Constrained Optimization Algorithms III

Yinyu Ye
Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

http://www.stanford.edu/~yyye

Chapters 14.5-14.7

The Lagrangian, ADMM and/or Primal-Dual Methods

$$\min \; f(\boldsymbol{x})$$
$$\text{s.t.} \; \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$$

Relax equality constraints into the objective function and update the Lagrange multipliers, together with the primal decision variables $\boldsymbol{x}$.

Alternate Direction Method of Multipliers is to decompose variables $\boldsymbol{x}$ into multi-blocks and update them alternatively.

For simplicity, we most time consider linear equality constraints

$$\min \; f(\boldsymbol{x})$$
$$\text{s.t.} \; \boldsymbol{A}(\boldsymbol{x})-\boldsymbol{b} = \boldsymbol{0}$$

Largrangian Method of Multipliers (LMM) for ECOP

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \end{aligned}$$

Lagrange Function

$L(\mathbf{x},\mathbf{y}) = f(\boldsymbol{x}) - \boldsymbol{h}(\boldsymbol{x})^T \boldsymbol{y}$

The gradient of the dual objective $\phi(\boldsymbol{y})$ is

$$\nabla \phi(\boldsymbol{y}) = -\boldsymbol{h}(\boldsymbol{x}^*)$$

where x^* is the minimizer of the Lagrange function

Lagrangian Method of Multipliers: start from an initial pair $(\boldsymbol{x}^0\ \boldsymbol{y}^0)$ and do the steepest ascent with a step-size

$$\begin{aligned} x^{k+1} &= \text{argmin} \quad L(x, y^k); \\ y^{k+1} &= y^k - \alpha^k \cdot \mathbf{h}(x^{k+1}) \end{aligned}$$

Augmented Largrangian Method (ALM) for ECOP

Lagrange Function

$L(\mathbf{x},\mathbf{y}) = f(\boldsymbol{x}) - \boldsymbol{h}(\boldsymbol{x})^T\boldsymbol{y}$

Augmented Lagrange Function

$L_A(\mathbf{x},\mathbf{y}) = f(\boldsymbol{x}) - \boldsymbol{h}(\boldsymbol{x})^T\boldsymbol{y} + 0.5\beta||\boldsymbol{h}(\boldsymbol{x})||^2$

Augmented Lagrangian Method:
The dual objective $\phi(\boldsymbol{y})$ is Lipschitz $1/\beta$ and $\nabla\phi(\boldsymbol{y}^k) = -\boldsymbol{h}(\boldsymbol{x}^{k+1})$

$$x^{k+1} = \arg\min L_A(x, y^k);$$
$$y^{k+1} = y^k - \beta \cdot \mathbf{h}(x^{k+1})$$

Analysis of the algorithm for the case that $f(\boldsymbol{x})$ is a convex function and $\boldsymbol{h}(\boldsymbol{x}) = A\boldsymbol{x} - \boldsymbol{b}$. First, if $\boldsymbol{y}^0 = \boldsymbol{y}^*$, then we must have $\boldsymbol{x}^1 = \boldsymbol{x}^*$.

KKT Condition of Lagrangian minimization

$$f(x^1) - f(x^*) \geq \nabla f(x^*)(x^1 - x^*)$$
$$f(x^*) - f(x^1) \geq \nabla f(x^1)(x^* - x^1)$$
$$0 \geq (x^1 - x^*)(\nabla f(x^*) - \nabla f(x^1))$$

$$\nabla f(x^*) = A^T y^*, Ax^* = b$$
$$\nabla f(x^1) = A^T y^* - \beta A^T(Ax^1 - b)$$
$$\nabla f(x^*) - \nabla f(x^1) = \beta A^T(Ax^1 - b)$$

$$0 \geq (x^1 - x^*)(\beta A^T(Ax^1 - b)) \geq \beta(Ax^1 - Ax^*)^T(Ax^1 - b) = \beta\|(Ax^1 - b)\|^2$$

Convergence of Augmented Lagrangian Method

Analysis continued: Suppose we are not so lucky. The KKT condition implies

$$0 = \nabla f(x^{k+1}) - A^T y^k + \beta A^T (Ax^{k+1} - b) = \nabla f(x^{k+1}) - A^T (y^k - \beta(Ax^{k+1} - b))$$

Then, from the update rule of the multipliers we have

$$0 = \nabla f(x^{k+1}) - A^T y^{k+1}$$

Thus, we only need to be concerned about where or not ($A\boldsymbol{x}^k$ - $\boldsymbol{b}$) converges to zero or not, and how fast if it does?

Theorem Consider the case that f($\boldsymbol{x}$) is a convex function and $\boldsymbol{h}(\boldsymbol{x}) = A\boldsymbol{x} - \boldsymbol{b}$. Let $\boldsymbol{y}^*$ be any optimal multiplier. Then

$$\left\| Ax^{k+1} - b \right\|^2 \le \frac{(1/\beta)}{k+1} \left\| y^0 - y^* \right\|^2$$

QP Example I

$$\begin{aligned}\min_{(x_1,x_2)} & (x_1)^2 + 2(x_2)^2 - 2x_1x_2 - .5x_1 - .5x_2 \\ \text{s.t.} \quad & x_1 + x_2 = 1,\end{aligned}$$

Augmented Lagrange Function with β=1

$$\min_{(x_1,x_2)} (x_1)^2 + 2(x_2)^2 - 2x_1x_2 - .5x_1 - .5x_2 - y_1(x_1 + x_2 - 1) + 0.5(x_1 + x_2 - 1)^2$$

KKT System of ALM for given y

$$\begin{pmatrix} 3 & -1 \\ -1 & 5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 1.5 \\ 1.5 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} y_1 = 0$$

ALM for Generic Equality Constrained QP

Augmented Lagrangian Method (ALM) for Equality QP with β=1

```
function [x,y] = ALMqp(A,b,Q,c);
[m,n]=size(A);
y=zeros(m,1);
x=randn(n,1);
for k=1:50,
  r=A'*(b+y)-c;
  x=(Q+A'*A)\r;
  y=y-(A*x-b);
end;
% Augmented Largrangian Method for solving EQP (beta=1)
%
%     minimize    0.5x'*Q*x+c'*x
%     subject to    A x = b
%
%     Output: decision vars x and multipliers y
```

Alternating Direction Method of Multipliers (ADMM) I

min $\quad f_1(\boldsymbol{x}_1)+f_2(\boldsymbol{x}_2)$

s.t. $\quad \boldsymbol{h}(\boldsymbol{x}_1,\boldsymbol{x}_2) = \boldsymbol{0}$

Augmented Lagrange Function

$$L_A(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}) = f_1(\boldsymbol{x}_1)+f_2(\boldsymbol{x}_2) - \boldsymbol{h}(\boldsymbol{x}_1,\boldsymbol{x}_2)^T\boldsymbol{y} + 0.5\beta\| \boldsymbol{h}(\boldsymbol{x}_1,\boldsymbol{x}_2)\|^2$$

ADMM:

$$x_1^{k+1} = \operatorname{argmin} L_A\,(x_1, x_2^k, y^k\,);$$
$$x_2^{k+1} = \operatorname{argmin} L_A\,(x_1^{k+1}, x_2, y^k)$$
$$y^{k+1} = y^k - \beta \cdot \mathbf{h}(x_1^{k+1}, x_2^{k+1})$$

Theorem Consider the case that both $f_1(\boldsymbol{x}_1)$ and $f_2(\boldsymbol{x}_2)$ are convex functions and $\boldsymbol{h}(\boldsymbol{x}_1,\boldsymbol{x}_2) = \mathrm{A}\boldsymbol{x}_1+\mathrm{A}_2\boldsymbol{x}_2 - \boldsymbol{b}$. Then ADMM generates a sequence of solutions that converges to an KKT point of the optimization problem.

QP Example II

$$\begin{aligned}\min_{(x_1,x_2)} \ & (x_1)^2 + 2(x_2)^2 - .5x_1 - .5x_2 \\ \text{s.t.} \ & x_1 + x_2 = 1,\end{aligned}$$

Augmented Lagrange Function with β=1

$$\min_{(x_1,x_2)} (x_1)^2 + 2(x_2)^2 - .5x_1 - .5x_2 - y_1(x_1 + x_2 - 1) + 0.5(x_1 + x_2 - 1)^2$$

KKT System of ALM for given y

$$\begin{pmatrix} 3 & 1 \\ 1 & 5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 1.5 \\ 1.5 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} y_1 = 0$$

ADMM for Linear Equality Constrained QP with Two Blocks

ADMM for Equality QP with β=1

```
function [x,y] = ADMMqp(A1,A2,b,Q1,Q2,c1,c2);
[m,n]=size(A1);
y=zeros(m,1);
x1=randn(n,1);
[m,n]=size(A2);
x2=randn(n,1);
for k=1:20,
  r=A1'*(b+y-A2*x2)-c1;
  x1=(Q1+A1'*A1)\r;
  r=A2'*(b+y-A1*x1)-c2;
  x2=(Q2+A2'*A2)\r;
  y=y-(A1*x1+A2*x2-b);
end;
x=[x1;x2];
% ADMM for solving EQP (beta=1)
%
%     minimize    0.5x1'*Q1*x1+c1'*x1+0.5x2'*Q2*x2+c2'*x2
%     subject to    A1*x1 + A2*x2  = b
```

Augmented Largrangian Method (ALM) with $\boldsymbol{x} \geq \mathbf{0}$

min $f(\boldsymbol{x})$

s.t. $A\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0}$

$X \geq 0$

Augmented Lagrange Function

$L_A(\mathbf{x},\mathbf{y}) = f(\boldsymbol{x}) - (A\boldsymbol{x} - \boldsymbol{b})^T\boldsymbol{y} + 0.5\beta\|A\boldsymbol{x}-\boldsymbol{b}\|^2,\ \boldsymbol{x} \geq \mathbf{0}$

Augmented Largrangian Method:

$$x^{k+1} = \arg\min_{x \geq 0} L_A(x, y^k);$$
$$y^{k+1} = y^k - \beta \cdot (Ax^{k+1} - b)$$

The algorithm complexity is an $O(1/\varepsilon)$ for the case that $f(\boldsymbol{x})$ is a convex function

ADMM with partial $\boldsymbol{x} \geq \mathbf{0}$

min	$f_1(\boldsymbol{x}_1)+f_2(\boldsymbol{x}_2)$
s.t.	$\boldsymbol{A}_1\boldsymbol{x}_1+\boldsymbol{A}_2\boldsymbol{x}_2 - \boldsymbol{b} = \boldsymbol{0}$
	$\boldsymbol{x}_2 \geq \boldsymbol{0}$

Augmented Lagrange Function

$$L_A(\mathbf{x}_1,\mathbf{x}_2\geq\mathbf{0},\mathbf{y}) = f_1(\boldsymbol{x}_1)+f_2(\boldsymbol{x}_2)-(A_1\boldsymbol{x}_1+A_2\boldsymbol{x}_2-\boldsymbol{b})^T\boldsymbol{y}+0.5\beta||A_1\boldsymbol{x}_1+A_2\boldsymbol{x}_2-\boldsymbol{b}||^2$$

ADMM:

$$\begin{aligned} x_1^{k+1} &= \arg\min \quad L_A(x_1, x_2^k, y^k); \\ x_2^{k+1} &= \arg\min_{x_2 \geq 0} L_A(x_1^{k+1}, x_2, y^k); \\ y^{k+1} &= y^k - \beta\cdot(A_1 x^{k+1} + A_2 x_2^{k+1} - b) \end{aligned}$$

The algorithm complexity is an $O(1/\varepsilon)$ for the case that $f_1(\boldsymbol{x}_1)+f_2(\boldsymbol{x}_2)$ are convex functions.

Application: the Dual Linear Program I

$$\begin{array}{ll} \max & b^T y \\ \text{s.t.} & A^T y + r = c, \\ & r \geq 0, \end{array}$$

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b, \\ & x \geq 0, \end{array}$$

$$\begin{array}{ll} \min & -b^T y \\ \text{s.t.} & A^T y + r = c, \\ & r \geq 0, \end{array} \quad \boxed{x}$$

$$L_A(y, r \geq 0, x)$$
$$= -b^T y - x^T (A^T y + r - c) + \frac{\beta}{2} ||A^T y + r - c||^2$$

Application: the Dual Linear Program II

$$-b^T y - x^T(A^T y + r - c) + \frac{\beta}{2}||A^T y + r - c||^2$$

The formula to update ***y*** for given ***r*** and ***x***

$$y = \frac{1}{\beta}(AA^T)^{-1}(b + A(x - \beta(r - c)))$$

The formula to update ***r*** for given ***y*** and ***x***

$$r = \max\{0, c - A^T y + \frac{1}{\beta} x\}$$

The formula to update ***x***, for given ***y*** and ***r***

$$x = x - \beta(A^T y + r - c)$$

Application: the Dual Linear Program III

```
function [x,y,r] = ADMMDualLP(A,b,c);
[m,n]=size(A);
y=zeros(m,1);
r=c;
x=ones(n,1);
AAT=inv(A*A');
beta=1;
for k=1:100,
% Update y
  bb=b+A*(x-beta*(r-c));
  y=(1/beta)*(AAT*bb);
% Update s
  cc=c-A'*y+(1/beta)*x;
  r=max(0,cc);
% Update multipliers
  x=min(0,x-beta*(A'*y+r-c));
%
end;
x=-x;
% ADMM for solving the dual linear program in standard
% form
```

Application: The Fisher Equilibrium Price Problem

$$\min \quad -5\log(2x_1+x_3) - 8\log(3x_2+x_4)$$

$$\text{s.t.} \quad x_1 + x_2 = 1,$$

$$x_3 + x_4 = 1,$$

$$\text{all } x \geq 0$$

$$\min \quad -5\log(u_1) - 8\log(u_2)$$

$$\text{s.t.} \quad x_1 + x_2 = 1$$

$$x_3 + x_4 = 1$$

$$2x_1+x_3-u_1 = 0$$

$$3x_2+x_4-u_2 = 0$$

$$x_1-s_1 = 0$$

$$x_2-s_2 = 0$$

$$x_3-s_3 = 0$$

$$x_4-s_4 = 0$$

$$s_1,s_2,s_3,s_4 \geq 0$$

Two-Block ADMM for the Fisher Price

$$
\begin{aligned}
\min\quad & -5\log(u_1) - 8\log(u_2) \\
\text{s.t.}\quad & x_1 + x_2 = 1 \\
& x_3 + x_4 = 1 \\
& 2x_1 + x_3 - u_1 = 0 \\
& 3x_2 + x_4 - u_2 = 0 \\
& x_1 - s_1 = 0 \\
& x_2 - s_2 = 0 \\
& x_3 - s_3 = 0 \\
& x_4 - s_4 = 0 \\
& s_1, s_2, s_3, s_4 \ge 0
\end{aligned}
$$

Apply the two-block ADMM where the first block, $\boldsymbol{X}_1$, including (x_1,x_2,x_3,x_4) and the second block, $\boldsymbol{X}_2$, including the rest: $(u_1,u_2, s_1,s_2,s_3,s_4)$, and they are independent in the augmented Lagrangian

The update of the first block $\boldsymbol{X}_1$, would be similar to **y**-variables in the dual LP case since they are free, the update of (s_1,s_2,s_3,s_4) in the second block $\boldsymbol{X}_2$, is identical to the ***r***-variables in the dual LP case.

The update of each (u_1,u_2) would involving a one variable minimization:

$$\min_u \ -w\log(u) + au + \beta(u-c)^2/2$$

where the minimizer is the positive root of the univariate quadratic equation

$$\beta u^2-(\beta c-a)u-w=0$$

Two-Block ADMM Solver for the Fisher Price

```
A1=[1 1 0 0;0 0 1 1;2 0 1 0;0 3 0 1;eye(4)];
A2=[zeros(2,6);-eye(6)];
A1INV=inv(A1'*A1);
b=[1;1;zeros(6,1)];
w=[5;8];
beta=1;
x=[1/2;1/2;1/2;1/2;3/2;2;1/2;1/2;1/2;1/2];
y=zeros(8,1);
for k=1:50;
 b1=b-A2*x(5:10);
 x(1:4)=A1INV*(A1'*(b1+y/beta));
 b2=A1*x(1:4)-b;
 tempb=beta*b2(3)-y(3);
 x(5)=(tempb+sqrt(tempb^2+4*beta*w(1)))/(2*beta);
 tempb=beta*b2(4)-y(4);
 x(6)=(tempb+sqrt(tempb^2+4*beta*w(2)))/(2*beta);
 x(7:10)=max(0,b2(5:8)-y(5:8)/beta);
 y=y-beta*(A1*x(1:4)+A2*x(5:10)-b);
end;
```

Alternating Direction Method of Multipliers (ADMM)

min $f_1(\boldsymbol{x}_1)+f_2(\boldsymbol{x}_2)+f_3(\boldsymbol{x}_3)$

s.t. $\boldsymbol{h}(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3)=\boldsymbol{0}$

Augmented Lagrange Function

$$L_A(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{y}) = f_1(\boldsymbol{x}_1)+f_2(\boldsymbol{x}_2)+f_3(\boldsymbol{x}_3) - \boldsymbol{h}(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3)^T\boldsymbol{y} + 0.5\beta||\boldsymbol{h}(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3)||^2$$

ADMM(?):
Works if the $\mathbf{x}_2$ and $\mathbf{x}_3$ blocks variables are independent to each other.

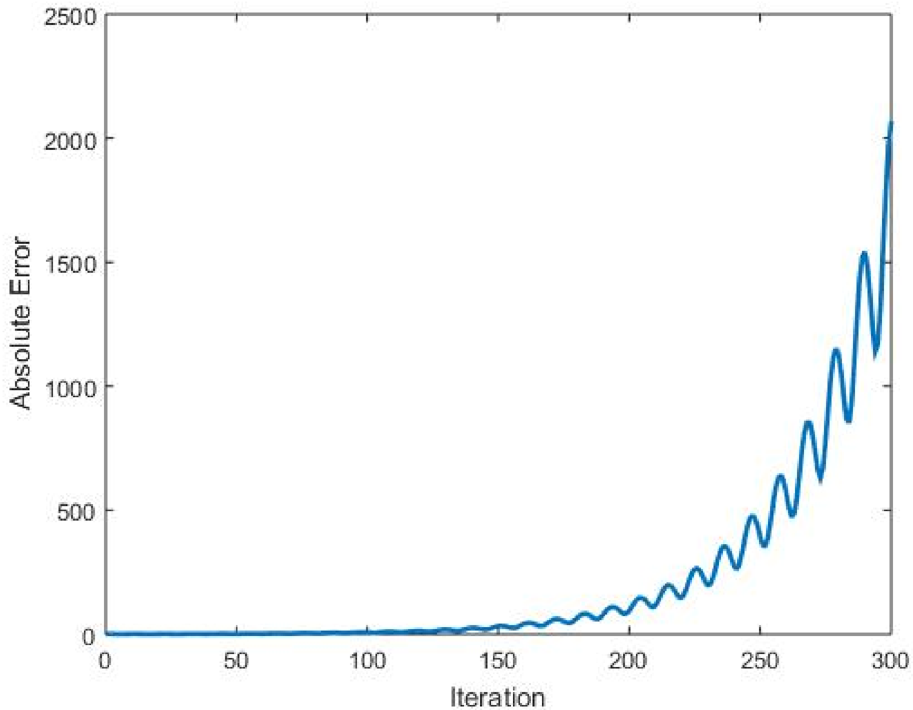
$$x_1^{k+1} = \arg\min L_A(x_1, x_2^k, x_3^k y^k);$$
$$x_2^{k+1} = \arg\min L_A(x_1^{k+1}, x_2, x_3^k, y^k)$$
$$x_3^{k+1} = \arg\min L_A(x_1^{k+1}, x_2^{k+1}, x_3, y^k)$$
$$y^{k+1} = y^k - \beta\cdot\mathbf{h}(x_1^{k+1}, x_2^{k+1}, x_3^{k+1})$$

Theorem There is an example where $f_i(\boldsymbol{x}_i)$'s are linear functions and $\boldsymbol{h}(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3) = A\boldsymbol{x}_1+A_2\boldsymbol{x}_2+A_3\boldsymbol{x}_3-\boldsymbol{b}$, for which the ADMM generates a sequence of solutions that diverges with probability one when starting from a random initial point. However, when the update order of $\boldsymbol{x}$'s is randomly permuted, then the sequence converges in expectation.

Diverging Example of ADMM with Three Blocks

$$\min_{(x_1,x_2,x_3)} \quad 0\cdot x_1 + 0\cdot x_2 + 0\cdot x_3$$

$$\text{s.t.} \qquad \begin{pmatrix}1\\1\\1\end{pmatrix} x_1 + \begin{pmatrix}1\\1\\2\end{pmatrix} x_2 + \begin{pmatrix}1\\2\\2\end{pmatrix} x_3 = \begin{pmatrix}0\\0\\0\end{pmatrix},$$



$$\begin{pmatrix}x^{k+1}\\y^{k+1}\end{pmatrix} = M\begin{pmatrix}x^{k}\\y^{k}\end{pmatrix}$$

$$\rho(M) > 1$$

Randomly Permuted ADMM

Why have to follow the updating order $1 \to 2 \to 3$ for **x**'s?

RP-ADMM: In each cycle of ADMM, randomly select an index, say б(1) from {1,2,3}, and let $\mathbf{x}_{б(1)}$ be the first block to update; then randomly select an index б(2) from the remaining two indexes and let $\mathbf{x}_{б(2)}$ be the second block to update; then denote the remaining index as б(3) update $\mathbf{x}_{б(3)}$; finally update the multiplier **y** as in the regular ADMM

Theorem The RP-ADMM generate a random sequence of $\{\mathbf{x}^k,\mathbf{y}^k\}$ that converges, in expectation, to the optimal solution of the equality constrained QP optimization problem for any number of blocks.

RP-ADMM with Three Blocks

$$\min_{(x_1,x_2,x_3)} \quad 0\cdot x_1 + 0\cdot x_2 + 0\cdot x_3$$

$$\text{s.t.} \quad \begin{pmatrix}1\\1\\1\end{pmatrix} x_1 + \begin{pmatrix}1\\1\\2\end{pmatrix} x_2 + \begin{pmatrix}1\\2\\2\end{pmatrix} x_3 = \begin{pmatrix}0\\0\\0\end{pmatrix},$$