CH2.2: DUAL DECOMPOSITION

Monday, April 1, 2024 5:37 PM

- The major benefit of dual ascent method is that it can sometimes lead to a decentralized algorithm.
- Suppose that the **objective function f is separable** such that the input variables can be paritioned off into subvectors.

ex) $f(x) = \sum_{i=1}^{N} f_i(x_i)$, $x = (x_1, x_2, \cdots, x_N)$, $\underbrace{x_i \in \mathbb{R}^{n_i}}$

objective function

subvectors of x.

$$x = [\leftarrow x_1 \rightarrow \leftarrow x_2 \rightarrow \ldots \leftarrow x_N \rightarrow]^T$$

Separable: $f(x) = \underbrace{f_1(x_1)} + f_2(x_2) + \ldots + f_N(x_N)$

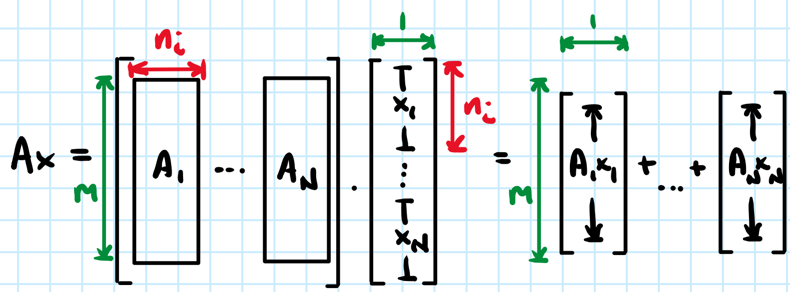
f_1 only depends on x_1, f_2 on x_2, etc.

$Ax = b$: $A \in \mathbb{R}^{m \times n}$, $A_i \in \mathbb{R}^{m \times n_i}$ (submatrix)

Partition matrix A as:

$A = [A_1, A_2, \ldots, A_N]$ so that

$\hookrightarrow$ $Ax = \sum_{i=1}^{N} A_i x_i$



Lagrangian can be written as:

$$L(x,y) = \sum_{i=1}^{N} L_i(x_i, y)$$

N subvectors x_i in x

$$= \sum_{i=1}^{N} \Big(\underbrace{f_i(x_i)}_{\mathbb{R}} + \underbrace{\underbrace{y^T}_{\mathbb{R}^m} \cdot \underbrace{A_i x_i}_{\mathbb{R}^m}}_{\mathbb{R}} - \frac{1}{N} \underbrace{y^T b}_{\mathbb{R}} \Big) \in \mathbb{R}$$

Also separable in x

Derivation: $f(x) = \sum_{i=1}^{N} f_i(x_i)$

LAGRANGIAN: $L(x,y) = f(x) + y^T(Ax - b)$

$$L(x,y) = \sum_{i=1}^{N} f_i(x_i) + y^T\left(\sum_{i=1}^{N} A_i x_i - b\right)$$
$$= \sum_{i=1}^{N}\left(f_i(x_i) + y^T A_i x_i\right) - y^T b$$
$$= \sum_{i=1}^{N}\left(f_i(x_i) + y^T A_i x_i - \frac{1}{N} y^T b\right) \quad \in \mathbb{R} \checkmark$$

X-minimization step can be parallelized:

NEW ALGORITHM:

$$x_i^{k+1} := \operatorname{argmin}_{x_i} L_i(x_i, y^k) \rightarrow \text{DUAL DECOMPOSITION!}$$
$$y^{k+1} := y^k + \alpha^k (Ax^{k+1} - b) \rightarrow \text{Split work among N processors.}$$