Inexact Alternating Direction Method of Multipliers for Separable Convex Optimization

Hongchao Zhang
hozhang@math.lsu.edu

Department of Mathematics
Center for Computation and Technology
Louisiana State University

U.S.-Mexico Workshop on Optimization and Applications
January 8th, 2016

Problem

Constrained Separable Convex Optimization

$$\min \sum_{i=1}^{m} f_i(\mathbf{x}_i) + h_i(\mathbf{x}_i) \quad \text{s. t.} \quad \sum_{i=1}^{m} \mathbf{A}_i \mathbf{x}_i = \mathbf{b}$$

where $m \geq 2$ and

- $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$ is convex, Lipschitz continuously differentiable.
- h_i is a simple proper closed convex function on $\mathbb{R}^{n_i}$, but not necessarily smooth.
- To put constraint $\mathbf{x}_i \in \mathcal{X}_i$, let h_i be the indicator function for closed convex set $\mathcal{X}_i \subseteq \mathbb{R}^{n_i}$.
- Many applications in image processing, statistical learning and compressive sensing, etc.

Motivation $m = 2$

Total variation image reconstruction

$$\min \; H(\mathbf{u}) + \phi(\mathbf{Bu})$$

which is equivalent to

$$\min \; H(\mathbf{u}) + \phi(\mathbf{w}) \quad \text{s. t.} \quad \mathbf{Bu} = \mathbf{w},$$

where

- $H(\mathbf{u}) = \frac{1}{2}\|\mathbf{Au} - \mathbf{f}\|^2$ with $\mathbf{A}$ large, dense and ill-conditioned
- $\phi(\mathbf{Bu}) = \alpha\|\mathbf{u}\|_{TV} = \alpha \sum_{i=1}^{N} \|(\nabla \mathbf{u})_i\|$ with $\alpha > 0$
- Augmented Lagrangian

$$\mathcal{L}^{\rho}(\mathbf{u}, \mathbf{w}, \mathbf{b}) = H(\mathbf{u}) + \phi(\mathbf{w}) + \langle \mathbf{b}, \mathbf{Bu} - \mathbf{w} \rangle + \frac{\rho}{2}\|\mathbf{Bu} - \mathbf{w}\|^2$$

Motivation $m = 2$

- Augmented Lagrangian Method (Method of Multipliers)

$$
\begin{aligned}
(\mathbf{u}^{k+1}, \mathbf{w}^{k+1}) &= \operatorname*{argmin}_{\mathbf{u},\mathbf{w}} \; \mathcal{L}^{\rho}(\mathbf{u}, \mathbf{w}, \mathbf{b}^{k}), \\
\mathbf{b}^{k+1} &= \mathbf{b}^{k} + \rho(\mathbf{B}\mathbf{u}^{k+1} - \mathbf{w}^{k+1}).
\end{aligned}
$$

- Alternating direction method of multipliers (ADMM)

$$
\begin{aligned}
\mathbf{u}^{k+1} &= \operatorname*{argmin}_{\mathbf{u}} \; \frac{1}{2}\|\mathbf{A}\mathbf{u} - \mathbf{f}\|^2 + \frac{\rho}{2}\|\mathbf{B}\mathbf{u} - \mathbf{w}^{k} + \rho^{-1}\mathbf{b}^{k}\|^2, \\
\mathbf{w}^{k+1} &= \operatorname*{argmin}_{\mathbf{w}} \; \phi(\mathbf{w}) + \frac{\rho}{2}\|\mathbf{B}\mathbf{u}^{k+1} - \mathbf{w} + \rho^{-1}\mathbf{b}^{k}\|^2, \\
\mathbf{b}^{k+1} &= \mathbf{b}^{k} + \rho(\mathbf{B}\mathbf{u}^{k+1} - \mathbf{w}^{k+1}).
\end{aligned}
$$

Motivation $m = 2$

- The "**u**" subproblem:

$$\mathbf{u}^{k+1} = \underset{\mathbf{u}}{\operatorname{argmin}} \ \frac{1}{2}\|\mathbf{Au} - \mathbf{f}\|^2 + \frac{\rho}{2}\|\mathbf{Bu} - \mathbf{w}^k + \rho^{-1}\mathbf{b}^k\|^2$$

- Bregman operator splitting (BOS) scheme: $\delta_k > \|\mathbf{A}^\mathsf{T}\mathbf{A}\|$

$$\mathbf{u}^{k+1} = \underset{\mathbf{u}}{\operatorname{argmin}} \ \delta_k\|\mathbf{u}-\mathbf{u}^k+\delta_k^{-1}\mathbf{A}^\mathsf{T}(\mathbf{Au}^k-\mathbf{f})\|^2+\rho\|\mathbf{Bu}-\mathbf{w}^k+\rho^{-1}\mathbf{b}^k\|^2$$

- Bregman operator splitting with variable stepsize (BOSVS): Apply nonmonotone line search with initial

$$\hat{\delta}_k = \max\left\{\Delta_k, \ \frac{\|\mathbf{A}(\mathbf{u}^k - \mathbf{u}^{k-1})\|^2}{\|\mathbf{u}^k - \mathbf{u}^{k-1}\|^2}\right\},$$

where $\Delta_k > 0$ is a lower bound and is adaptively increased.

BOSVS Algorithm

Parameters: $\tau,\ \eta > 1,\ \beta,\ C \geq 0,\ \rho,\ \delta_{\min} > 0,\ \xi, \sigma \in (0,1),\ \delta_0 = 1.$

Initialization: w^1, u^1 and b^1. Set $Q_1 = 0$ and $\Delta_1 = \delta_{\min}$.

For $k = 1, 2, \ldots$

Step 1. Set $\delta_k = \eta^j \hat{\delta}_k$ where $j \geq 0$ is the smallest integer such that

$Q_{k+1} \geq -\frac{C}{k^2}$ where $Q_{k+1} := \xi_k Q_k + \Omega_k$ with

$\Omega_k := \sigma(\delta_k \|u^{k+1} - u^k\|^2 + \rho \|Bu^{k+1} - w^k\|^2) - \|A(u^{k+1} - u^k)\|^2$, and

$$u^{k+1} = \operatorname*{argmin}_u \left\{ \delta_k \|u - u^k + \delta_k^{-1} A^{\mathsf{T}}(Au^k - f)\|^2 + \rho \|w^k - Bu + \frac{b^k}{\rho}\|^2 \right\},$$

$0 \leq \xi_k \leq \min\{\xi, (1 - k^{-1})^2\}$.

Step 2. If $\delta_k > \max\{\delta_{k-1}, \Delta_k\}$ when $k > 1$, $\Delta_{k+1} := \tau \Delta_k$.

Step 3. $$w^{k+1} = \operatorname*{argmin}_w \left\{ \phi(w) + \frac{\rho}{2} \|w - Bu^{k+1} + b^k/\rho\|^2 + \frac{\beta}{2} \|w - w^k\|^2 \right\}.$$

Step 4. $b^{k+1} = b^k + \rho(Bu^{k+1} - w^{k+1})$.

Step 5. If a stopping criterion is satisfied, stop.

End For

Convergence of BOSVS

Theorem. If the minimizer exists, the sequence $\mathbf{x}^k = (\mathbf{u}^k, \mathbf{w}^k, \mathbf{b}^k)$ generated by BOSVS converges to a solution $\mathbf{x}^* = (\mathbf{u}^*, \mathbf{w}^*, \mathbf{b}^*)$ satisfying the first-order optimality conditions

$$\nabla H(\mathbf{u}^*) - \mathbf{B}^\mathsf{T}\mathbf{b}^* = \mathbf{0}, \quad -\mathbf{b}^* \in \partial\phi(\mathbf{w}^*), \quad \mathbf{w}^* = \mathbf{B}\mathbf{u}^*.$$

Moreover, the ergodic mean $\mathbf{u}_K$ given by $\mathbf{u}_K := \frac{1}{K}\sum_{k=1}^K \mathbf{u}^k$ satifies

$$\phi(\mathbf{B}\mathbf{u}_K) + H(\mathbf{u}_K) - \min_{\mathbf{u}}\{\phi(\mathbf{B}\mathbf{u}) + H(\mathbf{u})\} = \mathcal{O}(\frac{1}{K}).$$

Numerical Experiments (BOSVS)

Partially Parallel Imaging (PPI) with $L = 8$ coils

- Optimize
$$\min_{\mathbf{u}} \frac{1}{2} \sum_{l=1}^{L} \|\mathcal{F}_p(\mathbf{s}_l \odot \mathbf{u}) - \mathbf{f}_l\|^2 + \alpha \|\mathbf{u}\|_{TV},$$
where $\mathcal{F}_p$ is the undersampled Fourier transform, $\mathbf{S}_l$ is the sensitivity map of the l-th channel, and $\alpha = 10^{-5}$.

- Three data sets:
  * data 1 and data 2 uses a random Poisson mask (25% Fourier coefficients);
  * data 3 uses a radial mask (34% Fourier coefficients).

- Plot the error in the objective function versus the iteration number ($\rho = 10^{-4}$).

Numerical Experiments (BOSVS)

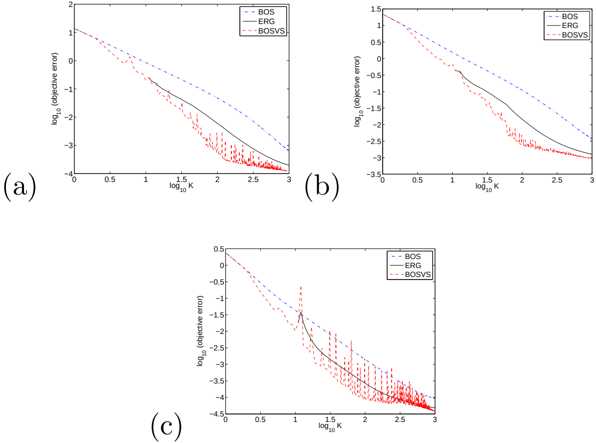


Figure: Plots of the objective error for $\mathrm{data}\ 1$, $\mathrm{data}\ 2$, and $\mathrm{data}\ 3$.

Least squares fit of $y = cK^{-p}$ give p as 1.6, 1.2, and 0.9 for data 1, data 2, and data 3 respectively.

Separable Convex Optimization $m \geq 3$

Optimize

$$\min \sum_{i=1}^{m} f_i(\mathbf{x}_i) + h_i(\mathbf{x}_i) \quad \text{s. t.} \quad \sum_{i=1}^{m} \mathbf{A}_i \mathbf{x}_i = \mathbf{b}$$

where

- $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$ is convex, Lipschitz continuously differentiable.
- h_i is a simple proper closed convex function on $\mathbb{R}^{n_i}$, but not necessarily smooth.

We assume

- $\mathbf{A}_i^\top \mathbf{A}_i$ is nonsingular and the solution set of the problem is nonempty.

Separable Convex Optimization $m \geq 3$

- Direct extention of ADMM:

$$\begin{cases} \text{For } i = 1, \ldots, m \\ \qquad \mathbf{x}_i^{k+1} = \arg\min \mathcal{L}^{\rho}(\mathbf{x}_1^{k+1}, \ldots, \mathbf{x}_{i-1}^{k+1}, \mathbf{x}_i, \mathbf{x}_{i+1}^{k}, \ldots, \mathbf{x}_m^{k}; \boldsymbol{\lambda}^k); \\ \textit{End} \\ \boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}), \end{cases}$$

  where $\mathcal{L}^{\rho}(\mathbf{x}_1, \ldots, \mathbf{x}_m; \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^{\mathsf{T}}(\mathbf{A}\mathbf{x} - \mathbf{b}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$, $f = \sum_{i=1}^{m} f_i + h_i$, $\mathbf{A}\mathbf{x} := \sum_{i=1}^{m} \mathbf{A}_i\mathbf{x}_i$ and $\rho > 0$ is a parameter.

- Practical efficiency has been observed in many recent applications.

- However, not necessarily converge ! (Chen, He, Ye, Yuan 2013)

- Moreover, each subproblem is solved exactly, which could be very expensive, not practical or even impossible when no closed formula exists.

Separable Convex Optimization $m \geq 3$

Literature

- Han and Yuan (2012):
  * Assume all f_i strongly convex and ρ in a specific range
  * Each subproblem needs to be solved exactly.

- He, Tao, Xu and Yuan (2013)
  * Block Gaussian backward substitution is used to ensure convergence.
  * Each subproblem needs to be solved exactly.

- Hong and Luo (2013)

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \alpha\rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}),$$

  * Stepsize α needs to be sufficiently small.
  * f_i and h_i needs to satisfy certain local error bound conditions.
  * Linear convergence rate is achieved.

- Much more recent works: randomization, $m-2$ strongly convex assumption, ...

Separable Convex Optimization $m \geq 3$

Motivation

- Extend BOSVS to the multi-block case to solve the subproblems inexactly.
- Analogous to the standard Augmented Lagrangian Method (ALM), i.e. $m = 1$, solve the subproblems to the adaptive accuracy relative to the current KKT error. (Do not require summable error !
Ex. Eckstein-Bertsekas 1992: $\|\mathbf{x}_i^k - \mathbf{x}_i^*\| \leq \eta^k$ with $\sum_k^\infty \eta^k < \infty$)
- Apply accelerated optimal gradient methods to solve the subproblems.
- Global convergence is guaranteed.

Notation

- Let us define $H = \mathrm{diag}(\mathbf{A}_2^\top \mathbf{A}_2, \mathbf{A}_3^\top \mathbf{A}_3, \ldots, \mathbf{A}_m^\top \mathbf{A}_m)$ and

$$M = \begin{pmatrix} \mathbf{A}_2^\top \mathbf{A}_2 & 0 & \cdots & 0 \\ \mathbf{A}_3^\top \mathbf{A}_2 & \mathbf{A}_3^\top \mathbf{A}_3 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_m^\top \mathbf{A}_2 & \mathbf{A}_m^\top \mathbf{A}_3 & \cdots & \mathbf{A}_m^\top \mathbf{A}_m \end{pmatrix}.$$

  Note H is positive definite and M is nonsingular.

- Let us define

$$\Phi_i^k(\mathbf{u}, \bar{\mathbf{u}}, \delta) = f_i(\bar{\mathbf{u}}) + \nabla f_i(\bar{\mathbf{u}})^\top (\mathbf{u} - \bar{\mathbf{u}}) + \frac{\delta}{2}\|\mathbf{u} - \bar{\mathbf{u}}\|^2 + h_i(\mathbf{u}) + \frac{\rho}{2}\|\mathbf{A}_i \mathbf{u} - \mathbf{b}_i^k + \boldsymbol{\lambda}^k/\rho\|^2,$$

  where $\mathbf{b}_i^k = \mathbf{b} - \sum_{j<i} \mathbf{A}_j \bar{\mathbf{x}}_j^k - \sum_{j>i} \mathbf{A}_j \tilde{\mathbf{x}}_j^k$.

Linearized ADMM with Gaussian backward substitution

Parameters: $\rho,\ \theta_1,\ \theta_2,\ \theta_3 \in \mathbb{R}^{++}$; $0 < \delta_{min} << \delta_{\max}$, $\alpha \in (0,1)$, $0 < \sigma < 1 < \eta$, $\{\epsilon^k\} \subset \mathbb{R}^+$ with $\sum_{k=1}^{\infty} \epsilon^k < \infty$.

Step 0: Initialize $\boldsymbol{\lambda}^1$. For $i = 1, \ldots, m$, set $\delta_{\min,i} = \delta_{\min}$, initialize $\mathbf{x}_i^1$ and set $\tilde{\mathbf{x}}_i^1 = \bar{\mathbf{x}}_i^1 = \mathbf{x}_i^1$. Let $k = 1$.

Step 1: For $i = 1, \ldots, m$

Choose $\delta_i^k = \eta^j \delta_{i,0}^k$, where $j \geq 0$ is the smallest integer such that

$$f_i(\mathbf{x}_i^k) + \langle \nabla f_i(\mathbf{x}_i^k), \bar{\mathbf{x}}_i^k - \mathbf{x}_i^k \rangle + \frac{(1-\sigma)\delta_i^k}{2} \|\bar{\mathbf{x}}_i^k - \mathbf{x}_i^k\|^2 + \epsilon^k \geq f_i(\bar{\mathbf{x}}_i^k),$$

where $\delta_{\min} \leq \delta_{i,0}^k \leq \delta_{\max}$ and $\bar{\mathbf{x}}_i^k = \mathrm{Arg\,min}_{\mathbf{u} \in \mathbb{R}^{n_i}} \Phi_i^k(\mathbf{u}, \mathbf{x}_i^k, \delta_i^k)$.

Let $\mathbf{x}_i^{k+1} = \bar{\mathbf{x}}_i^k$ and $\bar{r}_i^k = (1/\delta_i^k)\|\bar{\mathbf{x}}_i^k - \mathbf{x}_i^k\|^2$.

If $\delta_i^k > \max\{\delta_i^{k-1}, \delta_{\min,i}\}$ when $k > 1$, $\delta_{\min,i} := \eta\delta_{\min,i}$.

End

Step 2: Let $e^k = \theta_1\|\bar{\mathbf{v}}^k - \tilde{\mathbf{v}}^k\| + \theta_2\|\sum_{i=1}^m \mathbf{A}_i\bar{\mathbf{x}}_i^k - \mathbf{b}\| + \theta_3\sqrt{\sum_{i=1}^m \bar{r}_i^k}$.
If e^k is sufficiently small, stop the algorithm.

Step 3: Let $\tilde{\mathbf{x}}_1^{k+1} = \bar{\mathbf{x}}_1^k$ and $\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \alpha\rho(\sum_{i=1}^m \mathbf{A}_i\bar{\mathbf{x}}_i^k - \mathbf{b})$.
Compute $\tilde{\mathbf{v}}^{k+1} = \tilde{\mathbf{v}}^k + \alpha M^{-\mathrm{T}} H(\bar{\mathbf{v}}^k - \tilde{\mathbf{v}}^k)$. (Backward Substitution)
Let $k := k + 1$ and go to Step 1.

Figure: L-ADMM-G Algorithm

Global Convergence of L-ADMM-G

Theorem. Let $\tilde{\mathbf{w}}^k = (\tilde{\mathbf{x}}_1^k, \tilde{\mathbf{x}}_2^k, \ldots, \tilde{\mathbf{x}}_n^k, \boldsymbol{\lambda}^k)$, $\bar{\mathbf{w}}^k = (\bar{\mathbf{x}}_1^k, \bar{\mathbf{x}}_2^k, \ldots, \bar{\mathbf{x}}_n^k, \boldsymbol{\lambda}^k)$ be the iterates generated by the L-ADM-G algorithm. Then,

$$\lim_{k\to\infty} \tilde{\mathbf{w}}^k = \lim_{k\to\infty} \bar{\mathbf{w}}^k = \mathbf{w}^*,$$

where $\mathbf{w}^* = (\mathbf{x}_1^*, \mathbf{x}_2^*, \ldots, \mathbf{x}_,^*, \boldsymbol{\lambda}^*)$ is an optimal primal-dual solution pair.

Proof. *Denote* $E_k = \rho\|\tilde{\mathbf{v}}_e^k\|_G^2 + \frac{1}{\rho}\|\boldsymbol{\lambda}_e^k\|^2 + \alpha\sum_{i=1}^m(\delta_i^k\|\mathbf{x}_{i,e}^k\|^2)$. *We can show*

$$E_k \geq E_{k+1} + \tau_k,$$

for large k, *where*

$$\tau_k = \tilde{c}\sum_{i=1}^m \|\bar{\mathbf{x}}_i^k - \mathbf{x}_i^k\|^2 + \bar{c}(\rho\|\tilde{\mathbf{v}}^k - \bar{\mathbf{v}}^k\|_H^2 + \frac{1}{\rho}\|\boldsymbol{\lambda}^k - \bar{\boldsymbol{\lambda}}^k\|^2) - \hat{c}\epsilon^k.$$

Inexact ADMM with Gaussian backward substitution

Parameters: $\rho,\ \theta_1,\ \theta_2,\ \theta_3 \in \mathbb{R}^{++}$, $\alpha \in (0,1)$, $\{\epsilon^k\} \subset \mathbb{R}^{+}$ with $\sum_{k=1}^{\infty} \epsilon^k < \infty$.

Step 0: Initialize $\boldsymbol{\lambda}^1$. For $i = 1, \ldots, m$, initialize $\mathbf{x}_i^1$ and set $\tilde{\mathbf{x}}_i^1 = \bar{\mathbf{x}}_i^1 = \mathbf{x}_i^1$. Let $e^0 = \infty$, $\Gamma_i^0 = 0$ and $k = 1$.

Step 1: For $i = 1, \ldots, m$

Use the Accelerated optimal Gradient (AOG) method to solve $\min_{\mathbf{u} \in \mathbb{R}^{n_i}} L_i^k(\mathbf{u})$ inexactly.

End

Step 2: Let $e^k = \theta_1 \|\bar{\mathbf{v}}^k - \tilde{\mathbf{v}}^k\| + \theta_2 \| \sum_{i=1}^m \mathbf{A}_i \bar{\mathbf{x}}_i^k - \mathbf{b}\| + \theta_3 \sqrt{\sum_{i=1}^m \bar{r}_i^k}$.
If e^k is sufficiently small, stop the algorithm.

Step 3: Let $\tilde{\mathbf{x}}_1^{k+1} = \bar{\mathbf{x}}_1^k$ and $\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \alpha\rho(\sum_{i=1}^m \mathbf{A}_i \bar{\mathbf{x}}_i^k - \mathbf{b})$.
Compute $\tilde{\mathbf{v}}^{k+1} = \tilde{\mathbf{v}}^k + \alpha M^{-\mathsf{T}} H(\bar{\mathbf{v}}^k - \tilde{\mathbf{v}}^k)$. (Backward Substitution)
Let $k := k + 1$ and go to Step 1.

Figure: I-ADMM-G Algorithm

Accelerated Optimal Gradient Method

Notation

- For any h_i and $\mathbf{z}_i \in \mathbb{R}^{n_i}$, we define its proximal mapping

$$\mathrm{prox}_{h_i}(\mathbf{z}_i) = \mathrm{Arg}\min_{\mathbf{u}\in\mathbb{R}^{n_i}} h_i(\mathbf{u}) + \frac{1}{2}\|\mathbf{z}_i - \mathbf{u}\|^2,$$

  and define $\tilde{L}_i^k(\mathbf{u}) := L_i^k(\mathbf{u}) - h_i(\mathbf{u})$, where

$$L_i^k(\mathbf{u}) := L^\rho(\bar{\mathbf{x}}_1^k, \ldots, \bar{\mathbf{x}}_{i-1}^k, \mathbf{u}, \tilde{\mathbf{x}}_{i+1}^k, \ldots, \tilde{\mathbf{x}}_m^k; \boldsymbol{\lambda}^k).$$

- We have

$$\mathbf{x}_{i,k}^* = \mathrm{Arg}\min_{\mathbf{u}\in\mathbb{R}^{n_i}} \; L_i^k(\mathbf{u})$$

  if and only if

$$\|\mathrm{prox}_{h_i}(\mathbf{x}_{i,k}^* - \nabla\tilde{L}_i^k(\mathbf{x}_{i,k}^*)) - \mathbf{x}_{i,k}^*\| = 0.$$

- Define $\psi : \mathbb{R} \to \mathbb{R}^+$ to be any function having the property that

$$\lim_{t\to 0}\psi(t) = 0 \quad \text{and} \quad \psi(t) = 0 \text{ if and only if } t = 0.$$

Accelerated Optimal Gradient Method

Figure: Solve $\min_{\mathbf{u}\in\mathbb{R}^{n_i}} \; L_i^k(\mathbf{u})$ inexactly

Parameters: $0 \le \sigma < 1, \quad \epsilon^k > 0, \quad \{\omega^\ell\} \subset \mathbb{R}^+$ with $\sum_{\ell=1}^{\infty} \omega^\ell < \infty$.

Let $r_i^0 = 0,\ \mathbf{y}_i^0 = \mathbf{u}_i^0 = \mathbf{x}_i^k,\ \alpha^1 = 1,\ \ell = 1$ and $\mathit{flag} = \mathit{true}$.

while $(\mathit{flag} = \mathit{true})$

Choose $\delta^\ell > 0$ and $\alpha^\ell \in (0,1)$ for $\ell \ge 2$ such that

$$f_i(\bar{\mathbf{y}}_i^\ell) + \langle \nabla f_i(\bar{\mathbf{y}}_i^\ell), \mathbf{y}_i^\ell - \bar{\mathbf{y}}_i^\ell \rangle + \frac{(1-\sigma)\delta^\ell}{2\alpha^\ell} \|\mathbf{y}_i^\ell - \bar{\mathbf{y}}_i^\ell\|^2 + \pi^\ell \ge f_i(\mathbf{y}_i^\ell), \qquad \text{(II)}$$

where $\bar{\mathbf{y}}_i^\ell = (1-\alpha^\ell)\mathbf{y}_i^{\ell-1} + \alpha^\ell \mathbf{u}_i^{\ell-1},\ \mathbf{y}_i^\ell = (1-\alpha^\ell)\mathbf{y}_i^{\ell-1} + \alpha^\ell \mathbf{u}_i^\ell,$

$$\mathbf{u}_i^\ell = \arg\min_{\mathbf{u}\in\mathbb{R}^{n_i}} \langle \nabla f_i(\bar{\mathbf{y}}_i^\ell), \mathbf{u} \rangle + \frac{\delta^\ell}{2}\|\mathbf{u} - \mathbf{u}_i^{\ell-1}\|^2 + \frac{\rho}{2}\|\mathbf{A}_i\mathbf{u} - \mathbf{b}_i^k + \boldsymbol{\lambda}^k/\rho\|^2 + h_i(\mathbf{u}),$$

and $\pi^\ell = \epsilon^k \omega^\ell / (2\gamma^l)$ with $\gamma^1 = 1/\delta^1$ and $\gamma^\ell = \gamma^{\ell-1}/(1-\alpha^\ell)$ for $\ell \ge 2$.

Let $r_i^\ell = r_i^{\ell-1} + \delta^\ell \alpha^\ell \gamma^\ell \|\mathbf{u}_i^\ell - \mathbf{u}_i^{\ell-1}\|^2$.

If $\gamma^\ell \ge \Gamma_i^{k-1}$ and $\|\mathrm{prox}_{h_i}(\mathbf{y}_i^\ell - \nabla \tilde{L}_i^k(\mathbf{y}_i^\ell)) - \mathbf{y}_i^\ell\| \le \psi(e^{k-1})$,

set $\mathbf{x}_i^{k+1} = \mathbf{u}_i^\ell,\ \Gamma_i^k = \gamma^\ell,\ \bar{\mathbf{x}}_i^k = \mathbf{y}_i^\ell,\ \bar{r}_i^k = r_i^\ell/\Gamma_i^k$ and $\mathit{flag} = \mathit{false}$;

else $\ell := \ell + 1$.

end

Accelerated Optimal Gradient Method

- If the Lipschitz constant ζ_i of ∇f_i is known, we could simply set

$$\delta^\ell = \frac{1}{(1-\sigma)}\frac{2\zeta_i}{\ell} \quad \text{and} \quad \alpha^\ell = \frac{2}{\ell+1} \in (0,1].$$

Comment:

  * We have

$$\frac{(1-\sigma)\delta^\ell}{\alpha^\ell} = \frac{(l+1)\zeta_i}{l} > \zeta_i.$$

  Hence, the condition (II) in the AG algorithm will be satisfied.

  * In addition, we have

$$\gamma^\ell = \frac{1}{\delta^1}\left[\prod_{j=2}^{\ell}(1-\alpha^j)\right]^{-1} = \frac{\ell(\ell+1)}{2\delta^1} \quad \text{and} \quad \xi^\ell := \delta^\ell\alpha^\ell\gamma^\ell = 1.$$

Accelerated Optimal Gradient Method

- When the Lipschitz constant of ∇f_i is unknown, let $\tau^\ell = 1/(\delta_0^\ell \eta^j)$, where $\delta_0^\ell \in [\delta_{\min}, \delta_{\max}]$ and $j \geq 0$ is the smallest integer such that with

$$\delta^\ell = \frac{2}{\tau^\ell + \sqrt{(\tau^\ell)^2 + 4\tau^\ell \Lambda^{\ell-1}}} \quad \text{and} \quad \alpha^\ell = \frac{1}{1 + \delta^\ell \Lambda^{\ell-1}} \in (0, 1],$$

the condition (II) in the AG algorithm is satisfied, where $\Lambda^{\ell-1} = \sum_{i=1}^{\ell-1} \frac{1}{\delta^i}$.

Comment:

* Then, we have

$$\frac{(1-\sigma)\delta^\ell}{\alpha^\ell} = \frac{1-\sigma}{\theta^\ell} = (1-\sigma)\delta_0^\ell \eta^j.$$

Hence, the condition (II) in the AG algorithm will be satisfied as $j \to \infty$.

* In addition, we have

$$\gamma^\ell = \Lambda^\ell \geq \frac{\ell^2}{4(\eta\zeta_i/(1-\sigma) + \delta_{\max})} \quad \text{and} \quad \xi^\ell := \delta^\ell \alpha^\ell \gamma^\ell = 1.$$

Accelerated Optimal Gradient Method

Lemma. Suppose the accelerated optimal gradient method with the previous line search rule is applied to solve the subproblem. Then we have

$$\|\mathbf{y}_i^\ell - \mathbf{x}_{i,k}^*\|^2 \le \frac{1}{\nu_i \rho} \frac{\left(\|\mathbf{u}_i^0 - \mathbf{x}_{i,k}^*\|^2 + \epsilon^k \sum_{j=1}^{\ell} \omega^j\right)}{\gamma^\ell},$$

where $\mathbf{x}_{i,k}^* = \operatorname{Arg\,min}\ L_i^k(\mathbf{u})$ and ν_i is the minimum eigenvalue of $\mathbf{A}_i^\top \mathbf{A}_i$.

Comment:

- We have optimal convergence rate

$$\|\mathbf{y}_i^\ell - \mathbf{x}_{i,k}^*\|^2 \le \mathcal{O}(\frac{1}{\ell^2}).$$

- The subproblem stopping condition

$$\gamma^\ell \ge \Gamma_i^{k-1} \ \text{ and } \ \|\mathrm{prox}_{h_i}(\mathbf{y}_i^\ell - \nabla \tilde{L}_i^k(\mathbf{y}_i^\ell)) - \mathbf{y}_i^\ell\| \le \psi(e^{k-1}),$$

will be satisfied in finite number of steps.

Global Convergence of I-ADMM-G

Theorem. Let $\tilde{\mathbf{w}}^k = (\tilde{\mathbf{x}}_1^k, \tilde{\mathbf{x}}_2^k, \ldots, \tilde{\mathbf{x}}_n^k, \boldsymbol{\lambda}^k)$, $\bar{\mathbf{w}}^k = (\bar{\mathbf{x}}_1^k, \bar{\mathbf{x}}_2^k, \ldots, \bar{\mathbf{x}}_n^k, \boldsymbol{\lambda}^k)$ be the iterates generated by the I-ADMM-G algorithm. Suppose

- $\lim_{\ell\to\infty} \gamma^l = \infty$;
- $\xi^\ell := \delta^\ell \alpha^\ell \gamma^\ell$ is constant ;

Then,

$$\lim_{k\to\infty} \tilde{\mathbf{w}}^k = \lim_{k\to\infty} \bar{\mathbf{w}}^k = \mathbf{w}^*,$$

where $\mathbf{w}^* = (\mathbf{x}_1^*, \mathbf{x}_2^*, \ldots, \mathbf{x}_,^*, \boldsymbol{\lambda}^*)$ is an optimal primal-dual solution pair.

Proof. Denote $E_k = \rho\|\tilde{\mathbf{v}}_e^k\|_G^2 + \frac{1}{\rho}\|\boldsymbol{\lambda}_e^k\|^2 + \alpha\sum_{i=1}^m \frac{\|\mathbf{x}_{i,e}^k\|^2}{\Gamma_i^k}$. *We can show*

$$E_k \geq E_{k+1} + \tau_k,$$

for large k*, where*

$$\tau_k = \tilde{c}\sum_{i=1}^m \frac{1}{\Gamma_i^k}\sum_{\ell=1}^{I_i^k}\|\mathbf{u}_{i,k}^\ell - \mathbf{u}_{i,k}^{\ell-1}\|^2 + \bar{c}(\rho\|\tilde{\mathbf{v}}^k - \bar{\mathbf{v}}^k\|_H^2 + \frac{1}{\rho}\|\boldsymbol{\lambda}^k - \bar{\boldsymbol{\lambda}}^k\|^2) - \alpha\sum_{i=1}^m \frac{\epsilon^k}{\Gamma_i^k}\sum_{\ell=1}^{I_i^k}\omega^\ell.$$

Numerical Experiments

An image deblurring problem for the Cameraman image

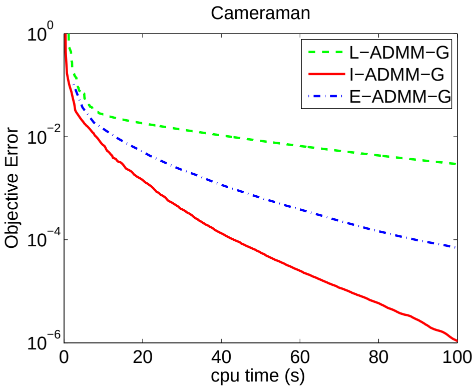
- Optimize
$$\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{A}\mathbf{u}-\mathbf{b}\|^2 + \alpha\|\mathbf{u}\|_{TV} + \beta\|\Phi^T\mathbf{u}\|_1,$$
which is equivalent to
$$\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{A}\mathbf{u}-\mathbf{b}\|^2 + \alpha\|\mathbf{w}\|_{1,2} + \beta\|\mathbf{z}\|_1, \text{ s. t. } \mathbf{B}\mathbf{u} = \mathbf{w}, \Phi^T\mathbf{u} = \mathbf{z}$$
where Φ is the wavelet transform, $\alpha = 0.005$ and $\beta = 0.001$.

- size 256×256, 9×9 uniform blur and Gaussian noise $SNR = 40$

- Relative accuracy for the subproblem: $\psi(e) = \min\{0.1e, e^{1.1}\}$

- Plot the relative objective function error versus the CPU time ($\rho = 5 \times 10^{-4}$).

Numerical Experiments



Plots of the relative objective error of the Cameraman image (size 256×256, 9×9 uniform blur and Gaussian noise $SNR = 40$)

Numerical Experiments

Three Partially Parallel Imaging (PPI) problems

- Optimize

$$\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{A}\mathbf{u}-\mathbf{b}\|^2 + \alpha\|\mathbf{u}\|_{TV} + \beta\|\Phi^{\mathsf{T}}\mathbf{u}\|_1,$$

  which is equivalent to

$$\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{A}\mathbf{u}-\mathbf{b}\|^2 + \alpha\|\mathbf{w}\|_{1,2} + \beta\|\mathbf{z}\|_1, \text{ s. t. } \mathbf{B}\mathbf{u}=\mathbf{w}, \Phi^{\mathsf{T}}\mathbf{u}=\mathbf{z}$$

  where Φ is the wavelet transform, $\alpha = 10^{-5}$ and $\beta = 10^{-6}$.

- Relative accuracy for the subproblem: $\psi(e) = \min\{0.1e, e^{1.1}\}$
- Plot the relative objective function error versus the CPU time ($\rho = 10^{-3}$).

Numerical Experiments

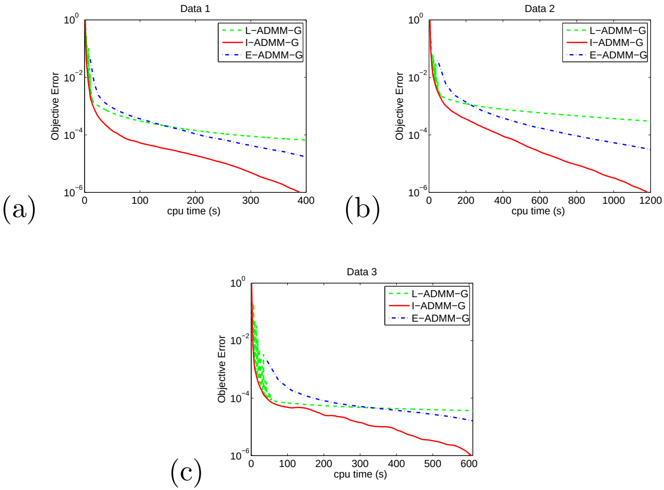


Figure: Plots of the relative objective error for $\mathrm{data}\ 1$, 2 and 3.