Arbitrary Shrinkage Rules for Approximation Schemes with Sparsity Constraints

Martin Ehler* Simone Geisel†

February 3, 2009

Finding a sparse representation of a possibly noisy signal can be modeled as a variational minimization with ℓ_q-sparsity constraints for q less than one. Especially for real-time and on-line applications, one requires fast computations of these minimizers. However, there are no sufficiently fast algorithms, and to circumvent this limitation, we consider minimization up to a constant factor. We verify that q-dependent modifications of shrinkage rules provide closed formulas for such minimizers, and we introduce a new shrinkage rule which is adapted to q.

To support the concept of shrinkage rules and minimizers up to a constant factor, we finally apply different shrinkage rules to wavelet-based variational image denoising. We verify in our numerical experiments that the H-curve criterion which is a parameter selection method already being successfully applied to soft-thresholding can yield better results if we replace soft- by other shrinkage rules.

Key words: shrinkage, variational problems, sparsity, frames, denoising, H-curve criterion

1 Introduction

Decomposing signals into simple building blocks and reconstructing from shrinked coefficients are used in signal representation and processing such as noise removal, compression as well as texture and boundary enhancement. For instance, wavelet shrinkage is applied to remote and subsurface sensing, where shrinking wavelet coefficients is used for noise and clutter reduction in speckled Synthetic Aperture Radar images, improving the performance of detection systems, cf. [26]. Statistical approaches and Bayesian objectives for noise removal make use of various shrinkage strategies, cf. [9, 16, 17, 29]. Variational models as in [5] justify shrinkage by smoothness estimates of the unperturbed signal. Other shrinkage rules are derived from a diffusion approach in [25].

Signal approximation with sparsity constraints leads to variational minimization problems, and the denoising approach in [5] is a particular case, see also Section 7. The expression to be minimized is a

*Department of Mathematics, Norbert Wiener Center, University of Maryland, College Park, MD 20742-4015, ehlermar@math.umd.edu, and Section on Medical Biophysics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, 9 Memorial Dr, Bethesda MD 20892-0924, ehlermar@mail.nih.gov

†Forschungsgesellschaft angewandter Naturwissenschaften, Forschungsinstitut fuer Hochfrequenzphysik und Radartechnik, Neuenahrer Str. 20, 53343 Wachtberg/Germany, simone.geisel@fgan.de

Dagstuhl Seminar Proceedings 08492
Structured Decompositions and Efficient Algorithms
http://drops.dagstuhl.de/opus/volltexte/2009/1881

sum of an approximation error and a penalty term which involves weighted ℓ_q-constraints, see Section 2.1. In [7], iteratively shrinking coefficients of an orthonormal basis expansion provides a sequence converging towards the minimizer. The method covers the convex case $q \in [1,2]$, but sparse signal representation, coding, and signal analysis requires the consideration of redundant basis-like systems and the nonconvex case $q \in [0,1)$ as well, see for instance [2, 6, 18]. By using hard-shrinkage, the iterative algorithm in [7] converges towards a local minimum for $q = 0$, cf. [3]. However, it could be far off the global minimum, it does not cover $q \in (0,1)$, and, for applications where computation time is crucial, iterative algorithms might be too time-consuming.

In the present paper, we obtain complementary results for $q \in [0,1)$ in terms of minimization up to a constant factor. In fact, we verify that such a minimization can be derived from q-dependent modifications of shrinkage rules. This means we have a closed formula for these minimizers, which allows for a fast computation. It is a good initial guess of the exact solution. We also introduce new shrinkage rules which are adapted to q. They could substitute the hard-shrinkage of the iterative algorithm in [3] to improve the initial guess for $q \in (0,1)$.

Beside shrinkage rules, the choice of the shrinkage parameter is essential for good results. For wavelet-based variational image denoising with $q = 1$ as proposed in [5], the H-curve criterion is a shrinkage parameter selection method which was adapted from the L-curve method in regularization theory of inverse problems to wavelet-based image denoising, cf. [24]. Although the method does not cover scale dependent parameters, it is a powerful heuristic parameter selection since it is not restricted to white noise. It can be applied to many other noise characteristics as well, and it has already been successfully applied to soft-shrinkage, cf. [15, 24]. As far as we know, the H-curve citerion has not yet been applied to shrinkage rules beyond soft-shrinkage. To support the concept of minimizers up to a constant factor, we finally apply different shrinkage rules to wavelet-based variational image denoising. In our numerical experiments, we verify that the H-curve criterion can be applied to shrinkage rules beyond soft-shrinkage and that there are shrinkage rules, whose combination with the H-curve criterion outperforms soft-shrinkage.

The outline is as follows: In Section 2, we present the variational problems under consideration and we recall the concept of frames. We introduce shrinkage rules in Section 3. The main results are presented in Section 4, and in Section 5 we apply the results to sparse signal representation. We introduce a new family of shrinkage rules in Section 6. Wavelet-based variational image denoising is addressed in Section 7, and conclusions are given in Section 8.

2 Variational Problems and Frames

2.1 Variational Minimization Problems

Let L be a bounded operator between two Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$, and let $\{\tilde{f}_n\}_{n\in\mathcal{N}}$ be a countable collection in $\mathcal{H}$. Given $h \in \mathcal{H}'$, we consider the minimization problem

$$\min_{g\in\mathcal{H}} \big(\|h - Lg\|^2_{\mathcal{H}'} + \sum_{n\in\mathcal{N}} \alpha_n |\langle g, \tilde{f}_n\rangle|^q\big), \tag{1}$$

where $q \in (0,2]$, $(\alpha_n)_{n\in\mathcal{N}}$ is a sequence of nonnegative numbers, and $\langle\cdot,\cdot\rangle$ denotes the inner product. This makes also sense for $q = 0$ with $0^0 = 0$, and the penalty term then counts the nonzero entries of $(\langle g, \tilde{f}_n\rangle)_{n\in\mathcal{N}}$ weighted by $(\alpha_n)_{n\in\mathcal{N}}$. For $\mathcal{H} = \mathcal{H}'$ and $L = \mathrm{id}_{\mathcal{H}}$, problem (1) is relevant in wavelet based signal denoising. There, $\{\tilde{f}_n\}_{n\in\mathcal{N}}$ is a wavelet system, and the sparsity constraint on the right hand side of (1) is related to the Besov regularity of the signal to be recovered, see [5] for details. Our approach is neither restricted to L being the identity nor must L be injective. However, we assume throughout the short note that it has a **bounded pseudo inverse**, i.e. there is a bounded operator $L^{\#} : \mathcal{H}' \mapsto \mathcal{H}$ such that $LL^{\#}L = L$. The iterative shrinkage for $q \in [1,2]$ in [7] does not require this assumption.

The sequence $(\alpha_n)_{n\in\mathcal{N}}$ is a collection of variable parameters which must be fitted to h and L. In case $(\alpha_n)_{n\in\mathcal{N}} = \alpha$,

$$\alpha \mapsto \big(\|h - Lg^{\alpha}\|^2_{\mathcal{H}'}, \sum_{n\in\mathcal{N}} |\langle g^{\alpha}, \tilde{f}_n\rangle|^q\big) \tag{2}$$

is considered as a curve in $\mathbb{R}^2$, where g^{α} is a minimizer of (1). Finally, one chooses α according to a point of outstanding curvature, see [21] and [24] for the *L-curve* and *H-curve criterion*, respectively. It requires to compute minimizers g^{α} for many different values of α, and g^{α} must be efficiently computable. This is another motivation for avoiding costly iterative minimization schemes beside real-time and on-line applications.

Handling nonstationary noise requires $(\alpha_n)_{n\in\mathcal{N}} \neq \alpha$, but it is often still reasonable to assume that there are positive constants a and b such that

$$a \leq \alpha_n \leq b, \text{ for all } n \in \mathcal{N}. \tag{3}$$

2.2 Bi-frames

The singular value decomposition of L is considered in [22] to address $q \in [0,1)$. Then $\{\tilde{f}_n\}$ in (1) is supposed to be an orthonormal basis for $\mathcal{H}$ which diagonalizes L. However, this is of limited interest in practical applications. We will consider redundant basis-like systems and L is not required to be diagonalized: a countable collection $\{f_n\}_{n\in\mathcal{N}}$ in $\mathcal{H}$ is a *frame* if there are two positive constants A, B such that

$$A\|g\|^2_{\mathcal{H}} \leq \sum_{n\in\mathcal{N}} |\langle g, f_n\rangle|^2 \leq B\|g\|^2_{\mathcal{H}}, \text{ for all } g \in \mathcal{H}. \tag{4}$$

If $\{f_n\}_{n\in\mathcal{N}}$ is a frame, then its *synthesis operator*

$$F : \ell_2(\mathcal{N}) \to \mathcal{H}, \quad (c_n)_{n\in\mathcal{N}} \mapsto \sum_{n\in\mathcal{N}} c_n f_n, \tag{5}$$

is onto. Each $g \in \mathcal{H}$ then has a series expansion, but we still have to find its coefficients. The synthesis operator's adjoint

$$F^* : \mathcal{H} \to \ell_2(\mathcal{N}), \quad g \mapsto (\langle g, f_n\rangle)_{n\in\mathcal{N}} \tag{6}$$

is called *analysis operator*, $S = FF^*$ is invertible, and $\{S^{-1}f_n\}_{n\in\mathcal{N}}$ is called *canonical dual frame* and expands

$$g = \sum_{n\in\mathcal{N}} \langle g, S^{-1}f_n\rangle f_n, \text{ for all } g \in \mathcal{H}.$$

The inversion of S is complicated, and, since F need not be injective, there could be 'better' coefficients than $\langle g, S^{-1}f_n\rangle$. This motivates the following: two frames $\{f_n\}_{n\in\mathcal{N}}$ and $\{\tilde{f}_n\}_{n\in\mathcal{N}}$ are called a *pair of dual frames* (or a *bi-frame*) if

$$g = \sum_{n\in\mathcal{N}} \langle g, \tilde{f}_n\rangle f_n, \text{ for all } g \in \mathcal{H}, \tag{7}$$

i.e., $F\widetilde{F}^* = \mathrm{id}_{\mathcal{H}}$, where $\widetilde{F}^*$ is the dual frame's analysis operator. For instance, the canonical dual of a wavelet frame may not have the wavelet structure as well, but it can possibly replaced by an alternative dual wavelet frame, cf. [11, 14].

Throughout the paper while considering (1), we suppose that $\{f_n\}_{n\in\mathcal{N}}$ and $\{\tilde{f}_n\}_{n\in\mathcal{N}}$ are a **bi-frame for** $\mathcal{H}$.

3 Shrinkage Rules

To solve (1), shrinkage plays a crucial role. Following ideas in [27], we call a function $\varrho : \mathbb{C} \times \mathbb{R}_{\geq 0} \to \mathbb{C}$ a *shrinkage rule* if there are constants $C_1, C_2, \rho, D > 0$ such that both conditions

$$|x - \varrho(x,\alpha)| \leq C_1 \min(|x|,\alpha), \text{ for all } \alpha \geq 0,\, x \in \mathbb{C}, \tag{8}$$

$$|\varrho(x,\alpha)| \leq C_2 |x| \Big|\frac{x}{\alpha}\Big|^{\rho}, \text{ for all } \alpha > 0,\, |x| \leq D\alpha, \tag{9}$$

are satisfied. A shrinkage rule ϱ is called a *thresholding rule* if there is a constant $C_3 > 0$ such that $|x| \leq C_3\alpha$ implies $\varrho(x,\alpha) = 0$. A thresholding rule allows for $\rho = \infty$ in (9), where we use $a^\infty = 0$ if $0 \leq a < 1$. We will recall a few common shrinkage rules and we restrict us to $x \in \mathbb{R}$: *Soft-shrinkage* is given by $\varrho_s(x,\alpha) = (x - \frac{x}{|x|}\alpha)\mathbf{1}_{\{|x|>\alpha\}}$. Contrary to soft- and *hard-shrinkage* $\varrho_h(x,\alpha) = x\mathbf{1}_{\{|x|>\alpha\}}$, the *nonnegative garotte-shrinkage* rule $\varrho_g(x,\alpha) = (x - \frac{\alpha^2}{x})\mathbf{1}_{\{|x|>\alpha\}}$ is continuous and large coefficients are left almost unaltered. It has been successfully applied to image denoising in [19]. Similar properties has *hyperbolic shrinkage* $\varrho^{hy}(x,\alpha) = \operatorname{sign}(x)\sqrt{x^2-\alpha^2}\mathbf{1}_{\{|x|>\alpha\}}(x)$, cf. [27].

The *n-degree garotte shrinkage* rule is given by $\varrho^n(x,\alpha) = \frac{x^{2n+1}}{x^{2n}+\alpha^{2n}}$, see [27]. For $k \in \mathbb{N}$, the twice differentiable rule

$$\varrho_k(x,\alpha) = \begin{cases} \frac{x^{2k+1}}{(2k+1)\alpha^{2k}}, & |x| \leq \alpha \\ x - \operatorname{sign}(x)(\alpha - \frac{\alpha}{2k+1}), & |x| > \alpha \end{cases} \tag{10}$$

is considered in [29]. Both rules are shrinkage rules with $\rho = 2k = 2n$ and $C_2 = 1$. The rules $\varrho(x,\alpha) = x\big(1 - \sqrt{\frac{\alpha^2}{\alpha^2+2x^2}}\big)$ and $\varrho(x,\alpha) \approx x\exp(-0.2\frac{\alpha^8}{x^8})$ are based on diffusion, see [25], and one verifies that both are shrinkage rules with $\rho = 1$.

Bruce and Gao proposed *firm-shrinkage* $\varrho_f(x,\alpha_1,\alpha_2) = x\mathbf{1}_{\{|x|>\alpha_2\}} + \frac{x}{|x|}\frac{\alpha_2(|x|-\alpha_1)}{\alpha_2-\alpha_1}\mathbf{1}_{\{\alpha_1\leq|x|\leq\alpha_2\}}$ in [20]. For fixed α_1, the mapping $(x,\alpha) \mapsto \varrho_f(x,\alpha_1,\alpha)$ is a thresholding rule.

4 Main Results

For $q \in [0,2]$, let $\ell_q^{(\alpha_n)}(\mathcal{N})$ be the space of complex-valued sequences $(\omega_n)_{n\in\mathcal{N}}$ such that $\|\omega\|^q_{\ell_q^{(\alpha_n)}} := \sum_{n\in\mathcal{N}} \alpha_n|\omega_n|^q$ is finite. One observes that $\sum_{n\in\mathcal{N}} \alpha_n |\langle g, \tilde{f}_n\rangle|^q = \|\widetilde{F}^* g\|^q_{\ell_q^{(\alpha_n)}}$, and to shorten notation, we denote

$$\mathcal{J}_q(h,g) = \|h - Lg\|^2_{\mathcal{H}'} + \|\widetilde{F}^* g\|^q_{\ell_q^{(\alpha_n)}}.$$

The idea for the following main result is to replace a shrinkage rule $\varrho(x,\alpha)$ by its q-dependent expression $\varrho(x,\alpha|x|^{q-1})$. Due to (8), it vanishes as $x \neq 0$ goes to 0, and we apply $\varrho(x,\alpha|x|^{q-1}) = 0$ for $x = 0$. If $\rho = \infty$, we use $\frac{1}{\rho} = 0$. Since $\varrho_g(x,\alpha) = \varrho_s(x,\alpha^2|x|^{-1})$, the nonnegative garotte is q-dependent soft-shrinkage for $q = 0$ and α replaced by α^2.

Theorem 4.1. *Let ϱ be a shrinkage rule with $\rho \in [\frac{1}{2},\infty]$. Suppose that $\widetilde{F}^* L^\# LF$ is bounded on $\ell_{1/\rho}^{(\alpha_n)}(\mathcal{N})$. Let $q = \frac{1}{\rho}$, then there is a constant $C > 0$ such that for all $h \in \operatorname{range}(L)$, and for all $g \in \mathcal{H}$*

$$\mathcal{J}_q(h,\hat{g}) \leq C\mathcal{J}_q(h,g),$$

where $\hat{g} = L^\# LF\varrho(v_n,\alpha_n|v_n|^{q-1})_{n\in\mathcal{N}}$ with $v = \widetilde{F}^* L^\# h$.

If $\widetilde{F}^* F$ is also bounded on $\ell_{1/\rho}^{(\alpha_n)}(\mathcal{N})$, one can choose $\hat{g} = F\varrho(v_n,\alpha_n|v_n|^{q-1})_{n\in\mathcal{N}}$. If (3) holds, then the statements extend to all $q \in [\frac{1}{\rho},2]$, and C is independent of q.

Remark 4.2. If the bi-frame is biorthogonal and $F\ell_{1/\rho}^{(\alpha_n)} \subset \mathrm{range}(L^{\#}L)$, then $\widetilde{F}^*L^{\#}LF = \mathrm{id}_{\ell_{1/\rho}^{(\alpha_n)}}$, because $\widetilde{F}^*F$ is the identity and $L^{\#}L$ is the identity on its range. The boundedness condition is then trivially satisfied as it is for finite $\mathcal{N}$.

To prove Theorem 4.1, we consider a decoupled minimization problem: given $v \in \ell_2(\mathcal{N})$, we try to minimize

$$\mathcal{I}_q(v,\omega) = \big(\|v-\omega\|_{\ell_2}^2 + \sum_{n\in\mathcal{N}} \alpha_n|\omega_n|^q\big) \tag{11}$$

over $\omega \in \ell_2(\mathcal{N})$. It turns out that minimizing (1) and (11) up to a constant factor are equivalent:

Proposition 4.3. *Given $q \in [0,2]$, suppose that $\widetilde{F}^*L^{\#}LF$ is bounded on $\ell_q^{(\alpha_n)}(\mathcal{N})$. For $h \in \mathrm{range}(L)$, let $v = \widetilde{F}^*L^{\#}h$. If $\hat{\omega}$ minimizes (11) up to a constant factor, then $\hat{g} = L^{\#}LF\hat{\omega}$ minimizes (1) up to a constant factor.*

If $\widetilde{F}^*F$ is bounded on $\ell_q^{(\alpha_n)}(\mathcal{N})$, one may also choose $\hat{g} = F\hat{\omega}$. The reverse implication holds for $\hat{\omega} = \widetilde{F}^*L^{\#}L\hat{g}$ and $\hat{\omega} = \widetilde{F}^*\hat{g}$, respectively.

Given a parameter set Γ and two expressions $(a_\tau)_{\tau\in\Gamma}$ and $(b_\tau)_{\tau\in\Gamma}$ such that there is a constant $C > 0$ with $a_\tau \le Cb_\tau$ for all $\tau \in \Gamma$, we write $a_\tau \lesssim b_\tau$ in the following proof.

Proof of Proposition 4.3. Let $\hat{\omega}$ minimize (11) up to a constant factor, i.e., $\mathcal{I}_q(v,\hat{\omega}) \lesssim \mathcal{I}_q(v,\omega)$, for all $\omega \in \ell_2(\mathcal{N})$. Since $F\widetilde{F}^* = \mathrm{id}_{\mathcal{H}}$ and since $LL^{\#}L = L$ yields $LL^{\#}h = h$, we have $h = LFv$. Applying $LL^{\#}L = L$ implies $L\hat{g} = LF\hat{\omega}$, which leads to

$$\mathcal{J}_q(h,\hat{g}) = \|LFv - LF\hat{\omega}\|_{\mathcal{H}'}^2 + \|\widetilde{F}^*L^{\#}LF\hat{\omega}\|_{\ell_q^{(\alpha_n)}}^q.$$

Since $LF : \ell_2 \mapsto \mathcal{H}'$ is bounded and due to the boundedness of $\widetilde{F}^*L^{\#}LF$ on $\ell_q^{(\alpha_n)}$, this implies $\mathcal{J}_q(h,\hat{g}) \lesssim \mathcal{I}_q(v,\hat{\omega})$. Since $\hat{\omega}$ minimizes (11) up to a constant factor, we have $\mathcal{J}_q(h,\hat{g}) \lesssim \mathcal{I}_q(v,\widetilde{F}^*L^{\#}Lg)$, for all $g \in \mathcal{H}$. By applying that $\widetilde{F}^*L^{\#}$ is bounded and that $F\widetilde{F}^* = \mathrm{id}_{\mathcal{H}}$, we obtain, for all $g \in \mathcal{H}$,

$$\begin{aligned}\mathcal{J}_q(h,\hat{g}) &\lesssim \|\widetilde{F}^*L^{\#}h - \widetilde{F}^*L^{\#}Lg\|_{\ell_2}^2 + \|\widetilde{F}^*L^{\#}LF\widetilde{F}^*g\|_{\ell_q^{(\alpha_n)}}^q \\ &\lesssim \|h - Lg\|_{\ell_2}^2 + \|\widetilde{F}^*L^{\#}LF\widetilde{F}^*g\|_{\ell_q^{(\alpha_n)}}^q \lesssim \mathcal{J}_q(h,g),\end{aligned}$$

where we have used that $\widetilde{F}^*L^{\#}LF$ is bounded on $\ell_q^{(\alpha_n)}$.

Analogous arguments can be applied to the case $\hat{g} = F\hat{\omega}$, and the reverse implications follow in a similar way. □

Next, we obtain a solution of the discrete problem (11).

Proposition 4.4. *Let ϱ be a shrinkage rule with $\rho \in [\frac{1}{2},\infty]$. Then there is a constant $C > 0$ such that for all $q \in [\frac{1}{\rho},2]$, for all $v \in \ell_2(\mathcal{N})$, and for all $\omega \in \ell_2(\mathcal{N})$,*

$$\mathcal{I}_q(v,\hat{\omega}) \le C\mathcal{I}_q(v,\omega),$$

where $\hat{\omega} = \varrho(v_n, \alpha_n|v_n|^{q-1})_{n\in\mathcal{N}}$.

Remark 4.5. The exact minimizer of (11) for $q = 2$ is known to be $\big(\frac{1}{1+\alpha_n}v_n\big)_{n\in\mathcal{N}}$. However, $(x,\alpha) \mapsto \frac{1}{1+\alpha}x$ is not a shrinkage rule since (9) is violated. On the other hand, the rule $\varrho(x,\alpha) = \frac{1}{1+\frac{\alpha}{|x|}}x$ is a shrinkage rule with constant $\rho = 1$. The q-dependent expression $\varrho(x,\alpha|x|^{q-1})$ for $q = 2$ then yields the exact minimizer. In this sense the exact minimizer for $q = 2$ is still derived from shrinkage.

Proof of Proposition 4.4. First, we consider $\frac{1}{2} \leq \rho < \infty$. Due to (8), the sequence $\varrho(v_n, \alpha_n |v_n|^{q-1})_{n\in\mathcal{N}}$ is indeed contained in $\ell_2(\mathcal{N})$. Adapting results in [5] to our setting yields that the hard-shrinked sequence $\varrho_h(v_n, \alpha_n|v_n|^{q-1})_{n\in\mathcal{N}}$ minimizes (11) up to a constant factor. By using the short-hand notation

$$\begin{aligned} K_n &:= \left|v_n - \varrho_h(v_n, \alpha_n|v_n|^{q-1})\right|^2 + \alpha_n\left|\varrho_h(v_n, \alpha_n|v_n|^{q-1})\right|^q, \\ G_n &:= |v_n - \varrho(v_n, \alpha_n|v_n|^{q-1})|^2 + \alpha_n|\varrho(v_n, \alpha_n|v_n|^{q-1})|^q, \end{aligned}$$

we consider each n in the sequence norms separately. We aim to verify $G_n \lesssim K_n$ independently of n. For $v_n = 0$, we have $G_n = K_n$. Now, we suppose $v_n \neq 0$. Since (9) gets weaker as ρ and D decrease, we may assume that $q = \frac{1}{\rho}$ and $D \leq 1$. Case 1: For $|v_n| \leq D\alpha_n|v_n|^{q-1}$, (8) and (9) with $\rho = \frac{1}{q}$ yield

$$\begin{aligned} G_n &\leq C_1^2|v_n|^2 + \alpha_n C_2^q |v_n|^q \frac{|v_n|}{\alpha_n|v_n|^{q-1}} \\ &\leq C_1^2|v_n|^2 + C_2^q|v_n|^2 \lesssim |v_n|^2 = K_n. \end{aligned}$$

Case 2: For $|v_n| > D\alpha_n|v_n|^{q-1}$, we have $1/D > \alpha_n|v_n|^{q-2}$, and the estimate (8) yields

$$\begin{aligned} G_n &\leq C_1^2(\alpha_n|v_n|^{q-1})^2 + \alpha_n\big(|v_n| + C_1\min(|v_n|, \alpha_n|v_n|^{q-1})\big)^q \\ &\leq C_1^2\alpha_n|v_n|^q\alpha_n|v_n|^{q-2} + \alpha_n|v_n|^q(1 + C_1\alpha_n|v_n|^{q-2})^q \\ &\leq C_1^2\alpha|v_n|^q\tfrac{1}{D} + (1 + C_1/D)^q\alpha_n|v_n|^q \lesssim \alpha_n|v_n|^q \leq K_n/D. \end{aligned}$$

Hence, $G_n \lesssim K_n$ holds in both cases. Similar arguments verify the statement for $\rho = \infty$. □

Our main result follows from combining both propositions:

Proof of Theorem 4.1. According to Proposition 4.4, $\varrho(v_n, \alpha_n|v_n|^{q-1})_{n\in\mathcal{N}}$ is a minimizer of (11) up to a constant factor, where $v = \widetilde{F}^* L^{\#} f$. For $q = \frac{1}{\rho}$, Proposition 4.3 then implies Theorem 4.1 . If (3) holds, $\widetilde{F}^* L^{\#} L F$ and $\widetilde{F}^* F$ are bounded on $\ell_2^{(\alpha_n)}$. Interpolation between $\ell_{1/\rho}^{(\alpha_n)}$ and $\ell_2^{(\alpha_n)}$ yields uniform boundedness on $\ell_q^{(\alpha_n)}$, for $q \in [\frac{1}{\rho}, 2]$. □

5 Sparse Approximation

Given $h \in \mathcal{H}$ (possibly noisy) and a frame $\{f_n\}_{n\in\mathcal{N}}$ for $\mathcal{H}$, an important problem in sparse signal representation is to find the minimizer of

$$\min_{\omega\in\ell_2} \|\omega\|_{\ell_q} \text{ subject to } F\omega \approx h, \tag{12}$$

for $q \in [0, 1)$. Under additional requirements on $\{f_n\}_{n\in\mathcal{N}}$ and h, the solution for $q \in [0, 1)$ can be obtained from solving the much simpler convex problem with $q = 1$, cf. [4, 8]. However, these results are limited to finite $\mathcal{N}$, and the additional requirements are not satisfied in many situations.

The problem (12) is often replaced by a variational formulation, and one seeks to minimize

$$\mathcal{K}_q(h, \omega) = \|h - F\omega\|_{\mathcal{H}}^2 + \sum_{n\in\mathcal{N}} \alpha_n|\omega_n|^q$$

over $\omega \in \ell_2(\mathcal{N})$. For finite $\mathcal{N}$, ℓ_q-basis-pursuit as in [28], for instance, solves (12) by minimizing $\|F^\# h\|_{\ell_q}$ over all pseudo inverses $F^\#$. The associated variational formulation is

$$\min_{F^\#}\big(\min_{\omega\in\ell_2}\big(\|F^\# h-\omega\|_{\ell_2}+\sum_{n\in\mathcal{N}}\alpha_n|\omega_n|^q\big)\big).$$

We do not require $\mathcal{N}$ to be finite, and instead of minimizing over $F^\#$, we suppose to have a particular pseudo inverse $\widetilde{F}^*$ being the analysis operator of a dual frame $\{\tilde{f}_n\}_{n\in\mathcal{N}}$ such that $\widetilde{F}^*F$ is bounded on $\ell_q^{(\alpha_n)}(\mathcal{N})$:

Theorem 5.1. *Given a bi-frame $\{f_n\}_{n\in\mathcal{N}}$ and $\{\tilde{f}_n\}_{n\in\mathcal{N}}$, let ϱ be a shrinkage rule with $\rho\in[\frac{1}{2},\infty]$. Suppose that $\widetilde{F}^*F$ is bounded on $\ell_{1/\rho}^{(\alpha_n)}(\mathcal{N})$. Let $q=\frac{1}{\rho}$, then there is a constant $C>0$ such that for all $h\in\mathcal{H}$ and for all $\omega\in\ell_2(\mathcal{N})$*

$$\mathcal{K}_q(h,\hat{\omega})\le C\mathcal{K}_q(h,\omega),$$

where $\hat{\omega}=\widetilde{F}^*F\varrho(v_n,\alpha_n|v_n|^{q-1})_{n\in\mathcal{N}}$ with $v=\widetilde{F}^*h$ or $\hat{\omega}=\varrho(v_n,\alpha_n|v_n|^{q-1})_{n\in\mathcal{N}}$. If (3) holds, then the statement extends to all $q\in[\frac{1}{\rho},2]$, and C is independent of q.

Remark 5.2. For sufficiently smooth wavelet bi-frames with vanishing moments, the operator $\widetilde{F}^*F$ is bounded on $\ell_{1/\rho}^{(\alpha_n)}$ provided that $(\alpha_n)_{n\in\mathcal{N}}$ satisfies (3), cf. [10].

Proof. We replace $\mathcal{H}$, $\mathcal{H}'$, L, $L^\#$, and the bi-frame $\{f_n\}_{n\in\mathcal{N}}$, $\{\tilde{f}_n\}_{n\in\mathcal{N}}$ in (1) by $\ell_2(\mathcal{N})$, $\mathcal{H}$, F, $\widetilde{F}^*$, and the canonical basis $\{e_n\}_{n\in\mathcal{N}}$ for $\ell_2(\mathcal{N})$, respectively. The condition on $\widetilde{F}^*L^\#LF$ in Theorem 4.1 becomes '$\widetilde{F}^*F$ is bounded on $\ell_{1/\rho}^{(\alpha_n)}(\mathcal{N})$', and Theorem 4.1 implies Theorem 5.1. $\square$

6 Explicit Shrinkage Rules Between Hard- and Soft-Shrinkage

This section is dedicated to finding a family of shrinkage rules which is adapted to q in (11). For $q\in[0,1)$, we will use the constant $c_q=2^{q-2}\frac{(2-q)^{2-q}}{(1-q)^{1-q}}$. It is monotonically decreasing with $c_0=1$, and continuous extension yields $c_1=\frac{1}{2}$. Due to [1], the exact minimizer of (11) is sandwiched between soft- and hard-shrinkage, and we will verify that

$$\varrho_{h,s}^{(q)}(x,\alpha)=(x-\tfrac{x}{|x|}qc_q\alpha)\mathbf{1}_{\{|x|>\alpha c_q\}}$$

is well adapted to $q\in[0,1]$:

Theorem 6.1. *The sequence $\varrho_{h,s}^{(q)}(v_n,\alpha_n|v_n|^{q-1})_{n\in\mathcal{N}}$ is an exact minimizer of (11) at the endpoints $q=0$, $q=1$. It minimizes (11) up to a constant factor in between, and it coincides with the exact minimizer on $\{n\in\mathcal{N}:|v_n|<c_q^{\frac{1}{2-q}}\alpha_n^{\frac{1}{2-q}}\}$.*

Proof. Soft-shrinkage $\varrho_s(v_n,\frac{\alpha_n}{2})_{n\in\mathcal{N}}$ is the exact minimizer of (11), for $q=1$, cf. [5]. It equals $\varrho_{h,s}^{(1)}(v_n,\alpha_n)_{n\in\mathcal{N}}$. The exact minimizer for $q=0$ is hard-shrinkage $\varrho_h(v_n,\sqrt{\alpha_n})_{n\in\mathcal{N}}$, see [22], which is equal to $\varrho_{h,s}^{(0)}(v_n,\alpha_n|v_n|^{-1})_{n\in\mathcal{N}}$. The shrinkage rule $\varrho_{h,s}^{(q)}$ satisfies (9) for $\rho=\infty$. Hence due to Proposition 4.4, it minimizes (11) up to a constant factor.

We have $\varrho_{h,s}^{(q)}(v_n,\alpha_n|v_n|^{q-1})=0$ iff $|v_n|\le c_q\alpha_n|v_n|^{q-1}$. Since $|v_n|\le c_q\alpha_n|v_n|^{q-1}$ is equivalent to $|v_n|^{2-q}\le c_q\alpha_n$, it is also equivalent to $|v_n|\le c_q^{\frac{1}{2-q}}\alpha_n^{\frac{1}{2-q}}$, for $q\in(0,1)$. According to the results in [22], see also [1], each exact minimizer $(\hat{\omega}_n)_{n\in\mathcal{N}}$ satisfies $\hat{\omega}_n=0$ for $|v_n|<c_q^{\frac{1}{2-q}}\alpha_n^{\frac{1}{2-q}}$. $\square$

Due to Theorem 6.1, the rule $\varrho_{h,s}^{(q)}$ is an adaptation to $q \in [0,1]$. This might also be useful for parameter fitting: While $\alpha = (\alpha_n)_{n\in\mathcal{N}}$ can be fitted to f and L by considering (2), the new family $\varrho_{h,s}^{(q)}$ provides additional flexibility to optimize the choice of q as well. One optimizes $\alpha = \alpha(q)$ as in (2), one may then vary $q \in [0,1]$ and may optimize this sparsity parameter by analyzing the univariate curve $\alpha(q)$.

7 An Application to Variational Image Denoising With $q = 1$

To support minimization up to a constant factor as well as the general concept of shrinkage rules, we will apply results of the previous sections to wavelet-based image denoising.

7.1 Wavelet-based Variational Denoising

Following [5] and see also [23], we assume that the original signal $\bar{h}$ is contained in the Besov space $\dot{B}^s$, where $s = \frac{d}{2}$ and $\dot{B}^s = \dot{B}^s_1(L_1(\mathbb{R}^d))$, see [23] for a detailed introduction to Besov spaces and their application to image analysis. We suppose that the measured noisy signal h is still contained in $L_2(\mathbb{R}^d)$, but the noise pulls h out of $\dot{B}^s$. For fixed $\alpha > 0$, the minimizer g^α of

$$\min_{g\in\dot{B}^s} \left(\|h-g\|^2_{L_2} + \alpha\|g\|_{\dot{B}^s}\right) \tag{13}$$

approximates h in $L_2(\mathbb{R}^d)$ such that its norm in $\dot{B}^s$ is not too large. In other words, it constitutes a denoised signal, and the parameter α controls the emphasis of the penalty term $\|g^\alpha\|_{\dot{B}^s}$.

In order to choose a specific α such that g^α constitutes a well denoised signal, Montefusco and Papi proposed the so-called *H-curve criterion* in [24]: varying $\alpha > 0$ provides a curve

$$\left(\log\left(\|h-g^\alpha\|^2_{L_2}\right), \log\left(\|g^\alpha\|_{\dot{B}^s}\right)\right) \tag{14}$$

in $\mathbb{R}^2$, see also (2). Heuristically, the curve is concave on a reasonable range of α, cf. [24], and one can choose α_H according to the maximum absolute value of the curvature.

In order to solve the problem (13), we follow [5, 15] and we formulate it in terms of wavelet coefficients. For $\psi : \mathbb{R}^d \to \mathbb{C}$, let

$$\psi_{j,k}(x) := m^{\frac{j}{2}}\psi(M^j x - k), \quad \text{for } j \in \mathbb{Z}, k \in \mathbb{Z}^d,$$

where M is an integer matrix, whose eigenvalues are greater than one in modulus and $m := |\det(M)|$. We say a collection $\{\psi^{(1)}, \dots, \psi^{(n)}\}$ in $L_2(\mathbb{R}^d)$ *generates a wavelet frame* if $\{\psi^{(\mu)}_{j,k} : j \in \mathbb{Z}, k \in \mathbb{Z}^d,\ \mu = 1, \dots, n\}$ is a frame for $L_2(\mathbb{R}^d)$. Analoguously, we say two collections $\{\psi^{(1)}, \dots, \psi^{(n)}\}$ and $\{\widetilde{\psi}^{(1)}, \dots, \widetilde{\psi}^{(n)}\}$ generate a wavelet bi-frame if their dilates and shifts constitute a bi-frame for $L_2(\mathbb{R}^d)$.

For the remainder of the present paper, let $\{\psi^{(1)}, \dots, \psi^{(n)}\}$ and $\{\widetilde{\psi}^{(1)}, \dots, \widetilde{\psi}^{(n)}\}$ be compactly supported generators of a wavelet bi-frame with respect to an *isotropic* dilation matrix M, i.e., M can be diagonalized and all its eigenvalues have the same modulus. Moreover, the wavelets are supposed to be contained in the Sobolev space

$$W^k(L_\infty(\mathbb{R}^d)) = \{f \in L_\infty(\mathbb{R}^d) : \partial^\alpha f \in L_\infty(\mathbb{R}^d),\ \text{for all } |\alpha| \le k\},$$

where k is an integer strictly larger than $\frac{d}{2}$, and all dual wavelets have at least k *vanishing moments*, i.e.,

$$\int_{\mathbb{R}^d} x^\alpha \widetilde{\psi}^{(\mu)}(x)dx = 0, \quad \text{for all } |\alpha| < k,\ \mu = 1, \dots, n.$$

According to [10], the Besov norm $\|g\|_{\dot{B}^s}$ is then equivalent to $\|\widetilde{F}^* g\|_{\ell_1}$ where $\widetilde{F}^* : g \mapsto (\langle g, \widetilde{\psi}_\lambda \rangle)_{\lambda\in\Lambda}$ is the dual wavelet frame's analysis operator and where we have collected the indices μ, j, and k into one set Λ with $\widetilde{\psi}_\lambda = \widetilde{\psi}^{(\mu)}_{j,k}$, for $\lambda \in \Lambda$. Thus instead of minimizing (13), we consider

$$\min_{g\in L_2} \big(\|h-g\|^2_{L_2} + \alpha \sum_{\lambda\in\Lambda} |\langle g, \widetilde{\psi}^{(\mu)}_{j,k} \rangle| \big) \tag{15}$$

and due to the H-curve criterion, we choose α_H according to the maximal curvature of

$$\big(\log \big(\|\widetilde{F}^* h - w^\alpha\|^2_{\ell_2} \big), \log \big(\|w^\alpha\|_{\ell_1} \big) \big) \tag{16}$$

which is the discretization of the curve in (14). Note that Theorem 4.1 can now be applied to (15) with L being the identity, $\mathcal{N} = \Lambda$, and $\alpha = (\alpha_n)_{n\in\mathcal{N}}$ since $\widetilde{F}^* F$ is known to be continuous on $\ell_1(\Lambda)$ under the above mentioned assumptions on the wavelet bi-frame, cf. [10].

7.2 Numerical Results

While the H-curve method has already been successfully applied to soft-shrinkage in [24], this section is dedicated to verify that its combination with other shrinkage rules can lead to better results.

We consider the 8-bit grayscale image 'lena' of size 512×512, and we corrupt it by additive and multiplicative gaussian white noise as well as by salt&pepper noise with uniform spatial density of 15%. Since we know the original image, we evaluate the results of the different shrinkage rules by the mean square error (MSE) between denoised and original image. Soft-shrinkage has already been proven to provide good results, and it is the benchmark for other shrinkage rules.

We apply the strict k-shrinkage rule (10) for $k = 1, 2, 3$ and the nonnegative garotte-shrinkage among the thresholding rules. It should be mentioned that we only consider global shrinkage parameters. There are many bivariate wavelet bi-frames in literature which satisfy the assumptions of Section 7.1, see for instance [11, 13, 14], and we choose the bi-frame Laplace (2-2) from [12]. We also apply the Daubechies 3 and the Haar wavelet bases. It should be mentioned that the Haar wavelet does not satisfy the smoothness and vanishing moment conditions in Section 7.1, but since it is an orthonormal basis, $\widetilde{F}^* F$ is the identity and hence continuous on $\ell_1(\Lambda)$ such that Theorem 4.1 is still applicable to the minimization problem (15). Let us also mention that we maximize the curvature of (16) by an unsupervised golden section search and that the curve in (16) is concave on a reasonable range for all of the addressed shrinkage rules.

Table 1 shows the root of the MSE (RMSE) for the discretization with respect to the bi-frame Laplace (2-2). It turns out that nonnegative garotte-shrinkage performs better than soft-shrinkage for low noise such as additive with $\sigma = 10, 20$ and multiplicative noise with $\sigma = 0.1$. For stronger noise such as the addressed salt&pepper noise and multiplicative noise with $\sigma = 0.2$, the k-shrinkage rule (10) for $k = 1$ yields lower RMSE than soft-shrinkage.

We make analogous observations for the Haar and Daubechies 3 wavelets, see Table 2, while the differences with respect to low noise are smaller. We want to point out that the Laplace (2-2) bi-frame yields significantly lower RMSE than both orthogonal bases.

8 Conclusion

We have addressed variational problems with ℓ_q-constraints for $q \in (0, 1)$. In case that computation time is crucial as it is in any real-time and on-line application, there are no sufficiently fast algorithms to solve them. By considering minimization up to a constant factor, we have overcome this limitation.

shrinkage	additive noise $\sigma = 10$	additive noise $\sigma = 20$	additive noise $\sigma = 30$	multiplicative noise $\sigma = 0.1$	multiplicative noise $\sigma = 0.2$	salt&p 15%
soft	6.08	9.43	13.52	7.61	13.19	27.51
nn garotte	**5.56**	**9.38**	13.96	**7.34**	13.63	27.59
$\mathbf{k = 1}$	7.70	9.85	**11.44**	8.84	**11.53**	**21.33**
$k = 2$	7.34	9.06	12.28	8.66	12.08	24.39
$k = 3$	7.05	9.07	12.68	8.24	12.40	25.33

Table 1: Root mean square errors for Laplace (2-2)

shrinkage	additive noise $\sigma = 10$	additive noise $\sigma = 20$	additive noise $\sigma = 30$	multiplicative noise $\sigma = 0.1$	multiplicative noise $\sigma = 0.2$	salt&p 15%
soft	6.72	10.68	14.01	8.37	14.05	26.85
nn garotte	**6.45**	**10.61**	15.66	**8.30**	15.07	29.42
$\mathbf{k = 1}$	8.31	11.86	**13.71**	9.02	**13.71**	**26.05**
soft	5.98	10.09	14.45	7.72	13.85	27.77
nn garotte	**5.88**	10.48	15.44	**7.69**	14.56	29.59
$\mathbf{k = 1}$	7.21	**9.98**	**13.01**	8.01	**12.81**	**26.75**

Table 2: RMSE for Haar wavelets above, Daubechies 3 wavelets below

We avoid costly iterative schemes and derive closed formulas for such minimizers. This approach provides a tool which makes problems for $q < 1$ more feasible than until now. If exact solutions are required, those minimizers can initialize iterative schemes to speed up their convergence and to find an accurate local minimum. Moreover, the iterative shrinkage scheme for $q \in [1, 2]$ in [7] and for $q = 0$ in [3] could be modified by applying $\varrho^{(q)}_{h,s}(v_n, \alpha_n |v_n|^{q-1})$ to cover $q \in (0, 1)$ as well.

In Section 7, we have verified that the H-curve criterion can yield better results if we replace soft- by other shrinkage rules. These results support the general concept of shrinkage rules and minimization up to a constant factor.

It remains to find general conditions on L and on the bi-frame such that $\widetilde{F}^* L^{\#} L F$ is bounded on $\ell^{(\alpha_n)}_q$ and to compute the difference between $\varrho^{(q)}_{h,s}$ and the exact minimizer of (11). It also remains to precisely determine the arising constants and to evaluate the performance in numerical experiments for a variety of operators L.

References

[1] A. Antoniadis and J. Fan, *Regularization of wavelet approximations*, J. of the American Statistical Association **96** (2001), 939–967.

[2] R. Balan, J. Rosca, and S. Rickard, *Equivalence principle for optimization of sparse versus low-spread representations for signal estimation in noise*, Int. J. Imag. Syst. Tech. **15** (2005), no. 1, 10–17.

[3] T. Blumensath, M. Yaghoobi, and M. E. Davies, *Iterative hard thresholding and L0 regularization*, ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing **3** (2007), 877–880.

[4] E. Candès, J. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math. **59** (2006), no. 8, 1207–1223.

[5] A. Chambolle, R. A. DeVore, N. Y. Lee, and B. J. Lucier, *Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Trans. Image Process. **7** (1998), 319–335.

[6] R. Chartrand, *Exact reconstruction of sparse signals via nonconvex minimization*, IEEE Signal Processing Letters **14** (2007), no. 10, 707–710.

[7] I. Daubechies, M. Defrise, and C. DeMol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math. **57** (2004), 1413–1541.

[8] D. Donoho, M. Elad, and V. N. Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise*, IEEE Trans. Inform. Theory **52** (2006), no. 1, 6–18.

[9] D. Donoho and I. M. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika **81** (1994), 425–455.

[10] M. Ehler, *Nonlinear approximation associated with nonseparable wavelet bi-frames*, to appear in: J. Approx. Theory.

[11] ———, *On multivariate compactly supported bi-frames*, J. Fourier Anal. Appl. **13** (2007), no. 5, 511–532.

[12] ———, *The Construction of Nonseparable Wavelet Bi-Frames and Associated Approximation Schemes*, Logos Verlag, Berlin, 2007.

[13] ———, *Compactly supported multivariate pairs of dual wavelet frames obtained by convolution*, Int. J. Wavelets Multiresolut. Inf. Process. **6** (2008), no. 2, 183–208.

[14] M. Ehler and B. Han, *Wavelet bi-frames with few generators from multivariate refinable functions*, Appl. Comput. Harmon. Anal. **25** (2008), no. 3, 407–414.

[15] M. Ehler and K. Koch, *Variational image denoising with nonseparable wavelet bi-frames*, submitted, http://www.math.umd.edu/~ehlermar.

[16] M. Elad and M. Aharon, *Image denoising via sparse and redundant representations over learned dictionaries*, IEEE Trans. Image Process. **15** (2006), no. 12, 3736–3745.

[17] M. T. Figueiredo and R. D. Nowak, *Wavelet-based image estimation: an empirical Bayes approach using Jeffrey's noninformative prior*, IEEE Trans. Image Process. **10** (2001), no. 9, 1322–1331.

[18] M. Fornasier and H. Rauhut, *Recovery algorithms for vector valued data with joint sparsity constraints*, SIAM J. Numer. Anal. **46** (2008), no. 2, 577–613.

[19] H. Y. Gao, *Wavelet shrinkage denoising using the non-negative garotte*, J. Comput. Graph. Statist. **7** (1998), no. 4, 469–488.

[20] H. Y. Gao and Andrew G. Bruce, *Waveshrink with firm shrinkage*, Statistica Sinica **7** (1997), 855–874.

[21] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*, SIAM, Philadelphia, 1998.

[22] D. Lorenz, *Convergence rates and source conditions for Tikhonov regularization with sparsity constraints*, Journal of Inverse and Ill-Posed Problems **16** (2008), no. 5, 463–478.

[23] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*, American Mathematical Society, Boston, MA, USA, 2001.

[24] L. B. Montefusco and S. Papi, *A parameter selection method for wavelet shrinkage denoising*, BIT **43** (2003), 611–626.

[25] P. Mrázek, J. Weickert, and G. Steidl, *Correspondences between wavelet shrinkage and nonlinear diffusion*, Scale Space Methods in Computer Vision. Lecture Notes in Computer Science (M. Lillholm L. D. Griffin, ed.), 2003, pp. 101–116.

[26] A. Niedermeier, E. Romaneessen, and S. Lehner, *Detection of coastlines in SAR images using wavelet methods*, IEEE Trans. Geosci. Remote Sensing **38** (2000), no. 5, 2270–2281.

[27] T. Tao and B. Vidakovic, *Almost everywhere convergence of general wavelet shrinkage estimators*, Appl. Comput. Harmon. Anal. **9** (2000), 72–82.

[28] O. Yilmaz, R. Saab, R. Abugharbieh, and M. McKeown, *Underdetermined anechoic blind source separation via ℓ^q-basis-pursuit with $q < 1$*, IEEE Trans. Signal Process. **55** (2007), no. 8, 4004–4017.

[29] Xiao-Ping Zhang and Mita D. Desai, *Adaptive denoising based on SURE risk*, IEEE Signal Processing Letters **5** (1998), no. 10, 265–267.