Convex Optimization II

Lecture 9: Decomposition Methods

Hamed Shah-Mansouri

Department of Electrical Engineering

Sharif University of Technology

1400-2

References

[1] D.P. Palomar and M. Chiang, “A Tutorial on Decomposition Methods for Network Utility Maximization,” *IEEE J. on Selected Areas in Communications*, vol. 24, no. 8, pp. 1439-1452, Aug. 2006.

[2] S.H. Low and D.E. Lapsley, “Optimization flow control, I: Basic algorithm and convergence,” *IEEE/ACM Trans on Networking*, vol. 7, no. 6, pp. 861-874, Dec. 1999.

[3] F.P. Kelly, A. Maulloo, and D. Tan, “Rate control for communication networks: Shadow prices, proportional fairness, and stability,” *Journal of Operations Research Society*, vol. 49, no. 3, pp. 237-252, March 1998.

[4] S.H. Low, F. Paganini, and J. Doyle, “Internet congestion control,” *IEEE Control Systems Magazine*, Feb. 2002.

MOTIVATIONS

- Efficiently solving an optimization problem, three levels of understanding as to what it means
  - Global optimality
  - Computational properties (polynomial-time, practically fast, and scalable (but centralized) algorithms)
  - Distributed algorithms that converge to the global optimum (particularly attractive in large-scale problems where a centralized solution is infeasible)

- How can we use convex optimization to address the above issues?

- Decomposition theory naturally provides the mathematical language to build an analytic foundation for the design of modularized and distributed algorithms.
  - Dual Decomposition
  - Primal Decomposition

- Case Study: Reverse engineering of transmission control protocol (TCP)

Convex Optimization and Lagrange Duality

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p \end{array} \tag{1}$$

- Lagrangian :

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x). \tag{2}$$

- Lagrange dual function: $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$,

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \\ &= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right) \end{aligned}$$

- Lagrange Dual Problem :

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succcurlyeq 0 \end{array} \qquad (3)$$

- Gradient and subgradient methods can solve the dual problem iteratively. The primal problem can be solved correspondingly.

- If the problem is very large, it takes a quite long time to solve the problem in the centralized solver.

DECOMPOSITION METHODS

The basic idea of decomposition is to decompose the original large problem into distributively solvable subproblems which are then coordinated by a high-level master problem by means of some kind of signaling.

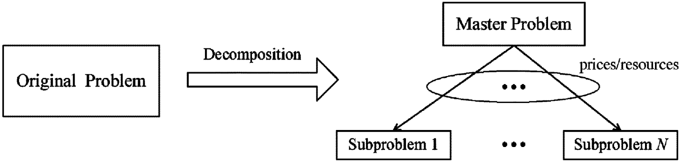

Figure: Decomposition of a problem into several subproblems controlled by a master problem [1].

Decomposition Methods

- Most of the existing decomposition techniques can be classified into
  - Primal Decomposition
  - Dual Decomposition
- The former is based on decomposing the original primal problem. Primal decomposition methods correspond to a direct decision making problem since the master problem decides about the variables by directly giving each subproblem a proper feasible value.
- The latter is based on decomposing the Lagrangian dual problem. Dual decomposition methods correspond to a decision making via pricing (dual variables), since the master problem sets the price for the variables (e.g., resources) to each subproblem, which has to decide the value (e.g., the amount of resources) based on the price.

Dual Decomposition

The dual decomposition is appropriate when the problem has a coupling constraint such that, when relaxed, the optimization problem decouples into several subproblems.
Consider the set $\mathcal{N}$ of firms (e.g., users, devices, and ...)

$$\begin{aligned} \text{maximize} \quad & \sum_{i\in\mathcal{N}} f_i(\mathbf{x}_i) \\ \text{subject to} \quad & \mathbf{x}_i \in \mathcal{X}_i, \forall i \in \mathcal{N} \\ & \sum_i h_i(\mathbf{x}_i) \leq \mathbf{c}. \end{aligned} \tag{4}$$

- The objective function is the sum of utility for different users, but it is separable.
- Constraint $\sum_i h_i(\mathbf{x}_i) \leq \mathbf{c}$ couples the allocation decisions (solution) together. Without this constraint, the problem could decouple.

Dual Decomposition

To mitigate the coupling arised by constraint $\sum_i h_i(\mathbf{x}_i) \leq \mathbf{c}$, form the Lagrangian by relaxing the coupling constraint as

$$L(\mathbf{x}, \lambda) = \sum_{i \in \mathcal{N}} f_i(\mathbf{x}_i) - \lambda^T \left(\sum_i h_i(\mathbf{x}_i) - \mathbf{c} \right). \tag{5}$$

The Lagrange dual function $g(\lambda)$ is obtained from

$$\begin{aligned} \text{maximize} \quad & \sum_{i \in \mathcal{N}} f_i(\mathbf{x}_i) - \lambda^T \left(\sum_i h_i(\mathbf{x}_i) - \mathbf{c} \right) \\ \text{subject to} \quad & \mathbf{x}_i \in \mathcal{X}_i, \forall i \in \mathcal{N}. \end{aligned} \tag{6}$$

The dual problem is

$$\begin{aligned} \text{minimize} \quad & g(\lambda) \\ \text{subject to} \quad & \lambda \succcurlyeq 0. \end{aligned}$$

Dual Decomposition

Separate the problems into two levels of optimization problems.
At the lower level, we have the subproblems (i.e., the Lagrangians), one for each i, in which problem (6) decouples.

$$\begin{array}{ll} \text{maximize} & f_i(\mathbf{x}_i) - \lambda^T h_i(\mathbf{x}_i) \\ \text{subject to} & \mathbf{x}_i \in \mathcal{X}_i. \end{array}$$

At the higher level, we have the master dual problem in charge of updating the dual variable by solving the dual problem:

$$\begin{array}{ll} \text{minimize} & g(\lambda) = \sum_i g_i(\lambda) + \lambda^T \mathbf{c} \\ \text{subject to} & \lambda \succcurlyeq 0. \end{array}$$

where $g_i(\lambda)$ is the dual function obtained as the maximum value of the Lagrangian in the above problem.

It will only give the optimal solution of primal problem if strong duality holds.

Dual Decomposition

Gradient method to solve the master problem (if g is differentiable)

$$\lambda^{(k+1)} = \left[\lambda^{(k)} - \alpha_k \nabla^T g\right]^+$$

The subproblems can be locally and independently solved with knowledge of λ.

Message passing is required to

- Update λ for the subproblems
- Update the gradient for the master problem

Application: Transmission control protocol

Primal Decomposition

A primal decomposition is appropriate when the problem has a coupling variable such that, when fixed to some value, the rest of the optimization problem decouples into several subproblems.

$$
\begin{array}{lll}
\underset{\mathbf{y},\mathbf{x}}{\text{maximize}} & \sum_{i \in \mathcal{N}} f_i(\mathbf{x}_i) & \\
\text{subject to} & \mathbf{x}_i \in \mathcal{X}_i, & \forall i \in \mathcal{N} \\
& \mathbf{A}_i \mathbf{x}_i \preceq \mathbf{y}, & \forall i \in \mathcal{N} \\
& \mathbf{y} \in \mathcal{Y}. &
\end{array}
\tag{7}
$$

If variables $\mathbf{y}$ were fixed, then the problem would decouple.

Decompose the problem into two levels of optimization.

PRIMAL DECOMPOSITION

Lower level: fix $\mathbf{y}$

$$\begin{array}{ll} \underset{\mathbf{x}_i}{\text{maximize}} & f_i(\mathbf{x}_i) \\ \text{subject to} & \mathbf{x}_i \in \mathcal{X}_i \\ & \mathbf{A}_i \mathbf{x}_i \preceq \mathbf{y}. \end{array}$$

Higher level: master problem in charge of updating the coupling variable $\mathbf{y}$.

$$\begin{array}{ll} \underset{\mathbf{y}}{\text{maximize}} & \sum_{i \in \mathcal{N}} f_i^*(\mathbf{y}) \\ & \mathbf{y} \in \mathcal{Y}. \end{array}$$

where $f_i^*(\mathbf{y})$ is the optimal objective value of lower level subproblem for a given $\mathbf{y}$.

Primal Decomposition

If the original problem is a convex optimization problem, so are the subproblems and the master problem.

The master problem can be solved using subgradient method. Each subproblem can be locally and independently solved with the knowledge of $\mathbf{y}$.

Primal Decomposition

Primal decomposition is naturally applicable to resource sharing scenarios where virtualization or slicing of the resources are carried out by dividing the total resource to multiple parts.

The points where the resources are divided can be represented by auxiliary variables in a primal decomposition.

If these variables are fixed, we would have a static slicing of the resources, which can be suboptimal.

If these variables are optimized by a master problem and used to coordinate the allocation of resources to the subproblems, we would have an optimal dynamic slicing.

Indirect Decomposition

Problems with coupling constraints are naturally suited for a dual decomposition, whereas problems with coupling variables are convenient for a primal decomposition.

However, this is not a strict rule as often the problem can be reformulated, and then more effective primal and dual decompositions can be indirectly applied.

Key element: Auxiliary variables

Indirect Decomposition

Problem (7) can be solved with an indirect dual decomposition by first introducing the additional auxiliary variables $\{\mathbf{y}_i\}$.

$$\begin{array}{lll} \underset{\mathbf{y},\{\mathbf{y}_i\},\mathbf{x}}{\text{maximize}} & \sum_{i\in\mathcal{N}} f_i(\mathbf{x}_i) & \\ \text{subject to} & \mathbf{x}_i \in \mathcal{X}_i, & \forall i \in \mathcal{N} \\ & \mathbf{A}_i\mathbf{x}_i \preceq \mathbf{y}_i, & \forall i \in \mathcal{N} \\ & \mathbf{y}_i = \mathbf{y}, & \forall i \in \mathcal{N} \\ & \mathbf{y} \in \mathcal{Y}, & \end{array}$$

then, relaxing the coupling constraints $\mathbf{y}_i = \mathbf{y}$ via a dual decomposition.

Indirect Decomposition

Consider now problem (4) which contains the coupling constraint $\sum_i h_i(\mathbf{x}_i) \leq \mathbf{c}$. Again, by introducing additional auxiliary variables $\{\mathbf{y}_i\}$, the problem becomes

$$
\begin{array}{ll}
\underset{\{\mathbf{y}_i\},\mathbf{x}}{\text{maximize}} & \displaystyle\sum_{i\in\mathcal{N}} f_i(\mathbf{x}_i) \\
\text{subject to} & \mathbf{x}_i \in \mathcal{X}_i, \qquad \forall i \in \mathcal{N} \\
& h_i(\mathbf{x}_i) \leq \mathbf{y}_i, \quad \forall i \in \mathcal{N} \\
& \displaystyle\sum_i \mathbf{y}_i \leq \mathbf{c}.
\end{array}
$$

The coupling variable $\mathbf{y}$ can be dealt with using a primal decomposition.

DECOMPOSITION

- Decomposition techniques are promising approaches for design of distributed algorithms.
- Dual decomposition technique is suitable for problems with coupling constraint.
- Primal decomposition technique is appropriate for problem with coupling variables.
- Both techniques can indirectly be applied to a problem by introducing auxiliary variables.