Identification of Overcomplete Dictionaries

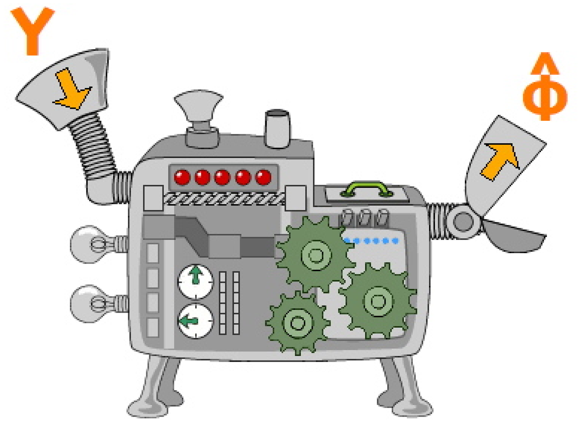
Karin Schnass
kschnass@uniss.it

Computer Vision Laboratory
University of Sassari

What do we want to do?

We have finally found intelligent life and now want to identify a dictionary Φ from the received sparse signals, $Y = \Phi X$.

$Y = \Phi X$



$\hat{\Phi} \simeq \Phi$?

We could use

ER-SpUD [7] - but we think the dictionary is overcomplete...

K-SVD [1, 5] - but we think that the dictionary is non-tight...

ℓ_1-minimisation [3, 2, 4] - but this might take a while...

...or think of something new [6].

A response maximisation principle

The reason why K-SVD might have trouble recovering non-tight frames is that for random sparse signals $\Phi_I x_I$ and an ε-perturbation Ψ the average of the largest squared response behaves like

$$\frac{1}{2}\left(1-\frac{\varepsilon^2}{2}+c(\Psi)\right)^2+\frac{1}{2}\left(1-\frac{\varepsilon^2}{2}-c(\Psi)\right)^2 = 1-\varepsilon^2+\frac{\varepsilon^4}{4}+c(\Psi)^2.$$

The term $c(\Psi)$ is constant over all dictionaries if (and only if ???) Φ is tight and therefore there is a local maximum at Φ.
But since the average of the largest *absolute* response should behave like

$$\frac{1}{2}\left|1-\frac{\varepsilon^2}{2}+c(\Psi)\right|+\frac{1}{2}\left|1-\frac{\varepsilon^2}{2}-c(\Psi)\right| = 1-\frac{\varepsilon^2}{2},$$

why not simply maximise the absolute norm of the S-largest responses,

$$(P_{R1}) \qquad \max_{\Psi\in\mathcal{D}} \sum_n \max_{|I|=S} \|\Psi_I^\star y_n\|_1. \qquad (1)$$

A definition

A probability measure ν on the unit sphere $S^{K-1} \subset \mathbb{R}^K$ is called symmetric if for all measurable sets $\mathcal{X} \subseteq S^{K-1}$, for all sign sequences $\sigma \in \{-1,1\}^K$ and all permutations p we have

$$\nu(\sigma\mathcal{X}) = \nu(\mathcal{X}), \quad \text{where} \quad \sigma\mathcal{X} := \{(\sigma_1 x_1,\dots,\sigma_K x_K) : x\in\mathcal{X}\} \qquad (2)$$

$$\nu(p(\mathcal{X})) = \nu(\mathcal{X}), \quad \text{where} \quad p(\mathcal{X}) := \{(x_{p(1)},\dots,x_{p(K)}) : x\in\mathcal{X}\} \qquad (3)$$

The signal model

We assume that our signals are generated as

$$y_n = \frac{\Phi x_n + r_n}{\sqrt{1+\|r_n\|_2^2}} \qquad (4)$$

where the coefficients x_n are drawn from a symmetric probability distribution ν on the unit sphere and $r_n = (r_n(1)\dots r_n(d))$ is a centred random subgaussian noise-vectors with parameter ρ, ie. the $r_n(i)$ are independent and we have $\mathbb{E}[\exp(t r_n(i))] \leq \exp(t^2\rho^2/2)$.

The result

Let Φ be a unit norm frame with frame constants $A \leq B$ and coherence μ and let the signals follow the model above. If there is a gap between the S and $S+1$-largest coefficients $c_S(x_n)$ and $c_{S+1}(x_n)$ such that for some $0 < q < 1/4$ the number of samples N, the noiselevel ρ and the coherence μ satisfy

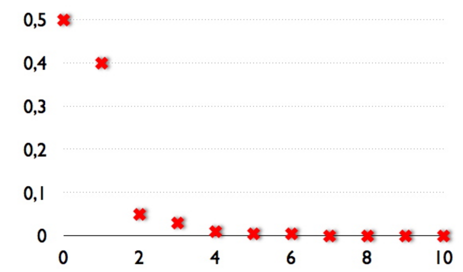
$$\max\{\mu,\rho,N^{-q}\} < \mathcal{O}\left(\frac{c_S(x_n)-c_{S+1}(x_n)}{\sqrt{\log K}}\right), \quad a.s. \qquad (5)$$

then except with probability

$$\exp\left(-\mathcal{O}\left(\frac{N^{1-4q}}{K^2}\right)+\mathcal{O}\left(Kd\log(NK)\right)\right)$$

there is a local maximum of (1) near Φ.

This means:
Our machine succeeds if we have a **gap** between the relevant and the irrelevant coefficients and $\mathcal{O}(K^3 d)$ **training signals**.



The algorithm

We derive an Iterative Thresholding & K-Means algorithm (ITKM) using LaGrange multipliers. We have

$$\frac{\partial}{\partial\psi_k}\left(\sum_n \max_{|I|=S}\|\Psi_I^\star y_n\|_1\right) = \sum_{n:k\in I(\Psi,y_n)} \operatorname{sign}(\langle\psi_k,y_n\rangle)y_n^\star,$$

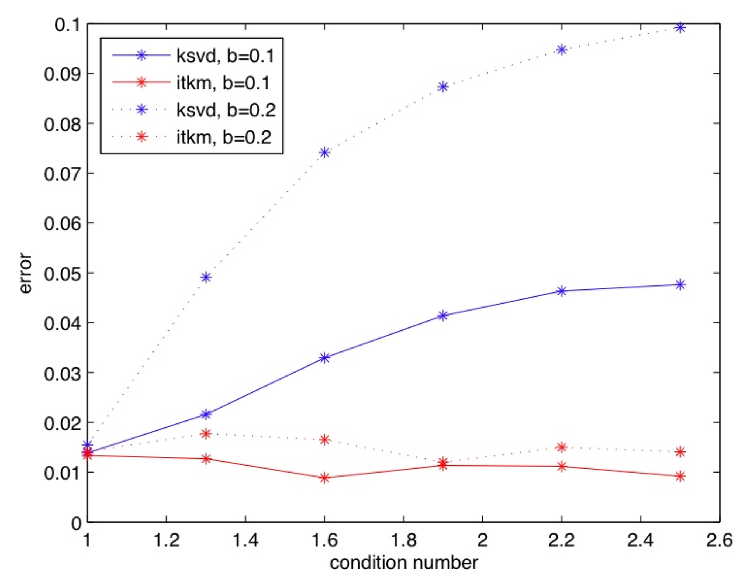
$$\frac{\partial}{\partial\psi_k}\left(\|\psi_k\|_2^2\right) = 2\psi_k^\star, \quad \text{where} \quad I(\Psi,y_n) := \arg\max_{|I|=S}\|\Phi_I^\star y_n\|_1,$$

which leads to the update rule,

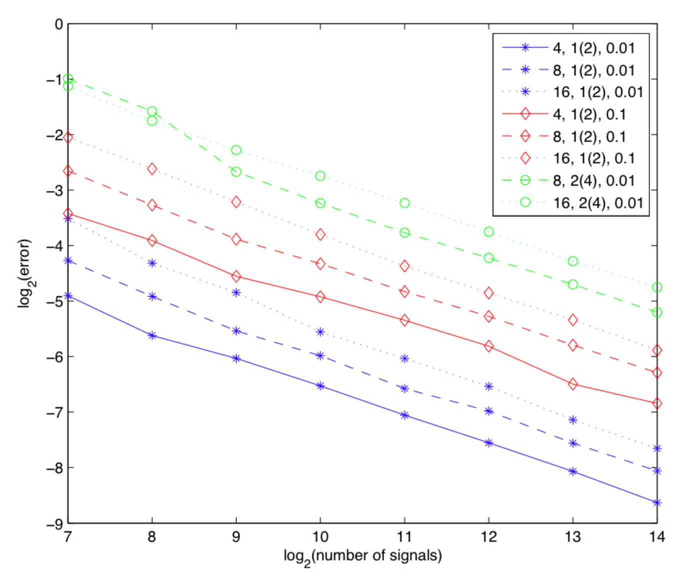
$$\psi_k^{new} = \lambda_k \cdot \sum_{n:k\in I(\Psi^{old},y_n)} \operatorname{sign}(\langle\psi_k^{old},y_n\rangle)y_n \qquad (6)$$

where λ_k is a scaling parameter ensuring that $\|\psi_k^{new}\|_2 = 1$.

Pretty Pictures



Local recovery error for K-SVD and ITKM for bases with varying condition numbers in $\mathbb{R}^3$, symmetrisation of c_1 uniformly distributed in $[1-b,1]$, $c_2 = \sqrt{1-c_1^2}$ and $c_3 = 0$.



Decay of recovery error of a (Dirac+$\frac{1}{2}$ Hadamard)-dictionary with increasing number of training signals and for various coefficient distributions.

Comparison & Discussion

The proposed principle compares quite favorably to classic machines:

Machine	ℓ_1-min.	ER-SpUD	K-SVD-princ.	P_{R1}
Overcomplete Φ	✓	☹	✓	✓
Non-tight Φ	✓	✓	☹/?	✓
Sparsity $O(\cdot)$	μ^{-1}	$\sqrt{d}$	μ^{-1}	μ^{-2}
Samples $O(\cdot)$	$K^3 d$	K^2	$K^3 d$	$K^3 d$
Noise stability	✓	?	✓	✓
Fast algorithm	☹	✓	?	✓
local = global	?	✓	?	☹

To make it really useful in practice we should figure out ways how to get to the global optimum and extend the result to the unit norm signal model.

References

[1] M. Aharon, M. Elad, and A.M. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing.*, 54(11):4311–4322, November 2006.

[2] Q. Geng, H. Wang, and J. Wright. On the local correctness of ℓ^1-minimization for dictionary learning. *arXiv:1101.5672*, 2011.

[3] R. Gribonval and K. Schnass. Dictionary identifiability - sparse matrix-factorisation via l_1-minimisation. *IEEE Transactions on Information Theory*, 56(7):3523–3539, July 2010.

[4] R. Jenatton, F. Bach, and R. Gribonval. Local stability and robustness of sparse dictionary learning in the presence of noise. *preprint*, 2012.

[5] K. Schnass. On the identifiability of overcomplete dictionaries via the minimisation principle underlying K-SVD. *submitted*, 2013.

[6] K. Schnass. Identification of overcomplete dictionaries. *in preparation since forever.*

[7] D. Spielman, H. Wang, and J. Wright. Exact recovery of sparsely-used dictionaries. In *Conference on Learning Theory (arXiv:1206.5882)*, 2012.