$\boxed{\text{Topic 7}}$ Interior-point methods

+) Optimization problem with inequality constraints

$$\min f_0(x)$$
$$\text{s.t. } f_i(x) \le 0, \; i = 1 \ldots m$$
$$Ax = b$$

- $f_0, f_1 \ldots f_m$ are all convex and twice differentiable.
  $A \in \mathbb{R}^{p \times n}$, $p \le n$, rank $A = p$.
- Assume $x^*$ exists and is attained, $p^* = f_0(x^*)$.
- Assume also that problem is strictly feasible, i.e.
  $\exists \, \tilde{x} \in \mathrm{dom} f_0$ s.t. $f_i(\tilde{x}) < 0$, $A\tilde{x} = b$.
  $\to$ Strong duality holds.
- Note differentiability may require a reformulation of the problem.

+) The KKT optimality conditions:

$$Ax^* = b, \quad f_i(x^*) \le 0, \quad i = 1 \ldots m$$
$$\lambda^* \geq 0$$
$$\lambda^* f_i(x^*) = 0, \quad i = 1 \ldots m$$
$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x^*) + A^T \nu^* = 0$$

We will study an interior-point algorithm to solve this set of KKT conditions called the barrier method. The idea is to reduce the problem w/ inequality constraints to a sequence of problem with only equality constraints.

## t) Logarithmic barrier

- Reformulate the original problem as

$$\min \; f_0(x) + \sum_{i=1}^{m} I\_(f_i(x))$$
$$\text{s.t.} \quad Ax = b$$

where $I\_ : \mathbb{R} \to \mathbb{R}$ is the indicator function:

$$I\_(u) = \begin{cases} 0 & u \leq 0 \\ \infty & u > 0 \end{cases}$$

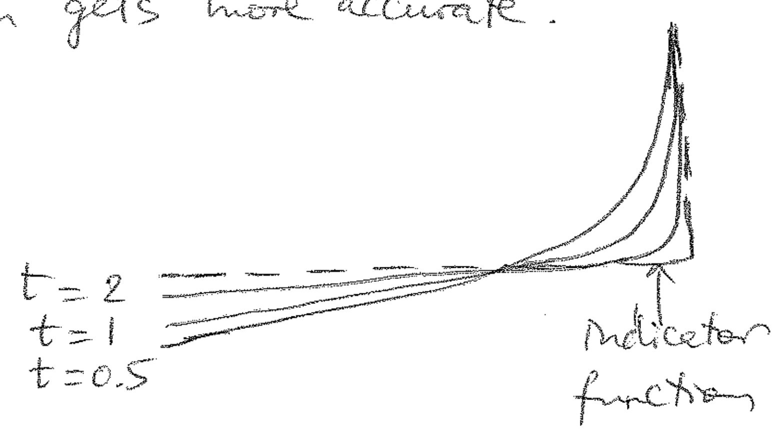
- The reformulation has no inequality constraints but the objective is not differentiable.

- We want to approximate $I\_$ by some other function $\hat{I}\_$ that is differentiable:

$$\hat{I}\_(u) = -\frac{1}{t}\log(-u), \quad \operatorname{dom} \hat{I}\_ = -\mathbb{R}_{++}.$$

where $t > 0$ is a parameter that affects the approximation. As $t \uparrow$, the approximation gets more accurate.

The function $\hat{I}\_(u)$ is differentiable.




- New approximated problem:

$$\min \; f_0(x) + \sum_{i=1}^{m} -\frac{1}{t}\log(-f_i(x))$$
$$\text{s.t.} \quad Ax = b$$

This problem is convex since $-\frac{1}{t}\log(-u)$ is convex in $u$ and increasing in $u$ (recall composition property).

- We can use Newton's method (or any descent method) to solve the approximated problem.

Let $\phi(x) = -\sum_{i=1}^{m} \log(-f_i(x))$

$\operatorname{dom} \phi = \{x \mid f_i(x) < 0 \ \forall i = 1 \dots m\}$.

then

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

$$\nabla^2\phi(x) = \sum_{i=1}^{m} \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T - \sum_{i=1}^{m} \frac{1}{f_i(x)} \nabla^2 f_i(x)$$

These can be used to solve the problem

$$\min \ t f_0(x) + \phi(x)$$
$$\text{s.t. } Ax = b$$

+) <u>Central path:</u>

- For each $t > 0$, define $x^*(t)$ to be the solution of

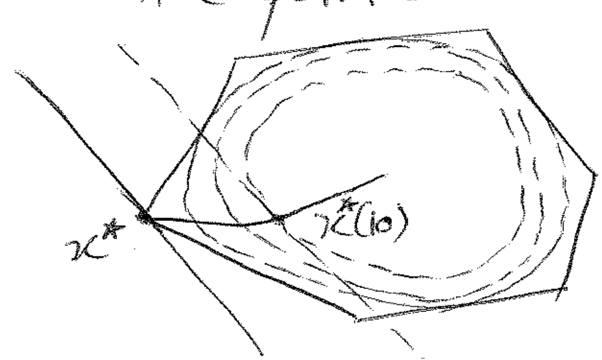
$$\min \ t f_0(x) + \phi(x)$$
$$\text{s.t. } Ax = b.$$

(Assume $x^*(t)$ exists and is unique for each $t$).

- The central path is the set of all $x^*(t)$ for $t > 0$. Each $x^*(t)$ is called a central point.

+) <u>Example:</u>

$$\min \ c^T x$$
$$\text{s.t. } Ax \preceq b$$
$$A \in \mathbb{R}^{6 \times n} \ (m = 6)$$



+) Necessary and sufficient conditions for central points:

$$A x^*(t) = b \;, \quad f_i(x^*(t)) < 0 \qquad \text{(strictly feasib}$$

$\exists \hat{\nu} \in \mathbb{R}^p$ s.t.:

$$0 = t \nabla f_0(x^*(t)) - \sum_{i=1}^{m} \frac{1}{f_i(x^*(t))} \nabla f_i(x^*(t)) + A^T\hat{\nu}$$

- Define $$\lambda_i^*(t) = -\frac{1}{t f_i(x^*(t))} \;, \quad \nu^*(t) = \frac{1}{t}\hat{\nu}$$

then $x^*(t)$ is the solution that minimizes the Lagrangia

$$\mathcal{L}(x, \lambda^*(t), \nu^*(t))$$
$$= f_0(x) + \sum \lambda_i^*(t) f_i(x) + (Ax-b)^T \nu^*(t)$$

- The dual function is finite:

$$g(\lambda^*(t), \nu^*(t)) = f_0(x^*(t)) + \underbrace{\sum_{i=1}^{m} \lambda_i^*(t) f_i(x^*(t))}_{= -\frac{m}{t}} + \underbrace{(Ax^*(t)-b)^T \nu}_{=0}$$

$\leftarrow$ (from definition of $\lambda_i^*(t)$)

$$= f_0(x^*(t)) - \frac{m}{t}$$

Thus the gap to optimality for each $t$ is

$$f_0(x^*(t)) - p^* \leq \frac{m}{t} .$$

This confirms that as $t \to \infty$, $x^*(t)$ will converge to the optimal point.

+) Interpretation of central path via KKT conditions:
At each $t$, $x = x^*(t)$, $\lambda = \lambda^*(t)$, $\nu = \nu^*(t)$ satisfy a continuous deformation of the KKT conditions as

$$\left.\begin{aligned} Ax = b,\ f_i(x) &\le 0 \\ \lambda &\succeq 0 \\ \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + A^T\nu &= 0 \end{aligned}\right\} \begin{array}{l}\text{(feasible conditions)} \\ \text{Same as in the original KKT.}\end{array}$$

$$-\lambda_i f_i(x) = \frac{1}{t} \quad \text{(approximate slackness)}$$

The only difference is the complementary slackness condition.

Thus each central point "almost" satisfies the original KKT and only violates the slackness constraint by $\frac{1}{t}$. As $t \to \infty$ the central point will be the optimal point.

+) Barrier method: Solve a sequence of unconstrained minimization problems (or with linear equality constraints only).
We increase $t$ at each problem until $t \geq \frac{m}{\varepsilon}$.

Given strictly feasible $x$, $t := t^{(0)}$, $\mu > 1$, tolerance $\varepsilon > 0$
repeat

1. Centering step
   Compute $x^*(t)$ by solving
   $$\begin{aligned} \min\ & t f_0 + \phi \\ \text{s.t. }\ & Ax = b \end{aligned}$$
   with starting point $x$.
2. Update: $x := x^*(t)$
3. Stopping criterion: quit if $\frac{m}{t} < \varepsilon$
4. Increase $t$: $t = \mu t$.

- For centering steps, usually use Newton's method, start at the current $x$. These are called inner iterations
- The choice of $\mu$ involves a tradeoff:
  - large $\mu$: few outer iterations but more Newton steps (inner iterations)
  - typical $\mu \approx 10-20$

- Each inner step produces a primal feasible point
At the end of each outer step we have a dual feasible p

- The choice of $t^{(0)}$ is also important.
  - too large $t^{(0)}$ makes the first outer iteration taking a long time.
  - too small $t^{(0)}$ will require extra outer iterations.

- Heuristic methods for picking $t^{(0)}$:
  - If a dual feasible point is known $(\lambda, \nu)$ then compute the duality gap $\eta = f_0(x^{(0)}) - g(\lambda, \nu)$ and use $t^{(0)} = m/$
  - Can also find $t^{(0)}$ (and $\nu$) that minimizes the central path condition

$$\min_{t,\nu} \| t \nabla f_0(x^{(0)}) + \nabla \phi(x^{(0)}) + A^T \nu \|_2$$

which is a quadratic minimization and has closed form solu or solved via a least square problem.

+) Convergence analysis:

$$\# \text{ outer (centering) iteration} = \left\lceil \frac{\log(m/(\varepsilon t^{(0)}))}{\log \mu} \right\rceil$$ exactl

For each centering step

$$\min \ t f_0(x) + \phi(x)$$
$$\text{s.t. } Ax = b$$

Each centering step will have the convergence analysis the same as of Newton's method (requires technical conditions)

- Can also use infeasible start for each centering step, the center point $x^*(t)$ will be feasible.

- Numerical evidence suggests that each centering step takes a nearly constant number of Newton steps.
  So centering step doesnot become more difficult as $t$ increases, since the previous step gives a good starting point for the next one.

See figures 11.4, 11.5, 11.6 in the text for example of convergence rate from reel problems. Also 11.7, 11.8.

Lecture 25

- For centering steps, usually use Newton's method, starting at the current $x$. These are called inner iterations.
- The choice of $\mu$ involves a tradeoff:
  - large $\mu$ : few outer iterations but more Newton steps (inner iterations)
  - typical $\mu \approx 10-20$.

- Each inner step produces a primal feasible point
At the end of each outer step we have a dual feasible point.

- The choice of $t^{(0)}$ is also important.
  - too large $t^{(0)}$ makes the first outer iteration taking a long time.
  - too small $t^{(0)}$ will require extra outer iterations.

- Heuristic methods for picking $t^{(0)}$:
  - If a dual feasible point is known $(\lambda, \nu)$ then compute the duality gap $\eta = f_0(x^{(0)}) - g(\lambda, \nu)$ and use $t^{(0)} = m/\eta$.
  - Can also find $t^{(0)}$ (and $\nu$) that minimizes the central path condition

$$\min_{t,\nu} \| t \nabla f_0(x^{(0)}) + \nabla \phi(x^{(0)}) + A^T \nu \|_2$$

which is a quadratic minimization and has closed form solution or solved via a least square problem.

+) Convergence analysis:

$$\# \text{ outer (centering) iteration} = \left\lceil \frac{\log(m/(\varepsilon t^{(0)}))}{\log \mu} \right\rceil \quad \text{exactly.}$$

For each centering step

$$\min \; t f_0(x) + \phi(x)$$
$$\text{s.t. } Ax = b$$

Each centering step will have the convergence analysis the same as of Newton's method (requires technical conditions)

- Can also use infeasible start for each centering step, the center point $x^*(t)$ will be feasible.

- Numerical evidence suggests that each centering step takes a nearly constant number of Newton steps.
  So centering step does not become more difficult as $t$ increases, since the previous step gives a good starting point for the next one.

See figures 11.4, 11.5, 11.6 in the text for example of convergence rate from real problems. Also 11.7, 11.8.

+) Feasibility and phase I methods

- The barrier method requires a strictly feasible starting point $x^{(0)}$
- If we do not know this point, then we need to find (compute) it! This stage is called phase I which proceed the barrier method.

Phase I - compute a strictly feasible point, or find that problem is infeasible.

Phase II - barrier method.

+) Feasibility problem: minimize the max infeasibility

$$\begin{aligned} \min \quad & s \\ \text{s.t.} \quad & f_i(x) \le s \quad i = 1 \dots m. \\ & Ax = b \end{aligned}$$

Each centering step will have the convergence analysis the same as of Newton's method (requires technical conditions)

- Can also use infeasible start for each centering step, the center point $x^*(t)$ will be feasible.

- Numerical evidence suggests that each centering step takes a nearly constant number of Newton steps.
So centering step doesnot become more difficult as t increases, since the previous step gives a good starting point for the next one.

See figures 11.4, 11.5, 11.6 in the text for example of convergence rate from real problems. Also 11.7, 11.8.

Lecture 24:

+) Feasibility and phase I methods

- The barrier method requires a strictly feasible starting point $x^{(0)}$
- If we donot know this point, then we need to find (compute) it. This stage is called phase I which proceed the barrier method.

Phase I - compute a strictly feasible point, or find that problem is infeasible.

Phase II - barrier method.

+) Feasibility problem: minimize the max infeasibility

$$\begin{aligned} \min \quad & s \\ \text{s.t.} \quad & f_i(x) \le s \quad i=1\ldots m. \\ & Ax = b \end{aligned}$$

This feasibility problem is always feasible:

- Assume we are given $x^{(0)}$ such that

$$Ax^{(0)} = b$$

and $x^{(0)} \in \mathrm{dom} f_1 \cap \mathrm{dom} f_2 \dots \cap \mathrm{dom} f_m$.

- Then start with $x^{(0)}$, take $s > \max\{f_i(x^{(0)})\}$
$\rightarrow$ we got a strictly feasible starting point!

Thus we can apply the barrier method on this problem

- Three cases of optimal $\bar{p}^*$ of the feasibility problem
  - (i) if $\bar{p}^* < 0$: then we have a strictly feasible point for the original problem.
  - (ii) if $\bar{p}^* > 0$: original problem is infeasible. Can construct a dual feasible point with positive dual objective to prove it
  - (iii) if $\bar{p}^* = 0$: If $\bar{p}^*$ is attained and $s^* = 0$ then the set of inequality is feasible but not strictly feasible.
    If $\bar{p}^*$ is not attained $\rightarrow$ inequalities are infeasible.

In practice we cannot determine if $\bar{p}^* = 0$, but can only determine upto $|\bar{p}^*| < \varepsilon$ for some small $\varepsilon > 0$

+) Some variation of phase I method:

$$\min \ 1^T s \qquad \text{sum of infeasibility.}$$

$$\text{s.t.} \quad f_i(x) \leq s_i$$

$$Ax = b$$

$$s \succeq 0$$

Optimal value of $s$ is $0$ when the original system is feasible.

When the system is infeasible, sum of infeasibilities give the number of inequalities that are satisfied, and the number of inequalities that are infeasible.

This indication of number of feasible / infeasible inequalities is usually better than using the max infeasibility method.

+) Phase I via infeasible start Newton method: Rewrite the original as:

$$\min f_0(x)$$
$$\text{s.t. } f_i(x) \leq s \quad i = 1 \dots m$$
$$Ax = 0, \ s = 0$$

Then start the barrier method

$$\min t^{(0)} f_0(x) - \sum_{i=1}^{m} \log(s - f_i(x))$$
$$\text{s.t. } Ax = b, \ s = 0$$

and use the infeasible start Newton's method with any initial $x \in D$ and $s > \max_i f_i(x)$

Provided that the problem is strictly feasible, this infeasible start Newton's method will eventually take a full step and produce $s = 0$ and $x$ strictly feasible.

+) <u>Example of barrier / central path method:</u>

$$\min c^T x$$
$$\text{s.t. } Ax \leq b.$$

The log barrier

$$\phi(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x), \quad \operatorname{dom} \phi = \{x \mid Ax = b\}$$

+) Phace I: We need to solve the feasibility problem:

$$\begin{aligned} &\min \; s \\ &\text{s.t. } Ax \leq b + s.1 \end{aligned}$$

(basic phase I meth... ⇒ use barrier method

or solve

$$\min \; -\sum_{i=1}^{m} \log s_i$$

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

using infeasible start Newton's method. If problem is feasible it will produce $s > 0$ and $Ax < b$.
If the problem is on the boundary of feasibily and infeasibility, the computational complexity (# iterations) grows fast. If the problem is exactly feasible but not strictly feasible, the computational complexity is infinity.

+) Phase II: Assume that now we have identified a strict feasible point $x^{(0)}$.
Gradiant and Hessian of $\phi$:

$$\nabla \phi(x) = \sum_{i=1}^{m} \frac{1}{b_i - a_i^T x} a_i \;, \quad \nabla^2 \phi(x) = \sum_{i=1}^{m} \frac{1}{(b_i - a_i^T x)^2} a_i a_i^T$$

or as $\nabla \phi(x) = A^T d \;, \quad \nabla^2 \phi(x) = A^T \operatorname{diag}(d)^2 A$

where $d_i = \dfrac{1}{b_i - a_i^T x}, \quad d \succ 0$

The centrality condition is:

$$0 = tc + \nabla \phi(x) = tc + A^T d = 0$$

Geometric interpretation:

- $\nabla \phi(x)$ must be parallel to $(-c)$ (since $t > 0$)
- $\nabla \phi(x^*(t))$ normal to level set of $\phi$ through $x^*(t)$

→ $c^T x = c^T x^*(t)$ must be tangent to level set of $\phi$ through $x^*(t)$ (hyperplane)

+) Primal - dual interior point methods:

Primal-dual interior point method is another type of interior point methods.

- Only one loop or iteration - no distinction between inner and outer iterations
- At each iteration, both the primal and dual variables are updated, by solving the (modified) KKT equations directly.
- Primal and dual iterates are not necessarily feasible
- Often more efficient than barrier method, better than linear convergence, especially when high accuracy is required.
- Still a topic of research for non-linear convex probs.

+) Primal - dual search direction:

- Similar to the barrier method, we start with the modified KKT conditions

$$Ax = b, \quad f_i(x) \le 0 \quad i = 1 \ldots m$$

$$\lambda \succeq 0$$

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + A^T \nu = 0$$

$$-\lambda_i f_i(x) = \frac{1}{t}, \quad i = 1 \ldots m$$

Rewrite these as

$$0 = r_t(x, \lambda, \nu) = \begin{bmatrix} \nabla f_0(x) + Df(x)^T \lambda + A^T \nu \\ -\operatorname{diag}(\lambda) f(x) - \frac{1}{t} \mathbf{1} \\ Ax - b \end{bmatrix} \triangleq - \begin{bmatrix} r_{dual} \\ r_{cent} \\ r_{pri} \end{bmatrix}$$

residual vector

where $f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix}$ and $Df(x) = \begin{bmatrix} \nabla f_1(x)^T \\ \vdots \\ \nabla f_m(x)^T \end{bmatrix}$
$(f: \mathbb{R}^n \to \mathbb{R}^m)$

If $x, \lambda, \nu$ satisfy $r_t(x, \lambda, \nu) = 0$ then we obtain a central point

$$x = x^*(t), \quad \lambda = \lambda^*(t), \quad \nu = \nu^*(t)$$

which are primal and dual feasible with duality gap $\frac{m}{t}$.

- Our goal is to find the Newton step for solving this equation

$$r_t(x, \lambda, \nu) = 0.$$

- Denote $y = (x, \lambda, \nu)$ current point
  $\Delta y = (\Delta x, \Delta\lambda, \Delta\nu)$ Newton step

and linearize the KKT equation:

$$r_t(y + \Delta y) \approx r_t(y) + Dr_t(y) \cdot \Delta y = 0$$

we get

$$\begin{bmatrix} \nabla^2 f_0(x) + \sum_{i=1}^m \lambda_i \nabla^2 f_i(x) & Df(x)^T & A^T \\ -\operatorname{diag}(\lambda) Df(x) & -\operatorname{diag}(f(x)) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda \\ \Delta\nu \end{bmatrix} = - \begin{bmatrix} r_{dual} \\ r_{cent} \\ r_{pri} \end{bmatrix}$$

Solving the above linear equation gives the Newton step $\Delta y$ which includes both primal and dual updates.

- The search direction in PD method is similar to that in the barrier method but not the same.

- PD iterates $x^{(k)}, \lambda^{(k)}, \nu^{(k)}$ are not necessarily feasible, except in the limit as the algorithm converges.

+) Surrogate duality gap:

Since $x^{(k)}, \lambda^{(k)}, \nu^{(k)}$ are not necessarily feasible, we cannot easily evaluate a duality gap $\eta^{(k)}$.

Instead we define the surrogate duality gap, for any $x$ s.t. $f(x) < 0$ and $\lambda \geq 0$, as

$$\hat{\eta}(x, \lambda) = -f(x)^T \lambda$$

If $x$ and $\lambda, \nu$ were primal and dual feasible then $\hat{\eta}$ would be the duality gap.

+) Primal-dual interior-point method.

given $x$ strictly feasible ($f_i(x) < 0$), $\lambda > 0$, $\mu > 1$, $\varepsilon_{feas} > 0$, $\varepsilon > 0$

repeat

1. Determine $t$: Set $t = \frac{\mu m}{\hat{\eta}}$
2. Compute primal-dual search direction $\Delta y_{pd}$ (by solving the linear equation that approximate $r_t(x, \lambda, \nu) = 0$.
3. Line search and update: Determine step length $s > 0$ and set $y := y + s \Delta y_{pd}$

until $\|r_{pri}\|_2 \leq \varepsilon_{feas}$, $\|r_{dual}\|_2 \leq \varepsilon_{feas}$, and $\hat{\eta} < \varepsilon$

+ Line search: This is usually the standard backtracking line search, modified to ensure $\lambda > 0$ and $f(x) < 0$

∘ First, compute the largest positive step length ($\leq 1$) such that $\lambda^+ \geq 0$:

$x^+ = x + s\Delta x_{pd}$, $\lambda^+ = \lambda + s\Delta\lambda_{pd}$, $\nu^+ = \nu + s\Delta\nu_{pd}$.

Choose

$$s^{max} = \sup\{s \in [0,1] \mid \lambda + s\Delta\lambda \geq 0\}$$
$$= \min\left\{1, \min_i\left\{-\frac{\lambda_i}{\Delta\lambda_i} \mid \Delta\lambda_i < 0\right\}\right\}.$$

Then start backtracking with initial step as

$s = 0.99\, s^{max}$

and backtrack a $s := s\beta$ ($\beta \in (0,1)$) until

$$\| r_t(x^+, \lambda^+, \nu^+)\|_2 \leq (1 - \alpha s)\| r_t(x, \lambda, \nu)\|_2$$

Typical values for $\alpha, \beta$ are similar to before: $\alpha \in [0.01, 0.1]$, $\beta \in [0.3, 0.8]$.

- Each iteration of the PD method is the same as one step of the infeasible Newton method.

Each iteration solves $F_t(x, \lambda, \nu) = 0$ approximately, while ensuring $\lambda > 0$ and $f(x) < 0$.

- Examples of actual problems: See Fig. 11.21, 11.22, 11.23.