COMPUTATIONAL METHODS IN OPTIMIZATION

December 4, 2023

Today: Final lecture on implementing IPMs

Next sessions: Lab / course project

- Tuesday Dec 5 3.30 - 6.45 pm
- Monday Dec 11 8.30 - 11.45 am
- Tuesday Dec 12 3.30 - 5 pm
- Thursday Dec 14 1.45 - 3.15 pm

IMPLEMENTING INTERIOR-POINT METHODS

↳ IPMs are well implemented when
- efficient linear algebra techniques are used
- the algorithm is tailored to the problem of interest
- the iterates remain in the interior of the primal-dual feasible set (central path neighborhood)

↳ Today: illustration of this for LPs (but this is even more important for conic programming in general)

1 Linear algebra in IPMs for LPs

Consider a linear program in standard form and its dual:

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

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

IPMs:
- Start with (x^0, y^0, s^0) such that $x^0 > 0$, $s^0 > 0$
- For $k = 0, 1, \ldots$
  - Solve

$I_n = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

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

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

- Set $(x^{k+1}, y^{k+1}, s^{k+1}) = (x^k, y^k, s^k) + \alpha^k (\Delta x^k, \Delta y^k, \Delta s^k)$
  where α^k is chosen so that $x^{k+1} > 0$ and $s^{k+1} > 0$
  (+ possibly additional conditions)

- Advanced variants: Replace $X^k S^k e$ by $X^k S^k e - \sigma^k \mu^k e$, with $\mu^k = \frac{(x^k)^T s^k}{n}$ and $\sigma^k > 0$

↳ The main cost of an IPM iteration is solving the linear system (size $(m+2n) \times (m+2n)$)

$\Rightarrow$ In practice, this system is never solved in the form above, but it is reduced to a smaller system of linear equations

Augmented system

(For simplicity, we drop the k superscripts)

System

$$\begin{array}{c|c} & \begin{array}{ccc} \overset{n}{\longleftrightarrow} & \overset{m}{\longleftrightarrow} & \overset{n}{\longleftrightarrow} \end{array} \\ \begin{array}{c} m \updownarrow \\ n \updownarrow \\ n \updownarrow \end{array} & \begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I_n \\ S & 0 & X \end{bmatrix} \end{array} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} r_b \\ r_c \\ r_{xs} \end{bmatrix} \begin{array}{c} \updownarrow m \\ \updownarrow n \\ \updownarrow n \end{array}$$

$S = \text{diag}(s) = \begin{bmatrix} s_1 & & 0 \\ & \ddots & \\ 0 & & s_n \end{bmatrix}$

$$r_b = b - Ax, \quad r_c = c - A^T y - s, \quad r_{xs} = -XSe$$

$X = \text{diag}(x)$

↳ The last n equations of the system give

$$S \Delta x + X \Delta s = r_{xs}$$

Needed for the IPM iteration

Since $x > 0$ and $s > 0$, S and X are invertible matrices, hence $S^{-1} = \begin{bmatrix} 1/s_1 & & 0 \\ & \ddots & \\ 0 & & 1/s_n \end{bmatrix}$ and $X^{-1} = \begin{bmatrix} 1/x_1 & & 0 \\ & \ddots & \\ 0 & & 1/x_n \end{bmatrix}$

are well-defined.

Thus, we can express Δs as a function of Δx as follows

$$\boxed{\Delta s = X^{-1} r_{xs} - X^{-1} S \Delta x.}$$

Substituting the expression for Δs in the main linear system, we reduce it to a system of size $m+n$ called the augmented system:

$$\begin{bmatrix} -\Theta^{-1} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} r_c - X^{-1} r_{xs} \\ r_b \end{bmatrix}$$

where $\Theta = X S^{-1} = \begin{bmatrix} x_1/s_1 & & 0 \\ & \ddots & \\ 0 & & x_n/s_n \end{bmatrix}$ is diagonal

Idea: • If we solve the augmented system for $(\Delta x, \Delta y)$, we get Δs immediately by computing $\underbrace{X^{-1}}\, r_{xs} - \underbrace{X^{-1} S}\, \Delta x$

diagonal matrices

• The augmented system is of size $m+n$ whereas the original system was of size $m+2n$

$\Rightarrow$ the augmented system is in general easier to solve computationally

$\longrightarrow$ The augmented system approach applies beyond LP (e.g. for QPs, the only difference with the approach above is that $-\Theta^{-1}$ term becomes $-Q-\Theta^{-1}$ where Q is the matrix representing the quadratic term in the objective)

Normal equations

↳ One step further than the augmented system, typically used for LPs but harder to use on more general classes of problems

↳ Starts from the augmented system

$$\begin{array}{c} \\ \updownarrow m+n \end{array} \begin{bmatrix} -\Theta^{-1} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} r_c - X^{-1} r_{xs} \\ r_b \end{bmatrix}$$

$\longleftrightarrow n+m$

First n equations

$$-\Theta^{-1}\Delta x + A^T \Delta y = r_c - X^{-1} r_{xs}$$

$$\Longleftrightarrow \quad -\Delta x + \Theta A^T \Delta y = \Theta r_c - \Theta X^{-1} r_{xs}$$

$$\Longleftrightarrow \quad \Delta x = \Theta A^T \Delta y - \Theta r_c + \Theta X^{-1} r_{xs}$$

↳ Plugging the expression for Δx in the last m equations, we obtain

$$A \Delta x = r_b$$

$$\Longleftrightarrow \quad A \Theta A^T \Delta y - A\Theta r_c + A \Theta X^{-1} r_{xs} = r_b$$

$$\Longleftrightarrow \quad \boxed{\underbrace{A \Theta A^T}_{\in \mathbb{R}^{m\times m}} \Delta y = r_b + A\Theta r_c - A \Theta X^{-1} r_{xs}}$$

"Normal equations"

Linear system of m equations with m unknowns

↳ Idea: If we solve the normal equations for Δy, we can then compute Δx (and thus Δs) without having to solve another linear system.

$\Rightarrow$ The normal equations are typically easier to solve numerically than the augmented system

② Exploiting structure

Motivating (LP) example

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad c^T x \quad \text{s.t.} \quad Ax = b, \quad 0 \le x \le e \qquad e = \begin{bmatrix} 1 \\ | \\ 1 \end{bmatrix} \in \mathbb{R}^n$$

↳ If we reformulate the problem in standard form, we would obtain a problem in $2n$ variables with $m + 3n$ dual variables.

$$\underset{\substack{x \in \mathbb{R}^n \\ w \in \mathbb{R}^n}}{\text{minimize}} \quad c^T x \quad \text{s.t.} \quad \begin{aligned} m\,(\ & Ax = b \\ n\,(\ & x \ge 0 \\ n\,(\ & w = e - x \quad (\Leftrightarrow x + w = e) \\ n\,(\ & w \ge 0 \end{aligned}$$

$\Rightarrow$ At every iteration of IPM, we would need to solve a linear system of $2n + m + 3n = m + 5n$ equations.

$\Rightarrow$ With the normal equations reformulation, we can reduce it to a system of $m + n$ equations ($m + 5n \to m + 3n$ augmented system $\to m + n$ normal equations)

↳ Generic approach for LPs in standard form, does not use the structure of this specific problem

↳ We can reduce the size of the linear system to be solved by exploiting the particular structure of the problem

$$\underset{\substack{x\in\mathbb{R}^n\\ w\in\mathbb{R}^n}}{\text{minimize}} \quad c^T x \quad \text{s.t.} \quad \begin{aligned} Ax &= b \\ x+w &= e \\ x &\geq 0 \\ w &\geq 0 \end{aligned}$$

The dual of this problem is

$$\underset{\substack{y\in\mathbb{R}^m\\ z\in\mathbb{R}^n\\ s\in\mathbb{R}^n\\ t\in\mathbb{R}^n}}{\text{maximize}} \quad b^T y \quad \text{s.t.} \quad \begin{aligned} A^T y + z + s &= c \\ z + t &= 0 \\ s &\geq 0 \\ t &\geq 0 \end{aligned}$$

(Standard LP dual for $\underset{\hat{x}\in\mathbb{R}^{2n}}{\text{minimize}} \ \hat{c}^T\hat{x}$ s.t. $\hat{A}\hat{x} = \hat{b}$, $\hat{x} \geq 0$

with $\hat{x} = \begin{bmatrix} x \\ w \end{bmatrix}$, $\hat{c} = \begin{bmatrix} c \\ 0 \end{bmatrix}$, $\hat{A} = \begin{bmatrix} A & 0 \\ I & I \end{bmatrix} \in \mathbb{R}^{\hat{m}\times 2n}$, $\hat{b} = \begin{bmatrix} b \\ e \end{bmatrix} \begin{matrix} \updownarrow m \\ \updownarrow n \end{matrix}$, $\hat{m} = m+n$

Dual

$$\underset{\substack{\hat{y}\in\mathbb{R}^{m+n}\\ \hat{s}\in\mathbb{R}^{2n}}}{\text{maximize}} \quad \hat{b}^T\hat{y} \quad \text{s.t.} \quad \begin{aligned} \hat{A}^T\hat{y} + \hat{s} &= \hat{c} \\ \hat{s} &\geq 0 \end{aligned}$$

$\hat{y} = \begin{bmatrix} y \\ z \end{bmatrix}$ $\longleftarrow$

$\hat{s} = \begin{bmatrix} s \\ t \end{bmatrix}$)

$\longrightarrow$ In the dual, the variables z can be eliminated since $z = -t$

$\Rightarrow$ we end up with a dual of the form

$$\underset{\substack{y\in\mathbb{R}^m\\ s,t\in\mathbb{R}^n}}{\text{maximize}} \quad b^T y \quad \text{s.t.} \quad \begin{aligned} A^T y - t + s &= c \\ s &\geq 0 \\ t &\geq 0 \end{aligned}$$

The reasoning from (1) can be applied to reduce the IPM iteration to a solve of a system of m linear equations (in y)

$$\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta w \\ \Delta y \\ \Delta s \\ \Delta t \end{bmatrix} = \ldots$$

$$\Downarrow$$

$$A \bar{H}^{-1} A^T \Delta y = \ldots$$

$$\bar{H} = X^{-1} S + W^{-1} T$$

$W = \mathrm{diag}(w)$

$T = \mathrm{diag}(t)$