Journal of Computational Mathematics
Vol.42, No.3, 2024, 679–704.
http://www.global-sci.org/jcm
doi:10.4208/jcm.2204-m2021-0288

SPARSE RECOVERY BASED ON THE GENERALIZED ERROR FUNCTION*

Zhiyong Zhou
Department of Statistics and Data Science, and Institute of Digital Finance, Hangzhou City University, Hangzhou 310015, China
Email: zhiyongzhou@zucc.edu.cn

Abstract

In this paper, we offer a new sparse recovery strategy based on the generalized error function. The introduced penalty function involves both the shape and the scale parameters, making it extremely flexible. For both constrained and unconstrained models, the theoretical analysis results in terms of the null space property, the spherical section property and the restricted invertibility factor are established. The practical algorithms via both the iteratively reweighted ℓ_1 and the difference of convex functions algorithms are presented. Numerical experiments are carried out to demonstrate the benefits of the suggested approach in a variety of circumstances. Its practical application in magnetic resonance imaging (MRI) reconstruction is also investigated.

Mathematics subject classification: 94A12, 94A20.
Key words: Sparse recovery, Generalized error function, Nonconvex regularization, Iterative reweighted L1, Difference of convex functions algorithms.

1. Introduction

High dimensionality is a basic feature of big data, that is, the number of features measured can be very large and are often considerable larger than the number of observations. To overcome the "curse of dimensionality" and reduce the redundancy, the vector of parameters to be estimated or the signal to be recovered is often assumed to be sparse (i.e., it has only a few nonzero entries) either by itself or after a proper transformation. How to exploit the sparsity to help estimating the underlying vector of parameters or recovering the unknown signal of interest, namely sparse recovery, has become a core research issue and gained immense popularity in the past decades [1].

Generally, sparse recovery aims to estimate an unknown sparse $\mathbf{x} \in \mathbb{R}^N$ from few noisy linear observations or measurements $\mathbf{y} = A\mathbf{x} + \boldsymbol{\varepsilon} \in \mathbb{R}^m$ where $A \in \mathbb{R}^{m\times N}$ with $m \ll N$ is the design or measurement matrix, and $\|\boldsymbol{\varepsilon}\|_2 \leq \eta$ is the vector of noise. It arises in many scientific research fields including high-dimensional linear regression [2] and compressed sensing [3–5]. Naturally, this sparse $\mathbf{x}$ can be recovered by solving a constrained ℓ_0-minimization problem

$$\min_{\mathbf{z}\in\mathbb{R}^N} \|\mathbf{z}\|_0 \quad \text{subject to} \quad \|A\mathbf{z} - \mathbf{y}\|_2 \leq \eta, \tag{1.1}$$

or the unconstrained ℓ_0-penalized least squares problem $\min_{\mathbf{z}\in\mathbb{R}^N} \frac{1}{2}\|A\mathbf{z}-\mathbf{y}\|_2^2+\lambda\|\mathbf{z}\|_0$ [6], where $\lambda > 0$ is a tuning parameter. However, due to the nonsmoothness and nonconvexity of the ℓ_0-norm, these are combinatorial problems which are known to be related to the selection of best

* Received October 11, 2021 / Revised version received February 16, 2022 / Accepted May 6, 2022 / Published online April 3, 2023 /

subset and are computationally NP-hard to solve [7]. Instead, a widely used solver is the following constrained ℓ_1-minimization problem (also called Basis Pursuit Denoising) [3]:

$$\min_{\mathbf{z}\in\mathbb{R}^N} \|\mathbf{z}\|_1 \quad \text{subject to} \quad \|A\mathbf{z}-\mathbf{y}\|_2 \leq \eta, \tag{1.2}$$

or the well-known Lasso $\min_{\mathbf{z}\in\mathbb{R}^N} \frac{1}{2}\|A\mathbf{z}-\mathbf{y}\|_2^2+\lambda\|\mathbf{z}\|_1$ [8]. The ℓ_1-minimization acts as a convex relaxation of ℓ_0-minimization. Although it enjoys attractive theoretical properties and has achieved great success in practice, it is biased and suboptimal. The Lasso does not have the oracle property [9] (means that in the asymptotic sense it performs as well as the case when the support is known in advance), since the ℓ_1-norm is just a loose approximation of the ℓ_0-norm.

To remedy this problem, many nonconvex sparse recovery methods have been employed to better approximate the ℓ_0-norm and enhance sparsity. They include ℓ_p $(0 < p < 1)$ [10–12], smoothed L0 (SL0) [13], Capped-L1 [14], transformed ℓ_1 (TL1) [15], smooth clipped absolute deviation (SCAD) [9], minimax concave penalty (MCP) [16], nonconvex shrinkage methods, [17], exponential-type penalty (ETP) [18, 19], error function (ERF) method [20], $\ell_1 - \ell_2$ [21, 22], $\ell_r^r - \alpha\ell_1^r$ $(\alpha \in [0,1], r \in (0,1])$ [23], ℓ_1/ℓ_2 [24,25], q-ratio sparsity minimization [26] and smoothed ℓ_p-over-ℓ_q (SPOQ) [27], among others. For a more comprehensive view, please see the survey on nonconvex regularization [28] and the references therein. And it should be pointed out that all the nonconvex regularization methods we mentioned here are only a small part of this field, because there are too many related studies and they are constantly developing. These parameterized nonconvex methods result in the difficulties of theoretical analysis and computational algorithms due to the nonconvexity of the penalty functions, but do outperform the convex ℓ_1-minimization in various scenarios. For example, it has been reported that ℓ_p gives superior results for incoherent measurement matrices (i.e., matrices with small coherence such as Gaussian random matrices), while $\ell_1 - \ell_2$, ℓ_1/ℓ_2 and q-ratio sparsity minimization are better choices for highly coherent measurement matrices (e.g., oversampled discrete cosine transform matrices). Meanwhile, TL1 is a robust choice no matter whether the measurement matrix is coherent or not.

The resulting nonconvex sparse recovery methods have the following general constrained form:

$$\min_{\mathbf{z}\in\mathbb{R}^N} R_\theta(\mathbf{z}) \quad \text{subject to} \quad \|A\mathbf{z}-\mathbf{y}\|_2 \leq \eta, \tag{1.3}$$

where $R_\theta(\cdot) : \mathbb{R}^N \to \mathbb{R}_+ := [0,\infty)$ denotes a nonconvex penalty or regularization function with an approximation parameter θ (it can be a vector of parameters), or a formulation of its corresponding unconstrained version $\min_{\mathbf{z}\in\mathbb{R}^N} \frac{1}{2}\|A\mathbf{z}-\mathbf{y}\|_2^2+\lambda R_\theta(\mathbf{z})$ with $\lambda > 0$ being the tuning parameter. Several nonconvex methods and their corresponding penalty functions are shown in Table 1.1. Basically, in practice a separable and concave on $\mathbb{R}_+^N$ penalty function is desired in order to facilitate the theoretical analysis and solving algorithms.

There are a great number of theoretical recovery results for the nonconvex methods listed above for their global or local optimal solutions (please refer to the original paper for a specific method). Moreover, some unified recovery analysis results have also been obtained for these nonconvex sparse recovery methods. For instance, when $R_\theta(\mathbf{z}) = \sum_{j=1}^N F_\theta(|z_j|)$ with $F_\theta(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying the subadditive property, both exact and robust reconstruction conditions were obtained in [29]. While, [30] established a theoretical recovery guarantee through unified null space properties. A general theoretical framework was presented in [31] based on regularity conditions. It shows that under appropriate conditions, the global solution of concave