Convergence of the Red-TOWER Method for Removing Noise From Data

Umberto Amato and Qi-nian Jin

Abstract—By coupling the wavelet transform with a particular nonlinear shrinking function, the Red-telescopic optimal wavelet estimation of the risk (TOWER) method is introduced for removing noise from signals. It is shown that the method yields convergence of the $L_2$ risk to the actual solution with optimal rate. Moreover, the method is proved to be asymptotically efficient when the regularization parameter is selected by the generalized cross validation criterion (GCV), the Mallows criterion, or the Nearest-neighbor criterion. Numerical experiments on synthetic data are provided to compare the performance of the Red-TOWER method with the original approach, the shrink-thresholding, and the nonparametric shrinking function. Furthermore, the numerical tests are also performed when the TOWER method is applied to the hard-thresholding, soft-thresholding, and neighbor thresholding. Furthermore, the full numerical results are still open.

Index Terms—Convergence rate, denoising, generalized cross validation, Mallows criterion, nonparametric regression, regularization, thresholding, TOWER, wavelets.

I. INTRODUCTION

Removing noise from signals is a well-known problem in signal processing. Due to the wide range of practical applications, a number of methods has been developed to deal with this problem under different frameworks (engineering, statistics, approximation theory, inverse problems). Currently, several well-optimized methods are available to give both efficient algorithms and an optimal convergence rate. Nevertheless, the research is still active in this field since a better cleaning of the signals is highly desirable in those applications where the current methods are ineffective because of the intrinsic trouble in analyzing the signals, e.g., low signal-to-noise ratio (SNR) and artifacts in signals introduced during measurement.

Although the technique of wavelet transform and the use of shrinking functions are well-known in the classical literature independently, they were never used together for removing noise problems until the beginning of 1990s. Recently, a class of methods, which is developed by coupling both techniques, attracts the attention of many researchers; it gives rise to a recipe that has proved successful in a wide range of applications. The pioneering work of such kinds of methods are described in a series of celebrated papers [20]–[23] by Donoho and Johnstone, in which the hard and soft shrinking functions were introduced. The key point for the success of these methods lies in the nice features of the wavelet transform and the shrinking functions (both are simple and easy to be realized in ICs, and the wavelet transform is time-frequency adaptive and more efficient in computation than the fast Fourier transform). Since then, several other shrinking functions have been introduced, each one with its own pros and cons. Among others, we mention a continuous version of hard-thresholding [18], firm [9], and garrote [25]. A linear shrinking function, based on the regularization approach, has been developed in [2] and [4]. Moreover, several variations have been suggested for some of the above-mentioned methods to deal with the more general cases on the signals and noise, e.g., the correlated noise [3], [28], the shift-invariant transform [10], [15], and the nonperiodic functions [14]. We also mention that other nonlinear wavelet methods were developed that did not rely on a shrinking function (e.g., [6], [8], [29]). The most recent shrinking functions are concerned with local block thresholding rules that threshold the empirical wavelet coefficients in groups rather than individually [11]–[13], [24], [26]. They show the best performance among the known wavelet estimators.

Recently, a new method, which is called TOWER, was introduced in [1] from an estimate of the optimal $L_2$ risk. This method enables us to produce a new approximate solution by “correcting” the “predicted” solution (called “predictor” in the sequel) obtained by any existing method. Comparing with the predictor, the preliminary numerical results in [1] confirm an improvement of the denoising properties. However, theoretical convergence results were missing.

In this paper, we will consider the Red-TOWER method, which is a kind of TOWER method where the predictor is obtained by the regularization technique [2], [4]. Our main aim is to show that the Red-TOWER method has an optimal convergence property in the framework of nonparametric regression. This gives the full theoretical justification to the Red-TOWER method, which was an open problem in [1]. Moreover, we intend to apply the TOWER method to other predictors, evaluating its performance from the numerical point of view. The organization of this paper is as follows. In Section II, we introduce some background and notations used in the paper. In Section III, we give the full convergence results for the Red-TOWER method. We perform the numerical tests in Section IV and make some comparisons with other known methods in the literature; moreover, we also perform the numerical experiments when the TOWER method is applied to other nonlinear shrinking functions for which the theoretical analysis is more difficult and the proof of convergence is still open. Some final conclusions are drawn in Section V.
II. BACKGROUND AND NOTATIONS

Let us consider the noisy model
\[ y_i = f(x_i) + \varepsilon_i, \quad 1 \leq i \leq N \]  
(1)
where
- \( x_i \) equispaced nodes in \([0, 1]\);
- \( f \) in some Sobolev space \( H^p \) with regularity \( p \);
- \( \varepsilon_i \) white noise with common variance \( \sigma^2 \).

We suppose \( N = 2^J \). In the following, we denote by \( \{F_{j,\ell}\}_{j=0, \ldots, J, \ell=0, \ldots, 2^j} \) the set of the discrete wavelet transform (DWT) coefficients of the sequence \( f / \sqrt{N} \equiv \{f(x_i) / \sqrt{N}\}_{i=1}^{N} \) with respect to a particular wavelet system. It is an approximation of the (continuous) wavelet transform of \( f \). Such an approximation was justified in several papers [19], [33] and is currently accepted in all applications; it would admit better approximation property in case the coiflets are considered as a basis [7]. Analogously, we denote by \( \{Y_{j,\ell}\}_{j=0, \ldots, J, \ell=0, \ldots, 2^j} \) the set of the DWT coefficients of the signal \( y / \sqrt{N} \equiv \{y_i / \sqrt{N}\}_{i=1}^{N} \) with respect to the same wavelet system. For ease of notation, we will simply denote the DWT coefficients by \( \{E_k\}_{k=0}^{N-1} \) and \( \{Y_k\}_{k=0}^{N-1} \), with \( k \) relating to the scales and translation parameters by \( k = 0 \) for \( j = 0, \ell = 0 \), and \( k = 2^j + \ell \) for \( 0 \leq j < J, 0 \leq \ell < 2^j \). Obviously, there is a bijection between the sets \( Z := \{0, 0 \leq k < N\} \) and \( \{j, \ell\}, 0 \leq j < J, 0 \leq \ell < 2^j \) with \( j, \ell \) and \( k \) related as above, and \( k \equiv 2j(2^j)0 \equiv 2^j \ell \). Moreover, we denote by \( W \) the wavelet transform matrix such that \( Y = Wf / \sqrt{N} \).

In [2] and [4], a wavelet-based regularization method was introduced that shrinks the wavelet coefficients according to the rule
\[ R_k := R_k(Y_k) = sR(\lambda)Y_k \]  
(2)
where \( \lambda \) is the regularization parameter, and \( sR(\lambda) \) is the regularization shrinking function \( sR(\lambda) = (1 + a \lambda^2)^{-1} \) (here, superscript \( R \) stands for regularization). Since \( sR \) does not depend on \( Y_k \), \( R_k \) is linear with respect to \( Y_k \). This regularization shrinking function is obtained by the penalization of the \( L_2 \) risk \( \rho := E[|\hat{\lambda} - \lambda|^2] \) through the Sobolev norm of the regularized solution and the introduction of a regularization parameter \( \lambda \) that controls the tradeoff between fidelity to data and smoothness of the solution, where \( R := \{R_k\}_{k=0}^{N-1} \). Wavelet regularization was rediscovered recently in [16] and [17], where more arguments were provided concerning the potential of the regularization method with respect to thresholding methods in terms of convergence rate, moderate size sample, and the performance on abrupt changes or discontinuities. Examples were also provided to enhance these arguments on an experimental basis.

Several convergence properties of the wavelet regularization method have been proved in [2] and [4], and the optimal convergence rate \( O(N^{-2p/(2p+1)}) \) of the risk has been demonstrated. Moreover, two different criteria were considered for the choice of the regularization parameter. The generalized cross validation (GCV) criterion [35] was considered in [2] to choose \( \lambda \) by
\[ \min_{\lambda \geq 0} V_N(\lambda) = \frac{\left\| (I - D_N(\lambda))Y \right\|^2}{\left\| \frac{1}{n} \text{Tr}(I - D_N(\lambda)) \right\|^2} \]  
(3)
where \( D_N(\lambda) \) is the diagonal matrix with non-null entries \( (1 + a \lambda^2)^{-1} \), \( 0 \leq k < N \). The regularization parameter was also estimated in [4] by the Mallows criterion endowed with an estimate of variance based on the finite differences of the sample \( \hat{y} \).

The theoretical convergence results are currently available for both criteria.

We also mention that a full extension to the case of correlated noise has been made in [3]. If \( S_N \) denotes the input error covariance matrix, then after transforming the input data to the wavelet domain, the resulting error covariance matrix of the wavelet coefficients \( \hat{S}_N \) will be given by \( \hat{S}_N = WSW^T \), where prime means transpose (i.e., inverse since \( W \) is orthogonal). In this case, the shrinking function takes the form
\[ sR(\lambda) = \frac{1}{1 + a \lambda^2 s_k^2 \lambda} \]  
(4)
and the GCV criterion is modified as
\[ \min_{\lambda \geq 0} \hat{V}_N(\lambda), \quad \hat{V}_N(\lambda) = \frac{|| (I - D_N(\lambda))Y ||^2}{\frac{1}{n} \text{Tr}(\text{diag}(\hat{S}_N)(I - D_N(\lambda)))} \]  
(5)
where \( s_k^2 \) are the diagonal elements of \( \hat{S}_N \) satisfying \( 0 < b_1 \leq s_k^2 \leq b_2 < +\infty \) for all \( k, N \) with two positive constants \( b_1, b_2 \), and diag \( (\hat{S}_N) \) stands for the diagonal matrix with diagonal entries of \( \hat{S}_N \). The bounds on \( s_k \) correspond to the reasonable hypothesis from the physical point of view of finite and non-null noise affecting data, and \( s_k = 1 \) in the case of white noise.

In [1], a shrinking function (TOWER) was introduced directly from an estimate of the optimal \( L_2 \) risk with respect to the diagonal shrinking functions (oracle). It is given by
\[ T_k := T_k(Y_k) = s^T(Y_k; \hat{Y}_k^{\text{f.g.}})Y_k \]  
(6)
where
\[ s^T(Y_k; \hat{Y}_k^{\text{f.g.}}) := \begin{cases} 0, & \text{if } |Y_k| \leq 2 \sigma s_k \sqrt{N} \text{ or } |Y_k^{\text{f.g.}}| \leq Q_k^T \\ Q_k^T, & \text{otherwise} \end{cases} \]  
(7)
(8)
where \( s_k^2 \) is the predictor obtained by some existing method; the superscript \( f.g. \) means that \( \hat{Y}_k^{\text{f.g.}} \) can be considered as a first guess of a suitable iterative process that converges to the solution (6) (see [1] for details).

In practice, the TOWER method goes through the following.

Algorithm 1

**Input:** The sample \( y_k, \ 0 \leq k < N \).

**Step 1** Compute the wavelet transform \( Y \) of the sequence \( \{y_k / \sqrt{N}\}_{0 \leq k < N} \).

**Step 2** Choose a first guess \( \hat{Y}_k^{\text{f.g.}} \) by solving problem (1) in the wavelet domain by any method.
Step 3) Apply TOWER to $\sum_{k} T_k$ by (5) and (6) to obtain $T_k$.
Step 4) Perform the inverse wavelet transform on the sequence $\{\sqrt{N}T_k\}_{0 \leq k < N}$ to yield the required solution of problem (1).

The above exposition indicates that the TOWER shrinking function has to be coupled with a predictor that is (hopefully) corrected after the TOWER step. Although any method could be considered to produce the predictor, in this paper, we mainly choose regularization for this purpose that gives rise to the RED-TOWER method. There are two reasons for our choice. On the one hand, the regularization shrinking function is the least signal adaptive one: the damping of the wavelet coefficients depends only on the global parameter $\lambda$ and on the factors $a_k^{2p}$; therefore, the regularization method is the best candidate to benefit from the intrinsic adaptivity in the TOWER method. One the other hand, the regularization method is linear so that the theoretical analysis of convergence is easier to handle.

III. CONVERGENCE OF THE RED-TOWER METHOD

Once a set of DWT coefficients is given, we can construct the corresponding finite-dimensional function in the original (physical) domain by an expansion in terms of wavelets and scaling function. For example, if we consider the coefficients $R_k$, $0 \leq k < N$, defined by (2), then the corresponding solution $r_N$ determined by wavelet regularization is given by

$$r_N(x) = R_{-1,0}\varphi_0(x) + \sum_{j=0}^{J-2} \sum_{\ell=0}^{2^j-1} R_{j,\ell} \psi_{j,\ell}(x)$$

where $\{\varphi_0, \psi_{j,\ell}, 0 \leq j < J, 0 \leq \ell < 2^j\}$ is the wavelet system we use. Analogously, we denote by $f_N$, the finite-dimensional actual solution (in the original domain) of (1) when no noise affects data, i.e.,

$$f_N(x) = F_{-1,0}\varphi_0(x) + \sum_{j=0}^{J-2} \sum_{\ell=0}^{2^j-1} F_{j,\ell} \psi_{j,\ell}(x).$$

Moreover, we use $t_N^\lambda$ to denote the finite-dimensional RED-TOWER solution (in the original domain), which is given by

$$t_N^\lambda(x) = T_{-1,0}\varphi_0(x) + \sum_{j=0}^{J-2} \sum_{\ell=0}^{2^j-1} T_{j,\ell} \psi_{j,\ell}(x)$$

where the coefficients $T_k$ are defined by (5) with the first guess given by the wavelet regularization method.

Let us first consider the $L_2$ risk $E_N(\lambda) := E[|r_N - f_N|^2]$ for the wavelet regularization method. We have the expression

$$E_N(\lambda) = G_N(\lambda) + \frac{\sigma^2}{N} H_N(\lambda)$$

with

$$G_N(\lambda) = \lambda^2 \sum_{k \in Z} \frac{a_k^{2p+1} R_k^2}{(1 + a_k^{2p+2} \lambda^{2p})^2}$$

and

$$H_N(\lambda) = \sum_{k \in Z} \frac{\sigma^2}{(1 + a_k^{2p+2} \lambda^{2p})^2}.$$ 

It is well known [3] that if $f \in H^p$ with $p > 1/4$, then

$$G_N(\lambda) \leq C_p \|f\|_{H^p}^2$$

and

$$H_N(\lambda) \leq \tilde{C}_p \|f\|_{H^p}^2$$

with positive constants $C_p$ and $\tilde{C}_p$ depending on $p$ only.

**Lemma 1:** Let $f \in H^p$ with $p > 1/2$, and let $\lambda_N$ be chosen by the GCV criterion. Then

$$\lambda_N \geq C_p \left( \frac{\sigma^2}{N \|f\|_{H^p}} \right)^{2p/(2p+1)}$$

and

$$G_N(\lambda_N) \leq \tilde{C}_p \left( \frac{\sigma^2}{N \|f\|_{H^p}} \right)^{2p/(2p+1)}$$

with positive constants $C_p$, $\tilde{C}_p$, and $\tilde{C}_p$ depending on $p$ only.

**Proof:** If $f = 0$, then $\lambda_N = \infty$, and the results are trivial. Let us assume $f \neq 0$ in the following. As shown in [3, Th. 8], we have $\lambda_N \to 0$ under stated conditions. Therefore, we can follow the estimate of [3, Eq. (12)] to obtain

$$\frac{d_1 \lambda^{1/2p}}{\lambda^{1/2p}} \leq H_N(\lambda_N) \leq \frac{d_2 \lambda^{1/2p}}{\lambda^{1/2p}}$$

with positive constants $d_1$ and $d_2$ depending on $p$ only. By invoking [3, Th. 9], we get

$$\frac{d_1 \sigma^2}{N \lambda^{1/2p}} \leq E_N(\lambda_N) \sim \min_{\lambda \geq 0} E_N(\lambda) \leq D_p \left( \frac{\sigma^2}{N} \right)^{2p/(2p+1)}$$

with some constant $D_p > 0$. Solving this equation, we prove (10).

Since $G_N(\lambda_N) \leq E_N(\lambda_N)$, we obtain (11) immediately.

Next, we consider the $L_2$ loss between $t_N^\lambda$ and $r_N^\lambda$. We then have the following theorem.

**Theorem 1:** Let $r_N^\lambda$ and $t_N^\lambda$ be defined by (7) and (8), respectively. If $f \in H^p$ with $p > 1/4$, then for all $\lambda > 0$, we have

$$E[|r_N^\lambda - t_N^\lambda|^2] \leq c_1 G_N(\lambda) + \frac{\sigma^2}{N^{1/2p}}$$

and

$$E[|r_N^\lambda - t_N^\lambda|^2] \leq c_2 \left( \frac{\sigma^2}{N} \right)^{2p/(2p+1)}$$

in particular, we have

$$E[|r_N^\lambda - t_N^\lambda|^2] \leq c_1 \lambda \|f\|_{H^p}^2 + c_2 \left( \frac{\sigma^2}{N} \right)^{2p/(2p+1)}$$

where $c_1$, $c_2$, and $c_3$ are positive constants depending on $p$ only.
The proof of Theorem 1 is deferred to the Appendix, where the following two lemmas are required.

**Lemma 2:** Let \( Z_1 \) be defined by \( \{ Z_k \} \geq \sigma_N k / \sqrt{N}, 0 \leq k < N \} \) and denote by \( | Z_1 | \) the number of elements in \( Z_1 \). If \( f \in H^p \), then

\[
| Z_1 | \leq C_p \left( \frac{N|f|_{H^p}^2}{\sigma^2} \right)^{1/(2p+1)}
\]

with a constant \( C_p > 0 \) depending on \( p \) only.

**Proof:** Since \( f \in H^p \) and \( b_1 \leq \sigma^p \leq b_2 \), we have

\[
| f |_{H^p}^2 \geq \sum_{k \in Z} a_k^2 | f_k |^2 \geq \sum_{k \in Z_1} a_k^2 | f_k |^2 \geq \frac{b_0 \sigma^2}{N} \sum_{k \in Z_1} a_k^2.
\]

If we denote by \( Z_1 \) the set \( \{ j, k : -1 \leq j < \log_2 | Z_1 |, 0 \leq k < a_j \} \), then \( | Z_1 | = | Z_1 | \) and

\[
| f |_{H^p}^2 \geq \frac{b_0 \sigma^2}{N} \sum_{k \in Z_1} a_k^2 \geq \frac{b_0 \sigma^2}{N} \log_2 | Z_1 | = \frac{2^p(2p+1) \max(0,j)}{2^p(2p+1)}
\]

and hence, the proof is complete.

**Lemma 3:** For any fixed integer \( l \), there exists a constant \( C_l \) such that

\[
\int_{|\eta| \geq x} \eta^2 \exp \left( -\frac{\eta^2}{2} \right) d\eta \leq C_l \frac{x^2}{x^2} \text{ for all } x \geq 1.
\]

**Proof:** Since \( x \geq 1 \), we have

\[
\int_{|\eta| \geq x} \eta^2 \exp \left( -\frac{\eta^2}{2} \right) d\eta = 2 \int_{x}^{\infty} \eta^2 \exp \left( -\frac{\eta^2}{2} \right) d\eta \leq 2 \int_{x}^{\infty} \eta^2 \exp \left( -\frac{\eta^2}{2} \right) d\eta = (2x^2 + 4) \exp \left( -\frac{x^2}{2} \right).
\]

Thus, for fixed integer \( l \), \( x \int_{|\eta| \geq x} \eta^2 \exp(-\eta^2/2) d\eta \) is always bounded, which implies the assertion.

By invoking Theorem 1, now we can give the convergence results for the Red-TOWER method.

**Theorem 2:** Let \( t_{\lambda} \) be defined by (8).

i) If \( f \in H^p \) with \( p > 1/4 \) and if \( \lambda \) is chosen as \( \lambda \sim (\sigma^2/N)|f|_{H^p}^2/2^{p(2p+1)} \), then

\[
E[|f^N - f^N - f_N|^2] \leq C \left( \frac{\sigma^2}{N} \right)^{2p/(2p+1)} |f|_{H^p}^{2p/(2p+1)}.
\]

ii) If \( f \in H^p \) with \( p > 1/2 \) and if \( \lambda_N \) is chosen by the GCV criterion, then

\[
E[|f^N - f_N|^2] \leq C \left( \frac{\sigma^2}{N} \right)^{2p/(2p+1)} |f|_{H^p}^{2p/(2p+1)}.
\]

**Proof:**

i) If \( \lambda \) is chosen as stated, then (13) gives the estimate

\[
E[|f^N - f_N|^2] \leq C \left( \frac{\sigma^2}{N} \right)^{2p/(2p+1)} |f|_{H^p}^{2p/(2p+1)}.
\]

Since [3, Th. 2] gives the same estimate for \( E[|r^\lambda_N - f_N|^2] \), we can use the inequality

\[
E(|X_1 + X_2|^2) \leq 2E X_1^2 + 2E X_2^2
\]

which is valid for arbitrary two random variables \( X_1 \) and \( X_2 \) to obtain (14).

ii) We first use [3, Th. 9] to derive

\[
E[|f^N - f_N|^2] \leq C \left( \frac{\sigma^2}{N} \right)^{2p/(2p+1)} |f|_{H^p}^{2p/(2p+1)}.
\]

Next, invoking Lemma 1 and (12), we can obtain the similar estimate for \( E[|f^N - f_N|^2] \). Therefore, invoking (16) again, we can conclude the proof of (15).

We have already mentioned that GCV is only one of the methods for estimating the optimal regularization parameter. This criterion could be substituted by a three-step one, where a consistent estimate of the variance of noise is provided, for example, by finite differences [31] as in [4] or by median absolute deviation (MAD) of the wavelet coefficients at the finest scale as in [21]; then, based on the variance just calculated, the optimal regularization parameter is estimated by minimizing a suitable functional, e.g., Mallows criterion [30], as in [4], and the approximate solution is evaluated for the regularization parameter just chosen.

Since the Mallows criterion (the same remark also holds for the Stein unbiased risk estimator (SURE) [32]) is a consistent estimate of the \( L_2 \) risk, the regularization parameter determined by this criterion should minimize the \( L_2 \) risk, and thus, it shares all the convergence properties of the GCV estimator. Therefore, the above convergence argument also applies when the GCV criterion is replaced by the Mallows criterion.

**IV. NUMERICAL EXPERIMENTS**

In order to evaluate the performance of TOWER methods, in this section, we report the numerical results based on synthetic data. Although the main concern of this paper is the Red-TOWER method, for which a full convergence proof is given, the performance is also evaluated for some other TOWERizations, in which the first guess is chosen by some well-known and consolidated estimators.

Let us formulate those TOWER methods we will consider in some detail.

**Hard-TOWER:** In this method, the first guess is chosen by the hard-thresholding method [21], and the solution \( T^H_k(t) \) obtained by this method is given as

\[
T^H_k(t) = s^T(Y_k, Y^H_k(t)Y_k), 0 \leq k < N
\]

where \( s^T \) is defined by (6), and \( t \) is the threshold. The threshold \( t \) is chosen such that the \( L_2 \) loss is minimized, which is possible since the actual solution is known; in practice, the solution (17) is computed for several values of the threshold, and the actual \( t \) is chosen as

\[
t := \arg \min_{t \geq 0} ||T^H_k(t) - f||^2.
\]
**Soft-TOWER:** In this method, the first guess is chosen by the soft-thresholding method [21], and the solution \( T^S_k(t) \) obtained by this method is given as
\[
\begin{align*}
T^S_k(t) &= s^T(Y_k, Y^S_k(t))_{+}, \\
Y^S_k(t) &= \text{sgn}(Y_k)(|Y_k| - t)_{+}
\end{align*}
\]
with \( s^T \) defined by (6). The optimal threshold \( t \) is chosen in the same way as in the Hard-TOWER method.

**Neigh-TOWER:** Here, the first guess is obtained by the neigh–coeff thresholding method [5], [13], and the approximate solution \( T^B_k \) is given by
\[
\begin{align*}
T^B_k &= s^T(Y_k, Y^B_k)_{+}, \\
Y^B_k &= Y_k (1 - \sigma^2/N)_{+}
\end{align*}
\]
with \( \sigma^2 \) estimated as in [4]. \( s^T \) is defined by (6), \( S^2_k = (Y^2_k - Y^{(j-1)}_k + Y^{(j)}_k)/2 \) for \( 1 < j < J, 0 < \ell < 2^j - 1 \) and at the boundaries of the scales \( S^2_k = Y^2_k, 0 \leq k \leq 3 \), \( S^2_{b0} = (Y^2_{b0} + Y^2_{p1} + Y^2_{p2})/2 \), and \( S^2_{b2} = (Y^2_{b2} + Y^2_{p3} + Y^2_{p4} + Y^2_{p5})/2 \) for \( 1 < j < J \). Neigh–coeff thresholding is considered here as a best performing prototype of wavelet blocking methods [13] (superscript \( B \) stands for blocking).

**Red-TOWER:** This method is dealt with in the present paper. Here, the first guess is given by a wavelet regularization method endowed with the GCV criterion, and the approximate solution \( T^R_k \) is given by
\[
T^R_k = s^T(Y_k, Y^R_k)_{+}, \\
Y^R_k = Y_k (1 - \sigma^2/N)_{+}
\]
where \( s^T \) is given by (6), \( Y^R_k \) is given by (2), and the regularization parameter is estimated by GCV criterion [see (4)].

We will not consider the objective (i.e., data driven) choice of the threshold \( t \) because it is beyond the scope of the paper. Such an efficient estimate exists, e.g., for soft-thresholding, and it is based on the GCV criterion [27]. We also mention that due to the nonlinearity of the TOWER method, a criterion that is optimal (or efficient) for devising a first guess may not be optimal when TOWER is applied. The only affirmative exception to this rule is given just by the Red-TOWER method, for which the present paper provides a full convergence analysis both for the method and for the criteria for choosing the regularization parameter.

We consider four sets of test functions (random-blocks, random-spires, random-Heavisine, and Doppler), representing the typical 1-D signals in many applications. These functions are defined in [34] as a generalization of [21]. We suppose that the value of these functions is known in the equispaced points in the interval \( [0, 1] \) according to the model (1) and that the noise affecting data is white with variance \( \sigma^2 \) such that the SNR is 7, i.e., \( \sum_{k=1}^{N-1} Y^2_k/\sigma^2 = 7^2 \).

In order to estimate the error of the retrieved solution, we define for each test function the performance index \( I \)
\[
I = \sqrt{\frac{\sum_{k=0}^{N-1} (F^\text{retrieved}_k - F_k)^2}{\sum_{k=0}^{N-1} F^2_k}}
\]
(18)

with \( F^\text{retrieved}_k \) being the wavelet coefficients retrieved by any method. Index \( I \) gives an estimate of the relative \( L_2 \) loss, and therefore, it is less sensitive to the particular function. According to [34], in order to simulate the variability of functions within each test set and to damp the influence of the random noise, 200 different functions and realizations of noise were generated, and index \( I \) was averaged over them for each test set. Of course, for a fair comparison, all the methods were tested on the same functions and noise sequences.

Tables I–IV show the index \( I \) [see (18)] together with its standard deviation over the repetitions for the considered methods (both plain and TOWERed) with several values of \( N \) for the four sets of functions.

Let us make some observations about the plain methods, i.e., methods without TOWER procedure. The tables confirm the expected worse behavior of the wavelet regularization method for the test functions when compared with the other methods generated by nonlinear shrinking functions. The performance of soft-thresholding seems worse than hard-thresholding. Moreover, the neigh–coeff thresholding shows the best performance. When we consider the TOWER versions of the above-mentioned methods, the performance is improved significantly, especially for soft-thresholding and wavelet regularization. Furthermore, TOWERization is even able to slightly improve the neigh–coeff thresholding method, which was claimed to perform best among wavelet estimators [13]. We also note that the performance of hard-soft- and Red-TOWER is quite similar, in particular, for the first two methods.

As examples, Figs. 1–4 show the plots of the soft-thresholding and the wavelet regularization retrievals, together with their TOWER versions, for a sample of the four test sets with size \( N = 2^2 \). The analysis of figures shows that the reduction of noise produced by the TOWER procedure with respect to the plain methods is clearly visible through much less jittering of the TOWERed solution around the actual ones.

V. CONCLUSIONS

This paper concerns the TOWER method, which was introduced in [1], to solve the problem of removing noise from data. By introducing a nonlinear shrinking function with double thresholds, this method could correct the predictor obtained by some wavelet-based methods.
A full convergence result is provided for the Red-TOWER method, where the TOWER procedure is performed with the predictor given by the wavelet regularization method. The objective (i.e., data-driven) choice is also considered for the regularization parameter estimated by the GCV criterion or by the Mallows criterion endowed with a consistent estimate of variance.

Numerical experiments are provided for some significant test functions available in the literature. The results confirm the strong improvement of the regularized solution by TOWERization. Moreover, numerical results are also reported when the TOWER method is applied to the most consolidated denoising methods such as the hard-thresholding, the soft-thresholding, and the neigh–coeff thresholding methods for which the convergence analysis is not available. The results also show the improvement by means of TOWERization. In particular, TOWERization is able to slightly improve the neigh–coeff

### Table II

<table>
<thead>
<tr>
<th>Method</th>
<th>$2^{16}$</th>
<th>$2^{12}$</th>
<th>$2^{11}$</th>
<th>$2^{10}$</th>
<th>$2^{8}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard</td>
<td>0.075±0.006</td>
<td>0.046±0.003</td>
<td>0.026±0.002</td>
<td>0.015±0.001</td>
<td>0.008±0.005</td>
</tr>
<tr>
<td>Hard-TOWER</td>
<td>0.075±0.006</td>
<td>0.046±0.003</td>
<td>0.026±0.002</td>
<td>0.015±0.001</td>
<td>0.008±0.005</td>
</tr>
<tr>
<td>Soft</td>
<td>0.066±0.005</td>
<td>0.058±0.003</td>
<td>0.037±0.002</td>
<td>0.025±0.001</td>
<td>0.013±0.007</td>
</tr>
<tr>
<td>Soft-TOWER</td>
<td>0.066±0.005</td>
<td>0.058±0.003</td>
<td>0.037±0.002</td>
<td>0.025±0.001</td>
<td>0.013±0.007</td>
</tr>
<tr>
<td>Neigh-Coef</td>
<td>0.072±0.006</td>
<td>0.042±0.003</td>
<td>0.024±0.002</td>
<td>0.013±0.009</td>
<td>0.007±0.004</td>
</tr>
<tr>
<td>Neigh-TOWER</td>
<td>0.072±0.006</td>
<td>0.042±0.003</td>
<td>0.024±0.002</td>
<td>0.013±0.009</td>
<td>0.007±0.004</td>
</tr>
<tr>
<td>Regularization</td>
<td>0.10±0.02</td>
<td>0.062±0.01</td>
<td>0.037±0.007</td>
<td>0.023±0.004</td>
<td>0.017±0.005</td>
</tr>
<tr>
<td>Red-TOWER</td>
<td>0.096±0.007</td>
<td>0.046±0.005</td>
<td>0.026±0.004</td>
<td>0.015±0.002</td>
<td>0.009±0.002</td>
</tr>
</tbody>
</table>

### Table III

<table>
<thead>
<tr>
<th>Method</th>
<th>$2^{16}$</th>
<th>$2^{12}$</th>
<th>$2^{11}$</th>
<th>$2^{10}$</th>
<th>$2^{8}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard</td>
<td>0.084±0.007</td>
<td>0.030±0.004</td>
<td>0.018±0.002</td>
<td>0.011±0.001</td>
<td>0.008±0.005</td>
</tr>
<tr>
<td>Hard-TOWER</td>
<td>0.084±0.007</td>
<td>0.030±0.004</td>
<td>0.018±0.002</td>
<td>0.011±0.001</td>
<td>0.008±0.005</td>
</tr>
<tr>
<td>Soft</td>
<td>0.057±0.006</td>
<td>0.026±0.004</td>
<td>0.015±0.001</td>
<td>0.014±0.007</td>
<td></td>
</tr>
<tr>
<td>Soft-TOWER</td>
<td>0.057±0.006</td>
<td>0.026±0.004</td>
<td>0.015±0.001</td>
<td>0.014±0.007</td>
<td></td>
</tr>
<tr>
<td>Neigh-Coef</td>
<td>0.049±0.008</td>
<td>0.030±0.004</td>
<td>0.018±0.002</td>
<td>0.010±0.008</td>
<td>0.007±0.005</td>
</tr>
<tr>
<td>Neigh-TOWER</td>
<td>0.049±0.008</td>
<td>0.030±0.004</td>
<td>0.018±0.002</td>
<td>0.010±0.008</td>
<td>0.007±0.005</td>
</tr>
<tr>
<td>Regularization</td>
<td>0.052±0.01</td>
<td>0.039±0.008</td>
<td>0.027±0.006</td>
<td>0.019±0.004</td>
<td>0.017±0.006</td>
</tr>
<tr>
<td>Red-TOWER</td>
<td>0.049±0.009</td>
<td>0.031±0.005</td>
<td>0.020±0.003</td>
<td>0.013±0.002</td>
<td>0.009±0.002</td>
</tr>
</tbody>
</table>

### Table IV

<table>
<thead>
<tr>
<th>Method</th>
<th>$2^{16}$</th>
<th>$2^{12}$</th>
<th>$2^{11}$</th>
<th>$2^{10}$</th>
<th>$2^{8}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard</td>
<td>0.063±0.004</td>
<td>0.037±0.002</td>
<td>0.020±0.001</td>
<td>0.011±0.001</td>
<td>0.006±0.003</td>
</tr>
<tr>
<td>Hard-TOWER</td>
<td>0.063±0.004</td>
<td>0.037±0.002</td>
<td>0.020±0.001</td>
<td>0.011±0.001</td>
<td>0.006±0.003</td>
</tr>
<tr>
<td>Soft</td>
<td>0.074±0.004</td>
<td>0.046±0.002</td>
<td>0.030±0.008</td>
<td>0.019±0.005</td>
<td>0.012±0.002</td>
</tr>
<tr>
<td>Soft-TOWER</td>
<td>0.074±0.004</td>
<td>0.046±0.002</td>
<td>0.030±0.008</td>
<td>0.019±0.005</td>
<td>0.012±0.002</td>
</tr>
<tr>
<td>Neigh-Coef</td>
<td>0.058±0.003</td>
<td>0.034±0.002</td>
<td>0.018±0.001</td>
<td>0.010±0.004</td>
<td>0.005±0.002</td>
</tr>
<tr>
<td>Neigh-TOWER</td>
<td>0.058±0.003</td>
<td>0.034±0.002</td>
<td>0.018±0.001</td>
<td>0.010±0.004</td>
<td>0.005±0.002</td>
</tr>
<tr>
<td>Regularization</td>
<td>0.086±0.003</td>
<td>0.054±0.001</td>
<td>0.033±0.000</td>
<td>0.017±0.000</td>
<td>0.009±0.002</td>
</tr>
<tr>
<td>Red-TOWER</td>
<td>0.086±0.003</td>
<td>0.054±0.001</td>
<td>0.033±0.000</td>
<td>0.017±0.000</td>
<td>0.009±0.002</td>
</tr>
</tbody>
</table>
method, which was claimed to be the best performing wavelet estimator.

APPENDIX

PROOF OF THEOREM 1

The claim (13) is an immediate consequence of (12) if we use the estimate (9), Therefore, we need only to prove the claim (12). In the following, we always use $C$ to denote a generic constant depending on $p$ only.

Note that

$$E[|t_N^\lambda - r_N^\lambda|^2] = \sum_{k \in Z} E(T_k(Y_k) - R_k(Y_k))^2.$$ 

If we use $\gamma_k(y)$ to denote the density function of the random variable $Y_k \sim N(F_k, (\sigma^2 \alpha_k^2/N))$, then

$$E[|t_N^\lambda - r_N^\lambda|^2] = \sum_{k \in Z} \int_{-\infty}^{\infty} (T_k(y) - R_k(y))^2 \gamma_k(y) \, dy.$$ 

According to the definition of $T_k$, if we introduce the sets

$$S_1^k = \left\{ y \mid |y| < \frac{2\sigma \alpha_k}{\sqrt{N}} \right\},$$

$$S_2^k = \left\{ y \mid |y| \geq \frac{2\sigma \alpha_k}{\sqrt{N}}, \left| R_k(y) \right| > \left| Q_k^+ \right| \right\},$$

$$S_3^k = \left\{ y \mid |y| \geq \frac{2\sigma \alpha_k}{\sqrt{N}}, \left| R_k(y) \right| < \left| Q_k^- \right| \right\},$$

$$S_4^k = \left\{ y \mid |y| \geq \frac{2\sigma \alpha_k}{\sqrt{N}}, \left| Q_k^- \right| \leq \left| R_k(y) \right| \leq \left| Q_k^+ \right| \right\},$$

then

$$E[|t_N^\lambda - r_N^\lambda|^2] = \sum_{k \in Z} \left\{ \left( \int_{S_1^k} + \int_{S_2^k} \right) R_k(y)^2 \gamma_k(y) \, dy \right. \left. + \left( \int_{S_3^k} + \int_{S_4^k} \right) (R_k(y) - Q_k^\pm)^2 \gamma_k(y) \, dy \right\}$$

$$= \sum_{k \in Z} \{ I_1^k + I_2^k + I_3^k + I_4^k \}.$$ 

In the following, we estimate the four terms separately.

For $I_1^k$, we have

$$I_1^k = \int_{S_1^k} \frac{y^2}{(1 + \alpha_k^2 \sigma_k^2 \lambda^2)^2} \gamma_k(y) \, dy \leq \frac{4\sigma^2 \alpha_k^2}{N(1 + \alpha_k^2 \sigma_k^2 \lambda^2)^2}.$$ 

Therefore, using the estimate (9), we can obtain

$$\sum_{k \in Z} I_1^k \leq \frac{4\sigma^2}{N} \sum_{k \in Z} \frac{s_k^2}{(1 + \alpha_k^2 \sigma_k^2 \lambda^2)^2} \leq C \frac{\sigma^2}{N^{1/2p}}.$$

For $I_2^k$, on one hand, we have

$$I_2^k \leq \int_{-\infty}^{\infty} R_k(y)^2 \gamma_k(y) \, dy = \frac{R_k^2 + \sigma^2 s_k^2}{N(1 + \alpha_k^2 \sigma_k^2 \lambda^2)^2}$$

and on the other hand, noting that $|R_k(y)| < |Q_k(y)| \leq \sigma/s_k \sqrt{N}$, we also have

$$I_2^k \leq \frac{\sigma^2 s_k^2}{N} \int_{S_2^k} \gamma_k(y) \, dy \leq \frac{\sigma^2 s_k^2}{N}.$$
Let us divide the set $Z$ into the two subsets $Z_1$ and $Z_2 := Z - Z_1$; then, invoking Lemma 2, we have

$$
\sum_{k \in Z} I_k^2 \leq \sum_{k \in Z_1} \frac{\sigma_k^2}{N} + \sum_{k \in Z_2} \frac{F_k^2 + \sigma_k^2 S_k^2/N}{(1 + \sigma_k^2 S_k^2 \lambda)^2} \\
\leq C a_k^2 |Z_1| \frac{\sigma_k^2}{N} + \frac{2a_k^2}{N} \sum_{j=1}^{2^p} \sum_{\ell=0}^{j-1} \left(1 + \frac{\sigma_k^2 S_k^2 \lambda}{N}\right)^2 \\
\leq C_p \left(\frac{\sigma_k^2}{N}\right)^{2/(2p+1)} ||f||_{L^p}^{2/(2p+1)} + \frac{C a_k^2 \sigma_k^2}{N \lambda^{1/2p}}. \tag{19}
$$

For $I_3$, it is easy to know

$$
I_3 = \int_{S_k^2} y^2 \left(1 - \frac{1}{1 + \sigma_k^2 S_k^2 \lambda} - \frac{1}{2} \sqrt{1 - \frac{4a_k^2 S_k^2}{N y^2} \sigma_k^2 S_k^2 \lambda} \right)^2 \gamma_k(y) dy \\
= \int_{S_k^2} y^2 \left(1 - \frac{1}{1 + \sigma_k^2 S_k^2 \lambda} - \frac{1}{2} \sqrt{1 - \frac{4a_k^2 S_k^2}{N y^2} \sigma_k^2 S_k^2 \lambda} \right)^2 \gamma_k(y) dy \\
\leq \int_{S_k^2} y^2 \left(1 - \frac{1}{1 + \sigma_k^2 S_k^2 \lambda} - \frac{1}{2} \sqrt{1 - \frac{4a_k^2 S_k^2}{N y^2} \sigma_k^2 S_k^2 \lambda} \right)^2 \gamma_k(y) dy \\
\leq \min \left\{ \frac{\sigma_k^2}{4N} \frac{F_k^2 + \sigma_k^2 S_k^2/N}{(1 + \sigma_k^2 S_k^2 \lambda)^2} \right\}.
$$

Hence, similar to the derivation of (19), we can obtain

$$
\sum_{k \in Z} I_k^2 \leq C_p \left(\frac{\sigma_k^2}{N}\right)^{2/(2p+1)} ||f||_{L^p}^{2/(2p+1)} + \frac{C \sigma_k^2 |Z_1|}{N \lambda^{1/2p}}.
$$

Finally, we estimate the term $\sum_{k \in Z} I_k^2$. We first note that

$$
I_k^2 = \int_{S_k^2} y^2 \left(1 + \frac{1}{2} \sqrt{1 - \frac{4a_k^2 S_k^2}{N y^2} - \frac{1}{1 + \sigma_k^2 S_k^2 \lambda}} \right)^2 \gamma_k(y) dy \\
\leq \left(\frac{a_k^2 S_k^2 \lambda}{1 + \sigma_k^2 S_k^2 \lambda}\right)^2 \int_{S_k^2} y^2 \gamma_k(y) dy \\
\leq \left(\frac{a_k^2 S_k^2 \lambda}{1 + \sigma_k^2 S_k^2 \lambda}\right)^2 \left(F_k^2 + \sigma_k^2 S_k^2/N \right).
$$

Therefore, by invoking Lemma 2, it follows that

$$
\sum_{k \in Z} I_k^2 \leq G_N(\lambda) + \sum_{k \in Z_1} \frac{\sigma_k^2}{N} \leq G_N(\lambda) + C \frac{|Z_1|}{N} \\
geq G_N(\lambda) + C \left(\frac{\sigma_k^2}{N}\right)^{2/(2p+1)} ||f||_{L^p}^{2/(2p+1)}.
$$

Now, we have to estimate $\sum_{k \in Z_2} I_k^2$. We distinguish two cases: $\sigma_k^2 S_k^2 \lambda < 1$ and $\sigma_k^2 S_k^2 \lambda \geq 1$. For the first case, we always have $j(k) < 1/2p \log_2(1/b_k \lambda)$. Hence, from (20), we obtain

$$
\sum_{k \in Z_2, \sigma_k^2 S_k^2 \lambda < 1} I_k^2 \leq \sum_{j=1}^{1/2p \log_2(1/b_k \lambda)} \sum_{\ell=0}^{j-1} \frac{2a_k^2 \sigma_k^2 \lambda}{N} \leq C \frac{\sigma_k^2 \lambda^{-1/2p}}{N^2}.
$$

In the following, we consider the second case. Since $|R_k(y)| \geq \Omega_k^2(y)$, we have

$$
\frac{1}{1 + \sigma_k^2 S_k^2 \lambda} \geq \frac{1}{2} - \frac{1}{2} \sqrt{1 - \frac{4a_k^2 S_k^2}{N y^2}}.
$$

Solving this inequality gives

$$
y^2 \geq \frac{a_k^2 S_k^2 \lambda}{N} \left(1 + \frac{\sigma_k^2}{N} \right)^2 \left(\geq \frac{4a_k^2 S_k^2}{N}\right).
$$

Therefore

$$
I_4^k = \int_{|y| \geq \left(\sigma_k^2 + \frac{a_k^2 S_k^2 \lambda}{(\sqrt{N} a_k^2 S_k^2 \lambda)^{1/2}}\right)} y^2 \gamma_k(y) dy \\
\leq \left(\frac{a_k^2 S_k^2 \lambda}{1 + \sigma_k^2 S_k^2 \lambda}\right)^2 \int_{|y| \geq \left(\sigma_k^2 + \frac{a_k^2 S_k^2 \lambda}{(\sqrt{N} a_k^2 S_k^2 \lambda)^{1/2}}\right)} y^2 \gamma_k(y) dy \\
\leq \left(\frac{a_k^2 S_k^2 \lambda}{1 + \sigma_k^2 S_k^2 \lambda}\right)^2 \left(F_k + \sigma_k^2 S_k^2/N \right).$$

Using the expression

$$
\gamma_k(y) = \frac{\sqrt{N}}{2\pi \sigma_k^2} \exp\left(-\frac{(y - F_k)^2}{2\sigma_k^2 S_k^2/N}\right)
$$

and making the transformation $t = \sqrt{N}(y - F_k)/\sigma_k$, we have

$$
I_4^k \leq \frac{1}{\sqrt{2\pi}} \left(\frac{F_k + \sigma_k^2 S_k^2 \lambda}{1 + \sigma_k^2 S_k^2 \lambda}\right)^2 \\
\cdot \int_{|t| \geq \left(\sigma_k^2 + \frac{a_k^2 S_k^2 \lambda}{(\sqrt{N} a_k^2 S_k^2 \lambda)^{1/2}}\right)} \left(F_k + \sigma_k^2 S_k^2/N \right) \exp\left(-\frac{t^2}{2}\right) dt \\
\leq \frac{2F_k^2 a_k^2 S_k^2 \lambda^2}{(1 + \sigma_k^2 S_k^2 \lambda)^2} + \frac{2a_k^2 S_k^2}{\sqrt{2\pi} N} \int_{|t| \geq \left(\sigma_k^2 + \frac{a_k^2 S_k^2 \lambda}{(\sqrt{N} a_k^2 S_k^2 \lambda)^{1/2}}\right)} \exp\left(-\frac{t^2}{2}\right) dt.
$$

Since $|\sqrt{N}F_k/(\sigma_k S_k)| \leq 1$ and $(1 + \sigma_k^2 S_k^2 \lambda)/(\sigma_k^2 S_k^2 \lambda^{1/2}) \geq 2$, we can use Lemma 3 to obtain

$$
0$$
\[
\frac{P_k}{l_k} \leq \frac{2\tau^2 \alpha_k^2}{(1 + \alpha_k^2) s_k^2 \lambda^2} + \frac{2\tau^2 s_k^2}{\sqrt{2\pi} N} \cdot \int_{|t| \geq (1 + \alpha_k^2) s_k^2 \lambda^2} t^2 \exp\left(-\frac{t^2}{2}\right) dt \leq \frac{2\tau^2 \alpha_k^2}{(1 + \alpha_k^2) s_k^2 \lambda^2} + C \frac{\sigma^2}{N} \frac{s_k^2}{(1 + \alpha_k^2) s_k^2 \lambda^2} \leq \frac{2\tau^2 \alpha_k^2}{(1 + \alpha_k^2) s_k^2 \lambda^2} + C \frac{\sigma^2}{N} \frac{s_k^2}{(1 + \alpha_k^2) s_k^2 \lambda^2}^{l/2}.
\]

Thus, by taking \( l = 4 \), we have
\[
\sum_{k \in Z^2; \alpha_k^2 s_k^2 \lambda \geq 1} \frac{P_k}{l_k} \leq 2G_N(\lambda) + \sum_{k \in Z} \frac{C\sigma^2}{N} \frac{s_k^2}{(1 + \alpha_k^2) s_k^2 \lambda}^{l/2} \leq 2G_N(\lambda) + C \frac{\sigma^2}{N\lambda^{l/2}}.
\]

Combining the above analysis, we complete the proof.

**Acknowledgment**

The authors would like to thank the referees for their useful comments that improved this paper.

**References**


Umberto Amato received the degree in physics from Napoli University, Napoli, Italy, in 1981. Since 1986, he has been with the Istituto per Applicazioni della Matematica, Italian National Research Council, Napoli, as a Researcher. His research interests have been in the areas of computational fluid dynamics (simulation of air motion inside cylinders of engines and shallow water models in the atmosphere) and inverse problems (maximum entropy regularization and aerosol retrieval from solar radiation measurements). Presently, he is involved in signal processing (Fourier smoothing, linear and nonlinear wavelet denoising, classification, and compression).

Qi-nian Jin received the B.S. degree from the Department of Mathematics, Anhui Normal University, Anhui, China, in 1992 and the M.S. degree from the Shanghai Institute of Applied Mathematics, Shanghai, China, in 1994. He is currently pursuing the Ph.D. degree with the Department of Mathematics, Purdue University, West Lafayette, IN.

He was a faculty member with the Department of Mathematics at Nanjing University, Nanjing, China. From October through December 1998, he visited the Istituto per Applicazioni della Matematica CNR, Napoli, Italy, where he worked on smoothing data problems and signal processing. His current research interest is inverse and ill-posed problems.