Interior Point Methods for Convex Optimization

Luiz-Rafael Santos

l.r.santos@ufsc.br

@lrsantos11

https://lrsantos11.github.io/

CME 307 / MS&E 311

UFSC, Brazil & MS&E/Stanford, USA

Outline

IPM for Linear and Quadratic Programs

IPM for Convex nonlinear programming

IPM for Conic Optimization

Convex optimization

Convex optimization problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0 \quad (s \in \mathbb{R}^p) \\ & Ax = b \qquad (y \in \mathbb{R}^m) \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^p$ are smooth and convex, and $A \in \mathbb{R}^{m \times n}$ is full rank.

- KKT conditions

$$\begin{aligned} \nabla f(x) + \sum_{i=1}^{m} y_i a_i + \sum_{j=1}^{p} s_j \nabla g_j(x) &= 0 \\ Ax &= b \\ g(x) &\leq 0 \\ s &\geq 0 \\ s_j g_j(x) &= 0 \quad \forall j = 1, \ldots, p \end{aligned}$$

IPM for Linear and Quadratic Programs

Linear/Quadratic Program

$$\begin{aligned} \min \quad & c^\top x + \tfrac{1}{2} x^\top Q x \\ \text{s.t.} \quad & Ax = b, \\ & x \geq 0, \end{aligned}$$

where $Q \in \mathbb{S}^n_+$, and $A \in \mathbb{R}^{m \times n}$ is full-rank.

- $\mathcal{P} := \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$ is a polyhedron.
- If $Q = 0$, then we have a linear program.

How to solve LP/QP problems?

If we ask Pelé, perhaps he would say "Go through the middle!".

Building blocks of IPM

What do we need to derive the Interior Point Method?

- Duality theory: Lagrangian function; KKT (first order optimality) condition.
- Logarithmic barriers.
- Newton method (with a good linear solver)

Then we will enjoy fantastic convergence properties:

- Theoretical: $O(\sqrt{n}\log(1/\varepsilon))$ iterations
- Practical: $O(\log n \log(1/\varepsilon))$ iterations (but the per-iteration cost may be high)

What is the core of an IPM?

IPM procedure

- replace inequalities with log barriers;
- form the Lagrangian;
- write down the KKT conditions of the perturbed problem;
- find one or more directions based on Newton method applied to KKT system;
- smartly combine the directions and compute a stepsize.

Duality and KKT conditions

Primal-dual pairs of QP

Primal problem

$$\begin{aligned} \min \quad & c^\top x + \tfrac{1}{2} x^\top Q x \\ \text{s.t.} \quad & Ax = b, \\ & x \geq 0, \end{aligned}$$

Dual problem

$$\begin{aligned} \max \quad & b^\top y - \tfrac{1}{2} x^\top Q x \\ \text{s.t.} \quad & A^\top y + s - Qx = c, \\ & s \geq 0, \end{aligned}$$

KKT conditions

$$\begin{aligned} Ax &= b \\ A^\top y + s - Qx &= c \\ XSe &= 0 \quad (i.e., x_j \cdot s_j = 0 \ \forall j) \ \text{complementarity} \\ (x, s) &\geq 0 \end{aligned}$$

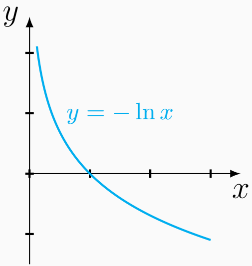
where for $X = \operatorname{diag}(x_1, \ldots, x_n), S = \operatorname{diag}(s_1, \ldots, s_n) \in \mathbb{R}^{n \times n}$ and $e = (1, \ldots, 1) \in \mathbb{R}^n$.

Logarithmic barrier

$$-\ln x_j$$

"replaces" the inequality

$$x_j \geq 0$$



- Them minimization of $-\sum_{j=1}^{n} \ln x_j$ is equivalent to the maximization of the product of distances from all hyperplanes defining the positive orthant:

  it prevents all x_j from approaching zero.

$$\min \mathrm{e}^{-\sum_{j=1}^{n} \ln x_j} \Longleftrightarrow \max \prod_{1 \leq j \leq n} x_j$$

Self-concordant logarithmic barrier

1st step

Replace the primal QP

$$\begin{aligned} \min \quad & c^\top x + \tfrac{1}{2} x^\top Q x \\ \text{s.t.} \quad & Ax = b, \\ & x \geq 0, \end{aligned}$$

with the barrier primal QP

$$\begin{aligned} \min \quad & c^\top x + \tfrac{1}{2} x^\top Q x - \mu \sum_{j=1}^{n} \ln x_j \\ \text{s.t.} \quad & Ax = b, \end{aligned}$$

Logarithmic barrier and stationary

2nd step: Lagrangian function

$$\mathcal{L}(x, y, \mu) = c^\top x + \frac{1}{2} x^\top Q x - y^\top (Ax - b) - \mu \sum_{j=1}^{n} \ln x_j$$

Conditions for a stationary point of the lagrangian

$$\nabla_x \mathcal{L}(x, y, \mu) = c + Qx - A^\top y - \mu X^{-1} e = 0$$
$$\nabla_y \mathcal{L}(x, y, \mu) = Ax - b = 0$$

with $X^{-1} = \operatorname{diag}(x_1^{-1}, \ldots, x_n^{-1}) \in \mathbb{R}^{n \times n}, (x_j > 0)$.

KKT conditions for barrier problem

- Define $s := \mu X^{-1} e$, which implies $XSe = \mu e$, to get

KKT$_\mu$

$$
\begin{aligned}
Ax &= b \\
A^\top y + s - Qx &= c \\
XSe &= \mu e \\
(x, s) &> 0
\end{aligned}
$$

KKT

$$
\begin{aligned}
Ax &= b \\
A^\top y + s - Qx &= c \\
XSe &= 0 \\
(x, s) &\geq 0
\end{aligned}
$$

KKT$_\mu$ $\to$ KKT as $\mu \to 0$.

Central path (LP case)

- Parameter μ controls the distance to optimality

$$c^\top x - b^\top y = c^\top x - x^\top A^\top y = x^\top s = n\mu$$

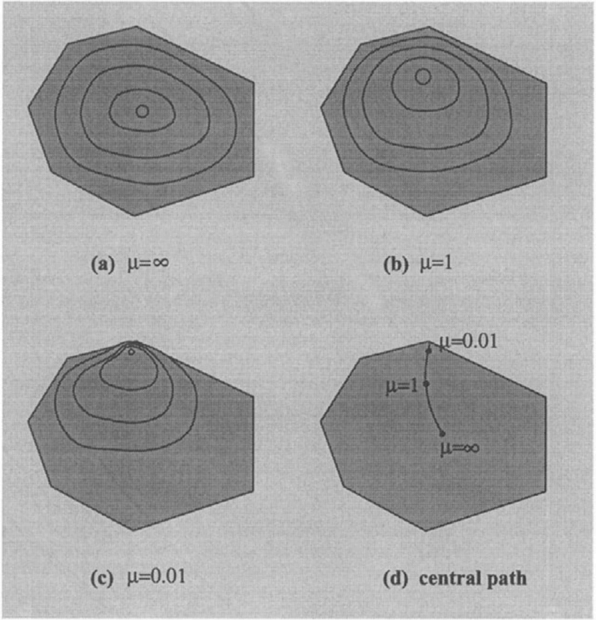
- Analytic center (μ-center): a (unique) point

$$(x(\mu), y(\mu), s(\mu)), \quad x(\mu) > 0, s(\mu) > 0$$

  that satisfies the KKT conditions.
- The curve

$$\mathcal{C}_\mu = \{(x(\mu), y(\mu), s(\mu)) \mid \mu > 0\}$$

  is called the primal-dual central path.

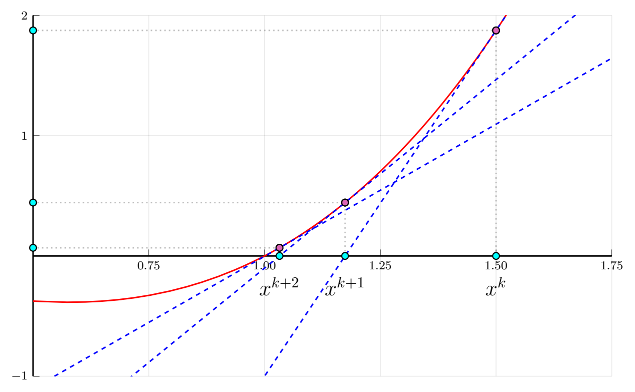


(a) μ=∞
(b) μ=1
(c) μ=0.01
μ=0.01
μ=1
μ=∞
(d) central path

Newton Method

- For $F : \mathbb{R}^n \to \mathbb{R}^n$ smooth, solve $F(x) = 0$.
- Newton method:

$$x^{k+1} = x^k - \alpha_k J_F(x^k)^{-1} F(x^k)$$



Apply Newton Method to KKT$_\mu$

- The first order optimality conditions for the barrier problem form a large system of nonlinear equations

$$F(x, y, s) = 0$$

where $F : \mathbb{R}^{2n+m} \mapsto \mathbb{R}^{2n+m}$ is an application defined as follows:

$$F(x, y, s) = \begin{bmatrix} Ax & -b \\ A^\top y + s - Q & -c \\ XSe & -\mu e \end{bmatrix}$$

- Actually, the first two terms of it are linear; only the last one, corresponding to the complementarity condition, is nonlinear. Note that

$$J_F(x, y, s) = \begin{bmatrix} A & 0 & 0 \\ -Q & A^\top & I \\ S & 0 & X \end{bmatrix}$$

Interior-point QP Algorithm

IPM Framework

We fix the barrier parameter μ and make only one (**damped**) Newton step towards the solution of FOC. We do not solve the current FOC exactly. Instead, we immediately reduce the barrier parameter μ (to ensure progress towards optimality) and repeat the process.

- Given $(x_0, y_0, s_0) \in \mathcal{F}^0$, $\mu_0 = \frac{1}{n} \cdot (x^0)^\top s^0$
- For $k = 0, 1, 2, \ldots$
  - $k = k + 1$
  - $\mu_k = \sigma \mu_{k-1}$, where $\sigma \in (0, 1)$
  - Find Newton direction $(\Delta x^k, \Delta y^k, \Delta s^k)$ by solving

$$\begin{bmatrix} A & 0 & 0 \\ -Q & A^\top & I \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{bmatrix} = \begin{bmatrix} b - Ax^k \\ c - A^\top y^k - s^k + Qx^k \\ \mu_k e - X^k S^k e \end{bmatrix}$$

  - Find step length α_k such that $(x^k + \alpha_k \Delta x^k, y^k + \alpha_k \Delta y^k, s^k + \alpha_k \Delta s^k) \in \mathcal{F}^0$.
  - Make step $(x^{k+1}, y^{k+1}, s^{k+1}) = (x^k + \alpha_k \Delta x^k, y^k + \alpha_k \Delta y^k, s^k + \alpha_k \Delta s^k)$.

Path-following algorithm

- **Short-step path-following method:** $\mathcal{O}(\sqrt{n})$ complexity result

Theorem ([Gondzio, 2012, Thm. 3.1])

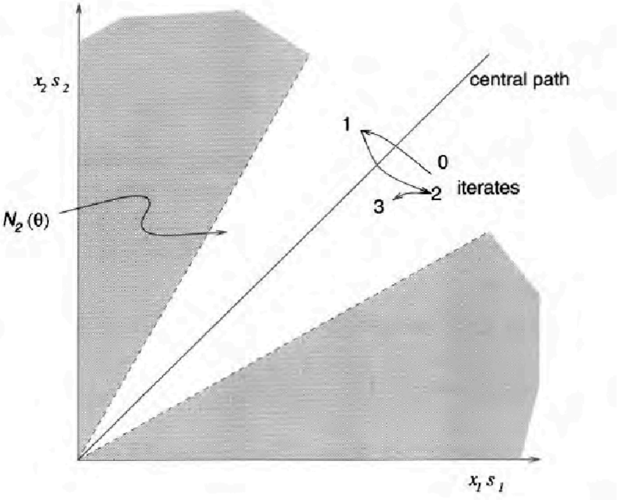
Given $\epsilon > 0$, *suppose that a feasible starting point* $(x^0, y^0, s^0) \in \mathcal{N}_2(0.1)$ *satisfies*

$$\left(x^0\right)^\top s^0 = n\mu^0, \text{ where } \mu^0 \leq 1/\epsilon^\kappa,$$

for some positive constant κ. *Then there exists an index* K *with* $K = \mathcal{O}(\sqrt{n}\ln(1/\epsilon))$ *such that*

$$\mu^k \leq \epsilon, \quad \forall k \geq K$$

- θ-neighborhood of the central path:
  $\mathcal{N}_2(\theta) := \{(x, y, s) \in \mathcal{F}^0 \mid \|XSe - \mu e\| \leq \theta\mu\}$, with $\mu = \frac{1}{n}x^\top s$.
- Slow progress towards optimality



$x_2 s_2$
central path
1
0
iterates
2
3
$N_2(\theta)$
$x_1 s_1$

LP as extension of QP

Newton direction

$$\begin{bmatrix} A & 0 & 0 \\ -Q & A^\top & I \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} b - Ax \\ c - A^\top y - s + Qx \\ \mu_k e - XSe \end{bmatrix} = \begin{bmatrix} \xi_p \\ \xi_d \\ \xi_\mu \end{bmatrix}$$

- Since $\Delta s = X^{-1}\xi_\mu - X^{-1}S\Delta x$, we get $(-Q - X^{-1}S)\Delta x + A^\top \Delta y = \xi_d - X^{-1}\xi_\mu$, so

Augmented system

$$\begin{bmatrix} -Q - \Theta^{-1} & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} \xi_d - X^{-1}\xi_\mu \\ \xi_p \end{bmatrix}$$

- $\Theta = XS^{-1}$ (ill-conditioned matrix)
- QP is a natural extension of LP

LP: Augmented vs Normal Equations

Augmented system

$$\begin{bmatrix} -\Theta^{-1} & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} \xi_d - X^{-1}\xi_\mu \\ \xi_p \end{bmatrix} =: \begin{bmatrix} g \\ \xi_p \end{bmatrix}$$

Normal equations

Eliminate Δx from the first equations gets us the normal equations

$$(A\Theta A^\top)\Delta y = A\Theta g + \xi_p$$

- One can use normal equations in LP, but not in QP.
- Normal equations in QP $(A(Q + \Theta)A^\top)\Delta y = g$ may become almost completely dense even for sparse matrices A and Q.
- In QP, usually the indefinite augmented system form is used.

IPM for NLP

- Convex NLP

$$\min f(x) \quad \text{s.t.} \quad g(x) + z = 0, z \geq 0$$

- Replace inequality $z \geq 0$ with logarithmic barrier

$$\min f(x) - \mu \sum_{i=1}^{m} \ln(z_i) \quad \text{s.t.} \quad g(x) + z = 0$$

- Write out Lagrangian

$$L(x, y, z, \mu) = f(x) + y^\top (g(x) + z) - \mu \sum_{i=1}^{m} \ln(z_i)$$

IPM for NLP

- Write conditions for stationary point

$$\begin{aligned}\nabla_x L(x,z,y) &= \nabla f(x) + J_g(x)^\top y = 0\\ \nabla_y L(x,z,y) &= g(x) + z = 0\\ \nabla_z L(x,z,y) &= y - \mu Z^{-1} e = 0\end{aligned}$$

- Write KKT system

$$\begin{aligned}\nabla f(x) + J_g(x)^\top y = 0,\\ g(x) + z = 0\\ YZe = \mu e\end{aligned}$$

Newton for KKT of NLP

- Apply Newton method for KKT system
- Jacobian matrix of KKT system

$$J_F(x,z,y) = \begin{bmatrix} Q(x,y) & J_g(x)^\top & 0 \\ J_g(x) & 0 & I \\ 0 & Z & Y \end{bmatrix}$$

where $Q(x,y) = \nabla^2 f(x) + \sum_{i=1}^m y_i \nabla^2 g_i(x)$ is the Hessian of L

- Newton step for KKT system

$$\begin{bmatrix} Q(x,y) & J_g(x)^\top & 0 \\ J_g(x) & 0 & I \\ 0 & Z & Y \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} -\nabla f(x) - J_g(x)^\top y \\ -g(x) - z \\ \mu e - YZe \end{bmatrix}$$

From QP to NLP

- Newton direction for NLP

$$\begin{bmatrix} Q(x,y) & J_g(x)^\top & 0 \\ J_g(x) & 0 & I \\ 0 & Z & Y \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} -\nabla f(x) - J_g(x)^\top y \\ -g(x) - z \\ \mu e - YZe \end{bmatrix}$$

- Augmented system for NLP

$$\begin{bmatrix} Q(x,y) & J_g(x)^\top \\ J_g(x) & -ZY^{-1} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} -\nabla f(x) - J_g(x)^\top y \\ -g(x) - \mu Y^{-1} e \end{bmatrix}$$

- NLP is a natural extension of QP
- Computation of $Q(x,y)$ and $J_g(x)$ at each iteration (Automatic differentiation(?))
- Caveat: using trust region method to choose stepsize

Self-concordant function

Definition

We call function f self-concordant if there exists a constant $M_f \geq 0$ such that the inequality

$$\nabla^3 f(x)[u,u,u] \leq M_f \|u\|^{3/2}_{\nabla^2 f(x)}$$

holds for any $x \in \operatorname{dom} f$ and $u \in R^n$.

- A self-concordant function is always well approximated by a quadratic model because the error of such an approximation can be bounded by the $\|u\|^{3/2}_{\nabla^2 f(x)}$

Theorem ([Boyd and Vandenberghe, 2004, Section 11.5])

Newton's method with line search finds an ε approximate solution in less than $T :=$ constant $\times (f(x_0) - f^\star) + \log_2 \log_2 \frac{1}{\varepsilon}$ iterations.

Log-barrier is self-concordant

Theorem

The barrier function $-\ln(x)$ is self-concordant in $\mathbb{R}_+$.

Proof.

Consider $f(x) = -\ln(x)$, then

$$f'(x) = -\frac{1}{x}, \quad f''(x) = \frac{1}{x^2}, \quad f'''(x) = -\frac{2}{x^3}$$

Complete and check that self-concordance condition holds with $M_f = 2$. ☐

- $-\ln(1/x^\alpha)$, with $\alpha \in (0, \infty)$ is not self-concordant in $\mathbb{R}_+$.
- $\exp(1/x)$ is not self-concordant in $\mathbb{R}_+$.

Conic optimization

- Consider the optimization problem

$$\begin{array}{ll} \min & c^\top x \\ \text{s.t.} & Ax = b \\ & x \in K \end{array}$$

  where K is a convex closed cone.

- Tha associated dual is

$$\begin{array}{ll} \max & b^\top y \\ \text{s.t.} & A^\top y + s = c \\ & x \in K^* \text{ (Dual cone)} \end{array}$$

- Weak duality

$$c^\top x - b^\top y = x^\top (c - A^\top y) = x^\top s \geq 0$$

- Conic optimization can be solved in polynomial time with IPMs

Second-order conic optimization

- $K = \mathbb{L} := \{(x,t) \mid x \in \mathbb{R}^{n-1}, t \in \mathbb{R}, \|x\|_2 \leq t, t \geq 0\}$ (Lorenz or second-order cone)
- Logarithmic barrier function for the second-order cone

$$f(x,t) = \begin{cases} -\ln(t^2 - \|x\|_2^2) & \text{if } \|x\| < t \\ +\infty & \text{otherwise} \end{cases}$$

Theorem

The barrier function $f(x,t)$ is self-concordant on $\mathbb{L}$.

Exercise: Prove in case $n = 2$.

Semidefinite programming

- Variable now is a symmetric matrix $X \in \mathbb{S}^n$
- $K = \mathbb{S}^n_+$ (Semi-definite cone)

SDPs and its dual

$$\begin{array}{ll} \min & C \bullet X \\ \text{s.t.} & A_i \bullet X = b_i, i = 1, \ldots, m \\ & X \succeq 0 \end{array} \qquad \begin{array}{ll} \max & b^\top y \\ \text{s.t.} & \sum_{i=1}^m y_i A_i + S = C \\ & S \succeq 0 \end{array}$$

- $A_i, C \in \mathbb{S}^n$ and $b \in \mathbb{R}^m$ given, and $X, S \in \mathbb{S}^n$ and $y \in \mathbb{R}^m$ unknown.
- $X \bullet Y = \operatorname{tr}(X^\top Y)$.

Theorem (Weak duality for SDP)

If X is primal feasible and (y, S) is dual feasible, then

$$C \bullet X - b^\top y = X \bullet S \geq 0$$

Logarithmic barrier for SDP

- Logarithmic barrier function for the semi-definite cone

$$f(X) = \begin{cases} -\ln(\det(X)) & \text{if } X \succ 0 \\ +\infty & \text{otherwise} \end{cases}$$

- Facts (for small t):
  - $\det(I + tU) = 1 + t\,\mathrm{tr}(U) + \mathcal{O}(t^2)$
  - $\ln(1 + t\,\mathrm{tr}(U)) \approx t\,\mathrm{tr}(U)$
- Let $X \succ 0$ and $H \in \mathbb{S}^n$ be given. Then

$$\begin{aligned} f(X + tH) &= -\ln(\det(X + tH)) = -\ln(\det(X(I + tX^{-1}H))) \\ &= -\ln(\det(X)) - \ln(\det(I + tX^{-1}H)) \\ &= -\ln(\det(X)) - \ln(1 + t\,\mathrm{tr}(X^{-1}H) + \mathcal{O}(t^2)) \\ &= f(X) - tX^{-1} \bullet H + \mathcal{O}(t^2) \end{aligned}$$

Derivatives of Logarithmic barrier for SDP

- First derivative of $f(X)$

$$f'(X) = \lim_{t \to 0} \frac{f(X + tH) - f(X)}{t} = -X^{-1}$$

So $Df(X)[H] = -X^{-1} \bullet H$.

- Second derivative of $f(X)$

$$\begin{aligned} f'(X + tH) &= -[X(I + tX^{-1}H)]^{-1} = -[I - tX^{-1}H + \mathcal{O}(t^2)]X^{-1} \\ &= f'(X) + tX^{-1}HX^{-1} + \mathcal{O}(t^2) \end{aligned}$$

so $f''(X)[H] = X^{-1}HX^{-1}$ and $D^2 f(X)[H, G] = X^{-1}HX^{-1} \bullet G$.

- $f'''(X)[H, G] = -X^{-1}HX^{-1}GX^{-1} - X^{-1}GX^{-1}HX^{-1}$

Characterization of self-concordance for SDP

Theorem

The function $f(X) = -\ln\det X$ *is a convex barrier for* $\mathbb{S}^n_+$.

Proof sketch.

Let $\varphi(t) = F(X + tH)$. Then, prove that $\varphi''(t) \geq 0$ for $t > 0$ such that $X + tH \succ 0$. Therefore, when $X \succ 0$ approaches a singular matrix, its determinant approaches zero, and the function $f(X) \to +\infty$. □

Theorem ([Nestervov and Nemirovskii, 1994])

The barrier function $f(X) = -\ln\det X$ *is self-concordant on* $\mathbb{S}^n_+$.

Solving SDPs with IPMs

- Replace the primal SDP

$$\begin{aligned} \min \quad & C \bullet X \\ \text{s.t.} \quad & \mathcal{A}X = b, \\ & X \succeq 0, \end{aligned}$$

  with the primal barrier SDP

$$\begin{aligned} \min \quad & C \bullet X + \mu f(X) \\ \text{s.t.} \quad & \mathcal{A}X = b, \end{aligned}$$

  (with a barrier parameter $\mu \geq 0$).
- Formulate the Lagrangian

$$L(X, y, S) = C \bullet X + \mu f(X) - y^T(\mathcal{A}X - b),$$

  with $y \in \mathcal{R}^m$, and write the first order conditions (FOC) for a stationary point of L:

$$C + \mu f'(X) - \mathcal{A}^* y = 0$$

Solving SDPs with IPMs (cont'd)

- Use $f(X) = -\ln \det X$ and $f'(X) = -X^{-1}$ to obtain

$$C - \mu X^{-1} - \mathcal{A}^* y = 0$$

- Denote $S = \mu X^{-1}$, i.e., $XS = \mu I$. Then, the FOC can be written as

$$\begin{aligned} \mathcal{A} X &= b \\ A^* y + S &= C \\ XS &= \mu I \end{aligned}$$

with $X, S \in \mathbb{S}^n_{++}$.

Newton direction

The differentiation in the above system is a nontrivial operation. The direction is the solution of the system:

$$\begin{bmatrix} \mathcal{A} & 0 & 0 \\ 0 & \mathcal{A}^* & \mathcal{I} \\ \mu\left(X^{-1} \odot X^{-1}\right) & 0 & \mathcal{I} \end{bmatrix} \cdot \begin{bmatrix} \Delta X \\ \Delta y \\ \Delta S \end{bmatrix} = \begin{bmatrix} \xi_b \\ \xi_C \\ \xi_\mu \end{bmatrix}.$$

We introduce a useful notation $P \odot Q$ for $n \times n$ matrices P and Q is the Kronecker product. This defines a linear operator from $\mathbb{S}^n$ to $\mathbb{S}^n$ given by

$$(P \odot Q)U = \frac{1}{2}\left(PUQ^T + QUP^T\right).$$

Interior point methods bird-view

- Logarithmic barrier functions for SOCP and SDP Self-concordant barriers
  - polynomial complexity (predictable behaviour)
- Unified view of optimization
  - from LP via QP to NLP, SOCP and SDP
- Efficiency
- good for SOCP
- problematic for SDP because solving the problem of size n involves linear algebra operations in dimension n^2
  - and this requires n^6 flops!

Thanks for your attention!

Check my webpage

https://lrsantos11.github.io.

@lrsantos11

l.r.santos@ufsc.br / lrsantos11@gmail.com