Structured Sparsity

- group testing & compressed sensing
- a norm that induces structured sparsity

(*arxiv.org/abs/1110.0413*
Obozinski, G., Jacob, L., & Vert, J.-P. ,
October 2011)

Mittagsseminar
2011 / 11 / 10
Martin Jaggi

Sparse Solutions to
underdetermined Linear Systems

$$b = Ax$$

$b \in \mathbb{R}^m$
$x \in \mathbb{R}^d$

$m \ll d$

$$\|x\|_0 \leq k$$

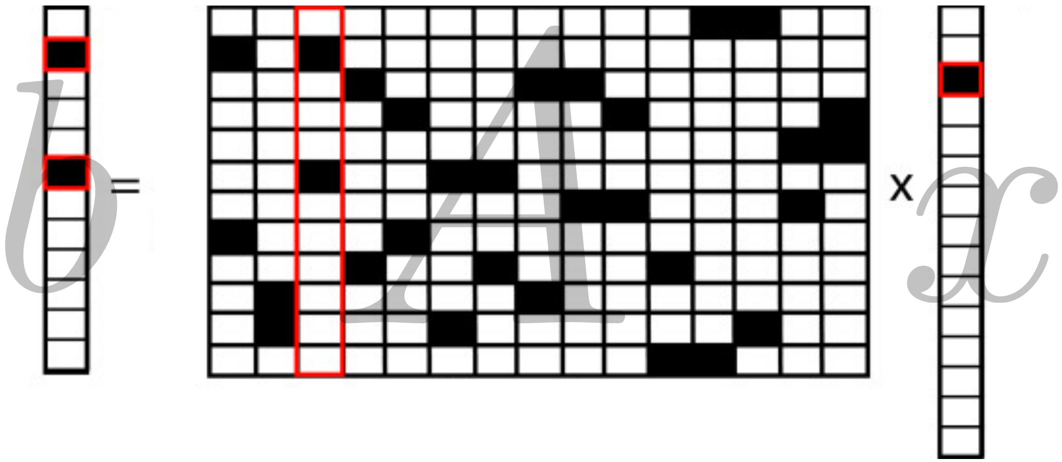
Group Testing

$b \in \mathbb{R}^m$ $\quad m \ll d$

$x \in \mathbb{R}^d$

$\|x\|_0 \leq k$

Dorfman, R. (1943). **The Detection of Defective Members of Large Populations**. *Annals of Mathematical Statistics*

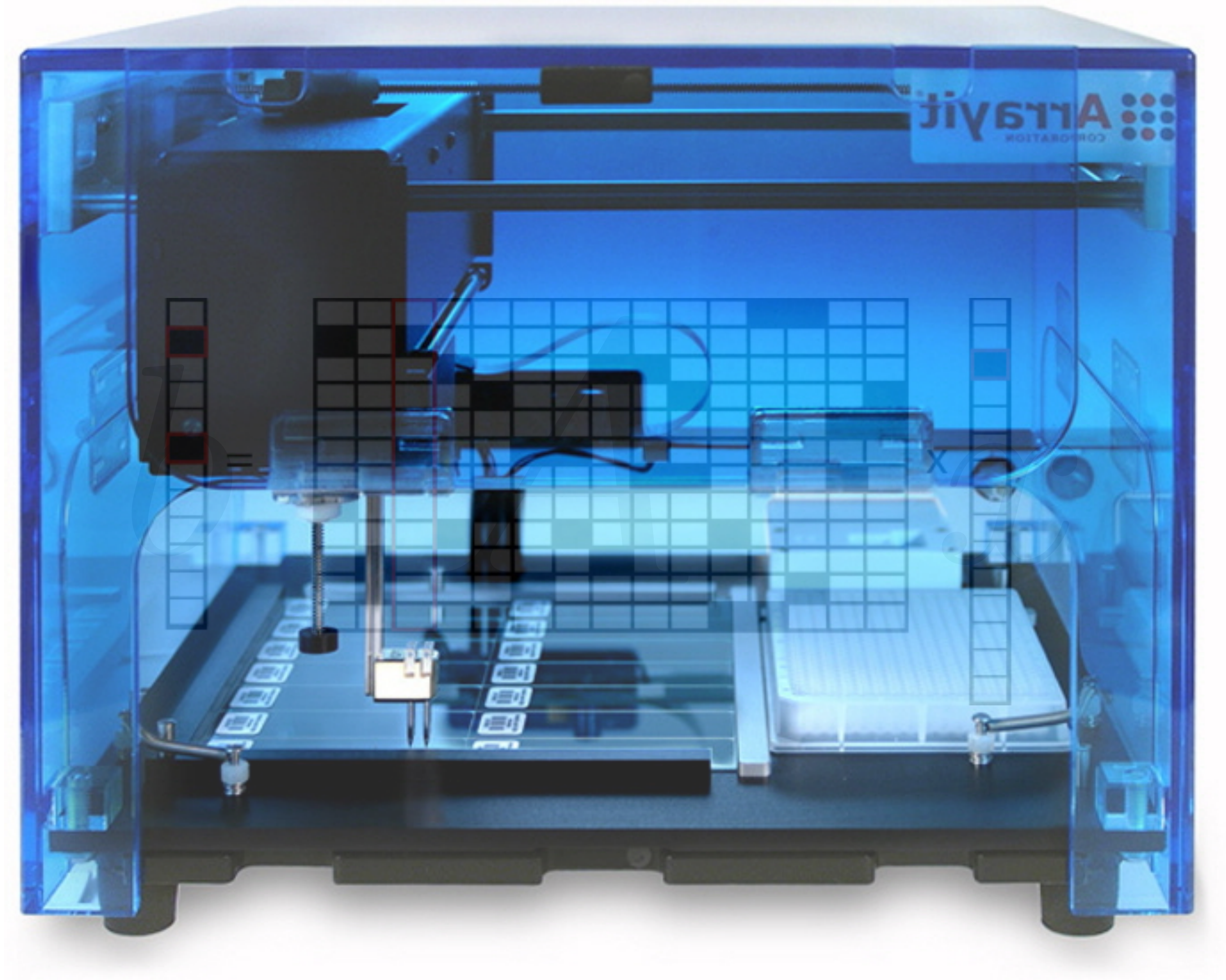


Kainkaryam et al. (2010). **poolMC: Smart pooling of mRNA samples in microarray experiments**. *BMC Bioinformatics*

Coding theory interpretation: A is the *parity check matrix* of a linear code

$$b \in \mathbb{R}^m \qquad m \ll d$$
$$x \in \mathbb{R}^d$$
$$\|x\|_0 \leq k$$

Group Testing

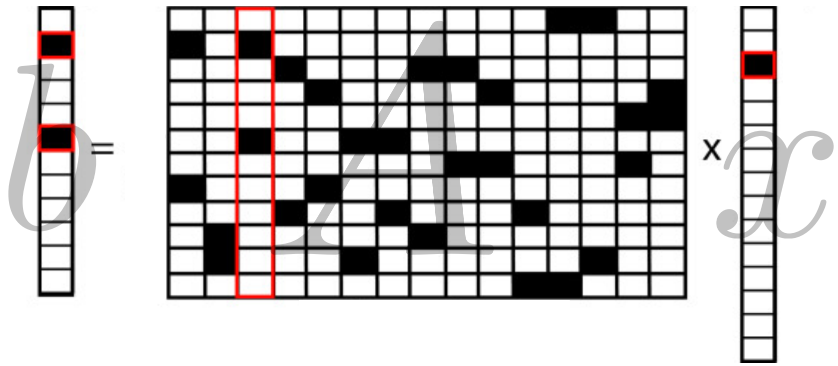


$b \in \mathbb{R}^m$ $\quad m \ll d$

$x \in \mathbb{R}^d$

$\|x\|_0 \leq k$

Compressed sensing



Theorem

$$m = O(k \ \log d)$$ is enough.

k-sparse signals are recovered by taking the solution of smallest ℓ_1-norm

Donoho, D. L. (2006). **Compressed sensing**. *IEEE Transactions on Information Theory* (3400 citations)

Candes, E. J., & Tao, T. (2005). **Decoding by Linear Programming**. *IEEE Transactions on Information Theory* (1200 citations)

hard ℓ_0

$$\min_x \ \|Ax - b\|^2 + \lambda\|x\|_0$$

easy ℓ_1

$$\min_x \ \|Ax - b\|^2 + \lambda\|x\|_1$$

Santosa and Symes (1983)

Phase transition

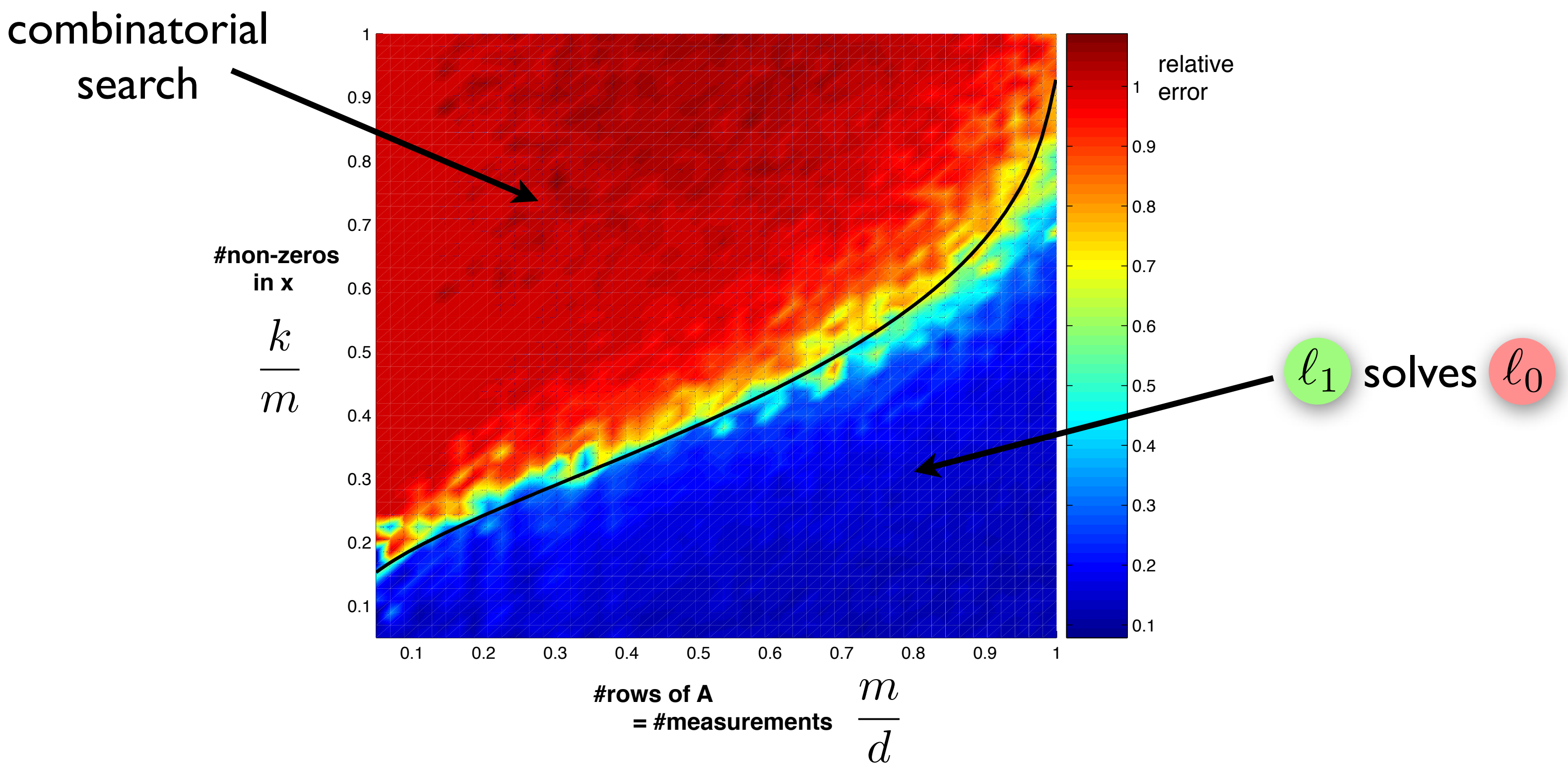
$$b \in \mathbb{R}^m \qquad m \ll d$$
$$x \in \mathbb{R}^d$$
$$\|x\|_0 \le k$$

hard ℓ_0

$$\min_x \; \|Ax - b\|^2 + \lambda \|x\|_0$$

easy ℓ_1

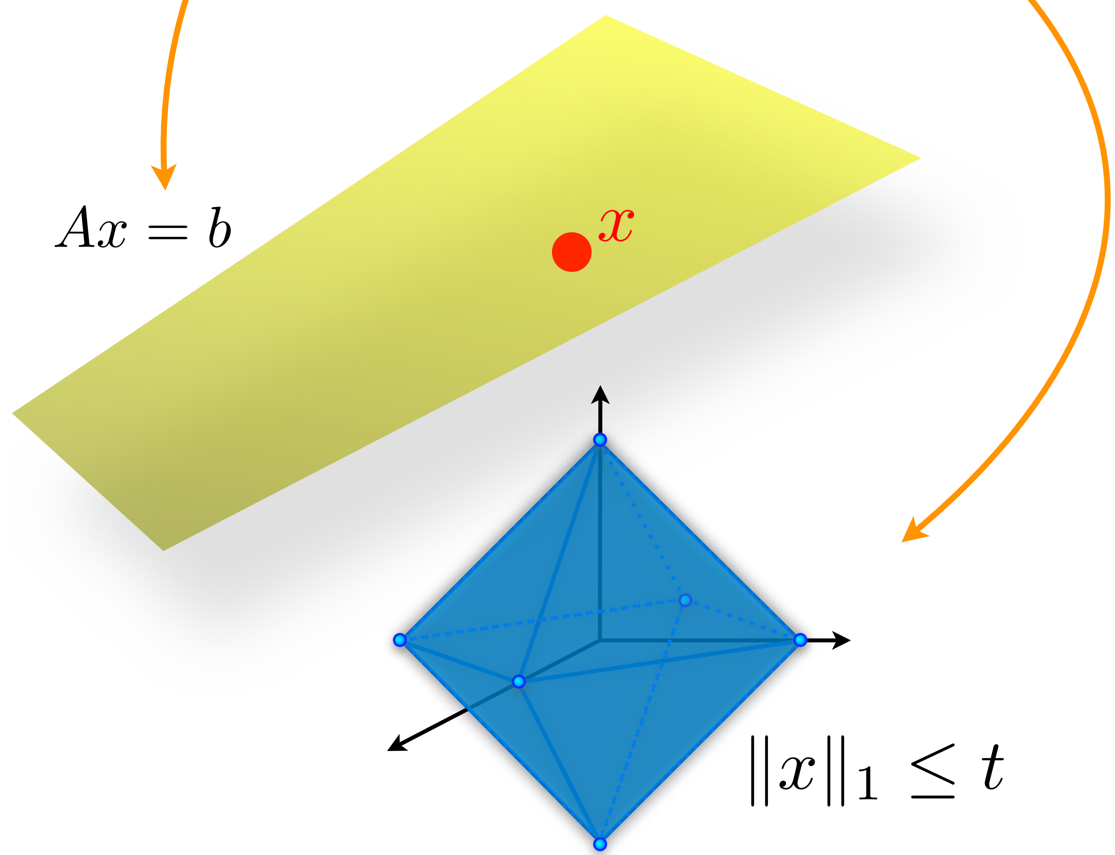
$$\min_x \; \|Ax - b\|^2 + \lambda \|x\|_1$$



Stodden, V. (2006). **Model Selection When The Number Of Variables Exceeds The Number Of Observations**. PhD thesis. stanford.edu.

Why does $\min_x \|Ax - b\|^2 + \lambda\|x\|_1$ have sparse solutions?

easy ℓ_1



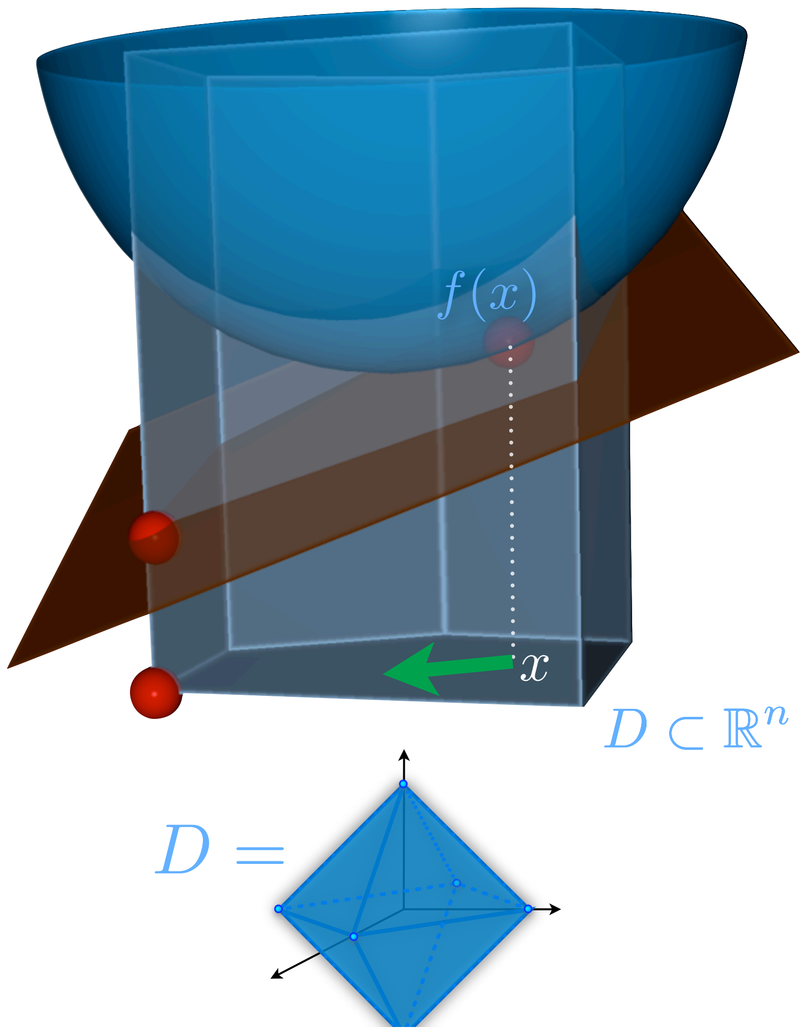
How to solve $\min_x \|Ax - b\|^2 + \lambda \|x\|_1$ (easy ℓ_1) in practice?

Linear Program

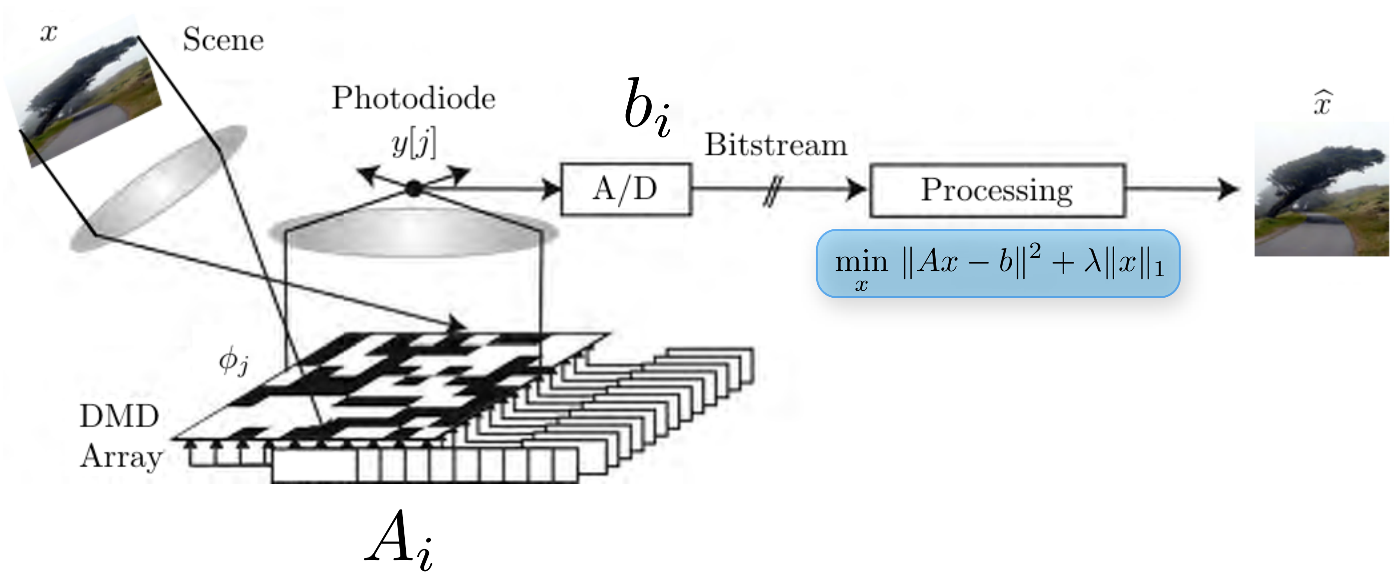
$$\begin{aligned} \min_x \quad & \|x\|_1 \\ s.t. \quad & Ax - b = 0 \end{aligned}$$

Frank-Wolfe (Sparse Greedy)

$$\begin{aligned} \min_x \quad & \|Ax - b\|^2 \\ s.t. \quad & \|x\|_1 \le t \end{aligned}$$



Single Pixel Camera



Duarte et al. **Single-Pixel Imaging via Compressive Sampling**.
IEEE Signal Processing Magazine

Computer Vision

Background Subtraction

$$\min_{X} \|B - X\|_* + \lambda \|X\|_1$$

(a) Original frames B (b) $B - X$ (c) Sparse X

Candes, E. J. et al. (2011). **Robust principal component analysis**
Journal of the ACM

A Structured Norm

Obozinski, G., Jacob, L., & Vert, J.-P. (October 2011).
Group Lasso with Overlaps: the Latent Group Lasso approach.
arXiv stat.ML.

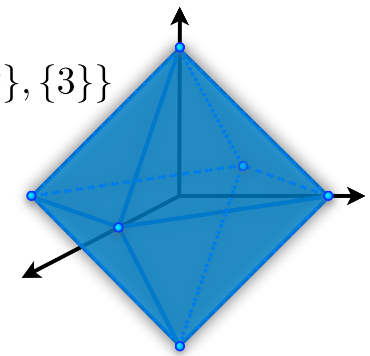
$\mathcal{G}$ is a collection of subsets $g \subseteq [d]$

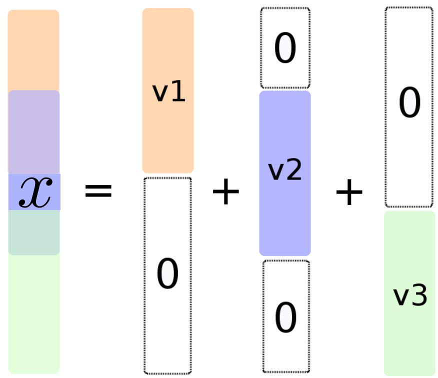
$$\bigcup_{g \in \mathcal{G}} g = [d]$$

$$\|x\|_1 := \min_{(v^i)} \sum_{i \in [d]} |v^i| \quad s.t. \quad x = \sum_{i \in [d]} v^i \quad supp(v^i) = \{i\}$$

$$\|x\|_{\mathcal{G}} := \min_{(v^g)} \sum_{g \in \mathcal{G}} \|v^g\|_g \quad s.t. \quad x = \sum_{g \in \mathcal{G}} v^g \quad supp(v^g) \subseteq g$$

$\mathcal{G} = \{\{1\}, \{2\}, \{3\}\}$





A Structured Norm

$\mathcal{G}$ is a collection of subsets $g \subseteq [d]$

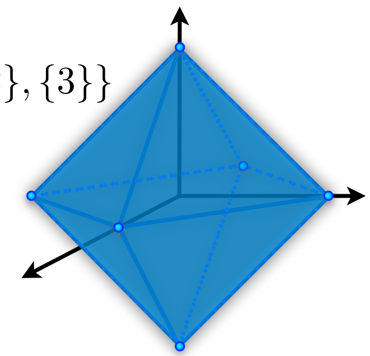
$$\bigcup_{g \in \mathcal{G}} g = [d]$$

Obozinski, G., Jacob, L., & Vert, J.-P. (October 2011).
Group Lasso with Overlaps: the Latent Group Lasso approach.
arXiv stat.ML.

$$\|x\|_1 := \min_{(v^i)} \sum_{i \in [d]} |v^i| \quad s.t. \quad x = \sum_{i \in [d]} v^i, \quad supp(v^i) = \{i\}$$

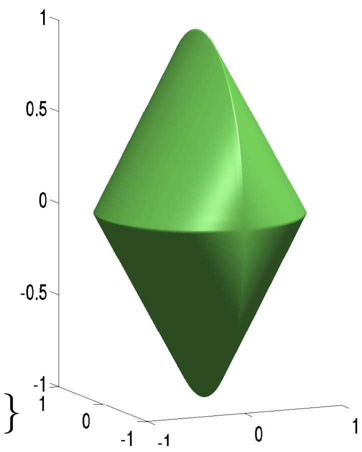
$$\|x\|_{\mathcal{G}} := \min_{(v^g)} \sum_{g \in \mathcal{G}} \|v^g\|_g \quad s.t. \quad x = \sum_{g \in \mathcal{G}} v^g, \quad supp(v^g) \subseteq g$$

$\mathcal{G} = \{\{1\}, \{2\}, \{3\}\}$



Lemma: the unit ball of $\|\cdot\|_{\mathcal{G}}$ is the convex hull of the union of disks

$$\mathcal{D}^g = \left\{ v \in \mathbb{R}^d \;\middle|\; \begin{array}{c} supp(v^g) \subseteq g, \\ \|v^g\|_g \leq 1 \end{array} \right\}$$



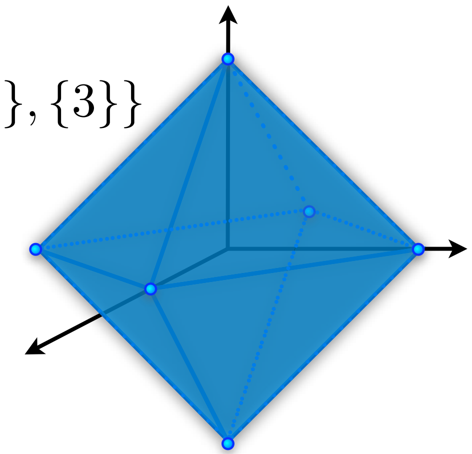
$\mathcal{G} = \{\{1, 2\}, \{2, 3\}\}$

A Structured Norm

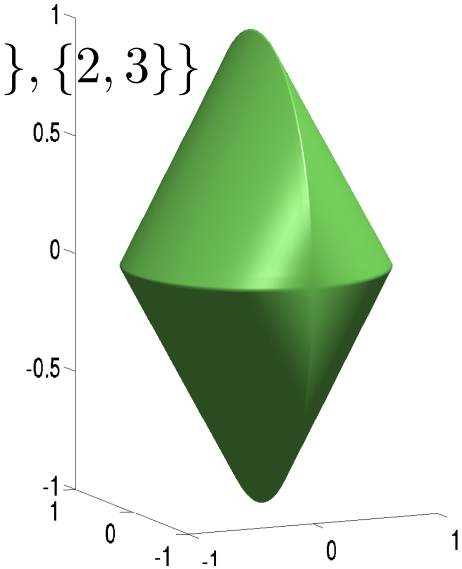
Obozinski, G., Jacob, L., & Vert, J.-P. (October 2011).
Group Lasso with Overlaps: the Latent Group Lasso approach.
arXiv stat.ML.

$$\|x\|_{\mathcal{G}} := \min_{(v^g)} \sum_{g\in\mathcal{G}} \|v^g\|_g \quad s.t. \quad x = \sum_{g\in\mathcal{G}} v^g$$

$$\mathcal{G} = \{\{1\},\{2\},\{3\}\}$$

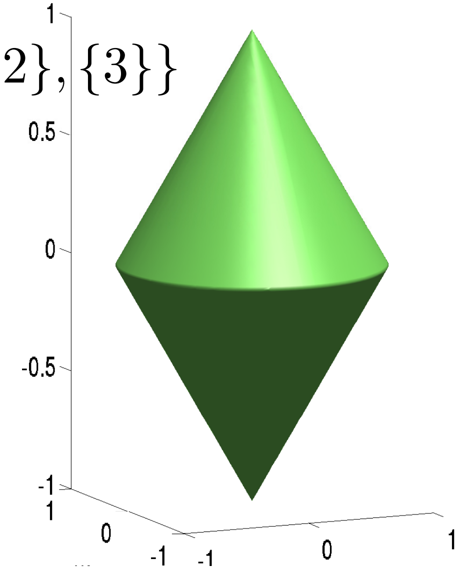


$$\mathcal{G} = \{\{1,2\},\{2,3\}\}$$

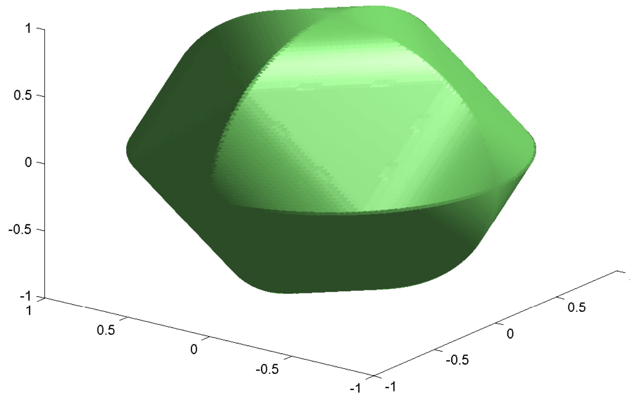


Lemma: the unit ball of $\|.\|_{\mathcal{G}}$ is the convex hull of a union of disks

$$\mathcal{G} = \{\{1,2\},\{3\}\}$$



$$\mathcal{G} = \{\{1,2\},\{1,3\},\{2,3\}\}$$

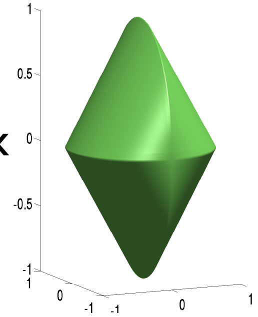


Optimizing with the Structured Norm

$$\|x\|_{\mathcal{G}} := \min_{(v^g)} \sum_{g \in \mathcal{G}} \|v^g\|_g \quad s.t. \quad x = \sum_{g \in \mathcal{G}} v^g$$

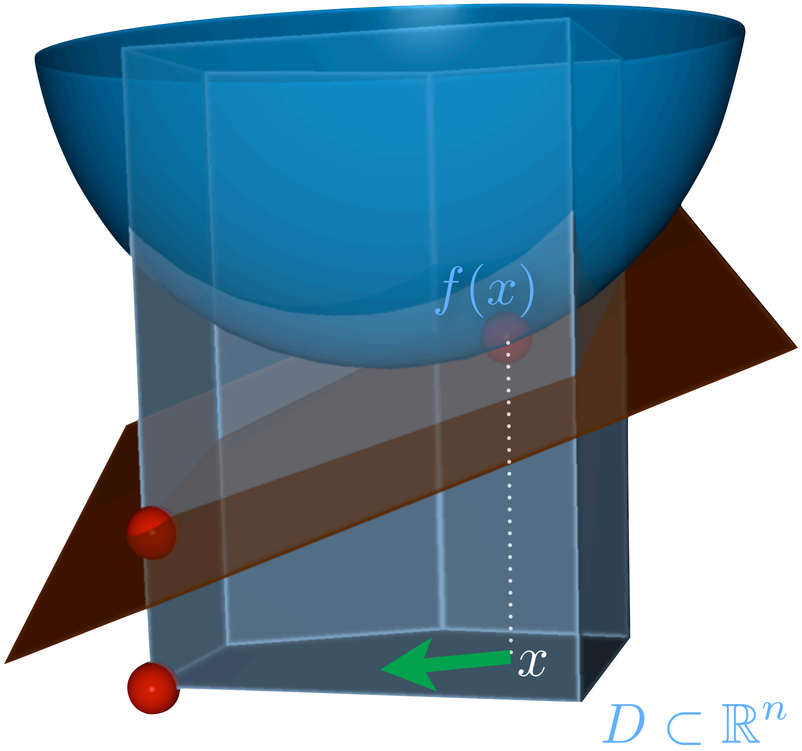
$$\min_x \; f(x) + \lambda \|x\|_{\mathcal{G}}$$

Lemma: the unit ball of $\|.\|_{\mathcal{G}}$ is the convex hull of a union of disks



Frank-Wolfe (Sparse Greedy)

$$\min_x \quad f(x)$$
$$s.t. \quad \|x\|_{\mathcal{G}} \leq t$$



Relation to Set-Cover

$\mathcal{G}$ is a collection of subsets $g \subseteq [d]$

$$\bigcup_{g \in \mathcal{G}} g = [d]$$

$$\begin{aligned} \|x\|_1 := \min_{(v^i)} \quad & \sum_{i \in [d]} |v^i| \\ s.t. \quad & x = \sum_{i \in [d]} v^i \\ & supp(v^i) = \{i\} \end{aligned}$$

$$\begin{aligned} \|x\|_{\mathcal{G}} := \min_{(v^g)} \quad & \sum_{g \in \mathcal{G}} \|v^g\|_g \\ s.t. \quad & x = \sum_{g \in \mathcal{G}} v^g \\ & supp(v^g) \subseteq g \end{aligned}$$

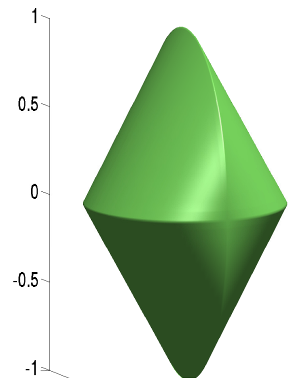
$$\begin{aligned} \|x\|_0 := \min_{(v^i)} \quad & \sum_{i \in [d]} \mathbf{1}_{v^i \neq 0} \\ s.t. \quad & x = \sum_{i \in [d]} v^i \\ & supp(v^i) = \{i\} \end{aligned}$$

$$\begin{aligned} \|x\|_{\substack{\mathcal{G}-\text{set} \\ \text{cover}}} := \min_{(v^g)} \quad & \sum_{g \in \mathcal{G}} \mathbf{1}_{v^g \neq 0} \\ s.t. \quad & x = \sum_{g \in \mathcal{G}} v^g \\ & supp(v^g) \subseteq g \end{aligned}$$

Open Questions

$$\|x\|_{\mathcal{G}} := \min_{(v^g)} \sum_{g \in \mathcal{G}} \|v^g\|_g \quad s.t. \quad x = \sum_{g \in \mathcal{G}} v^g$$

- More applications (related to set-cover?)
- Phase transition phenomenon when applied to the combinatorial set-cover?
- Is it the "closest" convex function to set-cover?



$\mathcal{G} = \{\{1,2\},\{2,3\}\}$