1.1) $B = \{ x \in \mathbb{R}^d \mid \|x\|_2^2 \leq 1 \}$

~~Let~~ $\Pi_B(x) = \min_{z \in B} \|z - x\|_2$

<u>if $\|x\|_2 \leq 1$</u> i.e. $x \in B$

then, $\cancel{\|z\| \leq \|z - x\|_2}$

$\|z - x\|_2 \geq 0$

and the minimum is achieved

$\|z - x\| = 0$

when $\boxed{z = x} \Rightarrow \boxed{\Pi_B(x) = x}$

if $\|x\|_2 > 1$

Let $\Pi_B(x) = z \neq \frac{x}{\|x\|_2}$

using the property,

$\langle y' - \Pi_D(y'), \Pi_D(y') - y \rangle \geq 0$

where D is a convex set,

$y' \notin D$, $\forall y \in D$

~~Let $y' = x$ s.t. $\|x\|_2 > 1$ and $D = B$ and $\Pi_D(x) = \Pi_B$~~

Let $D = B$, $y' = x \in B$

and $\Pi_D(y') = \Pi_B(x) = z \neq \frac{x}{\|x\|_2}$

$\Rightarrow$ If $\Pi_B(x) = z$ is correct projection then

$\langle x - z, z - y \rangle \geq 0$

or $\langle x - z, y - z \rangle \leq 0 \quad \forall y \in B$

LHS $\rightarrow$ $\langle x, y \rangle - \langle x - z \rangle - \langle z - y \rangle + \langle z, z \rangle$

Let $y = \frac{x}{\|x\|_2}$

$\Rightarrow$ $\langle x, y \rangle - \langle x, z \rangle - \langle z, y \rangle + \|z\|_2^2$

$= \langle x, \frac{x}{\|x\|_2} \rangle - \langle x, z \rangle - \frac{1}{\|x\|_2} \langle z, x \rangle + \|z\|_2^2$

$= \|x\|_2 + \langle x, z \rangle \left(\frac{1}{\|x\|_2} - 1 \right) + \|z\|_2^2$

$= \|x\|_2 + \|z\|_2^2 + \langle x, z \rangle \left(\frac{1}{\|x\|_2} - 1 \right) - ①$

Scanned with CamScanner

Now given $\|x\|_2 \geq 1$

$$\langle x, z\rangle\left(\frac{1}{\|x\|_2}-1\right) \leq \|x\|_2\|z\|_2\left(\frac{1}{\|x\|_2}-1\right)$$

$$\cancel{\leq \|x\|_2 \|}$$

substituting in (1)

~~LHS~~ $\Rightarrow LHS \geq \|x\|_2 + \|z\|_2^2 + \|x\|_2\|z\|_2\left(\frac{1}{\|x\|_2}-1\right)$

$$\geq \|x\|_2 + \|z\|_2^2 + \|z\|_2 - \|x\|_2\|z\|_2$$

given $z \in B \Rightarrow \|z\|_2 \leq 1$

$\Rightarrow LHS \geq \cancel{\|x\|_2} + \|z\|_2^2 + \|z\|_2 - \cancel{\|x\|_2}$

$LHS \geq \|z\|_2^2 + \|z\|_2 \geq 0$

$\Rightarrow$ Our assumption that $\cancel{\frac{x}{\|x\|}}$, $z \neq \frac{x}{\|x\|_2}$ is not true.

~~for at least $y = \frac{x}{\|x\|_2}$~~

as it violates, ~~$\langle x,$~~

$$\langle x-z, y-z\rangle \geq 0$$

or $\langle x-z, z-y\rangle \leq 0$ for at least $y = \frac{x}{\|x\|_2}$

But does $z = \frac{x}{\|x\|_2}$ satisfies this. $\forall y \in B$

$$\langle x - z, z - y \rangle = \langle x - \frac{x}{\|x\|_2}, \frac{x}{\|x\|_2} - y \rangle$$

$$= \langle x, \frac{x}{\|x\|_2} \rangle - \langle x, y \rangle - \langle \frac{x}{\|x\|_2}, \frac{x}{\|x\|_2} \rangle + \langle \frac{x}{\|x\|_2}, y \rangle$$

$$= \|x\|_2 + \langle x, y \rangle \left(\frac{1}{\|x\|_2} - 1 \right) - 1$$

$$= (\|x\|_2 - 1) + \langle x, y \rangle \left(\frac{1}{\|x\|_2} - 1 \right)$$

~~given $\|x\|_2 \geq 1$~~

and $\langle x, y \rangle \left(\frac{1}{\|x\|_2} - 1 \right) \geq \|x\|_2 \|y\|_2 \left(\frac{1}{\|x\|_2} - 1 \right)$

we can get

~~$\langle x - z, z - y \rangle \geq \|y\|_2 (1 - \|x\|_2)$~~

$$\langle x - z, z - y \rangle \geq (\|x\|_2 - 1) + \|y\|_2 (1 - \|x\|_2)$$

$$\geq (\|x\|_2 - 1)(1 - \|y\|_2)$$

given $\|x\|_2 \geq 1$ and $\|y\|_2 \leq 1$ as $y \in B$

Scanned with CamScanner

$$\Rightarrow \quad \langle x - z, z - y \rangle \geq (\|x\|_2 - 1) \cancel{\Pi y} (1 - \|y\|_2) \geq 0$$

$$\Rightarrow \quad \cancel{z}\ \Pi_B(x) = z = \frac{x}{\|x\|_2}$$ is the only projection which satisfies the given condition.

$$\Rightarrow \quad \Pi_B(x) = \begin{cases} x & , \text{ if } \|x\|_2^2 \leq 1 \\ \dfrac{x}{\|x\|_2} & , \text{ if } \|x\|_2^2 > 1 \end{cases}$$

1.2) $S = \{ x \in \mathbb{R}^d \mid U^T x = b \}$

s.t. $U^T U = I$

$b \in \mathbb{R}^k$

$\Pi_S(y) = \underset{x \in S}{\operatorname{argmin}} \|x - y\|_2 = \underset{x \in S}{\operatorname{argmin}} \frac{1}{2}\|x-y\|_2^2$

Lagrangian for the constraint minimization,

$f(x) = \frac{1}{2} \|x - y\|_2^2$

$U^T x - b = 0$

$L(x, \alpha) = f(x) + \alpha^T (U^T x - b) \quad \{ \alpha \in \mathbb{R}^{k \times 1} \}$

Using KKT conditions:-

- Stationarity $\quad \nabla_x L(x, \alpha) = \frac{1}{2} \cdot 2 (x - y) + U\alpha = 0$

$\Rightarrow$ ~~$z = x + U\alpha$ — (1)~~

$\Rightarrow x = y - U\alpha$ — (1)

- Primal feasibility

~~$x \in \mathbb{R}$~~

$U^T x - b = 0$

$\Rightarrow U^T (y - U\alpha) - b = 0$

$\Rightarrow U^T y - U^T U \alpha - b = 0$

using $U^T U = I$

$\alpha = U^T y - b$ — ②

Put ② in ①

$x = y - U(-b + U^T y)$

$x = y + Ub - UU^T y$

~~using $U^T U = UU$ I~~

$\boxed{x = Ub}$

$\Rightarrow \boxed{\cancel{\Pi_S(y) = Ub}}$

$\Rightarrow \quad x = (I - UU^T) y + Ub$

$\Rightarrow \boxed{\Pi_S(y) = (I - UU^T) y + Ub}$

Scanned with CamScanner

$$\Delta := \{x \in \mathbb{R}^d \mid \forall i \in [d],\ x_i \geq 0,\ \sum_{i \in [d]} x_i = 1\}$$

$$\Pi_\Delta(y) = \underset{x \in \cancel{\mathbb{R}^d} \Delta}{\operatorname{argmin}} \|x - y\|_2^2 = \underset{x \in \Delta}{\operatorname{argmin}} \frac{1}{2} \|x - y\|_2^2$$

$$f(x) = \frac{1}{2} \|x - y\|_2^2$$

$$\forall i \in [d] \quad -x_i \leq 0$$

$$\sum_{i=1}^{d} x_i = 1$$

Assume ~~g~~ the dimensions of y are arranged in sorted increasing order.

Lagrangian

$$L(x, u, v) = \frac{1}{2} \|x - y\|_2^2 - \sum_{i=1}^{d} u_i x_i + v\left(\sum_{i=1}^{d} x_i - 1\right)$$

Using KKT conditions

~~$\forall i \in [d]$~~

Stationarity:

$$\forall i \in [d] \quad \nabla_{x_i} L(x, u, v) = (x_i - y_i) - u_i + v = 0 \quad - (1)$$

Complimentry slackness

$$\forall i \in [d] \quad -u_i x_i = 0 \quad - (2)$$

Dual feasibility $\forall i \in [d] \quad u_i \geq 0 \quad - (3)$

Scanned with CamScanner

Primal feasibility :-

$$\sum_{i=1}^{d} x_i = 1 \quad -(4) ; \quad x_i \geq 0 \;\; \forall i \in [d] \quad -(5)$$

from (1)

$$x_i = y_i + u_i - v$$

using (5) we get 2 cases,

Case 1 $x_i > 0$ $\Rightarrow$ from (2) $\Rightarrow$ $u_i = 0$
(Let there be ρ such i) $\Rightarrow$ ~~from~~ $\Rightarrow$ $x_i = y_i - v$

Case 2 $x_i = 0$ $\Rightarrow$ $u_i > 0$

Now using (4)

$$\sum_{i=1}^{d} x_i = \sum_{i} 1$$

$$\sum_{i=d-\rho}^{d} x_i = \sum_{i=d-\rho+1}^{d} y_i - \rho v = 1$$

$$\Rightarrow \quad v = \frac{1}{\rho}\left(\sum_{i=d-\rho+1}^{d} y_i - 1\right)$$

$$\Rightarrow \cancel{x_i = y_i - \frac{1}{\rho}\left(\sum_{i=d-\rho}^{d} y_i - 1\right) \;\; \forall i \in \{d-\rho, \dots d\}}$$

$\Rightarrow$ ~~$\forall i \in \{1, \ldots d-\rho-1$~~

$$\Rightarrow \begin{cases} \forall i \in \{1, \ldots d-\rho\} & x_i = 0 \\ \forall i \in \{d-\rho+1, \ldots, d\} & x_i = y_i - \frac{1}{\rho}\left(\sum_{i=d-\rho}^{d} y_i - 1\right) \end{cases}$$

Bonus:- ~~For every~~

The function $\dfrac{\sum_{i=d-\rho}^{d} y_i - 1}{\rho}$ is monotonically increasing ~~for~~ wrt. ρ.

- Sort x_i with merge sort - $O(d \log d)$
- For every $\rho \in \{1, 2 \ldots, d\}$ - $O(d)$

  ~~S, Calculate~~ ~~$\sum_{i=d-\rho}^{d} y_i = \sum_{i=d-(\rho-1)}^{d} y_i + y_{d-\rho}$~~

  $$S_\rho = \sum_{i=d-\rho}^{d} y_i = S_{\rho-1} + y_{d-\rho}$$

  ~~$O(n)$~~
- Do binary search on ρ $\leftarrow O(\log d)$

  $\forall (i)$ Calculate x_i - $O(d)$ time as $\sum_{i=d-\rho}^{d} y_i$ is known

  Check $x \in \Delta$, $\sum_{i=1}^{d} x_i = 1$ and $x_i \geq 0 \ \forall i$ - $O(d)$

  if $x \in \Delta$; return x.

$\Rightarrow$ Total $O(d \log d)$ time complexity.

2.1) $\mathbb{E}[\tilde{\nabla} f(x_t)] = \nabla f(x)$ and $\mathbb{E}[\|\tilde{\nabla} f(x)\|_2^2] \leq G$

update rule :- $x_{t+1} = x_t - \eta_t \tilde{\nabla} f(x_t)$

Starting from the basic mirror descent lemma for SGD.

$$f(x_t) \leq f(x) + \mathbb{E}\left[\frac{1}{2\eta}\left(\|x_t - x\|_2^2 - \|x_{t+1} - x\|_2^2 + \|x_t - x_{t+1}\|_2^2\right)\right]$$

replace η with η_t

$$f(x_t) \leq f(x) + \mathbb{E}\left[\frac{1}{2\eta_t}\left(\|x_t - x\|_2^2 - \|x_{t+1} - x\|_2^2 + \|x_t - x_{t+1}\|_2^2\right)\right]$$

$$\Rightarrow \eta_t f(x_t) \leq \eta_t f(x) + \mathbb{E}\left[\frac{1}{2}\left(\|x_t - x\|_2^2 - \|x_{t+1} - x\|_2^2 + \|x_t - x_{t+1}\|_2^2\right)\right]$$

Taking telescopic sum

$$\sum_{t=0}^{T-1} \eta_t \mathbb{E}[f(x_t)] \leq f(x) \sum_{t=0}^{T-1} \eta_t + \mathbb{E}\left[\frac{1}{2}\left(\|x_0 - x^*\|_2^2 + \sum_{t=0}^{T-1} \|x_t - x_{t+1}\|_2^2\right)\right]$$

$$\cancel{\frac{1}{\sum_{s=0}^{T-1}\eta_t}\sum_{t=0}^{T-1}\mathbb{E}[f(x_t)] \le f(x) + \mathbb{E}\left(\frac{\|x_0-x\|_2^2}{2\eta} + \mathbb{E}\right.}$$

$$\left(\frac{1}{\sum_{s=0}^{T-1}\eta_s}\right)\left(\sum_{t=0}^{T-1}\eta_t\,\mathbb{E}[f(x_t)]\right) \le f(x) + \frac{\|x_0-x\|_2^2 + \mathbb{E}\left[\sum_{t=0}^{T-1}\|x_t-x_{t+1}\|_2^2\right]}{2\sum_{t=0}^{T-1}\eta_t}$$

using the update rule :

$$x_t - x_{t+1} = -\eta_t\tilde{\nabla}f(x_t)$$

$$\Rightarrow \left(\frac{1}{\sum_{s=0}^{T-1}\eta_s}\right)\left(\sum_{t=0}^{T-1}\eta_t\,\mathbb{E}[f(x_t)]\right) \le f(x) + \frac{\|x_0-x\|_2^2 + \sum_{t=0}^{T-1}\mathbb{E}\left[\|\tilde{\nabla}f(x_t)\|_2^2\right]\eta_t^2}{2\sum_{t=0}^{T-1}\eta_t}$$

given $\mathbb{E}\left[\|\tilde{\nabla}f(x_t)\|_2^2\right] \le G$

$$\boxed{\Rightarrow \left(\frac{1}{\sum_{s=0}^{T-1}\eta_s}\right)\left(\sum_{t=0}^{T-1}\eta_t\,\mathbb{E}[f(x_t)]\right) \le f(x) + \frac{\|x_0-x\|_2^2 + G\sum_{t=0}^{T-1}\eta_t^2}{2\sum_{t=0}^{T-1}\eta_t}}$$

Scanned with CamScanner

2.2) Starting from the second last step of Q.(2.1)

and setting $\eta_t = \eta \Rightarrow \sum_{t=0}^{T-1} \eta_t = \eta T$

$$\frac{1}{\eta T}\,\eta\left(\sum_{t=0}^{T-1} \mathbb{E}\left[f(x_t)\right]\right) \leq f(x) + \frac{\|x_0 - x\|_2^2 + \eta^2\, \mathbb{E}\left[\sum_{t=0}^{T-1} \|\tilde{\nabla} f(x_t)\|_2^2\right]}{2\eta T}$$

$$\frac{1}{T}\left(\mathbb{E}\left[\sum_{t=0}^{T-1} f(x_t)\right]\right) \leq f(x) + \frac{\left(\|x_0 - x\|_2^2 \cancel{+ \eta\, 2\eta\, \|\nabla f(x} + \eta^2 \sum_{t=0}^{T-1} \|\nabla f(x_t)\|_2^2\right)}{2\eta T}$$

Now f being a convex function,

$$\mathbb{E}\left(\frac{\sum_{t=0}^{T-1} f(x_t)}{T}\right) \geq f(\bar{x}_T)$$

where $\bar{x}_T = \sum_{t=0}^{T-1} \frac{x_t}{T}$

$$\Rightarrow \mathbb{E}\left[f(\bar{x}_T)\right] \leq f(x) + \frac{\|x_0 - x\|_2^2}{2\eta T} + \frac{2\eta}{T} \sum_{t=0}^{T-1} \|\nabla f(x_t)\|_2^2 \quad -(1)$$

using ~~tower~~ Quadratic upper bound

$$f(x_{t+1}) \leq f(x_t) + \langle \nabla f(x_t), (x_{t+1} - x_t) \rangle + \frac{L}{2} \|x_{t+1} - x_t\|_2^2$$

using the update rule

$$x_{t+1} = x_t - \eta \tilde{\nabla} f(x_t)$$

$$f(x_{t+1}) \leq f(x_t) - \eta \langle \nabla f(x_t), \tilde{\nabla} f(x_t) \rangle + \frac{L}{2} \eta^2 \|\tilde{\nabla} f(x_t)\|_2^2$$

Taking expectation

$$\mathbb{E}[f(x_{t+1})] \leq \mathbb{E}[f(x_t)] - \eta \, \mathbb{E}(\nabla f(x_t))^T \, \mathbb{E}[\tilde{\nabla} f(x_t)] + \frac{L}{2} \eta^2 \, \mathbb{E}[\|\tilde{\nabla} f(x_t)\|_2^2] \quad (1)$$

given,

$$\mathbb{E}[\tilde{\nabla} f(x_t)] = \nabla f(x_t)$$

and $\mathbb{E}[\|\tilde{\nabla} f(x_t)\|_2^2] \leq 2 \|\nabla f(x_t)\|_2^2$

$$\Rightarrow \mathbb{E}[f(x_{t+1})] \leq \mathbb{E}[f(x_t)] - \eta \|\nabla f(x_t)\|_2^2 + L\eta^2 \|\nabla f(x_t)\|_2^2$$

for $\eta \leq \frac{1}{2L} \Rightarrow L \leq \frac{1}{2\eta}$

$$\mathbb{E}[f(x_{t+1})] \leq \mathbb{E}[f(x_t)] - \frac{L\eta}{2} \|\nabla f(x_t)\|_2^2 \quad -(2)$$

Taking telescopic sum,

$$\sum_{t=0}^{T-1} \mathbb{E}[f(x_{t+1})] \le \sum_{t=0}^{T-1} \mathbb{E}[f(x_t)] - \frac{\eta}{2} \sum_{t=0}^{T-1} \|\nabla f(x_t)\|_2^2$$

$$\mathbb{E}[f(x_T)] - \mathbb{E}[f(x_0)] \le -\frac{\eta}{2} \sum_{t=0}^{T-1} \|\nabla f(x_t)\|_2^2$$

$$\Rightarrow \sum_{t=0}^{T-1} \|\nabla f(x_t)\|_2^2 \le \left(\frac{2}{\eta}\right)\left(f(x_0) - \mathbb{E}[f(x_T)]\right)$$

now, $f(x^*) \le f(x_T)$ $\{x^* \text{ is the minimizer}\}$

$$\Rightarrow \mathbb{E}[f(x^*)] = f(x^*) \le \mathbb{E}[f(x_T)]$$

$$\Rightarrow \sum_{t=0}^{T-1} \|\nabla f(x_t)\|_2^2 \le \frac{2}{\eta}\left(f(x_0) - f(x^*)\right) \quad -(3)$$

~~Now using Lipschitz smooth~~

Now using upper quadratic bound on x_0 and x^*

$$f(x_0) \le f(x^*) + \langle \nabla f(x^*), x_0 - x^* \rangle + \frac{L}{2}\|x_0 - x^*\|_2^2$$

given x^* is the minimizer

$$\Rightarrow \nabla f(x^*) = 0$$

$$\Rightarrow f(x_0) - f(x^*) \le \frac{L}{2}\|x_0 - x^*\|_2^2$$

$$\Rightarrow \sum_{t=0}^{T-1} \|\nabla f(x_t)\|_2^2 \le \frac{L}{\eta}\|x_0 - x^*\|_2^2 \quad -(4)$$

Scanned with CamScanner

using (4) in (1)

$\Rightarrow \cancel{\frac{1}{T} \mathbb{E}[f(x_T)]}$

$$\mathbb{E}[f(\bar{x}_T)] \leq f(x) + \frac{\|x_0 - x\|_2^2}{2\eta T} + \frac{2\eta}{T} \frac{L}{\eta} \|x_0 - x^*\|_2^2$$

given $\eta \leq \cancel{1/2} \, 1/2L$

$$\leq f(x) + \frac{\|x_0 - x\|_2^2}{2\eta T} + \frac{\|x_0 - x^*\|_2^2}{\eta T}$$

if ~~x_0 is~~ $x \in \mathbb{R}^d$ is farther away from x_0 than x^*

$$\Rightarrow \|x_0 - x\|_2^2 \geq \|x_0 - x^*\|_2^2$$

$$\Rightarrow \boxed{\mathbb{E}[f(\bar{x}_T)] \leq f(x) + \frac{3}{2} \frac{\|x_0 - x\|_2^2}{\eta T}}$$

~~$\Rightarrow$ The convergence~~

$$\Rightarrow \boxed{\mathbb{E}[f(\bar{x}_T)] \leq f(x) + O\left(\frac{\|x_0 - x\|_2^2}{\eta T}\right)}$$

Scanned with CamScanner

~~#1.2~~

2.3)

$$f(w) = \frac{1}{N} \sum_{i=1}^{N} \log\left(1 + \exp(-y_i \langle w, x_i \rangle)\right)$$

$$\nabla_w f(w) = \frac{1}{N} \sum_{i=1}^{N} \frac{\exp(-y_i \langle w, x_i \rangle)\, \nabla_w\left(-y_i \langle w, x_i \rangle\right)}{1 + \exp(-y_i \langle w, x_i \rangle)}$$

$$\boxed{\nabla_w f(w) = \frac{1}{N} \sum_{i=1}^{N} \frac{\exp(-y_i \langle w, x_i \rangle)\,(-y_i x_i)}{1 + \exp(-y_i \langle w, x_i \rangle)}}$$

Observations :-

- Lower batch size gives a less smooth graph as the gradients have ~~are~~ ~~noisy~~ high variance for lower batch size
- For a given batch size as we increase the learning rate the loss drops more quickly compared to lower ones.
- As the batch size increases, the member of operations required to calculate gradients will increase and ~~there~~ thus there is an ~~slight~~ increase in training time for higher batch sizes.

Scanned with CamScanner

3.1) $f(w) = f_1(w) + f_2(w)$

$f_1(w) = 2(w-1)^2 - w^2$
$f_2(w) = 4(w+1)^2 - w^2$

Global minimum of f

$\nabla_w f(w) = \nabla f_1(w) + \nabla_w f_2(w) = 0$

$= 4(w-1) - 2w + 8(w+1) - 2w = 0$

$= 2w - 4 + 6w + 8 = 0$

$\Rightarrow \boxed{w = -\frac{4}{8} = -1/2}$

Naive ADMM :-

$$w_j^{(t+1)} = \underset{w}{\arg\min} \left\{ f_j(w) + \lambda \| w - w^{(t)} \|_2^2 \right\}$$

iteration 1:-

~~for j = 1~~

$\cancel{w_1^{(1)} = \underset{w_1}{\arg\min} \left\{ 2(w_1 - 1)^2 - w_1^2 + \lambda \| w_1 - w^{(0)} \|_2^2 \right\}}$

$\cancel{\nabla_{w_1} f_1}$

iteration 1 : $w^{(0)} = 0$, $\lambda = 1$

$$w_1^{(1)} = \underset{w}{\text{argmin}} \left(2(w-1)^2 - w^2 + \|w - w^{(0)}\|_2^2 \right)$$

setting $\nabla_w = 4(w-1) - 2w + 2(w - w^{(0)}) = 0$

$$2w + 2w - 4 = 0$$

$$w = 1$$

$$\Rightarrow w_1^{(1)} = 1$$

$$w_2^{(1)} = \underset{w}{\text{argmin}} \left(4(w+1)^2 - w^2 + \|w - w^{(0)}\|_2^2 \right)$$

setting $\nabla_w = 8(w+1) - 2w + 2(w - w^{(0)}) = 0$

$$\Rightarrow 8w + 8 = 0$$

$$\Rightarrow w = -1$$

$$\Rightarrow w_2^{(1)} = -1$$

$$\Rightarrow w^{(1)} = \frac{1}{2}\left(w_1^{(1)} + w_2^{(1)}\right) = 0$$

$$\Rightarrow \boxed{w^{(1)} = 0}$$

as $w^{(1)} = w^{(0)}$

$\Rightarrow w_1^{(2)} = w_1^{(1)} = 1$

and $w_2^{(2)} = w_2^{(1)} = -1$

Scanned with CamScanner

$\Rightarrow \quad w_2^{(2)} = 0$

$\Rightarrow \quad w^{(i)} = 0 \quad \forall\, i \in \mathbb{N}$

i.e. w converges to $w = 0$

$\Rightarrow$ Clearly the naive ADMM does not converge to the global minimum.

3.2) $\min_{\forall w_j, w} \frac{1}{m} \sum_{j=1}^{m} \left(f_j(w_j) + \lambda \|w_j - w\|_2^2 \right)$

s.t. $w_j = w \quad \forall j$

Dual problem

$$L(\alpha, w_j, w) = \frac{1}{m} \sum_{j=1}^{m} \left[\left(f_j(w_j) + \lambda \|w_j - W\|_2^2 \right) + \alpha_j (w_j - w) \right]$$

$$\nabla_{w_1} L = 4w_1 - 4 - 2w_1 + 2(w_1 - w) + \alpha_1 = 0$$

$$\Rightarrow \quad \alpha_1 = 2w - 4w_1 + 4 \quad - (1)$$

$$\nabla_{w_2} L = 8w_2 + 8 - 2w_2 + 2(w_2 - w) + \alpha_2 = 0$$

$$\Rightarrow \quad \alpha_2 = 2w - 8w_2 - 8 \quad - (2)$$

$$\nabla_w L = 2(w_1 - w) + 2(w_2 - w) - \alpha_1 - \alpha_2 = 0 \quad - (3)$$

Put $w_1 = w_2 = w$ in (1), ~~and~~ (2), and (3)

$$\Rightarrow \quad \alpha_1 = -2w + 4 \quad - (4)$$

$$\alpha_2 = -6w - 8 \quad - (5)$$

$$\alpha_1 = -\alpha_2 \quad - (6)$$

Scanned with CamScanner

using (4) and (5) in (6)

$$-2w + 4 = 6w + 8$$

$$\Rightarrow \quad 8w = -4$$

$$\Rightarrow \quad \boxed{w = -1/2}$$

and $\boxed{\alpha_1 = 5}$ and $\boxed{\alpha_2 = -5}$

Clearly the optimal dual solution $\boxed{w = -1/2}$ is the same as the primal optimal solution obtained in (3.1) when $f(w)$ was directly minimized.

4.1) $h(x) = \lambda|x| \quad , \quad \lambda > 0$

$$\text{prox}_h(x) = \underset{z}{\text{argmin}} \left\{ h(z) + \frac{1}{2}\|z - x\|_2^2 \right\}$$

$$= \underset{z}{\text{argmin}} \left\{ \lambda|z| + \frac{1}{2}\|z - x\|_2^2 \right\}$$

Let $f(z) = \lambda|z| + \frac{1}{2}\|z-x\|_2^2$

$\Rightarrow$ ~~$\nabla_z f(z)$~~ for $z \neq 0$

$$\nabla_z f(z) = \lambda \, \text{sgn}(z) + (z - x) = 0$$

$$\Rightarrow z = x - \lambda \, \text{sgn}(z)$$

Case 1 $\quad x > \lambda$

$\Rightarrow \quad z = x - \lambda \, \text{sgn}(z) > \lambda - \lambda \, \text{sgn}(z)$

if $z > 0$

$\Rightarrow \quad z > \lambda - \lambda \Rightarrow \boxed{z > 0}$

~~or sgn(z) = sgn(x)~~

if $z < 0$

$\Rightarrow z > \lambda + \lambda \Rightarrow z > 2\lambda > 0$

but z has to be < 0 (contradiction)

$\Rightarrow$ for $x > \lambda$, $z > 0$

~~$\Rightarrow \text{sgn}(x) = \text{sgn}(z)$~~

$\Rightarrow \text{sgn}(z) = \text{sgn}(x)$

Case 2 $\cancel{\beta}\ x < -\lambda$

$$\Rightarrow \quad z = \cancel{\beta}\ x - \cancel{\text{sgn}}(z) < -\lambda - \text{sgn}(z)$$

if $z > 0$

$$\Rightarrow \quad z < -\lambda - \lambda \quad \Rightarrow z < -2\lambda < 0$$

contradiction

if $z \cancel{>} < 0$

$$\Rightarrow z < -\lambda + \lambda \Rightarrow z < 0$$

$\Rightarrow$ for $x < -\lambda$, $z < 0$

$$\Rightarrow \text{sgn}(z) = \text{sgn}(x)$$

Case 3 when $-\lambda \leq x \leq \lambda$

$$\Rightarrow -\lambda - \lambda\,\text{sgn}(z) \leq z = x - \lambda\,\text{sgn}(z) \leq \lambda - \lambda\,\text{sgn}(z)$$

if $z > 0$

$$\Rightarrow \quad \cancel{-2\lambda} \leq z \leq 0 \quad \text{(contradiction)}$$

if $z < 0$

$$\Rightarrow \quad 0 \leq z \leq 2\lambda \quad \text{(contradiction)}$$

$\Rightarrow$ There is no valid z which can be obtained when $x \in [-\lambda, \lambda]$ when we set $\nabla_z f(z) = 0$

As both $z<0$ and $z>0$ ~~a~~ can't be minimizer of $f(z)$ when $x \in [-\lambda, \lambda]$

$\Rightarrow$ $z=0$ has to be the minimizer when $x \in [-\lambda, \lambda]$

$$\Rightarrow \boxed{\operatorname{prox}_h(x) = \begin{cases} x - \lambda \operatorname{sgn}(x) & , x > |\lambda| \\ 0 & , \text{otherwise} \end{cases}}$$

Scanned with CamScanner

4.2) $h(x) = \lambda \|x\|_1$, $x \in \mathbb{R}^d$

update rule $\rightarrow$ $x_{t+1} = \text{prox}_{\eta h}(x_t)$

~~given~~ From Lecture 10, slide 15,

$$\forall i \in [d]; \quad [x_{t+1}]_i = \left[\text{prox}_{\eta h}(x_t)\right]_i = \begin{cases} [x_t]_i - \eta\lambda \,\text{sgn}([x_t]_i), & |[x_t]_i| > \eta\lambda \\ 0 & , \text{otherwise} \end{cases}$$

$\Rightarrow \forall i \in [d]$

if $|[x_t]_i| > \eta\lambda$, $\Rightarrow$ $[x_{t+1}]_i = [x_t]_i - \eta\lambda\,\text{sgn}([x_t]_i)$

$\cancel{\Rightarrow |[x_{t+1}]_i| = |[x_t]_i|}$

and as soon as $|[x_t]_i| \leq \eta\lambda$

$|[x_{t+1}]_i| = 0$

So starting from some $|[x_0]_i|$, in m number of steps where $m = \left\lceil \frac{|[x_0]_i|}{\eta\lambda} - 1 \right\rceil$ and $\lceil y \rceil$ is the ceiling function,

$$|[x_m]_i| \leq \eta\lambda$$

$$\Rightarrow |[x_{m+1}]_i| = 0$$

Scanned with CamScanner

~~and once $\|[x_{t+1}]\|_1 = 0$~~

$\Rightarrow$ ~~the~~ It will take $\left\lceil \frac{\max_i |[x_0]_i|}{\eta\lambda} - 1 \right\rceil$ steps to take ~~reduce~~ all dimensions to ~~0~~ exact 0.

$\Rightarrow$ It will take $O\left(\frac{\max_i |[x_0]_i|}{\eta\lambda}\right)$ steps.

Now, $\frac{\max_i |[x_0]_i|}{\eta\lambda} = \|x_0\|_\infty \leq \|x_0\|_2$

$\Rightarrow$ It will take $O\left(\frac{\|x_0\|_2}{\eta\lambda}\right)$ ~~St~~ steps

to converge to exact 0.

Using gradient descent

In this case ~~the~~ the ~~gradient~~ update is still the same for $|[x_t]_i| > \eta\lambda$ but when $|[x_t]_i| \leq \eta\lambda$

the next $|[x_{t+1}]_i|$ is not set to 0 but rather follows the same update.

In this case the dimension can overshoot the exact ~~mi~~ minimizer and keep on oscillating.

eg:- Let $x \in \mathbb{R}^2$, $x_0 = [2, -4]$

and Let $\eta\lambda = 5/3$

$\Rightarrow$ $i = 1$

$$[x_1]_1 = 2 - 5/3 = \frac{1}{3}$$

$$[x_1]_2 = -4 + 5/3 = -7/3$$

$i = 2$

$$[x_2]_1 = 1/3 - 5/3 = -4/3$$

$$[x_2]_2 = -7/3 + 5/3 = -2/3$$

$i = 3$

$$[x_3]_1 = -4/3 + 5/3 = 1/3$$

$$[x_3]_2 = -2/3 + 5/3 = 1$$

$i = 4$

$$[x_3]_1 = 1/3 - 5/3 = -4/3$$

$$[x_3]_2 = 1 - 5/3 = -2/3$$

Clearly both the dimensions have started oscillating around 0 without converging.