Computational Harmonic Analysis meets Imaging Sciences
Part I

Gitta Kutyniok
(Technische Universität Berlin)

BMS Summer School
Berlin, July 25 – August 5, 2016
Due to the data deluge, the area of imaging science is of tremendous importance in today’s world.

Main Tasks

- Acquisition
- Preprocessing
  - Denoising, Inpainting, ...
- Analysis
  - Feature Detection, ...
- Storing
  - Compression, ...
Due to the data deluge, the area of imaging science is of tremendous importance in today’s world.

**Main Tasks**

- Acquisition
- Preprocessing
  - Denoising, Inpainting, ...
- Analysis
  - Feature Detection, ...
- Storing
  - Compression, ...

*What has Computational Harmonic Analysis to offer?*
Representation systems designed by Computational Harmonic Analysis concepts have established themselves as a standard tool in applied mathematics, computer science, and engineering.

Examples:
- Wavelets.
- Ridgelets.
- Curvelets.
- Shearlets.
- ...

Key Property:

Fast Algorithms combined with Sparse Approximation Properties!
Outdent

1. Computational Harmonic Analysis
   - Decomposition
   - Sparse Representations

2. Frame Theory

3. Desiderata for Imaging Science
   - Model Situation
   - Benchmark Result

4. Wavelets

5. Shearlets

6. General Framework for Sparse Approximation
Exploit a carefully designed representation system \((\psi_\lambda)_{\lambda \in \Lambda} \subseteq \mathcal{H}\):

\[
\mathcal{H} \ni f \mapsto (\langle f, \psi_\lambda \rangle)_{\lambda \in \Lambda} \mapsto \sum_{\lambda \in \Lambda} \langle f, \psi_\lambda \rangle \psi_\lambda = f.
\]

Desiderata:

- Special features encoded in the “large” coefficients \(|\langle f, \psi_\lambda \rangle|\).
- Efficient representations:

\[
f \approx \sum_{\lambda \in \Lambda_N} \langle f, \psi_\lambda \rangle \psi_\lambda, \quad #(\Lambda_N) \text{ small}
\]

Goals:

- Modification of the coefficients according to the task.
- Derive high compression by considering only the “large” coefficients.
Two Main Viewpoints

Decomposition:

\[ \mathcal{H} \ni f \rightarrow (\langle f, \psi_\lambda \rangle)_{\lambda \in \Lambda}. \]

- Preprocessing (e.g. denoising).
- Analysis (e.g. feature detection).
- Clustering/Classification.
- ...

Efficient/Sparse Representations:

\[ f = \sum_{\lambda \in \Lambda} c_\lambda \psi_\lambda. \]

- Compression.
- Regularization of inverse problems.
- Ansatz functions for PDE solvers.
- ...

Decomposition

Denoising (Preprocessing):

\[ \mathcal{H} \ni f \quad \xrightarrow{\quad} \quad (\langle f, \psi_\lambda \rangle)_{\lambda \in \Lambda} \quad \xrightarrow{\quad} \quad (\langle f, \psi_\lambda \rangle)_{\lambda \in \Lambda \setminus \Lambda_0} \quad \xrightarrow{\sim} \quad \tilde{f}. \]
Decomposition

Denoising (Preprocessing):

\[ \mathcal{H} \ni f \rightarrow (\langle f, \psi_\lambda \rangle)_{\lambda \in \Lambda} \rightarrow (\langle f, \psi_\lambda \rangle)_{\lambda \in \Lambda \setminus \Lambda_0} \leadsto \tilde{f}. \]

Edge Detection (Analysis):

\[ \mathcal{H} \ni f \rightarrow (\langle f, \psi_\lambda \rangle)_{\lambda \in \Lambda} \rightarrow (\langle f, \psi_\lambda \rangle)_{\lambda \in \Lambda_1} \leadsto \tilde{f}. \]
Two Main Viewpoints

Decomposition:

\[ \mathcal{H} \ni f \rightarrow (\langle f, \psi_\lambda \rangle)_{\lambda \in \Lambda}. \]

- Preprocessing (e.g. denoising).
- Analysis (e.g. feature detection).
- Clustering/Classification.
- ...

Efficient/Sparse Representations:

\[ f = \sum_{\lambda \in \Lambda} c_\lambda \psi_\lambda. \]

- Compression.
- Regularization of inverse problems.
- Ansatz functions for PDE solvers.
- ...

Gitta Kutyniok (TU Berlin)
Representation Systems

Functional Analytic Properties:

- $(\psi_\lambda)_\lambda$ can be an orthonormal basis.
  - Unique expansion.
  - Optimal stability.

- $(\psi_\lambda)_\lambda$ can form a frame.
  - Non-unique/redundant expansions.
  - Flexibility in expansions $x = \sum_{\lambda \in \Lambda} c_\lambda \psi_\lambda$.
  - Stability.
  - Robustness against loss of coefficients $\langle x, \psi_\lambda \rangle$.

Definition: A sequence $(\psi_\lambda)_\lambda \in \Lambda \subset H$ is a frame for $H$ with frame bounds $0 < A \leq B < \infty$, if

$$A \| x \|_2 \leq \sum_{\lambda \in \Lambda} |\langle x, \psi_\lambda \rangle| \leq B \| x \|_2$$

for all $x \in H$.

We call a frame tight, if $A = B$, and Parseval, if $A = B = 1$. 
Representation Systems

Functional Analytic Properties:

- $(\psi_\lambda)_\lambda$ can be an orthonormal basis.
  - Unique expansion.
  - Optimal stability.

- $(\psi_\lambda)_\lambda$ can form a frame.
  - Non-unique/redundant expansions.
  - Flexibility in expansions $x = \sum_{\lambda \in \Lambda} c_\lambda \psi_\lambda$.
  - Stability.
  - Robustness against loss of coefficients $\langle x, \psi_\lambda \rangle$.

Definition: A sequence $(\psi_\lambda)_{\lambda \in \Lambda} \subset \mathcal{H}$ is a frame for $\mathcal{H}$ with frame bounds $0 < A \leq B < \infty$, if

$$A \|x\|^2 \leq \sum_{\lambda \in \Lambda} |\langle x, \psi_\lambda \rangle|^2 \leq B \|x\|^2 \quad \text{for all } x \in \mathcal{H}. $$

We call a frame tight, if $A = B$, and Parseval, if $A = B = 1$. 
Frame Theory

Analysis Operator: $$T : \mathcal{H} \rightarrow \ell_2(\Lambda), \ x \mapsto (\langle x, \psi_\lambda \rangle)_{\lambda \in \Lambda}$$

Synthesis Operator: $$T^* : \ell_2(\Lambda) \rightarrow \mathcal{H}, \ (c_\lambda)_{\lambda \in \Lambda} \mapsto \sum_{\lambda \in \Lambda} c_\lambda \psi_\lambda$$

Frame Operator: $$S = T^* T : \mathcal{H} \rightarrow \mathcal{H}, \ x \mapsto \sum_{\lambda \in \Lambda} \langle x, \psi_\lambda \rangle \psi_\lambda$$

Theorem: The frame operator is a positive, self-adjoint, and invertible operator and satisfies $A \cdot \text{Id} \leq S \leq B \cdot \text{Id}$. Thus, the following reconstruction/expansion formula holds:

$$x = \sum_{\lambda \in \Lambda} \langle x, \psi_\lambda \rangle S^{-1} \psi_\lambda = \sum_{\lambda \in \Lambda} \langle x, S^{-1} \psi_\lambda \rangle \psi_\lambda.$$
Frame Theory

Analysis Operator:

\[ T : \mathcal{H} \to \ell_2(\Lambda), \ x \mapsto (\langle x, \psi_\lambda \rangle)_{\lambda \in \Lambda} \]

Synthesis Operator:

\[ T^* : \ell_2(\Lambda) \to \mathcal{H}, \ (c_\lambda)_{\lambda \in \Lambda} \mapsto \sum_{\lambda \in \Lambda} c_\lambda \psi_\lambda \]

Frame Operator:

\[ S = T^* T : \mathcal{H} \to \mathcal{H}, \ x \mapsto \sum_{\lambda \in \Lambda} \langle x, \psi_\lambda \rangle \psi_\lambda \]

Theorem: The frame operator is a positive, self-adjoint, and invertible operator and satisfies \( A \cdot Id \leq S \leq B \cdot Id \). Thus, the following reconstruction/expansion formula holds:

\[ x = \sum_{\lambda \in \Lambda} \langle x, \psi_\lambda \rangle S^{-1} \psi_\lambda = \sum_{\lambda \in \Lambda} \langle x, S^{-1} \psi_\lambda \rangle \psi_\lambda. \]
Sparse Representations

Situation of Orthonormal Bases:

\[ f = \sum_{\lambda \in \Lambda} c_{\lambda} \psi_{\lambda} = \sum_{\lambda \in \Lambda} \langle f, \psi_{\lambda} \rangle \psi_{\lambda}, \]

with rapidly decaying \((\langle f, \psi_{\lambda} \rangle)_{\lambda \in \Lambda}\).

Situation of Frames:

\[ f = \sum_{\lambda \in \Lambda} c_{\lambda} \psi_{\lambda} = \sum_{\lambda \in \Lambda} \langle x, \psi_{\lambda} \rangle S^{-1} \psi_{\lambda} \]

with rapidly decaying \((c_{\lambda})_{\lambda \in \Lambda}\) or \((\langle f, \psi_{\lambda} \rangle)_{\lambda \in \Lambda}\).
Novel Paradigm:

For each class of data, there exists a sparsifying system!
Sparsity

Novel Paradigm:

For each class of data, there exists a sparsifying system!

Two Viewpoints of ‘Sparsifying System’:
Let $C \subseteq \mathcal{H}$ and $(\psi_\lambda)_\lambda \subseteq \mathcal{H}$.

- **Decay of Coefficients.** Consider the decay for $n \to \infty$ of the sorted sequence of coefficients

  $$(\|\langle x, \psi_{\lambda_n}\rangle\|)_n \quad \text{for all } x \in C.$$

- **Approximation Properties.** Consider the decay for $N \to \infty$ of the error of best $N$-term approximation, i.e.,

  $$\inf_{\#\Lambda_N=N,(c_\lambda)_\lambda} \left\| x - \sum_{\lambda \in \Lambda_N} c_\lambda \psi_\lambda \right\| \quad \text{for all } x \in C.$$
Notion of Optimality

Two Viewpoints of Optimality of $\psi_{\lambda}$: Let $C \subseteq \mathcal{H}$.

- **Decay of Coefficients.** $\beta > 0$ is largest (for all systems) with
  
  $$|\langle x, \psi_{\lambda_n} \rangle| \lesssim n^{-\beta} \text{ as } n \to \infty, \text{ for all } x \in C.$$ 

- **Approximation Properties.** $\gamma > 0$ is largest (for all systems) with
  
  $$\inf_{\#\Lambda_{N} = N,(c_{\lambda})_{\lambda}} \left\| x - \sum_{\lambda \in \Lambda_{N}} c_{\lambda} \psi_{\lambda} \right\| \lesssim N^{-\gamma} \text{ as } N \to \infty, \text{ for all } x \in C.$$
Notion of Optimality

Two Viewpoints of Optimality of $(\psi_\lambda)_\lambda$: Let $C \subseteq \mathcal{H}$.

- **Decay of Coefficients.** $\beta > 0$ is largest (for all systems) with
  \[ |\langle x, \psi_\lambda_n \rangle| \lesssim n^{-\beta} \text{ as } n \to \infty, \quad \text{for all } x \in C. \]

- **Approximation Properties.** $\gamma > 0$ is largest (for all systems) with
  \[
  \inf_{\#\Lambda_N=N,(c_\lambda)_{\lambda}} \left\| x - \sum_{\lambda \in \Lambda_N} c_\lambda \psi_\lambda \right\| \lesssim N^{-\gamma} \text{ as } N \to \infty, \quad \text{for all } x \in C.
  \]

Situation of an ONB: For the best $N$-term approximation $x_N$ of $x$, we have
\[
\|x - x_N\|^2 = \sum_{\lambda \not \in \Lambda_N} |c_\lambda|^2 = \sum_{n > N} |\langle x, \psi_\lambda_n \rangle|^2
\]
Notion of Optimality

Two Viewpoints of Optimality of \((\psi_\lambda)_\lambda\): Let \(C \subseteq \mathcal{H}\).

- **Decay of Coefficients.** \(\beta > 0\) is largest (for all systems) with
  \[|\langle x, \psi_{\lambda_n} \rangle| \lesssim n^{-\beta} \text{ as } n \to \infty, \quad \text{for all } x \in C.\]

- **Approximation Properties.** \(\gamma > 0\) is largest (for all systems) with
  \[\inf_{\#\Lambda_N=N,(c_\lambda)_\lambda} \left\| x - \sum_{\lambda \in \Lambda_N} c_\lambda \psi_\lambda \right\| \lesssim N^{-\gamma} \text{ as } N \to \infty, \quad \text{for all } x \in C.\]

**Situation of an ONB:** For the best \(N\)-term approximation \(x_N\) of \(x\), we have
  \[\|x - x_N\|^2 = \sum_{\lambda \not\in \Lambda_N} |c_\lambda|^2 = \sum_{n > N} |\langle x, \psi_{\lambda_n} \rangle|^2\]

**Situation of a Frame:** For the \(N\)-term approximation \(x_N = \sum_{\lambda \in \Lambda_N} \langle x, \psi_\lambda \rangle \tilde{\psi}_\lambda\) of \(x\) consisting of the \(N\) largest coefficients \(|\langle x, \psi_\lambda \rangle|\), we only have
  \[\|x - x_N\|^2 \leq \frac{1}{A} \sum_{n > N} |\langle x, \psi_{\lambda_n} \rangle|^2.\]
Two Main Viewpoints

Decomposition:

\[ \mathcal{H} \ni f \rightarrow (\langle f, \psi_\lambda \rangle)_{\lambda \in \Lambda}. \]

- Preprocessing (e.g. denoising).
- Analysis (e.g. feature detection).
- Clustering/Classification.
- ...

Efficient/Sparse Representations:

\[ f = \sum_{\lambda \in \Lambda} c_\lambda \psi_\lambda. \]

- Compression.
- Regularization of inverse problems.
- Ansatz functions for PDE solvers.
- ...
Regularization of Inverse Problems

General Setting:
Given $K : X \to Y$ and $y \in Y$, compute $x \in X$ with $Kx = y$.

Well-Posedness Conditions (Hadamard):
- **Existence**: For each $y \in Y$, there exists some $x \in X$ with $Kx = y$.
- **Uniqueness**: Such an $x \in X$ is unique.
- **Stability**: $\lim_{n \to \infty} Kx_n \to Kx$ implies $\lim_{n \to \infty} x_n \to x$.

Ill-Posed Inverse Problems:
Need for regularization!

Regularization Strategy:
A family of linear and bounded operators $R_\alpha : Y \to X$, $\alpha > 0$, such that $\lim_{\alpha \to 0} R_\alpha Kx = x$ for all $x \in X$. 
Regularization of Inverse Problems

General Setting:
Given $K : X \to Y$ and $y \in Y$, compute $x \in X$ with $Kx = y$.

Well-Posedness Conditions (Hadamard):
- **Existence:** For each $y \in Y$, there exists some $x \in X$ with $Kx = y$.
- **Uniqueness:** Such an $x \in X$ is unique.
- **Stability:** $\lim_{n \to \infty} Kx_n \to Kx$ implies $\lim_{n \to \infty} x_n \to x$.

Ill-Posed Inverse Problems:

*Need for regularization!*
Regularization of Inverse Problems

General Setting:
Given $K : X \to Y$ and $y \in Y$, compute $x \in X$ with $Kx = y$.

Well-Posedness Conditions (Hadamard):
- **Existence**: For each $y \in Y$, there exists some $x \in X$ with $Kx = y$.
- **Uniqueness**: Such an $x \in X$ is unique.
- **Stability**: $\lim_{n \to \infty} Kx_n \to Kx$ implies $\lim_{n \to \infty} x_n \to x$.

Ill-Posed Inverse Problems:

Need for regularization!

Regularization Strategy:
A family of linear and bounded operators $R_\alpha : Y \to X$, $\alpha > 0$, such that

$$\lim_{\alpha \to 0} R_\alpha Kx(=: x^\alpha) = x \quad \text{for all } x \in X.$$
Tikhonov Regularization

**Standard Tikhonov Regularization:**
Given an ill-posed inverse problem $Kx = y$, where $K : X \to Y$, an approximate solution $x^\alpha \in X$, $\alpha > 0$, can be determined by minimizing

$$J_\alpha(x) := \|Kx - y\|^2 + \alpha\|x\|^2, \quad x \in X.$$
Tikhonov Regularization

Standard Tikhonov Regularization:
Given an ill-posed inverse problem \( Kx = y \), where \( K : X \to Y \), an approximate solution \( x^\alpha \in X \), \( \alpha > 0 \), can be determined by minimizing

\[
J_\alpha(x) := \|Kx - y\|^2 + \alpha \|x\|^2, \quad x \in X.
\]

Generalization:

\[
\tilde{J}_\alpha(x) := \|Kx - y\|^2 + \alpha \mathcal{P}(x), \quad x \in X.
\]

The penalty term \( \mathcal{P} \)

- ensures continuous dependence on the data,
- incorporates properties of the solution.

Some Examples for \( \mathcal{P} \):

\[
\|x\|_{TV}, \quad \|x\|_{H^s}, \quad \|(\langle x, \psi_\lambda \rangle)_\lambda\|_1, \ldots
\]
Tikhonov Regularization

Standard Tikhonov Regularization:
Given an ill-posed inverse problem $Kx = y$, where $K : X \to Y$, an approximate solution $x^\alpha \in X$, $\alpha > 0$, can be determined by minimizing

$$J_\alpha(x) := \|Kx - y\|^2 + \alpha\|x\|^2, \quad x \in X.$$ 

Generalization:

$$\tilde{J}_\alpha(x) := \|Kx - y\|^2 + \alpha P(x), \quad x \in X.$$ 

The penalty term $P$
- ensures continuous dependence on the data,
- incorporates properties of the solution.

Some Examples for $P$:

$$\|x\|_{TV}, \quad \|x\|_{H^s}, \quad \|\langle x, \psi_\lambda \rangle\|_1, \ldots$$
Compressed Sensing (Candès, Romberg, Tao and Donoho; 2006)

Main Goal: Solve an underdetermined linear problem

\[ y = Ax, \quad A \text{ an } n \times N\text{-matrix with } n \ll N, \]

for a solution \( x \in \mathbb{R}^N \) admitting a sparsifying system \((\psi_\lambda)_\lambda\).

Approach: Recover \( x \) by the \( \ell_1 \)-analysis minimization problem

\[
\min_{\tilde{x}} \| (\langle \tilde{x}, \psi_\lambda \rangle)_\lambda \|_1 \text{ subject to } y = A\tilde{x}
\]
Main Goal: Solve an underdetermined linear problem

\[ y = Ax, \quad A \text{ an } n \times N\text{-matrix with } n \ll N, \]

for a solution \( x \in \mathbb{R}^N \) admitting a sparsifying system \((\psi_\lambda)_\lambda\).

Approach: Recover \( x \) by the \( \ell_1 \)-analysis minimization problem

\[
\min_{\tilde{x}} \| (\langle \tilde{x}, \psi_\lambda \rangle)_\lambda \|_1 \quad \text{subject to } y = A\tilde{x}
\]

Why \( \ell_1 \)?
Main Goal: Solve an underdetermined linear problem

\[ y = Ax, \quad A \text{ an } n \times N\text{-matrix with } n \ll N, \]

for a solution \( x \in \mathbb{R}^N \) admitting a sparsifying system \((\psi_\lambda)\_\lambda\).

Approach: Recover \( x \) by the \( \ell_1 \)-analysis minimization problem

\[
\min_{\tilde{x}} \| \langle \tilde{x}, \psi_\lambda \rangle \_\lambda \|_1 \text{ subject to } y = A\tilde{x}
\]

Why \( \ell_1 \)?

Meta-Result: If \((\langle x, \psi_\lambda \rangle)\_\lambda\) is sufficiently sparse, and \( A \) is sufficiently incoherent, then \( x \) can be recovered from \( Ax \) by \( \ell_1 \) minimization.
Two Main Viewpoints

Decomposition:

\[ \mathcal{H} \ni f \longrightarrow (\langle f, \psi_\lambda \rangle)_{\lambda \in \Lambda}. \]

- Preprocessing (e.g. denoising).
- Analysis (e.g. feature detection).
- Clustering/Classification.
- ...

Efficient/Sparse Representations:

\[ f = \sum_{\lambda \in \Lambda} c_\lambda \psi_\lambda. \]

- Compression.
- Regularization of inverse problems.
- Ansatz functions for PDE solvers.
- ...

Gitta Kutyniok (TU Berlin)
Desiderata:

- Multiscale representation system.
- Convenient structure: Operators applied to one generating function.
- Partition of Fourier domain.
- Space/frequency localization.
- Fast algorithms: $x \mapsto (\langle x, \psi_\lambda \rangle)_\lambda \leadsto x$.
- Optimality for the considered class.

*In this Talk:* Modeling natural images!
Modelling Anisotropic Structures
What is an Image?

Intuitively edges are main structure. Justified by neurophysiology. Field et al., 1993
What is an Image?

Intuitively edges are main structure.
Justified by neurophysiology.
Field et al., 1993
What is an Image?

Intuitively edges are main structure. Justified by neurophysiology. (Field et al., 1993)
What is an Image?

- Intuitively edges are main structure.
- Justified by neurophysiology.

Field et al., 1993
Fitting Model

Definition (Donoho; 2001):
The set of cartoon-like functions $\mathcal{E}^2(\mathbb{R}^2)$ is defined by

$$\mathcal{E}^2(\mathbb{R}^2) = \{ f \in L^2(\mathbb{R}^2) : f = f_0 + f_1 \cdot \chi_B \},$$

where $\emptyset \neq B \subset [0, 1]^2$ simply connected with $C^2$-boundary and bounded curvature, and $f_i \in C^2(\mathbb{R}^2)$ with $\text{supp } f_i \subseteq [0, 1]^2$ and $\| f_i \|_{C^2} \leq 1$, $i = 0, 1$. 

\[\begin{align*}
&\text{Gitta Kutyniok (TU Berlin)} \\
&\text{Computational Harmonic Analysis} \\
&\text{BMS Summer School’16} \\
&22 / 59
\end{align*}\]
Fitting Model

Definition (Donoho; 2001):
The set of cartoon-like functions $\mathcal{E}^2(\mathbb{R}^2)$ is defined by

$$\mathcal{E}^2(\mathbb{R}^2) = \{ f \in L^2(\mathbb{R}^2) : f = f_0 + f_1 \cdot \chi_B \},$$

where $\emptyset \neq B \subset [0, 1]^2$ simply connected with $C^2$-boundary and bounded curvature, and $f_i \in C^2(\mathbb{R}^2)$ with $\text{supp} f_i \subseteq [0, 1]^2$ and $\| f_i \|_{C^2} \leq 1$, $i = 0, 1$.

Theorem (Donoho; 2001):
Let $(\psi_\lambda)_\lambda \subset L^2(\mathbb{R}^2)$. Allowing only polynomial depth search, we have the following optimal behavior for $f \in \mathcal{E}^2(\mathbb{R}^2)$:

$$\| f - f_N \|_2^2 \asymp N^{-2} \quad \text{and} \quad | \langle f, \psi_\lambda_n \rangle | \lesssim n^{-\frac{3}{2}} \quad \text{as } N, n \to \infty.$$
Review of 2-D Wavelets

Definition (1D): Let $\phi \in L^2(\mathbb{R})$ be a scaling function and $\psi \in L^2(\mathbb{R})$ be a wavelet. Then the associated wavelet system is defined by

$$
\{\phi(x - m) : m \in \mathbb{Z}\} \cup \{2^{j/2} \psi(2^j x - m) : j \geq 0, m \in \mathbb{Z}\}.
$$

Theorem: Wavelets provide optimally sparse approximations for functions $f \in L^2(\mathbb{R}^2)$, which are $C^2$ apart from point singularities:

$$
\|f - f_N\|_2 \approx N^{-1}, \quad N \to \infty.
$$
Review of 2-D Wavelets

Definition (1D): Let \( \phi \in L^2(\mathbb{R}) \) be a scaling function and \( \psi \in L^2(\mathbb{R}) \) be a wavelet. Then the associated wavelet system is defined by

\[
\{ \phi(x - m) : m \in \mathbb{Z} \} \cup \{ 2^{j/2} \psi(2^j x - m) : j \geq 0, m \in \mathbb{Z} \}.
\]

Definition (2D): A wavelet system is defined by

\[
\{ \phi^{(1)}(x - m) : m \in \mathbb{Z}^2 \} \cup \{ 2^j \psi^{(i)}(2^j x - m) : j \geq 0, m \in \mathbb{Z}^2, i = 1, 2, 3 \},
\]

where

\[
\psi^{(1)}(x) = \phi(x_1)\psi(x_2), \quad \psi^{(2)}(x) = \psi(x_1)\phi(x_2), \quad \psi^{(3)}(x) = \psi(x_1)\psi(x_2).
\]

Theorem: Wavelets provide optimally sparse approximations for functions \( f \in L^2(\mathbb{R}^2) \), which are \( C^2 \) apart from point singularities:

\[
\| f - f_N \|_2^2 \asymp N^{-1}, \quad N \to \infty.
\]
Wavelet Decomposition: JPEG2000
Wavelet Decomposition: JPEG2000

Original

25% Compression

5% Compression
What can Wavelets do?

Problem:
- For $f \in \mathcal{E}^2(\mathbb{R}^2)$, wavelets only achieve $\|f - f_N\|_2^2 \asymp N^{-1}$, $N \to \infty$.
- **Isotropic** structure of wavelets:
  $$\{2^j \psi\left(\begin{pmatrix} 2^j & 0 \\ 0 & 2^j \end{pmatrix} x - m\right) : j \geq 0, m \in \mathbb{Z}^2\}.$$
- Wavelets **cannot** sparsely represent cartoon-like functions.

Intuitive explanation:
Main Goal

Design a Representation System which...

- fits into the framework of affine systems,
- provides an optimally sparsifying system for cartoons,
- allows for compactly supported analyzing elements,
- is associated with fast decomposition algorithms,
- treats the continuum and digital ‘world’ uniformly.
Main Goal

Design a Representation System which...

- fits into the framework of affine systems,
- provides an optimally sparsifying system for cartoons,
- allows for compactly supported analyzing elements,
- is associated with fast decomposition algorithms,
- treats the continuum and digital ‘world’ uniformly.

Non-Exhaustive List of Approaches:

- Ridgelets (Candès and Donoho; 1999)
- Curvelets (Candès and Donoho; 2002)
- Contourlets (Do and Vetterli; 2002)
- Bandlets (LePennec and Mallat; 2003)
- Shearlets (K and Labate; 2006)
What is a Shearlet?
Scaling and Orientation

Parabolic scaling (‘width \(\approx\) length\(^2\)’):

\[
A_{2^j} = \begin{pmatrix}
2^j & 0 \\
0 & 2^{j/2}
\end{pmatrix}, \quad j \in \mathbb{Z}.
\]

Historical remark:

Scaling and Orientation

Parabolic scaling (‘width $\approx$ length$^2$’):

$$A_{2j} = \begin{pmatrix} 2^j & 0 \\ 0 & 2^{j/2} \end{pmatrix}, \quad j \in \mathbb{Z}.$$  

Historical remark:


Orientation via shearing:

$$S_k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}, \quad k \in \mathbb{Z}.$$  

Advantage:

- Shearing leaves the digital grid $\mathbb{Z}^2$ invariant.
- Uniform theory for the continuum and digital situation.
Shearlet Systems

Affine systems:

\[ \{ \frac{1}{\sqrt{\det M}} \psi(M \cdot -m) : M \in G \subseteq GL_2, \ m \in \mathbb{Z}^2 \}. \]

Definition (K, Labate; 2006):

For \( \psi \in L^2(\mathbb{R}^2) \), the associated shearlet system is defined by

\[ \{ 2^{\frac{3j}{4}} \psi(S_k A_{2^i} \cdot -m) : j, k \in \mathbb{Z}, m \in \mathbb{Z}^2 \}. \]

Remarks:

- Advantage: Generated by a unitary representation of the locally compact group \( (\mathbb{R}^+ \times \mathbb{R}) \ltimes \mathbb{R}^2 \), the so-called shearlet group.
- Disadvantage: Non-uniform treatment of directions.
Example of Classical (Band-Limited) Shearlet

Let $\psi \in L^2(\mathbb{R}^2)$ be defined by

$$\hat{\psi}(\xi) = \hat{\psi}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1) \hat{\psi}_2(\frac{\xi_2}{\xi_1}),$$

where

- $\psi_1$ wavelet, $\text{supp}(\hat{\psi}_1) \subseteq [-2, -\frac{1}{2}] \cup [\frac{1}{2}, 2]$ and $\hat{\psi}_1 \in C^\infty(\mathbb{R})$,
- $\psi_2$ ‘bump function’, $\text{supp}(\hat{\psi}_2) \subseteq [-1, 1]$ and $\hat{\psi}_2 \in C^\infty(\mathbb{R})$. 
Example of Classical (Band-Limited) Shearlet

Let $\psi \in L^2(\mathbb{R}^2)$ be defined by

$$\hat{\psi}(\xi) = \hat{\psi}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1) \hat{\psi}_2\left(\frac{\xi_2}{\xi_1}\right),$$

where

- $\psi_1$ wavelet, $\text{supp}(\hat{\psi}_1) \subseteq [-2, -\frac{1}{2}] \cup [\frac{1}{2}, 2]$ and $\hat{\psi}_1 \in C^\infty(\mathbb{R})$,
- $\psi_2$ ‘bump function’, $\text{supp}(\hat{\psi}_2) \subseteq [-1, 1]$ and $\hat{\psi}_2 \in C^\infty(\mathbb{R})$.

Induced tiling of Fourier domain:
Example of Classical (Band-Limited) Shearlet

Let $\psi \in L^2(\mathbb{R}^2)$ be defined by

$$\hat{\psi}(\xi) = \hat{\psi}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1) \hat{\psi}_2(\frac{\xi_2}{\xi_1}),$$

where

- $\psi_1$ wavelet, $\operatorname{supp}(\hat{\psi}_1) \subseteq [-2, -\frac{1}{2}] \cup [\frac{1}{2}, 2]$ and $\hat{\psi}_1 \in C^\infty(\mathbb{R})$,
- $\psi_2$ ‘bump function’, $\operatorname{supp}(\hat{\psi}_2) \subseteq [-1, 1]$ and $\hat{\psi}_2 \in C^\infty(\mathbb{R})$.

Induced tiling of Fourier domain:
(Cone-adapted) Shearlet Systems

Definition (K, Labate; 2006):
The (cone-adapted) shearlet system $\mathcal{SH}(c; \phi, \psi, \tilde{\psi})$, $c > 0$, generated by $\phi \in L^2(\mathbb{R}^2)$ and $\psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ is the union of

\[
\{ \phi(\cdot - cm) : m \in \mathbb{Z}^2 \},
\]
\[
\{ 2^{3j/4} \psi(S_k A_{2j} \cdot - cm) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2 \},
\]
\[
\{ 2^{3j/4} \tilde{\psi}(\tilde{S}_k \tilde{A}_{2j} \cdot - cm) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2 \}.
\]
(Cone-adapted) Shearlet Systems

Definition (K, Labate; 2006):
The (cone-adapted) shearlet system $\mathcal{SH}(c; \phi, \psi, \tilde{\psi})$, $c > 0$, generated by $\phi \in L^2(\mathbb{R}^2)$ and $\psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ is the union of

\[
\{ \phi(\cdot - cm) : m \in \mathbb{Z}^2 \}, \\
\{ 2^{3j/4} \psi(S_k A_{2j} \cdot - cm) : j \geq 0, |k| \leq [2^{j/2}], m \in \mathbb{Z}^2 \}, \\
\{ 2^{3j/4} \tilde{\psi}(\tilde{S}_k \tilde{A}_{2j} \cdot - cm) : j \geq 0, |k| \leq [2^{j/2}], m \in \mathbb{Z}^2 \}.
\]

Theorem (K, Labate, Lim, Weiss; 2006):
For $\psi, \tilde{\psi}$ classical shearlets, $\mathcal{SH}(1; \phi, \psi, \tilde{\psi})$ is a Parseval frame for $L^2(\mathbb{R}^2)$:

\[
A \|f\|_2^2 \leq \sum_{\sigma \in \mathcal{SH}(\phi, \psi, \tilde{\psi})} |\langle f, \sigma \rangle|^2 \leq B \|f\|_2^2 \quad \text{for all } f \in L^2(\mathbb{R}^2)
\]

holds for $A = B = 1$. 
Proof of Parseval Frame Property

Specific conditions on a classical shearlet $\hat{\psi}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1) \hat{\psi}_2(\frac{\xi_2}{\xi_1})$:

- Wavelet: $\sum_{j \in \mathbb{Z}} |\hat{\psi}_1(2^{-j} \xi)|^2 = 1$ for a.e. $\xi \in \mathbb{R}$.
- ‘Bump Function’: $\sum_{k=-1,0,1} |\hat{\psi}_2(\xi + k)|^2 = 1$ for a.e. $\xi \in [-1,1]$.

By the above properties of $\psi_1$ and $\psi_2$, we have

$$\sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} |\hat{\psi}(S^T_{-k} A_{-j} \xi)|^2 = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} |\hat{\psi}(2^{-j} \xi_1, 2^{-j/2} \xi_2 - 2^{-j} \xi_1 k)|^2$$

$$= \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} |\hat{\psi}_1(2^{-j} \xi_1)|^2 |\hat{\psi}_2(2^{j/2} \xi_2 \frac{\xi}{\xi_1} - k)|^2$$

$$= \sum_{j \in \mathbb{Z}} |\hat{\psi}_1(2^{-j} \xi_1)|^2 = 1.$$
Compactly Supported Shearlets

Theorem (Kittipoom, K, Lim; 2012):
Let \( \phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2) \) be compactly supported, and let \( \hat{\phi}, \hat{\psi}, \hat{\tilde{\psi}} \) satisfy certain decay conditions. Then there exists \( c_0 \) such that \( \mathcal{SH}(c; \phi, \psi, \tilde{\psi}) \) forms a shearlet frame with controllable frame bounds for all \( c \leq c_0 \).

Remark: Exemplary class with \( B/A \approx 4 \).
Compactly Supported Shearlets

Theorem (Kittipoom, K, Lim; 2012):
Let $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ be compactly supported, and let $\hat{\phi}, \hat{\psi}, \hat{\tilde{\psi}}$ satisfy certain decay conditions. Then there exists $c_0$ such that $\mathcal{SH}(c; \phi, \psi, \tilde{\psi})$ forms a shearlet frame with controllable frame bounds for all $c \leq c_0$.

Remark: Exemplary class with $B/A \approx 4$.

Theorem (Guo, Labate; 2007)(K, Lim; 2011):
Let $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ be compactly supported, and let $\hat{\phi}, \hat{\psi}, \hat{\tilde{\psi}}$ satisfy certain decay conditions. Then $\mathcal{SH}(c; \phi, \psi, \tilde{\psi}) = (\sigma_\eta)_\eta$ provides an optimally sparsifying system for $f \in \mathcal{E}^2(\mathbb{R}^2)$, i.e., for $N, n \to \infty$,

$$\|f - f_N\|_2^2 \lesssim N^{-2}(\log N)^3 \quad \text{and} \quad |\langle f, \sigma_{\eta_n} \rangle| \lesssim n^{-\frac{3}{2}}(\log n)^{\frac{3}{2}}.$$
Heuristic Argument

Estimate:

\[ \| f - f_N \|_2^2 \lesssim \sum_{n > N} (|\langle f, \sigma \eta_n \rangle|)^2 \lesssim \sum_{n > N} (n^{-\frac{3}{2}})^2 \lesssim N^{-2}. \]

Case 1:

\[ |\langle f, \sigma \eta \rangle| \text{ negligible!} \]

Case 2:

\[ |\langle f, \sigma \eta \rangle| \text{ negligible!} \]

Case 3:

\[ |\langle f, \sigma \eta \rangle| \leq \| f \|_\infty \| \sigma \eta \|_1 \lesssim 2^{-\frac{3}{4}j} \]

\[ \implies |\langle f, \sigma \eta_n \rangle| \lesssim n^{-\frac{3}{2}} \]
Heuristic Argument

Estimate:

\[ \| f - f_N \|_2^2 \lesssim \sum_{n > N} (|\langle f, \sigma \eta_n \rangle|)^2 \lesssim \sum_{n > N} (n^{-\frac{3}{2}})^2 \lesssim N^{-2}. \]

Case 1:

|\langle f, \sigma \eta \rangle| \text{ negligible!}

Case 2:

|\langle f, \sigma \eta \rangle| \text{ negligible!}

Case 3:

|\langle f, \sigma \eta \rangle| \leq \| f \|_{\infty} \| \sigma \eta \|_1 \lesssim 2^{-\frac{3}{4}j}

\sim \ |\langle f, \sigma \eta_n \rangle| \lesssim n^{-\frac{3}{2}}
Heuristic Argument

Estimate:

\[ \| f - f_N \|_2^2 \lesssim \sum_{n > N} (|\langle f, \sigma\eta_n \rangle|)^2 \lesssim \sum_{n > N} (n^{-3/2})^2 \lesssim N^{-2}. \]

Case 1:

|\langle f, \sigma\eta \rangle| \text{ negligible!}

Case 2:

|\langle f, \sigma\eta \rangle| \text{ negligible!}

Case 3:

|\langle f, \sigma\eta \rangle| \leq \| f \|_\infty \| \sigma\eta \|_1 \lesssim 2^{-3/4j}

\[ \sim |\langle f, \sigma\eta_n \rangle| \lesssim n^{-3/2} \]
Heuristic Argument

Estimate:
\[ \| f - f_N \|_2^2 \lesssim \sum_{n > N} (| \langle f, \sigma_n \rangle |)^2 \lesssim \sum_{n > N} (n^{-\frac{3}{2}})^2 \lesssim N^{-2}. \]

Case 1:
\[ | \langle f, \sigma_\eta \rangle | \text{ negligible!} \]

Case 2:
\[ | \langle f, \sigma_\eta \rangle | \text{ negligible!} \]

Case 3:
\[ | \langle f, \sigma_\eta \rangle | \leq \| f \|_\infty \| \sigma_\eta \|_1 \lesssim 2^{-\frac{3}{4}j} \]
\[ \leadsto | \langle f, \sigma_{\eta_n} \rangle | \lesssim n^{-\frac{3}{2}} \]
Recent Approaches to Fast Shearlet Transforms

www.ShearLab.org:
- Separable Shearlet Transform (*Lim*; 2009)
- Digital Shearlet Transform (*K, Shahram, Zhuang*; 2011)
- 2D&3D (parallelized) Shearlet Transform (*K, Lim, Reisenhofer*; 2013)

Additional Code:
- Filter-based implementation (*Easley, Labate, Lim*; 2009)
- Fast Finite Shearlet Transform (*Häuser, Steidl*; 2014)
- Shearlet Toolbox 2D&3D (*Easley, Labate, Lim, Negy*; 2014)

Theoretical Approaches:
- Adaptive Directional Subdivision Schemes (*K, Sauer*; 2009)
- Shearlet Unitary Extension Principle (*Han, K, Shen*; 2011)
- Gabor Shearlets (*Bodmann, K, Zhuang*; 2013)
What about Curvelets…?
Curvelets

Definition (Candès, Donoho; 2002):
Let
- $W \in C^\infty(\mathbb{R})$ be a wavelet with $\text{supp}(W) \subseteq \left(\frac{1}{2}, 2\right)$,
- $V \in C^\infty(\mathbb{R})$ be a ‘bump function’ with $\text{supp}(V) \subseteq (-1, 1)$.

Then the curvelet system $(\gamma_{j,l,k})_{(j,l,k)}$ is defined by

$$\hat{\gamma}_{j,0,0}(r, \omega) := 2^{-3j/4} W (2^{-j} r) V(2^{\lfloor j/2 \rfloor} \omega)$$

and

$$\gamma_{j,l,k}(\cdot) := \gamma_{j,0,0}(R_{\theta_{(j,l,k)}}(\cdot - x_{(j,l,k)})).$$

Theorem (Candès, Donoho; 2002):
The curvelet system forms a Parseval frame for $L^2(\mathbb{R}^2)$. 

Gitta Kutyniok (TU Berlin)  Computational Harmonic Analysis  BMS Summer School’16  38 / 59
Comparison with Shearlets, I

Main Differences to Shearlets:

- Not affine systems.
- Based on rotation in contrast to shearing.
- Only band-limited version available.

Performance on Separation:
Comparison with Shearlets, II

But there are also many Similarities...

Theorem (Candès, Donoho; 2002):
Curvelets provide optimally sparse approximations of $f \in \mathcal{E}^2(\mathbb{R}^2)$, i.e.,

$$\|f - f_N\|_2^2 \leq C \cdot N^{-2} \cdot (\log N)^3, \quad N \to \infty.$$
Towards a General Framework...
Introduce a Framework which...

- ...covers all systems known to sparsify cartoons.
- ...enables easy transfer of (sparsity) results between systems.
- ...allows categorization of systems with respect to sparsity behaviors.
- ...is general enough to allow construction of novel systems.
Introduce a Framework which...

- ...covers all systems known to sparsify cartoons.
- ...enables easy transfer of (sparsity) results between systems.
- ...allows categorization of systems with respect to sparsity behaviors.
- ...is general enough to allow construction of novel systems.

Crucial Ingredient: Parabolic scaling, i.e., a scaling matrix of the type

$$A_{2j} = \begin{pmatrix} 2^j & 0 \\ 0 & 2^{j/2} \end{pmatrix}, \quad j \in \mathbb{Z},$$

since from

$$E(x_2) \approx \frac{1}{2} \kappa x_2^2 \quad \text{and} \quad E(\ell) = w$$

follows

$$w \approx \frac{\kappa}{2} \ell^2 \quad (\text{‘width } \approx \text{ length}^2).$$
Parametrization

Parameter space:

\[ \mathbb{P} := \mathbb{R}_+ \times \mathbb{T} \times \mathbb{R}^2, \]

where \((s, \theta, x) \in \mathbb{P}\) describes scale \(2^s\), orientation \(\theta\), and location \(x\).

Definition: A parametrization is a pair \((\Lambda, \Phi_\Lambda)\), where \(\Lambda\) is a discrete index set and \(\Phi_\Lambda\) is a mapping

\[
\Phi_\Lambda : \left\{ \begin{array}{c}
\Lambda \\ \lambda \end{array} \right\} \rightarrow \mathbb{P}, \quad \lambda \mapsto (s_\lambda, \theta_\lambda, x_\lambda).
\]
Parametrization

Parameter space:

\[ \mathbb{P} := \mathbb{R}_+ \times \mathbb{T} \times \mathbb{R}^2, \]

where \((s, \theta, x) \in \mathbb{P}\) describes scale \(2^s\), orientation \(\theta\), and location \(x\).

Definition: A parametrization is a pair \((\Lambda, \Phi_\Lambda)\), where \(\Lambda\) is a discrete index set and \(\Phi_\Lambda\) is a mapping

\[ \Phi_\Lambda : \left\{ \begin{array}{ll} \Lambda & \rightarrow \mathbb{P}, \\ \lambda & \mapsto (s_\lambda, \theta_\lambda, x_\lambda). \end{array} \right. \]

Example: The canonical parametrization \((\Lambda^0, \Phi^0(\lambda))\) is defined by

\[ \Lambda^0 := \left\{ (j, \ell, k) \in \mathbb{Z}^4 : j \geq 0, \ \ell = -2^{\lfloor \frac{j}{2} \rfloor - 1}, \ldots, 2^{\lfloor \frac{j}{2} \rfloor - 1} \right\}, \]

and

\[ \Phi^0(j, \ell, k) = (s_\lambda, \theta_\lambda, x_\lambda) = (j, \ell 2^{-\lfloor j/2 \rfloor} \pi, R_{-\theta_\lambda} A_{2^{-s_\lambda}} k). \]
Parabolic Molecules

Definition (Grohs, K; 2014):
Let \((\Lambda, \Phi_\Lambda)\) be a parametrization. Then \((m_\lambda)_{\lambda \in \Lambda}\) is a system of parabolic molecules of order \((L, M, N_1, N_2) \in (\mathbb{Z}_+ \cup \{\infty\})^2 \times \mathbb{Z}_+^2\), if, for all \(\lambda \in \Lambda\),

\[
m_\lambda(x) = 2^{3s_\lambda/4} g^{(\lambda)}(A_2^{s_\lambda} R_{\theta_\lambda} (x - x_\lambda)), \quad \Phi_\lambda(\lambda) = (s_\lambda, \theta_\lambda, x_\lambda),
\]

such that, for all \(|\beta| \leq L\),

\[
\left| \partial^\beta \hat{g}^{(\lambda)}(\xi) \right| \lesssim \min \left(1, 2^{-s_\lambda} + |\xi_1| + 2^{-s_\lambda/2}|\xi_2| \right)^M \langle |\xi| \rangle^{-N_1} \langle \xi_2 \rangle^{-N_2}.
\]

Control Parameters:
- \(L\): Spatial localization.
- \(M\): Number of directional (almost) vanishing moments.
- \(N_1, N_2\): Smoothness of \(m_\lambda\).
Parabolic Molecules

Definition (Grohs, K; 2014):
Let $(\Lambda, \Phi_{\Lambda})$ be a parametrization. Then $(m_\lambda)_{\lambda \in \Lambda}$ is a system of parabolic molecules of order $(L, M, N_1, N_2) \in (\mathbb{Z}_+ \cup \{\infty\})^2 \times \mathbb{Z}_+^2$, if, for all $\lambda \in \Lambda$,

$$m_\lambda(x) = 2^{3s_\lambda/4} g^{(\lambda)}(A^{s_\lambda} R_{\theta_\lambda} (x - x_\lambda)), \quad \Phi_{\Lambda}(\lambda) = (s_\lambda, \theta_\lambda, x_\lambda),$$

such that, for all $|\beta| \leq L$,

$$\left| \partial^\beta \hat{g}^{(\lambda)}(\xi) \right| \lesssim \min \left( 1, 2^{-s_\lambda} + |\xi_1| + 2^{-s_\lambda / 2} |\xi_2| \right)^M \langle |\xi| \rangle^{-N_1} \langle \xi_2 \rangle^{-N_2}.$$
Special Cases

This framework includes...

- Parabolic Frame (Smith; 1998)
- Second Generation Curvelets (Candès and Donoho; 2002)
- Curvelet Molecules (Candès and Demanet; 2002)
- Bandlimited Shearlets (K and Labate; 2006)
- Frame Decompositions (Borup and Nielsen; 2007)
- Shearlet Molecules (Guo and Labate; 2008)
- Compactly Supported Shearlets (Kittipoom, K, and Lim; 2012)
- ...

Gitta Kutyniok (TU Berlin)  Computational Harmonic Analysis  BMS Summer School’16
What about Wavelets, Ridgelets, ...?
Main Idea:

- Introduction of a parameter $\alpha \in [0, 1]$ to measure the amount of anisotropy.
- For $a > 0$, define

\[
A_{\alpha,a} = \begin{pmatrix} a & 0 \\ 0 & a^{\alpha} \end{pmatrix}.
\]

Illustration:

$\alpha = 0$ \hspace{1cm} $\frac{1}{2}$ \hspace{1cm} 1

Ridgelets \hspace{1cm} Curvelets/Shearlets \hspace{1cm} Wavelets
\(\alpha\)-Molecules

Definition (Grohs, Keiper, K, Schäfer; 2016):
Let \(\alpha \in [0, 1]\), and let \((\Lambda, \Phi_\Lambda)\) be a parametrization. Then \((m_\lambda)_{\lambda \in \Lambda}\) is a system of \(\alpha\)-molecules of order \((L, M, N_1, N_2) \in (\mathbb{Z}_+ \cup \{\infty\})^2 \times \mathbb{Z}_+^2\), if, for all \(\lambda \in \Lambda\),

\[
m_\lambda(x) = s_\lambda^{(1+\alpha)/2} g(\lambda) \left(A_{\alpha, s_\lambda} R_{\theta_\lambda} (x - x_\lambda)\right), \quad \Phi_\Lambda(\lambda) = (s_\lambda, \theta_\lambda, x_\lambda),
\]
such that, for all \(|\beta| \leq L\),

\[
\left| \partial^\beta \hat{g}(\lambda)(\xi) \right| \lesssim \min \left(1, s_\lambda^{-1} + |\xi_1| + s_\lambda^{-(1-\alpha)} |\xi_2| \right)^M \langle \xi \rangle^{-N_1} \langle \xi_2 \rangle^{-N_2}.
\]
**α-Molecules**

**Definition (Grohs, Keiper, K, Schäfer; 2016):**

Let \( \alpha \in [0, 1] \), and let \((\Lambda, \Phi_{\Lambda})\) be a parametrization. Then \((m_{\lambda})_{\lambda \in \Lambda}\) is a system of \(\alpha\)-molecules of order \((L, M, N_1, N_2) \in (\mathbb{Z}_+ \cup \{\infty\})^2 \times \mathbb{Z}_+^2\), if, for all \(\lambda \in \Lambda\),

\[
m_{\lambda}(x) = s_{\lambda}^{(1+\alpha)/2} g(\lambda) \left( A_{\alpha, s_{\lambda}} R_{\theta_{\lambda}} (x - x_{\lambda}) \right), \quad \Phi_{\Lambda}(\lambda) = (s_{\lambda}, \theta_{\lambda}, x_{\lambda}),
\]

such that, for all \(|\beta| \leq L\),

\[
\left| \partial^{\beta} \hat{g}(\lambda)(\xi) \right| \lesssim \min \left( 1, s_{\lambda}^{-1} + |\xi_1| + s_{\lambda}^{-(1-\alpha)} |\xi_2| \right)^M \langle |\xi| \rangle^{-N_1} \langle \xi_2 \rangle^{-N_2}.
\]

**Examples:**

- Wavelets \((\alpha = 1)\)
- Ridgelets \((\alpha = 0)\)
- Shearlets, parabolic molecules in general \((\alpha = \frac{1}{2})\)
- \(\alpha\)-Curvelets \((\alpha \in [0, 1])\)
Metric Properties of Parametrizations

Definition:
Let \( \alpha \in [0, 1] \), and let \((\Lambda, \Phi_{\Lambda})\) and \((\Delta, \Phi_{\Delta})\) be parametrizations. For \( \lambda \in \Lambda \) and \( \mu \in \Delta \), we define the index distance by

\[
\omega_{\alpha}(\lambda, \mu) := \omega_{\alpha}(\Phi_{\Lambda}(\lambda), \Phi_{\Delta}(\mu)) := \max \left\{ \frac{s_{\lambda}}{s_{\mu}}, \frac{s_{\mu}}{s_{\lambda}} \right\} (1 + d_{\alpha}(\lambda, \mu)),
\]

with \( d_{\alpha}(\lambda, \mu) \) defined by

\[
s_0^{2(1-\alpha)}|\theta_{\lambda} - \theta_{\mu}|^2 + s_0^{2\alpha}|x_{\lambda} - x_{\mu}|^2 + \frac{s_0^2}{1 + s_0^{2(1-\alpha)}|\theta_{\lambda} - \theta_{\mu}|^2}|\langle e_{\lambda}, x_{\lambda} - x_{\mu}\rangle|^2.
\]

where \( s_0 = \min\{s_{\lambda}, s_{\mu}\} \) and \( e_{\lambda} = (\cos(\theta_{\lambda}), -\sin(\theta_{\lambda}))^T \).

Remark: \( d_{\frac{1}{2}} \) is the Hart Smith’s phase space metric on \( \mathbb{T} \times \mathbb{R}^2 \).
Main Result: Decay of Cross-Grammian

Theorem (Grohs, Keiper, K, Schäfer; 2016):
Let $\alpha \in [0, 1]$, $N > 0$, and let $(m_\lambda)_{\lambda \in \Lambda}$, $(p_\mu)_{\mu \in \Delta}$ be systems of $\alpha$-molecules of order $(L, M, N_1, N_2)$ with
\[
L \geq 2N, \quad M > 3N - \frac{3 - \alpha}{2}, \quad N_1 \geq N + \frac{1 + \alpha}{2}, \quad N_2 \geq 2N.
\]
Then, for all $\lambda \in \Lambda$ and $\mu \in \Delta$,
\[
|\langle m_\lambda, p_\mu \rangle| \lesssim \omega_\alpha (\lambda, \mu)^{-N}.
\]
...towards Sparse Approximation Properties!
Sparsity Equivalence

**Definition:**
Let \((m_\lambda)_{\lambda \in \Lambda}\) and \((p_\mu)_{\mu \in \Delta}\) be systems of \(\alpha\)-molecules of order \((L, M, N_1, N_2)\) and \((\tilde{L}, \tilde{M}, \tilde{N}_1, \tilde{N}_2)\), respectively, and let \(0 < p \leq 1\). If

\[
\left\| \left( \langle m_\lambda, p_\mu \rangle \right)_{\lambda \in \Lambda, \mu \in \Delta} \right\|_{\ell^p \to \ell^p} < \infty,
\]
then \((m_\lambda)_{\lambda \in \Lambda}\) and \((p_\mu)_{\mu \in \Delta}\) are sparsity equivalent in \(\ell^p\).
Sparsity Equivalence

Definition:
Let \((m_\lambda)_{\lambda \in \Lambda}\) and \((p_\mu)_{\mu \in \Delta}\) be systems of \(\alpha\)-molecules of order \((L, M, N_1, N_2)\) and \((\tilde{L}, \tilde{M}, \tilde{N}_1, \tilde{N}_2)\), respectively, and let \(0 < p \leq 1\). If
\[
\left\| \left( \langle m_\lambda, p_\mu \rangle \right)_{\lambda \in \Lambda, \mu \in \Delta} \right\|_{\ell^p \rightarrow \ell^p} < \infty,
\]
then \((m_\lambda)_{\lambda \in \Lambda}\) and \((p_\mu)_{\mu \in \Delta}\) are sparsity equivalent in \(\ell^p\).

Definition:
Let \(\alpha \in [0, 1]\) and \(k > 0\). Two parametrizations \((\Lambda, \Phi_\Lambda)\) and \((\Delta, \Phi_\Delta)\) are \((\alpha, k)\)-consistent, if
\[
\sup_{\lambda \in \Lambda} \sum_{\mu \in \Delta} \omega_\alpha (\lambda, \mu)^{-k} < \infty \quad \text{and} \quad \sup_{\mu \in \Delta} \sum_{\lambda \in \Lambda} \omega_\alpha (\lambda, \mu)^{-k} < \infty.
\]
Sufficient Condition for Sparsity Equivalence

Theorem (Grohs, Keiper, K, Schäfer; 2016):
Let $0 < p \leq 1$, and let $(m_\lambda)_{\lambda \in \Lambda}$ and $(p_\mu)_{\mu \in \Delta}$ be frames of $\alpha$-molecules of order $(L, M, N_1, N_2)$ with $(\alpha, k)$-consistent parametrizations $(\Lambda, \Phi_\Lambda)$ and $(\Delta, \Phi_\Delta)$ for some $k > 0$. If

$$L \geq 2 \frac{k}{p}, \quad M > 3 \frac{k}{p} - \frac{3 - \alpha}{2}, \quad N_1 \geq \frac{k}{p} + \frac{1 + \alpha}{2}, \quad N_2 \geq 2 \frac{k}{p},$$

then $(m_\lambda)_{\lambda \in \Lambda}$ and $(p_\mu)_{\mu \in \Delta}$ are sparsity equivalent in $\ell^p$. 
Strategy

α-Curvelets

Shearlets

\( \alpha = 0 \)

\( \alpha = \frac{1}{2} \)

\( \alpha = 1 \)

Wavelets

α-Curvelets

Shearlet Molecules

Ridgelets

\( \alpha = \frac{1}{2} \)

Shearlet Molecules

\( \alpha = 0 \)

\( \alpha = \frac{1}{2} \)
Generalized Image Model

Definition (K, Lemvig, Lim; 2012), (Keiper; 2012):
The set of cartoon-like functions $\mathcal{E}^\beta(\mathbb{R}^2)$, $\beta \in (1, 2]$ is defined by

$$\mathcal{E}^\beta(\mathbb{R}^2) = \{ f \in L^2(\mathbb{R}^2) : f = f_0 + f_1 \cdot \chi_B \}$$

where $B \subset [0, 1]^2$ with $\partial B$ a closed $C^\beta$-curve, $f_0, f_1 \in C^\beta_0([0, 1]^2)$. 
Generalized Image Model

Definition (K, Lemvig, Lim; 2012), (Keiper; 2012): The set of cartoon-like functions $E^\beta(\mathbb{R}^2)$, $\beta \in (1, 2]$ is defined by

$$E^\beta(\mathbb{R}^2) = \{ f \in L^2(\mathbb{R}^2) : f = f_0 + f_1 \cdot \chi_B \},$$

where $B \subset [0, 1]^2$ with $\partial B$ a closed $C^\beta$-curve, $f_0, f_1 \in C^\beta_0([0, 1]^2)$.

Theorem (Grohs, Keiper, K, Schäfer; 2016): Let $\alpha \in [\frac{1}{2}, 1)$, $\beta = \alpha^{-1}$. The Parseval frame of $\alpha$-curvelets provides an optimally sparse approximation of $f \in E^\beta(\mathbb{R}^2)$, i.e.,

$$\| f - f_N \|_2^2 \leq C \cdot N^{-\beta} \cdot (\log N)^{\beta+1}, \quad N \to \infty.$$
Sparse Approximation with $\alpha$-Molecules

Theorem (Grohs, Keiper, K, Schäfer; 2016):

Let $\alpha \in [\frac{1}{2}, 1)$, $\beta = \alpha^{-1}$, and let $(m_\lambda)_{\lambda \in \Lambda}$ be a system of $\alpha$-molecules of order $(L, M, N_1, N_2)$ such that

(i) $(m_\lambda)_{\lambda \in \Lambda}$ constitutes a frame for $L^2(\mathbb{R}^2)$,

(ii) $(\Lambda, \Phi_\Lambda)$ is $(\alpha, k)$-consistent with the parametrization of $\alpha$-curvelets for all $k > 0$,

(iii) it holds that

$$L \geq k(1+\beta), \quad M \geq \frac{3k}{2}(1+\beta) + \frac{\alpha - 3}{2}, \quad N_1 \geq \frac{k}{2}(1+\beta) + \frac{1 + \alpha}{2}, \quad N_2 \geq k(1+\beta).$$

Then, for any $\varepsilon > 0$ and for any $f \in \mathcal{E}^\beta(\mathbb{R}^2)$, $(m_\lambda)_{\lambda \in \Lambda}$ satisfies

$$\|f - f_N\|_2^2 \leq C \cdot N^{-\beta + \varepsilon}, \quad N \to \infty,$$
Let’s conclude...
What to take Home...?

- **Computational Harmonic Analysis** provides various representation systems such as wavelets, ridgelets, curvelets, and shearlets.
- They provide **sparse approximation** for certain classes of images, leading to
  - Efficient decompositions for, e.g., the analysis/processing of images.
  - Sparse representations for, e.g., regularization of inverse problems.
- **Shearlets** provide an optimally sparsifying system for a model class of functions being governed by **anisotropic features**.
- **α-Molecules** provide a general framework for various systems from computational harmonic analysis.
- **Sparse approximation** results can be derived in a unified manner.
THANK YOU!

References available at:

www.math.tu-berlin.de/~kutyniok

Code available at:

www.ShearLab.org

Related Books:

- Y. Eldar and G. Kutyniok
  *Compressed Sensing: Theory and Applications*
- G. Kutyniok and D. Labate
  *Shearlets: Multiscale Analysis for Multivariate Data*