Convergence analysis of proximal symmetric alternating direction method of multipliers

Yu Chen

Nanjing University

September 24, 2017

Outline

1. Introduction
2. Main result
3. Convergence analysis
4. Numerical result
5. Conclusion

1 Introduction

2 Main result

3 Convergence analysis

4 Numerical result

5 Conclusion

Introduction

Consider the convex optimization problem:

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

- $\theta_1 : \Re^{n_1} \to \Re$ and $\theta_2 : \Re^{n_2} \to \Re$ are continuous convex function
- $A \in \Re^{m \times n_1}, B \in \Re^{m \times n_2}, b \in \Re^m$
- $\mathcal{X} \subset \Re^{n_1}, \mathcal{Y} \subset \Re^{n_2}$ are nonempty closed convex sets

The augmented Lagrangian function:

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

Example

Basis pursuit:

$$\min_{x \in \Re^n} \|x\|_1 + \frac{1}{2\mu}\|Ax - b\|_2^2$$

Example

Basis pursuit:

$$\min_{x \in \Re^n} \|x\|_1 + \frac{1}{2\mu}\|Ax - b\|_2^2$$

$$\downarrow$$

$$\min \left\{ \frac{1}{2\mu}\|Ax - b\|_2^2 + \|y\|_1 | x - y = 0, x, y \in \Re^n \right\}$$

ADMM

Glowinski and Marroco(1975), Gabay and Mercier(1976)

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

ADMM

Glowinski and Marroco(1975), Gabay and Mercier(1976)

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

Glowinski(1984)

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

ADMM

Glowinski and Marroco(1975), Gabay and Mercier(1976)

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

Glowinski(1984)

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

- $s \in (0, \frac{1+\sqrt{5}}{2})$

ADMM

Glowinski and Marroco(1975), Gabay and Mercier(1976)

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

Glowinski(1984)

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

- $s \in (0, \frac{1+\sqrt{5}}{2})$
- This larger step size is crucial to getting better numerical results.

Symmetric ADMM

Peaceman-Rachford(1955), Lions and Mercier(1979)

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \mid y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b). \end{cases}$$

- It may not converge.
- He, Liu, Wang, Yuan (2014) introduce a parameter $\alpha \in (0, 1)$ to shrink the step size in updating the Lagrange multiplier steps.

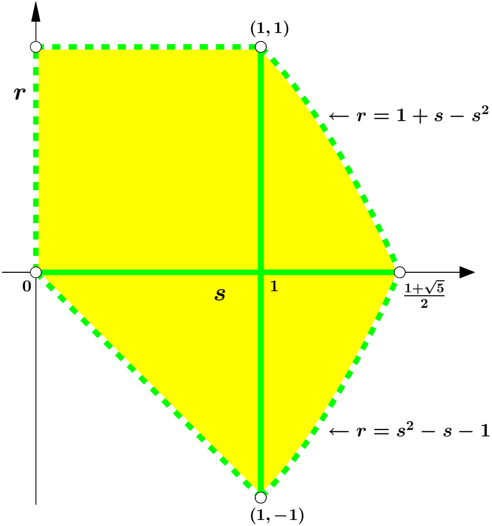
He, Ma, Yuan (2015)

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \mid y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(Ax^{k+1} + By^{k+1} - b), \end{cases}$$

The parameters r and s are restricted into the domain

$$\mathcal{D} = \{(r, s) \mid s \in \left(0, \tfrac{1+\sqrt{5}}{2}\right), r \in (-1, 1),\ r + s > 0,\ |r| < 1 + s - s^2\}.$$

Domain $\mathcal{D}$



Main work

Proximal Symmetric ADMM

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) + \frac{1}{2}\|x - x^k\|_R^2 \mid x \in \mathcal{X}\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) + \frac{1}{2}\|y - y^k\|_T^2 \mid y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(Ax^{k+1} + By^{k+1} - b), \end{cases}$$

- R and T are positive semidefinite matrices
- The parameters r and s are restricted into the domain

$$\mathcal{D} = \{(r, s) \mid s \in \left(0, \frac{1+\sqrt{5}}{2}\right), r \in (-1, 1),\ r + s > 0,\ |r| < 1 + s - s^2\}.$$

1 Introduction

2 Main result

3 Convergence analysis

4 Numerical result

5 Conclusion

Optimality condition

Proposition 3.1

Let $\mathcal{X} \subset \Re^n$ *be a closed convex set,* $\theta(x) : \Re^n \to \Re$ *and* $f(x) : \Re^n \to \Re$ *be convex functions. In addition,* $f(x)$ *is differentiable. We assume that the solution set of the minimization problem* $\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\}$ *is nonempty. Then,*

$$x^* = \arg\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\},$$

if and only if

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geqslant 0, \quad \forall\, x \in \mathcal{X}.$$

Denote:

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \; u = \begin{pmatrix} x \\ y \end{pmatrix}, \; F(w) := \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix},$$

$$\theta(u) = \theta_1(x) + \theta_2(y), \; \tilde{x}^k = x^{k+1}, \; \tilde{y}^k = y^{k+1},$$

$$\tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b).$$

prediction step

Lemma 1

Let w^{k+1} be generated by the proximal symmetric ADMM. Then, we have

$$\tilde{w}^k \in \Omega, \ \ \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \geqslant (w-\tilde{w}^k)^T Q(w^k-\tilde{w}^k), \ \ \forall w \in \Omega,$$

where

$$Q = \begin{pmatrix} R & 0 & 0 \\ 0 & \beta B^T B + T & -rB^T \\ 0 & -B & \frac{1}{\beta} I_m \end{pmatrix}.$$

correction step

Lemma 2

Let w^{k+1} be generated by the proximal symmetric ADMM. Then, we have

$$w^{k+1} = w^k - M(w^k - \tilde{w}^k),$$

where

$$M = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & -s\beta B & (r+s)I_m \end{pmatrix}.$$

correction step

Lemma 2

Let w^{k+1} be generated by the proximal symmetric ADMM. Then, we have

$$w^{k+1} = w^k - M(w^k - \tilde{w}^k),$$

where

$$M = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & -s\beta B & (r+s)I_m \end{pmatrix}.$$

$$(w - \tilde{w}^k)^T Q(w^k - \tilde{w}^k) = (w - \tilde{w}^k)^T H(w^k - w^{k+1})$$

correction step

Lemma 2

Let w^{k+1} be generated by the proximal symmetric ADMM. Then, we have

$$w^{k+1} = w^k - M(w^k - \tilde{w}^k),$$

where

$$M = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & -s\beta B & (r+s)I_m \end{pmatrix}.$$

$$(w - \tilde{w}^k)^T Q(w^k - \tilde{w}^k) = (w - \tilde{w}^k)^T H(w^k - w^{k+1})$$

where $H = QM^{-1}$

Main result

Theorem 3

For the sequence $\{w^k\}$ *generated by the proximal symmetric ADMM, we have*

$$\begin{aligned}&\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) \\ \geqslant\ & \frac{1}{2}\left(\|w - w^{k+1}\|_H^2 - \|w - w^k\|_H^2\right) + \frac{1}{2}\|w^k - \tilde{w}^k\|_G^2, \ \ \forall w \in \Omega,\end{aligned}$$

where $H = \begin{pmatrix} R & 0 & 0 \\ 0 & (1 - \frac{rs}{r+s})\beta B^T B + T & -\frac{r}{r+s} B^T \\ 0 & -\frac{r}{r+s} B & \frac{1}{(r+s)\beta} I_m \end{pmatrix}$ *and*

$$G = \begin{pmatrix} R & 0 & 0 \\ 0 & (1-s)\beta B^T B + T & -(1-s)B^T \\ 0 & -(1-s)B & \frac{1}{\beta}(2 - (r+s)) I_m \end{pmatrix}$$

Using optimality condition, we can obtain:

$$\|w^{k+1} - w^*\|_H^2 \leqslant \|w^k - w^*\|_H^2 - \|w^k - \tilde{w}^k\|_G^2, \quad \forall w^* \in \Omega^*$$

- H is a positive semidefinite matrix for $(s, r) \in \mathcal{D}$
- G is a positive semidefinite matrix if $s < 1$
- G may not be a positive semidefinite matrix if $s > 1$

1 Introduction

2 Main result

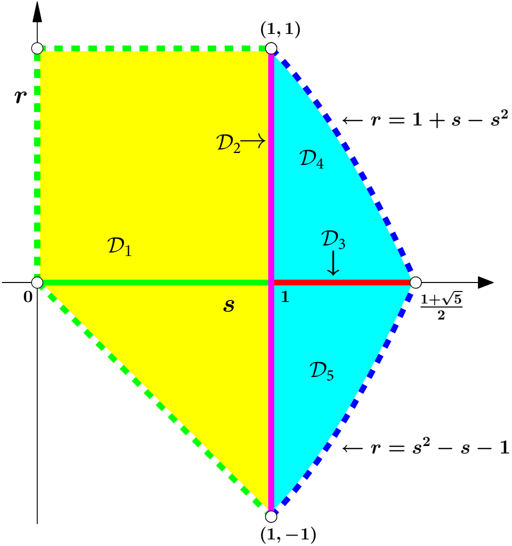
3 Convergence analysis

4 Numerical result

5 Conclusion

Convergence analysis

$$
\begin{cases}
\mathcal{D}_1 & = & \big\{(s,r)\,|\,s\in(0,1),\ r\in(-1,1),\ r+s>0\big\},\\
\mathcal{D}_2 & = & \big\{(s,r)\,|\,s=1,\ r\in(-1,1)\big\},\\
\mathcal{D}_3 & = & \big\{(s,r)\,|\,s\in\big(1,\frac{1+\sqrt{5}}{2}\big),\ r=0\big\},\\
\mathcal{D}_4 & = & \big\{(s,r)\,|\,s\in\big(1,\frac{1+\sqrt{5}}{2}\big),\ r\in(0,1),\ r<1+s-s^2\big\},\\
\mathcal{D}_5 & = & \big\{(s,r)\,|\,s\in\big(1,\frac{1+\sqrt{5}}{2}\big),\ r\in(-1,0),\ -r<1+s-s^2\big\}.
\end{cases}
$$



(1, 1)
r
$\leftarrow r = 1 + s - s^2$
$\mathcal{D}_2 \rightarrow$
$\mathcal{D}_4$
$\mathcal{D}_3$
$\mathcal{D}_1$
0
s
1
$\frac{1+\sqrt{5}}{2}$
$\mathcal{D}_5$
$\leftarrow r = s^2 - s - 1$
(1, −1)

Theorem 4

Let $\{w^k\}$ be generated by the proximal symmetric ADMM. Then, there exist constants C_i, $i = 0, 1, 2, 3$, such that

$$\begin{array}{rcl} \|w^k - \tilde{w}^k\|_G^2 & \geqslant & C_0\beta\big(\|Ax^{k+1} + By^{k+1} - b\|^2 - \|Ax^k + By^k - b\|^2\big) \\ & + & C_1\beta\|B(y^k - y^{k+1})\|^2 + C_2\beta\|Ax^{k+1} + By^{k+1} - b\|^2 \\ & + & C_3(\|y^k - y^{k+1}\|_T^2 - \|y^{k-1} - y^k\|_T^2), \end{array}$$

where

- $C_1, C_2 > 0$, $C_0 = C_3 = 0$ *if* $(r, s) \in \mathcal{D}_1$,
- $C_0 = 0$, C_1, C_2, $C_3 > 0$ *if* $(r, s) \in \mathcal{D}_2$, *and*
- $C_0, C_1, C_2, C_3 > 0$ *if* $(r, s) \in \mathcal{D}_3 \cup \mathcal{D}_4 \cup \mathcal{D}_5$.

Global convergence

Theorem 5

For the sequence $\{w^k\}$ *generated by the proximal symmetric ADMM, we have*

$$\lim_{k\to\infty}\left(\|B(y^k-y^{k+1})\|^2+\|Ax^{k+1}+By^{k+1}-b\|^2\right)=0.$$

Moreover, if the matrices A, B *are assumed to be full column rank, then the sequence* $\{w^k\}$ *converges to a solution point* $w^\infty\in\Omega^*$.

- stopping criterion:

$$\max\{\|B(y^k-y^{k+1})\|^2,\|Ax^{k+1}+By^{k+1}-b\|^2\}\leqslant\varepsilon$$

which $\varepsilon>0$ is the tolerance specified by the user.

Convergence rate

Worse-case $O(1/t)$ convergence rate in the ergodic sense:

Theorem 6

Let the sequence $\{w^k\}$ *be generated by the proximal symmetric ADMM. Then, for* $(s, r) \in \mathcal{D}$ *and any integer number* $t > 0$*, we have*

$$\begin{aligned}&\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w)\\ \leqslant&\frac{1}{2t}\left(\|w - w^1\|_H^2 + C_0\|Ax^1 + By^1 - b\|^2 + C_3\|y^0 - y^1\|_T^2\right) \quad \forall w \in \Omega,\end{aligned}$$

where

$$\tilde{w}_t = \frac{1}{t}\sum_{k=1}^{t} \tilde{w}^k.$$

1 Introduction

2 Main result

3 Convergence analysis

4 Numerical result

5 Conclusion

Numerical result

Basis pursuit

$$\min\left\{\frac{1}{2\mu}\|Ax-b\|_2^2+\|y\|_1 | x-y=0, x,y\in\Re^n\right\}$$

- $A\in\Re^{m\times n}$: a random Gaussian matrix,
- x^*: original signal, has p nonzero elements whose positions are arranged randomly
- b: sampling signal with white noise

Numerical result

Basis pursuit

$$\min\left\{\frac{1}{2\mu}\|Ax-b\|_2^2+\|y\|_1|x-y=0,x,y\in\Re^n\right\}$$

- $A\in\Re^{m\times n}$: a random Gaussian matrix,
- x^*: original signal, has p nonzero elements whose positions are arranged randomly
- b: sampling signal with white noise

iterative scheme:

$$\begin{cases} x^{k+1}=(A^TA/\mu+\beta I+R)^{-1}(A^Tb/\mu+\lambda^k+\beta y^k+Rx^k),\\ \lambda^{k+\frac{1}{2}}=\lambda^k-r\beta(x^{k+1}-y^k),\\ y^{k+1}=\mathcal{S}_{\frac{1}{\beta}}(x^{k+1}-\lambda^{k+\frac{1}{2}}/\beta),\\ \lambda^{k+1}=\lambda^{k+\frac{1}{2}}-s\beta(x^{k+1}-y^{k+1}), \end{cases}$$

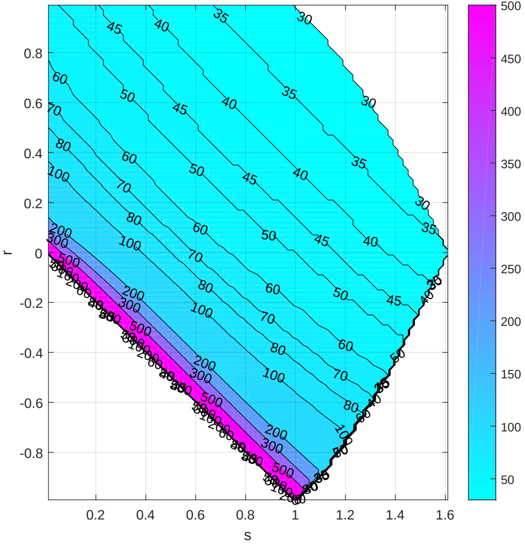


Figure: $R = \left(0.01\beta + \frac{1.01}{2}\lambda_{\max}(A^T A)\right) \cdot I - \frac{A^T A}{2}$

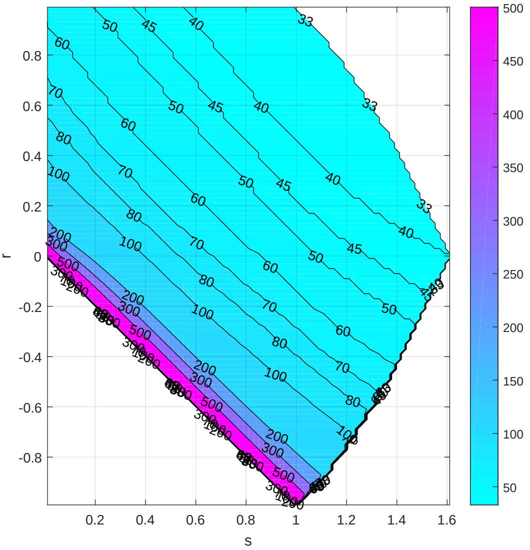


Figure: $R = \lambda_{\max}(A^T A) \cdot I - A^T A,\ T = 0$

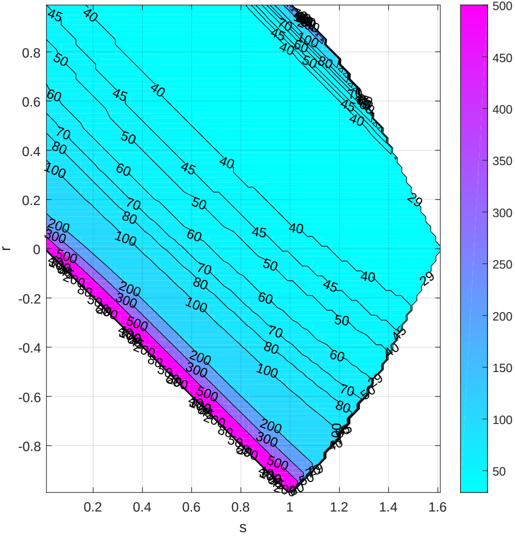

r
s
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
500
450
400
350
300
250
200
150
100
50

Figure: $R = 0, \ T = 0$

1 Introduction

2 Main result

3 Convergence analysis

4 Numerical result

5 Conclusion

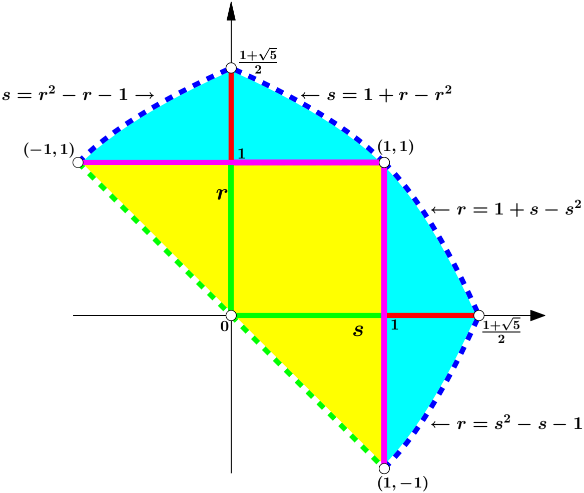
Extended domain

Noticing that the parameter r and s play the equal role in proximal symmetric ADMM scheme, clearly, the scheme of proximal symmetric ADMM can be written as:

$$\begin{cases} y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^k, y, \lambda^k) + \frac{1}{2}\|y - y^k\|_T^2 \mid y \in \mathcal{Y}\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - s\beta(Ax^k + By^{k+1} - b), \\ x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^{k+1}, \lambda^{k+\frac{1}{2}}) + \frac{1}{2}\|x - x^k\|_R^2 \mid x \in \mathcal{X}\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - r\beta(Ax^{k+1} + By^{k+1} - b), \end{cases}$$

The domain of the step sizes can be extended as:

$$\mathcal{D}^{\mathsf{sym}} = \{(r,s) \mid r+s > 0,\ |r| < 1+s-s^2\} \cup \{(r,s) \mid r+s > 0,\ |s| < 1+r-r^2\}$$


$\frac{1+\sqrt{5}}{2}$
$s = r^2 - r - 1 \rightarrow$
$\leftarrow s = 1 + r - r^2$
$(-1, 1)$
$(1, 1)$
1
r
$\leftarrow r = 1 + s - s^2$
0
s
1
$\frac{1+\sqrt{5}}{2}$
$\leftarrow r = s^2 - s - 1$
$(1, -1)$

Thank you!