Numerical approximation of spatially varying blur operators.

P. Weiss and P. Escande
S. Anthoine, C. Chaux, C. Mélot

CNRS - ITAV - IMT

Winter School on Computational Harmonic Analysis with Applications to Signal and Image Processing, CIRM, 23/10/2014

Disclaimer

- ✗ We are newcomers to the field of computational harmonic analysis and we are still living in the previous millenium!
- ✔ Do not hesitate to ask further questions to the pillars of time-frequency analysis present in the room.
- ✔ A rich and open topic!

Main references for this presentation

Beylkin, G. (1992). "On the representation of operators in bases of compactly supported wavelets". In: *SIAM Journal on Numerical Analysis* 29.6, pp. 1716–1740.

Beylkin, G., R. Coifman, and V. Rokhlin (1991). "Fast Wavelet Transform and Numerical Algorithm". In: *Commun. Pure and Applied Math.* 44, pp. 141–183.

Cohen, A. (2003). *Numerical analysis of wavelet methods*. Vol. 32. Elsevier.

Cohen, Albert et al. (2003). "Harmonic analysis of the space BV". In: *Revista Matematica Iberoamericana* 19.1, pp. 235–263.

Coifman, R. and Y. Meyer (1997). *Wavelets, Calderón-Zygmund and multilinear operators*. Vol. 48.

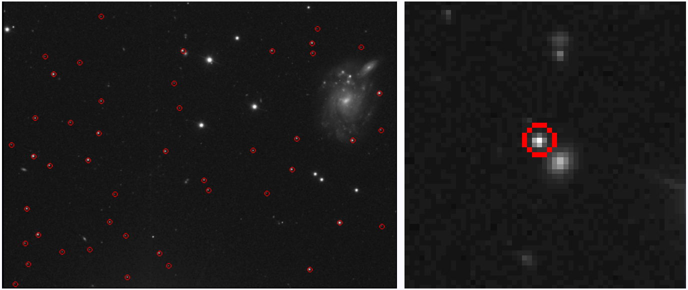
Escande, P. and P. Weiss (2014). "Numerical Computation of Spatially Varying Blur Operators A Review of Existing Approaches with a New One". In: *arXiv preprint arXiv:1404.1023*.

Wahba, G. (1990). *Spline models for observational data*. Vol. 59. Siam.

Outline

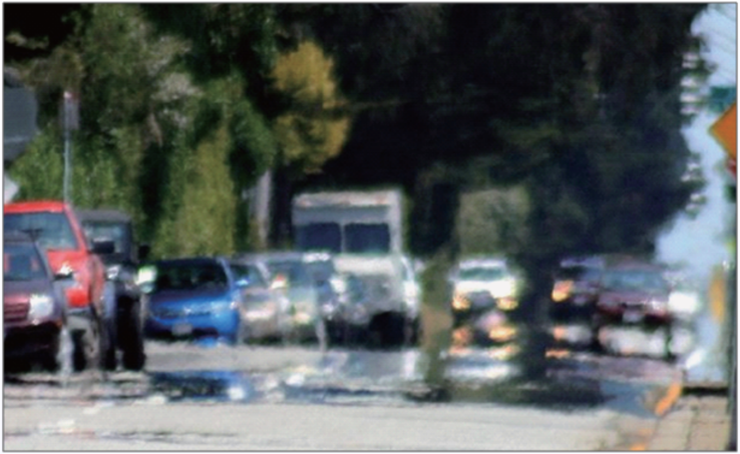
- **Part I:** Spatially varying blur operators.
  1. Examples and definitions.
  2. Challenges.
  3. Existing computational methods.
- **Part II:** Sparse representations in wavelet bases.
  1. Meyer's and Beylkin-Coifman-Rokhlin's results.
  2. Finding good sparsity patterns.
  3. Numerical results.
- **Part III:** Estimation/interpolation.
  1. An inverse problem on operators.
  2. A tractable numerical scheme using wavelets.
  3. Numerical results.

Part I: Spatially varying blur operators.

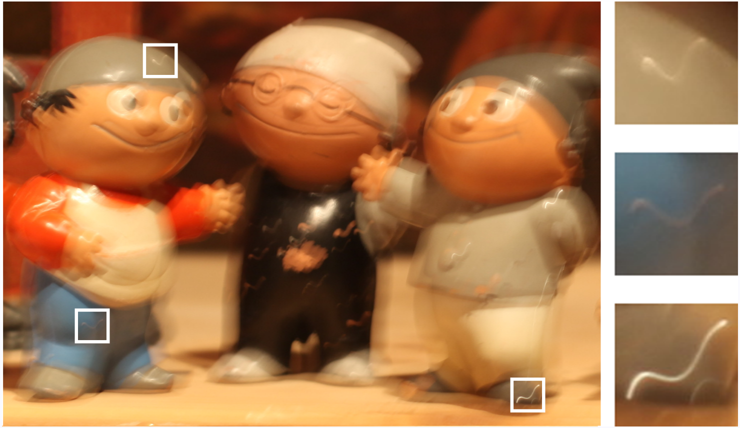


STARS IN ASTRONOMY. HOW TO IMPROVE THE RESOLUTION OF GALAXIES?
SLOAN DIGITAL SKY SURVEY http://www.sdss.org/.

Motivating examples - Imaging under turbulence (2D)

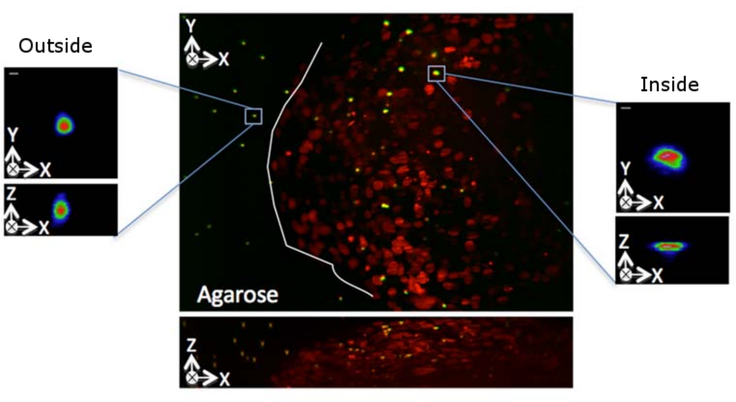


Variations of refractive indices due to air heat variability.



A real camera shake. How to remove blur?
(by courtesy of M. Hirsch, ICCV 2011)

Motivating examples - Microscopy (3D)



Fluorescence microscopy. Micro-beads are inserted in the sample. (Biological images we are working with).

Other potential applications

- ODFM (orthogonal frequency-division multiplexing) systems (see Hans Feichtinger).
- Geophysics and seismic data analysis (see Caroline Chaux).
- Solutions of PDE's $\mathrm{div}(c\nabla u) = f$ (see Philipp Grohs).
- ...

A number of challenges...

A standard inverse problem?

In all the imaging examples, we observe:

$$u_0 = Hu + b$$

where H is a blur operator, b is some noise and u is a clean image.

What makes it more difficult?

1. Few studies for such operators (compared to the huge amount dedicated to convolutions).
2. Images are large vectors. How to store an operator?
3. How to numerically evaluate products Hu and $H^T u$?
4. In practice, H is partially or completely unknown. How to retrieve the operator H?

Notation

- We work on $\Omega = [0,1]^d$, $d \in \mathbb{N}$ is the space dimension.
- A grayscale image $u : \Omega \to \mathbb{R}$ is viewed as an element of $L^2(\Omega)$.
- Operators are in capital letters (e.g. H), functions in lower-case (e.g. u), bold is used for matrices (e.g. **H**).

Spatially varying blur operators

In this talk, we model the spatially varying blur operator H as a linear integral operator:

$$Hu(x) = \int_\Omega K(x,y)u(y)\,dy$$

The function $K : \Omega \times \Omega \to \mathbb{R}$ is called kernel.

Important note

By the Schwartz (Laurent) kernel theorem, H can be *any* linear operator if K is a generalized function.

Definition of the PSF

The point spread function or impulse response at point $y \in \Omega$ is defined by

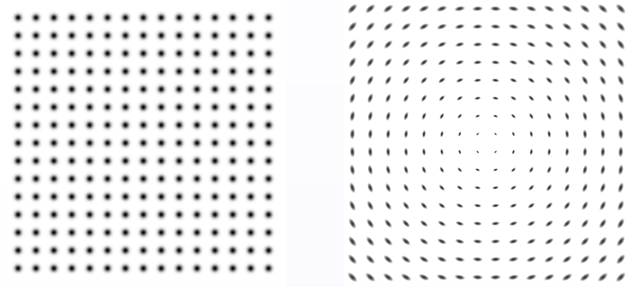
$$H\delta_y = K(\cdot, y), \qquad \text{(if } K \text{ is continuous)}$$

where δ_y denotes the Dirac at $y \in \Omega$.

An example

Assume that $K(x, y) = k(x - y)$.
Then H is a convolution operator.
The PSF at y is the function $k(\cdot - y)$.

Some PSF examples - Synthetic (2D)



Examples of 2D PSF fields (H applied to the dirac comb). Left: convolution operator (stationary). Right: spatially varying operator (unstationary).

What is specific to blurring operators?

Properties of blurring operators

- Boundedness of the operator:

$$H : L^2(\Omega) \to L^2(\Omega)$$

  is a bounded operator (the energy stays finite).
- Spatial decay: In most systems, PSFs satisfy:

$$|K(x,y)| \leq \frac{C}{\|x-y\|_2^\alpha}$$

  for a certain $\alpha > 0$.
  Examples: Motion blurs, Gaussian blurs, Airy patterns.

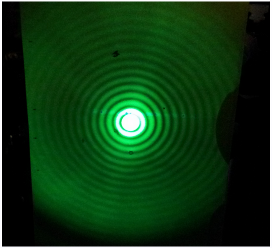
The Airy pattern

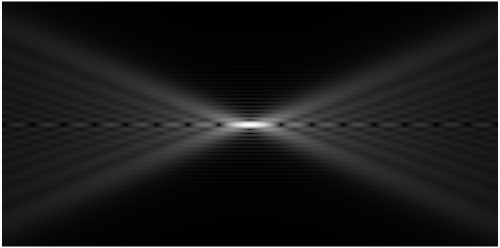
The most “standard” PSF is the Airy pattern (diffraction of light in a circular pinhole):

$$k(x) \simeq I_0 \left(\frac{2J_1(\|x\|_2)}{\|x\|_2} \right)^2,$$

where J_1 is the Bessel function of the first kind.

What is specific to blurring operators?





Axial and longitudinal views of an Airy pattern.

3D RENDERING OF MICROBEADS.

Motivating examples - Microscopy (3D)

3D rendering of microbeads.

What is specific to blurring operators?

More properties of blurring operators

- PSF smoothness: $\forall y \in \Omega$, $x \mapsto K(x,y)$ is C^M and
$$|\partial_x^m K(x,y)| \leq \frac{C}{\|x-y\|_2^\beta}, \ \forall m \leq M.$$
for some $\beta > 0$.
- PSF varies smoothly:
$\forall x \in \Omega$, $y \mapsto K(x,y)$ is C^M and
$$|\partial_x^m K(x,y)| \leq \frac{C}{\|x-y\|_2^\gamma}, \ \forall m \leq M.$$
for some $\gamma > 0$.

Other potential hypotheses (not assumed in this talk)

- Positivity: $K(x,y) \geq 0, \forall x, y$ (not necessarily true, e.g. echography).
- Mass conservation: $\forall y \in \Omega$, $\int_\Omega K(x,y)\,dx = 1$ (not necessarily true when attenuation occurs, e.g. microscopy).

Discretization

Let $\mathbf{\Omega} = \{k/N\}^d_{1\leq k\leq N}$ denote a Euclidean discretization of Ω.
We can define a discretized operator $\mathbf{H}$ by:

$$\mathbf{H}(i,j) = \frac{1}{N^d} K(x_i, y_j)$$

where $(x_i, y_j) \in \mathbf{\Omega}^2$. By the rectange rule $\forall x_i \in \mathbf{\Omega}$:

$$Hu(x_i) \approx (\mathbf{Hu})(i).$$

Typical sizes

For an image of size 1000×1000, $\mathbf{H}$ contains $10^6 \times 10^6 = 8$ TeraBytes.
For an image of size $1000 \times 1000 \times 1000$, $\mathbf{H}$ contains $10^9 \times 10^9 = 8$ ExaBytes.
The total amount of data of Google is estimated at 10 ExaBytes in 2013.

$\mathbf{H}$ can be viewed either as a $N^d \times N^d$ matrix, or as a
$\underbrace{N \times N \times \ldots \times N}_{d \text{ times}} \times \underbrace{N \times N \times \ldots \times N}_{d \text{ times}}$ array.

The naive numerical approach

Complexity of a matrix vector product

A matrix-vector multiplication is an $O(N^{2d})$ algorithm.
With a 1GHz computer (if the matrix was storable in RAM), a matrix-vector product would take:

- ✗ 18 minutes for a 1000×1000 image.
- ✗ 33 years for a $1000 \times 1000 \times 1000$ image.

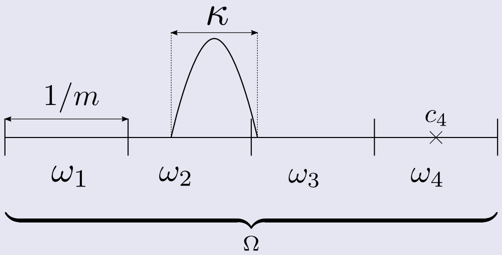
Bounded supports of PSFs help?

- ✔ Might work for very specific applications (astronomy).
- ✔ 0.4 seconds 20×20 PSF and 1000×1000 images.
- ✗ 2 hours for $20 \times 20 \times 20$ PSFs and $1000 \times 1000 \times 1000$ images.
- ✗ If the diameter of the largest PSF is $\kappa \in (0, 1]$, matrix $\mathbf{H}$ contains $O(\kappa^d N^{2d})$ non-zero elements.
- ✗ Whenever super-resolution is targeted, this approach is doomed.

The mainstream approach: piecewise convolutions

Main idea: approximate H by an operator H_m defined by the following process:

1. Partition Ω in squares $\omega_1, \ldots, \omega_m$.
2. On each subdomain ω_k, approximate the blur by a spatially invariant operator.



Theorem (A complexity result (1D) Escande and Weiss 2014)

Let K denote a Lipschitz kernel that is not a convolution. Then:

- *The complexity of an evaluation $\mathbf{H}_m\mathbf{u}$ using FFTs is*
$$(N + \kappa N m) \log (N/m + \kappa N).$$
- *For $1 \ll m < N$, there exists constants $0 < c_1 \leq c_2$ s.t.*
$$\|\mathbf{H} - \mathbf{H}_m\|_{2\to 2} \leq \frac{c_2}{m}$$
$$\|\mathbf{H} - \mathbf{H}_m\|_{2\to 2} \geq \frac{c_1}{m}$$
- *For sufficiently large N and sufficiently small $\epsilon > 0$ the number of operations necessary to obtain $\|\mathbf{H} - \mathbf{H}_m\|_{2\to 2} \leq \epsilon$ is proportional to*
$$\frac{L\kappa N \log(\kappa N)}{\epsilon}.$$

Definition of the spectral norm $\|\mathbf{H}\|_{2\to 2} := \sup_{\|\mathbf{u}\|_2 \leq 1} \|\mathbf{H}\mathbf{u}\|_2$.

Pros and cons

- ✔ Very simple conceptually.
- ✔ Simple to implement with FFTs.
- ✔ More than 100 papers using this technique (or slightly modified).
- ✗ The method is insensitive to higher degrees of regularity of the kernel.
- ✗ The dependency in ϵ is not appealing.

Part II: Sparse representations in wavelets bases.

Notation (1D)

We work on the interval $\Omega = [0, 1]$.
The Sobolev space $W^{M,p}(\Omega)$ is defined by

$$W^{M,p}(\Omega) = \{f^{(m)} \in L^p(\Omega),\ \forall\ 0 \leq m \leq M\}.$$

We define the semi-norm $|f|_{W^{M,p}} = \|f^{(M)}\|_{L^p}$.
Let ϕ and ψ denote the scaling function and mother wavelet.
We assume that $\psi \in W^{M,\infty}$ has M vanishing moments:

$$\forall 0 \leq m \leq M, \quad \int_{[0,1]} t^m \psi(t)\, dt = 0.$$

Every $u \in L^2(\Omega)$ can be decomposed as

$$u = \sum_{j \geq 0} \sum_{0 \leq m < 2^j} \langle u, \psi_{j,m} \rangle \psi_{j,m} + \langle u, \phi \rangle \phi.$$

where (apart from the boundaries of $[0, 1]$)

$$\psi_{j,m} = 2^{j/2} \psi(2^j \cdot - m).$$

Shorthand notation

$$u = \sum_{\lambda} \langle u, \psi_\lambda \rangle \psi_\lambda.$$

with $\lambda = (j, m)$, $j \geq 0$, $0 \leq m < 2^j$ and $|\lambda| = j$. The scalar product with ϕ is included in the sum.

Decomposition/Reconstruction operators

We let $\Psi : \ell^2 \to L^2([0,1])$ and $\Psi^* : L^2([0,1]) \to \ell^2$ denote the reconstruction/decomposition transforms:
Given a sequence in $\alpha \in \ell^2$,

$$\Psi u = \sum_{\lambda} \alpha_\lambda \psi_\lambda$$

Given a function $u \in L^2([0,1])$,

$$\Psi^* u = (u_\lambda)_\lambda$$

with

$$u_\lambda = \langle u, \psi_\lambda \rangle.$$

Decomposition of the operator on a wavelet basis

Let $u \in L^2(\Omega)$ and $v = Hu$.

$$
\begin{aligned}
v &= \sum_{\lambda} \langle Hu, \psi_\lambda \rangle \psi_\lambda \\
&= \sum_{\lambda} \left\langle H \left(\sum_{\lambda'} \langle u, \psi_{\lambda'} \rangle \psi_{\lambda'} \right), \psi_\lambda \right\rangle \psi_\lambda \\
&= \sum_{\lambda} \sum_{\lambda'} \langle u, \psi_{\lambda'} \rangle \langle H\psi_{\lambda'}, \psi_\lambda \rangle \psi_\lambda .
\end{aligned}
$$

The action of H is completely described by the (infinite) matrix

$$\Theta = (\theta_{\lambda,\lambda'})_{\lambda,\lambda'} = \left(\langle H\psi_{\lambda'}, \psi_\lambda \rangle\right)_{\lambda,\lambda'} .$$

With these notation

$$H = \Psi \Theta \Psi^* .$$

Definition - (Nonsingular) Calderón-Zygmund operators

An integral operator $H : L^2(\Omega) \to L^2(\Omega)$ with a kernel $K \in W^{M,\infty}(\Omega \times \Omega)$ is a Calderón-Zygmund operator of regularity $M \geq 1$ if

$$|K(x,y)| \leq \frac{C}{\|x-y\|_2^d}$$

and

$$|\partial_x^m K(x,y)| + |\partial_y^m K(x,y)| \leq \frac{C}{\|x-y\|_2^{d+m}}, \ \forall m \leq M.$$

Important notes

The above definition is simplified.
Calderón-Zygmund operators may be singular on the diagonal $x = y$.
For instance, the Hilbert transform corresponds to $K(x,y) = \frac{1}{x-y}$.

Take home message

Our blurring operators are simple Calderón-Zygmund operators.

Theorem (Decrease of $\theta_{\lambda,\lambda'}$ in 1D)

Assume that H belongs to the Calderón-Zygmund class and that the mother wavelet ψ is compactly supported with M vanishing moments. Set $\lambda = (j, m)$ and $\lambda' = (k, n)$. Then

$$|\theta_{\lambda,\lambda'}| \leq C_M 2^{-(M+1/2)|j-k|} \left(\frac{2^{-k} + 2^{-j}}{2^{-k} + 2^{-j} + |2^{-j}m - 2^{-k}n|} \right)^{M+1}$$

where C_M is a constant independent of j, k, n, m.

Take home message

- ✔ The coefficients decrease exponentialy with scales differences $2^{-(M+1/2)|j-k|}$.
- ✔ The coefficients decrease polynomialy with shift differences $\left(\frac{2^{-k}+2^{-j}}{2^{-k}+2^{-j}+|2^{-j}m-2^{-k}n|} \right)^{M+1}$.
- ✔ The kernel regularity M plays a key role.

Polynomial approximation - Annales de l'institut Fourier, Deny-Lions 1954

Let $f \in W^{M,p}([0,1])$. For $1 \leq p \leq +\infty$, $M \in \mathbb{N}^*$ and $I_h \subset [0,1]$ an interval of length h:

$$\inf_{g \in \Pi_{M-1}} \|f - g\|_{L^p(I_h)} \leq C h^M |f|_{W^{M,p}(I_h)}, \tag{1}$$

where C is a constant that depends on M and p only.

Let $I_{j,m} = \mathrm{supp}(\psi_{j,m}) = [2^{-j}(m-1), 2^{-j}(m+1)]$. Assume that $j \leq k$:

$$\begin{aligned}
&|\langle H\psi_{j,m}, \psi_{k,n}\rangle| \\
&= \left| \int_{I_{k,n}} \int_{I_{j,m}} K(x,y)\psi_{j,m}(y)\psi_{k,n}(x)\, dy\, dx \right| \\
&= \left| \int_{I_{j,m}} \int_{I_{k,n}} K(x,y)\psi_{j,m}(y)\psi_{k,n}(x)\, dx\, dy \right| \quad \text{(Fubini)} \\
&= \left| \int_{I_{j,m}} \inf_{g \in \Pi_{M-1}} \int_{I_{k,n}} \left(K(x,y) - g(x)\right) \psi_{j,m}(y)\psi_{k,n}(x)\, dx\, dy \right| \quad \text{(Vanishing moments)} \\
&\leq \int_{I_{j,m}} \inf_{g \in \Pi_{M-1}} \|K(\cdot,y) - g\|_{L^\infty(I_{k,n})} \|\psi_{k,n}\|_{L^1(I_{k,n})} |\psi_{j,m}(y)|\, dy \quad \text{(Hölder)}
\end{aligned}$$

Therefore:

$$
\begin{aligned}
&|\langle H\psi_{j,m}, \psi_{k,n}\rangle| \\
&\lesssim 2^{-kM} \|\psi_{k,n}\|_{L^1(I_{k,n})} \|\psi_{j,m}\|_{L^1(I_{j,m})} \operatorname*{esssup}_{y\in I_{j,m}} |K(\cdot,y)|_{W^{M,\infty}(I_{k,n})} \quad \text{(Hölder again)} \\
&\lesssim 2^{-kM} 2^{-\frac{j}{2}} 2^{-\frac{k}{2}} \operatorname*{esssup}_{y\in I_{j,m}} |K(\cdot,y)|_{W^{M,\infty}(I_{k,n})}.
\end{aligned}
$$

Controlling $\operatorname{esssup}_{y\in I_{j,m}} |K(\cdot,y)|_{W^{M,\infty}(I_{k,n})}$ can be achieved using the fact that derivatives of Calderón-Zygmund operator decay polynomially away from the diagonal. We obtain (not direct):

$$
\operatorname*{esssup}_{y\in I_{j,m}} |K(\cdot,y)|_{W^{M,\infty}(I_{k,n})} \lesssim \left(\frac{1+2^{j-k}}{2^{-j}+2^{-k}+|2^{-j}m-2^{-k}n|} \right)^{M+1} \square
$$

Theorem (Decrease of $\theta_{\lambda,\lambda'}$ in d-dimensions)

Assume that H belongs to the Calderón-Zygmund class and that the mother wavelet ψ is compactly supported with M vanishing moments. Set $\lambda = (j, m)$ and $\lambda' = (k, n)$. Then

$$|\theta_{\lambda,\lambda'}| \leq C_M 2^{-(M+d/2)|j-k|} \left(\frac{2^{-k} + 2^{-j}}{2^{-k} + 2^{-j} + |2^{-j}m - 2^{-k}n|} \right)^{M+d}$$

where C_M is a constant independent of j, k, n, m.

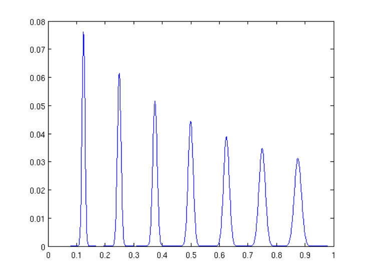
A practical example (1D)

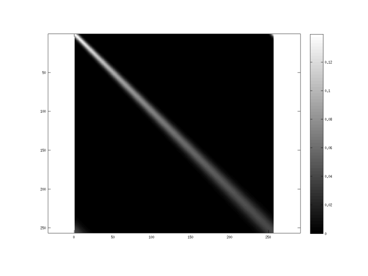
We set:

$$K(x,y) = \frac{1}{\sigma(y)\sqrt{2\pi}} \exp\left(-\frac{(x-y)^2}{2\sigma(y)^2}\right)$$

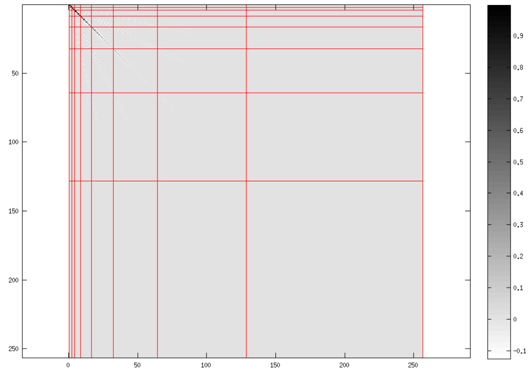
with

$$\sigma(y) = 4 + 10y.$$

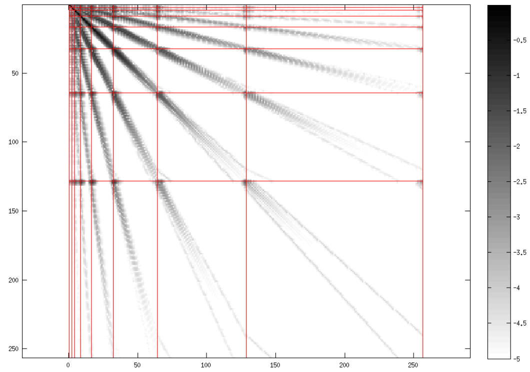




A FIELD OF PSFs AND THE DISCRETIZED MATRIX $\mathbf{H}$ WITH $N = 256$.



The matrix $\boldsymbol{\Theta}$ (usual scale).



THE MATRIX Θ ($\log_{10}$-SCALE).

Summary

Calderón-Zygmund operators are compressible in the wavelet domain !

Question

Can these results be used for fast computations?

A word on Galerkin approximations

Numerically, it is impossible to use infinite dimensional matrices.
We can therefore truncate the matrix Θ by setting a maximum scale J:

$$\boldsymbol{\Theta} = (\theta_{\lambda,\lambda'})_{0\leq \lambda,\lambda' \leq J}$$

Let

$$\Psi_J : \begin{cases} \mathbb{R}^{2^{J+1}} & \to L^2(\Omega) \\ \alpha & \mapsto \sum_{|\lambda|\leq J} \alpha_\lambda \psi_\lambda \end{cases}$$

$$\Psi_J^* : \begin{cases} L^2(\Omega) & \to \mathbb{R}^{2^{J+1}} \\ u & \mapsto (\langle u, \psi_\lambda \rangle)_{|\lambda|\leq J} \end{cases}$$

We obtain an approximation H_J of H defined by:

$$H_J = \Psi_J \boldsymbol{\Theta} \Psi_J^* = \Pi_J H \Pi_J,$$

where the operator $\Pi_J = \Psi_J \Psi_J^*$ is a projector on $\mathrm{span}(\{\psi_\lambda, |\lambda| \leq J\})$.

A word on Galerkin approximation

Standard results in approximation theory state that if u belong to some Banach space $\mathcal{B}$

$$\|u - \Pi_J(u)\|_2 = O(N^{-\alpha}), \tag{2}$$

where α depends on $\mathcal{B}$ and $N = 2^{J+1}$.
If we assume that H is regularizing, meaning that for any u satisfying (2)

$$\|Hu - \Pi_J(Hu)\|_2 = O(N^{-\beta}), \quad \text{with } \beta \geq \alpha.$$

Then:

$$\begin{aligned}\|Hu - H_J u\|_2 &= \|Hu - \Pi_J H(u - \Pi_J u - u)\|_2 \\ &\leq \|Hu - \Pi_J Hu\|_2 + \|\Pi_J H(\Pi_J u - u)\|_2 = O(N^{-\alpha}).\end{aligned}$$

Examples

- For $u \in H^1([0,1])$, $\alpha = 2$.
- For $u \in W^{1,1}([0,1])$ or $u \in BV([0,1])$, $\alpha = 1$.
- For $u \in W^{1,1}([0,1]^2)$ or $u \in BV([0,1]^2)$, $\alpha = 1/2$.

The main idea

Most coefficients in $\mathbf{\Theta}$ are small.

One can "threshold" it to obtain a sparse approximation $\mathbf{\Theta}_P$, where P denotes the number of nonzero coefficients.

We get an approximation $\mathbf{H}_P = \mathbf{\Psi}\mathbf{\Theta}_P\mathbf{\Psi}^*$.

Numerical complexity

A product $\mathbf{H}_P\mathbf{u}$ costs:

- 2 wavelet transforms of complexity $O(N)$.
- A matrix-vector product with $\mathbf{\Theta}_P$ of complexity $O(P)$.

The overall complexity for is $O(\max(P, N))$.

This is to be compared to the usual $O(N^2)$ complexity.

Theorem (theoretical foundations Beylkin, Coifman, and Rokhlin 1991)

Let $\mathbf{\Theta}_\eta$ be the matrix obtained by zeroing all coefficients in $\mathbf{\Theta}$ such that

$$\left(\frac{2^{-j}+2^{-k}}{2^{-j}+2^{-k}+|2^{-j}m-2^{-k}n|}\right)^{M+1} \leq \eta.$$

Let $\mathbf{H}_\eta = \mathbf{\Psi}\mathbf{\Theta}_\eta\mathbf{\Psi}^*$ denote the resulting operator. Then:

i) The number of non zero coefficients in $\mathbf{\Theta}_\eta$ is bounded above by

$$C'_M N \log_2(N)\, \eta^{-\frac{1}{M+1}}.$$

ii) The approximation $\mathbf{H}_\eta$ satisfies $\|\mathbf{H}-\mathbf{H}_\eta\|_{2\to 2} \lesssim \eta^{\frac{M}{M+1}}$.

iii) The complexity to obtain an ϵ-approximation $\|\mathbf{H}-\mathbf{H}_\eta\|_{2\to 2} \leq \epsilon$ is bounded above by $C''_M N \log_2(N)\, \epsilon^{-\frac{1}{M}}$.

Proof outline

1. Since $\mathbf{\Psi}$ is orthogonal,

$$\|\mathbf{H}_\eta - \mathbf{H}\|_{2\to 2} = \|\mathbf{\Theta} - \mathbf{\Theta}_\eta\|_{2\to 2}.$$

2. Let $\mathbf{\Delta}_\eta = \mathbf{\Theta} - \mathbf{\Theta}_\eta$. Use the Schur test

$$\|\mathbf{\Delta}_\eta\|_{2\to 2}^2 \leq \|\mathbf{\Delta}_\eta\|_{1\to 1}\|\mathbf{\Delta}_\eta\|_{\infty\to\infty}.$$

3. Majorize $\|\mathbf{\Delta}_\eta\|_{1\to 1}$ using Meyer's upper-bound.
   Note : the 1-norm has a simple explicit expression contrarily to the 2-norm.

Piecewise convolutions VS wavelet sparsity

	Piecewise convolutions	Wavelet sparsity
Simple theory	Yes	No
Simple implementation	Yes	No
Complexity	$O\left(N\log_2(N)\,\epsilon^{-1}\right)$	$O\left(N\log_2(N)\,\epsilon^{-\frac{1}{M}}\right)$
Adaptivity/universality	No	Yes

Geometrical intuition of the method

Link with the SVD

Let $\mathbf{\Psi} = (\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_N) \in \mathbb{R}^{N\times N}$ denote a discrete wavelet transform.
The change of basis $\mathbf{H} = \mathbf{\Psi}\mathbf{\Theta}\mathbf{\Psi}^*$ can be rewritten as:

$$\mathbf{H} = \sum_{\lambda,\lambda'} \theta_{\lambda,\lambda'} \boldsymbol{\psi}_\lambda \boldsymbol{\psi}_{\lambda'}^T.$$

The $N \times N$ matrix $\boldsymbol{\psi}_\lambda \boldsymbol{\psi}_{\lambda'}^T$ is rank-1.
Matrix $\mathbf{H}$ is therefore decomposed as the sum of N^2 rank-1 matrices.

By "thresholding" $\mathbf{\Theta}$ one can obtain an ϵ-approximation with $O(N \log_2(N)\epsilon^{-\frac{1}{M}})$ rank-1 matrices.

The SVD is a sum of N rank-1 matrices (which can also be compressed for compact operators).

Take home message

Tensor products of wavelets can be used to produce approximations of regularizing operators by sums of rank-1 matrices.

Geometrical intuition of the method

Illustration of the space decomposition with a naive thresholding.

How to choose the sparsity patterns?

First reflex - Hard thresholding

Construct $\mathbf{\Theta}_P$ by keeping the P largest coefficients of $\mathbf{\Theta}$.

This choice is optimal in the sense that it minimizes

$$\min_{\mathbf{\Theta}_P \in \mathbb{S}_P} \|\mathbf{\Theta} - \mathbf{\Theta}_P\|_F^2 = \|\mathbf{H} - \mathbf{H}_P\|_F^2$$

where $\mathbb{S}_P$ is the set of $N \times N$ matrices with at most P nonzero coefficients.

Problem: the Frobenius norm is not an operator norm.

Second reflex - Optimizing the $\|\cdot\|_{2\to 2}$-norm

In most (if not all) publications on wavelet compression of operators:

$$\min_{\mathbf{\Theta}_P \in \mathbb{S}_P} \|\mathbf{\Theta} - \mathbf{\Theta}_P\|_{2\to 2}.$$

This problem has no easily computable solution.

Approximate solutions lead to unsatisfactory approximation results.

A (much) better strategy Escande and Weiss 2014

Main idea: minimize an operator norm adapted to images.

Most signals/images are in $BV(\Omega)$ (or $B_1^{1,1}(\Omega)$), therefore (in 1D) Cohen et al. 2003:

$$\sum_{j\geq 0}\sum_{m=0}^{2^j-1} 2^j |\langle u, \psi_{j,m}\rangle| < +\infty.$$

This motivates to define a norm $\|\cdot\|_X$ on vectors:

$$\|\mathbf{u}\|_X = \|\mathbf{\Sigma}\mathbf{\Psi}^*\mathbf{u}\|_1$$

where $\mathbf{\Sigma} = diag(\sigma_1, \ldots, \sigma_N)$ and $\sigma_i = 2^{j(i)}$ where $j(i)$ is the scale of the i-th wavelet.

It leads to the following variational problem:

$$\min_{\mathbf{S}_P \in \mathbb{S}_P} \sup_{\|\mathbf{u}\|_X \leq 1} \|(\mathbf{H}-\mathbf{H}_P)\mathbf{u}\|_2 = \|\mathbf{H}-\mathbf{H}_P\|_{X\to 2}.$$

How to choose the sparsity patterns?

Optimization algorithm

Main trick : use the fact that signals and operators are sparse in the same wavelet basis. Let $\mathbf{\Delta}_P = \mathbf{\Theta} - \mathbf{\Theta}_P$. Then

$$\begin{aligned}
\max_{\|\mathbf{u}\|_X \leq 1} \|(\mathbf{H} - \mathbf{H}_P)\mathbf{u}\|_2 &= \max_{\|\mathbf{u}\|_X \leq 1} \|(\mathbf{\Psi}(\mathbf{\Theta} - \mathbf{\Theta}_P)\mathbf{\Psi}^*)\mathbf{u}\|_2 \\
&= \max_{\|\mathbf{\Sigma z}\|_1 \leq 1} \|\mathbf{\Delta}_P \mathbf{z}\|_2 \\
&= \max_{\|\mathbf{z}\|_1 \leq 1} \|\mathbf{\Delta}_P \mathbf{\Sigma}^{-1} \mathbf{z}\|_2 \\
&= \max_{1 \leq i \leq N} \frac{1}{\sigma_i} \|\mathbf{\Delta}_P^{(i)}\|_2.
\end{aligned}$$

This problem can be solved exactly using a greedy algorithm with quicksort.

Complexity

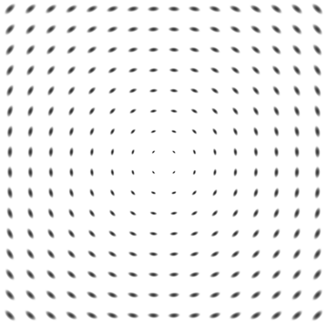
- If $\mathbf{\Theta}$ is known: $O(N^2 \log(N))$.
- If only Meyer's bound is known: $O(N \log(N))$.

Geometrical intuition of the method

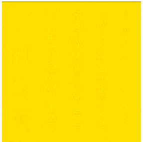
Optimal space decomposition minimizing $\|\mathbf{H} - \mathbf{H}_P\|_{X\to 2}$.

OPTIMAL SPACE DECOMPOSITION MINIMIZING $\|\mathbf{H} - \mathbf{H}_P\|_{X\to 2}$ WHEN ONLY AN UPPER-BOUND ON $\boldsymbol{\Theta}$ IS KNOWN.

Test case image

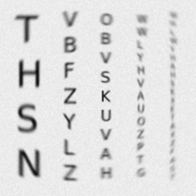


Rotational blur

	Piece. Conv.	Difference	Algorithm	Difference	$l =$
4×4 0.17 s	38.49 dB 		45.87 dB 		30 0.040s
8×8 0.36 s	44.51 dB 		50.26 dB 		50 0.048s

Blurred images using approximating operators and differences with the exact blurred image. We set the sparsity $P = lN^2$.

TV-L2 based deblurring



Degraded Image

Exact Operator
28.97dB – 2 hours

Wavelet
28.02dB – 8 seconds

Piece. Conv.
27.12dB – 35 seconds

Conclusion of the 2nd part

- ✔ Calderón-Zygmund operators are highly compressible in the wavelet domain.
- ✔ Evaluation of Calderón-Zygmund operators can be handled efficiently numerically in the wavelet domain.
- ✔ Wavelet compression outperforms piecewise convolution both theoretically and experimentally.

- ✔ Calderón-Zygmund operators are highly compressible in the wavelet domain.
- ✔ Evaluation of Calderón-Zygmund operators can be handled efficiently numerically in the wavelet domain.
- ✔ Wavelet compression outperforms piecewise convolution both theoretically and experimentally.

The devil was hidden!

Until now, we assumed that $\mathbf{\Theta}$ was known.
In 1D, the change of basis $\mathbf{\Theta} = \mathbf{\Psi}^*\mathbf{H}\mathbf{\Psi}$ has complexity $O(N^3)$!
We had to use 12 cores and 8 hours to compute $\mathbf{\Theta}$ and obtain the previous 2D results.

A dead end?

Part III: Operator reconstruction (ongoing work).

The setting

Assume we only know a few PSFs at points $(y_i)_{1\leq i\leq n} \in \Omega^n$.

The "inverse problem" we want to solve is:

Reconstruct K knowing $k_i = K(\cdot, y_i) + \eta_i$, where η_i is noise.

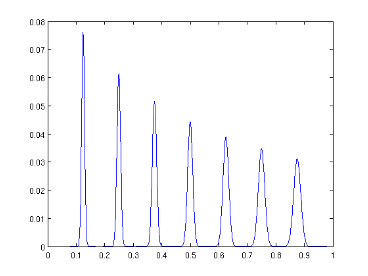
Severely ill-posed!

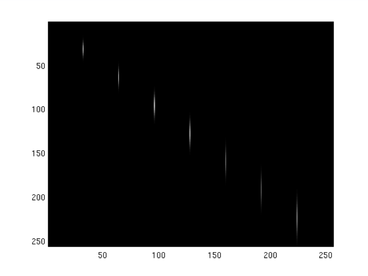
A variational formulation:

$$\inf_{K\in\mathcal{K}} \frac{1}{2}\sum_{i=1}^{n} \|k_i - K(\cdot, y_i)\|_2^2 + \lambda R(K).$$

How can we choose the regularization functional R and the space $\mathcal{K}$?

Examples





Problem illustration: some known PSFs and the associated matrix.

The first regularizer

From the first part of the talk, we know that blur operators can be approximated by matrices of type:

$$\mathbf{H}_P = \mathbf{\Psi}\mathbf{\Theta}_P\mathbf{\Psi} \tag{3}$$

where $\mathbf{\Theta}_P$ is a P sparse matrix with a known sparsity pattern $\mathbb{P}$.
We let $\mathbb{H}$ denote the space of matrices of type (3).
This is a first natural regularizer.

- ✔ Reduces the number of degrees of freedom.
- ✔ Compresses the matrix.
- ✔ Allow fast matrix-vector multiplication.
- ✗ Not sufficient to regularize the problem: we still have to find $O(N\log(N))$ coefficients.

Assumption: two neighboring PSFs are similar

From a formal point of view:

$$K(\cdot, y) \approx \tau_{-h} K(\cdot, y + h),$$

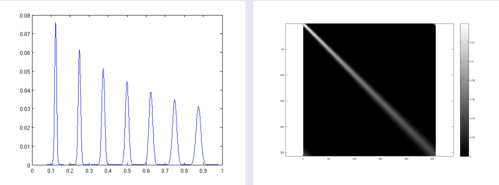
for sufficiently small h, where τ_{-h} denotes the translation operator.

Alternative formulation: the mappings

$$y \mapsto K(x + y, y)$$

should be smooth for all $x \in \Omega$.

Interpolation/approximation of scattered data



Some known PSFs and the associated matrix.

A variational formulation in 1D

Spline-based approximation of functions $\Omega = [0,1]$

Let $f : [0,1] \to \mathbb{R}$ denote a function such that $f(y_i) = \gamma_i + \eta_i$, $1 \leq i \leq n$.
A variational formulation to obtain piecewise linear approximations:

$$\inf_{g \in H^1([0,1])} \frac{1}{2} \sum_{i=1}^{n} \|g(y_i) - \gamma_i\|_2^2 + \frac{\lambda}{2} \int_{[0,1]} (g'(x))^2 \, dx.$$

From functions to operators $\Omega = [0,1]$

This motivates us to consider the problem

$$\inf_{K \in \mathcal{K}} \frac{1}{2} \sum_{i=1}^{n} \|k_i - K(\cdot, y_i)\|_2^2 + \frac{\lambda}{2} \underbrace{\int_\Omega \int_\Omega \langle \nabla K(x,y), (1;1) \rangle^2 \, dy \, dx}_{R(K)}$$

Discretization

Let $\mathbf{k}_i \in \mathbb{R}^N$ denote the discretization of $K(\cdot, y_i)$. The discretized variational problem can be rewritten:

$$\inf_{\mathbf{H}\in\mathbb{H}} \frac{1}{2}\sum_{i=1}^{n} \|\mathbf{k}_i - \mathbf{H}(\cdot, y_i)\|_2^2 + \frac{\lambda}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\mathbf{H}(i+1, j+1) - \mathbf{H}(i,j))^2.$$

The devil is still there!

This is an optimization problem over the space of $N \times N$ matrices!

Bad news...

We are now working with **HUGE** operators:

- A matrix $\mathbf{H} \in \mathbb{R}^{N\times N}$ can be thought of as a vector of size N^2.
- We need the translation operator $\mathbf{T}_{1,1} : \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times N}$ that maps $\mathbf{H}(i,j)$ to $\mathbf{H}(i+1,j+1)$.
- This way

$$\begin{aligned} R(\mathbf{H}) &= \sum_{i=1}^{N}\sum_{j=1}^{N}(\mathbf{H}(i+1,j+1)-\mathbf{H}(i,j))^2 \\ &= \left\| \left(\mathbf{T}_{1,1} - \mathbf{I}\right)(\mathbf{H})\right\|_F^2 . \end{aligned}$$

A first trick

Main observation: the shift operator $\mathbf{T}_{1,1} = \mathbf{T}_{1,0} \circ \mathbf{T}_{0,1}$.

- the shift in the vertical direction can be encoded by an $N \times N$ matrix:

$$\mathbf{T}_{1,0}(\mathbf{H}) = \mathbf{T}_1 \cdot \mathbf{H}$$

  where $\mathbf{T}_1 \in \mathbb{R}^{N\times N}$ is N-sparse.
- Similarly:

$$\mathbf{T}_{0,1}(\mathbf{H}) = (\mathbf{T}_1 \cdot \mathbf{H}^T)^T = \mathbf{H} \cdot \mathbf{T}_{-1}.$$

  Note that $\mathbf{T}_1$ is orthogonal, therefore $\mathbf{T}_1^T = \mathbf{T}_1^{-1} = \mathbf{T}_{-1}$.
- Overall $\mathbf{T}_{1,1}(\mathbf{H}) = \mathbf{T}_1 \cdot \mathbf{H} \cdot \mathbf{T}_{-1}$.

From space to wavelet domain

Theorem Beylkin 1992

The shift matrix $\mathbf{S}_1 = \mathbf{\Psi}^* \mathbf{T}_1 \mathbf{\Psi}$ contains $O(N \log N)$ non-zero coefficients. Moreover, $\mathbf{S}_1$ can be computed efficiently with an $O(N \log N)$ algorithm.

Consequences for numerical analysis

The regularization term can be computed efficiently in the wavelet domain:

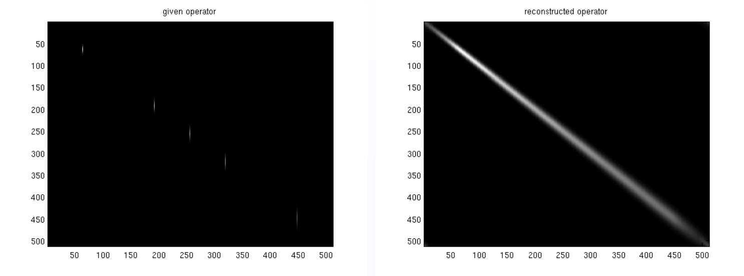
$$\begin{aligned} R(\mathbf{H}) &= \|\mathbf{T}_1 \mathbf{H} \mathbf{T}_{-1} - \mathbf{H}\|_F^2 \\ &= \|\mathbf{\Psi} \mathbf{S}_1 \mathbf{\Psi}^* \mathbf{H} \mathbf{\Psi} \mathbf{S}_{-1} \mathbf{\Psi}^* - \mathbf{H}\|_F^2 \\ &= \|\mathbf{S}_1 \mathbf{\Theta} \mathbf{S}_{-1} - \mathbf{\Theta}\|_F^2 . \end{aligned}$$

The overall problem is now formulated only in the wonderful sparse world:

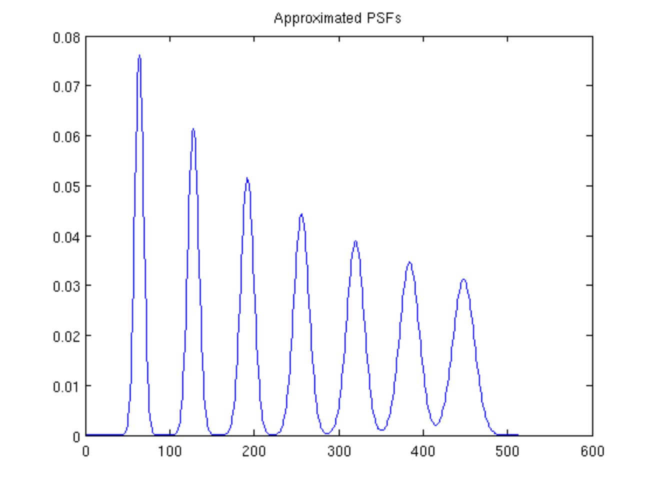
$$\min_{\mathbf{\Theta}_P \in \mathbf{\Xi}} \frac{1}{2} \sum_{i=1}^{n} \|\mathbf{k}_i - \mathbf{\Psi} \mathbf{\Theta}_P \mathbf{\Psi}^* \boldsymbol{\delta}_{y_i}\|_2^2 + \frac{\lambda}{2} \|\mathbf{S}_1 \mathbf{\Theta}_P \mathbf{S}_{-1} - \mathbf{\Theta}_P\|_F^2 .$$

where $\mathbf{\Xi}$ is the space of $N \times N$ matrices with fixed sparsity pattern $\mathbb{P}$.

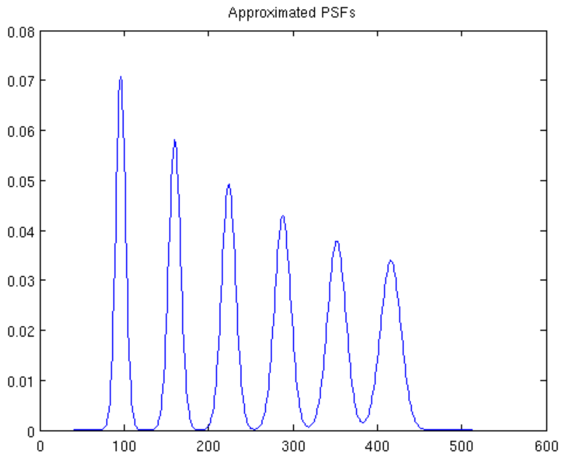
Can be solved using a projected conjugate gradient descent.



Left: known matrix. Right: reconstructed matrix.



Approximated PSF field at known locations.



Approximated PSF field at shifted known locations.

Spline approximation in higher dimensions

Scattered data in $\mathbb{R}^d$ can be interpolated or approximated using higher-order variational problems.
For instance one can use biharmonic splines:

$$\inf_{g \in H^2([0,1]^2)} \frac{1}{2} \sum_{i=1}^{n} \|g(y_i) - \gamma_i\|_2^2 + \frac{\lambda}{2} \int_{[0,1]^2} (\Delta g(y))^2 \, dx.$$

A basic reference: Wahba 1990.

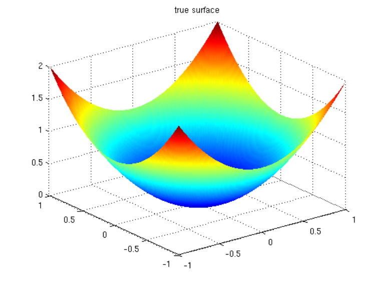
Why use higher orders?

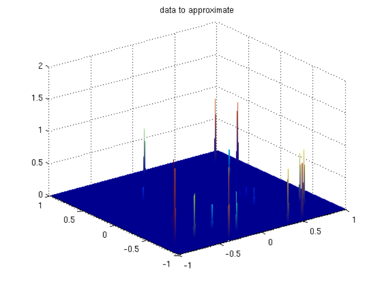
Let $B = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 \leq 1\}$. The function

$$u(x, y) = \log(|\log(\sqrt{x^2 + y^2})|)$$

belongs to $H^1(B) = W^{1,2}(B)$ and is unbounded at 0.

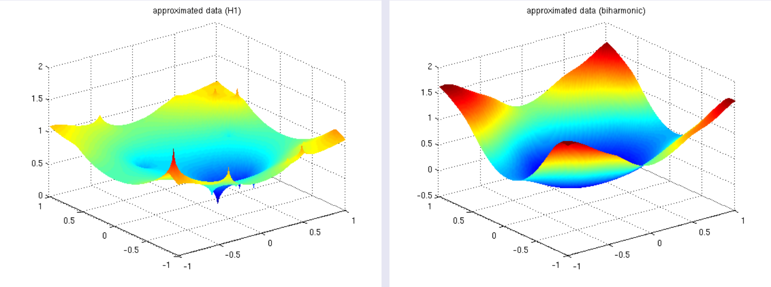
Illustration





LEFT: KNOW SURFACE $(x, y) \mapsto x^2 + y^2$. RIGHT: KNOW VALUES.

Illustration



LEFT: H_1 RECONSTRUCTION. RIGHT: BIHARMONIC RECONSTRUCTION.

The case $\Omega = [0,1]^2$

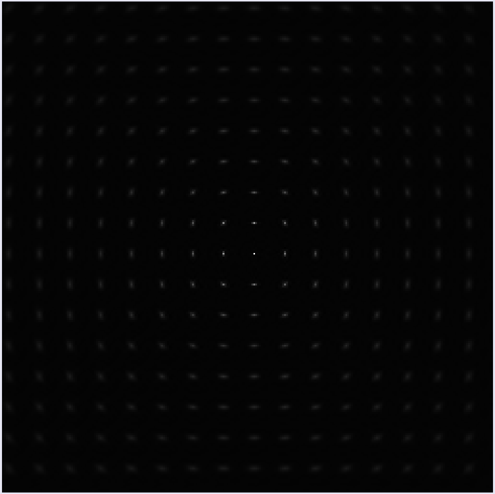
In 2D, one can solve the following variational problem:

$$\inf_{K \in H^1(\Omega\times\Omega)} \frac{1}{2}\sum_{i=1}^{n} \|k_i - K(\cdot, y_i)\|_2^2 + \frac{\lambda}{2} \underbrace{\int_\Omega \int_\Omega \Delta_y(L)^2(x,y)\, dy\, dx}_{R(K)}$$

where $L(x,y) := R(x+y, y)$.

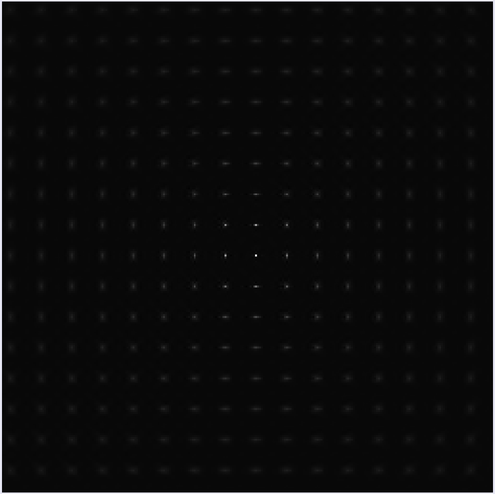
Using similar tricks as in the previous part, this problem can be entirely reformulated in the space of sparse matrices.

Reconstruction of an operator



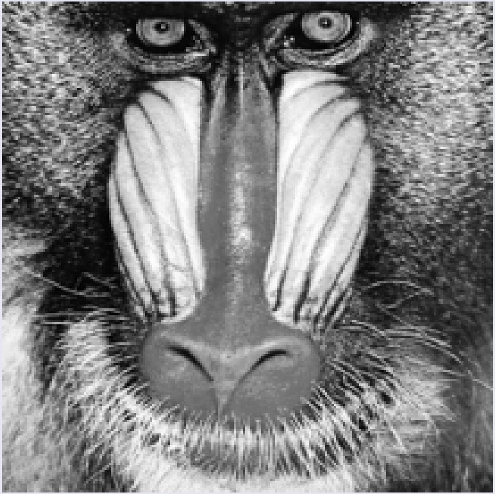
True operator (applied to the Dirac comb).

Reconstruction of an operator



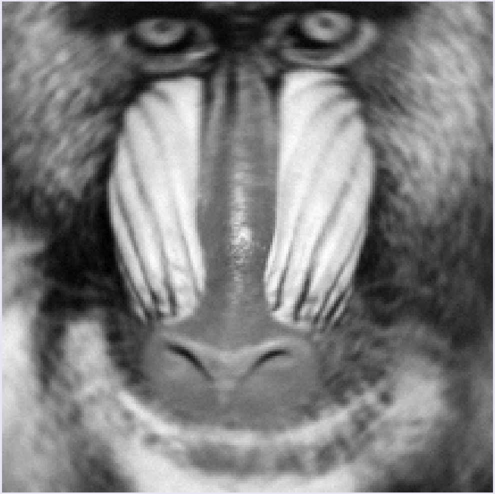
Operator reconstruction.

A complete deconvolution example



ORIGINAL IMAGE.

A complete deconvolution example



Blurry and noisy image (with the exact operator).

A complete deconvolution example

Restored image (with the reconstructed operator). Operator reconstruction = 40 minutes. Image reconstruction = 3 seconds (100 iterations of a descent algorithm).

Overall conclusion

Main facts

- ✔ Operators are highly compressible in wavelet domain.
- ✔ Operators can be computed efficiently in the wavelet domain.
- ✔ Possibility to formulate inverse problems on operator spaces.
- ✔ Regarding spatially varying deblurring:
  - ✔ numerical results are promising.
  - ✔ versatile method allowing to handle PSFs on non cartesian grids.
- ✗ Results are preliminary. Operator reconstruction takes too long.

A nice research topic

- ✔ Not much has been done.
- ✔ Plenty of work in theory.
- ✔ Plenty of work in implementation.
- ✔ Plenty of potential applications.

References

Beylkin, G. (1992). "On the representation of operators in bases of compactly supported wavelets". In: *SIAM Journal on Numerical Analysis* 29.6, pp. 1716–1740.

Beylkin, G., R. Coifman, and V. Rokhlin (1991). "Fast Wavelet Transform and Numerical Algorithm". In: *Commun. Pure and Applied Math.* 44, pp. 141–183.

Cohen, A. (2003). *Numerical analysis of wavelet methods*. Vol. 32. Elsevier.

Cohen, Albert et al. (2003). "Harmonic analysis of the space BV". In: *Revista Matematica Iberoamericana* 19.1, pp. 235–263.

Coifman, R. and Y. Meyer (1997). *Wavelets, Calderón-Zygmund and multilinear operators*. Vol. 48.

Escande, P. and P. Weiss (2014). "Numerical Computation of Spatially Varying Blur Operators A Review of Existing Approaches with a New One". In: *arXiv preprint arXiv:1404.1023*.

Wahba, G. (1990). *Spline models for observational data*. Vol. 59. Siam.