An Introduction to ADMM: Alternating Directions Method of Multipliers

Saurabh Khanna,
Signal Processing for Communication, ECE, IISc

Outline

- Solving constrained convex optimization problem
  - Lagrangian method and KKT conditions
  - Dual function and its properties
  - Dual problem and its significance
  - Fenchel's duality (a geometric perspective)
- Introduction to ADMM
  - Dual Ascent
  - Dual Decomposition
  - Augmented Lagrangian and Method of Multipliers
  - Alternating Directions Method of Multipliers

PART-I: Solving constrained convex optimization problem

Conjugate functions

- For a convex function $f : \mathbb{R}^n \to \mathbb{R}$, its convex conjugate f^* is defined as:

$$f^*(\mathbf{z}) = \sup_{\mathbf{x} \in \text{dom } f} \left(\mathbf{z}^T \mathbf{x} - f(\mathbf{x}) \right)$$

- Geometric interpretation:
  $f^*(\mathbf{z})$ is the negative intercept on y-axis made by tangent to curve $y = f(\mathbf{x})$ with slope $\mathbf{z}$.

Lagrangian method

- Standard constraint optimization problem (P):

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0, \qquad 1 \leq i \leq n \\ & h_i(\mathbf{x}) = 0, \qquad 1 \leq i \leq m \end{aligned}$$

  where $\mathbf{x} \in \mathbb{R}^n$, and $f, g_i, h_i : \mathbb{R}^n \to \mathbb{R}$.

- Let p^* denote the primal optimal value.

- Lagrangian function L is given by:

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\mathbf{x}) + \sum_{i=1}^{n} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{m} \nu_i h_i(\mathbf{x})$$

  - $\lambda_i > 0$ are Lagrange multipliers associated with $g_i(\mathbf{x}) \leq 0$.
  - ν_i are Lagrange multipliers associated with $h_i(\mathbf{x}) = 0$.

Karush-Kuhn-Tucker (KKT) conditions

- If
  1. Slater's conditions hold.
  2. f, g_i and h_i are differentiable

  then, optimal values of $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ must satisy:

  - primal feasibility constraints: $g_i(\mathbf{x}^*) \leq 0$ and $h_i(\mathbf{x}^*) = 0$.
  - dual feasibility constraints: $\lambda_i^* \geq 0$.
  - complementary slackness: $\lambda_i^* g_i(\mathbf{x}^*) = 0$.
  - gradient of Lagrangian with respect to **x** is zero i.e.,

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{n} \lambda_i^* \nabla g_i^*(\mathbf{x}^*) + \sum_{i=1}^{m} \nu_i^* \nabla h_i^*(\mathbf{x}^*) = 0$$

Lagrange dual function

- Lagrange dual function $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is defined as:

$$
\begin{aligned}
g(\boldsymbol{\lambda}, \boldsymbol{\nu}) &= \inf_{\mathbf{x} \in dom\ f} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \\
&= \inf_{\mathbf{x} \in dom\ f} \left(f(\mathbf{x}) + \sum_{i=1}^{n} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{m} \nu_i h_i(\mathbf{x}) \right)
\end{aligned}
$$

  where $\boldsymbol{\lambda} = (\lambda_1 \dots \lambda_n)$ and $\boldsymbol{\nu} = (\nu_1 \dots \nu_n)$ are the dual variables.

- Dual function g is **always concave** w.r.t. λ_i and ν_i.

- **Lower bound property** of dual function:

$$\text{If } \boldsymbol{\lambda} \succeq 0 \text{ then, } g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*.$$

  where p^* is the optimal value of objective function in primal problem (P).

The dual problem

- Lagrange dual problem is given by:

$$\begin{aligned} &\text{maximize } g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ &\text{subject to } \boldsymbol{\lambda} \succeq 0 \end{aligned}$$

- The dual problem find the best lower bound on p^*.
- $\boldsymbol{\lambda}, \boldsymbol{\nu}$ are dual feasible if $\boldsymbol{\lambda} \succeq 0$ and $(\boldsymbol{\lambda}, \boldsymbol{\nu}) \in \operatorname{dom} g$
- Question: Why should we care about the dual problem?

Weak and strong duality

- **Weak duality:** $d^* \leq p^*$
  - always holds (for convex and nonconvex problems) can be used to find nontrivial lower bounds for difficult problems
- **Strong duality:** $d^* = p^*$
  - does not hold in general
  - usually holds for convex problems
  - conditions that guarantee strong duality in convex problems are called *constraint qualifications*.

Obtaining primal solution

- Let $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ be the solution to the dual problem:

$$(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = \underset{\boldsymbol{\lambda}, \boldsymbol{\nu}}{\text{argmax}}\ g(\boldsymbol{\lambda}, \boldsymbol{\nu})$$

- Then, $\mathbf{x}^*$, the solution to primal problem is obtained by solving the minimization problem:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\text{argmin}}\ L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$$

A simple example

- Minimum norm solution to an underdetermined system of linear equations

$$\begin{aligned} &\text{minimize } \mathbf{x}^T\mathbf{x} \\ &\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

- Lagrangian: $L(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T\mathbf{x} + \mathbf{y}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$.

- Dual function: $g(\mathbf{y}) = \inf_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}) = \inf_{\mathbf{x}} \left(\mathbf{x}^T\mathbf{x} + \mathbf{y}^T(\mathbf{A}\mathbf{x} - \mathbf{b})\right)$
$= -\frac{1}{4}\mathbf{y}^T\mathbf{A}\mathbf{A}^T\mathbf{y} - \mathbf{y}^T\mathbf{b}$.

- Dual optimal $\mathbf{y}^* = \operatorname{argmax}_{\mathbf{y}} g(\mathbf{y}) = -2\left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{b}$.

- Primal optimal $\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}^*) = \mathbf{A}^T\left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{b}$.

Lagrange dual function and conjugate function

- Construction of dual problem is simplified if conjugate of objective function is known.

- For example, consider the convex optimation problem:

$$\begin{aligned} &\text{minimize } f(\mathbf{x}) \\ &\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

- Lagrangian $L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) + \mathbf{y}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$.

- Dual function $g(\mathbf{y}) = \inf_{\mathbf{x}} \left(f(\mathbf{x}) + \mathbf{y}^T(\mathbf{A}\mathbf{x} - \mathbf{b})\right)$

$$\begin{aligned} &= \inf_{\mathbf{x}} \left(f(\mathbf{x}) - (-\mathbf{A}^T\mathbf{y})\mathbf{x})\right) - \mathbf{y}^T\mathbf{b} \\ &= f^*(-\mathbf{A}^T\mathbf{y}) - \mathbf{y}^T\mathbf{b} \end{aligned}$$

- Recall definition of convex conjugate $f^*(\mathbf{y}) = \sup_{\mathbf{x} \in \text{dom } f} \left(\mathbf{y}^T\mathbf{x} - f(\mathbf{x})\right)$.

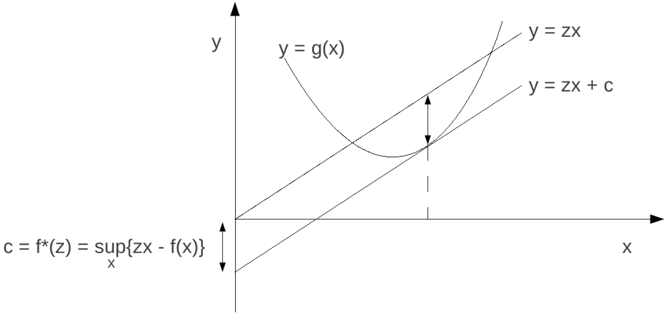
Fenchel's duality - conjugate functions

- For a convex function $f : \mathbb{R}^p \to \mathbb{R}$, its convex conjugate f^* is defined as:
$$f^*(\mathbf{z}) = \underset{\mathbf{x}\in\mathbb{R}^P}{sup}(\mathbf{z}^T\mathbf{x} - f(\mathbf{x})) \tag{1}$$

- For a concave function $f : \mathbb{R}^p \to \mathbb{R}$, its concave conjugate g_* is defined as:
$$g_*(\mathbf{z}) = \underset{\mathbf{x}\in\mathbb{R}^P}{inf}(\mathbf{z}^T\mathbf{x} - g(\mathbf{x})) \tag{2}$$

Fenchel's duality - conjugate functions

- Geometric interpretation of conjuate function.



- Conjugate function $f^*(\mathbf{z})$ is the negative intercept on y-axis made by tangent to curve $y = f(\mathbf{x})$ with slope $\mathbf{z}$

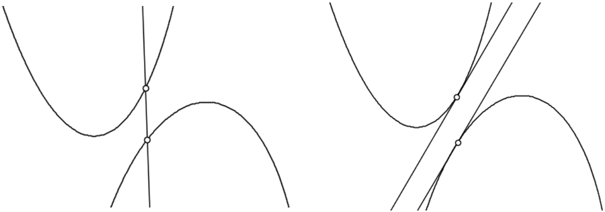
Fenchel's duality

Fenchel's duality theorem

For any convex function f and concave function g, we have,

$$\min_{\mathbf{x}\in\mathbb{R}^p}(f(\mathbf{x}) - g(\mathbf{x})) = \max_{\mathbf{z}\in\mathbb{R}^p}(g_*(\mathbf{z}) - f^*(\mathbf{z}))$$

- Geometric interpretation of Fenchel's duality theory.



(a) $\min_{\mathbf{x}\in\mathbb{R}^p}(f(\mathbf{x}) - g(\mathbf{x}))$

(b) $\max_{\mathbf{z}\in\mathbb{R}^p}(g_*(\mathbf{z}) - f^*(\mathbf{z}))$

Complementary slackness condition

- Let $\mathbf{x}^*$ be the primal optimal and $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ be the dual optimal for standard convex optimization problem (P).

- If strong duality holds, we have,

$$
\begin{aligned}
f(\mathbf{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \leq L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= f(\mathbf{x}^*) + \sum_{i=1}^{n} \lambda_i^* g_i(\mathbf{x}^*) + \sum_{i=1}^{m} \nu_i^* h_i(\mathbf{x}^*) \\
&= f(\mathbf{x}^*) + \sum_{i=1}^{n} \lambda_i^* g_i(\mathbf{x}^*) \\
\Rightarrow 0 &\leq \sum_{i=1}^{n} \lambda_i^* g_i(\mathbf{x}^*)
\end{aligned}
$$

- Since $\sum_{i=1}^{n} \lambda_i^* g_i(\mathbf{x}^*) \leq 0$, we have $\sum_{i=1}^{n} \lambda_i^* g_i(\mathbf{x}^*) = 0$.

Strong duality tells relation between primal and dual solutions

- If Slater's conditions (*constraint qualifications*) hold, strong duality holds.

- If strong duality holds, we have

$$\begin{aligned} f(\mathbf{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\ f(\mathbf{x}^*) &= \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \end{aligned}$$

- Since $\mathbf{x}^*$ is the unique minimizer of f in the given feasibility set, following must hold:

$$\mathbf{x}^* = \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$$

PART-II: Introduction to ADMM

Dual Ascent (1/3)

- Consider the convex optimization

$$\begin{aligned} &\text{minimize } f(\mathbf{x}) \\ &\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function.

- The Lagrangian is given by

$$L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) + \mathbf{y}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

where $\mathbf{y}$ is the dual variable or Lagrangian multiplier.

- The dual function is given by

$$g(\mathbf{y}) = \inf_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}) = f^*(-\mathbf{A}^T\mathbf{y}) - \mathbf{b}^T\mathbf{y}$$

where f^* is convex conjugate of f.

Dual Ascent (2/3)

- The dual problem is

$$\max_{\mathbf{y}} g(\mathbf{y})$$

- The primal optimal point $\mathbf{x}^*$ can be found from a dual optimal point as

$$\mathbf{x}^* = \operatorname*{argmin}_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}^*)$$

- A unique minimizer exists if f is strictly convex.

Dual Ascent (3/3)

- *Dual Ascent* method is as follows:

$$\mathbf{x}^{k+1} = \underset{\mathbf{x}}{\operatorname{argmin}}\ L(\mathbf{x}, \mathbf{y}^k)$$
$$\mathbf{y}^{k+1} = \mathbf{y}^k + \alpha^k(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b})$$

- For proper choice of stepsize α^k, the value of dual function increases in each iteration.
  - α^k is a non-increasing sequence.

- Under assumptions on f, $\mathbf{y}^k$ converges to dual optimal $\mathbf{y}^*$ and $\mathbf{x}^k$ converges to primal optimal $\mathbf{x}^*$, as $k \to \infty$.

Dual decomposition (1/2)

- If objective function f is separable, them dual ascent method can lead to a decentralized algorithm.
- Say f is separable such that,

$$f = \sum_{i=1}^{N} f_i(\mathbf{x}_i) \tag{3}$$

where $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2 \ldots \mathbf{x}_N)$ and the variables $\mathbf{x}_i \in \mathbb{R}^{n_i}$ are subvectors of $\mathbf{x}$.

- The equality constraint $\mathbf{Ax} = \mathbf{b}$ can also be split as:

$$\sum_{i=1}^{N} (\mathbf{A}_i \mathbf{x}_i - \frac{1}{N}\mathbf{b}) = 0$$

where $\mathbf{A} = [\mathbf{A}_1 \ldots\ldots \mathbf{A}_N]$.

Dual decomposition (2/2)

- The Lagrangian can be written in split form as

$$L(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{N} L_i(\mathbf{x}_i, \mathbf{y}) = \sum_{i=1}^{N} \left(f_i(\mathbf{x}) + \mathbf{y}^T(\mathbf{A}_i \mathbf{x}_i - \frac{1}{N}\mathbf{y}^T \mathbf{b}) \right)$$

- The dual ascent method leads to a decentralized algorithm:

$$\mathbf{x}_i^{k+1} = \underset{\mathbf{x}_i}{\operatorname{argmin}}\, L_i(\mathbf{x}_i, \mathbf{y}^k) \qquad 1 \leq i \leq N$$
$$\mathbf{y}^{k+1} = \mathbf{y}^k + \alpha^k(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b})$$

- **Decentralized Implementation:**
  1. Each node performs primal update step.
  2. Each node broadcasts its residual $\mathbf{A}_i \mathbf{x}_i - \frac{1}{N}\mathbf{b}$ to other nodes.
  3. Each node sums the residuals from individual nodes and performs dual update step.

Augmented Lagrangian and Method of Multipliers (1/2)

- Consider primal problem (P1):

$$\begin{aligned} &\text{minimize } f(\mathbf{x}) \\ &\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

- We construct *Augmented Lagrangian*:

$$L_\rho(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) + \mathbf{y}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) + \frac{\rho}{2}||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2$$

- Augmented Lagrangian can be viewed as the Lagrangian for a different primal problem (P2)

$$\begin{aligned} &\text{minimize } f(\mathbf{x}) + \frac{\rho}{2}||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 \\ &\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

- Primal problems (P1) and (P2) have same optimal point but (P2) has a more well behaved cost function.

Augmented Lagrangian and Method of Multipliers (2/2)

- By applying dual ascent method:

$$\mathbf{x}^{k+1} = \underset{\mathbf{x}}{\operatorname{argmin}}\, L_\rho(\mathbf{x}, \mathbf{y}^k)$$
$$\mathbf{y}^{k+1} = \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b})$$

- By using ρ as stepsize in dual ascent step, the iterate $(\mathbf{x}^{k+1}, \mathbf{y}^{k+1})$ is dual feasible.
  Proof: We work out.

- **Positives:** Convergence under more relaxed conditions.
  - f need not be unbounded or strictly convex

- **Negatives:** Due to quadratic penalty term in augmented Lagrangian, separability of f no longer results in a decentralized algorithm!

ADMM: Alternating Directions Method of Multipliers (1/2)

- ADMM problem setup:

$$\begin{aligned}&\text{minimize } f(\mathbf{x}) + g(\mathbf{z})\\ &\text{subject to } \mathbf{Ax} + \mathbf{Bz} = \mathbf{c}\end{aligned}$$

  where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{z} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{p\times n}$, $\mathbf{B} \in \mathbb{R}^{p\times m}$ and $\mathbf{c} \in \mathbb{R}^p$.

- f and g are convex functions.

- Augmented Lagrangian is given by:

$$L_\rho(\mathbf{x},\mathbf{z},\mathbf{y}) = f(\mathbf{x}) + g(\mathbf{z}) + \mathbf{y}^T(\mathbf{Ax}+\mathbf{Bz}-\mathbf{c}) + \frac{\rho}{2}||\mathbf{Ax}+\mathbf{Bz}-\mathbf{c}||_2^2$$

ADMM: Alternating Directions Method of Multipliers (2/2)

- The primal and dual update equations are given by:

$$
\begin{aligned}
\mathbf{x}^{k+1} &= \underset{\mathbf{x}}{\operatorname{argmin}}\ L_\rho(\mathbf{x}, \mathbf{z}^k, \mathbf{y}^k) \\
\mathbf{z}^{k+1} &= \underset{\mathbf{z}}{\operatorname{argmin}}\ L_\rho(\mathbf{x}^{k+1}, \mathbf{z}, \mathbf{y}^k) \\
\mathbf{y}^{k+1} &= \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c})
\end{aligned}
$$

- Primal variable update equation is executed in *Gauss Siedel* fashion.

- Dual variable update equation is similar to Method of Multipliers.

- If f is separable, a decentralized algorithm is possible.

Some questions..

- Does this iterative algorithm converge?
- If the algorithm converges, does it converge to correct value?
- How fast is the convergence?
  - How does primal gap $||f(\mathbf{x}^k) - f(\mathbf{x}^*)||_2$ decays with each iteration.
- What is a reasonable stopping criterion?
  - $||\mathbf{x}^k - \mathbf{x}^{k-1}||_2 \leq \epsilon$ is an overkill !
- How sensitive is the algorithm with respect to changes in algorithm parameters?
  - Sensitivity of ADMM's convergence with respect to augmented Lagrangian parameter ρ.

Convergence of ADMM

- Under assumptions:
  1. The functions f and g are closed, proper and convex.
  2. The unaugmented Lagrangian L_0 has a saddle point.

- We have:
  - *Residual convergence:*

$$\text{as } k \to 0, \ \mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} \to 0.$$

  - *Objective convergence:*

$$\text{as } k \to 0, \ f(\mathbf{x}^k) + g(\mathbf{z}^k) \to p^*).$$

  - *Dual variable convergence:*

$$\text{as } k \to 0, \ \mathbf{y}^k \to \mathbf{y}^*.$$

    where $\mathbf{y}^*$ is the dual optimal point.

ADMM and optimality conditions (1/2)

- Optimality conditions for ADMM problem consists of three conditions:

  1. Primal feasibility condition

  $$\mathbf{A}\mathbf{x}^* + \mathbf{B}\mathbf{z}^* - \mathbf{c}$$

  2. First dual feasibility condition:

  $$0 \in \partial f(\mathbf{x}^*) + \mathbf{A}^T\mathbf{y}^*$$

  3. Second dual feasibility condition:

  $$0 \in \partial g(\mathbf{z}^*) + \mathbf{B}^T\mathbf{y}^*$$

ADMM and optimality conditions (2/2)

- Primal and first dual feasibility are achieved as $k \to \infty$.
- $(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}, \mathbf{y}^{k+1})$ always satisfy second dual feasibility condition.

Proof.

From primal update equation for $\mathbf{z}$, we have:

$$\begin{aligned} 0 &\in \partial g(\mathbf{z}^{k+1}) + \mathbf{B}^T\mathbf{y}^k + \rho\mathbf{B}^T\left(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c}\right) \\ \Rightarrow 0 &\in \partial g(\mathbf{z}^{k+1}) + \mathbf{B}^T\left(\mathbf{y}^k + \rho\left(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c}\right)\right) \\ \Rightarrow 0 &\in \partial g(\mathbf{z}^{k+1}) + \mathbf{B}^T\mathbf{y}^{k+1} \end{aligned}$$

□

Stopping criterion for ADMM

- Primal gap at k^{th} iteration can be upper bounded as:

$$f(\mathbf{x}^k) + g(\mathbf{z}^k) - p^* \leq -(\mathbf{y}^k)^T \mathbf{r}^k + (\mathbf{x}^k - \mathbf{x}^*)^T \mathbf{s}^k \tag{4}$$

  where $\mathbf{r}^k$ is the **primal residual** and $\mathbf{s}^k$ is the **residual for first dual optimal condition.**

  Proof: We work out..

- Upper bound on primal gap can be used to design stopping criterion.

PART-III: Distributed optimization using ADMM - A simple example

Distributed optimization using ADMM (1/5)

- Consider an unconstrained convex optimization problem (P1):

$$\min_{x \in \mathbb{R}} f(x)$$

- Goal is to minimize the f using multiple computing nodes in a distributed fashion.

- Say, f is separable as: $f(x) = \sum_{j=1}^{L} f_j(x)$.

- Then we can formulate an equivalent constrained optimization problem (P2):

$$\min_{x_1 \ldots x_L} \sum_{j=1}^{L} f_j(x_j)$$

$$\text{subject to } x_j = x_{j'} \quad \forall j, j^{'} \in (1, 2 \ldots L)$$

Distributed optimization using ADMM (2/5)

- Use auxilliary variables to express (P2) as a standard ADMM problem (P3):

$$\min_{x_1 \ldots x_L} \sum_{j=1}^{L} f_j(x_j)$$

$$\text{subject to } x_j = z_b \quad \forall j \in (1, 2 \ldots L),\ b \in \mathcal{B}_j$$

  where $\mathcal{B}_j$ is the set of bridge/anchor nodes connected to node j.

- Augmented Lagrangian can be split w.r.t $x_1, x_2 \ldots x_L$!

Distributed optimization using ADMM (3/5)

- A more compact representation of (P3):

$$\min_{\mathbf{x}} f_{\text{ext}}(\mathbf{x})$$
$$\text{subject to } \mathbf{E}_1\mathbf{x} + \mathbf{E}_2\mathbf{z} = 0$$

  where
  - $\mathbf{x} = (x_1, x_2 \ldots)$ and $\mathbf{z} = (z_1, z_2 \ldots)$ are concatenated vectors.
  - the rows of $\mathbf{E}_1\mathbf{x} + \mathbf{E}_2\mathbf{z} = 0$ correspond to individual constraints in (P2).

Distributed optimization using ADMM (4/5)

- Let $\{\mathbf{x}^*, \mathbf{z}^*\}$ and $\boldsymbol{\lambda}^*$ denote the unique primal and dual optimal solutions, then the following holds
  1. Sequence $\mathbf{u}^k$ is Q-linearly convergent to $\mathbf{u}^*$ i.e.,

     $$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|_{\mathbf{G}} \leq \frac{1}{1+\delta}\|\mathbf{u}^k - \mathbf{u}^*\|_{\mathbf{G}}$$

     where $\mathbf{u}$ is constructed as $\mathbf{u} = [(\mathbf{E}_2\mathbf{z})^T \ \ \boldsymbol{\lambda}^T]^T$ and δ is evaluated as

     $$\delta = \min_{\mu \geq 1, \nu \geq 1} \left(\frac{2m_f}{\frac{\nu M_f^2}{\rho(\nu-1)\sigma_{\text{min}}^2} + \mu\rho\sigma_{\text{max}}^2}, \frac{\sigma_{\text{min}}^2}{\nu\sigma_{\text{max}}^2}, \frac{\mu-1}{\mu} \right).$$

  2. The primal sequence $\mathbf{x}^k$ is R-linearly convergent to $\mathbf{x}^*$, i.e.,

     $$\|\mathbf{x}^{k+1} - \mathbf{x}^*\|_2 \leq \frac{1}{2m_f}\|\mathbf{u}^k - \mathbf{u}^*\|_{\mathbf{G}}$$

  where $\|\cdot\|_{\mathbf{G}}$ is the matrix norm with respect to the diagonal matrix $\mathbf{G} = \mathrm{diag}(\rho I_{n|\mathcal{B}|}, \rho^{-1} I_{N_C})$, m_f is the strong convexity constant of f_{ext} and M_f is the Lipschitz constant of ∇f_{ext}.

Distributed optimization using ADMM (5/5)

- To speedup convergence, ρ is chosen such that δ is maximized.
- Optimized values of ρ and corresponding δ are given by:

$$\rho_{opt} = \frac{M_f}{\sigma_{\max}\sigma_{\min}} \left(\frac{\sqrt{(\kappa-1)^2 + 4\kappa\kappa_f^2} + (\kappa-1)}{\sqrt{(\kappa-1)^2 + 4\kappa\kappa_f^2} - (\kappa-1)} \right)^{\frac{1}{2}}$$

$$\text{and } \delta_{opt} = 2\left(\kappa + 1 + \sqrt{(\kappa-1)^2 + 4\kappa\kappa_f^2}\right)^{-1}$$

- $\kappa_f = \frac{M_f}{m_f}$ denotes the condition number of the objective function f_{ext}.
- $\kappa = \frac{\sigma^2_{\max}}{\sigma^2_{\min}} = \frac{\text{max no. of bridge nodes connections per node}}{\text{min no. of bridge nodes connections per node}}$

References

1. *Alternating Direction Method of Multipliers*, Stephen Boyd, Course Slides (EE364b), Stanford University.
2. *Distributed Optimization and Statistical Learning via the Alternating Directions Method of Multipliers*, Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato and Jonathan Eckstein, Foundation and Trends in Machine Learning, Volume 3, No. 1, 2010.
3. *Convex Optimization with Sparsity-Inducing Norms* Francis Bach, et al. Book chapter-1, Optimization for Machine Learning. MIT press, 2011.
4. Lecture slides on convex optimization, Boyd Vandenberghe, Stanford University.

Backup slides

Linear convergence of a sequence

- Suppose a sequence x_k converges to L.
- x_k is said to be *Q-linearly* convergent to L, if there exists $\mu \in (0, 1)$ such that

$$\lim_{k\to\infty} \frac{|x_{k+1} - L|}{|x_k - L|} = \mu$$

- x_k is said to be *R-linearly* convergent to L, if there exists Q-linearly convergent sequence y_k which converges to zero such that

$$\lim_{k\to\infty} |x_k - L| \leq y_k$$