Structured Sparsity
in Gabor Analysis

Dominik Fuchs
University of Vienna
Faculty of Mathematics

WS 2012/13
1 Motivation
Contents

1 Motivation

2 Sparse Regularization
   • Problem
   • Penalty Measure
Contents

1 Motivation

2 Sparse Regularization
   • Problem
   • Penalty Measure

3 Thresholding and Iterative Algorithms
   • Thresholding
   • (F)ISTA
Contents

1 Motivation

2 Sparse Regularization
   • Problem
   • Penalty Measure

3 Thresholding and Iterative Algorithms
   • Thresholding
   • (F)ISTA

4 Persistence and Neighborhood
   • Neighborhood
   • Empirical Wiener
Motivation

Sparse Regularization
- Problem
- Penalty Measure

Thresholding and Iterative Algorithms
- Thresholding
- (F)ISTA

Persistence and Neighborhood
- Neighborhood
- Empirical Wiener

Matlab Examples
Structure

1. Motivation
2. Sparse Regularization
   - Problem
   - Penalty Measure
3. Thresholding and Iterative Algorithms
   - Thresholding
   - (F)ISTA
4. Persistence and Neighborhood
   - Neighborhood
   - Empirical Wiener
5. Matlab Examples
‘Natural’ signals often yield unwanted noise disturbing the original sound.
‘Natural’ signals often yield unwanted noise disturbing the original sound.

Consider:
Narural Signals

‘Natural’ signals often yield unwanted noise disturbing the original sound.

Consider:

- ‘air rustle’ or even background noise in microphoned signals
‘Natural’ signals often yield unwanted noise disturbing the original sound.

Consider:

- ‘air rustle’ or even background noise in microphoned signals
- directly recorded music with electric instruments
‘Natural’ signals often yield unwanted noise disturbing the original sound.

Consider:

- ‘air rustle’ or even background noise in microphoned signals
- directly recorded music with electric instruments
- clipping
Example 1: There could be some background noise.

Microphoned Signal
Example 2a: Recordings with bad input can cause some noise.
Example 2b: Or even more noise.
Example 3: Too loud input signals can lead to clipping.

Clipped Signal
Motivation

Sparse Regularization

Thresholding and Iterative Algorithms

Persistence and Neighborhood

Matlab Examples

1 Motivation

2 Sparse Regularization
   • Problem
   • Penalty Measure

3 Thresholding and Iterative Algorithms
   • Thresholding
   • (F)ISTA

4 Persistence and Neighborhood
   • Neighborhood
   • Empirical Wiener

5 Matlab Examples
We will start by formulating the problem.
We will start by formulating the problem.

For our natural signal $y$, we get

$$y = f + e$$

where $f$ is the clean/wanted signal and $e$ the additional noise.
For checking the variance of our signal under the synthesis we define the discrepancy.
Discrepancy

For checking the variance of our signal under the synthesis we define the discrepancy

**Definition**

*The Discrepancy is defined by*

\[ \Delta(c) := \frac{1}{2} \| y - \Phi c \|_2^2 \]

*with synthesis operator* \( \Phi : \mathcal{H}_c \to \mathcal{H}_s, \Phi = (\varphi_1, \ldots, \varphi_\gamma, \ldots) \), *signal y and coefficients* \( c \in \mathcal{H}_c \).

We need to find coefficients, s.t. \( \Delta(c) \) of the data \( y \) and the image of \( c \) is minimized.
Problem:

- solution is not unique
- not continuously dependent on the data
Problem:

- solution is not unique
- not continuously dependent on the data

We need to take additional constraints on the coefficients into account.

Definition

The regularized functional called **Lagrangian** is defined by

\[ \mathcal{L}(c) := \mathcal{L}_{y,\lambda}(c) := \Delta(c) + \lambda \Psi(c) \]

where \( \Psi : \mathcal{H}_c \to \mathbb{R}_0^+ \) is the so called **penalty measure** and \( \lambda > 0 \) the **Lagrange-multiplier** resp. **sparsity level**.
Our aim is to seek \( \hat{c} \in \mathcal{H}_c \) such that

\[
\hat{c} = \arg\min_c L(c)
\]

This is the first general formulation of our problem for sparse regularization!
Non-convex Problem

For sparsity we want to minimize the number of non-zero coefficients

\[ \|c\|_0 := \#\{c_\gamma : c_\gamma \neq 0\}, \text{ i.e. } \Psi = \|c\|_0 \]
For sparsity we want to minimize the number of non-zero coefficients

\[ \| c \|_0 := \# \{ c_\gamma : c_\gamma \neq 0 \}, \quad i.e. \quad \Psi = \| c \|_0 \]

!! Problem not solvable in finite time !!
For sparsity we want to minimize the number of non-zero coefficients

\[ \|c\|_0 := \# \{c_\gamma : c_\gamma \neq 0\}, \text{ i.e. } \Psi = \|c\|_0 \]

!! Problem not solvable in finite time !!

It was shown that $\ell_1$-minimization uniquely recovers the $\ell_0$-solution. By using the $\ell_1$-norm instead of the $\ell_0$-norm we get the convex minimization problem $\rightarrow$
LASSO

\[ \hat{c} = \arg\min_c \left\{ \frac{1}{2} \| y - \Phi c \|_2^2 + \lambda \| c \|_1 \right\} \]

This Problem is known as the so called **LASSO - least absolute shrinkage and selection operator**. An equivalent formulation was given in the field of statistics with the name **Basis Pursuit Denoising**.
Mixed Norms

Viewing problem in Gabor-analysis ⇒ atoms ordered along two dimensions
Viewing problem in Gabor-analysis ⇒ atoms ordered along two dimensions

⇒ makes sense to split indices into groups and members
Mixed Norms

Viewing problem in Gabor-analysis ⇒ atoms ordered along two dimensions

⇒ makes sense to split indices into groups and members

Realizable by replacing our penalty by a weighted mixed norm.
Definition (Weighted Mixed Norm)

Let \( \Gamma \) be a doubly labelled index set and \( K, J_1, J_2, \ldots, J_k, \ldots \) be countable index sets such that \( \Gamma_k := \{(k,j) : j \in J_k\} \) \( \forall k \in K \) we have \( \Gamma = \bigcup_{k \in K} \Gamma_k \).

Let \( w = (w_\gamma)_{\gamma \in \Gamma} \) be a positive sequence of weights. The weighted mixed norm \( \ell_{w,p,q} \) on \( \mathcal{H}_c \) for \( 1 \leq p, q < \infty \) is defined by

\[
\|c\|_{w,p,q} := \left( \sum_{k \in K} \left( \sum_{j \in J_k} w_{k,j} |c_{k,j}|^p \right)^{q/p} \right)^{1/q}
\]
By reformulating the Lagrangian by setting $\Psi(c) = \frac{1}{q} \|c\|_{w,p,q}^q$ we get

$$\mathcal{L}_{w,p,q}(c) := \frac{1}{2} \|y - \Phi c\|_2^2 + \frac{1}{q} \|c\|_{w,p,q}^q.$$

and the resulting sparse recovery problem

$$\operatorname{argmin}_c \mathcal{L}_{w,p,q}(c).$$
The mixed norms will consider in the further context and their problems names are:

- $\ell_w, 2, 1$ ..... Group Lasso (GL)
- $\ell_w, 1, 2$ ..... Elitist Lasso (EL)
Structure

1 Motivation

2 Sparse Regularization
   - Problem
   - Penalty Measure

3 Thresholding and Iterative Algorithms
   - Thresholding
   - (F)ISTA

4 Persistence and Neighborhood
   - Neighborhood
   - Empirical Wiener

5 Matlab Examples
We introduce a new thresholding operator \( S_\xi \) that fulfills

\[
\hat{c} = S_{w,p,q}(a) = \arg\min_c \left( \frac{1}{2} \| c - a \|^2_2 + \frac{1}{q} \| c \|_{w,p,q}^q \right)
\]

for any \( a \in \mathcal{H}_c \). In fact \( a = \Phi^* y \).
Definition (Generalized Thresholding Operator)

For $z, w \in \mathcal{H}_c$, $w_\gamma > \delta > 0$ and a non-negative function $\xi = \xi_{\gamma,w} : \mathcal{H}_c \rightarrow [0, \infty]$, the generalized thresholding operator is defined component-wise by

$$\mathbb{S}_\xi(z_\gamma) := z_\gamma (1 - \xi_{\gamma,w}(z))^+$$

where $b \in \mathbb{R}$, $b^+ := \max(b, 0)$. $\xi$ is called the threshold function.
Threshold functions for Mixed Norms

\[ p = q = 1 : \xi_w(z_\gamma) = \frac{w_\gamma}{|z_\gamma|} \quad \text{(Lasso)} \]

\[ p = q = 2 : \xi_w(z_\gamma) = \frac{w_\gamma}{1 + w_\gamma} \quad \text{(Tykhonov Regularization)} \]

\[ p = 2, q = 1 ; w_{k,j} = w_k \forall k,j : \xi_w(z_{k,j}) = \frac{\sqrt{w_{k,j}}}{\|z_k\|_2} \quad \text{(Group-Lasso)} \]

\[ p = 1, q = 2 : \xi_w(z_{k,j}) = \frac{w_{k,j}}{1 + W_{w_k}} \frac{\|z_k\|_w}{|z_{k,j}|} \quad \text{(Elitist-Lasso)} \]

where \( W_{w_k} := \sum_{j_k=1}^{J_k} w_{k,j_k}^2 \), and \( \|z_k\|_w = \sum_{j_k=1}^{J_k} w_{k,j_k} |z_{k,j_k}| \) and for any \( k, j_k \) is a sequence of indices such that \( r_{k,j_k} := \frac{|z_{k,j_k}|}{w_{k,j_k}} \) is decreasing in \( j_k \), and \( J_k \) is the quantity verifying

\[ r_{k, J_k + 1} \leq \sum_{j_k=1}^{J_k + 1} w_{k,j_k}^2 (r_{k,j_k} - r_{k, J_k + 1}) \quad \text{and} \quad r_{k, J_k} > \sum_{j_k=1}^{J_k} w_{k,j_k}^2 (r_{k,j_k} - r_{k, J_k}) \]
For $\Phi : \mathcal{H}_c \to \mathcal{H}_s$ bounded we get a sequence

$$\chi(c) = \mathbb{S}_w(c + \Phi^*(y - \Phi c))$$

that converges to a minimizer of the Lagrangian $\mathcal{L}$. 
For $\Phi : \mathcal{H}_c \to \mathcal{H}_s$ bounded we get a sequence

$$\chi(c) = S_w(c + \Phi^*(y - \Phi c))$$

that converges to a minimizer of the Lagrangian $\mathcal{L}$.

This leads us directly to the soft-thresholding algorithm (ISTA), also called thresholded Landweber iteration:

$$c^{n+1} = (\chi^{n+1}(c^0))_{n+1} = S_w(c^n + \Phi^*(y - \Phi c^n))$$
Since the ISTA - Algorithm with the iteration step \( c^n = \mathcal{S}(b^n) \) with

\[
b^n = c^{n-1} + \Phi^*(y - \Phi c^{n-1})
\]

converges very slowly we want to improve our algorithm. We handle this by modify the choice of \( b^n \) by a linear combination of \( c^n \) and \( c^{n-1} \).
Algorithm:

For $S = S_{w,p,q}$, let $c^0 = b^1 \in \mathcal{H}_s$ and $t_1 = 1$.

Do

\[
\begin{align*}
    c^n &= S(b^n + \Phi^*(y - \Phi b^n)) \\
    t_{n+1} &= \frac{1}{2} \left( 1 + \sqrt{1 + 4t_n^2} \right) \\
    b^{n+1} &= c^n + \left( \frac{t_{n-1}}{t_{n+1}} \right) (c^n - c^{n-1})
\end{align*}
\]

Until convergence
Structure

1 Motivation

2 Sparse Regularization
   • Problem
   • Penalty Measure

3 Thresholding and Iterative Algorithms
   • Thresholding
   • (F)ISTA

4 Persistence and Neighborhood
   • Neighborhood
   • Empirical Wiener

5 Matlab Examples
Often there may occur some sharp cut-off’s. The idea of "smoothing" the coefficients leads to the definition of neighborhood weights.
Often there may occur some sharp cut-off’s. The idea of "smoothing" the coefficients leads to the definition of neighborhood weights.

By convolving a threshold function $\xi$ with a neighborhood-smoothing functional $\eta$. 
Definition (Time-Frequency Neighborhood)

For the countable index set $\Lambda$ the time-frequency neighborhood weights are defined as the non-negative sequences $v_\gamma = v_\gamma(\tilde{\gamma}) \geq 0 \ \forall \gamma, \tilde{\gamma} \in \Lambda$ which fulfill the following properties:

\[
\|v_\gamma\|_2 = 1, \quad \sum_{\tilde{\gamma}} v_\gamma(\tilde{\gamma}) \leq C < \infty \quad \forall \gamma, \quad v_\gamma(\gamma) > 0 \quad \forall \gamma
\]

$N_\gamma := \text{supp}(v_\gamma) = \{ \tilde{\gamma} \in \Gamma : v_\gamma(\tilde{\gamma}) > 0 \}$ is called the time-frequency neighborhood of $\gamma$. 
Neighborhood-smoothing functional

**Definition**

For given neighborhood weights $v_\gamma$, let the neighborhood-smoothing functional $\eta : \mathcal{H}_c \to \mathbb{R}_0^+$ be defined component-wise by

$$
\eta(c_\gamma) := \left( \sum_{\tilde{\gamma} \in \Gamma} v_\gamma(\tilde{\gamma}) |c_{\tilde{\gamma}}|^2 \right)^{1/2}
$$

For $c \in \mathcal{H}_c$, we set $\eta(c) := (\eta(c_\gamma))_{\gamma \in \Gamma}$.
We have now a couple of operators:

\[ \xi^L = \xi_{1,1} \quad \text{Lasso (L)} \]
\[ \xi^{GL} = \xi_{2,1} \quad \text{Group - Lasso (GL)} \]
\[ \xi^{EL} = \xi_{1,2} \quad \text{Elitist - Lasso (EL)} \]
\[ \xi^{WGL} = \xi^*_{1,1} = \xi^L \ast \eta_N \quad \text{windowed Group Lasso (WGL)} \]
\[ \xi^{PGL} = \xi^*_{2,1} = \xi^{GL} \ast \eta_N \quad \text{persistent Group Lasso (PGL)} \]
\[ \xi^{PEL} = \xi^*_{1,2} = \xi^{EL} \ast \eta_N \quad \text{persistent Elitist Lasso (PEL)} \]
Until now the shrinkage level $\lambda$ has been linear. Of course the idea of changing this value during the iteration would come up. For this the theory of fusing structured sparsity with empirical Wiener filtering has been introduced.
Until now the shrinkage level $\lambda$ has been linear. Of course the idea of changing this value during the iteration would come up. For this the theory of fusing structured sparsity with empirical Wiener filtering has been introduced.

We skip the derivation and just consider the relation to our operators so far.

To read more about this theory, take a look into the article

*Kai Siedenburg, "Persistent Empirical Wiener estimation with adaptive threshold selection for audio denoising", ÖFAI, 2012*
By taking our Soft-thresholding operator $S_\xi$ and taking the second power of the threshold function $\xi$ we get the Empirical Wiener operator

$$S^{EW} := S_\xi^\alpha (z_\gamma) := z_\gamma (1 - \xi_{\gamma,w}(z)^\alpha)^+$$

where $\alpha = 2$. 
Structure

1. Motivation
2. Sparse Regularization
   - Problem
   - Penalty Measure
3. Thresholding and Iterative Algorithms
   - Thresholding
   - (F)ISTA
4. Persistence and Neighborhood
   - Neighborhood
   - Empirical Wiener
5. Matlab Examples
In the following examples we see some applications performed with the *StrucAudioToolbox*.

(Download: [http://homepage.univie.ac.at/monika.doerfler/StrucAudio.html](http://homepage.univie.ac.at/monika.doerfler/StrucAudio.html))

The images we will see, are visualizations of the ‘active’ coefficients, not to be confounded with the spectogram.
Denoising: Flanger

The noisy 'Flanger' from above denoised with an PEW.
Denoising: Microphoned signal

The microphoned signal with background noise from a party denoised with an GL and group label in the frequency domain and 5 iteration steps.

![Denoised signal](image1)

![Original signal](image2)
Denoising: Microphoned signal

By setting the grouplabel in time, we might get reconstructed signals like this.
Multi-layer Decomposition

For Musical Clock with an addition generated noise we can perform a separation into tonal and transients parts and add them together afterwards.