# A Matrix Factorization Approach for Learning Semidefinite-Representable Regularizers

Yong Sheng Soh, Venkat Chandrasekaran

Regularization techniques are widely employed in the solution of inverse problems in data analysis and scientific computing due to their effectiveness in addressing difficulties due to ill-posedness. Frequently, these methods take the form of penalty functions added to the objective in optimization-based approaches for solving inverse problems. The purpose of the penalty function is to induce a desired structure in the solution, and these functions are specified based on prior domain-specific expertise. In this work, we study the question of *learning* suitable regularization functions from data in settings in which precise domain knowledge is not directly available. The regularizers obtained using our framework are specified as convex functions that can be computed efficiently via semidefinite programming, and therefore they can be employed in tractable convex optimization approaches for solving inverse problems.

We begin by highlighting the geometric aspects of regularizers that make them effective in promoting a desired structure. We focus on a family of convex regularizers that are useful for inducing a general form of sparsity in solutions to inverse problems. Concretely, suppose $\mathcal{A} \subset \mathbb{R}^d$ is a (possibly infinite) collection of elementary building blocks or atoms. Then $\mathbf{y} \in \mathbb{R}^d$ is said to have a sparse representation using the atomic set $\mathcal{A}$ if $\mathbf{y}$ can be expressed as follows:

$$\mathbf{y} = \sum_{i=1}^{k} c_i \mathbf{a}_i, \quad \mathbf{a}_i \in \mathcal{A}, c_i \geq 0,$$

for a relatively small number $k$. An important virtue of sparse descriptions based on $\mathcal{A}$ is that employing the *atomic norm* induced by $\mathcal{A}$ — the gauge function of the atomic set $\mathcal{A}$ — is effective at promoting sparse descriptions in solutions with respect to $\mathcal{A}$ [2]. The reason is that the low-dimensional faces of the convex hull of $\mathcal{A}$ contain points that have a sparse description using $\mathcal{A}$. Indeed, in many contemporary data analysis applications the solutions of regularized optimization problems with generic input data tend to lie on low-dimensional faces of sublevel sets of the regularizer [3]–[5].

The difficulty with employing an atomic norm regularizer in practice is that one requires prior domain knowledge of the atomic set $\mathcal{A}$. While such information may be available based on domain expertise in some problems, identifying a suitable atomic set is challenging for many contemporary datasets that are high-dimensional and are presented to an analyst in an unstructured fashion. We study the question of learning a suitable regularizer directly from observations $\{\mathbf{y}^{(j)}\}_{j=1}^{n} \subset \mathbb{R}^d$. Specifically, we want to obtain an atomic set $\mathcal{A}$ such that each $\mathbf{y}^{(j)}$ has a sparse representation using $\mathcal{A}$; the corresponding regularizer is simply the atomic norm induced by $\mathcal{A}$.

### A. Relating Dictionary Learning to Polyhedral Regularizers

The problem of learning a suitable polyhedral regularizer from data points $\{\mathbf{y}^{(j)}\}_{j=1}^{n}$ corresponds to identifying an appropriate *finite* atomic set to concisely describe each $\mathbf{y}^{(j)}$. This problem is equivalent to the question of 'dictionary learning' (also called 'sparse coding') [6]. To see this connection, suppose that we parametrize a finite atomic set via a matrix $L \in \mathbb{R}^{d \times p}$ so that the columns of $L$ and their negations specify the atoms. The associated atomic norm ball is the image under $L$ of the $\ell_1$ ball in $\mathbb{R}^p$. With this parametrization, learning a polyhedral regularizer may be viewed as obtaining a matrix $L$ such that each $\mathbf{y}^{(j)}$ is well-approximated as $L\mathbf{x}^{(j)}$ for a vector $\mathbf{x}^{(j)} \in \mathbb{R}^p$ with few nonzero coordinates. Computing such a representation of the data is precisely the objective in dictionary learning.

### B. From Polyhedral to Semidefinite Regularizers

The objective of this work is to investigate the problem of learning more general non-polyhedral atomic norm regularizers; in other words, the set of atoms may be *infinite*. On the approximation-theoretic front, infinite atomic sets offer the possibility of concise descriptions of data sets with much richer types of structure than those using finite atomic sets. Formally, we consider atomic sets in $\mathbb{R}^d$ that are images of rank-one matrices:

$$\mathcal{A}(\mathcal{L}) = \left\{ \mathcal{L}(\mathbf{u}\mathbf{v}') \mid \mathbf{u}, \mathbf{v} \in \mathbb{R}^q, \ \|\mathbf{u}\|_{\ell_2} = 1, \|\mathbf{v}\|_{\ell_2} = 1 \right\},$$

where $\mathcal{L} : \mathbb{R}^{q \times q} \to \mathbb{R}^d$ specifies a linear map. The corresponding atomic norm ball is given by:

$$\operatorname{conv}\left(\mathcal{A}(\mathcal{L})\right) = \left\{ \mathcal{L}(X) \mid X \in \mathbb{R}^{q \times q}, \ \|X\|_\star \leq 1 \right\},$$

where $\|X\|_\star := \sum_i \sigma_i(X)$. As the nuclear norm ball has a tractable semidefinite description [5], [7], the atomic norm induced by $\mathcal{A}(\mathcal{L})$ can be computed efficiently using semidefinite programming. The problem of learning a semidefinite-representable regularizer may be phrased as one of *matrix factorization* whereby our objective is to obtain a linear map $\mathcal{L}$ such that each $\mathbf{y}^{(j)}$ is well-approximated as $\mathcal{L}(X^{(j)})$ for a low-rank $X^{(j)} \in \mathbb{R}^{q \times q}$.

We develop an alternating update algorithm to compute such a factorization. Our approach is a generalization of methods that are widely employed in dictionary learning. With $\mathcal{L}$ fixed, updating the $X^{(j)}$'s entails the solution of affine rank minimization problems. Although this problem is intractable in general [8], tractable heuristics have been developed and proven to succeed under suitable conditions [5], [9], [10]. With the $X^{(j)}$'s fixed, $\mathcal{L}$ is updated by solving a least-squares problem. To address non-uniqueness issues that arises in our matrix factorization instance, we apply the Operator Sinkhorn iterative procedure to put the map $\mathcal{L}$ in a canonical form. Operator Sinkhorn iteration was developed by Gurvits to solve certain quantum matching problems [11], and can be viewed as an operator analog of the diagonal congruence scaling technique for nonnegative matrices developed by Sinkhorn [12]. Under suitable conditions on the input data, our algorithm provides a locally linearly convergent method for identifying the correct regularizer that promotes the type of structure contained in the data. Our analysis is based on the stability properties of Operator Sinkhorn scaling and their relation to geometric aspects of determinantal varieties. We demonstrate the utility of our framework with a series of experimental results on synthetic as well as real data.

The extended version of this work is available as an electronic preprint [1].

## REFERENCES


[1] Y. S. Soh and V. Chandrasekaran, "A Matrix Factorization Approach for Learning Semidefinite-Representable Regularizers," *CoRR*, vol. abs/1701.01207, 2017.

[2] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The Convex Geometry of Linear Inverse Problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[3] E. J. Candès and T. Tao, "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, 2006.

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

[5] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[6] B. A. Olshausen and D. J. Field, "Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images," *Nature*, vol. 381, 1996.

[7] M. Fazel, "Matrix Rank Minimization with Applications," Ph.D. dissertation, Department of Electrical Engineering, Stanford University, 2002.

[8] B. K. Natarajan, "Sparse Approximate Solutions to Linear Systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1993.

[9] D. Goldfarb and S. Ma, "Convergence of Fixed-Point Continuation Algorithms for Matrix Rank Minimization," *Foundations of Computational Mathematics*, vol. 11, pp. 183 – 210, 2011.

[10] P. Jain, R. Meka, and I. S. Dhillon, "Guaranteed Rank Minimization via Singular Value Projection," in *Advances in Neural Information Processing Systems 23*, vol. abs/0909.5457, 2009.

[11] L. Gurvits, "Classical Complexity and Quantum Entanglement," *Journal of Computer and Systems Sciences*, vol. 69, no. 3, pp. 448 – 484, 2004.

[12] R. Sinkhorn, "A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices," *The Annals of Mathematical Statistics*, vol. 35, no. 2, pp. 876–879, 1964.