y^k, s^k) strictly feasible, solve

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I_n \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \\ \\ \\ \end{bmatrix} = - \begin{bmatrix} 0 \\ 0 \\ X^k S^k e \end{bmatrix} = - \begin{bmatrix} Ax^k - b \\ A^T y^k + s^k - c \\ X^k S^k e \end{bmatrix}$$

→ Newton's method applied to the system of nonlinear equations

$$F_0(x, y, s) = \begin{bmatrix} Ax - b \\ A^T y + s - c \\ XSe \end{bmatrix} = 0_{\mathbb{R}^{2n+m}}$$

$\approx$ KKT equations assuming $x \geq 0$, $s \geq 0$

Newton's method

$F(z) = 0_{\mathbb{R}^\ell}$ — $z \leftarrow z + \Delta z$ where $J_F(z)\Delta z = -F(z)$

$F: \mathbb{R}^\ell \to \mathbb{R}^\ell$ — $J_F(z) \in \mathbb{R}^{\ell \times \ell}$: Jacobian of F evaluated at z

Remarks:

- The matrix $\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S & 0 & X \end{bmatrix}$ has a block structure, that should be exploited by the linear system solver.
- When $n >> 1$ (n large) or $m >> 1$, the solver must exploit the block structure and the sparsity (number of zero coefficients) of the matrix to solve the linear system
- In the second case, the linear system is only solved in an approximate fashion.

↳ In the notebook implementation, we replaced $F_0(\cdot)$ by

$$F_{\sigma^k}(x,y,s) = \begin{bmatrix} Ax - b \\ A^T y + s - c \\ XSe - \sigma^k e \end{bmatrix} \quad \text{with } \sigma^k = 0.1\,\mu^k = \frac{0.1}{n}(x^k)^T(s^k)$$

Idea: • Solving $F_0(x,y,s) = 0$ is not guaranteed to produce directions and iterates such that $x > 0$.. $s > 0$

- To promote these conditions, a better approach is to consider a sequence of functions $\{F_{\sigma^k}(\cdot)\}_k$ with $\sigma^k \xrightarrow[k \to \infty]{} 0$ (typically σ^k proportional to μ^k)

- Mathematical justification: provided by studying the central path

$$C = \left\{ (x_\sigma, y_\sigma, s_\sigma) \in \mathbb{R}^{n+m+n} \;\middle|\; F_\sigma(x,y,s) = 0 \right\}_{\sigma > 0}$$