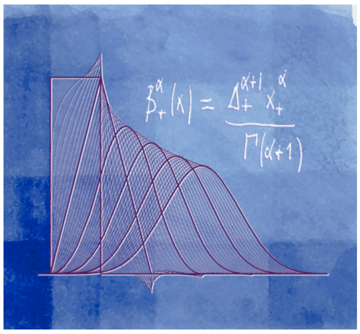
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Sparse stochastic processes

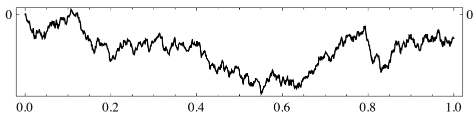
Part 3: Wavelet decomposition of sparse processes and ICA



Prof. Michael Unser, LIB

EPFL Doctoral School EDEE, Course EE-726, Spring 2017

20th Century SP: From AR(1) to coding standards



$$\Phi_s(\omega) = \frac{\sigma_0^2}{|j\omega - \alpha_1|^2}$$

- Generation by analog filtering of Gaussian white noise

$s(t) = (\rho_{\alpha_1} * w)(t)$ with $\rho_{\alpha_1}(t) = \mathcal{F}^{-1}\left\{\frac{1}{j\omega - \alpha_1}\right\} = \mathbb{1}_+(t)e^{\alpha_1 t}$

$$\Leftrightarrow \quad s = (\mathrm{D} - \alpha_1 \mathrm{I})^{-1} w$$

- Discrete AR(1) and linear prediction ☒ **Linear Predictive Coding**

$s[k] - e^{\alpha_1} s[k-1] = w_{\mathrm{d}}[k]$ is i.i.d. Gaussian for $k \in \mathbb{Z}$

- Optimality of sinusoidal transforms ☒ **Transform coding**

DCT is asymptotically equivalent to Karhunen-Loève transform (KLT)

Coding standards

Linear Predictive Coding

Transform Coding

DCT vs. **Wavelets**

Joint Photographic Experts Group

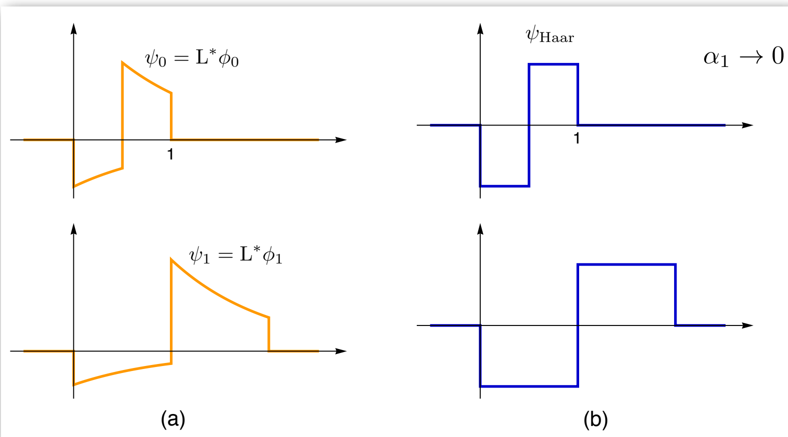
Wavelets and CAR(1) processes

Stability: $\mathrm{Re}(\alpha_1) < 0$

Innovation model: $\mathrm{L}s = w \quad \Leftrightarrow \quad s = \mathrm{L}^{-1}w$ with $\mathrm{L} = (\mathrm{D} - \alpha_1 \mathrm{I})$

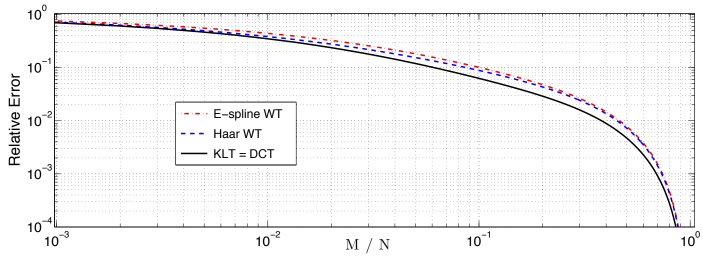
Operator-like wavelet: $\psi_i = \mathrm{L}^* \phi_i$ with ϕ_i: smoothing kernel

(Khalidov-U., 2006)



Gaussian vs. non-Gaussian AR(1)

M term approximation



(a): Gaussian

Correlation: $e^{\alpha_1} = 0.9$

DCT/KLT era:

(Gray, 1972;
Ahmed-Rao, 1975
U., 1984)

Wavelet era:

(Mallat, 1989;
Donoho, 1992)

Statistical, spline-based justification of **jpeg2000**?

OUTLINE

- Introduction: M-term approximation ✔
- *From legos to wavelets*
- Decoupling and sampling of CAR(1) processes
  - Generalized innovation model
  - Discrete innovation model
  - Transform-domain statistics for SαS processes
- Independent component analysis (ICA)
  - Mutual information
  - Comparison of transforms

Journal of Microscopy

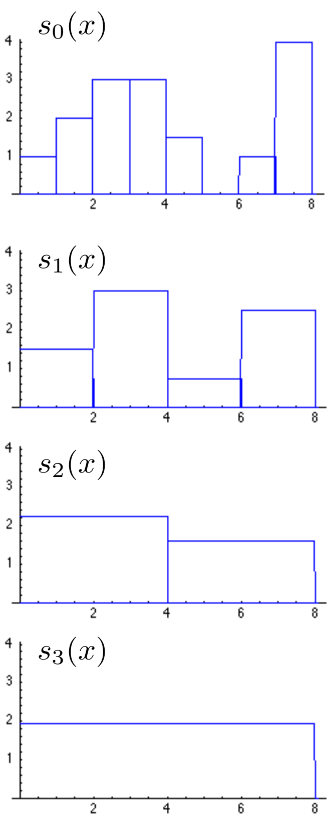
Journal of Microscopy, Vol. 255, Issue 3 2014, pp. 123–127
Received 6 March 2014; accepted 3 June 2014

Wavelets: on the virtues and applications of the mathematical microscope

MICHAEL UNSER
Biomedical Imaging Group, EPFL, Lausanne, Switzerland

Key words. Deconvolution, denoising, wavelets, image analysis, multiresolution, computational imaging.

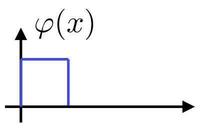
Wavelets: Haar transform revisited



Signal representation

$$s_0(x) = \sum_{k \in \mathbb{Z}} c[k] \varphi(x - k)$$

Scaling function



Multi-scale signal representation

$$s_i(x) = \sum_{k \in \mathbb{Z}} c_i[k] \varphi_{i,k}(x)$$

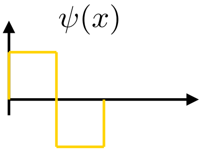
Multi-scale basis functions

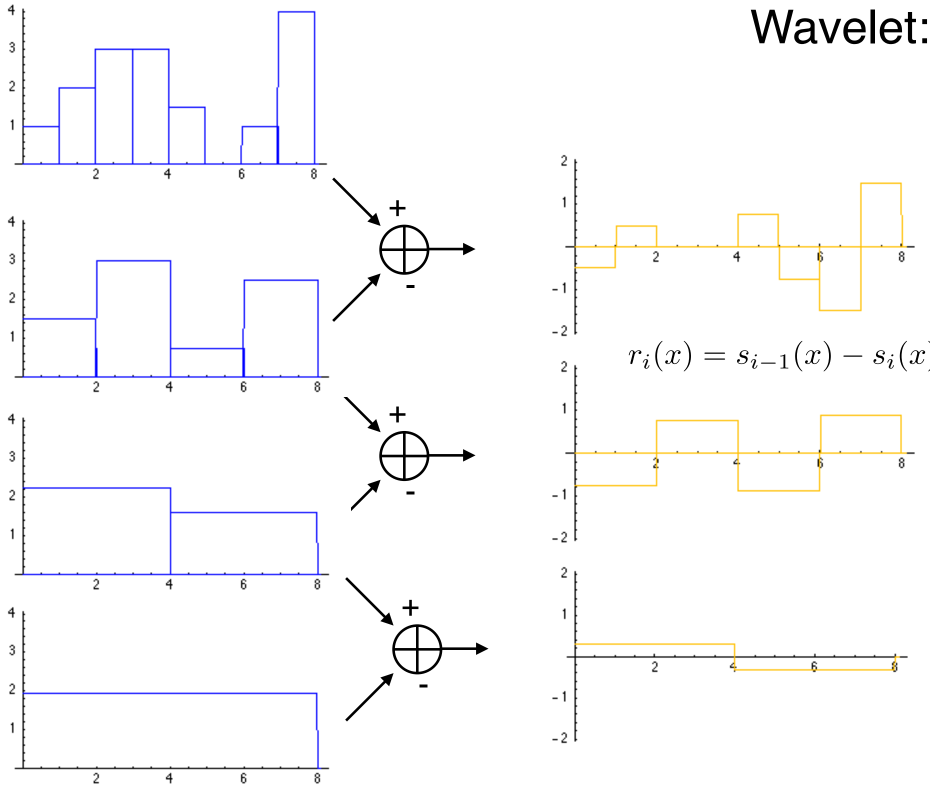
$$\varphi_{i,k}(x) = \varphi\left(\frac{x - 2^i k}{2^i}\right)$$

Wavelets: Haar transform revisited

Wavelet:

$\psi(x)$





$$r_i(x) = s_{i-1}(x) - s_i(x)$$

Wavelets: Haar transform revisited

$$r_1(x) = \sum_k w_1[k]\psi_{1,k}$$

$$+$$

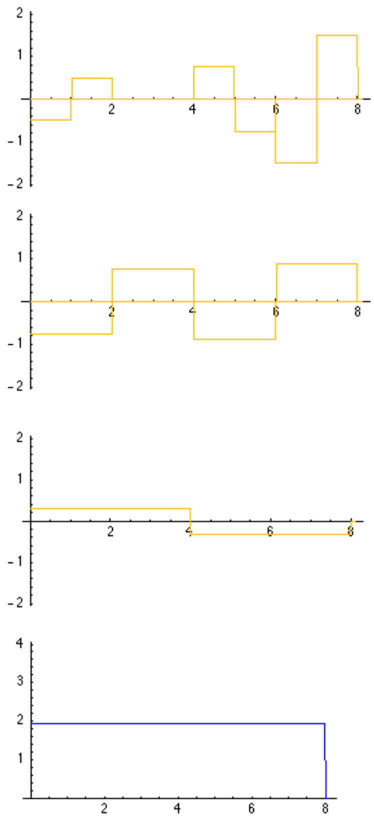
$$r_2(x) = \sum_k w_2[k]\psi_{2,k}$$

$$+$$

$$r_3(x) = \sum_k w_3[k]\psi_{3,k}$$

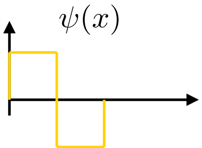
$$+$$

$$s_3(x) = \sum_k c_3[k]\varphi_{3,k}$$

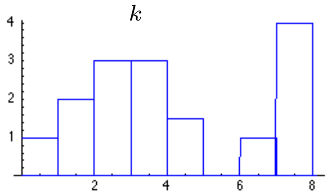


Wavelet:

$\psi(x)$



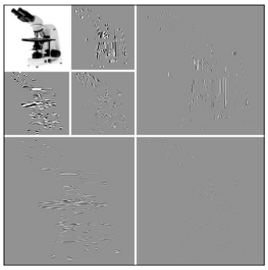
$$= \quad s(x) = \sum_k c[k]\varphi(x-k)$$



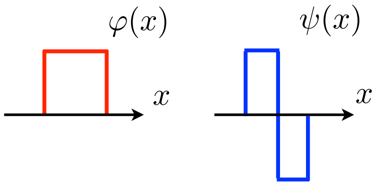
Haar wavelet and 2D basis functions

$$f(x,y) = \sum_{i,\boldsymbol{k}} w_{i,\boldsymbol{k}}\, \psi_{i,\boldsymbol{k}}(x,y)$$

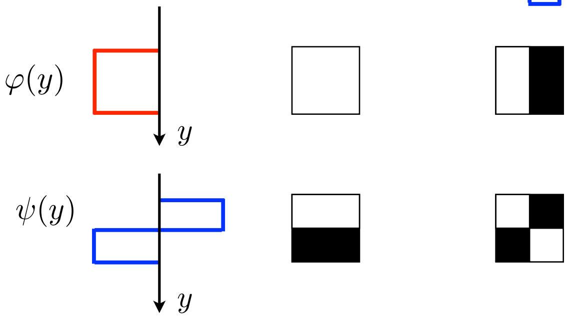
Expansion coefficients



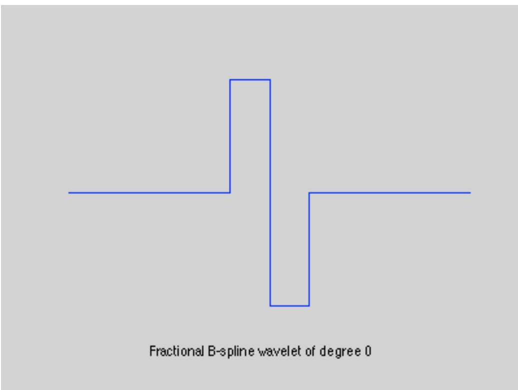
Tensor-product basis functions

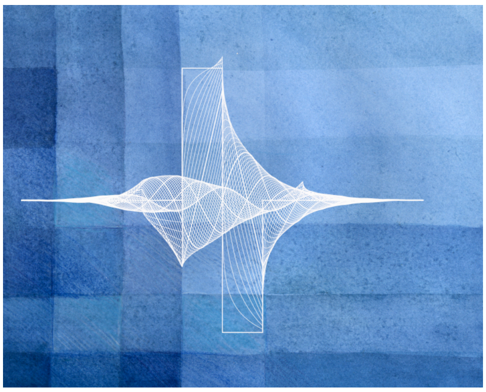


$$\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} s \\ d \end{bmatrix}$$



Beyond legos: Fractional B-spline wavelets





(Unser & Blu, *SIAM Rev,* 2000)

- Remarkable property

Each of these wavelets generates a Riesz basis of $L_2(\mathbb{R})$

$$\psi_+^\alpha(x/2) = \sum_{k\in\mathbb{Z}} \frac{(-1)^k}{2^\alpha} \sum_{n\in\mathbb{N}} \binom{\alpha+1}{n} \beta_*^{2\alpha+1}(n+k-1) \frac{\Delta_+^{\alpha+1}(x-k)_+^\alpha}{\Gamma(\alpha+1)}$$

Only known wavelet bases that have an explicit time-domain formula !

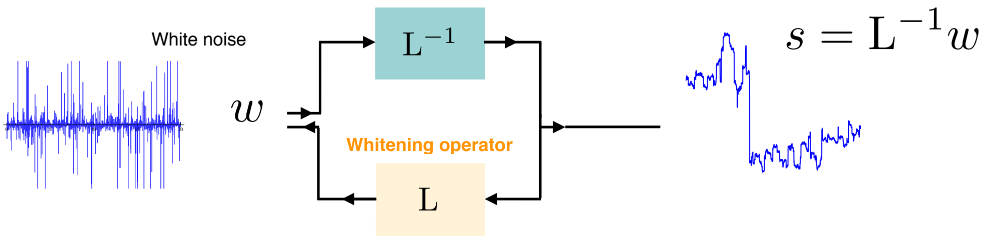
Sparse processes and generalized innovation model

Generic test function $\varphi \in \mathcal{S}$ plays the role of index variable

Formal characterization

Lévy process: $\mathrm{L} = \mathrm{D}$

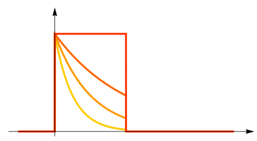
$$\mathbb{E}\{e^{j\langle w,\varphi\rangle}\} = \widehat{\mathscr{P}}_w(\varphi) \quad = e^{-\|\varphi\|_{L_\alpha}^{\alpha}} \text{ (S}\alpha\text{S)}$$



Observation: $X = \langle w, \varphi\rangle$ $\qquad Y = \langle s, \psi\rangle = \langle w, \mathrm{L}^{-1*}\psi\rangle$

$$\Rightarrow \quad \hat{p}_X(\omega) = \mathbb{E}\{e^{j\omega X}\} = \widehat{\mathscr{P}}_w(\omega\varphi) \quad = e^{-|s_0\omega|^{\alpha}} \text{ with } s_0 = \|\varphi\|_{L_\alpha}$$

Sampling and decoupling of CAR(1) processes



Exponential B-spline: $\beta_{\alpha_1}(x) = \mathbb{1}_{[0,1)}(x)e^{\alpha_1 x}$

Discrete operator: $u(x) = \Delta_{\alpha_1} s(x) = s(x) - e^{\alpha_1} s(x-1)$

$$= \big(\beta_{\alpha_1} * w\big)(x)$$

$$s_0 = \|\beta_{\alpha_1}\|_{L_\alpha}$$

- Sampling: AR(1) process

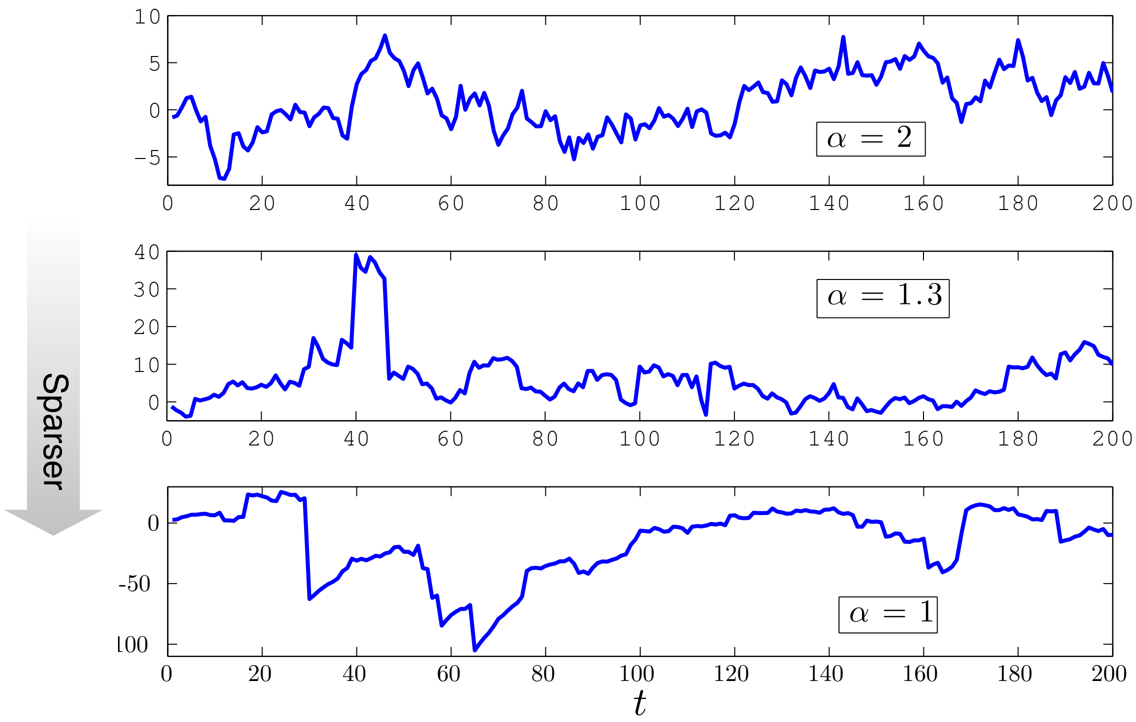
$s[k] = s(x)|_{x=k}$ such that $\boxed{s[k] = e^{\alpha_1} s[k-1] + u[k]}$

SαS

$u[k] = \langle w, \beta^{\vee}_{\alpha_1}(\cdot - k)\rangle \quad \Rightarrow \quad u[\cdot]$ are i.i.d. infinite divisible
with **modified Lévy exponent**

$$f_{\beta^{\vee}_{\alpha_1}}(\omega) = \int_{\mathbb{R}} f\big(\omega\beta_{\alpha_1}(-x)\big)\mathrm{d}x \quad = -s_0|\omega|^{\alpha}$$

= prediction error of **LPC** (perfect decoupling)



Examples of SαS AR(1) processes

Independent component analysis for SαS AR(1)

- Discrete innovation model for AR(1)

$$\mathbf{u} = (U_1, \dots, U_N) = \mathbf{L}\mathbf{s} \quad \text{with} \quad U_n = \langle w, \beta_{\alpha_1}(n - \cdot)\rangle \ \text{i.i.d. S}\alpha\text{S}$$

Differential entropy: $H_U = -\int_{\mathbb{R}} p_U(x) \log p_U(x) \mathrm{d}x$

- Orthogonal transform and mutual information

$$\mathbf{y} = \mathbf{T}\mathbf{s} \quad \text{with} \quad \mathbf{T}^T\mathbf{T} = \mathbf{I}$$

$$I(Y_1, \dots, Y_N) = \left(\sum_{n=1}^{N} H_{Y_n}\right) - \overbrace{H_{(Y_1:Y_N)}}^{H_{(S_1:S_N)} = N H_U}$$

- ICA: minimum mutual information

$\mathbf{z} = (Z_1, \dots, Z_N) = \mathbf{T}_{\mathrm{ICA}}\mathbf{x}$: *"most independent"* representation

Stability w.r.t. to linear combinations

- Characteristic function of SαS random variable

$$\hat{p}_U(\omega) = e^{-|s_0\omega|^\alpha}$$

- Linear combination of i.i.d. SαS random variables

$$Z = \sum_{n=1}^{N} a_n U_n = \mathbf{a}^T\mathbf{u} \quad \Leftrightarrow \quad \hat{p}_Z(\omega) = \prod_{n=1}^{N} \hat{p}_U(a_n\omega)$$

$$= \exp\left(-\sum_{n=1}^{N} |a_n s_0 \omega|^\alpha\right) = e^{-|r_0\,\omega|^\alpha}$$

$\Rightarrow \quad Z$ is SαS with parameter $r_0 = s_0 \left(\sum_{n=1} |a_n|^\alpha\right)^{1/\alpha} = s_0\|\mathbf{a}\|_\alpha$

ICA: From entropy to ℓ_α-norms

Orthogonality constraint: $\mathbf{T}^T\mathbf{T} = \mathbf{I}$ $\qquad$ $\mathbf{A} = \mathbf{T}\mathbf{L}^{-1} = [\mathbf{a}_1 \cdots \mathbf{a}_N]^T$

- Min-entropy solution: $\displaystyle\sum_{n=1}^{N} H_{Y_n} = I(Y_1, \dots, Y_N) - \underbrace{H_{(Y_1:Y_N)}}_{\text{Const}}$

- Special case of SαS processes: $f(\omega) = -|s_0\omega|^\alpha$

$$\hat{p}_{Y_n}(\omega) = \exp\left(\sum_{m=1}^{N} -|s_0 a_{n,m}\omega|^\alpha\right) = \mathrm{e}^{-|s_n\omega|^\alpha}$$

$\Longrightarrow \quad Y_n$ is SαS with dispersion $\quad s_n = s_0\|\mathbf{a}_n\|_{\ell_\alpha} = s_0 \left(\sum_{m=1}^{N} |a_{n,m}|^\alpha\right)^{\frac{1}{\alpha}}$

$$\Longrightarrow \quad H_{Y_n} = H_U - \log\|\mathbf{a}_n\|_{\ell_\alpha}$$

$$\Longrightarrow \quad I(\mathbf{T};\alpha) = -\sum_{n=1}^{N} \log\|\mathbf{a}_n\|_{\ell_\alpha} = -\sum_{n=1}^{N} \frac{1}{\alpha} \log\left(\sum_{m=1}^{N} \left|[\mathbf{T}\mathbf{L}^{-1}]_{n,m}\right|^\alpha\right)$$

ICA for discrete SαS processes

- Discrete innovation model

$$\mathbf{s} = \mathbf{L}^{-1}\mathbf{u} \quad \text{with} \quad \mathbf{u} = (U_1, \dots, U_N) \text{ i.i.d. S}\alpha\text{S}$$

- Orthogonal transform: $\mathbf{y} = \mathbf{T}\mathbf{s} = \mathbf{T}\mathbf{L}^{-1}\mathbf{u}$

- Minimization of mutual information

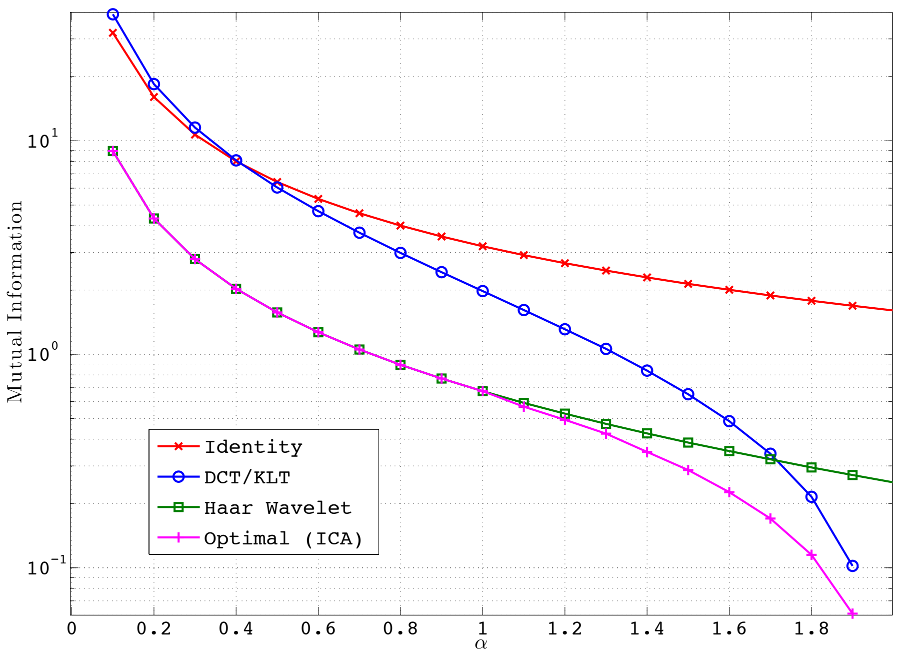
$$\arg\min_{\mathbf{T}} \left\{ -\sum_{n=1}^{N} \frac{1}{\alpha} \log\left(\sum_{m=1}^{N} [\mathbf{T}\mathbf{L}^{-1}]^{\alpha}_{n,m} \right) \right\} \quad \text{s.t. } \mathbf{T}^T\mathbf{T} = \mathbf{I}$$

Iterative gradient-based algorithm

- Update using analytical computation of gradient
- Projection on space of orthogonal matrices

Orthogonal expansion of a stable Lévy process

$\alpha_1 = 0$



sparser

Gaussian

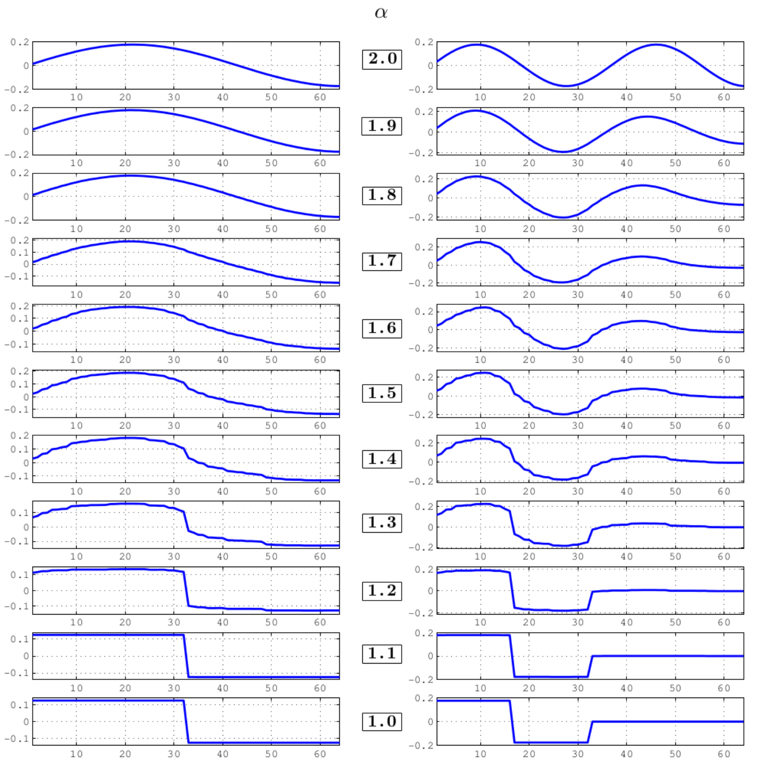
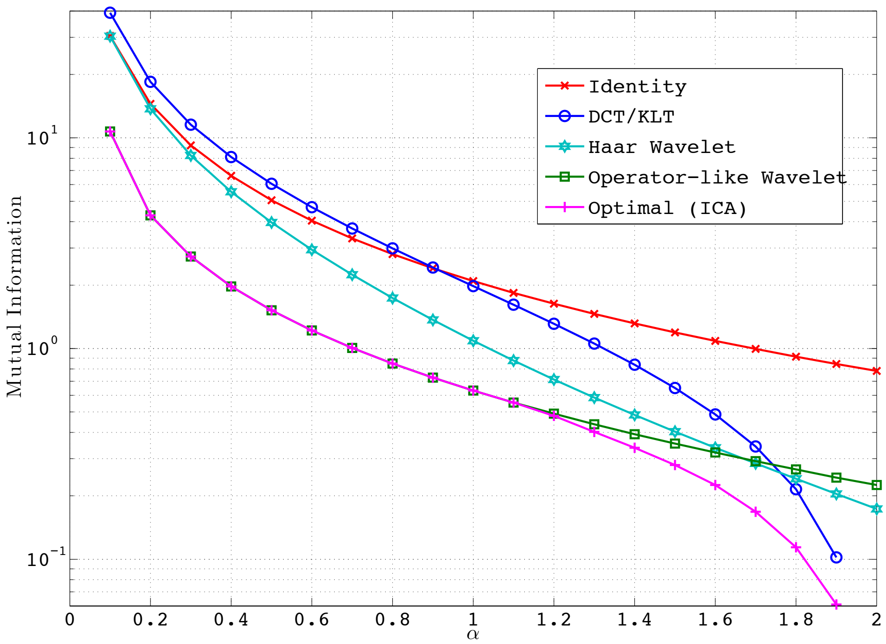


Fig. 6: Two rows of the optimal $\mathbf{H}$ (ICA) for $\alpha = 2$ down to 1 when $N = 64$. In each row, we see the evolution from sinusoidal waves to Haar wavelets by increasing the sparsity of the underlying innovation process.

Orthogonal expansion of a stable AR(1) process



sparser

Gaussian

$e^{\alpha_1} = 0.9$, $M = 64$

(Pad-U., ICASSP'13, ArXiv)

Addendum: Compressibility of sparse processes

- Stochastic ODE driven by periodic SαS noise

$$\mathrm{L}s = w_\alpha \quad \Rightarrow \quad s = \mathrm{L}^{-1} w_\alpha \qquad \mathrm{L} = \frac{\mathrm{d}^n}{\mathrm{d}t^n} + a_{n-1}\frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}} + \cdots + a_1 \frac{\mathrm{d}}{\mathrm{d}t} + a_0 \mathrm{I}$$

$s \in B^{n+1/\alpha-1}_{\alpha,\infty}(\mathbb{T})$ with probability 1

Sparsity index: $\alpha \in (0, 2]$

$B^{\tau}_{p,q}(\mathbb{T})$: Besov space over the circle

- M-term approximation (in matched wavelet basis)

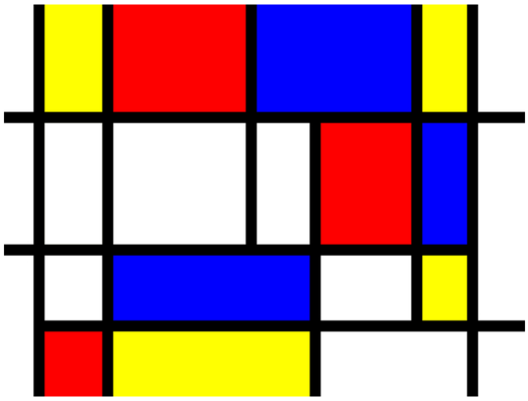
$\|s - s_M\|_{L_2(\mathbb{T})} = O(M^{-\tau_0})$ with $\tau_0 = n + \frac{1}{\alpha} - 1 - \epsilon$

(Fageot-U.-Ward, ACHA 2016)

CONCLUSION

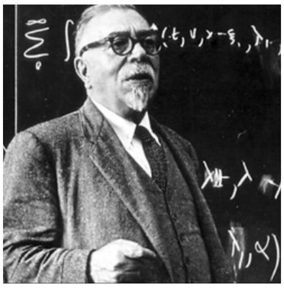
- Unifying continuous-domain stochastic model
  - Backward compatibility with classical Gaussian theory
  - Operator-based formulation: Lévy-driven SDEs or SPDEs
  - **Gaussian** vs. **sparse** (generalized Poisson, student, SαS)
- Decoupling and regularization
  - Sparsification via “operator-like” behavior (whitening)
  - Specific family of id potential functions (typ., non-convex)
- Conceptual framework for sparse signal recovery
  - New statistically-founded sparsity priors
  - Derivation of optimal estimators (MAP, MMSE)
  - Principled approach for the development of novel algorithms
  - Implementation of MMSE solution (belief propagation)

Gaussian vs. Sparse

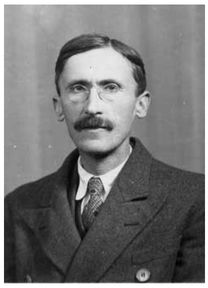


Fourier analysis

Wavelet analysis



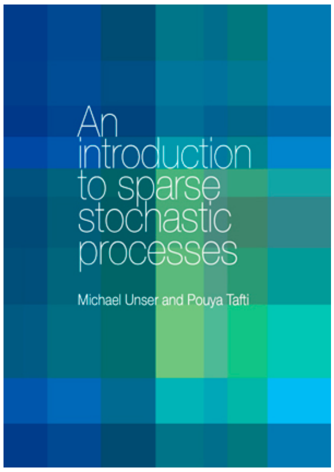
Norbert Wiener



Paul Lévy

References

CAMBRIDGE UNIVERSITY PRESS



http://www.sparseprocesses.org

- M. Unser and P.D. Tafti, *An Introduction to Sparse Stochastic Processes*, e-book at `http://www.sparseprocesses.org`.

- M. Unser, P.D. Tafti, and Q. Sun, "A unified formulation of Gaussian vs. sparse stochastic processes—Part I: Continuous-domain theory," *IEEE Trans. Information Theory*, available at http://arxiv.org/abs/1
- P. Pad, M. Unser, "On the Optimality of Operator-Like Wavelets for Sparse AR(1) Processes," Proc. IEEE Int. Conf. Acoust. Speech Sig. Proc. (ICASSP'13), Vancouver, Canada, May, 2013.
- I. Khalidov and M. Unser, "From Differential Equations to the Construction of New Wavelet-Like Bases," *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1256–1267, April 2006.

Acknowledgments

Many thanks to

- Dr. Pouya Tafti
- **Pedram Pad**
- Prof. Arash Amini
- Dr. John-Paul Ward
- Dr. Ulugbek Kamilov
- Emrah Bostan
- Dr. Masih Nilchain
- Julien Fageot

- **Members of EPFL's Biomedical Imaging Group**

erc

- Preprints and demos: http://bigwww.epfl.ch/