AN OPTIMIZED FIRST-ORDER METHOD FOR IMAGE RESTORATION

Donghwan Kim and Jeffrey A. Fessler

University of Michigan, EECS Department, Ann Arbor, MI, 48109, USA

ABSTRACT

First-order methods are used widely for large scale optimization problems in signal/image processing and machine learning, because their computation depends mildly on the problem dimension. Nesterov's fast gradient method (FGM) has the optimal convergence rate among first-order methods for smooth convex minimization; its extension to nonsmooth case, the fast iterative shrinkage-thresholding algorithm (FISTA), also satisfies the optimal rate; thus both algorithms have gained great interest. We recently introduced a new optimized gradient method (OGM) (for smooth convex functions) having a theoretical convergence speed that is $2\times$ faster than Nesterov's FGM. This paper further discusses the convergence analysis of OGM and explores its fast convergence on an image restoration problem using a smoothed total variation (TV) regularizer. In addition, we empirically investigate the extension of OGM to nonsmooth convex minimization for image restoration with l_1-sparsity regularization.

Index Terms— First-order methods, iterative shrinkage-thresholding, optimized gradient method, image restoration.

1. INTRODUCTION

An observed signal/image $\boldsymbol{b} \in \mathbb{R}^d$ is often modeled as:

$$\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}_{\text{true}} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{A} \in \mathbb{R}^{d\times p}$ is a system matrix such as a blur matrix, $\boldsymbol{x}_{\text{true}} \in \mathbb{R}^p$ is an unknown signal/image, and $\boldsymbol{\epsilon} \in \mathbb{R}^d$ is noise. There has been extensive research on recovering $\boldsymbol{x}_{\text{true}}$ from $\boldsymbol{b}$ by formulating optimization problems, and this paper discusses efficient *first-order* optimization methods.

We are interested in minimizing a *convex* function:

$$\boldsymbol{x}_* \in X_*(F) = \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathbb{R}^d} \left\{ F(\boldsymbol{x}) \triangleq f(\boldsymbol{x}) + g(\boldsymbol{x}) \right\}, \tag{2}$$

where $\boldsymbol{x}_*$ denotes a (possibly non-unique) minimizer, $X_*(F)$ denotes the set of minimizers of $F(\boldsymbol{x})$, function $f(\boldsymbol{x})$ is convex and *smooth*, *i.e.*, continuously differentiable with L-Lipschitz continuous gradient:

$$||\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})|| \le L||\boldsymbol{x} - \boldsymbol{y}||, \tag{3}$$

and $g(\boldsymbol{x})$ is convex and possibly *nonsmooth*.

For image restoration problems, one can use analysis models [1] to define the cost function:

$$f(\boldsymbol{x}) \triangleq ||\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}||_2^2 + \lambda R(\boldsymbol{x}), \;\text{ and }\; g(\boldsymbol{x}) = 0 \tag{4}$$

where λ is a regularization parameter, and $R(\boldsymbol{x})$ penalizes differences between neighboring pixels using a *smooth* potential function such as the Huber function. To promote sparsity of the finite-differences of an image $\boldsymbol{x}$, total variation (TV) regularization [1, 2] is used extensively:

$$f(\boldsymbol{x}) \triangleq ||\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}||_2^2, \;\text{ and }\; g(\boldsymbol{x}) = \lambda R_{\text{TV}}(\boldsymbol{x}). \tag{5}$$

Alternatively, one can consider synthesis models [3]:

$$f(\boldsymbol{z}) \triangleq ||\boldsymbol{b} - \boldsymbol{A}\boldsymbol{W}\boldsymbol{z}||_2^2, \;\text{ and }\; g(\boldsymbol{z}) = \lambda||\boldsymbol{z}||_1, \tag{6}$$

where $\boldsymbol{x} = \boldsymbol{W}\boldsymbol{z}$ and the columns of $\boldsymbol{W}$ consist of a basis such as a wavelet basis.

To solve image restoration problems like (4)-(6), many first-order methods suitable for large-scale problems have been developed. This paper focuses on first-order methods that minimize the primal problem (2), such as the gradient method (GM), Nesterov's FGM [4, 5, 6], the iterative shrinkage-thresholding algorithm (ISTA) [3] and accelerated ISTA methods such as TwIST [7] and FISTA [8]; other first-order methods [1, 9] tackle the dual or primal-dual formulation of (2).

Both FGM [4] and FISTA [8] decrease the cost function with rate $O(1/n^2)$, where n counts the number of iterations, achieving the optimal rate [5] and thus were thought to be the best first-order algorithms until recently. Recently, the papers [10, 11] studied a novel framework for developing best-performing first-order methods and described a new method, called OGM [11], that converges twice as fast as FGM (for unconstrained smooth convex minimization).

This paper improves the OGM convergence analysis and investigates its convergence speed on image restoration problems. (Preliminary work in [12] applied OGM to a tomography problem.) Even though the current convergence theory of OGM supports only smooth convex minimization, here we empirically extend OGM to minimize nonsmooth cases such as (5) and (6), investigating possible acceleration from OGM-type algorithms compared to the well-known FISTA method.

This work was supported in part by NIH grant U01 EB 018753 and equipment donations from Intel Corporation.

2. NESTEROV'S FAST GRADIENT METHOD

1: Initialize $\boldsymbol{x}_0 = \boldsymbol{y}_0$ and $t_0 = 1$.
2: **for** $n = 0, 1, \ldots$
3: $\quad \boldsymbol{x}_{n+1} = \boldsymbol{y}_n - \frac{1}{L}\nabla f(\boldsymbol{y}_n)$
4: $\quad t_{n+1} = \frac{1+\sqrt{1+4t_n^2}}{2}$
5: $\quad \boldsymbol{y}_{n+1} = \boldsymbol{x}_{n+1} + \frac{t_n - 1}{t_{n+1}}\left(\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\right)$

Table 1. Nesterov's fast gradient method [4].

Table 1 reviews Nesterov's FGM for smooth convex minimization, *i.e.*, $g(\boldsymbol{x}) = 0$; it is a simple modification of the classical gradient descent method (GM) with an extra momentum term in line 5. FGM is easily transformed to FISTA [8] by replacing line 3 of Table 1 with the following proximal update:

$$\boldsymbol{x}_{n+1} \triangleq \arg\min_{\boldsymbol{x}} \left\{ g(\boldsymbol{x}) + \frac{L}{2} \left\| \boldsymbol{x} - \left(\boldsymbol{y}_n - \frac{1}{L}\nabla f(\boldsymbol{y}_n) \right) \right\|^2 \right\}. \tag{7}$$

In Table 1, both FGM and FISTA reduce to GM and ISTA (or a proximal method in general) when $t_n = 1$ for all n.

A sequence $\{\boldsymbol{x}_n\}$ generated by either FGM (for $g(\boldsymbol{x}) = 0$) or FISTA satisfies the known convergence bound [4, 8]:

$$F(\boldsymbol{x}_n) - F(\boldsymbol{x}_*) \le \frac{2L||\boldsymbol{y}_0 - \boldsymbol{x}_*||^2}{(n+1)^2} \tag{8}$$

for all $n = 0, 1, \ldots$. (The sequence $\{\boldsymbol{y}_n\}$ of FGM is also shown to satisfy the bound (8) in [11].) The $O(1/n^2)$ rate in (8) is known to be optimal because Nesterov exhibited a smooth convex function $f(\boldsymbol{x})$ that satisfies the following lower bound [5]:

$$\frac{3L||\boldsymbol{y}_0 - \boldsymbol{x}_*||^2}{32(n+1)^2} \le f(\boldsymbol{x}_n) - f(\boldsymbol{x}_*) \tag{9}$$

for all $n = 1, \cdots, \left\lfloor \frac{p-1}{2} \right\rfloor$ and *any* first-order method having the general form

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \frac{1}{L}\sum_{k=0}^{n} h_{n+1,k} \nabla f(\boldsymbol{x}_k), \quad n = 0, 1, \ldots \tag{10}$$

with any "step-size" factors $\{h_{n+1,k}\}$.

Even though the $1/n^2$ rate *order* in (8) is optimal, the gap in the *constants* between (8) and (9) (*i.e.*, 2 vs $3/32$) remains non-negligible. The next section summarizes recent first-order methods [10, 11] that provide faster convergence (in the constant factor) than FGM, by optimizing the choice of $\{h_{n+1,k}\}$ in (10).

3. OPTIMIZED GRADIENT METHOD

3.1. Prior work

Seeking to improve on FGM for unconstrained smooth convex minimization, the paper [10] formulated the following interesting approach to designing the best-performing first-order method of the form (10):

$$\min_{\{h_{n+1,k}\}} \max_{\substack{f \in \mathcal{F}_L(\mathbb{R}^p), \\ \boldsymbol{y}_0, \ldots, \boldsymbol{y}_N \in \mathbb{R}^p, \\ \boldsymbol{x}_* \in X_*(f)}} f(\boldsymbol{y}_N) - f(\boldsymbol{x}_*) \tag{P0}$$

$$\text{s.t. } ||\boldsymbol{y}_0 - \boldsymbol{x}_*|| \le C, \text{ and for } n = 0, \ldots, N-1:$$

$$\boldsymbol{y}_{n+1} = \boldsymbol{y}_n - \frac{1}{L}\sum_{k=0}^{n} h_{n+1,k} \nabla f(\boldsymbol{y}_k),$$

where $\mathcal{F}_L(\mathbb{R}^p)$ consists of all p-dimensional real functions that are convex and smooth with L-Lipschitz continuous gradient, and C is a positive constant. Because the problem (P0) seems intractable, [10] relaxed the functional constraint $f \in \mathcal{F}_L(\mathbb{R}^p)$ with the (necessary condition) inequalities [5]:

$$\frac{1}{2L}||\nabla f(\boldsymbol{y}_n) - \nabla f(\boldsymbol{y}_k)||^2 \le f(\boldsymbol{y}_n) - f(\boldsymbol{y}_k) - \langle \nabla f(\boldsymbol{y}_k), \boldsymbol{y}_n - \boldsymbol{y}_k \rangle$$

for all $n, k = 0, \ldots, N, *$, and introduced a series of relaxations to simplify (P0). The relaxed version of (P0) was solved numerically using semidefinite programming [10], but the corresponding first-order algorithm was inefficient computationally because the optimal $\{h_{n+1,k}\}$ values seemed arbitrary in [10].

We recently found an analytic recursive rule for optimizing the relaxed version of (P0) [11], and showed how to write the update (10) with the optimized $\{h_{n+1,k}\}$ of the relaxed (P0) in an equivalent and efficient recursive form as shown in Table 2. Remarkably, the additional computation required for OGM in Table 2 compared to FGM in Table 1 is minimal.

1: Initialize $\boldsymbol{x}_0 = \boldsymbol{y}_0$ and $\theta_0 = 1$.
2: **for** $n = 0, 1, \ldots, N-1$
3: $\quad \boldsymbol{x}_{n+1} = \boldsymbol{y}_n - \frac{1}{L}\nabla f(\boldsymbol{y}_n)$
4: $\quad \theta_{n+1} = \begin{cases} \frac{1+\sqrt{1+4\theta_n^2}}{2}, & n \le N-2 \\ \frac{1+\sqrt{1+8\theta_n^2}}{2}, & n = N-1 \end{cases}$
5: $\quad \boldsymbol{y}_{n+1} = \boldsymbol{x}_{n+1} + \frac{\theta_n - 1}{\theta_{n+1}}\left(\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\right) + \frac{\theta_n}{\theta_{n+1}}\left(\boldsymbol{x}_{n+1} - \boldsymbol{y}_n\right)$

Table 2. Previous optimized gradient method (OGM) [11].

The final iterate $\boldsymbol{y}_N$ generated by OGM in Table 2 satisfies

$$f(\boldsymbol{y}_N) - f(\boldsymbol{x}_*) \le \frac{L||\boldsymbol{y}_0 - \boldsymbol{x}_*||^2}{(N+1)^2}, \tag{11}$$

showing that the cost function (bound) decreases twice as fast for OGM [11] than for the bound (8) of Nesterov's FGM. In other words, OGM requires about $\frac{1}{\sqrt{2}}$-times fewer iterations than FGM to guarantee reaching the same cost function value. However, the inequality (11) only bounds the last iterate $\boldsymbol{y}_N$ and no convergence analysis was provided for the sequence $\{\boldsymbol{x}_n\}$ in [11]. This paper extends [11] by analyzing (*i.e.*, bounding) the convergence speed of $\{\boldsymbol{x}_n\}$ for *all* iterations by formulating an optimization problem with respect to $\boldsymbol{x}_n$ that is inspired by (P0) but differs slightly.

3.2. Proposed work

Here, we propose a new approach to optimizing the choice of the "step-size" factors $\{h_{n+1,k}\}$ in (10) by considering the following alternative to (P0) that also considers the sequence $\{\boldsymbol{x}_n\}$ in the formulation:

$$\min_{\{h_{n+1,k}\}} \max_{\substack{f\in\mathcal{F}_L(\mathbb{R}^p),\\ \boldsymbol{x}_1,\ldots,\boldsymbol{x}_{N+1}\in\mathbb{R}^p,\\ \boldsymbol{y}_0,\ldots,\boldsymbol{y}_N\in\mathbb{R}^p,\\ \boldsymbol{x}_*\in X_*(f)}} f(\boldsymbol{x}_{N+1}) - f(\boldsymbol{x}_*) \quad \text{(P1)}$$

$$\text{s.t. } \boldsymbol{x}_{n+1} = \boldsymbol{y}_n - \frac{1}{L}\nabla f(\boldsymbol{y}_n), \quad n = 0,\ldots,N,$$
$$||\boldsymbol{y}_0 - \boldsymbol{x}_*|| \le C, \text{ and for } n = 0,\ldots,N-1:$$
$$\boldsymbol{y}_{n+1} = \boldsymbol{y}_n - \frac{1}{L}\sum_{k=0}^{n} h_{n+1,k}\nabla f(\boldsymbol{y}_k),$$

Using the inequality [5]:

$$f\left(\boldsymbol{x} - \frac{1}{L}\nabla f(\boldsymbol{x})\right) \le f(\boldsymbol{x}) - \frac{1}{2L}||\nabla f(\boldsymbol{x})||^2,$$

we relax (P1) to the following form:

$$\min_{\{h_{n+1,k}\}} \max_{\substack{f\in\mathcal{F}_L(\mathbb{R}^p),\\ \boldsymbol{y}_0,\ldots,\boldsymbol{y}_N\in\mathbb{R}^p,\\ \boldsymbol{x}_*\in X_*(f)}} f(\boldsymbol{y}_N) - \frac{1}{2L}||\nabla f(\boldsymbol{y}_N)||^2 - f(\boldsymbol{x}_*) \quad \text{(P2)}$$

$$\text{s.t. } ||\boldsymbol{y}_0 - \boldsymbol{x}_*|| \le C, \text{ and for } n = 0,\ldots,N-1:$$
$$\boldsymbol{y}_{n+1} = \boldsymbol{y}_n - \frac{1}{L}\sum_{k=0}^{n} h_{n+1,k}\nabla f(\boldsymbol{y}_k).$$

We have solved the relaxed version of (P2) using methods similar to those used in [11] for solving the relaxed version of (P0). Remarkably, the solution for the optimized step-size factors is very similar to the solution in [11]. Again there is an efficient recursive implementation that is even simpler than OGM in Table 2, as shown in Table 3. The detailed proof will be available in [13], in which we show that sequences $\{\boldsymbol{x}_n\}$ generated by the proposed method in Table 3 satisfy the following inequality:

$$f(\boldsymbol{x}_n) - f(\boldsymbol{x}_*) \le \frac{L||\boldsymbol{y}_0 - \boldsymbol{x}_*||^2}{(n+1)^2} \quad (12)$$

1: Initialize $\boldsymbol{x}_0 = \boldsymbol{y}_0$ and $t_0 = 1$.
2: **for** $n = 0, 1, \ldots$
3: $\quad \boldsymbol{x}_{n+1} = \boldsymbol{y}_n - \frac{1}{L}\nabla f(\boldsymbol{y}_n)$
4: $\quad t_{n+1} = \frac{1+\sqrt{1+4t_n^2}}{2}$
5: $\quad \boldsymbol{y}_{n+1} = \boldsymbol{x}_{n+1} + \frac{t_n - 1}{t_{n+1}}\left(\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\right) + \frac{t_n}{t_{n+1}}\left(\boldsymbol{x}_{n+1} - \boldsymbol{y}_n\right)$

Table 3. Proposed optimized gradient method (OGM).

for all $n = 0, 1, \ldots$. Because of its simplicity and convergence speed, the method in Table 3 is our new recommended first-order algorithm for smooth convex minimization problems. We believe it is the fastest known first-order method for such problems.

We would like to extend the OGM method in Table 3, using the proximal update (7), to handle *nonsmooth* problems such as (5) and (6), just as FISTA is such an extension of FGM. We do not yet have a convergence proof for OGM in nonsmooth cases. In practice one could use heuristic corner rounding methods or more advanced smoothing techniques [6] to convert nonsmooth problems into smooth problems and then apply OGM. However, tight corner rounding corresponds to a large Lipschitz constant L that could slow convergence, so we would like to try to extend OGM to nonsmooth problems.

Even without a convergence proof for the nonsmooth case, it is interesting to extend empirically OGM in Table 3 to nonsmooth optimization by simply replacing line 3 of Table 3 with the proximal update (7). We conjecture that the resulting algorithm will satisfy the convergence bound (12), because a similar relationship holds between FGM and FISTA. The next section illustrates experimentally the potential acceleration provided by an OGM-type update for nonsmooth image restoration optimizations, as well as for a smooth restoration problem where acceleration is expected based on the factor of 2 ratio between (8) and (12).

4. DEBLURRING RESULT

4.1. Experimental setup

We used a 256×256 cameraman image $\boldsymbol{x}_{\text{true}}$ that was normalized to have values within $[0\ 1]$. The noisy blurred data $\boldsymbol{b}$ was generated as described in [8, Section 5.1], where system matrix $\boldsymbol{A}$ is a blur matrix with a (rotationally symmetric) Gaussian filter of size 9×9 and standard deviation 4, and ϵ is zero-mean white Gaussian noise with standard deviation 10^{-3}. Fig. 1 shows both $\boldsymbol{x}_{\text{true}}$ and $\boldsymbol{b}$.

The next two subsections investigate deblurring the image $\boldsymbol{b}$ with a known blur matrix $\boldsymbol{A}$ using the smooth analysis model problem (4) with a Huber regularizer and using the

(a) True image ($\boldsymbol{x}_{\text{true}}$) (b) Noisy blurred image ($\boldsymbol{b}$)

Fig. 1. Noisy deblurred cameraman image.

nonsmooth synthesis model problem (6) with a sparsity regularizer, respectively.

4.2. Smooth analysis model deblurring

Instead of using $|s|$ in the (anistropic) TV-regularizer $R_{\text{TV}}(\boldsymbol{x})$, we used the (smooth) Huber function

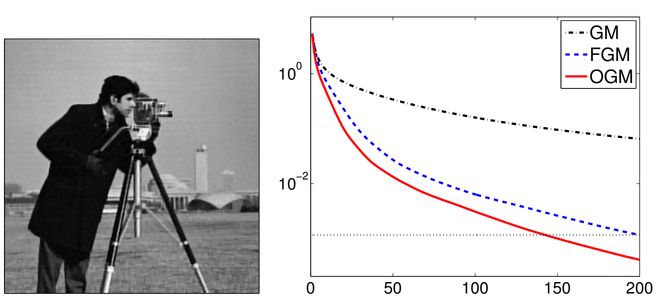
$$|s|_\delta \triangleq \begin{cases} \frac{|s|^2}{2\delta}, & |s| \leq \delta \\ |s| - \frac{\delta}{2}, & |s| > \delta \end{cases} \tag{13}$$

for the finite-difference regularizer

$$R(\boldsymbol{x}) = \sum_{i=1}^{p} \sum_{j \in \mathcal{N}_i} |x_i - x_j|_\delta, \tag{14}$$

where $\mathcal{N}_i$ is a set consisting of two neighboring pixel indices in the horizontal and vertical directions. This approach is known to reduce staircasing artifacts that may be introduced by TV, by properly choosing the constant δ in (13). We chose $\delta = 10^{-2}$ and $\lambda = 10^{-3}$ empirically in this experiment. The overall function is smooth and convex (with Lipschitz constant $L = 1 + 8\lambda$), so we used OGM to minimize the function (4) with (13) and (14), and compared with GM and FGM.

Fig. 2 shows the deblurred image using (4) and the convergence plot of GM, FGM and OGM for minimizing (4).



(a) Deblurred image $\boldsymbol{x}_*$ (b) $F(\boldsymbol{x}_n) - F(\boldsymbol{x}_*)$ vs Iteration (n)

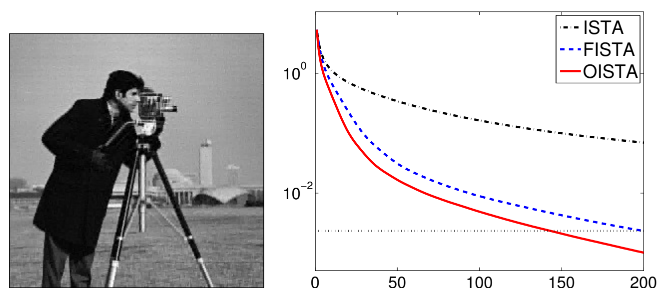
Fig. 2. Smooth analysis model deblurring.

FGM and OGM are clearly faster than GM, and OGM provides speedup compared to FGM, with negligible extra computation. To reach the same cost function value of FGM after 200 iterations, OGM required about 144 iterations, agreeing with the theoretical ratio between (8) and (12).

4.3. Nonsmooth synthesis model deblurring

As a nonsmooth example, we investigated the $l_2 + l_1$ problem (6), where the synthesis matrix $\boldsymbol{W}$ corresponds to a three stage Haar wavelet transform and $\lambda = 2 \times 10^{-5}$ [8, Section 5.1]. We illustrate the acceleration of using our empirical extension of OGM, which we call the optimized iterative shrinkage-thresholding algorithm (OISTA), and compared with ISTA and FISTA. (The proximal update (7) for (6) is a simple shrinkage operator [3].)

Fig. 3 illustrates the deblurred image using (6) and the convergence performance of ISTA, FISTA and OISTA. Even though the proposed OISTA lacks convergence theory, the results show empirically that OISTA converges faster than FISTA as predicted. Similar to Fig. 2, OISTA required about 144 iterations to reach the same cost function value of FISTA after 200 iterations. This suggests that OISTA might satisfy the convergence bound (12) in theory; we leave finding a convergence bound for OISTA as promising future work.



(a) Deblurred image $\boldsymbol{x}_*$ (b) $F(\boldsymbol{x}_n) - F(\boldsymbol{x}_*)$ vs Iteration (n)

Fig. 3. Nonsmooth synthesis model deblurring.

5. CONCLUSION

We discussed the convergence rate analysis of a new first-order method called OGM (Table 3) and applied it to various image restoration problems. The results agreed with the theory that OGM provides two-fold acceleration over FGM (in terms of reducing the cost function value) with minimal additional computation. Empirically we extended OGM to nonsmooth optimization and found potential acceleration from an OGM-type update over the widely used FISTA method. We recommend using OGM in place of FGM in any large-scale first-order optimization problem. Future work is to extend the convergence bounds of OGM to nonsmooth optimization problems.

6. REFERENCES

[1] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Im. Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[2] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithm," *Physica D*, vol. 60, no. 1-4, pp. 259–68, Nov. 1992.

[3] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–57, Nov. 2004.

[4] Y. Nesterov, "A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$," *Dokl. Akad. Nauk. USSR*, vol. 269, no. 3, pp. 543–7, 1983.

[5] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, Kluwer, 2004.

[6] Y. Nesterov, "Smooth minimization of non-smooth functions," *Mathematical Programming*, vol. 103, no. 1, pp. 127–52, May 2005.

[7] J. M. Bioucas-Dias and M. A. T. Figueiredo, "A new twIST: two-step iterative shrinkage/Thresholding algorithms for image restoration," *IEEE Trans. Im. Proc.*, vol. 16, no. 12, pp. 2992–3004, Dec. 2007.

[8] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[9] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Im. Proc.*, vol. 18, no. 11, pp. 2419–34, Nov. 2009.

[10] Y. Drori and M. Teboulle, "Performance of first-order methods for smooth convex minimization: A novel approach," *Mathematical Programming*, vol. 145, no. 1-2, pp. 451–82, June 2014.

[11] D. Kim and J. A. Fessler, "Optimized first-order methods for smooth convex minimization," 2014, arxiv 1406.5468.

[12] D. Kim and J. A. Fessler, "Optimized momentum steps for accelerating X-ray CT ordered subsets image reconstruction," in *Proc. 3rd Intl. Mtg. on image formation in X-ray CT*, 2014, pp. 103–6.

[13] D. Kim and J. A. Fessler, "On the convergence analysis of the optimized gradient method," In preparation.