A good reason for using OMP: average case results

Karin Schnass
Department of Mathematics, University of Innsbruck
Technikerstraße 13, 6020 Innsbruck, Austria
Email: karin.schnass@uibk.ac.at

I. Extended Abstract

In sparse approximation the goal is to approximate a given signal $y \in \mathbb{R}^d$ by a linear combination of a small number $S \ll d$ of elements $\varphi_i \in \mathbb{R}^d$, called atoms, out of a given larger set, such as a basis or a frame, called the dictionary. Storing the normalised atoms as columns in the dictionary matrix $\Phi = (\varphi_1 \dots, \varphi_K)$, and denoting the restriction to the columns indexed by a set I by Φ_I, we can write informally,

$$\text{find} \quad y \approx \sum\nolimits_{k\in I} \varphi_k x_k = \Phi_I x_I \quad \text{s.t.} \quad |I| = S \ll d. \tag{1}$$

Finding the smallest error for a given sparsity level S and the corresponding support set I, which determines x_I via $x_I = \Phi_I^\dagger y$ for $\Phi_I^\dagger$ the Moore-Penrose pseudo inverse, becomes an NP-hard problem in general unless the dictionary is an orthonormal system. In this case thresholding, meaning choosing as I the indices of the atoms having the S-largest inner products with the signal in magnitude, will succeed. For all other cases, one had to find algorithms which are more efficient, if less optimal, than an exhaustive search through all possible supports sets I with subsequent projection $P(\Phi_I)y := \Phi_I \Phi_I^\dagger y$. The two most investigated directions are greedy methods and convex relaxation techniques - the two golden classics being Orthogonal Matching Pursuit (OMP), [1], and Basis Pursuit (BP), [2], respectively.

The interesting question concerning both schemes is when they are successful, assuming that the signal y is known to be S-sparse, meaning $y = \Phi_I x_I$ with $|I| = S$. It was first studied in [3], [4] and for dictionaries with coherence $\mu := \max_{j\neq k} |\langle \varphi_j, \varphi_k\rangle|$ a sufficient condition for both schemes is that $2S\mu < 1$, which is quite restrictive, especially considering the much better performance in practice. This led people to look at the average performance when modelling the signals as generated via

$$y = \sum\nolimits_k \sigma_k c_k \varphi_{p(k)}, \tag{2}$$

where $(\sigma_k)_k$ is a Rademacher sequence, the coefficient sequence c is non-increasing, $c_k \geq c_{k+1} \geq 0$, and $c_k = 0$ for $k > S$ and p is some permutation such that the support $I = \{p(1), \dots, p(S)\}\}$ satisfies $\delta_I := \|\Phi_I^\top \Phi_I - \mathbb{I}_d\|_{2,2} \leq \frac{1}{2}$.

It was shown, [5], that BP recovers the true support except with probability $2K^{1-2m}$ as long as $16\mu^2 S \cdot m \log K \leq 1$. The fact that for OMP a similar result could only be found in a multi-signal scenario, [6], started to give OMP the reputation of being weaker than BP. This was further increased by the advent of Compressed Sensing (CS), [7], which can be seen as sparse approximation with design freedom for the dictionary. While for BP-type schemes in combination with randomly chosen dictionaries strong results appeared very early, [8], [9], comparable results for OMP and its variants took longer to develop and are weaker in general, [10], [11]. Still, thanks to its computational advantages and flexibility, e.g. concerning the stopping criteria, OMP remained popular in signal processing - the only difference being that users had a defensive statement a la 'of course BP will perform even better' ready at all times.

In [12] (resp. its extendend version [13]) we provide the long missing analysis of the average performance of OMP and show that on average neither BP nor OMP are stronger, but confirm folklore wisdom, that OMP works better for signals with decaying coefficients while BP is better for equally sized coefficients.

Concretely for the noiseless case our result reads as follows:

Theorem I.1. *Assume that the signals follow the model in* (2) *and that for $i \leq S$ the coefficients satisfy $c_{i+t}/c_i \leq 1 - \frac{\lambda}{S}$ for $t, \lambda > 0$. Then, except with probability $2SK^{1-2m}$, OMP will recover the full support as long as*

$$\left(t \left\lceil \frac{S}{\lambda} \right\rceil + \sqrt{\frac{mtS\log K}{\lambda}} + 1 \right) S\mu^2 \leq \frac{1}{20}. \tag{3}$$

The idea that the performance of OMP improves for decaying coefficients has already been used in [14] and the simplified result states that if the sorted absolute coefficients form a geometric sequence with decay $\alpha \leq \frac{1}{2}$, then OMP is guaranteed to succeed for all sparsity levels S with $S\mu < 1$.

Our result, specialised to the case $t = 1$, meaning $c_{i+1}/c_i \leq \alpha < 1$, essentially says that OMP will recover the support except with probability $2SK^{1-2m}$ as long as $S\mu^2 \lesssim 1 - \alpha$ and $S\mu^2\sqrt{m \log K} \lesssim \sqrt{1-\alpha}$. So giving up certainty for high probability allows to relax the bound in [14] by an order of magnitude.

Further comparison to BP - $S\mu^2 m \log K \lesssim 1$ for failure probability $2K^{1-2m}$ - shows that OMP has the advantage that the admissible sparsity level has a milder dependence on the dictionary size and success probability while BP has the advantage of being independent of the coefficient decay.

We will further present results for (partial) support recovery in the case of sub-Gaussian noise and support our findings with numerical simulations, Fig. 1/2.

II. Figures

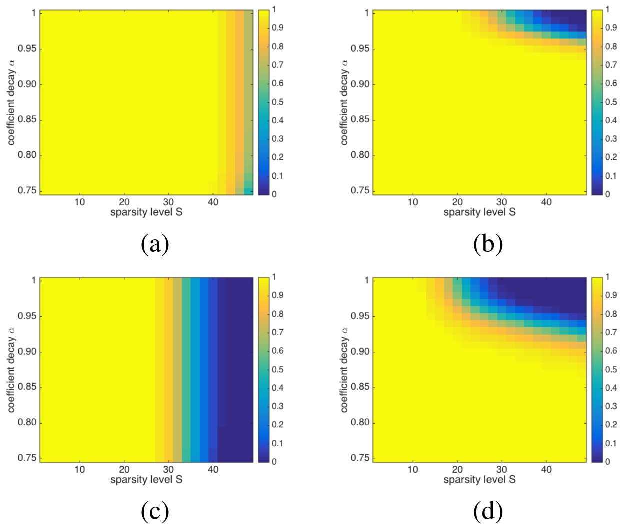

Fig. 1. Percentage of correctly recovered supports for noiseless signals with various sparsity and coefficient decay parameters via BP (a,c) and OMP (b,d) in the Dirac-DCT dictionary (a,b) and the Dirac-DCT-random dictionary (c,d).

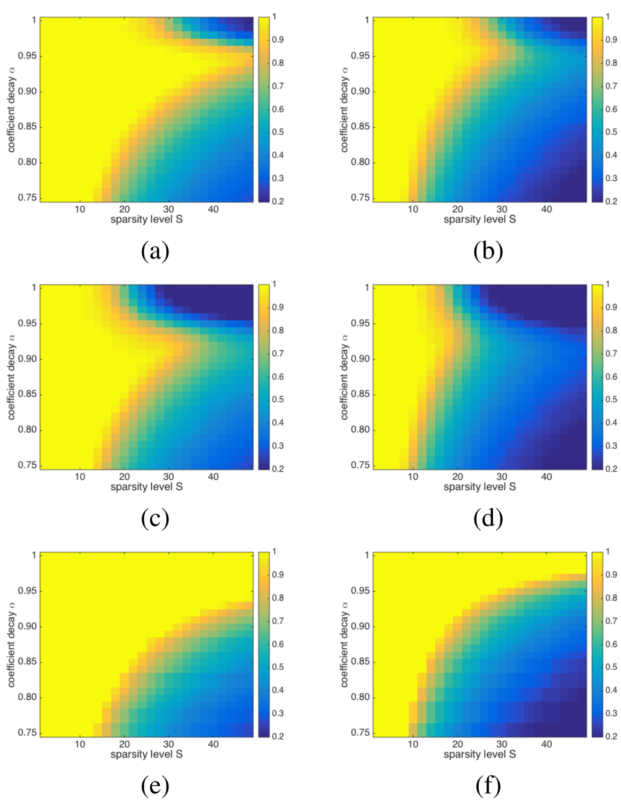

Fig. 2. Percentage of correctly recovered atoms before recovery of first wrong atom via OMP for signals with various sparsity levels and coefficient decay parameter contaminated with Gaussian noise corresponding to $\mathrm{SNR} = 256$ (a,c) and $\mathrm{SNR} = 16$ (b,d) in the Dirac-DCT dictionary (a,b) and the Dirac-DCT-random dictionary (c,d), as well as the percentage of correctly recoverable atoms for $\mathrm{SNR} = 256$ and $\mathrm{SNR} = 16$ (e,f).

Acknowledgments

This work was supported by the Austrian Science Fund (FWF) under Grant no. Y760. In addition the computational results presented have been achieved (in part) using the HPC infrastructure LEO of the University of Innsbruck.

References

[1] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal Matching Pursuit: recursive function approximation with application to wavelet decomposition," in *Asilomar Conf. on Signals Systems and Comput.*, 1993.

[2] D. Donoho and M. Elad, "Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ_1 minimization," *Proc. Nat. Aca. Sci.*,, vol. 100, no. 5, pp. 2197–2202, March 2003.

[3] J. Tropp, "Greed is good: Algorithmic results for sparse approximation." *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, October 2004.

[4] J. J. Fuchs, "Extension of the pisarenko method to sparse linear arrays," *IEEE Transactions on Signal Processing*, vol. 45, no. 2413-2421, October 1997.

[5] J. Tropp, "On the conditioning of random subdictionaries," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1-24, 2008.

[6] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, "Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 655–687, 2008.

[7] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[8] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[9] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[10] A. Gilbert and J. Tropp, "Signal recovery from random measurements via Orthogonal Matching Pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, December 2007.

[11] D. Needell and R. Vershynin, "Uniform Uncertainty Principle and signal recovery via Regularized Orthogonal Matching Pursuit," *Foundations of Computational Mathematics*, DOI: 10.1007/s10208-008-9031-3.

[12] K. Schnass, "Average performance of Orthogonal Matching Pursuit (OMP) for sparse approximation," *IEEE Signal Processing Letters*, vol. 25, no. 12, pp. 1865–1869, 2018.

[13] ——, "Average performance of Orthogonal Matching Pursuit (OMP) for sparse approximation," *arXiv:1809.06684*, 2018.

[14] C. Herzet, A. Drémeau, and C. Soussen, "Relaxed recovery conditions for OMP/OLS by exploiting both coherence and decay," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 459 – 470, 2016.