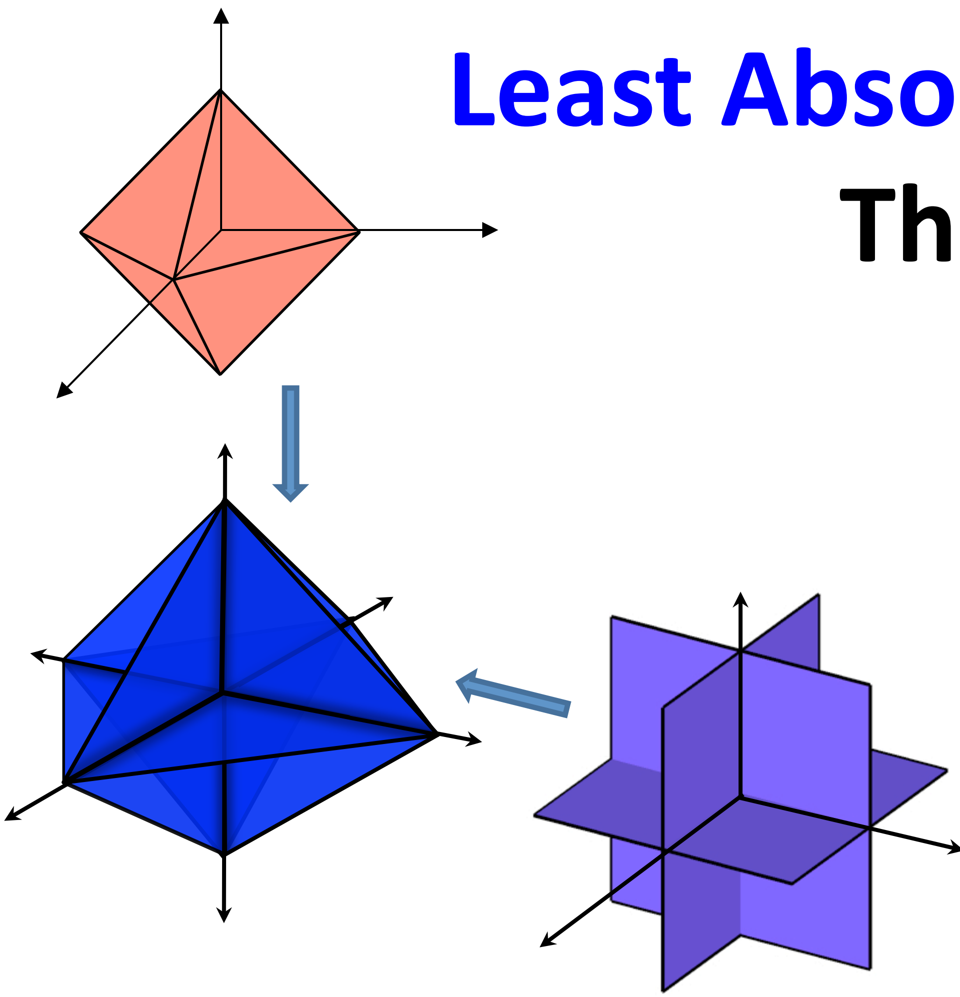
Combinatorial Selection and Least Absolute Shrinkage via The *CLASH* Operator



Volkan Cevher

*Laboratory for Information and Inference Systems – **LIONS / EPFL***

http://lions.epfl.ch

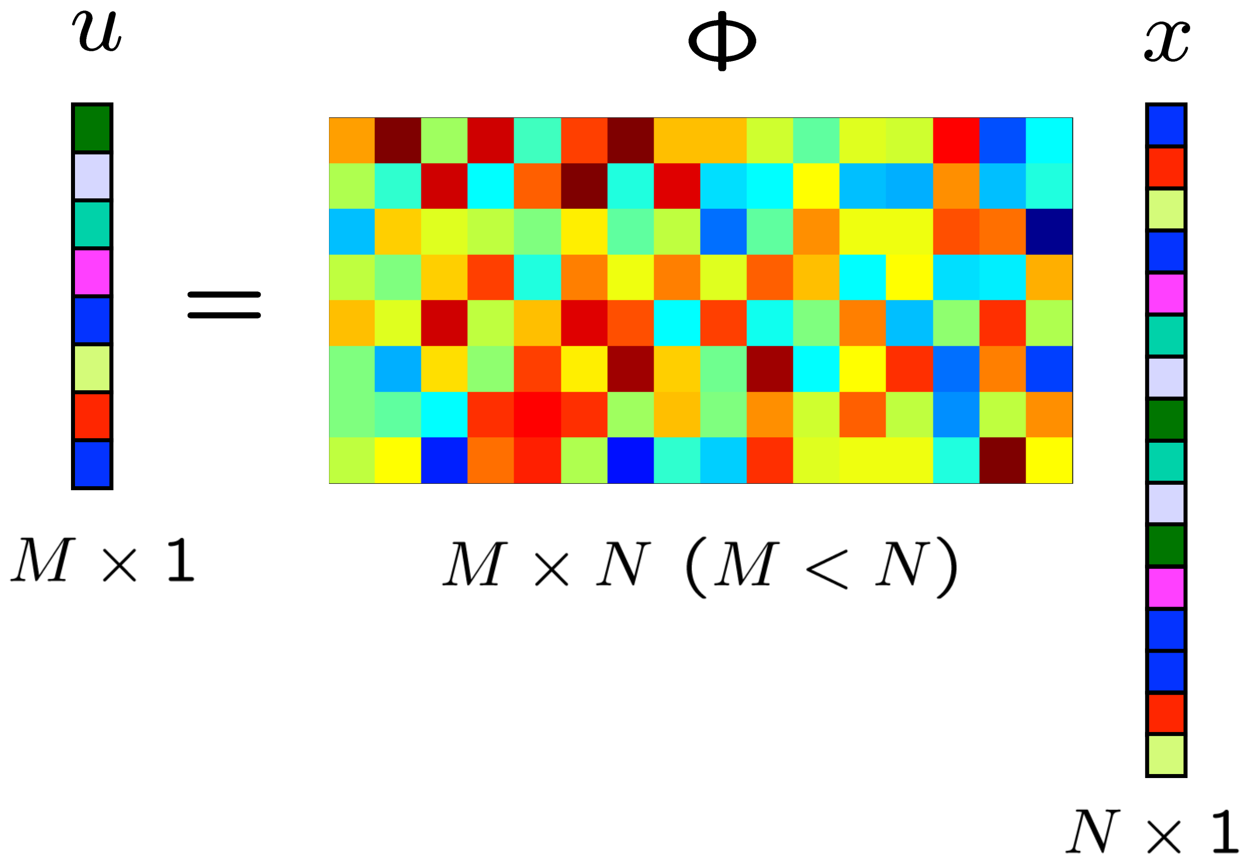
& Idiap Research Institute

joint work with my PhD student

Anastasios Kyrillidis @ EPFL

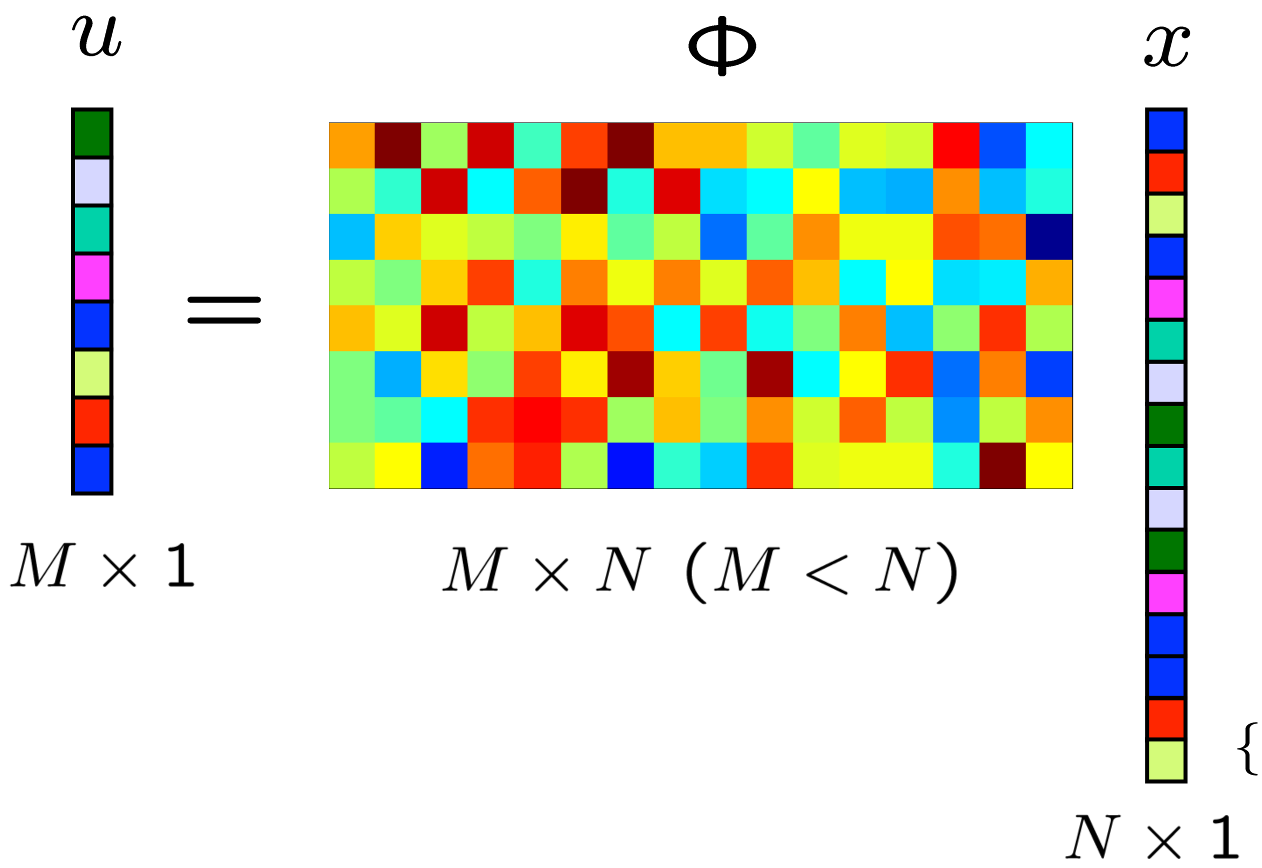
Duke—SAHD Meeting

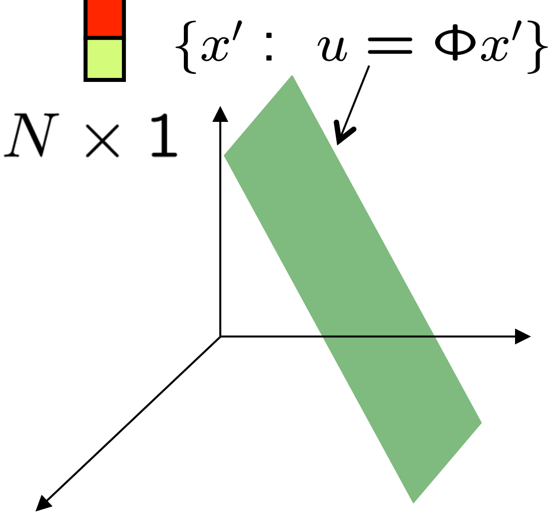
Linear Inverse Problems



compressive sensing	non-adaptive measurements
machine learning	dictionary of features
communications	MIMO user detection
theoretical computer science	sketching matrix / expander

Linear Inverse Problems

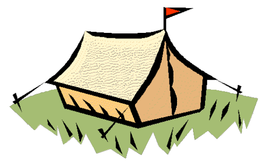




- **Challenge:** Null space of Φ: $\mathcal{N}(\Phi)$

$$x' = x + v \Rightarrow u, \;\; \forall v \in \mathcal{N}(\Phi)$$

Approaches



	Deterministic	**Probabilistic**
Prior	Sparsity	$f(x)$
Metric	ℓ_p-norm*	likelihood/ posterior

$* : \|x\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$

A Deterministic View

(with *a Model-based CS Flavor)*

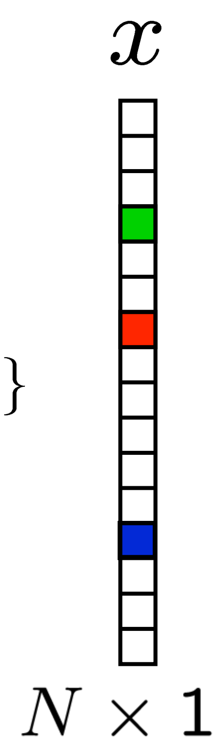
A Signal Prior

- **Sparse** signal: only K out of N coordinates nonzero
  - model: union of all K-dimensional subspaces aligned w/ coordinate axes

support:

$$\Omega = \{i : x_i \neq 0\}$$

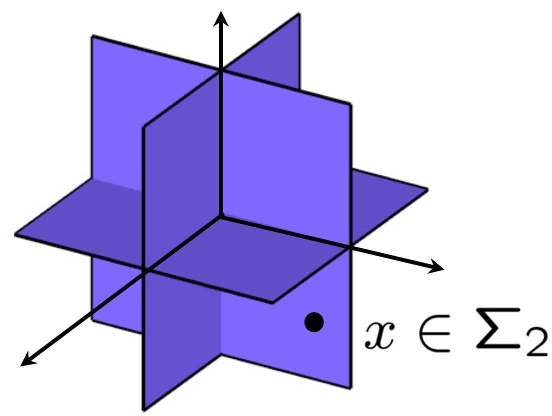
$$|\Omega| = K$$



Example: 2-sparse in 3-dimensions

$K = 2$

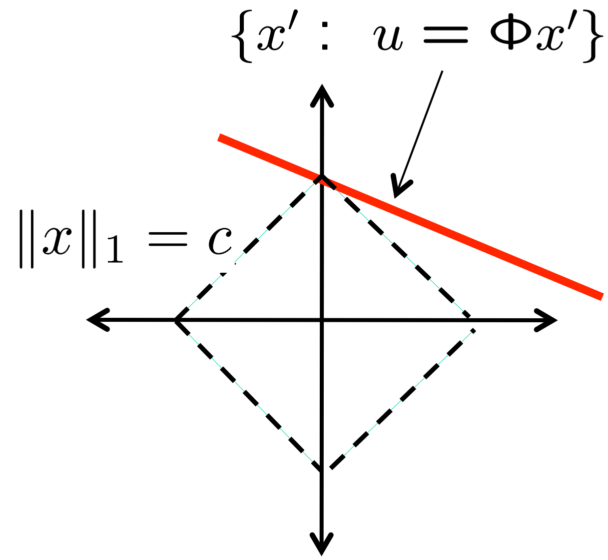
$\mathbf{R}^3$



Importance of Geometry

$\widehat{x} = \arg\min \|x\|_0 \text{ s.t. } u = \Phi x$

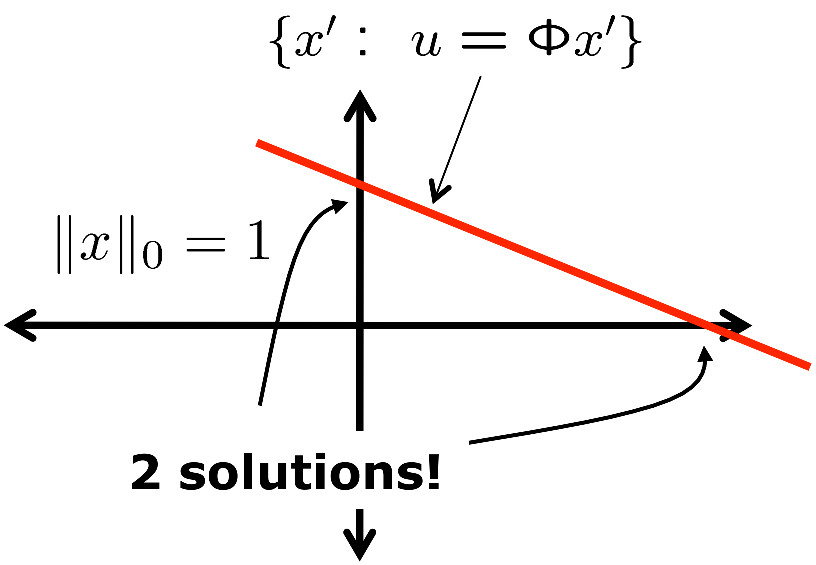
$\widehat{x} = \arg\min \|x\|_1 \text{ s.t. } u = \Phi x$



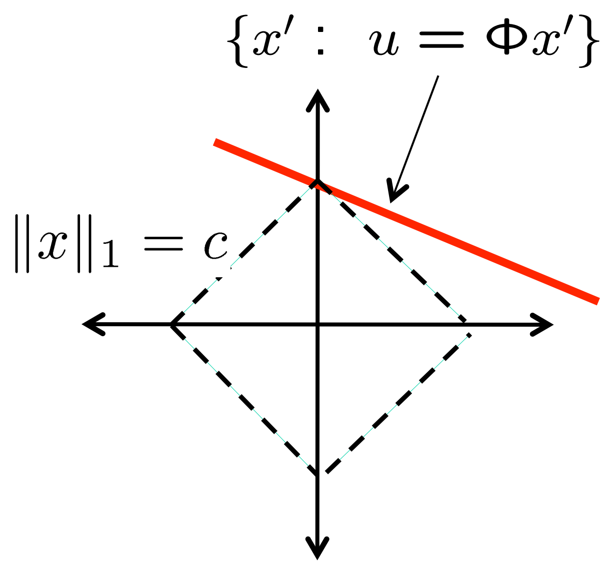
Importance of Geometry

- A subtle issue

$$\widehat{x} = \arg\min \|x\|_0 \text{ s.t. } u = \Phi x$$



$$\widehat{x} = \arg\min \|x\|_1 \text{ s.t. } u = \Phi x$$

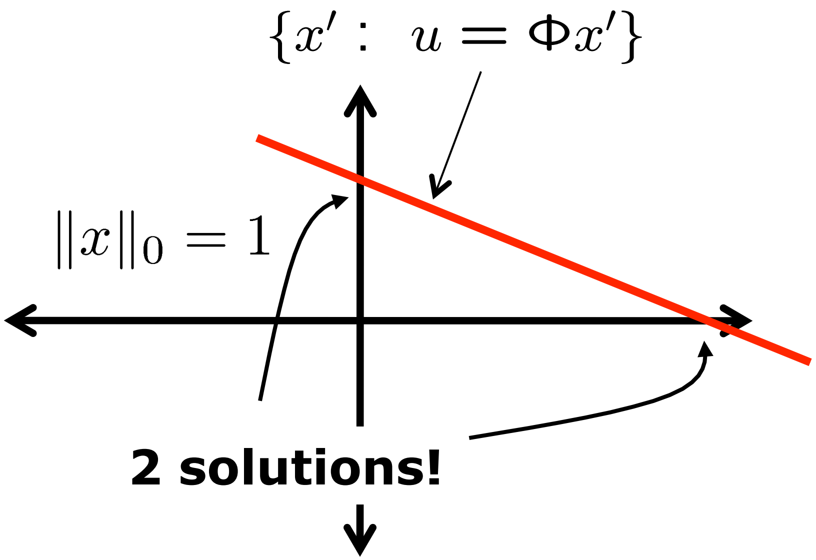


Which one is correct?

Importance of Geometry

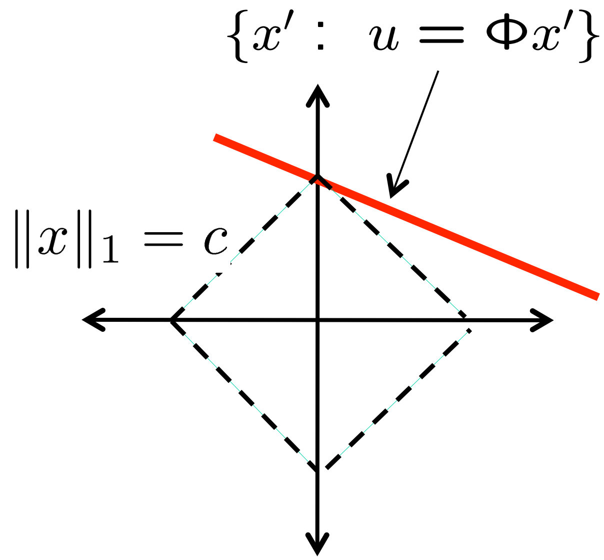
- A subtle issue

$$\widehat{x} = \arg\min \|x\|_0 \text{ s.t. } u = \Phi x$$

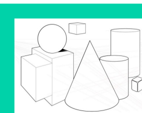


Which one is correct?

$$\widehat{x} = \arg\min \|x\|_1 \text{ s.t. } u = \Phi x$$



Sparse Recovery Algorithms

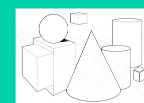
	Combinatorial $\binom{N}{K}$	**Geometric** 	**Probabilistic** 
Encoding	non-convex union-of-subspaces	atomic norm / convex relaxation	compressible / sparse priors
Example	$\min_{x:\|x\|_0 \leq K} \|u - \Phi x\|^2$	$\min_{x:\|x\|_1 \leq \lambda} \|u - \Phi x\|^2$	$E\{x\|u\}$
Algorithm	IHT, CoSaMP, SP, ALPS, OMP…	Basis pursuit, Lasso, basis pursuit denoising…	Variational Bayes, EP, Approximate message passing (AMP)…

$\|x\|_0 = \#\{x_i \neq 0\}$

http://lions.epfl.ch/ALPS

Sparse Recovery Algorithms

The Clash Operator

	Combinatorial $\binom{N}{K}$	Geometric 	Probabilistic 
Encoding	non-convex union-of-subspaces	atomic norm / convex relaxation	compressible / sparse priors
Example	$\min_{x:\|x\|_0 \leq K} \|\|u - \Phi x\|\|^2$	$\min_{x:\|x\|_1 \leq \lambda} \|\|u - \Phi x\|\|^2$	$E\{x\|u\}$
Algorithm	IHT, CoSaMP, SP, ALPS, OMP…	Basis pursuit, Lasso, basis pursuit denoising…	Variational Bayes, EP, Approximate message passing (AMP)…

$$\widehat{x}_{\text{Clash}} = \arg\min_{x:\|x\|_0 \leq K, \|x\|_1 \leq \lambda} \|u - \Phi x\|^2$$

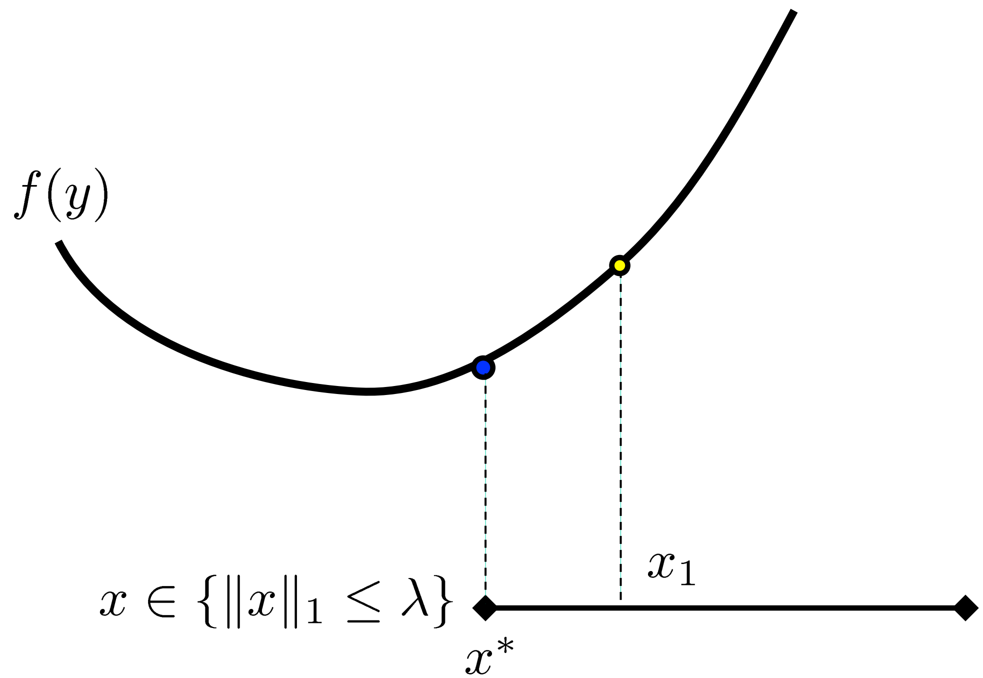
$$\|x\|_0 = \#\{x_i \neq 0\}$$

A Tale of Two Algorithms

- Soft thresholding

$$f(x) = \|u - \Phi x\|^2$$

$$\min_{x:\|x\|_1 \le \lambda} f(x)$$

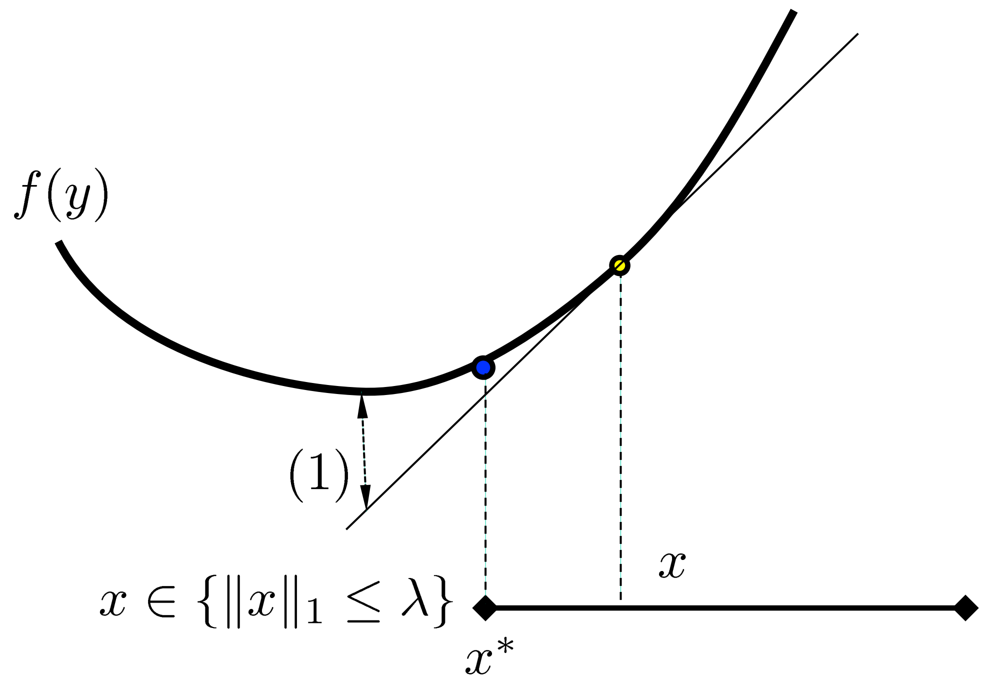


A Tale of Two Algorithms

- Soft thresholding

$$f(x) = \|u - \Phi x\|^2$$

$\min_{x:\|x\|_1 \leq \lambda} f(x)$



Structure in optimization:

Bregman distance

$$(1) \qquad \underbrace{f(y) - f(x) - \langle \nabla f(x), y - x \rangle}_{\text{Bregman distance}} = \|\Phi(y - x)\|^2 \qquad \forall x, y \in \mathcal{R}^N,$$

A Tale of Two Algorithms

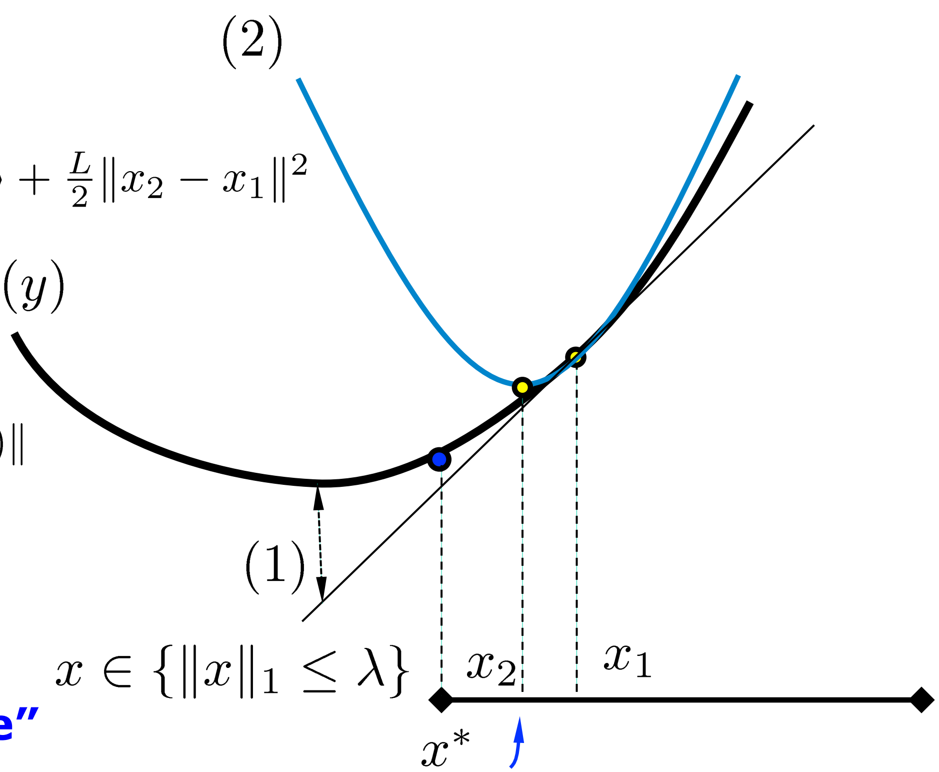
- Soft thresholding

$$f(x) = \|u - \Phi x\|^2$$

$\min_{x:\|x\|_1 \leq \lambda} f(x)$

$$U(x_2, x_1) = f(x_1) + \langle \nabla f(x_1), x_2 - x_1 \rangle + \tfrac{L}{2}\|x_2 - x_1\|^2$$

majorization-minimization



$$\arg\min_{\|x\|_1 \leq \lambda} U(x, y) = \arg\min_{\|x\|_1 \leq \lambda} \|x - (y - \frac{1}{L}\nabla f(y))\|$$

$$\mathrm{St}_{\{\|x\|_1 \leq \lambda\}}(t) = \arg\min_{\|x\|_1 \leq \lambda} \|x - t\|$$

$$x_{i+1} = \mathrm{St}_{\{\|x\|_1 \leq \lambda\}}\left(x_i - \tfrac{1}{L}\nabla f(x_i)\right)$$

Key actor: "least absolute shrinkage"

Bregman distance

$$\begin{array}{llcll}
(1) & f(y) - f(x) - \langle \nabla f(x), y - x \rangle & = & \|\Phi(y - x)\|^2 & \forall x, y \in \mathcal{R}^N, \\
(2) & f(y) - f(x) - \langle \nabla f(x), y - x \rangle & \leq & \frac{L}{2}\|y - x\|^2 & L = 2\|\Phi\|, \forall x, y \in \mathcal{R}^N,
\end{array}$$

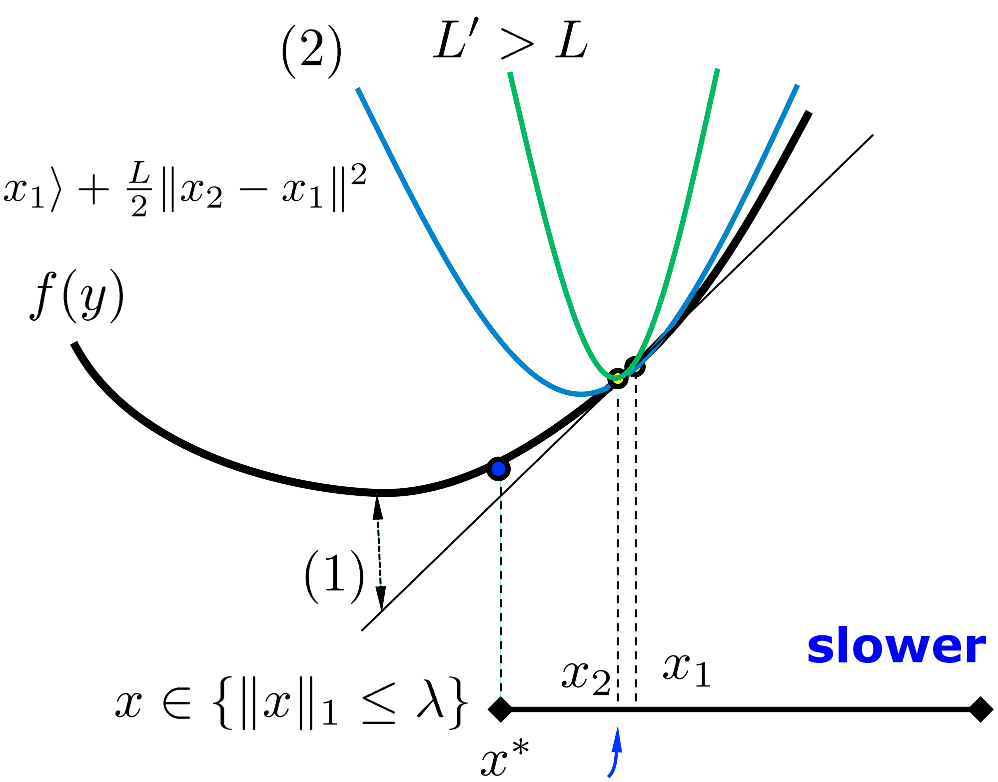
A Tale of Two Algorithms

- Soft thresholding

$$f(x) = \|u - \Phi x\|^2$$

$$\min_{x:\|x\|_1 \leq \lambda} f(x)$$

$$U(x_2, x_1) = f(x_1) + \langle \nabla f(x_1), x_2 - x_1 \rangle + \frac{L}{2}\|x_2 - x_1\|^2$$



$$x_{i+1} = \mathrm{St}_{\{\|x\|_1 \leq \lambda\}}\left(x_i - \frac{1}{L'}\nabla f(x_i)\right)$$

Bregman distance

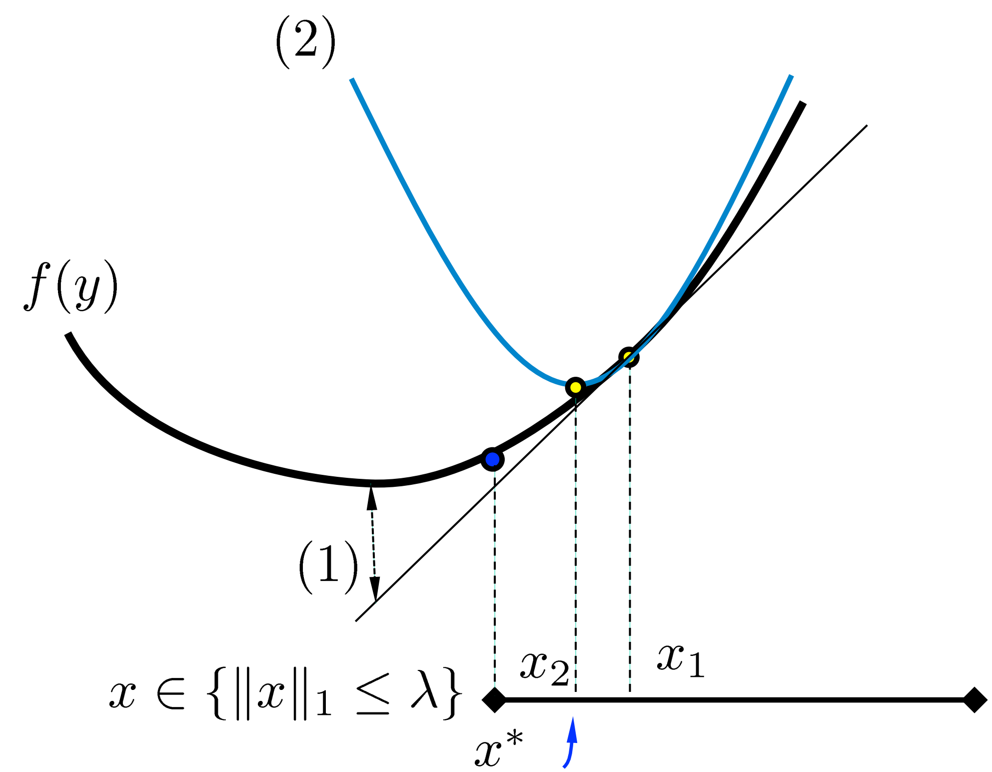
$$
\begin{array}{llll}
(1) & f(y) - f(x) - \langle \nabla f(x), y - x \rangle = \|\Phi(y - x)\|^2 & & \forall x, y \in \mathcal{R}^N, \\
(2) & f(y) - f(x) - \langle \nabla f(x), y - x \rangle \leq \frac{L}{2}\|y - x\|^2 & & L = 2\|\Phi\|, \forall x, y \in \mathcal{R}^N,
\end{array}
$$

A Tale of Two Algorithms

- Soft thresholding

$$f(x) = \|u - \Phi x\|^2$$

$\min_{x:\|x\|_1 \le \lambda} f(x)$



$$x_{i+1} = \mathrm{St}_{\{\|x\|_1 \le \lambda\}}\left(x_i - \frac{1}{L}\nabla f(x_i)\right)$$

- Is x* what we are looking for?

  local “unverifiable” assumptions:

  – ERC/URC condition

  – compatibility condition …

(local → global / dual certification / random signal models)

$\binom{N}{K}$ A Tale of Two Algorithms

- Hard thresholding

$f(x) = \|u - \Phi x\|^2$

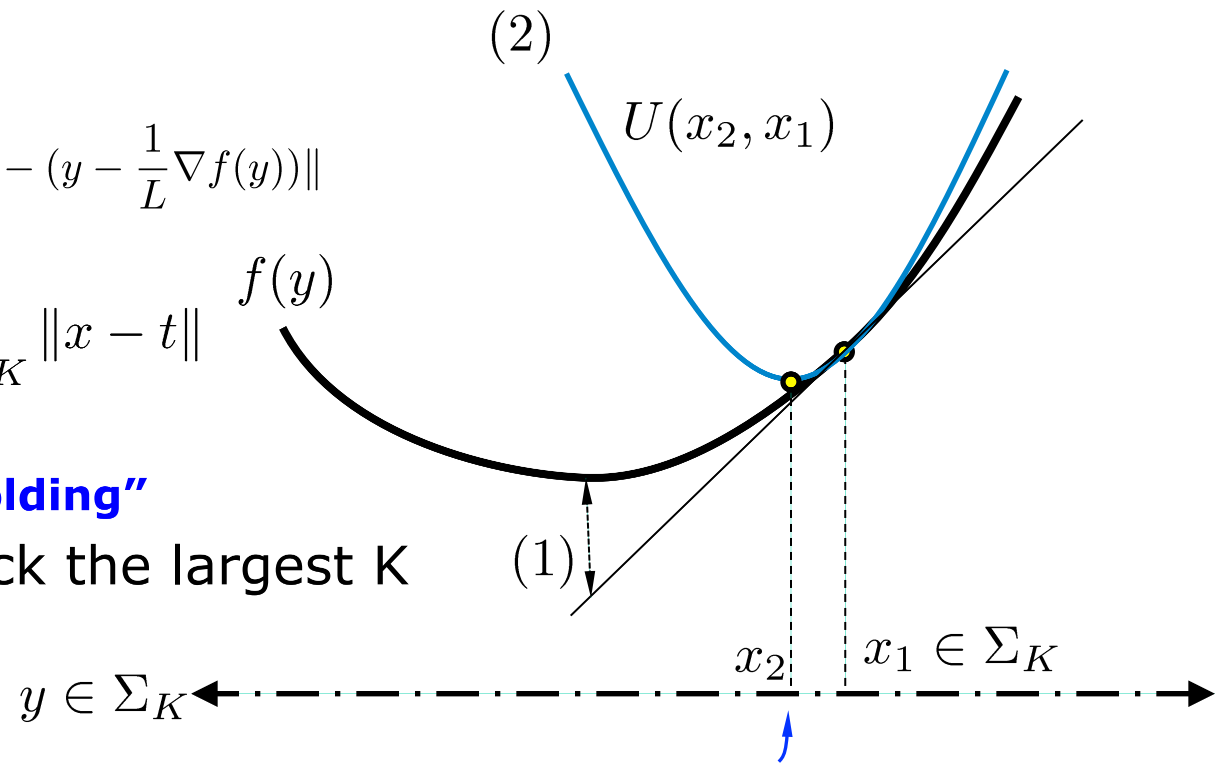
$\min_{x:\|x\|_0 \leq K} f(x)$

$$\arg\min_{\|x\|_0 \leq K} U(x,y) = \arg\min_{\|x\|_0 \leq K} \|x - (y - \frac{1}{L}\nabla f(y))\|$$

$$\mathrm{H}_{\{\|x\|_0 \leq K\}}(t) = \arg\min_{\|x\|_0 \leq K} \|x - t\|$$

Key actor: "hard thresholding"

ALGO: sort and pick the largest K



$$x_{i+1} = \mathrm{H}_{\{\|x\|_0 \leq K\}}\left(x_i - \tfrac{1}{L}\nabla f(x_i)\right)$$

$\binom{N}{K}$ A Tale of Two Algorithms

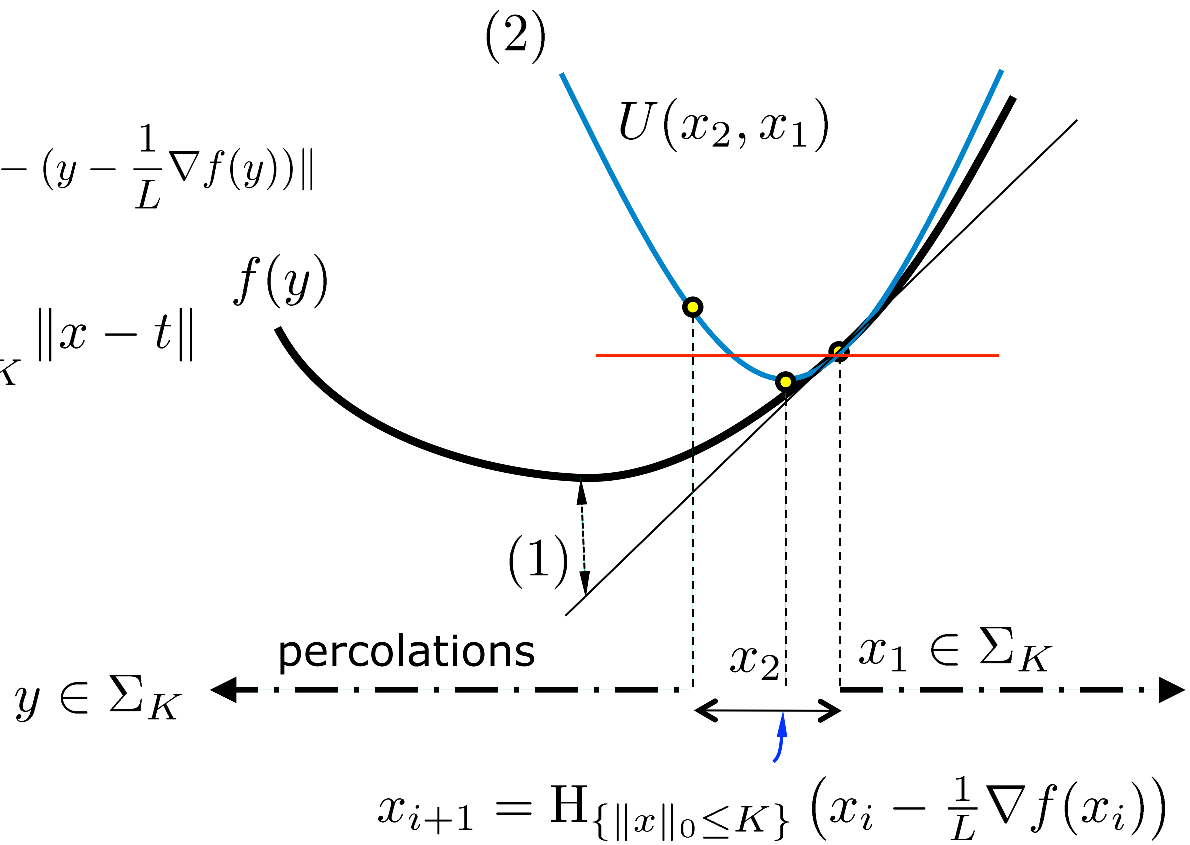
- Hard thresholding

$\min_{x:\|x\|_0\leq K} f(x)$

$$\arg\min_{\|x\|_0\leq K} U(x,y) = \arg\min_{\|x\|_0\leq K} \|x - (y - \frac{1}{L}\nabla f(y))\|$$

$$\mathrm{H}_{\{\|x\|_0\leq K\}}(t) = \arg\min_{\|x\|_0\leq K} \|x - t\|$$

$f(x) = \|u - \Phi x\|^2$



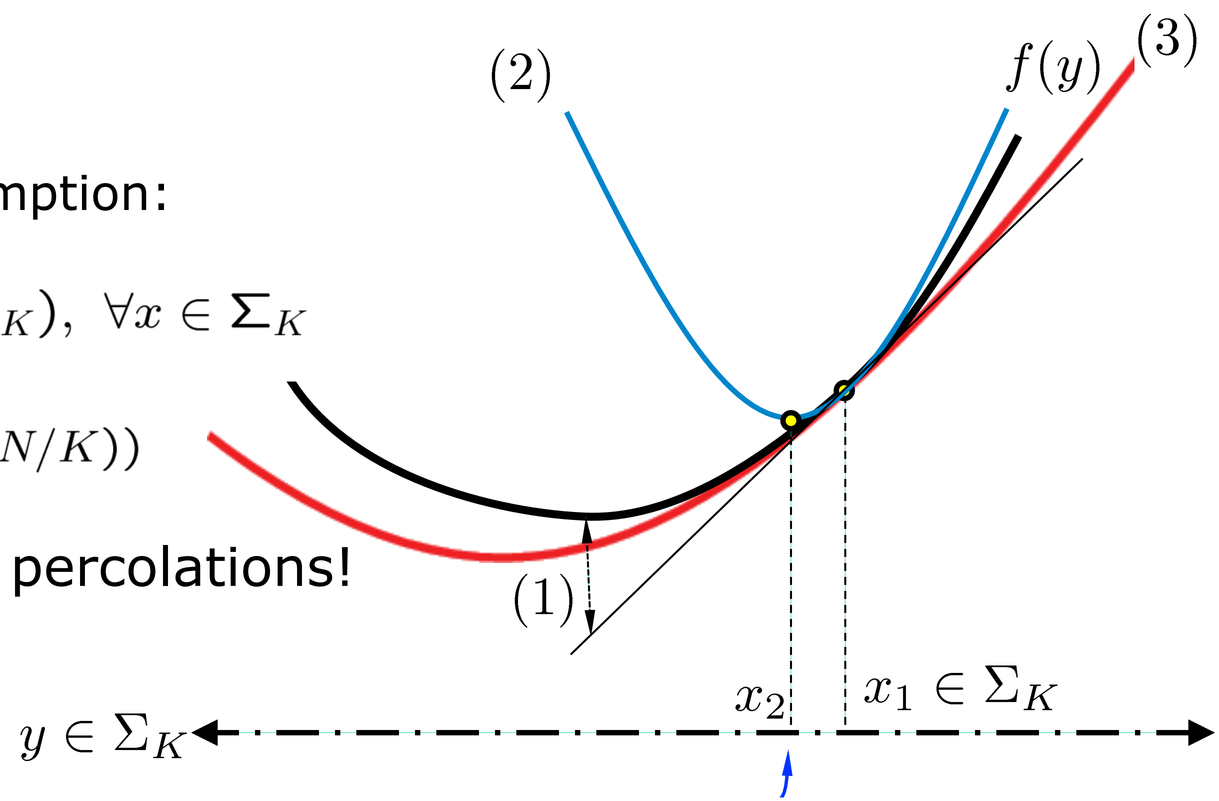
What could possibly go wrong with this naïve approach?

$\binom{N}{K}$ A Tale of Two Algorithms

- Hard thresholding

$f(x) = \|u - \Phi x\|^2$

$\min_{x:\|x\|_0 \le K} f(x)$



Global "unverifiable" assumption:

$$(1-\delta_K) \le \frac{\|\Phi x\|_2^2}{\|x\|_2^2} \le (1+\delta_K), \ \forall x \in \Sigma_K$$

RIP condition $M = O(K \log(N/K))$

$\Longrightarrow$ we can tiptoe among percolations!

$\delta_{2K} < 1/3$

another variant has $\delta_{3K} < 1/2$

GraDes: $x_{i+1} = \mathrm{H}_{\{\|x\|_0 \le K\}}\left(x_i - \frac{1}{L_{2K}} \nabla f(x_i)\right)$

$$\begin{array}{llll}
(1) & f(y) - f(x) - \langle \nabla f(x), y-x \rangle = \|\Phi(y-x)\|^2 & & \forall x, y \in \mathcal{R}^N, \\
(2) & f(y) - f(x) - \langle \nabla f(x), y-x \rangle \le \frac{L_{2K}}{2}\|y-x\|^2 & L_{2K} = 2(1+\delta_{2K}), & \forall x, y \in \Sigma_K, \\
(3) & f(y) - f(x) - \langle \nabla f(x), y-x \rangle \ge \frac{\mu_{2K}}{2}\|y-x\|^2 & \mu_{2K} = 2(1-\delta_{2K}), & \forall x, y \in \Sigma_K,
\end{array}$$

$|\mathcal{M}_K|$ A Model-based CS Algorithm

- Model-based hard thresholding $f(x) = \|u - \Phi x\|^2$

$\min_{x: x \in \Sigma_{\mathcal{M}_K}} f(x)$

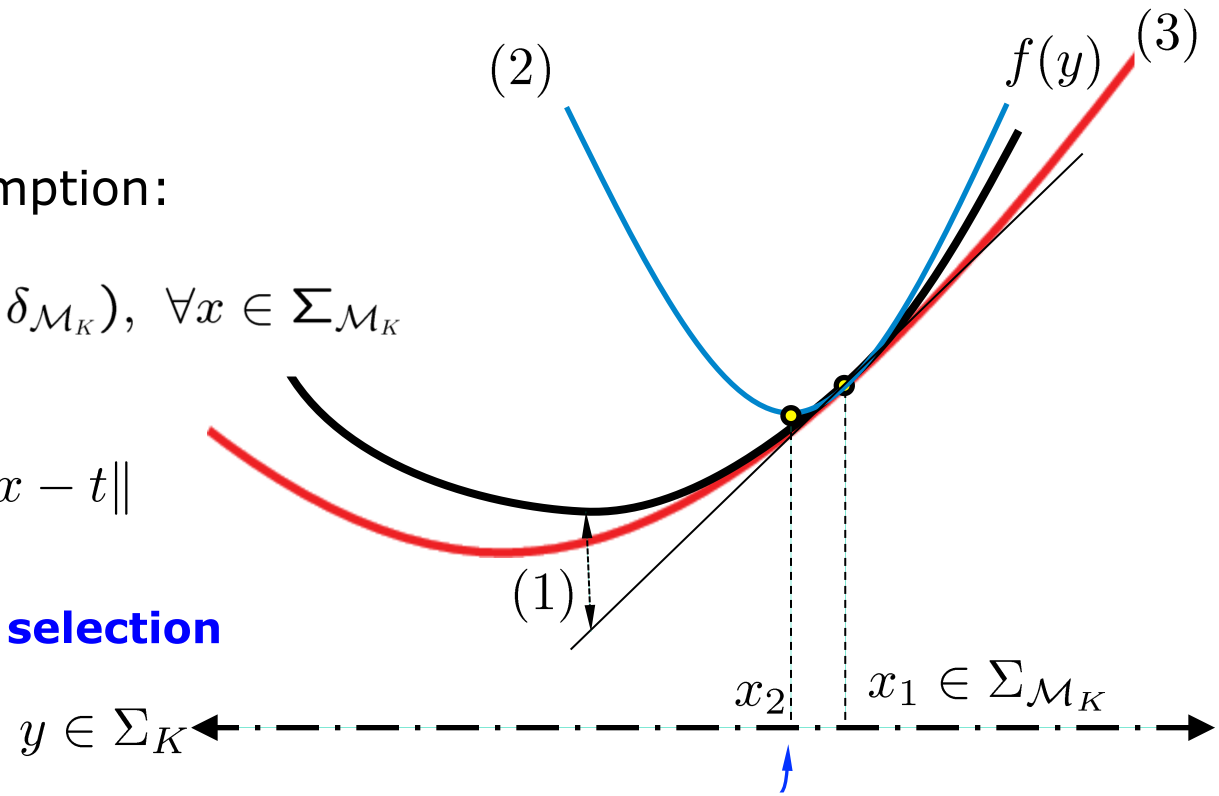
Global "unverifiable" assumption:

$$(1 - \delta_{\mathcal{M}_K}) \leq \frac{\|\Phi x\|_2^2}{\|x\|_2^2} \leq (1 + \delta_{\mathcal{M}_K}), \ \forall x \in \Sigma_{\mathcal{M}_K}$$

$$\mathrm{H}_{\Sigma_{\mathcal{M}_K}}(t) = \arg \min_{x: x \in \mathcal{M}_K} \|x - t\|$$

Key actor: combinatorial selection

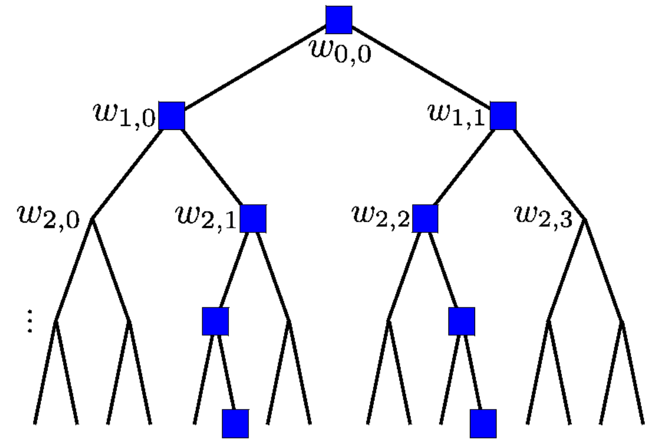
$\delta_{\mathcal{M}_{2K}} < 1/3$



$$x_{i+1} = \mathrm{H}_{\Sigma_{\mathcal{M}_K}}\left(x_i - \frac{1}{L_{\mathcal{M}_{2K}}} \nabla f(x_i)\right)$$

$$\begin{array}{llcll}
(1) & f(y) - f(x) - \langle \nabla f(x), y - x \rangle & = & \|\Phi(y - x)\|^2 & \forall x, y \in \mathcal{R}^N, \\
(2) & f(y) - f(x) - \langle \nabla f(x), y - x \rangle & \leq & \frac{L_{\mathcal{M}_{2K}}}{2}\|y - x\|^2 & L_{\mathcal{M}_{2K}} = 2(1 + \delta_{\mathcal{M}_{2K}}), \forall x, y \in \Sigma_{\mathcal{M}_{2K}}, \\
(3) & f(y) - f(x) - \langle \nabla f(x), y - x \rangle & \geq & \frac{\mu_{\mathcal{M}_{2K}}}{2}\|y - x\|^2 & \mu_{\mathcal{M}_{2K}} = 2(1 - \delta_{\mathcal{M}_{2K}}), \forall x, y \in \Sigma_{\mathcal{M}_{2K}},
\end{array}$$

Tree-Sparse



- **Model:** K-sparse coefficients
  + significant coefficients lie on a rooted subtree

$$M = O(K) < O(K \log(N/K))$$

- **Sparse approx:** find best set of coefficients
  - sorting
  - hard thresholding

- **Tree-sparse approx:** find best rooted subtree of coefficients
  - condensing sort and select [Baraniuk]
  - dynamic programming [Donoho]

Model CS in Context

- Basis pursuit and Lasso

exploit geometry	< >	interplay of ℓ_1 ball and ℓ_2 error
arbitrary selection	< >	*difficulty of interpretation*
		cannot leverage further structure

- Structured-sparsity ***inducing*** norms

"customize" geometry for selection	< >	"mixing" of norms over groups / Lovasz extension of submodular set functions
		inexact selections

- Model CS / structured-sparsity via OMP

exploit combinatorics	< >	exact selections
		cannot leverage geometry

Model CS in Context

- Basis pursuit and Lasso

exploit geometry	< >	interplay of ℓ_1 ball and ℓ_2 error
arbitrary selection	< >	*difficulty of interpretation*
		cannot leverage further structure

- Structured-sparsity ***inducing*** norms

"customize" geometry for selection	< >	"mixing" of norms over groups /
		Lovasz extension of submodular set functions
		inexact selections

- Model CS / structured-sparsity via OMP

exploit combinatorics	< >	exact selections

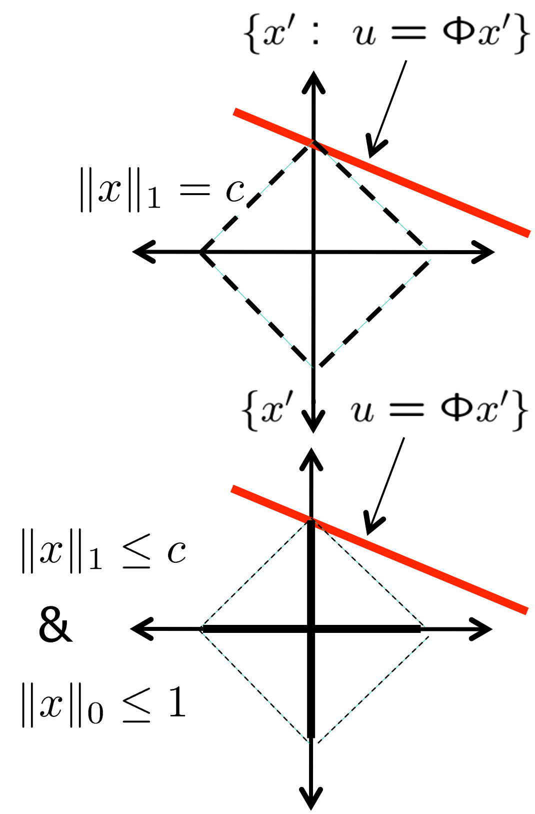
Or, can it?

Enter
CLASH

http://lions.epfl.ch/CLASH

CLASH Pseudocode

- Algorithm code @ http://lions.epfl.ch/CLASH
  - Active set expansion
  - Greedy descend
  - Combinatorial selection
  - Least absolute shrinkage
  - De-bias with convex constraint



$$\widehat{x}_{\text{Clash}} = \arg\min_{x:\|x\|_0 \leq K, \|x\|_1 \leq \lambda} \|u - \Phi x\|^2$$

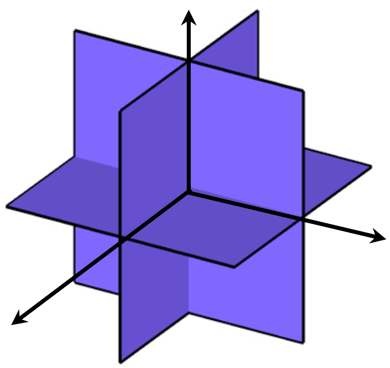
Minimum 1-norm solution still makes sense!

Geometry of CLASH

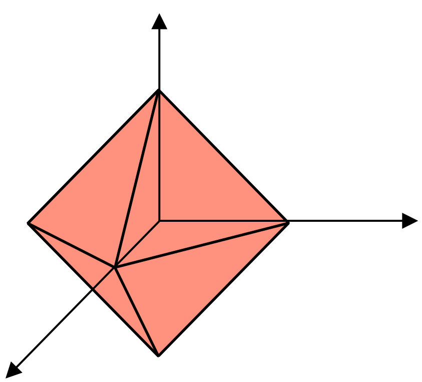
combinatorial selection
+
least absolute shrinkage

$$\mathrm{H}_{\{\|x\|_0 \leq K\}}(t) = \arg \min_{\|x\|_0 \leq K} \|x - t\|$$

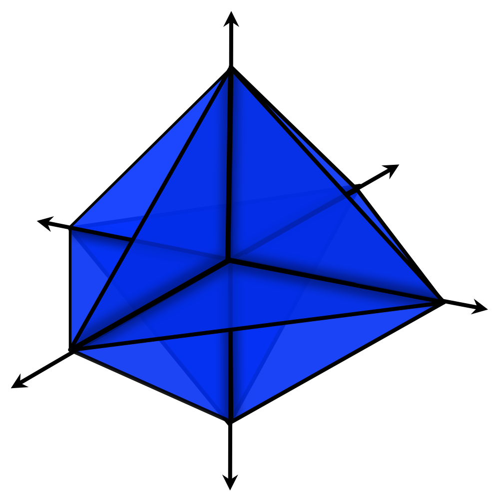
$$\mathrm{St}_{\{\|x\|_1 \leq \lambda\}}(t) = \arg \min_{\|x\|_1 \leq \lambda} \|x - t\|$$



$+$



$\approx$



Geometry of CLASH

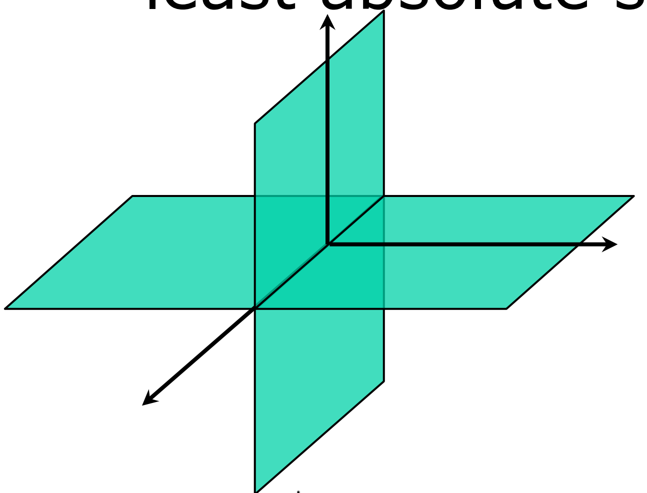
combinatorial selection

+

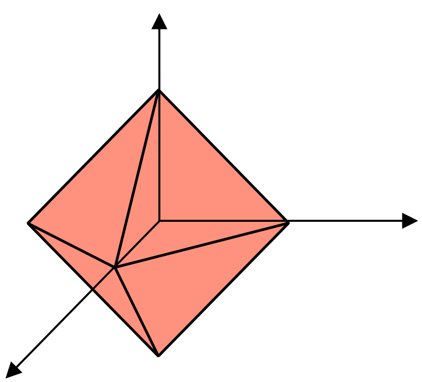
least absolute shrinkage

$$\mathrm{H}_{\Sigma_{\mathcal{M}_K}}(y) = \arg \min_{x : x \in \Sigma_{\mathcal{M}_K}} \|x - y\|$$

$$\mathrm{St}_{\{\|x\|_1 \leq \lambda\}}(t) = \arg \min_{\|x\|_1 \leq \lambda} \|x - t\|$$

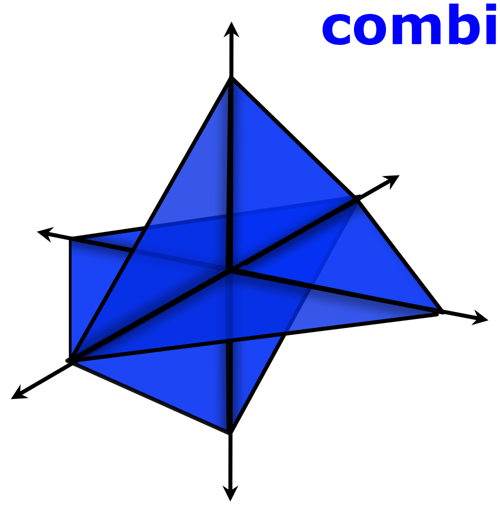


$+$



$\approx$

combinatorial origami



Combinatorial Selection

- A different view of the model-CS workhorse

$$\mathrm{H}_{\Sigma_{\mathcal{M}_K}}(y) = \arg \min_{x : x \in \Sigma_{\mathcal{M}_K}} \|x - y\|$$

(Lemma) support of the solution <> modular approximation problem

$$\operatorname{supp}\left(\arg\min_{x:\operatorname{supp}(x)\in\mathcal{M}_K} \|x-y\|_2^2\right) = \arg\max_{\mathcal{S}:\mathcal{S}\in\bar{\mathcal{M}}_K} F(S;y)$$

indexing set

where $F(S;y) = \sum_{i\in\mathcal{S}} |y_i|^2.$

PMAP

- An algorithmic generalization of union-of-subspaces

  Polynomial time modular epsilon-approximation property: PMAP_ϵ

- Sets with PMAP-0 $F(\widehat{\mathcal{S}}_\epsilon; y) \geq (1-\epsilon) \max_{\mathcal{S} \in \bar{\mathcal{M}}_K} F(\mathcal{S}; y)$

  - **Matroids**

uniform matroids	<>	regular sparsity
partition matroids	<>	block sparsity (disjoint groups)
cographic matroids	<>	rooted connected tree group adapted hull model

  - **Totally unimodular systems**

mutual exclusivity	<>	neuronal spike model
interval constraints	<>	sparsity within groups

Model-CS is applicable for all these cases!

PMAP

- An algorithmic generalization of union-of-subspaces

  Polynomial time modular epsilon-approximation property: PMAP_ϵ

- Sets with PMAP-epsilon $F(\widehat{\mathcal{S}}_\epsilon; y) \geq (1-\epsilon)\max_{\mathcal{S}\in\bar{\mathcal{M}}_K} F(\mathcal{S}; y)$

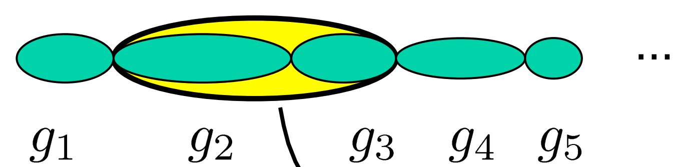
  - **Knapsack**

    multi-knapsack constraints

    weighted multi-knapsack

    quadratic knapsack (?)

  - Define algorithmically!



$$\sum_i w_i z_i \leq G$$

selector variables

$$z_i \in \{0,1\}, \forall i$$

$$z_2 + z_3 \leq 1$$

$$z_2 + z_4 \leq 1$$

$$z_3 + z_4 \leq 1$$

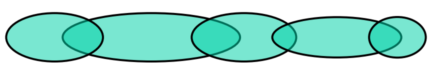
PMAP

- An algorithmic generalization of union-of-subspaces

  Polynomial time modular epsilon-approximation property: PMAP_ϵ

- Sets with PMAP-epsilon $F(\widehat{\mathcal{S}}_\epsilon; y) \geq (1-\epsilon)\max_{\mathcal{S}\in\bar{\mathcal{M}}_K} F(\mathcal{S}; y)$
  - **Knapsack**
  - Define algorithmically!

- Sets with PMAP-???



  - pairwise overlapping groups <> mincut with

    cardinality constraint

$$\max_{\mathcal{S}:\mathcal{S}\in\bar{\mathcal{M}}_K} F(S;\beta) = -\min\left\{\sum_{i>j}\|(\beta)_{g_i\cap g_j}\|_2^2 z_i z_j - \sum_i \|(\beta)_{g_i}\|_2^2 z_i : \sum_i z_i \leq G\right\}.$$

CLASH Approximation Guarantees

- (Theorem) PMAP / downward compatibility

$$\frac{\|x_{i+1} - x^*\|_2}{\|x^*\|_2} \leq \rho \frac{\|x_i - x^*\|_2}{\|x^*\|_2} + \frac{c_1(\delta_{2K}, \delta_{3K}, \epsilon)}{\text{SNR}} + c_2(\delta_{2K}, \delta_{3K}, \epsilon) + c_3(\delta_{2K}, \delta_{3K}, \epsilon)\sqrt{\frac{1}{\text{SNR}}}$$

$$\rho = \frac{\delta_{3K}+\delta_{2K}+\sqrt{\epsilon}(1+\delta_{2K})}{\sqrt{1-\delta_{2K}^2}} \sqrt{\frac{1+\left((1-\epsilon)+2\sqrt{1-\epsilon}\right)\delta_{3K}^2+2\delta_{3K}\sqrt{\epsilon}+\epsilon}{1-\delta_{3K}^2}}$$

$$c_2(\delta_{2K}, \delta_{3K}, \epsilon) = O(\delta_{3K}\sqrt{\epsilon} + \epsilon)$$

$$\text{SNR} = \frac{\|x^*\|}{\sqrt{f(x^*)}}$$

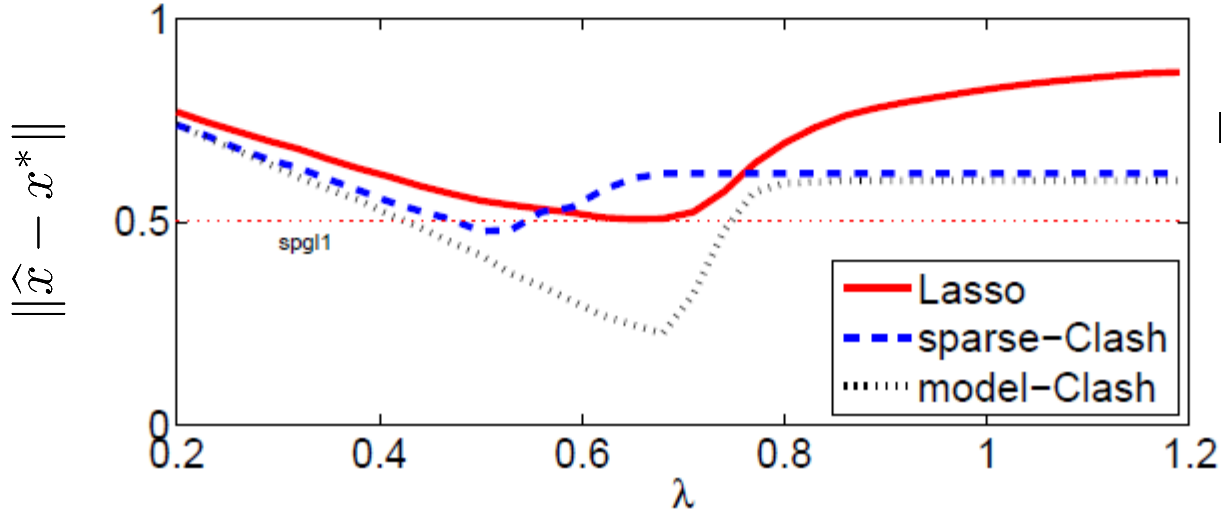
  - precise formulae are in the paper

    http://lions.epfl.ch/CLASH

- Isometry requirement (PMAP-0) <> $\delta_{3K} < 0.3658$

Examples

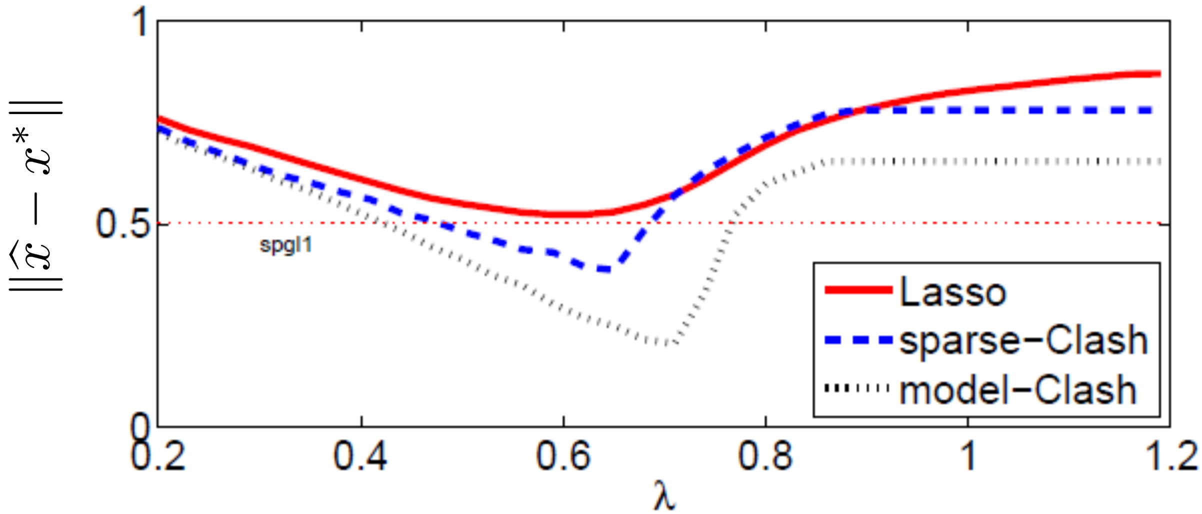
sparse matrix



Model: (K,C)-clustered model

O(KCN) – per iteration

~10-15 iterations

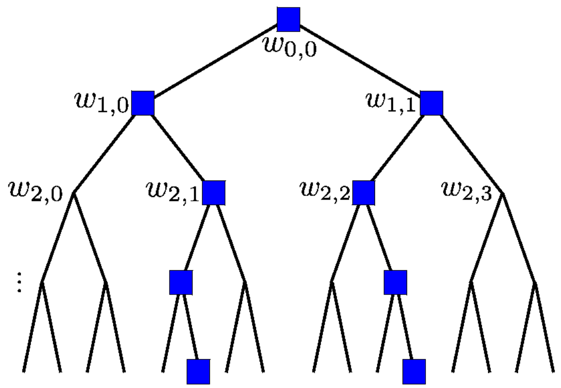


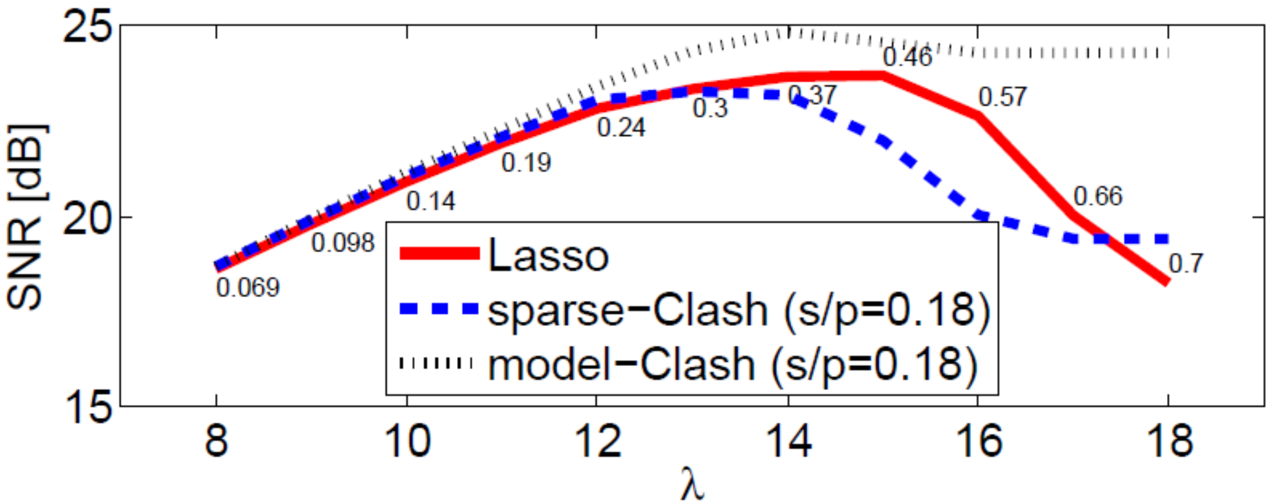
Model: partition model / TU

LP – per iteration

~20-25 iterations

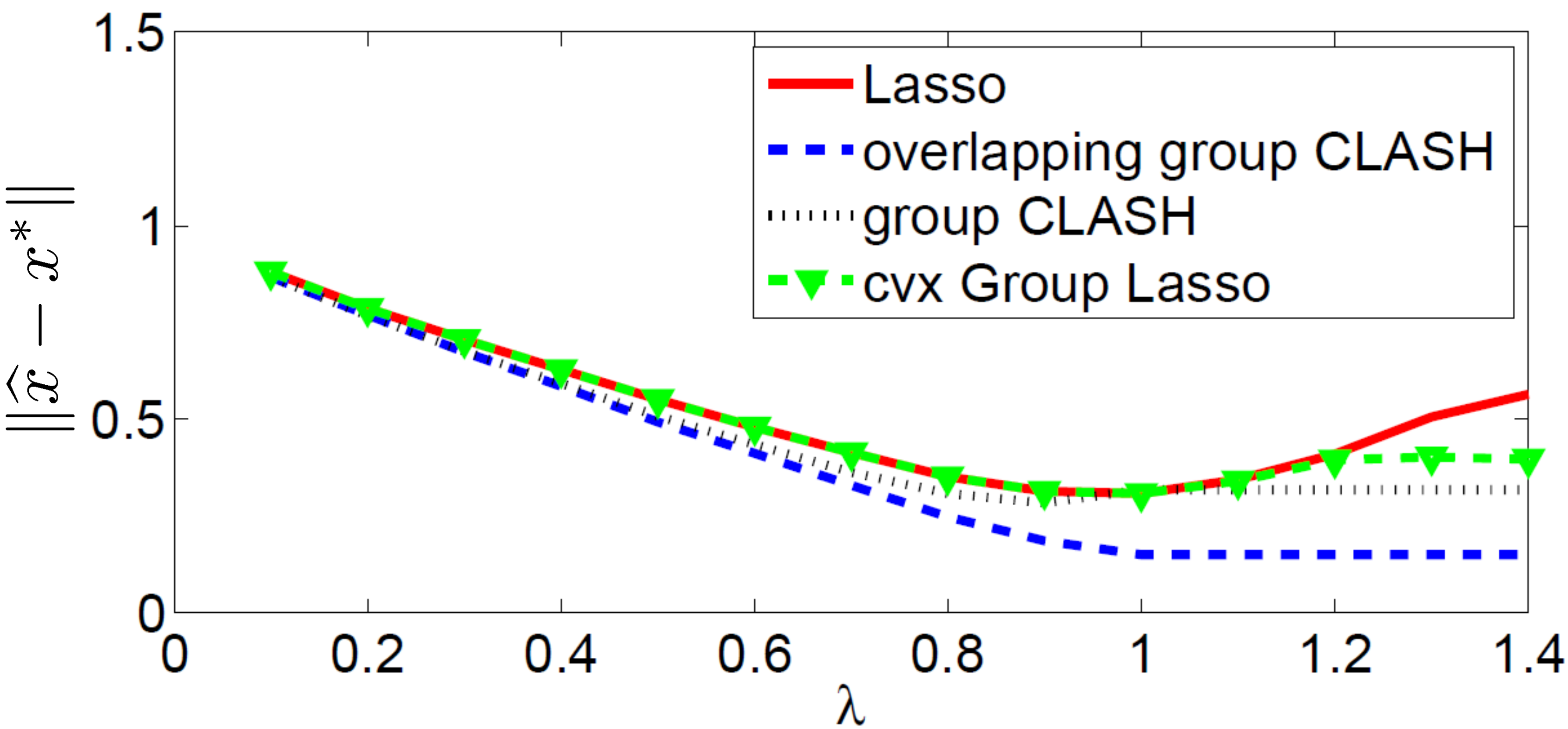
Examples





CCD array readout via noiselets

Examples

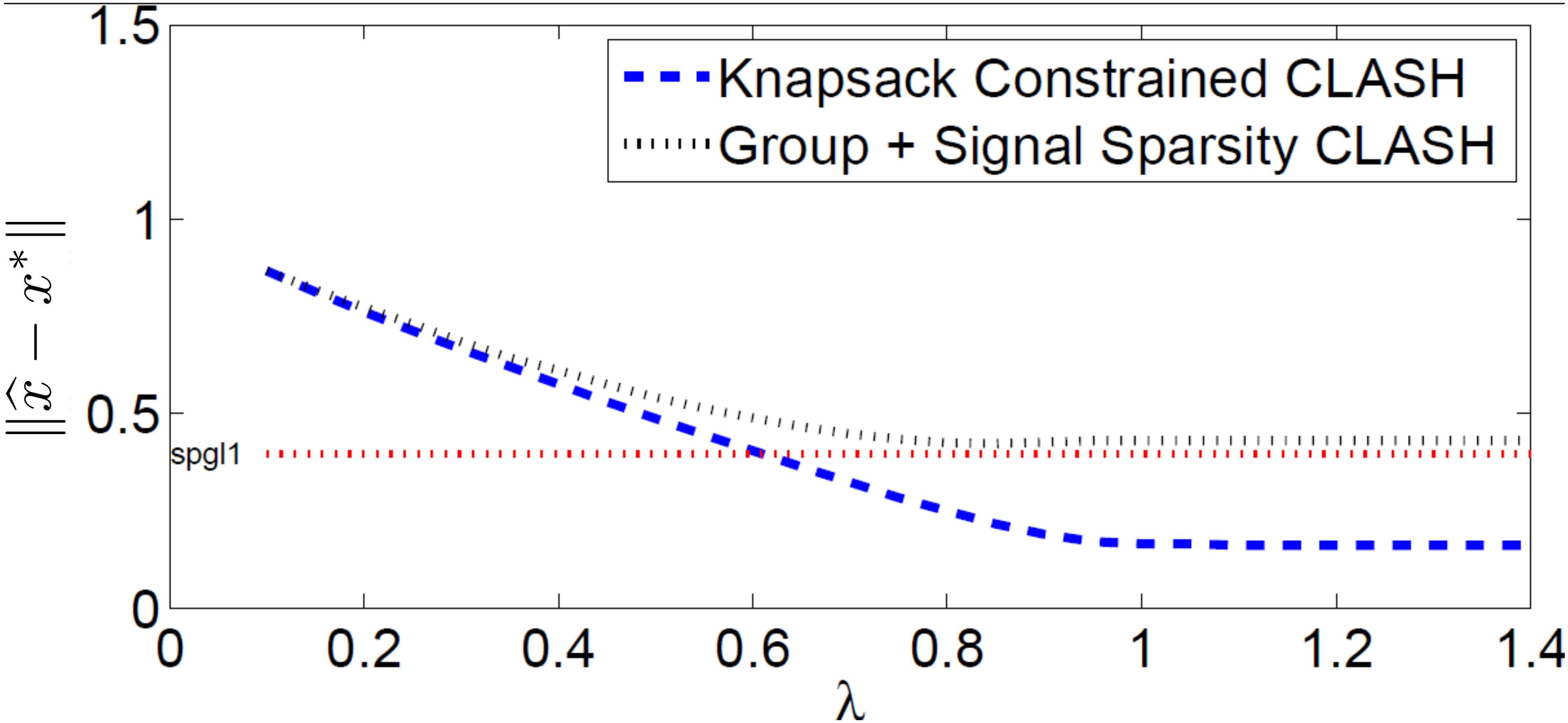


1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
λ
$\|\widehat{x} - x^*\|$
Lasso
overlapping group CLASH
group CLASH
cvx Group Lasso

Examples

g1 + g2 + g17 <= 1
g4 + g5 + g6 + g18 <= 1
g7 + g10 + g11 + g19 <= 1
g13 + g15 + g20 <= 1
... (more constraints with random weights)

```
%      |-------g17-------|      |-----------g18----------|--g19---|
%     ---------------------------------------------------------------
%  x = |  g1  |  g2  |  g3  |  g4  |  g5  |  g6  |  g7  |  g8  |
%     ---------------------------------------------------------------
%
%          |---g19---|---g19---|      |--g20--|      |--g20--|
%     ---------------------------------------------------------------
% ... |  g9  |  g10  |  g11  |  g12  |  g13  |  g14  |  g15  |  g16  |
%     ---------------------------------------------------------------
```



Conclusions

- CLASH <> combinatorial selection
  \+
  convex geometry
  $\lambda \to \infty \Rightarrow$ model-CS

- PMAP-epsilon <> inherent difficulty in combinatorial selection
  - beyond simple selection towards provable solution quality
    \+
    runtime/space bounds
  - algorithmic definition of sparsity + many models
    matroids, TU, knapsack,…

- Other norms / constraints <> TV-norm,…

- Postdoc positions @ LIONS / EPFL

contact: volkan.cevher@epfl.ch