Variational Justification of Cycle Spinning for Wavelet-Based Solutions of Inverse Problems

Ulugbek S. Kamilov, *Student Member, IEEE*, Emrah Bostan, *Student Member, IEEE*, and Michael Unser, *Fellow, IEEE*

***Abstract*—Cycle spinning is a widely used approach for improving the performance of wavelet-based methods that solve linear inverse problems. Extensive numerical experiments have shown that it significantly improves the quality of the recovered signal without increasing the computational cost. In this letter, we provide the first theoretical convergence result for cycle spinning for solving general linear inverse problems. We prove that the sequence of reconstructed signals is guaranteed to converge to the minimizer of some global cost function that incorporates all wavelet shifts.**

***Index Terms*—Cycle spinning, linear inverse problems, wavelet regularization.**

I. Introduction

The problem of estimating an unknown signal from noisy linear observations is fundamental in signal processing. The estimation task is often formulated as the linear inverse problem

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1}$$

where the goal is to reconstruct the unknown signal $\mathbf{x} \in \mathbb{R}^N$ from the noisy measurements $\mathbf{y} \in \mathbb{R}^M$. The matrix $\mathbf{H} \in \mathbb{R}^{M\times N}$ models the response of the acquisition device; the vector $\mathbf{n}$ represents the measurement noise, which is often assumed to be i.i.d. Gaussian. When Problem (1) is ill-posed, the standard approach is to rely on the estimator

$$\widehat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\min} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda\phi(\mathbf{x}) \right\}, \tag{2}$$

where the function ϕ is a regularizer that promotes solutions with desirable properties, and where $\lambda > 0$ controls the degree of regularization. In the wavelet-based framework, regularization is achieved by favoring solutions that have a "sparse" wavelet expansion. A popular approach is to use the non-smooth convex function $\phi(\mathbf{x}) = \|\mathbf{W}\mathbf{x}\|_1$, where $\mathbf{W} \in \mathbb{R}^{N\times N}$ represents a wavelet transform [1]–[3]. Although (2) generally does not admit a closed-form solution with non-quadratic regularizers, it can still be computed efficiently using the iterative shrinkage/thresholding algorithm (ISTA) [1]–[3]. Based on the definition of our forward model, the algorithm can be expressed as

$$\mathbf{z}^t = \widehat{\mathbf{x}}^{t-1} - \gamma_t \mathbf{H}^T\left(\mathbf{H}\widehat{\mathbf{x}}^{t-1} - \mathbf{y}\right) \tag{3a}$$

$$\widehat{\mathbf{x}}^t = \mathbf{W}^T \mathcal{T}\left(\mathbf{W}\mathbf{z}^t; \gamma_t\lambda\right), \tag{3b}$$

where $\gamma_t > 0$ is a step-size that can be determined a priori to ensure convergence of the algorithm [4]. Iterations in (3) combine gradient-descent steps (3a) with pointwise soft-thresholding operations (3b), of the form

$$\mathcal{T}(w;\lambda) = (|w| - \lambda)_+ \mathrm{sgn}(w). \tag{4}$$

Because of its simplicity, ISTA has become the method of choice for finding sparse solutions. More recently, several accelerated versions of ISTA have been developed that provide state-of-the-art rates of convergence [4]–[6].

The theory of wavelet-regularized reconstruction is often formulated with orthogonal wavelet transforms. However, in order to make regularized wavelet-based methods truly competitive, one needs to make the transform shift-invariant. The concept was first introduced in wavelet-based denoising under the name of cycle spinning [7]–[9]. The recursive approach to cycle spinning was presented and analyzed in [10]. Cycle spinning was then applied to more-general linear inverse problems by Figueiredo and Nowak [1]. Currently, it is used in the majority of wavelet-based reconstruction algorithms to obtain higher-quality solutions with less-blocky artifacts [11]–[14]. However, it is rarely accounted for in the accompanying theory.

Let the matrix $\mathbf{W}_k$ denote an orthogonal wavelet transform with k-th shift applied to all the basis functions in $\mathbf{W}$. We consider the K different shifts $\mathbf{W}_1, \dots, \mathbf{W}_K$ that are required to get a shift-invariant version of $\mathbf{W}$. Then, a simple way to implement cycle spinning, without increasing the memory usage, is to consider

$$\mathbf{z}^t = \widehat{\mathbf{x}}^{t-1} - \gamma_t \mathbf{H}^T\left(\mathbf{H}\widehat{\mathbf{x}}^{t-1} - \mathbf{y}\right) \tag{5a}$$

$$\widehat{\mathbf{x}}^t = \mathbf{W}_{k_t}^T \mathcal{T}\left(\mathbf{W}_{k_t}\mathbf{z}^t; \gamma_t\lambda\right), \tag{5b}$$

where $k_t = 1 + (t-1 \bmod K)$ is an iteration-dependent circular shift of the signal. Although Iteration (5) has nearly the same computational cost as (3), it yields results of significantly higher quality, as illustrated in Fig. 2. Also note that (5) is different from the original formulation of cycle spinning in [7], where the thresholded wavelet-coefficients corresponding to different shifts are simply averaged. The scheme in (5) is conceptually closer to recursive cycle spinning that was developed for signal denoising in [10].

Manuscript received May 09, 2014; revised June 23, 2014; accepted June 26, 2014. Date of current version July 09, 2014. This work was supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement 267439. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xudong Jiang.

The authors are with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne VD, Switzerland (e-mail: ulugbek.kamilov@epfl.ch; emrah.bostan@epfl.ch; michael.unser@epfl.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2014.2334306

II. Main Result

The apparent limitation of the cycle-spinning algorithm (5) lies in its greedy nature. At each iteration, the algorithm simply makes a locally optimal step towards the minimization of

$$F_{k_t}(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda\|\mathbf{W}_{k_t}\mathbf{x}\|_1, \tag{6}$$

instead of using the information available from all possible shifts. We are not aware of any prior analysis of the convergence properties of such a scheme. Our main theorem below establishes the first result.

Theorem 1: Let $f(\mathbf{x}) = (1/K)\sum_{k=1}^{K} F_k(\mathbf{x})$, where

$$F_k(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda\|\mathbf{W}_k\mathbf{x}\|_1. \tag{7}$$

Assume that the feasible set $\mathcal{X} \subseteq \mathbb{R}^N$ is nonempty, convex, bounded, and closed. Set $\gamma_t = 1/(L\sqrt{t})$, where L is any constant such that $L > \lambda_{\max}(\mathbf{H}^T\mathbf{H})$. Let $\widehat{\mathbf{x}}^0$ be an arbitrary vector in $\mathcal{X}$, and $(\widehat{\mathbf{x}}^t)_{t\in\mathbb{N}}$ be the sequence generated by (5). Then,

$$\lim_{t\to\infty} f(\widehat{\mathbf{x}}^t) = f^*, \tag{8}$$

where $f^* = \min_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x})$.

The proof is provided in the Appendix.

When $K = 1$, cycle spinning reduces to the standard ISTA, which is known to converge [4]. What our theorem proves is that, by iteratively cycling through K orthogonal wavelets, we are minimizing a cost function that penalizes ℓ_1-norm of all the shifts simultaneously, with the advantage that the underlying regularizer is truly shift-invariant.

III. Experiments

We illustrate the theorem with two simple examples. In the first example, we consider the estimation of a piecewise constant signal of length $N = 128$ corrupted by AWGN corresponding to an input signal-to-noise ratio (SNR) of 30 dB. An interesting property of such signals is that they can be sparsified with the finite-difference operator, which justifies the use of TV regularization [6], [15]. Since the TV regularizer corresponds to an ℓ_1-penalty applied to the finite differences of the signal, our theorem indicates that it can also be minimized with cycle spinning when $\mathbf{W}$ corresponds to the Haar-wavelet basis with one level of decomposition and a zero weight in the lowpass. In Fig. 1, we plot the per-iteration gap $(f(\widehat{\mathbf{x}}^t) - f^*)$, where $\widehat{\mathbf{x}}^t$ is computed with cycle spinning and f is the TV-regularized least-squares cost. We set $\lambda = 0.05$ and, following our analysis, we set the step-size to $\gamma_t = 1/(4\sqrt{t})$. As expected, we observe that, as $t \to \infty$, we have that $(f(\widehat{\mathbf{x}}^t) - f^*) \to 0$. Moreover, we note that, for large t, the slope of $(f(\widehat{\mathbf{x}}^t) - f^*)$ in log-log domain tends to $-1/2$, which indicates the asymptotic rate of convergence $\mathcal{O}(1/\sqrt{t})$.

In the second example, we consider an image-deblurring problem where the Cameraman image of size 256×256 is blurred with a 7×7 Gaussian kernel of standard deviation 2 with the addition of AWGN of variance $\sigma^2 = 10^{-5}$. In Fig. 2, we present the result of the reconstruction with three different methods: standard Haar-domain ℓ_1-regularization, anisotropic TV [6], and cycle spinning with 1D Haar-basis functions applied horizontally and vertically to imitate TV. The regularization parameters for the standard wavelet approach and TV were optimized for the least-error performance. The regularization parameter of cycle spinning was set by rescaling the regularization parameter of TV according to $\lambda_{\mathrm{CS}} = \sqrt{2}K\lambda_{\mathrm{TV}}$, with $K = 4$[1]. Therefore, we expect cycle spinning to again match the TV solution. It is clear from Fig. 2 that cycle spinning outperforms the standard wavelet regularization (improvement of at least 2 dB). As expected, the solution obtained by cycle spinning exactly matches that of TV both visually and in terms of SNR.

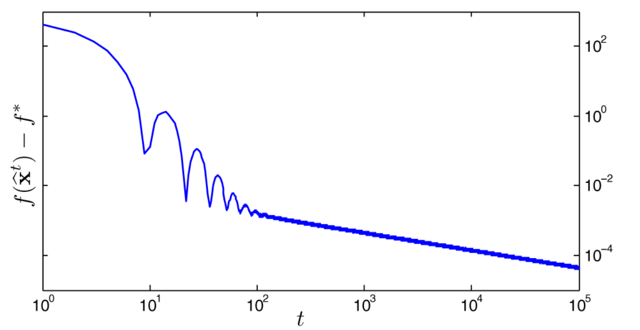

Fig. 1. Estimation of a sparse signal from noisy measurements. We plot the gap $(f(\widehat{\mathbf{x}}^t) - f^*)$ against the iteration number t. The plot illustrates the convergence of cycle spinning to the minimizer of the cost function f.

Fig. 2. Reconstruction of *Cameraman* from blurry and noisy measurements. (a) original, (b) blurry, (c) standard wavelet-based ISTA ($\mathrm{SNR} = 19.10$ dB), (d) reconstruction with TV ($\mathrm{SNR} = 21.58$ dB), (e) wavelet-based reconstruction with cycle spinning ($\mathrm{SNR} = 21.58$ dB).

[1]The goal is for cycle spinning to be minimizing exactly the same cost function as TV. Thus, the factor K is due to the number of shifts, while the factor $\sqrt{2}$ is due to the normalization of the Haar wavelets.

IV. Conclusion and Future Work

We have established the convergence result of the popular cycle-spinning technique for solving linear inverse problems. We have proved that the algorithm converges to the minimizer of a regularized least-squares cost function where the regularizer is the ℓ_1-norm of translation-invariant wavelet coefficients in the analysis form [16].

One can imagine numerous possible extensions of our results. The analysis presented in this letter was deterministic; an interesting avenue for future research would be to see if it also holds in the stochastic setting, where k_t would be generated randomly at each iteration. Our analysis indicates that the rate of convergence of cycle spinning is no worse than $\mathcal{O}(1/\sqrt{t})$. A possible direction of research is to search for a faster convergence rate by using various standard acceleration techniques for ISTA algorithms [4].

V. Appendix

A. Useful Facts from Convex Analysis

Before embarking on the actual proof of Theorem 1, it is convenient to summarize a few facts that will be used next.

A subgradient of a convex function f at $\mathbf{x}$ is any vector $\mathbf{g}$ that satisfies the inequality $f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle$, for all $\mathbf{y}$. When f is differentiable, the only possible choice for $\mathbf{g}$ is $\nabla f(\mathbf{x})$. The set of subgradients of f at $\mathbf{x}$ is the subdifferential of f at $\mathbf{x}$, denoted $\partial f(\mathbf{x})$. The condition that $\mathbf{g}$ be a subgradient of f at $\mathbf{x}$ can then be written $\mathbf{g} \in \partial f(\mathbf{x})$.

The proximal operator is defined as

$$\begin{aligned} \mathbf{x} &= \mathrm{prox}_{\gamma r}(\mathbf{z}) \\ &= \underset{\mathbf{x} \in \mathbb{R}^N}{\arg\min} \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 + \gamma r(\mathbf{x}) \right\}, \end{aligned} \tag{9}$$

where $\gamma > 0$ and r is a convex continuous function. The proximal operator is characterized by the following inclusion, for all $\mathbf{x}, \mathbf{z} \in \mathbb{R}^N$:

$$\mathbf{x} = \mathrm{prox}_{\gamma r}(\mathbf{z}) \Leftrightarrow \mathbf{z} - \mathbf{x} \in \gamma \partial r(\mathbf{x}). \tag{10}$$

We say that an operator $\mathrm{T} : \mathcal{X} \to \mathcal{X}$ is nonexpanding if, for all $\mathbf{x}, \mathbf{z} \in \mathcal{X}$, it satisfies $\|\mathrm{T}\mathbf{x} - \mathrm{T}\mathbf{z}\|_2 \leq \|\mathbf{x} - \mathbf{z}\|_2$. Note that the proximal operator is nonexpanding.

Next, we present the Browder-Göhde-Kirk's fixed-point theorem. It is a standard theorem in convex analysis (see [17, Theorem 4.19])

Theorem 2: Let $\mathcal{X}$ be a nonempty bounded closed convex subset of $\mathbb{R}^N$ and let $\mathrm{T} : \mathcal{X} \to \mathcal{X}$ be a nonexpansive operator. Then, the operator T has a nonempty set of fixed points or that $\mathrm{Fix}\ \mathrm{T} \neq \emptyset$.

The Krasnoselskii–Mann theorem will also be of use (see [17, Theorem 5.14]).

Theorem 3: Let $\mathcal{X}$ be a nonempty closed convex subset of $\mathbb{R}^N$, let $\mathrm{T} : \mathcal{X} \to \mathcal{X}$ be a nonexpansive operator such that $\mathrm{Fix}\ \mathrm{T} \neq \emptyset$, let $(\theta_t)_{t \in \mathbb{N}}$ be a sequence in (0,1) such that $\sum_{t \in \mathbb{N}} \theta_t (1 - \theta_t) = \infty$, and let $\mathbf{x}^0 \in \mathcal{X}$. Set $\widehat{\mathbf{x}}^t = \widehat{\mathbf{x}}^{t-1} + \theta_t (\mathrm{T}\widehat{\mathbf{x}}^{t-1} - \widehat{\mathbf{x}}^{t-1})$. Then, the sequence $(\widehat{\mathbf{x}}^t)_{t \in \mathbb{N}}$ converges to a point in $\mathrm{Fix}\ \mathrm{T}$.

B. General Model

For the purpose of our analysis, we consider the cost function

$$f(\mathbf{x}) = \frac{1}{K} \sum_{k=1}^{K} F_k(\mathbf{x}) = d(\mathbf{x}) + \frac{1}{K} \sum_{k=1}^{K} R_k(\mathbf{x}). \tag{11}$$

It is more general than the cost function of Theorem 1, which it includes as a special case. We shall also make the following assumptions:

- The feasible set $\mathcal{X} \subseteq \mathbb{R}^N$ is nonempty, convex, bounded, and closed. There exists $D > 0$ such that for all $\mathbf{x}, \mathbf{z} \in \mathcal{X}, \|\mathbf{x} - \mathbf{z}\|_2 \leq D$.
- $d : \mathbb{R}^N \to \mathbb{R}$ is convex and differentiable, with a Lipschitz-continuous gradient. There exists $L > 0$ such that, for all $\mathbf{x}, \mathbf{z} \in \mathcal{X}, \|\nabla d(\mathbf{x}) - \nabla d(\mathbf{z})\|_2 \leq L \|\mathbf{x} - \mathbf{z}\|_2$.
- For each k, R_k is a continuous, convex function that is possibly nondifferentiable.
- The gradient of d and the subgradients of R_k are bounded. There exists $G > 0$ such that, for all k and all $\mathbf{x} \in \mathcal{X}$, $\|\nabla d(\mathbf{x})\|_2 \leq G$ and $\|\partial_x R_k(\mathbf{x})\|_2 \leq G$.

Note that for $R_k(\mathbf{x}) = \|\mathbf{W}_k \mathbf{x}\|_1$, we have $\partial_x \|\mathbf{W}_k \mathbf{x}\|_1 = \{\mathbf{W}_k^T \mathbf{g} | \|\mathbf{g}\|_\infty \leq 1, \langle \mathbf{W}_k^T \mathbf{g}, \mathbf{x} \rangle = \|\mathbf{W}_k \mathbf{x}\|_1\}$, which indeed implies that the subgradients are bounded [18]. Then, the Algorithm (5) is a special case of the proximal-gradient iteration

$$\widehat{\mathbf{x}}^t = \mathrm{prox}_{\gamma_t R_{k_t}} \left(\widehat{\mathbf{x}}^{t-1} - \gamma_t \nabla d(\widehat{\mathbf{x}}^{t-1}) \right), \tag{12}$$

where $\mathrm{prox}_{\gamma_t R_{k_t}}$ is the proximal-map associated with $\gamma_t R_{k_t}$.

C. Main Technical Lemmas

The proof of Theorem 1 relies on two lemmas that we shall prove now.

D. Lemma 1

For all $t = 1, 2, \ldots$, and for any $\mathbf{x}^* \in \mathcal{X}$, we have that

$$\begin{aligned} &F_{k_t}\left(\widehat{\mathbf{x}}^t\right) - F_{k_t}\left(\mathbf{x}^*\right) \\ &\leq \frac{1}{2\gamma_t} \left(\|\widehat{\mathbf{x}}^{t-1} - \mathbf{x}^*\|_2 - \|\widehat{\mathbf{x}}^t - \mathbf{x}^*\|_2 \right) + 6\gamma_t G^2. \end{aligned} \tag{13}$$

Proof: The optimality condition of (12) implies that there must exist a vector $\mathbf{g}^t \in \partial_x R_{k_t}(\widehat{\mathbf{x}}^t)$ such that

$$\widehat{\mathbf{x}}^t = \widehat{\mathbf{x}}^{t-1} - \gamma_t \left(\nabla d(\widehat{\mathbf{x}}^{t-1}) + \mathbf{g}^t \right). \tag{14}$$

Then, we can write

$$\begin{aligned} \|\widehat{\mathbf{x}}^t - \mathbf{x}^*\|_2^2 &= \|\widehat{\mathbf{x}}^{t-1} - \gamma_t \left(\nabla d(\widehat{\mathbf{x}}^{t-1}) + \mathbf{g}^t \right) - \mathbf{x}^*\|_2^2 \\ &= \|\widehat{\mathbf{x}}^{t-1} - \mathbf{x}^*\|_2^2 - 2\gamma_t \langle \nabla d(\widehat{\mathbf{x}}^{t-1}) + \mathbf{g}^t, \widehat{\mathbf{x}}^{t-1} - \mathbf{x}^* \rangle \\ &+ \gamma_t^2 \|\nabla d(\widehat{\mathbf{x}}^{t-1}) + \mathbf{g}^t\|_2^2. \end{aligned} \tag{15}$$

By using the triangle inequality, we can bound the last term as

$$\|\nabla d(\widehat{\mathbf{x}}^{t-1}) + \mathbf{g}^t\|_2^2 \leq 4G^2. \tag{16}$$

To bound the second term we proceed in two steps. We first write that

$$\begin{aligned}\langle\nabla d(\widehat{\mathbf{x}}^{t-1}),\widehat{\mathbf{x}}^{t-1}-\mathbf{x}^*\rangle &\overset{\text{(a)}}{\geq} d(\widehat{\mathbf{x}}^{t-1})-d(\mathbf{x}^*)\\ &\overset{\text{(b)}}{\geq} d(\widehat{\mathbf{x}}^{t})-\langle\nabla d(\widehat{\mathbf{x}}^{t}),\widehat{\mathbf{x}}^{t}-\widehat{\mathbf{x}}^{t-1}\rangle-d(\mathbf{x}^*)\\ &= d(\widehat{\mathbf{x}}^{t})-d(\mathbf{x}^*)-\langle\nabla d(\widehat{\mathbf{x}}^{t}),-\gamma_t\left(\nabla d(\widehat{\mathbf{x}}^{t-1})+\mathbf{g}^t\right)\rangle\\ &\overset{\text{(c)}}{\geq} d(\widehat{\mathbf{x}}^{t})-d(\mathbf{x}^*)-2\gamma_t G^2,\end{aligned}\tag{17}$$

where in (a) and (b) we used the convexity of d, in (c) we used the Cauchy-Schwarz inequality and the bound on the gradient. Then, in a similar fashion, we can also write that

$$\begin{aligned}\langle\mathbf{g}^t,\widehat{\mathbf{x}}^{t-1}-\mathbf{x}^*\rangle &= \langle\mathbf{g}^t,\widehat{\mathbf{x}}^{t}-\mathbf{x}^*\rangle-\langle\mathbf{g}^t,\widehat{\mathbf{x}}^{t}-\widehat{\mathbf{x}}^{t-1}\rangle\\ &\overset{\text{(a)}}{\geq} R_{k_t}(\widehat{\mathbf{x}}^{t})-R_{k_t}(\mathbf{x}^*)-\langle\mathbf{g}^t,-\gamma_t\left(\nabla d(\widehat{\mathbf{x}}^{t-1})+\mathbf{g}^t\right)\rangle\\ &\overset{\text{(b)}}{\geq} R_{k_t}(\widehat{\mathbf{x}}^{t})-R_{k_t}(\mathbf{x}^*)-2\gamma_t G^2,\end{aligned}\tag{18}$$

where in (a) we used the convexity of R_{k_t}, in (b) we used the Cauchy-Schwarz inequality followed with a bound on the gradient.

By plugging (16), (17), and (18) into (15), by reorganizing the terms, and by using the definition $F_{k_t}(\mathbf{x})=d(\mathbf{x})+R_{k_t}(\mathbf{x})$, we obtain the claim. ■

Lemma 2: With $(\widehat{\mathbf{x}}^t)_{t\in\mathbb{N}}$ in $\mathcal{X}$ and $\bar{\mathbf{x}}\in\mathcal{X}$, let $\widehat{\mathbf{x}}^t\to\bar{\mathbf{x}}$. Then,

$$\lim_{n\to\infty}\left\{\frac{1}{nK}\sum_{t=1}^{nK}F_{k_t}(\widehat{\mathbf{x}}^t)\right\}=f(\bar{\mathbf{x}}).\tag{19}$$

Proof: We introduce the shorthand notation $\delta_t = F_{k_t}(\widehat{\mathbf{x}}^t)-F_{k_t}(\bar{\mathbf{x}})$. The convergence of $\widehat{\mathbf{x}}^t$ and the continuity of F_k imply that, for a given $\epsilon>0$, there exists an m such that, for all $t>mK$, $|\delta_t|=|F_{k_t}(\widehat{\mathbf{x}}^t)-F_{k_t}(\bar{\mathbf{x}})|<\epsilon/2$. Then,

$$\begin{aligned}\left|\frac{1}{nK}\sum_{t=1}^{nK}F_{k_t}(\widehat{\mathbf{x}}^t)-f(\bar{\mathbf{x}})\right| &= \left|\frac{1}{nK}\sum_{t=1}^{nK}\left(F_{k_t}(\widehat{\mathbf{x}}^t)-F_{k_t}(\bar{\mathbf{x}})\right)\right|\\ &\leq \frac{1}{nK}\left|\sum_{t=1}^{mK}\delta_t\right|+\frac{1}{nK}\left|\sum_{t=mK+1}^{nK}\delta_t\right|\\ &\leq \frac{1}{nK}\sum_{t=1}^{mK}|\delta_t|+\frac{1}{nK}\sum_{t=mK+1}^{nK}|\delta_t|\\ &\leq \frac{m}{n}\left(\max_{t\in[1\ldots mK]}|\delta_t|\right)+\left(1-\frac{m}{n}\right)\left(\max_{t\in[mK+1\ldots nK]}|\delta_t|\right).\end{aligned}\tag{20}$$

Now, considering $\tilde{n}\geq(2m/\epsilon)(\max_{t\in[1\ldots mK]}|\delta_t|)$ and realizing that the second term is bounded by $\epsilon/2$, we conclude that, for a given $\epsilon>0$, there exists $\tilde{n}$ such that, for all $n>\tilde{n}$,

$$\left|\frac{1}{nK}\sum_{t=1}^{nK}F_{k_t}(\widehat{\mathbf{x}}^t)-f(\bar{\mathbf{x}})\right|<\epsilon.\tag{21}$$

■

E. Proof of Theorem 1

Let $\mathbf{x}^*$ denote a minimizer of f. We introduce the shorthand notation $\Delta_t=\|\widehat{\mathbf{x}}^t-\mathbf{x}^*\|_2^2$. By following an approach similar to [19], we sum the bound in Lemma 1

$$\begin{aligned}\sum_{t=1}^{nK}\left(F_{k_t}(\widehat{\mathbf{x}}^t)-F_{k_t}(\mathbf{x}^*)\right) &= \sum_{t=1}^{nK}F_{k_t}(\widehat{\mathbf{x}}^t)-nKf(\mathbf{x}^*)\\ &\leq \frac{1}{2}\sum_{t=1}^{nK}\frac{1}{\gamma_t}\left(\Delta_{t-1}-\Delta_t\right)+6G^2\sum_{t=1}^{nK}\gamma_t\\ &\leq \frac{1}{2\gamma_1}\Delta_0+\frac{1}{2}\sum_{t=1}^{nK-1}\left(\frac{1}{\gamma_{t+1}}-\frac{1}{\gamma_t}\right)\Delta_t+6G^2\sum_{t=1}^{nK}\gamma_t\\ &\overset{\text{(a)}}{\leq} \frac{D^2}{2\gamma_1}+\frac{D^2}{2}\sum_{t=1}^{nK-1}\left(\frac{1}{\gamma_{t+1}}-\frac{1}{\gamma_t}\right)+6G^2\sum_{t=1}^{nK}\gamma_t\\ &\leq \frac{D^2}{2\gamma_{nK}}+6G^2\sum_{t=1}^{nK}\gamma_t,\end{aligned}\tag{22}$$

where in (a) we used the the boundedness of $\mathcal{X}$. Then, by choosing $\gamma_t=1/(L\sqrt{t})$, by using the bound $\sum_{t=1}^{T}1/\sqrt{t}\leq 2\sqrt{T}$, and by dividing the two sides of inequality by nK, we obtain

$$\frac{1}{nK}\sum_{t=1}^{nK}F_{k_t}(\widehat{\mathbf{x}}^t)-f(\mathbf{x}^*)\leq\frac{C}{\sqrt{n}},\tag{23}$$

where the constant C is given by

$$C=\frac{LD^2}{2\sqrt{K}}+\frac{12G^2}{L\sqrt{K}}.\tag{24}$$

To complete the proof, we argue that the sequence $(\widehat{\mathbf{x}}^t)_{t\in\mathbb{N}}$ converges to a fixed point in $\mathcal{X}$. On one hand, note that, due to the nonexpansiveness of $\mathrm{prox}_{\gamma_t R_{k_t}}$ and the Lipschitz property of d, the operator $\mathrm{T}_k:\mathcal{X}\to\mathcal{X}$,

$$\mathrm{T}_k\mathbf{x}=\mathrm{prox}_{\gamma R_k}\left(\mathbf{x}-\gamma\nabla d(\mathbf{x})\right)\tag{25}$$

is nonexpanding for any $\gamma\in(0,1/L)$. Therefore, the composition $\mathrm{T}=\mathrm{T}_K\cdots\mathrm{T}_1$ is also nonexpanding. Then, from Theorem 2, we know that there exists at least one fixed-point of T in $\mathcal{X}$. On the other hand, for our choice $\gamma_t=1/(L\sqrt{t})$, Theorem 3 implies that the subsequence generated via $\widehat{\mathbf{x}}^{Kt}=\mathrm{T}\widehat{\mathbf{x}}^{K(t-1)}$ converges to the fixed-point $\bar{\mathbf{x}}\in\mathrm{Fix}\ \mathrm{T}$. Since, $\gamma_t\to 0$, we have that $(\widehat{\mathbf{x}}^{t-1}-\mathrm{T}_{k_t}\widehat{\mathbf{x}}^{t-1})\to 0$ and conclude that $\widehat{\mathbf{x}}^t\to\bar{\mathbf{x}}$. Finally, this allows us to show that

$$\begin{aligned}0 &\overset{\text{(a)}}{\leq} f(\bar{\mathbf{x}})-f(\mathbf{x}^*)\\ &\overset{\text{(b)}}{=} \lim_{n\to\infty}\left\{\frac{1}{nK}\sum_{t=1}^{nK}F_{k_t}(\widehat{\mathbf{x}}^t)\right\}-f^*\overset{\text{(c)}}{\leq}0,\end{aligned}\tag{26}$$

where (a) comes from the optimality of $\mathbf{x}^*$, (b) from Lemma 2, and (c) from the upper bound (23).

ACKNOWLEDGMENT

The existence of this letter owes much to intense discussions with M. Nilchian. Also, P. Pad and J. P. Ward were extremely helpful during inspiring exchanges related to convex analysis. The authors are also grateful for A. Amini's help and encouragement.

REFERENCES

[1] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.

[2] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.

[3] J. Bect, L. Blanc-Feraud, G. Aubert, and A. Chambolle, "A ℓ_1-unified variational framework for image restoration," in *Proc. ECCV*, 2004, vol. 3024, pp. 1–13.

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

[5] J. M. Bioucas-Dias and M. A. T. Figueiredo, "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2992–3004, Dec. 2007.

[6] A. Beck and M. Teboulle, "ch. gradient-based algorithms with applications to signal recovery problems," in *Convex Optimization in Signal Processing and Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2009, pp. 42–88.

[7] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," in *Springer Lecture Notes in Statistics*. Berlin, Germany: Springer-Verlag, 1995, pp. 125–150.

[8] U. S. Kamilov, E. Bostan, and M. Unser, "Wavelet shrinkage with consistent cycle spinning generalizes total variation denoising," *IEEE Signal Process. Lett.*, vol. 19, no. 4, pp. 187–190, Apr. 2012.

[9] A. Kazerouni, U. S. Kamilov, E. Bostan, and M. Unser, "Bayesian denoising: From MAP to MMSE using consistent cycle spinning," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 249–252, Mar. 2013.

[10] A. K. Fletcher, K. Ramchandran, and V. K. Goyal, "Wavelet denoising by recursive cycle spinning," in *Proc. Int. Conf. Image Processing*, 2002.

[11] C. Vonesch and M. Unser, "A fast thresholded Landweber algorithm for wavelet-regularized multidimensional deconvolution," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 539–549, Apr. 2008.

[12] C. Vonesch and M. Unser, "A fast multilevel algorithm for wavelet-regularized image restoration," *IEEE Trans. Image Process.*, vol. 18, no. 3, pp. 509–523, Mar. 2009.

[13] M. Guerquin-Kern, M. Häberlin, K. P. Prüssmann, and M. Unser, "A fast wavelet-based reconstruction method for magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 30, no. 9, pp. 1649–1660, Sep. 2011.

[14] S. Ramani and J. A. Fessler, "A hybrid regularizer combining orthonormal wavelets and finite differences for statistical reconstruction in 3-D CT," in *Proc. 2nd Int. Mtg. on Image Formation in X-ray CT*, Salt Lake City, UT, USA, 2012, pp. 348–351.

[15] U. S. Kamilov, P. Pad, A. Amini, and M. Unser, "MMSE estimation of sparse Lévy processes," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 137–147, Jan. 2013.

[16] I. W. Selesnick and M. A. T. Figueiredo, "Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors," in *Proc. SPIE 7446, 74460D*, 2009.

[17] H. H. Bauschke and P. L. Combettes*, Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Berlin, Germany: Springer, 2010.

[18] S. Boyd and L. Vandenberghe, "Subgradients," April 2008 class notes for Convex Optimization II [Online]. Available: http://see.stanford.edu/materials/lsocoee364b/01-subgradients_notes.pdf

[19] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *Proc. 20th Int. Conf. Machine Learning (ICML)*, Washington, DC, USA, Aug. 21–24, 2003.