Lecture Outline

- Indicator Functions and Barrier transformation
- Optimality Conditions for the Barrier Transformed Problem
- Primal Interior Point Method (IPM)
- A primal-dual reformulation of the optimality conditions
- IPM for linear programmes;
- IPM for general convex problems;

You should be able to …

- understand the origins of interior point methods (IPM);
- Formulate a primal IPM and identify its limitations;
- Formulate a primal-dual version of IPM;
- Observe how Newton's method is applied in IPM iterations;
- Write down an IPM for a general convex problem;

Interior-point methods

$$
\begin{aligned}
\min \quad & f(x) \\
\text{s.t.} \quad & c_i(x) \geq 0, i \in \mathcal{I} \\
& a_i^T x - b_i = 0, \quad i \in \mathcal{E}
\end{aligned}
$$

- f is convex, c_i are concave and all are twice differentiable.
- $\mathcal{I} \cap \mathcal{E} = \emptyset$ and $\mathcal{I} \cup \mathcal{E} = \{1, \ldots, m\}$
- The optimal value exists and can be attained.
- The feasible set has an interior, in other words the Slaters' conditions are satisfied. $\Rightarrow$ The strong duality holds.
- Other constraint qualifications are enough as well, e.g. LICQ, $c_{\mathcal{I}}$ being affine, etc.
- Examples are LPs, QPs, QCQPs, SDPs, ...

Indicator Functions and Barrier transformation

Reformulation via indicator functions:

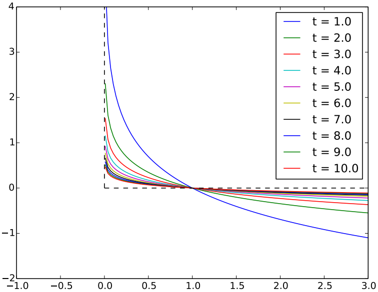
$$\begin{aligned} \min \quad & f(x) + \sum_{i \in \mathcal{I}} \mathbb{I}(c_i(x)) \\ \text{s.t.} \quad & a_i^T x - b_i = 0, \quad i \in \mathcal{E} \end{aligned}$$

$$\mathbb{I}(x) = \begin{cases} 0, & x \geq 0 \\ \infty, & \text{otherwise} \end{cases}$$

- A smooth way of approximating indicator functions is to use logarithmic barrier functions; replace $\mathbb{I}$ with $-\log(c_i(x))$
- Other barrier terms are possible. The logarithmic barrier term is the *canonical* choice.
- The effect is a perturbed problem where an infinite cost is incurred as $c_i(x) \to 0$.
- The weight on the barrier term is denoted by μ and is referred to as the *barrier parameter*.
- The combinatorial effect vs. a perturbation of the original

Barrier Transformation

$$\begin{aligned} \min \quad & f(x) - \mu \sum_{i \in \mathcal{I}} \log(c_i(x)) \\ \text{s.t.} \quad & a_i^T x - b_i = 0, \quad i \in \mathcal{E} \end{aligned}$$



- For concave $c_i(x)$, $-\log(c_i(x))$ is a convex function on the convex set $\{x | c_i(x) > 0,\ i \in \mathcal{I}\}$
- The approximation gets better as $\mu \to 0$ ($t \to \infty,\ \mu = 1/t$).

Optimality Conditions for the Barrier Transformed Problem

- The first order necessary conditions for the transformed problem (KKT):

$$\nabla f(x) - \sum_{i \in \mathcal{I}} \nabla c_i(x) \frac{\mu}{c_i(x)} - \sum_{i \in \mathcal{E}} a_i \lambda_i = 0$$

$$a_i^T x - b_i = 0, \quad i \in \mathcal{E}$$

- When x is close to the minimizer $x(\mu)$ and μ is small the optimal Lagrange multipliers $\lambda_i^\star$, $i \in \mathcal{I}$, can be estimated as

$$\lambda_i^\star = \mu / c_i(x), \quad i \in \mathcal{I}$$

Primal IPM

Algorithm:

Primal IPM

Choose $\mu_0 > 0$, a sequence $\{\tau_k\}$ with $\tau_k \to 0$

while A termination condition is not satisfied **do**

Find an approximate minimizer x_k of the barrier transformed problem. $\triangleright$ Terminate when $\max(\|\nabla f(x) - \sum_{i\in\mathcal{I}} \nabla c_i(x) \frac{\mu}{c_i(x)} - \sum_{i\in\mathcal{E}} a_i\lambda_i\|, \|a_i^T x - b_i\|) \leq \tau_k$

$[\lambda_k]_i \leftarrow \mu_k / c_i(x_k)$, $i \in \mathcal{I}$

Choose $\mu_{k+1} < \mu_k$

$k \leftarrow k + 1$

end while

- Proposed Frisch in the 1950s and was analyzed and popularized by Fiacco and McCormick in the late 1960s.
- Fell out of favour: $x(\mu)$ becomes prohibitively difficult to find as $\mu \downarrow 0$ because of the nonlinearity of the barrier transformed problem.

A primal-dual reformulation of the optimality conditions

- Let $\lambda_i = \mu / c_i(x)$, $i \in \mathcal{I}$.
- If $c_i(x) > 0$ then $\lambda_i - \mu / c_i(x) = 0$ iff $c_i(x)\lambda_i - \mu = 0$:

$$\begin{aligned} \nabla f(x) - \sum_{i \in \mathcal{I}} \nabla c_i(x) \lambda_i - \sum_{i \in \mathcal{E}} a_i \lambda_i = 0& \\ a_i^T x - b_i = 0, \quad i \in \mathcal{E}& \\ c_i(x)\lambda_i - \mu = 0, \quad i \in \mathcal{I}& \end{aligned}$$

- The equations are called *primal-dual nonlinear equations*.
- A method based on approximately solving these equations is called a *primal-dual interior point method*.
- Perturbed version of the KKT of the original problem:

$$\begin{aligned} \nabla f(x) - \sum_{i \in \mathcal{I}} \nabla c_i(x) \lambda_i - \sum_{i \in \mathcal{E}} a_i \lambda_i = 0& \\ a_i^T x - b_i = 0, \quad i \in \mathcal{E}& \\ \lambda_i \geq 0, \quad c_i(x) \geq 0, \quad i \in \mathcal{I}& \\ c_i(x)\lambda_i = 0, \quad i \in \mathcal{I}& \end{aligned}$$

Interior Point Methods (IPM)

- The term interior point methods is used as a common name for methods of barrier type for nonlinear optimization.
- Barrier methods in primal form are from the 60s. They have some less desirable properties due to ill-conditioning. The methods were revived in 1984 for linear programming.
- Primal-dual interior point methods are methods of the 90s. They have better behavior.
- We will consider the special case of linear programming, for simplicity.

Primal LP:

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Dual LP:

$$\begin{aligned} \max \quad & b^T y \\ \text{s.t.} \quad & A^T y + s = c \\ & s \geq 0 \end{aligned}$$

The primal-dual nonlinear equations

- Perturb the complementary condition $x_j s_j = 0$ to $x_j s_j = \mu$ for a positive barrier parameter.

$$\begin{aligned} Ax &= b, \\ A^T y + s &= c, \\ x_j s_j &= \mu, \quad j = 1, \ldots, n \end{aligned}$$

- $x \geq 0$ and $s \geq 0$ are kept "implicitly".

Theorem: *The primal-dual nonlinear equations are well defined and have a unique solution with $x > 0$ and $s > 0$ for all $\mu > 0$ if $\{x | Ax = b, x > 0\} \neq \emptyset$ and $\{(y, s) : A^T y + s = c, s > 0\} \neq \emptyset$. Denote this solution by $x(\mu)$, $y(\mu)$ and $s(\mu)$.*

The primal-dual nonlinear equations

- In matrix form:

$$\begin{aligned} Ax &= b, \\ A^T y + s &= c, \\ XS\mathbf{1} &= \mu\mathbf{1} \end{aligned}$$

- $X = \mathrm{diag}(x)$, $S = \mathrm{diag}(s)$, and $\mathbf{1} = [1, \ldots, 1]^T \in \mathbb{R}^n$.

Theorem: *A solution $x(\mu)$, $y(\mu)$ and $s(\mu)$ is such that $x(\mu)$ is primal feasible and $y(\mu)$ and $s(\mu)$ are dual feasible with duality gap $n\mu$.*

Primal Perspective

- $x(\mu)$ solves:

$$\begin{aligned} \min \quad & c^T x - \mu \sum_{i=1}^{n} \log x_i \\ \text{s.t.} \quad & Ax = b, \quad x \geq 0 \end{aligned}$$

- $y(\mu)$ Lagrange multiplier for $Ax = b$.
- Optimality conditions:

$$\begin{aligned} c_i - \mu / x_i &= a_i^T y, \quad i = 1, \ldots, n \\ Ax &= b \\ x &> 0 \end{aligned}$$

Dual Perspective

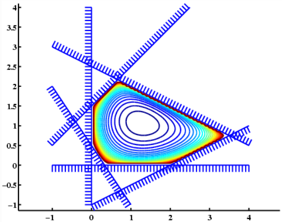
- $y(\mu)$ and $s(\mu)$ solve:

$$\begin{aligned} \min \quad & -b^T y - \mu \sum_{i=1}^{n} \log s_i \\ \text{s.t.} \quad & A^T y + s = c, \quad s \geq 0 \end{aligned}$$

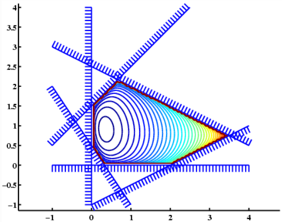
- $-x(\mu)$ Lagrange multiplier for $A^T y + s = c$.
- Optimality conditions:

$$\begin{aligned} b &= Ax \\ \frac{\mu}{s_i} &= x_i, \quad i = 1, \ldots, n \\ A^T y + s &= c \\ s &> 0 \end{aligned}$$

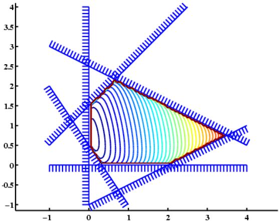
Primal Barrier Function for different μ



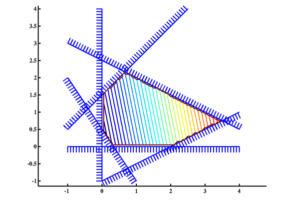
$\mu = 5$



$\mu = 1$



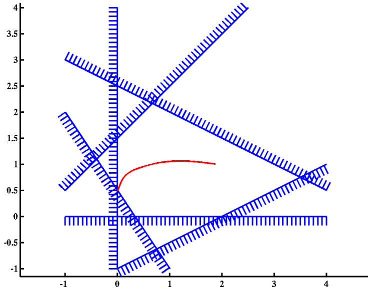
$\mu = 0.3$



$\mu = 10^{-16}$

Barrier Trajectory

- The *barrier trajectory* is defined as the set $\{[x(\mu), y(\mu), s(\mu)] | \mu > 0\}$.
- The primal-dual system of nonlinear equations is preferred.
- Pure primal and pure dual point of view gives high nonlinearity.



Theorem: *If the barrier trajectory is well defined, then* $\lim_{\mu \to 0} x(\mu) = x^\star$, $\lim_{\mu \to 0} y(\mu) = y^\star$, $\lim_{\mu \to 0} s(\mu) = s^\star$, *where* $x^\star$ *is primal optimal, and* $(y^\star, s^\star)$ *are dual optimal.*

Primal-Dual Interior Point Method Iterations

- Based on Newton-iterations on the perturbed optimality conditions.
- For a given point (x, y, s), with $x > 0$ and $s > 0$ a suitable value of μ is chosen. The Newton-iteration then becomes

$$\boxed{\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = - \begin{bmatrix} Ax - b \\ A^T y + s - c \\ XS\mathbf{1} - \mu \mathbf{1} \end{bmatrix}}$$

- Common choice: $\boxed{\mu = \sigma \frac{x^T s}{n}}$ for some $\sigma \in [0, 1]$
- Note that $Ax = b$ and $A^T y + s = c$ do not need to be satisfied initially. They will be satisfied in the next step.

Primal-Dual Interior Point Method For LPs

Algorithm: Primal-Dual IPM for LPs

$x_0 > 0,\ s_0 > 0,\ y_0,\ k \leftarrow 0$

while A termination condition is not satisfied **do**

  $\mu \leftarrow \sigma x_k^T s_k / n$

  Compute Δx, Δy, and Δs

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S_k & 0 & X_k \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} Ax_k - b \\ A^T y_k + s_k - c \\ X_k S_k \mathbf{1} - \mu \mathbf{1} \end{bmatrix}$$

  $\alpha_{\max} \leftarrow \max\ \alpha,\ \text{s.t.} x + \alpha\Delta x \geq 0,\ s + \alpha \Delta s \geq 0.$

  $\alpha = \min(1, (1-\tau)\alpha_{\max})$ $\triangleright$ $0 < \tau \ll 1$. This step rule is very simplistic; no convergence guarantees.

  $x_{k+1} \leftarrow x_k + \alpha \Delta x$;

  $y_{k+1} \leftarrow y_k + \alpha \Delta y$;

  $s_{k+1} \leftarrow s_k + \alpha \Delta s$;

  $k \leftarrow k + 1$

end while

Primal-Dual IPM For LPs: Choosing σ

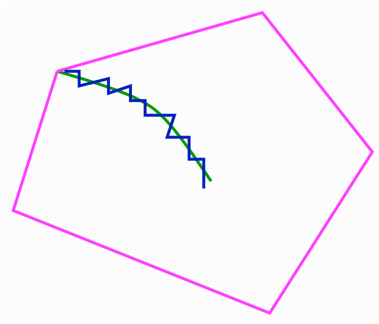
Theorem: *Assume that x satisfies $Ax = b$, $x > 0$, and assume that y and s satisfies $A^T y + s = c$, $s > 0$, and let $\mu = \sigma x^T s/n$. Then*

$$(x + \alpha\Delta x)^T (s + \alpha\Delta s) = (1 - \alpha(1 - \sigma))x^T s.$$

- It is desirable to have σ small (accuracy) and α large (speed). These goals are in general contradictory.
- Strategies for σ:
    1. Short-step method, $\sigma \ll 1$.
    2. Long-step method, $\sigma \approx 1$.
    3. Predictor-corrector method, $\sigma_k = 0$, $k \in \mathbb{E}$ (even iterations) and $\sigma_k = 1$, $k \in \mathbb{O}$ (odd iterations)

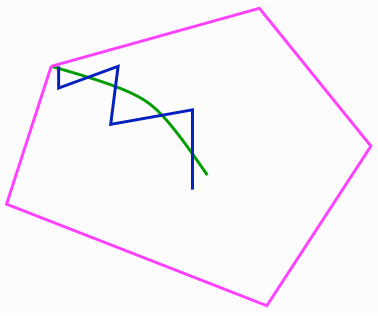
Primal-Dual IPM For LPs: Short-step method

- $\sigma_k = 0.1$, $\alpha_k = 1$.
- The iterates remain close to the barrier trajectory.



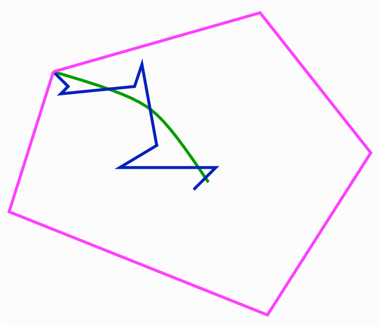
Primal-Dual IPM For LPs: Long-step method

- $\sigma_k = 1$, α_k is selected based on the proximity to the barrier trajectory.



Primal-Dual IPM For LPs: Predictor-corrector method

- The same as short-step for $k \in \mathbb{E}$ and as long-step for $k \in \mathbb{O}$.



- LP IPM converges in few iterations, in the order of 20.
- Typically does not grow with problem size.
- The iterates become more computationally expensive as the problem size increases.
- No clear efficient warm starting.

Solving the system of linear equations in LP-IPM

- Recall the system:

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = - \begin{bmatrix} Ax - b \\ A^T y + s - c \\ XS\mathbf{1} - \mu\mathbf{1} \end{bmatrix}$$

- It is equivalent to solving:

$$\begin{bmatrix} X^{-1}S & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ -\Delta y \end{bmatrix} = - \begin{bmatrix} c - \mu X^{-1}\mathbf{1} - A^T y \\ Ax - b \end{bmatrix}$$

- In turn it is equivalent to:

$$AXS^{-1}A^T \Delta y = AXS^{-1}(c - \mu X\mathbf{1} - A^T y) + b - Ax$$

IPM for Convex Problems

- A primal-dual IPM approximately solves:

$$\begin{aligned} \nabla f(x) - \sum_{i \in \mathcal{I}} \nabla c_i(x)\lambda_i - \sum_{i \in \mathcal{E}} a_i \lambda_i = 0 \\ a_i^T x - b_i = 0, \quad i \in \mathcal{E}, \\ c_i(x)\lambda_i - \mu = 0, \quad i \in \mathcal{I} \end{aligned}$$

- with implicit requirement: $c_i(x) > 0$ and $\lambda_i > 0$, $i \in \mathcal{I}$
- Newton iterations:

$$\begin{bmatrix} \nabla^2_{xx}\mathbf{L}(x,\lambda) & A^T & \nabla c_{\mathcal{I}}(x)^T \\ A & 0 & 0 \\ \Lambda_{\mathcal{I}} \nabla c_{\mathcal{I}}(x) & 0 & -C_{\mathcal{I}}(x) \end{bmatrix} \begin{bmatrix} \Delta x \\ -\Delta \lambda_{\mathcal{E}} \\ -\Delta \lambda_{\mathcal{I}} \end{bmatrix} = - \begin{bmatrix} \nabla_x \mathbf{L}(x,\lambda) \\ Ax - b \\ C_{\mathcal{I}}(x)\lambda_{\mathcal{I}} - \mu \mathbf{1} \end{bmatrix}$$

- $\Lambda_{\mathcal{I}} = \operatorname{diag}(\lambda_{\mathcal{I}})$ and $C_{\mathcal{I}}(x) = \operatorname{diag}(c_{\mathcal{I}}(x))$

IPM for Convex Problems: Solving the System of Linear Equations

- As seen before, one way of solving the problem is via factorisation.
- Let K be the coefficient:

$$P^T K P = L B L^T$$

- where L is lower triangular and B is block diagonal, and P is a matrix of row and column permutations.
- The permutation matrix P seeks a compromise between the goals of preserving sparsity and ensuring numerical stability.
- One can eliminate $-\Delta\lambda_{\mathcal{I}}$ to obtain a smaller problem with coefficient matrix

$$\begin{bmatrix} \nabla^2 \mathbf{L}(x, \lambda) + \nabla c_{\mathcal{I}}(x)^T C_{\mathcal{I}}(x)^{-1} \nabla c_{\mathcal{I}}(x) & A^T \\ A & 0 \end{bmatrix}$$

- Note that all these systems are ill-conditioned due to the nature of the problem.

A Dual-Primal IPM for Convex Problems

Algorithm: Dual-Primal IPM

A feasible x such that $c_{\mathcal{I}}(x_0) > 0$, y_0, $z_0 > 0$, $\mu_0 > 0$ and $\sigma > 0$, $k \leftarrow 0$ $\triangleright$ y stands in for $\lambda_{\mathcal{E}}$ and z for $\lambda_{\mathcal{I}}$

while A termination condition is not satisfied **do**

Compute Δx, Δy, and Δz

$$\begin{bmatrix} \nabla^2 \mathbf{L}(x_k, y_k, z_k) & A^T & \nabla c_{\mathcal{I}}(x_k)^T \\ -A & 0 & 0 \\ -Z_k \nabla c_{\mathcal{I}}(x_k) & 0 & C_{\mathcal{I}}(x_k) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} -\nabla \mathbf{L}(x_k, y_k, z_k) \\ A x_k - b \\ Z_k c_{\mathcal{I}}(x_k) - \mu_k \mathbf{1} \end{bmatrix}$$

Choose α_k such that $c_{\mathcal{I}}(x + \alpha_k \Delta x) > 0, \quad z_k + \alpha_x \Delta z_k > 0$

$x_{k+1} \leftarrow x_k + \alpha \Delta x$; $y_{k+1} \leftarrow y_k + \alpha \Delta y$; $z_{k+1} \leftarrow s_k + \alpha \Delta z$;

if $\max(\|\nabla f(x) - \sum_{i \in \mathcal{I}} \nabla c_i(x) \lambda_i - \sum_{i \in \mathcal{E}} a_i \lambda_i\|, \|Ax - b\|, \|c_i(x)\lambda_i - \mu\|) \leq \epsilon$ **then** $\triangleright$ desired accuracy

$\mu_{k+1} \leftarrow \sigma \mu_k$ $\triangleright$ σ can be 0.1 or time-varying

end if

$k \leftarrow k + 1$

end while

IPM for Convex Problems with Slack Variables

- Having $c_{\mathcal{I}}(x) > 0$ might be troublesome.
- An alternative is to add slack variables:

$$\begin{aligned} \nabla f(x) - \sum_{i\in\mathcal{I}} \nabla c_i(x)\lambda_i - \sum_{i\in\mathcal{E}} a_i\lambda_i &= 0 \\ a_i^T x - b_i &= 0, \quad i \in \mathcal{E}, \\ c_i(x) - s_i &= 0, \quad i \in \mathcal{I}, \\ s_i\lambda_i - \mu &= 0, \quad i \in \mathcal{I} \end{aligned}$$

- with implicit requirement: $s_i > 0$ and $\lambda_i > 0$, $i \in \mathcal{I}$
- Newton iterations:

$$\begin{bmatrix} \nabla^2\mathbf{L}(x,\lambda) & A^T & \nabla c_{\mathcal{I}}(x)^T \\ A & 0 & 0 \\ \Lambda_{\mathcal{I}}\nabla c_{\mathcal{I}}(x) & 0 & -S_{\mathcal{I}} \end{bmatrix} \begin{bmatrix} \Delta x \\ -\Delta\lambda_{\mathcal{E}} \\ -\Delta\lambda_{\mathcal{I}} \end{bmatrix} = - \begin{bmatrix} \nabla\mathbf{L}(x,\lambda) \\ Ax - b \\ \Lambda_{\mathcal{I}} c_{\mathcal{I}}(x) - \mu\mathbf{1} \end{bmatrix}$$

- $\Lambda_{\mathcal{I}} = \operatorname{diag}(\lambda_{\mathcal{I}})$ and $S_{\mathcal{I}} = \operatorname{diag}(s_{\mathcal{I}})$
- $\Delta s_{\mathcal{I}} = -s_{\mathcal{I}} + c_{\mathcal{I}}(x) + \nabla c_{\mathcal{I}}(x)\Delta x$
- Ensure that $s_{\mathcal{I}} > 0$ and $\lambda_{\mathcal{I}} > 0$ in the linesearch. We obtain $c_{\mathcal{I}}(x) \geq 0$ asymptotically.