Lecture 4
Duality and Decomposition Techniques

Jie Lu (jielu@kth.se)

Richard Combes
Alexandre Proutiere

Automatic Control, KTH

September 19, 2013

Lagrange Duality

- Consider the primal problem

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, 2, \ldots, m, \\ & Ax + b = 0, \\ & x \in X. \end{aligned}$$

  - Each $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 0, 1, \ldots, m$ is convex.
  - $A \in \mathbb{R}^{p \times n}$ and $b \in \mathbb{R}^p$.
  - $X \subset \mathbb{R}^n$ is closed and convex.
  - There exists a primal optimum $x^\star$

- Lagrangian function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \nu^T (Ax + b).$$

  - $\lambda = [\lambda_1, \ldots, \lambda_m]^T \in \mathbb{R}^m$ and $\nu \in \mathbb{R}^p$: Lagrange multipliers

Dual Function and Dual Problem

- Dual function $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

$$g(\lambda, \nu) = \min_{x \in X} \mathcal{L}(x, \lambda, \nu) = \min_{x \in X} f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \nu^T(Ax + b)$$

  - g is concave even if the primal problem is not convex.
  - $g(\lambda, \nu) \leq f^\star$, $\forall \lambda \geq 0$, where $f^\star$ is the primal optimal value.
  - $\partial g(\lambda, \nu) = \{[f_1(z), \ldots, f_m(z), (Az + b)^T]^T : z \in \arg\min_{x \in X} \mathcal{L}(x, \lambda, \nu)\}$, $\lambda \geq 0$.
  - g is differentiable at every (λ, ν), $\lambda \geq 0$ if f_0 is strongly convex.

- Lagrange Dual problem

$$\begin{array}{ll} \underset{\lambda, \nu}{\text{maximize}} & g(\lambda, \nu) \\ \text{subject to} & \lambda \geq 0. \end{array}$$

Strong Duality

- Weak duality:

  $g^\star \leq f^\star$, where $g^\star$ is the dual optimal value.
  (always hold, even nonconvex)

- Strong duality:

  $g^\star = f^\star$, i.e., no duality gap.
  (convex+more assumptions)

- **Theorem [Slater's constraint qualifications]:**
  If there exists $\tilde{x} \in \text{rel int}\, X$ such that $f_i(\tilde{x}) < 0$, $\forall i = 1, \ldots, m$ and $A\tilde{x} + b = 0$, then strong duality holds and there is at least one optimal Lagrange multiplier/dual optimum $(\lambda^\star, \nu^\star)$.

Optimality Conditions (Under Strong Duality)

- From now on, suppose strong duality holds.
- **Theorem [Primal optimality condition]:**
  Let $(\lambda^\star, \nu^\star)$ be a dual optimum. Then, $x^\star$ is a primal optimum if and only if $x^\star$ is primal feasible and satisfies
  $$\mathcal{L}(x^\star, \lambda^\star, \nu^\star) = \min_{x \in X} \mathcal{L}(x, \lambda^\star, \nu^\star), \quad \sum_{i=1}^m \lambda_i^\star f_i(x^\star) = 0.$$
- **Theorem [Lagrangian saddle point theorem]:**
  $(x^\star, \lambda^\star, \nu^\star)$ forms a primal-dual optimal pair if and only if $(x^\star, \lambda^\star, \nu^\star)$ is a saddle point of $\mathcal{L}$ in the sense that $x^\star \in X$, $\lambda^\star \geq 0$, and
  $$\mathcal{L}(x^\star, \lambda, \nu) \leq \mathcal{L}(x^\star, \lambda^\star, \nu^\star) \leq \mathcal{L}(x, \lambda^\star, \nu^\star), \quad \forall x \in X, \ \forall \lambda \geq 0, \ \forall \nu \in \mathbb{R}^p.$$
- **Theorem [Necessary and sufficient optimality condition]:**
  $(x^\star, \lambda^\star, \nu^\star)$ forms a primal-dual optimal pair if and only if
  $$x^\star \text{ is primal feasible,} \quad \text{(Primal feasibility)}$$
  $$\lambda^\star \geq 0, \quad \text{(Dual feasibility)}$$
  $$\mathcal{L}(x^\star, \lambda^\star, \nu^\star) = \min_{x \in X} \mathcal{L}(x, \lambda^\star, \nu^\star), \quad \text{(Lagrangian optimality)}$$
  $$\sum_{i=1}^m \lambda_i^\star f_i(x^\star) = 0. \quad \text{(Complementary Slackness)}$$

Primal Function

- Consider the previous problem with inequality constraints only:

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, 2, \ldots, m, \\ & x \in X. \end{array}$$

- Primal function

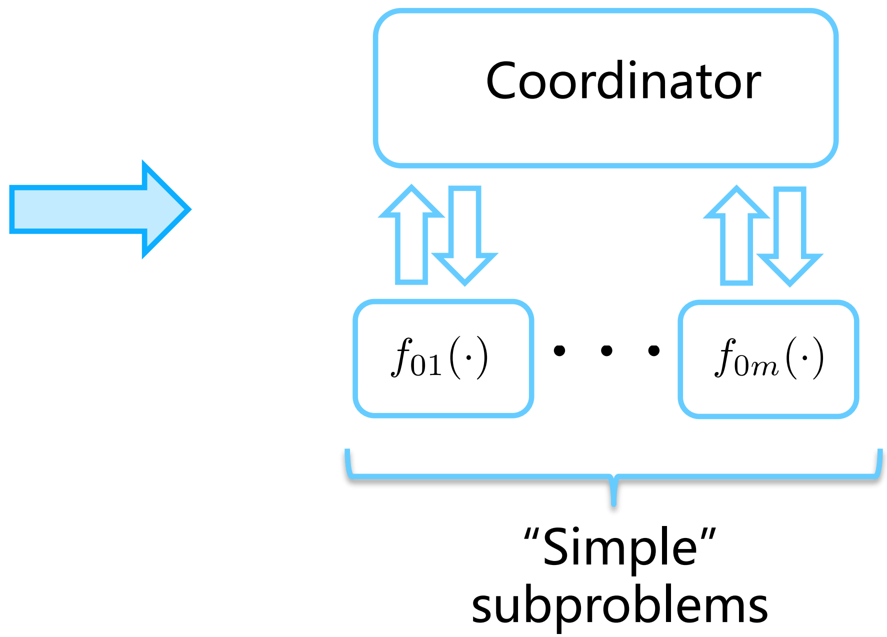
$$p(u) = \inf_{x \in X:\ f_i(x) \leq u_i,\ \forall i = 1, \ldots, m} f_0(x)$$

  - $p(u)$: optimal value of the primal problem with perturbed constraints
  - $p(u)$ is convex
  - Strong duality holds if and only if p is lower semicontinuous at $u = 0$.
  - Let $\lambda^\star \geq 0$ be a dual optimum and suppose strong duality holds. Then, $-\lambda^\star \in \partial p(0)$.

Decomposition Techniques

- **Basic idea:** Decompose one complex problem into many small:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \in X \end{array}$$



The Trivial Case

- Separable objectives and constraints

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n} f_{0i}(x_i) \\ \text{subject to} & x_i \in X_i \end{array}$$

- Trivially separates into n decoupled subproblems

$$\begin{array}{ll} \text{minimize} & f_{0i}(x_i) \\ \text{subject to} & x_i \in X_i \end{array}$$

that can be solved in parallel and combined.

More Interesting Ones

- Problems with **coupling constraints**

$$\begin{array}{ll} \text{minimize} & f_{01}(x_1) + f_{02}(x_2) \\ \text{subject to} & x_1 + x_2 \leq c \end{array}$$

- Problems with **coupled objectives**

$$\text{minimize} \quad f_{01}(x_1, x_{12}) + f_{02}(x_{12}, x_2)$$

- Coupled objectives can be cast as a problem of coupling constraints:

$$\begin{array}{ll} \text{minimize} & f_{01}(x_1, z_{12}) + f_{02}(z_{21}, x_2) \\ \text{subject to} & z_{12} = z_{21} \end{array}$$

Dual Decomposition

- **Basic idea:** decouple problem by relaxing coupling constraints.

$$\begin{array}{ll} \text{minimize} & f_{01}(x_1) + f_{02}(x_2) \\ \text{subject to} & x_1 + x_2 \leq c \end{array}$$

- Dual function

$$\mathcal{L}(x, \lambda) = f_{01}(x_1) + f_{02}(x_2) + \lambda(x_1 + x_2 - c)$$

$$g(\lambda) = \min_{x} \mathcal{L}(x, \lambda) = -\lambda c + \min_{x_1}\{f_{01}(x_1) + \lambda x_1\} + \min_{x_2}\{f_{02}(x_2) + \lambda x_2\}$$

- Dual problem

$$\begin{array}{ll} \text{maximize} & g_1(\lambda) + g_2(\lambda) \\ \text{subject to} & \lambda \geq 0 \end{array}$$

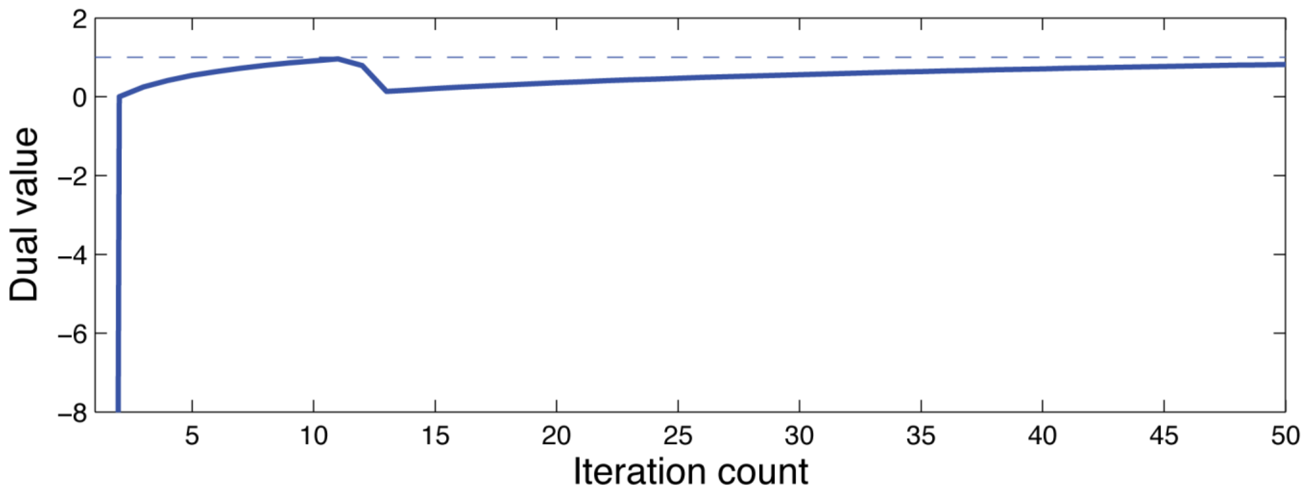
  - Additive (hence, can be evaluated in parallel) with simple constraints
  - Subgradient of the dual function: $x_1^\star(\lambda) + x_2^\star(\lambda) - c$
  - Can be solved using subgradient projection method

Example of Dual Decomposition

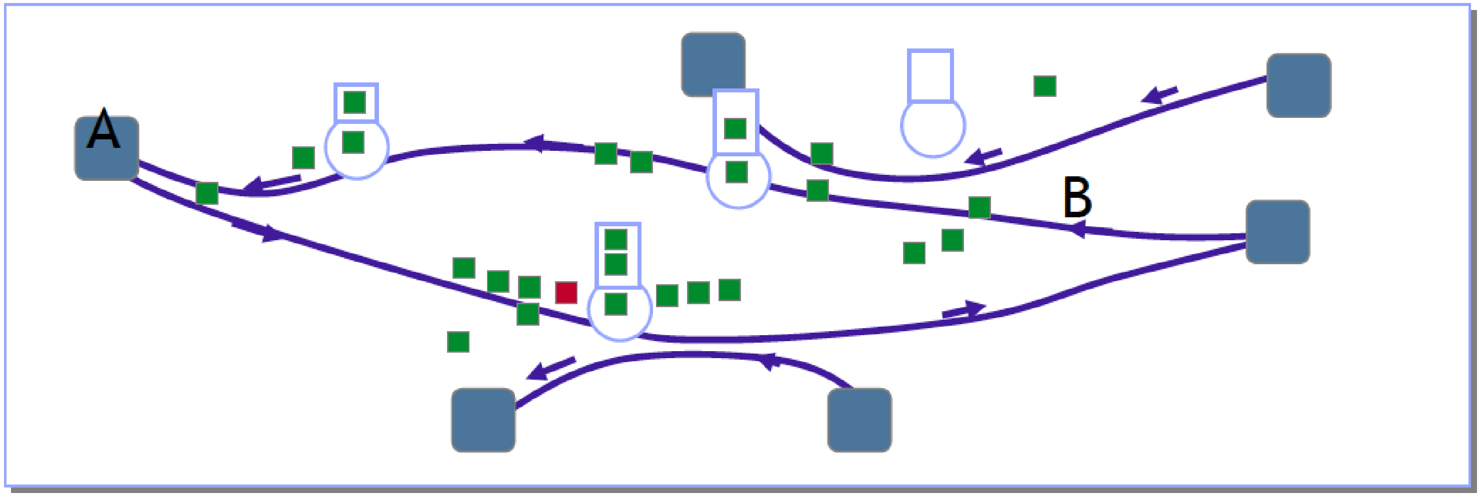
- Problem

$$\begin{array}{ll} \text{minimize} & |x_1 - 1| + |x_2 - 1| \\ \text{subject to} & x_1 + x_2 \leq 1 \\ & x_i \in [0, 10] \end{array}$$

  - Optimal value $f_0^\star = 1$
  - Primal optimum $x_1^\star = 1 - x_2^\star, \quad x_2^\star \in [0, 1]$



Network Utility Maximization (NUM)



- Adjust end-to-end rates fairly to share limited network capacity

$$\begin{array}{ll} \text{maximize}_{x \in \mathbb{R}^N} & \sum_{i=1}^{N} u_i(x_i) \\ \text{subject to} & \sum_{i \in \mathcal{P}(\ell)} x_i \leq c_\ell, \quad \forall \ell \in L, \\ & x_i \in X_i. \end{array}$$

  - u_i : utility function (concave)
    - Fairness by appropriate utility functions, e.g., $u_i(x_i) = \log(x_i)$

Network Utility Maximization (NUM)

- Rewrite the problem in vector form

$$\begin{array}{ll} \text{maximize}_{x\in\mathbb{R}^N} & \sum_{i=1}^N u_i(x_i) \\ \text{subject to} & Rx \leq c, \\ & x \in X. \end{array}$$

  - *R*: Routing matrix
    - Row *i* of *R* indicates which flows share link *i*

$$\sum_{i=1}^N r_{\ell i} x_i = \sum_{i\in\mathcal{P}(\ell)} x_i \leq c_\ell$$

- Rewrite on our standard form ($f_i = -u_i$)

$$\begin{array}{ll} \text{minimize}_{x\in\mathbb{R}^N} & \sum_{i=1}^N f_i(x_i) \\ \text{subject to} & Rx \leq c, \\ & x \in X \end{array}$$

Dual Decomposition for NUM

- Form Lagrangian

$$\mathcal{L}(x,\lambda) = \sum_{i=1}^{N} f_i(x_i) - \lambda^T(Rx - c) = \lambda^T c + \sum_{i=1}^{N} f_i(x_i) - x_i \sum_{\ell \in L(i)} \lambda_\ell$$

- Dual function is additive

$$g(\lambda) = \lambda^T c + \sum_{i=1}^{N} \min_{x_i \in X_i} \left\{ f_i(x_i) - x_i \sum_{\ell \in L(i)} \lambda_\ell \right\}$$

- Each source *i* can adjust its rate x_i based on feedback of $\sum_{\ell \in L(i)} \lambda_\ell$

- Use the dual subgradient projection method

$$\lambda(t+1) = P_+[\lambda(t) + \alpha(Rx(t) - c)] \text{ (dual iterate)}$$
$$x(t) = x^\star(\lambda(t)) \in \arg\min_{x \in X} \mathcal{L}(x, \lambda(t)) \text{ (primal iterate)}$$

$$\Updownarrow$$

$$\lambda_\ell(t+1) = \max\{0, \lambda_\ell(t) + \alpha(\sum_{i \in P(\ell)} x_i(t) - c_\ell)\}, \quad \forall \ell \in L$$

$$x_i(t) = \arg\min_{x_i \in X} f_i(x_i) - x_i \sum_{\ell \in L(i)} \lambda_\ell, \quad \forall i = 1, 2, \ldots, N$$

  - Use locally available information at the router queues

Drawback of Dual Decomposition

- The dual iterates can converge to some dual optimum.
- However, in general, the primal iterates can only reach sub-optimality and violate constraints.
- Suppose strong duality holds. Feasibility and primal optimality recovered in the limit.
- How to evaluate the primal optimality and feasibility of each primal iterate? How fast do primal iterates converge to primal optimum?

Evaluate Optimality of Primal Iterates

- Strongly convex objective function ($\mu > 0$) and linear constraints

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f_0(x) \\ \text{subject to} & Ax + b \leq 0, \\ & x \in X \end{array}$$

- Dual function is differentiable
- Dual function has Lipschitz gradient with Lipschitz constant $\frac{\lambda_{\max}(A^T A)}{\mu}$
  - Guaranteed convergence rate of the dual iterates
- $\|x^\star(\lambda) - x^\star\| \leq \frac{\sqrt{\lambda_{\max}(A^T A)}}{\mu} \|\lambda - \lambda^\star\|$
- $\|x^\star(\lambda) - x^\star\| \leq \sqrt{\frac{g^\star - g(\lambda)}{\mu}}$
- Allow the linear constraints to be both inequality and equality

Primal Convergence of Running Average

- Running average of primal iterates

$$\overline{x}^\star(t) = \frac{1}{t}\sum_{k=0}^{t} x^\star(\lambda(t))$$

- Using subgradient projection method,

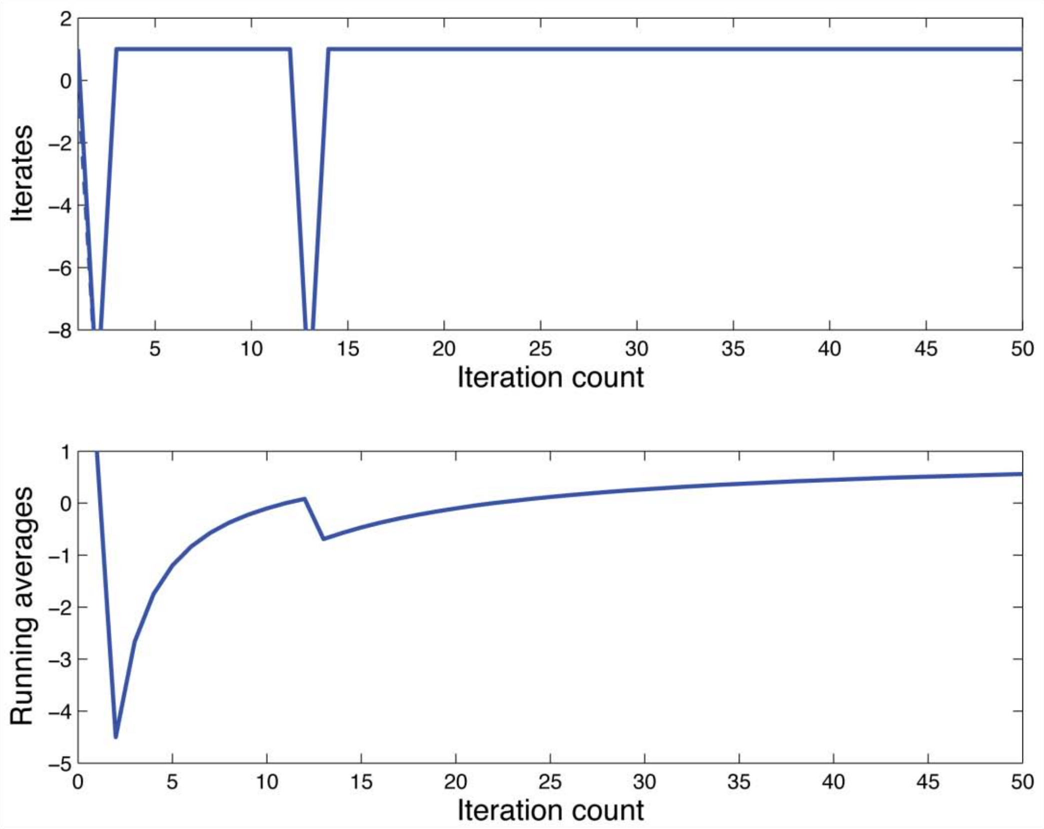
$$f_0(\overline{x}^\star(t)) \le f^\star + \frac{\alpha L^2}{2} + \frac{\|\lambda(0)\|^2}{2t\alpha}$$

  - **Note:** L is not Lipschitz constant (but an upper bound on constraint violation of the primal iterates)
  - For more details,

    A. Nedic and A. Ozdaglar, *Approximate Primal Solutions and Rate Analysis for Dual Subgradient Methods*, SIAM Journal on Optimization 19 (4) 1757-1780, 2009.

Example

- Simple example as before



Augmented Lagrangian

- Consider

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f_0(x) \\ \text{subject to} \quad & Ax + b = 0. \end{aligned}$$

- Augmented Lagrangian

$$\begin{aligned} \mathcal{L}_\rho(x, \lambda) &= \mathcal{L}_0(x, \lambda) + \frac{\rho}{2}\|Ax + b\|^2 \\ &= f_0(x) + \lambda^T(Ax + b) + \frac{\rho}{2}\|Ax + b\|^2 \end{aligned}$$

  - $\rho > 0$: penalty parameter
  - The unaugmented Lagrangian of the equivalent problem

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f_0(x) + \tfrac{\rho}{2}\|Ax + b\|^2 \\ \text{subject to} \quad & Ax + b = 0. \end{aligned}$$

Method of Multipliers

- The dual function associated with the augmented Lagrangian is differentiable
- The dual methods lead to convergence under more general conditions
- Method of multipliers

$$x(t+1) = \arg\min_x \mathcal{L}_\rho(x, \nu(t))$$
$$\nu(t+1) = \nu(t) + \rho(Ax(t+1) + b)$$

  - The dual gradient method applied to the dual associated with the augmented Lagrangian (with step-size ρ, Why?)
  - Does not need f_0 to be strongly convex to guarantee convergence
  - Dual and primal convergence rates can be derived

Alternating Direction Method of Multipliers (ADMM)

- Consider

$$\begin{aligned} \underset{x\in\mathbb{R}^n, z\in\mathbb{R}^m}{\text{minimize}} \quad & f(x) + g(z) \\ \text{subject to} \quad & Ax + Bz + c = 0. \end{aligned}$$

- The augmented Lagrangian

$$\mathcal{L}_\rho(x, z, \nu) = f(x) + g(z) + \nu^T(Ax + Bz + c) + \frac{\rho}{2}\|Ax + Bz + c\|^2$$

- Drawback of the method of multipliers : not parallelizable in

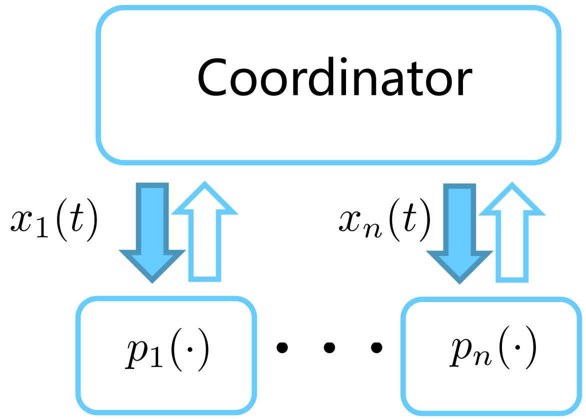
$$(x(t+1), z(t+1)) = \text{arg min}_{x,z}\, \mathcal{L}_\rho(x, z, \nu(t))$$

- ADMM

$$\begin{aligned} x(t+1) &= \text{arg min}_x\, \mathcal{L}_\rho(x, z(t), \nu(t)) \\ z(t+1) &= \text{arg min}_z\, \mathcal{L}_\rho(x(t+1), z, \nu(t)) \\ \nu(t+1) &= \nu(t) + \rho(Ax(t+1) + Bz(t+1) + c) \end{aligned}$$

Primal Decomposition

- Basic idea: coordinator immediately allocates primal variables



- Feasibility of primal iterates guaranteed throughout.

Primal Decomposition

- Consider the resource allocation problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq c_i \\ & \sum_i c_i \leq c_{\text{tot}} \end{array}$$

- Rewrite as

$$\begin{array}{ll} \text{minimize} & p(c) \\ \text{subject to} & \sum_i c_i \leq c_{\text{tot}} \end{array}$$

  - $p(c) = \inf_x \{f_0(x) \mid f_i(x) \leq c_i,\ i = 1, \ldots, n\}$

Primal Decomposition

- $p(c) = \inf_x \{f_0(x) \mid f_i(x) \le c_i,\ i = 1, \ldots, n\}$ is the primal function of

$$\begin{array}{ll} \text{minimize}_{x\in\mathbb{R}^n} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \ldots, n \end{array}$$

- $p(c)$ is convex because the above problem is convex

- A subgradient of $p(c)$ is $-\lambda^\star(c)$, where $\lambda^\star(c)$ is the dual optimal solution of

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le c_i,\ i = 1, \ldots, n \end{array}$$

  - $p(c)$ is differentiable if the optimal Lagrange multiplier is unique

- Hence, coordinator can use subgradient projection method to allocate resources

Distributed Primal Decomposition

- Primal decomposition advantageous when primal function has the form

$$p(c) = \sum_i p_i(c_i) = \sum_i \inf_{x_i} \{f_{0i}(x_i) \mid f_i(x_i) \leq c_i\}$$

- The projection in the update

$$c(t+1) = P_C\{c(t) - \alpha(t)\lambda^\star(c(t))\}$$

can be computed in a distributed manner. (more about this in Lecture 4)

Modeling for Decomposition

- Clever introduction of new variables enable distribution of dual, primal

- **Example:** Transforming coupling variables into coupling constraints.

$$\text{minimize} \quad f_{01}(x_1, x_{12}) + f_{02}(x_{12}, x_2)$$

$$\Downarrow$$

$$\begin{array}{ll} \text{minimize} & f_{01}(x_1, z_{12}) + f_{02}(z_{21}, x_2) \\ \text{subject to} & z_{12} = z_{21} \end{array}$$

- **Example:** Making problem that is a clear candidate for dual decomposition

$$\begin{array}{ll} \text{minimize} & \sum_i f_{0i}(x_i) \\ \text{subject to} & \sum_i f_i(x_i) \le c_{\text{tot}} \end{array} \quad \Longrightarrow \quad \begin{array}{ll} \text{minimize} & \sum_i f_{0i}(x_i) \\ \text{subject to} & f_i(x_i) \le c_i \\ & \sum_i c_i \le c_{\text{tot}} \end{array}$$

into the standard form for primal decomposition

Summary

- Duality
  - strong duality
  - primal and dual optimality
  - primal function
- Decomposition: subdivide large problem into many small
  - coupling constraints, coupling variables
- Dual decomposition
  - relax coupling constraints to make dual function additive/distributable
  - dual problem possibly non-smooth, might require central coordinator
  - primal iterates not always well-behaved, might need primal recovery
- Primal decomposition
  - coordinator immediately allocates resources to subsystems
  - feasibility of primal iterates guaranteed
  - can sometimes be distributed (more in next lecture)

References

- Duality theory
  - D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1999.
  - D. P. Bertsekas, A. Nedich, and A. Ozdaglar, *Convex Analysis and Optimization*. Belmont, MA: Athena Scientific, 2003.
  - S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY: Cambridge University Press, 2004
- Decomposition techniques
  - B. Johansson, P. Soldati and M. Johansson, *Mathematical decomposition techniques for distributed cross-layer optimization of data networks*, IEEE Journal on Selected Areas in Communications, Vol. 24, No. 8, pp. 1535-1547, August 2006.
- ADMM
  - S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, Foundations and Trends in Machine Learning, 3(1):1–122, 2011.