Sparse & Redundant Representation

Iterated Shrinkage Algorithms

Wavelet thresholding

We will Talk About

- The minimization of the functional

$$f(\mathbf{a}) = \frac{1}{2}\|\mathbf{D}\mathbf{a} - \mathbf{x}\|_2^2 + \lambda\rho(\mathbf{a})$$

- Substitution: $\mathbf{u} = \mathbf{D}\mathbf{a}$

  and assume **D** is invertible $\mathbf{B} = \mathbf{D}^{-1}$: $\mathbf{B}\mathbf{D} = \mathbf{I}$

$$f(\mathbf{u}) = \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|_2^2 + \lambda\rho(\mathbf{B}\mathbf{u})$$

  This is our original denoising functional, if

$$\mathbf{B} = \begin{bmatrix}\partial_x \\ \partial_y\end{bmatrix}$$

$$\mathbf{B}^{+}\mathbf{B} = \mathbf{I}$$

$$\mathbf{D} = \mathbf{B}^{+} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$$

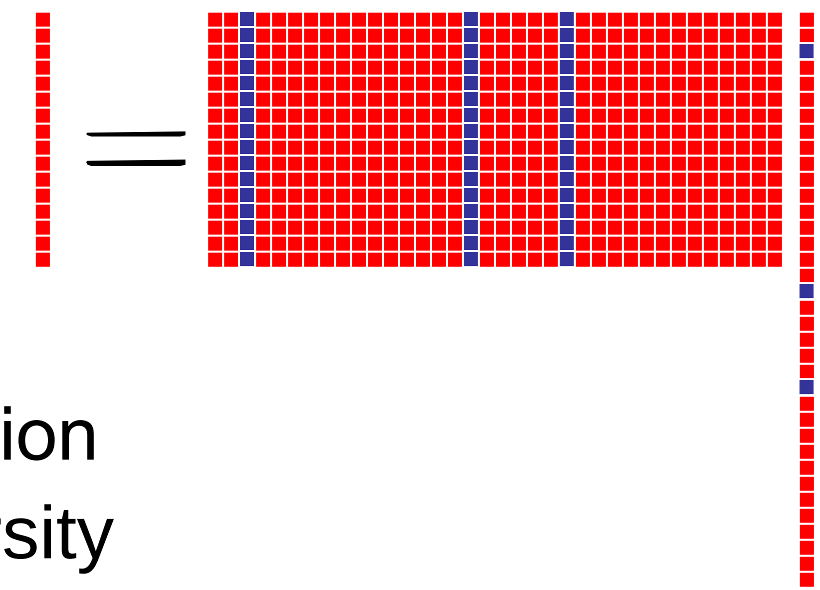
Moore-Penrose pseudoinverse

A Closer Look at the Functional

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \frac{1}{2}\|\mathbf{D}\mathbf{a} - \mathbf{x}\|_2^2 + \lambda\rho(\mathbf{a})$$

- $\mathbf{D}$… rectangular, overcomplete dictionary -> underdetermined system

$$\mathbf{D}\mathbf{a} = \mathbf{x} =$$

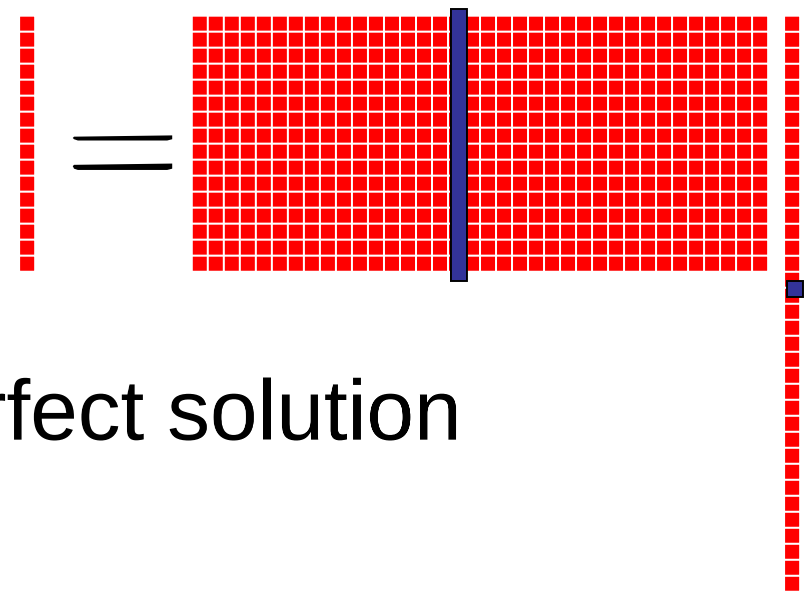


- $\mathbf{a}$… sparse representation
- $\rho(\mathbf{a})$… measure of sparsity

Why Seeking Sparse Representation?

- Assume $\mathbf{x}$ is 1-sparse signal in the dictionary $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_p]$

$$\mathbf{Da} = \mathbf{x} =$$



- $\max_i \langle \mathbf{d}_i, \mathbf{x} \rangle$ is the perfect solution

- Overcomplete dictionaries => sparsity

Measure of Sparsity

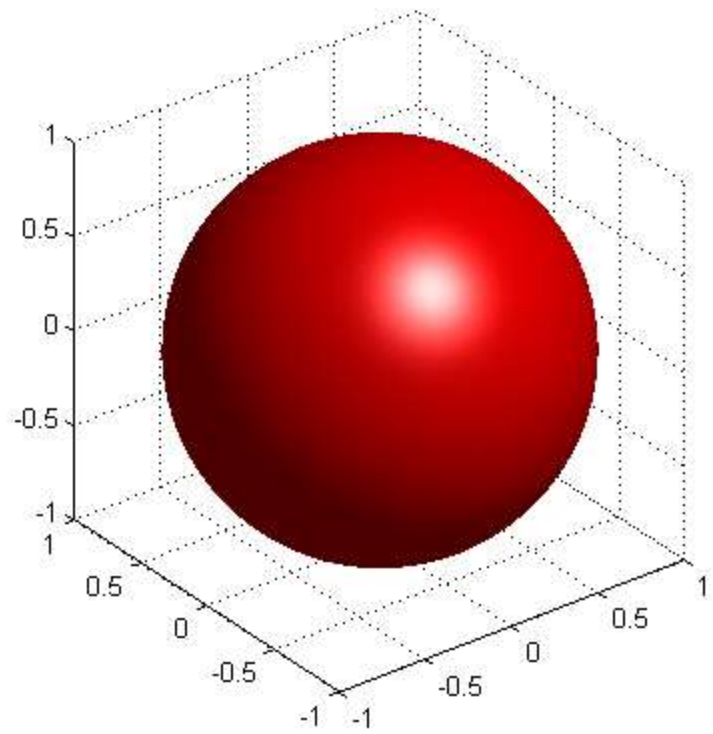
- $l_p\,,\quad 0 < p \leq 1$ norms ($\|\mathbf{a}\|_p^p$)

$$\|\mathbf{a}\|_p = \left(\sum_i |a_i|^p\right)^{\frac{1}{p}}$$

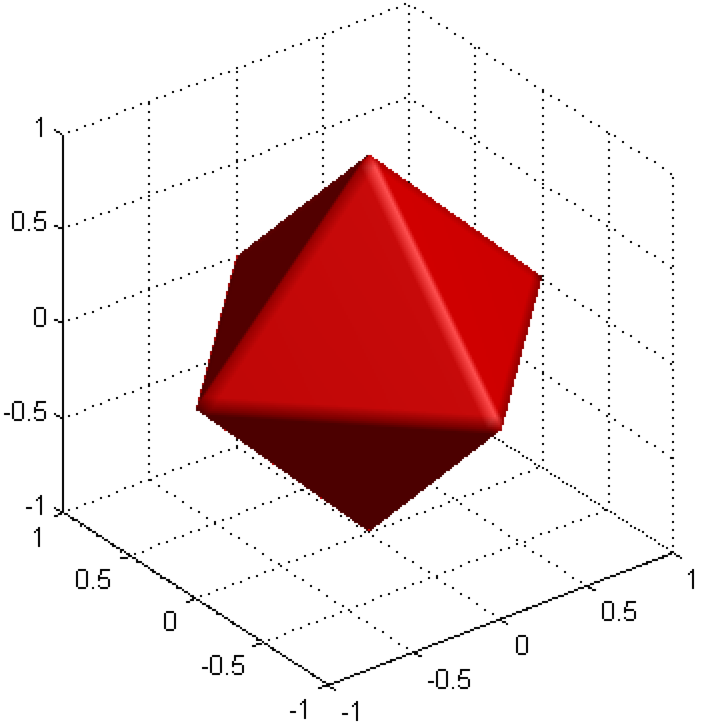
- l_0 norm, counts nonzero elements

- many other sparsity measures
  – smooth l_1 $\quad \rho(\mathbf{a}) = \|\mathbf{a}\|_1 - \epsilon \log\left(1 + \frac{\|\mathbf{a}\|_1}{\epsilon}\right)$

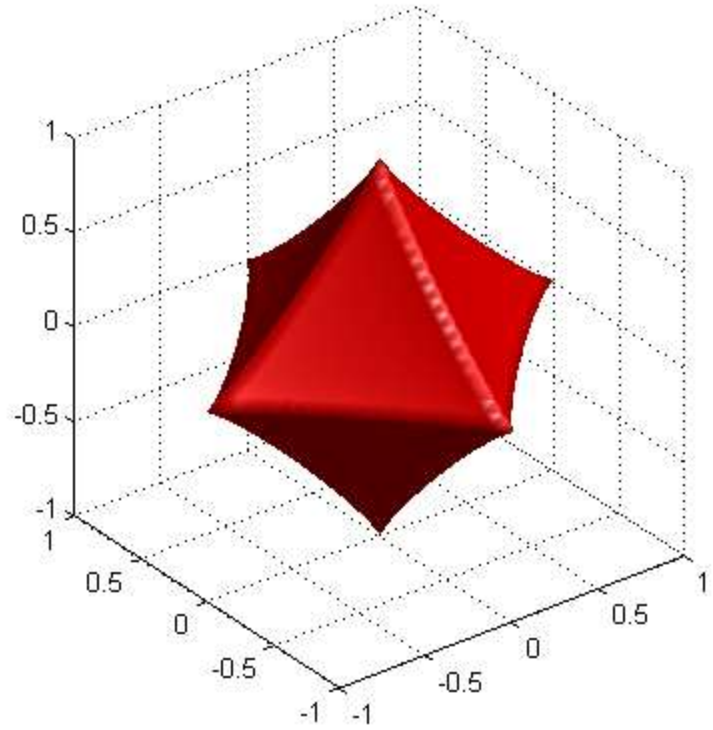
l_2 unit ball



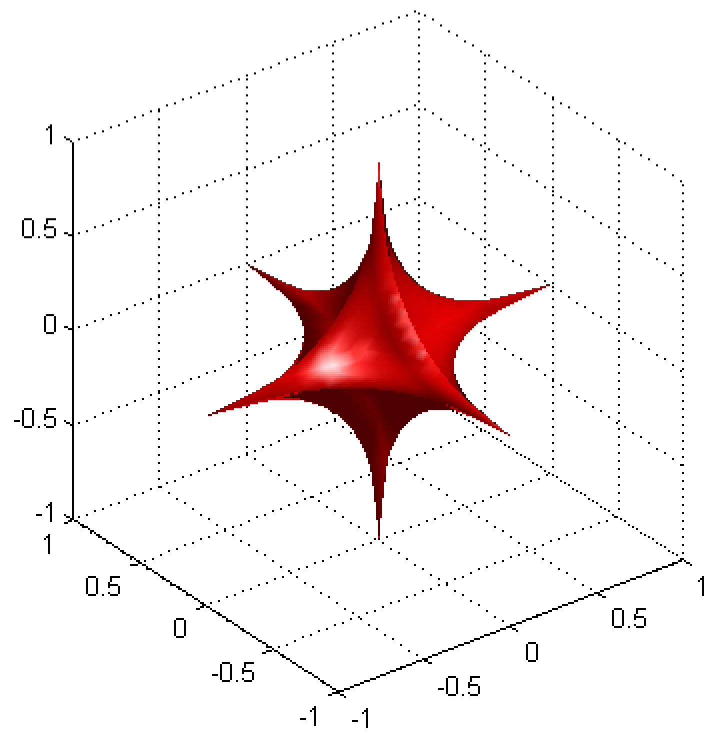
l_1 unit ball



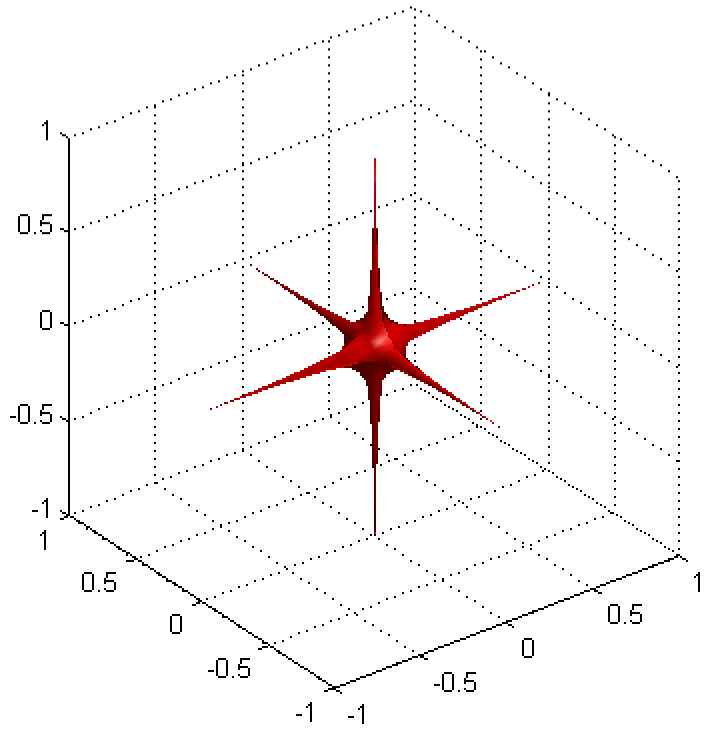
$l_{0.9}$ unit ball



$l_{0.5}$ unit ball



$l_{0.3}$ unit ball



Similar formulation

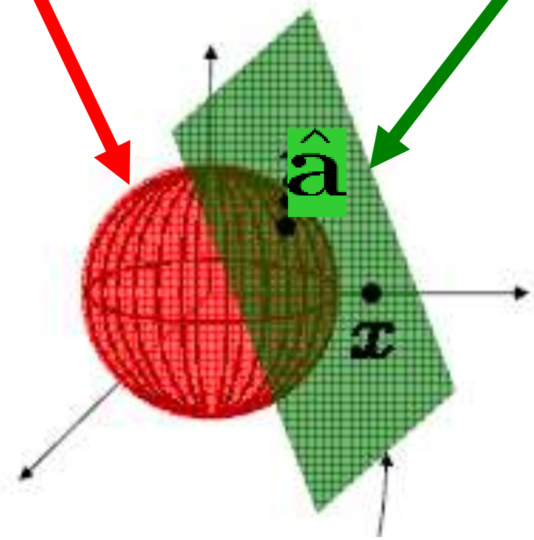
$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \frac{1}{2} \|\mathbf{D}\mathbf{a} - \mathbf{x}\|_2^2 + \lambda \rho(\mathbf{a})$$

- Method of Lagrange multipliers

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \rho(\mathbf{a}) \quad \text{subject to} \quad \mathbf{D}\mathbf{a} = \mathbf{x}$$

l_2-norm

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \|\mathbf{a}\|_2^2 \quad \text{subject to} \quad \mathbf{D}\mathbf{a} = \mathbf{x}$$

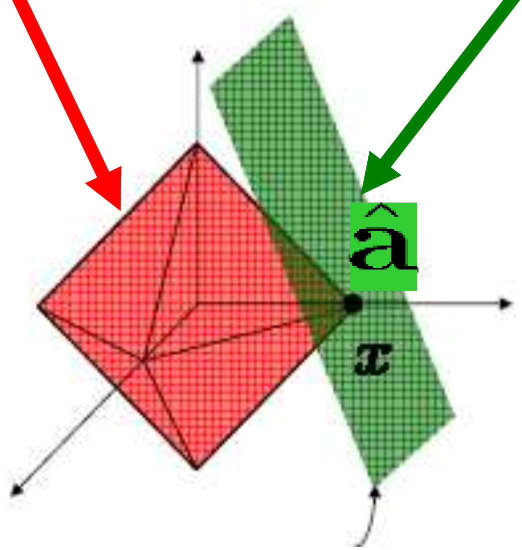


Solution via pseudoinverse is simple but not sparse:

$$\hat{\mathbf{a}} = \mathbf{D}^{+}\mathbf{x}$$

l_1-norm

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \|\mathbf{a}\|_1 \quad \text{subject to} \quad \mathbf{D}\mathbf{a} = \mathbf{x}$$



Classical Solvers

$$f(\mathbf{a}) = \frac{1}{2}\|\mathbf{Da} - \mathbf{x}\|_2^2 + \lambda\rho(\mathbf{a})$$

- Steepest Descent:

$$\hat{\mathbf{a}}_{i+1} = \hat{\mathbf{a}}_i - \mu\nabla f(\mathbf{a})|_{\hat{\mathbf{a}}_i} = \hat{\mathbf{a}}_i - \mu[\mathbf{D}^T(\mathbf{D}\hat{\mathbf{a}}_i - \mathbf{x}) + \lambda\rho'(\hat{\mathbf{a}}_i)]$$

- Hessian:

$$\nabla^2 f(\mathbf{a}) = \mathbf{D}^T\mathbf{D} + \lambda\rho''(\mathbf{a})$$

- Convergence depends on the condition number κ of the Hessian: $(\kappa - 1)/(\kappa + 1)$
- Hessian $\nabla^2 f$ is ill-conditioned

  => poor convergence

The Unitary Case ($\mathbf{DD}^{\mathrm{T}} = \mathbf{I}$)

$$f(\mathbf{a}) = \frac{1}{2}\|\mathbf{Da} - \mathbf{x}\|_2^2 + \lambda\rho(\mathbf{a})$$

Use

$\mathbf{DD}^T = \mathbf{I}$

$$f(\mathbf{a}) = \frac{1}{2}\|\mathbf{Da} - \mathbf{DD}^T\mathbf{x}\|_2^2 + \lambda\rho(\mathbf{a})$$

Define

$\mathbf{b} = \mathbf{D}^T\mathbf{x}$

$$f(\mathbf{a}) = \frac{1}{2}\|\mathbf{D}(\mathbf{a} - \mathbf{b})\|_2^2 + \lambda\rho(\mathbf{a})$$

l_2 is unitarily invariant

$$f(\mathbf{a}) = \frac{1}{2}\|\mathbf{a} - \mathbf{b}\|_2^2 + \lambda\rho(\mathbf{a})$$

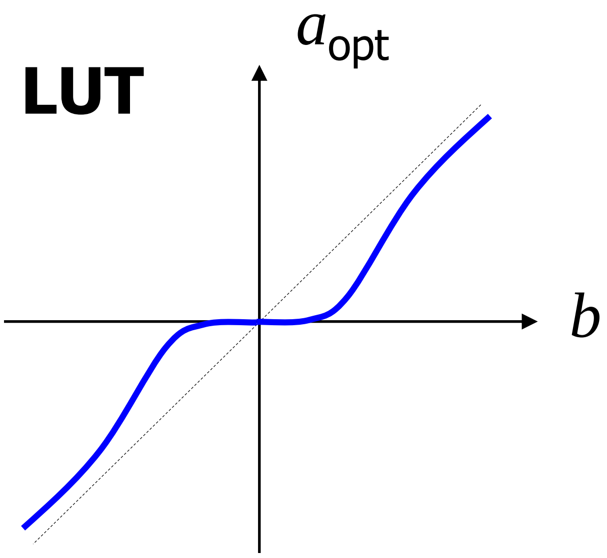
$$= \sum_{i=1}^{m}\left[\frac{1}{2}(a_i - b_i)^2 + \lambda\rho(a_i)\right]$$

We have m independent 1D optimization problems

The 1D Task

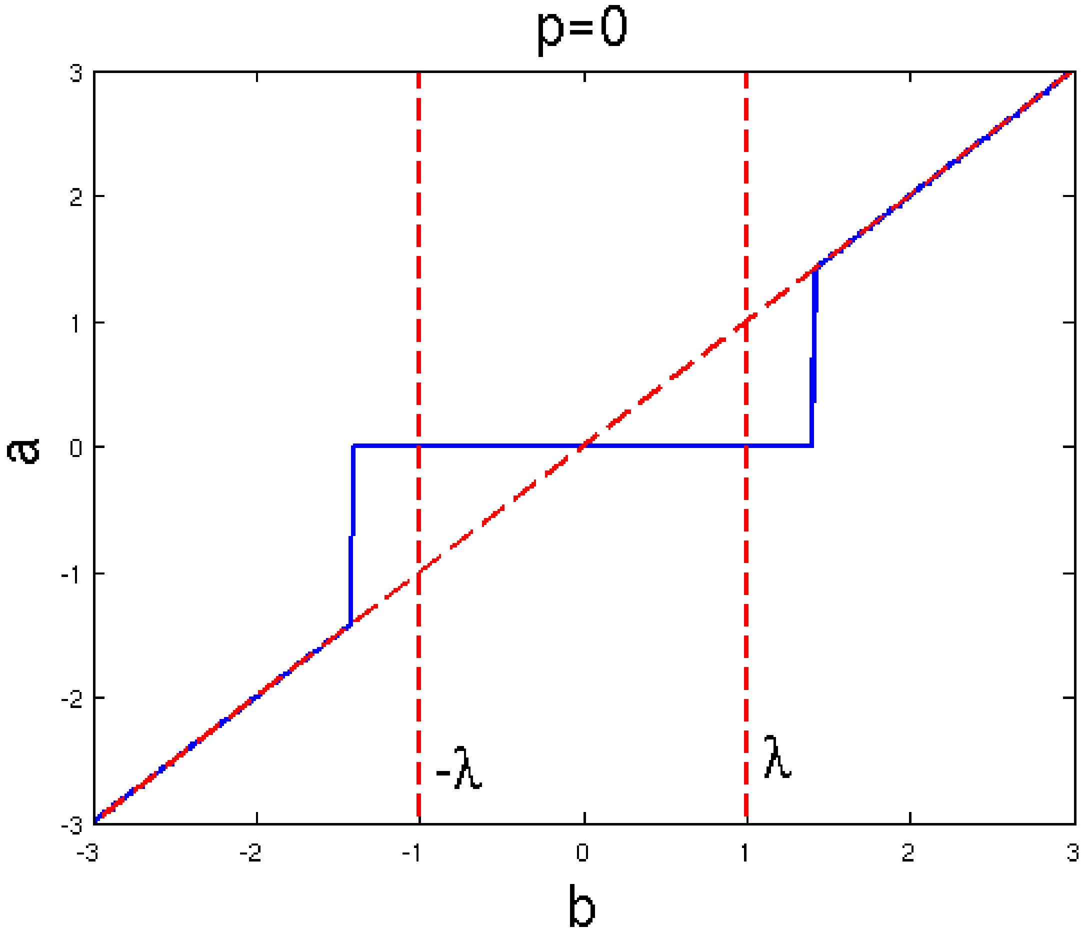
- We need to solve the following 1D problem:

$$a_{\text{opt}} = \arg\min_a \frac{1}{2}(a-b)^2 + \lambda\rho(a)$$

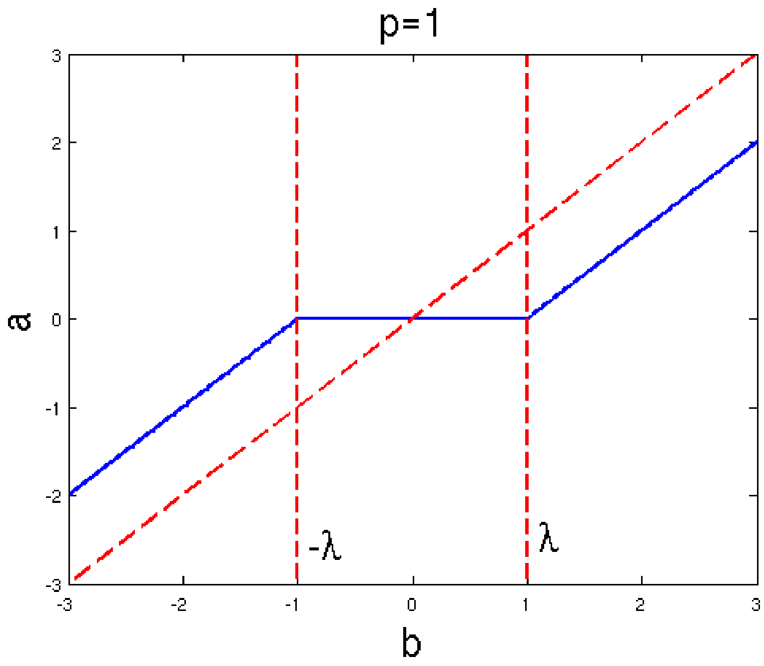


- Such a Look-Up-Table (LUT) $a_{\text{opt}} = S_{\rho,\lambda}(b)$ can be built for **ANY** sparsity measure function $\rho(a)$, including non-convex ones and non-smooth ones (e.g., l_0 norm).

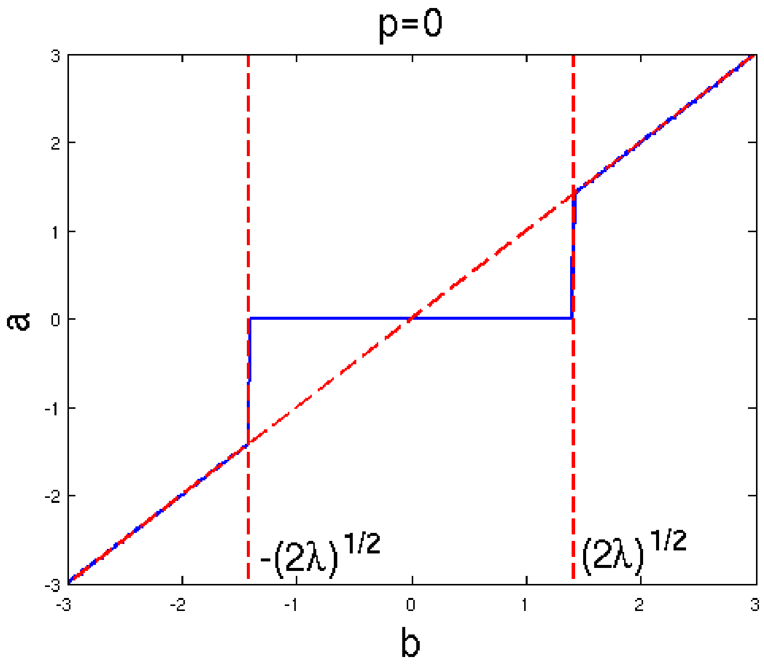
Shrinkage for $\rho(a)=|a|^p$



Soft & Hard Thresholding



Soft thresholding



Hard thresholding

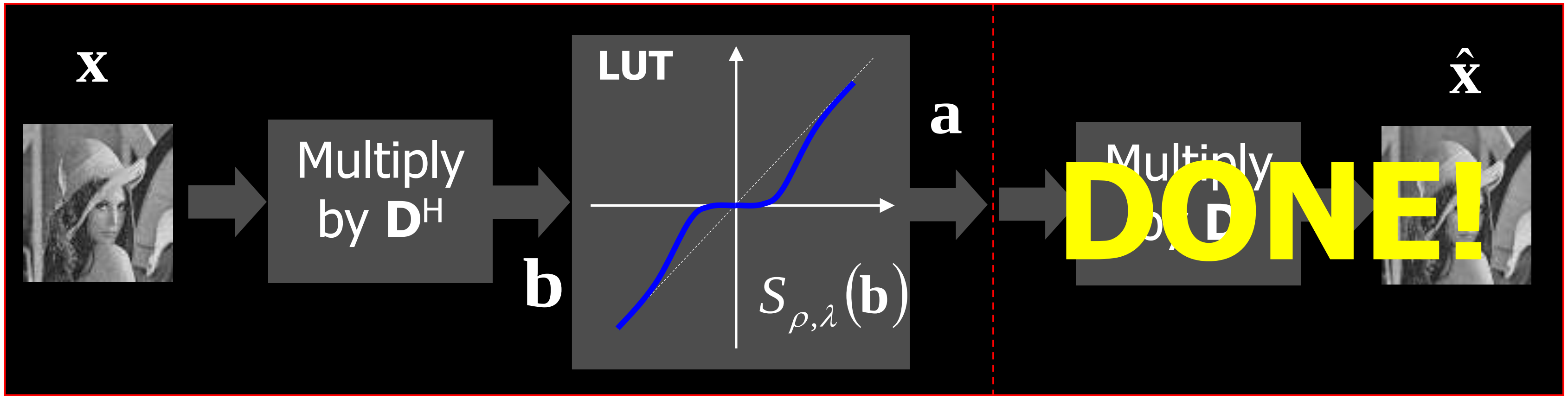
The Unitary Case: Summary

- Minimizing

$$f(\mathbf{a}) = \frac{1}{2}\|\mathbf{D}\mathbf{a} - \mathbf{x}\|_2^2 + \lambda\rho(\mathbf{a}) = \frac{1}{2}\|\mathbf{a} - \mathbf{b}\|_2^2 + \lambda\rho(\mathbf{a})$$

$$\mathbf{b} = \mathbf{D}^T\mathbf{x}$$

is done by:



Separable case: we achieve GLOBAL minimizer of $f(\mathbf{a})$, even if $f(\mathbf{a})$ is non-convex.

The minimization of

$$f(\mathbf{a}) = \frac{1}{2}\|\mathbf{D}\mathbf{a} - \mathbf{x}\|_2^2 + \lambda\rho(\mathbf{a})$$

Leads to two very **Contradicting Observations**:

1. The problem is **quite hard** – classic optimization find it hard.
2. The problem is **trivial** for the case of unitary **D**.

Solution: Proximal Algorithms

Proximal Operator

$$a^* = \arg\min_a \frac{1}{2\lambda} \|a - b\|_2^2 + \rho(a) = \mathbf{prox}_{\lambda\rho}(b)$$

ex.: indicator function

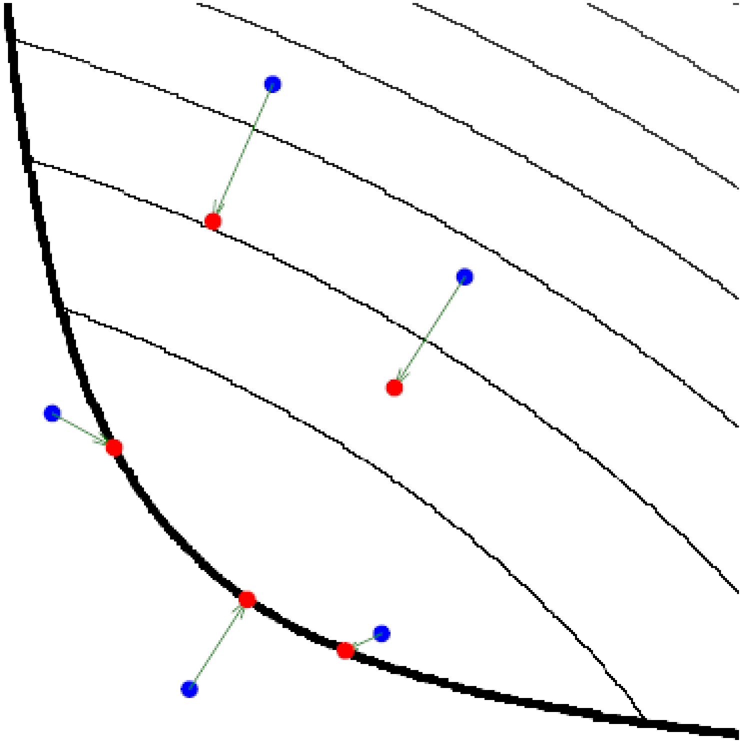
- If $\rho(a)$ closed proper convex

=>

$\mathbf{prox}_\rho(b)$ strictly convex

=>

unique minimizer



Examples of **prox** op.

- L1 norm ->

soft thresholding

$$\rho = \|\quad\|_1 \quad \rightarrow \quad \mathbf{prox}(b) := S_\lambda(b)$$

- Indicator function of a convex set C ->

projection onto C

$$\rho = I_C \quad \rightarrow \quad \mathbf{prox}(b) := \Pi_C(b)$$