Recall ADMM algorithm for

$$\min_x \frac{1}{2}\|Ax-b\|^2 + \lambda g(x)$$

The algorithm is

$$\begin{cases} \text{(i)} \quad x^{k+1} = \arg\min_x \frac{1}{2}\|Ax-b\|^2 + u^T(x-v) + \frac{\rho}{2}\|x-v\|^2 \\ \text{(ii)} \quad v^{k+1} = \arg\min_v \lambda g(v) + u^T(x-v) + \frac{\rho}{2}\|x-v\|^2 \\ \text{(iii)} \quad u^{k+1} = u^k + \rho(x^{k+1} - v^{k+1}) \end{cases}$$

<u>Claim:</u> The iterations are equivalent to

$$\begin{cases} x^{k+1} = \arg\min_x \frac{1}{2}\|Ax-b\|^2 + \frac{\rho}{2}\|x-\tilde{x}^k\|^2, & \tilde{x}^k = v^k - u^k \\ v^{k+1} = \arg\min_v \lambda g(v) + \frac{\rho}{2}\|v-\tilde{v}^k\|^2, & \tilde{v}^k = x^{k+1} + u^k \end{cases}$$

<u>Proof:</u> (i) $\frac{1}{2}\|Ax-b\|^2 + u^T(x-v) + \frac{\rho}{2}\|x-v\|^2$

$$= \frac{1}{2}\|Ax-b\|^2 + \frac{\rho}{2}\left[\|x\|^2 - 2x^Tv + \|v\|^2 + \frac{2}{\rho}u^Tx\right] + \text{const.}$$

$$= \frac{1}{2}\|Ax-b\|^2 + \frac{\rho}{2}\left[\|x\|^2 - 2x^T\left(v-\frac{u}{\rho}\right) + \|v\|^2\right] + \text{const.}$$

$$= \frac{1}{2}\|Ax-b\|^2 + \frac{\rho}{2}\left[\|x\|^2 - 2x^T\left(v-\frac{u}{\rho}\right) + \|v-\frac{u}{\rho}\|^2 - \|v-\frac{u}{\rho}\|^2 + \|v\|^2\right]$$

$$= \frac{1}{2}\|Ax-b\|^2 + \frac{\rho}{2}\|x - \underbrace{\left(v-\frac{u}{\rho}\right)}_{\overset{\text{def}}{=}\tilde{x}^k}\|^2 + \text{const.}$$

(ii) $\lambda g(v) + u^T(x-v) + \frac{\rho}{2}\|x-v\|^2$

$$= \lambda g(v) + \frac{\rho}{2}\left[\|x\|^2 - 2x^Tv + \|v\|^2 - \frac{2}{\rho}u^Tx\right] + \text{const.}$$

$$= \lambda g(v) + \frac{\rho}{2}\left[\|x\|^2 - 2v^T\left(x+\frac{u}{\rho}\right) + \|v\|^2 - \|x+\frac{u}{\rho}\|^2 - \|x+\frac{u}{\rho}\|^2\right] + \text{const.}$$

$$= \lambda g(v) + \frac{\rho}{2}\|\underbrace{v - \left(x+\frac{u}{\rho}\right)}_{\overset{\text{def}}{=}\tilde{v}^k}\|^2 + \text{const}$$

(iii) $\frac{u^{k+1}}{\rho} = \frac{u^k}{\rho} + (x^{k+1} - v^{k+1})$.

Interpretation of the equivalent form:

$$x^{k+1} = \underset{x}{\operatorname{argmin}} \underbrace{\frac{1}{2}\|Ax-b\|^2}_{\text{data term}} + \underbrace{\frac{\rho}{2}\|x-\tilde{x}^k\|^2}_{\substack{\text{quadratic regularization} \\ \text{(or Gaussian prior)}}}$$

$$v^{k+1} = \underset{v}{\operatorname{argmin}} \underbrace{\lambda g(v)}_{\text{regularization}} + \underbrace{\frac{\rho}{2}\|v-\tilde{v}^k\|^2}_{\text{data term}}$$

For example, if $g(v) = \|v\|_{TV}$, then v-subproblem is Total-variation denoising.

Plug-and-Play ADMM (Bouman et al. 2013)

$$\begin{cases} x^{k+1} = \underset{x}{\operatorname{argmin}} \frac{1}{2}\|Ax-b\|^2 + \frac{\rho}{2}\|x-\tilde{x}^k\|^2 \\ v^{k+1} = \mathcal{D}(\tilde{v}^k) \quad \leftarrow \text{denoiser} \\ u^{k+1} = u^k + (x^{k+1} - v^{k+1}). \end{cases}$$

Denoiser Parameter:

$$\lambda g(v) + \frac{\rho}{2}\|v-\tilde{v}^k\|^2$$

$$= g(v) + \frac{1}{2\left(\frac{\lambda}{\rho}\right)}\|v-\tilde{v}^k\|^2$$

So the denoiser "strength" is $\sqrt{\frac{\lambda}{\rho}}$.

CNN-based ADMM (Zhang et al 2017)

$$\begin{cases} x^{k+1} = \underset{x}{\operatorname{argmin}} \frac{1}{2}\|Ax-b\|^2 + \frac{\rho}{2}\|x-\tilde{x}^k\|^2 \\ v^{k+1} = \mathrm{CNN}(\tilde{v}^k) \\ u^{k+1} = u^k + (x^{k+1} - v^{k+1}). \end{cases}$$

Interior Point Method

Goal: To solve an inequality constrained problem

$$(1) \quad \begin{aligned} &\min \ f_0(x) \\ &\text{s.t. } f_i(x) \leq 0 \quad , \ i = 1, 2, \dots, m. \\ &\qquad Ax = b \end{aligned}$$

Assumptions: (i) $f_0, f_1, \dots, f_m$ are twice differentiable, convex

(ii) $A \in \mathbb{R}^{p \times n}$, $\text{rank}(A) = p$.

(iii) x^* exists.

(iv) slater condition holds, so dual optimal λ^*, ν^* exist.

The KKT conditions:

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

Barrier Function

Define an indicator function

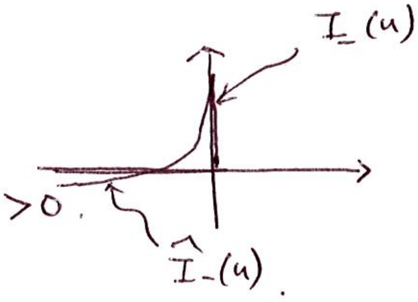
$$I_-(u) = \begin{cases} 0, & u \leq 0 \\ \infty, & u > 0. \end{cases}$$

Then write (1) as

equality constrained problem

$$\begin{aligned} &\min \ f_0(x) + \sum_{i=1}^{m} I_-(f_i(x)) \\ &\text{s.t. } Ax = b. \end{aligned}$$

To approximate $I_-(u)$, we choose



$$\hat{I}_-(u) = -\left(\frac{1}{t}\right)\log(-u), \quad t>0.$$

(i) $\hat{I}_-(u)$ is convex; non-decreasing

(ii) goes to ∞ as $u \to 0$.

Therefore, the optimization becomes

$$(2) \text{ ——— } \begin{array}{ll} \min & f_0(x) + \sum_{i=1}^{m} -\left(\frac{1}{t}\right)\log(-f_i(x)) \\ \text{s.t.} & Ax=b. \end{array}$$

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

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

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

Central Path

(2) can be written as

$$\begin{array}{ll} \min & t f_0(x) + \phi(x) \\ \text{s.t.} & Ax=b. \end{array} \text{ ——— } (3)$$

Let $x^*(t)$ be the solution of (3).

central point: $x^*(t)$

central path: $\{x^*(t) \mid t>0\}$.

Conditions for $x^*(t)$ to be a central point:

(i) $Ax^*(t)=b$, $f_i(x^*(t))<0$, $i=1,2,\dots,m$

(ii) $\exists\, \hat{v} \in \mathbb{R}^p$ s.t.

$$t\nabla f_0(x^*(t)) + \nabla\phi(x^*(t)) + A^T\hat{v} = 0.$$

Example $\min c^T x$

s.t. $Ax \le b$

The barrier function is

$$\phi(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x).$$

We can show that

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

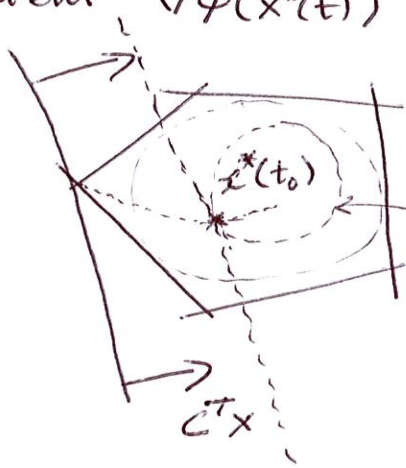
or more compactly if we define $d_i = \dfrac{1}{b_i - a_i^T x}$,

$$\nabla\phi(x) = A^T d, \qquad \nabla^2\phi(x) = A^T \mathrm{diag}(d) A.$$

The central path condition is then

$$tc + \sum_{i=1}^{m} \frac{1}{b_i - a_i^T x} a_i = tc + A^T d = 0$$

Geometric interpretation: For any central point $x^*(t)$, the gradient $\nabla\phi(x^*(t))$ must be parallel to $-c$.



barrier function at $t = t_0$

At $x^*(t_0)$, the normal of $\nabla\phi(x)$ is parallel to the normal of $f_0(x)$.

Dual point from central path

Central path condition implies

$$t \nabla f_0(x^*(t)) + \nabla \phi(x^*(t)) + A^T \hat{v} = 0$$

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

Define $\lambda_i = \dfrac{-1}{t f_i(x^*(t))}$, $i = 1, 2, \ldots, m$, $\quad v^*(t) = \dfrac{\hat{v}}{t}$.

Then central path condition is equivalent to

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

So we can define

$$\mathcal{L}(x, \lambda, v) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + v^T (Ax - b),$$

and $x^*(t)$ will minimize this Lagrangian, for $\lambda = \lambda^*(t)$, $v = v^*(t)$.

So $(\lambda^*(t), v^*(t))$ is a dual feasible pair. So the dual function $g(\lambda^*(t), v^*(t))$ is finite, and

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

$$= f_0(x^*(t)) + \sum_{i=1}^{m} \left(\frac{-1}{t}\right) + 0$$

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

Therefore, the dual gap is $\frac{m}{t}$, and

$$f_0(x^*(t)) - p^* \le \frac{m}{t} \longrightarrow 0 \text{ as } t \to \infty.$$

Interpretation via KKT:

(i) $\nabla f_0(x) + \sum \lambda_i \nabla f_i(x) + A^T v = 0$

(ii) $Ax = b$, $f_i(x) \le 0$

(iii) $\lambda \ge 0$

(iv) $\boxed{-\lambda_i f_i(x) = \frac{1}{t}}$

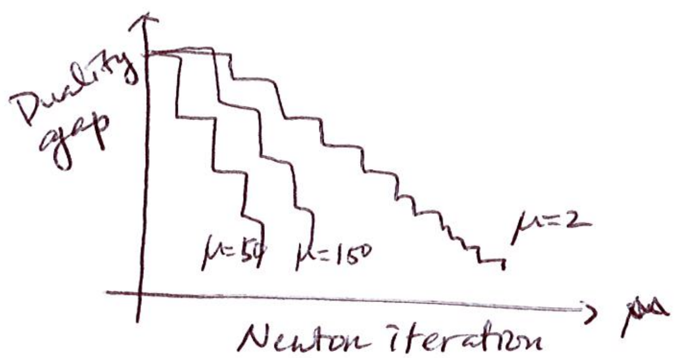
Complementary slackness is relaxed.

The Barrier Method

Given strictly feasible x, $t = t^{(0)} > 0$, $\mu > 1$ (parameter (gain factor)), $\epsilon > 0$.

repeat

1. centering step
   Compute $x^*(t)$ by minimizing $t f_0(x) + \phi(x)$
   s.t. $Ax = b$, starting at x.
2. Update $x = x^*(t)$
3. Stop if $m/t < \epsilon$
4. $t \longleftarrow \mu t$.



Convergence

Assuming that $t f_0 + \phi$ can be minimized by the Newton method for $t = t^{(0)}, \mu t^{(0)}, \mu^2 t^{(0)}, \ldots,$ then the duality gap after the initial centering step, and k additional centering steps, is

$$\frac{m}{\mu^k t^{(0)}}.$$

So if we want $\frac{m}{\mu^k t^{(0)}} < \epsilon$, we have

$$k > \frac{\log\left(\frac{m}{\epsilon t^{(0)}}\right)}{\log \mu}.$$

Newton Step

In the Barrier method, the Newton step Δx_{nt} is

$$\begin{bmatrix} t\nabla^2 f_0(x) + \nabla^2\phi(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{nt} \\ \Delta \nu_{nt} \end{bmatrix} = -\begin{bmatrix} t\nabla f_0(x) + \nabla\phi(x) \\ 0 \end{bmatrix}$$

Why? Start from the KKT conditions:

$$\begin{cases} \nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) + A^T\nu = 0 \\ -\lambda_i f_i(x) = \frac{1}{t} \\ Ax = b \end{cases}$$

$$\Rightarrow \begin{cases} \nabla f_0(x) + \sum_{i=1}^m \frac{1}{-tf_i(x)} \nabla f_i(x) + A^T\nu = 0 \\ Ax = b \end{cases}$$

Taylor expansion:

$$\nabla f_0(x+v) + \sum_{i=1}^m \lambda_i \nabla f_i(x+v)$$

$$= \nabla f_0(x+v) + \sum_{i=1}^m \frac{1}{-t f_i(x+v)} \nabla f_i(x+v)$$

$$\cong \underbrace{\nabla f_0(x) + \sum_{i=1}^m \frac{1}{-t f_i(x)} \nabla f_i(x)} + \underline{\nabla^2 f_0(x) v}$$

$$+ \underline{\sum_{i=1}^m \frac{1}{-tf_i(x)} \nabla^2 f_i(x) v} + \underline{\sum_{i=1}^m \frac{1}{t f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T v}$$

$$\Rightarrow \quad Hv + A^T\nu = -g, \quad Av = 0$$

Observe that $H = \nabla^2 f_0(x) + \frac{1}{t}\nabla^2\phi(x)$,

$g = \nabla f_0(x) + \frac{1}{t}\nabla\phi(x)$.

So we obtain the Newton step equation.

(if we not substitute λ into the stationarity equation, then we will obtain a primal-dual direction if we solve the KKT directly.)

Primal-Dual Interior Point Method

In Barrier method, the KKT conditions are:

$$
\begin{cases}
\text{(i)} & \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + A^T v = 0 \\
\text{(ii)} & Ax = b \,, \quad f_i(x) \leq 0 \\
\text{(iii)} & \lambda \geq 0 \\
\text{(iv)} & -\lambda_i f_i(x) = \frac{1}{t}
\end{cases}
$$

This set of conditions can be rewritten as

$$
r_t(x, \lambda, v) = \begin{bmatrix} \nabla f_0(x) + (Df)(x)^T \lambda + A^T v \\ -\operatorname{diag}(\lambda) f(x) - \left(\frac{1}{t}\right) 1 \\ Ax - b \end{bmatrix}
\begin{matrix} \overset{\text{def}}{=} r_{dual} \text{ (dual residue)} \\ \overset{\text{def}}{=} r_{cent} \text{ (centrality residue)} \\ \overset{\text{def}}{=} r_{pri} \text{ (primal residue)} \end{matrix}
$$

where $f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix}$ $\quad (Df)(x) = \begin{bmatrix} \nabla f_1(x)^T \\ \vdots \\ \nabla f_m(x)^T \end{bmatrix}$.

if (x, λ, v) satisfies $r_t(x, \lambda, t) = 0$, then $x = x^*(t)$, $\lambda = \lambda^*(t)$ and $v = v^*(t)$.

Now, consider Newton step for solving the nonlinear equation $\quad r_t(x, \lambda, v) = 0$.

Denote: current point $y = (x, \lambda, v)$

Newton step $\Delta y = (\Delta x, \Delta \lambda, \Delta v)$.

Newton step is characterized by

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

$$\Rightarrow \quad \Delta y = -Dr_t(y)^{-1} r_t(y).$$

So in terms of x, λ, ν, we have

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

The search direction $\Delta y = \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta \nu \end{bmatrix}$ is called the primal dual search direction.