Accelerated Proximal Gradient Algorithm for Frame-Based Image Restorations

Sangwoon Yun

Computational Sciences
Korea Institute for Advanced Study

December 3, 2010
KSIAM 2010 Annual Meeting

Outline

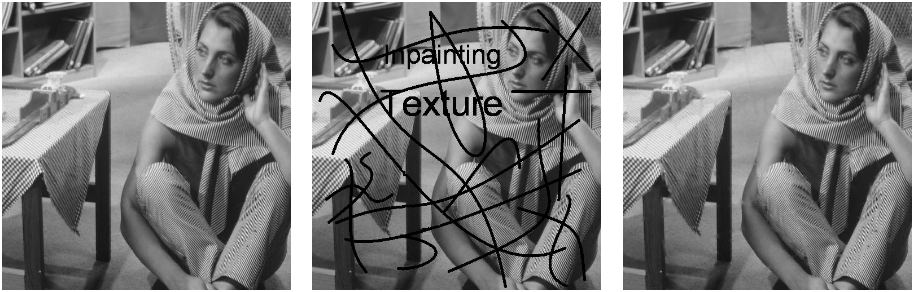
- Image Restorations
- Optimization Formulations
- Accelerated Proximal Gradient Algorithm
- Numerical Experience
- Conclusions

Image Restorations

Figure: Image Deblurring

(a) original: 512×512 (b) motion blurred image (c) recovered image

Figure: Image Inpainting



(a) original: 512×512 (b) scratched image (c) recovered image

Figure: Image Denoising

(a) original: 512 × 512 (b) noised image (c) recovered image

Inverse Problem

Image restorations such as image deconvolution, image inpainting and image denoising is often formulated as an inverse problem.

$$b = Au + \eta,$$

where

- unknown true image $u \in \Re^n$.
- observed image (or measurements) $b \in \Re^\ell$.
- η is a white Gaussian noise with variance σ^2.
- $A \in \Re^{\ell \times n}$ is a linear operator, typically a convolution operator in image deconvolution, a projection in image inpainting and the identity in image denoising.

Frame-Based Approach

- Our proposed approach for image restoration is based on tight frames.
- Tight frames are redundant system in $\Re^n$.
- Suppose $W \in \Re^{m \times n}$ (with $m \geq n$) satisfies $W^T W = I$. Then, the rows of W form a tight frame in $\Re^n$.
- For every vector $u \in \Re^n$,

$$u = W^T(Wu).$$

- The components of the vector Wu are called the canonical coefficients representing u.
- The tight frame system W used is generated from piecewise linear B-spline framelet constructed via the unitary extension principle.
- Mapping from the image u to its coefficients is not one-to-one, i.e., the representation of u in the frame domain is not unique.

Optimization Formulations

Formulations for the sparse approximation of the underlying images

- Analysis Approach: analyzed coefficient vector Wu can be sparsely approximated

$$\min_{u \in \Re^n} \frac{1}{2}\|Au - b\|^2 + \lambda^T |Wu| = \min_{x \in \mathrm{Range}(W)} \frac{1}{2}\|AW^T x - b\|^2 + \lambda^T |x|$$

where λ is a given positive weight vector.

- Synthesis Approach: underlying image u is assumed to be synthesized from a sparse coefficient vector x with $u = W^T x$.

$$\min_{x \in \Re^m} \frac{1}{2}\|AW^T x - b\|^2 + \lambda^T |x|.$$

Balanced Approach

$$\min_{x \in \Re^m} \frac{1}{2}\|AW^T x - b\|^2 + \frac{\kappa}{2}\|(I - WW^T)x\|^2 + \lambda^T |x|,$$

where $\kappa > 0$.

- Objective function of the balanced approach is convex but may not be strictly convex.
- Optimal solution need not be unique.

We consider the Tikhonov regularized balanced approach:

$$\min_{x \in \Re^m} \frac{1}{2}\|AW^T x - b\|^2 + \frac{\kappa}{2}\|(I - WW^T)x\|^2 + \frac{\alpha}{2}\|x\|^2 + \lambda^T |x|,$$

where $\alpha \geq 0$.

Accelerated Proximal Gradient Algorithm

Proximal point mapping of the weighted ℓ_1 norm function

Given g and $\tau > 0$, we consider the proximal point mapping of $\lambda^T|x|$:

$$s_{\lambda/L}(g) \;=\; \arg\min_x \frac{L}{2}\|x - g\|^2 + \lambda^T|x|,$$

where $g = y - \nabla f(y)/L$ with

$$f(y) = \frac{1}{2}\|AW^T x - b\|^2 + \frac{\kappa}{2}\|(I - WW^T)x\|^2 + \frac{\alpha}{2}\|x\|^2.$$

Then

$$s_{\lambda/L}(g) = \mathrm{sgn}(g) \odot \max\{|g| - \lambda/L, 0\},$$

where $\odot$ denotes the component-wise product, i.e., $(x \odot y)_i = x_i y_i$, and

$$\mathrm{sgn}(t) := \begin{cases} +1 & \text{if } t > 0; \\ 0 & \text{if } t = 0; \\ -1 & \text{if } t < 0. \end{cases}$$

The APG algorithm we adapted has the following template:

$$
\begin{aligned}
y^k &= x^k + \tfrac{t^{k-1}-1}{t^k}(x^k - x^{k-1}) \\
g^k &= y^k - \nabla f(y^k)/L \\
x^{k+1} &= s_{\lambda/L}(g^k) \\
t^{k+1} &= \tfrac{1+\sqrt{1+4(t^k)^2}}{2}.
\end{aligned}
$$

- (Beck and Teboulle 08) proposed a fast iterative shrinkage-thresholding algorithm (FISTA) to solve ℓ_1-regularized linear least squares problems.
- It belongs to the class of accelerated proximal gradient algorithms studied earlier by Nesterov, Nemirovski, and others.
- Recently, (Tseng 08) gave a unified treatment.

- When the APG algorithm with $t^k = 1$ for all k, it is the popular iterative shrinkage/thresholding algorithm and it is also the proximal forward-backward splitting algorithm.
- IST and PFBS only require gradient evaluations and soft-thresholding operations, so the computation at each iteration is very cheap.
- But, for any $\epsilon > 0$, those algorithms terminate in $O(L/\epsilon)$ iterations with an ϵ-optimal solution.
- Hence the sequence $\{x^k\}$ converges slowly.

The APG algorithms have an attractive iteration complexity of $O(1/\sqrt{\epsilon})$ for achieving ϵ-optimality.

Convergence Results

Iteration Complexity: Let $\{x^k\}$, $\{y^k\}$, $\{t^k\}$ be the sequences generated by Algorithm APG. Then, for any $k \geq 1$, we have

$$f(x^k) + \lambda^T|x^k| - f(x^*_\alpha) - \lambda^T|x^*_\alpha| \leq \frac{2L\|x^*_\alpha - x^0\|^2}{(k+1)^2}, \qquad \forall x^*_\alpha \in \mathcal{X}^*_\alpha.$$

where $\mathcal{X}^*_\alpha$ is the set of optimal solutions.

Global convergence: Let $\{x^k\}$ be the sequence generated by Algorithm APG. Then $\{x^k\}$ is bounded and each cluster point of the sequence $\{x^k\}$ is an optimal solution. In addition, if $\alpha > 0$, then the sequence $\{x^k\}$ converges to a unique optimal solution.

Minimal Euclidean Norm Solution when $\alpha = 0$

Let x_α^* be the unique solution of Tikhonov regularized balanced approach ($\alpha > 0$) and x_0^* be the unique solution of the following problem:

$$\min_{x \in \mathcal{X}_0^*} \|x\|.$$

Then $\|x_\alpha^*\|$ is a nonincreasing function of α and

$$\lim_{\alpha \downarrow 0} \|x_\alpha^* - x_0^*\| = 0.$$

Numerical Experience

Lipschitz Constant

The Lipschitz constant L of ∇f has the following upper bound:

$$L \leq \lambda_{\max}(A^T A) + \kappa + \alpha.$$

For the inpainting problem, the Lipschitz constant L of ∇f has the following upper bound:

$$L \leq \max\{1, \kappa\} + \alpha.$$

Note that A is a diagonal matrix with diagonals equal to 1 if the corresponding pixel values are known, but 0 otherwise.

Stopping Condition

- The natural stopping condition: $\delta(x) := \mathrm{dist}(0; \partial(f(x) + \lambda^T|x|))$ is sufficiently small, where $\partial(\cdot)$ denotes the sub-differential and

$$\partial|x_i| = \begin{cases} \{+1\} & \text{if } x_i > 0; \\ [-1, 1] & \text{if } x_i = 0; \\ \{-1\} & \text{if } x_i < 0. \end{cases}$$

- At the k-th iteration, we can observe that

$$L(g^k - x^{k+1})(= L(y^k - x^{k+1}) - \nabla f(y^k)) \in \partial(\lambda^T|x^{k+1}|).$$

  Thus we have

$$L(y^k - x^{k+1}) + \nabla f(x^{k+1}) - \nabla f(y^k) \in \partial(f(x^{k+1}) + \lambda^T|x^{k+1}|).$$

- Good upper bound on $\delta(x)$ without incurring extra computational cost:

$$\delta(x^{k+1}) \leq \|L(y^k - x^{k+1}) + \nabla f(x^{k+1}) - \nabla f(y^k)\| \leq 2L\|y^k - x^{k+1}\|.$$

- Condition 1:

$$\frac{2L\|y^{k-1} - x^k\|}{\max\{1, \|x^k\|\}} \leq \mathrm{Tol}.$$

- Condition 2:

$$\frac{|\|AW^T x^k - b\| - \|AW^T x^{k-1} - b\||}{\|AW^T x^k - b\|} \leq \mathrm{Tol}.$$

- Condition 3:

$$\frac{\|x^k - x^{k-1}\|}{\max\{1, \|x^k\|\}} \leq \mathrm{Tol}.$$

For our tests, $\mathrm{Tol} = 5 \times 10^{-4}$ except for the image deblurring problems ($0.2 \times \mathrm{Tol}$ instead of Tol for conditi0n 2 in order to prevent our algorithm from stopping prematurely).

- Continuation Strategy: solving a sequence of ℓ_1LS problems defined by a decreasing sequence $\{\lambda^0, \lambda^1, \ldots, \lambda^\ell = \lambda\}$.
- Initially, we set $\lambda^0 = 10\lambda$, and update $\lambda^k = \max\{0.8\lambda^{k-1}, \lambda\}$ once at every 3 consecutive iterations or whenever the condition 3 is satisfied with $\mathrm{Tol} = 10^{-2}$.
- We set $\alpha = 0.1 e^T \lambda / m^2$.
- $\kappa = 1$.

- Real images usually have two layers, referring to cartoons (the piecewise smooth part of the image) and textures (the oscillating pattern part of the image).
- The layers usually have sparse approximations under different tight frame systems.
- Formulation:

$$\min_{x_1 \in \Re^{m_1}, x_2 \in \Re^{m_2}} \quad \frac{1}{2}\|A(\sum_{i=1}^{2} W_i^T x_i) - b\|^2 + \sum_{i=1}^{2}\left(\frac{\kappa_i}{2}\|(I - W_i W_i^T)x_i\|^2 + \frac{\alpha_i}{2}\|x_i\|^2 + \lambda_i^T |x_i|\right).$$

where, for $i = 1, 2$, $W_i^T W_i = I$, $\kappa_i > 0$, $\alpha_i \geq 0$, $\lambda_i > 0$.

Figure: Image Decomposition

(a) original: 512×512

(b) cartoon part

(c) texture part

Numerical Comparison

image	problem type	APG on Tbal iter/psnr/time	[1] split Bregman on ana	linearized Bregman on syn	proximal forward backward on bal
pepper256	inpainting $\lambda = 0.03$	22/33.69/3.4	51/33.86/10.1		329/33.82/46.0
barbara512	inpainting cartoon & texture $\lambda = 0.01$	31/33.82/34.1	67/33.77/71.5		
barbara512	denoising cartoon & texture $\sigma = 20, \ \lambda = 0.08$	29/28.39/31.5	55/29.01/71.6		
goldhill256	deblur average, 9 $\sigma = 3, \ \lambda = 0.03$	27/26.41/5.0	19/26.40/14.6	11/26.21/7.1	171/26.21/107.3
boat256	deblur disk, 4 $\sigma = 3, \ \lambda = 0.03$	28/25.46/5.8	18/25.30/13.8	12/25.32/7.7	155/25.00/99.9

$$\text{PSNR} = -20 \log_{10} \frac{\|u - \tilde{u}\|}{255N}$$

where u and $\tilde{u}$ are the original and restored images, respectively, and N is the total number of pixels in u.

[1] J.-F. Cai, S. Osher, and Z. Shen, *Split Bregman methods and frame based image restoration*, Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal, to appear.

Conclusions

- Convergence speed of the APG algorithm is competitive to that of the linearized Bregman iteration for the synthesis based approach and the split Bregman iteration for the analysis based approach.
- Demonstrate that this single algorithmic framework can universally handle several image restoration problems, such as image deblurring, denoising, inpainting, and cartoon-texture decomposition.
- The algorithms we implemented are able to restore 512×512 images in various image restoration problems in less than 50 seconds on a modest PC.

Thank You!