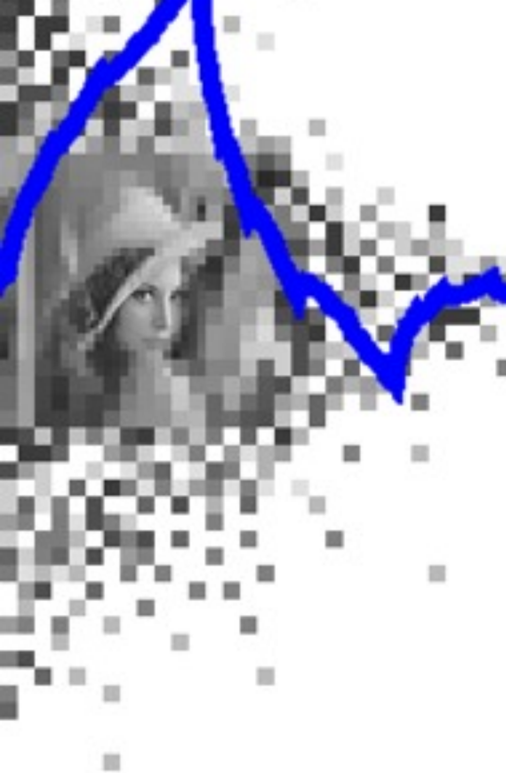


Low Complexity Regularization of Inverse Problems

Cours #1
Inverse Problems

Gabriel Peyré

CNRS

www.numerical-tours.com

DAUPHINE UNIVERSITÉ PARIS

Overview of the Course

- Course #1: Inverse Problems
- Course #2: Recovery Guarantees
- Course #3: Proximal Splitting Methods

Overview

- Inverse Problems
- Compressed Sensing
- Sparsity and L1 Regularization

Inverse Problems

Recovering $x_0 \in \mathbb{R}^N$ from noisy observations

$$y = \Phi x_0 + w \in \mathbb{R}^P$$

$\Phi : \mathbb{R}^N \mapsto \mathbb{R}^P$ with $P \ll N$ (missing information)

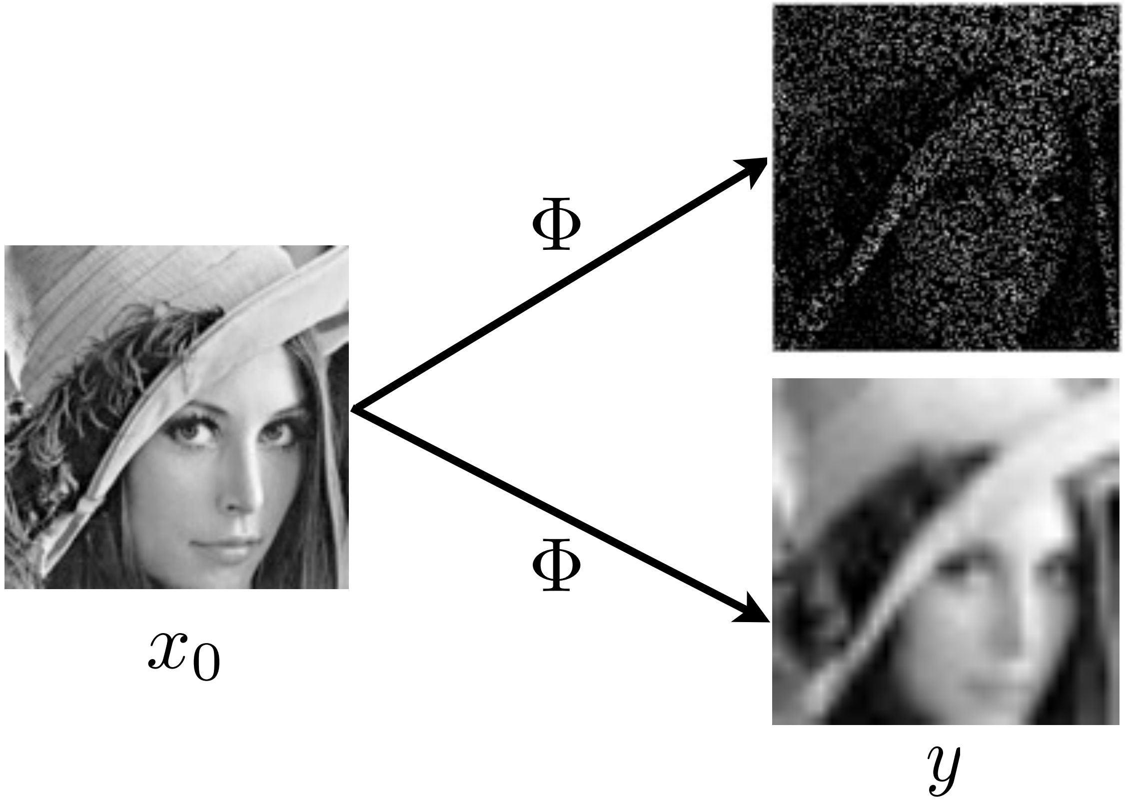
Inverse Problems

Recovering $x_0 \in \mathbb{R}^N$ from noisy observations

$$y = \Phi x_0 + w \in \mathbb{R}^P$$

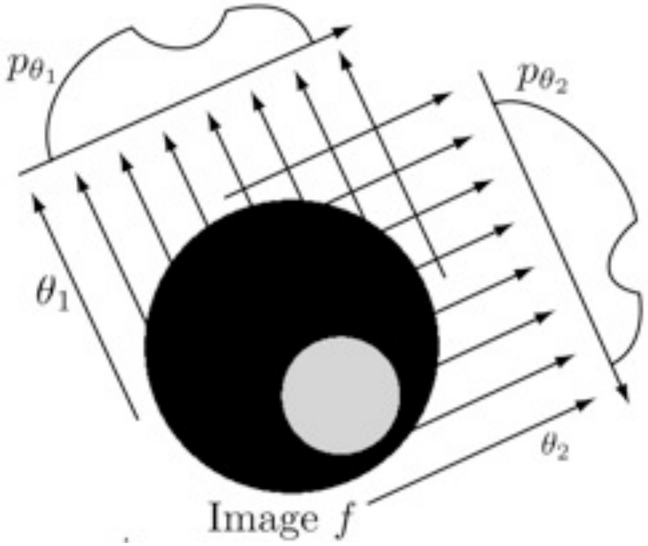
$\Phi : \mathbb{R}^N \mapsto \mathbb{R}^P$ with $P \ll N$ (missing information)

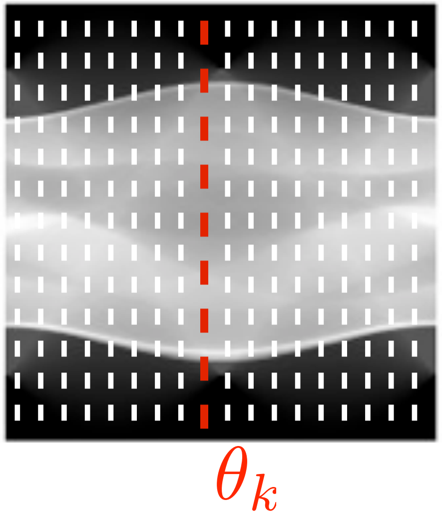
Examples: Inpainting, super-resolution, …



Inverse Problems in Medical Imaging

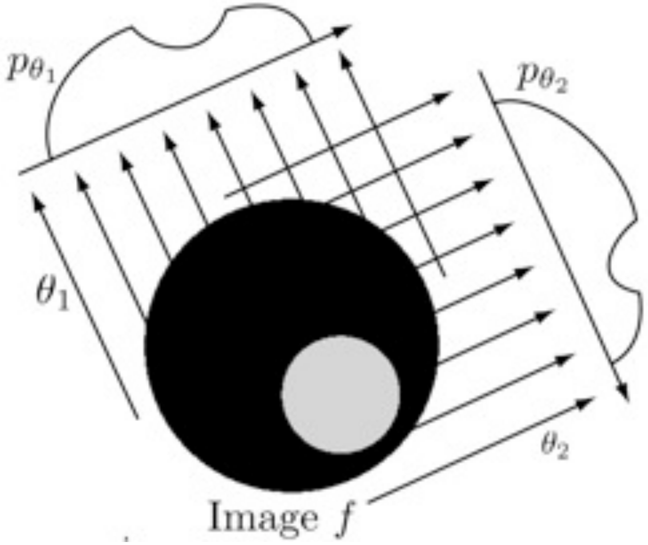
Tomography projection: $\Phi x = (p_{\theta_k})_{1 \leqslant k \leqslant K}$

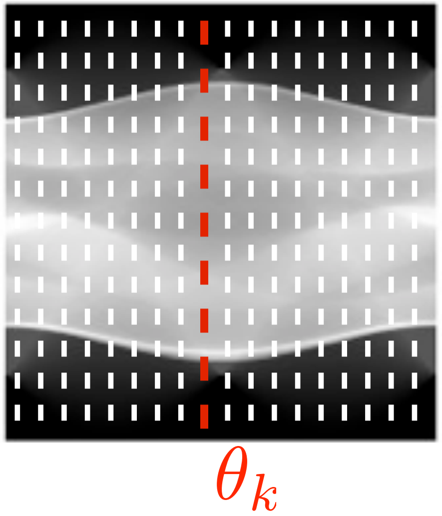




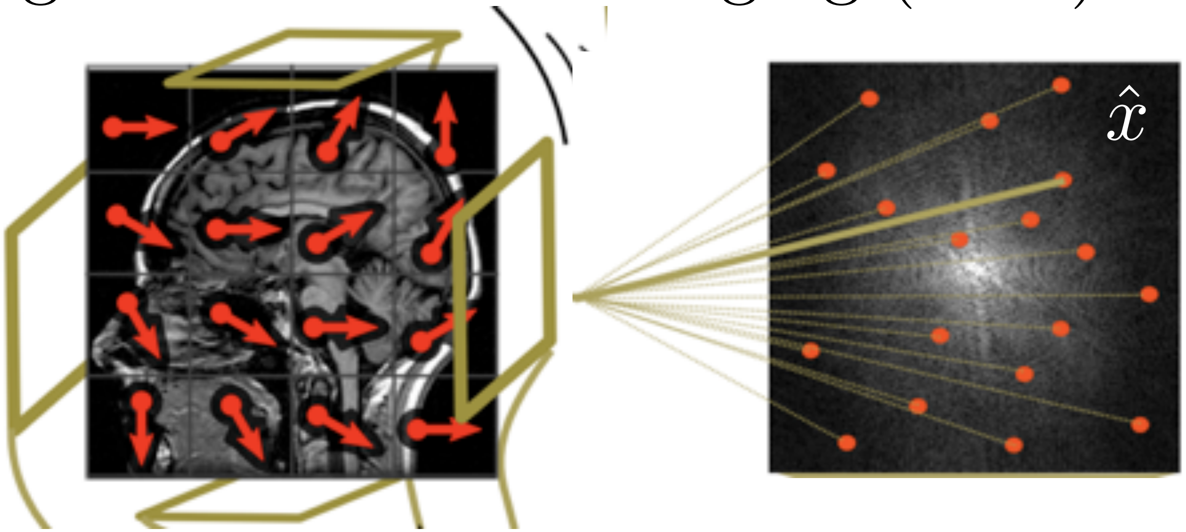
Inverse Problems in Medical Imaging

Tomography projection: $\Phi x = (p_{\theta_k})_{1 \leqslant k \leqslant K}$



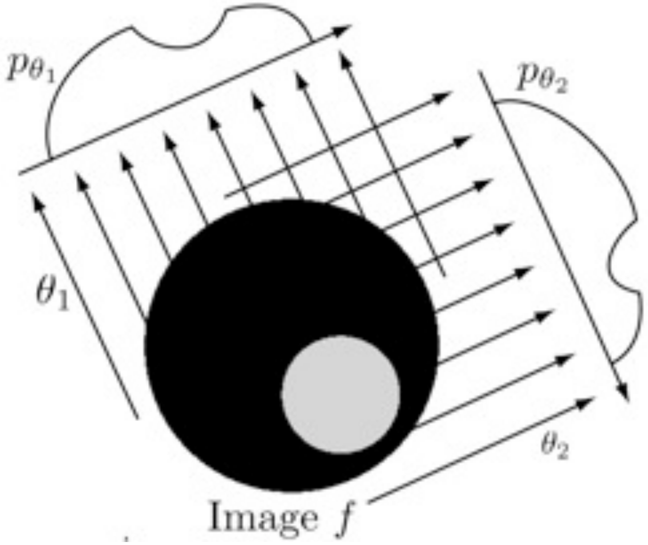


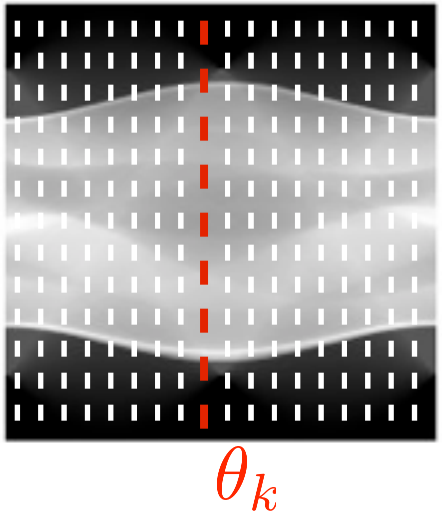
Magnetic resonance imaging (MRI): $\Phi x = (\hat{f}(\omega))_{\omega \in \Omega}$



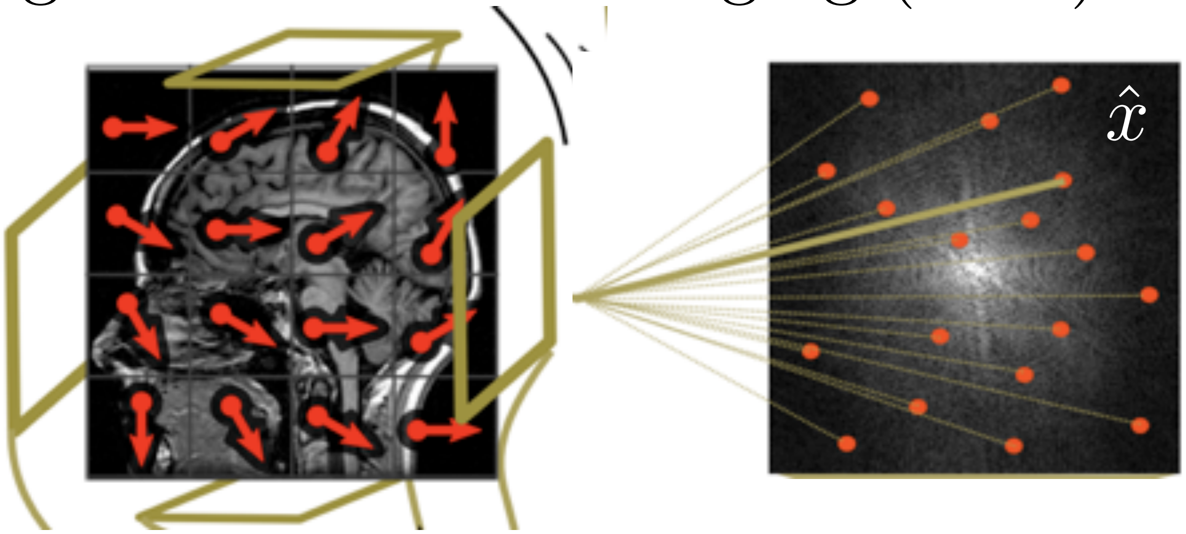
Inverse Problems in Medical Imaging

Tomography projection: $\Phi x = (p_{\theta_k})_{1 \leqslant k \leqslant K}$





Magnetic resonance imaging (MRI): $\Phi x = (\hat{f}(\omega))_{\omega \in \Omega}$



Other examples: MEG, EEG, ...

Inverse Problem Regularization

Observations: $y = \Phi x_0 + w \in \mathbb{R}^P$.

Estimator: $x(y)$ depends only on

- observations y
- parameter λ

Inverse Problem Regularization

Observations: $y = \Phi x_0 + w \in \mathbb{R}^P$.

Estimator: $x(y)$ depends only on

observations y

parameter λ

Example: variational methods

$$x(y) \in \underset{x \in \mathbb{R}^N}{\text{argmin}} \underbrace{\boxed{\frac{1}{2}\|y - \Phi x\|^2}}_{\text{Data fidelity}} + \lambda \underbrace{\boxed{J(x)}}_{\text{Regularity}}$$

Inverse Problem Regularization

Observations: $y = \Phi x_0 + w \in \mathbb{R}^P$.

Estimator: $x(y)$ depends only on
- observations y
- parameter λ

Example: variational methods

$$x(y) \in \underset{x\in\mathbb{R}^N}{\operatorname{argmin}} \underset{\text{Data fidelity}}{\boxed{\frac{1}{2}\|y - \Phi x\|^2}} + \lambda \underset{\text{Regularity}}{\boxed{J(x)}}$$

Choice of λ: tradeoff

$$\boxed{\begin{array}{c}\text{Noise level}\\ \|w\|\end{array}} \longleftrightarrow \boxed{\begin{array}{c}\text{Regularity of } x_0\\ J(x_0)\end{array}}$$

Inverse Problem Regularization

Observations: $y = \Phi x_0 + w \in \mathbb{R}^P$.

Estimator: $x(y)$ depends only on

- observations y
- parameter λ

Example: variational methods

$$x(y) \in \underset{x\in\mathbb{R}^N}{\text{argmin}} \; \underset{\text{Data fidelity}}{\boxed{\frac{1}{2}\|y-\Phi x\|^2}} + \lambda \; \underset{\text{Regularity}}{\boxed{J(x)}}$$

Choice of λ: tradeoff

$$\boxed{\begin{array}{c}\text{Noise level}\\ \|w\|\end{array}} \longleftrightarrow \boxed{\begin{array}{c}\text{Regularity of } x_0\\ J(x_0)\end{array}}$$

No noise: $\lambda \to 0^+$, minimize $\quad x^\star \in \underset{x\in\mathbb{R}^Q, \Phi x = y}{\text{argmin}} \; J(x)$

Inverse Problem Regularization

Observations: $y = \Phi x_0 + w \in \mathbb{R}^P$.

Estimator: $x(y)$ depends only on
- observations y
- parameter λ

Example: variational methods

$$x(y) \in \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \underbrace{\boxed{\frac{1}{2}\|y - \Phi x\|^2}}_{\text{Data fidelity}} + \lambda \underbrace{\boxed{J(x)}}_{\text{Regularity}}$$

Choice of λ: tradeoff $\boxed{\text{Noise level } \|w\|} \longleftrightarrow \boxed{\text{Regularity of } x_0 \; J(x_0)}$

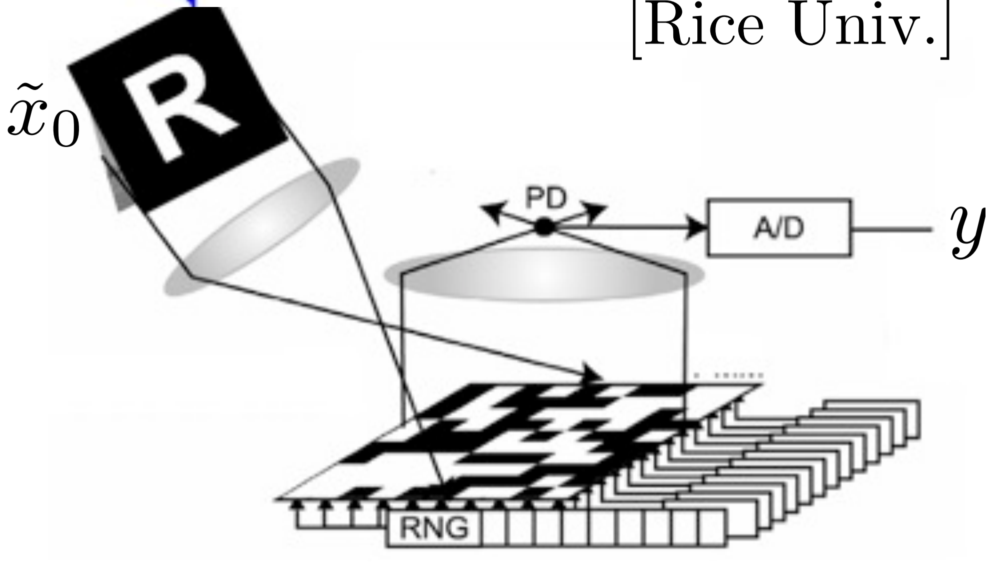
No noise: $\lambda \rightarrow 0^+$, minimize $\quad x^\star \in \underset{x \in \mathbb{R}^Q, \Phi x = y}{\operatorname{argmin}} J(x)$

This course:
- **Performance analysis.**
- **Fast computational scheme.**

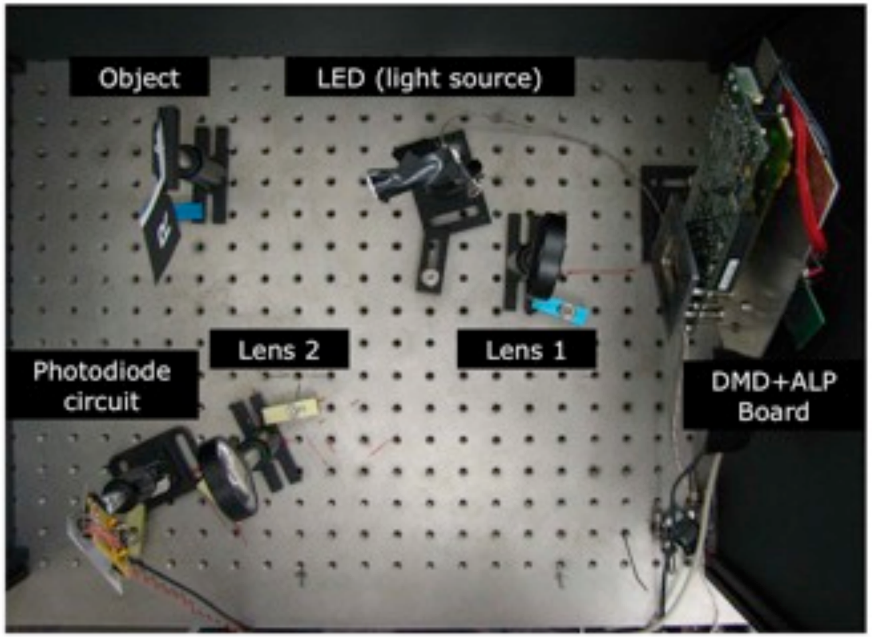
Overview

- Inverse Problems
- **Compressed Sensing**
- Sparsity and L1 Regularization

Compressed Sensing

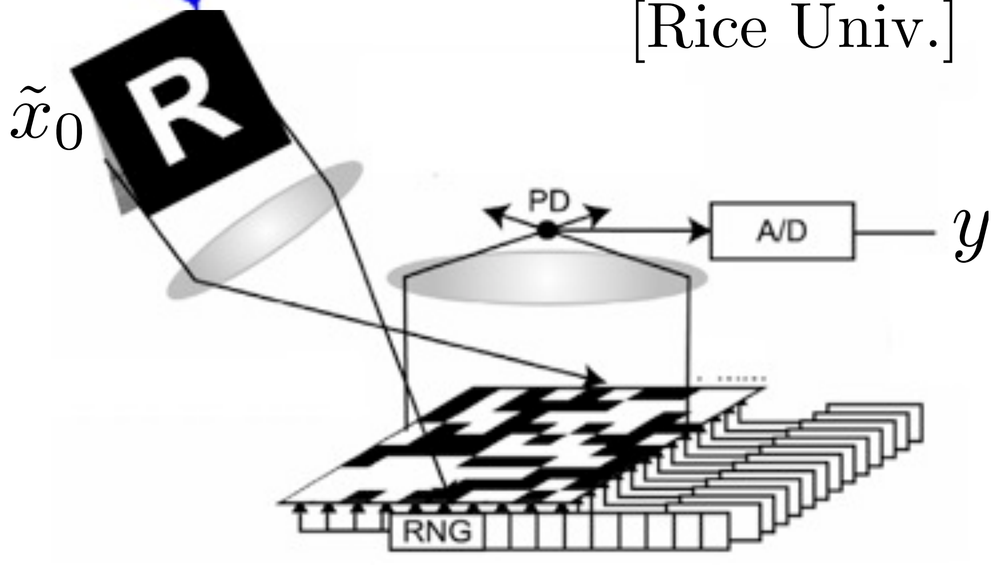


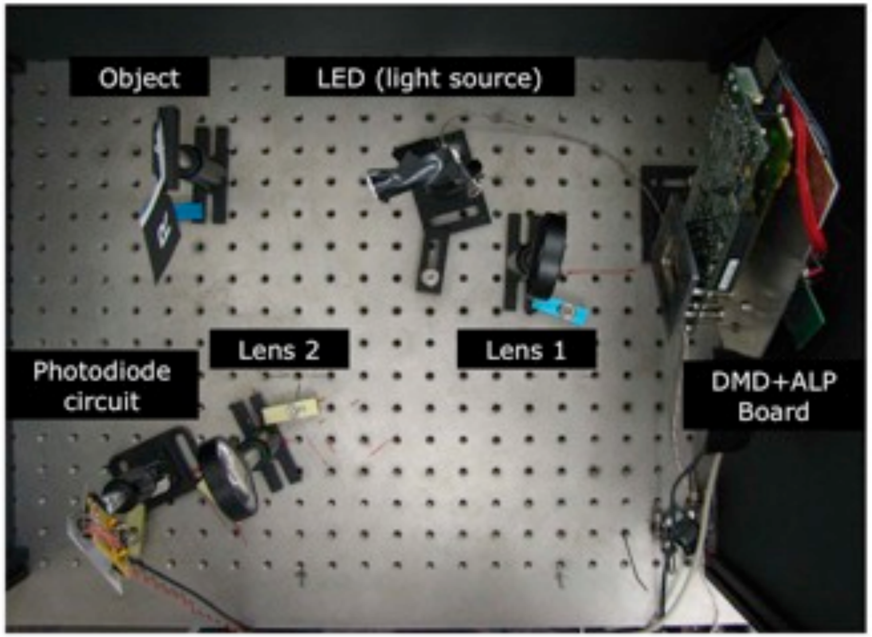
[Rice Univ.]



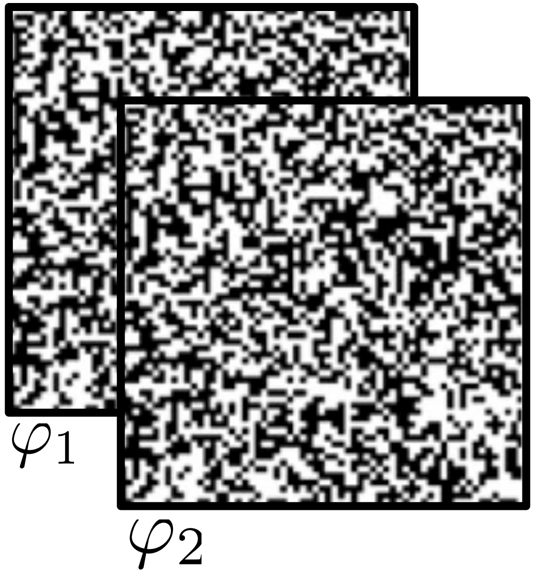
Compressed Sensing

[Rice Univ.]





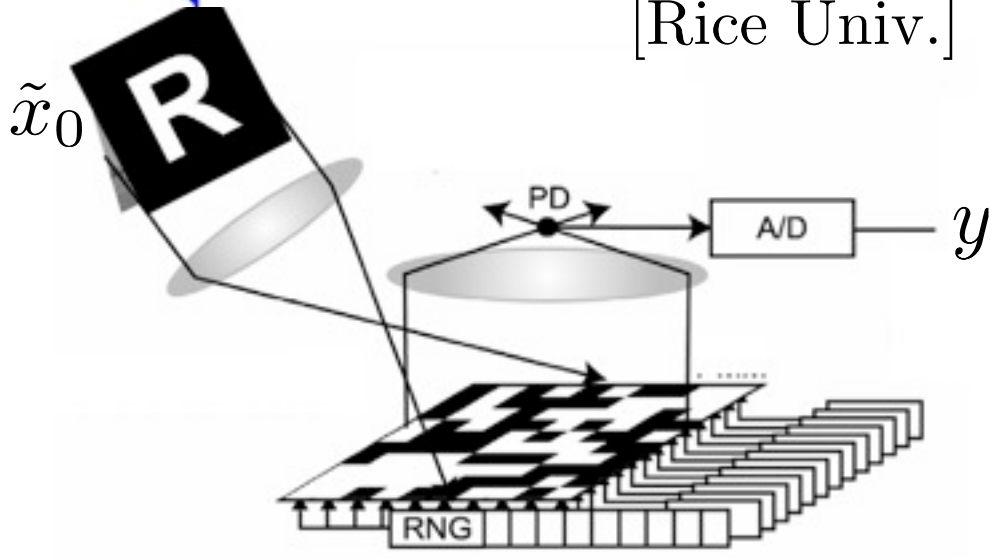
$$y[i] = \langle x_0,\, \varphi_i \rangle$$

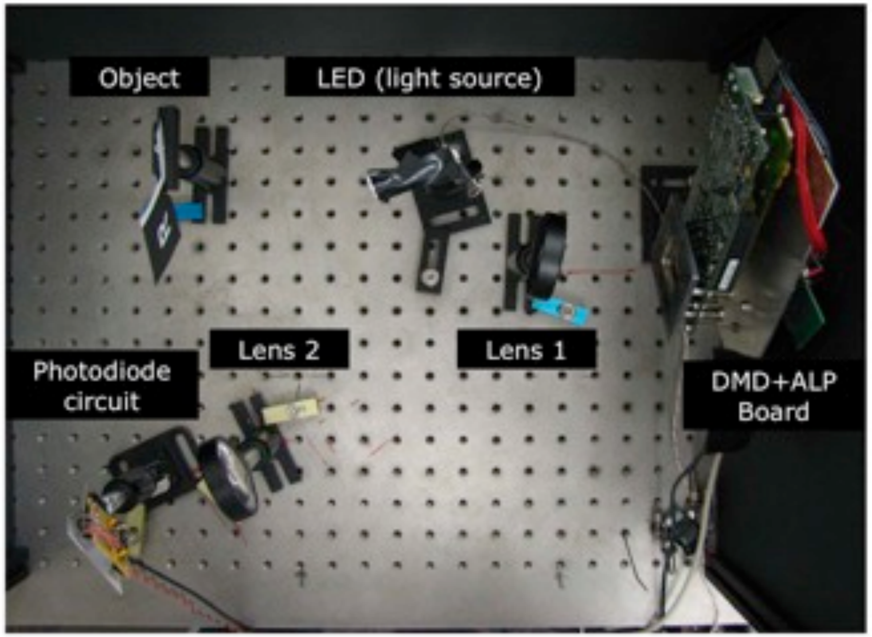


P measures $\ll N$ micro-mirrors

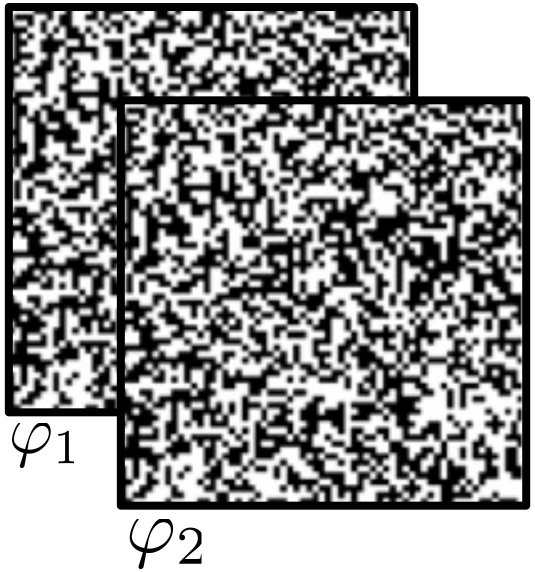
Compressed Sensing

[Rice Univ.]





$$y[i] = \langle x_0,\, \varphi_i \rangle$$



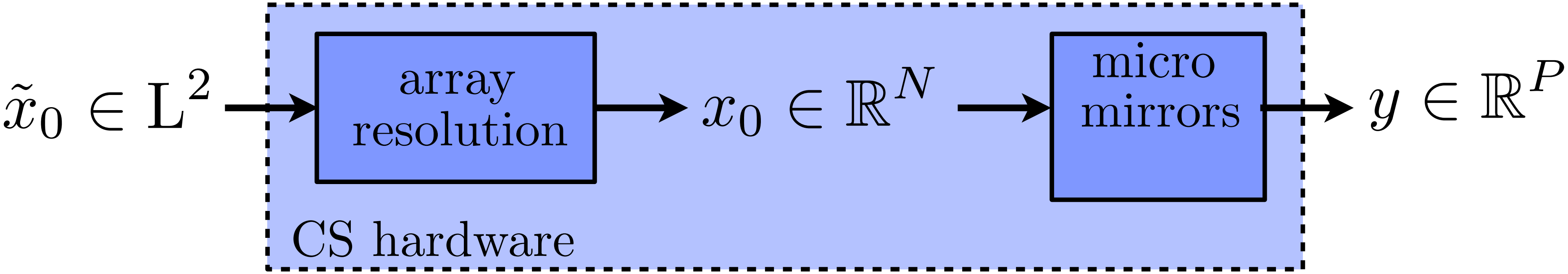
P measures $\ll N$ micro-mirrors

$P/N = 1$ $\quad$ $P/N = 0.16$ $\quad$ $P/N = 0.02$

CS Acquisition Model

CS is about designing hardware: input signals $\tilde{f} \in \mathrm{L}^2(\mathbb{R}^2)$.

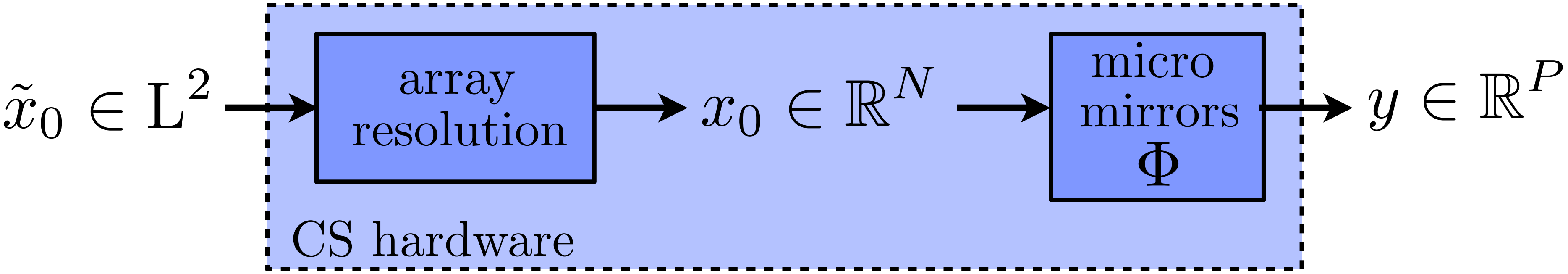
Physical hardware resolution limit: target resolution $f \in \mathbb{R}^N$.

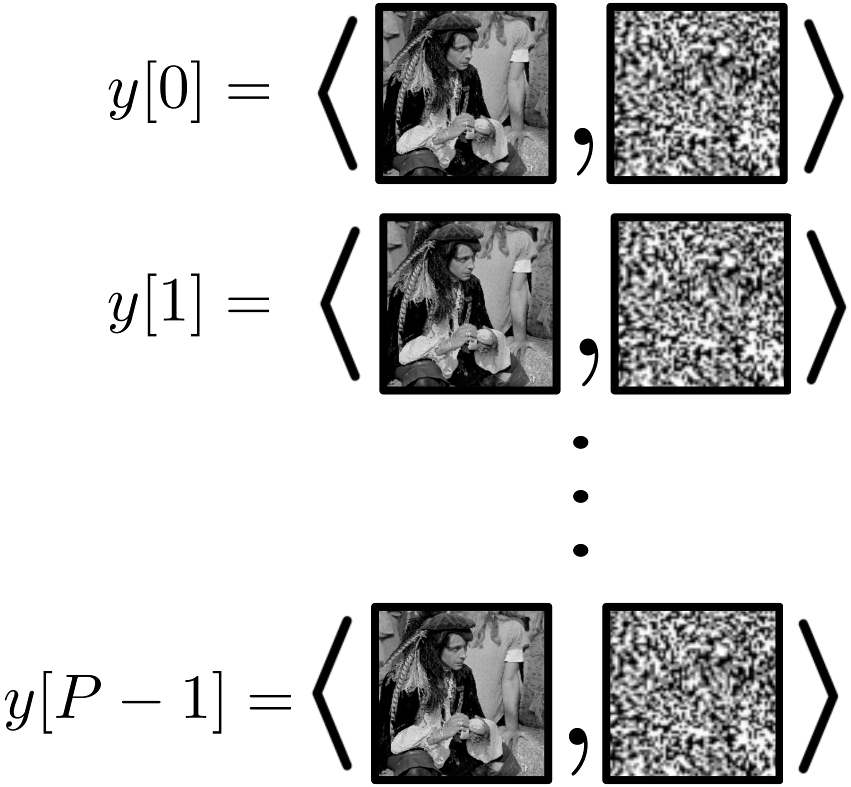


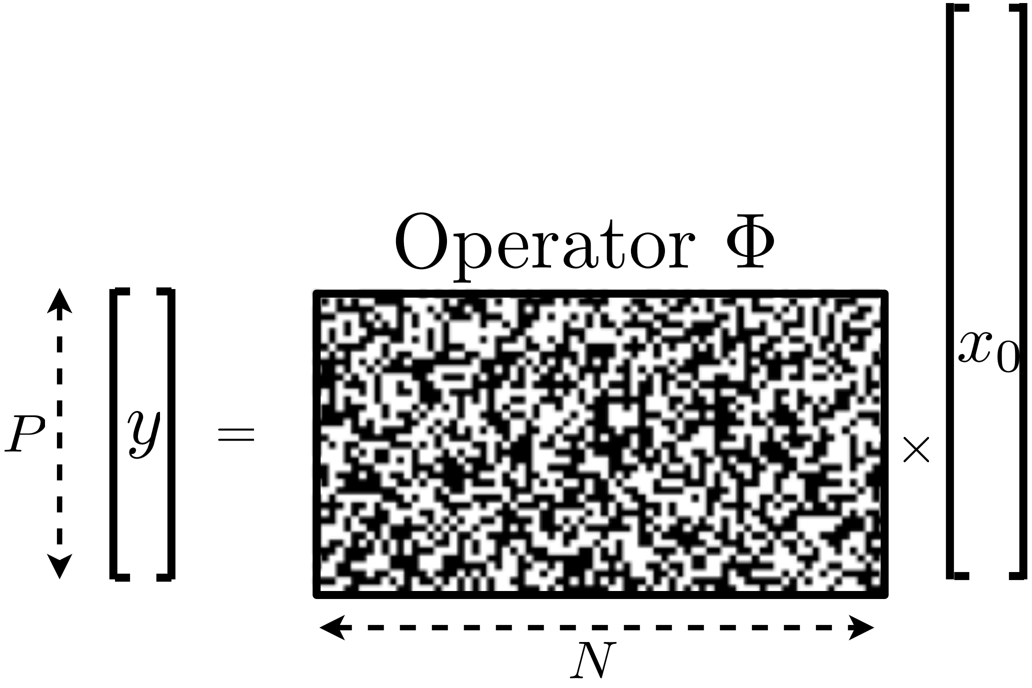
CS Acquisition Model

CS is about designing hardware: input signals $\tilde{f} \in \mathrm{L}^2(\mathbb{R}^2)$.

Physical hardware resolution limit: target resolution $f \in \mathbb{R}^N$.





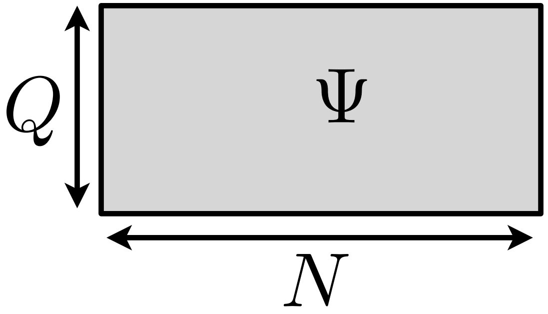


Overview

- Inverse Problems
- Compressed Sensing
- **Sparsity and L1 Regularization**

Redundant Dictionaries

Dictionary $\Psi = (\psi_m)_m \in \mathbb{R}^{Q \times N}$, $N \geqslant Q$.

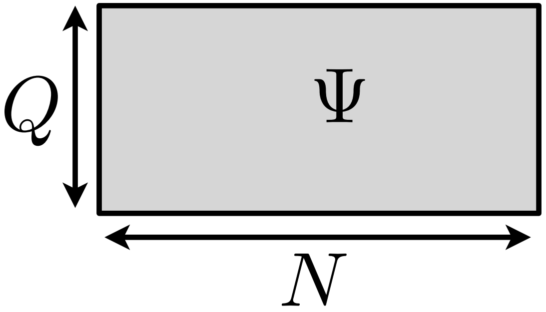


Redundant Dictionaries

Dictionary $\Psi = (\psi_m)_m \in \mathbb{R}^{Q \times N}$, $N \geqslant Q$.

Fourier: $\psi_m = e^{i\langle \cdot,\, m \rangle}$

frequency



Redundant Dictionaries

Dictionary $\Psi = (\psi_m)_m \in \mathbb{R}^{Q \times N}$, $N \geqslant Q$.

$$m = ({\color{blue}j}, {\color{red}\ell}, {\color{green}n})$$

scale

orientation

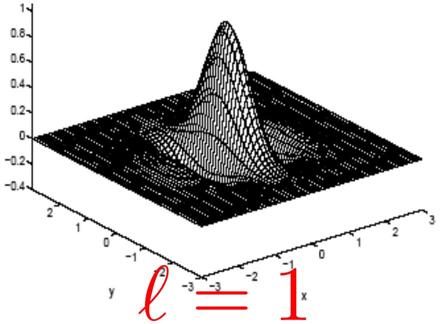
position

Fourier: $\psi_m = e^{i\langle \cdot,\, {\color{red}m}\rangle}$

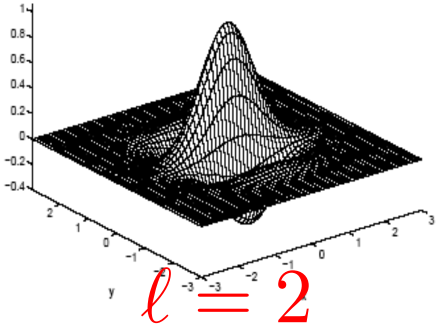
frequency

Wavelets:

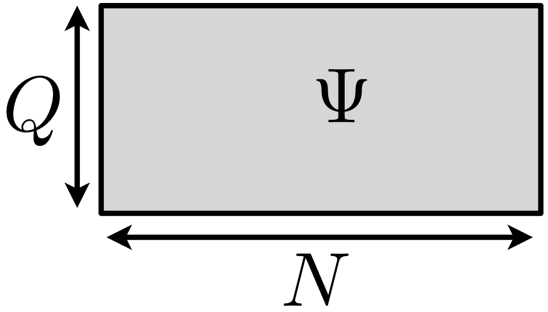
$$\psi_m = \psi({\color{blue}2^{-j}} {\color{red}R_{\theta_\ell}} x - {\color{green}n})$$



$\ell = 1$



$\ell = 2$



Redundant Dictionaries

Dictionary $\Psi = (\psi_m)_m \in \mathbb{R}^{Q \times N}$, $N \geqslant Q$.

$$m = (j, \ell, n)$$

scale

orientation

position

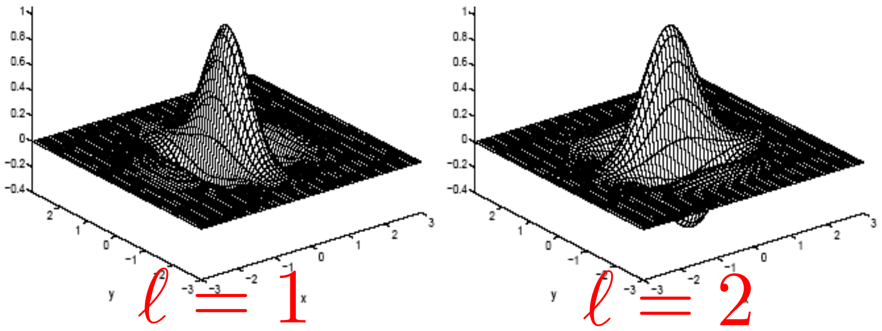
Fourier: $\psi_m = e^{i\langle \cdot, m \rangle}$

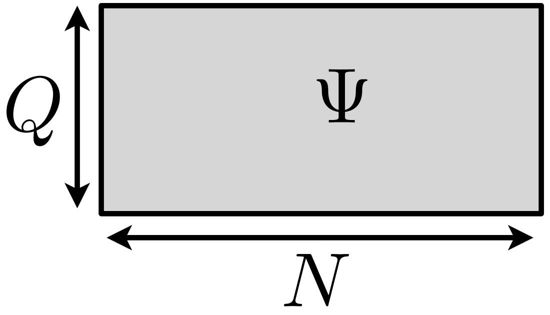
frequency

Wavelets:

$$\psi_m = \psi(2^{-j} R_{\theta_\ell} x - n)$$

DCT, Curvelets, bandlets, …


$\ell = 1$ $\ell = 2$



Redundant Dictionaries

Dictionary $\Psi = (\psi_m)_m \in \mathbb{R}^{Q \times N}$, $N \geqslant Q$.

$m = (j, \ell, n)$

scale

orientation

position

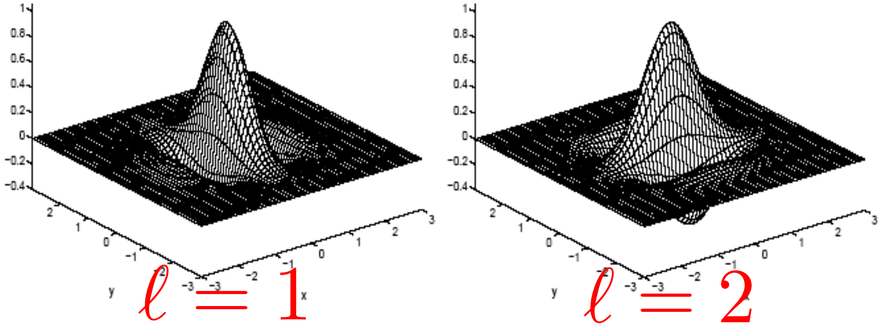
Fourier: $\psi_m = e^{i\langle \cdot, m \rangle}$

frequency

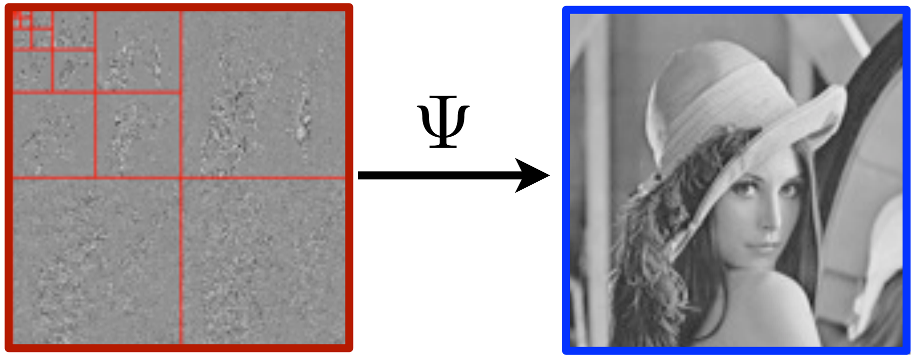
Wavelets:

$$\psi_m = \psi(2^{-j} R_{\theta_\ell} x - n)$$

DCT, Curvelets, bandlets, ...

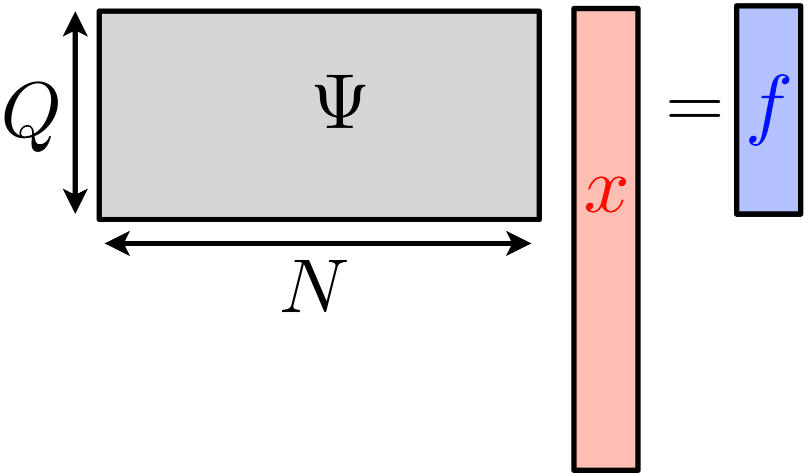


Synthesis: $f = \sum_m x_m \psi_m = \Psi x$.



Coefficients x

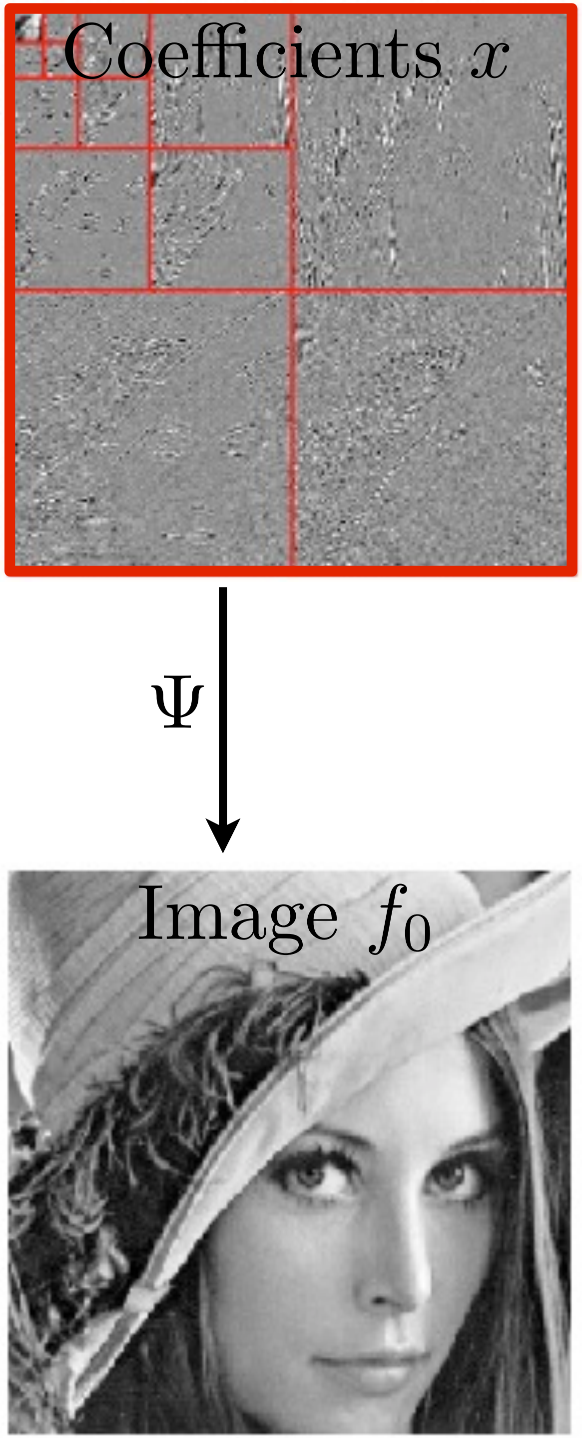
Image $f = \Psi x$



Sparse Priors

Ideal sparsity: for most m, $x_m = 0$.

$$J_0(x) = \# \{ m \setminus x_m \neq 0 \}$$



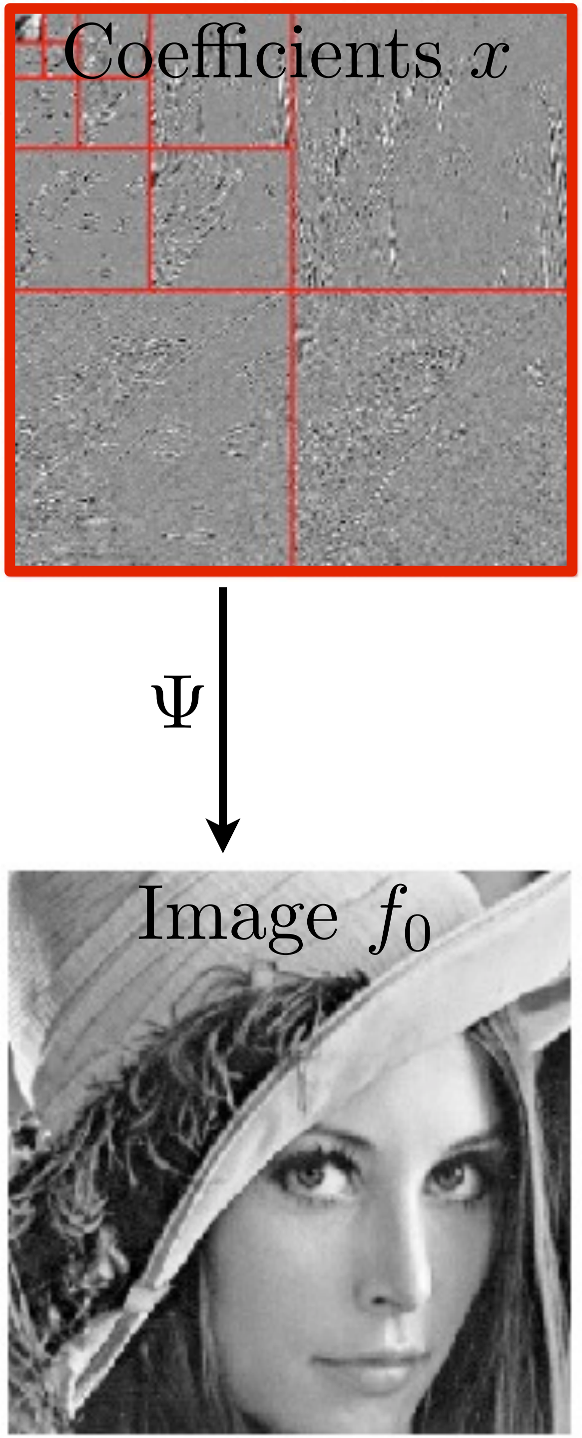
Sparse Priors

Ideal sparsity: for most m, $x_m = 0$.

$$J_0(x) = \#\{m \setminus x_m \neq 0\}$$

Sparse approximation: $f = \Psi x$ where

$$\underset{x \in \mathbb{R}^N}{\text{argmin}} \|f_0 - \Psi x\|^2 + T^2 J_0(x)$$



Sparse Priors

Ideal sparsity: for most m, $x_m = 0$.

$$J_0(x) = \#\{m \setminus x_m \neq 0\}$$

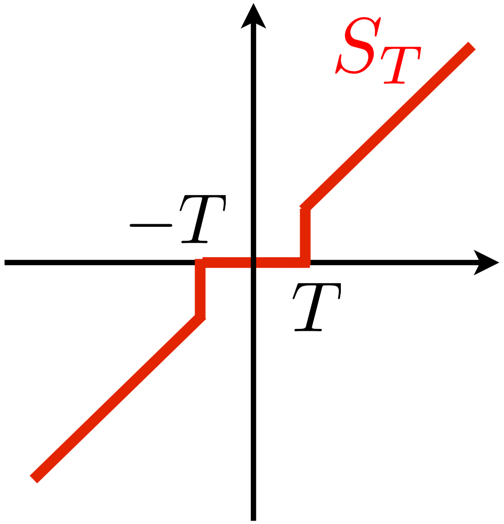
Sparse approximation: $f = \Psi x$ where

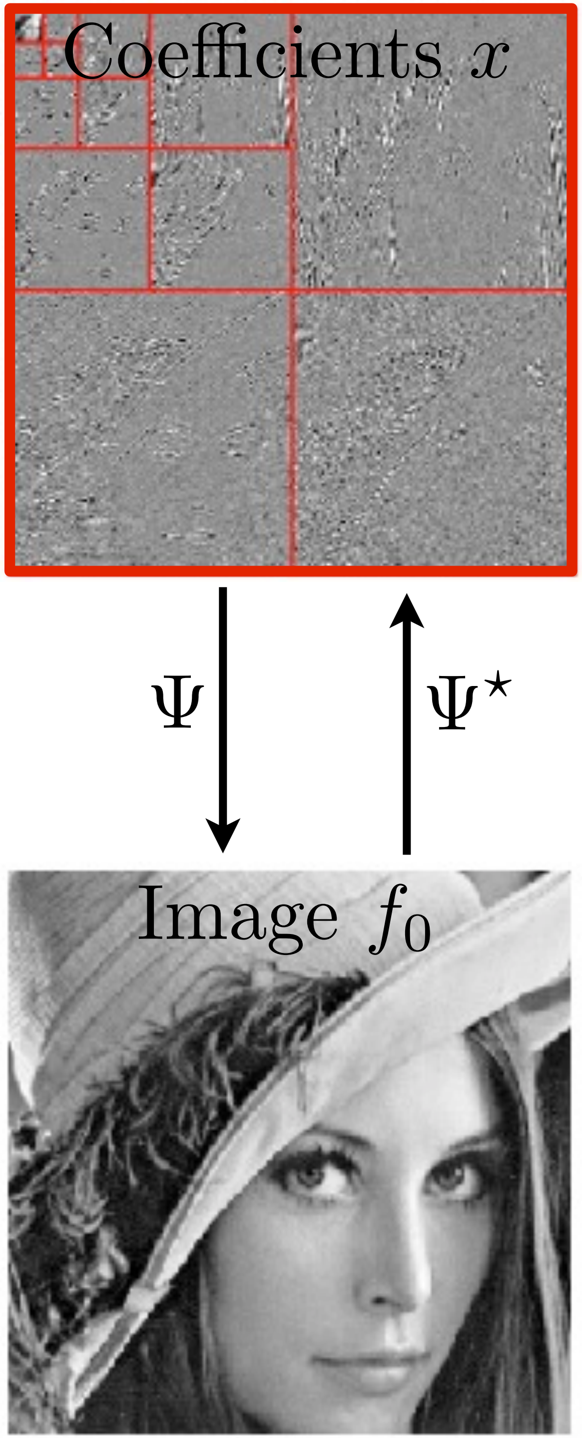
$$\underset{x \in \mathbb{R}^N}{\text{argmin}} \; \|f_0 - \Psi x\|^2 + T^2 J_0(x)$$

Orthogonal Ψ*:* $\Psi^* \Psi = \Psi \Psi^* = \text{Id}_N$

$$x_m = \begin{cases} \langle f_0, \psi_m \rangle & \text{if} \quad |\langle f_0, \psi_m \rangle| > T, \\ 0 & \text{otherwise.} \end{cases}$$

$$f = \Psi \circ S_T \circ \Psi^*(f_0)$$





Sparse Priors

Ideal sparsity: for most m, $x_m = 0$.

$$J_0(x) = \#\{m \setminus x_m \neq 0\}$$

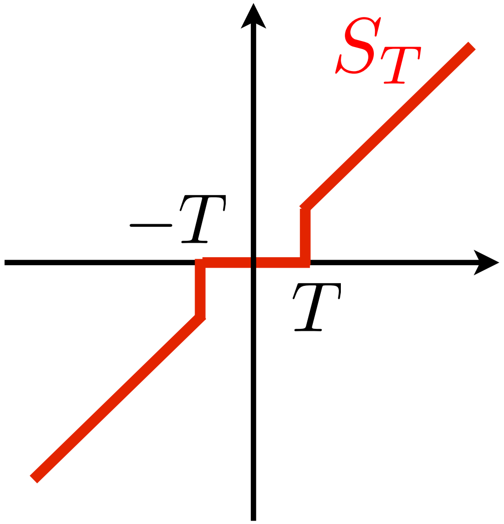
Sparse approximation: $f = \Psi x$ where

$$\underset{x \in \mathbb{R}^N}{\text{argmin}} \, \|f_0 - \Psi x\|^2 + T^2 J_0(x)$$

Orthogonal Ψ*:* $\Psi^* \Psi = \Psi \Psi^* = \text{Id}_N$

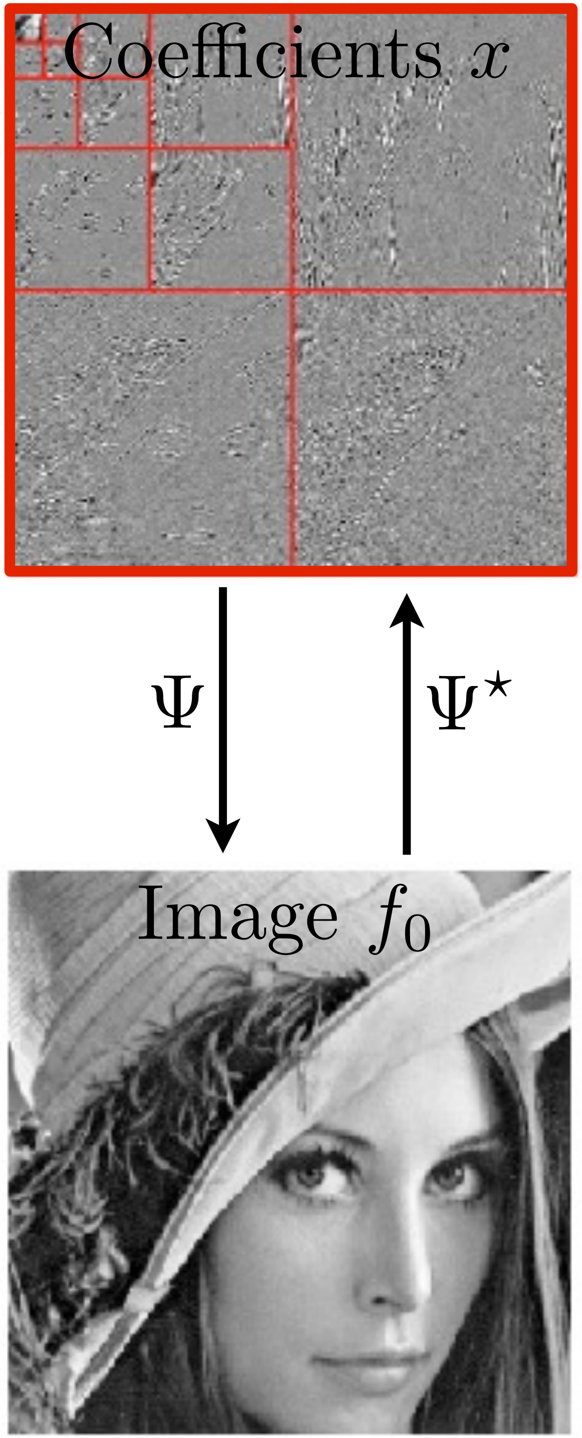
$$x_m = \begin{cases} \langle f_0, \psi_m \rangle & \text{if} \quad |\langle f_0, \psi_m \rangle| > T, \\ 0 & \text{otherwise.} \end{cases}$$

$$f = \Psi \circ S_T \circ \Psi^*(f_0)$$



Non-orthogonal Ψ*:*

$\longrightarrow$ NP-hard.

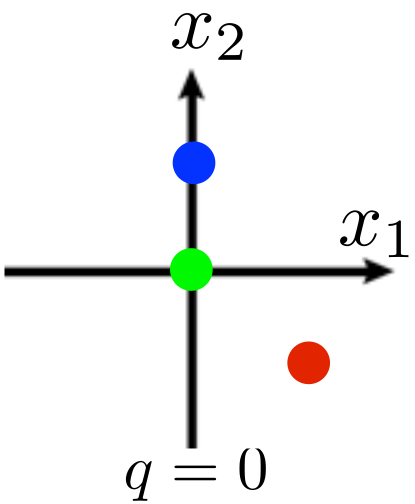


Convex Relaxation: L1 Prior

"Ideal" sparsity prior: $J_0(x) = \# \{m \setminus x_m \neq 0\}$

Image with 2 pixels:

- $J_0(x) = 0 \longrightarrow$ null image.
- $J_0(x) = 1 \longrightarrow$ sparse image.
- $J_0(x) = 2 \longrightarrow$ non-sparse image.

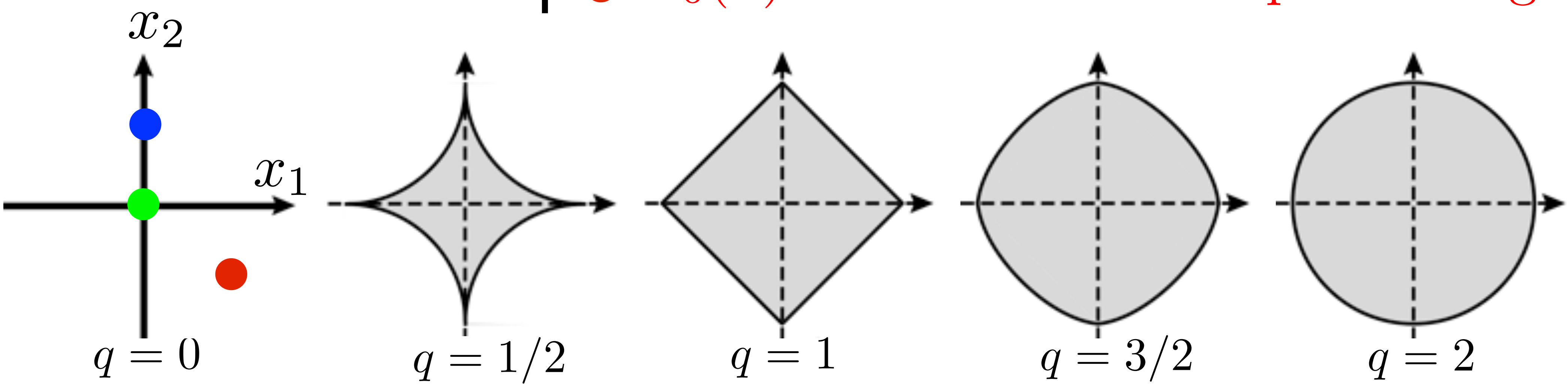


Convex Relaxation: L1 Prior

"Ideal" sparsity prior: $J_0(x) = \#\{m \setminus x_m \neq 0\}$

Image with 2 pixels:

- ● $J_0(x) = 0 \longrightarrow$ null image.
- ● $J_0(x) = 1 \longrightarrow$ sparse image.
- ● $J_0(x) = 2 \longrightarrow$ non-sparse image.



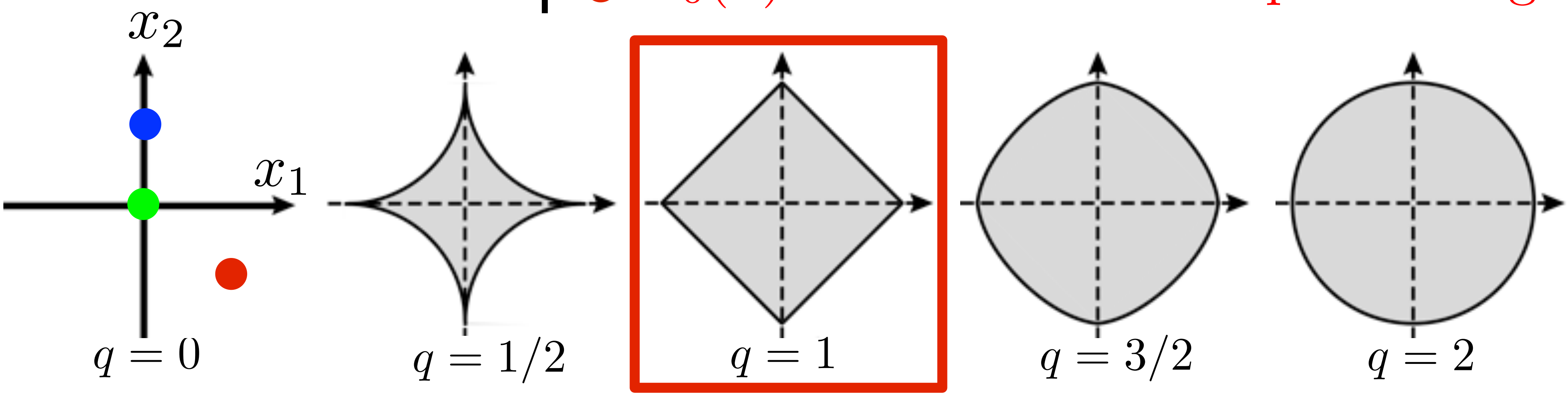
ℓ^q priors: $J_q(x) = \sum_m |x_m|^q$ (convex for $q \geqslant 1$)

Convex Relaxation: L1 Prior

"Ideal" sparsity prior: $J_0(x) = \# \{m \setminus x_m \neq 0\}$

Image with 2 pixels:

- ● $J_0(x) = 0 \longrightarrow$ null image.
- ● $J_0(x) = 1 \longrightarrow$ sparse image.
- ● $J_0(x) = 2 \longrightarrow$ non-sparse image.



ℓ^q priors: $J_q(x) = \sum_m |x_m|^q$ (convex for $q \geqslant 1$)

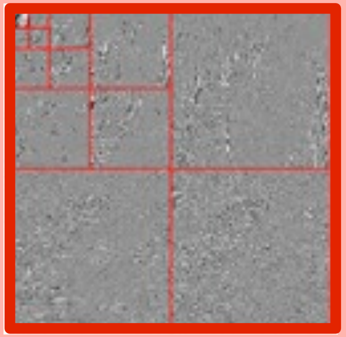
ℓ^1 norm: ℓ^q norm the "closest" to the ℓ^0 ideal sparsity.

Sparse ℓ^1 prior: $$\boxed{J_1(x) = \sum_m |x_m|}$$

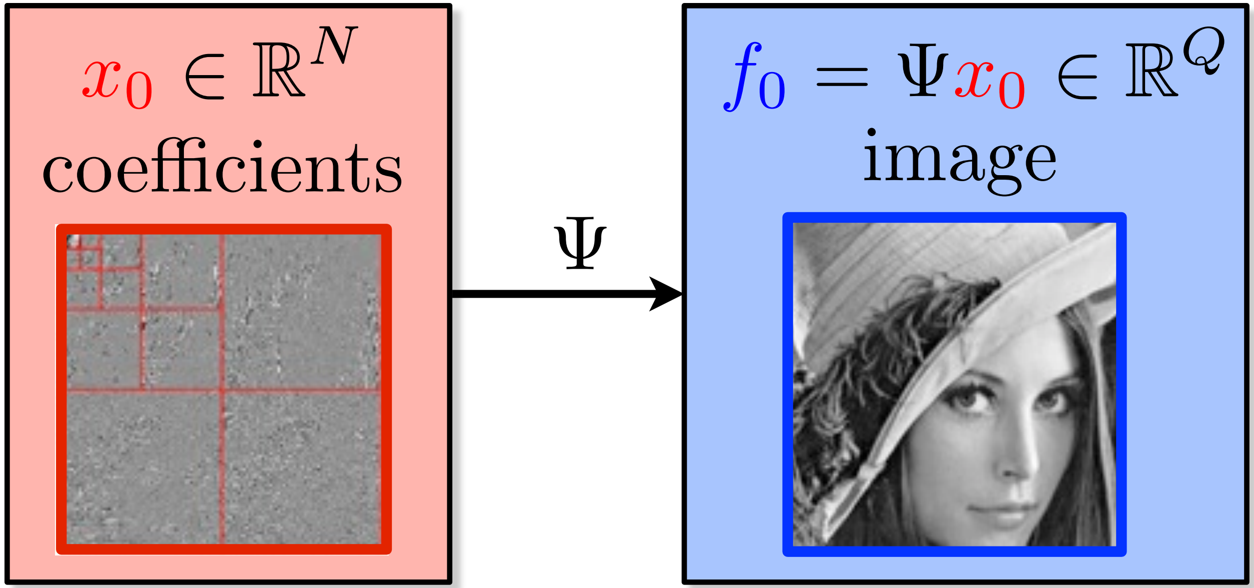
L1 Regularization

$x_0 \in \mathbb{R}^N$

coefficients



L1 Regularization



$x_0 \in \mathbb{R}^N$

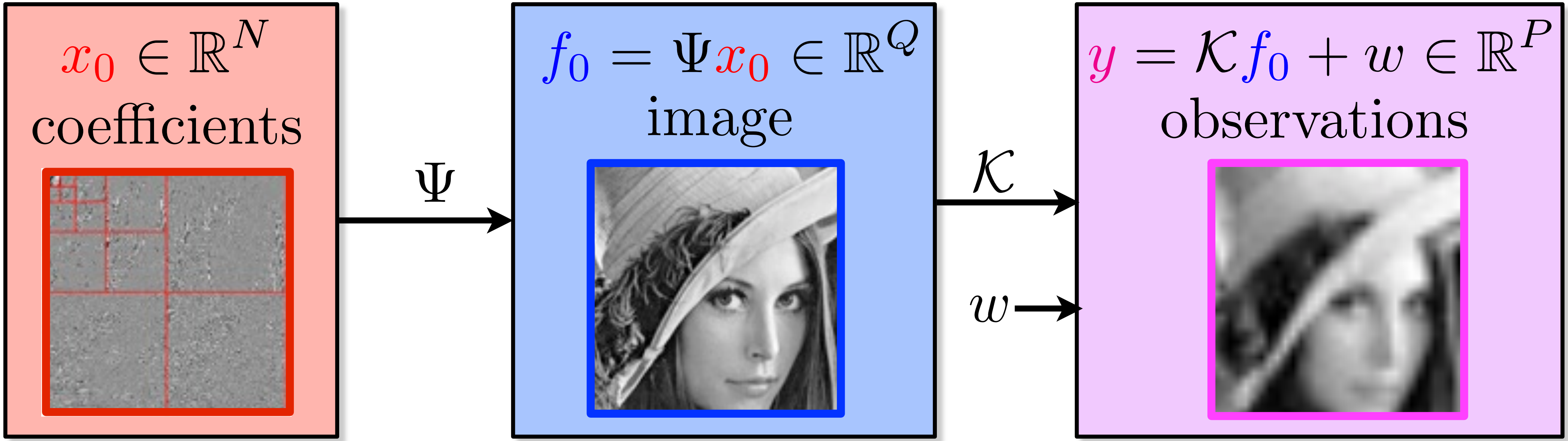
coefficients

Ψ

$f_0 = \Psi x_0 \in \mathbb{R}^Q$

image

L1 Regularization



$x_0 \in \mathbb{R}^N$

coefficients

$\xrightarrow{\Psi}$

$f_0 = \Psi x_0 \in \mathbb{R}^Q$

image

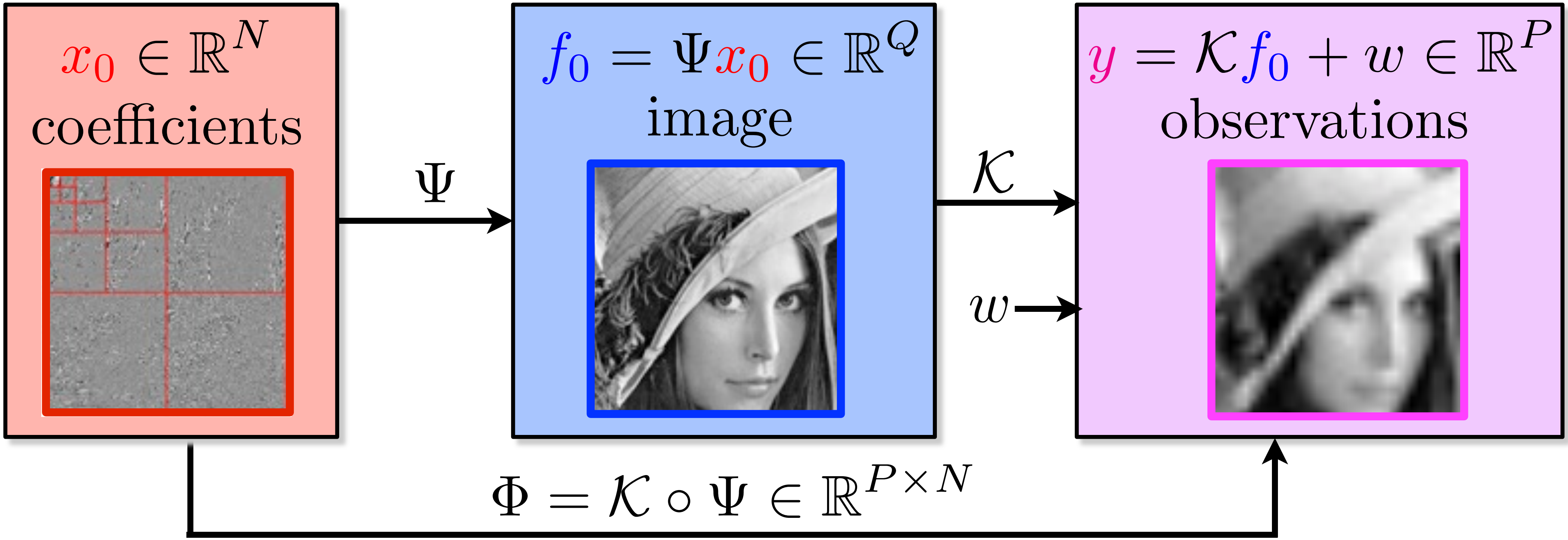
$\xrightarrow{\mathcal{K}}$

$w \rightarrow$

$y = \mathcal{K} f_0 + w \in \mathbb{R}^P$

observations

L1 Regularization



$x_0 \in \mathbb{R}^N$

coefficients

Ψ

$f_0 = \Psi x_0 \in \mathbb{R}^Q$

image

$\mathcal{K}$

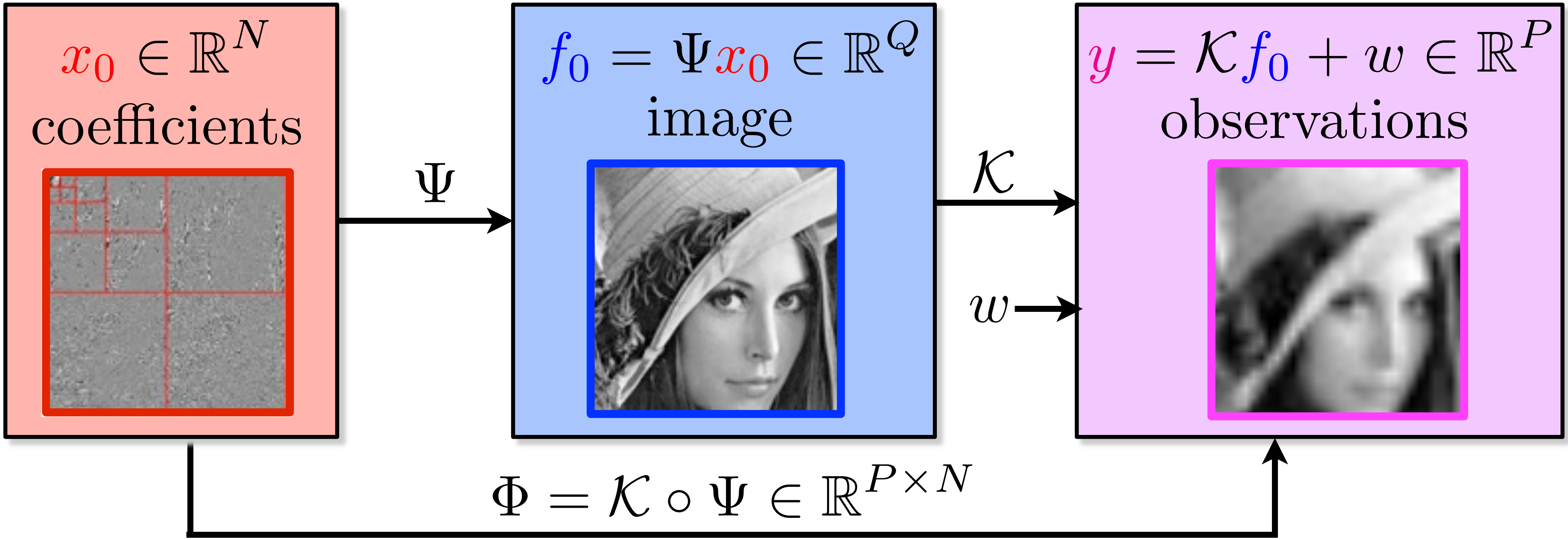
w

$y = \mathcal{K} f_0 + w \in \mathbb{R}^P$

observations

$\Phi = \mathcal{K} \circ \Psi \in \mathbb{R}^{P \times N}$

L1 Regularization



$x_0 \in \mathbb{R}^N$

coefficients

Ψ

$f_0 = \Psi x_0 \in \mathbb{R}^Q$

image

$\mathcal{K}$

w

$y = \mathcal{K} f_0 + w \in \mathbb{R}^P$

observations

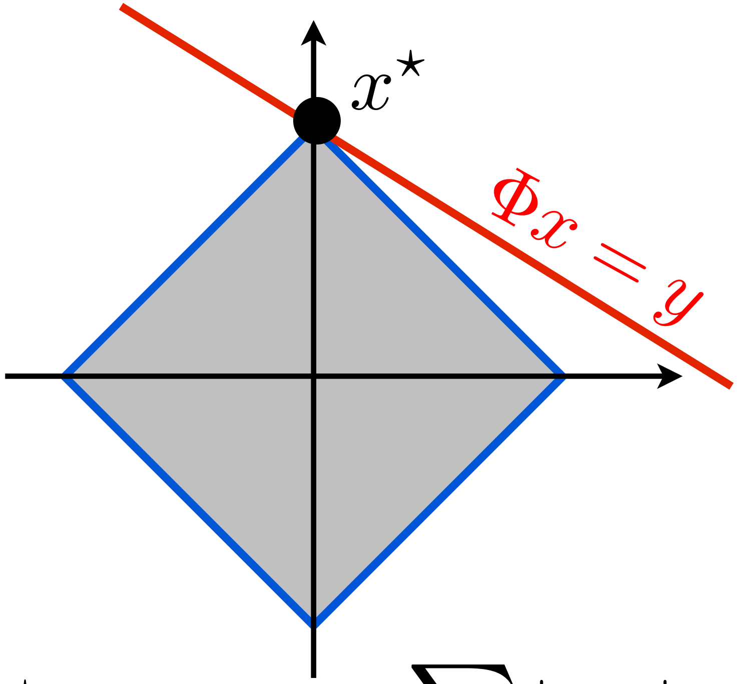
$\Phi = \mathcal{K} \circ \Psi \in \mathbb{R}^{P \times N}$

Sparse recovery: $f^\star = \Psi x^\star$ where $x^\star$ solves

$$\min_{x \in \mathbb{R}^N} \underbrace{\boxed{\frac{1}{2}\|y - \Phi x\|^2}}_{\text{Fidelity}} + \underbrace{\boxed{\lambda \|x\|_1}}_{\text{Regularization}}$$

Noiseless Sparse Regularization

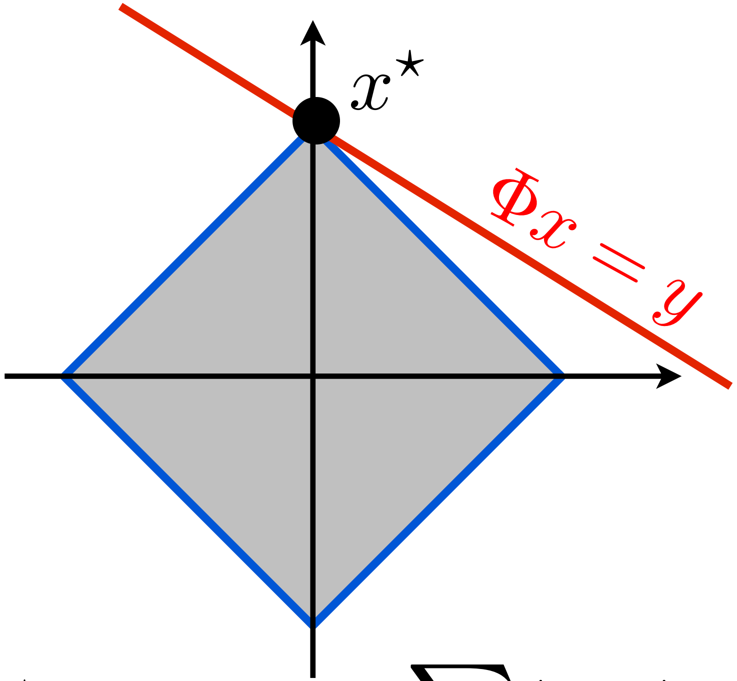
Noiseless measurements: $y = \Phi x_0$



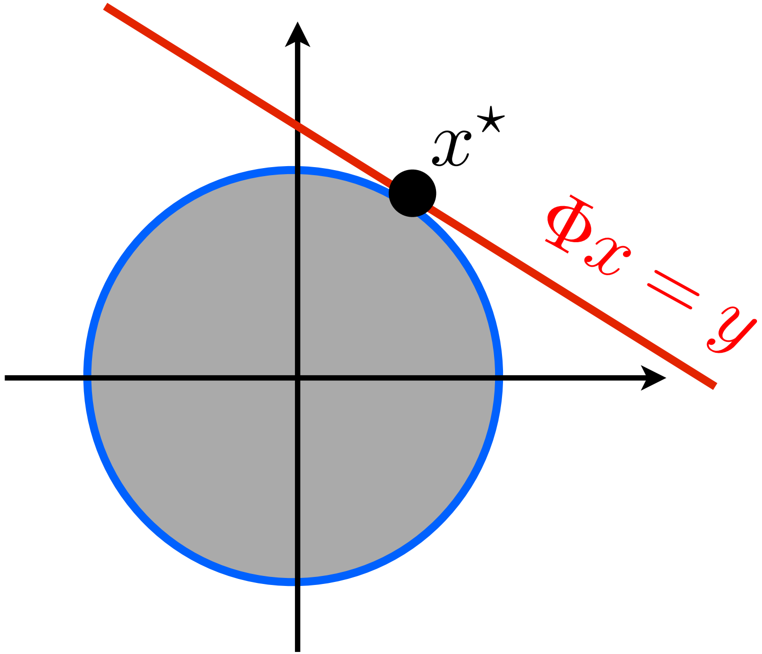
$$x^\star \in \underset{\Phi x = y}{\text{argmin}} \sum_m |x_m|$$

Noiseless Sparse Regularization

Noiseless measurements: $y = \Phi x_0$



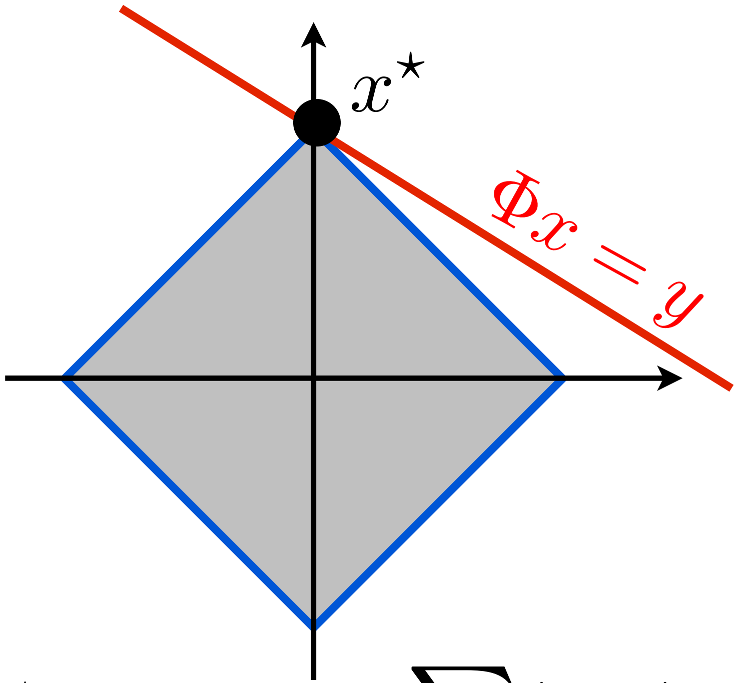
$$x^\star \in \underset{\Phi x = y}{\operatorname{argmin}} \sum_m |x_m|$$

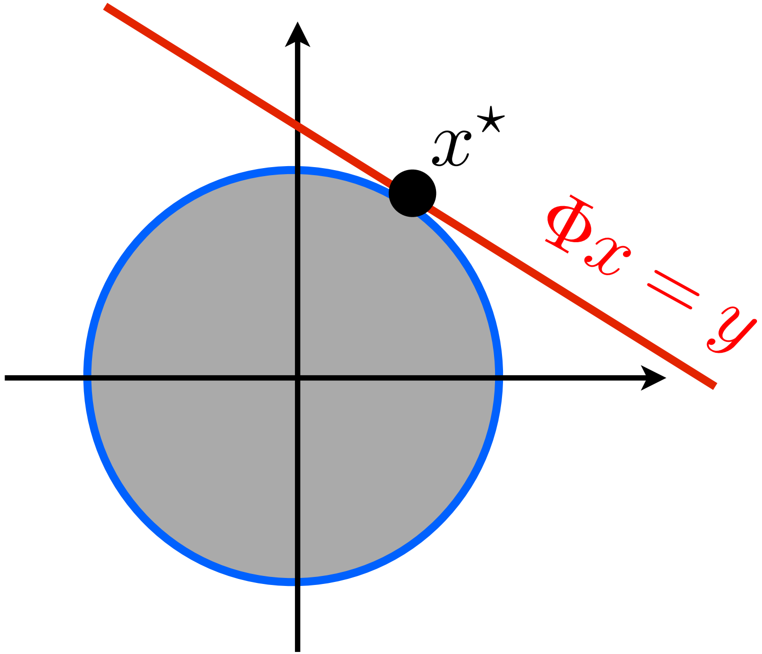


$$x^\star \in \underset{\Phi x = y}{\operatorname{argmin}} \sum_m |x_m|^2$$

Noiseless Sparse Regularization

Noiseless measurements: $y = \Phi x_0$





$$x^\star \in \underset{\Phi x = y}{\text{argmin}} \sum_m |x_m| \qquad x^\star \in \underset{\Phi x = y}{\text{argmin}} \sum_m |x_m|^2$$

Convex linear program.

- → Interior points, cf. [Chen, Donoho, Saunders] "basis pursuit".
- → Douglas-Rachford splitting, see [Combettes, Pesquet].

Noisy Sparse Regularization

Noisy measurements: $y = \Phi x_0 + w$

$$x^\star \in \underset{x \in \mathbb{R}^Q}{\operatorname{argmin}} \underset{\text{Data fidelity}}{\boxed{\frac{1}{2}\|y - \Phi x\|^2}} + \lambda \underset{\text{Regularization}}{\boxed{\|x\|_1}}$$

Noisy Sparse Regularization

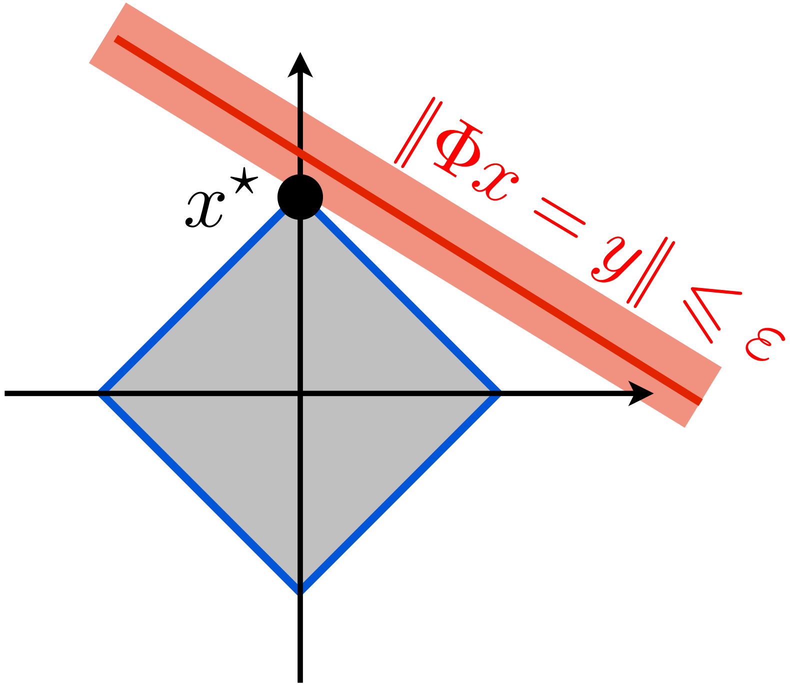
Noisy measurements: $y = \Phi x_0 + w$

$$x^\star \in \underset{x\in\mathbb{R}^Q}{\operatorname{argmin}} \boxed{\frac{1}{2}\|y - \Phi x\|^2} + \lambda \boxed{\|x\|_1}$$

Data fidelity Regularization

Equivalence
$\varepsilon \leftrightarrow \lambda$

$$x^\star \in \underset{\|\Phi x - y\| \leqslant \varepsilon}{\operatorname{argmin}} \|x\|_1$$



Noisy Sparse Regularization

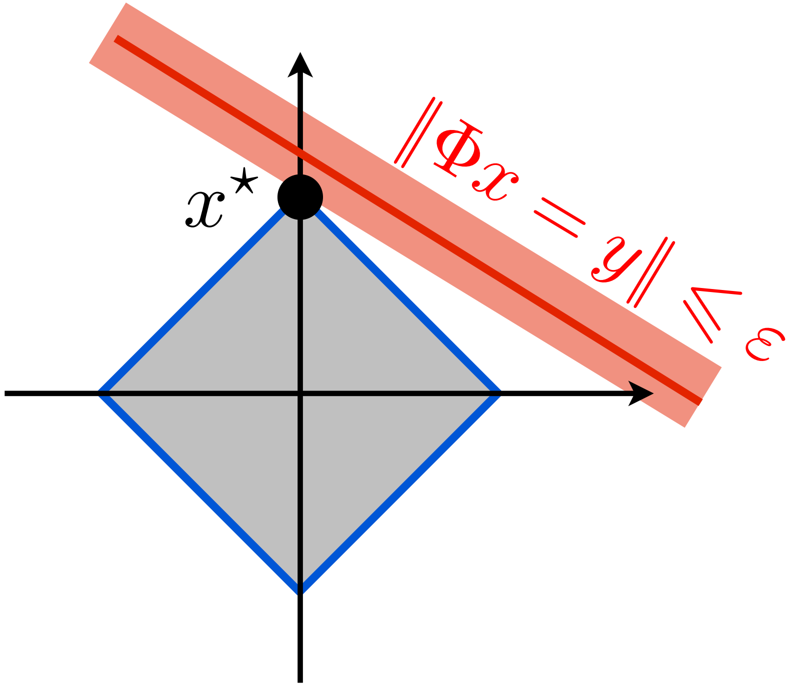
Noisy measurements: $y = \Phi x_0 + w$

$$x^\star \in \underset{x \in \mathbb{R}^Q}{\text{argmin}} \boxed{\frac{1}{2}\|y - \Phi x\|^2} + \lambda \boxed{\|x\|_1}$$

Data fidelity Regularization

Equivalence
$\varepsilon \leftrightarrow \lambda$

$$x^\star \in \underset{\|\Phi x - y\| \leqslant \varepsilon}{\text{argmin}} \|x\|_1$$



Algorithms:

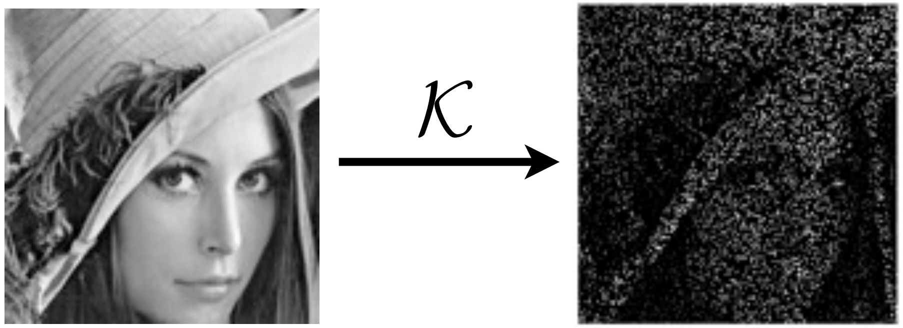
- → Iterative soft thresholding
  $\Longleftrightarrow$ Forward-backward splitting

see [Daubechies et al], [Pesquet et al], etc

- → Nesterov multi-steps schemes.

Inpainting Problem

Inpainting: set $\Omega \subset \{0, \ldots, N-1\}$ of missing pixels, $P = N - |\Omega|$.



$$(\mathcal{K}f)_i = \begin{cases} 0 & \text{if} \quad i \in \Omega, \\ f_i & \text{if} \quad i \notin \Omega. \end{cases}$$

Measures: $y = \mathcal{K} f_0 + w$


Sobolev, 20.8dB


Wavelets orth, 16.6dB


Wavelets TI, 23.6dB