ACCELERATED DUAL GRADIENT-BASED METHODS FOR TOTAL VARIATION IMAGE DENOISING/DEBLURRING PROBLEMS

Donghwan Kim and Jeffrey A. Fessler

University of Michigan
Dept. of Electrical Engineering and Computer Science

ABSTRACT

We study accelerated dual gradient-based methods for image denoising/deblurring problems based on the total variation (TV) model. For the TV-based denoising problem, combining the dual approach and Nesterov's fast gradient projection (FGP) method has been found effective. The corresponding denoising method minimizes the dual function with FGP's optimal rate $O(1/k^2)$ where k denotes the number of iterations, and guarantees a rate $O(1/k)$ for the primal function decrease. Considering that the dual projected gradient decrease is closely related to the primal function decrease, this paper proposes new accelerated gradient projection methods that decrease the projected gradient norm with a fast rate $O(1/k^{1.5})$ and that are as efficient as FGP. The proposed approach also decreases the primal function with a faster rate $O(1/k^{1.5})$. We provide preliminary results on image denoising/deblurring problems with a TV regularizer, where the fast and efficient denoising solver is iteratively used for solving a deblurring problem as the inner proximal update of a fast iterative shrinkage/thresholding algorithm (FISTA).

***Index Terms*—** Fast dual gradient-based methods, image deblurring, image denoising, total variation.

1. INTRODUCTION

Image denoising and deblurring problems based on the total variation (TV) regularizer [1] have been successfully used in various image processing applications, since TV removes noise in a given image while preserving important edge information. However, unlike image denoising/deblurring problems with a (separable) ℓ_1-regularizer that have a simple iterative shrinkage/thresholding update [2], a TV regularizer is more difficult to optimize due to its non-separability and non-smoothness.

The dual formulation of a TV-regularized least-squares denoising problem becomes a constrained smooth convex separable problem, so using a gradient projection (GP) method leads to a simple update in the dual domain [3]. In [4], that dual method is accelerated using Nesterov's fast gradient projection (FGP) method [5] that decreases the dual function with a fast rate $O(1/k^2)$ where k denotes the number of iterations, improving on the rate $O(1/k)$ of GP. That dual-based FGP provides a $O(1/k)$ rate for the primal function decrease [6, 7] (and for the primal distance decrease [6]). While one cannot improve the rate $O(1/k^2)$ of the dual function decrease [5], one can improve the $O(1/k)$ rate of the primal function decrease of dual gradient-based methods to a faster $O(1/k^{1.5})$ rate as discussed in [7]; this paper applies the new fast dual gradient method of [7] to TV-based imaging problems.

Inspired by [8], we showed in [7] that the dual projected gradient decrease is directly related to the primal function decrease. As a consequence, applying accelerated gradient projection methods that decrease the projected gradient norm with rate $O(1/k^{1.5})$ in [9] to the dual problem decreases the primal function with a fast rate $O(1/k^{1.5})$ [7]. This paper studies the convergence behavior of such dual-based methods for TV-based denoising problems. In addition, we empirically extend the optimized gradient method (OGM) [10] and OGM-OG (OG for optimized over gradient) in [11] to the dual constrained problem. Although there is no known convergence rate theorem for these methods for constrained problems, both have the best known bounds for decreasing the function and gradient of the unconstrained smooth convex problem respectively [11].

In [4], a fast iterative shrinkage/thresholding algorithm (FISTA) is used for solving a TV-regularized deblurring problem, which requires iteratively solving a denoising subproblem. A fast and efficient dual-based FGP solves such denoising subproblems in [4], and we propose to use our accelerated denoising solvers with rate $O(1/k^{1.5})$ to accelerate overall convergence speed for solving the image deblurring problem. We provide preliminary results on TV-based image denoising/deblurring problems.

2. PROBLEM AND METHOD

2.1. Total variation regularization model

We consider a linear model $\mathbf{b} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon}$, where $\mathbf{b} \in \mathbb{R}^{MN}$ is an observed (blurred and noisy) image, $\mathbf{A} \in \mathbb{R}^{MN\times MN}$ is a blurring matrix, $\mathbf{x} = \{x_{m,n}\} \in \mathbb{R}^{MN}$ is a true image,

This research was supported in part by NIH grant U01 EB018753.

and $\boldsymbol{\epsilon} \in \mathbb{R}^{MN}$ is an additive noise, where we assume the (standard) reflexive boundary conditions. Here for simplicity, we represent a 2D image with size $M \times N$ as a vector with length MN.

To recover an image $\mathbf{x}$ from a known $\mathbf{A}$ and an observed $\mathbf{b}$, we consider solving the following least-squares problem with a TV regularizer [1]:

$$\min_{\mathbf{x}} \left\{ \Phi(\mathbf{x}) := \frac{1}{2}||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 + \lambda||\mathbf{x}||_{\mathrm{TV}} \right\}, \tag{P0}$$

where an anisotropic TV semi-norm regularizer is defined as

$$\begin{aligned}||\mathbf{x}||_{\mathrm{TV}} &= \sum_{m=1}^{M-1}\sum_{n=1}^{N-1}\{|x_{m,n} - x_{m+1,n}| + |x_{m,n} - x_{m,n+1}|\} \\ &\quad + \sum_{m=1}^{M-1}|x_{m,N} - x_{m+1,N}| + \sum_{n=1}^{N-1}|x_{M,n} - x_{M,n+1}|.\end{aligned}$$

Note that the analysis in this paper can be easily extended to an isotropic TV regularizer.

2.2. FISTA for a deblurring problem (P0)

For a large-scale image, *i.e.,* large M and N, in (P0), one would prefer using first-order optimization algorithms that are mildly dependent on the problem dimension [12]. One such representative "algorithm" is a proximal gradient method that has the following update:

$$\begin{aligned}\mathbf{x}_i &= \mathrm{p}_L(\mathbf{x}_{i-1}) \\ &:= \mathrm{prox}_{\frac{\lambda}{L}||\cdot||_{\mathrm{TV}}}\left(\mathbf{x}_{i-1} - \frac{1}{L}\mathbf{A}^\top(\mathbf{A}\mathbf{x}_{i-1} - \mathbf{b})\right),\end{aligned} \tag{1}$$

where i denotes the number of iterations, and the Lipschitz constant L is chosen to satisfy $L \geq ||\mathbf{A}||_2^2$ to guarantee a descent update, and the proximity operator is defined as

$$\mathrm{prox}_h(\boldsymbol{\eta}) := \arg\min_{\mathbf{x}} \left\{ h(\mathbf{x}) + \frac{1}{2}||\mathbf{x} - \boldsymbol{\eta}||_2^2 \right\}.$$

In addition, by adding only few additional operations as shown next, its rate $O(1/i)$ for the function decrease can be improved to an optimal rate $O(1/i^2)$ [13]; the corresponding "algorithm" named FISTA [13] is used for image deblurring problem (P0) in [4].

Algorithm 1 FISTA for image deblurring (P0)

1: **Input:** $\mathbf{x}_0 = \boldsymbol{\eta}_0, t_0 = 1, L \geq ||\mathbf{A}||_2^2$.
2: **for** $i \geq 1$ **do**
3: $\quad \mathbf{x}_i = \mathrm{p}_L(\boldsymbol{\eta}_{i-1})$
4: $\quad t_i = \frac{1+\sqrt{1+4t_{i-1}^2}}{2}$
5: $\quad \boldsymbol{\eta}_i = \mathbf{x}_i + \frac{t_{i-1}-1}{t_i}(\mathbf{x}_i - \mathbf{x}_{i-1})$

Even though each proximal gradient update (1) only involves the operations $\mathbf{A}$ and $\mathbf{A}^\top$ rather than an expensive inverse of $\mathbf{A}^\top\mathbf{A}$, the proximal mapping with the TV regularizer is difficult to tackle due to its non-separability and non-smoothness. This computational burden can be circumvented using dual approach in [3], which we review in Section 3.

By defining $\bar{\mathbf{b}} := \boldsymbol{\eta}_{i-1} - \frac{1}{L}\mathbf{A}^\top(\mathbf{A}\boldsymbol{\eta}_{i-1} - \mathbf{b})$ and $\bar{\lambda} := \frac{\lambda}{L}$, the proximal gradient update in line 3 of Alg. 1 (and similarly in (1)) becomes the following TV-based denoising problem:

$$\min_{\mathbf{x}} \left\{ H(\mathbf{x}) := \frac{1}{2}||\mathbf{x} - \bar{\mathbf{b}}||_2^2 + \bar{\lambda}||\mathbf{x}||_{\mathrm{TV}} \right\}. \tag{P1}$$

The next section reviews fast and efficient dual gradient-based methods (dual-based GP and FGP) in [3, 4] for solving "inner" TV-based denoising problem (P1).

3. DUAL-BASED GP AND FGP FOR DENOISING

3.1. Dual approach for "inner" denoising problem (P1)

TV regularizer can be rewritten as $||\mathbf{x}||_{\mathrm{TV}} = ||\mathbf{D}\mathbf{x}||_1$ by defining the differencing matrix $\mathbf{D}\mathbf{x} = \mathbf{z} := \{z_{m,n}^{\mathrm{v}}, z_{m,n}^{\mathrm{h}}\} \in \mathbb{R}^{(M-1)N+M(N-1)}$, where

$$z_{m,n}^{\mathrm{v}} = x_{m,n} - x_{m+1,n},\ m = 1,\ldots,M-1,\ n = 1,\ldots,N,$$
$$z_{m,n}^{\mathrm{h}} = x_{m,n} - x_{m,n+1},\ m = 1,\ldots,M,\ n = 1,\ldots,N-1.$$

Note that the adjoint to $\mathbf{D}$ is given by

$$[\mathbf{D}^\top\mathbf{z}]_{m,n} = z_{m,n}^{\mathrm{v}} - z_{m-1,n}^{\mathrm{v}} + z_{m,n}^{\mathrm{h}} - z_{m,n-1}^{\mathrm{h}}.$$

Using the matrix $\mathbf{D}$, solving (P1) is equivalent to solving

$$\min_{\mathbf{x},\mathbf{z}} \left\{ \tilde{H}(\mathbf{x},\mathbf{z}) := \frac{1}{2}||\mathbf{x} - \bar{\mathbf{b}}||_2^2 + \bar{\lambda}||\mathbf{z}||_1 \ :\ \mathbf{D}\mathbf{x} = \mathbf{z} \right\}, \tag{P1$'$}$$

which has the following (constrained smooth convex) dual problem [6, 7]:

$$\min_{\mathbf{y}} \{\tilde{q}(\mathbf{y}) := F(\mathbf{y}) + G(\mathbf{y})\} \tag{D1}$$

with the corresponding primal variable $\mathbf{x}(\mathbf{y})$ of the dual variable $\mathbf{y}$ defined as [6, 7]:

$$\mathbf{x}(\mathbf{y}) = \mathbf{D}^\top\mathbf{y} + \bar{\mathbf{b}},$$

where $F(\mathbf{y}) = \frac{1}{2}||\mathbf{D}^\top\mathbf{y} + \bar{\mathbf{b}}||_2^2 - \frac{1}{2}||\bar{\mathbf{b}}||_2^2$,

$$G(\mathbf{y}) = \begin{cases} 0, & \mathbf{y} \in \mathcal{Y}_{\bar{\lambda}} := \{\mathbf{y} \ :\ |y_l| \leq \bar{\lambda},\ \forall l\}, \\ \infty, & \text{otherwise}, \end{cases}$$

and the projection operator onto a set $\mathcal{Y}_{\bar{\lambda}}$ is defined as $\mathrm{P}_{\mathcal{Y}_{\bar{\lambda}}}(\mathbf{y}) = \min\{|y_l|, \bar{\lambda}\}\mathrm{sgn}\{y_l\}$. Solving (D1) is equivalent to maximizing the dual function $q(\mathbf{y}) := -\tilde{q}(\mathbf{y})$, and later sections discuss convergence rates on the primal-dual gap decrease $H(\mathbf{x}(\mathbf{y})) - q(\mathbf{y})$. Note that the convergence rate of $H(\mathbf{x}(\mathbf{y})) - q(\mathbf{y})$ determines the convergence rate of the primal function decrease $H(\mathbf{x}(\mathbf{y})) - H(\mathbf{x}_*)$, where $\mathbf{x}_*$ is a solution of (P1).

3.2. Dual-based GP and FGP

The GP update for the dual (D1) is as follows [3, 4, 6, 7]:

$$\mathbf{y}_k = \mathrm{p}_{L_q}^q(\mathbf{y}_{k-1}) := \mathrm{P}_{\mathcal{Y}_{\bar{\lambda}}}\left(\mathbf{y}_{k-1} - \frac{1}{L_q}\mathbf{D}(\mathbf{D}^\top \mathbf{y}_{k-1} + \bar{\mathbf{b}})\right)$$

for $L_q \geq ||\mathbf{D}||_2^2$, which only involves simple operations, where k denotes the number of "inner" iterations. This dual-based GP is accelerated in [4] using Nesterov's FGP [5] that is an instance of FISTA [13], as shown below.

Algorithm 2 DualFGP ($\bar{\mathbf{b}}, \bar{\lambda}, K$) for image denoising (P1)

1: **Initialize:** $\mathbf{y}_0 = \mathbf{w}_0$, $t_0 = 1$, $L_q \geq ||\mathbf{D}||_2^2$.
2: **for** $1 \leq k \leq K$ **do**
3: $\quad\mathbf{y}_k = \mathrm{p}_{L_q}^q(\mathbf{w}_{k-1})$
4: $\quad t_k = \frac{1+\sqrt{1+4t_{k-1}^2}}{2}$
5: $\quad\mathbf{w}_k = \mathbf{y}_k + \frac{t_{k-1}-1}{t_k}(\mathbf{y}_k - \mathbf{y}_{k-1})$
6: **Output:** $\mathbf{x}(\mathbf{y}_K)$

This dual-based FGP (DualFGP) [4] decreases the dual function with an optimal rate $O(1/k^2)$ that is faster than the rate $O(1/k)$ of dual-based GP [3]. However, the convergence rate for the primal function decrease is not discussed in [3, 4]; the next two subsections and Section 4 adapt the analysis in [7] to characterize the convergence rate of DualFGP and the proposed methods.

3.3. Relation between the dual gradient norm and the primal-dual gap

In [7, Lemma 4.4], the primal-dual gap decrease has the following bound, using the fact that the subgradient $\mathbf{d} \in \partial||\mathbf{z}||_1$ is bounded as $||\mathbf{d}||_2 \leq \sqrt{(M-1)N + M(N-1)}$.

Theorem 1. *For any* $\mathbf{y}$ *and a solution* $\mathbf{y}_*$ *of* (D1)*, the primal-dual gap of* (P1) *has the following bound:*

$$H(\mathbf{x}(\mathrm{p}_{L_q}^q(\mathbf{y}))) - q(\mathrm{p}_{L_q}^q(\mathbf{y})) \leq 2L_q\Big(||\mathrm{p}_{L_q}^q(\mathbf{y}) - \mathbf{y}_*||_2 + ||\mathbf{y}_*||_2 + \sqrt{(M-1)N + M(N-1)}\Big)\,||\mathrm{p}_{L_q}^q(\mathbf{y}) - \mathbf{y}||_2.$$

The iterates $\{\mathbf{y}_k\}$ of both GP and FGP satisfy $||\mathrm{p}_{L_q}^q(\mathbf{y}_k) - \mathbf{y}_*|| \leq ||\mathbf{y}_0 - \mathbf{y}_*||$ [7], so the term $||\mathrm{p}_{L_q}^q(\mathbf{y}) - \mathbf{y}||_2$ becomes the factor that affects the primal-dual gap decrease in Thm. 1. We next discuss the primal convergence analysis of dual-based GP and FGP based on Thm. 1.

3.4. Convergence rates of dual-based GP and FGP

We showed in [9] that both GP and FGP decrease the projected gradient norm with rate $O(1/k)$, *i.e.*,

$$||\mathrm{p}_{L_q}^q(\mathbf{y}_k) - \mathbf{y}_k||_2 \leq \frac{2||\mathbf{y}_0 - \mathbf{y}_*||_2}{k}. \qquad (2)$$

This means that the dual-based FGP guarantees a primal function rate of decrease that is no better than that of the dual-based GP, which is curious in light of the empirical acceleration of dual-based FGP for the primal function decrease. (In [6], the primal function rate for dual-based GP was implied to be $O(1/\sqrt{k})$, unlike [7].) This practical acceleration may be explained as follows. First, comparing the algorithms based on worst-case bounds such as (2) (that are not guaranteed to be tight) may not apply to the specific problem (D1). Second, we showed in [11] that Nesterov's fast gradient method (FGP for unconstrained problems) decreases the gradient norm (that is the smallest among all iterates) with rate $O(1/k^{1.5})$, and this rate might hold for the convergence behavior of DualFGP. Whether or not FGP decreases the projected gradient norm with rate $O(1/k^{1.5})$ is an open question, while the rate $O(1/k)$ for GP cannot be improved [9].

3.5. FISTA with DualFGP for image deblurring

DualFGP$\big(\bar{\mathbf{b}}(\boldsymbol{\eta}_{i-1}), \lambda/L, K\big)$ was used in [4] as an approximation of $\mathrm{p}_L(\boldsymbol{\eta}_{i-1})$ in Alg. 1, as described in Alg. 3 below, where the approximation accuracy depends on the choice of the total number of (inner) iterations K.

Algorithm 3 FISTA for (P0) with DualFGP for (P1)

1: **Initialize:** $\mathbf{x}_0 = \boldsymbol{\eta}_0$, $t_0 = 1$, $L \geq ||\mathbf{A}||_2^2$, K.
2: **for** $i \geq 1$ **do**
3: $\quad\mathbf{x}_i = \mathrm{DualFGP}\big(\bar{\mathbf{b}}(\boldsymbol{\eta}_{i-1}), \lambda/L, K\big)$
4: $\quad t_i = \frac{1+\sqrt{1+4t_{i-1}^2}}{2}$
5: $\quad\boldsymbol{\eta}_i = \mathbf{x}_i + \frac{t_{i-1}-1}{t_i}(\mathbf{x}_i - \mathbf{x}_{i-1})$

Due to the inexactness for finite K, there is no known convergence rate bound for Alg. 3, but one could expect roughly $O(1/i^2)$ as for Alg. 1 if K is "sufficiently large". On the other hand, a monotone version of FISTA (MFISTA) was proposed in [4] to heuristically improve stability of Alg. 3 with additional computation of the function value $\Phi(\cdot)$ at each iteration. To further accelerate DualFGP (and thus FISTA with DualFGP in Alg. 3), the next section considers algorithms that decrease the primal-dual gap of (P1) with rate $O(1/k^{1.5})$.

4. PROPOSED METHODS

4.1. Dual-based Generalized FGP

We consider applying the generalized FGP (GFGP) algorithm in [9] to the dual problem (D1); these methods with particular choices of t_k decrease the primal-dual gap with rate $O(1/k^{1.5})$ based on Thm. 1. We call the proposed algorithm the dual-based GFGP (DualGFGP) in this paper and the outline is shown as Alg. 4 on the next page.

The iterates $\mathbf{y} \in \{\mathbf{y}_k, \mathbf{w}_k\}$ of DualGFGP satisfy $||\mathrm{p}_{L_q}^q(\mathbf{y}) - \mathbf{y}_*|| \leq ||\mathbf{y}_0 - \mathbf{y}_*||$ and have the projected gradient bound [9]:

$$\min_{\mathbf{y} \in \{\{\mathbf{w}_l\}_{l=0}^{k-1}, \mathbf{y}_k\}} ||\mathrm{p}_{L_q}^q(\mathbf{y}) - \mathbf{y}||_2 \leq \frac{||\mathbf{y}_0 - \mathbf{y}_*||_2}{\sqrt{\sum_{l=0}^{k-1}(T_l - t_l^2) + T_{k-1}}}. \qquad (3)$$

Algorithm 4 DualGFGP$(\bar{\mathbf{b}}, \bar{\lambda}, K)$ for image denoising (P1)

Initialize: $\mathbf{y}_0 = \mathbf{w}_0$, $t_0 = T_0 = 1$, $L_q \geq ||\mathbf{D}||_2^2$.
for $1 \leq k \leq K$ **do**
  $\mathbf{y}_k = \mathrm{p}_{L_q}^q(\mathbf{w}_{k-1})$
  Choose t_k s.t. $t_k > 0$ and $t_k^2 \leq T_k := \sum_{i=0}^k t_i$.
  $\mathbf{w}_k = \mathbf{y}_k + \frac{(T_{k-1}-t_{k-1})t_k}{t_{k-1}T_k}(\mathbf{y}_k - \mathbf{y}_{k-1})$
  $+ \frac{(t_{k-1}^2 - T_{k-1})t_k}{t_{k-1}T_k}(\mathbf{y}_k - \mathbf{w}_{k-1})$
Output: $\mathbf{x}(\mathbf{y}_K)$

In [9], the dual-based GFGP with the following choice:

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

minimizes the upper bound (3) and guarantees the best known bounds with a rate $O(1/k^{1.5})$ for the projected gradient norm decrease, which we denote dual-based FGP-OPG (OPG for optimized over projected gradient). While we use the FGP-OPG as a representative of the GFGP in the result section, when one does not want to select K in advance one could consider using $t_k = \frac{k+a}{a}$ for any $a > 2$ that also guarantees a rate $O(1/k^{1.5})$ [9]. (Note that DualGFGP reduces to DualFGP when $t_k^2 = T_k$.) Section 5 explores using DualGFGP with $O(1/k^{1.5})$ instead of DualFGP for acceleration.

4.2. Other accelerated dual gradient-based methods

For unconstrained smooth convex problems, OGM [10] and OGM-OG [11] have the best known bounds for the function and gradient norm decrease respectively. We empirically extend them for the constrained problem (D1), and denote their projection versions as OGP and OGP-OG respectively. The resulting dual-based OGP and OGP-OG simply replace the $\mathbf{w}_k$ update of DualGFGP by $\mathbf{w}_k = \mathbf{y}_k + \frac{(T_{k-1}-t_{k-1})t_k}{t_{k-1}T_k}(\mathbf{y}_k - \mathbf{y}_{k-1}) + \frac{(2t_{k-1}^2 - T_{k-1})t_k}{t_{k-1}T_k}(\mathbf{y}_k - \mathbf{w}_{k-1})$, where OGP uses t_k in line 4 of DualFGP and OGP-OG uses t_k in (4).

5. RESULT

5.1. Denoising

We use the (normalized) 512×512 Lena image for the TV-based denoising problem (P1). The noisy image $\bar{\mathbf{b}}$ is generated by adding noise that is normally distributed with zero mean and standard deviation 0.1, and we use $\bar{\lambda} = 0.1$ in (P1). Fig. 1 shows the convergence plot of the primal-dual gap decrease of GP, FGP, FGP-OPG, OGP and OGP-OG for solving the dual (P1). Interestingly, FGP and OGP (OGM) that only guarantees decreasing the (projected) gradient norm with rate $O(1/k)$ [9, 11] is faster than FGP-OPG and OGP-OG with a rate $O(1/k^{1.5})$ respectively. In addition, we see an empirical acceleration using OGP and OGP-OG over FGP and FGP-OPG (and thus GP). We made similar plots for other noise realizations and $\bar{\lambda}$ values and observed the same trends.

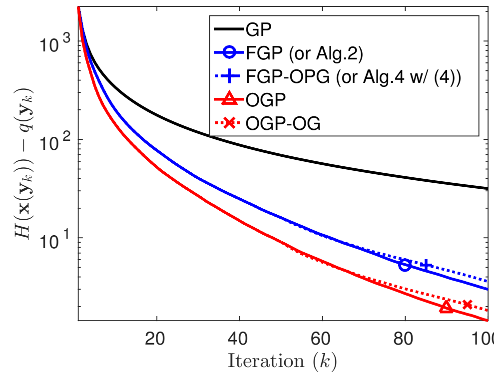

Fig. 1. Image denoising: $H(\mathbf{x}(\mathbf{y}_k)) - q(\mathbf{y}_k)$ vs. Iteration (k)

5.2. Deblurring

We use the same Lena image for the TV-based deblurring problem (P0) where the operator $\mathbf{A}$ uses 19×19 Gaussian filter with standard deviation 4 and a normally distributed noise $\boldsymbol{\epsilon}$ has zero mean and standard deviation 0.001. We choose $\lambda = 0.005$ in (P0). Fig. 2 shows the convergence plot of the function decrease $\Phi(\mathbf{x}_i)$ of Alg. 3 using dual-based GP, FGP-OPG, OGP and OGP-OG as well as dual-based FGP in [4] for solving denoising subproblems with a fixed $K = 10$ (and a warm-start). Since the accuracy of solving the subproblem affects the overall convergence of Alg. 3, OGP behaves the best among all algorithms as expected. Either using MFISTA or increasing K will improve convergence of other algorithms.

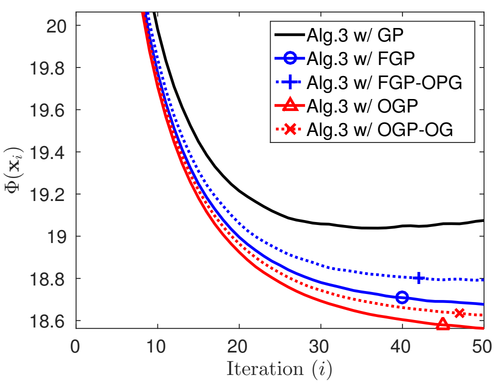

Fig. 2. Image deblurring: $\Phi(\mathbf{x}_i)$ vs. Iteration (i)

6. CONCLUSION

We analyzed the convergence rate of dual gradient-based methods for TV-based image denoising/deblurring problems. In particular, we showed that the proposed dual approach guarantees decreasing the primal function with rate $O(1/k^{1.5})$ that is superior to the existing $O(1/k)$ bound for the dual-based GP and FGP. Although it was unexpected that the proposed approach did not yield any practical acceleration over dual-based FGP despite its improved theoretical bounds, we found empirically that using dual-based OGP and OGP-OG provides practical acceleration over the "conventional" combination of FISTA with DualFGP.

7. REFERENCES

[1] 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.

[2] 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.

[3] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Im. Vision*, vol. 20, no. 1-2, pp. 89–97, Jan. 2004.

[4] 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.

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

[6] A. Beck and M. Teboulle, "A fast dual proximal gradient algorithm for convex minimization and applications," *Operations Research Letters*, vol. 42, no. 1, pp. 1–6, Jan. 2014.

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

[8] Y. Nesterov, "How to make the gradients small," *Optima*, vol. 88, pp. 10–11, 2012.

[9] D. Kim and J. A. Fessler, "Another look at the "Fast Iterative Shrinkage/Thresholding Algorithm (FISTA)"," 2016, arxiv 1608.03861.

[10] D. Kim and J. A. Fessler, "On the convergence analysis of the optimized gradient methods," *J. Optim. Theory Appl.*, 2016.

[11] D. Kim and J. A. Fessler, "Generalizing the optimized gradient method for smooth convex minimization," 2016, arxiv 1607.06764.

[12] V. Cevher, S. Becker, and M Schmidt, "Convex optimization for big data: scalable, randomized, and parallel algorithms for big data analytics," *IEEE Sig. Proc. Mag.*, vol. 31, no. 5, pp. 32–43, Sept. 2014.

[13] 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.