Interior Points Methods

V. Leclère (ENPC)

May 31, 2024

Why should I bother to learn this stuff?

- Interior point methods are competitive with simplex methods for linear programs
- Interior point methods are state of the art for most conic (convex) problems (inc. SOCP, SDP, etc.)
- $\Longrightarrow$ useful for
  - understanding what is used in numerical solvers (like MOSEK, Gurobi, SCS, etc.)
  - specialization in optimization

Contents

Convex differentiable optimization problem

We consider the following convex optimization problem

$$
\begin{aligned}
(\mathcal{P}) \quad \min_{x \in \mathbb{R}^n} \quad & f(x) \\
\text{s.t.} \quad & Ax = b \\
& h_i(x) \leq 0 \qquad\qquad \forall i \in [n_I]
\end{aligned}
$$

where A is a $n_E \times n$ matrix, and all functions f and h_i are assumed convex, real valued and twice differentiable.

Introducing the Lagrangian

$$
\begin{aligned}
(\mathcal{P}) \qquad \min_{x \in \mathbb{R}^n} \quad & f(x) \\
\text{s.t.} \quad & Ax = b \\
& h_i(x) \leq 0 \qquad\qquad \forall i \in [n_I]
\end{aligned}
$$

is equivalent to

$$
\min_{x \in \mathbb{R}^n} \quad f(x) + \mathbb{I}_{\{Ax-b=0\}} + \sum_{i=1}^{n_I} \mathbb{I}_{\{h_i(x) \leq 0\}}
$$

which we rewrite

$$
\min_{x \in \mathbb{R}^n} \quad f(x) + \sup_{\lambda \in \mathbb{R}^{n_E}} \lambda^\top (Ax - b) + \sum_{i=1}^{n_I} \sup_{\mu_i \geq 0} \mu_i h_i(x)
$$

Introducing the Lagrangian

$$
(\mathcal{P}) \quad \min_{x \in \mathbb{R}^n} \quad f(x)
$$

$$
\text{s.t.} \quad Ax = b
$$

$$
h_i(x) \leq 0 \qquad \forall i \in [n_I]
$$

is equivalent to

$$
\min_{x \in \mathbb{R}^n} \quad f(x) + \mathbb{I}_{\{Ax-b=0\}} + \sum_{i=1}^{n_I} \mathbb{I}_{\{h_i(x)\leq 0\}}
$$

which we rewrite

$$
\min_{x \in \mathbb{R}^n} \sup_{\lambda \in \mathbb{R}^{n_E}, \mu \in \mathbb{R}_+^{n_I}} \quad f(x) + \lambda^\top (Ax - b) + \sum_{i=1}^{n_I} \mu_i h_i(x)
$$

Introducing the Lagrangian

$$(\mathcal{P}) \quad \min_{x\in\mathbb{R}^n} \sup_{\lambda\in\mathbb{R}^{n_E},\mu\in\mathbb{R}_+^{n_I}} \underbrace{f(x) + \lambda^\top(Ax-b) + \sum_{i=1}^{n_I}\mu_i h_i(x)}_{:=\mathcal{L}(x;\lambda,\mu)}$$

$$(\mathcal{D}) \quad \sup_{\lambda\in\mathbb{R}^{n_E},\mu\in\mathbb{R}_+^{n_I}} \min_{x\in\mathbb{R}^n} f(x) + \lambda^\top(Ax-b) + \sum_{i=1}^{n_I}\mu_i h_i(x)$$

As for any function ϕ we always have

$$\sup_y \inf_x \phi(x,y) \leq \inf_x \sup_y \phi(x,y)$$

we have that (weak duality)

$$val(\mathcal{D}) \leq val(\mathcal{P}).$$

Introducing the Lagrangian II ♡

$$(\mathcal{P}) \quad \min_{x\in\mathbb{R}^n} \sup_{\lambda\in\mathbb{R}^{n_E},\mu\in\mathbb{R}_+^{n_I}} \underbrace{f(x) + \lambda^\top(Ax-b) + \sum_{i=1}^{n_I} \mu_i h_i(x)}_{:=\mathcal{L}(x;\lambda,\mu)}$$

$$(\mathcal{D}) \quad \sup_{\lambda\in\mathbb{R}^{n_E},\mu\in\mathbb{R}_+^{n_I}} \min_{x\in\mathbb{R}^n} f(x) + \lambda^\top(Ax-b) + \sum_{i=1}^{n_I} \mu_i h_i(x)$$

As for any function ϕ we always have

$$\sup_y \inf_x \phi(x,y) \leq \inf_x \sup_y \phi(x,y)$$

we have that (weak duality)

$$val(\mathcal{D}) \leq val(\mathcal{P}).$$

Lower bounds from duality

Define the dual function

$$d(\lambda,\mu) := \inf_{x} \mathcal{L}(x;\lambda,\mu)$$

Then we have $val(\mathcal{D}) = \sup_{\lambda\in\mathbb{R}^{n_E},\mu\in\mathbb{R}^{n_I}_+} d(\lambda,\mu)$.

Thus, we can compute a lower bound to $val(\mathcal{D}) \leq val(\mathcal{P})$ by choosing any admissible dual points $\lambda \in \mathbb{R}^{n_E}, \mu \in \mathbb{R}^{n_I}_+$ and solving the unconstrained problem

$$d(\lambda,\mu) = \inf_{x\in\mathbb{R}^n} f(x) + \lambda^\top(Ax-b) + \sum_{i=1}^{n_I} \mu_i h_i(x)$$

Lower bounds from duality

Define the dual function

$$d(\lambda, \mu) := \inf_{x} \mathcal{L}(x; \lambda, \mu)$$

Then we have $val(\mathcal{D}) = \sup_{\lambda \in \mathbb{R}^{n_E}, \mu \in \mathbb{R}_+^{n_I}} d(\lambda, \mu)$.

Thus, we can compute a lower bound to $val(\mathcal{D}) \leq val(\mathcal{P})$ by choosing any admissible dual points $\lambda \in \mathbb{R}^{n_E}, \mu \in \mathbb{R}_+^{n_I}$ and solving the unconstrained problem

$$d(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} f(x) + \lambda^\top (Ax - b) + \sum_{i=1}^{n_I} \mu_i h_i(x)$$

Constraint qualification

Recall that, for a convex differentiable optimization problem, the constraints are qualified if *Slater's condition* is satisfied :

$$\exists x_0 \in \mathbb{R}^n, \qquad Ax_0 = b, \quad \forall i \in [n_I], \quad h_i(x_0) < 0$$

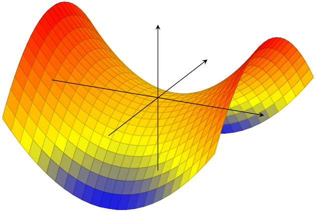
i.e., there exists a strictly admissible feasible point

Saddle point

♡

If $(\mathcal{P})$ is a convex optimization problem with qualified constraints, then

- $val(\mathcal{D}) = val(\mathcal{P})$
- any optimal solution $x^\sharp$ of $(\mathcal{P})$ is part of a saddle point $(x^\sharp; \lambda^\sharp, \mu^\sharp)$ of $\mathcal{L}$
- $(\lambda^\sharp, \mu^\sharp)$ is an optimal solution of $(\mathcal{D})$



Karush Kuhn Tucker conditions

If Slater's condition is satisfied, then $x^\sharp$ is an optimal solution to (P) if and only if there exists optimal multipliers $\lambda^\sharp \in \mathbb{R}^{n_E}$ and $\mu^\sharp \in \mathbb{R}^{n_I}$ satisfying

$$\begin{cases} \nabla f(x^\sharp) + A^\top \lambda^\sharp + \sum_{i=1}^{n_I} \mu_i^\sharp \nabla h_i(x^\sharp) = 0 & \text{first order condition} \\ Ax^\sharp = b & \text{primal admissibility} \\ h(x^\sharp) \leq 0 & \\ \mu^\sharp \geq 0 & \text{dual admissibility} \\ \mu_i^\sharp h_i(x^\sharp) = 0, \quad \forall i \in [n_I] & \text{complementarity} \end{cases}$$

The three last conditions are sometimes compactly written

$$0 \geq h(x^\sharp) \perp \mu \geq 0$$

Karush Kuhn Tucker conditions

If Slater's condition is satisfied, then $x^\sharp$ is an optimal solution to (P) if and only if there exists optimal multipliers $\lambda^\sharp \in \mathbb{R}^{n_E}$ and $\mu^\sharp \in \mathbb{R}^{n_I}$ satisfying

$$\begin{cases} \nabla f(x^\sharp) + A^\top \lambda^\sharp + \sum_{i=1}^{n_I} \mu_i^\sharp \nabla h_i(x^\sharp) = 0 & \text{first order condition} \\ Ax^\sharp = b & \text{primal admissibility} \\ h(x^\sharp) \leq 0 & \\ \mu^\sharp \geq 0 & \text{dual admissibility} \\ \mu_i^\sharp h_i(x^\sharp) = 0, \quad \forall i \in [n_I] & \text{complementarity} \end{cases}$$

The three last conditions are sometimes compactly written

$$0 \geq h(x^\sharp) \perp \mu \geq 0$$

Contents

Intuition for Newton's method: unconstrained case ♡

Newton's method is an iterative optimization method that minimizes a quadratic approximation of the objective function at the current point $x^{(k)}$.
Consider the following unconstrained optimization problem (f smooth):

$$\min_{x \in \mathbb{R}^n} f(x)$$

At $x^{(k)}$ we have

$$f(x^{(k)} + d) = f(x^{(k)}) + \nabla f(x^{(k)})^\top d + \frac{1}{2} d^\top \nabla^2 f(x^{(k)}) d + o(\|d\|^2)$$

And the direction $d^{(k)}$ minimizing the quadratic approximation is given by solving for d

$$\nabla f(x^{(k)}) + \nabla^2 f(x^{(k)}) d = 0.$$

Intuition for Newton's method: constrained case

Approximate the linearly constrained optimization problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

by

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & f(x^{(k)}) + \nabla f(x^{(k)})^\top d + \frac{1}{2} d^\top \nabla^2 f(x^{(k)}) d \\ \text{s.t.} \quad & A(x^{(k)} + d) = b \end{aligned}$$

Which is equivalent to solving (for given admissible $x^{(k)}$)

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & \nabla f(x^{(k)})^\top d + \frac{1}{2} d^\top \nabla^2 f(x^{(k)}) d \\ \text{s.t.} \quad & Ad = 0 \end{aligned}$$

Finding Newton's direction

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & \nabla f(x^{(k)})^\top d + \frac{1}{2} d^\top \nabla^2 f(x^{(k)}) d \\ \text{s.t.} \quad & Ad = 0 \end{aligned}$$

By KKT the optimal $d^{(k)}$ is given by solving for (d, λ)

$$\begin{cases} \nabla f(x^{(k)}) + \nabla^2 f(x^{(k)}) d + A^\top \lambda = 0 \\ Ad = 0 \end{cases}$$

Or in a matricial form

$$\begin{pmatrix} \nabla^2 f(x^{(k)}) & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x^{(k)}) \\ 0 \end{pmatrix}$$

Finding Newton's direction

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & \nabla f(x^{(k)})^\top d + \frac{1}{2} d^\top \nabla^2 f(x^{(k)}) d \\ \text{s.t.} \quad & Ad = 0 \end{aligned}$$

By KKT the optimal $d^{(k)}$ is given by solving for (d, λ)

$$\begin{cases} \nabla f(x^{(k)}) + \nabla^2 f(x^{(k)}) d + A^\top \lambda = 0 \\ Ad = 0 \end{cases}$$

Or in a matricial form

$$\begin{pmatrix} \nabla^2 f(x^{(k)}) & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x^{(k)}) \\ 0 \end{pmatrix}$$

Newton's algorithm: equality constrained case

Data: Initial admissible point x_0
Result: quasi-optimal point
$k = 0$;
while $\|\nabla f(x^{(k)}) + A^\top \lambda\| \geq \varepsilon$ **do**

Solve for d

$$\begin{pmatrix} \nabla^2 f(x^{(k)}) & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x^{(k)}) \\ 0 \end{pmatrix}$$

Line-search for $\alpha \in [0, 1]$ on $f(x^{(k)} + \alpha d^{(k)})$
$x^{(k+1)} = x^{(k)} + \alpha d^{(k)}$
$k \leftarrow k + 1$

Algorithm 1: Newton's algorithm

Contents

Video explanation

A short video introduction to the content of this and the next section.

`https://www.youtube.com/watch?v=MsgpSl5JRbI`

Contents

Constrained optimization problem

We now want to consider a convex differentiable optimization problem with equality and inequality constraints.

$$
\begin{aligned}
(\mathcal{P}_\infty) \qquad \min_{x\in\mathbb{R}^n} \quad & f(x) \\
\text{s.t.} \quad & Ax = b \\
& h_i(x) \leq 0 \qquad\qquad \forall i \in [n_I]
\end{aligned}
$$

where all functions f and h_i are assumed convex, finite valued and twice differentiable.

Which we rewrite

$$
\begin{aligned}
\min_{x\in\mathbb{R}^n} \quad & f(x) + \sum_{i=1}^{n_I} \mathbb{I}_{\mathbb{R}^-}(h_i(x)) \\
\text{s.t.} \quad & Ax = b
\end{aligned}
$$

Constrained optimization problem

We now want to consider a convex differentiable optimization problem with equality and inequality constraints.

$$
\begin{aligned}
(\mathcal{P}_\infty) \qquad \min_{x\in\mathbb{R}^n} \quad & f(x) \\
\text{s.t.} \quad & Ax = b \\
& h_i(x) \leq 0 & \forall i \in [n_I]
\end{aligned}
$$

where all functions f and h_i are assumed convex, finite valued and twice differentiable. Which we rewrite

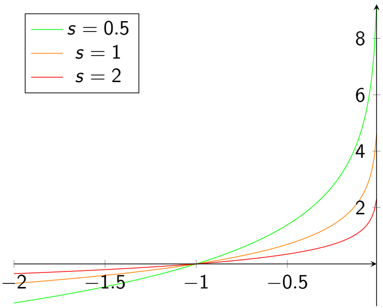
$$
\begin{aligned}
\min_{x\in\mathbb{R}^n} \quad & f(x) + \sum_{i=1}^{n_I} \mathbb{I}_{\mathbb{R}^-}(h_i(x)) \\
\text{s.t.} \quad & Ax = b
\end{aligned}
$$

The negative log function

- The idea of barrier method is to replace the indicator function $\mathbb{I}_{\mathbb{R}^-}$ by a smooth function.
- We choose the function $z \mapsto -1/s \log(-z)$
- Note that they also take value $+\infty$ on $\mathbb{R}^+$

➥ In particular these barrier functions ensure strict feasibility of the iterates.

Illustration of barrier functions



Calculus

- We define

$$\phi(x) := \begin{cases} -\sum_{i=1}^{n_I} \ln\big(-h_i(x)\big), & \text{if } h_i(x) < 0 \ \forall i \in [n_I], \\ +\infty, & \text{otherwise.} \end{cases}$$

- Thus we have $\frac{1}{s}\phi(x) \xrightarrow[s\to+\infty]{} \mathbb{I}_{\{h_i(x)<0,\ \forall i\in[n_I]\}}$
- On the domain $\{x : h_i(x) < 0\}$, we have:

$$\nabla\phi(x) =$$

$$\nabla^2\phi(x) =$$

Calculus

- We define

$$\phi(x) := \begin{cases} -\sum_{i=1}^{n_I} \ln\big(-h_i(x)\big), & \text{if } h_i(x) < 0 \ \forall i \in [n_I], \\ +\infty, & \text{otherwise.} \end{cases}$$

- Thus we have $\frac{1}{s}\phi(x) \xrightarrow[s \to +\infty]{} \mathbb{I}_{\{h_i(x)<0,\ \forall i \in [n_I]\}}$
- On the domain $\{x : h_i(x) < 0\}$, we have:

$$\nabla \phi(x) = \sum_{i=1}^{n_I} -\frac{1}{h_i(x)} \nabla h_i(x)$$

$$\nabla^2 \phi(x) =$$

Calculus

- We define

$$\phi(x) := \begin{cases} -\sum_{i=1}^{n_I} \ln\big(-h_i(x)\big), & \text{if } h_i(x) < 0 \ \forall i \in [n_I], \\ +\infty, & \text{otherwise.} \end{cases}$$

- Thus we have $\frac{1}{s}\phi(x) \xrightarrow[s\to+\infty]{} \mathbb{I}_{\{h_i(x)<0,\ \forall i\in[n_I]\}}$
- On the domain $\{x : h_i(x) < 0\}$, we have:

$$\nabla\phi(x) = \sum_{i=1}^{n_I} -\frac{1}{h_i(x)}\nabla h_i(x)$$

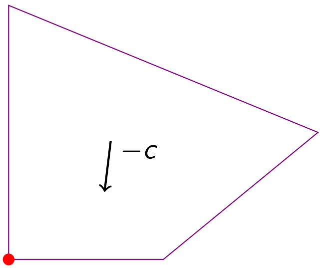
$$\nabla^2\phi(x) = \sum_{i=1}^{n_I} \left[\frac{1}{h_i^2(x)}\nabla h_i(x)\nabla h_i(x)^\top - \frac{1}{h_i(x)}\nabla^2 h_i(x)\right]$$

Penalized problem

We consider

$$(\mathcal{P}_\infty) \quad \min_{x \in \mathbb{R}^n} \; f(x)$$
$$\text{s.t.} \quad Ax = b$$

with optimal solution $x^\sharp$.



Penalized problem

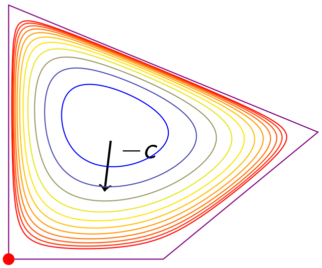
♡

We consider

$$(\mathcal{P}_s) \quad \min_{x \in \mathbb{R}^n} \; f(x) + \frac{1}{s}\phi(x)$$
$$\text{s.t.} \quad Ax = b$$

with optimal solution x_s.

scaling objective function by s does not change the solution.



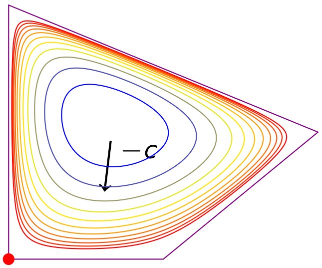
Penalized problem

♡

We consider

$$(\mathcal{P}_s) \quad \min_{x \in \mathbb{R}^n} s f(x) + \phi(x)$$
$$\text{s.t.} \quad Ax = b$$

with optimal solution x_s.



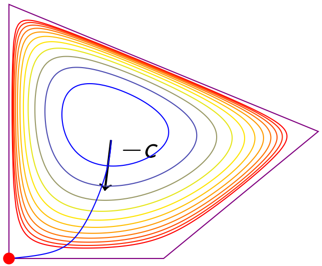
Letting s goes to $+\infty$ make x_s go to the solution of $(\mathcal{P})$ along the central path.

Penalized problem

We consider

$$(\mathcal{P}_s) \quad \min_{x \in \mathbb{R}^n} s f(x) + \phi(x)$$
$$\text{s.t.} \quad Ax = b$$

with optimal solution x_s.



Letting s goes to $+\infty$ make x_s go to the solution of $(\mathcal{P})$ along the central path.

Characterizing central path

x_s is solution of

$$(\mathcal{P}_s) \quad \min_{x \in \mathbb{R}^n} s f(x) + \phi(x)$$
$$\text{s.t.} \quad Ax = b$$

if and only if, there exists $\lambda_s \in \mathbb{R}^{n_E}$, such that

Characterizing central path

x_s is solution of

$$(\mathcal{P}_s) \quad \min_{x \in \mathbb{R}^n} s f(x) + \phi(x)$$
$$\text{s.t.} \quad Ax = b$$

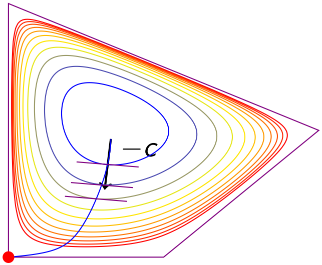
if and only if, there exists $\lambda_s \in \mathbb{R}^{n_E}$, such that

$$\begin{cases} Ax_s = b & \\ h_i(x_s) < 0 & \forall i \in [n_I] \\ s\nabla f(x_s) + \nabla\phi(x_s) + A^\top \lambda = 0 & \end{cases}$$

Characterizing central path II ♢

$$\begin{cases} Ax_s = b \\ h(x_s) < 0 \\ s\nabla f(x_s) + \nabla\phi(x_s) + A^\top \lambda = 0 \end{cases}$$

If $A = 0$ it means that $\nabla f(x_s)$ is orthogonal to the level lines of ϕ



Contents

Duality

Recall the original optimization problem

$$(\mathcal{P}_\infty) \qquad \min_{x\in\mathbb{R}^n} \quad f(x)$$
$$\text{s.t.} \quad Ax = b$$
$$h_i(x) \leq 0 \qquad \forall i \in [n_I]$$

with Lagrangian

$$\mathcal{L}(x;\lambda,\mu) := f(x) + \lambda^\top(Ax-b) + \sum_{i=1}^{n_I} \mu_i h_i(x)$$

and dual function

$$d(\lambda,\mu) := \inf_{x\in\mathbb{R}^n} \mathcal{L}(x;\lambda,\mu).$$

For any admissible dual point $(\lambda,\mu) \in \mathbb{R}^{n_E} \times \mathbb{R}_+^{n_I}$, we have

$$d(\lambda,\mu) \leq val(\mathcal{P}_\infty)$$

Duality

Recall the original optimization problem

$$(\mathcal{P}_\infty) \qquad \min_{x \in \mathbb{R}^n} \quad f(x)$$

$$\text{s.t.} \quad Ax = b$$

$$h_i(x) \leq 0 \qquad \forall i \in [n_I]$$

with Lagrangian

$$\mathcal{L}(x; \lambda, \mu) := f(x) + \lambda^\top (Ax - b) + \sum_{i=1}^{n_I} \mu_i h_i(x)$$

and dual function

$$d(\lambda, \mu) := \inf_{x \in \mathbb{R}^n} \mathcal{L}(x; \lambda, \mu).$$

For any admissible dual point $(\lambda, \mu) \in \mathbb{R}^{n_E} \times \mathbb{R}_+^{n_I}$, we have

$$d(\lambda, \mu) \leq val(\mathcal{P}_\infty)$$

Getting a lower bound

For given admissible dual point $(\lambda, \mu) \in \mathbb{R}^{n_E} \times \mathbb{R}_+^{n_I}$, a point $x^\sharp(\lambda, \mu)$ minimizing $\mathcal{L}(\cdot, \lambda, \mu)$, is characterized by first order conditions

$$\nabla f(x^\sharp(\lambda, \mu)) + A^\top \lambda + \sum_{i=1}^{n_I} \mu_i \nabla h_i(x^\sharp(\lambda, \mu)) = 0$$

which gives

$$d(\lambda, \mu) = \mathcal{L}(x^\sharp(\lambda, \mu); \lambda, \mu) \leq val(\mathcal{P}_\infty)$$

Dual point on the central path

Now recall that x_s, solution of $(\mathcal{P}_s)$, is characterized by

$$\begin{cases} Ax_s = b, h(x_s) < 0 \\ s\nabla f(x_s) + \nabla \phi(x_s) + A^\top \lambda_s = 0 \end{cases}$$

And we have seen that

$$\nabla \phi(x) = \sum_{i=1}^{n_I} \frac{1}{-h_i(x)} \nabla h_i(x)$$

Thus,

$$\nabla f(x_s) + A^\top \lambda_s/s + \sum_{i=1}^{n_I} \underbrace{\frac{1}{-sh_i(x_s)}}_{(\mu_s)_i} \nabla h_i(x_s) = 0$$

which means that $x_s = x^\sharp(\lambda_s/s, \mu_s)$.

Dual point on the central path

Now recall that x_s, solution of $(\mathcal{P}_s)$, is characterized by

$$\begin{cases} Ax_s = b, h(x_s) < 0 \\ s\nabla f(x_s) + \nabla\phi(x_s) + A^\top \lambda_s = 0 \end{cases}$$

And we have seen that

$$\nabla\phi(x) = \sum_{i=1}^{n_I} \frac{1}{-h_i(x)} \nabla h_i(x)$$

Thus,

$$\nabla f(x_s) + A^\top \lambda_s / s + \sum_{i=1}^{n_I} \underbrace{\frac{1}{-s h_i(x_s)}}_{(\mu_s)_i} \nabla h_i(x_s) = 0$$

which means that $x_s = x^\sharp(\lambda_s/s, \mu_s)$.

Bounding the error

Let x_s be a primal point on the central path satisfying

$$\exists \lambda_s \in \mathbb{R}^{n_E}, \qquad s\nabla f(x_s) + \nabla \phi(x_s) + A^\top \lambda_s = 0$$

We define a dual point $(\mu_s)_i = \frac{1}{-s h_i(x_s)} > 0$. We have

$$\begin{aligned} d(\mu_s, \lambda_s/s) &= \mathcal{L}(x_s, \mu_s, \lambda_s/s) \\ &= f(x_s) + \frac{1}{s}{\lambda_s}^\top \underbrace{(Ax_s - b)}_{=0} + \sum_{i=1}^{n_I} \frac{1}{-s h_i(x_s)} h_i(x_s) \\ &= f(x_s) - \frac{n_I}{s} \leq val(\mathcal{P}_\infty) \end{aligned}$$

➥ x_s is an n_I/s-optimal solution of $(\mathcal{P}_\infty)$.

Bounding the error

Let x_s be a primal point on the central path satisfying

$$\exists \lambda_s \in \mathbb{R}^{n_E}, \qquad s\nabla f(x_s) + \nabla \phi(x_s) + A^\top \lambda_s = 0$$

We define a dual point $(\mu_s)_i = \frac{1}{-s h_i(x_s)} > 0$. We have

$$\begin{aligned} d(\mu_s, \lambda_s/s) &= \mathcal{L}(x_s, \mu_s, \lambda_s/s) \\ &= f(x_s) + \frac{1}{s}\lambda_s{}^\top \underbrace{(Ax_s - b)}_{=0} + \sum_{i=1}^{n_I} \frac{1}{-s h_i(x_s)} h_i(x_s) \\ &= f(x_s) - \frac{n_I}{s} \leq val(\mathcal{P}_\infty) \end{aligned}$$

➥ x_s is an n_I/s-optimal solution of $(\mathcal{P}_\infty)$.

Contents

Interpretation through KKT condition ♡

A point x_s is on the central path iff it is strictly admissible and there exists $\lambda \in \mathbb{R}^{n_E}$ such that

$$\nabla f(x_s) + A^\top \lambda + \sum_{i=1}^{n_I} \underbrace{\frac{1}{-sh_i(x)}}_{(\mu_s)_i} \nabla h_i(x) = 0$$

which can be rewritten

$$\begin{cases} \nabla f(x) + A^\top \lambda + \sum_{i=1}^{n_I} \mu_i \nabla h_i(x) = 0 \\ Ax = b,\, h_i(x) \leq 0 \\ \mu \geq 0 \\ -\mu_i h_i(x) = \frac{1}{s} & \forall i \in [n_I] \end{cases}$$

Contents

Taking a step back

♡

- We saw that we can extend Newton's method to solve linearly constrained optimization problem.
- We saw that we can approximate inequality constraints through the use of logarithmic barrier $-1/s \sum_i \ln(-h_i(x))$.
- We proved that x_s is an n_I/s-optimal solution.
- The trade-off with s is : larger s means x_s closer to optimal solution x_∞ but the approximate problem $(\mathcal{P}_s)$ have worse conditionning.

Barrier method

Data: increase $\rho > 1$, error $\varepsilon > 0$,
initial s

Result: ε-optimal point

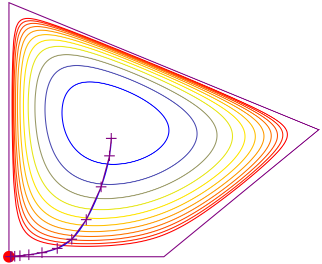
solve $(\mathcal{P}_s)$ and set $x = x_s$;

while $n_I/t \geq \varepsilon$ **do**

increase t: $t = \rho t$ centering step:

solve $(\mathcal{P}_s)$ starting at x ;

update : $x = x_s$



Barrier method

Data: increase $\rho > 1$, error $\varepsilon > 0$,
initial s

Result: ε-optimal point

solve $(\mathcal{P}_s)$ and set $x = x_s$;

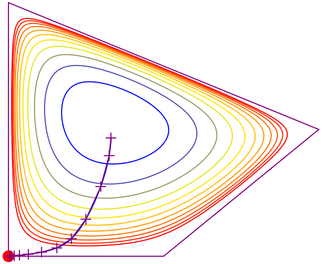
while $n_I / t \geq \varepsilon$ **do**

increase t*:* $t = \rho t$ *centering step:*

solve $(\mathcal{P}_s)$ starting at x ;

update : $x = x_s$

Question : why solve $(\mathcal{P}_s)$ to optimality ?



Solving $(\mathcal{P}_s)$ with Newton's method

$$
(\mathcal{P}_s) \quad \min_{x \in \mathbb{R}^n} \quad s f(x) + \phi(x) \\
\text{s.t.} \quad Ax = b
$$

is a linearly constrained optimization problem that can be solved by Newton's method. More precisely we have $x^{(k+1)} = x^{(k)} + d^{(k)}$ with $d^{(k)}$ a solution of

$$
\begin{pmatrix} s\nabla^2 f(x^{(k)}) + \nabla^2 \phi(x^{(k)}) & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -s\nabla f(x^{(k)}) - \nabla \phi(x^{(k)}) \\ 0 \end{pmatrix}
$$

Path following interior point method

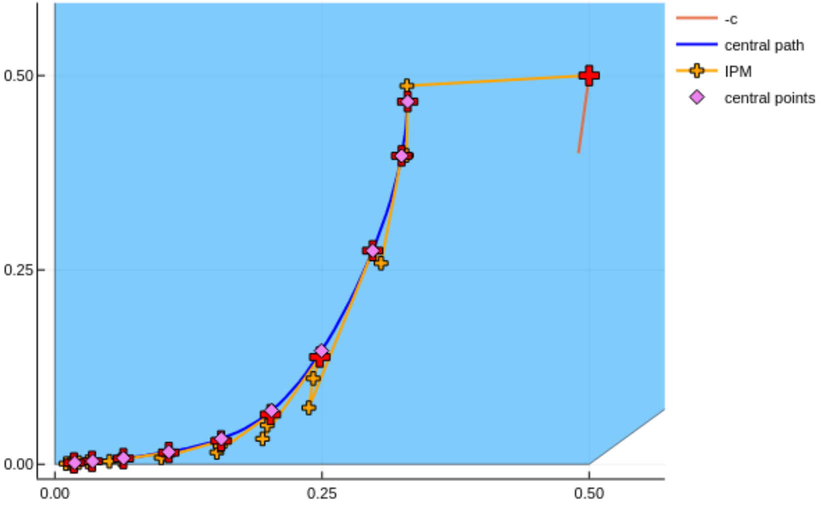
Data: increase $\rho > 1$, error $\varepsilon > 0$, initial s_0
initial strictly feasible point x_0
$k = 0$
$x \leftarrow x_0$, $s \leftarrow s_0$
for $k \in \mathbb{N}$ **do** // Outer step
  for $\kappa \in [K]$ **do** // Inner step
    solve for d ; // Newton step for $(\mathcal{P}_s)$

$$\begin{pmatrix} s_k \nabla^2 f(x) + \nabla^2 \phi(x) & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -s_k \nabla f(x) - \nabla \phi(x) \\ 0 \end{pmatrix}$$

    reduce α from 1 until $f(x + \alpha d) \leq f(x)$;
    $x \leftarrow x + \alpha d$;
  $s \leftarrow \rho s$;

Algorithm 2: Path following algorithm

Path following algorithm



Video explanation

A longer presentation to watch at a later time

`https://www.youtube.com/watch?v=zm4mfr-QT1E`

Contents

A linear problem - inequality form

We consider the following LP

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^\top x \\ \text{s.t.} \quad & a_i^\top x \leq b_i & & \forall i \in [n_I] \end{aligned}$$

Where $a_i^\top = A[i,:]$ is the row of matrix A, such that the constraints can be written $Ax \leq b$.
Thus, x_s is the solution of

A linear problem - inequality form

We consider the following LP

$$\min_{x \in \mathbb{R}^n} \quad c^\top x$$
$$\text{s.t.} \quad a_i^\top x \leq b_i \qquad \forall i \in [n_I]$$

Where $a_i^\top = A[i, :]$ is the row of matrix A, such that the constraints can be written $Ax \leq b$.
Thus, x_s is the solution of

$$\min_{x \in \mathbb{R}^n} \quad sc^\top x + \phi(x)$$

where

$$\phi(x) :=$$

A linear problem - inequality form

We consider the following LP

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^\top x \\ \text{s.t.} \quad & a_i^\top x \leq b_i & \forall i \in [n_I] \end{aligned}$$

Where $a_i^\top = A[i,:]$ is the row of matrix A, such that the constraints can be written $Ax \leq b$.
Thus, x_s is the solution of

$$\min_{x \in \mathbb{R}^n} \quad sc^\top x + \phi(x)$$

where

$$\phi(x) := -\sum_{i=1}^{n_I} \ln(b_i - a_i^\top x)$$

Calculus

$$\phi(x) = -\sum_{i=1}^{n_I} \ln(b_i - a_i^\top x)$$

$$\nabla \phi(x) =$$

$$\nabla^2 \phi(x) =$$

Calculus

$$\phi(x) = -\sum_{i=1}^{n_I} \ln(b_i - a_i^\top x)$$

$$\nabla\phi(x) = \sum_{i=1}^{n_I} \frac{1}{b_i - a_i^\top x} a_i$$

$$\nabla^2\phi(x) =$$

Calculus

$$\phi(x) = -\sum_{i=1}^{n_I} \ln(b_i - a_i^\top x)$$

$$\nabla\phi(x) = \sum_{i=1}^{n_I} \frac{1}{b_i - a_i^\top x} a_i$$

$$\nabla^2\phi(x) = \sum_{i=1}^{n_I} \frac{1}{(b_i - a_i^\top x)^2} a_i a_i^\top$$

Calculus

$$\phi(x) = -\sum_{i=1}^{n_I} \ln(b_i - a_i^\top x)$$

$$\nabla \phi(x) = \sum_{i=1}^{n_I} \frac{1}{b_i - a_i^\top x} a_i$$

$$\nabla^2 \phi(x) = \sum_{i=1}^{n_I} \frac{1}{(b_i - a_i^\top x)^2} a_i a_i^\top$$

This can be written in matrix form, using the vector $d \in \mathbb{R}^{n_I}$ defined by $d_i = \frac{1}{b_i - a_i^\top x}$

$$\nabla \phi(x) =$$

$$\nabla^2 \phi(x) =$$

Calculus

$$\phi(x) = -\sum_{i=1}^{n_I} \ln(b_i - a_i^\top x)$$

$$\nabla \phi(x) = \sum_{i=1}^{n_I} \frac{1}{b_i - a_i^\top x} a_i$$

$$\nabla^2 \phi(x) = \sum_{i=1}^{n_I} \frac{1}{(b_i - a_i^\top x)^2} a_i a_i^\top$$

This can be written in matrix form, using the vector $d \in \mathbb{R}^{n_I}$ defined by $d_i = \frac{1}{b_i - a_i^\top x}$

$$\nabla \phi(x) = A^\top d$$

$$\nabla^2 \phi(x) =$$

Calculus

$$\phi(x) = -\sum_{i=1}^{n_I} \ln(b_i - a_i^\top x)$$

$$\nabla\phi(x) = \sum_{i=1}^{n_I} \frac{1}{b_i - a_i^\top x} a_i$$

$$\nabla^2\phi(x) = \sum_{i=1}^{n_I} \frac{1}{(b_i - a_i^\top x)^2} a_i a_i^\top$$

This can be written in matrix form, using the vector $d \in \mathbb{R}^{n_I}$ defined by $d_i = \frac{1}{b_i - a_i^\top x}$

$$\nabla\phi(x) = A^\top d$$

$$\nabla^2\phi(x) = A^\top diag(d)^2 A$$

Newton step

Starting from x, the Newton direction for $(\mathcal{P}_s)$ is

$$dir_s(x) =$$

which, in algebraic form, yields

$$dir_s(x) =$$

with $d_i = 1/(b_i - a_i^\top x)$.

Newton step

Starting from x, the Newton direction for $(\mathcal{P}_s)$ is

$$dir_s(x) = -\left(\nabla^2 \phi(x)\right)^{-1}\left(sc + \nabla \phi(x)\right)$$

which, in algebraic form, yields

$$dir_s(x) =$$

with $d_i = 1/(b_i - a_i^\top x)$.

Newton step

Starting from x, the Newton direction for $(\mathcal{P}_s)$ is

$$dir_s(x) = -(\nabla^2\phi(x))^{-1}(sc + \nabla\phi(x))$$

which, in algebraic form, yields

$$dir_s(x) = -[A^\top diag(d)^2 A]^{-1}(sc + A^\top d)$$

with $d_i = 1/(b_i - a_i^\top x)$.

Newton step

Starting from x, the Newton direction for $(\mathcal{P}_s)$ is

$$dir_s(x) = -(\nabla^2\phi(x))^{-1}(sc + \nabla\phi(x))$$

which, in algebraic form, yields

$$dir_s(x) = -[A^\top diag(d)^2 A]^{-1}(sc + A^\top d)$$

with $d_i = 1/(b_i - a_i^\top x)$.

Theory tell us to use a step-size of 1 for Newton's method.

Newton step

Starting from x, the Newton direction for $(\mathcal{P}_s)$ is

$$dir_s(x) = -\left(\nabla^2\phi(x)\right)^{-1}\left(sc + \nabla\phi(x)\right)$$

which, in algebraic form, yields

$$dir_s(x) = -\left[A^\top diag(d)^2 A\right]^{-1}\left(sc + A^\top d\right)$$

with $d_i = 1/(b_i - a_i^\top x)$.

Theory tell us to use a step-size of 1 for Newton's method.

Practice teach us to use a smaller step-size (or linear-search).

Interior Point Method for LP pseudo code

Data: Initial admissible point x_0, initial penalization $s_0 > 0$;
parameter: $\rho > 1$, $N_{in} \geq 1$, $N_{out} \geq 1$;
Result: quasi-optimal point

$x = x0$, $s = s_0$;
for $k = 1..N_{out}$ **do**
    for $\kappa = 1..N_{in}$ **do**
        Compute d, with $d_i = 1/(b_i - a_i^T x)$;
        Solve for dir

$$A^\top \mathrm{diag}(d)^2 A\mathrm{dir} = -(sc + A^\top d)$$

        reduce α from 1 until[a] $f(x + \alpha\mathrm{dir}) \leq f(x)$;
        update $x \leftarrow x + \alpha\mathrm{dir}$;
    update $s \leftarrow \rho s$;

Algorithm 3: Interior Point Method for LP

[a] simplest condition described here

Contents

What you have to know

- IPM are state of the art algorithms for LP and more generally conic optimization problem
- That logarithmic barrier are a useful inner penalization method

What you really should know

- That Newton's algorithm can be applied with equality constraints
- What is the central path
- That IPM work with inner and outer optimization loop