# ADMM

Aaron Myers, Sameer Tharakan

University of Texas at Austin

October 14, 2014

# Overview

1. ADMM Overview
   - Mulipliers and the dual problem
2. Consensus and Sharing
   - Consensus Problem
3. Distributed Model Fitting
   - Splitting Across Examples
   - Splitting Across Features

# Purpose

We can think about ADMM as a method for solving problems in which the objective function and constraints are distributed across multiple threads

ADMM can also be applied to any convex problem, where gradient descent or conjugate gradient descent requires differntiable and unconstrained conditions.

ADMM Combines two concepts:

1. Method of Multipliers
2. Dual Ascent

# Multipliers and Descent

Typical Lagrange multiplier problem where f(x) is the objective function, y is the multiplier, h(x)=0 is the constraint:

$$L(x, y, b) = f(x) + y^T h(x) \tag{1}$$

Descent Problem, with $\gamma$ as the step size

$$x_{n+1} = x_n - \gamma_n \nabla f(x_n) \tag{2}$$

Method of Multipliers

$$x^{k+1} = \quad argmin_x L_\rho(x, y^k) \tag{3}$$

$$y^{k+1} = y^k + \rho h(x) \tag{4}$$

# What's the Reasoning behind the updates?

Two variable example

$$
\begin{aligned}
Ax^* - b = 0 \quad \nabla f(x^*) + A^T y^* = 0 && (5)\\
0 = \nabla_x L_\rho(x^{k+1}, y^k) && (6)\\
= \nabla_x f(x^{k+1}) + A^T(y^k + \rho(Ax^{k+1} - b)) && (7)\\
= \nabla_x f(x^{k+1}) + A^T y^{k+1} && (8)
\end{aligned}
$$

Similar motivation for two variables z, y

# Example ADMM Framing

Original minimization problem

$$min \quad f(x) = \|Ax - b\|_2^2 \tag{9}$$

$$x \geq 0 \tag{10}$$

ADDM form:

$$min \quad \|Ax - b\|_2^2 + I_+(z) \tag{11}$$

$$x - z = 0 \tag{12}$$

ADMM Augmented Lagrangian:

$$L_\rho(x, z, y) = f(x) + g(z) + y^T(x - z) + \frac{\rho}{2}\|x - z\|_2^2 \tag{13}$$

# ADMM General Problem and Algorithm

General ADMM form

$$min \quad f(x) + g(z) \tag{14}$$
$$Ax + Bz = c \tag{15}$$

**Theorem (ADMM Algorithm)**

$x^{k+1} = argmin_x \left(L_\rho(x, z^k, y^k)\right)$
$z^{k+1} = argmin_z \left(L_\rho(x^{k+1}, z, y^k)\right)$
$y^{k+1} = y^k + \rho(Ax^{k+1} + Bz^{k+1} - c)$

We can set u$= \frac{1}{\rho}$y and simplify the algorithm

# Consensus Problem

Global Consensus Problem

$$min \quad f(x) = \sum_{i=1}^{N} f_i(x_i) \tag{16}$$

$$x_i - z = 0 \quad \forall i \tag{17}$$

We separate x into $x_i$'s (local variables) and add the constraint associated with the global variable z.

This problem shows up in networks and signal processing.

# Consensus General Form with regularization

$$min \quad \sum_{i=1}^{N} f_i(x_i) + g(z) \tag{18}$$

$$x_i - z = 0 \quad \forall i \tag{19}$$

Resulting ADMM algorithm

$$x_i^{k+1} = argmin_x \left( f_i(x_i) + y_i^{kT}(x_i - z^k) + \frac{\rho}{2}\|x_i - z^k\|_2^2 \right) \tag{20}$$

$$z^{k+1} = argmin_z \left( g(z) + \sum_{i=1}^{N} \left( -y_i^{kT} z + \frac{\rho}{2}\|x_i^{k+1} - z\|_2^2 \right) \right) \tag{21}$$

$$y_i^{k+1} = y_i^k + \rho(x_i^{k+1} - z^{k+1}) \tag{22}$$

# Sharing Problem

General Sharing problem (this has a dual realtionship with the consensus problem)

$$min \quad \sum_{i=1}^{N} f_i(x_i) + g(\sum_{i=1}^{N} x_i) \tag{23}$$

Sharing problem in ADMM form:

$$min \quad \sum_{i=1}^{N} f_i(x_i) + g(\sum_{i=1}^{N} z_i) \tag{24}$$

$$x_i - z_i = 0 \quad \forall i \tag{25}$$

# Resulting algorithm - Scaled Form

$$x_i^{k+1} = argmin_x \left( f_i(x_i) + \frac{\rho}{2}\|x_i - z_i^k + u_i^k\|_2^2 \right) \tag{26}$$

$$z^{k+1} = argmin_z \left( g(\sum_{i=1}^{N} z_i) + (\rho/2)\sum_{i=1}^{N}\left(\|z_i - x_i^{k+1} - u_i^k\|_2^2\right)\right) \tag{27}$$

$$u_i^{k+1} = u_i^k + x_i^{k+1} - z^{k+1} \tag{28}$$

# Exchange Problem

Exchange Problem:

$$min \quad \sum_{i=1}^{N} f_i(x_i) \tag{29}$$

$$\sum_{i=1}^{N} x_i = 0 \tag{30}$$

Here g would be the indicator function on the set $\{0\}$

The Exchange ADMM problem can be veiwed as a form of the tatonnement or prices adjustment process

# Algorithm for Exchange problem

$$x_i^{k+1} = argmin_x \left( f_i(x_i) + y^{kT} x_i + \frac{\rho}{2} \| x_i - (x_i^k - \bar{x}^k) \|_2^2 \right) \quad (31)$$

$$y^{k+1} = y^k + \rho \bar{x}^{k+1} \quad (32)$$

Here we can interperet y to be the price vector when applied to competitive market problems in economics. Note: we use the average $\bar{x}$ rather than $x^{k+1}$.

# General Form for model fit

General Model fit problem where l is the loss function

$$min \quad l(Ax - b) + r(x) \tag{33}$$

$$l(Ax - b) = \sum_{i=1}^{m} l_i(a_i^T x - b_i) \tag{34}$$

Here r is the regularization term and could be (among others):

ridge penalty $\lambda \|x\|_2^2$
lasso penalty $\lambda \|x\|_1$

# Model Fit examples

1) Regeression

$$b_i = a_i^T x + v_i \tag{35}$$

Where $v_i$ are the error terms

2) Classification

$$sign(p_i^T w + v) = q_i \tag{36}$$

The equation above is referred to as the discriminant function

# Splitting Across Examples

Splitting across examples is a way to handle model fit problems with a modest number of features but a very large number of training examples

$$\text{A}=[A_1 \dots A_N]^T$$
$$\text{b}=[b_1 \dots b_N]^T$$

$$min \quad l_i(A_i x - b_i) + r(z) \tag{37}$$
$$x_i - z = 0 \;\; \forall i \tag{38}$$

Applies to Lasso, Log Regression, SVM

# Scaled ADMM algorithm for example splitting

The below scaled version of ADMM is much easier to work with than the unscaled version

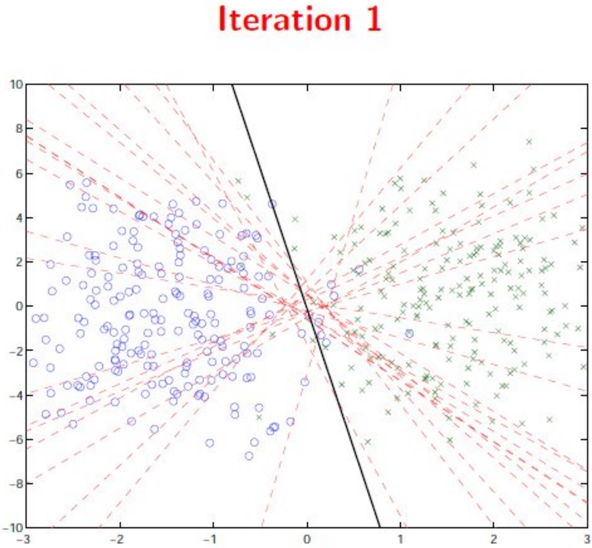
$$x_i^{k+1} = argmin_x \left( l_i(A_i x_i - b_i) + (\rho/2)\|x_i - x^k + u_i^k\|_2^2 \right) \tag{39}$$

$$z^{k+1} = argmin_x \left( r(z) + (N\rho/2)\|z - \bar{x}^{k+1} - \bar{u}^k\|_2^2 \right) \tag{40}$$

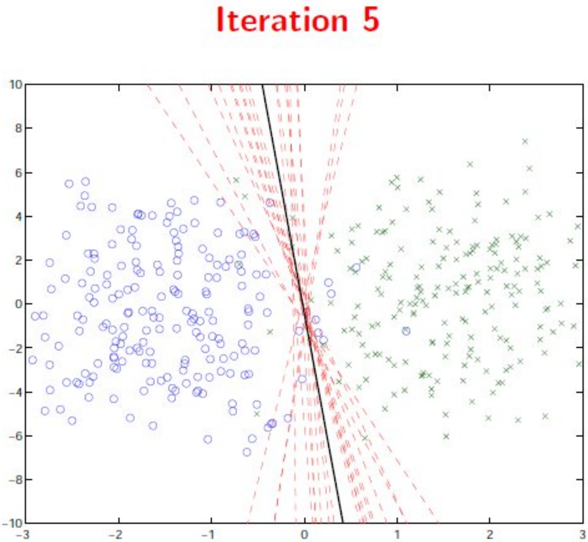
$$u_i^{k+1} = u_i^k + x_i^{k+1} - z^{k_1} \tag{41}$$

Now we will show an example with: N=400, d=2, split into 20 groups in the worst way (each group contatins only elements from one class)

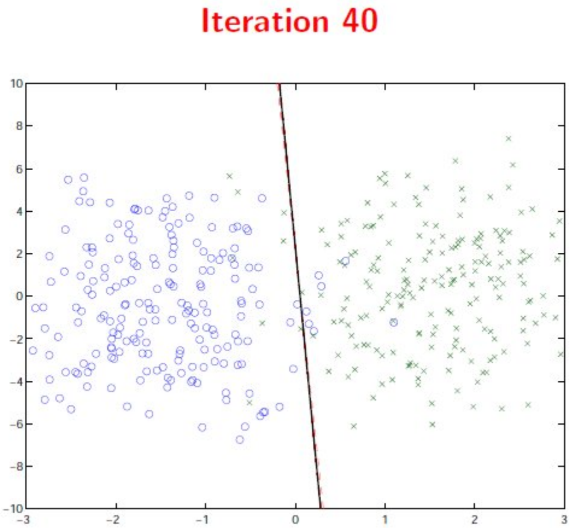
# Consensus Example



# Consensus Example



# Consensus Example



# Splting Across Features

This is for model problems with a modest number of examples but a large number of features

$$\mathrm{x} = [x_1 \cdots x_N]$$

$$min \quad l(\sum_{i=1}^{N} A_i x_i - b) + \sum_{i=1}^{N} r_i(x_i) \tag{42}$$

This method can be applied to lasso, group lasso, sparse log regression, SVM

# Sharing problem seen as splitting across features

The sharing problem from eariler can be expressed as a feature splitting problem

$$min \;\; l\left(\sum_{i=1}^{N} z_i - b\right) + \sum_{i=1}^{N} r_i(x_i) \tag{43}$$

$$A_i x_i - z_i = 0 \;\; \forall i \tag{44}$$

Below are comparison results for ADMM applied to elastic net regularization:

$$min_u \quad \lambda_1|u| + \frac{\lambda_2}{2}\|u\|^2 + \frac{1}{2}\|Mu - f\|^2 \tag{45}$$

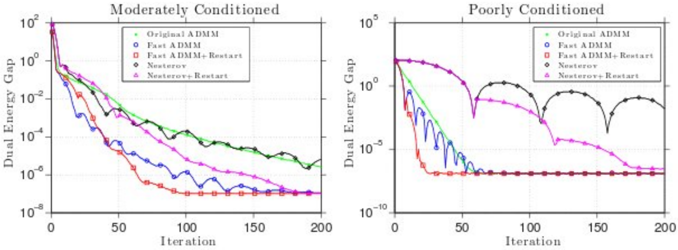
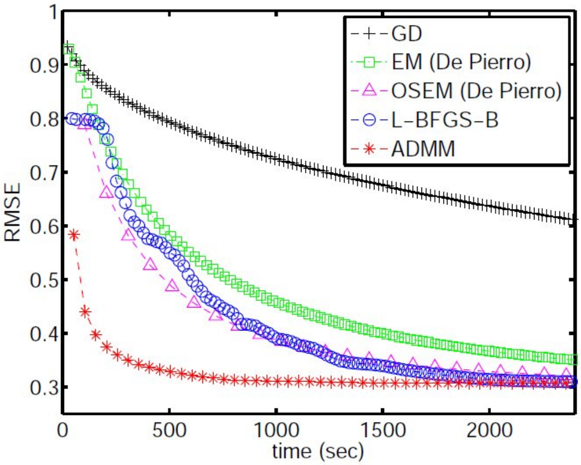



FIGURE 1. Convergence curves for the elastic net problem.

# ADMM Results

Performance Comparison with other methods



# References


- Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers Boyd, et al. (2010)
- Fast Alternating Direction Optimization Methods Goldsteing et al.
- Alternating Direction Method of Multipliers for Tomography with Nonlocal Regularizers S.T. Chun, et al (2014)

# Questions