Convex Quadratic Programming: Restricted Wolfe Dual and Symmetric Gauss-Seidel Decomposition Theorem

Defeng Sun

Department of Applied Mathematics

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

Shandong University
(The Fourth Youth Forum of China Operations Research)
June 28, 2020

Linear programming

Consider the following linear programming

$$\begin{aligned} \min \quad & \langle c, x \rangle \\ \text{s.t.} \quad & Ax = b, \\ & x \geq 0, \end{aligned}$$

where $A : \Re^n \to \Re^m$, $b \in \Re^m$ and $c \in \Re^n$. Define the Lagrange function as follows:

$$L(x; y, s) := \langle c, x \rangle + \langle y, b - Ax \rangle - \langle s, x \rangle.$$

Then the (Lagrange) dual of linear programming is defined as

$$\max_{y \in \Re^m, s \geq 0} \quad \left\{ \inf_{x \in \Re^n} L(x; y, s) \right\}.$$

Lagrange dual of linear programming

By noting that for any $(x, y, s) \in \Re^n \times \Re^m \times \Re^n$,

$$L(x; y, s) := \langle x, c - A^* y - s \rangle + \langle y, b \rangle,$$

we get an explicit formula for the dual problem as in the following

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

or equivalently

$$\begin{array}{ll} \max & \langle b, y \rangle \\ \text{s.t.} & A^* y \leq c. \end{array}$$

No body will question the above (Lagrange) dual!

Linear conic programming and its dual

In the same vein, we can write down the (linear) conic programming

$$\begin{aligned} \min \quad & \langle c, x \rangle \\ \text{s.t.} \quad & Ax = b, \\ & x \in \mathcal{K} \end{aligned}$$

and its (Lagrange) dual

$$\begin{aligned} \max \quad & \langle b, y \rangle \\ \text{s.t.} \quad & A^* y + s = c, \\ & s \in \mathcal{K}^*, \end{aligned}$$

where $\mathcal{K}$ is a closed convex cone and $\mathcal{K}^*$ is its dual, e.g., $\mathcal{K}$ is the second-order-cone or the PSD (positive and semidefinte) cone.

Convex quadratic programming

Now let us turn to the convex quadratic programming (CQP)

$$\min_{x \in \mathcal{X}} \left\{ \frac{1}{2}\langle x, \mathcal{Q}x\rangle + \langle c, x\rangle + \psi(x) \;\Big|\; \mathcal{A}x = b \right\}$$

- $\psi : \mathcal{X} \to (-\infty, +\infty]$ is a closed proper convex polyhedral function, e.g., $\psi(\cdot) = \delta_P(\cdot)$, the indicator function over a convex polyhedral set [simple]
- $\mathcal{Q} : \mathcal{X} \to \mathcal{X}$ satisfying $\mathcal{Q} = \mathcal{Q}^*$, $\mathcal{Q} \succeq 0$
- $\mathcal{A} : \mathcal{X} \to \mathcal{Y}$ is a given linear mapping
- $b \in \mathcal{Y}$ is a given vector
- $c \in \mathcal{X}$ is given
- $\mathcal{X}$ and $\mathcal{Y}$ are two finite-dimensional real Euclidean spaces.

Convex quadratic programming (continued)

Equivalently,

$$\min_{u,x\in\mathcal{X}} \left\{ \frac{1}{2}\langle x, \mathcal{Q}x\rangle + \langle c, x\rangle + \psi(u) \;\middle|\; \mathcal{A}x = b, \quad x - u = 0 \right\}$$

The corresponding Lagrange function is

$$\begin{aligned} L(u, x; y, s) \quad &:= \quad \tfrac{1}{2}\langle x, \mathcal{Q}x\rangle + \langle c, x\rangle + \psi(u) + \langle y, b - \mathcal{A}x\rangle + \langle s, u - x\rangle \\ &= \quad \tfrac{1}{2}\langle x, \mathcal{Q}x\rangle + \langle x, c - \mathcal{A}^*y - s\rangle + \psi(u) + \langle s, u\rangle + \langle y, b\rangle \end{aligned}$$

and the Lagrange dual of CQP takes the form of

$$\max_{y\in\mathcal{Y}, s\in\mathcal{X}} \left\{ \inf_{u\in\mathcal{X}, x\in\mathcal{X}} L(u, x; y, s) \right\}$$

or

$$\max_{y\in\mathcal{Y}, s\in\mathcal{X}} \left\{ \langle y, b\rangle + \inf_{u\in\mathcal{X}} \left\{ \psi(u) + \langle s, u\rangle \right\} + \inf_{x\in\mathcal{X}} \left\{ \frac{1}{2}\langle x, \mathcal{Q}x\rangle + \langle x, c - \mathcal{A}^*y - s\rangle \right\} \right\}$$

Wolfe dual

By simplifying, we get the following Lagrange dual

$$\max_{y\in\mathcal{Y},s\in\mathcal{X}} \Big\{ -\psi^*(-s) + \langle y, b\rangle + \theta(y,s)\Big\},$$

where

$$\theta(y,s) := \inf_{x\in\mathcal{X}} \Big\{ \frac{1}{2}\langle x, \mathcal{Q}x\rangle + \langle x, c - \mathcal{A}^* y - s\rangle \Big\}$$

and $\psi^*(\cdot)$ is the Fenchel conjugate of ψ defined by

$$\psi^*(s) := \sup_{u\in\mathcal{X}} \{\langle s, u\rangle - \psi(u)\}.$$

Since the computation of $\theta(y,s)$ is complicated, instead one normally considers the following Wolfe dual

$$\max_{s\in\mathcal{X},x\in\mathcal{X},y\in\mathcal{Y}} \Big\{ -\psi^*(-s) - \frac{1}{2}\langle x, \mathcal{Q}x\rangle + \langle y, b\rangle \;\Big|\; s - \mathcal{Q}x + \mathcal{A}^* y = c \Big\}.$$

Limitations of the Wolfe dual

Note that in the Wolfe dual (in the minimization format)

$$\min_{s\in\mathcal{X},x\in\mathcal{X},y\in\mathcal{Y}} \Big\{\psi^*(-s) + \frac{1}{2}\langle x, \mathcal{Q}x\rangle - \langle y, b\rangle \;\Big|\; s - \mathcal{Q}x + \mathcal{A}^*y = c\Big\},$$

the primal variable x is also involved. But more seriously, its solution set, if nonempty, is always unbounded as long as $\mathcal{Q} \not\succ 0$ (the null space of $\mathcal{Q}$ is uncontrollable).

- It lacks unification with linear conic programming
- The deep equivalent connections between (LICQ) strict MFCQ of the primal and the (strong) second order sufficient conditions of the restricted Wolfe dual are lost
- The dual based approaches such as the augmented Lagrangian method cannot be used. For example, the software SDPNAL (Zhao, Sun and Toh; SIOPT 2010) for solving the dual of semidefinite programming cannot be extended to the Wolfe dual
- In a word, neither theory nor computation supports the Wolfe dual

Restricted Wolfe dual

Our remedy is to consider the following restricted Wolfe dual (in the minimization format)

$$\min_{s\in\mathcal{X},x'\in\mathcal{X}',y\in\mathcal{Y}} \left\{ \psi^*(-s) + \frac{1}{2}\langle x', \mathcal{Q}x'\rangle - \langle y, b\rangle \;\middle|\; s - \mathcal{Q}x' + \mathcal{A}^*y = c \right\},$$

where $\mathcal{X}'$ is the range space of $\mathcal{Q}$, i.e.,

$$\mathcal{X}' := \mathrm{Range}\,(\mathcal{Q}).$$

One can easily check that $\mathcal{Q} : \mathcal{X}' \to \mathcal{X}'$ is self-adjoint and positive definite even if $\mathcal{Q} : \mathcal{X} \to \mathcal{X}$ is not positive definite. Note that if $\mathcal{Q} = 0$, then $\mathcal{X}' = \{0\}$ (in this case $\mathcal{Q}$ is still positive definite on $\mathcal{X}'$ – using definition to verify it!). Also note that x' in the dual is differen from x in the primal, which does not need to stay in $\mathrm{Range}\,(\mathcal{Q})$.

Nice properties of the restricted Wolfe dual

Different from the Wolfe dual, the restricted Wolfe dual possesses the following nice properties:

- It unifies with linear conic programming
- The wonderful equivalent connections between the (LICQ) strict MFCQ of the primal and the (strong) second order sufficient conditions of the restricted Wolfe dual are kept
- The dual based approaches such as the augmented Lagrangian method can be employed. For example, Li, Sun and Toh (MPC, 2018) has successfully extended the software SDPNAL for solving the dual of semidefinite programming to the restricted Wolfe dual of the convex quadratic semidefinite programming (software QSDPNAL)
- In a word, both theory and computation favor the restricted Wolfe dual

Too good to be true? — the key is to keep in $\mathrm{Range}\,(\mathcal{Q})$. But, how?

Augmented Lagrange function of restricted Wolfe dual

For given $\sigma > 0$, the augmented Lagrange function of the restriced Wolfe dual of the convex quadratic programming can be written as

$$\begin{aligned} L_\sigma(s, x', y; x) \quad := \quad & \psi^*(-s) + \frac{1}{2}\langle x', \mathcal{Q}x'\rangle - \langle y, b\rangle \\ & + \langle x, s - \mathcal{Q}x' + \mathcal{A}^*y - c\rangle + \frac{\sigma}{2}\|s - \mathcal{Q}x' + \mathcal{A}^*y - c\|^2, \end{aligned}$$

which, fixing the dual variable x, is a proper closed convex function in the first block variable s plus a convex quadratic function in terms of (s, x', y). Note that y can be further split into many pieces as you please.

The above property of the augmented Lagrange function is also true for the CQP in the primal form.

Augmented Lagrange function of the primal CQP

For given $\sigma > 0$, the augmented Lagrange function of the CQP (primal)

$$\min_{u,x\in\mathcal{X}} \left\{ \frac{1}{2}\langle x, \mathcal{Q}x\rangle + \langle c, x\rangle + \psi(u) \;\middle|\; \mathcal{A}x = b, \quad x - u = 0 \right\}$$

takes the form of

$$\begin{aligned} L_\sigma(u, x; y, s) \quad := \quad & \psi(u) + \tfrac{1}{2}\langle x, \mathcal{Q}x\rangle + \langle c, x\rangle + \langle y, b - \mathcal{A}x\rangle + \langle s, u - x\rangle \\ & + \tfrac{\sigma}{2}\|b - \mathcal{A}x\|^2 + \tfrac{\sigma}{2}\|u - x\|^2, \end{aligned}$$

which, fixing the dual variables y and s, is a proper closed convex function in the first block variable u plus a convex quadratic function in terms of (u, x). Note that x can be further split into as many pieces as you like.

Note that the augmented Lagrange function for the convex QP in the primal form does not contain a nonsmooth term for x even if it is non-separable for x.

The (mysterious?) forms of the two augmented Lagrange functions lead to the discovery of the symmetric Gauss-Seidel decomposition theorem!

Convex composite quadratic programming

Actually, we can consider more general convex composite quadratic programming (CCQP)

$$\min_{x \in \mathcal{X}} \left\{ \psi(x) + \frac{1}{2}\langle x, \mathcal{Q}x \rangle - \langle c, x \rangle \;\middle|\; \mathcal{A}_E x = b_E,\ \mathcal{A}_I x - b_I \in \mathcal{K} \right\}$$

- $\psi : \mathcal{X} \to (-\infty, +\infty]$ is a closed proper convex function [simple]
- $\mathcal{Q} : \mathcal{X} \to \mathcal{X}$ satisfying $\mathcal{Q} = \mathcal{Q}^*,\ \mathcal{Q} \succeq 0$
- $\mathcal{A}_E : \mathcal{X} \to \mathcal{Z}_1$ and $\mathcal{A}_I : \mathcal{X} \to \mathcal{Z}_2$, given linear mappings
- $b = (b_E; b_I) \in \mathcal{Z} := \mathcal{Z}_1 \times \mathcal{Z}_2$, given vector
- $c \in \mathcal{X}$ is given.
- $\mathcal{K} \subseteq \mathcal{Z}_2$ is a closed convex set (cone) [simple]
- $\mathcal{X}$, $\mathcal{Z}_1$, and $\mathcal{Z}_2$ are finite-dimensional real Euclidean spaces

CCQP and its restricted Wolfe dual

Equivalently,

$$\min_{x\in\mathcal{X},x'\in\mathcal{Z}_2} \left\{\psi(x) + \delta_{\mathcal{K}}(x') + \frac{1}{2}\langle x, \mathcal{Q}x\rangle - \langle c, x\rangle \;\middle|\; \begin{pmatrix} \mathcal{A}_E & 0 \\ \mathcal{A}_I & -\mathcal{I} \end{pmatrix} \begin{pmatrix} x \\ x' \end{pmatrix} = b \right\},$$

whose restricted Wolfe dual (in the minimization format) is

$$\min_{\substack{s\in\mathcal{Y},z\in\mathcal{Z} \\ y'\in\mathrm{Range}(\mathcal{Q})}} \left\{ p(s) + \frac{1}{2}\langle y', \mathcal{Q}y'\rangle - \langle b, z\rangle \;\middle|\; s + \begin{pmatrix} \mathcal{Q} \\ 0 \end{pmatrix} y' - \begin{pmatrix} \mathcal{A}_E^* & \mathcal{A}_I^* \\ 0 & -\mathcal{I} \end{pmatrix} z = \begin{pmatrix} c \\ 0 \end{pmatrix} \right\}$$

- $s := (u, v) \in \mathcal{Y} := \mathcal{X} \times \mathcal{Z}_2$
- $p(s) := p(u, v) = \psi^*(u) + \delta_{\mathcal{K}}^*(v)$
- $\delta_{\mathcal{K}}(\cdot)$ is the indicator function over $\mathcal{K}$

Gauss-Seidel method for solving $Qx = b$

- Update only one element of the variable $\mathbf{x}$ in each iteration.

Input: $Q \in \Re^{n \times n}, \mathbf{b} \in \Re^n$ and $\mathbf{x}^0 \in \Re^n$

for $k = 0, 1, \ldots$
  for $i = 1, \ldots, n$

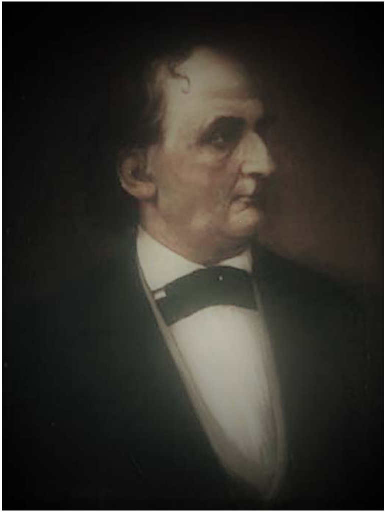
$$\mathbf{x}_i^{k+1} := Q_{ii}^{-1} \left(\mathbf{b}_i - \sum_{j=1}^{i-1} Q_{ij} \mathbf{x}_j^{k+1} - \sum_{j=i+1}^{n} Q_{ij} \mathbf{x}_j^k \right)$$

  end for
end for

- Converges if Q is diagonally dominant, or symmetric positive definite.
- Was it really the original form invented by Gauss or Seidel? We were taught so in the textbooks...

F. Gauss and L. Seidel

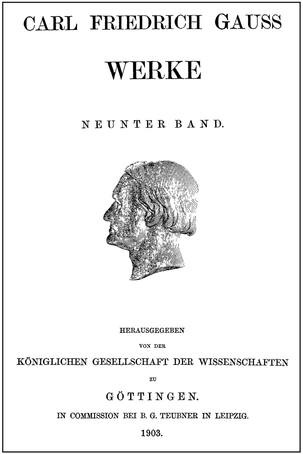
Johann Carl Friedrich Gauß
(30 April 1777--23 February 1855)



Philipp Ludwig von Seidel
(23 October 1821--13 August 1896)

*Photos from Wikipedia

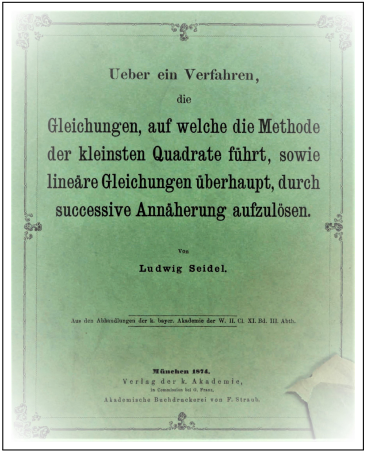
Mentioned in a private letter[1] from Gauss to Gerling in 1823.
A publication was not delivered before 1874 by Seidel.


CARL FRIEDRICH GAUSS

WERKE

NEUNTER BAND.

HERAUSGEGEBEN
VON DER
KÖNIGLICHEN GESELLSCHAFT DER WISSENSCHAFTEN
ZU
GÖTTINGEN.
IN COMMISSION BEI B. G. TEUBNER IN LEIPZIG.
1903.


Ueber ein Verfahren,
die
Gleichungen, auf welche die Methode der kleinsten Quadrate führt, sowie lineäre Gleichungen überhaupt, durch successive Annäherung aufzulösen.

Von
Ludwig Seidel.

Aus den Abhandlungen der k. bayer. Akademie der W. II. Cl. XI. Bd. III. Abth.

München 1874.
Verlag der k. Akademie,
in Commission bei G. Franz.
Akademische Buchdruckerei von F. Straub.

[1] In Carl Friedrich Gauss Werke 9, Geodäsie, 278-281 (1903). English translation in J.-L. Chabert (Ed.), A History of Algorithms, Springer-Verlag, Berlin, Heidelberg, 297–298 (1999).

Gauss' letter to Gerling

[6.]

[Über Stationsausgleichungen.]

Gauss an Gerling. Göttingen, 26. December 1823.

Mein Brief ist zu spät zur Post gekommen und mir zurückgebracht. Ich erbreche ihn daher wieder, um noch die praktische Anweisung zur Elimination beizufügen. Freilich gibt es dabei vielfache kleine Localvortheile, die sich nur ex usu lernen lassen.

......

......

Die Bedingungsgleichungen sind also:

$$\begin{aligned}
0 &= +6 + 67a - 13b - 28c - 26d \\
0 &= -7558 - 13a + 69b - 50c - 6d \\
0 &= -14604 - 28a - 50b + 156c - 78d \\
0 &= +22156 - 26a - 6b - 78c + 110d;
\end{aligned}$$

$$\text{Summe} = 0.$$

......

- Gauss considered a 4 dimensional symmetric positive semidefinite but singular linear equation.
- Starting from $(a,b,c,d) = (0,0,0,0)$, update exactly one variable from $\{a,b,c,d\}$ each time via a certain rule.
- Gauss worked with integers: an inexact iterative method!

The conditional equations are:

$$
\begin{aligned}
0 &= +\quad 6 + 67a - 13b - 28c - 26d \\
0 &= -\ 7558 - 13a + 69b - 50c - 6d \\
0 &= -\ 14604 - 28a - 50b + 156c - 78d \\
0 &= +\ 22156 - 26a - 6b - 78c + 110d \\
\hline
\text{sum} &= 0
\end{aligned}
$$

To eliminate indirectly now, I notice that, if three of the quantities *a*, *b*, *c*, *d* are set equal to 0, the fourth will take the greatest value if *d* is chosen as the fourth quantity. Naturally each quantity has to be determined from its own equation, and so *d* from the fourth one. So I put $d = -201$ and substitute this value. The constant terms then become: + 5232, – 6352, + 1074, +46; the rest remaining unchanged.

Now I let *b* be next, and I find $b = +92$, I substitute it, and I find the constant terms: + 4036, – 4, – 3526, – 506. I continue in this way until there is nothing left to correct. But in actual fact, for the whole of this calculation, I merely write out the following table:

Gauss' algorithm and conclusion

	$d = -201$	$b = +92$	$a = -60$	$c = +12$	$a = +5$	$b = -2$	$a = -1$
+ 6	+ 5232	+ 4036	+ 16	− 320	+ 15	+ 41	− 26
− 7558	− 6352	− 4	+ 776	+ 176	+ 111	− 27	− 14
− 14604	+ 1074	− 3526	− 1846	+ 26	− 114	− 14	+ 14
+ 22156	+ 46	− 506	+ 1054	+ 118	− 12	0	+ 26

In that I am only taking the calculation to the next 1/2000th of a second, I see that there is now nothing more to correct. I collect up the terms:

$a = -60$	$b = +92$	$c = +12$	$d = -201$
+ 5 − 1	− 2		
− 56	+ 90	+ 12	− 201

Almost
prove. Co
pleasant c
remains to
again hav
knowns. T
about othe

Seidel's method

To solve the linear equation

$$\boxed{Ax = b} \quad \text{with} \quad A \in \Re^{m\times n}, b \in \Re^{m},$$

Seidel defined the quadratic function

$$q(x) := \frac{1}{2}\|Ax - b\|_2^2 = \frac{1}{2}\langle x, (A^*A)x\rangle - \langle b, Ax\rangle + \frac{1}{2}\|b\|^2$$

to solve the corresponding normal equation

$$\mathrm{Q}x = A^*b \quad \text{with} \quad \mathrm{Q} := A^*A.$$

- Update only one component of the vector x each step to reduce the value of q.
- The most rational thing (according to Seidel): choose the index that brings the maximum update (decrease) of q.

Gauss-Seidel

The now well-known Gauss-Seidel iterative method:

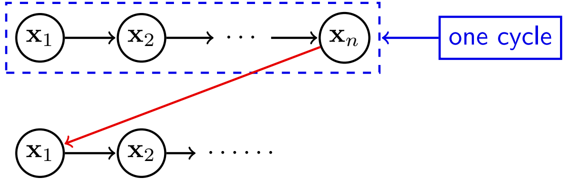
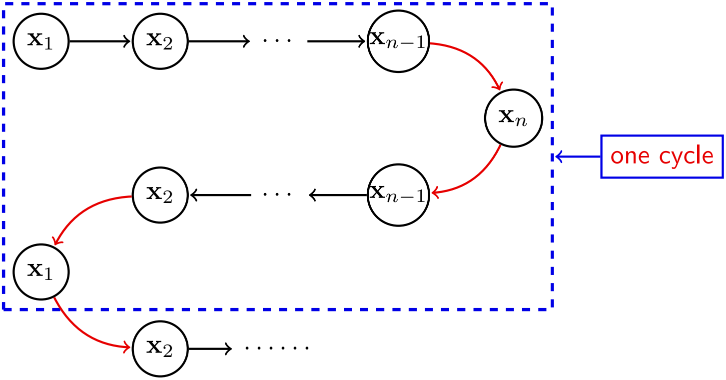
- Forget the "optimal" choice indicated by Gauss and Seidel.
- Changes are carried "cyclically".
- Successively update the elements of x in a fixed order.
- Turn to the first one if the last one is updated.
- What have been taking about is actually only the sequential Gauss-Seidel method – a much over simplified one.

How about turning to the penultimate one and so on after the last one is updated

- such as the symmetric Gauss-Seidel (sGS) iterative method[2]?
- Note that for $n = 2$, GS $\equiv$ sGS, which means that the two-block case is indeed special. Maybe sGS is the real tool?

[2] R.E. Bank, T.F. Dupont, and H. Yserentant, "The hierarchical basis multigrid method", Numerische Mathematik 52, 427–458 (1988).

Comparison: GS vs. sGS

GS	
sGS	

A general form of symmetric Gauss-Seidel iteration

Consider the **block** vector $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_s) \in \mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_s$. Given a positive semidefinite linear operator $\mathcal{Q}$ such that

$$\mathcal{Q}\mathbf{x} \equiv \begin{pmatrix} \mathcal{Q}_{11} & \mathcal{Q}_{12} & \cdots & \mathcal{Q}_{1s} \\ \mathcal{Q}_{12}^* & \mathcal{Q}_{22} & \cdots & \mathcal{Q}_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{Q}_{1s}^* & \mathcal{Q}_{2s}^* & \cdots & \mathcal{Q}_{ss} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_s \end{pmatrix}, \quad \mathcal{Q}_{ii} \succ 0.$$

Let $p : \mathcal{X}_1 \to (-\infty, +\infty]$ be a given closed proper convex function. Let the quadratic function

$$q(\mathbf{x}) := \frac{1}{2}\langle \mathbf{x},\, \mathcal{Q}\mathbf{x} \rangle - \langle \mathbf{r},\, \mathbf{x} \rangle.$$

Consider the problem $\boxed{\min_{\mathbf{x}\in\mathcal{X}} p(\mathbf{x_1}) + q(\mathbf{x})}$

- Block GS and block sGS (no conditions) are applicable.
- For sGS, one can get iteration complexity + linear convergence under error bounds with no efforts
- But, more importantly, block sGS can be used together with the celebrated acceleration technique of Nesterov[3].

[3] Yu. E. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$", Soviet Mathematics Doklady 27(2), 372–376 (1983).

An inexact block sGS iteration

Consider the following block decomposition:

$$\mathcal{U}\mathbf{x} \equiv \begin{pmatrix} 0 & \mathcal{Q}_{12} & \cdots & \mathcal{Q}_{1s} \\ & \ddots & & \vdots \\ & & \ddots & \mathcal{Q}_{(s-1)s} \\ & & & 0 \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_s \end{pmatrix}.$$

Then $\mathcal{Q} = \mathcal{U}^* + \mathcal{D} + \mathcal{U}$, where $\mathcal{D}\mathbf{x} = (\mathcal{Q}_{11}\mathbf{x}_1, \ldots, \mathcal{Q}_{ss}\mathbf{x}_s)$.

Let $\hat{\delta} \equiv (\hat{\delta}_1, \ldots, \hat{\delta}_s)$ and $\delta^+ \equiv (\delta_1^+, \ldots, \delta_s^+)$ with $\hat{\delta}_1 = \delta_1^+$ being given error tolerance vectors. Define

$$\Delta(\hat{\delta}, \delta^+) := \delta^+ + \mathcal{U}\mathcal{D}^{-1}(\delta^+ - \hat{\delta}), \ \mathcal{T} := \mathcal{U}\mathcal{D}^{-1}\mathcal{U}^* \ \text{(sGS decomp. op.)}.$$

Note that $\mathcal{T} \succeq 0$ is NOT positive definite. Let $\mathbf{y} \in \mathcal{X}$ be given. Define

$$\mathbf{x}^+ := \operatorname*{arg\,min}_{\mathbf{x} \in \mathcal{X}} \left\{ p(\mathbf{x}_1) + q(\mathbf{x}) + \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_{\mathcal{T}}^2 - \langle \Delta(\hat{\delta}, \delta^+), \mathbf{x} \rangle \right\}. \qquad (1)$$

(1) looks complicated, but is much easier to solve!

An inexact block sGS decomposition theorem

Theorem (Li-Sun-Toh, MP 2019)

Given $\mathbf{y}$*. For* $i = s, \ldots, 2$*, define*

$$\begin{aligned} \hat{\mathbf{x}}_i &:= \arg\min_{\mathbf{x}_i} \{ p(\mathbf{y}_1) + q(\mathbf{y}_{\le i-1}, \mathbf{x}_i, \hat{\mathbf{x}}_{\ge i+1}) - \langle \hat{\delta}_i, \mathbf{x}_i \rangle \} \\ &= \mathcal{Q}_{ii}^{-1} \big(\mathrm{r}_i + \hat{\delta}_i - \textstyle\sum_{j=1}^{i-1} \mathcal{Q}_{ji}^* \mathbf{y}_j - \sum_{j=i+1}^{s} \mathcal{Q}_{ij} \hat{\mathbf{x}}_j \big) \end{aligned}$$

computed in the backward GS cycle. **The optimal solution $\mathbf{x}^+$ in (1) can be obtained exactly via**

$$\begin{aligned} \mathbf{x}_1^+ &= \arg\min_{\mathbf{x}_1} \{ p(\mathbf{x}_1) + q(\mathbf{x}_1, \hat{\mathbf{x}}_{\ge 2}) - \langle \delta_1^+, \mathbf{x}_1 \rangle \}, \\ \mathbf{x}_i^+ &= \arg\min_{\mathbf{x}_i} \{ p(\mathbf{x}_1^+) + q(\mathbf{x}_{\le i-1}^+, \mathbf{x}_i, \hat{\mathbf{x}}_{\ge i+1}) - \langle \delta_i^+, \mathbf{x}_i \rangle \} \\ &= \mathcal{Q}_{ii}^{-1} (\mathrm{r}_i + \delta_i^+ - \textstyle\sum_{j=1}^{i-1} \mathcal{Q}_{ji}^* \mathbf{x}_j^+ - \sum_{j=i+1}^{s} \mathcal{Q}_{ij} \hat{\mathbf{x}}_j), \quad i \ge 2, \end{aligned}$$

where $\mathbf{x}_i^+$*,* $i = 1, 2, \ldots, s$*, is computed in the forward GS cycle.*

Reduces to the classical block sGS if both $p(\cdot) \equiv 0$ and $\delta = 0$.
Caution: Such a theorem is not available for GS even if $p(\cdot) \equiv 0$.

An inexact APG (accelerated proximal gradient)

Consider

$$\min\{F(x) := p(\mathbf{x}) + f(\mathbf{x}) \mid \mathbf{x} \in \mathcal{X}\}$$

with $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{x}')\| \leq L\|\mathbf{x} - \mathbf{x}'\| \quad \forall\, \mathbf{x}, \mathbf{x}' \in \mathcal{X}$.

Algorithm. Input $\mathbf{y}^1 = \mathbf{x}^0 \in \mathrm{dom}(p)$, $t_1 = 1$. Iterate

1. Find an approximate minimizer $\mathbf{x}^k$ to

$$\min_{\mathbf{y}\in\mathcal{X}}\Big\{p(\mathbf{y}) + f(\mathbf{y}^k) + \langle \nabla f(\mathbf{y}^k), \mathbf{y} - \mathbf{y}^k\rangle + \frac{1}{2}\langle \mathbf{y} - \mathbf{y}^k, \mathcal{H}_k(\mathbf{y} - \mathbf{y}^k)\rangle\Big\},$$

   where $\mathcal{H}_k \succ 0$ is a priorily given linear operator.

2. Compute $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$, $\mathbf{y}^{k+1} = \mathbf{x}^k + \left(\frac{t_k - 1}{t_{k+1}}\right)(\mathbf{x}^k - \mathbf{x}^{k-1})$.

An inexact APG

Consider the following admissible conditions

$$F(\mathbf{x}^k) \ \le \ p(\mathbf{x}^k) + f(\mathbf{y}^k) + \langle \nabla f(\mathbf{y}^k), \, \mathbf{x}^k - \mathbf{y}^k \rangle + \tfrac{1}{2} \langle \mathbf{x}^k - \mathbf{y}^k, \, \mathcal{H}_k(\mathbf{x}^k - \mathbf{y}^k) \rangle,$$

$$\nabla f(\mathbf{y}^k) + \mathcal{H}_j(\mathbf{x}^k - \mathbf{y}^k) + \gamma^k =: \delta^k \quad \text{with } \|\mathcal{H}_k^{-1/2} \delta^k\| \le \frac{\epsilon_k}{\sqrt{2} t_k},$$

where $\gamma^k \in \partial p(\mathbf{x}^k)$ = the set of subgradients of p at $\mathbf{x}^k$, $\{\epsilon_k\}$ is a nonnegative summable sequence. Note $t_k \approx k/2$ for k large.

Theorem (Jiang-Sun-Toh, SIOPT 2012)

Suppose that the above conditions hold and $\mathcal{H}_{k-1} \succeq \mathcal{H}_k \succ 0$ for all k. Then

$$0 \le F(\mathbf{x}^k) - F(\mathbf{x}^*) \le \frac{2}{(k+1)^2} \Big[\Big(\|\mathbf{x}^0 - \mathbf{x}^*\|_{\mathcal{H}_1} + \sqrt{6} \sum_{j=1}^{k} \epsilon_j \Big)^2 \Big].$$

An inexact APG

Apply the inexact APG to

$$\min\{F(\mathbf{x}) := p(\mathbf{x}_1) + f(\mathbf{x}) \mid \mathbf{x} \in \mathcal{X}\}.$$

Since $\nabla f(\cdot)$ is Lipschitz continuous, $\exists$ a symmetric PSD linear operator $\mathcal{Q} : \mathcal{X} \to \mathcal{X}$ such that

$$\mathcal{Q} \ \succeq \ \mathcal{M}, \quad \forall\, \mathcal{M} \in \partial^2 f(\mathbf{x}),\ \forall\, \mathbf{x} \in \mathcal{X}$$

and $\mathcal{Q}_{ii} \succ 0$ for all i.
Given y^k, we have for all $\mathbf{x} \in \mathcal{X}$,

$$f(\mathbf{x}) \ \leq \ q_k(\mathbf{x}) := f(\mathbf{y}^k) + \langle \nabla f(\mathbf{y}^k),\, \mathbf{x} - \mathbf{y}^k \rangle + \frac{1}{2}\langle \mathbf{x} - \mathbf{y}^k,\, \mathcal{Q}(\mathbf{x} - \mathbf{y}^k)\rangle.$$

APG subproblem: need to solve a nonsmooth composite QP of the form

$$\min_{\mathbf{x}\in\mathcal{X}}\{p(\mathbf{x}_1) + q_k(\mathbf{x})\}, \quad x = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_s),$$

which is not easy to solve!
Idea: add an additional proximal term to make it easier (too easy bad too)!

Elimination of one block via the Danskin theorem

Let $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_s) \in \mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_s$ and the corresponding optimization problem

$$\begin{aligned} & \min\{p(\mathbf{x}_1) + \varphi(\mathbf{z}) + \phi(\mathbf{z}, \mathbf{x}) \mid \mathbf{z} \in \mathcal{Z},\ \mathbf{x} \in \mathcal{X}\} \\ = & \boxed{\min\{p(\mathbf{x}_1) + f(\mathbf{x}) \mid \mathbf{x} \in \mathcal{X}\}}, \end{aligned}$$

where $p(\cdot)$, $\varphi(\cdot)$ are convex functions (possibly nonsmooth), and

$$\begin{aligned} f(\mathbf{x}) &= \min\{\varphi(\mathbf{z}) + \phi(\mathbf{z}, \mathbf{x}) \mid \mathbf{z} \in \mathcal{Z}\}, \\ z(\mathbf{x}) &= \operatorname{argmin}\{\ldots\}. \end{aligned}$$

Assume that φ, ϕ satisfy the conditions in the next theorem, then f has Lipschitz continuous gradient $\nabla f(\mathbf{x}) = \nabla_x \phi(\mathbf{z}(\mathbf{x}), \mathbf{x})$.

A Danskin-type theorem

- $\varphi : \mathcal{Z} \to (-\infty, \infty]$ is a closed proper convex function.
- $\phi(\cdot, \cdot) : \mathcal{Z} \times \mathcal{X} \to \Re$ is a convex function.
- $\phi(\mathbf{z}, \cdot) : \Omega \to \Re$ is continuously differentiable on Ω for each $\mathbf{z}$.
- $\nabla_x \phi(\mathbf{z}, \mathbf{x})$ is continuous on $\mathrm{dom}(\varphi) \times \Omega$.

Consider $f : \Omega \to [-\infty, +\infty)$ defined by

$$f(x) = \inf_{\mathbf{z} \in \mathcal{Z}} \{\varphi(\mathbf{z}) + \phi(\mathbf{z}, \mathbf{x})\}, \quad \mathbf{x} \in \Omega.$$

Condition: The minimizer $\mathbf{z}(\mathbf{x})$ is unique for each $\mathbf{x}$ and is bounded on a compact set.

A Danskin-type theorem

Theorem

(i) *If* $\exists$ *an open neighborhood* $\mathcal{N}_{\mathbf{x}}$ *of* $\mathbf{x}$ *such that* $\mathbf{z}(\cdot)$ *is bounded on any compact subset of* $\mathcal{N}_{\mathbf{x}}$*, then the convex function* f *is differentiable on* $\mathcal{N}_{\mathbf{x}}$ *and*

$$\nabla f(\mathbf{x}') = \nabla_{\mathbf{x}}\phi(\mathbf{z}(\mathbf{x}'), \mathbf{x}') \quad \forall\, \mathbf{x}' \in \mathcal{N}_{\mathbf{x}}.$$

(ii) *Suppose that* $\mathbf{z}(\cdot)$ *is bounded on any nonempty compact subset of* $\mathcal{Z}$*. Assume that for any* $\mathbf{z} \in \mathrm{dom}(\varphi)$*,* $\nabla_{\mathbf{x}}\phi(\mathbf{z}, \cdot)$ *is Lipschitz continuous on* $\mathcal{Z}$ *and* $\exists\ \Sigma \succeq 0$ *such that for all* $\mathbf{x} \in \mathcal{X}$ *and* $\mathbf{z} \in \mathrm{dom}(\varphi)$*,*

$$\Sigma \succeq \mathcal{H} \quad \forall\, \mathcal{H} \in \partial^2_{\mathbf{xx}}\phi(\mathbf{z}, \mathbf{x}).$$

Then, $\nabla f(\cdot)$ *is Lipschitz continuous on* $\mathcal{X}$ *with the Lipschitz constant* $||\Sigma||_2$ *(the spectral norm of* Σ*) and for any* $\mathbf{x} \in \mathcal{X}$*,*

$$\Sigma \succeq \mathcal{G} \quad \forall\, \mathcal{G} \in \partial^2 f(\mathbf{x}),$$

where $\partial^2 f(\mathbf{x})$ *denotes the generalized Hessian of* f *at* $\mathbf{x}$*.*

$$\min\{p(\mathbf{x}_1) + \varphi(\mathbf{z}) + \phi(\mathbf{z}, \mathbf{x}) \mid \mathbf{z} \in \mathcal{Z},\ \mathbf{x} \in \mathcal{X}\}$$

Algorithm 2. Input $\mathbf{y}^1 = \mathbf{x}^0 \in \mathrm{dom}(p) \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_s$, $t_1 = 1$. Let $\{\epsilon_k\}$ be a nonnegative summable sequence. Iterate

1. Suppose δ_i^k, $\hat{\delta}_i^k \in \mathcal{X}_i$, $i = 1, \ldots, s$, with $\hat{\delta}_1^k = \delta_1^k$, are error vectors such that

$$\max\{\|\delta^k\|, \|\hat{\delta}^k\|\} \le \epsilon_k / (\sqrt{2} t_k),$$

$$\mathbf{z}^k = \arg\min_{\mathbf{z}} \left\{\varphi(\mathbf{z}) + \phi(\mathbf{z}, \mathbf{y}^k)\right\}, \quad \text{(elimination via Danskin)}$$

$$\mathbf{x}^k = \arg\min_{\mathbf{x}} \left\{p(\mathbf{x}_1) + q_k(\mathbf{x}) + \frac{1}{2}\|\mathbf{x} - \mathbf{y}^k\|_{\mathcal{T}}^2 - \langle \Delta(\hat{\delta}^k, \delta^k), \mathbf{x}\rangle\right\}.$$

(inexact sGS)

2. Compute $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$, $\mathbf{y}^{k+1} = \mathbf{x}^k + \left(\frac{t_k - 1}{t_{k+1}}\right)(\mathbf{x}^k - \mathbf{x}^{k-1})$.

An inexact accelerated block coordinate gradient descent

Theorem

Let $\mathcal{H} = \mathcal{Q} + \mathcal{T}$ *and* $\beta = 2\|\mathcal{D}^{-1/2}\| + \|\mathcal{H}^{-1/2}\|$. *The sequence* $\{(\mathbf{z}^k, \mathbf{x}^k)\}$ *generated by Algorithm 2 satisfies*

$$0 \leq F(\mathbf{x}^k) - F(\mathbf{x}^*) \leq \frac{2}{(k+1)^2}\left[\left(\|\mathbf{x}^0 - \mathbf{x}^*\|_{\mathcal{H}} + \sqrt{6}\beta \sum_{j=1}^{k} \epsilon_j\right)^2\right].$$

Consider the convex optimization model:

$$\begin{array}{ll} \min & \theta(y_1) + f(y_1, y_2, \dots, y_s) \\ \text{s.t.} & \mathcal{A}_1^* y_1 + \mathcal{A}_2^* y_2 + \cdots + \mathcal{A}_s^* y_s = c\,. \end{array} \tag{2}$$

Linear mappings: $\mathcal{A}_i,\ i = 1, \dots, s,\ \mathcal{A}^* y = \sum_{i=1}^s \mathcal{A}_i^* y_i,\ y := (y_1, \dots, y_s)$. Closed proper convex function $\theta : \mathcal{Y}_1 \to (-\infty, +\infty]$ and convex quadratic function $f(y) = \frac{1}{2}\langle y,\, \mathcal{Q} y\rangle - \langle b,\, y\rangle$. Then, (2) can be written compactly as

$$\min\{\theta(y_1) + f(y) \mid \mathcal{A}^* y = c\},$$

which is a very general CCQP.
Given $\sigma > 0$, the augmented Lagrangian function of the CCQP is

$$\mathcal{L}_\sigma(y; x) = \theta(y_1) + \underbrace{f(y) + \langle x,\, \mathcal{A}^* y - c\rangle + \frac{\sigma}{2}\|\mathcal{A}^* y - c\|^2}_{\text{quadratic}}\,.$$

The proximal augmented Lagrangian method (pALM) for the CCQP:

Given (y^0, x^0) in the domain and $\tau \in (0,2)$. For $k = 0, 1, \ldots$

Step 1. $y^{k+1} \approx \arg\min \mathcal{L}_\sigma(y; x^k) + \frac{1}{2}\|y - y^k\|^2_{\mathcal{T}}$

$$= \arg\min_y \left\{\theta(y_1) + f(y) + \langle x^k, \mathcal{A}^* y - c\rangle + \frac{\sigma}{2}\|\mathcal{A}^* y - c\|^2 + \frac{1}{2}\|y - y^k\|^2_{\mathcal{T}}\right\}.$$

Step 2. $x^{k+1} = x^k + \tau\sigma(\mathcal{A}^* y^{k+1} - c)$.

- $\mathcal{T}$ is the block sGS decomposition operator of $\mathcal{Q} + \sigma\mathcal{A}\mathcal{A}^*$, which does not need to be formulated explicitly. Note that $\mathcal{T} \succeq 0$ but $\mathcal{T} \not\succ 0$. So it is not a classical pALM, but a "semiproximal" ALM.
- y^{k+1} is obtained via the inexact block sGS procedure [s blocks in total].
- In practice, the dual step-length τ is often chosen in $[1.618, 1.95]$, e.g, $\tau = 1.9$.

Extensions (1)

- In the sGS procedure, the coefficient matrices of the linear systems to be solved only need to be factorized once at the start of the procedure. The additional costs of the repetitions are minimal and can be offset by the larger step length $\tau \in (0, 2)$.
- There are many applications that can be "solved" via block sGS + pALM if the solution accuracy is not a big concern.
- We can also deal with problems whose objective functions involving non-quadratic smooth functions via majorizations.
- To make the algorithms even faster, we often introduce indefinite proximal terms with guaranteed convergence for $\tau \in (0, 2)$.

Extensions (2)

One can deal with TWO nonsmooth blocks plus many smooth blocks: use the sGS decomposition theorem + (indefinite-) semiproximal ADMM (pADMM) (now $\tau \in (0, (1+\sqrt{5})/2)$:

$$\begin{array}{cl} \min\limits_{x \in \mathcal{X}, y \in \mathcal{Y}} & \{p(x_1) + f(x_1, \ldots, x_m) + q(y_1) + g(y_1, \ldots, y_n)\} \\ \text{s.t.} & \sum_{i=1}^{m} \mathcal{A}_i^* x_i + \sum_{j=1}^{n} \mathcal{B}_j^* y_j = c \quad \rightarrow \; \mathcal{A}^* x + \mathcal{B}^* y = c \end{array} \tag{3}$$

$x = (x_1, \ldots, x_m) \in \mathcal{X} := \mathcal{X}_1 \times \ldots \times \mathcal{X}_m$

$y = (y_1, \ldots, y_n) \in \mathcal{Y} := \mathcal{Y}_1 \times \ldots \times \mathcal{Y}_n$

$p : \mathcal{X}_1 \rightarrow (-\infty, \infty]$, $q : \mathcal{Y}_1 \rightarrow (-\infty, \infty]$ are proper closed convex;

$f : \mathcal{X} \rightarrow \Re$, $g : \mathcal{Y} \rightarrow \Re$ are continuously differentiable convex functions with Lipschitz continuous gradients; $\mathcal{A}_i : \mathcal{Z} \rightarrow \mathcal{X}_i$, $\mathcal{B}_j : \mathcal{Z} \rightarrow \mathcal{Y}_i$ are given linear maps.

Note that problem (3) is extremely general: multi-block ADMM is naturally introduced with guaranteed convergence.

Example: QP with Birkhoff polytope constraints

Convex QP:

$$(\mathbf{P}) \quad \min \left\{ \frac{1}{2}\langle X,\, \mathcal{Q}X\rangle + \langle G,\, X\rangle \mid X \in \mathfrak{B}_n \right\},$$

Self-adjoint linear operator $\mathcal{Q} \succeq 0$ and the and Birkhoff polytope:

$$\mathfrak{B}_n := \{X \in \Re^{n\times n} \mid Xe = e,\ X^T e = e, X \geq 0\}$$

$e \in \Re^n$: the vector of all ones.

$$(\mathbf{D}) \min \left\{ \delta^*_{\mathfrak{B}_n}(Z) + \frac{1}{2}\langle W,\, \mathcal{Q}W\rangle \mid Z + \mathcal{Q}W + G = 0,\ W \in \mathrm{Range}(\mathcal{Q}) \right\}$$

$\delta^*_{\mathfrak{B}_n}$: the conjugate of the indicator function $\delta_{\mathfrak{B}_n}$

ALM function for (**D**), given $\sigma > 0$

$$\mathcal{L}_\sigma(Z, W; X) = \delta^*_{\mathfrak{B}_n}(Z) + \frac{1}{2}\langle W,\, \mathcal{Q}W\rangle - \langle X,\, Z + \mathcal{Q}W + G\rangle + \frac{\sigma}{2}\|Z + \mathcal{Q}W + G\|^2$$

ALM for (D)

Algorithm ALM: An augmented Lagrangian method for (D).

Given $\sigma_0 > 0$, iterates $k = 0, 1, \ldots$

Step 1. Compute

$$(Z^{k+1}, W^{k+1}) \approx \operatorname{argmin} \left\{ \begin{array}{l} \Psi_k(Z, W) := \mathcal{L}_{\sigma_k}(Z, W; X^k) \\ \mid (Z, W) \in \Re^{n\times n} \times \text{Range}(\mathcal{Q}) \end{array} \right\}.$$

Step 2. Compute

$$X^{k+1} = X^k - \sigma_k(Z^{k+1} + \mathcal{Q}W^{k+1} + G).$$

Update $\sigma_{k+1} \uparrow \sigma_\infty \leq \infty$.

Convex piecewise linear-quadratic minimization:

error bound holds $\Longrightarrow$ ALM converges asymptotically superlinearly

Semismooth Newton-CG method for inner problem

For any $W \in \text{Range}(\mathcal{Q})$,

$$\psi(W) := \inf_Z \mathcal{L}_\sigma(Z, W; \widehat{X}), \quad Z(W) := \widehat{X} - \sigma(\mathcal{Q}W + G)$$

Subproblem solution $(\overline{Z}, \overline{W})$:

$$\overline{W} = \arg\min \left\{ \psi(W) \mid W \in \text{Range}(\mathcal{Q}) \right\},$$
$$\overline{Z} = \sigma^{-1}\big(Z(\overline{W}) - \Pi_{\mathfrak{B}_n}(Z(\overline{W}))\big)$$

For all $W \in \text{Range}(\mathcal{Q})$,

$$\nabla\psi(W) = \mathcal{Q}W - \mathcal{Q}\Pi_{\mathfrak{B}_n}(Z(W))$$

Semismooth Newton CG solves nonsmooth piecewise affine equation

$$\nabla\psi(W) = 0, \quad W \in \text{Range}(\mathcal{Q}).$$

Semismooth Newton-CG method for inner problem

Given $\widehat{W}$, linear operator $\mathcal{M} : \Re^{n\times n} \to \Re^{n\times n}$

$$\mathcal{M}(\Delta W) := (\mathcal{Q} + \sigma \mathcal{Q} P_{HS} \mathcal{Q})\Delta W, \quad \forall \Delta W \in \Re^{n\times n}$$

P_{HS}: the HS-Jacobian of $\Pi_{\mathfrak{B}_n}$ at $Z(\widehat{W})$

j-th iter., solve linear system (CG)

$$\mathcal{M}_j dW + \nabla\psi(W^j) = 0, \quad dW \in \mathrm{Range}(\mathcal{Q})$$

Global convergence: Line search (using $\psi(W)$)

Local convergence:
positive definiteness of $\mathcal{M}$ on $\mathrm{Range}(\mathcal{Q})$ $\Longrightarrow$ at least superlinear

Given $A, B \in \mathcal{S}^n$, quadratic assignment problem (QAP):

$$\min\{\langle X,\, AXB\rangle \mid X \in \{0,1\}^{n\times n} \cap \mathfrak{B}_n\}$$

Convex relaxation [Anstreicher et al. MP, 2001]:

$$\min\{\langle X,\, \mathcal{Q}X\rangle \mid X \in \mathfrak{B}_n\}$$

Self-adjoint linear operator $\mathcal{Q}(X) := AXB - SX - XT, \quad \mathcal{Q} \succeq 0$

Matrices $S, T \in \mathcal{S}^n$ obtained from [Anstreicher et al. MP, 2001]

Relative KKT residual:

$$\eta = \frac{\|X - \Pi_{\mathfrak{B}_n}(X - \mathcal{Q}X)\|}{1 + \|X\| + \|\mathcal{Q}X\|}$$

Matrices A, B from QAPLIB

Numerical results for QAP

"a": Gurobi, "b": ALM

		iter	η	time
problem	n	a | b (itersub)	a|b	a|b
lipa80a	80	11 | 25 (68)	1.3-6 | 7.3-8	2:46 | 01
lipa90a	90	11 | 20 (54)	2.7-6 | 8.8-8	5:32 | 01
sko100a	100	14 | 26 (95)	8.5-6 | 8.5-8	2:06 | 11
tai100a	100	11 | 18 (52)	1.3-6 | 9.5-8	10:31 | 02
tai100b	100	11 | 27 (98)	1.3-6 | 9.1-8	10:31 | 13
tai80b	80	11 | 27 (98)	1.2-6 | 8.5-8	2:36 | 07
tai256c	256	* | 2 (4)	* | 2.1-16	* | 00
tai150b	150	19 | 27 (94)	4.3-7 | 9.3-8	2:46:17 | 13
tho150	150	16 | 24 (96)	5.6-6 | 9.9-8	18:52 | 22

"*": Gurobi out of memory (128 G RAM)

"tai150b": Gurobi reports error, "small positive term" needed

Collaborators

- NUS: Kim-Chuan Toh; Xin Yee Lam; Meixia Lin; Yancheng Yuan
- Beijing University of Technology: Xinyuan Zhao
- Chinese Academy of Sciences: Chao Ding
- Fudan University: Xudong Li
- University of Minnesota: Ying Cui
- Xinan Jiaotong University: Chengjing Wang
- Nanjing University: Min Li
- Hunan University: Liang Chen
- Dongguan University of Technology: Ning Zhang
- Dalian University of Technology: Liwei Zhang
- Beihang University: Deren Han
- DBS Bank: Kaifeng Jiang
- Alibaba: Liuqin Yang
- There are many more ...

References: Part 1

- Xinyuan Zhao, Defeng Sun, and Kim Chuan Toh, A Newton-CG augmented Lagrangian method for semidefinite programming, SIAM Journal on Optimization 20 (2010) 1737–1765.
- Chengjing Wang, Defeng Sun, and Kim Chuan Toh, Solving log-determinant optimization problems by a Newton-CG proximal point algorithm, SIAM Journal on Optimization 20 (2010) 2994–3013.
- Kaifeng Jiang, Defeng Sun, and Kim Chuan Toh, An inexact accelerated proximal gradient method for large scale linearly constrained convex SDP, SIAM Journal on Optimization 22 (2012) 1042–1064.
- Liuqin Yang, Defeng Sun, and Kim Chuan Toh, SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints, Mathematical Programming Computation Vol. 7, Issue 3 (2015) 331-366.
- Min Li, Defeng Sun, and Kim Chuan Toh, A convergent 3-block semi-proximal ADMM for convex minimization problems with one strongly convex block, Asia-Pacific Journal of Operational Research 32 (2015) 1550024 (19 pages).
- Defeng Sun, Kim Chuan Toh, and Liuqin Yang, A convergent 3-block semi-proximal alternating direction method of multipliers for conic programming with 4-type constraints, SIAM Journal on Optimization Vol. 25, No. 2 (2015) 882-915.

References: Part 2

- Xudong Li, Defeng Sun, and Kim Chuan Toh, A Schur complement based semi-proximal ADMM for convex quadratic conic programming and extensions, Mathematical Programming 155 (2016) 333-373.
- Meixia Lin, Defeng Sun, Kim-Chuan Toh, Efficient algorithms for multivariate shape-constrained convex regression problems, arXiv:2002.11410, February, 2020.
- Min Li, Defeng Sun, and Kim Chuan Toh, A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization, SIAM Journal on Optimization 26 (2016) 922–950.
- Defeng Sun, Kim Chuan Toh, and Liuqin Yang, An efficient inexact ABCD method for least squares semidefinite programming, SIAM Journal on Optimization 26 (2016) 1072–1100.
- Chao Ding, Defeng Sun, and Liwei Zhang, Characterization of the robust isolated calmness for a class of conic programming problems, SIAM Journal on Optimization 27 (2017) 67–90.
- Liang Chen, Defeng Sun, and Kim Chuan Toh, An efficient inexact symmetric Gauss-Seidel based majorized ADMM for high-dimensional convex composite conic programming, Mathematical Programming 161 (2017) 237–270.

- Xin Yee Lam, J.S. Marron, Defeng Sun, and Kim Chuan Toh, Fast algorithms for large scale generalized distance weighted discrimination, Journal of Computational and Graphical Statistics 27 (2018) 368–379.
- Xudong Li, Defeng Sun, and Kim Chuan Toh, QSDPNAL: A two-phase augmented Lagrangian method for convex quadratic semidefinite programming, Mathematical Programming Computation, 10 (2018) 703–743.
- Deren Han, Defeng Sun, and Liwei Zhang, Linear rate convergence of the alternating direction method of multipliers for convex composite programming, Mathematics of Operations Research 43 (2018) 622–637.
- Ying Cui, Defeng Sun, and Kim Chuan Toh, On the R-superlinear convergence of the KKT residuals generated by the augmented Lagrangian method for convex composite conic programming, Mathematical Programming 178 (2019) 381–415.
- Xudong Li, Defeng Sun, and Kim Chuan Toh, A block symmetric Gauss-Seidel decomposition theorem for convex composite quadratic programming and its applications, Mathematical Programming 175 (2019) 395–418.

References: Part 4

- Liang Chen, Defeng Sun, Kim Chuan Toh, Ning Zhang, A unified algorithmic framework of symmetric Gauss-Seidel decomposition based proximal ADMMs for convex composite programming, Journal of Computational Mathematics 37 (2019) 739–757.
- Liang Chen, Xudong Li, Defeng Sun, and Kim Chuan Toh, On the equivalence of inexact proximal ALM and ADMM for a class of convex composite programming, Mathematical Programming 18X (2020).
- Xudong Li, Defeng Sun, and Kim Chuan Toh, On the efficient computation of a generalized Jacobian of the projector over the Birkhoff polytope, Mathematical Programming 179 (2020) 419–446.
- Defeng Sun, Kim Chuan Toh, Yancheng Yuan, Xin-Yuan Zhao, SDPNAL+: A Matlab software for semidefinite programming with bound constraints (version 1.0), Optimization Methods and Software 35 (2020) 87–115.

Thank you for your attention!