Adaptive Structured Block Sparsity via Dyadic Partitioning

Jalal Fadili
Gabriel Peyré and Christophe Chesneau

1GREYC, CNRS-ENSICAEN-Univ. Caen
2CEREMADE, CNRS-Univ. Paris Dauphine
3Department of Mathematics, Univ. Caen

EUSIPCO 2011
Today’s talk is about ...

- Structured (block) sparsity.
- Adaptivity.
- Dyadic partitioning: CART algorithm.
- Risk estimation: bias-variance trade off.
- Non-iterative fast algorithms.
Outline

Problem statement.

Block estimators and risks:
  - Block estimators.
  - Estimators risks.

Block sparsity by dyadic partitioning.

Experimental results.

Conclusion and future work.
Outline

Problem statement.

Block estimators and risks:
  - Block estimators.
  - Estimators risks.

Block sparsity by dyadic partitioning.

Experimental results.

Conclusion and future work.
Statistical estimation

\[ s = Hf + \varepsilon, \]
\[ \varepsilon \sim \mathcal{N}(0, \sigma^2). \]

\[ f \text{ is compressible in } \Phi : \Phi^* f = x \in w\ell_p(C), \ p \leq 1. \]

\[ \text{e.g. denoising : } H = I. \]

\[ \text{arginf}_{D_{\Phi}} \mathbb{E}_{\varepsilon} \left[ \| D_{\Phi}(s) - f \| ^2 \right] \]

\[ \hat{f} := D_{\Phi}(s) = \Phi \circ S_\lambda \circ \Phi^*(s) \]
Statistical estimation

\[ s = H f + \varepsilon, \]

\( \varepsilon \sim \mathcal{N}(0, \sigma^2). \)

\( f \) is compressible in \( \Phi : \Phi^* f = x \in \text{wl}_p(C), \ p \leq 1. \)

e.g. denoising : \( H = I. \)

\[ f \] is compressible in \( \Phi : \Phi^* f = x \in \text{wl}_p(C), \ p \leq 1. \)

\[ f \text{ sarginf}_{D\Phi} \mathbb{E}_\varepsilon \left[ ||D\Phi(s) - f||^2 \right] \]

\[ \hat{f} := D\Phi(s) = \Phi \circ S_\lambda \circ \Phi^*(s) \]
Block Thresholding

\[ y \sim \sum_{j} \phi_j(x) \]

\[ \phi_j(x) = \begin{cases} \frac{1}{\sqrt{2^j}} \phi \left( \frac{x}{2^j} \right) & \text{for } j > 0 \\ \phi(x) & \text{for } j = 0 \end{cases} \]

\[ \phi = \text{Wavelets} \]

\[ \Phi = \text{Curvelets} \]

\[ s \]

\[ \hat{x}_b \]

\[ \Phi = \text{Wavelets} \]

\[ \Phi = \text{Curvelets} \]
**Fixed block sparsity**

With appropriate **fixed** block size, block thresholding achieves the **optimal minimax** convergence rates over several functional spaces [Cai 99, Cai and Silverman 01, Chesneau-F.-Starck 09].

Extended to inverse problems by [Chesneau-F.-Starck 10].

Adaptive choice of the block size (and threshold) using the SURE [Cai and Zhou 09].

Practical performance confirmed by several authors.
Adaptive block sparsity

Learning the block structure with (heuristic) agglomerative clustering [Rosenblum et al. 10]: NP-hard problem.

Contributions:

- A novel method to adapt the block-sparsity structure to the observed noisy data.
- The block-sparsity is dyadically organized in a tree.
- A fast algorithm that yields the best partition by minimizing exactly a new data-driven estimator of the risk.
Outline

- Problem statement.
- Block estimators and risks:
  - Block estimators.
  - Estimators risks.
- Block sparsity by dyadic partitioning.
- Experimental results.
- Conclusion and future work.
Block Estimators

A block segmentation \( B \) is a disjoint union of the indices set

\[
\{0, \ldots, N - 1\} = \bigcup_{b \in B} b, \quad b \cap b' = \emptyset, \; \forall b \neq b'.
\]

A block-based thresholding estimator of \( x \) is \( \hat{x} = S_{\lambda,B}(y) \)

\[
\forall b \in B, \; i \in b, \quad \hat{x}[i] = \rho_{\lambda}(\|y_b\|)y[i], \quad \|y_b\|^2 = \frac{1}{|b|} \sum_{i \in b} |y[i]|^2.
\]

We consider two popular thresholding rules: soft and James-Stein (JS)

\[
\rho_{\lambda}^{\text{Soft}}(a) = \max \left( 0, 1 - \frac{\lambda}{\|a\|} \right), \quad \rho_{\lambda}^{\text{JS}}(a) = \max \left( 0, 1 - \frac{\lambda^2}{\|a\|^2} \right).
\]

Other thresholding rules could be considered as well if \( a \mapsto \rho_{\lambda}(a) \) is weakly differentiable.
Outline

- Problem statement.
- Block estimators and risks:
  - Block estimators.
  - Estimators risks.
- Block sparsity by dyadic partitioning.
- Experimental results.
- Conclusion and future work.
Estimators risks

Stein Unbiased Risk Estimator (recall $\varepsilon \sim \mathcal{N}(0, \sigma^2)$):

**Lemme (Stein’s lemma)** Suppose $x + \varepsilon = y \in \mathbb{R}^n \mapsto S(y)$ is weakly differentiable, and let

$$J(y, S) = n\sigma^2 + \|y - S(y)\|^2 + 2\sigma^2 \text{div}(S - \text{Id})(y),$$

where $\text{div}(f)(x) = \sum_i \frac{\partial f_i}{\partial x_i}(x)$. The SURE is an unbiased estimator of the risk, i.e.

$$\mathbb{E}_\varepsilon(\|S(y) - x\|^2) = \mathbb{E}_\varepsilon(J(y, S)).$$

**Proposition** The SURE on each block $b \in B$ corresponding to the soft and JS estimators are

$$J^{\text{Soft}}(y_b, \lambda, \sigma) = |b|\sigma^2 + \left(|b|\|y_b\|^2 - 2|b|\sigma^2\right)I(\|y_b\| < \lambda) + \left(|b|\lambda^2 - 2\sigma^2(|b| - 1)\frac{\lambda}{\|y_b\|}\right)I(\|y_b\| \geq \lambda),$$

$$J^{\text{JS}}(y_b, \lambda, \sigma) = |b|\sigma^2 + \left(|b|\|y_b\|^2 - 2|b|\sigma^2\right)I(\|y_b\| < \lambda) + \frac{|b|\lambda^2 - 2\sigma^2(|b| - 2)}{\|y_b\|^2 / \lambda^2}I(\|y_b\| \geq \lambda).$$
Estimators risks

Stein Unbiased Risk Estimator (recall $\varepsilon \sim \mathcal{N}(0, \sigma^2)$):

**Lemme (Stein’s lemma)** Suppose $x + \varepsilon = y \in \mathbb{R}^n \mapsto S(y)$ is weakly differentiable, and let

$$J(y, S) = n\sigma^2 + \|y - S(y)\|^2 + 2\sigma^2 \text{div}(S - \text{Id})(y),$$

where $\text{div}(f)(x) = \sum_i \frac{\partial f_i}{\partial x_i}(x)$. The SURE is an unbiased estimator of the risk, i.e.

$$\mathbb{E}_\varepsilon(\|S(y) - x\|^2) = \mathbb{E}_\varepsilon(J(y, S)).$$

**Proposition** The SURE on each block $b \in B$ corresponding to the soft and JS estimators are

$$J^{\text{Soft}}(y_b, \lambda, \sigma) = |b|\sigma^2 + \left(|b| \|y_b\|^2 - 2|b|\sigma^2\right) I(\|y_b\| < \lambda) + \left(|b| \lambda^2 - 2\sigma^2(|b| - 1) \frac{\lambda}{\|y_b\|}\right) I(\|y_b\| \geq \lambda),$$

$$J^{\text{JS}}(y_b, \lambda, \sigma) = |b|\sigma^2 + \left(|b| \|y_b\|^2 - 2|b|\sigma^2\right) I(\|y_b\| < \lambda) + \frac{|b| \lambda^2 - 2\sigma^2(|b| - 2)}{\|y_b\|^2 / \lambda^2} I(\|y_b\| \geq \lambda).$$

Infer the block structure and $\lambda$ by minimizing $\sum_{b \in B} J^*(y_b, \lambda, \sigma)$ with respect to them.
A new risk estimator

- The SURE is an unbiased estimator of the risk, but may exhibit a non-negligible variance.
- This is the case for the SURE of the block Soft and JS estimators.
- Our proposal: Monte-Carlo integration noting that $|b| \|y_b\|^2 / \sigma^2 \sim \chi^2_{|b|}(|b| \|x_b\|^2 / \sigma^2)$.
- A new risk estimator:

$$\hat{J}^{JS}(y_b, \lambda, \sigma) = \frac{1}{K} \sum_{i=1}^{K} \left( |b| \sigma^2 + (z_i - 2|b|\sigma^2)I(z_i < \lambda^2|b|) - \frac{\lambda^2 |b| (\lambda^2 |b| - 2\sigma^2(|b| - 2))}{z_i} \frac{I(z_i \geq \lambda^2|b|)}{|z_i|} \right),$$

$$J^{JS}(y_b, \lambda, \sigma) = |b| \sigma^2 + \left( |b| \|y_b\|^2 - 2|b|\sigma^2 \right)I(\|y_b\| \leq \lambda) + \frac{\lambda^2 \left( |b| \lambda^2 - 2\sigma^2(|b| - 2) \right)}{\|y_b\|^2} I(\|y_b\| \geq \lambda).$$

where $z_i \sim \sigma^2 \chi^2_{|b|}((\hat{\mu}))$, $\hat{\mu} = |b| \max \left( 0, \frac{\|y_b\|^2}{\sigma^2} - 1 \right)$.

- This turns to be an asymptotically unbiased, with a lower variance, estimator of the risk.
Outline

Problem statement.

Block estimators and risks:
  - Block estimators.
  - Estimators risks.

Block sparsity by dyadic partitioning.

Experimental results.

Conclusion and future work.
We seek

\[ (\lambda^*, B^*) \in \text{Argmin}_{\lambda, B} \sum_{b \in B} J^\bullet(y_b, \lambda, \sigma). \]

This is an intractable combinatorial problem if no additional structure on the set of allowable partitions.

A fast algorithm inspired by the CART methodology assuming that the blocks are organized in a tree structure by a recursive dyadic subdivision of the blocks.

A block partition \( B \) is associated to a subtree \( T \) of the whole \( 2^d \)-tree containing all possible partitions.

The blocks are the leaves \( \mathcal{L}(T) \) of \( T \).
**Dyadic Block Partitioning Algorithm**

\[
\min_{T,\lambda} \sum_{b \in \mathcal{L}(T)} \hat{J}(y_b, \lambda, \sigma)_{\lambda,B}.
\]

\(\lambda\) is a scalar \(\rightarrow\) its value optimized by any dichotomic-like (e.g. Golden Section) search algorithm over the cost function marginally minimized with respect to \(T\).

---

**Compute the risk on each block.** For each possible dyadic block \(b_{j,i}\), compute

\[
J_{j,i} = \hat{J}(y_{b_{j,i}}, \lambda, \sigma).
\]

**Best blocks selection.** A bottom-up recursive pruning of the complete tree. 

**for** depth \(j = J - 1 \ldots, 0\) **and** position \(0 \leq i < 4^j\) **do**

\[
\tilde{J}_{j,i} = \min(J_{j,i}, J_{j,i}^c), \quad J_{j,i}^c = \sum_{l=0}^{3} \tilde{J}_{j+1,4i+l}.
\]

**if** \(\tilde{J}_{j,i} = J_{j,i}\) **then**

node \((j, i)\) is a leaf;

**else**

\((j, i)\) is an interior node.

---

**Complexity** \(O(KN)\) operations.
Outline

- Problem statement.
- Block estimators and risks:
  - Block estimators.
  - Estimators risks.
- Block sparsity by dyadic partitioning.
- Experimental results.
- Conclusion and future work.
Synthetic images

PSNR=20 dB
Real images

PSNR=20 dB

QT with $\hat{J}^{JS}$

QT with $\mathbb{E}(\|\rho^{JS}(y) - x\|^2)$

QT with $\hat{J}^{\text{Soft}}$

QT with $\mathbb{E}(\|\rho^{\text{Soft}}(y) - x\|^2)$

Risk (dB) vs $\lambda/\sigma$
Denoising

Original | Noisy PSNR=10 dB | QT with $\hat{J}_{JS}$
PSNR=29.5 dB | QT with $\hat{J}_{Soft}$
PSNR=28.94 dB

Fixed $4 \times 4 \rho_{JS}$
PSNR=28.17 dB | Fixed $4 \times 4 \rho_{Soft}$
PSNR=28.32 dB | QT with $\mathbb{E}(\|\rho_{JS}(y) - x\|^2)$
PSNR=30.66 dB | QT with $\mathbb{E}(\|\rho_{Soft}(y) - x\|^2)$
PSNR=29.56 dB
**Denoising**

<table>
<thead>
<tr>
<th>Original</th>
<th>Noisy PSNR=10 dB</th>
<th>QT with $\widehat{J}_{JS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td>PSNR=26.62 dB</td>
</tr>
<tr>
<td>Fixed $4 \times 4 \rho_{JS}$</td>
<td></td>
<td>QT with $\widehat{J}_{Soft}$</td>
</tr>
<tr>
<td>PSNR=26.19 dB</td>
<td></td>
<td>PSNR=26.4 dB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original</th>
<th>Noisy PSNR=10 dB</th>
<th>QT with $\widehat{J}_{JS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3" alt="Image" /></td>
<td><img src="image4" alt="Image" /></td>
<td>PSNR=26.62 dB</td>
</tr>
<tr>
<td>Fixed $4 \times 4 \rho_{Soft}$</td>
<td></td>
<td>QT with $\widehat{J}_{Soft}$</td>
</tr>
<tr>
<td>PSNR=26.24 dB</td>
<td></td>
<td>PSNR=27.40 dB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original</th>
<th>Noisy PSNR=10 dB</th>
<th>QT with $\widehat{J}_{JS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
<td>PSNR=26.62 dB</td>
</tr>
<tr>
<td>Fixed $4 \times 4 \rho_{JS}$</td>
<td></td>
<td>QT with $\widehat{J}_{Soft}$</td>
</tr>
<tr>
<td>PSNR=26.19 dB</td>
<td></td>
<td>PSNR=26.4 dB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original</th>
<th>Noisy PSNR=10 dB</th>
<th>QT with $\widehat{J}_{JS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
<td>PSNR=26.62 dB</td>
</tr>
<tr>
<td>Fixed $4 \times 4 \rho_{Soft}$</td>
<td></td>
<td>QT with $\widehat{J}_{Soft}$</td>
</tr>
<tr>
<td>PSNR=26.24 dB</td>
<td></td>
<td>PSNR=27.40 dB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original</th>
<th>Noisy PSNR=10 dB</th>
<th>QT with $\widehat{J}_{JS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image9" alt="Image" /></td>
<td><img src="image10" alt="Image" /></td>
<td>PSNR=26.62 dB</td>
</tr>
<tr>
<td>Fixed $4 \times 4 \rho_{JS}$</td>
<td></td>
<td>QT with $\widehat{J}_{Soft}$</td>
</tr>
<tr>
<td>PSNR=26.19 dB</td>
<td></td>
<td>PSNR=26.4 dB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original</th>
<th>Noisy PSNR=10 dB</th>
<th>QT with $\widehat{J}_{JS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image11" alt="Image" /></td>
<td><img src="image12" alt="Image" /></td>
<td>PSNR=26.62 dB</td>
</tr>
<tr>
<td>Fixed $4 \times 4 \rho_{Soft}$</td>
<td></td>
<td>QT with $\widehat{J}_{Soft}$</td>
</tr>
<tr>
<td>PSNR=26.24 dB</td>
<td></td>
<td>PSNR=27.40 dB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original</th>
<th>Noisy PSNR=10 dB</th>
<th>QT with $\widehat{J}_{JS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image13" alt="Image" /></td>
<td><img src="image14" alt="Image" /></td>
<td>PSNR=26.62 dB</td>
</tr>
<tr>
<td>Fixed $4 \times 4 \rho_{JS}$</td>
<td></td>
<td>QT with $\widehat{J}_{Soft}$</td>
</tr>
<tr>
<td>PSNR=26.19 dB</td>
<td></td>
<td>PSNR=26.4 dB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original</th>
<th>Noisy PSNR=10 dB</th>
<th>QT with $\widehat{J}_{JS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image15" alt="Image" /></td>
<td><img src="image16" alt="Image" /></td>
<td>PSNR=26.62 dB</td>
</tr>
<tr>
<td>Fixed $4 \times 4 \rho_{Soft}$</td>
<td></td>
<td>QT with $\widehat{J}_{Soft}$</td>
</tr>
<tr>
<td>PSNR=26.24 dB</td>
<td></td>
<td>PSNR=27.40 dB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original</th>
<th>Noisy PSNR=10 dB</th>
<th>QT with $\widehat{J}_{JS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image17" alt="Image" /></td>
<td><img src="image18" alt="Image" /></td>
<td>PSNR=26.62 dB</td>
</tr>
<tr>
<td>Fixed $4 \times 4 \rho_{JS}$</td>
<td></td>
<td>QT with $\widehat{J}_{Soft}$</td>
</tr>
<tr>
<td>PSNR=26.19 dB</td>
<td></td>
<td>PSNR=26.4 dB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original</th>
<th>Noisy PSNR=10 dB</th>
<th>QT with $\widehat{J}_{JS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image19" alt="Image" /></td>
<td><img src="image20" alt="Image" /></td>
<td>PSNR=26.62 dB</td>
</tr>
<tr>
<td>Fixed $4 \times 4 \rho_{Soft}$</td>
<td></td>
<td>QT with $\widehat{J}_{Soft}$</td>
</tr>
<tr>
<td>PSNR=26.24 dB</td>
<td></td>
<td>PSNR=27.40 dB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original</th>
<th>Noisy PSNR=10 dB</th>
<th>QT with $\widehat{J}_{JS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image21" alt="Image" /></td>
<td><img src="image22" alt="Image" /></td>
<td>PSNR=26.62 dB</td>
</tr>
<tr>
<td>Fixed $4 \times 4 \rho_{JS}$</td>
<td></td>
<td>QT with $\widehat{J}_{Soft}$</td>
</tr>
<tr>
<td>PSNR=26.19 dB</td>
<td></td>
<td>PSNR=26.4 dB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original</th>
<th>Noisy PSNR=10 dB</th>
<th>QT with $\widehat{J}_{JS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image23" alt="Image" /></td>
<td><img src="image24" alt="Image" /></td>
<td>PSNR=26.62 dB</td>
</tr>
<tr>
<td>Fixed $4 \times 4 \rho_{Soft}$</td>
<td></td>
<td>QT with $\widehat{J}_{Soft}$</td>
</tr>
<tr>
<td>PSNR=26.24 dB</td>
<td></td>
<td>PSNR=27.40 dB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original</th>
<th>Noisy PSNR=10 dB</th>
<th>QT with $\widehat{J}_{JS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image25" alt="Image" /></td>
<td><img src="image26" alt="Image" /></td>
<td>PSNR=26.62 dB</td>
</tr>
<tr>
<td>Fixed $4 \times 4 \rho_{JS}$</td>
<td></td>
<td>QT with $\widehat{J}_{Soft}$</td>
</tr>
<tr>
<td>PSNR=26.19 dB</td>
<td></td>
<td>PSNR=26.4 dB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original</th>
<th>Noisy PSNR=10 dB</th>
<th>QT with $\widehat{J}_{JS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image27" alt="Image" /></td>
<td><img src="image28" alt="Image" /></td>
<td>PSNR=26.62 dB</td>
</tr>
<tr>
<td>Fixed $4 \times 4 \rho_{Soft}$</td>
<td></td>
<td>QT with $\widehat{J}_{Soft}$</td>
</tr>
<tr>
<td>PSNR=26.24 dB</td>
<td></td>
<td>PSNR=27.40 dB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original</th>
<th>Noisy PSNR=10 dB</th>
<th>QT with $\widehat{J}_{JS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image29" alt="Image" /></td>
<td><img src="image30" alt="Image" /></td>
<td>PSNR=26.62 dB</td>
</tr>
<tr>
<td>Fixed $4 \times 4 \rho_{JS}$</td>
<td></td>
<td>QT with $\widehat{J}_{Soft}$</td>
</tr>
<tr>
<td>PSNR=26.19 dB</td>
<td></td>
<td>PSNR=26.4 dB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original</th>
<th>Noisy PSNR=10 dB</th>
<th>QT with $\widehat{J}_{JS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image31" alt="Image" /></td>
<td><img src="image32" alt="Image" /></td>
<td>PSNR=26.62 dB</td>
</tr>
<tr>
<td>Fixed $4 \times 4 \rho_{Soft}$</td>
<td></td>
<td>QT with $\widehat{J}_{Soft}$</td>
</tr>
<tr>
<td>PSNR=26.24 dB</td>
<td></td>
<td>PSNR=27.40 dB</td>
</tr>
</tbody>
</table>
Outline

- Problem statement.
- Block estimators and risks:
  - Block estimators.
  - Estimators risks.
- Block sparsity by dyadic partitioning.
- Experimental results.
- Conclusion and future work.
Take-away messages

- A framework to adapt the block-sparsity structure from noisy data: SURE and CART.
- Grounded framework.
- Fast algorithm.
- Promising results.
- Anisotropic blocks.
- Inverse problems.
- Other noise models.
- Other sparsity structures.
- Theoretical guarantees.
Papers and code available at
http://www.greyc.ensicaen.fr/~jfadili

Thanks
Any questions?