Primal-dual subgradient method

6:31 PM

(Boyd notes on subgradient methods)

$\min f_0(x)$ → { f_i not necessarily differentiable }

s.t. $\forall_{i\in\{1,\dots,m\}}\ f_i(x)\le 0$

$Ax=b$

* Equality constrained problems ∵ for any feasible x $\|Ax-b\|_2^2$ is 0, so objective value will not change

$$\begin{pmatrix} \min f(x) \\ \text{s.t. } \underset{\in\mathbb{R}^{m\times n}}{A}x=b \end{pmatrix} = \begin{pmatrix} \min f(x)+\frac{\rho}{2}\|Ax-b\|_2^2 \\ \text{s.t. } Ax=b \end{pmatrix}$$

$$L(x,\nu)=f(x)+\frac{\rho}{2}\|Ax-b\|_2^2+\nu^T(Ax-b)=f(x)+(A^T\nu)^Tx-\nu^Tb+\frac{\rho}{2}\|Ax-b\|_2^2$$

$$T(x,\nu)\ \{\text{KKT operator}\}=\begin{bmatrix}\partial_x L(x,\nu)\\ -\partial_\nu L(x,\nu)\end{bmatrix}=\begin{bmatrix}\partial f(x)+A^T\nu+\rho A^T(Ax-b)\\ -(Ax-b)\end{bmatrix} \quad \frac{\partial}{\partial x}\left[f(Ax+b)\right]=A^T\left[\frac{\partial}{\partial x}f(x)\right]_{x:=Ax+b}$$

this ∂ is subgradient operator, not gradient

$$=\begin{bmatrix}\partial f(x)+A^T\nu+\rho A^T(Ax-b)\\ b-Ax\end{bmatrix}$$

[illegible] structurally this is a set

KKT condition [illegible] vanishing gradient of Lagrangian and primal feasibility law [illegible] optimal (x^*,ν^*) [illegible]

$$\left(0\in\partial_{x^*}L(x^*,\nu^*),\ -\partial_{\nu^*}L(x^*,\nu^*)=b-Ax^*=0\right)\ \overset{k}{=}\ 0\in T(x^*,\nu^*)$$

(*)

Primal dual subgradient method:

let. $z^{(k)}=\begin{bmatrix}x^{(k)}\\ \nu^{(k)}\end{bmatrix}$

$$z^{(k+1)}=z^{(k)}-\underbrace{\alpha_k}_{>0}T^{(k)} \quad \#\ T^{(k)}\in T(z^{(k)})=T(x^{(k)},\nu^{(k)}) \ \therefore \begin{bmatrix}g_x+A^T\nu+\rho A^T(Ax-b)\\ b-Ax\end{bmatrix}\in\begin{bmatrix}\partial f(x)+A^T\nu+\rho A^T(Ax-b)\\ b-Ax\end{bmatrix}$$

$$\begin{bmatrix}x^{(k+1)}\\ \nu^{(k+1)}\end{bmatrix}=\begin{bmatrix}x^{(k)}\\ \nu^{(k)}\end{bmatrix}+\alpha_k\begin{bmatrix}\overbrace{g_{x^{(k)}}}^{g^{(k)}}+A^T\nu^{(k)}+\rho A^T(Ax^{(k)}-b)\\ b-Ax^{(k)}\end{bmatrix}$$

$$\Leftrightarrow \boxed{\begin{aligned}x^{(k+1)}&=x^{(k)}+\alpha_k\left(g^{(k)}+A^T\nu^{(k)}+\rho A^T(Ax^{(k)}-b)\right)\\ \nu^{(k+1)}&=\nu^{(k)}+\alpha_k(Ax^{(k)}-b)\end{aligned}}$$

Primal Dual Subgradient Method

\# Note that $x^{(k)}$ is not necessarily feasible, i.e. $Ax^{(k)}-b$ can be $\ne 0$

Now want to show: • the algorithm converges for stepsize rule $\alpha_k=\frac{\gamma_k}{\|T^{(k)}\|_2}$, $\gamma_k\ \{>0,\ \sum_{k=1}^{\infty}\gamma_k=\infty,\ \sum_{k=1}^{\infty}\gamma_k^2=S<\infty\}$

$$\lim_{k\to\infty}f(x^{(k)})=p^*,\quad \lim_{k\to\infty}\|Ax^{(k)}-b\|_2=0$$

\# positive square summable but not summable

Goal Primal Dual Alg Conv

• Proof: Assumption: • $\exists_{R>0}\ \|z^*\|_2\le R,\ \|z^{(1)}\|_2\le R$

• $\exists_{\bar{R}}\ \|g_{\{\in\partial f\}}\|_2\le\bar{R}$

$$\|z^{(k+1)}-z^*\|_2^2=\|z^{(k)}-\alpha_kT^{(k)}-z^*\|_2^2$$

$$=\|(z^{(k)}-z^*)-\alpha_kT^{(k)}\|_2^2$$

$$=\|z^{(k)}-z^*\|_2^2+\underbrace{\alpha_k^2\|T^{(k)}\|_2^2}_{\gamma_k^2}-2\underset{\downarrow\ \gamma_k/\|T^{(k)}\|}{\alpha_k}T^{(k)T}(z^{(k)}-z^*)=\|z^{(k)}-z^*\|_2^2+\gamma_k^2-2\frac{\gamma_k}{\|T^{(k)}\|_2}T^{(k)T}(z^{(k)}-z^*)=\|z^{(k-1)}-z^*\|_2^2+\sum_{i=k-1}^{k}\gamma_i^2-2\sum_{i=k-1}^{k}\frac{\gamma_i}{\|T^{(i)}\|_2}T^{(i)T}(z^{(i)}-z^*)$$

$$-\nu^{(k)}(Ax^{(k)}-b)+\nu^*(Ax^{(k)}-b)$$

$\ne$

$= \|z^{(k)}-z^*\|_2^2 + \underbrace{\alpha_k^2 \|T^{(k)}\|_2^2}_{\gamma_k^2} - 2\underset{\downarrow\ \gamma_k/\|T^{(k)}\|_2}{\alpha_k} T^{(k)T}(z^{(k)}-z^*) = \|z^{(k)}-z^*\|_2^2 + \gamma_k^2 - 2\frac{\gamma_k}{\|T^{(k)}\|_2} T^{(k)}(z^{(k)}-z^*) = \|z^{(k-1)}-z^*\|_2^2 + \sum_{i=k-1} \gamma_i^2 - 2\sum_{i=k-1} \frac{\gamma_i}{\|T^{(i)}\|_2} T^{(i)T}(z^{(i)}-z^*)$

$\vdots$

$= \|z^{(0)}-z^*\|_2^2 + \sum_{i=0}^{k} \gamma_i^2 - 2\sum_{i=0}^{k} \frac{\gamma_i}{\|T^{(i)}\|_2} \underbrace{T^{(i)T}(z^{(i)}-z^*)}_{\geq 0}$ { ∵ T is a monotonic operator $\leftrightarrow (T^{(i)} - \underbrace{T^*}_{\in 0})^T(z^{(i)}-z^*) \geq 0 \rightarrow T^{(i)T}(z^{(i)}-z^*) \geq 0$ }

similarly

$\|z^{(k)}-z^*\|_2^2 = \|z^{(k-1)}-z^*\|_2^2 + \gamma_{k-1}^2 - 2\frac{\gamma_{k-1}}{\|T^{(k-1)}\|_2} T^{(k-1)T}(z^{(k)}-z^*)$

$\vdots$

$\|z^{(1)}-z^*\|_2^2 = \|z^{(0)}-z^*\|_2^2 + \gamma_0^2 - 2\frac{\gamma_0}{\|T^{(0)}\|_2} T^{(0)T}(z^{(k)}-z^*$

$\rightarrow \underbrace{\|z^{(k+1)}-z^*\|_2^2}_{\geq 0} + \underbrace{2\sum_{i=0}^{k} \frac{\gamma_i}{\|T^{(i)}\|_2^2} T^{(i)T}(z^{(i)}-z^*)}_{\geq 0} = \|z^{(0)}-z^*\|_2^2 + \sum_{i=0}^{k} \gamma_i^2$

$\leq \underbrace{\|z^{(0)}-z^*\|_2^2}_{} + \underbrace{\sum_{i=0}^{\infty} \gamma_i^2}_{S} \leq 4R^2+S$

$\|z^{(0)}-z^*\|_2^2$

$= \underbrace{\|z^{(0)}\|_2^2}_{\leq R^2} + \underbrace{\|z^*\|_2^2}_{\leq R^2} \underbrace{-2z^{(0)T}z^*}_{\leq 2R^2}$ # Cauchy-Schwartz:

$\underbrace{|z^{(0)T}z^*|}_{\max\{z^{(0)T}z^*, -z^{(0)T}z^*\}} \leq \underbrace{\|z^{(0)}\|_2}_{\leq R} \underbrace{\|z^*\|_2}_{\leq R} \leq R^2$

$\rightarrow -z^{(0)T}z^* \leq R^2$

$\leq 4R^2$

$\therefore \|z^{(k+1)}-z^*\|_2^2 + 2\sum_{i=0}^{k} \frac{\gamma_i}{\|T^{(i)}\|_2^2} T^{(i)T}(z^{(i)}-z^*) \leq 4R^2+S$

compactness of the sequence z^{(k)}

{ কারণ L.H.S এর দুটো termই positive so যেকোনো একটা বাদ দিলে অসমতার দিক পাল্টাবে না }

$\therefore \forall_k \quad \|z^{(k+1)}-z^*\|_2^2 \leq 4R^2+S.$

$\forall_k \quad 2\sum_{i=0}^{k} \frac{\gamma_i}{\|T^{(i)}\|_2^2} T^{(i)T}(z^{(i)}-z^*) \leq 4R^2+S$ # Bounded series with in a compact set, but $\sum_k \gamma_k \to \infty$, so for convergence we need:

$\lim_{k\to\infty} \frac{1}{\|T^{(k)}\|_2} T^{(k)T}(z^{(k)}-z^*) = 0$

$\forall_k \ \|z^{(k)} - \alpha_k T^{(k)} - z^*\|_2^2 \leq 4R^2+S$

? # need to figure out later

$\Rightarrow \|T^{(k)}\|_2 \leq \square$

but $\|T^{(k)}\|_2 \leq \square$, so $\lim_{k\to\infty} T^{(k)T}(z^{(k)}-z^*) = 0$

an alternative proof: $T^{(k)T}(z^{(k)}-z^*) = \left(g^{(k)} + A^T\nu^{(k)} + \rho A^T(Ax^{(k)}-b)\right)^T(x^{(k)}-x^*) + \overbrace{(b-Ax^{(k)})^T(\nu^{(k)}-\nu^*)}^{\#} = g^{(k)T}(x^{(k)}-x^*) + \cancel{\nu^{(k)T}(Ax^{(k)}-b)} + \rho\|Ax^{(k)}-b\|_2^2 - \cancel{\nu^{(k)T}}(Ax^{(k)}-b) + \nu^{*T}(Ax^{(k)}-b)$

$\begin{bmatrix} g^{(k)} + A^T\nu^{(k)} + \rho A^T(Ax^{(k)}-b) \\ b - Ax^{(k)} \end{bmatrix}^T \begin{bmatrix} x^{(k)}-x^* \\ \nu^{(k)}-\nu^* \end{bmatrix}$

$= g^{(k)T}(x^{(k)}-x^*) + \nu^{(k)T}A(x^{(k)}-x^*) + \rho(Ax^{(k)}-b)^TA(x^{(k)}-x^*)$

$Ax^{(k)} - \underbrace{Ax^*}_{b} = Ax^{(k)}-b$

$Ax^{(k)} - \underbrace{Ax^*}_{b} = Ax^{(k)}-b^*$

$= g^{(k)T}(x^{(k)}-x^*) + \nu^{(k)T}(Ax^{(k)}-b) + \rho\|Ax^{(k)}-b\|_2^2$

$= g^{(k)T}(x^{(k)}-x^*) + \nu^{*T}(Ax^{(k)}-b) + \rho\|Ax^{(k)}-b\|_2^2$ # ∵ $g^{(k)} \in \partial f(x^{(k)})$ ∴ $\forall_{x\in \mathrm{dom} f}$ $f(x) \geq f(x^{(k)}) + g^{(k)T}(x-x^{(k)})$

$x := x^* \in \mathrm{dom} f \rightarrow f(x^*) \geq f(x^{(k)}) + g^{(k)T}(x^*-x^{(k)})$

$\Leftrightarrow -g^{(k)T}(x^*-x^{(k)}) = g^{(k)T}(x^{(k)}-x^*) \geq f(x^{(k)}) - \underbrace{f(x^*)}_{p^*}$

$\Leftrightarrow g^{(k)T}(x^{(k)}-x^*) \geq f(x^{(k)}) - p^*$

$\geq f(x^{(k)}) - p^* + \nu^{*T}(Ax^{(k)}-b) + \rho\|Ax^{(k)}-b\|_2^2$

$= f(x^{(k)}) + \nu^{*T}(Ax^{(k)}-b) + \frac{\rho}{2}\|Ax^{(k)}-b\|_2^2 - p^* + \frac{\rho}{2}\|Ax^{(k)}-b\|_2^2$

$L(x,\nu) = f(x) + \frac{\rho}{2}\|Ax-b\|_2^2 + \nu^T(Ax-b)$

∴ $L(x^{(k)},\nu^*) = f(x^{(k)}) + \frac{\rho}{2}\|Ax^{(k)}-b\|_2^2 + \nu^{*T}(Ax-b)$

$L(x^*,\nu^*) = f(x^*) + \frac{\rho}{2}\underbrace{\|Ax^*-b\|_2^2}_{0} + \nu^{*T}\underbrace{(Ax^*-b)}_{0} = f(x^*) = p^*$

$= \underbrace{L(x^{(k)},\nu^*) - L(x^*,\nu^*)}_{\geq 0} + \overbrace{\frac{\rho}{2}\|Ax^{(k)}-b\|_2^2}^{\geq 0} \geq 0$

$L(x^*,\nu^*) = \min_x(x,\nu^*) \leq L(x^{(k)},\nu)$

∴ $L(x^{(k)},\nu) - L(x^*,\nu^*) \geq 0$

$\therefore \forall_k \quad T^{(k)T}(z^{(k)}-z^*) \geq L(x^{(k)},\nu^*) - L(x^*,\nu^*) + \frac{\rho}{2}\|Ax^{(k)}-b\|_2^2 \geq 0$ inequlity for T^{(k)T} (z^{(k)}-z^*)

again. inequlity for T^{(k)T} (z^(k)-z^*)

[illegible]:

$\forall_k \quad 0 \leq L(x^{(k)},\nu^*) - L(x^*,\nu^*) + \frac{\rho}{2}\|Ax^{(k)}-b\|_2^2 \leq T^{(k)T}(z^{(k)}-z^*)$

$\xRightarrow{\lim_{k\to\infty}} \quad 0 \leq \lim_{k\to\infty}\left(\underbrace{L(x^{(k)},\nu^*) - L(x^*,\nu^*)}_{\geq 0} + \underbrace{\frac{\rho}{2}\|Ax^{(k)}-b\|_2^2}_{\geq 0}\right) \leq \lim_{k\to\infty} T^{(k)T}(z^{(k)}-z^*) = 0$

$\leftrightarrow \lim_{k\to\infty}\left(L(x^{(k)},\nu^*) - L(x^*,\nu^*) + \frac{\rho}{2}\|Ax^{(k)}-b\|_2^2\right) = 0$

$\lim_{k\to\infty} a(k)\{\geq 0\} + b(k)\{\geq 0\} = 0 \Rightarrow$

$\lim_{k\to\infty} a(k) = 0, \quad \lim_{k\to\infty} b(k) = 0$

$\mathrm{Lt}_{k\to\infty} \; a(k)\{\geq 0\} + b(k)\{\geq 0\} = 0 \Rightarrow \mathrm{Lt}_{k\to\infty} a(k) = 0, \; \mathrm{Lt}_{k\to\infty} b(k) = 0$

$\Leftrightarrow \mathrm{Lt}_{k\to\infty} \left(L(x^{(k)}, \nu^{*}) - L(x^{*}, \nu^{*}) \right) = 0 \quad \mathrm{Lt}_{k\to\infty} \|Ax^{(k)} - b\|_2^2 = 0 \Leftrightarrow \mathrm{Lt}_{k\to\infty} Ax^{(k)} = b$

$\wedge$

$\Leftrightarrow \mathrm{Lt}_{k\to\infty} L(x^{(k)}, \nu^{*}) = L(x^{*}, \nu^{*}) = p^{*}$

$\stackrel{\#}{=} \mathrm{Lt}_{k\to\infty} \left(f(x^{(k)}) + \nu^{*T}(Ax^{(k)} - b) + \frac{\nu^{*}}{2}\|Ax^{(k)} - b\|_2^2 \right)$

$= \mathrm{Lt}_{k\to\infty} f(x^{(k)}) + \mathrm{Lt}_{k\to\infty} \underbrace{\nu^{*T}(Ax^{(k)} - b)}_{\parallel \atop 0} + \underbrace{\frac{\nu^{*}}{2}\|Ax^{(k)} - b\|_2^2}_{\parallel \atop 0}$

$= \mathrm{Lt}_{k\to\infty} f(x^{(k)})$

$\Leftrightarrow \mathrm{Lt}_{k\to\infty} f(x^{(k)}) = p^{*}$

used proof strategy if we want to prove P=>R
then proving (P=>Q1, P=>Q2,...,P=>Qn, (Q1/\Q2/\.../\Qn)=>R) Suffices as this implies (P=>R
)

$\therefore \mathrm{Lt}_{k\to\infty} f(x^{(k)}) = p^{*}, \quad \mathrm{Lt}_{k\to\infty} Ax^{(k)} = b$ Goal Primal Dual Alg Conv reached!

(Proved)

$*$ Inequality constrained problems:

The previous algorithm can be extended to inequality constrained problem:

$$\begin{pmatrix} \min f_0(x) \\ \text{s.t.} \\ \forall_{i\in\{1,\dots,m\}} \; f_i(x) \leq 0 \\ x \in \mathbb{R}^n \end{pmatrix} = \begin{pmatrix} \min f_0(x) \\ \text{s.t.} \\ \forall_{i\in\{1,\dots,m\}} \; f_i(x)_+ = 0 \end{pmatrix}$$

\# define, $F(x) = \begin{bmatrix} f_1(x)_+ \\ \vdots \\ f_i(x)_+ \end{bmatrix} = \bar{0}$ $= \begin{pmatrix} \min f_0(x) \\ \text{s.t.} \; F(x) = 0 \end{pmatrix} = \begin{pmatrix} \min f_0(x) + \frac{\rho}{2}\|F(x)\|_2^2 \\ \text{s.t.} \; F(x) = 0 \end{pmatrix}$ Equality version of inequality constrained problem

however note that this is not affine anymore, so this is not in DCP format anymore

\# note that:

$f_i(x) \leq 0 \Leftrightarrow f_i(x)_+ = 0$

$(\rightarrow)$

$f_i(x) \leq 0$

then $f_i(x)_+ = \max\{f_i(x), 0\} = 0$

this is $\in \mathbb{R}_-$, so max with 0 will result in 0_+

$(\leftarrow)$ per absurdum

$f_i(x)_+ = \max\{f_i(x), 0\} = 0$ but $f_i(x) = 0_{++}$ then

$f_i(x)_+ = \max\{0_{++}, 0\} = 0_{++} > 0 \Rightarrow$ contradiction

Define, the augmented Lagrangian as follows: //similar as before

$$L(x, \lambda) = f_0(x) + \frac{\rho}{2}\|F(x)\|_2^2 + \lambda^T F(x) = f_0(x) + \frac{\rho}{2}\sum_{i=1}^{m} f_i(x)_+^2 + \sum_{i=1}^{m} \lambda_i f_i(x)_+$$

note % $D_x\left(f_i(x)_+^2\right) = D_{f_i(x_+)} f_i(x)_+^2 \; D_x f_i(x)_+$

% $= 2 \underbrace{f_i(x_+)}_{\text{number}} \underbrace{D_x f_i(x)_+}_{\text{row vector}}$

% $\nabla_x f_i(x)_+^2 = \left(D_x (f_i(x)_+^2) \right)^T = 2 f_i(x_+) \nabla_x f_i(x)_+$

$\therefore \partial_x f_i(x)_+^2 = 2 f_i(x)_+ \partial_x f_i(x)_+$

similarly, $\partial_x (\lambda_i f_i(x)_+) = \lambda_i \partial_x f_i(x)_+$

$f_i(x)_+ = \max\{0_{++}, 0\} = 0_{++} > 0 \Rightarrow$ contradiction

similarly, $\partial_x(\lambda_i f_i(x)_+) = \lambda_i \partial_x f_i(x)_+$

$$\partial_x L(x,\lambda) = \partial_x \left(f_0(x) + \frac{\rho}{2}\sum_{i=1}^m f_i(x)_+^2 + \sum_{i=1}^m \lambda_i f_i(x)_+ \right)$$

$$= \partial_x f_0(x) + \frac{\rho}{2}\sum_{i=1}^m \partial_x \left(f_i(x)_+^2\right) + \sum_{i=1}^m \lambda_i \partial_x (f_i(x)_+)$$

$$= \partial_x f_0(x) + \frac{\rho}{2}\sum_{i=1}^m 2 f_i(x)_+ \partial_x f_i(x)_+ + \sum_{i=1}^m \lambda_i \partial_x (f_i(x)_+)$$

$$= \partial_x f_0(x) + \sum_{i=1}^m (\rho f_i(x)_+ \partial_x f_i(x)_+) + \sum_{i=1}^m \lambda_i \partial_x (f_i(x)_+) = \partial_x f_0(x) + \sum_{i=1}^m (\rho f_i(x)_+ + \lambda_i)\, \partial_x f_i(x)_+$$

$\partial_\lambda L(x,\lambda)$ # $L(x,\lambda)$ is affine in λ, so differentiable #

$$= \nabla_\lambda L(x,\lambda) = \nabla_\lambda \left(f_0(x) + \frac{\rho}{2}\|F(x)\|_2^2 + \lambda^T F(x) \right) = F(x)$$

Define the KKT operator again

$$T(x,\lambda) = \begin{bmatrix} \partial_x L(x,\lambda) \\ -\partial_\lambda L(x,\lambda) \end{bmatrix} = \begin{bmatrix} \partial_x f_0(x) + \sum_{i=1}^m (\rho f_i(x)_+ + \lambda_i)\, \partial f_i(x)_+ \\ -F(x) \end{bmatrix}$$

The optimality condition is just as before:

comes from KKT condition:

$$T(x^*,\lambda^*) \ni \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Vanishing gradient of Lagrangian

Primal feasibility, note that

- λ is unrestricted
- in this formulation inequality constraint 가 사라졌기 때문에 dual feasibility $\lambda \geq 0$, λ_i complementary slackness 사라졌다.

The primal-dual pair (x^*, v^*) is a saddle point of augmented Lagrangian:

$$\forall_x \forall_\lambda \quad L(x^*,\lambda) \leq L(x^*,\lambda^*) \leq L(x,\lambda^*)$$

The primal-dual subgradient method is:

$$z^{(k+1)} = z^{(k)} - \alpha_k T^{(k)}$$

$z^{(k)} = \begin{bmatrix} x^{(k)} \\ \lambda^{(k)} \end{bmatrix}$, $T^{(k)} \in T(z^{(k)}) = T(x^{(k)},\lambda^{(k)}) = \begin{bmatrix} \partial_x L(x^{(k)},\lambda^{(k)}) \\ -\partial_\lambda L(x^{(k)},\lambda^{(k)}) \end{bmatrix} = \begin{bmatrix} \partial_x f_0(x^{(k)}) + \sum_{i=1}^m (\rho f_i(x^{(k)})_+ + \lambda_i^{(k)})\, \partial_x f_i(x^{(k)})_+ \\ -F(x^{(k)}) \end{bmatrix}$

expand

$\therefore T^{(k)} = \begin{bmatrix} g_0^{(k)} + \sum_{i=1}^m (\rho f_i(x^{(k)})_+ + \lambda_i^{(k)}) g_i^{(k)} \\ -F(x^{(k)}) \end{bmatrix}$ # $g_0^{(k)} \in \partial_x f_0(x^{(k)})$, $g_i^{(k)} \in \partial_x f_i(x^{(k)})_+$

expand

$$\# \; g_0^{(k)} \in \partial_x f_0(x^{(k)}),\; g_i^{(k)} \in \partial_x f_i(x^{(k)})_+$$

$$\begin{bmatrix} x^{(k+1)} \\ \lambda^{(k+1)} \end{bmatrix} = \begin{bmatrix} x^{(k)} \\ \lambda^{(k)} \end{bmatrix} - \alpha_k \begin{bmatrix} g_0^{(k)} + \sum_{i=1}^{m} \left(\lambda_i^{(k)} + \rho f_i(x^{(k)})_+\right) g_i^{(k)} \\ -F(x^{(k)}) \end{bmatrix}$$

$$\Leftrightarrow \begin{pmatrix} x^{(k+1)} = x^{(k)} - \alpha_k \left(g_0^{(k)} + \sum_{i=1}^{m} \left(\lambda_i^{(k)} + \rho f_i(x^{(k)})_+\right) g_i^{(k)}\right) \\ \lambda^{(k+1)} = \lambda^{(k)} + \alpha_k \underbrace{F(x^{(k)})}_{\begin{bmatrix} f_1(x^{(k)})_+ \\ \vdots \\ f_m(x^{(k)}) \end{bmatrix}} \Leftrightarrow \forall_{i \in \{1,\dots,m\}} \; \lambda_i^{(k+1)} = \lambda_i^{(k)} + \alpha_k f_i(x^{(k)})_+ \end{pmatrix}$$

$$\Leftrightarrow \boxed{\begin{gathered} x^{(k+1)} = x^{(k)} - \alpha_k \left(g_0^{(k)} + \sum_{i=1}^{m} \left(\lambda_i^{(k)} + \rho f_i(x^{(k)})_+\right) g_i^{(k)}\right) \\ \forall_{i \in \{1,\dots,m\}} \; \lambda_i^{(k+1)} = \lambda_i^{(k)} + \underbrace{\alpha_k}_{>0} \underbrace{f_i(x^{(k)})_+}_{\geq 0} \end{gathered}}$$

Primal dual subgradient method for inequality constraints

\# so, λ_i can only increase with iterations

- Convergence Proof:

Assume optimal solution $(x^*, \lambda^*) = z^*$, so $\quad T(x^*, \lambda^*) \ni 0 \Leftrightarrow F(x^*) = 0, \left[\partial_x L(x,\lambda)\right]_{x=x^*, \lambda=\lambda^*} = \partial_x (L(x^*, \lambda^*)) \ni 0$

$\downarrow$ assume $f(x^*) = p^*$

We want to prove: with step size rule: $\alpha_k = \gamma_k / \|T^{(k)}\|_2$ $\quad$ \# γ_k positive, square summable, but not summable

$$\lim_{k\to\infty} f_0(x^{(k)}) = p^*, \quad \lim_{k\to\infty} \|F(x^{(k)})\|_2 = 0,$$

At first note compactness of the sequence z^{(k)} [illegible] ([illegible] equality constraint case [illegible],
note that in the equations $T^{(k)}$ [illegible]
[illegible], so the conclusion will still be valid, except L.H.S [illegible] second term [illegible] positivity proof [illegible])

$$\forall_k \quad \|z^{(k+1)} - z^*\|_2^2 + 2 \sum_{i=0}^{k} \frac{\gamma_i}{\|T^{(i)}\|_2^2} T^{(i)T} (z^{(i)} - z^*) \leq 4R^2 + S$$

[illegible] positive

$$T_i^{(k)T} (z^{(i)} - z^*)$$

$$= \begin{bmatrix} g_0^{(k)} + \sum_{i=1}^{m} \left(\rho f_i(x^{(k)})_+ + \lambda_i^{(k)}\right) g_i^{(k)} \\ -F(x^{(k)}) = -\left[f_i(x^{(k)})_+\right]_{i=1}^{m} \end{bmatrix}^T \begin{bmatrix} x^{(k)} - x^* \\ \lambda^{(k)} - \lambda^* \end{bmatrix}$$

$$\# \; T^{(k)} = \begin{bmatrix} g_0^{(k)} + \sum_{i=1}^{m} \left(\rho f_i(x^{(k)})_+ + \lambda_i^{(k)}\right) g_i^{(k)} \\ -F(x^{(k)}) \end{bmatrix} \quad \# \; g_0^{(k)} \in \partial_x f_0(x^{(k)}),\; g_i^{(k)} \in \partial_x f_i(x^{(k)})_+$$

$$= \left(g_0^{(k)} + \sum_{i=1}^{m} \left(\rho f_i(x^{(k)})_+ + \lambda_i^{(k)}\right) g_i^{(k)}\right)^T (x^{(k)} - x^*) - F(x^{(k)})^T (\lambda^{(k)} - \lambda^*)$$

$\# \ \forall_{i \in \{1,\dots,m\}} \quad g_i^{(k)} \in \partial f_i(x^{(k)})_+ \leftrightarrow \forall_y \ f_i(y)_+ \geq f_i(x^{(k)})_+ + g_i^{(k)T}(y - x^{(k)})$

$y := x^* \Rightarrow f_i(x^*)_+ \geq f_i(x^{(k)})_+ + g_i^{(k)T}(x^* - x^{(k)})$

$\Leftrightarrow -g_i^{(k)T}(x^* - x^{(k)}) = g_i^{(k)T}(x^{(k)} - x^*) \geq f_i(x^{(k)})_+ - \underbrace{f_i(x^*)_+}_{=0} = f_i(x^{(k)})_+ \quad \#$ as x^* Ⓢ $\forall_{i \in \{1,\dots,m\}} f_i(x^*) \leq 0$

$\Rightarrow \forall_{i \in \{1,\dots,m\}} f_i(x^*)_+ = 0$

$$\boxed{\forall_{i \in \{1,\dots,m\}} \ g_i^{(k)T}(x^{(k)} - x^*) \geq f_i(x^{(k)})_+}$$

$g_0^{(k)} \in f_0(x^{(k)}) \leftrightarrow \forall_y \ f_0(y) \geq f_0(x^{(k)}) + g_0^{(k)T}(y - x^{(k)})$

$y := x^* \Rightarrow f_0(x^*) \geq f_0(x^{(k)}) + g_0^{(k)T}(x^* - x^{(k)})$

$\Leftrightarrow -g_0^{(k)T}(x^* - x^{(k)}) = g_0^{(k)T}(x^{(k)} - x^*) \geq f_0(x^{(k)}) - \underbrace{f_0(x^*)}_{p^*}$

$$\boxed{\therefore g_0^{(k)T}(x^{(k)} - x^*) \geq f_0(x^{(k)}) - p^*}$$

$$= \underbrace{g_0^{(k)T}(x^{(k)} - x^*)}_{\geq f_0(x^{(k)}) - p^*} + \sum_{i=1}^{m} \left(\rho f_i(x^{(k)})_+ + \lambda_i^{(k)}\right) \underbrace{g_i^{(k)T}(x^{(k)} - x^*)}_{\geq f_i(x^{(k)})_+} - \underbrace{F(x^{(k)})^T(\lambda^{(k)} - \lambda^*)}_{=}$$

$$\begin{bmatrix} f_1(x^{(k)})_+ \\ \vdots \\ f_m(x^{(k)})_+ \end{bmatrix}^T \begin{bmatrix} \lambda_1^{(k)} - \lambda_1^* \\ \vdots \\ \lambda_m^{(k)} - \lambda_m^* \end{bmatrix} = \sum_{i=1}^{m} f_i(x^{(k)})_+ (\lambda_i^{(k)} - \lambda_m)$$

$$\geq f_0(x^{(k)}) - p^* + \underbrace{\sum_{i=1}^{m} \left(\rho f_i(x^{(k)})_+ + \lambda_i^{(k)}\right) f_i(x^{(k)})_+ - \sum_{i=1}^{m} f_i(x^{(k)})_+ (\lambda_i^{(k)} - \lambda_m^*)}_{=}$$

$\# \ \underbrace{\sum_{i=1}^{m} \rho f_i(x^{(k)})_+^2}_{=} + \cancel{\sum_{i=1}^{m} \lambda_i^{(k)} f_i(x^{(k)})_+} - \cancel{\sum_{i=1}^{m} \lambda_i^{(k)} f_i(x^{(k)})} + \underbrace{\sum \lambda_m^* (f_i(x^{(k)})_+)}_{= \lambda^{*T} F(x^{(k)})}$

$\rho \sum_{i=1}^{m} f_i(x^{(k)})^2 = \rho \| F(x^{(k)}) \|_2^2$

$$= f_0(x^{(k)}) - p^* + \rho \| F(x^{(k)}) \|_2^2 + \lambda^{*T} F(x^{(k)})$$

$\# \ L(x, \lambda) = f_0(x) + \lambda^T F(x) + \frac{\rho}{2} \| F(x) \|_2^2$

$L(x^{(k)}, \lambda^*) = f_0(x^{(k)}) + \lambda^{*T} F(x^{(k)}) + \frac{\rho}{2} \| F(x^{(k)}) \|_2^2$

$L(x^*, \lambda^*) = f_0(x^*) + \lambda^{*T} \underbrace{F(x^*)}_{=0} + \frac{\rho}{2} \| \underbrace{F(x^*)}_{=0} \|_2^2 = f_0(x^*) = p^*$ at (x^*, λ^*) [Equality version of inequality constrained problem](#)

feasible

$$= \left(f_0(x^{(k)}) + \lambda^{*T} F(x^{(k)}) + \frac{\rho}{2} \| F(x^{(k)}) \|_2^2\right) - p^* + \frac{\rho}{2} \| F(x^{(k)}) \|_2^2$$

$$= \underbrace{L(x, \lambda^*) - L(x^*, \lambda^*)}_{\geq 0} + \underbrace{\frac{\rho}{2} \| F(x^{(k)}) \|_2^2}_{\geq 0}$$

$\# \ L(x^*, \lambda^*) = \min_x L(x, \lambda^*) \leq L(x, \lambda^*) \ \forall_x$

$\therefore L(x, \lambda^*) - L(x^*, \lambda^*) \geq 0$

≥ 0

$\therefore \; T^{(k)T}(z^{(k)} - z^*) \geq 0 \quad \forall$

rest of the proof proceeds exactly the same as the equality constrained case. replace $Ax^{(k)} - b$ with $F(x^{(k)})$