Decomposition Methods

Dr. Dingzhu Wen

School of Information Science and Technology (SIST)
ShanghaiTech University

wendzh@shanghaitech.edu.cn

April 1, 2024

Overview

Separable Problems

An example of separable problems

$$\begin{aligned} \min \;& f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2), \\ \text{s.t.} \;& \mathbf{x}_1 \in \mathbb{S}_1, \; \mathbf{x}_2 \in \mathbb{S}_2. \end{aligned}$$

- We can solve for $\mathbf{x}_1$ and $\mathbf{x}_2$ separately (in parallel or in a distributed way).
- Even if they are solved sequentially, this gives advantage if computational effort is superlinear in problem size.
- Generalizes to any objective of form $\Psi(f_1, f_2)$ with Ψ nondecreasing (e.g., max).

Complicating Variables

Consider the problem

$$\min \; f_1(\mathbf{x}_1, \mathbf{y}) + f_2(\mathbf{x}_2, \mathbf{y}).$$

- $\mathbf{y}$ is the complicating variable or coupling variable.
- When $\mathbf{y}$ is fixed, the problem is separable in $\mathbf{x}_1$ and $\mathbf{x}_2$.
- $\mathbf{x}_1$ and $\mathbf{x}_2$ are private or local variables. $\mathbf{y}$ is a public or interface or boundary variable between the two subproblems

Primal Decomposition

Methodology: Fix $\mathbf{y}$ and define:

$$\begin{aligned} \text{Subproblem 1:} \quad & \min_{\mathbf{x}_1} \quad f_1(\mathbf{x}_1, \mathbf{y}), \\ \text{Subproblem 2:} \quad & \min_{\mathbf{x}_2} \quad f_2(\mathbf{x}_2, \mathbf{y}), \end{aligned}$$

with optimal values $\phi_1(\mathbf{y})$ and $\phi_2(\mathbf{y})$.
Original problem is equivalent to **master problem**:

$$\min_{\mathbf{y}} \quad \phi_1(\mathbf{y}) + \phi_2(\mathbf{y}).$$

This is called primal decomposition, since master problem manipulates primal (complicating) variables.

Primal Decomposition

Properties

- If original problem is convex, so is master problem.
- Can solve master problem using
  - bisection (if y is scalar),
  - gradient or Newton method (if $\{\phi_i\}$ are differentiable),
  - subgradient, cutting-plane, or ellipsoid method.
- Each iteration of master problem requires solving the two subproblems (in parallel).
- If master algorithm converges fast enough and subproblems are sufficiently easier to solve than original problem, we get savings.

Primal Decomposition Algorithm

Using subgradient algorithm for master:

- **Repeat**
  - Solve the subproblems (in parallel).
    Find $\mathbf{x}_1$ that minimizes $f_1(\mathbf{x}_1, \mathbf{y}_k)$,
    Find a subgradient $\mathbf{g}_1 \in \partial\phi_1(\mathbf{y}_k)$.
    Find $\mathbf{x}_2$ that minimizes $f_2(\mathbf{x}_2, \mathbf{y_k})$,
    Find a subgradient $\mathbf{g}_2 \in \partial\phi_2(\mathbf{y}_k)$.
  - Update complicating variable:

$$\mathbf{y}_{k+1} = \mathbf{y}_k - \eta_k(\mathbf{g}_1 + \mathbf{g}_2).$$

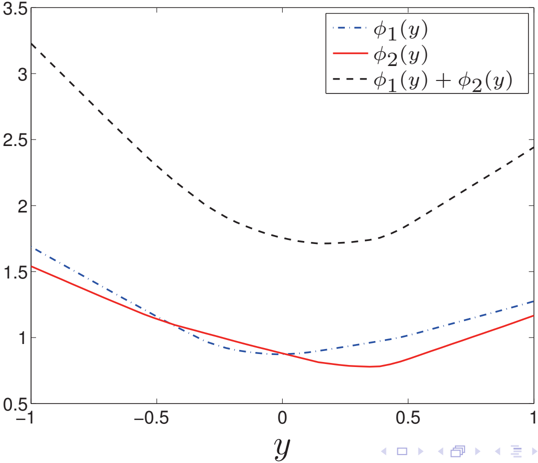
- **Until** convergence.

The step size η_k can be chosen in any of the standard ways.

An Example of Primal Decomposition

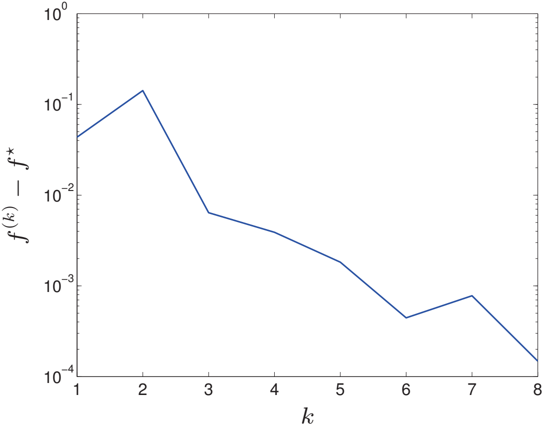
Settings

- $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^{20}$, $y \in \mathbb{R}$,
- f_i are PWL (max of 100 affine functions).



An Example of Primal Decomposition

Primal decomposition (using bisection on y)



Dual Decomposition

Optimization Problem: $\min \ f_1(\mathbf{x}_1, \mathbf{y}) + f_2(\mathbf{x}_2, \mathbf{y})$.
Step 1: Introduce new variables $\mathbf{y}_1, \mathbf{y}_2$

$$
\begin{aligned}
\min \ & f(\mathbf{x}) = f_1(\mathbf{x}_1, \mathbf{y}_1) + f_2(\mathbf{x}_2, \mathbf{y}_2), \\
\text{s.t.} \ & \mathbf{y}_1 = \mathbf{y}_2.
\end{aligned}
$$

- $\mathbf{y}_1$, $\mathbf{y}_2$ are local versions of complicating variable $\mathbf{y}$.
- $\mathbf{y}_1 = \mathbf{y}_2$ is the consensus constraint.

Dual Decomposition

Step 2: Form dual problem:

$$L(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_1, \mathbf{y}_2) = f_1(\mathbf{x}_1, \mathbf{y}_1) + f_2(\mathbf{x}_2, \mathbf{y}_2) + \boldsymbol{\nu}^T(\mathbf{y}_1 - \mathbf{y}_2),$$

which is separable and can minimize over $(\mathbf{x}_1, \mathbf{y}_1)$ and $(\mathbf{x}_2, \mathbf{y}_2)$ in parallel:

$$g_1(\boldsymbol{\nu}) = \inf_{\mathbf{x}_1, \mathbf{y}_1} \ f_1(\mathbf{x}_1, \mathbf{y}_1) + \boldsymbol{\nu}^T \mathbf{y}_1,$$
$$g_2(\boldsymbol{\nu}) = \inf_{\mathbf{x}_2, \mathbf{y}_2} \ f_2(\mathbf{x}_2, \mathbf{y}_2) - \boldsymbol{\nu}^T \mathbf{y}_2.$$

The dual problem is $\max_{\boldsymbol{\nu}} \ g(\boldsymbol{\nu}) = g_1(\boldsymbol{\nu}) + g_2(\boldsymbol{\nu})$.

- A subgradient of $-g$ is $\mathbf{y}_2 - \mathbf{y}_1$ (from solutions of subproblems).

Dual Decomposition Algorithm

Using subgradient algorithm for master:

- **Repeat**
  - Solve the subproblems (in parallel).
    Find $\mathbf{x}_1, \mathbf{y}_1$ that minimizes $f_1(\mathbf{x}_1, \mathbf{y}_1) + \boldsymbol{\nu}^T \mathbf{y}_1$,
    Find $\mathbf{x}_2, \mathbf{y}_2$ that minimizes $f_2(\mathbf{x}_2, \mathbf{y}_2) - \boldsymbol{\nu}^T \mathbf{y}_2$,
  - Update complicating variable:

$$\boldsymbol{\nu}_{k+1} = \boldsymbol{\nu}_k - \eta_k(\mathbf{y}_2 - \mathbf{y}_1).$$

- **Until** convergence.

The results are generally infeasible, i.e., $\mathbf{y}_2 \neq \mathbf{y}_1$.
Question: How to get feasible solution?

Dual Decomposition Algorithm

Finding Feasible Solutions

- Reasonable guess of feasible point from $(\mathbf{x}_1, \mathbf{y}_1)$ and $(\mathbf{x}_2, \mathbf{y}_2)$:

$$(\mathbf{x}_1, \bar{\mathbf{y}}), \qquad (\mathbf{x}_2, \bar{\mathbf{y}}), \qquad \bar{\mathbf{y}} = \frac{\mathbf{y}_1 + \mathbf{y}_2}{2}.$$

  - Projection onto feasible set $\mathbf{y}_1 = \mathbf{y}_2$.
  - Gives upper bound of solution: $p^* \leq f_1(\mathbf{x}_1, \bar{\mathbf{y}}) + f_2(\mathbf{x}_2, \bar{\mathbf{y}})$.
- A better feasible point: Replace $\mathbf{y}_1, \mathbf{y}_2$ with $\bar{\mathbf{y}}$ and solve primal subproblems:

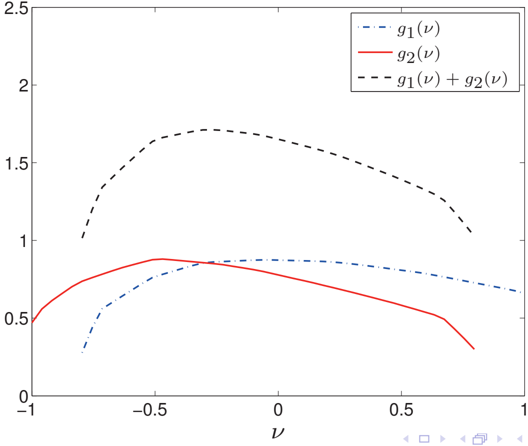
$$\min_{\mathbf{x}_1} \; f_1(\mathbf{x}_1, \bar{\mathbf{y}}) \text{ and } \min_{\mathbf{x}_2} \; f_1(\mathbf{x}_2, \bar{\mathbf{y}}).$$

  - Gives better (tighter) upper bound of solution.

An Example for Dual Decompostion

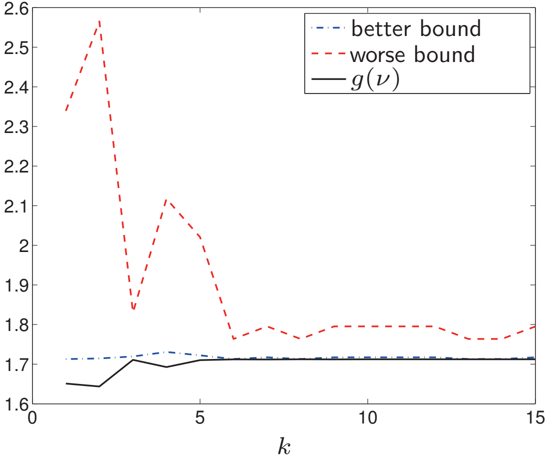
Settings

- $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^{20}$, $y \in \mathbb{R}$,
- f_i are PWL (max of 100 affine functions each).



An Example for Dual Decompostion

Dual decomposition convergence (using bisection on ν)



Dual Decomposition

Interpretation

- $\mathbf{y}_1$ is resources consumed by first unit,
- $\mathbf{y}_2$ is resources generated by second unit,
- $\mathbf{y}_1 = \mathbf{y}_2$ is consistency condition: Supply equals demand,
- $\boldsymbol{\nu}$ is a set of resource prices,
- Master algorithm adjusts prices at each step, rather than allocating resources directly (primal decomposition).

Recovering Primal Solutions from the Dual

- Iterates the dual decompostion algorithm till convergence.
- $\boldsymbol{\nu}_k \to \boldsymbol{\nu}^*$: Have price convergence,
- Subtlety: $\mathbf{y}_{1,k} = \mathbf{y}_{2,k}$ is not needed,
- The hammer: if $\{f_i\}$ are strictly convex, we have $\mathbf{y}_{1,k} = \mathbf{y}_{2,k}$,
- Master algorithm adjusts prices at each step, rather than allocating resources directly (primal decomposition).
- Can fix allocation, or add regularization terms.

Decomposition with Constraints

Optimization Problem:

$$\begin{aligned} \min \quad & f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2), \\ \text{s.t.} \quad & \mathbf{x}_1 \in \mathbb{S}_1,\ \mathbf{x}_2 \in \mathbb{S}_2, \\ & \mathbf{h}_1(\mathbf{x}_1) + \mathbf{h}_2(\mathbf{x}_2) \le \mathbf{0}, \end{aligned}$$

where $\le$ is element-wise operation.
Note: The problem has complicating constraints.

- $\{f_i\}$, $\{h_i\}$, and $\{\mathcal{S}_i\}$ are convex.
- $\mathbf{h}_1(\mathbf{x}_1) + \mathbf{h}_2(\mathbf{x}_2) \le \mathbf{0}$ is a set of complicating or coupling constraints, involving both $\mathbf{x}_1$ and $\mathbf{x}_2$.
- Can interpret coupling constraints as limits on resources shared between two subproblems.

Primal Decomposition with Constraints

For a given constant vector $\mathbf{c}$ and define

$$\text{Subproblem 1}: \quad \begin{aligned} &\min \;\; f_1(\mathbf{x}_1), \\ &\text{s.t.} \;\; \mathbf{x}_1 \in \mathbb{S}_1, \qquad \mathbf{h}_1(\mathbf{x}_1) \le \mathbf{c}. \end{aligned}$$

$$\text{Subproblem 2}: \quad \begin{aligned} &\min \;\; f_2(\mathbf{x}_2), \\ &\text{s.t.} \;\; \mathbf{x}_2 \in \mathbb{S}_2, \qquad \mathbf{h}_2(\mathbf{x}_2) \le -\mathbf{c}. \end{aligned}$$

- $\mathbf{c}$ decides the resorce allocation between two subproblems.
- Denote $\phi_1(\mathbf{c})$ and $\phi_2(\mathbf{c})$ as the optimal value of the problems.
- **Matser problem**: $\min_{\mathbf{c}} \;\; \phi_1(\mathbf{c}) + \phi_2(\mathbf{c})$.
- Subporblems can be solved in parallel.

Primal Decoposition Algorithm (With Constraints)

- **Repeat**
  - Solve the subproblems (in parallel).
    Solve subproblem 1, finding $\mathbf{x}_1$ and $\boldsymbol{\lambda}_1$.
    Solve subproblem 2, finding $\mathbf{x}_2$ and $\boldsymbol{\lambda}_2$.
  - Update resource allocation.

$$\mathbf{c}_{k+1} = \mathbf{c}_k - \eta_k(\boldsymbol{\lambda}_2 - \boldsymbol{\lambda}_1).$$

- **Until** convergence.
- $\boldsymbol{\lambda}_i$ is an optimal Lagrange multiplier associated with resource constraint in subproblems.
- $(\boldsymbol{\lambda}_2 - \boldsymbol{\lambda}_1) \in \partial[\phi_1(\mathbf{c}) + \phi_2(\mathbf{c})]$.
- All iterations are feasible when the subproblems are feasible.

Primal Decoposition Algorithm (With Constraints)

Theorem

Denote $p(\mathbf{z})$ as the optimal value of the following convex problem:

$$\begin{aligned} &\min \;\; f(\mathbf{x}), \\ &s.t. \;\; \mathbf{x} \in \mathbb{S}, \qquad \mathbf{h}(\mathbf{x}) \leq \mathbf{z}, \end{aligned}$$

where $\mathbf{z} \in \mathrm{dom} p$. Let $\boldsymbol{\lambda}(\mathbf{z})$ be an optimal dual variable vector associated with the constraint above.Then,

$$-\boldsymbol{\lambda}(\mathbf{z}) \in \partial p(\mathbf{z}).$$

Proof: Consider another point $\tilde{\mathbf{z}} \in \mathrm{dom} p$:

$$p(\tilde{\mathbf{z}}) = \sup_{\boldsymbol{\lambda} \geq 0} \inf_{\mathbf{x}} \left[f(\mathbf{x}) + \boldsymbol{\lambda}^T (\mathbf{h}(\mathbf{x}) - \tilde{\mathbf{z}}) \right],$$

Primal Decoposition Algorithm (With Constraints)

Proof(Continue): It follows that

$$
\begin{aligned}
p(\tilde{\mathbf{z}}) &= \sup_{\boldsymbol{\lambda} \geq 0} \inf_{\mathbf{x}} \left[f(\mathbf{x}) + \boldsymbol{\lambda}^T (\mathbf{h}(\mathbf{x}) - \tilde{\mathbf{z}}) \right], \\
&\geq \inf_{\mathbf{x}} \left[f(\mathbf{x}) + \boldsymbol{\lambda}(\mathbf{z})^T (\mathbf{h}(\mathbf{x}) - \tilde{\mathbf{z}}) \right], \\
&= \inf_{\mathbf{x}} \left[f(\mathbf{x}) + \boldsymbol{\lambda}(\mathbf{z})^T (\mathbf{h}(\mathbf{x}) - \mathbf{z} + \mathbf{z} - \tilde{\mathbf{z}}) \right], \\
&= \inf_{\mathbf{x}} \left[f(\mathbf{x}) + \boldsymbol{\lambda}(\mathbf{z})^T (\mathbf{h}(\mathbf{x}) - \mathbf{z}) \right] + \boldsymbol{\lambda}(\mathbf{z})^T (\mathbf{z} - \tilde{\mathbf{z}}), \\
&= p(\mathbf{z}) - \boldsymbol{\lambda}(\mathbf{z})^T (\tilde{\mathbf{z}} - \mathbf{z}).
\end{aligned}
$$

This holds of all points $\tilde{\mathbf{z}} \in \mathrm{dom} p$. Hence, $-\boldsymbol{\lambda}(\mathbf{z})$ is a subgradient of $p(\mathbf{z})$.

Dual Decomposition with Constraints

Optimization Problem:

$$\begin{aligned} \min \quad & f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2), \\ \text{s.t.} \quad & \mathbf{x}_1 \in \mathbb{S}_1,\ \mathbf{x}_2 \in \mathbb{S}_2, \\ & \mathbf{h}_1(\mathbf{x}_1) + \mathbf{h}_2(\mathbf{x}_2) \le \mathbf{0}, \end{aligned}$$

The Lagrangian function is

$$\begin{aligned} L &= f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) + \boldsymbol{\lambda}^T \left[\mathbf{h}_1(\mathbf{x}_1) + \mathbf{h}_2(\mathbf{x}_2)\right], \\ &= \left[f_1(\mathbf{x}_1) + \boldsymbol{\lambda}^T \mathbf{h}_1(\mathbf{x}_1)\right] + \left[f_2(\mathbf{x}_2) + \boldsymbol{\lambda}^T \mathbf{h}_2(\mathbf{x}_2)\right]. \end{aligned}$$

Dual Decomposition with Constraints

For fixed dual varaibles $\boldsymbol{\lambda}$ and define

$$\text{Subproblem 1:} \quad \begin{aligned} \min_{\mathbf{x}_1} \;& f_1(\mathbf{x}_1) + \boldsymbol{\lambda}^T \mathbf{h}_1(\mathbf{x}_1), \\ \text{s.t.} \;& \mathbf{x}_1 \in \mathbb{S}_1, \end{aligned}$$

$$\text{Subproblem 2:} \quad \begin{aligned} \min_{\mathbf{x}_2} \;& f_2(\mathbf{x}_2) + \boldsymbol{\lambda}^T \mathbf{h}_2(\mathbf{x}_2), \\ \text{s.t.} \;& \mathbf{x}_2 \in \mathbb{S}_2, \end{aligned}$$

with optimal values $g_1(\lambda)$ and $g_2(\lambda)$.

Dual Decomposition with Constraints

Master Problem:

$$\max_{\boldsymbol{\lambda}} \quad g(\boldsymbol{\lambda}) = g_1(\boldsymbol{\lambda}) + g_2(\boldsymbol{\lambda}).$$

Properties

- $-h_i(x_i^*) \in \partial(-g_i)(\boldsymbol{\lambda})$ with x_i^* being the optimal solutions of the subproblems.
- $-h_1(x_1^*) - h_2(x_2^*) \in \partial(-g_1 - g_2)(\boldsymbol{\lambda})$.
- The master algorithm updates $\boldsymbol{\lambda}$ using this subgradient.

Dual Decomposition Algorithm (with Constraints)

- **Repeat**
  - Solve the subproblems (in parallel).
    Solve subproblem 1, finding an optimal $\mathbf{x}_1^*$.
    Solve subproblem 2, finding an optimal $\mathbf{x}_2^*$.
  - Update dual variables.

$$\boldsymbol{\lambda}_{k+1} = \max\left\{\boldsymbol{\lambda}_k - \eta_k(h_1(x_1^*) + h_2(x_2^*)), 0\right\}.$$

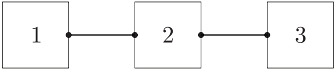
- **Until** convergence.
- η_k is an appropriate step size.
- Iterations are not needed to be feasible.
- Can construct feasible primal variables using projection.

General Decomposition Structures

- Multiple subsystems,
- Variable and/or constraint coupling between subsets of subsystems,
- Represent as hypergraph with subsystems as vertices, coupling as hyperedges or nets,
- Without loss of generality, can assume all coupling is via consistency constraints.

A Simple Example

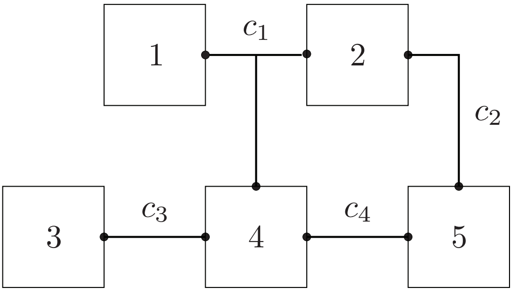
Simple example



- 3 subsystems, with private variables x_1, x_2, x_3, and public variables y_1, (y_2, y_3), and y_4
- 2 (simple) edges

$$\begin{array}{ll} \text{minimize} & f_1(x_1, y_1) + f_2(x_2, y_2, y_3) + f_3(x_3, y_4) \\ \text{subject to} & (x_1, y_1) \in \mathcal{C}_1, \quad (x_2, y_2, y_3) \in \mathcal{C}_2, \quad (x_3, y_4) \in \mathcal{C}_3 \\ & y_1 = y_2, \quad y_3 = y_4 \end{array}$$

A Complex Example



General Form

Optimization Problem:

$$\begin{aligned} \min \quad & \sum_{i=1}^{K} f_i(\mathbf{x}_i, \mathbf{y}_i), \\ \text{s.t.} \quad & (\mathbf{x}_i, \mathbf{y}_i) \in \mathbb{S}_i, \quad 1 \le i \le K, \\ & \mathbf{y}_i = \mathbf{E}_i \mathbf{z}, \qquad 1 \le i \le K. \end{aligned}$$

- Private variables $\mathbf{x}_i$, public variables $\mathbf{y}_i$.
- Net (hyperedge) variables $\mathbf{z}$: $\mathbf{z}_i$ is common value of public variables in net i.
- Matrices $\mathbf{E}_i$ give netlist or hypergraph.

Primal Decomposition

Subproblems: $\phi_i(\mathbf{y}_i)$ is the optimal value of the subproblem

$$\begin{aligned} \min \quad & f_i(\mathbf{x}_i, \mathbf{y}_i), \\ \text{s.t.} \quad & (\mathbf{x}_i, \mathbf{y}_i) \in \mathbb{S}_i. \end{aligned}$$

Master Problem:

$$\min \quad \phi(\mathbf{z}) = \sum_{i=1}^{K} \phi_i(\mathbf{E}_i \mathbf{z}).$$

Subgradient of $\phi(\mathbf{z})$:

$$\mathbf{g} = \sum_{i=1}^{K} \mathbf{E}_i^T \mathbf{g}_i,$$

where $\mathbf{g}_i \in \partial \phi_i(\mathbf{E}_i \mathbf{z})$.

Primal Decomposition

Algorithm

- **Repeat**
  - Distribute net variables to subsystems: $\mathbf{y}_i = \mathbf{E}_i \mathbf{z}, \qquad 1 \leq i \leq K.$
  - Optimal subsystems (in parallel).
    Solve the subproblem to find optimal $\mathbf{x}_i$ and $\mathbf{g}_i \in \partial \phi_i(\mathbf{y}_i)$.
  - Collect and sum subgradients for each net: $\mathbf{g} = \sum_{i=1}^{K} \mathbf{E}_i^T \mathbf{g}_i$.
  - Update net variables: $\mathbf{z}_{k+1} = \mathbf{z}_k - \eta_k \mathbf{g}$.
- **Until** convergence.

Dual Decomposition

Optimization Problem :
$$\begin{aligned} \min \quad & \sum_{i=1}^{K} f_i(\mathbf{x}_i, \mathbf{y}_i), \\ \text{s.t.} \quad & (\mathbf{x}_i, \mathbf{y}_i) \in \mathbb{S}_i, \quad 1 \leq i \leq K, \\ & \mathbf{y}_i = \mathbf{E}_i \mathbf{z}, \qquad 1 \leq i \leq K. \end{aligned}$$

Lagrange Function :
$$\begin{aligned} L &= \sum_{i=1}^{K} f_i(\mathbf{x}_i, \mathbf{y}_i) + \sum_{i=1}^{K} \boldsymbol{\nu}_i^T (\mathbf{y}_i - \mathbf{E}_i \mathbf{z}), \\ &= \sum_{i=1}^{K} \left[f_i(\mathbf{x}_i, \mathbf{y}_i) + \boldsymbol{\nu}_i^T \mathbf{y}_i \right] - \sum_{i=1}^{K} \boldsymbol{\nu}_i^T \mathbf{E}_i \mathbf{z}. \end{aligned}$$

To find the optimal solution, a necessary condition is to minimize over $\mathbf{z}$:

$$\frac{\partial L}{\partial \mathbf{z}} = 0, \Rightarrow \sum_{i=1}^{K} \mathbf{E}_i \boldsymbol{\nu}_i = \mathbf{0}.$$

Dual Decomposition

First, select $\{\boldsymbol{\nu}_i\}$ that satisfies $\sum_{i=1}^{K} \mathbf{E}_i \boldsymbol{\nu}_i = \mathbf{0}$.
Then, a **subproblem** can be written as

$$\begin{aligned} \min \quad & f_i(\mathbf{x}_i, \mathbf{y}_i) + \boldsymbol{\nu}_i^T \mathbf{y}_i, \\ \text{s.t.} \quad & (\mathbf{x}_i, \mathbf{y}_i) \in \mathbb{S}_i, \end{aligned}$$

whose optimal solution is denoted as $g_i(\boldsymbol{\nu}_i)$ and the solved $\mathbf{y}_i$ is a subgradient of $g_i(\boldsymbol{\nu}_i)$.
Next, the **master problem** is

$$\begin{aligned} \min \quad & g(\{\boldsymbol{\nu}_i\}) = \sum_{i=1}^{K} g_i(\boldsymbol{\nu}_i), \\ \text{s.t.} \quad & \sum_{i=1}^{K} \mathbf{E}_i \boldsymbol{\nu}_i = \mathbf{0}. \end{aligned}$$

Dual Decomposition

Algorithm

- **Repeat**
  - Optimize subsystems (in parallel).
    Solve subproblem to obtain $(\mathbf{x}_i, \mathbf{y}_i)$.
  - Compute average value of public variables over each net.

$$\hat{\mathbf{z}} = (\mathbf{E}^T\mathbf{E})^{-1}\mathbf{E}^T \sum_{i=1}^{K} \mathbf{y}_i.$$

    Solve the subproblem to find optimal $\mathbf{x}_i$ and $\mathbf{g}_i \in \partial\phi_i(\mathbf{y}_i)$.
  - Update dual variables.

$$\boldsymbol{\nu}_{i,k+1} = \boldsymbol{\nu}_{i,k} - \eta_k(\mathbf{y}_i - \mathbf{E}_i\hat{\mathbf{z}}).$$

- **Until** convergence.

Thank you!

wendzh@shanghaitech.edu.cn