Proximal Algorithms

Neal Parikh
Department of Computer Science
Stanford University
npparikh@cs.stanford.edu

Stephen Boyd
Department of Electrical Engineering
Stanford University
boyd@stanford.edu

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

N. Parikh and S. Boyd. *Proximal Algorithms*. Foundations and Trends® in Optimization, vol. 1, no. 3, pp. 127–239, 2014.

This Foundations and Trends® issue was typeset in LaTeX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-60198-717-4

Foundations and Trends® in Optimization
Volume 1, Issue 3, 2014

Editorial Scope

Topics

Foundations and Trends® in Optimization publishes survey and tutorial articles on methods for and applications of mathematical optimization, including 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
  - 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, 2013, Volume XX, 4 issues. ISSN paper version 2167-3888. ISSN online version 2167-3918. Also available as a combined paper and online subscription.

Foundations and Trends® in Optimization
Vol. 1, No. 3 (2014) 127–239

DOI: 10.1561/2400000003

Proximal Algorithms

Neal Parikh
Department of Computer Science
Stanford University
npparikh@cs.stanford.edu

Stephen Boyd
Department of Electrical Engineering
Stanford University
boyd@stanford.edu

Contents

Abstract

This monograph is about a class of optimization algorithms called *proximal algorithms*. Much like Newton's method is a standard tool for solving unconstrained smooth optimization problems of modest size, proximal algorithms can be viewed as an analogous tool for nonsmooth, constrained, large-scale, or distributed versions of these problems. They are very generally applicable, but are especially well-suited to problems of substantial recent interest involving large or high-dimensional datasets. Proximal methods sit at a higher level of abstraction than classical algorithms like Newton's method: the base operation is evaluating the *proximal operator* of a function, which itself involves solving a small convex optimization problem. These subproblems, which generalize the problem of projecting a point onto a convex set, often admit closed-form solutions or can be solved very quickly with standard or simple specialized methods. Here, we discuss the many different interpretations of proximal operators and algorithms, describe their connections to many other topics in optimization and applied mathematics, survey some popular algorithms, and provide a large number of examples of proximal operators that commonly arise in practice.

N. Parikh and S. Boyd. *Proximal Algorithms*. Foundations and Trends® in Optimization, vol. 1, no. 3, pp. 127–239, 2014.

DOI: 10.1561/2400000003.

1

Introduction

This monograph is about a class of algorithms, called *proximal algorithms*, for solving convex optimization problems. Much like Newton's method is a standard tool for solving unconstrained smooth minimization problems of modest size, proximal algorithms can be viewed as an analogous tool for nonsmooth, constrained, large-scale, or distributed versions of these problems. They are very generally applicable, but they turn out to be especially well-suited to problems of recent and widespread interest involving large or high-dimensional datasets.

Proximal methods sit at a higher level of abstraction than classical optimization algorithms like Newton's method. In the latter, the base operations are low-level, consisting of linear algebra operations and the computation of gradients and Hessians. In proximal algorithms, the base operation is evaluating the *proximal operator* of a function, which involves solving a small convex optimization problem. These subproblems can be solved with standard methods, but they often admit closed-form solutions or can be solved very quickly with simple specialized methods. We will also see that proximal operators and proximal algorithms have a number of interesting interpretations and are connected to many different topics in optimization and applied mathematics.

1.1 Definition

Let $f : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ be a closed proper convex function, which means that its *epigraph*

$$\mathbf{epi}\, f = \{(x,t) \in \mathbf{R}^n \times \mathbf{R} \mid f(x) \le t\}$$

is a nonempty closed convex set. The *effective domain* of f is

$$\mathbf{dom}\, f = \{x \in \mathbf{R}^n \mid f(x) < +\infty\},$$

i.e., the set of points for which f takes on finite values.

The *proximal operator* $\mathbf{prox}_f : \mathbf{R}^n \to \mathbf{R}^n$ of f is defined by

$$\mathbf{prox}_f(v) = \operatorname*{argmin}_x \left(f(x) + (1/2)\|x - v\|_2^2 \right), \tag{1.1}$$

where $\|\cdot\|_2$ is the usual Euclidean norm. The function minimized on the righthand side is strongly convex and not everywhere infinite, so it has a unique minimizer for every $v \in \mathbf{R}^n$ (even when $\mathbf{dom}\, f \subsetneq \mathbf{R}^n$).

We will often encounter the proximal operator of the scaled function λf, where $\lambda > 0$, which can be expressed as

$$\mathbf{prox}_{\lambda f}(v) = \operatorname*{argmin}_x \left(f(x) + (1/2\lambda)\|x - v\|_2^2 \right). \tag{1.2}$$

This is also called the proximal operator of f with parameter λ. (To keep notation light, we write $(1/2\lambda)$ rather than $(1/(2\lambda))$.)

Throughout this monograph, when we refer to the proximal operator of a function, the function will be assumed to be closed proper convex, and it may take on the extended value $+\infty$.

1.2 Interpretations

Figure 1.1 depicts what a proximal operator does. The thin black lines are level curves of a convex function f; the thicker black line indicates the boundary of its domain. Evaluating $\mathbf{prox}_f$ at the blue points moves them to the corresponding red points. The three points in the domain of the function stay in the domain and move towards the minimum of the function, while the other two move to the boundary of the domain and towards the minimum of the function. The parameter λ controls

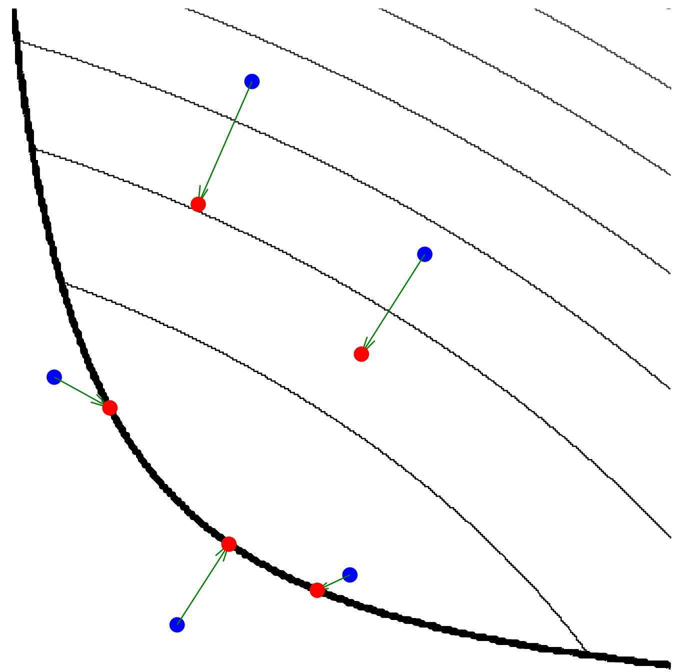

Figure 1.1: Evaluating a proximal operator at various points.

the extent to which the proximal operator maps points towards the minimum of f, with larger values of λ associated with mapped points near the minimum, and smaller values giving a smaller movement towards the minimum. It may be useful to keep this figure in mind when reading about the subsequent interpretations.

We now briefly describe some basic interpretations of (1.1) that we will revisit in more detail later. The definition indicates that $\mathbf{prox}_f(v)$ is a point that compromises between minimizing f and being near to v. For this reason, $\mathbf{prox}_f(v)$ is sometimes called a *proximal point* of v with respect to f. In $\mathbf{prox}_{\lambda f}$, the parameter λ can be interpreted as a relative weight or trade-off parameter between these terms.

When f is the *indicator function*

$$I_{\mathcal{C}}(x) = \begin{cases} 0 & x \in \mathcal{C} \\ +\infty & x \notin \mathcal{C}, \end{cases}$$

where $\mathcal{C}$ is a closed nonempty convex set, the proximal operator of f reduces to Euclidean projection onto $\mathcal{C}$, which we denote

$$\Pi_{\mathcal{C}}(v) = \underset{x \in \mathcal{C}}{\operatorname{argmin}} \|x - v\|_2. \tag{1.3}$$

Proximal operators can thus be viewed as generalized projections, and this perspective suggests various properties that we expect proximal operators to obey.

The proximal operator of f can also be interpreted as a kind of gradient step for the function f. In particular, we have (under some assumptions described later) that

$$\mathbf{prox}_{\lambda f}(v) \approx v - \lambda \nabla f(v)$$

when λ is small and f is differentiable. This suggests a close connection between proximal operators and gradient methods, and also hints that the proximal operator may be useful in optimization. It also suggests that λ will play a role similar to a step size in a gradient method.

Finally, the fixed points of the proximal operator of f are precisely the minimizers of f (we will show this in §2.3). In other words, $\mathbf{prox}_{\lambda f}(x^\star) = x^\star$ if and only if $x^\star$ minimizes f. This implies a close connection between proximal operators and fixed point theory, and suggests that proximal algorithms can be interpreted as solving optimization problems by finding fixed points of appropriate operators.

1.3 Proximal algorithms

A *proximal algorithm* is an algorithm for solving a convex optimization problem that uses the proximal operators of the objective terms. For example, the *proximal minimization algorithm*, discussed in more detail in §4.1, minimizes a convex function f by repeatedly applying $\mathbf{prox}_f$ to some initial point x^0. The interpretations of $\mathbf{prox}_f$ above suggest several potential perspectives on this algorithm, such as an approximate gradient method or a fixed point iteration. In Chapters 4 and 5 we will encounter less trivial and far more useful proximal algorithms.

Proximal algorithms are most useful when all the relevant proximal operators can be evaluated sufficiently quickly. In Chapter 6, we discuss how to evaluate proximal operators and provide many examples.

There are many reasons to study proximal algorithms. First, they work under extremely general conditions, including cases where the functions are nonsmooth and extended real-valued (so they contain implicit constraints). Second, they can be fast, since there can be simple proximal operators for functions that are otherwise challenging to handle in an optimization problem. Third, they are amenable to distributed optimization, so they can be used to solve very large scale problems. Finally, they are often conceptually and mathematically simple, so they are easy to understand, derive, and implement for a particular problem. Indeed, many proximal algorithms can be interpreted as generalizations of other well-known and widely used algorithms, like the projected gradient method, so they are a natural addition to the basic optimization toolbox for anyone who uses convex optimization.

1.4 What this paper is about

We aim to provide a readable reference on proximal operators and proximal algorithms for a wide audience. There are several novel aspects.

First, we discuss a large number of different perspectives on proximal operators, some of which have not previously appeared in the literature, and many of which have not been collected in one place. These include interpretations based on projection operators, smoothing and regularization, resolvent operators, and differential equations. Second, we place strong emphasis on practical use, so we provide many examples of proximal operators that are efficient to evaluate. Third, we have a more detailed discussion of distributed optimization algorithms than most previous references on proximal operators.

To keep the treatment accessible, we have omitted a few more advanced topics, such as the connection to monotone operator theory.

We also include source code for all examples, as well as a library of implementations of proximal operators, at

`http://www.stanford.edu/~boyd/papers/prox_algs.html`

We provide links to other libraries of proximal operators, such as those by Becker et al. and Vaiter, in the documentation for our own library.

1.5 Related work

We emphasize that proximal operators are not new and that there have been other surveys written on various aspects of this topic over the years. Lemaire [123] surveys the literature on the proximal point algorithm up to 1989. Iusem [110] reviews the proximal point method and its connection to augmented Lagrangians. An excellent recent reference by Combettes and Pesquet [63] discusses proximal operators and proximal algorithms in the context of signal processing problems. The lecture notes for Vandenberghe's EE 236C course [196] covers proximal algorithms in detail. Finally, the recent monograph by Boyd et al. [33] is about a particular algorithm (ADMM), but also discusses connections to proximal operators. We will discuss more of the history of proximal operators in the sequel.

1.6 Outline

In Chapter 2, we give some basic properties of proximal operators. In Chapter 3, we discuss a variety of interpretations of proximal operators. Chapter 4 covers some core proximal algorithms for solving convex optimization problems. In Chapter 5, we discuss how to use these algorithms to solve problems in a parallel or distributed fashion. Chapter 6 presents a large number of examples of different projection and proximal operators that can be evaluated efficiently. In Chapter 7, we illustrate these ideas with some examples and applications.

References

[1] K.J. Arrow and G. Debreu. Existence of an equilibrium for a competitive economy. *Econometrica: Journal of the Econometric Society*, 22(3):265–290, 1954.

[2] K.J. Arrow, L. Hurwicz, and H. Uzawa. *Studies in Linear and Nonlinear Programming*. Stanford University Press: Stanford, 1958.

[3] H. Attouch. Convergence de fonctions convexes, de sous-différentiels et semi-groupes. *C.R Acad. Sci. Paris*, 284:539–542, 1977.

[4] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski. Optimization with sparsity-inducing penalties. *Foundations and Trends in Machine Learning*, 4(1):1–106, 2011.

[5] S. Barman, X. Liu, S. Draper, and B. Recht. Decomposition methods for large scale LP decoding. In *Allerton Conference on Communication, Control, and Computing*, pages 253–260. IEEE, 2011.

[6] J. Barzilai and J.M. Borwein. Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8(1):141–148, 1988.

[7] H.H. Bauschke. *Projection Algorithms and Monotone Operators*. PhD thesis, Simon Fraser University, 1996.

[8] H.H. Bauschke and J.M. Borwein. Dykstra's alternating projection algorithm for two sets. *Journal of Approximation Theory*, 79(3):418–443, 1994.

[9] H.H. Bauschke and J.M. Borwein. On projection algorithms for solving convex feasibility problems. *SIAM Review*, 38(3):367–426, 1996.

[10] H.H. Bauschke and P.L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer-Verlag, 2011.

[11] H.H. Bauschke, P.L. Combettes, and D. Noll. Joint minimization with alternating Bregman proximity operators. *Pacific Journal on Mathematics*, 2:401–424, 2006.

[12] H.H. Bauschke, R. Goebel, Y. Lucet, and X. Wang. The proximal average: basic theory. *SIAM Journal on Optimization*, 19(2):766–785, 2008.

[13] H.H. Bauschke and V.R. Koch. Projection methods: Swiss army knives for solving feasibility and best approximation problems with halfspaces. See arXiv:1301.4506v1, 2013.

[14] H.H. Bauschke and S.G. Krug. Reflection-projection method for convex feasibility problems with an obtuse cone. *Journal of Optimization Theory and Applications*, 120(3):503–531, 2004.

[15] H.H. Bauschke, S.M. Moffat, and X. Wang. Firmly nonexpansive mappings and maximally monotone operators: correspondence and duality. *Set-Valued and Variational Analysis*, pages 1–23, 2012.

[16] A. Beck and M. Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175, 2003.

[17] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[18] A. Beck and M. Teboulle. Gradient-based algorithms with applications to signal recovery problems. In D.P. Palomar and Y.C. Eldar, editors, *Convex Optimization in Signal Processing and Communications*, pages 42–88. Cambribge University Press, 2010.

[19] A. Beck and M. Teboulle. Smoothing and first order methods: A unified framework. *SIAM Journal on Optimization*, 22(2):557–580, 2012.

[20] S. Becker, J. Bobin, and E.J. Candès. NESTA: A fast and accurate first-order method for sparse recovery. *SIAM Journal on Imaging Sciences*, 4(1):1–39, 2011.

[21] S. Becker and M. Fadili. A quasi-Newton proximal splitting method. *Advances in Neural Information Processing Systems*, 2012.

[22] S.R. Becker, E.J. Candès, and M.C. Grant. Templates for convex cone problems with applications to sparse signal recovery. *Mathematical Programming Computation*, pages 1–54, 2011.

[23] J.F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4:238–252, 1962.

[24] D.P. Bertsekas. Multiplier methods: A survey. *Automatica*, 12:133–145, 1976.

[25] D.P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, 1982.

[26] D.P. Bertsekas. *Nonlinear Programming*. Athena Scientific, second edition, 1999.

[27] D.P. Bertsekas and J.N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Prentice Hall, 1989.

[28] J.M. Bioucas-Dias and M.A.T. Figueiredo. A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image Processing*, 16(12):2992–3004, 2007.

[29] L.S. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, and R.C. Whaley. *ScaLAPACK user's guide*. SIAM: Philadelphia, 1997.

[30] M. Bogdan, E. van den Berg, W. Su, and E. Candès. Statistical estimation and testing via the sorted ℓ_1 norm. *arXiv:1310.1969*, 2013.

[31] J.F. Bonnans, J.C. Gilbert, C. Lemaréchal, and C.A. Sagastizábal. A family of variable metric proximal methods. *Mathematical Programming*, 68(1):15–47, 1995.

[32] S. Boyd, M. Mueller, B. O'Donoghue, and Y. Wang. Performance bounds and suboptimal policies for multi-period investment. To appear in *Foundations and Trends in Optimization*, 2013.

[33] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[34] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[35] S. Boyd and L. Vandenberghe. Localization and cutting-plane methods. From Stanford EE 364b lecture notes, 2007.

[36] K. Bredies and D. Lorenz. Linear convergence of iterative soft-thresholding. *Journal of Fourier Analysis and Applications*, 14(5-6):813–837, 2008.

[37] L.M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3):200–217, 1967.

[38] H. Brezis. *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert*. North-Holland: Amsterdam, 1973.

[39] F.E. Browder. Multi-valued monotone nonlinear mappings and duality mappings in Banach spaces. *Transactions of the American Mathematical Society*, 118:338–351, 1965.

[40] F.E. Browder. Nonlinear monotone operators and convex sets in Banach spaces. *Bull. Amer. Math. Soc*, 71(5):780–785, 1965.

[41] F.E. Browder. Convergence theorems for sequences of nonlinear operators in Banach spaces. *Mathematische Zeitschrift*, 100(3):201–225, 1967.

[42] F.E. Browder. Nonlinear maximal monotone operators in Banach space. *Mathematische Annalen*, 175(2):89–113, 1968.

[43] R.D. Bruck. An iterative solution of a variational inequality for certain monotone operator in a Hilbert space. *Bulletin of the American Mathematical Society*, 81(5):890–892, 1975.

[44] A.M. Bruckstein, D.L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51(1):34–81, 2009.

[45] R.S. Burachik and A.N. Iusem. *Set-Valued Mappings and Enlargements of Monotone Operators*. Springer, 2008.

[46] J.F. Cai, E.J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

[47] E.J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? See arXiv:0912.3599, 2009.

[48] E.J. Candès, C.A. Sing-Long, and J.D. Trzasko. Unbiased risk estimates for singular value thresholding and spectral estimators. See arXiv:1210.4139, 2012.

[49] E.J. Candès and M. Soltanolkotabi. Discussion of 'latent variable graphical model selection via convex optimization'. *Annals of Statistics*, pages 1997–2004, 2012.

[50] Y. Censor and S.A. Zenios. Proximal minimization algorithm with *D*-functions. *Journal of Optimization Theory and Applications*, 73(3):451–464, 1992.

[51] Y. Censor and S.A. Zenios. *Parallel Optimization: Theory, Algorithms, and Applications*. Oxford University Press, 1997.

[52] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[53] V. Chandrasekaran, P.A. Parrilo, and A.S. Willsky. Latent variable graphical model selection via convex optimization. *Annals of Statistics* (with discussion), 2012.

[54] V. Chandrasekaran, S. Sanghavi, P.A. Parrilo, and A.S. Willsky. Sparse and low-rank matrix decompositions. In *Allerton 2009*, pages 962–967. IEEE, 2009.

[55] V. Chandrasekaran, S. Sanghavi, P.A. Parrilo, and A.S. Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011.

[56] G.H.G. Chen. *Forward-backward splitting techniques: theory and applications*. PhD thesis, University of Washington, 1994.

[57] G.H.G. Chen and R.T. Rockafellar. Convergence rates in forward-backward splitting. *SIAM Journal on Optimization*, 7(2):421–444, 1997.

[58] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Review*, 43(1):129–159, 2001.

[59] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu. Epigraphical projection and proximal tools for solving constrained convex optimization problems: Part i. See arXiv:1210.5844, 2012.

[60] P. Combettes and J.-C. Pesquet. Proximal thresholding algorithm for minimization over orthonormal bases. *SIAM Journal on Optimization*, 18(4):1351–1376, 2007.

[61] P.L. Combettes. Solving monotone inclusions via compositions of non-expansive averaged operators. *Optimization*, 53(5-6), 2004.

[62] P.L. Combettes and J.-C. Pesquet. A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE Journal on Selected Topics in Signal Processing*, 1(4):564–574, 2007.

[63] P.L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212, 2011.

[64] P.L. Combettes and V.R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling and Simulation*, 4(4):1168–1200, 2006.

[65] A. Daniilidis, D. Drusvyatskiy, and A.S. Lewis. Orthogonal invariance and identifiability. arXiv:1304.1198, 2013.

[66] A. Daniilidis, A.S. Lewis, J. Malick, and H. Sendov. Prox-regularity of spectral functions and spectral sets. *Journal of Convex Analysis*, 15(3):547–560, 2008.

[67] A. Daniilidis, J. Malick, and H. Sendov. Locally symmetric submanifolds lift to spectral manifolds. arXiv:1212.3936, 2012.

[68] G.B. Dantzig and P. Wolfe. Decomposition principle for linear programs. *Operations Research*, 8:101–111, 1960.

[69] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57:1413–1457, 2004.

[70] C. Davis. All convex invariant functions of Hermitian matrices. *Archiv der Mathematik*, 8(4):276–278, 1957.

[71] C.A. Deledalle, S. Vaiter, G. Peyré, J. Fadili, and C. Dossal. Risk estimation for matrix recovery with spectral regularization. See arXiv:1205.1482, 2012.

[72] J.W. Demmel, M.T. Heath, and H.A. Van Der Vorst. *Parallel numerical linear algebra*. Computer Science Division (EECS), University of California, 1993.

[73] A.P. Dempster. Covariance selection. *Biometrics*, 28(1):157–175, 1972.

[74] N. Derbinsky, J. Bento, V. Elser, and J. Yedidia. An improved three-weight message-passing algorithm. *arXiv:1305.1961*, 2013.

[75] C.B. Do, Q.V. Le, and C.S. Foo. Proximal regularization for online and batch learning. In *International Conference on Machine Learning*, pages 257–264, 2009.

[76] D.L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41:613–627, 1995.

[77] J. Douglas and H.H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American Mathematical Society*, 82:421–439, 1956.

[78] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra. Efficient projections onto the ℓ_1-ball for learning in high dimensions. In *Proceedings of the 25th International Conference on Machine learning*, pages 272–279, 2008.

[79] R.L. Dykstra. An algorithm for restricted least squares regression. *Journal of the American Statistical Association*, 78(384):837–842, 1983.

[80] J. Eckstein. *Splitting methods for monotone operators with applications to parallel optimization*. PhD thesis, MIT, 1989.

[81] J. Eckstein. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Mathematics of Operations Research*, pages 202–226, 1993.

[82] E. Esser, X. Zhang, and T.F. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, 3(4):1015–1046, 2010.

[83] F. Facchinei and J.S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer-Verlag, 2003.

[84] M.C. Ferris. Finite termination of the proximal point algorithm. *Mathematical Programming*, 50(1):359–366, 1991.

[85] M.A.T. Figueiredo, J.M. Bioucas-Dias, and R.D. Nowak. Majorization–minimization algorithms for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 16(12):2980–2991, 2007.

[86] M.A.T. Figueiredo and R.D. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 12(8):906–916, 2003.

[87] M.A.T. Figueiredo and R.D. Nowak. A bound optimization approach to wavelet-based image deconvolution. In *IEEE International Conference on Image Processing*, volume 2, pages II–782. IEEE, 2005.

[88] G.F. Franklin, J.D. Powell, and A. Emami-Naeini. *Feedback Control of Dynamic Systems*, volume 3. Addison-Wesley: Reading, MA, 1994.

[89] S. Friedland. Convex spectral functions. *Linear and Multilinear Algebra*, 9(4):299–316, 1981.

[90] M. Fukushima and H. Mine. A generalized proximal point algorithm for certain non-convex minimization problems. *International Journal of Systems Science*, 12(8):989–1000, 1981.

[91] M. Fukushima and L. Qi. A globally and superlinearly convergent algorithm for nonsmooth convex minimization. *SIAM Journal on Optimization*, 6(4):1106–1120, 1996.

[92] D. Gabay. Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*. North-Holland: Amsterdam, 1983.

[93] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximations. *Computers and Mathematics with Applications*, 2:17–40, 1976.

[94] K.A. Gallivan, R.J. Plemmons, and A.H. Sameh. Parallel algorithms for dense linear algebra computations. *SIAM Review*, pages 54–135, 1990.

[95] A.M. Geoffrion. Generalized Benders decomposition. *Journal of Optimization Theory and Applications*, 10(4):237–260, 1972.

[96] R. Glowinski and A. Marrocco. Sur l'approximation, par elements finis d'ordre un, et la resolution, par penalisation-dualité, d'une classe de problems de Dirichlet non lineares. *Revue Française d'Automatique, Informatique, et Recherche Opérationelle*, 9:41–76, 1975.

[97] D. Goldfarb and K. Scheinberg. Fast first-order methods for composite convex optimization with line search. *preprint*, 2011.

[98] G.H. Golub and J.H. Wilkinson. Note on the iterative refinement of least squares solution. *Numerische Mathematik*, 9(2):139–148, 1966.

[99] A. Granas and J. Dugundji. *Fixed Point Theory*. Springer, 2003.

[100] M. Grant, S. Boyd, and Y. Ye. CVX: Matlab software for disciplined convex programming, ver. 1.1, build 630. Available at `www.stanford.edu/~boyd/cvx/`, April 2008.

[101] S.J. Grotzinger and C. Witzgall. Projections onto order simplexes. *Applied Mathematics and Optimization*, 12(1):247–270, 1984.

[102] O. Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM Journal on Control and Optimization*, 29:403, 1991.

[103] O. Güler. New proximal point algorithms for convex minimization. *SIAM Journal on Optimization*, 2:649, 1992.

[104] E.T. Hale, W. Yin, and Y. Zhang. Fixed-point continuation for ℓ_1-minimization: Methodology and convergence. *SIAM Journal on Optimization*, 19(3):1107–1130, 2008.

[105] P.T. Harker and J.S. Pang. Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. *Mathematical Programming*, 48(1):161–220, 1990.

[106] B. He and X. Yuan. On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. *SIAM Journal on Numerical Analysis*, 50(2):700–709, 2012.

[107] P.J. Huber. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35(1):73–101, 1964.

[108] D.R. Hunter and K. Lange. A tutorial on MM algorithms. *The American Statistician*, 58(1):30–37, 2004.

[109] S. Ibaraki, M. Fukushima, and T. Ibaraki. Primal-dual proximal point algorithm for linearly constrained convex programming problems. *Computational Optimization and Applications*, 1(2):207–226, 1992.

[110] A.N. Iusem. Augmented Lagrangian methods and proximal point methods for convex optimization. *Investigación Operativa*, 8:11–49, 1999.

[111] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach. Proximal methods for hierarchical sparse coding. See arXiv:1009.2139, 2010.

[112] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach. Proximal methods for sparse hierarchical dictionary learning. In *International Conference on Machine Learning*, 2010.

[113] R.I. Kachurovskii. Monotone operators and convex functionals. *Uspekhi Matematicheskikh Nauk*, 15(4):213–215, 1960.

[114] R.I. Kachurovskii. Non-linear monotone operators in Banach spaces. *Russian Mathematical Surveys*, 23(2):117–165, 1968.

[115] A. Kaplan and R. Tichatschke. Proximal point methods and nonconvex optimization. *Journal of Global Optimization*, 13(4):389–406, 1998.

[116] S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky. ℓ_1 trend filtering. *SIAM Review*, 51(2):339–360, 2009.

[117] B.W. Kort and D.P. Bertsekas. Multiplier methods for convex programming. In *IEEE Conference on Decision and Control*, volume 12, 1973.

[118] M. Kraning, E. Chu, J. Lavaei, and S. Boyd. Message passing for dynamic network energy management. To appear, 2012.

[119] M. Kyono and M. Fukushima. Nonlinear proximal decomposition method for convex programming. *Journal of Optimization Theory and Applications*, 106(2):357–372, 2000.

[120] L.S. Lasdon. *Optimization Theory for Large Systems*. MacMillan, 1970.

[121] J. Lee, B. Recht, R. Salakhutdinov, N. Srebro, and J.A. Tropp. Practical large-scale optimization for max-norm regularization. *Advances in Neural Information Processing Systems*, 23:1297–1305, 2010.

[122] B. Lemaire. Coupling optimization methods and variational convergence. *Trends in Mathematical Optimization, International Series of Numerical Mathematics*, 84, 1988.

[123] B. Lemaire. The proximal algorithm. *International Series of Numerical Mathematics*, pages 73–87, 1989.

[124] C. Lemaréchal and C. Sagastizábal. Practical aspects of the Moreau-Yosida regularization I: theoretical properties, 1994. INRIA Technical Report 2250.

[125] C. Lemaréchal and C. Sagastizábal. Practical aspects of the Moreau–Yosida regularization: Theoretical preliminaries. *SIAM Journal on Optimization*, 7(2):367–385, 1997.

[126] K. Levenberg. A method for the solution of certain problems in least squares. *Quarterly of Applied Mathematics*, 2:164–168, 1944.

[127] A.S. Lewis. The convex analysis of unitarily invariant matrix functions. *Journal of Convex Analysis*, 2(1):173–183, 1995.

[128] A.S. Lewis. Convex analysis on the Hermitian matrices. *SIAM Journal on Optimization*, 6(1):164–177, 1996.

[129] A.S. Lewis. Derivatives of spectral functions. *Mathematics of Operations Research*, 21(3):576–588, 1996.

[130] A.S. Lewis and J. Malick. Alternating projections on manifolds. *Mathematics of Operations Research*, 33(1):216–234, 2008.

[131] P.L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16:964–979, 1979.

[132] F.J. Luque. Asymptotic convergence analysis of the proximal point algorithm. *SIAM Journal on Control and Optimization*, 22(2):277–293, 1984.

[133] S. Ma, L. Xue, and H. Zou. Alternating direction methods for latent variable gaussian graphical model selection. See arXiv:1206.1275, 2012.

[134] S. Ma, L. Xue, and H. Zou. Alternating direction methods for latent variable Gaussian graphical model selection. *Neural Computation*, pages 1–27, 2013.

[135] W.R. Mann. Mean value methods in iteration. *Proceedings of the American Mathematical Society*, 4(3):506–510, 1953.

[136] D.W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[137] B. Martinet. Régularisation d'inéquations variationnelles par approximations successives. *Revue Française de Informatique et Recherche Opérationelle*, 1970.

[138] B. Martinet. Détermination approchée d'un point fixe d'une application pseudo-contractante. *C.R. Acad. Sci. Paris*, 274A:163–165, 1972.

[139] J. Mattingley and S. Boyd. CVXGEN: A code generator for embedded convex optimization. *Optimization and Engineering*, pages 1–27, 2012.

[140] N. Meinshausen and P.Bühlmann. High-dimensional graphs and variable selection with the lasso. *Annals of Statistics*, 34(3):1436–1462, 2006.

[141] G.J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Mathematical Journal*, 29(3):341–346, 1962.

[142] G.J. Minty. On the monotonicity of the gradient of a convex function. *Pacific J. Math*, 14(1):243–247, 1964.

[143] C.B. Moler. Iterative refinement in floating point. *Journal of the ACM (JACM)*, 14(2):316–321, 1967.

[144] J.-J. Moreau. Fonctions convexes duales et points proximaux dans un espace Hilbertien. *Reports of the Paris Academy of Sciences, Series A*, 255:2897–2899, 1962.

[145] J.-J. Moreau. Proximité et dualité dans un espace Hilbertien. *Bull. Soc. Math. France*, 93(2):273–299, 1965.

[146] A. Nedić and A. Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1):48–61, 2009.

[147] A. Nedić and A. Ozdaglar. Cooperative distributed multi-agent optimization. In D.P. Palomar and Y.C. Eldar, editors, *Convex Optimization in Signal Processing and Communications*. Cambridge University Press, 2010.

[148] A.B. Németh and S.Z. Németh. How to project onto an isotone projection cone. *Linear Algebra and its Applications*, 433(1):41–51, 2010.

[149] A.S. Nemirovsky and D.B. Yudin. *Problem Complexity and Method Efficiency in Optimization*. Wiley, 1983.

[150] Y. Nesterov. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Soviet Mathematics Doklady*, 27(2):372–376, 1983.

[151] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Springer, 2004.

[152] Y. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1):127–152, 2005.

[153] Y. Nesterov. Gradient methods for minimizing composite objective function. *CORE Discussion Paper, Catholic University of Louvain*, 76:2007, 2007.

[154] J.X.C. Neto, O.P. Ferreira, A.N. Iusem, and R.D.C. Monteiro. Dual convergence of the proximal point method with Bregman distances for linear programming. *Optimization Methods and Software*, pages 1–23, 2007.

[155] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer-Verlag, 1999.

[156] B. O'Donoghue and E. Candès. Adaptive restart for accelerated gradient schemes. See arXiv:1204.3982, 2012.

[157] B. O'Donoghue, G. Stathopoulos, and S. Boyd. A splitting method for optimal control. To appear in *IEEE Transactions on Control Systems Technology*, 2012.

[158] H. Ohlsson, L. Ljung, and S. Boyd. Segmentation of ARX-models using sum-of-norms regularization. *Automatica*, 46(6):1107–1111, 2010.

[159] H. Ouyang, N. He, and A. Gray. Stochastic ADMM for nonsmooth optimization. See arXiv:1211.0632, 2012.

[160] N. Parikh and S. Boyd. Block splitting for distributed optimization. Submitted, 2012.

[161] G.B. Passty. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *Journal of Mathematical Analysis and Applications*, 72(2):383–390, 1979.

[162] J.P. Penot. Proximal mappings. *Journal of Approximation Theory*, 94(2):203–221, 1998.

[163] R.A. Poliquin and R.T. Rockafellar. Prox-regular functions in variational analysis. *Transactions of the American Mathematical Society*, 348(5):1805–1838, 1996.

[164] B. Polyak. *Introduction to Optimization*. Optimization Software, Inc., 1987.

[165] B.T. Polyak. Iterative methods using Lagrange multipliers for solving extremal problems with constraints of the equation type. *USSR Computational Mathematics and Mathematical Physics*, 10(5):42–52, 1970.

[166] R. Polyak and M. Teboulle. Nonlinear rescaling and proximal-like methods in convex optimization. *Mathematical Programming*, 76(2):265–284, 1997.

[167] N. Pustelnik, C. Chaux, and J.-C. Pesquet. Parallel proximal algorithm for image restoration using hybrid regularization. *IEEE Transactions on Image Processing*, 20(9):2450–2462, 2011.

[168] N. Pustelnik, J.-C. Pesquet, and C. Chaux. Relaxing tight frame condition in parallel proximal methods for signal restoration. *IEEE Transactions on Signal Processing*, 60(2):968–973, 2012.

[169] A. Quattoni, X. Carreras, M. Collins, and T. Darrell. An efficient projection for $\ell_{1,\infty}$ regularization. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 857–864, 2009.

[170] P. Ravikumar, A. Agarwal, and M.J. Wainwright. Message-passing for graph-structured linear programs: Proximal methods and rounding schemes. *Journal of Machine Learning Research*, 11:1043–1080, 2010.

[171] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[172] R.T. Rockafellar. On the maximal monotonicity of subdifferential mappings. *Pacific J. Math.*, 33(1):209–216, 1970.

[173] R.T. Rockafellar. On the maximality of sums of nonlinear monotone operators. *Transactions of the American Mathematical Society*, 149(1):75–88, 1970.

[174] R.T. Rockafellar. A dual approach to solving nonlinear programming problems by unconstrained optimization. *Mathematical Programming*, 5(1):354–373, 1973.

[175] R.T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1:97–116, 1976.

[176] R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14:877, 1976.

[177] R.T. Rockafellar and R. J-B Wets. *Variational Analysis*. Springer-Verlag, 1998.

[178] J. Saunderson, V. Chandrasekaran, P.A. Parrilo, and A.S. Willsky. Diagonal and low-rank matrix decompositions, correlation matrices, and ellipsoid fitting. See arXiv:1204.1220, 2012.

[179] K. Scheinberg and D. Goldfarb. Fast first-order methods for composite convex optimization with large steps. Available online, 2012.

[180] K. Scheinberg, S. Ma, and D. Goldfarb. Sparse inverse covariance selection via alternating linearization methods. In *Advances in Neural Information Processing Systems*, 2010.

[181] H. Sendov. The higher-order derivatives of spectral functions. *Linear Algebra and its Applications*, 424(1):240–281, 2007.

[182] S. Sra. Fast projections onto $\ell_{1,q}$-norm balls for grouped feature selection. *Machine Learning and Knowledge Discovery in Databases*, pages 305–317, 2011.

[183] M. Teboulle. Entropic proximal mappings with applications to nonlinear programming. *Mathematics of Operations Research*, pages 670–690, 1992.

[184] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58(1):267–288, 1996.

[185] K.C. Toh and S. Yun. An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. *Preprint*, 2009.

[186] P. Tseng. Further applications of a splitting algorithm to decomposition in variational inequalities and convex programming. *Mathematical Programming*, 48(1):249–263, 1990.

[187] P. Tseng. Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM Journal on Control and Optimization*, 29(1):119–138, 1991.

[188] P. Tseng. Alternating projection-proximal methods for convex programming and variational inequalities. *SIAM Journal on Optimization*, 7:951–965, 1997.

[189] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, 38:431, 2000.

[190] P. Tseng. On accelerated proximal gradient methods for convex-concave optimization. *SIAM Journal on Optimization*, 2008.

[191] J.N. Tsitsiklis. *Problems in decentralized decision making and computation.* PhD thesis, Massachusetts Institute of Technology, 1984.

[192] H. Uzawa. Market mechanisms and mathematical programming. *Econometrica: Journal of the Econometric Society*, 28(4):872–881, 1960.

[193] H. Uzawa. Walras' tâtonnement in the theory of exchange. *The Review of Economic Studies*, 27(3):182–194, 1960.

[194] L. Vandenberghe. Fast proximal gradient methods. From http://www.ee.ucla.edu/~vandenbe/236C/lectures/fgrad.pdf, 2010.

[195] L. Vandenberghe. Lecture on proximal gradient method. From http://www.ee.ucla.edu/~vandenbe/shortcourses/dtu-10/lecture3.pdf, 2010.

[196] L. Vandenberghe. Optimization methods for large-scale systems. UCLA EE 236C lecture notes, 2010.

[197] J. von Neumann. Some matrix inequalities and metrization of matrix space. *Tomsk University Review*, 1:286–300, 1937.

[198] J. von Neumann. *Functional Operators, Volume 2: The Geometry of Orthogonal Spaces*. Princeton University Press: Annals of Mathematics Studies, 1950. Reprint of 1933 lecture notes.

[199] Z. Wen, D. Goldfarb, and W. Yin. Alternating direction augmented Lagrangian methods for semidefinite programming. Technical report, Department of IEOR, Columbia University, 2009.

[200] P. Whittle. *Risk-sensitive Optimal Control*. Wiley, 1990.

[201] S.J. Wright, R.D. Nowak, and M.A.T. Figueiredo. Sparse reconstruction by separable approximation. *IEEE Transactions on Signal Processing*, 57(7):2479–2493, 2009.

[202] K. Yosida. *Functional Analysis*. Springer, 1968.

[203] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006.

[204] E.H. Zarantonello. *Solving functional equations by contractive averaging*. Mathematics Research Center, United States Army, University of Wisconsin, 1960.

[205] E.H. Zarantonello. Projections on convex sets in Hilbert space and spectral theory. I. Projections on convex sets. In *Contributions to Nonlinear Functional Analysis (Proceedings of a Symposium held at the Mathematics Research Center, University of Wisconsin, Madison, Wis., 1971)*, pages 237–341, 1971.

[206] X. Zhang, M. Burger, X. Bresson, and S. Osher. Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. *SIAM Journal on Imaging Sciences*, 3(3):253–276, 2010.

[207] X. Zhang, M. Burger, and S. Osher. A unified primal-dual algorithm framework based on Bregman iteration. *Journal of Scientific Computing*, 46(1):20–46, 2011.

[208] P. Zhao, G. Rocha, and B. Yu. The composite absolute penalties family for grouped and hierarchical variable selection. *Annals of Statistics*, 37(6A):3468–3497, 2009.

[209] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67:301–320, 2005.