

Bregman generalized subgradient projected algorithms and applications

Xiaoyu Mao · Xianfu Wang

Received: date / Accepted: date

Abstract Projected subgradient methods were thought as ideal algorithms for constrained minimization problems. But they are often hampered by the computational complexity of the projection operator. The Bregman distance has been identified as a potential solution to this issue. The existing mirror descent algorithm, which is designed for one-block constrained minimization problems, has been proved an efficient approach. Previous research has revealed its equivalence with Bregman projected subgradient method. This paper expands the algorithm to two-block situations, thereby increasing its applicability. We also provide its convergence proof. Our algorithm has a strong correlation with another renowned method, the “proximal alternating linearized method,” which can be considered a specific variant of our approach. In the numerical experiments, we choose a variety of examples from various fields. The results consistently indicate that Bregman generalized algorithms significantly reduce computation time, particularly when the variable is constrained to the unit simplex.

Keywords Bregman distance · subgradient method · projection operator · alternating minimization

Mathematics Subject Classification (2010) 90C29 · 90C56

1 Introduction

The two-block minimization problem is an important subset of traditional optimization theory. Numerous practical applications can be abstracted into a two-block problem. For instance, matrix decomposition is a two-block problem, where each factorized matrix corresponds to a block. Sometimes even if the objective function is one-block, we introduce an instrumental variable and a constraint, transforming

X. Mao · X. Wang

Department of Mathematics, I.K. Barber Faculty of Science, Kelowna, BC, Canada.

E-mail: alex.mao@ubc.ca

X. Wang (✉)

E-mail: shawn.wang@ubc.ca

the problem into a two-block one. It can potentially simplify the process of finding a solution. This paper mainly considers a constrained two-block optimization problem that can be formulated by

$$\min F(x, y) \quad \text{s.t. } x \in \mathcal{X}, y \in \mathcal{Y}, \quad (1)$$

where $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^n$ are closed convex sets, and F satisfies the following assumptions:

- $F : \mathcal{Z} = \mathcal{X} \times \mathcal{Y} \rightarrow \overline{\mathbb{R}}$ is partially convex and Lipschitz continuous;
- The partial subgradients of F are computable;
- The minimizer set \mathcal{Z}^* is nonempty.

The subdifferentiability throughout this paper uses Frechet's definition. The partial subdifferentials in the second assumption means that for each fixed point $(z_1, z_2) \in \text{dom } F$, the subdifferentials $\partial_x F(x, z_2)$ and $\partial_y F(z_1, y)$ with respect to x and y in order are easy to compute.

Definition 1 (Subdifferential) Let $f : \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ be a proper and lower semicontinuous function

1. The Frechet subdifferential of function f at x , where $x \in \text{dom } f$, denoted by $\hat{\partial}f(x)$, is the collection of all the vectors $u \in \mathbb{R}^m$ which satisfies

$$\lim_{\substack{y \neq x, \\ y \rightarrow x}} \inf \frac{f(y) - f(x) - \langle u, y - x \rangle}{\|y - x\|} \geq 0.$$

When $x \notin \text{dom } f$, otherwise, we let $\hat{\partial}f(x) = \emptyset$.

2. The subdifferential (or limiting subdifferential) is defined through following process

$$\partial f(x) = \{u \in \mathbb{R}^m : \exists y^\nu \rightarrow x, f(y^\nu) \rightarrow f(x) \text{ and } u^\nu \in \hat{\partial}f(y^\nu) \rightarrow u\}.$$

The partially convexity and Lipschitz continuity are another two fundamental assumptions in this paper. They are defined in the following ways.

Definition 2 (Partially convex function) Let $f : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ be an extended-real-valued function. Then f is partially convex if $f(\cdot, z_2)$ and $f(z_1, \cdot)$ are two convex functions, i.e.

$$f(\alpha x + (1 - \alpha)y, z_2) \leq \alpha f(x, z_2) + (1 - \alpha)f(y, z_2),$$

for all $x, y \in \text{dom } f(\cdot, z_2)$, and

$$f(z_1, \alpha x + (1 - \alpha)y) \leq \alpha f(z_1, x) + (1 - \alpha)f(z_1, y),$$

for all $x, y \in \text{dom } f(z_1, \cdot)$, where $(z_1, z_2) \in \text{dom } f$ is a fixed point and $\alpha \in [0, 1]$.

Definition 3 (Partially Lipschitz function) Let $f : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ be a proper and lower semicontinuous function. Then f is partially Lipschitz if there exists two constants L_1 and L_2 such that for each fixed point $(z_1, z_2) \in \text{dom } f$, $f(\cdot, z_2)$ and $f(z_1, \cdot)$ are two Lipschitz functions, i.e.

$$|f(x, z_2) - f(y, z_2)| \leq L_1 \|x - y\|,$$

for all $x, y \in \text{int dom } f(\cdot, z_2)$ and

$$|f(z_1, x) - f(z_1, y)| \leq L_2 \|x - y\|,$$

for all $x, y \in \text{int dom } f(z_1, \cdot)$.

Note that a partially convex function does not indicate it is convex with respect to the whole blocks. In the other word, partial convexity is weaker than convexity. The similar relation also adopts to Lipschitz continuity. Please refer th following two examples.

Example 1 Let $f : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ be given by $f(x, y) = f_1(x)f_2(y)$ where $f_1 \geq 0$ and $f_2 \geq 0$ are two nonnegative convex functions. Then f may be not a convex function but is a partially convex function.

Example 2 Let $f : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ be $f(x, y) = \langle x, y \rangle$ defined on $\|x\| \leq 1$ and $\|y\| \leq 1$. It is a Lipschitz function with parameter $L = \sqrt{2}$ but a partially Lipschitz function with $L_1 = 1$ and $L_2 = 1$.

The conventional approach for solving (1) is known as the “alternating subgradient projected methods” [25, 13], which is straightforward and efficient. It updates the variables in cyclic order by

$$\begin{aligned} y^{k+1} &= \Pi_{\mathcal{Y}} \left(y^k - t_k \hat{g}_k(y^k) \right), \\ x^{k+1} &= \Pi_{\mathcal{X}} \left(x^k - t_k \hat{h}_k(x^k) \right), \end{aligned}$$

where $\hat{g}_k(y) \in \partial_y F(x_k, y)$ and $\hat{h}_k(x) \in \partial_x F(x, y_{k+1})$ are subgradients, $t_k > 0$ are step sizes and $\Pi_{\mathcal{C}}$ is an Euclidean projection onto a closed and nonempty convex set \mathcal{C} defined by $\Pi_{\mathcal{C}}(x) := \operatorname{argmin}_{y \in \mathcal{C}} \{1/2\|y - x\|^2\}$. As a gradient (subgradient) based algorithm, it has an oracle complexity independent from the dimension [13]. When the decision variable is high dimensional, the convergence rate does not get slower even in high-dimensional settings. This characteristic has led to its widespread application in various fields, including large-scale machine learning and high-dimensional statistics.

Theoretically, this algorithm is suitable for any constrained problems, as long as the projection is well-defined. But in practice, the form of projection operator plays a crucial role and needs to be carefully considered. For simple convex sets, like box (rectangle) sets, affine sets and balls, the computation of the projection operator is relatively straightforward. But for some complicated sets, the form of the projection operator may be implicit. A typical example is unit simplex. In \mathbb{R}^m , it is defined by

$$\mathcal{S}_m := \left\{ (x_1, x_2, \dots, x_m) \mid x_i \geq 0, \sum_{i=1}^m x_i = 1 \right\}.$$

Indeed, the calculation of its corresponding Euclidean projection $\Pi_{\mathcal{S}}$ is challenging. Based on the induction in [15] (also see the Appendix), it can be written as

$$\Pi_{\mathcal{S}_m}(x) := x - \operatorname{argmin}_y \left\{ \max_{1 \leq i \leq m} \{y_i\} + \frac{1}{2}\|y - x\|^2 \right\}.$$

Unfortunately, the minimization problem contained in the expression does not have an explicit solution. If applying this projection operator in the subgradient algorithms, we have to solve a nested programming problem at each step k . For \mathbb{R}^m , the paper [15] offers a quick and implementable algorithm to project an arbitrary point onto a unit simplex. The time complexity of this algorithm is primarily determined by the internal ascending sorting procedure. In computer science and combinatorial

optimization theory, common sorting algorithms based on swapping, such as quick sort, merge sort, or bucket sort, have a lower-bound time complexity of $m \log(m)$. While there have been proposals for non-swapping or parallel sorting algorithms, the fundamental limitations of sorting algorithms remain challenging to overcome. When dealing with a high-dimensional vector, this procedure can be time-consuming and restricts the performance of algorithms. Our motivation improving the algorithm comes from this issue. We seek to find a better approach to avoid this problem.

The existing mirror descent algorithm offers some inspiration, although it is designed for single-block problems. It enhances the subgradient projected algorithm by applying the concept of Bregman distance. Precisely, they substitute the Euclidean projection operators in the algorithm by Bregman ones. Bregman projections have some exclusive properties that are helpful to computation. For instance, a Bregman projection onto a unit simplex is considerably simpler than a Euclidean one [5]. This advantage helps to mitigate the shortcomings of traditional projected methods.

Our contribution is to extend this algorithm to two-block situations and establish the convergence with different assumptions. The paper is organized as the follows:

Notation: Throughout this paper, we use \mathbb{R}^m to denote the m -dimensional Euclidean space equipped with the inner product defined by $\langle x, y \rangle := x^\top y$. The norm is denoted by $\|\cdot\|$ and the induced dual norm is $\|\cdot\|_*$. The norms most frequently used in this paper are l_p norms defined by $\|x\|_p = (\sum_{i=1}^m |x_i|^p)^{1/p}$, where $x = (x_1, x_2, \dots, x_m)$ and $p \geq 1$. When $p = 2$, the norm becomes a Euclidean norm $\|x\|_2 = \sqrt{\langle x, x \rangle}$. In the rest of the paper, if the type of the norm is not specified, we will default to $\|\cdot\|_2$. The nonnegative orthant is $\mathbb{R}_+^m := \{x \in \mathbb{R}^m | x_i \geq 0, i = 1, 2, \dots, m\}$ and the positive orthant is $\mathbb{R}_{++}^m := \{x \in \mathbb{R}^m | x_i > 0, i = 1, 2, \dots, m\}$. More generally, a closed box is denoted by $[a, b]^m := \{x \in \mathbb{R}^m | a \leq x_i \leq b, i = 1, 2, \dots, m\}$ and an open box is denoted by $(a, b)^m := \{x \in \mathbb{R}^m | a < x_i < b, i = 1, 2, \dots, m\}$. The half-open boxes can be represented by $[a, b)^m := \{x \in \mathbb{R}^m | a \leq x_i < b, i = 1, 2, \dots, m\}$ and $(a, b]^m := \{x \in \mathbb{R}^m | a < x_i \leq b, i = 1, 2, \dots, m\}$. Suppose that \mathcal{C} is a set in \mathbb{R}^m . The notations $\text{int } \mathcal{C}$ and $\text{bd } \mathcal{C}$ are the interior and boundary of \mathcal{C} . In term of an extended-real-valued, proper and lower semicontinuous function $f : \mathbb{R}^m \rightarrow \bar{\mathbb{R}} = (-\infty, \infty]$ in variational analysis, $\text{dom } f$ and $\text{epi } f$ are its effective domain and epigraph.

2 Bregman distance and related concepts

The Bregman distance was initially introduced by L.M. Bregman's paper [12] in 1967. It is a distance-like function that is nonnegative but usually not symmetric. Similar to a distance, it satisfies a generalized Pythagorean identity. Based on that, it is possible to define a series of concepts, like projection and Pythagorean theorem, and apply them to the algorithms. Bregman distance-based algorithms sometimes exhibit specific advantages compared to the traditional Euclidean distance. An illustrative example is "mirror descent" [5], where the Euclidean projection operator in the subgradient projected method is replaced by a Bregman one. When the set associated to the operator is a unit simplex, Bregman projection has a explicit form expression while the Euclidean does not. This characteristic makes mirror decent more efficient than traditional algorithms. In this section, we will introduce in detail.

2.1 Bregman distance and basic properties

A Bregman distance is a distance-like function with two input variables associated to a function $\psi : \mathcal{I} \subset \mathbb{R}^m \rightarrow \mathbb{R}$. The formal definition (also see [31, Chapter 26]) is

Definition 4 (Bregman distance) Let $\psi : \mathcal{I} \rightarrow \mathbb{R}$ be a continuously differentiable and strictly convex function on $\text{int } \mathcal{I}$ with $\|\nabla \psi(x^k)\| \rightarrow \infty$ for any sequence $\{x^k\} \subset \text{int } \mathcal{I} \rightarrow \text{bd } \mathcal{I}$ when $k \rightarrow \infty$. Then a function $B_\psi : \mathcal{I} \times \text{int } \mathcal{I} \rightarrow \mathbb{R}$ defined by

$$B_\psi(x, y) = \psi(x) - \psi(y) - \langle \nabla \psi(y), x - y \rangle,$$

for any $x \in \mathcal{I}$ and $y \in \text{int } \mathcal{I}$ is called Bregman distance induced by ψ . Such a function ψ is also a “Legendre function” [1].

The function ψ is also called the kernel of Bregman distance sometimes. Since ψ is a convex function, one can verify that B_ψ is nonnegative. Another important property is that when $\psi(x) = 1/2\|x\|^2$, B_ψ reduces to the square of a Euclidean distance. Therefore, B_ψ is a generalization of Euclidean distance. But for most other situations, B_ψ is not always symmetric, i.e. $B_\psi(x, y) \neq B_\psi(y, x)$. Similar to triangle inequality for Euclidean distance, Bregman distance also satisfies a relation called the generalized Pythagorean theorem. The proof can be seen in [5, Lemma 4.1].

Proposition 1 (Generalized Pythagorean theorem) Let $\mathcal{I} \subset \mathbb{R}^m$ be a closed set. If $\psi : \mathcal{I} \rightarrow \mathbb{R}$ is a continuously differentiable function on \mathcal{I} , then we have

$$B_\psi(c, a) + B_\psi(a, b) - B_\psi(c, b) = \langle \nabla \psi(b) - \nabla \psi(a), c - a \rangle,$$

for all points $a, b \in \text{int } \mathcal{I}$ and $c \in \mathcal{I}$.

A lot of distance-like functions are actually Bregman distances. They are applied in many fields related to science or engineering used to measure the difference of two objects. The following three examples are just some samples of them.

Example 3 (Squared Mahalanobis distance) When the kernel $\psi_Q(x) = 1/2\langle x, Qx \rangle$ is a quadratic function and $\mathcal{I} = \mathbb{R}^m$, where Q is a positive definite matrix, the corresponding Bregman distance $B_{\psi_Q}(x, y) = 1/2\langle x - y, Q(x - y) \rangle$ is called squared Mahalanobis distance [14]. When $Q = I$, it reduces to the squared Euclidean distance.

Example 4 (Generalized Kullback–Leibler divergence) When the kernel $\psi_e(x) = \sum_{i=1}^m x_i \log(x_i)$ is Boltzmann–Shannon entropy and $\mathcal{I} = \mathbb{R}_+^m$, the corresponding Bregman distance $B_{\psi_e}(x, y) = \sum_{i=1}^m x_i \log(x_i/y_i) - \sum_{i=1}^m (x_i - y_i)$ is called generalized Kullback–Leibler (KL) divergence [23]. When $x, y \in \mathcal{S}_m$, it turns into Kullback–Leibler divergence.

Example 5 (Itakura–Saito distance) When the kernel $\psi_b(x) = -\sum_{i=1}^m \log(x_i)$ is a Burg’s entropy and $\mathcal{I} = \mathbb{R}_{++}^m$, the corresponding Bregman distance $B_{\psi_b}(x, y) = \sum_{i=1}^m (x_i/y_i - \log(x_i/y_i) - 1)$ is called Itakura–Saito distance [20].

Figure 1 illustrates intuitively how the Bregman distances differ from Euclidean distance. What appear in the figure are three unit balls centered at $(3, 3)$ expressed by $B_\psi((x, y), (3, 3)) = 1$. However, it can be observed from the figure that Bregman balls are irregular shapes instead of a circle like Euclidean distance.

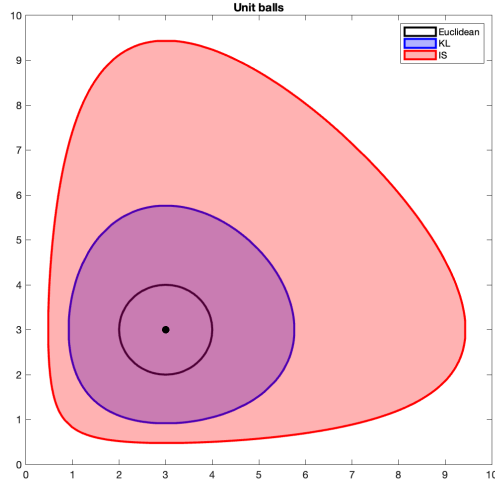


Fig. 1 The unit balls in \mathbb{R}^2 centered at $(3,3)$ based on Euclidean, Kullback-Leibler and Itakura-Saito distances.

2.2 Bregman projection operator

When we discussed projection operators in the last section, the definition relied on the “distance” in a Hilbert space. Now, we extend this concept to a more general situation.

Definition 5 (Bregman projection operator) Let $\mathcal{C} \subset \mathbb{R}^m$ be a nonempty and closed convex set. Let $B_\psi : \mathcal{I} \times \text{int } \mathcal{I} \rightarrow \mathbb{R}$ be a Bregman distance. For any $x \in \text{int } \mathcal{I}$, Bregman projection onto \mathcal{C} is defined by

$$\tilde{\Pi}_{\mathcal{C}}(x) := \arg \min_{y \in \mathcal{C} \cap \mathcal{I}} \{B_\psi(y, x)\}.$$

Clearly, when ψ is a half squared l_2 norm, it reduces to the projection operator under Euclidean distance definition. This type of projection operator is sometimes named “right Bregman projection operator” while the left one exchanges the positions of x and y in the Bregman distance.

Acknowledgment

References

1. H.H. Bauschke and J.M Borwein: Legendre functions and the method of random Bregman projections. *Journal of Convex Analysis* **4**(1) 27-67 (1997)
2. H.H. Bauschke, X. Wang, J. Ye and X. Yuan: Bregman distances and Chebyshev sets. *Journal of Approximation Theory* **159**(1) 3-25 (2009)

3. H.H. Bauschke, J. Bolte and M. Teboulle: A descent lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications. *Mathematics of Operations Research* **42(2)** 330-348 (2017)
4. M.S. Bazaraa, H.D. Sherali and C.M. Shetty: *Nonlinear Programming: theory and algorithms*. John Wiley and Sons (2013)
5. A. Beck and M. Teboulle: Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters* **31(3)** 167-175 (2003)
6. A. Beck and L. Tetruashvili: On the convergence of block coordinate descent type methods. *SIAM Journal on Optimization* **23(4)** 2037-2060 (2013)
7. A. Beck: *First-order methods in optimization*. Society for Industrial and Applied Mathematics (2017)
8. A. Ben-Tal, T. Margalit and A. Nemirovski: The ordered subsets mirror descent optimization method with applications to tomography. *SIAM Journal on Optimization* **12(1)** 79-108 (2001)
9. D.P. Bertsekas: *Nonlinear Programming*, 2nd Ed. Athena Scientific, Belmont (1999)
10. J. Bolte, S. Sabach and M. Teboulle: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming* **146(1)** 459-494 (2014)
11. S. Boyd and L. Vandenberghe: *Convex Optimization*. Cambridge University Press (2004)
12. 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)
13. S. Bubeck: *Convex Optimization: Algorithms and Complexity*. Foundations and Trends® in Machine Learning (2015)
14. M.P. Chandra: On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India* **2(1)** (1936)
15. Y. Chen and X. Ye: Projection onto a simplex. *arXiv preprint arXiv:1101.6081* (2011)
16. G.B. Dantzig: Maximization of a linear function of variables subject to linear inequalities. *Activity Analysis of Production and Allocation* **13** 339-347 (1951)
17. I.I. Dikin: Iterative solution of problems of linear and quadratic programming. *Doklady Akademii Nauk* **174(4)** (1967)
18. C. Emmanuel and T. Tao: The Dantzig selector: Statistical estimation when p is much larger than n . *The Annals of Statistics* **35(6)** 2313-2351 (2007)
19. B.R. Gaines, J. Kim and H. Zhou: Algorithms for fitting the constrained lasso. *Journal of Computational and Graphical Statistics* **27(4)** 861-871 (2018)
20. F. Itakura and S. Saito: Analysis synthesis telephony based on the maximum likelihood method. *Reports of the 6th International Congress on Acoustics* (1968)
21. C. Kan and W. Song: The Moreau envelope function and proximal mapping in the sense of the Bregman distance. *Nonlinear Analysis: Theory, Methods and Applications* **75(3)** 1385-1399 (2012)
22. T.G. Kolda and B.W. Bader: Tensor decompositions and applications. *SIAM Review* **51(3)** 455-500 (2009)
23. S. Kullback and R.A. Leibler: On information and sufficiency. *The Annals of Mathematical Statistics* **22(1)** 79-86 (1951)
24. Y. LeCun et al: Backpropagation applied to handwritten zip code recognition. *Neural Computation* **1(4)** 541-551 (1989)
25. C. Lemaréchal: Cauchy and the gradient method. *Documenta Mathematica Extra Volume* **251(254)** 10 (2012)
26. Z. Lu, T.K. Pong and Y. Zhang: An alternating direction method for finding Dantzig selectors. *Computational Statistics and Data Analysis* **56(12)** 4037-4046 (2012)
27. V. Nair and G.E. Hinton: Rectified linear units improve restricted Boltzmann machines. *ICML* (2010)
28. A. Nitanda: Stochastic proximal gradient descent with acceleration techniques. *Advances in Neural Information Processing Systems* **27** (2014)
29. D. Noll: Convergence of non-smooth descent methods using the Kurdyka-Łojasiewicz inequality. *Journal of Optimization Theory and Applications* **160(2)** 553-572 (2014)
30. P. Ochs, Y. Chen, T. Brox and T. Pock: iPiano: Inertial proximal algorithm for nonconvex optimization. *SIAM Journal on Imaging Sciences* **7(2)** 1388-1419 (2014)
31. R.T. Rockafellar. *Convex Analysis*. Princeton University Press, New Jersey (1970)
32. R.T. Rockafellar and R. J.-B. Wets: *Variational Analysis*. Springer, New York (1998)
33. R. Shefi and M. Teboulle: On the rate of convergence of the proximal alternating linearized minimization algorithm for convex problems. *EURO Journal on Computational Optimization* **4(1)** 27-46 (2016)

34. R. Tibshirani: Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58(1)** 267-288 (1996)
35. R. Tibshirani, M. Saunders, S. Rosset, J. Zhu and K. Knight: Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67(1)** 91-108 (2005)
36. R.J. Tibshirani and J. Taylor: The solution path of the generalized lasso. *The Annals of Statistics* **39(3)** 1335-1371 (2011)
37. P. Tseng: Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of optimization theory and applications* **109(3)** 475 (2001)
38. C.F. Van Loan and G. Golub: *Matrix Computations*. Johns Hopkins Studies in Mathematical Sciences (1996)
39. O. Vu-Thanh, N. Gillis and F. Lecron: Bounded Simplex-Structured Matrix Factorization. ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (2022)
40. Y. Wang and Y. Zhang: Nonnegative matrix factorization: A comprehensive review. *IEEE Transactions on Knowledge and Data Engineering* **25(6)** 1336-1353 (2012)
41. S. Wright: Coordinate descent algorithms. *Mathematical Programming* **151(1)** 3-34 (2015)
42. Y. Xu and W. Yin: A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on Imaging Sciences* **6(3)** 1758-1789 (2013)
43. 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)
44. H. Zou and T. Hastie: Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: series B (Statistical Methodology)* **67(2)** 301-320 (2005)
45. Y. Zhu: An augmented ADMM algorithm with application to the generalized lasso problem. *Journal of Computational and Graphical Statistics* **26(1)** 195-204 (2017)