Parallel and distributed algorithm

7:55 PM

Page1: KEY: split the objective into two terms
$\downarrow$
atleast one seperable $\rightarrow$ evaluate proximal operator in parallel

5.1. Problem Structure

$[n] = \{1, \ldots, n\}$

Parallel_and_distribu...

Parallel Distributed Alg Part 1

$C \subseteq \{1, \ldots, n\}$ # e.g. $C = \{1, 3, 5, 7\}$,

$x_C \in \mathbb{R}^{|C|}$ # e.g. $x_C = \{x_1, x_3, x_5, x_7\}$

$(\mathcal{P} = \{C_1, C_2, \ldots, C_N\} : C_i \subseteq \{1, \ldots, n\})$ {partition of $\{1, \ldots, n\}$} $\Leftrightarrow \bigcup_{i \in \text{indexset}(\mathcal{P})} C_i = \{1, \ldots, n\}$, $\forall_{i \neq j}\ C_i \cap C_j = \emptyset$, $\forall_i\ C_i \neq \emptyset$

$f\{: \mathbb{R}^n \to \mathbb{R}\}$ {$\mathcal{P}$ seperable} $\Leftrightarrow f(x) = \sum_{i=1}^{N} f_i(x_{C_i})$: $(f_i : \mathbb{R}^{|C_i|} \to \mathbb{R}$, $\underbrace{x_{C_i}}_{C_i \text{ \{scope of } f_i\}}$: subvector of x with indices in $C_i)$

e.g. $\underbrace{\mathcal{P}}_{\text{partition}} = \{\underbrace{\{1,3,5\}}^{C_1}, \underbrace{\{2,4\}}^{C_2}\}$, $n = 5$, $x_{C_1} = x_{\{1,3,5\}} = (x_1, x_3, x_5)$,
$x_{C_2} = x_{\{2,4\}} = (x_2, x_4)$

$f(x) = \underbrace{\frac{x_1^2 + x_3^2}{x_5}}_{f_1(x_{C_1})} + \underbrace{e^{x_2 + x_4}}_{f_2(x_{C_2})}$

Page2 *Full seperability: • key property of proximal mapping: $f(x,y) = \varphi(x) + \psi(y) \Rightarrow \mathrm{prox}_f(v,w) = \begin{bmatrix} \mathrm{prox}_{\varphi}(v) \\ \mathrm{prox}_{\psi}(w) \end{bmatrix}$ # $\mathrm{prox}_f(x) = \underset{z}{\mathrm{argmin}}\left(f(z) + \frac{1}{2}\|z - x\|_2^2\right)$

$(f$: fully seperable$) \Leftrightarrow (f = \sum_{i=1}^{N} f_i(x_{C_i}),\ C_i = \{i\})$ $f(x) = \sum_{i=1}^{N} f_i(x_i) \Rightarrow \mathrm{prox}_f(x) = \left[(\mathrm{prox}_f(x))_i\right]_{i=1}^{N} = \left[\mathrm{prox}_{f_i}(x_i)\right]_{i=1}^{N}$

implications of seperability: proximal operator breaks into N smaller operations, can be carried out independently in parallel:

for a $\mathcal{P}$-seperable function, $\mathrm{prox}_f(x) = \left[(\mathrm{prox}_{f_i}(x_{C_i}))_i\right]_{i=1}^{n} = \begin{bmatrix} \mathrm{prox}_{f_1}(\square_{C_1}) \\ \mathrm{prox}_{f_2}(\square_{C_2}) \\ \vdots \\ \mathrm{prox}_{f_N}(\square_{C_N}) \end{bmatrix}(x) = \begin{bmatrix} \mathrm{prox}_{f_1}(x_{C_1}) \\ \mathrm{prox}_{f_2}(x_{C_2}) \\ \vdots \\ \mathrm{prox}_{f_N}(x_{C_N}) \end{bmatrix}$

$: f(\square) = \sum_{i=1}^{N} f_i(\square_{C_i})$

Problem in consideration: epi $f : (\neq \emptyset, [\,], \triangledown)$

$\underset{x}{\mathrm{minimize}}\ f(x) + g(x)$: $(f, g : (\mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}, [\,], \text{proper}, \triangledown))$
often g is the indicator function of underlying convex set.

Assumption:

$\mathcal{P} = \{C_1, \ldots, C_N\}$: partition of $\{1, \ldots, n\}$

$\mathcal{Q} = \{d_1, \ldots, d_M\}$: " " $\{1, \ldots, n\}$

$f : (\mathcal{P}$ seperable$) \Leftrightarrow f(x) = \sum_{i=1}^{N} f_i(x_{C_i})$

$g : (\mathcal{Q}$ seperable$) \Leftrightarrow g(x) = \sum_{j=1}^{M} g_j(x_{d_j})$

$\underset{x}{\mathrm{minimize}} \sum_{i=1}^{N} f_i(x_{C_i}) + \sum_{j=1}^{M} g_j(x_{d_j})$ # $f_i : \mathbb{R}^{|C_i|} \to \mathbb{R} \cup \{+\infty\}$
$g_j : \mathbb{R}^{|d_j|} \to \mathbb{R} \cup \{+\infty\}$
i associated with f blocks
j " " g "

ADMM

(eg: ADMM from prox alg (1.1.2))

$x^{k+1} = \mathrm{prox}_{\lambda f}(z^k - u^k)$ # now $f(x) = \sum_{i=1}^{N} f_i(x_{C_i}) \Rightarrow \lambda f(x) = \sum_{i=1}^{N} \lambda f_i(x_{C_i})$: $\mathcal{P}$seperable $\Rightarrow \mathrm{prox}_{\lambda f}(\square) = \left[\mathrm{prox}_{\lambda f_i}(\square_{C_i})\right]_{i=1}^{n}$

$\therefore \mathrm{prox}_{\lambda f}(z^k - u^k) = \left[\mathrm{prox}_{\lambda f_i}(\underbrace{(z^k - u^k)_{C_i}}_{z_{C_i}^k - u_{C_i}^k})\right]_{i=1}^{N} = \left[\underbrace{\mathrm{prox}_{\lambda f_i}(z_{C_i}^k - u_{C_i}^k)}\right]_{i=1}^{N}$

$\Leftrightarrow$

$\forall_{i \in \{1, \ldots, N\}}$

$x_{C_i}^{k+1} = \mathrm{prox}_{\lambda f_i}(z_{C_i}^k - u_{C_i}^k)$

note that this one will produce the iterate associated with C_i indexed components of $x^{k+1} = \left[x_{C_i}^{k+1}\right]_{i=1}^{N}$

So, we have $\forall_{i \in \{1, \ldots, N\}}$ $x_{C_i}^{k+1} = \mathrm{prox}_{\lambda f_i}(z_{C_i}^k - u_{C_i}^k)$

$z^{k+1} = \mathrm{prox}_{\lambda g}(x^{k+1} + u^k)$ # similarly $\Leftrightarrow$ $\forall_{j \in \{1, \ldots, M\}}$ $z_{d_j}^{k+1} = \mathrm{prox}_{\lambda g_j}(x_{d_j}^{k+1} + u_{d_j}^k)$

$u^{k+1} = u^k + x^{k+1} - z^{k+1}$

So we have:

$x_{C_i}^{k+1} = \mathrm{prox}_{\lambda f_i}(z_{C_i}^k - u_{C_i}^k) \Big|_{i=1}^{N}$ # N updates carried out independently in parallely

$z_{d_j}^{k+1} = \mathrm{prox}_{\lambda g_j}(x_{d_j}^{k+1} + u_{d_j}^k) \Big|_{j=1}^{M}$ # M " " " " " "

$u^{k+1} = u^k + x^{k+1} - z^{k+1}$ # The final step is trivially parallelizable

[ADMM for distributed optimization]

Page3:

$\tilde{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$ [# $\forall_i\ x_i \in \mathbb{R}^n$] $= \begin{bmatrix} \tilde{x}[1:n] \\ \tilde{x}[n+1:2n] \\ \vdots \\ \tilde{x}[1+(N-1)n : Nn] \end{bmatrix}$ [# define $\forall_{i \in \{1, \ldots, N\}}\ C_i := [1 + (i-1)n, in]$] $= \begin{bmatrix} \tilde{x}_{C_1} \\ \tilde{x}_{C_2} \\ \vdots \\ \tilde{x}_{C_N} \end{bmatrix}$

$\underbrace{1+(N-1)n}$: jth term $= j + (j-1)n = 1 + (j-1)n$

[Consensus using Proximal ADMM]

Compare with
Consensus using Monotone Operator Splitting

Consensus: # Hello world! of parallel and distributed algorithm

consensus constraint: all the local variables have to agree

5.2.1. Global Consensus: $x \in \mathbb{R}^n$

$\left(\underset{x}{\mathrm{minimize}}\ f(x) = \sum_{i=1}^{N} f_i(x)\right) = \left(\underset{x_i}{\mathrm{minimize}} \underbrace{\sum_{i=1}^{N} f_i(x_i)}_{\text{seperable objective}}\right) : \overbrace{x_1 = x_2 = \ldots = x_N}$ $) : \forall_{i \in \{1, \ldots, N\}}\ x_i \in \mathbb{R}^n$ $\in \mathbb{R}^{nN}$

#standard form

in enumerated variable way: এখানে আমরা single vector $\tilde{x}$ use করব structurally এভাবে:

$x_1 = (\tilde{x}(1), \ldots, \tilde{x}(n))$, $x_2 = (\tilde{x}(n+1), \ldots, \tilde{x}(2n)), \ldots, x_N = (\tilde{x}(1 + (N-1)n), \ldots, \tilde{x}((N-1)n + n))$, so as a single variable, $\tilde{x} = (x_1, \ldots, x_N) = (\tilde{x}_{C_1}, \ldots, \tilde{x}_{C_N})$

Page4: $= \left(\underset{x}{\mathrm{minimize}} \sum_{i=1}^{N} f_i(x_i) + I_C(x_1, \ldots, x_N)\right) : C = \{(x_1, \ldots, x_N) \mid x_1 = \ldots = x_N\}$

then consensus constraint becomes: $x_1 = \ldots = x_i = \ldots = x_N \xrightarrow{\text{set}} (C = \{(x_1, \ldots, x_N) \mid x_1 = \ldots = x_N\})$

$\Leftrightarrow \tilde{x}_{C_1} = \ldots = \tilde{x}_{C_i} = \ldots = \tilde{x}_{C_N}$

= componentwise

$I_C(\tilde{x}) = \sum_{j=1}^{n} I_{G_j}(\tilde{x}_{d_j})$

$\left(\begin{array}{l} \tilde{x}(1) = \tilde{x}(n+1) = \ldots = \tilde{x}(1 + (N-1)n) \\ \vdots \\ \tilde{x}(i) = \tilde{x}(n+i) = \ldots = \tilde{x}(i + (N-1)n) \\ \vdots \\ \tilde{x}(n) = \tilde{x}(2n) = \ldots = \tilde{x}(n + (N-1)n) \end{array}\right) \xrightarrow{\text{sets}} \left(G_j = \{\underbrace{(\tilde{x}(j), \tilde{x}(j+n), \ldots, \tilde{x}(j + (N-1)n))}_{\tilde{x}_{d_j}} \mid \tilde{x}(j) = \tilde{x}(j+n) = \ldots = \tilde{x}(j + (N-1)n)\}\Big|_{j=1}^{n}\right)$

• # handling $f(x)$ thus designing $\mathcal{P}$

$\mathcal{P} = \{\underbrace{\{1, \ldots, n\}}_{[n]}, \underbrace{\{n+1, \ldots, 2n\}}_{n + \{1, \ldots, n\} = n + [n]}, \ldots, \underbrace{\{1 + (i-1)n, \ldots, in\}}_{(i-1)n + \{1, \ldots, n\} = (i-1)n + [n]}, \ldots, \underbrace{\{1 + (N-1)n, \ldots, Nn\}}_{(N-1)n + \{1, \ldots, n\} = (N-1)n + [n]}\}$

distribution of indices of $\tilde{x}$

1	$n+1$ .. $n(i-1)+1$	. . .	$(N-1)n+1$	$\to d_1$
2	$n+2$ .. $n(i-1)+2$		$(N-1)n+2$	$\to d_2$
⋮	⋮	⋮	⋮	
j	$n+j$ .. $n(i-1)+j$		$(N-1)n+j$	$\to d_j$
⋮	⋮	⋮	⋮	
n	$2n$ .. $n(i-1)+n$	..	$(N-1)n+n$	$\to d_n$
$\downarrow C_1$	$\downarrow C_2$	$\downarrow C_i$	$\downarrow C_N$	

$P = \{ \underbrace{\{1,\dots,n\}}_{[n]}, \underbrace{\{n+1,\dots,2n\}}_{n+\{1,\dots,n\} = n+[n]}, \dots, \underbrace{\{1+(i-1)n,\dots,in\}}_{(i-1)n+\{1,\dots,n\} = (i-1)n+[N]}, \dots, \underbrace{\{1+(N-1)n,\dots,Nn\}}_{(N-1)n+\{1,\dots,n\} = (N-1)n+[N]} \}$

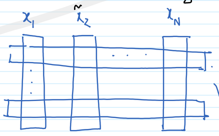
$= \{ [n], n+[n], \dots, \underset{\substack{\| \\ (i-1)n+[N]}}{(i-1)n+[n]}, \dots, (N-1)n+[n] \} = \{c_1, c_2, \dots, c_i, \dots, c_N\}$ $\therefore \sum_{i=1}^{N} f_i(x_i) = \sum_{i=1}^{N} f_i(\tilde{x}_{c_i})$

$$\begin{pmatrix} \tilde{x}(1) = \tilde{x}(n+1) = \dots = \tilde{x}(1+(N-1)n) \\ \vdots \\ \tilde{x}(i) = \tilde{x}(n+i) = \dots = \tilde{x}(i+(N-1)n) \\ \vdots \\ \tilde{x}(n) = \tilde{x}(2n) = \dots = \tilde{x}(n+(N-1)n) \end{pmatrix} \xrightarrow{\text{sets}} \left(G_j = \{ \underbrace{(\tilde{x}(j), \tilde{x}(j+n), \dots, \tilde{x}(j+(N-1)n))}_{\tilde{x}_{d_j}} \mid \tilde{x}(j) = \tilde{x}(j+n) = \dots = \tilde{x}(j+(N-1)n) \} \Big|_{j=1}^{n} \right)$$

$$\begin{array}{c} \begin{bmatrix} n & 2n \dots n(i-1)+n \dots (N-1)n+n \end{bmatrix} \to d_n \\ \downarrow \quad \downarrow \quad \downarrow \quad \downarrow \\ c_1 \quad c_2 \quad c_i \quad c_N \end{array}$$

- # Handling $g(x) = I_C(x)$, thus designing Q

$Q = \bigcup_{i=1}^{n} \{i, n+i, 2n+i, \dots, (N-1)n+i\} = \{d_1, d_2, \dots, d_j, \dots, d_n\}$



Column$^{(i)}$ index correspond to c_i

Row$_{(j)}$ index " " d_j

$\therefore \sum_{i=1}^{N} f_i(x_i) = \sum_{i=1}^{N} f_i(\tilde{x}_{c_i})$: $c_i = [n] + (i-1)n \quad \forall_{i \in \{1,\dots,N\}}$

and $I_C(x_1, \dots, x_N) = I_C(\tilde{x}) = \sum_{j=1}^{n} I_{G_j}(\tilde{x}_{d_j})$: $G_j = \{ \tilde{x}_{d_j} \mid \tilde{x}(j) = \tilde{x}(n+j) = \dots = \tilde{x}((N-1)n+j) \}$; $d_j = \{j, n+j, 2n+j, \dots, (N-1)n+j\} \quad \forall_{j \in \{1,\dots,n\}}$

so in the $\tilde{x}$ variable the optimization problem becomes (to write it in form)

$\min \sum_{i=1}^{N} \underset{\substack{\downarrow \\ f_i}}{f_i(\tilde{x}_{c_i})} + \sum_{j=1}^{n} \underset{\substack{\downarrow \\ g_j}}{I_{G_j}(\tilde{x}_{d_j})}$

$\rightleftarrows$ ADMM

first iteration $\to \forall_{i \in \{1,\dots,N\}}$ $\tilde{x}_{c_i}^{k+1} = \text{prox}_{\lambda f_i}(\tilde{z}_{c_i}^{k} - u_{c_i}^{k})$ as $\tilde{x}_{d_j}$ is the rowwise elements of $\tilde{x}$, so $\tilde{x}_{d_j}^{k+1}$ can be calculated from $\tilde{x}_{c_i}^{(k+1)}$

second iteration $\to \forall_{j \in \{1,\dots,n\}}$ $z_{d_j}^{k+1} = \text{prox}_{\lambda g_j}(\tilde{x}_{d_j}^{k+1} + u_{d_j}^{k})$ [$g_j = I_{G_j}$] [$\text{prox}_{\lambda I_\chi}(\square) = \Pi_\chi(\square)$] $= \Pi_{I_{G_j}}(x_{d_j}^{k+1} + u_{d_j}^{k})$ [$\Pi_{y_1 = \dots = y_n : y_j \in \mathbb{R}}(\xi) = \frac{1}{n}(1^T\xi)\mathbb{1}$] $= \frac{1}{|d_j|}\left(1^T(\tilde{x}_{d_j}^{k+1} + u_{d_j}^{k})\right)\mathbb{1} = \frac{1}{N} 1^T \left(\begin{bmatrix} \tilde{x}(j) \\ \tilde{x}(n+j) \\ \vdots \\ \tilde{x}((N-1)n+j) \end{bmatrix}^{k+1} + \begin{bmatrix} u(j) \\ u(n+j) \\ \vdots \\ u((N-1)n+j) \end{bmatrix}^{k} \right) \mathbb{1}$ [let $\left(\frac{1}{N}\sum_{i=1}^{N} \square\right)\mathbb{1} = \bar{\square}$]

$= \bar{\tilde{x}}_{d_j}^{k+1} + \bar{u}_{d_j}^{k}$ % remember this averages are vectors themselves

$d_j = \{0+j, n+j, 2n+j, \dots, (N-1)n+j\}$
$|d_j| = N$

note: $c_i = \{(N-1)i + [n]\} = \{(N-1)i + \{1,\dots,n\}\}$

$d_j = \{j, n+j, \dots, (N-1)i+j, \dots, (N-1)n+j\}$

distribution of indices of $\tilde{x} \in \mathbb{R}^{nN}$

$$\begin{array}{c} \begin{bmatrix} 1 & n+1 \dots n(i-1)+1 \dots (N-1)n+1 \\ 2 & n+2 \dots n(i-1)+2 \quad (N-1)n+2 \\ \vdots & \vdots \quad \vdots \quad \vdots \\ j & n+j \dots n(i-1)+j \quad (N-1)n+j \\ \vdots & \vdots \quad \vdots \quad \vdots \\ n & 2n \dots n(i-1)+n \dots (N-1)n+n \end{bmatrix} \begin{matrix} \to d_1 \\ \to d_2 \\ \\ \to d_j \\ \\ \to d_n \end{matrix} \\ \downarrow \quad \downarrow \quad \downarrow \quad \downarrow \\ c_1 \quad c_2 \quad c_i \quad c_N \end{array}$$

So, $z_{d_j}^{k+1}$ step এর করার আগে আমরা তা থেকে $z_{c_i}^{k}$ construct করতে পারবো:

then the third iteration :

$u^{k+1} = u^{k} + \tilde{x}^{k+1} - z^{k+1} \in \mathbb{R}^{nN}$

so we can partition them rowwise : # Row-wise partition korar uddeshsho holo $u_{d_j}^{k+1}$ e $z_{d_j}^{k+1}$ er value boshie $u_{d_j}^{k+1}$ er average vector shomkorke ekti surprising info ber kora, the end goal is iteration 3tir jaigai duti kora, both $x_{c_j}^{k+1}$, $u_{d_j}^{k+1}$ e $z_{d_j}^{k+1}$ er value bebohar kora

$\forall_{j \in \{1,\dots,n\}}$ $u_{d_j}^{k+1} = u_{d_j}^{k} + \tilde{x}_{d_j}^{k+1} - z_{d_j}^{k+1}$

$z_{d_j}^{k+1} = \bar{\tilde{x}}_{d_j}^{k+1} + \bar{u}_{d_j}^{k}$, so all components of $z_{d_j}^{k+1}$ are the same

$\therefore \bar{z}_{d_j}^{k+1} = z_{d_j}^{k+1}$

$\bar{\square} = \frac{1}{N}(1^T\square)\mathbb{1}$

$\bar{u}_{d_j}^{k+1} = \bar{u}_{d_j}^{k} + \bar{\tilde{x}}_{d_j}^{k+1} - \bar{z}_{d_j}^{k+1} = \bar{u}_{d_j}^{k} + \bar{\tilde{x}}_{d_j}^{k+1} - \bar{\tilde{x}}_{d_j}^{k+1} - \bar{u}_{d_j}^{k} = 0$

$\bar{z}_{d_j}^{k+1} = z_{d_j}^{k+1} = \bar{\tilde{x}}_{d_j}^{k+1} + \bar{u}_{d_j}^{k}$

$\therefore \forall_{k \in \{0,\dots\}}$ $\bar{u}_{d_j}^{k+1} = 0$

So, $z_{d_j}^{k+1} = \bar{\tilde{x}}_{d_j}^{k+1} + \bar{u}_{d_j}^{k} \quad \forall_{k \in \{0,\dots,\}}$

$\forall_{k \in \{0,\dots,\}}$ $z_{d_j}^{k+2} = \bar{\tilde{x}}_{d_j}^{k+2} + \bar{u}_{d_j}^{k+1} = \bar{\tilde{x}}_{d_j}^{k+2}$

and set $\bar{u}_{d_j}^{0} = 0$ # the first iterate is upto us.

$\therefore \forall_{k \in \{0,\dots,\}}$ $\bar{u}_{d_j}^{k} = \bar{0}$ // rowwise sum of u_{d_j} is $\bar{0}$ at every iteration

$\therefore z_{d_j}^{k+1} = \bar{\tilde{x}}_{d_j}^{k+1} + \bar{u}_{d_j}^{k} = \bar{\tilde{x}}_{d_j}^{k+1}$

$\forall_{i \in \{1,\dots,N\}}$ $\tilde{x}_{c_i}^{k+1} = \text{prox}_{\lambda f_i}(z_{c_i}^{k} - u_{c_i}^{k}) = \text{prox}_{\lambda f_i}(\bar{\tilde{x}}_{c_i}^{k} - u_{c_i}^{k})$

$\forall_{j \in \{1,\dots,n\}}$ $z_{d_j}^{k+1} = \bar{\tilde{x}}_{d_j}^{k+1}$ # After this iteration we have the rows of $z \in \mathbb{R}^{nN}$, so redistributing the elements columnwise we can construct $z_{c_i}^{k+1} = \text{ith column} \begin{bmatrix} - z_{d_1}^{T\,k+1} - \\ \vdots \\ - z_{d_n}^{T\,k+1} - \end{bmatrix} = \text{ith column} \begin{bmatrix} - \bar{\tilde{x}}_{d_1}^{(k+1)T} - \\ \vdots \\ - \bar{\tilde{x}}_{d_n}^{(k+1)T} - \end{bmatrix} = \bar{\tilde{x}}_{c_j}^{(k+1)}$ #

$z^{k+1} = \begin{bmatrix} (z_{d_1}^{k+1})^T \\ \vdots \\ (z_{d_n}^{k+1})^T \end{bmatrix} = \begin{bmatrix} -(\bar{\tilde{x}}_{d_j}^{k+1})^T- \\ \vdots \\ -(\bar{\tilde{x}}_{d_n}^{k+1})^T- \end{bmatrix} = \begin{bmatrix} \mid & & \mid \\ \bar{\tilde{x}}_{c_i} & \cdots & \bar{\tilde{x}}_{c_n} \\ \mid & & \mid \end{bmatrix}$

$\Downarrow$ reconstruct along c_i

$z_{c_i}^{k+1} = \bar{\tilde{x}}_{c_i}^{k+1}$ # Caution: in $\bar{\tilde{x}}_{c_i}^{k+1}$ all the terms are not the same, it just means that it is C_i wise construction of z^{k+1}

$$z_{c_i}^{k+1} = \bar{\tilde{x}}_{c_i}^{k+1}$$ # Caution: in $\bar{\tilde{x}}_{c_i}^{k+1}$ all the terms are not the same, it just means that it is C_i wise construction of $\bar{z}^{k+1}$

$$\forall_{j\in\{1,\dots,n\}} \quad u_{d_j}^{k+1} = u_{d_j}^{k} + \tilde{x}_{d_j}^{k+1} - \bar{\tilde{x}}_{d_j}^{k+1}$$

so, finally:

$$\forall_{i\in\{1,\dots,N\}} \quad \tilde{x}_{c_i}^{k+1} = \mathrm{prox}_{\lambda f_i}\left(\bar{\tilde{x}}_{c_i}^{k} - u_{c_i}^{k}\right)$$ # $= \underset{z}{\mathrm{argmin}} \left(\lambda f_i(z) + \frac{1}{2}\|z-\square\|_2^2\right)\Big|_{\square = \bar{\tilde{x}}_{c_i}^{k} - u_{c_i}^{k}}$: pulls the variables towards the average value while attempting to minimize each local f_i

$$\forall_{j\in\{1,\dots,n\}} \quad u_{d_j}^{k+1} = u_{d_j}^{k} + \tilde{x}_{d_j}^{k+1} - \bar{\tilde{x}}_{d_j}^{k+1}$$ #this is a measure of deviation of $\tilde{x}_{d_j}$ from average $\bar{\tilde{x}}_{d_j}$

running sum of

Parallel_and_distrib...
Parallel Dist Alg Part 2

[eq: ADMM for global consensus]

* General Consesus: [General Consensus]

$$\mathcal{V} \; f(x) = \sum_{i=1}^{N} f_i(x_{c_i})$$

$c_i : \subseteq \{1,\dots,n\}$, may overlap with each other

$\{c_1,\dots,c_N\}$ f cover]

e.g. $\{1,2,3\}, \{3,4\}, \{2,4\}$ is a cover of $\{1,2,3,4\}$

$x_1^2+x_2^2+x_3^2$ $\quad e^{x_3+x_4}$ $\quad \log(\exp(x_2)+\exp(x_4))$

so $\mathcal{V} \; (x_1^2+x_2^2+x_3^2) + e^{x_3+x_4} + \log(\exp(x_2)+\exp(x_4))$ is one such example

stuff 0 to make the vector a n dimensional one

$(x_{c_i}, 0)$ //rearrange components along index $\{1,2,\dots,n\}$

$= z_i \in \mathbb{R}^n$ with unused ones set to zero

there might be a typo in Boyd's note: z_i \in R^|c_i|

e.g.

$z_1 = (x_1, x_2, x_3, \cancel{x_4}) = x_{\{1,2,3\}} = x_{c_1} \in \mathbb{R}^{|\{1,2,3\}|} = \mathbb{R}^3$

$z_2 = (\cancel{x_1}, \cancel{x_2}, x_3, x_4) = x_{\{3,4\}} = x_{c_2} \in \mathbb{R}^{|\{3,4\}|} = \mathbb{R}^2$

$z_3 = (\cancel{x_1}, x_2, \cancel{x_3}, x_4) = x_{\{2,4\}} \in \mathbb{R}^{|\{2,4\}|} = \mathbb{R}^2$

note $c_1 \cap c_2 = \{1,2,3\}\cap\{3,4\} = \{3\}$ so: $(z_1)_3 = (z_2)_3$

$c_2 \cap c_3 = \{3,4\}\cap\{2,4\} = \{4\}$ $\quad (z_2)_4 = (z_3)_4$

$c_1 \cap c_3 = \{1,2,3\}\cap\{2,4\} = \{2\}$ $\quad (z_1)_2 = (z_3)_2$

$\mathcal{V} \; \|z_1\|_2^2 + e^{1^T z_2} + \exp(z_3)$

$\ni$ $(z_1)_3 = (z_2)_3$ ~ visually

$(z_2)_4 = (z_3)_4$

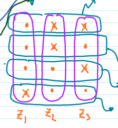
$(z_1)_2 = (z_3)_2$

$$\mathcal{V} \; \sum_{i=1}^{N} f_i(z_i)$$

$\ni \; (z_1,\dots,z_N) \in C$ # $C = \left\{(z_1,\dots,z_n) \,\middle|\, \forall_{i,j\neq i} \; \forall_{k\in c_i\cap c_j} (z_i)_k = (z_j)_k\right\}$

Visualization

no consistency constraint



x = means unused component

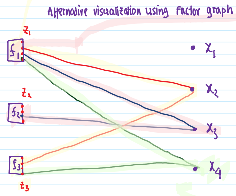
so, • on the same rows must agree

$(z_1)_2 = (z_3)_2 = x_2$ //associated with $f_1 = \|z_1\|_2^2$, $f_2 = e^{1^T z_2}$

$(z_1)_3 = (z_2)_3 = x_3$ // " " $f_1 = e^{1^T z_2}$, $f_3 = \log(\exp(x_2)+\exp(x_4))$

$(z_2)_4 = (z_3)_4 = x_4$ // " " f_1, f_3

Alternative visualization using factor graph



Each edge of the graph is a consistency constraint

Just like before we introduce local copies $\tilde{x}_{c_1}, \tilde{x}_{c_2}, \dots, \tilde{x}_{c_N}$ with unused components set to zero or any other value.

$$\mathcal{V} \; \sum_{i=1}^{n} f_i(\tilde{x}_{c_i}) + \sum_{j=1}^{n} g_j(\tilde{x}_{d_j}) \quad \text{where}$$

c_i = are columnwise non-crossed index

d_j = " rowwise " "

with rest of the constraints same as global consensus

[eq: ADMM for global consensus]

:

$$\forall_{i\in\{1,\dots,N\}} \quad \tilde{x}_{c_i}^{k+1} = \mathrm{prox}_{\lambda f_i}\left(\bar{\tilde{x}}_{c_i}^{k} - u_{c_i}^{k}\right)$$

$$\forall_{j\in\{1,\dots,n\}} \quad u_{d_j}^{k+1} = u_{d_j}^{k} + \tilde{x}_{d_j}^{k+1} - \bar{\tilde{x}}_{d_j}^{k+1}$$

So the only thing that changes is definition of c_i and d_j.

in $\tilde{x}$ notation

distribution of indices of $\tilde{x} \in \mathbb{R}^{nN}$

$$\begin{array}{ccccc|l}
1 & n+1 \dots & n(i-1)+1 & \dots (N-1)n+1 & & \to d_1 \\
\cancel{2} & n+2 \dots & n(i-1)+2 & (N-1)\cancel{n}+2 & & \to d_2 \\
\vdots & \vdots & \vdots & \vdots & & \\
j & \cancel{n+j} \dots & n(i-1)+j & (N-1)n+j & & \to d_j \\
\vdots & \vdots & \vdots & \vdots & & \\
n & 2n \dots & n(i-1)+n \dots & (N-1)n+n & & \to d_n \\
\hline
\downarrow & \downarrow & \downarrow & \downarrow & & \\
c_1 & c_2 & c_i & c_N & &
\end{array}$$

X means index $\tilde{x}$ component = 0 (unused)

* Exchange: # Has a game theory flavor

The exchange problem:

$$\left(\begin{array}{l}
\mathcal{V} \; \sum_{i=1}^{N} f_i(x_i) \\
\ni \; \sum_{i=1}^{N} x_i = \bar{0}, \\
\forall_{i\in\{1,\dots,n\}} \; x_i \in \mathbb{R}^n
\end{array}\right)$$

Origin of the name exchange:

x_i : quantities that are exchanged among N agents

$(x_i)_j > 0 \Rightarrow$ amount of commodity j received by agent i from the exchange

$(x_i)_j < 0 \Rightarrow$ " " " j supplied by " i to the "

$\sum_{i=1}^{N} x_i = \bar{0}$: amount of each commodity contributed by agents balances total

$(x_i)_j < 0 \Rightarrow$ " " " j supplied by " i to the "

$\sum_{i=1}^{N} x_i = \bar{0}$: amount of each commodity contributed by agents balances total amount taken by agents $\Leftrightarrow \sum_{i=1}^{N} (x_i)_j = 0 \quad \forall j \in \{1, \dots, n\}$

Exchange problem seeks the commodity quantities that minimizes the social cost. subject to the market clearing

Optimal dual variables : Set of equillibrium prices for the commodities

$$\min \sum_{i=1}^{N} f_i(x_i) + I_G(x_1, \dots, x_N)$$

$$\# \quad G = \{(x_1, \dots, x_N) \in \mathbb{R}^{nN} \mid x_1 + \dots + x_N = 0\}$$

$$= \{(x_1, \dots, x_N) \in \mathbb{R}^{nN} \mid [I \; I \cdots I] \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} = 0\}$$

now,

$$x_1 + x_2 + \dots + x_N = \bar{0} \xrightarrow{\text{componentwise}} (x_1)_j + (x_2)_j + \dots + (x_N)_j = 0 \quad \forall_{j \in \{1, \dots, n\}}$$

: introduce new variable $\tilde{x}$ just like before such that:

distribution of indices of $\tilde{x} \in \mathbb{R}^{nN}$

1	$n+1$ ..	$n(i-1)+1$ ...	$(N-1)n+1$	$\rightarrow d_1 \Rightarrow \tilde{x}_{d_1}$
2	$n+2$ ..	$n(i-1)+2$	$(N-1)n+2$	$\rightarrow d_2 \Rightarrow \tilde{x}_{d_2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	
j	$n+j$ ..	$n(i-1)+j$	$(N-1)n+j$	$\rightarrow d_j \Rightarrow \tilde{x}_{d_j}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	
n	$2n$ ..	$n(i-1)+n$ ...	$(N-1)n+n$	$\rightarrow d_n \Rightarrow \tilde{x}_{d_n}$
$\downarrow c_1$	$\downarrow c_2$	$\downarrow c_i$	$\downarrow c_N$	

then, $x_i = \tilde{x}_{c_i}$ and the constraints: $\forall_{j \in \{1, \dots, n\}} \; 1_N^T \tilde{x}_{d_j} = 0$

the objective: $\sum_{i=1}^{n} f_i(\tilde{x}_{c_i})$

$$\therefore \; I_G(x_1, \dots, x_N) = \sum_{j=1}^{n} I_{G_j}(\tilde{x}_{d_j}) \; : \; G_j = \{\tilde{x}_{d_j} \mid 1^T \tilde{x}_{d_j} = 0\}$$

So the optimization problem in the new variable becomes:

$$\min \sum_{i=1}^{N} f_i(\tilde{x}_{c_i}) + \sum_{j=1}^{n} I_{G_j}(\tilde{x}_{d_j})$$

$\downarrow$ ADMM form [ADMM for distributed optimization] [illegible]

$$\tilde{x}_{c_i}^{k+1} = \text{prox}_{\lambda f_i}(z_{c_i}^{k} - u_{c_i}^{k}) \Big|_{i=1}^{N}$$

$$z_{d_j}^{k+1} = \text{prox}_{I_{C_j}}(\tilde{x}_{d_j}^{k+1} + u_{d_j}^{k}) = \Pi_{C_j}(\tilde{x}_{d_j}^{k+1} + u_{d_j}^{k}) \Big|_{j=1}^{n} \quad \# \; \text{prox}_{\lambda I_\chi}(\boxdot) = \Pi_\chi(\boxdot)$$

$$u^{k+1} = u^{k} + x^{k+1} - z^{k+1}$$

now [illegible], from [eq: Eucledian_proj_on_hyperplane] : $\Pi_{a^T\square = b}(x) = x - \frac{a^T x - b}{\|a\|_2^2} a$

$$G_j = \{\square \mid 1^T \square = 0\}$$

$$\Pi_{G_j}(\boxdot) = \boxdot - \frac{\bar{1}^T \boxdot}{\|\bar{1}\|_2^2} \bar{1}$$

$\|\bar{1}\|_2^2 = (1^2 + 1^2 + \dots + \overset{N\text{th}}{1^2}) = N$

$$\Pi_{C_j}(\tilde{x}_{d_j}^{k+1} + u_{d_j}^{k}) = (\tilde{x}_{d_j}^{k+1} + u_{d_j}^{k}) - \frac{1^T(\tilde{x}_{d_j}^{k+1} + u_{d_j}^{k})}{N} \bar{1} \quad \# \text{ as we have defined earlier, } \bar{x} = \frac{1^T x}{\dim(x)} \bar{1}$$

$$= \tilde{x}_{d_j}^{k+1} + u_{d_j}^{k} - \bar{\tilde{x}}_{d_j}^{k+1} - \bar{u}_{d_j}^{k}$$

$$z_{d_j}^{k+1} = \tilde{x}_{d_j}^{k+1} + u_{d_j}^{k} - \bar{\tilde{x}}_{d_j}^{k+1} - \bar{u}_{d_j}^{k}$$

splitting rowwise:

$\forall_{j \in \{1, \dots, n\}}$

$$u_{d_j}^{k+1} = u_{d_j}^{k} + \tilde{x}_{d_j}^{k+1} - z_{d_j}^{k+1}$$

$$= u_{d_j}^{k} + \tilde{x}_{d_j}^{k+1} - (\tilde{x}_{d_j}^{k+1} + u_{d_j}^{k} - \bar{\tilde{x}}_{d_j}^{k+1} - \bar{u}_{d_j}^{k})$$

$$= \bar{\tilde{x}}_{d_j}^{k+1} + \bar{u}_{d_j}^{k}$$

$$= \bar{\tilde{x}}_{d_j}^{k+1} + u_{d_j}^{k}$$

\# now note that because $\bar{\tilde{x}}_{d_j}^{k+1} = \frac{1^T \tilde{x}_{d_j}^{k+1}}{|d_j|} \bar{1}$, $\bar{u}_{d_j}^{k} = \frac{1^T u_{d_j}^{k}}{|d_j|} \bar{1}$ are vectors with each component same, $\therefore$ if we start with $u_{d_j}^{0} = 0$ or same component vector $u_{d_j}^{k+1}$ will have all components same, so $\bar{u}_{d_j}^{k+1} = \frac{1^T u_{d_j}^{k+1}}{|d_j|} \bar{1} = u_{d_j}^{k+1}$

$$z_{d_j}^{k+1} = \tilde{x}_{d_j}^{k+1} + \underbrace{u_{d_j}^{k}}_{= \bar{u}_{d_j}^{k}} - \bar{\tilde{x}}_{d_j}^{k+1} - \bar{u}_{d_j}^{k} = \tilde{x}_{d_j}^{k+1} - \bar{\tilde{x}}_{d_j}^{k+1} \quad \forall_{j \in \{1, \dots, n\}}$$

redistributing the components over the columns we can con construct $z_{c_i}^{k+1}$

The final ADMM iterations become:

$$\tilde{x}_{c_i}^{k+1} = \text{prox}_{\lambda f_i}(z_{c_i}^{k} - u_{c_i}^{k}) \Big|_{i=1}^{N}$$

$$z_{d_j}^{k+1} = \tilde{x}_{d_j}^{k+1} - \bar{\tilde{x}}_{d_j}^{k+1} \Big|_{j=1}^{n}$$

$$u_{d_j}^{k+1} = \bar{\tilde{x}}_{d_j}^{k+1} + u_{d_j}^{k} \Big|_{j=1}^{n}$$

(ADMM iterations for exchange problem)

* Exchange and consensus problems are dual of each other.

*General form exchange:

Same math as the consensus case [General Consensus]

*Allocation problem:

$$\min \sum_{i=1}^{N} f_i(x_i)$$

$$\forall_{i\in\{1,\dots,n\}} \quad x_i \succeq 0 \;\rightarrow\; \forall_{j\in\{1,\dots,n\}} \; x_i(j) \geq 0$$

#Note that, the constraint looks very similar to unit simplex

$$\sum_{i=1}^{N} x_i = b \;\longrightarrow\; \forall_{j\in\{1,\dots,n\}} \sum_{i=1}^{n} x_i(j) = b_j$$

$$x_i \in \mathbb{R}^n$$

interpretation: n types of resources

distribution of indices of $\tilde{x} \in \mathbb{R}^{nN}$

1	$n+1 \dots n(i-1)+1 \dots (N-1)n+1$	$\rightarrow d_1 \Rightarrow \tilde{x}_{d_1}$
2	$n+2 \dots n(i-1)+2 \quad (N-1)n+2$	$\rightarrow d_2 \Rightarrow \tilde{x}_{d_2}$
$\vdots$	$\vdots$	
j	$n+j \dots n(i-1)+j \quad (N-1)n+j$	$\rightarrow d_j \Rightarrow \tilde{x}_{d_j}$
$\vdots$	$\vdots$	
n	$2n \dots n(i-1)+n \dots (N-1)n+n$	$\rightarrow d_n \Rightarrow \tilde{x}_{d_n}$
$\downarrow c_1$	$\downarrow c_2 \quad \downarrow c_i \quad \downarrow c_N$	

$x_1 = \tilde{x}_{c_1}$ $\quad x_2 = \tilde{x}_{c_2}$ $\quad x_i = \tilde{x}_{c_i}$ $\quad x_N = \tilde{x}_{c_N}$

So, in $\tilde{x}$ constraints become

$$G_j = \{\tilde{x}_{d_j} \succeq 0 \mid 1^T \tilde{x}_{d_j} = b_j\} \quad \forall_{j\in\{1,\dots,n\}}$$

Objective becomes:

$$\min \sum_{i=1}^{N} f_i(\tilde{x}_{c_i})$$

the initial problem becomes:

$$\min \sum_{i=1}^{N} f_i(\tilde{x}_{c_i}) + \sum_{j=1}^{n} I_{G_j}(\tilde{x}_{d_j})$$

[ADMM for distributed optimization] [illegible]

$$\tilde{x}_{c_i}^{k+1} = \operatorname{prox}_{\lambda f_i}(z_{c_i}^{k} - u_{c_i}^{k}) \Big|_{i=1}^{N}$$

$$z_{d_j}^{k+1} = \operatorname{prox}^{k+1}_{I_{c_j}}(\tilde{x}_{d_j}^{k+1} + u_{d_j}^{k}) = \Pi_{c_j}(\tilde{x}_{d_j}^{k+1} + u_{d_j}^{k}) \Big|_{j=1}^{n}$$

$\operatorname{prox}_{\lambda I_{\chi}}(\boxminus) = \Pi_{\chi}(\boxminus)$

$$u^{k+1} = u^{k} + x^{k+1} - z^{k+1}$$

now: [Projection onto the standard simplex] [illegible]:

$$\Pi_{\square \succeq 0,\, 1^T\square = b}(x) = \max\{0, x - v^* \bar{1}\} \quad \text{//elementwise maximum}$$

$$\downarrow$$

$$\sum_{i=1}^{N} \max\{0, x_i - v_i\} = b$$

$$z_{d_j}^{k+1} = \Pi_{c_j}(\tilde{x}_{d_j}^{k+1} + u_{d_j}^{k}) = \max\{0, \tilde{x}_{d_j}^{k+1} + u_{d_j}^{k} - v^* \bar{1}\} \left\{ \sum_{i=1}^{N} \max\{0, x_i - v_i\} = b \right\}$$

$\ast$ ADMM can parallelize when projection on a certain part of the constraint set is easy.

*Some tricks:

$$\underbrace{\underbrace{C_1}_{x_1}}_{\in\mathbb{R}^{n_1}},\; \underbrace{\underbrace{C_2}_{x_2}}_{\in\mathbb{R}^{n_2}},\; \underbrace{C_3}_{\text{function of both } x_1, x_2}$$

$$C_1 = \{x_1 \in \mathbb{R}^{n_1} \mid f_i^{(1)}(x_1) \leq 0 \quad \forall i\}$$

$$C_2 = \{x_2 \in \mathbb{R}^{n_2} \mid f_j^{(2)}(x_2) \leq 0 \quad \forall j\}$$

$$C_3 = \{(x_1, x_2) \in \mathbb{R}^{n_1+n_2} \mid f_k^{(3)}(x_1, x_2) \leq 0 \;\forall k\}$$

What does it mean saying projection is easy?

- → Closed form exists
- → can be calculated in a matrix-free manner (does not involve any sort of matrix inversion)
- → can be implemented by very basic coding (just if-else, for loop) → if it can be implemented in a relatively short length of codes in assembly languages