Atomic Decomposition via Polar Alignment

The Geometry of Structured Optimization

Other titles in Foundations and Trends® in Optimization

The Many Faces of Degeneracy in Conic Optimization
Dmitriy Drusvyatskiy and Henry Wolkowicz
ISBN: 978-1-68083-390-4

Multi-Period Trading via Convex Optimization
Stephen Boyd, Enzo Busseti, Steve Diamond, Ronald N. Kahn, Kwangmoo Koh, Peter Nystrup and Jan Speth
ISBN: 978-1-68083-298-3

Introduction to Online Convex Optimization
Elad Hazan
ISBN: 978-1-68083-170-2

Low-Rank Semidefinite Programming: Theory and Applications
Alex Lemon, Anthony Man-Cho So and Yinyu Ye
ISBN: 978-1-68083-136-8

Atomic Decomposition via Polar Alignment

The Geometry of Structured Optimization

Zhenan Fan
University of British Columbia
Canada
zhenanf@cs.ubc.ca

Halyun Jeong
University of British Columbia
Canada
clatar1@gmail.com

Yifan Sun
Stony Brook University
USA
yifan.sun@stonybrook.edu

Michael P. Friedlander
University of British Columbia
Canada
michael.friedlander@ubc.ca

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Optimization

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

Z. Fan, H. Jeong, Y. Sun and M. P. Friedlander. *Atomic Decomposition via Polar Alignment*. Foundations and Trends® in Optimization, vol. 3, no. 4, pp. 280–366, 2020.

ISBN: 978-1-68083-743-8

Foundations and Trends® in Optimization
Volume 3, Issue 4, 2020

Editorial Scope

Topics

Foundations and Trends® in Optimization publishes survey and tutorial articles in the following topics:

- algorithm design, analysis, and implementation (especially, on modern computing platforms
- models and modeling systems, new optimization formulations for practical problems
- applications of optimization in machine learning, statistics, and data analysis, signal and image processing, computational economics and finance, engineering design, scheduling and resource allocation, and other areas

Information for Librarians

Foundations and Trends® in Optimization, 2020, Volume 3, 4 issues. ISSN paper version 2167-3888. ISSN online version 2167-3918. Also available as a combined paper and online subscription.

Contents

Atomic Decomposition via Polar Alignment

Zhenan Fan[1], Halyun Jeong[2], Yifan Sun[3] and Michael P. Friedlander[4]

[1] *University of British Columbia, Canada; zhenanf@cs.ubc.ca*
[2] *University of British Columbia, Canada; clatar1@gmail.com*
[3] *Stony Brook University, USA; yifan.sun@stonybrook.edu*
[4] *University of British Columbia, Canada; michael.friedlander@ubc.ca*

ABSTRACT

Structured optimization uses a prescribed set of atoms to assemble a solution that fits a model to data. Polarity, which extends the familiar notion of orthogonality from linear sets to general convex sets, plays a special role in a simple and geometric form of convex duality. This duality correspondence yields a general notion of alignment that leads to an intuitive and complete description of how atoms participate in the final decomposition of the solution. The resulting geometric perspective leads to variations of existing algorithms effective for large-scale problems. We illustrate these ideas with many examples, including applications in matrix completion and morphological component analysis for the separation of mixtures of signals.

Zhenan Fan, Halyun Jeong, Yifan Sun and Michael P. Friedlander (2020), "Atomic Decomposition via Polar Alignment", Foundations and Trends® in Optimization: Vol. 3, No. 4, pp 280–366. DOI: 10.1561/2400000028.

1

Introduction

Convex optimization provides a valuable computational framework that renders many problems tractable because of the range of powerful algorithms that can be brought to the task. The key is that a certain mathematical structure—i.e., convexity of the functions and sets defining the problem—lays open an enormous range of theoretical and algorithmic tools that lend themselves astonishingly well to computation. There are limits, however, to the scalability of general-purpose algorithms for convex optimization. As has been recognized in the optimization and related communities for at least the past decade, significant efficiencies can be gained by acknowledging the latent structure in the solution itself, coupled with the overarching structure provided by convexity.

Structured optimization proceeds along these lines by using a prescribed set of atoms from which to assemble an optimal solution. In effect, the atoms selected to participate in forming a solution decompose the model into simpler parts, which offers opportunities for algorithmic efficiency in solving the optimization problem. From a modeling point of view, the particular atoms that constitute the computed solution often represent key explanatory components of a model. An atomic decomposition thus provides a description of the most informative features of a solution—in other words, a kind of generalized principal component analysis.

Our purpose with this monograph is to describe the rich convex geometry that underlies atomic decomposition. The path we follow builds on the duality inherent in convex cones: every convex cone is paired uniquely with another cone that is polar to it. The extreme rays of each cone in this pair are in some sense *aligned*. Brought into the context of atomic decomposition, this notion of alignment through the polar operation provides a theoretical framework that can be harnessed to identify the atoms that participate in a decomposition. This approach facilitates certain algorithmic design patterns that promote computational efficiency, as we demonstrate with concrete examples. Similar computational economies accrue within reduced-space active-set methods for optimization problems with inequality constraints, such as implemented by the MINOS software package [1].

Early work in structured optimization focused on problem formulations meant to produce sparse solution vectors, i.e., a solution with relatively few non-zero elements. Compressed sensing [2]–[4] and model selection [5], [6], with their many applications in signal processing and statistics, helped to establish sparse optimization as an important class of problems with a range of specialized algorithms. Generalizations that accommodated different notions of sparsity soon followed, including matrix problems with low-rank solutions (sparsity in the vector of singular values), fused index pairs (sparsity in terms of the norms of subgroups of variables), and sparsity in specialized dictionaries, such as mass spectrographs of simple molecules used to represent structures of more complicated molecules [7, Section 6.3.1].

Nonsmooth regularization functions that promote sparsity, such as the 1-norm for sparse vectors, or the nuclear norm for low-rank matrices, are key features of these formulations. Gauge functions, which significantly generalize the notion of a norm, were recognized as flexible regularization functions that promote a broad range of sparse structures. By defining a set of atoms from which to build a solution, an almost arbitrary set of solution structures can be considered. The gauge function to this set can be incorporated into a convex optimization problem in order to obtain a solution with the desired structure. The convex analysis of gauges and support functions, which are their dual counterparts, is rich in geometry and rife with opportunity for efficient algorithm implementations for high-dimensional problems. Our purpose with this monograph is to expose the basic elements of this theory and its many connections to sparse and structured optimization. To make it accessible to researchers who are not specialists in convex analysis, we chose a largely self-contained treatment and make a few modest assumptions that greatly simplify the derivations.

1.1 Applications and Prior Work

One of the main implications of our approach is its usefulness in adapting dual optimization methods for discovering atomic decompositions. With the tools of polar alignment, a dual optimization method can be interpreted as solving for an aligning dual vector z that exposes the support of a primal solution x. If the number of exposed atoms is small, a solution x of the primal problem can be obtained from a reduced problem defined over the exposed support, but without the nonsmooth atomic regularization. The resulting reduced problem is often computationally much cheaper [8] and better conditioned [9]. Alternatively, two-metric methods can be designed to act differently on a primal iterate's suspected support [10]. In many applications, such as feature selection, knowing the optimal support may itself be sufficient. As we illustrate through various examples, there are several important cases where the dual aligning vector z can be computed directly.

Machine Learning. The regularized optimization problems described in Section 5 frequently appear in applications of machine learning for the purpose of model complexity reduction. The most popular tools are the vector 1-norm in feature selection [5], its group-norm variant [11], and the nuclear norm in matrix completion [12]. Many other sparsity-promoting regularizers, however, appear in practice [13]. Although unconstrained formulations are most popular, particularly when the proximal operator is computationally convenient [14], the gauge-constrained formulation is frequently used and solved via the conditional gradient method [15]–[17]. Popular dual methods, which iterate over a dual variable $z^{(k)}$ but maintain the corresponding primal variable $x^{(k)}$ only implicitly, include bundle methods [18] and dual averaging [19], [20].

Linear Conic Optimization. Conic programs are a cornerstone of convex optimization. The nonnegative cone, the second-order cone and the semidefinite cone respectively, give rise to linear, second-order, and semidefinite programs. These problem classes capture an enormous range of important models, and can be solved efficiently by a variety of algorithms, including interior methods [21]–[23]. Conic programs and their associated solvers are key ingredients for general purpose optimization software packages such as YALMIP [24] and CVX [25]. The alignment conditions for these specific cones have been exploited in dual methods, such as in the spectral bundle method for large-scale semidefinite programming [26]. Example 3.6 demonstrates this alignment principle in the context of conic optimization.

Gauge Optimization. The class of gauge optimization problems, as defined by Freund's 1987 seminal work [27], can be simply stated: find the element of a convex set that is minimal with respect to a gauge function. These conceptually simple problems appear in a remarkable array of applications, and include parts of sparse optimization and all of conic optimization [28, Example 1.3]. This class of optimization problems admits a duality relationship different from classical Lagrange duality, and is founded on the polar inequality. In this context, the polar inequality provides an analogue to weak duality, well-known in Lagrange duality, which guarantees that any feasible primal value provides an upper bound for any feasible dual value. In the gauge optimization context, a primal-dual pair (x, z) is optimal if and only if the polar inequality holds as an equation, which under Definition 2.4 implies that x and z are aligned. The connection between polar alignment and optimality is discussed further in Subsection 5.2.

Two-Stage Methods. In sparse optimization, two-stage methods first identify the primal variable support, and then solve the problem over a reduced support [29], [30]. If the support is sparse enough, the second problem may be computationally much cheaper because it can allow for faster Newton-like methods. The atomic alignment principles we describe in Section 4 give a general recipe for extracting primal variable support from a computed dual variable, which at optimality is aligned with the primal variable; see Section 5. This property forms the basis for our approach to morphological component analysis, described in Subsection 7.4.

Method Interpretability. The connection between sparsity and alignment points to a likely "aligning behavior" in many of the most effective methods for sparse optimization [31]. Indeed, we show in Section 6 that this is true for a range of methods, including proximal gradient, conditional gradient, and cutting-plane methods. Surprisingly, we also find hints of aligning behavior in seemingly unrelated methods, such as augmented Lagrangian and bundle methods. The alignment point of view thus offers greater interpretability of commonly used methods in many modern optimization applications.

1.2 Basic Definitions and Notation

We work with n-vectors in $\mathbb{R}^n$ and p-by-n matrices in $\mathbb{R}^{p\times n}$. The restriction to real-valued vectors and matrices considerably simplifies our development, though many of the ideas set forth in this monograph extend to more general functional spaces, as described by Zălinescu [32] and Bauschke and Combettes [33].

Vectors are always denoted by lower-case letters; matrices by capital letters. A vector norm $\|x\|$ always refers to the 2-norm, unless otherwise specified. Matrix norms always refer to the Schatten norm, e.g., if $(s_1, s_2, \ldots)$ are the singular values of X, then

$$\|X\|_1 = \sum_i s_i, \quad \|X\|_2 = \Big(\sum_i s_i^2\Big)^{1/2}, \quad \text{and} \quad \|X\|_\infty = \max_i s_i.$$

Let e_i denote the ith canonical unit vector, i.e., the vector of all zeros except a single 1 in the ith position. The dot product of two n-vectors x and z is $\langle x, z\rangle = \sum_j x_j z_j$. The dot product of two p-by-n matrices

X and Z is the trace inner product $\langle X, Z\rangle = \mathrm{tr}(X^T Z) = \sum_{ij} X_{ij} Z_{ij}$. The adjoint F^* of any linear map F is the unique linear map that satisfies the relationship $\langle Fx, z\rangle = \langle x, F^* z\rangle$ for all x and z. Thus, for the linear map $F\colon \mathbb{R}^n \to \mathbb{R}^m$, the product of the adjoint and an m-vector y is $F^* y = \sum_{i=1}^m y_i(Fe_i)$. For the linear map $\mathcal{F}\colon \mathbb{R}^{p\times n} \to \mathbb{R}^m$, the forward and adjoint maps take the form

$$\mathcal{F}X = \begin{bmatrix} \langle F_1, X\rangle \\ \vdots \\ \langle F_m, X\rangle \end{bmatrix} \quad \text{and} \quad \mathcal{F}^* y = \sum_{i=1}^m y_i F_i, \tag{1.1}$$

where each $F_1, \dots, F_m$ is a p-by-n matrix. The notation $X \succeq 0$ indicates that X is symmetric positive definite.

Throughout the monograph, we use the symbol $\mathcal{C}$ to denote a convex set in $\mathbb{R}^n$. The convex hull of any set $\mathcal{D}$ in $\mathbb{R}^n$ contains all weighted averages of the elements of the set, denoted

$$\operatorname{conv} \mathcal{D} = \left\{ \sum_{i=1}^m \alpha_i x_i \;\middle|\; x_i \in \mathcal{D},\ \alpha_i \geq 0,\ \sum_{i=1}^m \alpha_i = 1 \right\},$$

for some positive integer m. Define the conic extension of $\mathcal{D}$ by

$$\operatorname{cone} \mathcal{D} = \{\alpha d \mid d \in \mathcal{D},\ \alpha \geq 0\}.$$

The closure, boundary and relative interior, respectively, of $\mathcal{D}$ denoted $\operatorname{cl} \mathcal{D}$, $\operatorname{bnd} \mathcal{D}$ and $\operatorname{ri} \mathcal{D}$. The indicator to $\mathcal{D}$ is the function

$$\delta_{\mathcal{D}}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{D}; \\ +\infty & \text{otherwise.} \end{cases}$$

The normal cone to the set $\mathcal{C}$ at $x \in \mathcal{C}$ is defined as

$$\mathcal{N}_{\mathcal{C}}(x) = \{d \mid \langle d, u - x\rangle \leq 0 \text{ for all } u \in \mathcal{C}\}.$$

The Euclidean projection onto the set $\mathcal{C}$ is denoted

$$\operatorname{proj}_{\mathcal{C}}(x) = \operatorname*{arg\,min}_{u \in \mathcal{C}} \|x - u\|_2,$$

which defines the distance of a point to the set $\mathcal{C}$, denoted by

$$\operatorname{dist}_{\mathcal{C}}(x) = \|x - \operatorname{proj}_{\mathcal{C}}(x)\|_2.$$

Let $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be any function. The domain is denoted $\operatorname{dom} f = \{x \mid f(x) < +\infty\}$, and the convex conjugate is denoted

$$f^*(z) = \sup_{x \in \mathbb{R}^n} \{\langle x, z\rangle - f(x)\}\,.$$

References

[1] B. A. Murtagh and M. A. Saunders, *MINOS 5.5 User's Guide.* Systems Optimization Laboratory. Tech. Rep. 83-20R, Department of Management Science and Engineering, Stanford University, Stanford, CA, 1983.

[2] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[3] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.

[4] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[5] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Stat. Soc. Ser. B*, pp. 267–288, 1996.

[6] R. Tibshirani, "The lasso method for variable selection in the Cox model," *Stat. Med.*, vol. 16, no. 4, pp. 385–395, 1997.

[7] L. Vandenberghe, "The CVXOPT linear and quadratic cone program solvers", 2010. [Online]. Available: http://www.seas.ucla.edu/~vandenbe/publications/coneprog.pdf.

[8] R. M. Freund, P. Grigas, and R. Mazumder, "An extended Frank–Wolfe method with 'in-face' directions, and its application to low-rank matrix completion," *SIAM J. Optim.*, vol. 27, no. 1, pp. 319–346, 2017.

[9] S. Negahban and M. J. Wainwright, "Restricted strong convexity and weighted matrix completion: Optimal bounds with noise," *J. Mach. Learn. Res.*, vol. 13, pp. 1665–1697, May 2012.

[10] E. M. Gafni and D. P. Bertsekas, "Two-metric projection methods for constrained optimization," *SIAM J. Control Optim.*, vol. 22, no. 6, pp. 936–964, 1984.

[11] L. Jacob, G. Obozinski, and J.-P. Vert, "Group lasso with overlap and graph lasso," in *Inter. Conf. Mach. Learning (ICML 2009)*, ACM, pp. 433–440, 2009.

[12] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.

[13] X. Zeng and M. A. Figueiredo, "The ordered weighted ℓ_1 norm: Atomic formulation, projections, and algorithms," *arXiv:1409.4271*, 2014.

[14] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 123–231, 2013.

[15] M. Frank and P. Wolfe, "An algorithm for quadratic programming," *Naval Research Logistics (NRL)*, vol. 3, no. 1–2, pp. 95–110, 1956.

[16] J. C. Dunn and S. Harshbarger, "Conditional gradient algorithms with open loop step size rules," *J. Math. Anal. Appl.*, vol. 62, no. 2, pp. 432–444, 1978.

[17] M. Jaggi, "Revisiting Frank–Wolfe: Projection-free sparse convex optimization," in *Inter. Conf. Mach. Learning (ICML 2013)*, pp. 427–435, 2013.

[18] C. Lemarechal, J.-J. Strodiot, and A. Bihain, "On a bundle algorithm for nonsmooth optimization," in *Nonlinear Programming 4*, Elsevier, 1981, pp. 245–282.

[19] L. Xiao, "Dual averaging methods for regularized stochastic learning and online optimization," *J. Mach. Learn. Res.*, vol. 11, pp. 2543–2596, Oct. 2010.

[20] J. C. Duchi, A. Agarwal, and M. J. Wainwright, "Dual averaging for distributed optimization: Convergence analysis and network scaling," *IEEE Trans. Automat. Control*, vol. 57, no. 3, pp. 592–606, 2012.

[21] N. Karmarkar, "A new polynomial-time algorithm for linear programming," in *Proceedings of the Sixteenth Annual ACM Symposium on Theory of Computing*, ACM, pp. 302–311, 1984.

[22] Y. E. Nesterov and A. Nemirovski, *Interior-Point Polynomial Algorithms in Convex Programming*, vol. 13, ser. Stud. Appl. Math. Philadelphia: Society of Industrial and Applied Mathematics, 1994.

[23] J. Renegar, *A Mathematical View of Interior-Point Methods in Convex Optimization*, ser. MPS/SIAM Series on Optimization. Philadelphia: Society of Industrial and Applied Mathematics, 2001.

[24] J. Lofberg, "Yalmip: A toolbox for modeling and optimization in matlab," in *2004 IEEE International Conference on Robotics and Automation*, IEEE, pp. 284–289, 2004.

[25] M. Grant and S. Boyd, *CVX: Matlab Software for Disciplined Convex Programming (Web Page and Software)*, http://cvxr.com, 2009.

[26] C. Helmberg and F. Rendl, "A spectral bundle method for semidefinite programming," *SIAM J. Optim.*, vol. 10, no. 3, pp. 673–696, 2000.

[27] R. M. Freund, "Dual gauge programs, with applications to quadratic programming and the minimum-norm problem," *Math. Program.*, vol. 38, no. 1, pp. 47–67, 1987.

[28] M. P. Friedlander, I. Macêdo, and T. K. Pong, "Gauge optimization and duality," *SIAM J. Optim.*, vol. 24, no. 4, pp. 1999–2022, 2014.

[29] M. Kocvara and J. Zowe, "An iterative two-step algorithm for linear complementarity problems," *Numerische Mathematik*, vol. 68, no. 1, pp. 95–106, 1994.

[30] A. Cristofari, M. De Santis, S. Lucidi, and F. Rinaldi, "A two-stage active-set algorithm for bound-constrained optimization," *J. Optim. Theory Appl.*, vol. 172, no. 2, pp. 369–401, 2017.

[31] W. Hare and A. S. Lewis, "Identifying active constraints via partial smoothness and prox-regularity," *J. Convex Anal.*, vol. 11, no. 2, pp. 251–266, 2004.
[32] C. Zălinescu, *Convex Analysis in General Vector Spaces.* World scientific, 2002.
[33] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, vol. 408. Springer, 2011.
[34] V. Chandrasekaran, B. Recht, P. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," English, *Found. Comput. Math.*, vol. 12, no. 6, pp. 805–849, 2012.
[35] R. T. Rockafellar, *Convex Analysis.* Princeton: Princeton University Press, 1970.
[36] J. von Neumann, "Some matrix inequalities and metrization of matric-space," in *Univ. Tomsk. Rev.* Ser. Collected Works, vol. IV, Oxford: Pergamon, 1962, pp. 205–218.
[37] A. S. Lewis, "The convex analysis of unitarily invariant matrix functions," *J. Convex Anal.*, vol. 2, no. 1, pp. 173–183, 1995.
[38] M. P. Friedlander and I. Macêdo, "Low-rank spectral optimization via gauge duality," *SIAM J. Sci. Comput.*, vol. 38, no. 3, A1616–A1638, 2016.
[39] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis.* New York, NY: Springer, 2001.
[40] D. P. Bertsekas, *Convex Optimization Theory.* Athena Scientific, 2009.
[41] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Sig. Proc.*, vol. 53, no. 7, pp. 2477–2488, 2005.
[42] I. F. Gorodnitsky, J. S. George, and B. D. Rao, "Neuromagnetic source imaging with FOCUSS: A recursive weighted minimum norm algorithm," *Electroencephalography and Clinical Neurophysiology*, vol. 95, no. 4, pp. 231–251, 1995.
[43] C. Ding, D. Zhou, X. He, and H. Zha, "R 1-PCA: Rotational invariant L1-norm principal component analysis for robust subspace factorization," in *Inter. Conf. Mach. Learning (ICML 2006)*, ACM, pp. 281–288, 2006.

[44] M. Bogdan, E. v. d. Berg, W. Su, and E. Candès, "Statistical estimation and testing via the sorted L1 norm," *arXiv:1310.1969*, 2013.

[45] H. D. Bondell and B. J. Reich, "Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with oscar," *Biometrics*, vol. 64, no. 1, pp. 115–123, 2008.

[46] N. Rao, H.-F. Yu, P. K. Ravikumar, and I. S. Dhillon, "Collaborative filtering with graph information: Consistency and scalable methods," in *Advances in Neural Information Processing Systems (NIPS 2015)*, pp. 2107–2115, 2015.

[47] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in Neural Information Processing Systems (NIPS 2002)*, pp. 585–591, 2002.

[48] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[49] Y. Chi and M. F. Da Costa, "Harnessing sparsity over the continuum: Atomic norm minimization for superresolution," *IEEE Sig. Proc. Mag.*, vol. 37, no. 2, pp. 39–57, 2020.

[50] J. W. McLean and H. J. Woerdeman, "Spectral factorizations and sums of squares representations via semidefinite programming," *SIAM J. Matrix Anal. Appl.*, vol. 23, no. 3, pp. 646–655, 2002.

[51] B. Dumitrescu, *Positive Trigonometric Polynomials and Signal Processing Applications*, vol. 103. Springer, 2007.

[52] E. Ndiaye, O. Fercoq, A. Gramfort, and J. Salmon, "Gap safe screening rules for sparse-group lasso," in *Advances in Neural Information Processing Systems (NIPS 2016)*, pp. 388–396, 2016.

[53] A. Y. Aravkin, J. V. Burke, D. Drusvyatskiy, M. P. Friedlander, and K. MacPhee, "Foundations of gauge and perspective duality," *SIAM J. Optim.*, vol. 28, no. 3, pp. 2406–2434, 2018.

[54] R. W. Harrison, "Phase problem in crystallography," *J. Opt. Soc. Am. A*, vol. 10, pp. 1046–1055, 1993.

[55] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Sig. Proc. Mag.*, vol. 32, no. 3, pp. 87–109, 2015.

[56] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, “Phase retrieval via matrix completion,” *SIAM J. Imag. Sci.*, vol. 6, no. 1, pp. 199–225, 2013.
[57] E. J. Candès, T. Strohmer, and V. Voroninski, “Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming,” *Comm. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, 2013.
[58] M. Teboulle, “Convergence of proximal-like algorithms,” *SIAM J. Optim.*, vol. 7, 1997.
[59] A. S. Nemirovsky and D. B. Yudin, *Problem Complexity and Method Efficiency in Optimization.* New York: Wiley, 1983.
[60] A. Beck and M. Teboulle, “Mirror descent and nonlinear projected subgradient methods for convex optimization,” *Oper. Res. Lett.*, vol. 31, no. 3, pp. 167–175, 2003.
[61] N. Littlestone and M. K. Warmuth, “The weighted majority algorithm,” *Inform. and Comput.*, vol. 108, no. 2, pp. 212–261, 1994.
[62] Y. Freund, R. Schapire, and N. Abe, “A short introduction to boosting,” *J. Jpn. Soc. Artif. Intell.*, vol. 14, no. 771–780, p. 1612, 1999.
[63] N. Rao, P. Shah, and S. Wright, “Forward–backward greedy algorithms for atomic norm regularization,” *IEEE Trans. Sig. Proc.*, vol. 63, no. 21, pp. 5798–5811, 2015.
[64] M. Jaggi and M. Sulovsk, “A simple algorithm for nuclear norm regularized problems,” in *Inter. Conf. Mach. Learning (ICML 2010)*, pp. 471–478, 2010.
[65] S. Shalev-Shwartz, A. Gonen, and O. Shamir, “Large-scale convex minimization with a low-rank constraint,” *arXiv:1106.1622*, 2011.
[66] K. Lee and Y. Bresler, “Efficient and guaranteed rank minimization by atomic decomposition,” in *IEEE International Symposium on Information Theory (ISIT 2009)*, IEEE, pp. 314–318, 2009.
[67] A. Yurtsever, M. Udell, J. A. Tropp, and V. Cevher, “Sketchy decisions: Convex low-rank matrix optimization with optimal storage,” in *International Conference on Artificial Intelligence and Statistics (AISTATS 2017)*, pp. 1188–1196, 2017.

[68] R. M. Bell and Y. Koren, "Lessons from the Netflix prize challenge," *ACM SIGKDD Explorations Newsletter*, vol. 9, no. 2, pp. 75–79, 2007.

[69] C. A. Holloway, "An extension of the Frank and Wolfe method of feasible irections," *Math. Program.*, vol. 6, no. 1, pp. 14–27, 1974.

[70] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[71] D. P. Bertsekas and H. Yu, "A unifying polyhedral approximation framework for convex optimization," *SIAM J. Optim.*, vol. 21, no. 1, pp. 333–360, 2011.

[72] J. E. Kelley Jr, "The cutting-plane method for solving convex programs," *J. Soc. Indust. Appl. Math.*, vol. 8, no. 4, pp. 703–712, 1960.

[73] Z. Fan, Y. Sun, and M. P. Friedlander, "Bundle methods for dual atomic pursuit," in *Asilomar Conference on Signals, Systems, and Computers (ACSSC 2019)*, IEEE, pp. 264–270, 2019.

[74] M. R. Hestenes, "Multiplier and gradient methods," *J. Optim. Theory Appl.*, vol. 4, no. 5, pp. 303–320, 1969.

[75] K. C. Kiwiel, "Proximity control in bundle methods for convex nondifferentiable minimization," *Math. Program.*, vol. 46, no. 1–3, pp. 105–122, 1990.

[76] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Advances in Neural Information Processing Systems (NIPS 2009)*, pp. 2080–2088, 2009.

[77] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. Assoc. Comput. Mach.*, vol. 58, no. 3, p. 11, 2011.

[78] J. Wright, A. Ganesh, K. Min, and Y. Ma, "Compressive principal component pursuit," *IMA Inform. Inference*, vol. 2, no. 1, pp. 32–68, 2013.

[79] M. B. McCoy and J. A. Tropp, "Sharp recovery bounds for convex demixing, with applications," *Found. Comput. Math.*, vol. 14, no. 3, pp. 503–567, 2014.

[80] S. Oymak and J. A. Tropp, "Universality laws for randomized dimension reduction, with applications," *IMA Inform. Inference*, vol. 7, no. 3, pp. 337–446, 2017.

[81] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 58, no. 2, pp. 1094–1121, 2012.

[82] M. P. Friedlander, I. Macêdo, and T. K. Pong, "Polar convolution," *SIAM J. Optim.*, vol. 29, no. 4, pp. 1366–1391, 2019.

[83] A. Seeger and M. Volle, "On a convolution operation obtained by adding level sets: Classical and new results," *RAIRO Recherche Opérationnelle*, vol. 29, pp. 131–154, 1995.