Accelerated Dual Gradient-Based Methods for Total Variation Image Denoising/Deblurring Problems (and other Inverse Problems)

Donghwan Kim and **Jeffrey A. Fessler**

EECS Department, University of Michigan

IEEE ICASSP 2017

March 7, 2017

Overview

- Total-variation (TV) regularization is useful in many inverse problems, such as "Large-Scale Computational Imaging with Wave Models."

- TV regularized optimization problems are challenging due to:
  - nonseparability of finite-difference operator,
  - nonsmoothness of ℓ_1 norm.

- Variable splitting methods + Proximal gradient methods
  - Split Bregman, ADMM, $\cdots$
  - **"FISTA + Gradient Projection (GP)"**

  [Beck and Teboulle, IEEE TIP, 2009]

- **Goal:** Provide faster convergence for **"FISTA + GP"**
  - eventually: for large-scale inverse problems
  - here: for TV-based image deblurring.

Consider the linear model:

$$\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{b} \in \mathbb{R}^{MN}$ is observed data, $\boldsymbol{A} \in \mathbb{R}^{MN \times MN}$ is a system matrix, $\boldsymbol{x} = \{x_{m,n}\} \in \mathbb{R}^{MN}$ is a true image, and $\boldsymbol{\varepsilon} \in \mathbb{R}^{MN}$ is additive noise.

To estimate image $\boldsymbol{x}$, solve the TV-regularized least-squares problem:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \Phi(\boldsymbol{x}), \quad \Phi(\boldsymbol{x}) := \frac{1}{2}||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||_2^2 + \lambda||\boldsymbol{x}||_{\text{TV}},$$

where the (anistropic) Total Variation (TV) semi-norm uses finite differences:

$$||\boldsymbol{x}||_{\text{TV}} := \sum_{m=1}^{M-1}\sum_{n=1}^{N-1} |x_{m,n} - x_{m+1,n}| + |x_{m,n} - x_{m,n+1}|.$$

FISTA for Inverse Problems

FISTA [Beck and Teboulle, SIIMS, 2009]

Initialize $\boldsymbol{x}_0 = \boldsymbol{\eta}_0$, $L_{\boldsymbol{A}} = ||\boldsymbol{A}||_2^2$, $t_0 = 1$, $\bar{\lambda} := \lambda / L_{\boldsymbol{A}}$.
For $i = 1, 2, \ldots$

$$\bar{\boldsymbol{b}}_i := \boldsymbol{\eta}_{i-1} - \frac{1}{L_{\boldsymbol{A}}} \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{\eta}_{i-1} - \boldsymbol{b}) \qquad \text{(gradient descent step)}$$

$$\boldsymbol{x}_i = \arg\min_{\boldsymbol{x}} \; H_i(\boldsymbol{x}), \qquad H_i(\boldsymbol{x}) = \frac{1}{2} \left\| \boldsymbol{x} - \bar{\boldsymbol{b}}_i \right\|_2^2 + \bar{\lambda} \, ||\boldsymbol{x}||_{\mathrm{TV}}^2$$

$$t_i = \frac{1}{2} \left(1 + \sqrt{1 + 4t_{i-1}^2} \right) \qquad \text{(momentum factors)}$$

$$\boldsymbol{\eta}_i = \boldsymbol{x}_i + \frac{t_{i-1} - 1}{t_i} (\boldsymbol{x}_i - \boldsymbol{x}_{i-1}) \qquad \text{(momentum update)}$$

FISTA decreases the cost function with the optimal rate $O(1/i^2)$:

$$\Phi(\boldsymbol{x}_i) - \Phi(\boldsymbol{x}_*) \leq \frac{2L_{\boldsymbol{A}} ||\boldsymbol{x}_0 - \boldsymbol{x}_*||_2^2}{(i+1)^2}, \text{ where } \boldsymbol{x}_* \text{ is an optimal solution.}$$

However, it is difficult to exactly compute the inner problem for TV.

Inner Denoising Problem of FISTA

For solving inverse problems with FISTA, the inner minimization problem is a "simpler" TV-regularized *denoising* problem:

FISTA's inner TV-regularized denoising problem

$$\boldsymbol{x}_i \approx \arg\min_{\boldsymbol{x}} H_i(\boldsymbol{x}), \quad H_i(\boldsymbol{x}) := \underbrace{\frac{1}{2}\left\|\boldsymbol{x} - \bar{\boldsymbol{b}}_i\right\|_2^2}_{\text{no } \boldsymbol{A}\,!} + \bar{\lambda}\,||\boldsymbol{x}||_{\mathrm{TV}}.$$

Still, no easy solution because

- nonseparability of finite differences,
- absolute value function in TV semi-norm is nonsmooth.

Beck and Teboulle [IEEE TIP, 2009] approach:

- write *dual* of FISTA's inner denoising problem (based on Chambolle [JMIV, 2004])
- apply iterative Gradient Projection (GP) method
- for a finite number of iterations.

Variable Splitting + Duality for Inner Denoising Proble

Rewrite the inner denoising problem of FISTA in composite form:

$$\arg\min_{\boldsymbol{x}} \{H(\boldsymbol{x}) := f(\boldsymbol{x}) + g(\boldsymbol{D}\boldsymbol{x})\}$$

$$f(\boldsymbol{x}) := \frac{1}{2}||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||_2^2, \quad g(\boldsymbol{z}) := \lambda||\boldsymbol{z}||_1,$$

where $g(\boldsymbol{D}\boldsymbol{x}) = \lambda||\boldsymbol{D}\boldsymbol{x}||_1 = \lambda||\boldsymbol{x}||_{\text{TV}}$.

Using variable splitting, an equivalent *constrained* problem is:

$$\arg\min_{\boldsymbol{x}} \min_{\boldsymbol{z}} \left\{\tilde{H}(\boldsymbol{x}, \boldsymbol{z}) := f(\boldsymbol{x}) + g(\boldsymbol{z}) \;:\; \boldsymbol{D}\boldsymbol{x} = \boldsymbol{z}\right\}.$$

Note that $H(\boldsymbol{x}) = \tilde{H}(\boldsymbol{x}, \boldsymbol{D}\boldsymbol{x})$, and $g(\boldsymbol{z})$ is separable unlike $g(\boldsymbol{D}\boldsymbol{x})$.

Variable splitting + Duality for Inner Denoising (cont'd)

To efficiently solve this constrained problem, consider the Lagrangian dual:

$$q(\boldsymbol{y}) := \inf_{\boldsymbol{x},\boldsymbol{z}} \mathcal{L}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{y}) = -f^*(\boldsymbol{D}^\top \boldsymbol{y}) - g^*(-\boldsymbol{y}),$$

$$\mathcal{L}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{y}) := f(\boldsymbol{x}) + g(\boldsymbol{z}) - \langle \boldsymbol{y},\, \boldsymbol{D}\boldsymbol{x} - \boldsymbol{z}\rangle \qquad \text{(Lagrangian)}$$

$$f^*(\boldsymbol{u}) = \max_{\boldsymbol{x}} \{\langle \boldsymbol{u},\, \boldsymbol{x}\rangle - f(\boldsymbol{x})\} \qquad \text{(convex conjugates)}$$

$$g^*(\boldsymbol{y}) = \max_{\boldsymbol{z}} \{\langle \boldsymbol{y},\, \boldsymbol{z}\rangle - g(\boldsymbol{z})\}.$$

For simplicity, define the following convex functions:

$$F(\boldsymbol{y}) := f^*(\boldsymbol{D}^\top \boldsymbol{y}) = \frac{1}{2}||\boldsymbol{D}^\top \boldsymbol{y} + \bar{\boldsymbol{b}}||_2^2 - \frac{1}{2}||\bar{\boldsymbol{b}}||_2^2, \qquad \text{(quadratic)}$$

$$G(\boldsymbol{y}) := g^*(-\boldsymbol{y}) = \begin{cases} 0, & \boldsymbol{y} \in \mathcal{Y}_{\bar{\lambda}} := \{\boldsymbol{y} \;:\; \|\boldsymbol{y}\|_\infty \le \bar{\lambda}\}, \\ \infty, & \text{otherwise}, \end{cases} \qquad \text{(separable)}$$

for an equivalent composite convex function:

$$\tilde{q}(\boldsymbol{y}) := -q(\boldsymbol{y}) = F(\boldsymbol{y}) + G(\boldsymbol{y}).$$

Gradient Projection (GP) for Dual Problem

Dual of inner problem equivalent to solving a constrained quadratic problem:

$$\min_{\boldsymbol{y} \in \mathcal{Y}_{\bar{\lambda}}} F(\boldsymbol{y}), \quad F(\boldsymbol{y}) := f^*(\boldsymbol{D}^\top \boldsymbol{y}) = \frac{1}{2}||\boldsymbol{D}^\top \boldsymbol{y} + \bar{\boldsymbol{b}}||_2^2 - \frac{1}{2}||\bar{\boldsymbol{b}}||_2^2.$$

Quadratic function $F(\boldsymbol{y})$ has Lipschitz continuous gradient with a constant $L := ||\boldsymbol{D}||_2^2$, *i.e.*, for any $\boldsymbol{y}, \boldsymbol{w}$ $||\nabla F(\boldsymbol{y}) - \nabla F(\boldsymbol{w})||_2 \leq L||\boldsymbol{y} - \boldsymbol{w}||_2$.

Separability of ℓ_∞ ball $\mathcal{Y}_{\bar{\lambda}} \Longrightarrow$ GP algorithm natural.

GP for Dual Problem [Chambole, EMMCVPR, 2005]

Initialize $\boldsymbol{y}_0$, $L = ||\boldsymbol{D}||_2^2$.
For $k = 1, 2, \ldots 1$

$$\nabla F(\boldsymbol{y}_{k-1}) = \boldsymbol{D}(\boldsymbol{D}^\top \boldsymbol{y}_{k-1} + \bar{\boldsymbol{b}})$$

$$\boldsymbol{y}_k = \mathrm{p}(\boldsymbol{y}_{k-1}) := \mathcal{P}_{\mathcal{Y}_{\bar{\lambda}}}\left(\boldsymbol{y}_{k-1} - \frac{1}{L}\nabla F(\boldsymbol{y}_{k-1})\right)$$

where $\mathcal{P}_{\mathcal{Y}_{\bar{\lambda}}}(\boldsymbol{y}) := [\min\{|y_l|, \bar{\lambda}\}\,\mathrm{sgn}\{y_l\}]$ projects $\boldsymbol{y}$ onto ℓ_∞ ball $\mathcal{Y}_{\bar{\lambda}}$.

Fast Gradient Projection (FGP) for Dual Problem

GP convergence rate is $O(1/k)$. To accelerate, use FGP (for dual problem).

FGP for Dual Problem [Beck and Teboulle, IEEE TIP, 2009]

Initialize $\boldsymbol{y}_0 = \boldsymbol{w}_0$, $t_0 = 1$.
For $k \geq 1$,

$$\boldsymbol{y}_k = \mathrm{p}(\boldsymbol{w}_{k-1})$$

$$t_k = \frac{1}{2}\left(1 + \sqrt{1 + 4t_{k-1}^2}\right)$$

$$\boldsymbol{w}_k = \boldsymbol{y}_k + \frac{t_{k-1} - 1}{t_k}(\boldsymbol{y}_k - \boldsymbol{y}_{k-1})$$

FGP decreases the dual function with the optimal rate $O(1/k^2)$, *i.e.*,

$$\tilde{q}(\boldsymbol{y}_k) - \tilde{q}(\boldsymbol{y}_*) \leq \frac{2L||\boldsymbol{y}_0 - \boldsymbol{y}_*||_2^2}{(k+1)^2}$$

for an optimal dual solution $\boldsymbol{y}_*$.

Convergence Analysis of the Inner Primal Sequence

More important is convergence rate of the inner primal sequence:

$$\boldsymbol{x}(\boldsymbol{y}) := \boldsymbol{D}^\top \boldsymbol{y} + \bar{\boldsymbol{b}}.$$

[Beck and Teboulle, ORL, 2014] showed the following bounds:

$$||\boldsymbol{x}(\boldsymbol{y}_k) - \boldsymbol{x}_*||_2 \le \left(2(\tilde{q}(\boldsymbol{y}_k) - \tilde{q}(\boldsymbol{y}_*))\right)^{1/2}$$

$$\underbrace{H(\boldsymbol{x}(\boldsymbol{y}_k)) - H(\boldsymbol{x}_*)}_{\therefore\, O(1/k) \text{ for FGP}} \le \gamma_H \Big(2 \underbrace{(\tilde{q}(\boldsymbol{y}_k) - \tilde{q}(\boldsymbol{y}_*))}_{O(1/k^2) \text{ for FGP}}\Big)^{1/2}$$

for $\gamma_H := \max_{\boldsymbol{x}} \max_{\boldsymbol{d} \in \partial H(\boldsymbol{x})} ||\boldsymbol{d}||_2 < \infty$.

FGP has optimal rate $O(1/k^2)$ for the inner dual function decrease.
$\Longrightarrow O(1/k)$ rate for the inner primal sequence.

Next: new algorithm that improves the convergence rate of the inner primal function $H(\boldsymbol{x}(\boldsymbol{y}_k)) - H(\boldsymbol{x}_*)$ to $O(1/k^{1.5})$.

Recap of FISTA for Inverse Problems

FISTA for solving inverse problems

- Momentum to provide fast $O(1/i^2)$ rate for outer loop
- Inner TV denoising problem (challenging)
  - Consider dual of inner denoising problem
  - Algorithms for inner dual problem:
    - GP (slow)
    - FGP (faster due to momentum)
    - Next: new momentum-type algorithms (FGP-OPG, OGP)

New Convergence Analysis of Inner Primal Sequence

New inner primal-dual gap bound (and the inner primal function bound):

$$\begin{aligned} H(\boldsymbol{x}(\mathrm{p}(\boldsymbol{y}))) - H(\boldsymbol{x}_*) &\leq H(\boldsymbol{x}(\mathrm{p}(\boldsymbol{y}))) - q(\mathrm{p}(\boldsymbol{y})) \\ &\leq 2L\,(||\,\mathrm{p}(\boldsymbol{y})||_2 + \gamma_g)\, \underbrace{||\,\mathrm{p}(\boldsymbol{y}) - \boldsymbol{y}||_2}_{\text{gradient projection norm}} \end{aligned}$$

for $\gamma_g := \max_{\boldsymbol{z}} \max_{\boldsymbol{d} \in \partial g(\boldsymbol{z})} ||\boldsymbol{d}||_2 < \infty$. [Kim and Fessler, arXiv:1609.09441]

Recall projected gradient is $p(\boldsymbol{y}) = \mathcal{P}_{\mathcal{Y}_{\bar{\lambda}}} \left(\boldsymbol{y} - \frac{1}{L}\nabla F(\boldsymbol{y})\right)$.

- The rate of decrease of the gradient projection norm $||\,\mathrm{p}(\boldsymbol{y}_k) - \boldsymbol{y}_k||_2$ of *both (!)* GP and FGP is only $O(1/k)$.
- Recent new algorithm FPG-OPG decreases gradient projection norm with rate $O(1/k^{1.5})$ and best known constant.
  [Kim and Fessler, arXiv:1608.03861]
- $\Longrightarrow$ FPG-OPG provider better rates above.

FGP-OPG for Dual Problem

FGP-OPG for Dual Problem [Kim and Fessler, arXiv:1608.03861]

Initialize $\boldsymbol{y}_0 = \boldsymbol{w}_0$, $t_0 = T_0 = 1$

For $k = 1, \ldots, N$,

$$\boldsymbol{y}_k = \mathrm{p}(\boldsymbol{w}_{k-1}) \quad \text{(gradient projection update)}$$

$$t_k = \begin{cases} \frac{1+\sqrt{1+4t_{k-1}^2}}{2}, & k = 1, \ldots, \lfloor \frac{N}{2} \rfloor - 1 \\ \frac{N-k+1}{2}, & \text{otherwise} \end{cases}$$

$$T_k = \sum_{i=0}^{k} t_i \quad \text{(new momentum factor)}$$

$$\boldsymbol{w}_k = \boldsymbol{y}_k + \frac{(T_{k-1} - t_{k-1})t_k}{t_{k-1}T_k}(\boldsymbol{y}_k - \boldsymbol{y}_{k-1}) + \frac{(t_{k-1}^2 - T_{k-1})t_k}{t_{k-1}T_k}(\boldsymbol{y}_k - \boldsymbol{w}_{k-1})$$

(new momentum update)

This becomes FGP for usual t_k choice where $t_k^2 = T_k$ for all k.

Gradient Projection Bound of FGP-OPG

FGP-OPG has the following bound for the "smallest" gradient projection norm:
[Kim and Fessler, arXiv:1608.03861]

$$\min_{\boldsymbol{y} \in \{\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1}, \boldsymbol{y}_N\}} ||\,\mathrm{p}(\boldsymbol{y}) - \boldsymbol{y}||_2 \leq \frac{||\boldsymbol{y}_0 - \boldsymbol{y}_*||_2}{\sqrt{\sum_{k=0}^{N-1}(T_k - t_k^2) + T_{N-1}}} \leq \frac{2\sqrt{6}||\boldsymbol{y}_0 - \boldsymbol{y}_*||_2}{N^{1.5}}.$$

Improves on $O(1/N)$ bound of GP and FGP.

(Using $t_k = \frac{k+a}{a}$ for any $a > 2$ also provides the rate $O(1/k^{1.5})$ without selecting N in advance, unlike FGP-OPG.)

Optimized Gradient Method (OGM)

For unconstrained problem, *i.e.* $G(\boldsymbol{y}) = 0$, the following OGM decreases the (dual) function faster than FGP (in the worst-case).

OGM [Kim and Fessler, Math. Prog., 2016]

Initialize $\boldsymbol{y}_0 = \boldsymbol{w}_0$, $\theta_0 = 1$

For $k = 1, \ldots, N$,

$$\boldsymbol{y}_k = \boldsymbol{w}_{k-1} - \frac{1}{L}\nabla F(\boldsymbol{w}_{k-1})$$

$$\theta_k = \begin{cases} \frac{1+\sqrt{1+4\theta_{k-1}^2}}{2}, & k = 1, \ldots, N-1, \\ \frac{1+\sqrt{1+8\theta_{k-1}^2}}{2}, & k = N, \end{cases}$$

$$\boldsymbol{w}_k = \boldsymbol{y}_k + \frac{\theta_{k-1} - 1}{\theta_k}(\boldsymbol{y}_k - \boldsymbol{y}_{k-1}) + \frac{\theta_{k-1}}{\theta_k}(\boldsymbol{y}_k - \boldsymbol{w}_{k-1})$$

For unconstrained problem, OGM satisfies better bound than FGM:

$$\tilde{q}(\boldsymbol{w}_k) - \tilde{q}(\boldsymbol{y}_*) \le \frac{1L||\boldsymbol{y}_0 - \boldsymbol{y}_*||_2^2}{(k+1)^2}.$$

Projection version of OGM (OGP) [Taylor et al., arXiv:1512.07516]

Initialize $\boldsymbol{y}_0 = \boldsymbol{w}_0 = \boldsymbol{u}_0$, $t_0 = 1$, $\zeta_0 = 1$

For $k = 1, \ldots, N$,

$$\boldsymbol{y}_k = \boldsymbol{w}_{k-1} - \frac{1}{L}\nabla F(\boldsymbol{w}_{k-1})$$

$$\boldsymbol{u}_k = \boldsymbol{y}_k + \frac{\theta_{k-1} - 1}{\theta_k}(\boldsymbol{y}_k - \boldsymbol{y}_{k-1}) + \frac{\theta_{k-1}}{\theta_k}(\boldsymbol{y}_k - \boldsymbol{w}_{k-1}) - \frac{\theta_{k-1} - 1}{\theta_k}\frac{1}{\zeta_{k-1}}(\boldsymbol{w}_{k-1} - \boldsymbol{u}_{k-1})$$

$$\boldsymbol{w}_k = \mathcal{P}_{\mathcal{Y}_{\bar{\lambda}}}(\boldsymbol{u}_k)$$

$$\zeta_k = 1 + \frac{\theta_{k-1} - 1}{\theta_k} + \frac{\theta_{k-1}}{\theta_k}$$

This OGP reduces to OGM when $G(\boldsymbol{y}) = 0$, *i.e.*, $\mathcal{P}_{\mathcal{Y}_{\bar{\lambda}}}(\boldsymbol{y}) = \boldsymbol{y}$.

OGP is "numerically" found to satisfy the bound similar to OGM as

$$\tilde{q}(\boldsymbol{w}_k) - \tilde{q}(\boldsymbol{y}_*) \lessapprox \frac{1L||\boldsymbol{y}_0 - \boldsymbol{y}_*||_2^2}{(k+1)^2}.$$

Image Denoising: Experimental Setup

Generated a noisy image $\boldsymbol{b}$ by adding noise $\epsilon \sim \mathcal{N}(0, 0.1^2)$ to a normalized 512×512 Lena image $\boldsymbol{x}_{\text{true}}$.

True image ($\boldsymbol{x}_{\text{true}}$)

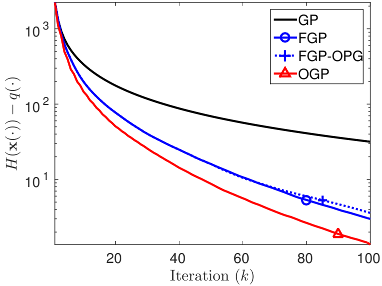
Noisy image ($\boldsymbol{b}$)

Denoise $\boldsymbol{b}$ by solving the following for $\lambda = 0.1$, using its dual:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} H(\boldsymbol{x}), \quad H(\boldsymbol{x}) := \frac{1}{2}||\boldsymbol{x} - \boldsymbol{b}||_2^2 + \lambda||\boldsymbol{x}||_{\text{TV}}.$$

Image Denoising: Primal-Dual Gap vs. Iteration

Denoised image



$H(\boldsymbol{x}(\boldsymbol{y}_k)) - q(\boldsymbol{y}_k)$
or $H(\boldsymbol{x}(\boldsymbol{w}_k)) - q(\boldsymbol{w}_k)$ vs. Iteration (k)

Known Rate	GP	FGP	FGP-OPG	OGP
Dual Function	$O(1/k)$	$O(1/k^2)$	$O(1/k^2)$	$O(1/k^2)$
Primal-Dual Gap	$O(1/k)$	$O(1/k)$	$O(1/k^{1.5})$	$O(1/k)$

- FGP(-OPG) and OGP are clearly faster than GP.
- FGP-OPG is slower than FGP and OGP unlike our worst-case analysis.
- OGP provides a speedup over FGP(-OPG).

Image Deblurring: Experimental Setup

Generated a noisy and blurred image $\boldsymbol{b}$ by using a blurring operator $\boldsymbol{A}$ of 19×19 Gaussian filter with standard deviation 4, and by adding noise $\epsilon \sim \mathcal{N}(0, 0.001^2)$ to a normalized 512×512 Lena image $\boldsymbol{x}_{\text{true}}$.

True image ($\boldsymbol{x}_{\text{true}}$)

Noisy and blurred image ($\boldsymbol{b}$)

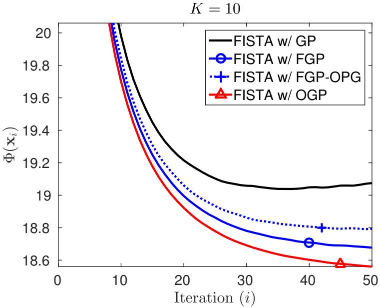
Deblur $\boldsymbol{b}$ by solving the following for $\lambda = 0.005$:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \Phi(\boldsymbol{x}), \quad \Phi(\boldsymbol{x}) := \frac{1}{2}||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||_2^2 + \lambda\,||\boldsymbol{x}||_{\text{TV}}.$$

Image Deblurring: Cost Function vs. Iteration

50 outer iterations (i) of FISTA with $K = 10$ inner iterations (k)

Deblurred image



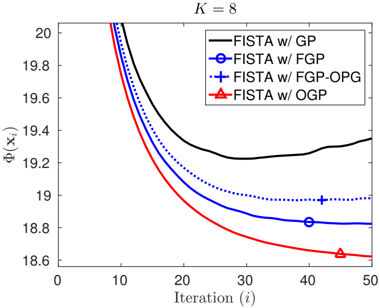
$\Phi(\boldsymbol{x}_i)$ vs. Iteration (i)

- FISTA converges faster with accelerated inner methods than with GP.
- FISTA with FGP-OPG is slower here than with FGP or OGP, unlike our worst-case analysis.
- FISTA with OGP is faster than with FGP(-OPG).
-

Image Deblurring: Cost Function vs. Iteration

50 outer iterations (i) of FISTA with $K = 8$ inner iterations (k)

Deblurred image



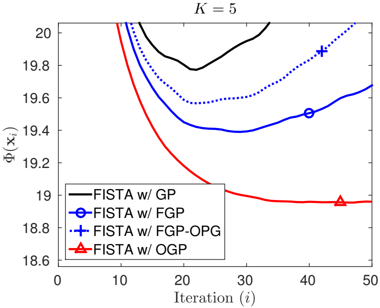
$\Phi(\boldsymbol{x}_i)$ vs. Iteration (i)

- FISTA converges faster with accelerated inner methods than with GP.
- FISTA with FGP-OPG is slower here than with FGP or OGP, unlike our worst-case analysis.
- FISTA with OGP is faster than with FGP(-OPG).
-

Image Deblurring: Cost Function vs. Iteration

50 outer iterations (i) of FISTA with $K = 5$ inner iterations (k)

Deblurred image



$\Phi(\boldsymbol{x}_i)$ vs. Iteration (i)

- FISTA converges faster with accelerated inner methods than with GP.
- FISTA with FGP-OPG is slower here than with FGP or OGP, unlike our worst-case analysis.
- FISTA with OGP is faster than with FGP(-OPG).
- FISTA unstable with too few inner iterations

Summary and Future Work

- We accelerated (in worst-case bound sense) solving the inner denoising problem of FISTA for inverse problems.
- For that inner denoising problem, standard FGP decreases the (inner) primal function with rate $O(1/k)$.
- Proposed FGP-OPG guarantees a faster rate $O(1/k^{1.5})$ for the (inner) primal-dual gap decrease.
- However, FGP-OPG was slower than FGP in the experiment.
- OGP provided acceleration over FGP(-OPG) in the experiment, possibly due to its fast decrease of the (inner) dual function.

- Future work
  - Develop faster gradient projection methods that decrease the function or the gradient projection.
  - Determine if $O(1/k^{1.5})$ is optimal rate for decreasing the gradient projection norm.
  - For TV, compare to parallel proximal algorithm of U. Kamilov. [10]

References

[1] A. Beck and M. Teboulle. "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems." In: *IEEE Trans. Im. Proc.* 18.11 (Nov. 2009), 2419–34.

[2] A. Beck and M. Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems." In: *SIAM J. Imaging Sci.* 2.1 (2009), 183–202.

[3] A. Chambolle. "An algorithm for total variation minimization and applications." In: *J. Math. Im. Vision* 20.1-2 (Jan. 2004), 89–97.

[4] A. Chambolle. "Total variation minimization and a class of binary MRF models." In: *Energy Minimization Methods in Computer Vision and Pattern Recognition, EMMCVPR*. Lecture Notes Comput. Sci., vol. 3757. 2005, 136–52.

[5] A. Beck and M. Teboulle. "A fast dual proximal gradient algorithm for convex minimization and applications." In: *Operations Research Letters* 42.1 (Jan. 2014), 1–6.

[6] D. Kim and J. A. Fessler. *Fast dual proximal gradient algorithms with rate $O(1/k^{1.5})$ for convex minimization*. arxiv 1609.09441. 2016.

[7] D. Kim and J. A. Fessler. *Another look at the "Fast iterative shrinkage/Thresholding algorithm (FISTA)*. arxiv 1608.03861. 2016.

[8] D. Kim and J. A. Fessler. "Optimized first-order methods for smooth convex minimization." In: *Mathematical Programming* 159.1 (Sept. 2016), 81–107.

[9] A. B. Taylor, J. M. Hendrickx, and François Glineur. *Exact worst-case performance of first-order algorithms for composite convex optimization*. arxiv 1512.07516. 2015.

[10] U. S. Kamilov. "A parallel proximal algorithm for anisotropic total variation