Method of Multiplier

Consider an equality constrained problem

$$\min f(x) \quad \text{s.t. } h(x)=0. \qquad \text{——— (1)}$$

Define the augmented Lagrangian function

$$\mathcal{L}_c(x,\lambda) = f(x) + \lambda^T h(x) + \frac{c}{2}\|h(x)\|^2$$

c: penalty parameter
λ: multiplier vector.

Two special cases:
(i) $\lambda = 0 \Rightarrow$ quadratic penalty
(ii) $c = 0 \Rightarrow$ Lagrange function

Approaches to solve (1):

Approach A. Solve a sequence of problems of the form

$$\min \mathcal{L}_{c_k}(x, \lambda_k),$$

with $0 < c_k < c_{k+1}$, $\forall k$, $c_k \to \infty$.

And set $\lambda_k = 0 \ \forall k$.

This is called the penalty method

Approach B: On top of the penalty method, let

$$\lambda_{k+1} = \lambda_k + c_k h(x_k),$$

make $c_k =$ constant $\forall k$.

(Hestenes 1969, Powell 1969).

Why does this method (both A & B) work?

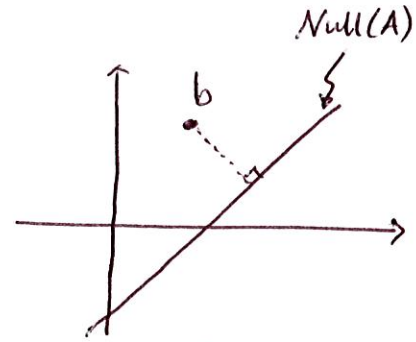
Bertsekas 1982

Let x_k be the global minimum of

$$\min L_{c_k}(x, \lambda_k),$$

where λ_k is bounded, and $c_k < c_{k+1}$ $\forall k$ and $c_k \to \infty$. Then the limit point of the sequence $\{x_k\}$ is also the global minimum of (1).

Example $\min \frac{1}{2}\|x-b\|^2$ s.t. $Ax = 0$



projecting b onto the null space of A.

E.g. $A = \boxed{}$

So there is non-trivial null space

The true solution:

$$\mathcal{L} = \frac{1}{2}\|x-b\|^2 + \lambda^T(Ax-0)$$

$$\Rightarrow \frac{\partial \mathcal{L}}{\partial x} = x - b + A^T\lambda = 0$$

So $x = b - A^T\lambda$

Want $Ax = 0$. So $Ax = Ab - AA^T\lambda = 0$

$$\Rightarrow \lambda = (AA^T)^{-1}Ab.$$

So
$$x^* = b - A^T(AA^T)^{-1}Ab = \left(I - A^T(AA^T)^{-1}A\right)b.$$

$f(x) + \lambda^T h(x) + \frac{c}{2}\|h(x)\|^2$ can set $\lambda = 0$.

The augmented Lagrangian solution:

$$\mathcal{L}_c = \frac{1}{2}\|x-b\|^2 + \lambda^T(Ax-0) + \frac{c}{2}\|Ax\|^2$$

Set $\lambda = 0$, $\mathcal{L}_c = \frac{1}{2}\|x-b\|^2 + \frac{c}{2}\|Ax\|^2$.

$$\Rightarrow \frac{\partial \mathcal{L}_c}{\partial x} = x - b + cA^TAx = 0$$

$$\Rightarrow x_c = (I + cA^TA)^{-1}b$$

Question: How close is x_c compared to x^*?

$$x_c = (I + cA^TA)^{-1}b = \left[I - A^T\left(\tfrac{1}{c}I + AA^T\right)^{-1}A\right]b$$

Sherman-Morrison formula

So
$$\|x_c - x^*\|^2 = \left\|A^T\left[\left(\tfrac{1}{c}I + AA^T\right)^{-1} - (AA^T)^{-1}\right]A\,b\right\|^2$$

$$\leq \|A\|_F^2\,\|b\|_2^2\,\left\|\left(\tfrac{1}{c}I + AA^T\right)^{-1} - (AA^T)^{-1}\right\|_F^2$$

Let $A = USV$ be the singular value decomposition of A.
Then

$$\left\| \left(\tfrac{1}{c}I + AA^T\right)^{-1} - (AA^T)^{-1} \right\|_F^2$$

$$= \left\| U\left(\left(\tfrac{1}{c}I + S^2\right)^{-1} - (S^2)^{-1}\right)V \right\|_F^2$$

Since S is diagonal, we consider the ith entry:

$$\left| \left(\tfrac{1}{c} + s_i^2\right)^{-1} - (s_i^2)^{-1} \right| = \left| \frac{1}{\frac{1}{c} + s_i^2} - \frac{1}{s_i^2} \right|$$

$$= \left| \frac{s_i^2 - \frac{1}{c} - s_i^2}{\left(\frac{1}{c} + s_i^2\right)(s_i^2)} \right|$$

$$= \left| \frac{\frac{1}{c}}{\left(\frac{1}{c} + s_i^2\right)(s_i^2)} \right| \leq \frac{\frac{1}{c}}{s_i^4} = \frac{1}{c s_i^4}$$

So $\left\| \left(\tfrac{1}{c}I + AA^T\right)^{-1} - (AA^T)^{-1} \right\|_F^2 \leq \|U\|_F^2 \|V\|_F^2 \left\| \left(\tfrac{1}{c}I + S^2\right)^{-1} - (S^2)^{-1} \right\|_F^2$

$$\leq \frac{1}{c^2 s_i^8}$$

$$\Rightarrow \quad \|x_c - x^*\|^2 \leq \frac{1}{c^2} \|A\|_F^2 \|b\|_2^2 \left(\sum_{i=1}^{n} \frac{1}{s_i^8} \right).$$

So As $c \longrightarrow \infty$, $\|x_c - x^*\|^2 \longrightarrow 0$.

Intuition:

As $c_k \to \infty$, x has to become a point such that $h(x)=0$.
When $h(x)=0$, $\lambda_k^T h(x_k) = 0$ no matter what λ_k we choose.
So $L_{c_k}(x, \lambda_k) \longrightarrow f(x)$. Hence by minimizing $L_c(x,\lambda)$ we also minimize $f(x)$.

The local minimum version is available. (See Bertsekas 1982)

Why approach B? Converges faster.

Consider a "primal functional"

$$P(u) = \min_x f(x), \quad \text{s.t. } h(x) = u. \qquad (2)$$

Property (i): $P(0) = f(x^*)$:

If $u=0$, then $P(0) = \min f(x)$ s.t. $h(x)=0$.

(ii) $\nabla P(0) = -\lambda^*$:

The Lagrange function of (2) is

$f(x) + \lambda h(x)$. Put $x \leftarrow x(u)$; $f(x(u)) + \lambda(u) h(x(u))$

$$\text{So} \quad \nabla_u f(x(u)) + \lambda^* \underbrace{\nabla_u h(x(u))}_{=\nabla_u u = 1} = 0$$

$$\Rightarrow \quad \nabla_u f(x(u)) = -\lambda^*.$$

$$\Rightarrow \quad \nabla P(0) = -\lambda^*.$$

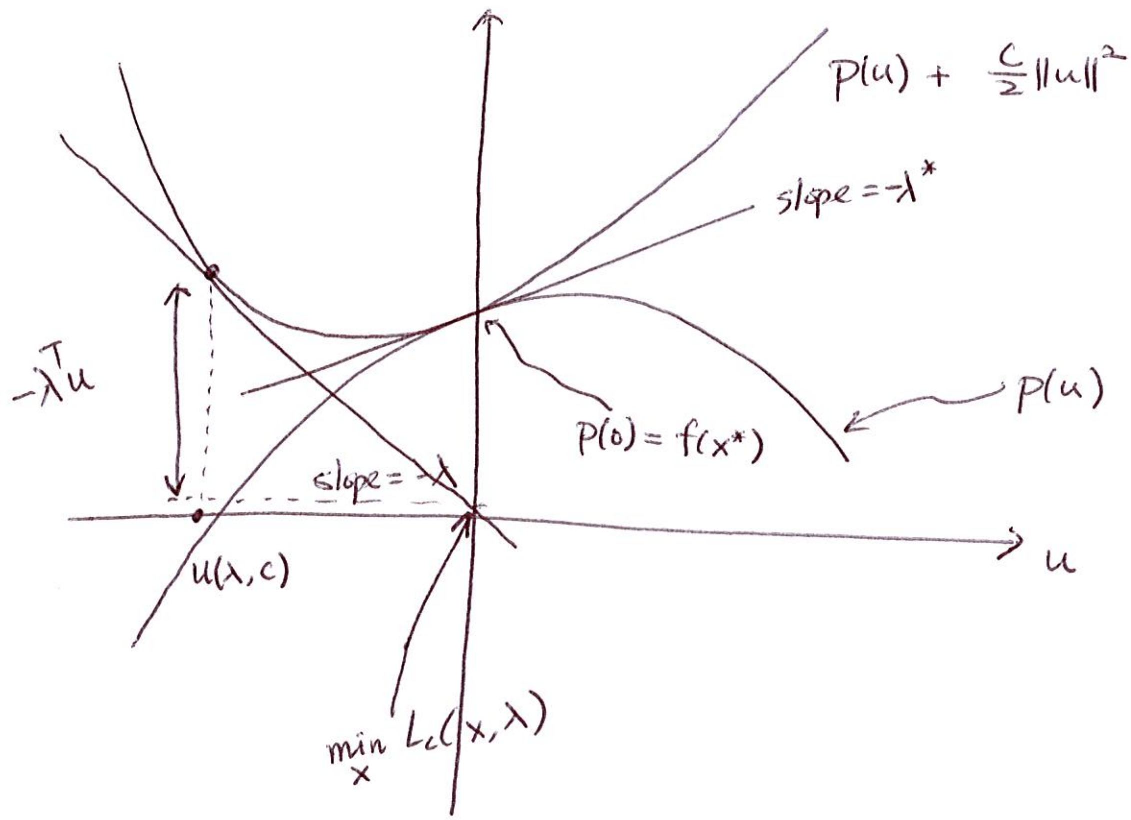
The minimization satisfies

$$\min_x L_c(x,\lambda) = \min_u \min_{h(x)=u} \left\{ f(x) + \lambda^T h(x) + \frac{c}{2}\|h(x)\|^2 \right\}$$

$$= \underbrace{\min_u \left\{ p(u) + \lambda^T u + \frac{c}{2}\|u\|^2 \right\}}_{\text{attains minimum when}}.$$

$$\nabla \left\{ p(u) + \lambda^T u + \frac{c}{2}\|u\|^2 \right\} = 0$$

$$\Longleftrightarrow \quad \nabla \left\{ p(u) + \frac{c}{2}\|u\|^2 \right\}\Big|_{u=u^*} = -\lambda$$



if we choose an arbitrary λ, then we can find $\min_x L_c(x,\lambda)$, which is the y-intersept of the straight line with slope $-\lambda$.
But if $\lambda = \lambda^*$, then we will obtain $f(x^*)$.

The update rule is

$$\lambda_{k+1} = \lambda_k + c_k h(x_k).$$

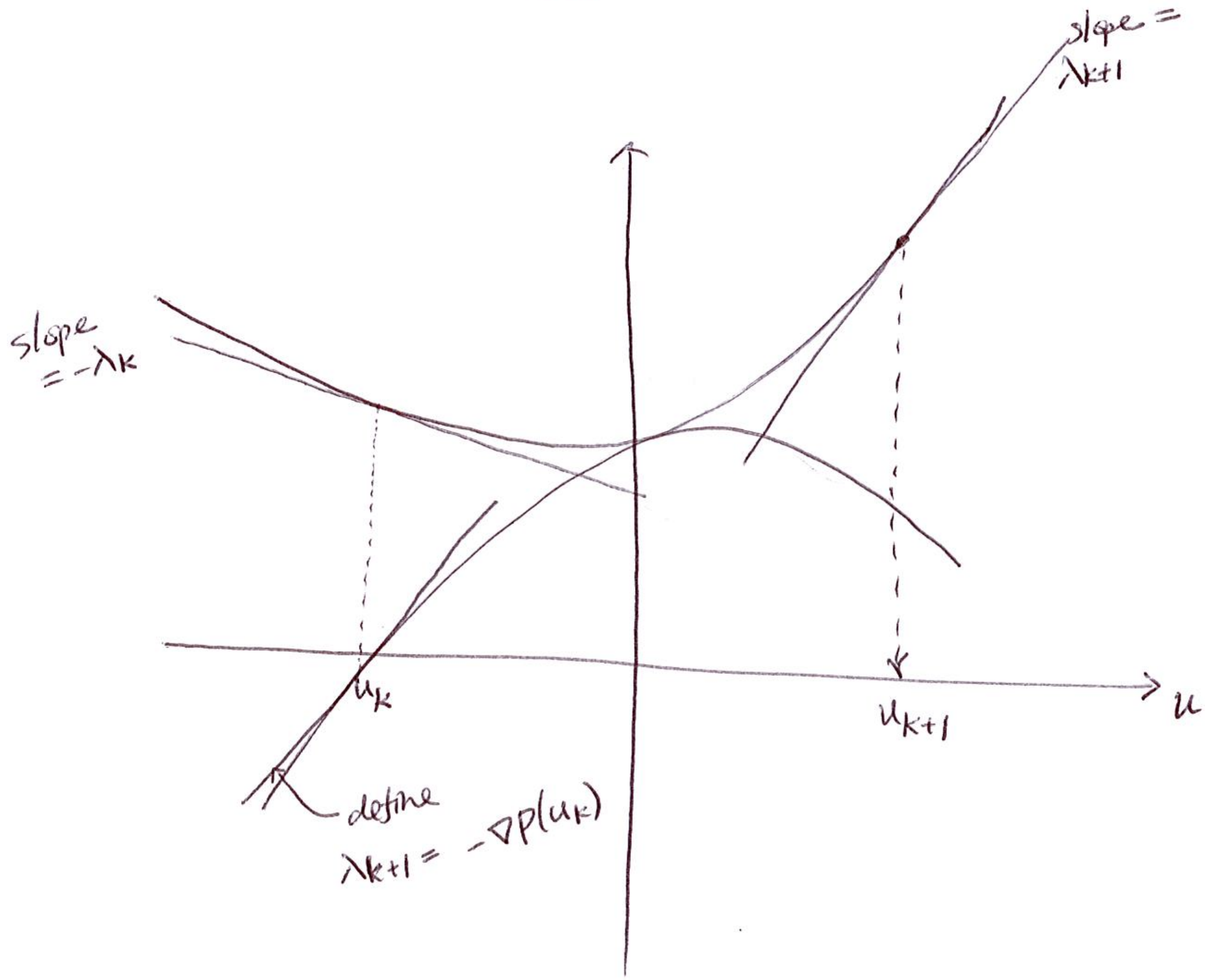
Note that

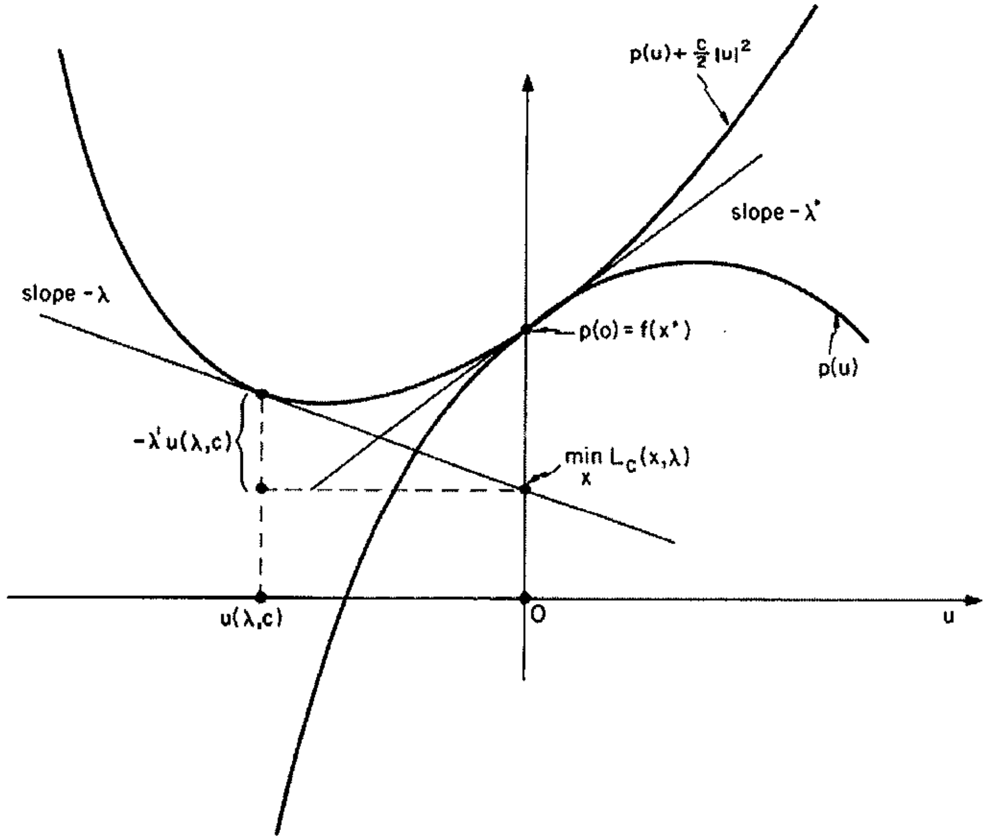
$$\nabla\left\{ p(u) + \frac{c}{2}\|u\|^2 \right\}\Big|_{u_k} = -\lambda_k$$

$$\Rightarrow \quad \nabla p(u_k) + c\, u_k = -\lambda_k$$

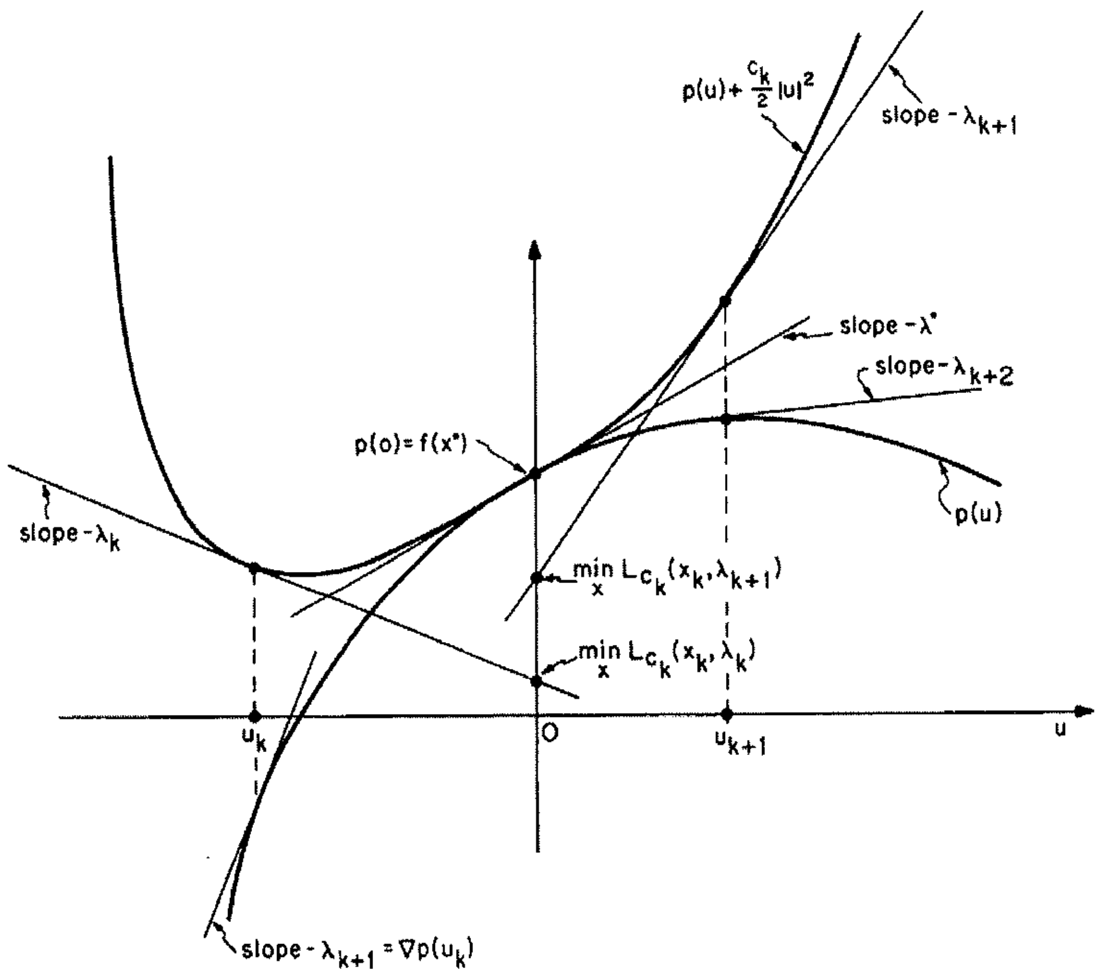
$$\Rightarrow \quad \nabla p(u_k) + c\, h(x_k) = -\lambda_k$$

$$\Rightarrow \quad \underbrace{\nabla p(u_k)}_{\overset{\text{def}}{=} -\lambda_{k+1}} = -\left[\lambda_k + c\, h(x_k)\right].$$





p(u)+ c/2 |u|^2
slope - λ*
slope - λ
p(o) = f(x*)
p(u)
-λ' u(λ,c)
min_x L_c(x,λ)
u(λ,c)
O
u



p(u)+ c_k/2 |u|^2
slope - λ_{k+1}
slope - λ*
slope - λ_{k+2}
p(o) = f(x*)
p(u)
slope - λ_k
min_x L_{c_k}(x_k, λ_{k+1})
min_x L_{c_k}(x_k, λ_k)
u_k
O
u_{k+1}
u
slope - λ_{k+1} = ∇p(u_k)

Augmented Lagrangian Method:

$$\min_{x} \frac{1}{2}\|Ax-b\|^2 + \alpha\|x\|_1$$

This unconstrained problem is equivalent to

$$\min_{x,v} \frac{1}{2}\|Ax-b\|^2 + \alpha\|v\|_1$$
$$\text{s.t.} \quad x - v = 0.$$

The augmented Lagrangian function is

$$\mathcal{L}(x,v,\lambda,c) = \frac{1}{2}\|Ax-b\|^2 + \alpha\|v\|_1 + \lambda^T(x-v) + \frac{c}{2}\|x-v\|^2$$

So the algorithm is

$$\begin{cases} (x^{k+1}, v^{k+1}) = \arg\min_{(x,v)} \mathcal{L}(x,v,\lambda^k,c) & \text{———— (1)} \\ \lambda^{k+1} = \lambda^k + c(x^{k+1} - v^{k+1}). \end{cases}$$

However, how to solve (1) ?

This leads to <u>alternating direction</u> method of multiplier:

$$\arg\min_{(x,v)} \frac{1}{2}\|Ax-b\|^2 + \alpha\|v\|_1 + \lambda^T(x-v) + \frac{c}{2}\|x-v\|^2$$

$$= \begin{cases} \arg\min_x \frac{1}{2}\|Ax-b\|^2 + \lambda^T x + \frac{c}{2}\|x-v\|^2 \\ \arg\min_v \alpha\|v\|_1 - \lambda^T v + \frac{c}{2}\|v-x\|^2. \end{cases}$$

$$\Rightarrow \begin{cases} \arg\min_x \frac{1}{2}\|Ax-b\|^2 + \frac{c}{2}\|x-(v-\frac{\lambda}{c})\|^2 \cancel{= \arg\min \frac{1}{2}\|Ax-b\|^2 + c} \\ \arg\min_v \alpha\|v\|_1 + \frac{c}{2}\|v-(x+\frac{\lambda}{c})\|^2 \end{cases}$$

$$\Rightarrow \begin{cases} \underset{x}{\operatorname{argmin}} \frac{1}{2}\|Ax-b\|^2 + \frac{c}{2}\|x-\tilde{x}\|^2 \\ \underset{v}{\operatorname{argmin}} \cancel{\tfrac{1}{2}\|} \alpha\|v\|_1 + \frac{c}{2}\|v-\tilde{v}\|^2 \end{cases}$$

$$\Rightarrow \begin{cases} x = (A^TA+cI)^{-1}(A^Tb + c\tilde{x}) \\ v = S_{\alpha/c}(\tilde{v}) = \max\{|\tilde{v}| - \frac{\alpha}{c}, 0\}\operatorname{sgn}(\tilde{v}). \end{cases}$$

Therefore, the overall algorithm is

$$\begin{cases} x^{k+1} = (A^TA+cI)^{-1}(A^Tb + c\tilde{x}^k) \\ v^{k+1} = \max\{|\tilde{v}^k| - \frac{\alpha}{c}, 0\}\operatorname{sgn}(\tilde{v}^k). \\ \lambda^{k+1} = \lambda^k + c(x^{k+1} - v^{k+1}). \end{cases}$$

This is the ADMM algorithm.

ADMM algorithm

$$\min_x f(x) + g(x)$$

can be written as

$$\min_{x,v} f(x) + g(v)$$
$$\text{s.t. } x = v.$$

augmented Lagrangian:

$$\mathcal{L}(x,v,u) = f(x) + g(v) + u^T(x-v) + \frac{\rho}{2}\|x-v\|^2$$

Then, solve

$$\begin{cases} x^{k+1} = \underset{x}{\operatorname{argmin}}\, f(x) + \frac{\rho}{2}\|x-\tilde{x}\|^2 & \tilde{x} = v - \frac{u}{c} \\ v^{k+1} = \underset{v}{\operatorname{argmin}}\, g(v) + \frac{\rho}{2}\|v-\tilde{v}\|^2 & \tilde{v} = x + \frac{u}{c} \\ u^{k+1} = u^k + \rho(x^{k+1} - v^{k+1}). \end{cases}$$

General ADMM algorithm

$$\min_{x,v} \; f(x) + g(v)$$
$$\text{s.t.} \quad Ax + Bv = c.$$

The augmented Lagrangian is

$$\mathcal{L}(x, v, u) = f(x) + g(v) + u^T(Ax + Bv - c) + \frac{\rho}{2}\|Ax + Bv - c\|^2.$$

Sub-problems:

$$x^{k+1} = \operatorname{argmin}_x \mathcal{L}(x, v^k, u^k)$$
$$v^{k+1} = \operatorname{argmin}_v \mathcal{L}(x^{k+1}, v, u^k)$$
$$u^{k+1} = u^k + \rho(x^{k+1} - v^{k+1}).$$

Convergence (Boyd)

Assumption 1: f and g are closed, proper, convex

$\Longrightarrow$ ensure subproblems can be solved.

Assumption 2: the unaugmented Lagrangian has a saddle point.

Then

(1): $r^k = Ax^k + Bv^k - c \longrightarrow 0$ as $k \to \infty$

primal residue converges

(2): $f(x^k) + g(v^k) \longrightarrow p^*$ as $k \to \infty$

primal objective value converges

(3): $u^k \longrightarrow u^*$

dual variable converges.

Example:

$$\min_{x} \; \frac{1}{2}\|Ax-b\|^2 + \lambda\|x\|_{TV}$$

Note that $\|x\|_{TV} = \|Dx\|_1$.

So
$$\begin{cases} \min\limits_{x,v} \; \frac{1}{2}\|Ax-b\|^2 + \lambda\|v\|_1 \\ \text{s.t.} \quad v = Dx. \end{cases}$$

$$\mathcal{L}(x,v,u) = \frac{1}{2}\|Ax-b\|^2 + \lambda\|v\|_1 + u^T(v-Dx) + \frac{\rho}{2}\|v-Dx\|^2$$

$$x^{k+1} = \operatorname*{argmin}_x \frac{1}{2}\|Ax-b\|^2 + u^T(v-Dx) + \frac{\rho}{2}\|v-Dx\|^2$$

$$= \operatorname*{argmin}_x \frac{1}{2}\|Ax-b\|^2 + \frac{\rho}{2}\|Dx-v\|^2 - u^TDx.$$

$$\frac{d}{dx}(\cdot) = 0 \Rightarrow A^T(Ax-b) + \rho D^T(Dx-v) - D^Tu = 0$$

$$\Rightarrow (A^TA + \rho D^TD)x = A^Tb + \rho D^Tv + D^Tu$$

$$\Rightarrow x = (A^TA + \rho D^TD)^{-1}\left[A^Tb + \rho D^Tv + D^Tu\right].$$

if A = circular matrix, and D = circular matrix, then

$$A = F\Lambda_A F^H, \quad D = F\Lambda_D F^H.$$

$$\Rightarrow A^TA + \rho D^TD = F\Big[\underbrace{|\Lambda_A|^2 + \rho|\Lambda_D|^2}_{\text{Fourier spectrum.}}\Big]F^H$$

$$\Rightarrow (A^TA + \rho D^TD)^{-1} = F\left[\frac{1}{|\Lambda_A|^2 + \rho|\Lambda_D|^2}\right]F^H.$$

So $$x = \underbrace{\mathcal{F}^{-1}}_{\text{Fourier transform}}\left[\frac{\mathcal{F}[A^T b + \rho D^T v + D^T u]}{|\mathcal{F}[A]|^2 + \rho|\mathcal{F}[D]|^2}\right]$$

Fourier transform

The v-subproblem is

$$v^{k+1} = \underset{v}{\operatorname{argmin}}\ \lambda\|v\|_1 + u^T v + \frac{\rho}{2}\|v - Dx\|^2$$

$$= \max\left\{\left|Dx + \frac{u}{\rho}\right| - \frac{\lambda}{\rho},\ 0\right\} \operatorname{sgn}\left(Dx + \frac{u}{\rho}\right).$$

<u>Example</u>:

Image-Superresolution:

$$\min_x \frac{1}{2}\|SHx - b\|^2 + \lambda\|x\|_{TV}$$

$$\begin{cases} H = \text{convolutional matrix} \\ S = \text{sampling matrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \text{ for example.} \end{cases}$$

Then the x-sub-problem is

$$x = (\underbrace{H^T S^T S H}_{\text{this matrix is not circular}} + \rho D^T D)^{-1}\left[H^T S^T b + D^T(\rho v + u)\right].$$

Sherman-Morrison-Woodbury:

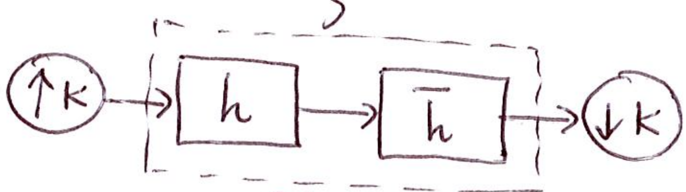
Let $r = H^TS^Tb + D^T(\rho v + u)$. $G = \cancel{H^TS}\ SH$

For simplicity, let's also consider $D = I$.

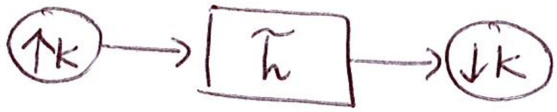
Then

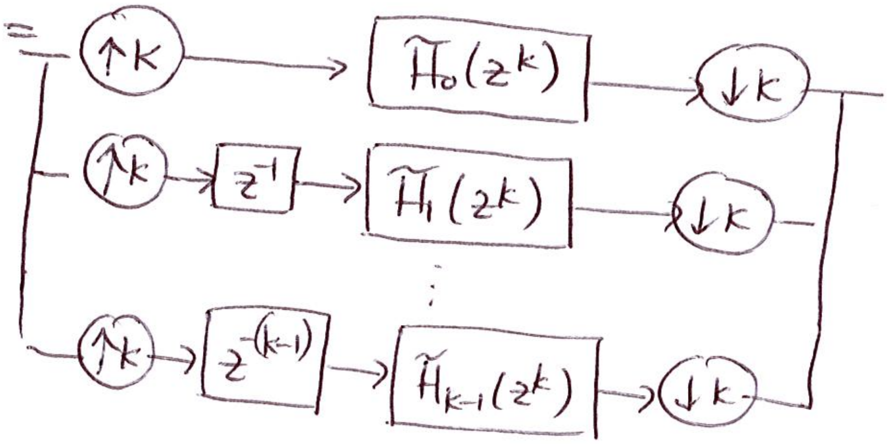
$$\left(G^TG + \rho I\right)^{-1} r$$

$$= \rho^{-1} r - \rho^{-1} G^T (\rho I + \underbrace{GG^T}_{= SHH^TS^T})^{-1} G r .$$

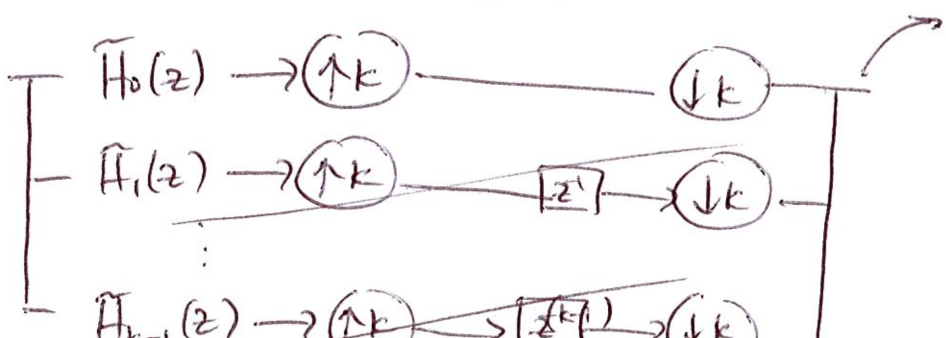


$= \tilde{h}$





$= \tilde{h}_0[n]$.



Example:

$$\min_x \tfrac{1}{2}x^T A x$$
$$\text{s.t. } \mathbb{1}^T x = 1, \quad x \geq 0.$$

$$\Longrightarrow \quad \min_{x,v} \tfrac{1}{2}x^T A x + \mathbb{1}\{v \geq 0\}$$
$$\text{s.t. } \mathbb{1}^T x = 1, \quad x = v.$$

$$\begin{cases} x^{k+1} = \operatorname*{argmin}_{\mathbb{1}^T x = 1} \tfrac{1}{2}x^T A x + u^T(x-v) + \tfrac{\rho}{2}\|x-v\|^2 \\ v^{k+1} = \operatorname*{argmin}_{v} \mathbb{1}\{v \geq 0\} + u^T(x-v) + \tfrac{\rho}{2}\|x-v\|^2 \\ u^{k+1} = u^k + \rho(x^{k+1} - v^{k+1}). \end{cases}$$

$$L = \tfrac{1}{2}x^T A x + u^T x + \tfrac{\rho}{2}\|x-v\|^2 + \gamma(\mathbb{1}^T x - 1)$$

$$\begin{cases} Ax + u + \rho(x-v) + \gamma\mathbb{1} = 0 \\ \mathbb{1}^T x - 1 = 0 \end{cases}$$

$$\begin{bmatrix} A+\rho I & +\mathbb{1} \\ \mathbb{1}^T & 0 \end{bmatrix} \begin{bmatrix} x \\ \gamma \end{bmatrix} = \begin{bmatrix} \rho v - u \\ 1 \end{bmatrix}$$

$$= P_{\Omega}\left(v + \frac{u}{\rho}\right).$$

Example

$$\min_x \ \|Ax-b\|_1 + \lambda\|Dx\|_1$$

$$\Longrightarrow \quad \min_{y,v,x} \ \|y-b\|_1 + \lambda\|v\|_1$$

$$\text{s.t.} \quad y = Ax \ , \quad v = Dx. \quad \Longrightarrow \begin{bmatrix} A \\ D \end{bmatrix} x = \begin{bmatrix} y \\ v \end{bmatrix}.$$

$$\begin{aligned} \mathcal{L}(x,y,v,u_1,u_2) = \ & \|y-b\|_1 + \lambda\|v\|_1 \\ & + u_1^T(\cancel{y=} Ax - y) + \frac{\rho_1}{2}\|Ax-y\|^2 \\ & + u_2^T(Dx-v) + \frac{\rho_2}{2}\|Dx-v\|^2. \end{aligned}$$

$$\begin{aligned} x^{k+1} &= \operatorname*{argmin}_x \ u_1^T Ax + \frac{\rho_1}{2}\|Ax-y\|^2 + u_2^T Dx + \frac{\rho_2}{2}\|Dx-v\|^2 \\ &= \operatorname*{argmin}_x \ \frac{\rho_1}{2}\|Ax-y\|^2 + \frac{\rho_2}{2}\|Dx-v\|^2 + (A^T u_1)^T x + (D^T u_2)^T x \\ &= (\rho_1 A^T A + \rho_2 D^T D)^{-1}(\rho_1 A^T y + \rho_2 D^T v + A^T u_1 + D^T u_2). \end{aligned}$$

$$\begin{aligned} v^{k+1} &= \operatorname*{argmin}_v \ \lambda\|v\|_1 + u_2^T(Dx-v) + \frac{\rho_2}{2}\|Dx-v\|^2 \\ &= \max\left(\left|Dx + \frac{u_2}{\rho_2}\right| - \frac{\lambda}{\rho_2}, 0\right) \operatorname{sgn}\left(Dx + \frac{u_2}{\rho_2}\right) \end{aligned}$$

$$\begin{aligned} y^{k+1} &= \operatorname*{argmin}_y \ \|y-b\|_1 + u_1^T(Ax-y) + \frac{\rho_1}{2}\|Ax-y\|^2 \\ &= \cancel{\max} \operatorname*{argmin}_y \ \|y-b\|_1 \cancel{+} - u_1^T y + \frac{\rho_1}{2}\|Ax-y\|^2 \\ &= \operatorname*{argmin}_{\tilde{y}} \ \|\tilde{y}\|_1 - u_1^T \tilde{y} + \frac{\rho_1}{2}\|Ax - \tilde{y} - b\|^2 \\ &= \max\left(\left|Ax-b-\frac{u_1}{\rho_1}\right| - \frac{\lambda}{\rho_1}, 0\right) \operatorname{sgn}\left(Ax-b-\frac{u_1}{\rho_1}\right). \end{aligned}$$