

# Sparse approximation

---

**Sparse approximation** (also known as **sparse representation**) theory deals with sparse solutions for systems of linear equations. Techniques for finding these solutions and exploiting them in applications have found wide use in image processing, signal processing, machine learning, medical imaging, and more.

## Contents

---

### Sparse decomposition

Noiseless observations

Noisy observations

### Variations

### Algorithms

### Applications

### See also

### References

## Sparse decomposition

---

### Noiseless observations

Consider a linear system of equations  $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$ , where  $\mathbf{D}$  is an underdetermined  $m \times p$  matrix ( $m < p$ ) and  $\mathbf{x} \in \mathbb{R}^m, \boldsymbol{\alpha} \in \mathbb{R}^p$ . The matrix  $\mathbf{D}$  (typically assumed to be full-rank) is referred to as the dictionary, and  $\mathbf{x}$  is a signal of interest. The core sparse representation problem is defined as the quest for the sparsest possible representation  $\boldsymbol{\alpha}$  satisfying  $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$ . Due to the underdetermined nature of  $\mathbf{D}$ , this linear system admits in general infinitely many possible solutions, and among these we seek the one with the fewest non-zeros. Put formally, we solve

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^p} \|\boldsymbol{\alpha}\|_0 \text{ subject to } \mathbf{x} = \mathbf{D}\boldsymbol{\alpha},$$

where  $\|\boldsymbol{\alpha}\|_0 = \#\{i : \alpha_i \neq 0, i = 1, \dots, p\}$  is the  $\ell_0$  pseudo-norm, which counts the number of non-zero components of  $\boldsymbol{\alpha}$ . This problem is known to be NP-hard with a reduction to NP-complete subset selection problems in combinatorial optimization.

Sparsity of  $\boldsymbol{\alpha}$  implies that only a few ( $k \ll m < p$ ) components in it are non-zero. The underlying motivation for such a sparse decomposition is the desire to provide the simplest possible explanation of  $\mathbf{x}$  as a linear combination of as few as possible columns from  $\mathbf{D}$ , also referred to as atoms. As such, the signal  $\mathbf{x}$  can be viewed as a molecule composed of a few fundamental elements taken from  $\mathbf{D}$ .

While the above posed problem is indeed NP-Hard, its solution can often be found using approximation algorithms. One such option is a convex relaxation of the problem, obtained by using the  $\ell_1$ -norm instead of  $\ell_0$ , where  $\|\boldsymbol{\alpha}\|_1$  simply sums the absolute values of the entries in  $\boldsymbol{\alpha}$ . This is known as the basis pursuit (BP)

algorithm, which can be handled using any linear programming solver. An alternative approximation method is a greedy technique, such as the matching pursuit (MP), which finds the location of the non-zeros one at a time.

Surprisingly, under mild conditions on  $D$  (using the spark (mathematics), the mutual coherence or the restricted isometry property) and the level of sparsity in the solution,  $k$ , the sparse representation problem can be shown to have a unique solution, and BP and MP are guaranteed to find it perfectly.<sup>[1] [2] [3]</sup>

## Noisy observations

Often the observed signal  $x$  is noisy. By relaxing the equality constraint and imposing an  $\ell_2$ -norm on the data-fitting term, the sparse decomposition problem becomes

$$\min_{\alpha \in \mathbb{R}^p} \|\alpha\|_0 \text{ subject to } \|x - D\alpha\|_2^2 \leq \epsilon^2,$$

or put in a Lagrangian form,

$$\min_{\alpha \in \mathbb{R}^p} \lambda \|\alpha\|_0 + \frac{1}{2} \|x - D\alpha\|_2^2,$$

where  $\lambda$  is replacing the  $\epsilon$ .

Just as in the noiseless case, these two problems are NP-Hard in general, but can be approximated using pursuit algorithms. More specifically, changing the  $\ell_0$  to an  $\ell_1$ -norm, we obtain

$$\min_{\alpha \in \mathbb{R}^p} \lambda \|\alpha\|_1 + \frac{1}{2} \|x - D\alpha\|_2^2,$$

which is known as the basis pursuit denoising. Similarly, matching pursuit can be used for approximating the solution of the above problems, finding the locations of the non-zeros one at a time until the error threshold is met. Here as well, theoretical guarantees suggest that BP and MP lead to nearly optimal solutions depending on the properties of  $D$  and the cardinality of the solution  $k$ .<sup>[4] [5] [6]</sup> Another interesting theoretical result refers to the case in which  $D$  is unitary. Under this assumption, the problems posed above (with either  $\ell_0$  or  $\ell_1$ ) admit closed-form solutions in the form of non-linear shrinkage.<sup>[4]</sup>

## Variations

---

There are several variations to the basic sparse approximation problem.

**Structured sparsity:** In the original version of the problem, any of the atoms in the dictionary can be picked. In the structured (block) sparsity model, instead of picking atoms individually, groups of them are to be picked. These groups can be overlapping and of varying size. The objective is to represent  $x$  such that it is sparse while forcing this block-structure.<sup>[7]</sup>

**Collaborative (joint) sparse coding:** The original version of the problem is defined for a single signal  $x$ . In the collaborative (joint) sparse coding model, a set of signals is available, each believed to emerge from (nearly) the same set of atoms from  $D$ . In this case, the pursuit task aims to recover a set of sparse representations that best describe the data while forcing them to share the same (or close-by) support.<sup>[8]</sup>

**Other structures:** More broadly, the sparse approximation problem can be cast while forcing a specific desired structure on the pattern of non-zero locations in  $\alpha$ . Two cases of interest that have been extensively studied are tree-based structure, and more generally, a Boltzmann distributed support.<sup>[9]</sup>

## Algorithms

---

As already mentioned above, there are various approximation (also referred to as **pursuit**) algorithms that have been developed for addressing the sparse representation problem:

$$\min_{\alpha \in \mathbb{R}^p} \|\alpha\|_0 \text{ subject to } \|x - D\alpha\|_2^2 \leq \epsilon^2.$$

We mention below a few of these main methods.

- Matching pursuit is a greedy iterative algorithm for approximately solving the above problem. It works by gradually finding the locations of the non-zeros in  $\alpha$  one at a time. The core idea is to find in each step the column (atom) in  $D$  that best correlates with the current residual (initialized to  $x$ ), and then updating this residual to take the new atom and its coefficient into account. Matching pursuit might pick the same atom multiple times.
- Orthogonal matching pursuit is very similar to matching pursuit, with one major difference: in each of the algorithm's step, all the non-zero coefficients are updated by a least squares. As a consequence, the residual is orthogonal to the already chosen atoms, and thus an atom cannot be picked more than once.
- Stage-wise greedy methods: Improved variations over the above are algorithms that operate greedily while adding two critical features: (i) the ability to add groups of non-zeros at a time (instead of one non-zero per round); and (ii) including a pruning step in each round in which several of the atoms are discarded from the support. Representatives of this approach are the Subspace-Pursuit algorithm and the CoSaMP.<sup>[10]</sup>
- Basis pursuit solves a convex relaxed version of the problem by replacing the  $\ell_0$  by an  $\ell_1$ -norm. Note that this only defines a new objective, while leaving open the question of the algorithm to use for getting the desired solution. Commonly considered such algorithms are the IRLS, LARS, and iterative soft-shrinkage methods.<sup>[11]</sup>
- There are several other methods for solving sparse decomposition problems: homotopy method, coordinate descent, iterative hard-thresholding, first order proximal methods, which are related to the above-mentioned iterative soft-shrinkage algorithms, and Dantzig selector.

## Applications

---

Sparse approximation ideas and algorithms have been extensively used in signal processing, image processing, machine learning, medical imaging, array processing, data mining, and more. In most of these applications, the unknown signal of interest is modeled as a sparse combination of a few atoms from a given dictionary, and this is used as the regularization of the problem. These problems are typically accompanied by a dictionary learning mechanism that aims to fit  $D$  to best match the model to the given data. The use of sparsity-inspired models has led to state-of-the-art results in a wide set of applications.<sup>[12][13][14]</sup> Recent work suggests that there is a tight connection between sparse representation modeling and deep-learning.<sup>[15]</sup>

## See also

---

- Compressed sensing
- Sparse dictionary learning

- K-SVD
- Lasso (statistics)
- Regularization (mathematics) and inverse problems

## References

---

1. Donoho, D.L. and Elad, M. (2003). "Optimally sparse representation in general (nonorthogonal) dictionaries via L1 minimization" ([https://www.cs.technion.ac.il/~elad/publications/journals/2003/15\\_New\\_Sparseness\\_IT.pdf](https://www.cs.technion.ac.il/~elad/publications/journals/2003/15_New_Sparseness_IT.pdf)) (PDF). *Proceedings of the National Academy of Sciences*. **100** (5): 2197–2202. Bibcode:2003PNAS..100.2197D (<https://ui.adsabs.harvard.edu/abs/2003PNA S..100.2197D>). doi:10.1073/pnas.0437847100 (<https://doi.org/10.1073%2Fpnas.0437847100>). PMC 153464 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC153464>). PMID 16576749 (<https://pubmed.ncbi.nlm.nih.gov/16576749>).
2. Tropp, J.A. (2004). "Greed is good: Algorithmic results for sparse approximation" (<https://authors.library.caltech.edu/9035/1/TROieetit04a.pdf>) (PDF). *IEEE Transactions on Information Theory*. **50** (10): 2231–2242. CiteSeerX 10.1.1.321.1443 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.321.1443>). doi:10.1109/TIT.2004.834793 (<https://doi.org/10.1109%2FTIT.2004.834793>).
3. Donoho, D.L. (2006). "For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution" (<http://www-stat.stanford.edu/~donoho/Reports/2004/l1l0approx.pdf>) (PDF). *Communications on Pure and Applied Mathematics*. **56** (6): 797–829. doi:10.1002/cpa.20132 (<https://doi.org/10.1002%2Fcpa.20132>).
4. Elad, M. (2010). *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer. CiteSeerX 10.1.1.331.8963 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.331.8963>). doi:10.1007/978-1-4419-7011-4 (<https://doi.org/10.1007%2F978-1-4419-7011-4>). ISBN 978-1441970107.
5. Donoho, D.L., Elad, M. and Templyakov, V. (2006). "Stable recovery of sparse overcomplete representations in the presence of noise" ([https://www.cs.technion.ac.il/~elad/publications/journals/2004/23\\_Stability\\_IEEE\\_TIT.pdf](https://www.cs.technion.ac.il/~elad/publications/journals/2004/23_Stability_IEEE_TIT.pdf)) (PDF). *IEEE Transactions on Information Theory*. **52** (1): 6–18. CiteSeerX 10.1.1.125.5610 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.125.5610>). doi:10.1109/TIT.2005.860430 (<https://doi.org/10.1109%2FTIT.2005.860430>).
6. Tropp, J.A. (2006). "Just relax: Convex programming methods for identifying sparse signals in noise" (<https://authors.library.caltech.edu/9040/1/TROieetit06.pdf>) (PDF). *IEEE Transactions on Information Theory*. **52** (3): 1030–1051. CiteSeerX 10.1.1.184.2957 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.184.2957>). doi:10.1109/TIT.2005.864420 (<https://doi.org/10.1109%2FTIT.2005.864420>).
7. Eldar, Y.C., Kuppinger, P. and Bolcskei, H. (2009). "Block-sparse signals: Uncertainty relations and efficient recovery". *IEEE Transactions on Signal Processing*. **58** (6): 3042–3054. arXiv:0906.3173 (<https://arxiv.org/abs/0906.3173>). Bibcode:2010ITSP...58.3042E (<https://ui.adsabs.harvard.edu/abs/2010ITSP...58.3042E>). doi:10.1109/TSP.2010.2044837 (<https://doi.org/10.1109%2FTSP.2010.2044837>).
8. Tropp, J.A., Gilbert, A.C. and Strauss, M.J. (2006). "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit". *Signal Processing*. **86** (3): 572–588. doi:10.1016/j.sigpro.2005.05.030 (<https://doi.org/10.1016%2Fj.sigpro.2005.05.030>).
9. Peleg, T., Eldar, Y.C. and Elad, M. (2012). "Exploiting Statistical Dependencies in Sparse Representations for Signal Recovery". *IEEE Transactions on Signal Processing*. **60** (5): 2286–2303. arXiv:1010.5734 (<https://arxiv.org/abs/1010.5734>). Bibcode:2012ITSP...60.2286P (<https://ui.adsabs.harvard.edu/abs/2012ITSP...60.2286P>). doi:10.1109/TSP.2012.2188520 (<https://doi.org/10.1109%2FTSP.2012.2188520>).

10. Needell, D. and Tropp, J.A. (2009). "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples". *Applied and Computational Harmonic Analysis*. **26** (3): 301–321. arXiv:0803.2392 (<https://arxiv.org/abs/0803.2392>). doi:10.1016/j.acha.2008.07.002 (<https://doi.org/10.1016%2Fj.acha.2008.07.002>).
11. Zibulevsky, M. and Elad, M. (2010). "L1-L2 optimization in signal and image processing" ([http://www.cs.technion.ac.il/~elad/publications/journals/2010/L1L2\\_IEEE\\_SPM\\_2010.pdf](http://www.cs.technion.ac.il/~elad/publications/journals/2010/L1L2_IEEE_SPM_2010.pdf)) (PDF). *IEEE Signal Processing Magazine*. **27** (3): 76–88. Bibcode:2010ISPM...27...76Z (<https://ui.adsabs.harvard.edu/abs/2010ISPM...27...76Z>). doi:10.1109/MSP.2010.936023 (<https://doi.org/10.1109%2FMSP.2010.936023>).
12. Baraniuk, R.G. Candes, E. Elad, M. and Ma, Y. (2010). "Applications of sparse representation and compressive sensing". *Proceedings of the IEEE*. **98** (6): 906–909. doi:10.1109/JPROC.2010.2047424 (<https://doi.org/10.1109%2FJPROC.2010.2047424>).
13. Elad, M. Figueiredo, M.A.T., and Ma, Y. (2010). "On the role of sparse and redundant representations in image processing" (<https://pdfs.semanticscholar.org/5393/46f7f35f83875d9e86851010ee35d1a47e69.pdf>) (PDF). *Proceedings of the IEEE*. **98** (6): 972–982. CiteSeerX 10.1.1.160.465 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.160.465>). doi:10.1109/JPROC.2009.2037655 (<https://doi.org/10.1109%2FJPROC.2009.2037655>).
14. Plumbley, M.D. Blumensath, T. Daudet, L. Gribonval, R. and Davies, M.E. (2010). "Sparse representations in audio and music: From coding to source separation". *Proceedings of the IEEE*. **98** (6): 995–1005. CiteSeerX 10.1.1.160.1607 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.160.1607>). doi:10.1109/JPROC.2009.2030345 (<https://doi.org/10.1109%2FJPROC.2009.2030345>).
15. Pappayan, V. Romano, Y. and Elad, M. (2017). "Convolutional Neural Networks Analyzed via Convolutional Sparse Coding" (<http://jmlr.org/papers/volume18/16-505/16-505.pdf>) (PDF). *Journal of Machine Learning Research*. **18** (83): 1–52. arXiv:1607.08194 (<https://arxiv.org/abs/1607.08194>). Bibcode:2016arXiv160708194P (<https://ui.adsabs.harvard.edu/abs/2016arXiv160708194P>).

---

Retrieved from "[https://en.wikipedia.org/w/index.php?title=Sparse\\_approximation&oldid=1013750533](https://en.wikipedia.org/w/index.php?title=Sparse_approximation&oldid=1013750533)"

---

This page was last edited on 23 March 2021, at 07:55 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.