Lecture: Algorithms for Compressed Sensing

Zaiwen Wen

Beijing International Center For Mathematical Research
Peking University

`http://bicmr.pku.edu.cn/~wenzw/bigdata2024.html`
wenzw@pku.edu.cn

Outline

- Proximal gradient method
- Accelerated gradient method
- Alternating direction methods of Multipliers (ADMM)
- Linearized Alternating direction methods of Multipliers
- Greedy methods

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

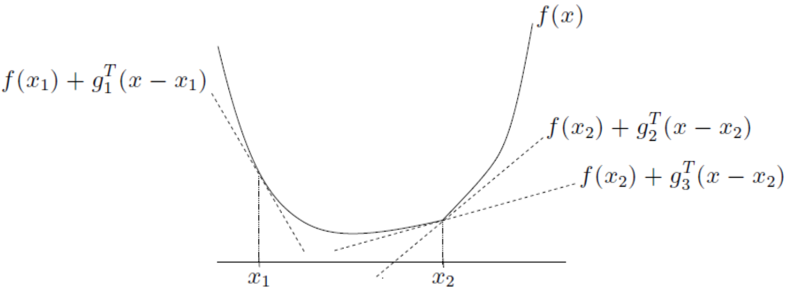
Subgradient

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 $x \in \mathrm{dom} f$ if

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

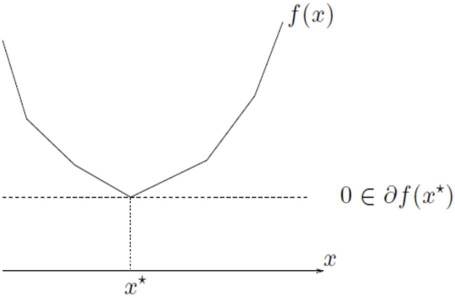


g_2, g_3 are subgradients at x_2, g_1 is a subgradient at x_1.

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
- $\lambda^* \geq 0$
- 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:

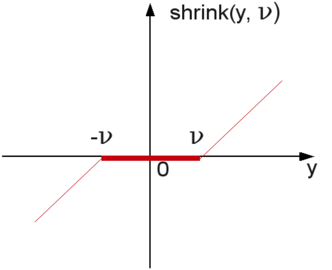
$$\begin{aligned} x^{k+1} &:= \arg\min_{x\in\mathbb{R}^n} \; \mu\|x\|_1 + (\nabla f(x^k))^\top(x - x^k) + \frac{1}{2\tau}\|x - x^k\|_2^2 \\ &= \arg\min_{x\in\mathbb{R}^n} \; \mu\|x\|_1 + \frac{1}{2\tau}\|x - (x^k - \tau\nabla f(x^k))\|_2^2 \\ &= \text{shrink}(x^k - \tau\nabla f(x^k), \mu\tau) \end{aligned}$$

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

Shrinkage (soft thresholding)

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

- 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

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

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

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

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

Proximal Operator

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

Convergence of proximal gradient method

to minimize $f + h$, choose x^0 and repeat

$$x^k = \mathrm{prox}_{t_k h}\left(x^{k-1} - t\nabla f(x^{k-1})\right), \quad k \geq 1$$

assumptions

- f convex with $\mathbf{dom}\, g = \mathbf{R}^n$; ∇f Lipschitz continuous with constant L:

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2 \quad \forall x, y$$

- h is closed and convex (so that prox_{th} is well defined)
- optimal value F^* is finite and attained at x^* (not necessarily unique)

convergence result: $1/k$ rate convergence with fixed step size $t_k = 1/L$

Gradient map

$$G_t(x) = \frac{1}{t}(x - \text{prox}_{th}(x - t\nabla f(x)))$$

$G_t(x)$ is the negative 'step' in the proximal gradient update

$$\begin{aligned} x^+ &= \text{prox}_{th}(x - t\nabla f(x)) \\ &= x - tG_t(x) \end{aligned}$$

- $G_t(x)$ is not a gradient or subgradient of $F = g + h$
- from subgradient definition of prox-operator

$$G_t(x) \in \partial f(x) + \partial h(x - tG_t(x))$$

- $G_t(x) = 0$ if and only if x minimizes $F(x) = f(x) + h(x)$

Consequences of Lipschitz assumption

recall upper bound (lecture on "gradient method") for convex f with Lipschitz continuous gradient

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

- substitute $y = x - tG_t(x)$:

$$f(x - tG_t(x)) \le f(x) - t\nabla f(x)^\top G_t(x) + \frac{t^2 L}{2}\|G_t(x)\|_2^2$$

- if $0 < t \le 1/L$, then

$$f(x - tG_t(x)) \le f(x) - t\nabla f(x)^\top G_t(x) + \frac{t}{2}\|G_t(x)\|_2^2 \qquad (1)$$

A global inequality

if the inequality (1) holds, then for all z,

$$F(x - tG_t(x)) \leq F(x) + G_t(x)^\top (x - z) - \frac{t}{2}\|G_t(x)\|_2^2 \qquad (2)$$

proof: (define $v = G_t(x) - \nabla f(x)$)

$$\begin{aligned}
F(x - tG_t(x)) &\leq f(x) - t\nabla f(x)^\top G_t(x) + \frac{t}{2}\|G_t(x)\|_2^2 + h(x - tG_t(x)) \\
&\leq f(z) + \nabla f(x)^\top (x - z) - t\nabla f(x)^\top G_t(x) + \frac{t}{2}\|G_t(x)\|_2^2 \\
&+ h(z) + v^\top (x - z - tG_t(x)) \\
&= f(z) + h(z) + G_t(x)^\top (x - z) - \frac{t}{2}\|G_t(x)\|_2^2
\end{aligned}$$

line 2 follows from convexity of f and h, and $v \in \partial h(x - tG_t(x))$

Progress in one iteration

$$x^+ = x - tG_t(x)$$

- inequality (2) with $z = x$ shows the algorithm is a descent method:

$$F(x^+) \le F(x) - \frac{t}{2}\|G_t(x)\|_2^2$$

- inequality (2) with $z = x^*$

$$\begin{aligned} F(x^+) - F^* &\le G_t(x)^\top (x - x^*) - \frac{t}{2}\|G_t(x)\|_2^2 \\ &= \frac{1}{2t}\left(\|x - x^*\|_2^2 - \|x - x^* - tG_t(x)\|_2^2\right) \\ &= \frac{1}{2t}\left(\|x - x^*\|_2^2 - \|x^+ - x^*\|_2^2\right) \end{aligned} \tag{3}$$

(hence, $\|x^+ - x^*\|_2^2 \le \|x - x^*\|_2^2$, *i.e.*, distance to optimal set decreases)

Analysis for fixed step size

add inequalities (3) for $x = x^{i-1}, x^+ = x^i, t = t_i = 1/L$

$$\begin{aligned}
\sum_{i=1}^{k}(F(x^i) - F^*) &\le \frac{1}{2t}\sum_{i=1}^{k}\left(\|x^{i-1} - x^*\|_2^2 - \|x^i - x^*\|_2^2\right) \\
&= \frac{1}{2t}\left(\|x^0 - x^*\|_2^2 - \|x^k - x^*\|_2^2\right) \\
&\le \frac{1}{2t}\|x^0 - x^*\|_2^2
\end{aligned}$$

since $f(x^i)$ is nonincreasing,

$$F(x^k) - F^* \le \frac{1}{k}\sum_{i=1}^{k}(F(x^i) - F^*) \le \frac{1}{2kt}\|x^0 - x^*\|_2^2$$

conclusion: reaches $F(x^k) - F^* \le \epsilon$ after $O(1/\epsilon)$ iterations

Outline: Accelerated Gradient Method

- 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*

FISTA: Accelerated proximal gradient

Consider the model

$$\min\ F(x) := f(x) + h(x).$$

Given $t = 1/L$, $y^1 = x_0$ and $\gamma^1 = 1$, compute:

$$\begin{aligned} x^k &= \operatorname{prox}_{th}(y^k - t\nabla f(y^k)) \\ \gamma_{k+1} &= \frac{1 + \sqrt{1 + 4\gamma_k^2}}{2} \\ y^{k+1} &= x^k + \frac{\gamma_k - 1}{\gamma_{k+1}}(x^k - x^{k-1}) \end{aligned}$$

Complexity results:

$$F(x^k) - F(x^*) \leq \frac{2L\|x^0 - x^*\|_2^2}{(k+1)^2}$$

APG Variant 1

Acclerated proximal gradient (APG):
Set $x^{-1} = x^0$ and $\theta_{-1} = \theta_0 = 1$:

$$
\begin{aligned}
y^k &= x^k + \theta_k(\theta_{k-1}^{-1} - 1)(x^k - x^{k-1}) \\
x^{k+1} &= \mathrm{prox}_{th}(y^k - t\nabla f(y^k)) \\
\theta_{k+1} &= \frac{\sqrt{\theta_k^4 + 4\theta_k^2} - \theta_k^2}{2}
\end{aligned}
$$

Question: what is the difference between θ_k and γ_k? Show $\theta_k \le \frac{2}{k+2}$ for all k.

Complexity:

$$F(x^k) - F(x^*) \le \frac{4L}{(k+1)^2}\|x^* - x^0\|_2^2$$

APG Variant 2

Another version of APG:

$$
\begin{aligned}
y^k &= (1-\theta_k)x^k + \theta_k z^k \\
z^{k+1} &= \text{prox}_{th}(z^k - t\nabla f(y^k)) \\
x^{k+1} &= (1-\theta_k)x^k + \theta_k z^{k+1} \\
\theta_{k+1} &= \frac{\sqrt{\theta_k^4 + 4\theta_k^2} - \theta_k^2}{2}
\end{aligned}
$$

y^k is a convex combination of x^k and z^k,
x^{k+1} is a convex combination of x^k and z^{k+1}.

Complexity:

$$F(x^k) - F(x^*) \le \frac{4L}{(k+1)^2}\|x^* - z^0\|_2^2$$

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

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

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

  (DL) $\quad \min_{\lambda,s}\ \mathcal{L}(\lambda, s, x^k),$ 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

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

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:

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

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:

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

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:

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

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:

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

$\nu, \rho \ge 0$, $A \in \mathbb{C}^{m\times n}$, $b \in \mathbb{C}^m$,$x \in \mathbb{C}^n$ 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:

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

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:

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

- 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

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

Bregmanized operator splitting:

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

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

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

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:

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

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:

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

Convergence of the linearized scheme?

Orthogonal Matching Pursuit, OMP

- $g^k = A^\top (Ax^{k-1} - b)$
- $x^k = \mathrm{argmin}_x\{\|Ax - b\|_2 : \mathrm{supp}(x) \subseteq \mathcal{S}^k\}$. 如果矩阵$A$满足相关RIP条件，则$A_{\mathcal{S}^k}^\top A_{\mathcal{S}^k}$实际上是可逆的．则等价于$\mathrm{argmin}_x\{\|A_{\mathcal{S}^k} x_{\mathcal{S}^k} - b\|_2 : \mathrm{supp}(x) \subseteq \mathcal{S}^k\}$. 显式解为$x_{\mathcal{S}^k} = (A_{\mathcal{S}^k}^\top A_{\mathcal{S}^k})^{-1} A_{\mathcal{S}^k}^\top b$

Algorithm 1 OMP算法框架

```
输入：A, b, x^0 ∈ R^n, S^0 = ∅, k = 1, 最大迭代次数 k_max.
while k < k_max do
    计算 r^k = A x^{k-1} − b.
    计算 g^k = A^⊤ r^k.
    计算 S^k = S^{k-1} ∪ argmax_i |g_i^k|.
    计算 x^k = argmin_x{‖Ax − b‖_2 : supp(x) ⊆ S^k}.
    k ← k + 1 .
end while
```

CoSaOMP

- $g^k_{2s} = \operatorname{argmin}\{\|x - g^k\|_2 : \|x\|_0 \le 2s\}$是$g^k$的$2s$-逼近

Algorithm 2 CoSaOMP算法框架

输入：$A, r^0, x^0 \in \mathbb{R}^n, \mathcal{S}^0 = \oslash, k = 1,$终止条件$\varepsilon$.
while $\|r^k\| \le \varepsilon$ **do**
    计算$r^k = Ax^{k-1} - b$.
    计算$g^k = A^\top r^k$.
    计算$\mathcal{S}^k = \operatorname{supp}(x^{k-1}) \cup \operatorname{supp}(g^k_{2s})$.
    计算$c = \operatorname{argmin}_x\{\|Ax - b\|_2 : \operatorname{supp}(x) \subseteq \mathcal{S}^k\}$.
    计算$x^k = c_s$.
    $k \leftarrow k + 1$.
end while