The Sparsity of Cycle Spinning for Wavelet-Based Solutions of Linear Inverse Problems

Rahul Parhi, *Member, IEEE*, and Michael Unser, *Fellow, IEEE*

***Abstract*—The usual explanation of the efficacy of wavelet-based methods hinges on the sparsity of many real-world objects in the wavelet domain. Yet, standard wavelet-shrinkage techniques for sparse reconstruction are not competitive in practice, one reason being that the lack of shift-invariance of the wavelet transform produces blocky artifacts. The standard remedy is cycle spinning, which results in a substantial reduction of these artifacts. In this letter, we propose a theoretical investigation of the sparsity of solutions to the cycle-spinning variant of wavelet-based resolutions of linear inverse problems. We derive a representer theorem that provides a complete characterization of the solution set. Our theorem indicates that the solutions are typically not sparse, where sparsity is measured with respect to the wavelet dictionary. This exposes that the role of sparsity in the success of wavelet-based solutions of linear inverse problems requires further investigation. We corroborate our theoretical results with numerical examples for the problem of image denoising.**

***Index Terms*—Cycle spinning, regularization, representer theorem, sparsity, wavelets.**

I. Introduction

THE resolution of a linear inverse problem involves the estimation of an unknown signal $\boldsymbol{x} \in \mathbb{R}^N$ from the data $\boldsymbol{y} \in \mathbb{R}^M$ coming from the measurement process

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

where $\mathbf{H} \in \mathbb{R}^{M \times N}$ is the system matrix that models the response of the acquisition device and $\boldsymbol{\varepsilon} \in \mathbb{R}^M$ denotes a perturbation, or noise term, which is usually assumed to be a vector of i.i.d. Gaussian random variables. One typically solves the inverse problem by considering the regularized least-squares problem

$$\min_{\boldsymbol{x} \in \mathbb{R}^N} \left(\frac{1}{2} \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|_2^2 + \lambda \Phi(\boldsymbol{x}) \right), \tag{2}$$

where $\lambda > 0$ controls the strength of the regularization and $\Phi : \mathbb{R}^N \to \mathbb{R}_{\geq 0}$ denotes the regularizer, which is chosen to promote certain desirable properties in the estimated signal.

Manuscript received 16 February 2023; revised 27 April 2023; accepted 10 May 2023. Date of publication 12 May 2023; date of current version 19 May 2023. This work was supported by European Research Council (ERC Project FunLearn) under Grant 101020573. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yipeng Liu. *(Corresponding author: Rahul Parhi.)*

The authors are with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland (e-mail: rahul.parhi@epfl.ch; michael.unser@epfl.ch).

Digital Object Identifier 10.1109/LSP.2023.3275916

Wavelet methods correspond to $\Phi(\boldsymbol{x}) = \|\mathbf{W}\boldsymbol{x}\|_1$ in (2), where $\mathbf{W} \in \mathbb{R}^{N \times N}$ is an orthogonal discrete wavelet transform (DWT) such that $\mathbf{W}^\mathsf{T}\mathbf{W} = \mathbf{W}\mathbf{W}^\mathsf{T} = \mathbf{I}_N$, where $\mathbf{I}_N$ is the $(N \times N)$ identity matrix. This setting is often referred to as *iterative wavelet shrinkage* [14]. The deployment of wavelets as the tool of choice for the resolution of linear inverse problems is motivated by the observation that the wavelet transform of many real-world objects (*e.g.*, natural images) tends to be very sparse. Indeed, many such objects are spatially inhomogeneous or exhibit singularities (*e.g.*, the edges in an image). As it turns out, wavelet-shrinkage estimators are able to automatically adapt to the spatial inhomogeneities or singularities of the unknown signal, while classical linear signal-reconstruction methods cannot [15]. This, combined with the theory of compressed sensing [1], [2], [7], [8], [10], [13], [19], provides a rigorous mathematical justification for the success of wavelet-based solutions of linear inverse problems.

The wavelet-shrinkage problem is typically solved by the iterative soft-thresholding algorithm (ISTA) [4], [12], [18] and its fast variants [3]. The simplest form of ISTA considers the iterates

$$\begin{aligned} \boldsymbol{z}_t &= \boldsymbol{x}_{t-1} - \gamma_t \mathbf{H}^\mathsf{T}(\mathbf{H}\boldsymbol{x}_{t-1} - \boldsymbol{y}) \\ \boldsymbol{x}_t &= \mathbf{W}^\mathsf{T}\mathcal{T}(\mathbf{W}\boldsymbol{z}_t; \gamma_t\lambda), \end{aligned} \tag{3}$$

where $\gamma_t > 0$ is a stepsize that can be chosen *a priori* to ensure convergence of the algorithm [3] and $\mathcal{T}(\,\cdot\,;\,\cdot\,)$ denotes the pointwise shrinkage/soft-thresholding operation

$$\mathcal{T}(\theta; \tau) = \operatorname{sgn}(\theta)(|\theta| - \tau)_+,$$

where $(\,\cdot\,)_+ = \max\{0, \cdot\}$. Wavelet shrinkage has been extensively studied theoretically. It is known to be a near-optimal signal-denoising estimator (in the minimax sense) for a wide variety of signal classes [15]. And yet, it is actually not competitive in practice due to artifacts associated with the lack of shift-invariance of the DWT [11], [22]. To alleviate this problem, *cycle spinning* was proposed for denoising problems in [11] and applied to more general linear-inverse problems in [18], [21]. In accordance with the setup of [20], [21], the main idea is to augment the original wavelet dictionary by also including shifted versions of the wavelet basis functions. One can then increase the shift-invariance of the transform by considering the K integer shifts of the DWT, which we denote $\mathbf{W}_k := \mathbf{W}\mathbf{S}_k$ for $k = 0, \ldots, (K-1)$, where $\mathbf{S}_k$ denotes the operator that applies a circular shift to a signal k times. Cycle spinning is implemented

1070-9908 © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on May 22,2023 at 14:56:20 UTC from IEEE Xplore. Restrictions apply.

by considering the modified ISTA iterates

$$\begin{aligned} \boldsymbol{z}_t &:= \boldsymbol{x}_{t-1} - \gamma_t \mathbf{H}^\mathsf{T}(\mathbf{H}\boldsymbol{x}_{t-1} - \boldsymbol{y}) \\ \boldsymbol{x}_t &:= \mathbf{W}_{k_t}^\mathsf{T} \mathcal{T}(\mathbf{W}_{k_t}\boldsymbol{z}_t; \gamma_t \lambda), \end{aligned} \tag{4}$$

where $k_t = t \bmod K$. We see that the algorithm iteratively *cycles* through each of the K shifts of the original DWT. The choice $K = N$ would correspond to full shift-invariance, while $K = 1$ would reduce the problem to classical wavelet shrinkage. Cycle spinning has nearly the same computational cost as the iteration in (3). Yet, it results in higher quality solutions with less blocky artifacts [21]. While cycle spinning is often included in the implementation of wavelet-based regularization methods (see, *e.g.*, the `restoration` module in `scikit-image` [29] or the `wdenoise2` command in MATLAB which has been built-in since version 9.6 [24]), there is little theory accompanying the procedure.

In this letter, we propose a theoretical investigation of the sparsity of solutions to the cycle-spinning procedure in (4) and compare it to the sparsity of solutions to the classical formulation in (3). To the best of our knowledge, this is the first theoretical investigation of the sparsity of cycle spinning. Our main contribution is to derive a novel *representer theorem* that completely characterizes the solution set of the cycle-spinning procedure. As a byproduct, we establish a quantitative bound on the sparsity of the solutions.

Interestingly, we find that the solutions offered by cycle spinning are typically not sparse at all. Thus, we see that there is a striking contrast between these solutions and those offered by classical wavelet shrinkage, which is known to admit very sparse solutions instead. Our result also quantifies the tradeoff between sparsity and shift-invariance found in the number K of shifts. We conclude that sparsity does not tell the entire story for the success of wavelet-based methods for linear inverse problems. We corroborate these theoretical findings with numerical examples for the problem of image denoising.

II. Main Results

Convergence results from [21, Theorem 1] have shown that the iterates $\boldsymbol{x}_t$ from the cycle-spinning procedure in (4) converge to a minimizer of the variational problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} \left(\frac{1}{2}\|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|_2^2 + \frac{\lambda}{K} \sum_{k=0}^{K-1} \|\mathbf{W}_k \boldsymbol{x}\|_1 \right). \tag{5}$$

Thus, the cycle-spinning procedure solves a variational problem that regularizes not only the DWT of the signal, but also the DWT of shifts of the signal. We refer to the problem in (5) as the *cycle-spinning variational problem*. To investigate the sparsity of cycle spinning, we provide a complete characterization of the solution set of (5) in Theorem 1. As a byproduct, this theorem provides a bound on the sparsity of the solutions.

If we let

$$\mathbf{V} = \begin{bmatrix} \mathbf{W}_0 \\ \vdots \\ \mathbf{W}_{K-1} \end{bmatrix} \in \mathbb{R}^{NK \times N}, \tag{6}$$

then we can rewrite (5) as the equivalent problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} \left(\frac{1}{2}\|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|_2^2 + \frac{\lambda}{K}\|\mathbf{V}\boldsymbol{x}\|_1 \right). \tag{7}$$

The transform $\mathbf{V}$ is a *redundant discrete wavelet transform*. In particular, the rows of $\mathbf{V}$ do not form a basis. They form instead a tight frame with frame bound $1/K$ [20, Eq. (4)]. We shall see in Theorem 1 that this redundancy plays an important role in the (lack of) sparsity of the solutions to (7).

The tight-frame property translates into

$$\begin{aligned} \mathbf{V}^\mathsf{T}\mathbf{V} &= \begin{bmatrix} \mathbf{W}_0^\mathsf{T} & \cdots & \mathbf{W}_{K-1}^\mathsf{T} \end{bmatrix} \begin{bmatrix} \mathbf{W}_0 \\ \vdots \\ \mathbf{W}_{K-1} \end{bmatrix} \\ &= \mathbf{W}_0^\mathsf{T}\mathbf{W}_0 + \cdots + \mathbf{W}_{K-1}^\mathsf{T}\mathbf{W}_{K-1} \\ &= K\,\mathbf{I}_N. \end{aligned}$$

Consequently, if $\boldsymbol{\theta} = \mathbf{V}\boldsymbol{x}$, then $\boldsymbol{x} = \frac{1}{K}\mathbf{V}^\mathsf{T}\boldsymbol{\theta}$. This motivates the *synthesis formulation* of the problem given by

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^{NK}} \left(\frac{1}{2}\left\|\boldsymbol{y} - \frac{1}{K}\mathbf{H}\mathbf{V}^\mathsf{T}\boldsymbol{\theta}\right\|_2^2 + \frac{\lambda}{K}\|\boldsymbol{\theta}\|_1 \right). \tag{8}$$

It is well-known that there exist sparse solutions to (8) such that $\boldsymbol{\theta}$ is M-sparse [28], where we recall that M is the number of measurements (*i.e.*, $\mathbf{H} \in \mathbb{R}^{M\times N}$). Unfortunately, the synthesis problem in (8) is not equivalent to the analysis problem in (7) due to a lack of surjectivity from the signal space $\mathbb{R}^N$ to $\mathbb{R}^{NK}$. To alleviate this issue, we must consider the constrained optimization problem

$$\begin{aligned} &\min_{\boldsymbol{\theta}\in\mathbb{R}^{NK}} \left(\frac{1}{2}\left\|\boldsymbol{y} - \frac{1}{K}\mathbf{H}\mathbf{V}^\mathsf{T}\boldsymbol{\theta}\right\|_2^2 + \frac{\lambda}{K}\|\boldsymbol{\theta}\|_1 \right) \\ &\quad \text{s.t.} \quad \frac{1}{K}\mathbf{V}\mathbf{V}^\mathsf{T}\boldsymbol{\theta} = \boldsymbol{\theta}, \end{aligned} \tag{9}$$

where the constraint ensures that the search space only includes coefficients that lie in the range of $\mathbf{V}$ (since $\frac{1}{K}\mathbf{V}\mathbf{V}^\mathsf{T}$ is the orthoprojector from $\mathbb{R}^{NK}$ onto the range of $\mathbf{V}$). This establishes an (isometric) isomorphism between the signal space $\mathcal{X} = \mathbb{R}^N$ equipped with the norm $\|\boldsymbol{x}\|_\mathcal{X} := \|\mathbf{V}\boldsymbol{x}\|_1$, and the coefficient space

$$\Theta = \left\{ \boldsymbol{\theta} \in \mathbb{R}^{NK} : \frac{1}{K}\mathbf{V}\mathbf{V}^\mathsf{T}\boldsymbol{\theta} = \boldsymbol{\theta} \right\},$$

equipped with the norm $\|\boldsymbol{\theta}\|_\Theta := \|\boldsymbol{\theta}\|_1$. In particular, these norms make $\mathcal{X}$ and Θ *bona fide* Banach spaces.[1]

If we compare (8) and (9), we see that there is a stark difference between the analysis and synthesis formulations when working with redundant signal representations [17]. In fact, it has been observed that the synthesis formulation may be too sensitive and that it is often outperformed by its analysis counterpart [9], [26], which explains why cycle spinning is preferred in practice.

[1] One can take advantage of the equivalence of norms in finite dimensions to check that $\|\cdot\|_\mathcal{X}$ has a trivial null space, since the rows of $\mathbf{V}$ form a tight frame.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on May 22,2023 at 14:56:20 UTC from IEEE Xplore. Restrictions apply.

The isometric isomorphism between $\mathcal{X}$ and Θ establishes that the cycle-spinning variational problem in (5) is equivalent to the *constrained* synthesis problem in (9). The isometric isomorphism from $\mathcal{X}$ to Θ is given by $\boldsymbol{x} \mapsto \mathbf{V}\boldsymbol{x}$, while the inverse mapping from Θ to $\mathcal{X}$ is $\boldsymbol{\theta} \mapsto \frac{1}{K}\mathbf{V}^{\mathsf{T}}\boldsymbol{\theta}$. Using this isomorphism, we can study the sparsity of the solutions to the problem in (9) to understand the sparsity of the solutions to (5). Our findings are summarized in a representer theorem.

Theorem 1: The solution set to the cycle-spinning variational problem

$$\mathcal{V} = \underset{\boldsymbol{x}\in\mathbb{R}^N}{\operatorname{argmin}} \left(\frac{1}{2}\|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|_2^2 + \frac{\lambda}{K}\sum_{k=0}^{K-1} \|\mathbf{W}_k\boldsymbol{x}\|_1 \right) \tag{10}$$

is nonempty, convex, and compact. Moreover, $\mathcal{V}$ is the closure of the convex hull of its extreme points, which can all be expressed as

$$\boldsymbol{x}_{\text{extreme}} = \sum_{s=1}^{S_\lambda} c_s \boldsymbol{v}_s,$$

where $S_\lambda \leq \min\{NK, MNK - MN + M\}$ depends on the regularization parameter λ, $c_s \in \mathbb{R}\setminus\{0\}$, and $\boldsymbol{v}_s$, is a transposed row of $\mathbf{V}$ defined in (6) so that $\boldsymbol{v}_s$ is a (shifted) wavelet. The regularization cost, which is common to all solutions is

$$\sum_{k=0}^{K-1} \|\mathbf{W}_k\boldsymbol{x}_{\text{extreme}}\|_1 = \sum_{s=1}^{S_\lambda} |c_s| = \|\boldsymbol{c}\|_1.$$

Proof: To prove the theorem, we need to characterize the solution set to the problem in (9). The result of the theorem will then follow from the equivalence between (9) and (10) given by the isometric isomorphism $\boldsymbol{\theta} \mapsto \frac{1}{K}\mathbf{V}^{\mathsf{T}}\boldsymbol{\theta}$. From recent abstract representer theorems studying the solution sets to convex optimization problems over Banach spaces with sparsity-promoting norms [27, Theorems 2 and 3] (see also [5], [6]), we know that the solution set to (9) is nonempty, convex, and compact. Moreover, its extreme points take the form of a linear combination (with nonzero coefficients) of $\leq M$ extreme points of the unit ball of the regularizer and the regularization cost is shared by all solutions. To characterize the extreme points of the unit ball of the regularizer, we must characterize the extreme points of

$$B_\Theta = \{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\theta}\|_1 \leq 1\}.$$

Let

$$B_{\ell^1} = \{\boldsymbol{\theta} \in \mathbb{R}^{NK} : \|\boldsymbol{\theta}\|_1 \leq 1\}$$

denote the unit ℓ^1-ball in $\mathbb{R}^{NK}$ and recall that

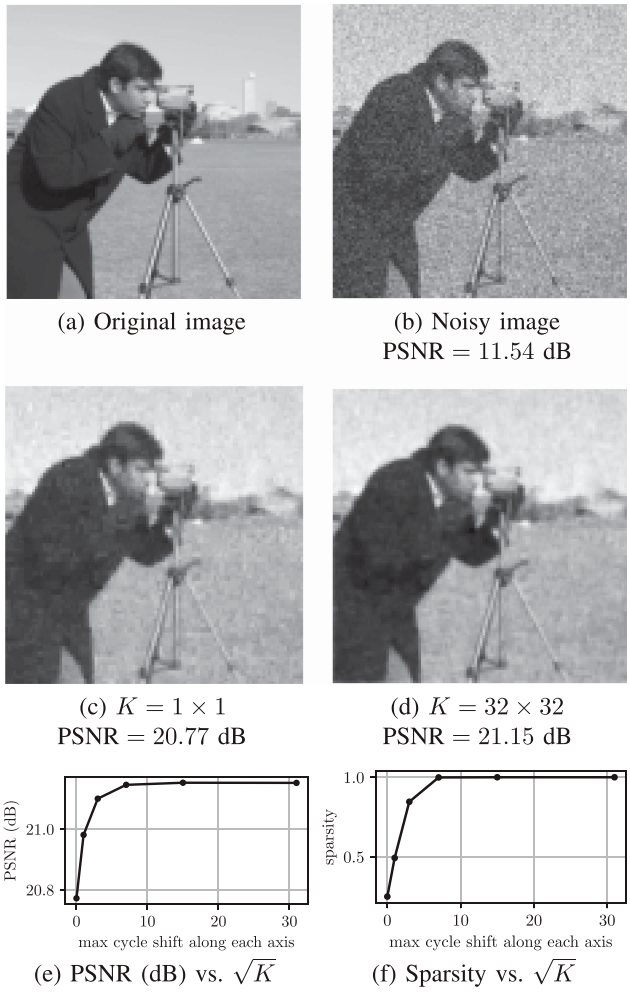
$$\Theta = \left\{ \boldsymbol{\theta} \in \mathbb{R}^{NK} : \frac{1}{K}\mathbf{V}\mathbf{V}^{\mathsf{T}}\boldsymbol{\theta} = \boldsymbol{\theta} \right\},$$

which is the range of $\mathbf{V}$. Clearly, Θ is a subspace of $\mathbb{R}^{NK}$ of dimension N since it is isomorphic to the signal space $\mathbb{R}^N$. Then, we have that

$$B_\Theta = B_{\ell^1} \cap \Theta.$$

We see that B_Θ is the intersection of the polytope B_{ℓ^1} with the subspace Θ of co-dimension $(NK - N)$. It then follows from the classical result of Dubins [16] that each extreme point of B_Θ is the convex combination of at most $(NK - N + 1)$ extreme points of B_{ℓ^1}. Since the extreme points of B_{ℓ^1} are the canonical unit vectors, we have that the extreme points of B_Θ are $(NK - N + 1)$-sparse vectors.

Finally, we conclude by noticing that a linear combination of $\leq M$ vectors that are $(NK - N + 1)$-sparse will be, at worst, $M(NK - N + 1)$-sparse if $M(NK - N + 1) \leq NK$, and NK-sparse otherwise. □



(a) Original image
(b) Noisy image PSNR = 11.54 dB
(c) $K = 1 \times 1$ PSNR = 20.77 dB
(d) $K = 32 \times 32$ PSNR = 21.15 dB
(e) PSNR (dB) vs. $\sqrt{K}$
(f) Sparsity vs. $\sqrt{K}$

Fig. 1. Comparison of denoising results for the (512×512) cameraman: (a) original image; (b) original image plus Gaussian noise with $\sigma^2 = 0.33^2$; (c) denoised image via classical wavelet shrinkage ($K = (1 \times 1)$); (d) denoised image via cycle spinning with $K = (32 \times 32)$; (e) PSNR vs. $\sqrt{K}$; (f) sparsity (proportion of nonzero coefficients) vs. $\sqrt{K}$.

A. Discussion

The key takeaway from Theorem 1 is that the solutions to the cycle-spinning variational problem are typically not sparse at all, where sparsity is measured with respect to the wavelet dictionary defined by the rows of $\mathbf{V}$ defined in (6). The extreme points of the solution set correspond to the sparsest solutions. Therefore, the sparsest solutions satisfy the sparsity bound $S_\lambda \leq$

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on May 22,2023 at 14:56:20 UTC from IEEE Xplore. Restrictions apply.

$\min\{NK, MNK - MN + M\}$. From this bound, we see that, as we increase the shift-invariance (increase K), the sparsity bound decreases. We also remark that increasing λ will result in sparser solutions. Indeed, as $\lambda \to \infty$ (10) will reduce to the minimization of the regularization term, which is minimized when $\boldsymbol{x} = \mathbf{0}$. Thus, as λ increases, S_λ decreases. This phenomenon has been investigated in the lasso literature [26]. When $K = 1$, the cycle spinning variational problem is the classical wavelet shrinkage problem, and we recover the well-known sparsity bound of $S_\lambda \leq M$. Thus, we see that Theorem 1 provides a quantification of the tradeoff between shift-invariance and sparsity. We also remark that the lack of sparsity of cycle spinning has been observed before in the literature [9], although only from an empirical perspective. Theorem 1 provides a theoretical justification for this observation.

III. Numerical Examples

In this section we verify numerically the sparsity results of Theorem 1. We focus on the problem of pure denoising which corresponds to $\mathbf{H} = \mathbf{I}_N$ in (1) and, so that $M = N$. We consider the denoising of images in $\mathbb{R}^{\sqrt{N}\times\sqrt{N}}$ using two-dimensional Haar wavelets with shifting $\sqrt{K} \times \sqrt{K}$ times (*i.e.*, $\sqrt{K}$ times in each axis, for a total of K shifts). Our experiments are implemented in Python using the PyWavelets package [23]. We choose λ in (4) by estimating the noise level of the image by the procedure from [14]. As in [20], we only threshold the detail coefficients. Our results are shown in Fig. 1.

From this figure, we see that cycle spinning increases the quality of the solutions at the price of sparsity. In particular, Fig. 1(d) is smoother and has less artifacts than Fig. 1(c), particularly in the cameraman's jacket. One also sees in Fig. 1(e) that the peak signal-to-noise (PSNR) increases as the number K of shifts increases. Moreover, we see that the PSNR is saturated long before the procedure is fully shift-invariant ($K = (512 \times 512)$), which shows that only a few shifts are required for high-quality solutions. The sparsity vs. $\sqrt{K}$ is plotted in Fig. 1(f), where sparsity is quantified by the proportion of *nonzero* coefficients from the wavelet dictionary $\mathbf{V}$. This confirms the property from Theorem 1 according to which, as the number of shifts K increases, the solutions become less sparse, eventually becoming completely dense.

IV. Conclusion

In this letter, we have investigated theoretically the sparsity of cycle spinning for wavelet-based solutions of linear inverse problems. We have derived a representer theorem that provides a complete characterization of the solutions to the cycle-spinning procedure. This theorem asserts that the solutions are typically not sparse with respect to the wavelet dictionary. We have corroborated our theoretical findings with numerical examples for the problem of image denoising. We have found that the denoised images with the highest peak signal-to-noise ratio are solutions with the least sparsity. This shows that sparsity is not solely responsible for explaining the success of wavelets in the resolution of linear inverse problems. Additional theoretical investigations of cycle spinning are required for a full understanding of the state of affairs.

V. Future Work

When $K = N$, the matrix $\mathbf{V}$ in (6) is the undecimated wavelet transform (UWT). The filter-bank implementation of the UWT only requires $N \log N$ filters, while $\mathbf{V}$ uses N^2 [18], [22]. This is because only $N \log N$ coefficients are unique. The authors of [9] take advantage of this fact to reformulate (9), in the case $K = N$, with a coefficient space of smaller dimension. Therefore, as K approaches N, we have reason to believe that the sparsity bound in Theorem 1 may be loose due to the nonuniqueness of coefficients. One direction of future work will be devoted to sharpening the sparsity bound as K approaches N.

For one-dimensional signals, a single iteration of the total variation (TV) denoising scheme from [25] is equivalent to cycle spinning with Haar wavelets. Meanwhile, for the problem of image reconstruction, cycle spinning with Haar wavelets and TV regularization results in very similar-looking reconstructed images [21]. We conjecture that solutions to cycle spinning with full shift-invariance ($K = N$) will be sparse in the sense of TV. Thus, another direction of future work will be devoted to the theoretical study of the relationship between cycle spinning and total-variation regularization.

References

[1] B. Adcock, S. Brugiapaglia, and M. King-Roskamp, "Do log factors matter? On optimal wavelet approximation and the foundations of compressed sensing," *Found. Comput. Math.*, vol. 22, no. 1, pp. 99–159, 2022.

[2] B. Adcock and A. C. Hansen, *Compressive Imaging: Structure, Sampling, Learning*. Cambridge, U.K.: Cambridge Univ. Press, 2021.

[3] 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.

[4] J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle, "A l^1-unified variational framework for image restoration," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2004, pp. 1–13.

[5] C. Boyer, A. Chambolle, Y. D. Castro, V. Duval, F. De Gournay, and P. Weiss, "On representer theorems and convex regularization," *SIAM J. Optim.*, vol. 29, no. 2, pp. 1260–1281, 2019.

[6] K. Bredies and M. Carioni, "Sparsity of solutions for variational inverse problems with finite-dimensional data," *Calculus Variations Partial Differ. Equ.*, vol. 59, no. 1, 2020, Art. no. 14.

[7] E. J. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, 2007, Art. no. 969.

[8] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[9] D. Carrera, G. Boracchi, A. Foi, and B. Wohlberg, "Sparse overcomplete denoising: Aggregation versus global optimization," *IEEE Signal Process. Lett.*, vol. 24, no. 10, pp. 1468–1472, Oct. 2017.

[10] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 211–231, 2009.

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

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

[13] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[14] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[15] D. L. Donoho and I. M. Johnstone, "Minimax estimation via wavelet shrinkage," *Ann. Statist.*, vol. 26, no. 3, pp. 879–921, 1998.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on May 22,2023 at 14:56:20 UTC from IEEE Xplore. Restrictions apply.

[16] L. E. Dubins, "On extreme points of convex sets," *J. Math. Anal. Appl.*, vol. 5, no. 2, pp. 237–244, 1962.
[17] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, 2007, Art. no. 947.
[18] 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.
[19] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing* (Applied and Numerical Harmonic Analysis Series). Berlin, Germany: Springer New York, 2013.
[20] U. 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.
[21] U. S. Kamilov, E. Bostan, and M. Unser, "Variational justification of cycle spinning for wavelet-based solutions of inverse problems," *IEEE Signal Process. Lett.*, vol. 21, no. 11, pp. 1326–1330, Nov. 2014.
[22] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, "Noise reduction using an undecimated discrete wavelet transform," *IEEE Signal Process. Lett.*, vol. 3, no. 1, pp. 10–12, Jan. 1996.
[23] G. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O'Leary, "PyWavelets: A. Python package for wavelet analysis," *J. Open Source Softw.*, vol. 4, no. 36, 2019, Art. no. 1237.
[24] MATLAB, version 9.6 (R2019a). Natick, Massachusetts: The MathWorks Inc., 2019.
[25] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk, "On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs," *SIAM J. Numer. Anal.*, vol. 42, no. 2, pp. 686–713, 2004.
[26] R. J. Tibshirani and J. Taylor, "The solution path of the generalized LASSO," *Ann. Statist.*, vol. 39, no. 3, pp. 1335–1371, 2011.
[27] M. Unser, "A unifying representer theorem for inverse problems and machine learning," *Found. Comput. Math.*, vol. 21, no. 4, pp. 941–960, 2021.
[28] M. Unser, J. Fageot, and H. Gupta, "Representer theorems for sparsity-promoting ℓ_1-regularization," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 5167–5180, Sep. 2016.
[29] S. Van der Walt et al., "Scikit-image: Image processing in python," *PeerJ*, vol. 2, 2014, Art. no. e453.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on May 22,2023 at 14:56:20 UTC from IEEE Xplore. Restrictions apply.