Title for Super Fancy Stuff

Generic Name

Some Super Fancy Institution

November 13, 2023

ToC

Frame Title

Formula Presented in Block

$$\min_x g(x) + h(x) \tag{1}$$

1. Throughout this presentation, we assume the objective of a function f is the sum of 2 functions.
2. We are interested in the paper: FISTA (Fast Iterative-Shrinkage Algorithm) by Beck and Teboulle [1].

1. When $h = \delta_Q$ with Q closed and convex with $Q \subseteq \text{ri} \circ \text{dom}(g)$, we use projected subgradient.
2. When g is **strongly smooth** and h is **closed convex proper** whose proximal oracle is easy to compute, we consider the use of FISTA.

Frame Title

Formula Presented in Block

$$\min_x g(x) + h(x) \tag{1}$$

1. Throughout this presentation, we assume the objective of a function f is the sum of 2 functions.
2. We are interested in the paper: FISTA (Fast Iterative-Shrinkage Algorithm) by Beck and Teboulle [1].
1. When $h = \delta_Q$ with Q closed and convex with $Q \subseteq \text{ri} \circ \text{dom}(g)$, we use projected subgradient.
2. When g is **strongly smooth** and h is **closed convex proper** whose proximal oracle is easy to compute, we consider the use of FISTA.

Frame Title

Definition (Definition of Something)

Let f be convex closed and proper, then the proximal operator parameterized by $\alpha > 0$ is a non-expansive mapping defined as:

$$\mathrm{prox}_{f,\alpha}(x) := \arg\min_{y} \left\{ f(y) + \frac{1}{2\alpha}\|y - x\|^2 \right\}.$$

Remark

When f is convex, closed, and proper,

Prox is the Resolvant of Subgradient

Lemma (The Lemma)

When the function f is convex closed and proper, the $\text{prox}_{\alpha,f}$ can be viewed as the following operator $(I + \alpha\partial f)^{-1}$.

Proof.

Minimizer satisfies zero subgradient condition,

$$\begin{aligned}
\mathbf{0} &\in \partial\left[f(y) + \frac{1}{2\alpha}\|y - x\|^2 \middle| y \right](y^+) \\
\mathbf{0} &\in \partial f(y^+) + \frac{1}{\alpha}(y^+ - x) \\
\frac{x}{\alpha} &\in (\partial f + \alpha^{-1} I)(y^+) \\
x &\in (\alpha\partial f + I)(y^+) \\
y &\in (\alpha\partial f + I)^{-1}(x).
\end{aligned}$$

□

Equivalence of Strong Smoothness and Lipschitz Gradient

Theorem (Lipschitz Gradient Equivalence under Convexity)

Suppose g is differentiable on the entire of $\mathbb{E}$. It is closed convex proper. It is strongly smooth with parameter α if and only if the gradient ∇g is globally Lipschitz continuous with a parameter of α and g is closed and convex.

$$\|\nabla g(x) - \nabla g(y)\| \leq \alpha \|x - y\| \quad \forall x, y \in \mathbb{E}$$

Proof.

Using line integral, we can prove Lipschitz gradient implies strong smoothness without convexity. The converse requires convexity and applying generalized Cauchy Inequality to (iv) in Theorem 5.8 for Beck's textbook [2]. □

A Major Assumption

Assumption (Convex Smooth Nonsmooth with Bounded Minimizers)

We will assume that $g : \mathbb{E} \mapsto \mathbb{R}$ *is* **strongly smooth** *with constant* L_g *and* $h : \mathbb{E} \mapsto \bar{\mathbb{R}}$ **is closed convex and proper**. *We define* $f := g + h$ *to be the summed function and* $ri \circ dom(g) \cap ri \circ dom(h) \neq \emptyset$*. We also assume that a set of minimizers exists for the function* f *and that the set is bounded. Denote the minimizer using* $\bar{x}$*.*

The Accelerated Proximal Gradient Method

Momentum Template Method

Algorithm Template Proximal Gradient Method With Momentum

Input: $x^{(0)}, x^{(-1)}, L, h, g$; 2 initial guesses and stepsize L
$y^{(0)} = x^{(0)} + \theta_k(x^{(0)} - x^{(-1)})$
for $k = 1, \cdots, N$ **do**
  $x^{(k)} = \mathrm{prox}_{h,L^{-1}}(y^{(k)} + L^{-1}\nabla g(y^{(k)})) =: \mathcal{P}^{g,h}_{L^{-1}}(y^{(k)})$
  $y^{(k+1)} = x^{(k)} + \theta_k(x^{(k)} - x^{(k-1)})$
end for

Some Fancy Results

The plot of Δ_k:

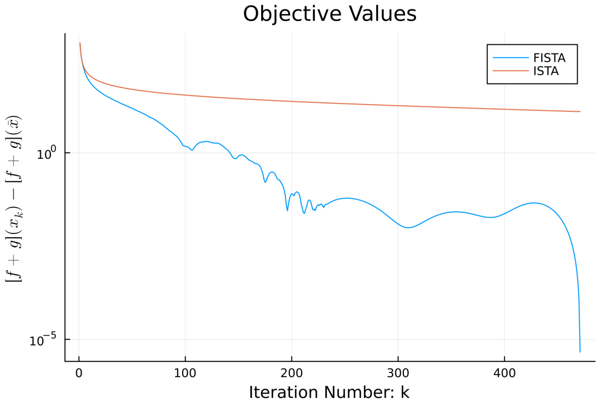
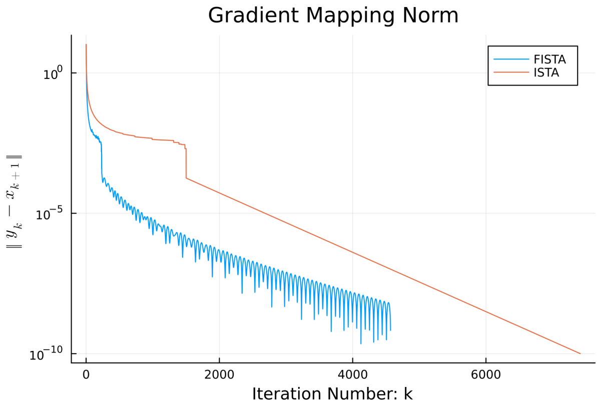


Figure: The left is the objective value of the function during all iterations.

Results

The plot of $\|y^{(k)} - x^{(k+1)}\|_\infty$:



Experiment Setup

Given an image that is convoluted by a Guassian kernel with some guassian noise, we want to recover the image, given the parameters for convolutions.

1. Guassian blur with a discrete 15 by 15 kernel is a linear transform represented by a sparse matrix A in the computer.
2. When an image is 500 by 500 with 3 color channels, A is 750000×750000.
3. Let the noise be on all normalized colors values with $N(0, 10^{-2})$
4. We let $\lambda = \alpha \times (3 \times 500^2)^{-1}$.
5. Implemented in Julia, and the code is too long to be shown here.

One Big Image for Some Fancy Results

We consider blurring the image of a pink unicorn that I own.

Figure: The image is blurred by the Gaussian Blurred matrix A with a tiny amount of noise on the level of 2×10^{-2} that is barely observable. Zoom in to observe the tiny amount of Gaussian noise on top of the blur.

Moer Image to Show Impressive Results

(a) Graph 1 (b) Graph 2 (c) Graph 3

Figure: (a) $\alpha = 0$, without any one norm penalty, is not robust to the additional noise. (b) $\alpha = 0.01$, there is a tiny amount of λ. (c) $\alpha = 0.1$, it is more penalty compared to (a).

References

A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009. [Online]. Available: https://doi.org/10.1137/080716542

A. Beck, *First-Order Methods in Optimization*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2017. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611974997