Lecture: Algorithms for Compressed Sensing

Zaiwen Wen

Beijing International Center For Mathematical Research
Peking University

http://bicmr.pku.edu.cn/~wenzw/bigdata2018.html
wenzw@pku.edu.cn

Acknowledgement: part of this slides is based on Prof. Lieven Vandenberghe and Prof. Wotao Yin's lecture notes

Outline

- Proximal gradient method
- Complexity analysis of Proximal gradient method
- Alternating direction methods of Multipliers (ADMM)
- Linearized Alternating direction methods of Multipliers
- Check lecture notes at
  `http://bicmr.pku.edu.cn/~wenzw/opt-2017-fall.html`

ℓ_1-regularized least square problem

Consider

$$\min \ \psi_\mu(x) := \mu\|x\|_1 + \frac{1}{2}\|Ax - b\|_2^2$$

Approaches:

- Interior point method: l1_ls
- Spectral gradient method: GPSR
- Fixed-point continuation method: FPC
- Active set method: FPC_AS
- Alternating direction augmented Lagrangian method
- Nesterov's optimal first-order method
- many others

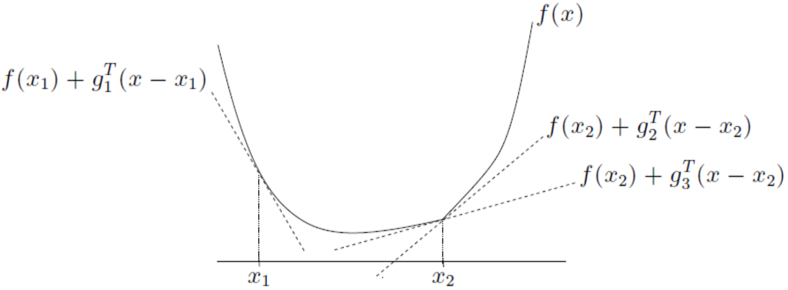
Subgradient

recall basic inequality for convex differentiable f:

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x).$$

g is a subgradient of a convex function f at $x \in \mathbf{dom} f$ if

$$f(y) \geq f(x) + g^\top (y - x), \forall y \in \mathbf{dom} f.$$

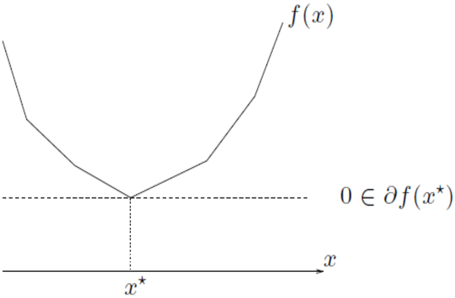


g_2, g_3 are subgradients at x_2, g_1 is a subgradient at x_1.

Optimality conditions — unconstrained

x^* minimizes $f(x)$ if and only

$$0 \in \partial f(x^*)$$



Proof: by definition

$$f(y) \geq f(x^*) + 0^\top (y - x^*) \text{ for all } y \quad \Longleftrightarrow 0 \in \partial f(x^*).$$

Optimality conditions — constrained

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0, i = 1, \ldots, m. \end{aligned}$$

From **Lagrange duality**: if strong duality holds, then x^*, λ^* are primal, dual optimal if and only if

- x^* is primal feasible
- $\lambda^* \ge 0$
- complementary: $\lambda_i^* f_i(x^*) = 0$ for $i = 1, \ldots, m$
- x^* is a minimizer of $\min \ \mathcal{L}(x, \lambda^*) = f_0(x) + \sum_i \lambda_i^* f_i(x)$, i.e.,

$$0 \in \partial_x \mathcal{L}(x, \lambda^*) = \partial f_0(x^*) + \sum_i \lambda_i^* \partial f_i(x^*)$$

Proximal Gradient Method

Let $f(x) = \frac{1}{2}\|Ax - b\|_2^2$. The gradient $\nabla f(x) = A^\top(Ax - b)$. Consider

$$\min \ \psi_\mu(x) := \mu\|x\|_1 + f(x).$$

- First-order approximation + proximal term:

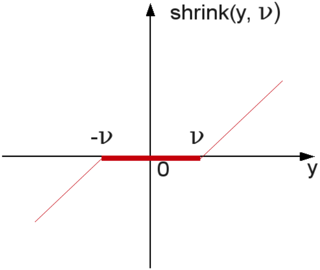
$$\begin{aligned} x^{k+1} &:= \arg\min_{x\in\mathbb{R}^n} \ \mu\|x\|_1 + (\nabla f(x^k))^\top(x - x^k) + \frac{1}{2\tau}\|x - x^k\|_2^2 \\ &= \arg\min_{x\in\mathbb{R}^n} \ \mu\|x\|_1 + \frac{1}{2\tau}\|x - (x^k - \tau\nabla f(x^k))\|_2^2 \\ &= \text{shrink}(x^k - \tau\nabla f(x^k), \mu\tau) \end{aligned}$$

- gradient step: bring in candidates for nonzero components
- shrinkage step: eliminate some of them by “soft” thresholding

Shrinkage (soft thresholding)

$$\text{shrink}(y,\nu) := \arg\min_{x\in\mathbb{R}} \; \nu\|x\|_1 + \frac{1}{2}\|x-y\|_2^2$$
$$= \text{sgn}(y)\max(|y|-\nu,0)$$
$$= \begin{cases} y-\nu\,\text{sgn}(y), & \text{if } |y|>\nu \\ 0, & \text{otherwise} \end{cases}$$

- Chambolle, Devore, Lee and Lucier
- Figueirdo, Nowak and Wright
- Elad, Matalon and Zibulevsky
- Hales, Yin and Zhang
- Darbon, Osher
- Many others



Proximal gradient method For General Problems

Consider the model

$$\min \; F(x) := f(x) + r(x)$$

- $f(x)$ is convex, differentiable
- $r(x)$ is convex but may be nondifferentiable

General scheme: linearize $f(x)$ and add a proximal term:

$$\begin{aligned} x^{k+1} &:= \arg\min_{x\in\mathbb{R}^n} \; r(x) + (\nabla f(x^k))^\top (x - x^k) + \frac{1}{2\tau}\|x - x^k\|_2^2 \\ &= \arg\min_{x\in\mathbb{R}^n} \; \tau r(x) + \frac{1}{2}\|x - (x^k - \tau\nabla f(x^k))\|_2 \\ &= \mathrm{prox}_{\tau r}(x^k - \tau\nabla f(x^k)) \end{aligned}$$

Proximal Operator

$$\mathrm{prox}_r(y) := \arg\min_x \; r(x) + \frac{1}{2}\|x - y\|_2^2.$$

Proximal mapping

if g is convex and closed (has a closed epigraph), then

$$\text{prox}_r(y) := \arg\min_x \; r(x) + \frac{1}{2}\|x-y\|_2^2$$

exists and is unique for all y.

- from optimality conditions of minimization in the definition:

$$x = \text{prox}_r(y) \quad \Longleftrightarrow \quad y - x \in \partial r(x)$$
$$\Longleftrightarrow \quad r(z) \geq r(x) + (y-x)^\top (z-x), \quad \forall z$$

subdifferential of a convex function is a monotone operator:

$$(u-v)^\top (x-y) \geq 0 \quad \forall x, y, u \in \partial f(x), v \in \partial f(y)$$

proof: combining the following two inequalities

$$f(y) \geq f(x) + u^\top (y-x), \quad f(x) \geq f(y) + v^\top (x-y)$$

Nonexpansiveness

if $u = \text{prox}_r(x)$, $v = \text{prox}_r(y)$, then

$$(u - v)^\top (x - y) \geq \|u - v\|^2$$

prox_r is is firmly nonexpansive, or co-coercive with constant 1

- follows from last page and monotonicity of subgradients

$$x - u \in \partial r(u), y - v \in \partial r(v) \Longrightarrow (x - u - y + v)^\top (u - v) \geq 0$$

- implies (from Cauchy-Schwarz inequality)

$$\|\text{prox}_r(x) - \text{prox}_r(y)\|_2 \leq \|x - y\|_2$$

Global convergence

- The proximal mapping is nonexpansive
- Under certain assumptions, $h(x) = x - \tau \nabla f(x)$ is nonexpansive:

$$\|h(x) - h(x')\| \leq \|x - x'\|,$$

  and $r(x) = r(x')$ whenever the equality holds.
- x is a fixed point if

$$\|\mathrm{prox}_{\tau r}(h(x)) - \mathrm{prox}_{\tau r}(h(x^*))\| \equiv \|\mathrm{prox}_{\tau r}(h(x)) - x^*\| = \|x - x^*\|$$

Proximal gradient method for ℓ_1-minimization

- Advantages
  - Low computational cost (first-order method)
  - Finite convergence of the nondegenerate support
  - Global q-linear convergence
- Disadvantages
  - Takes a lot of steps when the support is relatively large
  - Slow convergence: degenerate support identification or magnitude recovery?
- Strategies for improving shrinkage
  - Adjust λ dynamically (use λ^k)
  - Line search $x^{k+1} = x^k + \alpha^k(\text{shrink}(x^k - \lambda^k \nabla f^k, \mu\lambda^k) - x^k)$
  - Continuation: approximately solve for $\mu_0 > \mu_1 > \cdots > \mu_p = \mu$ (FPC)
  - Debiasing or subspace optimization

Line search approach

Let $x^k(\tau^k) = \text{shrink}(x^k - \tau^k \nabla f^k, \mu\tau^k)$. Then set

$$\begin{aligned} x^{k+1} &= x^k + \alpha^k (x^k(\tau^k) - x^k) \\ &= x^k + \alpha^k d^k \end{aligned}$$

- Choosing τ^k: Barzilai-Borwein method
  - $\nabla f^k = A^\top (Ax^k - b)$
  - $s^{k-1} = x^k - x^{k-1}$ and $y^{k-1} = \nabla f^k - \nabla f^{k-1}$
  - $\tau^{k,BB1} = \frac{(s^{k-1})^\top s^{k-1}}{(s^{k-1})^\top y^{k-1}}$ or $\tau^{k,BB2} = \frac{(s^{k-1})^\top y^{k-1}}{(y^{k-1})^\top y^{k-1}}$.
- Choosing α^k: Armijo-like Line search

$$\psi_\mu(x^k + \alpha^k d^k) \le C^k + \sigma \alpha^k \Delta^k$$

  - FPC: $\Delta^k := (\nabla f^k)^\top d^k$
  - FPC_AS: $\Delta^k := (\nabla f^k)^\top d^k + \mu \|x^k(\tau^k)\|_1 - \mu \|x^k\|_1$
  - $C^k = (\eta Q^{k-1} C^{k-1} + \psi_\mu(x^k))/Q^k$, $Q^k = \eta Q^{k-1} + 1$, $C^0 = \psi_\mu(x^0)$ and $Q^0 = 1$ (Zhang and Hager)

Outline: Complexity Analysis

- Amir Beck and Marc Teboulle, *A fast iterative shrinkage thresholding algorithm for linear inverse problems*
- Paul Tseng, *On accelerated proximal gradient methods for convex-concave optimization*
- Paul Tseng, *Approximation accuracy, gradient methods and error bound for structured convex optimization*
- N. S. Aybat and G. Iyengar, *A first-order augmented Lagrangian method for compressed sensing*

Proximal Gradient/ISTA/FPC

The following terminologies refer to the same algorithm:

- proximal gradient method
- ISTA: iterative shrinkage thresholding algorithm
- FPC: fixed-point continuation method

Consider

$$\min \ F(x) := \mu\|x\|_1 + \frac{1}{2}\|Ax - b\|_2^2$$

ISTA/FPC computes

$$\begin{aligned} x_{k+1} &:= \arg\min \mu\|x\|_1 + (\nabla f_k)^\top (x - x_k) + \frac{1}{2\tau}\|x - x_k\|_2^2 \\ &= \text{shrink}(x_k - \tau\nabla f_k, \mu\tau) \end{aligned}$$

- Complexity?
- How can we speed it up?

Generalization of ISTA

Consider a solvable model

$$\min \ F(x) := f(x) + r(x)$$

- $r(x)$ continuous convex, may be nonsmooth
- $f(x)$ continuously diff. with Lipschitz continuous gradient

$$\|\nabla f(x) - \nabla f(y)\|_2 \le L_f \|x - y\|_2$$

- The cost of the following problem is tractable

$$\begin{aligned} p_L(y) &:= \arg\min_x \ Q_L(x,y) := f(y) + \langle \nabla f(y), x - y \rangle + \frac{L}{2}\|x-y\|_2^2 + r(x) \\ &= \arg\min_x \ r(x) + \frac{L}{2}\left\| x - \left(y - \frac{1}{L}\nabla f(y) \right) \right\|_2^2 \\ &= \mathrm{prox}_{1/L}(y - \frac{1}{L}\nabla f(y)) \end{aligned}$$

ISTA

Basic iteration: $x_k = p_L(x_{k-1})$

Quadratic upper bound

Suppose ∇f is Lipschitz continuous with parameter L and $\mathbf{dom} f$ is convex. Then

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2}\|y - x\|_2^2, \quad \forall x, y.$$

Proof: Let $v = y - x$. Then

$$\begin{aligned}
f(y) &= f(x) + \nabla f(x)^\top v + \int_0^1 (\nabla f(x + tv) - \nabla f(x))^\top v dt \\
&\leq f(x) + \nabla f(x)^\top v + \int_0^1 \|\nabla f(x + tv) - \nabla f(x)\|_2 \|v\|_2 dt \\
&\leq f(x) + \nabla f(x)^\top v + \int_0^1 Lt\|v\|_2^2 dt \\
&= f(x) + \nabla f(x)^\top v + \frac{L}{2}\|y - x\|_2^2
\end{aligned}$$

The above inequality also implies that $F(p_L(y)) \leq Q(p_L(y), y)$.

Lemma: Assume $F(p_L(y)) \le Q(p_L(y), y)$, then for any $x \in \mathbb{R}^n$,

$$\begin{aligned} F(x) - F(p_L(y)) &\ge \frac{L}{2}\|p_L(y) - y\|^2 + L\langle y - x, p_L(y) - y\rangle \\ &= \frac{L}{2}\left(\|p_L(y) - x\|_2^2 - \|x - y\|_2^2\right) \end{aligned}$$

- **Proof**: $\nabla f(y) + L(p_L(y) - y) + z = 0$ for some $z \in \partial r(p_L(y))$.
- The convexity of f and g gives

$$\begin{aligned} f(x) &\ge f(y) + \nabla f(y)^\top (x - y) \\ r(x) &\ge r(p_L(y)) + z^\top (x - p_L(y)). \end{aligned}$$

- Using the definition of $Q(p_L(y), y)$ and z

$$\begin{aligned} F(x) - F(p_L(y)) &\ge F(x) - Q(p_L(y), y) \\ &\ge -\frac{L}{2}\|p_L(y) - y\|^2 + (x - p_L(y))^\top (\nabla f(y) + z) \end{aligned}$$

Complexity analysis of ISTA

Complexity results: $F(x_k) - F(x^*) \le \frac{L_f \|x_0 - x^*\|_2^2}{2k}$

- Let $x = x^*$, $y = x_j$, then

$$\frac{2}{L}(F(x^*) - F(x_{j+1})) \ge \|x^* - x_{j+1}\|_2^2 - \|x^* - x_j\|_2^2$$

Summing over $j := 0, \cdots, k-1$ gives

$$\frac{2}{L}(kF(x^*) - \sum_{j=0}^{k-1} F(x_{j+1})) \ge \|x^* - x^k\|_2^2 - \|x^* - x_0\|_2^2$$

- Let $x = y = x_j$ yields

$$\tfrac{2}{L}(F(x_j) - F(x_{j+1})) \ge \|x_j - x_{j+1}\|_2^2$$

Multiplying the last ineq. by j and summing over $j := 0, \cdots, k-1$ gives

$$\tfrac{2}{L}(-kF(x_k) + \textstyle\sum_{j=0}^{k-1} F(x_{j+1})) \ge \textstyle\sum_{j=0}^{k-1} j\|x_j - x_{j+1}\|_2^2$$

- Therefore

$$\tfrac{2k}{L}(F(x^*) - F(x_k)) \ge \|x^* - x^k\|_2^2 + \textstyle\sum_{j=0}^{k-1} j\|x_j - x_{j+1}\|_2^2 - \|x^* - x_0\|_2^2$$

FISTA: Accelerated proximal gradient

Given $y^1 = x_0$ and $t^1 = 1$, compute:

$$\begin{aligned} x_k &= p_L(y_k) \\ t_{k+1} &= \frac{1+\sqrt{1+4t_k^2}}{2} \\ y_{k+1} &= x_k + \frac{t_k - 1}{t_{k+1}}(x_k - x_{k-1}) \end{aligned}$$

Complexity results:

$$F(x_k) - F(x^*) \leq \frac{2L_f\|x_0 - x^*\|_2^2}{(k+1)^2}$$

Lipschitz constant L_f is unknown?
Choose L by backtracking so that $F(p_L(y_k)) \leq Q_L(p_L(y_k), y_k)$

Complexity analysis of FISTA

- Let $v_k := F(x_k) - F(x^*)$ and $u_k := t_k x_k - (t_k - 1)x_{k-1} - x^*$,

$$\frac{2}{L}\left(t_k^2 v_k - t_{k+1}^2 v_{k+1}\right) \geq \|u_{k+1}\|_2^2 - \|u_k\|_2^2$$

- If $a_k - a_{k+1} \geq b_{k+1} - b_k$, with $a_1 + b_1 \leq c$, then $a_k \leq c$
- $t_k \geq (k+1)/2$
- Let $a_k = \frac{2}{L} t_k^2 v_k$, $b_k = \|u_k\|_2^2$, and $c = \|x_0 - x^*\|_2^2$.
- Check $a_1 + b_1 \leq c$:

$$\begin{aligned}
F(x^*) - F(x_1) &= F(x^*) - F(p_L(y_1)) \\
&\geq \frac{L}{2}\|p_L(y_1) - y_1\|_2^2 + L\langle y_1 - x^*, p_L(y_1) - y_1\rangle \\
&= \frac{L}{2}(\|x_1 - x^*\|_2^2 - \|y_1 - x^*\|_2^2)
\end{aligned}$$

Complexity analysis of FISTA

- Let ($x = x_k$, $y = y_{k+1}$) and ($x = x_*$, $y = y_{k+1}$). Note $x_{k+1} = p_{L_{k+1}}(y_{k+1})$:

$$2L^{-1}(v_k - v_{k+1}) \geq \|x_{k+1} - y_{k+1}\|_2^2 + 2\langle x_{k+1} - y_{k+1}, y_{k+1} - x_k \rangle \quad (1)$$

$$-2L^{-1}v_{k+1} \geq \|x_{k+1} - y_{k+1}\|_2^2 + 2\langle x_{k+1} - y_{k+1}, y_{k+1} - x_* \rangle \quad (2)$$

- $((1) * (t_{k+1} - 1) + (2)) * t_{k+1}$ and using $t_k^2 = t_{k+1}^2 - t_{k+1}$:

$$\frac{2t_{k+1}}{L}((t_{k+1} - 1)v_k - t_{k+1}v_{k+1}) = \frac{2}{L}(t_k^2 v_k - t_{k+1}^2 v_{k+1})$$

$$\geq \|t_{k+1}(x_{k+1} - y_{k+1})\|_2^2 + 2t_{k+1}\langle x_{k+1} - y_{k+1}, t_{k+1}y_{k+1} - (t_{k+1} - 1)x_k - x^* \rangle$$

- $\|b - a\|_2^2 + 2\langle b - a, a - c \rangle = \|b - c\|_2^2 - \|a - c\|_2^2$
- $y_{k+1} = x_k + \frac{t_k - 1}{t_{k+1}}(x_k - x^{k-1})$:

$$\frac{2}{L}(t_k^2 v_k - t_{k+1}^2 v_{k+1})$$

$$\geq \|t_{k+1}x_{k+1} - (t_{k+1} - 1)x_k - x^*\|_2^2 - \|t_{k+1}y_{k+1} - (t_{k+1} - 1)x_k - x^*\|_2^2$$

$$= \|u_{k+1}\|_2^2 - \|u_k\|_2^2$$

Proximal gradient method

Consider the model

$$\min \ F(x) := f(x) + r(x)$$

Linearize $f(x)$:

$$\ell_f(x,y) := f(y) + \langle \nabla f(y), x - y \rangle + r(x)$$

Given a strictly convex function $h(x)$, Bregman distance:

$$D(x,y) := h(x) - h(y) - \langle \nabla h(y), x - y \rangle .$$

For example, take $h(x) = \|x\|_2^2$. Then $D(x,y) = \|x - y\|_2^2$.
Proximal gradient method can also be written as

$$x_{k+1} := \arg\min_x \ \ell_f(x, x^k) + \frac{L}{2} D(x, x^k)$$

APG Variant 1

Acclerated proximal gradient (APG):
Set $x_{-1} = x_0$ and $\theta_{-1} = \theta_0 = 1$:

$$\begin{aligned} y_k &= x_k + \theta_k(\theta_{k-1}^{-1} - 1)(x_k - x_{k-1}) \\ x_{k+1} &= \arg\min_x \ \ell_f(x, y^k) + \frac{L}{2}\|x - y_k\|_2^2 \\ \theta_{k+1} &= \frac{\sqrt{\theta_k^4 + 4\theta_k^2} - \theta_k^2}{2} \end{aligned}$$

Question: what is the difference between θ_k and t_k?

APG Variant 2

Replace $\frac{1}{2}\|x - y^k\|_2^2$ by Bregman distance $D(x, y_k)$
Set x_0, z_0 and $\theta_0 = 1$:

$$\begin{aligned}
y_k &= (1-\theta_k)x_k + \theta_k z_k \\
z_{k+1} &= \arg\min_x \; \ell_f(x, y^k) + \theta_k L D(x, z_k) \\
x_{k+1} &= (1-\theta_k)x_k + \theta_k z_{k+1} \\
\theta_{k+1} &= \frac{\sqrt{\theta_k^4 + 4\theta_k^2} - \theta_k^2}{2}
\end{aligned}$$

APG Variant 3

Set x_0, $z_0 := \arg\min h(x)$ and $\theta_0 = 1$:

$$
\begin{aligned}
y_k &= (1-\theta_k)x_k + \theta_k z_k \\
z_{k+1} &= \arg\min_x \sum_{i=0}^{k} \frac{\ell_f(x, y^i)}{\theta_i} + Lh(x) \\
x_{k+1} &= (1-\theta_k)x_k + \theta_k z_{k+1} \\
\theta_{k+1} &= \frac{\sqrt{\theta_k^4 + 4\theta_k^2} - \theta_k^2}{2}
\end{aligned}
$$

Complexity analysis

- Proximal gradient method

$$F(x_k) \le F(x) + \frac{1}{k} LD(x, x^0)$$

- APG1:

$$F(x_k) \le F(x) + \frac{4}{(k+1)^2} LD(x, x^0)$$

- APG2:

$$F(x_k) \le F(x) + \frac{4}{(k+1)^2} LD(x, z^0)$$

- APG3:

$$F(x_k) \le F(x) + \frac{4}{(k+1)^2} L(h(x) - h(x_0))$$

Complexity analysis

Let $f(x)$ is convex and assume $\nabla f(x)$ is Lipschitz cont.

$$F(x) \geq \ell_f(x,y) \geq F(x) - \frac{L}{2}\|x-y\|_2^2.$$

For any proper lsc convex function $\psi(x)$, if $z_+ = \arg\min_x \ \psi(x) + D(x,z)$ and $h(x)$ is differentiable at z_+, then

$$\psi(x) + D(x,z) \geq \psi(z_+) + D(x,z_+) + D(z_+,z).$$

Proof: The optimality at z_+ and definition of subgradient gives

$$\psi(x) + \nabla_x D(z_+,z)^\top (x - z_+) \geq \psi(z_+).$$

Note $\nabla_x D(z_+,z) = \nabla h(z_+) - \nabla h(z)$. Rearranging terms yields

$$\psi(x) - \nabla h(z)^\top (x-z) \geq \psi(z_+) - \nabla h(z)^\top (z_+ - z) - \nabla h(z_+)^\top (x - z_+).$$

Adding $h(x) - h(z)$ to both sides.

Complexity analysis of APG2

$$
\begin{aligned}
& F(x_{k+1}) \\
\leq \;& \ell_f(x_{k+1}, y^k) + \frac{L}{2}\|x_{k+1} - y_k\|_2^2 \\
= \;& \ell_f((1-\theta_k)x_k + \theta z_{k+1}, y_k) + \frac{L\theta_k^2}{2}\|z_{k+1} - z_k\|_2^2 \\
\leq \;& (1-\theta_k)\ell_f(x_k, y_k) + \theta_k \ell_f(z_{k+1}, y_k) + \theta_k^2 LD(z_{k+1}, z_k) \\
\leq \;& (1-\theta_k)\ell_f(x_k, y_k) + \theta_k \left(\ell_f(x, y_k) + \theta_k LD(x, z_k) - \theta_k LD(x, z_{k+1})\right) \\
\leq \;& (1-\theta_k)F(x_k) + \theta_k F(x) + \theta_k^2 LD(x, z_k) - \theta_k^2 LD(x, z_{k+1})
\end{aligned}
$$

The third inequality applies the inequality for $\ell_f(x,z) + \theta LD(x,z)$.

Let $e_k = F(x_k) - F(x)$ and $\Delta_k = LD(x, z_k)$,then

$$e_{k+1} \leq (1-\theta_k)e_k + \theta_k^2 \Delta_k - \theta_k^2 \Delta_{k+1}$$

Complexity analysis of APG2

Divide both sides by θ_k^2 and using $\frac{1}{\theta_k^2} = \frac{1-\theta_{k+1}}{\theta_{k+1}^2}$:

$$\frac{1-\theta_{k+1}}{\theta_{k+1}^2} e_{k+1} + \Delta_{k+1} \le \frac{1-\theta_k}{\theta_k^2} e_k + \Delta_k$$

Hence

$$\frac{1-\theta_{k+1}}{\theta_{k+1}^2} e_{k+1} + \Delta_{k+1} \le \frac{1-\theta_0}{\theta_0^2} e_0 + \Delta_0$$

Using $\frac{1}{\theta_k^2} = \frac{1-\theta_{k+1}}{\theta_{k+1}^2}$ and $\theta_0 = 1$:

$$\frac{1}{\theta_k^2} e_{k+1} \le \Delta_0 - \Delta_{k+1} \le \Delta_0 = LD(x, z_0)$$

An application to Basis Pursuit problem

Consider

$$\min \ \|x\|_1 \ \text{ s.t. } Ax = b$$

Augmented Lagragian (Bregman) framework ($b_1 = b$, $k = 1$):

$$\begin{aligned} x_k^* &:= \arg\min_x \ F_k(x) := \mu_k \|x\|_1 + \frac{1}{2}\|Ax - b_k\|_2^2 \\ b_{k+1} &:= b + \frac{\mu_{k+1}}{\mu_k}(b_k - Ax_k) \end{aligned}$$

- obtaining x_{k+1}^* exactly is difficult
- inexact solver: how do we control the accuracy?
- Analysis of Bregman approach (see wotao), no complexity
- Solving each x_{k+1}^* by the APG algorithms?

Outline: ADMM

- Alternating direction augmented Lagrangian methods
- Variable splitting method
- Convergence for problems with two blocks of variables

References

- Wotao Yin, Stanley Osher, Donald Goldfarb, Jerome Darbon, *Bregman Iterative Algorithms for l1-Minimization with Applications to Compressed Sensing*
- Junfeng Yang, Yin Zhang, *Alternating direction algorithms for l1-problems in Compressed Sensing*
- Tom Goldstein, Stanely Osher, *The Split Bregman Method for L1-Regularized Problems*
- B.S. He, H. Yang, S.L. Wang, *Alternating Direction Method with Self-Adaptive Penalty Parameters for Monotone Variational Inequalities*

Basis pursuit problem

$$\begin{aligned} \text{Primal:} \quad & \min \quad \|x\|_1, \;\; \text{s.t. } Ax = b \\ \text{Dual:} \quad & \max \quad b^\top \lambda, \;\; \text{s.t. } \|A^\top \lambda\|_\infty \le 1 \end{aligned}$$

The dual problem is equivalent to

$$\max \; b^\top \lambda, \;\; \text{s.t. } A^\top \lambda = s, \; \|s\|_\infty \le 1.$$

Augmented Lagrangian (Bregman) framework

Augmented Lagrangian function:

$$\mathcal{L}(\lambda, s, x) := -b^\top \lambda + x^\top (A^\top \lambda - s) + \frac{1}{2\mu}\|A^\top \lambda - s\|^2$$

Algorithmic framework

- Compute λ^{k+1} and s^{k+1} at k-th iteration

$$\text{(DL)} \quad \min_{\lambda,s} \; \mathcal{L}(\lambda, s, x^k), \quad \text{s.t. } \|s\|_\infty \le 1$$

- Update the Lagrangian multiplier:

$$x^{k+1} = x^k + \frac{A^\top \lambda^{k+1} - s^{k+1}}{\mu}$$

Pros and Cons:

- Pros: rich theory, well understood and a lot of algorithms
- Cons: $\mathcal{L}(\lambda, s, x^k)$ is not separable in λ and s, and the subproblem (DL) is difficult to minimize

An alternating direction minimization scheme

- Divide variables into different blocks according to their roles
- Minimize the augmented Lagrangian function with respect to one block at a time while all other blocks are fixed

ADMM

$$
\begin{aligned}
\lambda^{k+1} &= \arg\min_{\lambda} \ \mathcal{L}(\lambda, s^k, x^k) \\
s^{k+1} &= \arg\min_{s} \ \mathcal{L}(\lambda^{k+1}, s, x^k), \ \text{ s.t. } \|s\|_\infty \le 1 \\
x^{k+1} &= x^k + \frac{A^\top \lambda^{k+1} - s^{k+1}}{\mu}
\end{aligned}
$$

An alternating direction minimization scheme

Explicit solutions:

$$
\begin{aligned}
\lambda^{k+1} &= (AA^\top)^{-1}\left(\mu(Ax^k - b) + As^k\right) \\
s^{k+1} &= \arg\min \ \|s - A^\top\lambda^{k+1} - \mu x^k\|^2, \ \ \text{s.t. } \|s\|_\infty \le 1 \\
&= \mathcal{P}_{[-1,1]}(A^\top\lambda^{k+1} + \mu x^k) \\
x^{k+1} &= x^k + \frac{A^\top\lambda^{k+1} - s^{k+1}}{\mu}
\end{aligned}
$$

ADMM for BP-denoising

Primal:

$$\min \ \|x\|_1, \text{ s.t. } \|Ax - b\|_2 \leq \sigma$$

which is equivalent to

$$\min \ \|x\|_1, \text{ s.t. } Ax - b + r = 0, \ \|r\|_2 \leq \sigma$$

Lagrangian function:

$$\begin{aligned} \mathcal{L}(x, r, \lambda) &:= \|x\|_1 - \lambda^\top (Ax - b + r) + \pi(\|r\|_2 - \sigma) \\ &= \|x\|_1 - (A^\top \lambda)^\top x + \pi\|r\|_2 - \lambda^\top r + b^\top \lambda - \pi\sigma \end{aligned}$$

Hence, the dual problem is:

$$\max b^\top \lambda - \pi\sigma, \ \text{ s.t. } \|A^\top \lambda\|_\infty \leq 1, \ \|\lambda\|_2 \leq \pi$$

ADMM for BP-denoising

The dual problem is equivalent to:

$$\max b^\top \lambda - \pi\sigma, \ \ \text{s.t. } A^\top \lambda = s, \ \|s\|_\infty \le 1, \ \|\lambda\|_2 \le \pi$$

Augmented Lagrangian function is:

$$\mathcal{L}(\lambda, s, x) := -b^\top \lambda + \pi\sigma + x^\top (A^\top \lambda - s) + \frac{1}{2\mu}\|A^\top \lambda - s\|^2$$

ADMM scheme:

$$\begin{aligned}
\lambda^{k+1} &= \arg\min \ \frac{1}{2\mu}\|A^\top \lambda - s^k\|^2 + (Ax^k - b)^\top \lambda, \ \ \text{s.t. } \|\lambda\|_2 \le \pi^k \\
s^{k+1} &= \arg\min \ \|s - A^\top \lambda^{k+1} - \mu x^k\|^2, \ \ \text{s.t. } \|s\|_\infty \le 1 \\
&= \mathcal{P}_{[-1,1]}(A^\top \lambda^{k+1} + \mu x^k) \\
\pi^{k+1} &= \|\lambda^{k+1}\|_2 \\
x^{k+1} &= x^k + \frac{A^\top \lambda^{k+1} - s^{k+1}}{\mu}
\end{aligned}$$

ADMM for ℓ_1-regularized problem

Primal:

$$\min \ \mu\|x\|_1 + \frac{1}{2}\|Ax - b\|_2^2$$

which is equivalent to

$$\min \ \mu\|x\|_1 + \frac{1}{2}\|r\|_2^2, \ \text{ s.t. } Ax - b = r.$$

Lagrangian function:

$$\begin{aligned} \mathcal{L}(x, r, \lambda) \ &:= \ \mu\|x\|_1 + \frac{1}{2}\|r\|_2^2 - \lambda^\top(Ax - b - r) \\ &= \ \mu\|x\|_1 - (A^\top\lambda)^\top x + \frac{1}{2}\|r\|_2^2 + \lambda^\top r + b^\top\lambda \end{aligned}$$

Hence, the dual problem is:

$$\max b^\top\lambda - \frac{1}{2}\|\lambda\|^2, \ \text{ s.t. } \|A^\top\lambda\|_\infty \le \mu$$

ADMM for ℓ_1-regularized problem

The dual problem is equivalent to

$$\max\ b^\top \lambda - \frac{1}{2}\|\lambda\|^2, \ \ \text{s.t. } A^\top \lambda = s, \ \|s\|_\infty \le \mu.$$

Augmented Lagrangian function is:

$$\mathcal{L}(\lambda, s, x) := -b^\top \lambda + \frac{1}{2}\|\lambda\|^2 + x^\top (A^\top \lambda - s) + \frac{1}{2\mu}\|A^\top \lambda - s\|^2$$

ADMM scheme:

$$\begin{aligned}
\lambda^{k+1} &= (AA^\top + \mu I)^{-1}\left(\mu(Ax^k - b) + As^k\right) \\
s^{k+1} &= \arg\min\ \|s - A^\top \lambda^{k+1} - \mu x^k\|^2, \ \ \text{s.t. } \|s\|_\infty \le \mu \\
&= \mathcal{P}_{[-\mu,\mu]}(A^\top \lambda^{k+1} + \mu x^k) \\
x^{k+1} &= x^k + \frac{A^\top \lambda^{k+1} - s^{k+1}}{\mu}
\end{aligned}$$

Derive ADMM for the following problems:

$$
\begin{aligned}
\text{BP:} \quad & \min_{x\in\mathbb{C}^n} \quad \|Wx\|_{w,1}, \quad \text{s.t. } Ax = b \\
\text{L1/L1:} \quad & \min_{x\in\mathbb{C}^n} \quad \|Wx\|_{w,1} + \frac{1}{\nu}\|Ax-b\|_1 \\
\text{L1/L2:} \quad & \min_{x\in\mathbb{C}^n} \quad \|Wx\|_{w,1} + \frac{1}{2\rho}\|Ax-b\|_2^2 \\
\text{BP+:} \quad & \min_{x\in\mathbb{R}^n} \quad \|x\|_{w,1}, \quad \text{s.t. } Ax = b,\ x \geq 0 \\
\text{L1/L1+:} \quad & \min_{x\in\mathbb{R}^n} \quad \|x\|_{w,1} + \frac{1}{\nu}\|Ax-b\|_1, \quad \text{s.t. } x \geq 0 \\
\text{L1/L2+:} \quad & \min_{x\in\mathbb{R}^n} \quad \|x\|_{w,1} + \frac{1}{2\rho}\|Ax-b\|_2^2, \quad \text{s.t. } x \geq 0
\end{aligned}
$$

$\nu, \rho \geq 0$, $A \in \mathbb{C}^{m\times n}$, $b \in \mathbb{C}^m$,$x \in \mathbb{C}^n$ for the first three and $x \in \mathbb{R}^n$ for the last three, $W \in \mathbb{C}^{n\times n}$ is an unitary matrix serving as a sparsifying basis, and $\|x\|_{w,1} := \sum_{i=1}^n w_i|x_i|$.

Variable splitting

Given $A \in \mathbb{R}^{m\times n}$, consider $\min\ f(x) + g(Ax)$, which is

$$\min\ f(x) + g(y), \quad \text{s.t. } Ax = y$$

Augmented Lagrangian function:

$$\mathcal{L}(x, y, \lambda) = f(x) + g(y) - \lambda^\top (Ax - y) + \frac{1}{2\mu}\|Ax - y\|_2^2$$

ADMM

$$\begin{aligned}
(P_x):\quad & x^{k+1} := \arg\min_{x\in\mathcal{X}} \quad \mathcal{L}(x, y^k, \lambda^k), \\
(P_y):\quad & y^{k+1} := \arg\min_{y\in\mathcal{Y}} \quad \mathcal{L}(x^{k+1}, y, \lambda^k), \\
(P_\lambda):\quad & \lambda^{k+1} := \lambda^k - \gamma\frac{Ax^{k+1} - y^{k+1}}{\mu}
\end{aligned}$$

Variable splitting

split Bregman (Goldstein and Osher) for anisotropic TV:

$$\min \ \alpha\|Du\|_1 + \beta\|\Psi u\|_1 + \frac{1}{2}\|Au - f\|_2^2$$

Introduce $y = Du$ and $w = \Psi u$, obtain

$$\min \ \alpha\|y\|_1 + \beta\|w\|_1 + \frac{1}{2}\|Au - f\|_2^2, \ \text{ s.t. } y = Du, \ w = \Psi u$$

Augmented Lagrangian function:

$$\begin{aligned} \mathcal{L} \ := \ & \alpha\|y\|_1 + \beta\|w\|_1 + \frac{1}{2}\|Au - f\|_2^2 - p^\top(Du - y) + \frac{1}{2\mu}\|Du - y\|_2^2 \\ & -q^\top(\Psi u - w) + \frac{1}{2\mu}\|\Psi u - w\|_2^2 \end{aligned}$$

Variable splitting

- The variable u can be otained by

$$\left(A^\top A + \frac{1}{\mu}(D^\top D + I)\right) u = A^\top f + \frac{1}{\mu}(D^\top y + \Psi^\top w) + D^\top p + \Psi^\top q$$

If A and D are diagonalizable by FFT, then the computational cost is very cheap. For example, $A = R\mathcal{F}$, both R and D are circulant matrices.

- Variables y and w:

$$\begin{aligned} y &:= \operatorname{shrink}(Du - \mu p, \alpha\mu) \\ w &:= \operatorname{shrink}(\Psi u - \mu q, \alpha\mu) \end{aligned}$$

- apply a few iterations before updating the Lagrangian multipliers p and q

Exercise: isotropic TV

$$\min \ \alpha \|Du\|_2 + \beta \|\Psi u\|_1 + \frac{1}{2}\|Au - f\|_2^2$$

FTVd: Fast TV deconvolution

Wang-Yang-Yin-Zhang consider:

$$\min_u \sum \|D_i u\|_2 + \frac{1}{2\mu}\|Ku - f\|_2^2$$

Introducing w and quadratic penalty:

$$\min_{u,w} \sum \left(\|w_i\|_2 + \frac{1}{2\beta}\|w_i - D_i u\|_2^2 \right) + \frac{1}{2\mu}\|Ku - f\|_2^2$$

Alternating minimization:

- For fixed u, $\{w_i\}$ can be solved by shrinkage at $O(N)$
- For fixed $\{w_i\}$, u can be solved by FFT at $O(N \log N)$

Outline: Linearized ADMM

- Linearized Bregman and Bregmanized operator splitting
- ADMM + proximal point method
- Xiaoqun Zhang, Martin Burgerz, Stanley Osher, *A unified primal-dual algorithm framework based on Bregman iteration*

Review of Bregman method

Consider BP:

$$\min \ \|x\|_1, \ \text{ s.t. } Ax = b$$

Bregman method:

- $D_J^{p^k}(x, x^k) := \|x\|_1 - \|x^k\|_1 - \langle p^k, x - x^k \rangle$
- $x^{k+1} := \arg\min_x \ \mu D_J^{p^k}(x, x^k) + \frac{1}{2}\|Ax - b\|_2^2$
- $p^{k+1} = p^k + \frac{1}{\mu} A^\top (b - Ax^{k+1})$

Augmented Lagrangian (updating multiplier or b):

- $x^{k+1} := \arg\min_x \ \mu\|x\|_1 + \frac{1}{2}\|Ax - b^k\|_2^2$
- $b^{k+1} = b + (b^k - Ax^{k+1})$

They are equivalent, see Yin-Osher-Goldfarb-Darbon

Linearized approaches

Linearized Bregman method:

$$\begin{aligned} x^{k+1} &:= \arg\min\ \mu D_J^{p^k}(x, x^k) + (A^\top(Ax^k - b))^\top(x - x^k) + \frac{1}{2\delta}\|x - x^k\|_2^2, \\ p^{k+1} &:= p^k - \frac{1}{\mu\delta}(x^{k+1} - x^k) - \frac{1}{\mu}A^\top(Ax^k - b), \end{aligned}$$

which is equivalent to

$$\begin{aligned} x^{k+1} &:= \arg\min\ \mu\|x\|_1 + \frac{1}{2\delta}\|x - v^k\|_2^2 \\ v^{k+1} &:= v^k - \delta A^\top(Ax^{k+1} - b) \end{aligned}$$

Bregmanized operator splitting:

$$\begin{aligned} x^{k+1} &:= \arg\min\ \mu\|x\|_1 + (A^\top(Ax^k - b^k))^\top(x - x^k) + \frac{1}{2\delta}\|x - x^k\|_2^2 \\ b^{k+1} &= b + (b^k - Ax^{k+1}) \end{aligned}$$

Are they equivalent?

Linearized approaches

Linearized Bregman method:

$$
\begin{aligned}
x^{k+1} &:= \arg\min \ \mu D_J^{p^k}(x, x^k) + (A^\top(Ax^k - b))^\top(x - x^k) + \frac{1}{2\delta}\|x - x^k\|_2^2,\\
p^{k+1} &:= p^k - \frac{1}{\mu\delta}(x^{k+1} - x^k) - \frac{1}{\mu}A^\top(Ax^k - b),
\end{aligned}
$$

which is equivalent to

$$
\begin{aligned}
x^{k+1} &:= \text{shrink}(v^k, \mu\delta)\\
v^{k+1} &:= v^k - \delta A^\top(Ax^{k+1} - b)
\end{aligned}
$$

or

$$
\begin{aligned}
x^{k+1} &:= \text{shrink}(\delta A^\top b^k, \mu\delta)\\
b^{k+1} &:= b + (b^k - Ax^{k+1})
\end{aligned}
$$

Bregmanized operator splitting:

$$
\begin{aligned}
x^{k+1} &:= \text{shrink}(x^k - \delta(A^\top(Ax^k - b^k)), \mu\delta) = \text{shrink}(\delta A^\top b^k + x^k - \delta A^\top A x^k, \mu\delta)\\
b^{k+1} &= b + (b^k - Ax^{k+1})
\end{aligned}
$$

Linearized approaches

Linearized Bregman:

- If the sequence x^k converges and p^k is bounded, then the limit of x^k is the unique solution of

$$\min \ \mu\|x\|_1 + \frac{1}{2\delta}\|x\|_2^2 \ \text{ s.t. } Ax = b.$$

- For μ large enough, the limit solution solves BP.
- Exact regularization if $\delta > \bar{\delta}$

What about Bregmanized operator splitting?

Primal ADMM for ℓ_1-regularized problem

Primal: $\min \ \mu\|x\|_1 + \frac{1}{2}\|Ax - b\|_2^2$ which is equivalent to

$$\min \ \mu\|x\|_1 + \frac{1}{2}\|r\|_2^2, \ \text{ s.t. } Ax - b = r.$$

Augmented Lagrangian function:

$$\mathcal{L}(x, r, \lambda) = \mu\|x\|_1 + \frac{1}{2}\|r\|_2^2 - \lambda^\top(Ax - b - r) + \frac{1}{2\delta}\|Ax - b - r\|_2^2$$

ADMM scheme:

$$x^{k+1} = \arg\min_x \ \mu\|x\|_1 + \frac{1}{2\delta}\|Ax - b - r^k - \delta\lambda^k\|_2^2 \quad \text{original problem}$$

$$r^{k+1} = \arg\min_r \ \frac{1}{2}\|r\|_2^2 + \frac{1}{2\delta}\|Ax^{k+1} - b - r - \delta\lambda^k\|_2^2$$

$$\lambda^{k+1} = \lambda^k + \frac{Ax^{k+1} - b - r^{k+1}}{\delta}$$

Primal ADMM for ℓ_1-regularized problem

Primal: $\min \; \mu\|x\|_1 + \frac{1}{2}\|Ax-b\|_2^2$ which is equivalent to

$$\min \; \mu\|x\|_1 + \frac{1}{2}\|r\|_2^2, \quad \text{s.t. } Ax - b = r.$$

Augmented Lagrangian function:

$$\mathcal{L}(x, r, \lambda) = \mu\|x\|_1 + \frac{1}{2}\|r\|_2^2 - \lambda^\top(Ax - b - r) + \frac{1}{2\delta}\|Ax - b - r\|_2^2$$

ADMM scheme:

$$\begin{aligned}
x^{k+1} &= \arg\min_x \; \mu\|x\|_1 + (g^k)^\top(x - x^k) + \frac{1}{2\tau}\|x - x^k\|_2^2 \\
r^{k+1} &= \arg\min_r \; \frac{1}{2}\|r\|_2^2 + \frac{1}{2\delta}\|Ax^{k+1} - b - r - \delta\lambda^k\|_2^2 \\
\lambda^{k+1} &= \lambda^k + \frac{Ax^{k+1} - b - r^{k+1}}{\delta}
\end{aligned}$$

Convergence of the linearized scheme?

Generalized algorithm

Consider

$$\min f(x), \quad \text{s.t. } Ax = b$$

Proximal-point method:

$$\begin{aligned} x^{k+1} &:= \arg\min_x f(x) - \left\langle \lambda^k, Ax - b \right\rangle + \frac{1}{2\delta}\|Ax - b\|^2 + \frac{1}{2}\|x - x^k\|_Q^2 \\ C\lambda^{k+1} &:= C\lambda^k + (b - Ax^{k+1}) \end{aligned}$$

- Augmented Lagrangian or Bregman method if $Q = 0$ and $C = \delta$
- Proximal point method by Rockafellar if $Q = I$ and $C = \gamma$
- Bregmanized operator splitting if $Q = \frac{1}{\delta}(I - A^\top A)$

Theoretical results: $\lim_k \|Ax^k - b\|_2 = 0$, $\lim_k f(x^k) = f(\bar{x})$ and all limit point of (x^k, λ^k) are saddle points.

Convergence proof

From the x-subproblem:

$$\partial f(x^{k+1}) \ni s^{k+1} := A^\top \lambda^k - \frac{1}{\delta} A^\top (Ax^{k+1} - b) - Q(x^{k+1} - x^k)$$

Assume $(\bar{x}, \bar{\lambda})$ is a saddle point, then $A\bar{x} = b$, and $\bar{s} - A^\top \bar{\lambda} = 0$.
Let $s_e^{k+1} = s^{k+1} - \bar{s}$, $x_e^{k+1} = x^{k+1} - \bar{x}$ and $\lambda_e^{k+1} = \lambda^{k+1} - \bar{\lambda}$:

$$s_e^{k+1} + \frac{1}{\delta} A^\top A x_e^{k+1} + Q x_e^{k+1} = Q x_e^k + A^\top \lambda_e^k, \quad C\lambda_e^{k+1} = C\lambda_e^k - Ax_e^{k+1}$$

Taking the inner product with x_e^{k+1} on both sides of the first equality:

$$\frac{1}{2}\left(\|x_e^{k+1}\|_Q^2 + \|x^{k+1} - x^k\|_Q^2 - \|x_e^k\|_Q^2\right) = -\left\langle s_e^{k+1}, x_e^{k+1}\right\rangle - \frac{1}{\delta}\|Ax_e^{k+1}\|_2^2 + \left\langle \lambda_e^k, Ax_e^{k+1}\right\rangle$$

Taking the inner product with λ_e^{k+1} on both sides of the second equality:

$$\frac{1}{2}\left(\|\lambda_e^{k+1}\|_C^2 - \|\lambda_e^k\|_C^2 - \|Ax_e^{k+1}\|_{C^{-1}}^2\right) = -\left\langle \lambda_e^k, Ax_e^{k+1}\right\rangle$$

Adding the above inequality ($w^k = (x^k, \lambda^k)$):

$$\|w_e^{k+1}\|_Q^2 + \|x^{k+1} - x^k\|_Q^2 + 2\left\langle s_e^{k+1}, x_e^{k+1}\right\rangle + \frac{2}{\delta}\|Ax_e^{k+1}\|_2^2 - \|Ax_e^{k+1}\|_{C^{-1}}^2 = \|w_e^k\|_Q^2$$

Convergence proof

The convexity of $f(x)$ gives $\langle s_e^{k+1}, x_e^{k+1} \rangle \geq 0$. If $\frac{2}{\delta} > \frac{1}{\lambda_m^C}$, then

$$2 \left\langle s_e^{k+1}, x_e^{k+1} \right\rangle + \frac{2}{\delta} \|Ax_e^{k+1}\|_2^2 - \|Ax_e^{k+1}\|_{C^{-1}}^2 \geq 0$$

Hence, $\|w_e^{k+1}\|_Q^2 \leq \|w_e^k\|_Q^2$, which implies the boundedness of $w^k = (x^k, \lambda^k)$.
Furthermore,

$$\sum_k \left\{ 2 \left\langle s_e^{k+1}, x_e^{k+1} \right\rangle + \|x^{k+1} - x^k\|_Q^2 + \left(\frac{2}{\delta} \|Ax_e^{k+1}\|_2^2 - \|Ax_e^{k+1}\|_{C^{-1}}^2 \right) \right\}$$
$$\leq \quad \|w_e^0\|_Q^2 < \infty$$

Hence, $0 = \lim_k \|Ax_e^{k+1}\|_2^2 = \lim_k \|Ax^k - b\|$, and $\lim_k s^k - A^\top \lambda^k = 0$ follows from

$$s^{k+1} := A^\top \lambda^k - \frac{1}{\delta} A^\top (Ax^{k+1} - b) - Q(x^{k+1} - x^k)$$

Generalized algorithm

Consider

$$\min\ f(x)+g(z), \quad \text{s.t. } Bx=z$$

Proximal-point method:

$$\begin{aligned}
x^{k+1} &:= \arg\min_x\ f(x) - \left\langle \lambda^k, Bx \right\rangle + \frac{1}{2\delta}\|Bx - z^k\|^2 + \frac{1}{2}\|x - x^k\|_Q^2 \\
z^{k+1} &:= \arg\min_z\ g(z) - \left\langle \lambda^k, z \right\rangle + \frac{1}{2\delta}\|Bx^{k+1} - z\|^2 + \frac{1}{2}\|z - z^k\|_M^2 \\
C\lambda^{k+1} &:= C\lambda^k + (Bx^{k+1} - z^{k+1})
\end{aligned}$$

- ADMM if $Q=M=0$ and $C=\delta$
- Proximal point method by Rockafellar if $Q=I$ and $C=\gamma$
- Bregmanized operator splitting if $Q=\frac{1}{\delta}(I-A^\top A)$

Theoretical results: $\lim_k \|Bx^k - z^k\|_2 = 0$, $\lim_k f(x^k) = f(\bar{x})$, $\lim_k g(z^k) = g(\bar{z})$ and all limit point of (x^k, z^k, λ^k) are saddle points.