Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

A Shearlet-Based Fast Thresholded Landweber Algorithm for Deconvolution

P. Grohs and U. Wiesmann and Z. Kereta

Research Report No. 2014-09
March 2014

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

A Shearlet-Based Fast Thresholded Landweber Algorithm for Deconvolution

Philipp Grohs, Željko Kereta, Uwe Wiesmann*

ETH Zürich

Abstract

Image deconvolution is an important problem which has seen plenty of progress in the last decades. Due to its ill-posedness, a common approach is to formulate the reconstruction as an optimisation problem, regularised by an additional sparsity-enforcing term. This term is often modeled as an ℓ_1 norm measured in the domain of a suitable signal transform. The resulting optimisation problem can be solved by an iterative approach via Landweber iterations with soft thresholding of the transform coefficients. Previous approaches focused on thresholding in the wavelet-domain. In particular, recent work [1] has shown that the use of Shannon wavelets results in particularly efficient reconstruction algorithms. The present paper extends this approach to Shannon shearlets, which we also introduce in this work. We show that for anisotropic blurring filters, such as the motion blur, the novel shearlet-based approach allows for further improvement in efficiency. In particular, we observe that for such kernels using shearlets instead of wavelets improves the quality of image restoration and SERG when compared after the same number of iterations.

*philipp.grohs@sam.math.ethz.ch, zeljko.kereta@math.ethz.ch, wuwe@student.ethz.ch

1 Introduction

This paper proposes a new shearlet-based algorithm for the image-deconvolution problem of reconstructing a given signal $\mathbf{x}_{\text{orig}}$ from measurements

$$\mathbf{y} = \mathsf{H}\mathbf{x}_{\text{orig}} + \mathbf{b},$$

where H denotes the blurring operator, defined via a blurring kernel h, i.e. $\mathsf{H}\mathbf{x} = h * \mathbf{x}$ (the symbol '$*$' denoting convolution), and $\mathbf{b}$ represents measurement noise.

Several algorithms exist for this problem, we refer to [2] and the references therein for an overview of the literature on this problem.

The focus of the present paper is on frame-based regularization methods that assume the existence of a dictionary T in which the sought signal $\mathbf{x}_{\text{orig}}$ possesses a sparse representation. A common choice for such a dictionary are wavelets, which optimally compress functions with pointlike singularities [2].

Adding this sparse representability as an additional constraint, in form of an ℓ_1 norm of the transform coefficients, allows us to regularize the original restoration problem. Thus, we arrive at the optimization problem of estimating $\mathbf{x}_{\text{orig}}$ by minimizing the functional

$$J(\mathbf{x}) = \|\mathsf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathsf{T}\mathbf{x}\|_1. \tag{1}$$

Several algorithms for the numerical solution of optimization problems of this sort exist. Among those algorithms we shall focus on iterative methods that are based on thresholded Landweber iterations, as proposed in [3], which amount to a forward-backward splitting scheme for J.

From a computational perspective the bottleneck of this class of algorithms is their typically slow convergence speed, which enforces the computation of many iterations in order to arrive at a reasonable restoration quality. More precisely, the convergence speed of the thresholded Landweber algorithm is determined by the spectral radius of H, which might be large.

In order to overcome this limitation the paper [1] has introduced a method that takes Shannon wavelets as the underlying sparsifying dictionary T. The specific frequency structure of Shannon wavelets allows the splitting of the original optimization problem into separate problems, with each problem localized in a distinct frequency band. Furthermore, it turns out that for reasonable blurring operators, i.e., those with a smooth blurring kernel h, the convergence speed of the thresholded Landweber method can be greatly increased in the high-frequency regime. This is due to the fact that a restriction of H to some high frequency band typically has a much smaller spectral radius than the full operator H. Consequently, the approach proposed in [1] allows for a dramatic speedup in the reconstruction of $\mathbf{x}_{\text{orig}}$ due to the fact that far fewer iterations are needed in comparison with the usual thresholded Landweber method as described in [3]. The resulting algorithm is called the Fast Thresholded Landweber (FTL) algorithm [1].

The purpose of this paper is to extend the method of [1] to the case of Shannon shearlets which we shall introduce below in Section 3. Shearlets are a recent addition to the arsenal of sparse transforms used in image processing. Compared to wavelets, shearlets show a superior ability to resolve anisotropic structures such as edges, which makes them a good choice for the sparsifying transform T. Furthermore, as we shall

see, the finer frequency resolution of the Shannon shearlet construction, as compared to Shannon wavelets, allows for a further speedup of the FTL algorithm of [1], in terms of the number of iterations needed for reaching a desired reconstruction quality. This is especially true for anisotropic blur kernels such as the motion blur.

We present a brief outline of our work. We begin in Section 2 by describing the general setup of this paper. Building on ideas of [1] we introduce the FTL algorithm for general operators H, and Parseval frames T in which H assumes a block-diagonal form. Compared with [1], we allow T to be an overcomplete dictionary, which requires some care. In particular, we will not minimize the energy functional itself, as defined in (1), but rather its synthesis-based version, as defined in (2). In Section 3 we will introduce shearlets, first in the continuous setting and then in the discrete setting, focusing on *Shannon shearlets*. Section 4 is devoted to experiments, which are aimed to give evidence that the shearlet-based approach provides better image restoration quality than wavelet-based approach. Table 1 and Figure 4 seem to give evidence that this indeed is the case as shearlets outperform wavelets (measured through the signal-to-error ratio gain) in the experiments we conducted.

2 Problem Formulation and Algorithm Description

2.1 Preliminaries

The goal of image restoration is the recovery of an original image $\mathbf{x}_{\text{orig}}$ from its corrupted observation $\mathbf{y}$. Images, that is, signals, in question are d-dimensional objects; given N_i samples along dimension i the dimensionality of the variables is $N = N_1 \times N_2 \times \cdots \times N_d$. The observations come with several limitations: we can only get a hold of finitely many samples and furthermore, the observations are polluted by noise and/or blur. In the following we shall assume that the standard image acquisition model holds, pertaining to the equation

$$\mathbf{y} = \mathsf{H}\mathbf{x}_{\text{orig}} + \mathbf{b},$$

where $\mathbf{x}_{\text{orig}}, \mathbf{y}, \mathbf{b} \in \mathbb{R}^N$ are vectors representing the original signal, measured signal and the measurement error, respectively, and $\mathsf{H} \in \mathbb{R}^{N \times N}$ is the observation operator that approximates blurring (or convolution). More specifically, H is taken to be a (block-) circulant matrix.

Problems of this sort, recovery of $\mathbf{x}_{\text{orig}}$ from $\mathbf{y}$, are called image restoration problems, or image deconvolution. Generally speaking, a perfectly accurate recovery of the original signal $\mathbf{x}_{\text{orig}}$ from the observation $\mathbf{y}$ is impossible, since the error $\mathbf{b}$ is unknown and H is not necessarily invertible, at the very least it is ill-conditioned. In fact, recovery of $\mathbf{x}_{\text{orig}}$ from $\mathbf{y}$ is an ill-posed problem. To escape this unfortunate predicament the problem of estimating $\mathbf{x}_{\text{orig}}$ is formulated as an optimization problem, and it is carried out through minimizing a cost functional. The optimization in question combines regularization of the given variable and penalization of its deviation from predictions, stemming from prior information about the signal.

In this paper we are concerned with signals that have a good representation with respect to some regularization transform $\mathsf{T} \in \mathbb{C}^{M \times N}$, $M \geq N$, which is, in the discrete case, a matrix whose rows form a Parseval frame, that is

$$\mathsf{T}^\top \mathsf{T} = \mathsf{I}_N.$$

A popular method of choice for the estimation of the original signal $\mathbf{x}_{\text{orig}}$ is to define the estimate $\mathbf{x}_*$ as

$$\mathbf{x}_* = \mathsf{T}^\top \mathbf{d}_*,$$

with $\mathbf{d}_* \in \mathbb{C}^M$ being a minimiser of the functional

$$J(\mathbf{d}) := J_{data}(\mathbf{d}) + \lambda J_{reg}(\mathbf{d}), \tag{2}$$

where

$$J_{data}(\mathbf{d}) := \|\mathsf{H}\mathsf{T}^\top \mathbf{d} - \mathbf{y}\|_2^2,$$

and

$$J_{reg}(\mathbf{d}) := \|\mathbf{d}\|_1.$$

The choice of 1-norm in the regularization term leads to a nonlinear deconvolution algorithm. The rationale behind choosing 1-norm over the more standard 2-norm is that the 1-norm promotes sparse vectors, i.e., we are looking for solutions which possess a sparse representation in the transform domain induced by T.

The optimization strategy (2) is well-known under the name *synthesis-based ℓ_1-norm minimization* [4].

2.2 Forward-Backward Splitting of $J(\mathbf{d})$

In the general case, we cannot minimize $J(\mathbf{d})$ in a straightforward fashion because the convolution operator H makes the coefficients interdependent. Therefore, unless we are in a special case, we have to resort to other, bit more circuitous, methods of minimizing $J(\mathbf{d})$. *Bound optimization algorithms*, initially derived in [3], tackle this issue by minimizing a sequence of easy-to-minimize functionals $\tilde{J}_n(\mathbf{d})$. Given $\mathbf{d}^{(n)}$ these functionals satisfy the following properties

1. If $\mathbf{d} = \mathbf{d}^{(n)}$ then $\tilde{J}_n(\mathbf{d}) = J(\mathbf{d})$.
2. Otherwise, $\tilde{J}_n(\mathbf{d}) \geq J(\mathbf{d})$.

The minimiser $\mathbf{d}^{(n+1)}$ is then defined as

$$\mathbf{d}^{(n+1)} = \arg\min \tilde{J}_n(\mathbf{d}).$$

The convergence criteria for this class of algorithms have been readily derived in [3].

Following [1] we discuss a specific choice of surrogate functionals $\tilde{J}_n(\mathbf{d})$. First we note that, due to the fact that $\mathsf{T}^\top\mathsf{T} = \mathsf{I}$, we can equivalently write

$$J_{data}(\mathbf{d}) := \|\tilde{\mathsf{H}}\mathbf{d} - \tilde{\mathbf{y}}\|_2^2$$

with

$$\tilde{\mathsf{H}} := \mathsf{T}\mathsf{H}\mathsf{T}^\top \in \mathbb{C}^{M\times M} \quad \text{and} \quad \tilde{\mathbf{y}} := \mathsf{T}\mathbf{y} \in \mathbb{C}^M.$$

We can now define the surrogate functionals

$$\tilde{J}_n(\mathbf{d}) := \alpha\|\mathbf{d}^{(n)} - \mathbf{d}\|_2^2 + J(\mathbf{d}) - \|\tilde{\mathsf{H}}\mathbf{d}^{(n)} - \tilde{\mathsf{H}}\mathbf{d}\|_2^2.$$

It follows that whenever

$$\alpha > \rho(\tilde{\mathsf{H}}^\top\tilde{\mathsf{H}})$$

holds, where $\rho(\tilde{\mathsf{H}}^\top\tilde{\mathsf{H}})$ is the spectral radius of $\tilde{\mathsf{H}}^\top\tilde{\mathsf{H}}$, the surrogate functionals $\tilde{J}_n(\mathbf{d})$ satisfy conditions 1. and 2. from above, see [1]. In addition, we have the representation

$$\tilde{J}_n(\mathbf{d}) = \alpha\|\mathbf{d}^{(n)} + \alpha^{-1}\tilde{\mathsf{H}}^\top(\tilde{\mathbf{y}} - \tilde{\mathsf{H}}\mathbf{d}^{(n)}) - \mathbf{d}\|_2^2 + \lambda\|\mathbf{d}\|_1 + c \tag{3}$$

for a constant $c \in \mathbb{R}$. The minimization of (3) can be computed explicitly using the soft-thresholding function

$$\mathcal{T}_\nu(x) = \frac{x}{|x|}\left(|x| - \nu\right)_+, \tag{4}$$

where $t_+ = \max(0, t)$. The minimiser of the functional (3) is then given by

$$\mathbf{d}^{(n+1)} := \mathcal{T}_{\lambda\tau/2}\left(\mathbf{d}^{(n)} + \alpha^{-1}\tilde{\mathsf{H}}^\top(\mathbf{d} - \tilde{\mathsf{H}}\mathbf{d}^{(n)})\right),$$

where $\mathcal{T}_{\lambda\tau/2}(\cdot)$ is the component-wise modification of (4).

We thus arrive at Algorithm 1 for the iterative solution of $J(\mathbf{d})$ as given in (2).

Algorithm 1: Thresholded Landweber Iteration $\hat{\mathbf{d}}_\varepsilon = \mathbf{TL}\left(\varepsilon, \alpha, \tilde{H}, \tilde{y}\right)$

Data: Tolerance $\varepsilon > 0$, $\alpha > \rho(\tilde{\mathsf{H}}^\top \tilde{\mathsf{H}})$, matrix $\tilde{\mathsf{H}}$, right hand side $\tilde{\mathbf{y}}$
Result: Approximate minimiser $\mathbf{d}_{*,\varepsilon}$ of (2)
$\tau := \alpha^{-1}$;
$n := 0$;
$\mathbf{d}^{(n)} := 0$;
while $\|\boldsymbol{d}^{(n)} - \boldsymbol{d}^{(n-1)}\|_2 > \varepsilon$ **do**
  $\mathbf{d}^{(n+1/2)} := \mathbf{d}^{(n)} + \tau \tilde{\mathsf{H}}^\top (\mathbf{d} - \tilde{\mathsf{H}} \mathbf{d}^{(n)})$;
  $\mathbf{d}^{(n+1)} := \mathcal{T}_{\lambda\tau/2}(\mathbf{d}^{(n+1/2)})$;
  $n := n + 1$;
$\mathbf{d}_{*,\varepsilon} := \mathbf{d}^{(n)}$;

It is important to emphasize the fact that the convergence speed of Algorithm 1 crucially depends on the size τ of the timesteps: the larger the τ the fewer iterations are needed. In particular, it is desirable to have a representation system which makes the quantity $\rho := \rho(\tilde{\mathsf{H}}^\top \tilde{\mathsf{H}})$ as small as possible.

2.3 Block-Diagonal Operators

It has been demonstrated recently in [1] that it is possible to improve the performance of Algorithm 1 by considering Shannon wavelets. We would like to further expand that approach.

We shall assume that the transform T is built in such way that the matrix $\tilde{\mathsf{H}}$ has a block-diagonal structure. Denoting

$$[M] := \{1, \dots, M\}$$

we assume that there exists a disjoint partitioning

$$[M] = \bigcup_{l \in L} I_l$$

such that with

$$V_l := \left\{ \mathbf{d} \in \mathbb{C}^M : \, d_i = 0 \text{ for all } i \in [M] \setminus I_l \right\}$$

we have

$$\tilde{\mathsf{H}}(V_l) \subset V_l.$$

It follows that $\tilde{\mathsf{H}}$ induces mappings

$$\tilde{\mathsf{H}}_l : V_l \to V_l,$$

hence, the minimization problem (2) can be recast as

$$\mathbf{d}_* := \sum_{l \in L} \mathbf{d}_{*,l},$$

where $\mathbf{d}_{*,l}$ is the minimiser of the functional

$$J_l(\mathbf{d}) := J_{data,l}(\mathbf{d}) + \lambda J_{reg,l}(\mathbf{d}) : \ V_l \to \mathbb{R}, \tag{5}$$

where

$$J_{data,l}(\mathbf{d}) := \|\tilde{\mathsf{H}}_l \mathbf{d} - \tilde{\mathbf{y}}_l\|_2^2,$$

and

$$J_{reg,l}(\mathbf{d}) := \|\mathbf{d}\|_1$$

are both defined on V_l, while $\tilde{\mathbf{y}}_l$ denotes the orthogonal projection of $\tilde{\mathbf{y}}$ onto V_l. Therefore, we can apply Algorithm 1 to subproblems (5), for all $l \in L$.

Herein lies the crucial difference from the standard approach [3]: the stepsizes can be chosen independently for each subspace V_l. Analogously to previous considerations, constants α_l ought to satisfy

$$\alpha_l > \rho_l := \rho(\tilde{\mathsf{H}}_l^\top \tilde{\mathsf{H}}_l) = \max_{\|\mathbf{d}\|_2=1, \mathbf{d}\in V_l} \|\tilde{\mathsf{H}}\mathbf{d}\|_2^2.$$

Therefore, each term in (5) can be minimized independently, with parameters depending only on the governing subspace. Further elaboration regarding the details and consequences of this approach can be found in [3, 1].

In order to estimate the quantities ρ_l we employ the following lemma.

Lemma 2.1. *With $W_l := \mathsf{T}^\top V_l \subset \mathbb{C}^N$ we have*

$$\rho_l = \rho\left((\mathsf{H}|_{W_l})^\top \mathsf{H}|_{W_l}\right) = \max_{\|\boldsymbol{d}\|_2=1, \boldsymbol{d}\in W_l} \|\mathsf{H}\boldsymbol{d}\|_2^2.$$

Proof. By the definition of $\tilde{\mathsf{H}}$ and the fact that $\mathsf{T}^\top\mathsf{T} = I_N$ we have

$$\rho_l = \max_{\|\mathbf{d}\|_2=1, \mathbf{d}\in V_l} \|\mathsf{T}\mathsf{H}\mathsf{T}^\top \mathbf{d}\|_2^2 = \max_{\|\mathbf{d}\|_2=1, \mathbf{d}\in V_l} \|\mathsf{H}\mathsf{T}^\top \mathbf{d}\|_2^2$$

□

Solving each subproblem (5) separate, rather than solving the full problem (2), comes with two advantages. Firstly, splitting a big problem into subproblems lends itself to parallelisation in a trivial way. Secondly, in several cases we have the property that for most $l \in L$ the quantity ρ_l is significantly smaller than $\rho := \rho(\mathsf{H}^T\mathsf{H})$. Hence, for subproblems (5) the parameter α_l can be chosen to be much smaller than for the full problem, which results in bigger stepsizes and thus, in faster convergence.

We end up with the following algorithm.

Algorithm 2: Fast Thresholded Landweber Iteration $\mathbf{d}_{*,\varepsilon} = \mathbf{FTL}\left(\varepsilon, \alpha, \tilde{H}, \tilde{y}\right)$

Data: Tolerances $\varepsilon = (\varepsilon_l > 0)_{l\in L}$, $\alpha = (\alpha_l > \rho_l)_{l\in L}$, matrix $\tilde{\mathsf{H}}$, right hand side $\tilde{\mathbf{y}}$
Result: Approximate minimiser $\mathbf{d}_{*,\varepsilon}$ of (2)
for $l \in L$ **do**
  $\mathbf{d}_{*,\varepsilon,l} := \mathbf{TL}(\varepsilon_l, \alpha_l, \tilde{H}_l, \tilde{y}_l)$;
$\mathbf{d}_{*,\varepsilon} := \sum_{l\in L} \mathbf{d}_{*,\varepsilon,l}$;

2.4 Shannon Systems

In the remainder of this paper we will restrict our attention to blurring operators H which are defined via a convolution with a blurring kernel, i.e., operators given by

$$\mathsf{H}\mathbf{x} = h * \mathbf{x} = (\hat{h} \cdot \hat{\mathbf{x}})^{\vee},$$

where the symbol $\hat{\mathbf{x}}$ denotes the discrete Fourier transform and $(\hat{\mathbf{x}})^{\vee} = \mathbf{x}$. In the paper [1] the Fast Landweber Iteration, as in Algorithm 2, has been introduced and studied for T being a Shannon wavelet transform that naturally splits the operator H into different frequency subbands. Furthermore, the authors showed in in [1] that this approach greatly enhances the performance of the resulting deblurring algorithm due to the fact that large stepsizes can be chosen for high frequency bands provided the kernel h is sufficiently smooth.

For convolution operators H we describe a general class of representation systems which we call *Shannon systems*. We assume that our signals are sampled on a d-dimensional cube

$$\Omega := \prod_{i=1}^{d} [N_i]$$

with $N = N_1 \times N_2 \times \cdots \times N_d$.

Definition 2.1. We call T a *Shannon system* if there exists a disjoint partitioning $\Omega = \bigcup_{i \in L} \Omega_i$ and a disjoint partitioning $[M] = \bigcup_{i \in L} I_l$ such that the discrete Fourier transform of $\mathbf{t}_j$ is supported in Ω_l for all $j \in I_l$, where $\mathbf{t}_j$ denotes the j-th row of T.

It is easy to see that convolution operators H admits a block-diagonal structure in a Shannon system.

Lemma 2.2. *Let H be a blurring operator with kernel h and T a Shannon system. Then, with respect to the associated partitioning $[M] = \bigcup_{i \in L} I_l$, operator H possesses a block-diagonal form. Moreover, we have*

$$\rho_l \leq \sup_{k \in \Omega_l} |\hat{h}(k)|$$

for all $l \in L$.

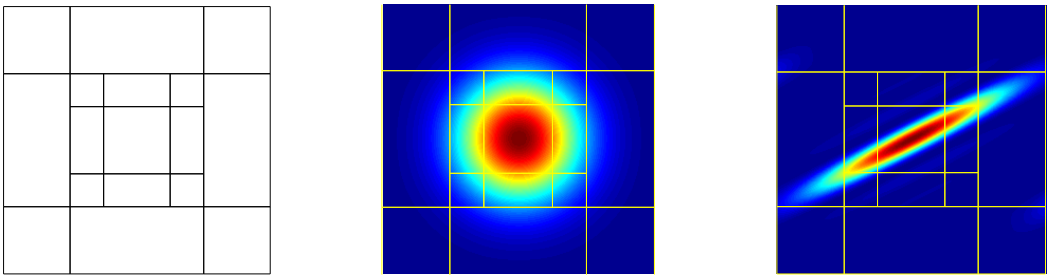
Proof. The proof follows from the fact that in the Fourier domain the blurring operator becomes a multiplication by $\hat{h}$ which implies

$$\widehat{\mathsf{H}\mathbf{x}}(k) = \hat{h}(k)\hat{\mathbf{x}}(k), \quad k \in \Omega.$$

Clearly, the Fourier transform of $\mathbf{t}_j$ is supported in Ω_l for all indices $j \in I_l$, i.e., we have $\operatorname{supp} \hat{\mathbf{x}} \subset \Omega_l$ for all $\mathbf{x} \in W_l$. Hence, by Plancherel's theorem, for $\mathbf{x} \in W_l$ we have

$$\|\mathsf{H}\mathbf{x}\|_2 = \|\widehat{h\mathbf{x}}\|_2 \leq \sup_{k \in \Omega_l} |\hat{h}(k)| \|\hat{x}\|_2 = \sup_{k \in \Omega_l} |\hat{h}(k)| \|x\|_2.$$

□



(a) *Frequency support*

(b) *Frequency support and the FFT of a Gaussian filter*

(c) *Frequency support and the FFT of a motion filter*

Figure 1: *Shannon wavelets*

A well-known example of a Shannon system are Shannon wavelets as described e.g. in [1]. The partitioning of Ω is illustrated in Figure 1a. In [1] it is observed that for smooth filters the quantities ρ_l associated to high frequencies become very small (see Figure 1c) which allows for a dramatic speedup in efficiency when using the FTL algorithm. However, for anisotropic filters h, such as motion blurs, the partitioning induced by the Shannon wavelet construction need not be optimal, see Figure 1c.

In the following sections we will extend this approach to what we call *Shannon shearlets*. The advantage of this approach lies in the fact that the finer frequency partitioning associated with the shearlet transform allows for better choices of parameters α_l in several cases. Furthermore, it is well-known that shearlets are superior to wavelets in terms of resolution of anisotropic phenomena such as edges.

3 Shearlets

After having described the setup and adapted the FTL to more general transforms, our next goal is to describe the shearlet transform that we are going to use to minimize the cost function $J(\mathbf{x})$. Shearlets are a part of fairly recent development in the multiscale analysis family of transforms, along with curvelets, ridgelets, and others [5, 6] . They provide optimally sparse representations of several classes of multivariate data exhibiting anisotropic behavior. Shearlets are a family of analyzing functions that, as opposed to wavelets, are defined not only with respect to scale and location parameters but take directionality into account as well. These changes allow us to consider and gain access to many notions beyond the reach of wavelets [7, 5, 8, 9].

An important trait distinguishing these families of transforms from one another is how they handle directionality. Shearlets adhere to *shearing*, a geometrical transformation that leaves all entries of a vector, except for one, unchanged. This remaining entry is displaced with respect to another, predetermined, entry according to the value of a given shearing parameter.

The crucial advantage of shearing is that it does not destroy the rectangular grid, hence, it allows a unified treatment of the continuous and the discrete settings. An interested reader is referred to a number of resources about this topic, e.g. [10].

3.1 Continuous Shearlets

Shearlets are an affine system, generated by a single mother shearlet and parametrized through scaling, translation and shearing. Directionality of singularities is handled through the aforementioned shearing, given by a shearing matrix $S_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$ for a real valued s. For $\psi \in L_2(\mathbb{R}^2)$ we define the analyzing function

$$\psi_{a,s,\mathbf{b}}(\mathbf{x}) = a^{-3/4}\psi\left(D_{a^{-1}}S_s^{-1}(\mathbf{x}-\mathbf{b})\right) = a^{-3/4}\psi\left(\frac{x_1 - b_1 - s(x_2-b_2)}{a}, \frac{x_2-b_2}{a^{1/2}}\right),$$

where $D_a = \mathrm{diag}(a, \sqrt{a})$ is the dilation matrix, $a > 0$ the dilation parameter and $\mathbf{b} \in \mathbb{R}^2$ denotes the location.

The *continuous shearlet transform* of a function $f \in L_2(\mathbb{R}^2)$ is then defined through

$$\mathcal{SH}(f)(a, s, \mathbf{b}) = \langle f, \psi_{a,s,\mathbf{b}} \rangle = \langle \hat{f}, \hat{\psi}_{a,s,\mathbf{b}} \rangle$$

where $(a, s, \mathbf{b}) \in \mathbb{S} = \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^2$. Additional assumptions are needed to ensure that continuous shearlet transform is an isometry. For example, ψ should posses $M \geq 1$ vanishing moments in the x_1 direction, by which we mean that ψ should satisfy

$$\int_{\mathbb{R}^2} \frac{|\hat{\psi}(\boldsymbol{\xi})|^2}{|\xi_1|^{2M}} d\boldsymbol{\xi} < \infty.$$

There is another issue that merits our attention. Continuous shearlets exhibit a certain directional bias, in the sense that if a function has a singularity along the x_1 axis then this singularity can in the shearlet domain only be detected as the shearing parameter s tends to infinity. It should not come as a surprise that behavior of this type can give rise

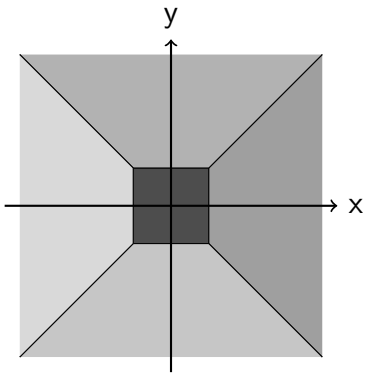


Figure 2: *Partioning of the (frequency) plane into a low-frequency box and four higher frequency cones*

to serious limitations when it comes to practical applications. Hence, it is an issue that needs to be addressed. A standard way to work around this issue is to split the Fourier domain into four high-frequency cones and one low-frequency square centred around the origin, as in Figure 2. It can be seen that by doing things this way we keep the shearing parameter in a finite set in each of the cones and that the corresponding shearlet system does not exhibit a directional bias. Shearlet systems of this type are eponymously called *Cone-Adapted*.

3.2 Discrete Shearlets

Discrete shearlet systems are defined by sampling continuous shearlets on a discrete subset of $\mathbb{S}$ [10]. For these discretization we need to discretize all of the involved parameters. This means choosing maximum number of scales $j_{\max}$, maximum number $k_{\max}(j)$ of shears on each given scale $0 \leq j < j_{\max}$, scaling parameters $a_j > 0$, shearing parameters $s_{j,k}$, and the grid scaling parameter γ. To avoid the consequences of shearlets' directional bias we will be using the cone-adapted shearlets which adds an additional parameter $\kappa \in \{1, 2\}$ indicating the appropriate cone. One of the underlying ideas is that piecing together Parseval frames of each respective cone yields a Parseval frame for the whole space.

For the sampling of the frequency domain we take

$$\Omega := \left\{ \left(\gamma \frac{m_1}{N_1}, \gamma \frac{m_2}{N_1} \right) \in \mathbb{R}^2 \;\middle|\; -\frac{N_1}{2} \leq m_1, m_2 < \frac{N_1}{2},\ m_1, m_2 \in \mathbb{Z} \right\}.$$

while the horizontal cone C_1 and the vertical cone C_2 are defined by

$$\begin{aligned} C_1 &:= \{\omega \mid \omega \in \Omega \setminus \{0\}, |\omega_2| \leq |\omega_1|\}, \\ C_2 &:= \{\omega \mid \omega \in \Omega \setminus \{0\}, |\omega_2| > |\omega_1|\}. \end{aligned}$$

Before we consider cone adapted shearlets, we will define the untranslated discretized shearlets $\psi_{j,k}$ in the frequency domain

$$\hat{\psi}_{j,k}(\omega) := \hat{\psi}_1(a_j \omega_1)\, \hat{\psi}_2\left(a_j^{-1/2} \left(\frac{\omega_2}{\omega_1} + s_{j,k} \right) \right). \tag{6}$$

The untranslated, cone-adapted, discretized shearlets $\psi_{j,k,\kappa}$ can now be defined in the frequency domain through

$$\hat{\psi}_{j,k,\kappa}(\omega) := \begin{cases} \hat{\psi}_{j,k}(\omega_1,\omega_2), & \text{if } \kappa = 1,\ |k| \neq k_{\max}(j),\ \omega \neq 0, \\ \hat{\psi}_{j,k}(\omega_2,\omega_1), & \text{if } \kappa = 2,\ |k| \neq k_{\max}(j),\ \omega \neq 0, \\ \hat{\psi}_{j,k}(\omega_1,\omega_2)\,\chi_{C_1}(\omega) + \hat{\psi}_{j,k}(\omega_2,\omega_1)\,\chi_{C_2}(\omega), & \text{if } |k| = k_{\max}(j),\ \omega \neq 0, \\ 0, & \text{if } \omega = 0. \end{cases}$$

for $\omega \in \Omega$, $0 \leq j < j_{\max}$ and $-k_{\max}(j) \leq k \leq k_{\max}(j)$. Furthermore, we define

$$k_{\max}(j,\kappa) := \begin{cases} k_{\max}(j), & \text{if } \kappa = 1, \\ k_{\max}(j) - 1, & \text{if } \kappa = 2 \end{cases}$$

and the set I

$$I := \{(a_j, s_{j,k}, \kappa) \mid 0 \leq j < j_{\max},\ -k_{\max}(j,\kappa) \leq k \leq k_{\max}(j,\kappa)\} \cup \{(0,0,0)\}.$$

The element $(0,0,0) \in I$ refers to the *scaling function* $\phi = \psi_{0,0,0}$.

What is left is to define the shearlet transformation operator $\mathsf{T}_{\mathcal{S}}$, which should leave us squarely in the framework set out in the previous section. First, we define diagonal matrices $\mathbf{\Psi}_i$ as

$$\mathbf{\Psi}_i := \operatorname{diag}\left((\hat{\psi}_i(\omega))_{\omega\in\Omega}\right) \in \mathbb{R}^{N\times N}.$$

The matrix $\mathbf{\Psi}$ is then defined through concatenation of matrices $\mathbf{\Psi}_i$, that is

$$\mathbf{\Psi} := (\mathbf{\Psi}_i)_{i\in I} \in \mathbb{R}^{M\times N}.$$

The discrete shearlet transform matrix $\mathsf{T}_{\mathcal{S}}$ can hence be written as

$$\mathsf{T}_{\mathcal{S}} := \mathbf{F}_{|I|}^{\top}\mathbf{\Psi}\mathbf{F}.$$

where $\mathbf{F}_{|I|} := \operatorname{diag}(\underbrace{\mathbf{F},\ldots,\mathbf{F}}_{|I| \text{ times}})$.

A salient detail missing thus far is with regards to the very functions ψ_1, ψ_2 and ϕ which are the building blocks of our shearlet system. Different shearlet discretizations offer various choices for these functions. The following theorem states the conditions functions ψ_1, ψ_2 and ϕ have to satisfy so that $\mathsf{T}_{\mathcal{S}}$ indeed forms a Parseval frame.

Theorem 3.1. *Assume*

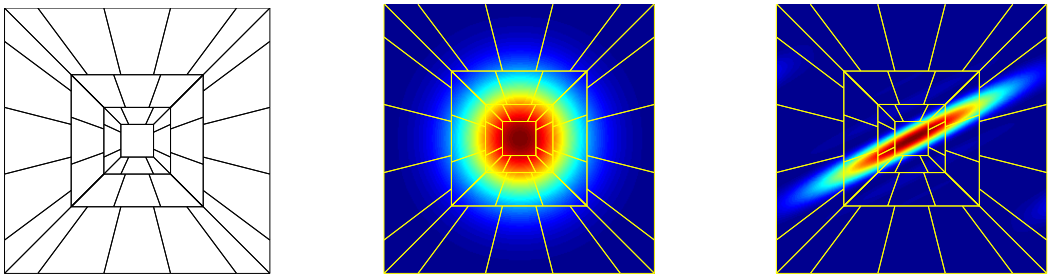
$$\left(\sum_{j=0}^{j_{max}-1} |\hat{\psi}_1(a_j\omega_\kappa)|^2\right) + |\phi(\omega)|^2 = 1 \quad \forall\ \kappa \in \{1,2\},\ \omega \in C_\kappa \tag{7}$$

for an appropriate choice of $\phi : \Omega \to \mathbb{R}$ with $\phi(0) = 1$ and also assume

$$\sum_{k=-k_{max}(j)}^{k_{max}(j)} \left|\hat{\psi}_2\left(a_j^{-1/2}\left(\frac{\omega_2}{\omega_1} + s_{j,k}\right)\right)\right|^2 = 1 \quad \forall\ \omega \in C_1,\ 0 \leq j < j_{max}. \tag{8}$$

Then

$$\mathsf{T}_{\mathcal{S}}^{\top}\mathsf{T}_{\mathcal{S}} = I_N.$$



(a) *Frequency support*

(b) *Frequency support and the FFT of a Gaussian filter*

(c) *Frequency support and the FFT of a motion filter*

Figure 3: *Shannon shearlets*

3.3 Discrete Shannon Shearlets

The shearlet system we are going to use in the FTL algorithm is what we will refer to as the *Shannon shearlet transform*. The Shannon shearlets have disjoint supports in the frequency domain and are generated by

$$\begin{aligned}\hat{\psi}_1(\omega) &= \chi_{[-1,-\frac{1}{2})\cup(\frac{1}{2},1]}(\omega),\\ \hat{\psi}_2(\omega) &= \chi_{[-\frac{1}{2},\frac{1}{2})}(\omega).\end{aligned}$$

The corresponding discretization parameters are given by

$$\begin{aligned}j_{\max} &\leq \lfloor \log_2 N_1 \rfloor\\ k_{\max}(j) &= \lceil 2^{\frac{j}{2}} \rceil\\ a_j &= 2^{-j}\\ s_{j,k} &= k2^{-\frac{j}{2}}\\ \gamma &= 2^{j_{\max}-1}.\end{aligned}$$

It can be shown that for an appropriate choice of ϕ, concerned with low-frequency cases, Shannon shearlet transform satisfies Parseval's property. Furthermore, shearlets in such a family have disjoint frequency supports. Therefore, the corresponding shearlet transform can be decomposed on a per-shearlet basis, i.e., as

$$\mathsf{T}_{\mathcal{S}} = \sum_{i\in I} \mathsf{T}_{\mathcal{S}}^{(i)}, \quad \mathsf{T}_{\mathcal{S}}^{(i)} := (0, 0, \ldots, \mathbf{F}\boldsymbol{\Psi}_i^{\top}\mathbf{F}^{\top}, \ldots, 0)^{\top} \in \mathbb{C}^{M\times N}.$$

Therefore, it fits perfectly into the FTL framework. The partitioning of Ω induced by Shannon shearlets is shown in Figure 3. One can see that for anisotropic filters, such as motion blurs, this partitioning is a better fit than Shannon wavelets, see Figure 3c.

4 Experiments

What is important for implementations of the FTL algorithm is how do we choose the stepsizes. In general, a stepsize is chosen as the maximum of the squared norm of frequency components, within a certain region, of the filter that has been used.

In the case of Shannon shearlets the stepsizes are chosen on a per-shearlet basis, that is, for each shearlet the region over which we choose its stepsize is the support of that given shearlet. For purposes of comparison we also implemented the FTL algorithm for undecimated Shannon wavelets. The algorithms work by performing the Landweber step in the DFT basis and computing the regularisation transform used therein.

Since H is diagonalised in the DFT basis the transforms can be easily performed using that basis. This is computationally more efficient than going back to the image space during the iterations. Furthermore, instead of using the same regularisation parameter λ for all coefficients during the soft thresholding step, the algorithms are extended to accept regularisation parameters λ_j, corresponding to each scale. This is a straightforward extension of the FTL algorithm that aims to improve algorithm's performance by offering more freedom in its treatment of different scales. The regularisation parameter for the scaling coefficients of shearlet (wavelet) transforms is independent of other regularisation parameters and is usually set to zero, which means that by default no soft thresholding takes place for these coefficients.

The discrete shearlet transforms themselves work very similar to the theory we have presented in Section 3.3. The matrix $\mathbf{\Psi}$ is far to large to work with in practice. Hence, it is represented through its non-zero components and their indices, which are computed in a preprocessing step. Since the standard FFT formulation assumes that high frequency components of an image in the frequency domain are located at the center, the indices are adjusted appropriately. Decomposition and reconstruction can then easily be achieved by computing $\mathbf{d}_* = \mathbf{F}_{i_{\max}}^\top \mathbf{\Psi} \mathbf{F} \mathbf{x}_*$ and $\mathbf{x}_* = \mathbf{F}^\top \mathbf{\Psi}^\top \mathbf{F}_{i_{\max}} \mathbf{d}_*$, with aforementioned reduced representation of $\mathbf{\Psi}$ and the help of the FFT and the inverse FFT in place of $\mathbf{F}$ and $\mathbf{F}^\top$. Note that this requires as many full-sized FFTs as there are shearlets. Together with large number of coefficients in the shearlet decomposition the soft thresholding operation is applied to, this makes the FTL algorithm for Shannon shearlets significantly slower than the FTL algorithm for Shannon wavelets. The number of shearlets is controlled with the maximum number of scales $j_{\max}$ and with an additional parameter $j_0 \in \mathbb{N}_0$, the *scale offset*, which was not included in the presented theory. The purpose of j_0 is to shift the range of scales j used during the discretisation, that is, to go from $0 \leq j < j_{\max}$ to $j_0 \leq j < j_{\max} + j_0$, with the effect that the number of shears at every scale is increased.

4.1 General Notes

The experiments which we undertook are synthetic experiments. This means that the blurred and noisy images for the algorithm input were generated by convolving a reference image with a predefined filter and applying white Gaussian noise with a given standard deviation σ. In order to measure algorithm's performance we used the *signal-to-error ratio gain* to facilitate quantitative comparisons. For the estimate $\mathbf{x}_*$, original signal $\mathbf{x}_{\text{orig}}$ and input signal $\mathbf{y}$ we define signal-to-error ratio gain as

$$\text{SERG}(\mathbf{x}_*) := \text{SER}(\mathbf{x}_*) - \text{SER}(\mathbf{y}),$$

where

$$\mathrm{SER}(\mathbf{v}) := 10\log_{10}\left(\frac{\|\mathbf{x}_{\mathrm{orig}}\|^2}{\|\mathbf{v}-\mathbf{x}_{\mathrm{orig}}\|^2}\right), \quad \mathbf{v}\in\mathbb{C}^N.$$

The SERG values listed in Table 1 are averages from 10 different algorithm outputs obtained through inputs with different noise realizations. If the "convergence rate" of one of the algorithms or regularisation methods is mentioned, it refers to the speed of SERG improvement with respect to the number of iterations. A caveat associated with using SERG, or similar measures of restoration quality, lies in the fact that it does not necessarily correspond well to the perceived restoration quality.

Unless the standard deviation σ of the Gaussian noise is explicitly given, it will be calculated by fixing the *blurred signal-to-noise ratio*, abbreviated as BSNR and a filter with convolution matrix H. It is defined as follows:

$$\mathrm{BSNR} = 10\log_{10}\left(\frac{\|\mathsf{H}\mathbf{x}_{\mathrm{orig}}\|-N\mathrm{mean}(\mathsf{H}\mathbf{x}_{\mathrm{orig}})^2}{N\sigma^2}\right)$$

The advantage of specifying the BSNR is that σ will be chosen according to the signal strength of the blurred original image without noise, and thus the degradation of image quality, caused by the noise, is similar across different filters and images, provided that BSNR level remains unchanged. The filter we have used for the experiments is a motion blur filter generated by the Matlab function `fspecial` with `motion` as the filter type, `12` as the motion length and `45` as the angle.

Although the complete program offers different choices for the initial estimate $\mathbf{x}^{(0)}$, we adhere to the Tikhonov estimate [11]

$$\mathbf{x}^{(0)} = (\tilde{\mathsf{H}}^\top\tilde{\mathsf{H}} + \beta\sigma^2\mathsf{I}_N)^{-1}\tilde{\mathsf{H}}^\top\mathbf{y}$$

where we have used $\beta = 10^{-3}$ in all experiments. Using the Tikhonov estimate helps to reduce unwanted artifacts produced by the algorithm during preliminary testings. Furthermore, the regularisation parameter was chosen uniformly for all scales in most experiments.

4.2 Comparison Between Regularisation Methods for the FTL Algorithm

The goal of the first and largest experiment we have conducted was to compare different versions of the FTL algorithm against each other. Algorithms were applied to blurred and noisy versions of the two standard test images "Lena" and "Cameraman". This was done for three different BSNR levels. Both images measure 256 pixels in each direction. The regularisation parameters were the same for every scale and chosen by hand to approximately yield the highest SER gain after 300 iterations. The performance of the algorithm was evaluated through the SERG after 10, 30 and 150 iterations.

The results regarding the motion filter are summarized in Table 1. Shannon shearlets outperform Shannon wavelets for any of the numbers of iterations tested, except at BSNR value of 30, where only the algorithm convergence is faster. A visual comparison for the "Lena" image with the Shannon wavelet and shearlet algorithms can be found in Figure 4, where the parameters again have been adjusted to reduce artifacts. Note that the

	BSNR	Method	SERG after n iterations		
			$n = 10$	$n = 30$	$n = 150$
Cameraman	30	sw	3.82	4.76	5.62
		ds	5.27	5.61	5.78
	40	sw	6.71	7.42	9.31
		ds	8.70	9.32	9.77
	50	sw	11.43	11.83	13.57
		ds	12.68	13.73	14.60
Lena	30	sw	3.47	4.23	5.49
		ds	5.04	5.35	5.48
	40	sw	6.63	7.07	8.71
		ds	8.33	8.94	9.34
	50	sw	11.53	11.84	13.16
		ds	12.37	13.06	14.03

Table 1: *Comparison of regularisation methods for the motion filter. sw: Shannon wavelets, ds: Shannon shearlets*

noise of the estimates obtained with the Shannon wavelet regularisation can of course be reduced through increasing λ, but only at the cost of making the estimate less sharp.

Figure 4: *Estimates after 10 and 30 iterations with Shannon wavelets (middle) and Shannon shearlets (bottom).*
Filter: motion, BSNR: 30

5 Conclusion

We have introduced discrete Shannon shearlet systems and generalized the Shannon wavelet-based Fast Thresholded Landweber deconvolution algorithm of [1] to Shannon shearlets. Due to the finer frequency resolution of these shearlet systems our approach achieves superior convergence rates for specific problems, such as motion deblurring, as compared to the approach of [1].

The results of [1] have been generalized to arbitrary wavelet systems in the paper [12]. In future work we aim to generalize the results of [12] to arbitrary (non-Shannon) shearlet systems.

References

[1] C. Vonesch and M. Unser. A Fast Thresholded Landweber Algorithm for Wavelet-Regularized Multidimensional Deconvolution. *IEEE Trans. Image Process*, 17(4):539–549, 2008.

[2] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, third edition, 2009.

[3] I. Debauchies, M. Defrise, and C. De Mol. An Iterative Thresholding Algorithm For Linear Inverse Problems With a Sparsity Constraint. *Comm. Pure Appl. Math*, 57(11):1413–1457, November 2004.

[4] B. Dong and Z. Shen. *MRA-Based Wavelet Frames and Applications*. IAS Lecture Notes Series, Park City Mathematics Institute, first edition, 2011.

[5] E.J. Candès and D. L. Donoho. Continuous Curvelet Transform: I. Resolution of the Wavefront Set . *Appl. Comput. Harmon. Anal.*, 19:162–197, 2003.

[6] Philipp Grohs, Sandra Keiper, Gitta Kutyniok, and Martin Schaefer. α-molecules: Wavelets, curvelets, shearlets, ridgelts and beyond. In *Proceedings SPIE 2013*, 2013.

[7] G. Kutyniok and D. Labate. Resolution of the Wavefront Set using Continuous Shearlets. *Trans. Amer. Math. Soc.*, 361:2719–2754, 2009.

[8] P. Grohs and G. Kutyniok. Parabolic molecules. *Foundations of Computational Mathematics*, 2014. in press.

[9] P. Grohs. Continuous shearlet frames and resolution of the wavefront set. *Monatshefte für Mathematik*, 164(4):393–426, 2011.

[10] G. Kutyniok and D. Labate. *Shearlets - Multiscale Analysis for Multivariate Data*. Birkhäuser, Boston, first edition, 2012.

[11] J. N. Franklin. On Tikhonov's Method for Ill-Posed Problems. *Math. Comp*, 28(128):889–907, 1978.

[12] Cedric Vonesh and Michel Unser. A fast multilevel algorithm for wavelet-regularized image restoration. *IEEE Transactions on Image Processing*, 18:509–523, 2009.

[13] M. A. T. Figueiredo and R. D. Nowak. An EM Algorithm for Wavelet Based Image Restoration. *IEEE Trans. Image Process*, 12(8):906–916, 2003.

[14] U. Wiesmann. The Shearlet-Regularized Fast Thresholded Landweber Algorithm. Master's thesis, ETH Zürich, 2013.

[15] S. Mallat and D. Labate. *A Wavelet Tour of Signal Processing*. Academic, San Diego, second edition, 1999.

[16] L. Landweber. An Iterative Formula for Fredholm Integral Equationes of the First Kind. *American Journal of Mathematics*, 73:615–624, 1951.

[17] D. L. Donoho. De-noising by Soft-Thresholding. *IEEE Trans. Inform. Theory*, 41(3):613–627, 1995.

[18] S. Häuser. Fast Finite Shearlet Transform: A Tutorial. *Preprint*, 2012.

[19] E.J. Candès J. L. Starck and D. L. Donoho. The Curvelet Transform for Image Denoising. *IEEE Transactions on Image Processing*, 11:670–684, 2000.

Recent Research Reports

Nr.	Authors/Title

2013-49 P. Grohs and S. Hosseini
ε-Subgradient Algorithms for Locally Lipschitz Functions on Riemannian Manifolds

2013-50 A. Andersson and R. Kruse and S. Larsson
Duality in refined Watanabe-Sobolev spaces and weak approximations of SPDE

2014-01 M. Eigel and C.J. Gittelson and Ch. Schwab and E. Zander
A convergent adaptive stochastic Galerkin finite element method with quasi-optimal spatial meshes

2014-02 R. Kaeppeli and S. Mishra
Structure preserving schemes

2014-03 K. Grella
Sparse tensor phase space Galerkin approximation for radiative transport

2014-04 A. Hiltebrand and S. Mishra
Efficient preconditioners for a shock capturing space-time discontinuous Galerkin method for systems of conservation laws

2014-05 X. Claeys and R. Hiptmair and C. Jerez-Hanckes and S. Pintarelli
Novel Multi-Trace Boundary Integral Equations for Transmission Boundary Value Problems

2014-06 X. Claeys and R. Hiptmair
Integral Equations for Acoustic Scattering by Partially Impenetrable Composite Objects

2014-07 P. Grohs and S. Keiper and G. Kutyniok and M. Schaefer
Cartoon Approximation with α -Curvelets

2014-08 P. Grohs and M. Sprecher and T. Yu
Scattered Manifold-Valued Data Approximation