Some recent advances in the linearized ALM, ADMM and Beyond

Relax the crucial parameter requirements

Bingsheng HE

Department of Mathematics, SUSTech, and
Department of Mathematics, Nanjing University

`Homepage: maths.nju.edu.cn/~hebma`

Based on some papers co-authored by F. Ma and X.M. Yuan

1 Introduction

Some linearly constrained convex optimization problems

1. Linearly constrained convex optimization $\min\{\theta(x)|Ax = b,\ x \in \mathcal{X}\}$

2. Convex optimization problem with separable objective function

$$\min\{\theta_1(x) + \theta_2(y)|Ax + By = b,\ x \in \mathcal{X},\ y \in \mathcal{Y}\}$$

3. Convex optimization problem with 3 separable objective functions

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z)|Ax+By+Cz = b, x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\}$$

There are some crucial parameters：

- Crucial parameter in the so called linearized ALM for the first problem,
- Crucial parameter in the so called linearized ADMM for the second problem,
- Crucial proximal parameter in the Proximal Parallel ADMM-like Method for the convex optimization problem with 3 separable objective functions.

2 Linearized Augmented Lagrangian Method

Consider the following convex optimization problem:

$$\min\{\theta(x) \mid Ax = b,\, x \in \mathcal{X}\}. \tag{2.1}$$

The augmented Lagrangian function of the problem (2.1) is

$$\mathcal{L}_\beta(x,\lambda) = \theta(x) - \lambda^T(Ax-b) + \frac{\beta}{2}\|Ax-b\|^2.$$

Starting with a given λ^k, the k-th iteration of the Augmented Lagrangian Method [11, 12] produces the new iterate $w^{k+1} = (x^{k+1}, \lambda^{k+1})$ via

$$\text{(ALM)} \quad \left\{ \begin{array}{ll} x^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x,\lambda^k) \;\big|\; x \in \mathcal{X}\big\}, & \text{(2.2a)} \\ \lambda^{k+1} = \lambda^k - \gamma\beta(Ax^{k+1} - b), \quad \gamma \in (0,2) & \text{(2.2b)} \end{array} \right.$$

In the classical ALM, the optimization subproblem (2.2a) is

$$\min\{\theta(x) + \frac{\beta}{2}\|Ax - (b + \frac{1}{\beta}\lambda^k)\|^2 | x \in \mathcal{X}\}.$$

Sometimes, because of the structure of the matrix A, we should simplify the subproblem (2.2a). Notice that

- Ignore the constant term in the objective function of $\mathcal{L}_\beta(x,\lambda^k)$, we have

$$
\begin{aligned}
&\arg\min\{\mathcal{L}_\beta(x,\lambda^k) \mid x\in\mathcal{X}\}\\
&\quad = \quad \arg\min\Big\{\theta(x)-(\lambda^k)^T(Ax-b)+\tfrac{\beta}{2}\|Ax-b\|^2 \Big| \; x\in\mathcal{X}\Big\}\\
&\quad = \quad \arg\min\left\{\begin{array}{c}\theta(x)-(\lambda^k)^T(Ax-b)+\\ \frac{\beta}{2}\|(Ax^k-b)+A(x-x^k)\|^2\end{array}\middle|\; x\in\mathcal{X}\right\}\\
&\quad = \quad \arg\min\left\{\begin{array}{c}\theta(x)-x^TA^T[\lambda^k-\beta(Ax^k-b)]\\ +\frac{\beta}{2}\|A(x-x^k)\|^2\end{array}\middle|\; x\in\mathcal{X}\right\}. \qquad (2.3)
\end{aligned}
$$

- In the so called **Linearized ALM** [14], the term $\frac{\beta}{2}\|A(x-x^k)\|^2$ is replaced with $\frac{r}{2}\|x-x^k\|^2$. In this way, the x-subproblem becomes

$$
x^{k+1}=\arg\min\Big\{\theta(x)-x^TA^T[\lambda^k-\beta(Ax^k-b)]+\frac{r}{2}\|x-x^k\|^2\Big|x\in\mathcal{X}\Big\}. \qquad (2.4)
$$

In fact, the linearized ALM simplifies the quadratic term $\frac{\beta}{2}\|A(x-x^k)\|^2$.

In comparison with (2.3), the simplified x-subproblem (2.4) is equivalent to

$$x^{k+1} = \arg\min\{\mathcal{L}_\beta(x,\lambda^k) + \tfrac{1}{2}\|x - x^k\|_{D_A}^2 \mid x \in \mathcal{X}\}, \tag{2.5}$$

where

$$D_A = rI - \beta A^T A. \tag{2.6}$$

In order to ensure the convergence, it **was** required that $\boldsymbol{r > \beta\|A^T A\|}$.

Thus, the mathematical form of the Linearized ALM can be written as

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x,\lambda^k) + \tfrac{1}{2}\|x - x^k\|_{D_A}^2 \;\big|\; x \in \mathcal{X}\}, & \text{(2.7a)} \\ \lambda^{k+1} = \lambda^k - \gamma\beta(Ax^{k+1} - b), \qquad \gamma \in (0,2). & \text{(2.7b)} \end{cases}$$

where D_A is defined by (2.6).

Large parameter r in (2.6) will lead a slow convergence !

Recent Advance. Bingsheng He, Feng Ma, Xiaoming Yuan:
Optimal proximal augmented Lagrangian method and its application to full Jacobian splitting for multi-block separable convex minimization problems, IMA Journal of Numerical Analysis. 39(2019).

Our new result in the above paper:

For the matrix D_A in (2.7a) with the form (2.6)

- if $r > \frac{2+\gamma}{4}\beta\|A^TA\|$ is taken in the method (2.7), it is still convergent;
- if $r < \frac{2+\gamma}{4}\beta\|A^TA\|$ is taken in the method (2.7), there is divergent example.

Especially, when $\gamma = 1$,

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x,\lambda^k) + \frac{1}{2}\|x - x^k\|_{D_A}^2 \mid x \in \mathcal{X}\}, & \text{(2.8a)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - b). & \text{(2.8b)} \end{cases}$$

According to our new result: For the matrix D_A in in (2.7a) with the form (2.6),

- if $r > \frac{3}{4}\beta\|A^TA\|$ is taken in the method (2.8), it is still convergent;
- if $r < \frac{3}{4}\beta\|A^TA\|$ is taken in the method (2.8), there is divergent example.

$r = 0.75$ **is the threshold factor in the matrix** D_A **for linearized ALM** (2.8) **!**

3 Linearized ADMM

Consider the convex optimization problem with separable objective function:

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\ x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{3.1}$$

The augmented Lagrangian function of the problem (3.1) is

$$\mathcal{L}_\beta^2(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b) + \frac{\beta}{2}\|Ax + By - b\|^2.$$

Starting with a given (y^k, λ^k), the k-th iteration of the classical ADMM [4, 5] generates the new iterate $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ via

$$\text{(ADMM)} \quad \left\{ \begin{array}{ll} x^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x, y^k, \lambda^k) \bigm| x \in \mathcal{X}\big\}, & \text{(3.2a)} \\ y^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \bigm| y \in \mathcal{Y}\big\}, & \text{(3.2b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(3.2c)} \end{array} \right.$$

In (3.2a) and (3.2a), the optimization subproblems are

$$\min\{\theta_1(x) + \frac{\beta}{2}\|Ax - p^k\|^2 | x \in \mathcal{X}\} \quad \text{and} \quad \min\{\theta_2(y) + \frac{\beta}{2}\|By - q^k\|^2 | y \in \mathcal{Y}\},$$

respectively. We assume that one of the minimization subproblems (without loss of the generality, say, (3.2b)) should be simplified. Notice that

- Using the notation $\mathcal{L}_\beta(x^{k+1}, y, \lambda^k)$ and ignoring the constant term in the objective function, we have

$$
\begin{aligned}
&\arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \mid y \in \mathcal{Y}\} \\
&= \arg\min\left\{ \begin{array}{c} \theta_2(y) - (\lambda^k)^T(Ax^{k+1} + By - b) \\ +\frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\} \\
&= \arg\min\left\{ \begin{array}{c} \theta_2(y) - (\lambda^k)^T By + \\ \frac{\beta}{2}\|(Ax^{k+1} + By^k - b) + B(y - y^k)\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\} \\
&= \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T[\lambda^k - \beta(Ax^{k+1} + By^k - b)] \\ +\frac{\beta}{2}\|B(y - y^k)\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\}. \quad (3.3)
\end{aligned}
$$

- In the so called **Linearized ADMM** [13, 14, 15], the term $\frac{\beta}{2}\|B(y - y^k)\|^2$ is replaced with $\frac{s}{2}\|y - y^k\|^2$. Thus, the y-subproblem becomes

$$y^{k+1} = \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T[\lambda^k - \beta(Ax^{k+1} + By^k - b)] \\ + \frac{s}{2}\|y - y^k\|^2 \end{array} \;\middle|\; y \in \mathcal{Y} \right\}. \tag{3.4}$$

In fact, the linearized ADMM simplifies the quadratic term $\frac{\beta}{2}\|B(y - y^k)\|^2$.

In comparison with (3.3), the simplified y-subproblem (3.4) is equivalent to

$$y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) + \tfrac{1}{2}\|y - y^k\|^2_{D_B} \mid y \in \mathcal{Y}\}, \tag{3.5}$$

where

$$D_B = sI - \beta B^T B. \tag{3.6}$$

In order to ensure the convergence, it **was** required that $\boldsymbol{s > \beta\|B^T B\|}$.

Thus, the mathematical form of the Linearized ADMM can be written as

$$\left\{ \begin{array}{ll} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, & \text{(3.7a)} \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) + \tfrac{1}{2}\|y - y^k\|^2_{D_B} \mid y \in \mathcal{Y}\}, & \text{(3.7b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b), & \text{(3.7c)} \end{array} \right.$$

where D_B is defined by (3.6).

A large parameter s will lead a slow convergence of the linearized ADMM.

最新进展：最优线性化因子的选择– OO6228 的结论

Recent Advance. Bingsheng He, Feng Ma, Xiaoming Yuan:
Optimal Linearized Alternating Direction Method of Multipliers for Convex Programming. http://www.optimization-online.org/DB_HTML/2017/09/6228.html

Our new result in the above paper: For the matrix D_B in (3.7b) with the form (3.6)

- if $s > \frac{3}{4}\beta\|B^TB\|$ is taken in the method (3.7), it is still convergent;
- if $s < \frac{3}{4}\beta\|B^TB\|$ is taken in the method (3.7), there is divergent example.

$s = 0.75$ is the threshold factor in the matrix D_B for linearized ADMM (3.7) **!**

Notice that the matrix D_B defined in (3.6) is indefinite for $s \in (0.75, 1)$!

4 Parameters improvements in the method for problem with 3 separable objective functions

For the problem with three separable objective functions

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z) | Ax+By+Cz=b,\ x\in\mathcal{X}, y\in\mathcal{Y}, z\in\mathcal{Z}\}, \quad (4.1)$$

the augmented Lagrangian function is

$$\begin{aligned}\mathcal{L}^3_\beta(x,y,z,\lambda) \quad &= \quad \theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^T(Ax+By+Cz-b) \\ &\quad\quad +\frac{\beta}{2}\|Ax+By+Cz-b\|^2.\end{aligned}$$

Using the **direct extension of ADMM** to solve the problem (4.1), the formula is

$$\left\{\begin{array}{l} x^{k+1} = \text{Argmin}\{\mathcal{L}^3_\beta(x,y^k,z^k,\lambda^k)\,|\,x\in\mathcal{X}\}, \\ y^{k+1} = \text{Argmin}\{\mathcal{L}^3_\beta(x^{k+1},y,z^k,\lambda^k)\,|\,y\in\mathcal{Y}\}, \\ z^{k+1} = \text{Argmin}\{\mathcal{L}^3_\beta(x^{k+1},y^{k+1},z,\lambda^k)\,|\,z\in\mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k-\beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b). \end{array}\right. \quad (4.2)$$

Unfortunately, the direct extension (4.2) **is not necessarily convergent [2] !**

ADMM + Parallel Splitting ALM

[强制 y, z 平等]

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta^3(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta^3(x^{k+1}, y, z^k, \lambda^k) \mid y \in \mathcal{Y}\}, \\ z^{k+1} = \arg\min\{\mathcal{L}_\beta^3(x^{k+1}, y^k, z, \lambda^k) \mid z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \end{cases}$$

平行处理 y, z 子问题, 各自为政, 不能保证方法收敛!

ADMM + Parallel-Prox Splitting ALM

各自为政, 过分自由. 给它们加个适当的正则项($\tau > 1$), 方法就能保证收敛.

$$x^{k+1} = \arg\min\{\mathcal{L}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \tag{4.3a}$$

$$\begin{cases} y^{k+1} = \arg\min\{\mathcal{L}(x^{k+1}, y, z^k, \lambda^k) + \frac{\tau}{2}\|B(y - y^k)\|^2 \mid y \in \mathcal{Y}\}, \\ z^{k+1} = \arg\min\{\mathcal{L}(x^{k+1}, y^k, z, \lambda^k) + \frac{\tau}{2}\|C(z - z^k)\|^2 \mid z \in \mathcal{Z}\}, \end{cases} \tag{4.3b}$$

$$\lambda^{k+1} = \lambda^k - (Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \tag{4.3c}$$

Notice that (4.3b) can be written as

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \end{pmatrix} = \arg\min\left\{ \mathcal{L}(x^{k+1}, y, z, \lambda^k) + \frac{1}{2}\left\| \begin{array}{c} y - y^k \\ z - z^k \end{array} \right\|_{D_{BC}}^2 \,\middle|\, \begin{array}{l} y \in \mathcal{Y} \\ z \in \mathcal{Z} \end{array} \right\},$$

where

$$D_{BC} = \begin{pmatrix} \tau B^T B & -B^T C \\ -C^T B & \tau C^T C \end{pmatrix}. \tag{4.4}$$

D_{BC} is positive semidefinite when $\tau \geq 1$.

However, the matrix D_{BC} is indefinite for $\tau \in (0, 1)$.

In other words, the scheme (4.3) can be rewritten as

$$\left\{ \begin{array}{rcl} x^{k+1} & = & \arg\min\{\mathcal{L}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ \begin{pmatrix} y^{k+1} \\ z^{k+1} \end{pmatrix} & = & \arg\min\left\{ \mathcal{L}(x^{k+1}, y, z, \lambda^k) + \frac{1}{2}\left\| \begin{array}{c} y - y^k \\ z - z^k \end{array} \right\|_{D_{BC}}^2 \,\middle|\, \begin{array}{l} y \in \mathcal{Y} \\ z \in \mathcal{Z} \end{array} \right\}, \\ \lambda^{k+1} & = & \lambda^k - (Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{array} \right.$$

The algorithm (4.3) can be rewritten in an equivalent form: $(\mu = \tau + 1 > 2)$.

$$\left\{\begin{array}{l} x^{k+1} = \arg\min\{\theta_1(x) + \frac{\beta}{2}\|Ax + By^k + Cz^k - b - \frac{1}{\beta}\lambda^k\|^2 \,|\, x \in \mathcal{X}\}, \\ \qquad \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b) \\ y^{k+1} = \arg\min\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T By + \frac{\mu\beta}{2}\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\}, \\ z^{k+1} = \arg\min\{\theta_3(z) - (\lambda^{k+\frac{1}{2}})^T Cz + \frac{\mu\beta}{2}\|C(z - z^k)\|^2 \,|\, z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{array}\right. \tag{4.5}$$

The related publication:

- B. He, M. Tao and X. Yuan, A splitting method for separable convex programming. IMA J. Numerical Analysis, 31(2015), 394-426.

In the above paper, in order to ensure the convergence, it **was** required

$\tau > 1$ (in (4.3)) which is equivalent to $\mu > 2$ (in (4.5)).

This method is accepted by Osher's research group

- E. Esser, M. Möller, S. Osher, G. Sapiro and J. Xin, A convex model for non-negative matrix factorization and dimensionality reduction on physical space, IEEE Trans. Imag. Process., 21(7), 3239-3252, 2012.

> tion refinement step. Due to the different algorithm used to solve the extended model, there is an additional numerical parameter μ, which for this application must be greater than two according to [34]. We set μ equal to 2.01. There are also model parame-

Thus, Osher's research group utilize the iterative formula (4.5), according to our previous paper, they set

$$\mu = 2.01, \qquad \text{it is only a pity larger than 2.}$$

Large parameter μ (or τ) will lead a slow convergence.

最新进展：最优正则化因子的选择– OO6235 的结论

Recent Advance in : Bingsheng He, Xiaoming Yuan: On the Optimal Proximal Parameter of an ADMM-like Splitting Method for Separable Convex Programming http://www.optimization-online.org/DB_HTML/2017/ 10/6235.html

Our new assertion: In (4.3)

- if $\tau > 0.5$, the method is still convergent;
- if $\tau < 0.5$, there is divergent example.

Equivalently in (4.5) :

- if $\mu > 1.5$, the method is still convergent;
- if $\mu < 1.5$, there is divergent example.

For convex optimization problem (4.1) with three separable objective functions, the parameters in the equivalent methods (4.3) and (4.5) :

- **0.5** is the threshold factor of the parameter τ in (4.3) !
- **1.5** is the threshold factor of the parameter μ in (4.5) !

References

[1] 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 Vol. 3, No. 1 (2010) 1 – 122.

[2] C.H. Chen, B.S. He, Y.Y. Ye and X. M. Yuan, The direct extension of ADMM for multi-block convex minimization problems is not necessary convergent, Mathematical Programming, 155 (2016) 57-79.

[3] E. Esser, M. Möller, S. Osher, G. Sapiro and J. Xin, A convex model for non-negative matrix factorization and dimensionality reduction on physical space, IEEE Trans. Imag. Process., 21(7), 3239-3252, 2012.

[4] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite-element approximations*, Comput. Math. Appli., 2 (1976), pp. 17-40.

[5] R. Glowinski, Numerical methods for Nonlinear Variational Problems, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.

[6] G.Y. Gu, B.S He and X. M. Yuan, Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a unified approach, Comput. Optim. Appl., 59, pp. 135-161, 2014.

[7] B. S. He, PPA-like contraction methods for convex optimization: a framework using

variational inequality approach. J. Oper. Res. Soc. China 3 (2015) 391 – 420.

[8] B.S. He, F. Ma and X.M. Yuan, Optimal proximal augmented Lagrangian method and its application to full Jacobian splitting for multi-block separable convex minimization problems, IMA Journal of Numerical Analysis. 39(2019)

[9] B.S. He, M. Tao and X.M. Yuan, A splitting method for separable convex programming, IMA J. Numerical Analysis 31 (2015), 394-426.

[10] B.S. He, and X. M. Yuan, A class of ADMM-based algorithms for three-block separable convex programming. Comput. Optim. Appl. 70 (2018), 791 – 826.

[11] M. R. Hestenes, Multiplier and gradient methods, JOTA 4, 303-320, 1969.

[12] M. J. D. Powell, A method for nonlinear constraints in minimization problems, in Optimization, R. Fletcher, ed., Academic Press, New York, NY, pp. 283-298, 1969.

[13] X. F. Wang and X. M. Yuan, *The linearized alternating direction method of multipliers for Dantzig selector*, SIAM J. Sci. Comput., 34 (2012), pp. A2792–A2811.

[14] J. F. Yang and X. M. Yuan, *Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization*, Math. Comp., 82 (2013), pp. 301-329.

[15] X. Q. Zhang, M. Burger, and S. Osher, *A unified primal-dual algorithm framework based on Bregman iteration*, J. Sci. Comput., 46 (2010), pp. 20-46.

[16] 何炳生, 我和乘子交替方向法 20 年, 《运筹学学报》22 卷第1期, pp. 1-31, 2018.