IPAM MGA Tutorial on Feature Extraction and Denoising: A Saga of $u + v$ Models

Naoki Saito

saito@math.ucdavis.edu

http://www.math.ucdavis.edu/~saito/

Department of Mathematics

University of California, Davis

Outline

- What Are Features and What Are Noise?
- Some History
- $u + v$ Models
- Very Briefly, Harlan-Claerbout-Rocca Model
- Mumford-Shah Model
- Rudin-Osher-Fatemi Model and Total Variation
- DeVore-Lucier Model and Besov Spaces
- Sparsity

Outline . . .

- Basis Pursuit Denoising
- Sparsity vs. Statistical Independence
- BV via Cohen-Dahmen-Daubechies-DeVore
- BV vs Besov
- Very Briefly, Meyer-Vese-Osher Model for Texture
- My Comments and Summary

Acknowledgment

- Yves Meyer
- Hyeokho Choi (Rice)
- Other authors of articles
- IPAM/UCLA
- NSF & ONR

What Are Features and What Are Noise?

- To answer those questions, we need to specify our aim:
- Approximation
- Compression
- Noise Removal (Denoising)
- Object Detection
- Classification/Discrimination
- Regression

Some History

- Satosi Watanabe (circa 1981) characterized pattern recognition as a quest for minimum entropy by saying, "the essential nature of pattern recognition is . . . a conceptual adaptation to the empirical data in order to see a form in them. The form means a structure which always entails small entropy values."
- Raphy Coifman (circa 1991) suggested that "noise" should be defined as incoherent components in data used to represent data whereas "signal" or "features" are coherent components both relative to waveform libraries. coherent $\approx$ sparse $\approx$ focused

$u + v$ Models

Yves Meyer (2001) describes the so-called $u + v$ models

- The u component is aimed at modeling the objects or important features
- The v component represents textures and noise
- The refined model is the $u + v + w$ model where v and w represent textures and noise, respectively.
- Examples include:
  - Harlan, Claerbout, & Rocca (1984)
  - Mumford & Shah (1985)
  - Rudin, Osher, & Fatemi (1991)

$u + v$ Models . . .

- DeVore & Lucier (1992)
- Chen, Donoho, & Saunders (1995)
- Olshausen & Field (1996)
- Coifman & Sowa (1998)
- Donoho, Huo, & Starck (2000)
- Cohen, Dahmen, Daubechies, DeVore (2000)
- Meyer, Vese, Osher (2002)
- many others . . .

Common Intuitions

- Should be able to represent recognizable patterns/structures in data efficiently and compactly via some invertible transform
- $u =$ signal $\approx$ features $\Longleftrightarrow$ sharply focused $\approx$ sparse
- $v =$ noise $\Longleftrightarrow$ defocused/diffused

What is u?

- Requires some regularity (e.g., smoothness), i.e., $\|u\|_B < C$, where B some appropriate function space, and $C > 0$.
- More general approach $\|Au\|_{B'} < C$, where $A : B \to B'$ is some invertible transform (e.g., $A =$ Radon transform)
- An important problem (modeling) is what B should be for various natural images.

Various Viewpoints

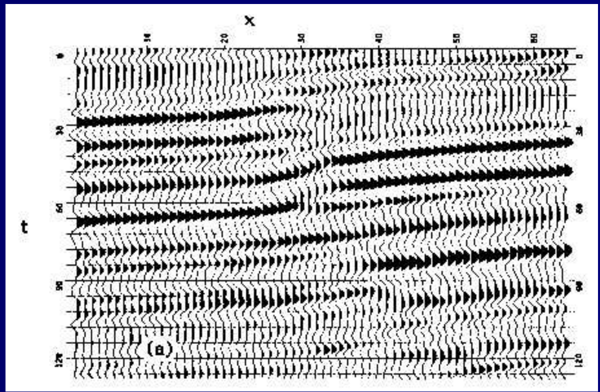
- Harmonic Analysis approach (Cohen, Coifman, Daubechies, DeVore, Donoho, Meyer, . . .)
- PDE approach (Chan, Osher, Meyer, Morel, Sapiro, Vese, . . .)
- Deterministic approach (Cohen, DeVore, Donoho, Osher, Terzopoulos, . . .)
- Stochastic approach (Mumford, Grenander, Donoho, Zhu, Wu, . . .)
- Highly active area and more and more interactions among various schools

Very Briefly, Harlan, Claerbout, & Rocca (1984)

A stacked seismic section = $\sum$ of

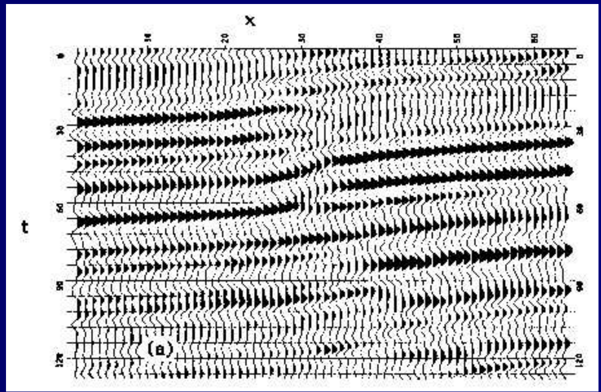
- Geologic component $\approx$ linear events
- Diffraction component $\approx$ hyperbolic events
- Noise component $\approx$ white Gaussian noise $+ \alpha$.

Very Briefly, Harlan, Claerbout, & Rocca (1984) . . .

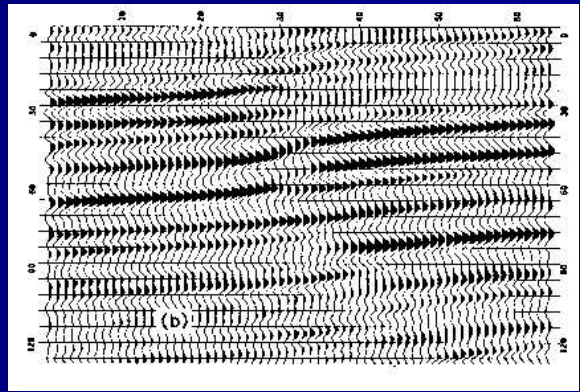


(a) Original

Very Briefly, Harlan, Claerbout, & Rocca (1984) . . .

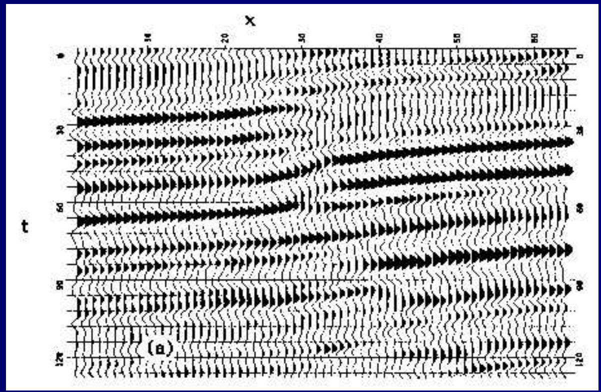


(a) Original

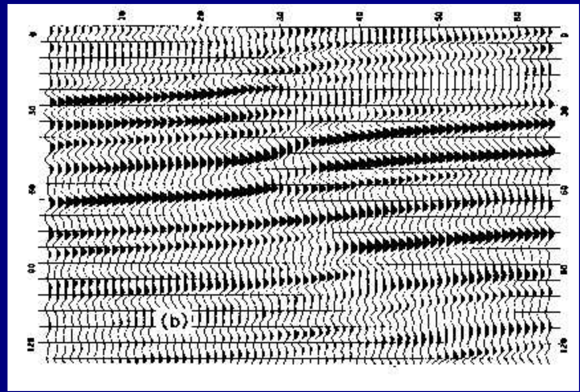


(b) Geology

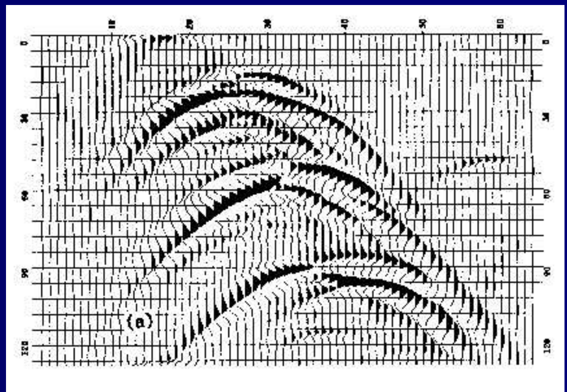
Very Briefly, Harlan, Claerbout, & Rocca (1984) . . .



(a) Original

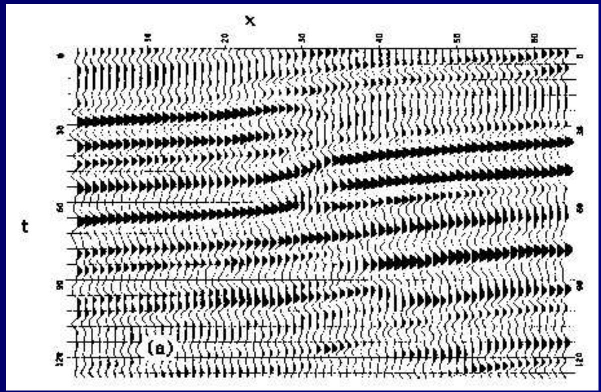


(b) Geology

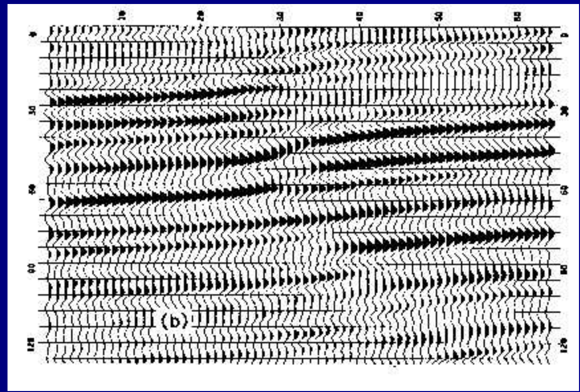


(c) Diffraction

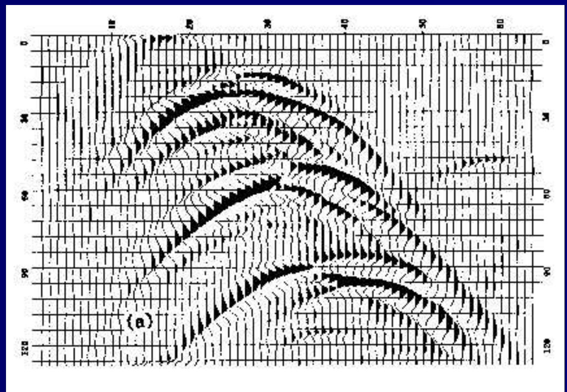
Very Briefly, Harlan, Claerbout, & Rocca (1984) . . .



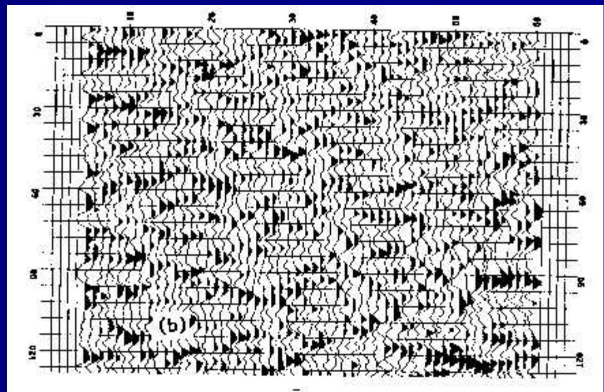
(a) Original



(b) Geology

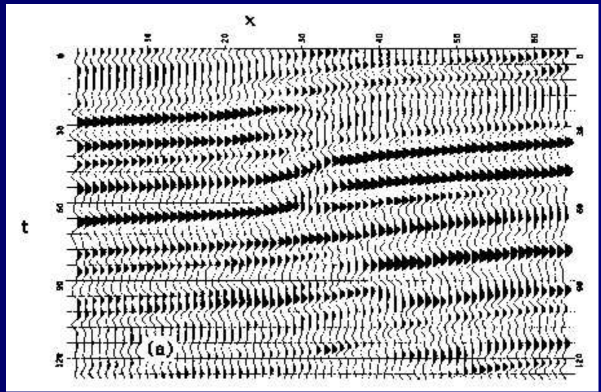


(c) Diffraction

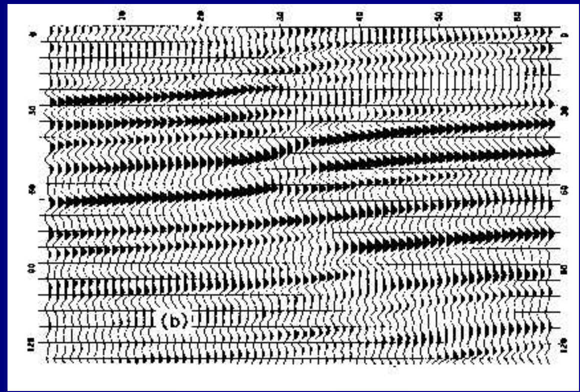


(d) Noise

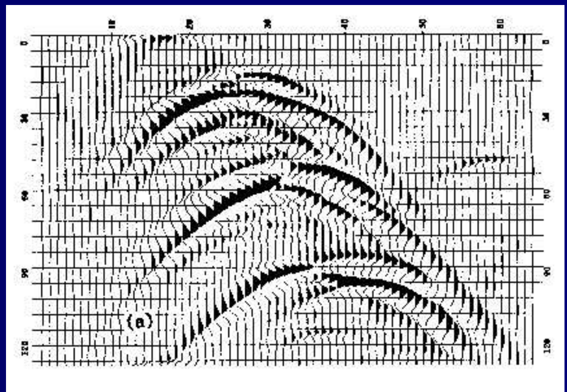
Very Briefly, Harlan, Claerbout, & Rocca (1984) . . .



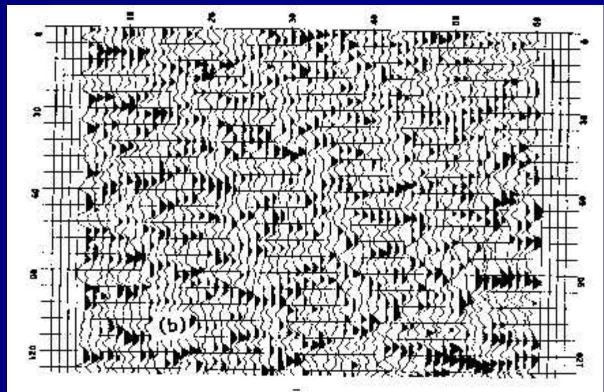
(a) Original



(b) Geology



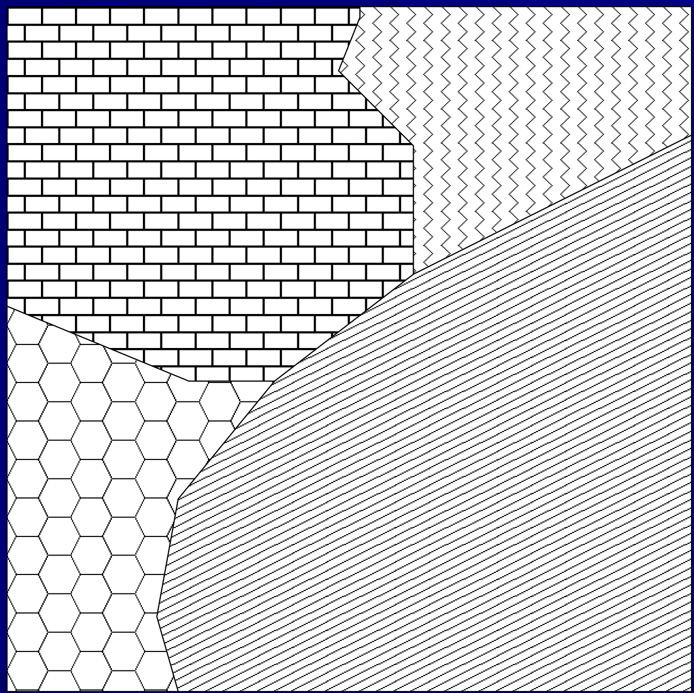
(c) Diffraction



(d) Noise

Mumford & Shah (1985)

- Motivation: simultaneous image segmentation and denoising
- Let $\Omega = [0,1] \times [0,1] \subset \mathbb{R}^2$



Mumford & Shah (1985) ...

- Let u component is smooth everywhere except on a compact set $K \in \Omega$, which is unknown.
- Find u and K from the data $f = u + v$ by minimizing:

$$J_{MS}(u, K) = \int_{\Omega} |f(x) - u(x)|^2 \, \mathrm{d}x + \lambda \int_{\Omega \setminus K} |\nabla u(x)|^2 \, \mathrm{d}x + \mu \mathcal{H}^1(K),$$

where λ, μ are positive weights and $\mathcal{H}^1$ is the 1D Hausdorff measure (total length) of K.

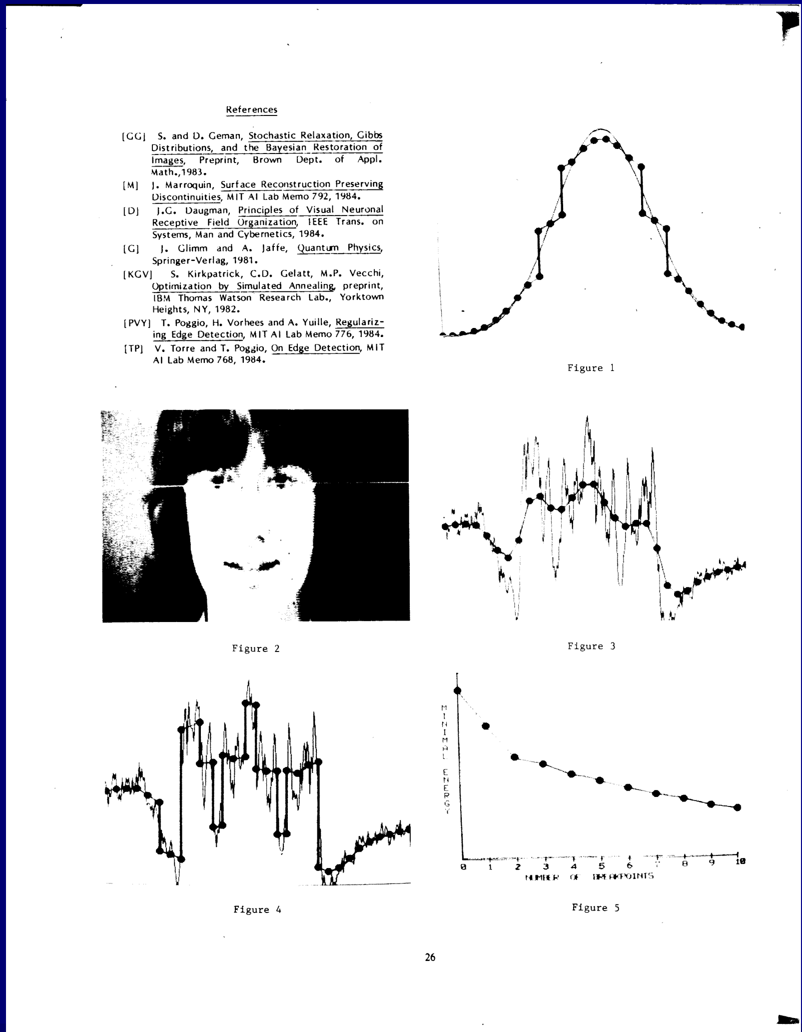
- Measures fidelity, smoothness of u, simplicity of K, respectively.

Mumford & Shah (1985) ...

- Since $u|_{\Omega \setminus K} \in H^1(\Omega \setminus K) = W^{1,2}(\Omega \setminus K)$, the objective is: Find $u \in L^2(\Omega)$ and $K \subset \Omega$ s.t.

$$\inf_{u \in L^2(\Omega)} \|f - u\|_{L^2(\Omega)} \quad \text{subject to } \|u\|_{H^1(\Omega \setminus K)} < C \text{ and } \mathcal{H}^1(K) < C'.$$

Mumford & Shah (1985) . . .



References

[GG] S. and D. Geman, Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, Preprint, Brown Dept. of Appl. Math.,1983.

[M] J. Marroquin, Surface Reconstruction Preserving Discontinuities, MIT AI Lab Memo 792, 1984.

[D] J.G. Daugman, Principles of Visual Neuronal Receptive Field Organization, IEEE Trans. on Systems, Man and Cybernetics, 1984.

[G] J. Glimm and A. Jaffe, Quantum Physics, Springer-Verlag, 1981.

[KGV] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimization by Simulated Annealing, preprint, IBM Thomas Watson Research Lab., Yorktown Heights, NY, 1982.

[PVY] T. Poggio, H. Vorhees and A. Yuille, Regularizing Edge Detection, MIT AI Lab Memo 776, 1984.

[TP] V. Torre and T. Poggio, On Edge Detection, MIT AI Lab Memo 768, 1984.

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

26

Mumford & Shah (1985) . . .

- Very influential in Computer Vision; refined by Ivan Leclerc using MDL formalism; faster numerical algorithm called 'Graduated Non-Convexity' (GNC) algorithm by Andrew Blake and Andrew Zisserman, . . .
- Numerical optimization was and still is an issue
- Choice of λ and μ
- Representation/basis functions for u were not used

Rudin, Osher, & Fatemi (1992)

- Motivation: image enhancement and denoising
- The MS functional is modified to:

$$\begin{aligned} J_{ROF}(u) &= \int_\Omega |f(x) - u(x)|^2 \, \mathrm{d}x + \lambda \int_\Omega |\nabla u(x)| \, \mathrm{d}x \\ &= \|f - u\|^2_{L^2(\Omega)} + \lambda |u|_{BV(\Omega)}, \end{aligned}$$

  where $|u|_{BV(\Omega)}$ is the so-called total variation of u.

- In other words, Find $u \in L^2(\Omega)$ s.t.

$$\inf_{u \in L^2(\Omega)} \|f - u\|_{L^2(\Omega)} \quad \text{subject to} \quad |u|_{BV(\Omega)} < C.$$

Rudin, Osher, & Fatemi (1992) . . .

- Solve the following Euler-Lagrange equation

$$u = f + \frac{\lambda}{2}\nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right),$$

by forming the evolution equation and compute the solution as $t \to \infty$.

- The coarea formula links MS to ROF: If $|u|_{BV(\Omega)} < \infty$,

$$|u|_{BV(\Omega)} = \int_{-\infty}^{\infty} \mathcal{H}^1(\partial E_t[u])\,\mathrm{d}t,$$

where $E_t[u] \triangleq \{x \in \mathbb{R}^n : u(x) > t\}$ is the level set of u at t and $\partial E_t[u]$ is its perimeter.

Rudin, Osher, & Fatemi (1992) . . .

- Boundaries K are not explicit.
- Choice of λ: dynamic, i.e., $\lambda(t)$
- Strictly speaking, total variation is defined more generally and $|u|_{BV(\Omega)} = \int_\Omega |\nabla u| \, \mathrm{d}x$ is true for $u \in L_1^1(\Omega) = W^{1,1}(\Omega) \subset BV(\Omega)$.
- More about $BV(\Omega)$ (the space of functions of bounded variation) will be discussed later.

Rudin, Osher, & Fatemi (1992) . . .

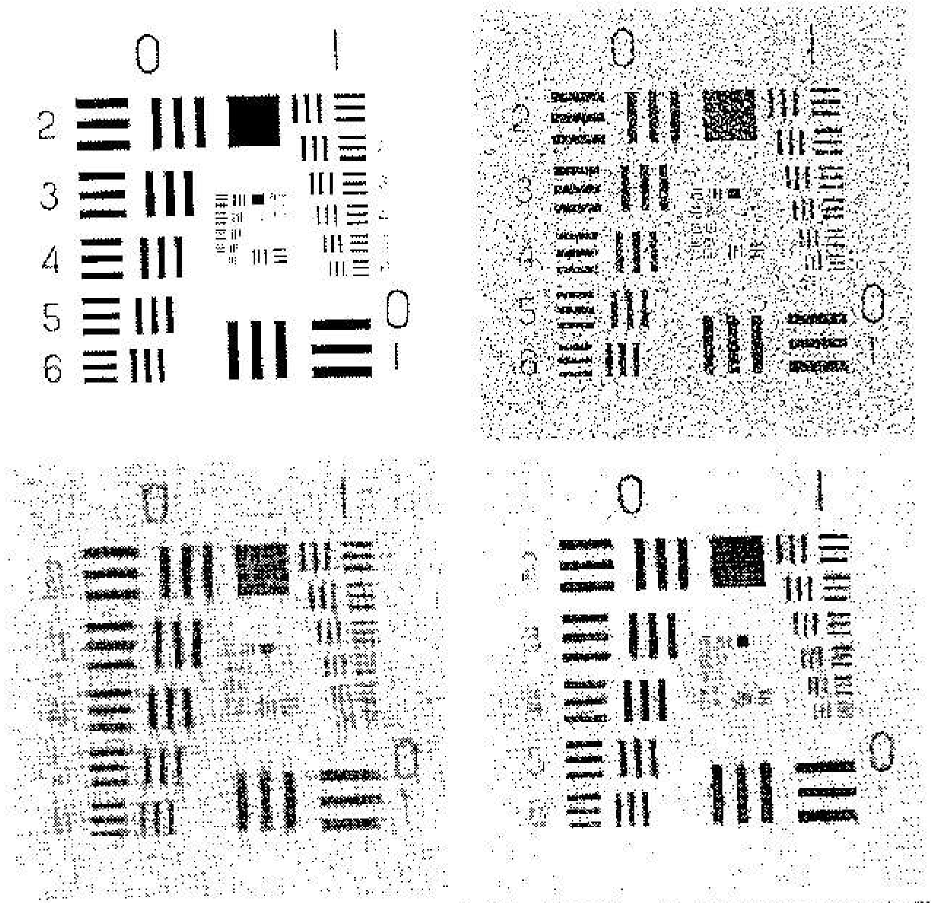


Fig. 3. (a) "Resolution Chart". (b) Noisy "Resolution Chart", SNR = 1.0. (c) Wiener filter reconstruction from (b). (d) TV reconstruction from (b).

Comments on MS and ROF models

- Very influential; generated a new field “PDE-based image processing”
- Characterization of the constraint space via basis functions were not used
- Improved over the years; yet still computationally intensive

DeVore & Lucier (1992)

- Now, the functional becomes:

$$J_{DL}(u) = \|f - u\|^2_{L^2(\Omega)} + \lambda \|c[u]\|_{\ell^1},$$

  where $c[u] = (c_\nu[u])_{\nu\in\Gamma}$ is the expansion coefficients of u relative to an orthonormal wavelet basis $\{\psi_\nu\}_{\nu\in\Gamma}$, i.e., $c_\nu[u] = \langle u, \psi_\nu \rangle$.

- In other words, Find $u \in L^2(\Omega)$ s.t.

$$\inf_{u\in L^2(\Omega)} \|f - u\|_{L^2(\Omega)} \quad \text{subject to} \quad \|c[u]\|_{\ell^1} < C.$$

DeVore & Lucier (1992) ...

- Optimization leads to the soft thresholding (or wavelet shrinkage) on the empirical wavelet coefficients
- Let $f(x) = \sum_{\nu\in\Gamma} c_\nu[f]\,\psi_\nu(x)$. Then,

$$J_{DL}(u) = \sum_{\nu\in\Gamma} \left((c_\nu[f] - c_\nu[u])^2 + \lambda |c_\nu[u]| \right),$$

whose minimization leads to:

$$c_\nu[u] = \begin{cases} c_\nu[f] + \lambda/2 & \text{if } c_\nu[f] < -\lambda/2, \\ 0 & \text{if } |c_\nu[f]| \le \lambda/2 \\ c_\nu[f] - \lambda/2 & \text{if } c_\nu[f] > \lambda/2. \end{cases}$$

DeVore & Lucier (1992) . . .

(a) Noisy Lena

(b) Linear

(c) Nonlinear

DeVore & Lucier (1992) . . .

- In the Besov space language:

$$\|c[u]\|_{\ell^1} \asymp \|u\|_{\dot{B}_1^{1,1}(\Omega)}.$$

  Thus, the constraint in optimization is $\|u\|_{\dot{B}_1^{1,1}(\Omega)} < C$.

- If v is WGN with mean 0, variance σ^2, then the choice of $\lambda \approx const \cdot \frac{\sigma}{N}\sqrt{\log N^2}$, where N is the number of samples in each direction in Ω.

- Lots of effort for deriving (near-)optimal threshold, e.g., DeVore-Lucier, Donoho-Johnstone, and others, most recently Johnstone-Silverman.

Besov Spaces

- Let $f \in L^p(\Omega)$, $0 < p \leq \infty$. Let $0 < \alpha < \infty$, and $0 < q \leq \infty$.
- Then, roughly speaking, functions belonging to the homogeneous Besov space $\dot{B}_p^{\alpha,q}(\Omega)$ has "α derivatives" measured in $L^p(\Omega)$. The parameter q makes finer distinctions in smoothness
- The inhomogeneous Besov space

$$B_p^{\alpha,q}(\Omega) = \dot{B}_p^{\alpha,q}(\Omega) \cap L^p(\Omega)$$

$$\|f\|_{B_p^{\alpha,q}(\Omega)} = \|f\|_{L^p(\Omega)} + \|f\|_{\dot{B}_p^{\alpha,q}(\Omega)}$$

Besov Spaces . . .

- Generalization of Lipschitz/Hölder and L^2-Sobolev spaces because:

$$B_2^{\alpha,2}(\Omega) = W^{\alpha,2}(\Omega) = H^\alpha(\Omega)$$

$$B_\infty^{\alpha,\infty}(\Omega) = \Lambda^\alpha(\Omega) = C^\alpha(\Omega)$$

- Easy to characterize via wavelet coefficients

$$\|f\|_{\dot{B}_\tau^{\alpha,\tau}} \asymp \|c[f]\|_{\ell^\tau} \quad \text{for } \tau = 2/(1+\alpha) \text{ in 2D.}$$

Besov Spaces . . .

- Thus, in the specific DL model with $\alpha = 1 = \tau$, we have:

$$\|u\|_{\dot{B}_1^{1,1}} \asymp \|c[u]\|_{\ell^1}\,.$$

- This norm equivalence means that this DL model is really seeking a function whose wavelet expansion is sparse since $\|c[u]\|_{\ell^1} < C$ is a form of sparsity constraint.

Sparsity via ℓ^p (quasi-) norm ($0 < p \leq 1$)

- Consider a vector or sequence $\boldsymbol{x} = (x_j)_{j \in \mathbb{N}}$.
- Then consider the so-called ℓ^0 (quasi-)norm as the measure of the sparsity of $\boldsymbol{x}$:

$$\|\boldsymbol{x}\|_{\ell^0} = \#\{j \in \mathbb{N} : x_j \neq 0\}.$$

- This counts a number of nonzero components in $\boldsymbol{x}$.
- Thus, under, say, $\|\boldsymbol{x}\|_{\ell^2} = 1$, the smaller $\|\boldsymbol{x}\|_{\ell^0}$ is, the sparser $\boldsymbol{x}$ is; a precise definition of sparsity.
- However, this norm is too fragile to use (e.g., sensitivity to noise).

Sparsity via ℓ^p (quasi-) norm ($0 < p \leq 1$) ...

- Thus, consider the ℓ^p (quasi-) norm $0 < p \leq 1$ instead:

$$\|\boldsymbol{x}\|_{\ell^p} = \left(\sum_{j \in \mathbb{N}} |x_j|^p \right)^{1/p}.$$

- Instead of explicitly saying the number of nonzeros in the sequence via ℓ^0 quasi-norm, we can say:

$$\|\boldsymbol{x}\|_p \leq C \Longrightarrow |x|_{(k)} < Ck^{-1/p},$$

($\because k|x|^p_{(k)} \leq \sum_{j=1}^k |x|^p_{(j)} < C^p$) which relates sparsity to the decay of the magnitudes of the rearranged sequence. The smaller p, the faster the decay.

Sparsity via ℓ^p (quasi-) norm ($0 < p \leq 1$) . . .

- $w\ell^p$, i.e., the weak ℓ^p space is defined as

$$w\ell^p(\mathbb{N}) \triangleq \{\boldsymbol{x} = (x_1, x_2, \ldots) : |x|_{(k)} \leq Ck^{-1/p}, \exists C > 0, \forall k \in \mathbb{N}\}$$

- Clearly, $\ell^p \subsetneq w\ell^p$, e.g., $x_n = n^{-1/p} \in w\ell^p$, but not in ℓ^p.
- Later $w\ell^1$ will be used to “almost” characterize the space $BV(\Omega)$.

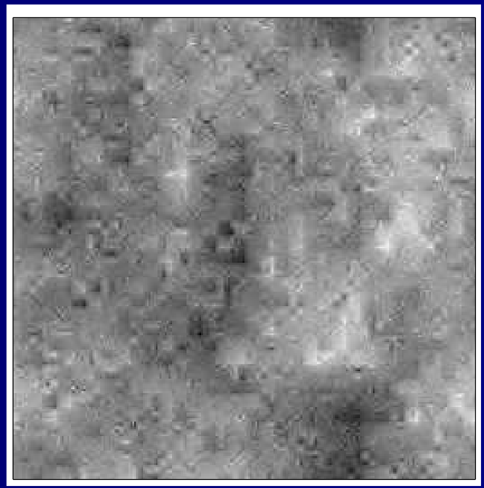
Disadvantages of $\dot{B}_1^{1,1}$

- Even very simple cartoon-like images such as $\chi_E(x)$, where ∂E is a smooth closed curve, do not belong to $\dot{B}_1^{1,1}$
- Oscillatory patterns do not belong to $\dot{B}_1^{1,1}$ either

Disadvantages of $\dot{B}_1^{1,1}$ ...

- Choi and Baraniuk on the Besov spaces vs wavelet-based statistical models (the parameters of the generalized Gaussian distribution $\Longleftrightarrow$ the Besov parameters)

(a) Original u



(b) Random shuffles of $c_{j,\cdot}$

(c) Random sign flips of c

Basis Pursuit Denoising (Chen, Donoho, & Saunders, 1995)

- Under the discrete setting, assume that $\boldsymbol{f} = \boldsymbol{u} + \boldsymbol{v} \in \mathbb{R}^n$, where $\boldsymbol{v} = \sigma \boldsymbol{z}$ is WGN vector with variance σ^2. The functional is similar to J_{DL}:

$$J_{BP}(\boldsymbol{u}) = \|\boldsymbol{f} - \boldsymbol{u}\|_{\ell^2}^2 + \lambda \|\boldsymbol{\alpha}[\boldsymbol{u}]\|_{\ell^1},$$

- The coefficient vector $\boldsymbol{\alpha}[\boldsymbol{u}] \in \mathbb{R}^p$ are in the form:

$$\boldsymbol{u} = \sum_{\gamma \in \Gamma} \alpha_\gamma[\boldsymbol{u}] \boldsymbol{\phi}_\gamma,$$

  where $\{\boldsymbol{\phi}_\gamma\}_{\gamma \in \Gamma}$, $|\Gamma| = p \geq n$ is a dictionary of bases (i.e., redundant) such as stationary wavelet, wavelet packets, local Fourier bases, or other frames, etc.

Basis Pursuit Denoising ...

- Choice of $\lambda = \sigma\sqrt{2 \log p}$.
- Solution of this convex non-quadratic optimization by "primal-dual log-barrier linear programming"
- Specific combination of basis dictionaries, the uncertainty principle, equivalence with ℓ^0 minimization problem $\Longrightarrow$ ask Donoho, Candès, Huo, Elad, Starck who are all participating in this program!

Sparsity vs. Statistical Independence

- Independent Component Analysis
- Stochastic setting; a collection of images
- Can find a basis (LSDB) that provides the least statistically-dependence (the best one) out of the wavelet packet library or local Fourier library
- Better off to pursue the sparsity than independence except the problems that really have statistically independent sources
- Read my articles as well as Donoho & Flesia for more info.

$BV(\Omega)$: Functions of Bounded Variation

- Definition of total variation

$$|u|_{BV(\Omega)} \triangleq \sup_{\boldsymbol{g}} \left\{ \int_\Omega u \nabla \cdot \boldsymbol{g} \, \mathrm{d}x : \boldsymbol{g} \in C_c^1(\Omega; \mathbb{R}^2), |\boldsymbol{g}(x)| \leq 1 \ \forall x \in \Omega \right\}$$

- $BV(\Omega) \subset L^1(\Omega)$ is a Banach space with the norm:

$$\|u\|_{BV(\Omega)} = \|u\|_{L^1(\Omega)} + |u|_{BV(\Omega)}.$$

- If $u \in W^{1,1}(\Omega) \subset BV(\Omega)$, then

$$|u|_{BV(\Omega)} = \int_\Omega |\nabla u(x)| \, \mathrm{d}x,$$

via integration by parts.

$BV(\Omega)$ ...

- In this tutorial, $\Omega \subset \mathbb{R}^2$ is bounded. Thus,

$$W^{1,1}(\Omega) = \mathcal{BV}(\Omega) \subset BV(\Omega) \subset L^2(\Omega) \subset L^1(\Omega).$$

- Unfortunately, $W^{1,1}(\Omega)$ does not contains cartoon-like images such as $\chi_E(x)$ where $E \subset \Omega$ and ∂E is smooth whereas $BV(\Omega)$ does.
- Minimizer u^* exists in $BV(\Omega)$ for the corrected version of the $J_{ROF}(u)$ (Vese, 2001)
- However, the original version of the ROF criterion does not guarantee to get $(u, v) = (\chi_E, 0)$ from $f = \chi_E$ (Y. Meyer 2001).

Besov vs BV

- Embedding (DeVore-Lucier, Donoho, Meyer, . . .):

$$\dot{B}_1^{1,1}(\Omega) \subset BV(\Omega) \subset \dot{B}_1^{1,\infty}(\Omega)$$

- $\dot{B}_1^{1,1}$ does not contain cartoon-like images while $BV(\Omega)$ does.
- On the other hand, $BV(\Omega)$ does not possess any unconditional basis while $\dot{B}_1^{1,1}$ does.

Unconditional Bases

- Let $\{\phi_\nu\}_{\nu\in\Gamma}$ be a basis for a Banach space B.
- Let $f = \sum_{\nu\in\Gamma} c_\nu \phi_\nu \in B$.
- Let $\|f\|_B$ be a functional norm of $f \in B$ and let $\|c[f]\|_b$ be a discrete sequence norm of $c[f] = (c_\nu[f])$.
- Suppose $\|f\|_B \asymp \|c[f]\|_b$.
- This is already not a trivial condition because . . .

Fourier is not an unconditional basis for $L^p(T)$, $p \neq 2$

- Let $f \in B = L^p(T)$, $T = [0, 2\pi)$, and $\phi_\nu(x) = \mathrm{e}^{\mathrm{i}\nu x}$. Let $c_\nu[f]$ be the Fourier coefficients of f.
- Then, $\|f\|_{L^2(T)} = \|c[f]\|_{\ell^2(\mathbb{Z})}$ (Plancherel)
- However, $\|c[f]\|_{\ell^p(\mathbb{Z})}$ does not tell information about $\|f\|_{L^p(T)}$ if $p \neq 2$.
- For example, $\|f\|_{L^4(T)}$ tells you some info about the distribution of the energy of f over T ($\sim$ kurtosis).
- However, $|c_\nu[f]|$ does not tell you anything about $\|f\|_{L^4(T)}$.
- $\Longrightarrow$ the Littlewood-Paley theory

Unconditional Bases . . .

- Then, $\{\phi_\nu\}_{\nu\in\Gamma}$ is called an unconditional basis of B if any sequence $\tilde{c} = (\tilde{c}_\nu)$ satisfying $|\tilde{c}_\nu| \le |c_\nu[f]|, \forall \nu \in \Gamma$, yields a new function $\tilde{f} = \sum_{\nu\in\Gamma} \tilde{c}_\nu \phi_\nu$ that belong to B.
- In other words, operations on the coefficients, such as shrinking, sign flips, do not change the membership of B.
- Examples: Fourier: L^2, Wavelets: $L^p, 1 < p < \infty$, $B^{\alpha,q}_p$, $\alpha > 0, 1 \le p, q \le \infty, \ldots$

Unconditional Bases . . .

- Advantages: $\{\phi_\nu\}_{\nu\in\Gamma} \Longleftrightarrow$ axes of symmetry for the ball in B, e.g., $\|f\|_B < C$.
- "Rotation" into a coordinate system where the norm is "diagonalized" even if the norm is not quadratic.
- Read articles by Donoho as well as Meyer's books!

Cohen, Dahmen, Daubechies, DeVore, Meyer, Petrushev, & Xu

- $BV(\Omega)$ does not have an unconditional basis; but we can say the following:

$$u \in BV(\Omega) \Longrightarrow c[u] \in w\ell^1(\Gamma).$$

- In other words, the sorted wavelet coefficients decay as $O(k^{-1})$.
- This implies that k-term approximation of a BV-function using wavelets is of $O(k^{-1/2})$.

Cohen, Dahmen, Daubechies, DeVore, Meyer, Petrushev, & Xu ...

- Embeddings in the sequence spaces:

$$\ell^1(\Gamma) \subset bv(\Gamma) \subset w\ell^1(\Gamma),$$

  where $bv(\Gamma)$ is a space of vectors consisting of the wavelet coefficients of $BV(\Omega)$ functions and its norm is defined to be the BV norm of the corresponding function.

- This allows wavelet shrinkage on the coefficients.

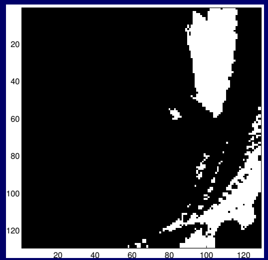
Cohen, Dahmen, Daubechies, DeVore, Meyer, Petrushev, & Xu . . .

- The Haar case by C-De-P-X (1998), and the general wavelet case by Meyer (1998).
- The stronger versions by C-Dah-Dau-De (2000).
- Of course, using other methods such as ridgelets and curvelets, one can get better decay $\Longrightarrow$ Lectures by Donoho and Candès tomorrow

BV for Image Modeling?

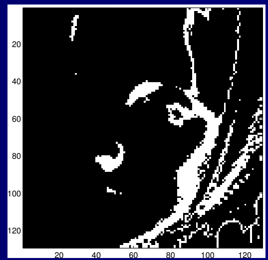
- Study of Gousseau & Morel
- $BV(\Omega)$ may be well adapted for large scale geometric structures
- But natural images are not in $BV(\Omega)$.
- $\because$ Natural images often contain too many small objects and textures $\Longrightarrow$ sum of the length of the perimeters of the level sets may blow up, i.e., $\int_{-\infty}^{\infty} \mathcal{H}^1(\partial E_t[u])\, \mathrm{d}t = \infty$.

BV for Image Modeling?



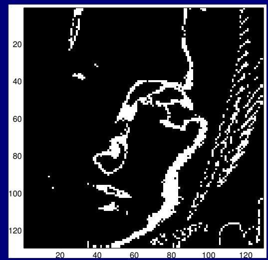
(a)

(23,43)



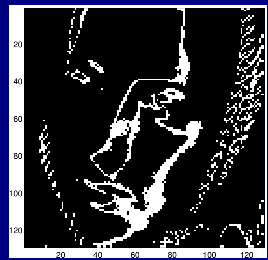
(b)

(43,63)



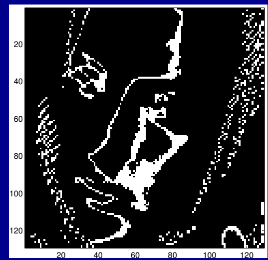
(c)

(63,83)



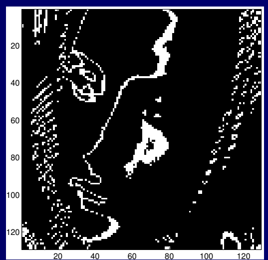
(d)

(83,103)



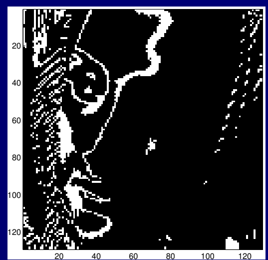
(e)

(103,123)



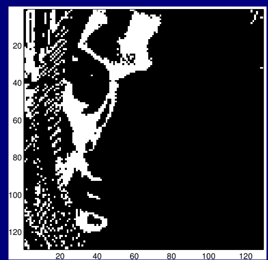
(f)

(123,143)



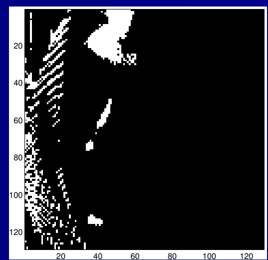
(g)

(143,163)



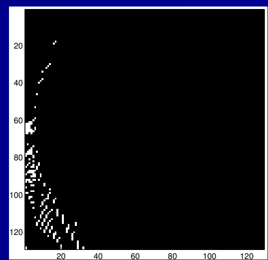
(h)

(163,183)



(i)

(183,203)



(j)

(203,226)

Very Briefly, Meyer, Vese, & Osher (2002)

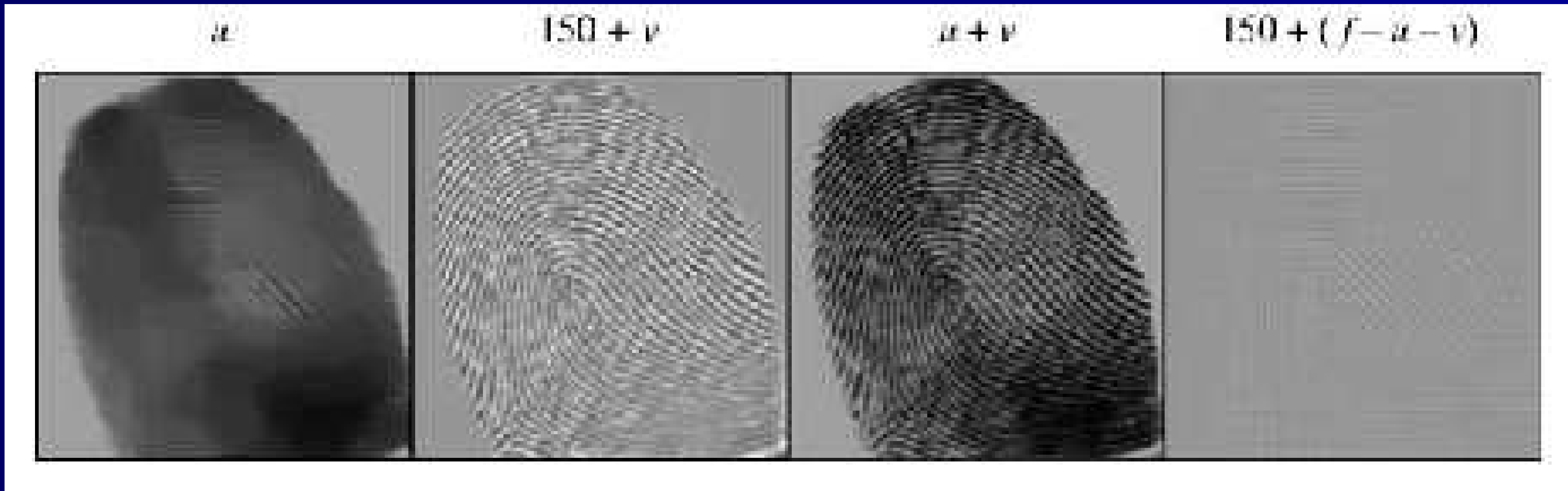
- All the previous models with $\|f-u\|^2_{L^2(\Omega)}$ anticipated WGN for the v component, which are also very much related to statistical estimation methods such as MLE, Bayes, MDL, etc. with prior information on the u component.
- Improve the ROF model by changing the L^2 norm of $v = f - u$ component to

$$J_{MVO}(u) = \|f-u\|_{G(\Omega)} + \lambda |u|_{BV(\Omega)},$$

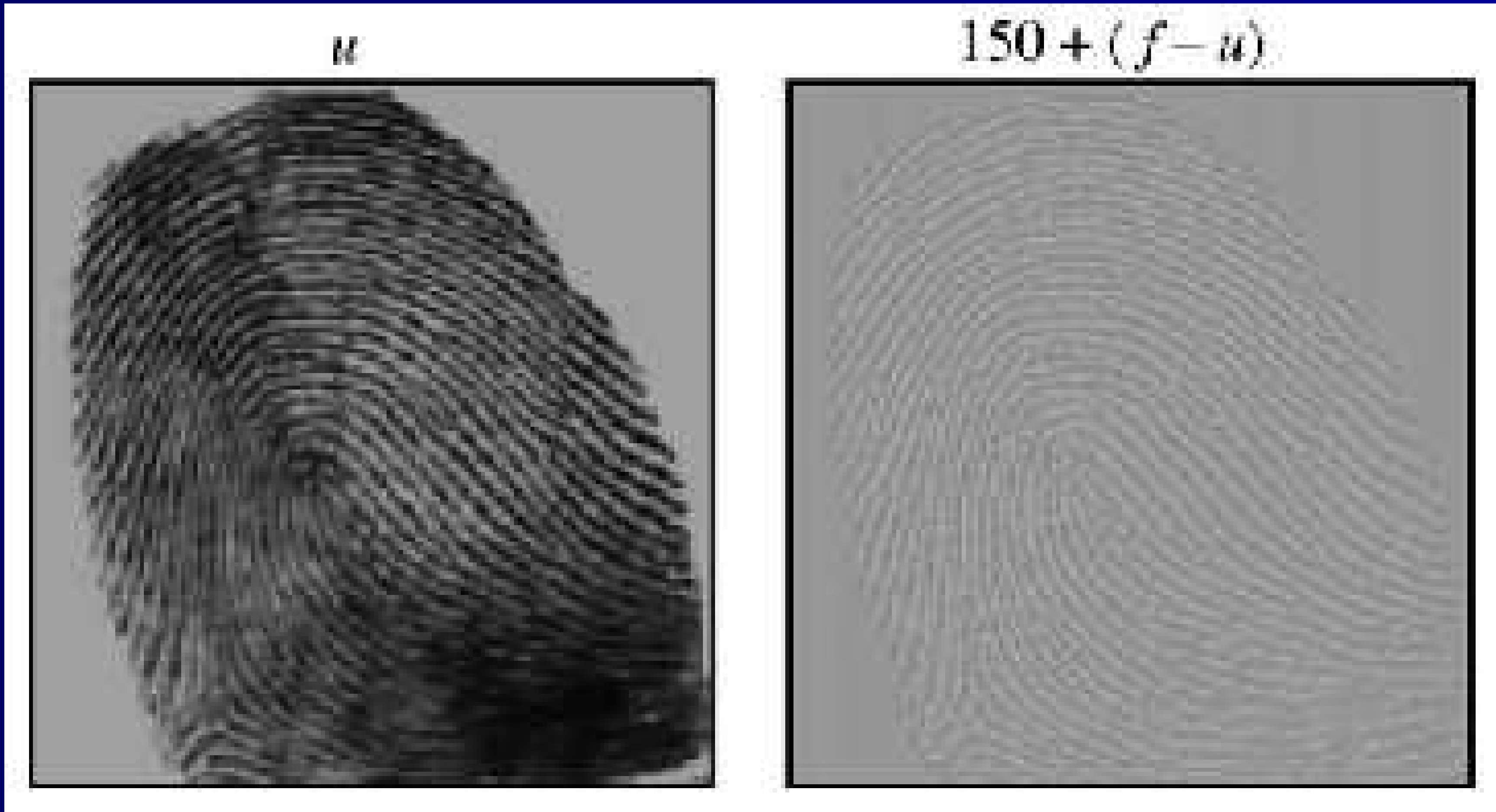
where $G(\Omega)$ is a dual space of $\mathcal{BV}(\Omega) = W^{1,1}(\Omega)$.

Very Briefly, Meyer, Vese, & Osher (2002)

- $G(\Omega)$ contains oscillatory patterns (textures).
- Y. Meyer's book for precise definition of $G(\Omega)$
- Vese-Osher (2003) for numerical algorithm.



Compare with the ROF model . . .



Summary

- Reviewed $u + v$ models
- u component is often modeled by the constraints $\|u\|_B < C$ for some function space B.
- This constraint corresponds to the prior information in Bayesian statistics and MDL formalism, and regularization term in the inverse problem.
- v component is often assumed to be i.i.d. WGN, yielding L^2 fidelity term in the functional to be optimized $\Longrightarrow$ not good for texture

Summary . . .

- Wavelet shrinkage: works very well for $B = \dot{B}_1^{1,1}$, and reasonably well for $B = BV$, and computationally very fast.
- PDE-based approach: works well for $B = W^{1,1}$, more computationally intensive, but allows more flexible modeling for non-L^2 error criterion for v.

My Comments

- Use of function spaces for image modeling:
  - Mathematically sound
  - Can get deep results
  - Extremely hard to find a good one for natural images
- Use of orthonormal bases:
  - Mathematically tractable
  - Good affinity with function spaces
  - Fast algorithms
  - But too restrictive

My Comments . . .

- Use of overcomplete dictionaries
  - Mathematically more challenging
  - Can get better results
  - Can develop fast algorithms
  - Still in the form of linear combinations in most cases

My Comments . . .

- $\|u\|_B < C$ is mathematically great, but restrictive for image modeling
- Need more interaction with stochastic modeling community
- Explore more about wavelet shrinkage after the invertible transform a la Harlan-Claerbout-Rocca
- Yves Meyer: “Sparsity does not open the gate to feature extraction.”
- My reaction: “Sparsity can still open the gate to feature extraction.”

References: Books and Survey Articles

- R. DeVore: “Nonlinear Approximation,” in *Acta Numerica*, Cambridge Univ. Press, 1998.
- D. Donoho, M. Vetterli, R. DeVore, & I. Daubechies: “Data compression and harmonic analysis,” IEEE Trans. Info. Theory, vol.44, pp.2435–2476, 1998.
- S. Jaffard, Y. Meyer, & R. D. Ryan: *Wavelets: Tools for Science & Technology*, SIAM, 2001.
- S. Mallat: *A Wavelet Tour of Signal Processing*, 2nd ed., Academic Press, 1999.
- Y. Meyer: *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, University Lecture Series Vol.22, AMS, 2001.

References: Articles

- B. Bénichou & N. Saito: "Sparsity vs. statistical independence in adaptive signal representations: A case study of the spike process," in *Beyond Wavelets* (G. V. Welland, ed.), Chap.9, pp.225–257, Academic Press, 2003.
- H. Choi & R. Baraniuk: "Wavelet statistical models and Besov spaces," in *Nonlinear Estimation and Classification* (D. Denison ed.), Springer-Verlag, 2003.
- A. Cohen, R. DeVore, P. Petrushev, & H. Xu: "Nonlinear approximation and the space $BV(\mathbb{R}^2)$," *Amer. J. Math.*, vol.121, pp.587–628, 1999.

References: Articles . . .

- A. Cohen, W. Dahmen, I. Daubechies, & R. DeVore: "Harmonic analysis of the space BV," *Revista Mathematica Iberoamericana*, vol.19, pp.235–263, 2003.
- S. Chen, D. L. Donoho, & M. A. Saunders: "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol.20, pp.33-61, 1999.
- R. A. DeVore & B. J. Lucier: "Fast wavelet techniques for near-optimal image processing," *IEEE Military Communications Conference Record*, pp.1129–1135, 1992.
- D. L. Donoho: "Unconditional bases are optimal bases for data compression and for statistical estimation," *Appl. Comput. Harm. Anal.*, vol.1, pp.100–115, 1993.

References: Articles . . .

- D. L. Donoho: “Sparse components of images and optimal atomic decomposition,” *Constr. Approx.*, vol.17, pp.353–382, 2001.
- D. L .Donoho & A. G. Flesia: “Can recent innovations in harmonic analysis ‘explain’ key findings in natural image statistics?” *Network: Comput. Neural Syst.*, vol.12, pp.371–393, 2001.
- Y. Gousseau & J.-M. Morel: “Are natural images of bounded variation?” *SIAM J. Math. Anal.*, vol.33, pp.634–648, 2001.
- W. S. Harlan, J. F. Claerbout, and F. Rocca: “Signal/noise separation and velocity estimation,” *Geophysics*, vol.49, pp.1869–1880, 1984.

References: Articles . . .

- D. Mumford & J. Shah: "Boundary detection by minimizing functionals, I" *IEEE Conf. Computer Vision & Pattern Recognition*, pp.22–26, 1985.
- L. Rudin, S. Osher, & E. Fatemi: "Nonlinear total variation based noise removal algorithms," Physica D, vol.60, pp.259–268, 1992.
- N. Saito: "Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion," in *Wavelets in Geophysics* (E. Foufoula-Georgiou and P. Kumar, eds.), chap. XI, pp.299–324, Academic Press, 1994.

References: Articles . . .

- N. Saito: "Image approximation and modeling via least statistically-dependent bases," Pattern Recognition, vol.34, pp.1765–1784, 2001.
- N. Saito: "The generalized spike process, sparsity, and statistical independence," in *Modern Signal Processing* (D. Rockmore and D. Healy, Jr. eds.), MSRI Publications, vol.46, pp.317–340, Cambridge Univ. Press, 2004.
- L. Vese: "A study in the BV space of a denoising-deblurring variational problem," *Applied Mathematics & Optimization*, vol.44, pp.131–161, 2001.

References: Articles . . .

- L. Vese & S. J. Osher: "Modeling textures with total variation minimization and oscillating patterns in image processing," *J. Sci. Comput.*, vol.19, pp.553–572, 2003.
- S. Watanabe: "Pattern recognition as a quest for minimum entropy," *Pattern Recognition*, vol.13, no.5, pp.381–387, 1981.