# Interior Points Methods

V. Leclère (ENPC)

May 21st, 2021

# Why should I bother to learn this stuff ?

- Interior point methods are competitive with simplex method for linear programm
- Interior point methods are state of the art for most conic (convex) problems
- $\Longrightarrow$ useful for
  - understanding what is used in numerical solvers
  - specialization in optimization

# Contents

# Convex differentiable optimization problem

We consider the following convex optimization problem

$$
\begin{aligned}
(\mathcal{P}) \quad \min_{x \in \mathbb{R}^n} \quad & f(x) \\
\text{s.t.} \quad & Ax = b \\
& g_i(x) \leq 0 \qquad\qquad \forall i \in [\![1, n_I]\!]
\end{aligned}
$$

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

# Introducing the Lagrangian



$$
\begin{aligned}
(\mathcal{P}) \quad \min_{x \in \mathbb{R}^n} \quad & f(x) \\
\text{s.t.} \quad & Ax = b \\
& g_i(x) \leq 0 \qquad\qquad \forall i \in [\![1, n_I]\!]
\end{aligned}
$$

is equivalent to

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

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



$$
\begin{aligned}
(\mathcal{P}) \quad & \min_{x \in \mathbb{R}^n} \quad f(x) \\
& \text{s.t.} \quad Ax = b \\
& \qquad\ \ g_i(x) \leq 0 \qquad\qquad \forall i \in [\![1, n_I]\!]
\end{aligned}
$$

is equivalent to

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

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 II ♡

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

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

As for any function $\phi$ 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



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

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

As for any function $\phi$ 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 an 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 an 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 :

$$\exists x_0 \in \mathbb{R}^n, \qquad Ax_0 = b, \quad \forall i \in [\![1, n_I]\!], \quad g_i(x_0) < 0$$

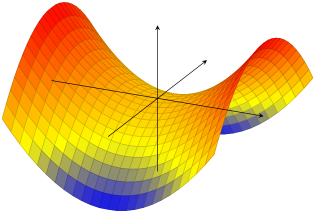
i.e.there exists a strictly admissible feasable point

# 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

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

The three last conditions are sometimes compactly written

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

# 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

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

The three last conditions are sometimes compactly written

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

# 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:

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

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 : eq. 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

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

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

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

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 $(d, \lambda)$

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

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** $|\nabla f(x^{(k)})| \geq \varepsilon$ **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 $f(x^{(k)} + \alpha d^{(k)})$
$x^{(k+1)} = x^{(k)} + \alpha d^{(k)}$
$k = k+1$

**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.

$$
\begin{aligned}
(\mathcal{P}_\infty) \qquad \min_{x\in\mathbb{R}^n} \quad & f(x) \\
\text{s.t.} \quad & Ax = b \\
& g_i(x) \leq 0 \qquad\qquad \forall i \in [\![1, n_I]\!]
\end{aligned}
$$

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

Which we rewrite

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

# Constrained optimization problem

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

$$(\mathcal{P}_\infty) \qquad \min_{x \in \mathbb{R}^n} \quad f(x)$$
$$\text{s.t.} \quad Ax = b$$
$$g_i(x) \leq 0 \qquad \forall i \in [\![1, n_I]\!]$$

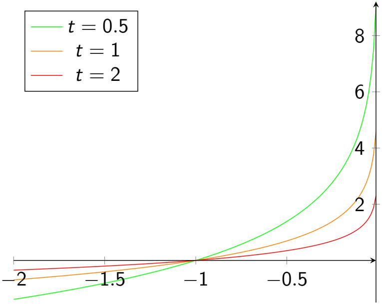
where all functions $f$ and $g_i$ are assumed convex, finite valued and twice differentiable.
Which we rewrite

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

# 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 $z \mapsto -1/t \log(-z)$
- Note that they also take value $+\infty$ on $\mathbb{R}^+$

Illustration of barrier functions



# Calculus

- We define

$$\phi : x \mapsto -\sum_{i=1}^{n_I} \ln(-g_i(x))$$

- Thus we have $\frac{1}{t}\phi(x) \xrightarrow[t\to+\infty]{} \mathbb{I}_{\{g_i(x)<0,\ \forall i\in[n_I]\}}$
- We have

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

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

# Calculus

- We define

$$\phi : x \mapsto -\sum_{i=1}^{n_I} \ln(-g_i(x))$$

- Thus we have $\frac{1}{t}\phi(x) \xrightarrow[t\to+\infty]{} \mathbb{I}_{\{g_i(x)<0,\ \forall i\in[n_I]\}}$
- We have

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

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

# Calculus

- We define
$$\phi : x \mapsto -\sum_{i=1}^{n_I} \ln(-g_i(x))$$
- Thus we have $\frac{1}{t}\phi(x) \xrightarrow[t\to+\infty]{} \mathbb{I}_{\{g_i(x)<0,\ \forall i\in[n_I]\}}$
- We have
$$\nabla\phi(x) = \sum_{i=1}^{n_I} -\frac{1}{g_i(x)}\nabla g_i(x)$$
$$\nabla^2\phi(x) = \sum_{i=1}^{n_I} \left[\frac{1}{g_i^2(x)}\nabla g_i(x)\nabla g_i(x)^\top - \frac{1}{g_i(x)}\nabla^2 g_i(x)\right]$$

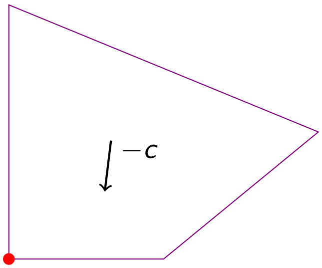
# Penalized problem

♡

We consider

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

with optimal solution $x^\sharp$.



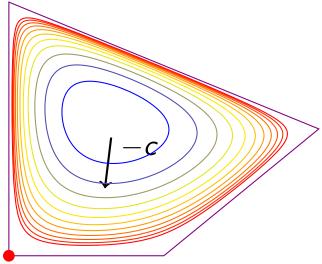
# Penalized problem

♡

We consider

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

with optimal solution $x_t$.



# Penalized problem

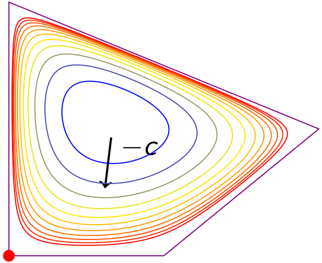
♡

We consider

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

with optimal solution $x_t$.

Letting $t$ goes to $+\infty$ get to solution of $(\mathcal{P})$ along the central path.



# Penalized problem

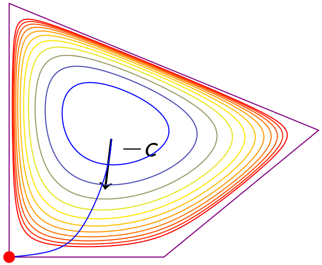
♡

We consider

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

with optimal solution $x_t$.

Letting $t$ goes to $+\infty$ get to solution of $(\mathcal{P})$ along the central path.



# Characterizing central path

$x_t$ is solution of

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

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

# Characterizing central path I $\diamondsuit$

$x_t$ is solution of

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

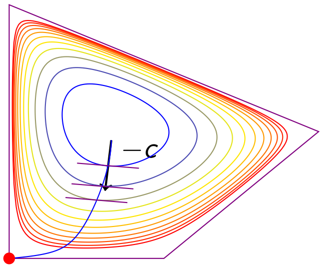
if and only if, there exists $\lambda_t \in \mathbb{R}^{n_E}$, such that

$$\begin{cases} Ax_t = b & \\ g_i(x_t) < 0 & \forall i \in [n_I] \\ t\nabla f(x_t) + \nabla\phi(x_t) + A^\top \lambda = 0 & \end{cases}$$

# Characterizing central path II ♢

$$\begin{cases} Ax_t = b \\ g(x_t) < 0 \\ t\nabla f(x_t) + \nabla \phi(x_t) + A^\top \lambda = 0 \end{cases}$$

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



# Contents

# Duality

Recall the original optimization problem

$$
\begin{aligned}
(\mathcal{P}_\infty) \qquad \min_{x\in\mathbb{R}^n} \quad & f(x) \\
\text{s.t.} \quad & Ax = b \\
& g_i(x) \leq 0 \qquad\qquad \forall i \in [\![1, n_I]\!]
\end{aligned}
$$

with Lagrangian

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

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

$$
\begin{aligned}
(\mathcal{P}_\infty) \qquad & \min_{x\in\mathbb{R}^n} \quad f(x) \\
& \text{s.t.} \quad Ax = b \\
& \qquad\quad g_i(x) \leq 0 \qquad\qquad \forall i \in [\![1, n_I]\!]
\end{aligned}
$$

with Lagrangian

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

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

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

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 $x_t$, solution of $(\mathcal{P}_t)$, is characterized by

$$\begin{cases} Ax_t = b, g(x_t) < 0 \\ t\nabla f(x_t) + \nabla \phi(x_t) + A^\top \lambda = 0 \end{cases}$$

And we have seen that

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

Thus,

$$\nabla f(x_t) + A^\top \lambda / t + \sum_{i=1}^{n_I} \underbrace{\frac{1}{-t g_i(x_t)}}_{(\mu_t)_i} \nabla g_i(x) = 0$$

which means that $x_t = x^\sharp(\lambda/t, \mu_t)$.

# Dual point on the central path

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

$$\begin{cases} Ax_t = b, g(x_t) < 0 \\ t\nabla f(x_t) + \nabla \phi(x_t) + A^\top \lambda = 0 \end{cases}$$

And we have seen that

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

Thus,

$$\nabla f(x_t) + A^\top \lambda / t + \sum_{i=1}^{n_I} \underbrace{\frac{1}{-t g_i(x_t)}}_{(\mu_t)_i} \nabla g_i(x) = 0$$

which means that $x_t = x^\sharp(\lambda / t, \mu_t)$.

# Bounding the error

Let $x_t$ be a primal point on the central path satisfying

$$\exists \lambda_t \in \mathbb{R}^{n_E}, \qquad t\nabla f(x_t) + \nabla \phi(x_t) + A^\top \lambda_t = 0$$

We define a dual point $(\mu_t)_i = \frac{1}{-t g_i(x_t)} > 0$. We have

$$\begin{aligned} d(\mu_t, \lambda_t/t) &= \mathcal{L}(x_t, \mu_t, \lambda_t/t) \\ &= f(x_t) + \frac{1}{t}\lambda_t^\top \underbrace{(Ax_t - b)}_{=0} + \sum_{i=1}^{n_I} \frac{1}{-t g_i(x_t)} g_i(x_t) \\ &= f(x_t) - \frac{n_I}{t} \leq val(\mathcal{P}_\infty) \end{aligned}$$

And in particular $x_t$ is an $n_I/t$-optimal solution of $(\mathcal{P}_\infty)$.

# Bounding the error

Let $x_t$ be a primal point on the central path satisfying

$$\exists \lambda_t \in \mathbb{R}^{n_E}, \qquad t\nabla f(x_t) + \nabla \phi(x_t) + A^\top \lambda_t = 0$$

We define a dual point $(\mu_t)_i = \frac{1}{-tg_i(x_t)} > 0$. We have

$$\begin{aligned} d(\mu_t, \lambda_t/t) &= \mathcal{L}(x_t, \mu_t, \lambda_t/t) \\ &= f(x_t) + \frac{1}{t}\lambda_t^\top \underbrace{(Ax_t - b)}_{=0} + \sum_{i=1}^{n_I} \frac{1}{-tg_i(x_t)} g_i(x_t) \\ &= f(x_t) - \frac{n_I}{t} \leq val(\mathcal{P}_\infty) \end{aligned}$$

And in particular $x_t$ is an $n_I/t$-optimal solution of $(\mathcal{P}_\infty)$.

# Contents

# Interpretation through KKT condition ♡

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

$$\nabla f(x_t) + A^\top \lambda + \sum_{i=1}^{n_I} \underbrace{\frac{1}{-tg_i(x)}}_{(\mu_t)_i} \nabla g_i(x) = 0$$

which can be rewritten

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

# Contents

# Taking a step back

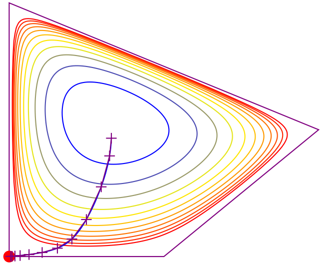
♡

- 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 $-1/t \sum_i \ln(-g_i(x))$.
- We proved that $x_t$ is an $n_I/t$-optimal solution.
- The trade-off with $t$ is : larger $t$ means $x_t$ closer to optimal solution $x_\infty$ but the approximate problem $(\mathcal{P}_t)$ have worse conditionning.

# Barrier method

♡

**Data:** increase $\rho > 1$, error $\varepsilon > 0$, initial $t$
**Result:** $\varepsilon$-optimal point
solve $(\mathcal{P}_t)$ and set $x = x_t$ ;
**while** $n_I/t \geq \varepsilon$ **do**
  *increase t:* $t = \rho t$
  *centering step:* solve $(\mathcal{P}_t)$ starting at $x$ ;
  *update :* $x = x_t$

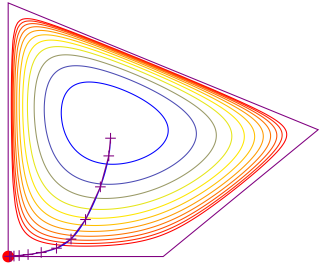


# Barrier method

♡

**Data:** increase $\rho > 1$, error $\varepsilon > 0$, initial $t$
**Result:** $\varepsilon$-optimal point
solve $(\mathcal{P}_t)$ and set $x = x_t$ ;
**while** $n_I/t \geq \varepsilon$ **do**
  *increase t:* $t = \rho t$
  *centering step:* solve $(\mathcal{P}_t)$ starting at $x$ ;
  *update :* $x = x_t$

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



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

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

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

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

# Path following interior point method

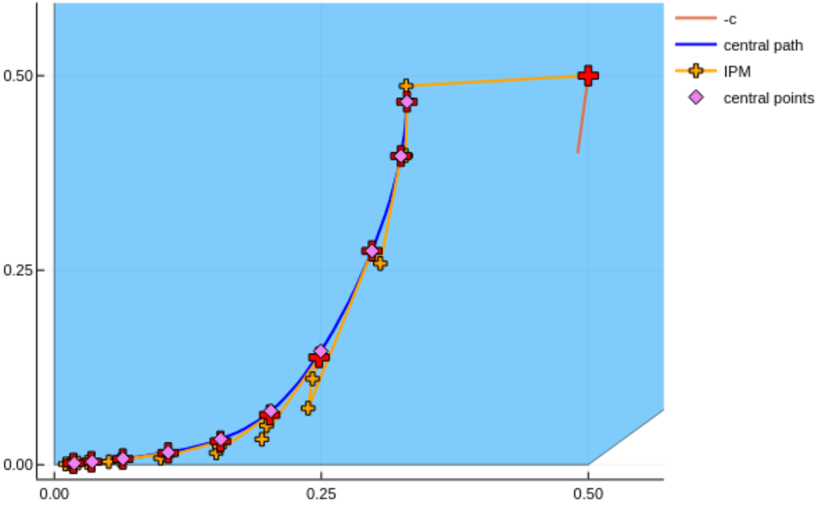
**Data:** increase $\rho > 1$, error $\varepsilon > 0$, initial $t_0$
initial strictly feasible point $x_0$
$k = 0$
**for** $k \in \mathbb{N}$ **do** `// Outer step`
  $x \leftarrow x_0$ , $t \leftarrow t_0$
  **for** $\kappa \in [K]$ **do** `// Inner step`
    solve for $d$ ; `// Newton step for` $(\mathcal{P}_t)$

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

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

**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, $x_t$ 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, $x_t$ is the solution of

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

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, $x_t$ is the solution of

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

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

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

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

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

# Calculus

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

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

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

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

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

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

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

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

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

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

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

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 $(\mathcal{P}_t)$ is

$$dir_t(x) =$$

which, in algebraic form, yields

$$dir_t(x) =$$

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

# Newton step

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

$$dir_t(x) = -(\nabla^2\phi(x))^{-1}(tc + \nabla\phi(x))$$

which, in algebraic form, yields

$$dir_t(x) =$$

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

# Newton step

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

$$dir_t(x) = -(\nabla^2\phi(x))^{-1}(tc + \nabla\phi(x))$$

which, in algebraic form, yields

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

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

# Newton step

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

$$dir_t(x) = -(\nabla^2\phi(x))^{-1}(tc + \nabla\phi(x))$$

which, in algebraic form, yields

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

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 $(\mathcal{P}_t)$ is

$$dir_t(x) = -(\nabla^2\phi(x))^{-1}(tc + \nabla\phi(x))$$

which, in algebraic form, yields

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

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

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

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 $t_0 > 0$;
**parameter:** $\rho > 1$, $N_{in} \geq 1$, $N_{out} \geq 1$;
**Result:** quasi-optimal point
$x = x0$, $t = t_0$;
**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 $\mathrm{dir}$

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

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

**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