Front. Comput. Sci., 2017, 11(3): 362–391
DOI 10.1007/s11704-016-5552-0

REVIEW ARTICLE

Optimization methods for regularization-based ill-posed problems: a survey and a multi-objective framework

Maoguo GONG (✉), Xiangming JIANG, Hao LI

Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an 710071, China

Abstract Ill-posed problems are widely existed in signal processing. In this paper, we review popular regularization models such as truncated singular value decomposition regularization, iterative regularization, variational regularization. Meanwhile, we also retrospect popular optimization approaches and regularization parameter choice methods. In fact, the regularization problem is inherently a multi-objective problem. The traditional methods usually combine the fidelity term and the regularization term into a single-objective with regularization parameters, which are difficult to tune. Therefore, we propose a multi-objective framework for ill-posed problems, which can handle complex features of problem such as non-convexity, discontinuity. In this framework, the fidelity term and regularization term are optimized simultaneously to gain more insights into the ill-posed problems. A case study on signal recovery shows the effectiveness of the multi-objective framework for ill-posed problems.

Keywords ill-posed problem, regularization, multi-objective optimization, evolutionary algorithm, signal processing

1 Introduction

The idea of ill-posed problem can be traced back to Jacques Hadamard's research [1] about one century ago. In his opinion, a problem is well-posed if its solution is *existent, unique and stable*. Otherwise, a problem which lacks any of these attributes is ill-posed. Most of the inverse problems in physics, geophysics, medicine, ecology, engineering and economics are ill-posed [2].

In this paper, we concentrate mainly on the optimization methods for regularization-based ill-posed problems within signal processing. A lot of inverse problems in signal processing are ill-posed such as signal recovery and convolution equation with analytic kernel [3]. Nowadays, researchers generally use sophisticated methods from geometry, differential geometry, convex analysis and numerical analysis to handle ill-posed problems in signal processing [4].

The key of handling ill-posed problems is to overcome instability of their solutions. By now, theoretical strategies and numerical tools have been developed to mitigate this instability, and *regularization theory* has been gradually formed as a mainstream. The idea of regularization is to replace the ill-posed problem pointing an oscillating solution with a well-posed problem for an approximate solution depending continuously on the given data [5]. Regularization aims at preventing over-fit to the given data by introducing priori information of problems to penalize the model. The priori information is usually a penalty for complexity, such as restrictions for smoothness in total variational regularization [6] and bounds on the norm of the space vector in Tikhonov regularization [7, 8]. In order to obtain better compromise between the loss term and the penalty term, the regularization parameters are introduced naturally to adjust the dominator in these two terms according to the application demands. Nowadays, the idea of regularization is used in many fields of science. For example, the least squares fit-to-data model:

Received December 21, 2015; accepted June 16, 2016

E-mail: gong@ieee.org

$\min_{\boldsymbol{u}} \|\boldsymbol{Au}-\boldsymbol{f}\|^2$ can be viewed as a loss term of the ill-posed problem: $\boldsymbol{Au}=\boldsymbol{f}$, where $\boldsymbol{A}$ is an ill-conditioned operator. By using appropriate iterative methods [9, 10], the least square fit-to-data model can be solved to approximate the original ill-posed model effectively. From the plot of the solution error norm against the iteration count in Ref. [11], we are informed that the iteration count plays the same role as the regularization parameter in iterative solving process, this is known as iterative regularization. There are also some other regularization ideas such as regularization by filtering (TSVD observably) [12].

Regularization ideas simply reform the original ill-posed model into regularization model by introducing regularization terms, but regularization model generally has an optimizing expression (see variational regularization models [6, 13]) which needs to be solved by effective optimization methods. Optimization methods used in regularization models can be roughly classified into three types: 1) methods utterly depending on differential characteristics of system such as gradient-based methods, 2) methods using a few about differential characteristics of system such as iterative-based methods, 3) methods using nothing about differential characteristics of system such as intelligent optimization algorithms.

Regularization models always need to get a trade-off between fidelity term and penalty term by adjusting the value of regularization parameter [14], but what compromise is the better decision we yearn for? In traditional solving process of regularization models, the regularization parameter must be selected at the very beginning, then solution corresponding to the fixed regularization parameter can be obtained by optimization methods. Once the regularization parameter is determined, a compromise is following. Thus, the decision-making (compromise) is prior to the calculation of solution. By contrast, we are more inclined to make decision from a series of compromises according to the actual application needs, namely, the decision-making is posterior to the calculation of solutions. Multi-objective optimization (MOO) happens to achieve this goal.

Many real-world problems have several objectives to be optimized at the same time [15]. For example, the battery is hoped to have huge volume and small size simultaneously, not simply huge volume or small size, for only one of these two properties is of little significance in real production. MOO has attracted great concern in recent two decades [15]. Evolutionary algorithms (EAs) are well suited to MOO because they push a set of current solutions forward simultaneously. The resultant methods are termed as evolutionary multi-objective optimization (EMO). The ability to handle complex problems [16], involving features such as discontinuities, non-convexity, non-convex feasible region, and noisy function evaluations, reinforces the underlying effectiveness of EAs in MOO [17–21]. The main advantages of EMO are as follows: firstly, EMO can optimize multiple objectives of real world problems simultaneously instead of scalarizing them together. Secondly, the regularization problem is essentially a multi-objective problem. Because multiple objectives can not be well handled by traditional optimization methods, they are usually reformulated as an unconstraint single-objective optimization problem. This treatment will lead to negative issues as described in previous paragraph. EMO can avoid these issues by obtaining a set of nondominated solutions which are mathematically equal in a single run, so the decision maker will have more options from these nondominated solutions according to their preference [22].

In this paper, we firstly explain the ill-posedness reasons for the original forward model and the modeling process of its regularization. Thereafter, we review some popular regularization models and corresponding optimization methods. Afterwards, a multi-objective framework for ill-posed problems is established under some in-depth analysis on regularization model. Finally, a case study of sparse reconstruction based on the proposed multi-objective framework is conducted to illustrate the effectiveness of EMO for signal recovery.

This paper is organized as follows: the second section shows a purpose-driven regularization modeling process for ill-posed problem. In the third section, we summarize popular regularization models and corresponding optimization strategies. In the fourth section, we establish a multi-objective framework for ill-posed problems. A case study is revealed in the fifth section. Conclusions and prospects are following in the final section.

2 The ill-posed forward model and its regularization

In this section, we mainly explain why and how to construct a regularization model for an original forward problem. Firstly, we briefly explain reasons for ill-posedness of the original forward model. Then, an example explaining the formulate process of regularization model is given under some assumptions about the solution and noise. Finally, we roughly analyze effects of the regularization parameters on the final solution.

2.1 Regularization model

Inverse problems can be generally formulated as:

$$\boldsymbol{y} = \boldsymbol{F}(\boldsymbol{x}) + \boldsymbol{\eta}, \tag{1}$$

where $\boldsymbol{y}$ is used to represent the collected noised data, $\boldsymbol{x}$ denotes the ideal quantities (desired solution) we pursue for, $\boldsymbol{\eta}$ represents the noise and other errors in collection process of $\boldsymbol{y}$, and $\boldsymbol{F}(\cdot)$ is related to data collection process and assumed to be continuous.

The lack of effective observation and bad algebraic spectrum distribution related to $\boldsymbol{F}(\cdot)$ may be reasons for ill-posedness of Eq. (1). An improper observation also leads to undetermined system whose solution is not unique or nonexistent, these characters are determined by algebraical structure of the matrix Eq. (1). Although the solution of Eq. (1) exists, it may oscillate frequently because of the bad algebraic spectrum distribution related to $\boldsymbol{F}(\cdot)$. In the actual date collection process, $\boldsymbol{F}(\cdot)$ is a continuous function, but we are more willing to tackle with the discrete form of problem Eq. (1), in consideration of computation convenience. However, the discretization of Eq. (1) will make components of solution more sensitive to the given data, and thus the discrete model is more likely to be ill-posed. Therefore, regularization ideas are needed to overcome these problems mentioned above.

Regularization, in a narrow sense, is a term introduced to inhibit oscillation of solution and prevent over-fit to noisy data. This term is usually termed as *penalty term*, which incarnates the prior information about the unknown data $\boldsymbol{x}$ with some essential characteristics (such as smoothness, sparseness or bounds on the vector space norm). By combining with a fidelity term, the general regularization model can be formulated as:

$$\min_{\boldsymbol{x}} : E(\boldsymbol{x}) = L\left[\boldsymbol{F}(\boldsymbol{x}), \boldsymbol{y}\right] + \lambda P(\boldsymbol{x}), \tag{2}$$

where the *fidelity term* $L\left[\boldsymbol{F}(\boldsymbol{x}), \boldsymbol{y}\right]$ evolved from the original forward model Eq. (1) forces $\boldsymbol{F}(\boldsymbol{x})$ to inherit most features from the noisy data $\boldsymbol{y}$, $\lambda > 0$ is a regularization parameter that controls trade-off between these two terms [23]. Occasionally, more than one penalty terms are structured in Eq. (2) for different intrinsic characteristics of data $\boldsymbol{x}$ (see Ref. [24]).

2.2 Formulate process of the regularization model: an example

We are going to explain the purpose-driven modeling process via specific example. Assuming that $\boldsymbol{F}(\boldsymbol{x})$ is a linear operator, i.e., $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{A}$ is a constant matrix. $\boldsymbol{\eta}$ is a discrete Gaussian random vector with zero mean and covariance $\sigma^2\boldsymbol{I}_n$, where σ^2 is the noise variance. If we are informed from the priori information that the solution should be smooth, what should we do to modify the original forward model? From the assumptions above and Eq. (1), we deduce that the desired solution satisfy the following equation:

$$\begin{aligned} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|^2 &= \left\|\boldsymbol{\eta}\right\|^2 \\ &= \boldsymbol{\eta}'\boldsymbol{\eta} = E\left[(\boldsymbol{\eta}' - \boldsymbol{E}\boldsymbol{\eta}')(\boldsymbol{\eta} - \boldsymbol{E}\boldsymbol{\eta})\right] \\ &= E(\boldsymbol{\eta}'\boldsymbol{\eta}) = n\sigma^2. \end{aligned} \tag{3}$$

The above equation can be regarded as constraint on the desired solution. Under this circumstance, we can design proper regularization matrix to achieve the smoothness of the desired solution. Firstly, the smoothness of the solution can be achieved via exerting restraints on its derivative. Secondly, the discrete approximation of the derivative operator can be modeled in form of matrix. Therefore, we can achieve smoothness of the desired solution via a matrix, which is called regularization matrix.

We can simply choose the identity matrix as the regularization matrix ($\boldsymbol{L}_0 = \boldsymbol{I}_n$) to control the magnitude of the components of the solution, but this strategy is powerless when it refers to smoothness of the solution. At this time, the possible forms for the regularization matrix can be

$$\boldsymbol{L}_1 = \begin{bmatrix} -1 & 1 & \cdots & 0 & 0 \\ 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & -1 \end{bmatrix}, \text{ or } \boldsymbol{L}_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{bmatrix}. \tag{4}$$

$\boldsymbol{L}_1$ is called the first-order difference regularization matrix. The use of discrete approximation of derivative operator instead of the identity matrix is under the following considerations: the rough oscillation of individual components of the solution caused by noisy data will provoke large $\|\boldsymbol{L}_1\cdot\|_2$, but not much for the standard norm $\|I\cdot\|_2$, and thus the former is more effective in restraining solution oscillation.

However, the discrete approximation of derivative operators $\|\boldsymbol{L}_1\cdot\|$ does not necessarily perform better than the standard norm $\|I\cdot\|$ under l_2-norm in restraining oscillation of the solution in some specific cases. Thus, for comprehensive consideration, we can construct the regularization matrix from the Cholesky factorization of several linearly combined derivative approximation matrices [25]:

$$\boldsymbol{L}'\boldsymbol{L} = \sum\nolimits_{i=1}^{n} \omega_i \boldsymbol{L}_i'\boldsymbol{L}_i, \tag{5}$$

where $\omega_i,\ i = 1, 2, \ldots, n$ are positive weighting factors satisfying $\sum_{i=1}^{n} \omega_i = 1$. $\boldsymbol{L}$ can also be constructed using other derivative approximation matrices such as the second-order difference regularization matrix, etc., or via other approaches such as statistical information [26].

Once the regularization matrix is constructed, the regularization model can be established as a constrained optimization problem:

$$\min_{\boldsymbol{x}} : \|\boldsymbol{L}\boldsymbol{x}\|^2, \text{ subject to } \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|^2 = n\sigma^2. \tag{6}$$

The constraint is deduced in Eq. (3) according to the priori information about the type and level of noise. The objective function in minimization problem Eq. (6) can characterize smoothness of the solution. The smaller value the function has, the stronger smoothness the solution will have. Therefore, we can derive a smooth solution satisfying the noisy constraint from regularization model Eq. (6). Except for the regularization matrix, one can construct a total variational term to pursue an edge-preserving smooth solution [6,27,28]. We can also construct other penalty terms according to the priori knowledge of the solution or noise. For example, a sparse term can be constructed to characterize the sparse degree of the solution if we are informed that the solution is sparse.

An unconstrained minimization problem can be constructed on the basis of the constrained minimization problem described in Eq. (6) using Euler-Lagrange method, its form is

$$\min_{\boldsymbol{x}} : E(\boldsymbol{x}) = \|\boldsymbol{L}\boldsymbol{x}\|^2 + \alpha\left(\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|^2 - n\sigma^2\right), \tag{7}$$

where α is a positive constant that stays to be chosen. This minimization problem is equal to

$$\min_{\boldsymbol{x}} : E(\boldsymbol{x}) = \|\boldsymbol{L}\boldsymbol{x}\|^2 + \alpha\left(\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|^2\right). \tag{8}$$

By contrast between Eq. (6) and Eq. (7) or Eq. (8), we find that the optimal solution of Eq. (6) may not be the optimal solution of Eq. (7) or Eq. (8), i.e., $E(\hat{\boldsymbol{x}}_1) \geqslant E(\hat{\boldsymbol{x}}_2)$ if $\hat{\boldsymbol{x}}_1$ and $\hat{\boldsymbol{x}}_2$ are optimal solution of Eq. (6) and Eq. (7), respectively. Because the optimal solution of Eq. (6) is located in the feasible region of Eq. (7), not the other way around, we can only achieve an approximate solution from Eq. (7) for Eq. (6) by altering parameter α.

In order to comply with people's habits and cognition, we can reformulate Eq. (8) with $\lambda = 1/\alpha$ as follows:

$$\min_{\boldsymbol{x}} : E(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|^2 + \lambda\|\boldsymbol{L}\boldsymbol{x}\|^2. \tag{9}$$

In the above text, we mainly analyze the establishment process of the regularization model from the original forward model and priori knowledge about the solution and noise. The construction process of regularization matrix for specific purpose, the relationship between the original forward model and the regularization model are also illustrated meanwhile. In fact, the penalty term can be constructed in other forms as long as certain purpose can be accomplished. For example, total variational term can achieve smooth solutions as the regularization matrix $\boldsymbol{L}$ does, and it also has its own edge-preserving character.

We can make a rough analysis on Eq. (9): a regularized solution can be obtained once a regularization parameter $\boldsymbol{\lambda}$ is given. If the value of parameter $\boldsymbol{\lambda}$ is large, the penalty term $\|\boldsymbol{L}\boldsymbol{x}\|^2$ will dominate the value of the objective function $E(\boldsymbol{x})$, in order to make the value of $E(\boldsymbol{x})$ as small as possible, the penalty term should be as small as possible which will provide guidance for selection of the solution. However, if $\boldsymbol{\lambda}$ is too large, the regularization model Eq. (9) will have little connection with the original model Eq. (1). Conversely, if the value of parameter $\boldsymbol{\lambda}$ is small, the fidelity term $\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|^2$ will dominate the value of the objective function $E(\boldsymbol{x})$, in order to make $E(\boldsymbol{x})$ as small as possible, the regularized solution $\boldsymbol{x}$ should be selected to make $\boldsymbol{A}\boldsymbol{x}$ close to $\boldsymbol{y}$. In contrast, if $\boldsymbol{\lambda}$ is too small, the solution will be over-fit and contain noise. Therefore, a proper regularization parameter is important for obtaining a precise and stable solution. The analysis of this paragraph can also be seen via the optimality condition of Eq. (9):

$$\begin{aligned} E'(\boldsymbol{x}) &= 2\boldsymbol{A}'(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}) + 2\lambda\boldsymbol{L}'\boldsymbol{L}\boldsymbol{x} \\ &= 2(\boldsymbol{A}'\boldsymbol{A} + \lambda\boldsymbol{L}'\boldsymbol{L})\boldsymbol{x} - 2\boldsymbol{A}'\boldsymbol{y} \\ &= 0. \end{aligned} \tag{10}$$

The regularization model Eq. (2) can be formulated similar to the modeling process of Eq. (9) for different purposes.

3 A review on the regularization models and their optimization methods

In this section, we are going to review popular regularization models and related optimization methods for ill-posed problems.

3.1 A general regularization model for linear ill-posed problems

The regularization models for linear ill-posed problems are much more developed than those for nonlinear ill-posed problems which are highly related to particular applications. What's more, studying linear ill-posed problems is more likely to obtain good intuition and present novel methods. For these reasons, linear ill-posed problems were firstly investigated by researchers, and regularization models and optimization approaches were developed in the meanwhile.

The original problem described in Eq. (1) in the previous

section is a linear problem if function $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{A}$ is a $M \times N$ linear operator. In this case, Eq. (1) becomes

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\eta}. \tag{11}$$

The task of regularization is to work out an approximation of the actual solution using the noised data $\boldsymbol{y}$ and the operator $\boldsymbol{A}$.

3.2 Truncated singular value decomposition (TSVD) regularization

The linear inverse problem described in Eq. (11) commonly belongs to rank-deficient or discrete ill-posed problem [25], which refers to the SVD knowledge of matrix $\boldsymbol{A}$. The solution of Eq. (11) can be represented by the SVD system of $\boldsymbol{A}$ as follows:

$$\boldsymbol{x} = \boldsymbol{A}^{\dagger}\boldsymbol{y} = \sum_{i=1}^{r} \frac{\boldsymbol{u}_i'\boldsymbol{y}\boldsymbol{v}_i}{\sigma_i} = \sum_{i=1}^{r} \frac{\boldsymbol{u}_i'\boldsymbol{y}_{\boldsymbol{exact}}\boldsymbol{v}_i}{\sigma_i} + \sum_{i=1}^{r} \frac{\boldsymbol{u}_i'\boldsymbol{\eta}\boldsymbol{v}_i}{\sigma_i}, \tag{12}$$

where $\boldsymbol{U} = (\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_m}) \in \boldsymbol{R}^{m\times m}$, $\boldsymbol{V} = (\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}) \in \boldsymbol{R}^{n\times n}$ and $\boldsymbol{\Sigma} = diag\left[diag(\sigma_i)_{r\times r}, diag\left(\boldsymbol{0}_{(m-r)\times(n-r)}\right)\right]$ constitute the SVD system $(\boldsymbol{U}, \boldsymbol{\Sigma}, \boldsymbol{V})$ of matrix $\boldsymbol{A}$, and $\boldsymbol{A}^{\dagger}$ is the pseudoinverse of matrix $\boldsymbol{A}$ [29].

From Eq. (12) of solution $\boldsymbol{x}$, we can clearly explain the oscillation reason for the solution. The two terms on right hand side of Eq. (12) will converge if the Picard condition [30]

$$\sum_{i=1}^{\infty} \left[\frac{(\boldsymbol{u}_i, \boldsymbol{y})}{\sigma_i}\right]^2 < \infty, \tag{13}$$

set up. The Picard condition Eq. (13) is identical to the condition that the given data $\boldsymbol{y}$ belongs to the range of $\boldsymbol{A}$, i.e., $\boldsymbol{y} \in R(\boldsymbol{A})$. However, we can deduce that $\boldsymbol{y}_{\boldsymbol{exact}} \in R(\boldsymbol{A})$ due to $\boldsymbol{y}_{\boldsymbol{exact}} = \boldsymbol{A}\boldsymbol{x}_{\boldsymbol{exact}}$, but we cannot expect the noise $\boldsymbol{\eta}$ belongs to the range of $\boldsymbol{A}$. Therefore, $\boldsymbol{y} = \boldsymbol{y}_{\boldsymbol{exact}} + \boldsymbol{\eta} \notin R(\boldsymbol{A})$, the solution $\boldsymbol{x}$ in Eq. (12) is not convergent all because of the oscillation caused by noise $\boldsymbol{\eta}$ especially amplified by the reciprocal of small singular value σ_i.

By avoiding the use of smaller singular values, one can calculate an approximate solution which is "close enough" to the exact solution [31]. The numerical $\varepsilon - rank$ was structured in Ref. [32] to determine the one from which the small singular values should be truncated. Thus, the oscillation of solution caused by noise $\boldsymbol{\eta}$ and amplified by the small singular values can be cut-off. Hansen's investigation [12] suggested that the truncated singular value decomposition can be a favorable alternative to standard-form regularization in the case of ill-conditioned matrix with well-determined rank. Under some extra assumptions about the noised data $\boldsymbol{y}$, the truncated singular value decomposition can still be a substitution of the standard-form regularization in case of ill-conditioned matrix with ill-determined rank. Therefore, the TSVD method is a regularization method for ill-posed problem from a theoretical as well as a practical point of view.

Although the TSVD method can tackle with the standard-form problem: $\min \|\boldsymbol{x}\|_2$ subject to $\|\boldsymbol{A_k}\boldsymbol{x} - \boldsymbol{y}\|_2$ *minimal*, where $\boldsymbol{A_k} \triangleq \sum_{i=1}^{k} \boldsymbol{u}_i\sigma_i\boldsymbol{v}_i^T$, it does powerless for the ill-posed problem with general form suggested by Hansen [33]: $\min \|\boldsymbol{L}\boldsymbol{x}\|_2$ subject to $\|\boldsymbol{A_k}\boldsymbol{x} - \boldsymbol{y}\|_2$ *minimal*. The generalized singular value decomposition (GSVD) proposed by Van Loan [34] can be used to calculate the GSVD of the matrix pair $(\boldsymbol{A}, \boldsymbol{L})$, then the truncated generalized singular value decomposition(TGSVD) [35] is capable to solve this general formal problem. The TGSVD method gives better approximation of the exact solution than TSVD method in the promotion of suitable chosen regularization matrix $\boldsymbol{L}$. However, the difficulty for constructing the regularization matrix and high computational efforts for computing the GSVD of the matrix pair $(\boldsymbol{A}, \boldsymbol{L})$ [36] limit the use of TGSVD method. Hansen proposed another approach namely modified TSVD (MTSVD) [33] combining SVD with approximation in the l_1-norm to avoid the shortcomings of TSVD, this approach is able to detect and provide accurate approximations of discontinuities in the exact solution. By using auxiliary knowledge about the exact solution in a novel and natural way, the TPSVD [37] method can obtain a better approximation of the exact solution than TSVD and TGSVD methods. A series of singular value decomposition generalizations such as ΠSVD [38] and RSVD [39,40] for gradually increased scale of matrices were subsequently proposed. One can find some applications of these generalized SVDs in signal processing in Refs. [41,42]. Recently, a novel modified TSVD method in which the operator $\boldsymbol{A}$ is replaced by a closest matrix in a unitarily invariant matrix norm was also proposed to yield higher quality approximate solution than original TSVD [43]. What's more, a rescaled GSVD aiming at minimizing the condition number of the nonsingular matrix $\boldsymbol{A}$ was put forward, and the derived new truncated GSVD method is competitive with the original GSVD method and Tikhonov regularization method with respect to the quality of computed approximate solution [44].

These methods can not do without an accurate SVD system which is most costly by dense linear algebraic methods. Edo et al. proposed two randomize algorithms which can be used for constructing low-rank approximation of SVD system of matrix $\boldsymbol{A}$ [45]. These two classes of algorithms can converge faster than classical methods while ensuring comparable accuracy. Afterwards, a large amount of researches sprung up for more high-efficiency randomized methods, e.g.,

Refs. [46–48].

The effective performance of the truncated singular value decomposition can not do without a proper criteria for truncation of the singular values. If a relatively larger eigenvalue is truncated, too much information involved in matrix $\boldsymbol{A}$ would be ignored, which leads to an inaccurate solution. Conversely, if a relatively smaller eigenvalue is not truncated, the noise contained in the given data $\boldsymbol{y}$ would be enlarged by the small eigenvalue, and thus a spotted and oscillatory solution might be obtained. Unfortunately, the truncation has been done simply in line with one's intuition (e.g., Refs. [49, 50]) for a long time. The first formal truncation method is to discard the eigenvalues on the basis of the F-statistic with different significance levels [51, 52], but because there is no practical guidelines in choosing proper significance levels, the F-statistic criterion often fails to truncate the eigenvalues and leads to unstable solution. The TSVD method by means of the L-curve criteria [53, 54] is the best in stabilizing the ill-posed problem but easily leads to over-cutting of the eigenvalues, and thus results in an inaccurate solution. By contrast, the quality-based TSVD method by minimizing the MSE of a solution outperforms the above two methods in terms of stability and other two indexes [55]. Recently, some new methods were proposed for choosing proper truncation index, e.g., Refs. [56, 57]. The method suggested in Ref. [56] has shown the least computational cost and the smallest mean square error compared with the GCV [58] and L-curve methods. The vector extrapolation was also used for the choice of truncation index of TSVD [59, 60], the approximate solution obtained by this method is not easy to deteriorate in quality as the truncation index increases.

With a proper truncation criterion, TSVD methods can obtain a regularized approximation of the exact solution easily on the whole.

3.3 Iterative regularization

For an ill-posed problem, the type of regularization, the extra constraints, and a variety of the optimization approaches should be chosen properly depending on the structure of matrix $\boldsymbol{A}$. In some cases such as small to moderately sized linear ill-posed problems, it is wise to use some efficient filtering approaches such as the TSVD in previous subsection, but more often it is favorable to use iterative methods in partial or entire solving process of large-scale problems. Iterative method can be used as a regularization method for the least squares fit-to-data model: $\min_{\boldsymbol{u}} \|\boldsymbol{Au} - \boldsymbol{f}\|^2$ or a pure optimization method for the regularization model described in Eq. (2). In this subsection, we mainly explain how the iterative methods can be used to regularize the ill-posed problem described in Eq. (1) and review some widely used iterative approaches.

From the plot of solution error norm against the iteration count shown in Fig. 1.6 of Ref. [61], numerical error analysis in Chapter 5 of Ref. [62], and Ref. [63], we can find the regularizing properties of iterative approaches: under the discrete Picard condition, over-smoothly approximate solution can be yielded for small value of iterative count; conversely, the solution would be highly oscillatory as the iterative count becomes too large, this phenomenon is known as semi-convergence. Therefore, the iterative count plays the role of regularization parameter, the iterative stop rule plays the role of parameter choice method, and the related schemes are known as iterative regularization methods.

Many iterative approaches can be used to regularize the ill-posed problem, typically, Landweber, semi-iterative, steepest descent, etc. Most iterative regularization methods can be included by a general framework used to solve the minimization problem:

$$\min\ f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x'A'Ax} - \boldsymbol{x'A'y}, \tag{14}$$

the iterative methods have the general form:

$$\boldsymbol{x_{k+1}} = \boldsymbol{x_k} + \rho_k \boldsymbol{M_k}(\boldsymbol{A'y} - \boldsymbol{A'Ax_k}) = \boldsymbol{x_k} + \rho_k \boldsymbol{M_k r_k}, \tag{15}$$

where $\boldsymbol{r_k} = (\boldsymbol{A'y} - \boldsymbol{A'Ax_k})$. By changing the materials ρ_k and $\boldsymbol{M_k}$ in Eq. (15) into specific forms, different iterative approaches can be obtained as follows:

- By taking $\rho_k = \rho$ and $\boldsymbol{M_k} = \boldsymbol{I}$ (the identity matrix), the iteration Eq. (15) represents the Landweber iterative method [9]. This method is always used to analyze the regularization properties of iterative methods due to its simple expression [62]. The regularizing effect of Landweber iterative methods can be reflected by a filtering factor which is determined by the iterative count [25], and thus the iterative count plays the role of regularization parameter. However, the Landweber iterative method is limited in use because of its slow convergent rate.
- By introducing different residual polynomials which are uniformly bounded on $[0, 1]$ and converge pointwise to 0 on $(0, 1]$, a series of semi-iterative methods, such as Chebyshev method of Stiefel [64], Nemirovskii and Polyak [65], the ν-method of Brakhage [66, 67] can be founded. Semi-iterative regularization methods can work more effectively than the classic Landeweber iterative methods, but it requires the scaling of matrix $\boldsymbol{A}$.

- If the matrix $\boldsymbol{M_k}$ is again defined as the identity matrix, and ρ_k is chosen as the minimizer of the problem: $\rho_k = \arg\min_{\rho_k>0} f(\boldsymbol{x_k} + \rho_k \boldsymbol{r_k})$, a steepest descent method will be produced. The convergence rate is upper-bounded related to the condition number of matrix $\boldsymbol{A}$ and the initial iteration point. For ill-conditioned matrix $\boldsymbol{A}$, this method is typically converge slowly unless with proper preconditioning. A two-point step size gradient method with better performance and cheaper computation compared to the steepest descent method was proposed in Ref. [68] at the cost of extra iteration and gradient storage.
- If the matrix is set that $\boldsymbol{M_0} = \boldsymbol{I}$ and $\boldsymbol{M_k} = \boldsymbol{I} - \boldsymbol{s_k h_k'} / \left(\boldsymbol{h_k' s_k}\right)$, where $\boldsymbol{s_k} = \boldsymbol{x_{k+1}} - \boldsymbol{x_k}$, $\boldsymbol{h_k} = \boldsymbol{A'A}(\boldsymbol{x_{k+1}} - \boldsymbol{x_k})$, ρ_k is chosen the same as the steepest descent method, we will obtain the conjugate gradient (CG) method which is a classical Krylov subspace iterative method. This method was initially developed for settling sparse systems of equations with a symmetric positive definite(SPD) coefficient matrix [69,70]. For linear equation $\boldsymbol{Ax} = \boldsymbol{y}$, where $\boldsymbol{A}$ is a SPD matrix, it is general to use a preconditioning $\boldsymbol{MAx} = \boldsymbol{My}$ to improve the convergence rate of this method by handling spectrum of matrix $\boldsymbol{A}$. However, for $\boldsymbol{A}$ is no longer SPD matrix in ill-posed problems, we are more willing to use a preconditioner to improve parts of eigenvalues of matrix $\boldsymbol{A}$ to promote the intrinsic regularizing effect [71, 72] of the CG method. The different implementations of the CG method will lead to different derivative methods such as CGLS [29], CGQR [36] (both have pluses and minuses in different applications), etc. By using the Ritz values derived from Rayleigh-Ritz method [73], the regularizing effects and the spectral filtering of the CG method can be well studied.

There are also many other Krylov subspace methods which have been studied for regularizing ill-posed problems such as GMRES, BiCG, QMR. The regularizing effects of these algorithms were presented in Refs. [74,75]. Together with GGLS, CGS and BiCGSTAB, the regularizing effects of these six algorithms were compared in Ref. [76]. Recently, the IDR(s) [77] and LSMR [78] algorithms which are state-of-art iterative methods for large-scale linear systems were compared with CGLS in regularizing ill-posed problems in Ref. [79]. It turned out that IDR(s) can give satisfactory solution with less computational cost with discrepancy principle as a stopping rule and LSMR can produce a more accurate solution with L-curve method as a stopping rule.

The above discussed algorithms are all impeded by a semi-convergence behavior, because of which the iterative count plays the role of regularization parameter. However, there is a class of iterative shrinkage-thresholding algorithms (ISTAs) [80] which can avoid undergoing the semi-convergence behaviour. The original ISTA is attractive because of its simplicity and capability for solving large-scale problems, but it is limited in use because of its slow convergence rate, which even becomes serious as the observation operator is ill-conditioned. Different acceleration strategies were successively exploited to overcome its convergent drawbacks and utilize its advantages mentioned above. The most popular accelerated ISTAs are fast iterative shrinkage-thresholding algorithm (FISTA) [81] and two-step iterative shrinkage/thresholding algorithm (TwIST) [82, 83]. The next iteration of these two algorithms depends on two or more previous iterations. Afterwards, more ISTAs such as subband-adaptive ISTA [84], over-relaxed variant of FISTA [85], improved ISTA [86], exponential wavelet ISTA [87, 88], aF-PISTA [89], MFISTA [90] were proposed successively. Furthermore, a general iterative shrinkage algorithm (GISA) [91] was proposed for non-convex l_p norm minimization by extending the soft-thresholding operator. Compared with the advanced algorithms, namely LUT [92], IRLS [93], ITM-l_p [94], GISA is easier to understand, implement and more reliable. The convergence result of GISA were also studied in Ref. [95]. The global and local convergence results of ISTA and FISTA were certificated in Refs. [96–99], which would be strong support for applications of ISTAs. As a necessary numerical tool, iterative methods have been a research focus all the time. Therefore, different classical iterative methods were enriched by introducing various preconditioning techniques, search schemes, termination criterion, hybrid schemes in different application environments. We are not intend to have a further discussion on them.

As we mentioned repeatedly above that iterative methods have the intrinsic regularizing effect, i.e., the spectral components associated with large eigenvalues tend to converge faster than those associated with remained eigenvalues. In order to prevent the noisy components from spotting the iterative solution, we are encouraged to stop the iteration before the iterative solution starts to diverge. The most widely-used stopping rule for iterative regularization methods is the discrepancy principle [100] which is terminated at the kth step

$$\left\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(k)}\right\|^2 \leqslant \tau\Delta^2 \leqslant \left\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(k-1)}\right\|^2, \tag{16}$$

where Δ^2 represents the noise level, $\tau > 1$ is a constant. One can refer to Refs. [101–103] for more variants and em-

ployments of the discrepancy principle. Error-free parameter choice methods can also be reformulated as the stopping rule. This type of methods needs to combine each iterate with the corresponding objective function such as generalized cross validation (GCV) [58] method, the knowledge such as the generalized inverse of the forward matrix $\boldsymbol{A}$ is also essential requirement. However, the iterative methods are preferred to tackle problem which has little details about the matrix. Thus, this type of methods is of little practical use in iterative regularization methods. The L-curve method is the most popular regularization parameter choice method in recent years. It can actually be modified as a stopping rule for iterative regularization methods, from the approximate smooth L-curve [53, 54] or log-log scale of the iterative solution norm versus the residual norm, the knee point can be generally found at the L-shaped corner. The iterative solution corresponding to the knee point performs better at balancing the residual error and the perturbation error. Recently, a few stopping rules were proposed [104, 105] and classic stopping rules were still investigated [106].

In this subsection, we mainly reviewed the well-known iterative methods and explained their regularizing effect with proper stopping rules. There exist some other iterative regularization methods and stopping rules not involved above, see Refs. [11, 107–109] for more further and systematic information.

3.4 Variational regularization

The general form of variational regularization method is shown as the model described in Eq. (2). Under the condition of the Gaussian White Noise which is always assumed in amount of literatures, the fidelity (residual) term has a quadratic form: $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$ [110]. Different types of fidelity terms can be formulated using the MAP framework presented in Ref. [110] under different noise settings, for the residual term is supposed to have the same probability density distribution as the noisy data from a Bayesian standpoint. In this subsection, we are mainly concerned about the variety of the penalty terms in model Eq. (2), for different types of penalty terms are designed for different purposes associated with the regularized solution. For example, a penalty term with a form $\|\boldsymbol{x}\|_2^2$ represents the solution is desired to have minimal Euclidean length.

The earliest use of variational regularization methods can be tracked back to Tikhonov's researches on the incorrect problems [7, 111]. As we analyzed before, the penalty term generally contains some of the priori knowledge about the desired solution. Therefore, the variational form of the penalty term has great flexibility, it can be constraints on norm, smoothness or other structure properties of the desired solution [112]. Because of the explosion developments of variational regularization methods in recent years, it is impossible for us to keep an overview of the analysis, employments and application results in this brief paper. Therefore, we will review the most popular penalty terms in a comprehensive and succinct way while other penalty terms briefly.

3.4.1 Tikhonov (l_2) regularization

The penalty term of Tikhonov regularization is of the form $\|\boldsymbol{L}\boldsymbol{x}\|_2^2$. In the second section, the formulate process of the Tikhonov regularization model concentrating on the construction of the regularization matrix and the relationship between the Tikhonov regularization model and the original forward model was analyzed in detail. The regularization matrix can be identity matrix, first or second-order difference regularization matrix or decomposition of linear combination of these individuals for different purposes on the desired solution. In recent years, many results associated with the Tikhonov regularization have been published including linear and nonlinear problems with different regularization parameter choice methods. Convergence rate results have been explored under specific conditions. For example, Refs. [113–116] and Refs. [117–122] involve priori and posterior regularization parameter choice methods, respectively. Because large amount of regularization parameter choice methods used for problems with different constraints or boundary conditions will generate various applications which we can not review one after another, we only review the analysis tools, regularization parameter choice methods and optimization strategies separately.

A variety of numerical and analysis tools have been designed so far to get more insight into a discrete ill-posed problem. To be specific, the SVD or GSVD of a matrix or matrix pair and associated low-rank approximation and dimensionality reduction are the most important numerical tools in analyzing the structure of the forward operator and the features of the regularized solution. These analysis tools including the filter factors, the discrete Picard condition, the errors in the state and data space, the mean square error matrix and the averaging kernels are also important tools in analyzing the effects of regularization parameter and accuracy of the regularized solution. See Refs. [25, 31, 123–127] for details.

The computation of a good approximation to the exact solution depends on a proper selection of regularization param-

eter, for the regularized solution will either too smooth or too oscillatory if the regularization parameter is chosen too large or too small. A regularization parameter choice method depending only on noise level Δ^2 is known as priori parameter choice method such as the expected error estimation method using information about the errors in state space. By contrast, the posteriori parameter choice methods are formulated by combining the noised data with the noisy level Δ^2, e.g., the discrepancy principle, the generalized discrepancy principle which overcomes the drawback of the discrepancy principle for giving too small regularization parameter and undersmooth solution, the error consistency method [128] and the unbiased predictive risk estimator method [129, 130]. In semi-stochastic setting, the noisy level Δ^2 of the first two parameter choice methods in deterministic setting should be replaced by statistical information $n\sigma^2$ of the noise in data, and the last two methods can only be formulated in semi-stochastic setting. The error consistency method is identical to the discrepancy principle for constant $\tau = 1$, and the regularization parameter chosen by error consistency method is smaller than that chosen by discrepancy principle under the condition $\tau > 1$, and thus the error consistency method shares the same defect as the discrepancy principle. The predictive error can replace the exact solution error and get accurate estimation using the unbiased predictive risk estimator method, but the valley of predictive risk estimator can be very flat for large value of regularization parameter λ, thus leading to difficulty in choosing a proper regularization parameter. No matter what setting it is, error-free parameter choice methods mentioned in above text can be formulated because these methods only need to know the noised data. To be specific, the generalized cross-validation method using the ratio of the residual norm and the degree of freedom for noise as function, which has a minimum because both of the numerator and denominator are increasing functions of regularization parameter, but it has the same phenomenon as the unbiased predictive risk estimator method if the discrete Picard condition is satisfied. The maximum likelihood estimation can be a favorable method in overcoming the shortcomings of the generalized cross-validation and the unbiased predictive risk estimator methods. The quasi-optimality criterion can also find a balanced solution between the smoothness and noise error under certain assumptions, but it is often caught in local minima. As revealed before, the L-curve method is not only an analytical but a visual method which can be carried on in an analytical or numerical way. Similarly, the so-called residual curve method [107], generalized residual curve method [64, 107, 131] and flattest slope method [132] are all this type of methods which chooses the regularization parameter corresponding to a point on special bends of the curve.

In recent few years, some new regularization choice methods such as quadratic programming-based method [133], balance discrepancy principle [134], heuristic rule [135], adaptive regularization choice method [136] were exploited. Some classic or new regularization choice methods were compared under the conditions of exact-estimate, underestimate and overestimate error levels [137]. The performance of some known or new regularization choice methods was also investigated for LSQR method under inaccurate estimation of noise level in Ref. [138]. Once the regularization parameter is determined, the regularized solution can be obtained.

In the presence of noise, a regularized solution with ideal accuracy might be unavailable by the Tikhonov regularization method, it can be improved by the so-called iterative Tikhonov regularization. The first iteration step of iterative Tikhonov regularization is as usual, while an improved solution step is formulated to modify the present solution starting from the second iteration step. For sufficiently large modify steps, the accuracy of the regularized solution can not be improved significantly by any other methods. The procedures of iterative Tikhonov regularization are available in Refs. [139–141]. Usually, the regularization parameter is fixed, but Brill and Schock had investigated a non-stationary iterative Tikhonov regularization method [142] in which the regularization parameter is a variable depending on the iteration steps. A linear convergence rate has been established under a geometric parameter sequence for non-stationary iterative Tikhonov regularization method [143], and the discrepancy principle has been used to obtain an optimal accuracy. What's more, the non-stationary iterative Tikhonov regularization method was shown to be superior than its stationary counterpart and sophisticated ordinary Tikhonov regularization on simplicity of implementation, computational complexity and accuracy of reconstruction. Recently, some effective projected nonstationary iterative Tikhonov regularization methods were also put forward for reducing dimension of large-scale problems [144–146]. All these iterative Tikhonov regularization methods are optimization methods for Tikhonov regularization model.

3.4.2 Total variational regularization

One of the original regularization methods proposed in image processing which minimizes the Sobolev norm of the resulting image $\int_{R^2} |\nabla \boldsymbol{u}|^2 dxdy$ performs well in reducing the oscillation of the resulting image, but it can do nothing in recov-

ering the discontinuities (edges) and oscillatory textured patterns. Rudin, Osher, and Fatemi were motivated to formulate a total variational (TV) regularizer [6] in bounded variational (BV) space, this *TV term* is a non-smooth convex regularizer which evaluates the sparse gradients (or sparse derivative) of the resulting data. Therefore, with the TV regularizer as a penalty term, the regularization model ought to promote edge-preserving in sharp regions and noise removal in flat regions simultaneously in theory, which was later confirmed to be true in Ref. [147]. The unconstrained version of the model in Ref. [6] was formulated under the condition of Gaussian noise [148] as

$$\min_{\boldsymbol{x}\in BV(\Omega)} \int_{\Omega} (\boldsymbol{x}-\boldsymbol{y})^2 d\Omega + \lambda \int_{\Omega} |\nabla \boldsymbol{x}|\, d\Omega, \tag{17}$$

the second term in Eq. (17) is the TV regularizer multiplied by the regularization parameter λ, which restrains the amount of oscillation of the resulting data. While the fidelity term in Eq. (17) measures the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{y}$, which forces the resulting data to inherit most features from the noisy data. Overall, the geometric regularity of the resulting data $\boldsymbol{x}$ can be well imposed by restraining the oscillation while preserving sharp edges with an altering regularization parameter λ.

The image restoration problem was solved using a deterministic least squares framework firstly by Hunt [149]. According to Hunt, the choice of regularization parameter λ requires a priori knowledge of the noise variance. However, a number of researches (e.g., Ref. [150]) have shown that the choice of λ oversmooths the solution. A set-theoretic formulation for image inverse problem was used by Katsaggelos et al. in Refs. [151, 152]. A ellipsoid bounding of the smoothed data was used in this set-theoretic method for choosing the regularization parameter, and a very good restoration of the noised data was then produced. However, the ellipsoid bounding was not known in most cases. A novel algorithm for TV-based image reconstruction and parameter estimation was proposed using variational distribution within the hierarchical Bayesian formulation, and the competitive performance without any extra assumptions was shown in Ref. [153]. Different iterative schemes were also designed for different image restoration tasks, see e.g., Refs. [154, 155]. During the iteration, the regularization parameter was automatically chosen to guarantee the current restored image satisfying the discrepancy principle. Similarly, inspired by the better performance of unbiased predictive risk error (UPRE) method in Tikhonov regularization-based image restoration, an extended UPRE which relies on linear relationship between regularized solution and data was proposed in Ref. [156]. As a regularization parameter choice method, a statistically unbiased estimate of mean square error-Stein's unbiased risk estimate (SURE) depending on the noisy data and estimated image was presented in Ref. [157]. Afterwards, a novel Monte-Carlo technique was used to enable the estimation of SURE via stochastic ways [158], this estimation is quite good with no exaggerated computational cost as presented in Ref. [159]. The performance of these above methods relies on an accurate estimation of the noisy level. Otherwise, the regularized solution will be oversmooth or retain noise. A TV-based image restoration method was exploited in Ref. [160], the generalized cross-validation method was used to determine the value of regularization parameter in each restoration step. Furthermore, multi-scale TV-based regularization model with multiple spatially dependent regularization parameters was formulated in Refs. [161, 162], for the regularization parameter in original ROF model is a global parameter which does not satisfy local piecewise smoothness constraints in all image regions.

Total variational term can not be minimized directly because it is nondifferentiable at zero. The first approach succeeding in regularizing is its smoothed form $\int_{\Omega} \sqrt{|\nabla \boldsymbol{x}|^2 + \beta^2} \mathrm{d}\boldsymbol{x}$ using the associated Euler-Lagrange equation, but the Euler-Lagrange equation is still very stiff to solve. Lots of numerical methods for TV minimization have been excavated till now. The ideas of these methods are inspirations from various fields. Some particular mentioned categories are Newton-type methods, dual/primal-dual methods, Bregman iterative methods, graph cut methods, and programming methods. The steepest descent method and Newton-type method can be applied to TV-based regularization model, but the search strategies of these two types of methods often need to compute the minimization of the penalty term in each iteration step, which is computational costly. An alternative to these above two types of methods is the lagged diffusivity fixed point iteration [163] which avoids the referred drawbacks above, but the hessian matrix of the penalty term which is difficult to obtain or does not exist is also needed for Newton-type methods and lagged diffusivity fixed point iteration. Dual and primal-dual methods are used frequently recently because they are very fast in practice. Many novel ideas and techniques can be induced from the numerical optimization of dual formulation. Primal-dual methods can make use of information from the primal and dual spaces. Actually, the dual and primal-dual problems have to face some inevitable numerical difficulties, e.g., the rank deficient objective and the constraints need to deal

with. Chan, Golub, and Mulet (CGM) firstly formulated a primal-dual system involving a primal and dual variable synchronously [164]. Because the component in Euler-Lagrange equation corresponding to the smoothed total variational term is singularity, Chan et al. introduced an auxiliary variable to eliminate this singularity and presented a global convergence using Newton's method coupled with a simple line search scheme. This method can be understood well intuitively. Chambolle [165] proposed a pure dual method which is the first to solve the exactly problem described in Eq. (17) rather than its approximation. By simplifying the KKT system of the discretized objective, Chambolle proposed a semi-implicit scheme which settles Eq. (17). This algorithm converges fast for denoising problem, but when confronting other problems in which $\boldsymbol{A}$ is usually ill-conditioned, the condition of the semi-implicit scheme will be too stringent. Huang et al. [166] and Bresson and Chan [167] proposed a spliting method which is modeled as

$$\min_{\boldsymbol{x},\boldsymbol{v}} G(\boldsymbol{x},\boldsymbol{v}) = \frac{1}{2}\|\boldsymbol{K}\boldsymbol{x}-\boldsymbol{y}\|^2 + \frac{\mu}{2}\|\boldsymbol{x}-\boldsymbol{v}\|^2 + \lambda\|\nabla\boldsymbol{v}\|_1, \quad (18)$$

the objective function can be minimized against $\boldsymbol{x}$ and $\boldsymbol{v}$ in turn, and thus the minimization problem described in Eq. (18) can be regarded as Tikhonov regularization and TV-based denoising problem in turn. Therefore, Eq. (18) can be solved using methods from Thikhonov regularization and TV-based denoising problem, Chambolle's scheme can still be used for TV-based denoising in part of the whole process. The KKT system can also be written as a semi-smooth system which can be solved using a semi-smooth Newton's method [168], and this method usually has a superlinearly convergence rate under some proper conditions. In order to use advantages of both primal and dual spaces, Zhu and Chan proposed the primal-dual hybrid gradient method which uses the dual form instead of the primal total variational term [169], and minimizes the objective function against the primal and dual variables alternatively. Iteration strategies and numerical performance are illustrated in Refs. [169, 170]. In regard to anisotropic discrete of TV norm with bilaterally constraint on dual variable, Hintermöller and Kunisch proposed the primal-dual active-set (PDAS) algorithm which has a superlinear convergence rate, but this method can only be applied to isotropic discrete of TV norm. A nonnegatively constrained CGM method in which the primal and dual variables are treated as in the PDAS and CGM methods separately was proposed in Refs. [171, 172]. The optimality conditions of primal-dual problem are shown to be semi-smooth and can be solved superlinearly using a semi-smooth Newton's method. This method can be applied to both the isotropic and anisotropic discrete forms of TV norm.

The Bregman iteration method proposed by Osher et al. [173] has been generalized for many convex inverse problems. The original Bregman iteration method can bring back the loss signal by solving a sequence of ROF models, but it may bring the noise back if the iteration is not terminated until convergence. The split Bregman iteration method proposed in Ref. [174] can avoid this drawback because it converges to the solution of the ROF model, and the split Bregman iteration method can be interpreted as a forward-backward splitting method from the view of spliting method. In Eq. (9), if $1/\lambda$ is an equivalent infinitesimal of $\|\boldsymbol{L}\boldsymbol{x}\|^2$, then $\lambda\|\boldsymbol{L}\boldsymbol{x}\|^2$ may have no obvious change along with the regularization parameter λ, therefore the trade-off role of λ will no longer obvious. The augmented Lagrangian method [175] can handle this awkward situation perfectly. What's more, many popular methods such as CGM and Chambolle's dual methods, are connected to the augmented Lagrangian method, the split Bregman iteration method is also equivalent to the augmented Lagrangian method, see Ref. [176] for details.

Graph cut methods Refs. [177, 178] are being popular recently, for these methods are fast and they can achieve global optimal solution even for nonconvex problem. Whether the objective function is levelable is crucial for using graph cut method, fortunately, the total variational function satisfies this condition. For anisotropic discrete TV norm, a quadratic objective function with linear constraints can be formulated by introducing two auxiliary variables [179], then solved by a standard primal-dual interior-point methods. However, if we encounter isotropic discrete TV norm, the above methods will not work. A second-order cone programming combined with a interior-point scheme was considered in Ref. [180] to handle this problem, but this method will incur too many variables and thus enlarge the scale of the problem. Except for these methods reviewed above, the majorization-minimization (or majorization-maximization) is also a well-studied method in optimization and it can be applied to TV-based inverse problem [181, 182], the effectiveness of this method depends largely on the choice of surrogate function.

As shown in the MAP framework presented in Ref. [110], the fidelity term in the total variational based model is determined by the noisy form contained in the collected data. The ROF model formulated under the Gaussian noise in the Bayesian framework is under the consideration of convenient analysis. However, we may encounter other forms of noise in some applications such as Laplace noise, Poisson noise, multiplicative noise. Total variational regularization models under these forms of noise and corresponding

numerical methods can be found in Refs. [183–189]. Different space norms for the fidelity term have also been investigated (cf. Refs. [190, 191]). There are several variants of total variational term for different desires: multichannel and matrix-valued TV were proposed for handling multiple perspectives, high dimension datum [192, 193] and discrete TV were designed for procedure implementation [194, 195], smoothed TV was needed for applying numerical techniques conveniently [164].

As we known, the classical local TV based image recovery inevitably confronts staircase edges and block artifacts. Some schemes such as improved steepest descent projection [196], a fourth-order TV based scheme [197] were initially presented to relieve the piecewise constant effect substantially. Thereafter, the nonlocal TV was formulated to eliminate these drawbacks more effectively [198, 199]. Recently, new high-order TV [200, 201], novel nonlocal TV [202], and fractional order TV [203] were successively proposed for better elimination of staircase effects. A non-convex hybrid TV [204] was proposed at almost the same time which can retain more valuable edges while eliminating staircase effect compared with high-order TV models. We are not intend to have in-depth analysis due to the large amount of literatures about TV regularization model, one can refer to relevant literatures for specific total variational regularization model.

3.4.3 Sparse and low-rank regularization

Compressive sensing (CS) has brought about many striking ideas to signal restoration since it was proposed by Donoho et al. [205–209]. We can get a thorough understanding of l_p norm-based regularization model for ill-posed problems from the sparsity of the regularized solution, which is the key target in compressive sensing. CS relies on the empirical observation that signals can be represented sparsely in terms of a suitable basis, commonly a sparse dictionary. With a proper sparse dictionary $\boldsymbol{\psi}$, the signal $\boldsymbol{z}$ can be represented by a sparse signal $\boldsymbol{\theta}$, i.e., $\boldsymbol{z} = \boldsymbol{\psi\theta}$. We make an observation on signal $\boldsymbol{z}$ with an underdetermined measurement operator $\boldsymbol{\Phi}$ shown as

$$\boldsymbol{H} = \boldsymbol{\Phi z}, \tag{19}$$

combining with the sparse representation, Eq. (19) becomes

$$\boldsymbol{H} = \boldsymbol{\Phi\psi\theta}, \tag{20}$$

by denoting $\boldsymbol{A^{CS}} = \boldsymbol{\Phi\psi}$, the problem recovers $\boldsymbol{z}$ from $\boldsymbol{H}$ is transformed into recovers $\boldsymbol{\theta}$ from $\boldsymbol{H}$. Although Eq. (20) is still an underdetermined problem, the unknown components of variable $\boldsymbol{\theta}$ decrease largely compared to variable $\boldsymbol{z}$ because of its sparsity. The opportunistic point of Eq. (20) is that the sparse dictionary $\boldsymbol{\psi}$ can be designed without knowing the signal $\boldsymbol{z}$ but to make it sparse just right. However, in order to ensure the existence of the solution for Eq. (20), the matrix $\boldsymbol{A^{CS}}$ should have at least $2K$ random linearly independent columns for K-sparse signal $\boldsymbol{\theta}$ [210]. If the signal $\boldsymbol{z}$ is compressible under sparse dictionary $\boldsymbol{\psi}$ and matrix $\boldsymbol{A^{CS}}$ meets the requirement above, we can obtain an unique approximate solution of Eq. (20) by solving the minimization problem

$$\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_0 \quad \text{s.t. } \boldsymbol{H} = \boldsymbol{\Phi z} = \boldsymbol{\Phi\psi\theta}, \tag{21}$$

considering the observed signal $\boldsymbol{H}$ is always noised, the constraint in Eq. (21) should be corrected as $\|\boldsymbol{H} - \boldsymbol{\Phi\psi\theta}\|^2 \leqslant \varepsilon$ for robust. the constraint can also be set up based on the l_1 norm for large and impulsive noise [211,212]. Unfortunately, the l_0-norm optimization is non-convex and the combinatorial minimization problem described in Eq. (21) is NP-hard in general [213]. Therefore, greedy strategies such as matching pursuit algorithm [214] and orthogonal matching pursuit [215], iterative thresholding algorithm [216], FOCUSS [217] and convex relaxation approach [207] are dug out to tackle this problem. The application of greedy algorithms is time-consuming and the global optimum can not be guaranteed. The iterative thresholding algorithm can be applied with little time expense, but it is sensitive to the noise in data and only guarantees to find local optimum. A proximal method which can be fast and converge to a stationary point was proposed in Ref. [218] for l_0 minimizations recently.

The convex relaxation l_1 regularization is a landmark work because it indeed promotes sparsity in a convex environment. In fact, the TV regularizer mentioned above is evidently that the l_1 norm acts on gradient domain of the recovery data. Thus, it can provide recovery data with sparse gradients, which is the reason why TV penalty term can promote edge-preserving in sharp regions and noise removal in flat regions simultaneously. The theoretical proofs for equivalence of l_1 and l_0 -norm in promoting sparsity under the frame of RIP and RIC were given in Refs. [207, 219–221]. Therefore, various sophisticated convex optimization techniques can be applied for the sparse solution. The l_1-minimization problem under Gaussian noise reads

$$\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_1 \quad \text{s.t. } \|\boldsymbol{H} - \boldsymbol{\Phi\psi\theta}\|^2 \leqslant \varepsilon, \tag{22}$$

which is also known as the basis pursuit problem [222], it can be transformed into an unconstrained minimization problem using Lagrange method

$$\min_{\boldsymbol{\theta}} E_\lambda(\boldsymbol{\theta}) = \frac{1}{2}\|\boldsymbol{H} - \boldsymbol{\Phi\psi\theta}\|^2 + \lambda\|\boldsymbol{\theta}\|_1. \tag{23}$$

In order to make sure an unique sparse solution of Eq. (23), the matrix $\boldsymbol{A}^{\boldsymbol{CS}}$ has to meet the restricted isometry property (RIP) which still has difficulty at present. Nevertheless, the combination of mutual coherence coefficient, spark [210] and RIC of matrix $\boldsymbol{A}^{\boldsymbol{CS}}$ in analyzing the recovery condition from sensing model described in Eq. (22) does make sense. Equation (23) can be solved by similar procedures as in Tikhonov and TV regularization models once the above condition is met. What's more, the minimizer of E_λ can be solved explicitly by soft-thresholding the components of $\boldsymbol{H}$ if the sensing matrix $\boldsymbol{A}^{\boldsymbol{CS}}$ is the identity. If the matrix $\boldsymbol{A}^{\boldsymbol{CS}}$ is a general linear operator, the solution can be approximately computed by soft thresholded Landweber iteration methods proposed in Refs. [223–225].

Unfortunately, the RIP condition of matrix $\boldsymbol{A}^{\boldsymbol{CS}}$ can hardly be satisfied when the undetermined measurement matrix $\boldsymbol{\Phi}$ is ill-conditioned. Therefore, based on this finding, Herrholz and Teschke [226] introduced a sampling function $\boldsymbol{F}_s$ together with a proper design of $\boldsymbol{\psi}$ to diagonalize $\boldsymbol{\Phi}$ into Λ, and consider the following problem

$$\min_{\boldsymbol{\theta}\in B(l_1,R)} E_\lambda(\boldsymbol{\theta}) = \frac{1}{2}\|\boldsymbol{G} - \boldsymbol{F}_s\boldsymbol{\Phi}\boldsymbol{\psi}\boldsymbol{\theta}\|^2 + \lambda\|\boldsymbol{\theta}\|^2, \tag{24}$$

where $B(l_1, R) = \left\{\boldsymbol{\theta} \in l_2 \left| \|\boldsymbol{\theta}\|_{l_1} \leqslant R \right.\right\}$ is a bounded set, $\boldsymbol{G} = \boldsymbol{F}_s\boldsymbol{H}$, and the selection of $\boldsymbol{F}_s$ and $\boldsymbol{\psi}$ should bring $\boldsymbol{F}_s\boldsymbol{\Phi}\boldsymbol{\psi}\boldsymbol{\theta} = \boldsymbol{F}_d\boldsymbol{\Lambda}\boldsymbol{\theta}$ into existence. This variational formulation allows the treatment of ill-posed problems in the context of compressively sampled data. A different combination way of l_1 and l_2 constraints used for the so-called elastic net regularization was given in Ref. [227].

If the signal itself is sparse, the sparse dictionary is no longer needed and the observation matrix $\boldsymbol{A}$ is equivalent to the sensing matrix. Therefore, the l_1-norm-based models

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \quad \text{s.t. } \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|^2 \leqslant \varepsilon, \tag{25}$$

and

$$\min_{\boldsymbol{x}} E_\lambda(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|^2 + \lambda\|\boldsymbol{x}\|_1, \tag{26}$$

were frequently used for sparse reconstruction recently, see e.g., [222, 228–230]. Equation (25) can be solved by the representative algorithms: gradient projection (GP) [228, 229], ISTAs [80, 81, 231], proximal gradient (PG) [232], homotopy [230, 233] and augmented Lagrange multiplier [212]. One can choose the regularization parameter of Eq. (26) in the same way as used in Tikhonov regularization and solve it using gradient projection sparse representation (GPSR) [228] and truncated Newton interior-point method (TNIPM) [229]. We also have [234] which explained the conflict between two terms of Eq. (26) and exploited a new way to solve this model using multi-objective optimization.

The non-Lipschitz l_p $(0 < p < 1)$ norm-based minimization is another technique to relieve the awkward situation that the RIP condition is difficult to meet. Chartrand [235] has generalized the result of Ref. [236] to demonstrate that non-Lipschitz l_p $(0 < p < 1)$ regularizer can recover sparse signal perfectly under weaker RIP conditions than l_1 regularizer in the noise-free setting. The experiments in Ref. [235] also indicated that non-Lipschitz l_p regularizer requires fewer measurements for exact reconstruction than l_1 regularizer, and decreasing p requires decreasing number of measurements. What's more, depending on the same RIP condition and the noise level, the non-Lipschitz l_p regularizer provides better theoretical guarantees in respect of stability and robustness than l_1 regularizer according to Ref. [237]. For example, the l_1 regularizer is untalented when the data errors have heavy tail process [238], and it often suffers from estimation bias [239] and sparsity insufficient problems. Fan and Li [240] urged that the penalty functions ought to be singular at the origin to promote sparsity, and its gradients should be zero for large variable values to promote unbiasedness. Following these guidelines, some fruitful penalty functions have been proposed for various purpose. For example, the smoothly clipped absolute deviation (SCAD) [240] function and the Zhang's function [239] are unbiased penalized estimators, the Logarithm and fraction penalty functions [241] are expert in edge-preserving, the non-Lipschitz l_p penalty has oracle property in statistical community [242, 243].

Due to recent ardour on sparse presentation and compressed sensing, we are mainly concerned with the research progress of the non-Lipschitz l_p regularizer which is a better alternative to l_1 regularizer according to the above statements. Although the non-Lipschitz l_p regularizer can promote competitive solution with less measurements and weaker RIP conditions compared with l_1 regularizer, it is difficult to solve efficiently because the l_p minimization model is nonsmooth and nonconvex. Although the theoretical guarantees for solution optimality and algorithm convergence do not seem to be available, some typical algorithms actually work well in practice for non-Lipschitz l_p regularization. According to Ref. [244], these typical algorithms can be classified into three kinds: iteratively reweighted l_1 minimization (IRL1) [245], iteratively reweighted least squares (IRLS) [246] (including its general form GIRLS), and iteratively thresholding method (ITM) [247]. The former two kinds of algorithms can be used to solve both unconstrained and constrained l_p minimizations, while the latter one can only solve unconstrained l_p

minimizations.

Through a clever shift, IRL1 and IRLS transform the l_p regularization problems into a series of reweighted l_1 and l_2 regularization problems, respectively. Then, the convex optimization algorithms can be used to handle these reweighted problems. In ITM, the iterative scheme for the final solution depends on the so-called thresholding function. However, the reweighted l_2 regularization problems can not promote sparse solution, which violates the original intention of using the non-Lipschitz l_p regularizer. Therefore, the IRLS algorithm will have poor achievements in promoting sparsity compared with the IRL1 and ITM unless a threshold strategy is used. What's more, the ITM without any stepsize search always maintains a global convergence property [94], while the IRL1 and IRLS can hardly guarantee a global convergence. The systematically comparison of the performance of these three algorithms was also conducted in Ref. [244].

Because decreasing p ($p \in (0, 1)$) requires decreasing measurements, weakening RIP conditions and increasing complexity, a proper p is important for better performance of l_p regularization. Xu et al. [248] revealed that $l_{1/2}$ regularizer has unbiasedness, sparsity and oracle properties, and their experiments [249] showed that $l_{1/2}$ regularizer is effective and efficient, and could be taken as a representative of the l_p regularizer. Because the $l_{1/2}$ regularizer is a specific l_p regularizer, algorithms for l_p minimization are all applicable to $l_{1/2}$ minimization.

Apart from the classical sparse constraint, the low-rank can be interpreted as a sparse constraint on matrix. For the fit-to-data model: $\min_{\boldsymbol{u}} \|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{f}\|^2$, the structural information of matrix $\boldsymbol{A}$ is not fully utilized. Considering of this, some rank minimization models such as regularized nuclear-norm minimization problem [250] and first order singular value thresholding (SVT) problem [251] were proposed. As fundamental problems, there are many algorithms can be applied to solve these problems such as alternating direction method of multipliers (ADMM) [252, 253], proximal point algorithm [254], alternating direction method [255] and so on. One can easily find more algorithms from a large amount of literatures.

In order to reduce the complexity in the solving process of Eq. (22), low-rank characteristic was imposed on the sparse learning dictionary by clustering the similar patches in Ref. [256]. The experiment results showed that their proposed LRSR algorithm performs better than or is comparable with some state-of-art algorithms in denoising and texture recovery. This idea from matrix decomposition can be used for signal recovery because corrupted two-dimensional signal can be conformed into a matrix, then the matrix can be decomposed into a low-rank part and a sparse part. Based on the robust principal component analysis (RPCA), Waters et al. proposed a SpaRCS algorithm [257] which can be used to separate out the noisy components from noised signal. For example, Li et al. supposed the degraded HSIs consist of a low-rank signal component and a structural sparse noise component, and proposed a novel algorithm based on RPCA [258]. The simulated and real data experiments demonstrated their assumptions and presented promissing denoising results. Zhou and Tao proposed a more general algorithm GoDec [259] which considers the noise part in their model. Based on this algorithm, Zhang et al. proposed a new Hyperspectral image restoration method [260] which can further remove the mixed noise such as Gaussian noise, impulse noise, stripes, etc.

It is worth mentioning that Zhang, et al. [261] roughly categorized the sparse representation models into five groups based on the norm used for regularizer, namely l_0, l_p ($p \in (0, 1)$), l_1, $l_{2,1}$, and l_2 group. The available algorithms for sparse representation were also classified into four groups, namely greedy strategies, constrained optimization methods, proximity algorithms, and homotopy algorithms. They also conducted a comparison study of these representative algorithms. For more details, readers may refer to Ref. [261]. Now that we have had a thorough retrospect on different regularization models and their optimization algorithms, We find and list some well-known packages for the implementation of the well-known algorithms in Table 1. The available website and description for each package are shown in this table, the cited references, demo files and user guides are also available at the corresponding website.

3.5 Discussion on fidelity term

So far, we have concentrated our mainly attention on various penalty terms, but seldom concerned with the form of the fidelity term. In fact, when non-Gaussian noise and nonlinear forward model are encountered, the fidelity term will no longer have quadratic or convex form. For example, when the noise in forward model follows a Laplacian, Poisson or Multiplicative distribution, various fidelity terms can be deduced by means of the MAP framework [110]. Readers can get insight into the concrete forms of these fidelity terms in Ref. [110]. In addition, several other fidelity terms also have been studied based on different motivations recently. For example, the dual norm of BV, the Wasserstein metric, the l_1 space norm and the dual Sobolev norm were used as a fidelity term respectively in Refs. [164, 191, 262, 263]. Different regularization models with various fidelity terms ought to have

Table 1 Available websites and brief descriptions for some well-known packages

Package name	Website	Brief description
YALL1 (YALL1-Group)	http://yall1.blogs.rice.edu/	Provides the implementation of alternating direction algorithm for l_1 minimizations (for group sparsity problems)
l_1_ls	http://www.stanford.edu/~boyd/l1_ls/	Developed for problems with large amount of variables or large density
Homotopy	http://users.ece.gatech.edu/sasif/homotopy/	Provides the implementation of homotopy techniques for some l_1 minimization problems
Sparselab	http://sparselab.stanford.edu	A well-known package for sparse representation and compressed sensing, it provides lots of famous algorithms for sparse modelings, such as Lasso, BP, MP, OMP, and IST
L1General	https://www.cs.ubc.ca/~schmidtm/Software/L1General.html	Mainly focuses on algorithms which require calculation or storing of Hessian or approximate Hessian, such as some gradient-based and soft-thresholding algorithms
SLEP	http://parnec.nuaa.edu.cn/jliu/Softwares.htm	Mainly provides the implementation of the families of Lasso algorithm, such as fused Lasso, group Lasso, and sparse group Lasso

different existence, uniqueness and convergence characteristics with their solutions, which becomes an active research field.

3.6 Summary on the regularization methods

In above subsections, we mainly reviewed the frequently used regularization methods for linear ill-posed problems, including TSVD, iterative regularization, variational regularization such as Tikhonov, TV-based and the sparse regularization. The truncation behavior of TSVD, iterative count of iteration regularization do play the same role as the regularization parameters in variational regularization methods in our analysis. Nonlinear problems are highly related to the particular applications, the practical and theoretical mansions are much more easily built up in linear cases, models for nonlinear problem are usually linearized and parts of the approaches for nonlinear problems are generalized or inspired from those for linear problems [4, 25, 62].

The function mechanism of the filter operator in TSVD regularization has been in-depth analyzed associated with matrix decomposition up to now. Its use in nonlinear problems is limited because nonlinear operator can hardly conduct a singular value decomposition. If the nonlinear operator is linearized, the analysis of TSVD may be inaccurate. As for the Tikhonov regularization for nonlinear ill-posed problems, the step-length and trust region methods base on Taylor expansion are proved to be effective. The step-length method needs the implementation of searching direction, computing step length and accurate terminating. The optimization methods of Tikhonov regularization for linear ill-posed problems can be used in each iterative step to improve stability of this method and reduce the total iteration steps. Thus, the solving process of a nonlinear ill-posed problem is equivalent to solving a sequence of ill-posed linearizations via Tikhonov regularization in each iterate [62]. The step-length method chooses a shorter step on a restrained direction, but the trust region method searches in any direction within a restrained region for the minimizer. The iterative regularization methods for linear ill-posed problems are also generalized and enriched for nonlinear problems such as nonlinear Landweber iteration, Newton-type methods, asymptotic regularization, and multilevel or level set methods [4, 62, 264].

The numerical tools and theoretical framework of iterative and Tikhonov regularization for linear and nonlinear ill-posed problems have tended to be mature. However, the total variation regularization used for nonlinear ill-posed problems remains to be explored, mainly including the existence and uniqueness of solution especially for nonconvex problems, the solving strategies and wide applications for different areas. As an active field, there are many directions in sparse regularization waiting to be explored, which can not be listed exhaustively. For example, the extension of sparse regularization towards nonlinear measurements and high-dimension problems has a strong need for developing highly adaptive algorithms, the measurement matrix and sparse dictionary which highly conform to the signal features are eager to get further research. Except for these practical aspects, theoretical guarantees for stable accomplishment of algorithms also need to be enriched, a necessary and sufficient condition for unique sparse solution of sparse regularization models is still a key problem. What's more, a combination of other signal priori and its sparse priori knowledge may exploit high quality recovery.

4 A multi-objective framework for ill-posed problems

In this section, we intend to introduce some foundations of

multi-objective optimization and construct a multi-objective framework for ill-posed problems under some rational analysis.

4.1 Introduction to multi-objective optimization

A multi-objective optimization problem (MOP) is normally defined as follows [17–19]:

$$\begin{aligned} &\min_{\boldsymbol{x}} \boldsymbol{F}(\boldsymbol{x}) = [f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_n(\boldsymbol{x})]^{\mathrm{T}} \\ &\text{s.t. } g_i(\boldsymbol{x}) \leqslant 0,\ i = 1, 2, \ldots, k; \\ &\quad h_l(\boldsymbol{x}) = 0,\ l = 1, 2, \ldots, e, \end{aligned} \tag{27}$$

where n is the number of objective functions, k and e are the number of inequality and equality constraints separately. $\boldsymbol{x} \in R^m$ is a vector of decision variables. $\boldsymbol{F}(\boldsymbol{x}) \in R^n$ is a vector of objective functions, $\boldsymbol{F}(\boldsymbol{x}) : R^m \to R^n$. The feasible decision space $\boldsymbol{X}$ is defined by the constraints as $\left\{\boldsymbol{x} \left| g_i(\boldsymbol{x}) \leqslant 0, i = 1, 2, \ldots, k;\ h_j(\boldsymbol{x}) = 0, j = 1, 2, \ldots, e. \right.\right\}$. The feasible cost space $\boldsymbol{Y}$ is defined as $\{\boldsymbol{F}(\boldsymbol{x}) | \boldsymbol{x} \in \boldsymbol{X}\}$.

Compared with the single-objective optimization, the outcome of multi-objective optimization process is more of a set of mathematically equally trade-offs than just a single solution. The set of trade-offs is defined as follows.

Definition 1 A point $\boldsymbol{x}^* \in \boldsymbol{X}$ is Pareto optimal if and only if there does not exist another point $\boldsymbol{x} \in \boldsymbol{X}$, such that $f_i(\boldsymbol{x}) \leqslant f_i(\boldsymbol{x}^*), \forall i = 1, 2, \ldots, n$, and $f_i(\boldsymbol{x}) < f_i(\boldsymbol{x}^*)$ for at least one function, i.e., the point $\boldsymbol{x}$ dominates $\boldsymbol{x}^*$ in general terms.

All these Pareto optimal points make up a set which is known as Pareto optimal set [265], and their corresponding Pareto optimal objectives are called the Pareto optimal front (PF) [265]. Dynamically to see, a Pareto optimal solution with some superior objectives inevitably has some inferior objectives compared to other Pareto optimal solutions. Thus, it can be deduced that objectives in multi-objective optimization problems are conflict to each other.

The multi-objective problem is solved by a variety of methods with a priori articulation of preferences traditionally [266]. However, these methods sometimes can not reflect decision maker's desires perfectly as is expected. By contrast, methods incorporating a posteriori articulation of preferences which are also called generate-first-choose-later approaches are more favorable to decision makers because of their capacity for handling preference information [266]. Remarkably, the EAs as global optimization techniques can generate a set of solutions simultaneously, it is suitable for searching Pareto optimal set of multi-objective optimization problems, and the methods generated by EAs used in multi-objective optimization are referred to as multi-objective evolutionary algorithms (MOEAs) [17–19].

According to our survey on multi-objective evolutionary algorithms in Refs. [267], MOEAs can be classified into three continuously improving stages. The MOGA [268], NSGA [269] and NPGA [270] are representatives of the first stage, where the selection of individuals and diversity of population are based on the Pareto dominance and fitness sharing mechanism, respectively. The MOEAs of second stage have been put forward since the use of external continuously updated population for storing nondominated solutions in SPEA [271]. The improved SPEA which was named as SPEA2 [272] was proposed two years later by Zitzler et al, and the SPEA2+ [273] was proposed soon after to improve the strength of SPEA2. Similarly, other methods such as the PAES [274] and its updated versions PESA [275] and PESA-II [276] were gradually proposed. Compared with these algorithms, the modified algorithm of NSGA namely NSGA-II has found its pervasive applications beacause of its low computational overhead, elitist and parameter-less sharing approach [277]. Most famously, a multi-objective evolutionary algorithm based on decomposition (MOEA/D) [278] can decompose the MOP into a number of subproblems, these subproblems can be optimized by scalar objective local search strategies naturally. Hybrid MOEAs are also useful algorithms which hybridize different techniques to utilize their advantages for tackling complicated MOPs. Since the put forward of the MOEA/D, a variety of variant algorithms with different decomposition methods [279, 280], different reproduction operators [281], different local search operators [282] and different dominance mechanisms [283] and so on were successively proposed for better performance. Recently, a reference-point based NSGA-II namely NSGA-III [284] was exploited to handle many objective problems. A number of well-spread reference points were supplied and updated adaptively for diversity of population members. By introducing non-dominated sorting scheme, a θ-NSGA-III [285] was obtained for better compromise between convergence and diversity in many objective optimization. Soon after, a unified evolutionary optimization algorithm U-NSGA-III [286] was proposed to unify mono-objective problems, bi-objective problems and three or more objective problems, the simulation results showed that this unification is excellent and worthy of further application. In a word, by introducing advanced schemes in different stages of evolvement, some efficient evolution algorithms with specific superiorities can be obtained such as the MOMA [287] with a novel path encoding scheme and a specific evolutionary operator, the MGFE

[288] with a novel fitness evaluation function and a pruning local search scheme, and the LSH-MOMA [289] with a collaborative strategy of MOEA and a locality-sensitive hashing (LSH) based search scheme. These algorithms are proposed for problems with different characteristics and all verified to be effective by the experiments therein.

4.2 Motivation for modeling ill-posed problems as MOPs

From the subsection of TSVD regularization and variational regularization, we know that the penalty term aims at using priori knowledge of the exact solution and indirectly restraining the noise amplified by the small singular values of observation matrix, while the regularization parameter is to expand this restraint. We are going to illustrate the relationship between fidelity term and penalty term together with the regularization parameter.

In order to analyze easily, we assume that the function $L(\boldsymbol{x})$ and $P(\boldsymbol{x})$ are convex. Because the regularization parameter is positive, the objective function $E(\boldsymbol{x})$ is a convex function as well. Several facts can be deduced as follows:

- Function $L(\boldsymbol{x})$ and $P(\boldsymbol{x})$ can not have the same minimizer. Otherwise, these two function will increase or decrease at the same time according to the properties of convex function. Therefore, the penalty term could not play any active role, not to mention its regularizing effect. Assuming $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are the minimizer of $L(\boldsymbol{x})$ and $P(\boldsymbol{x})$ separately ($\boldsymbol{x_1 \neq x_2}$), then the values of function $L(\boldsymbol{x})$ and $P(\boldsymbol{x})$ will change conversely for shifting points on the line segment $\boldsymbol{x}_1\boldsymbol{x}_2$. The global minimizer of function $E(\boldsymbol{x})$ will generate in regions where the values of $L(\boldsymbol{x})$ and $P(\boldsymbol{x})$ have negative correlation rather than positive correlation.
- In order to make our analysis intuitionistic, the fidelity term $L(\boldsymbol{x})$ and the penalty term $P(\boldsymbol{x})$ of Eq. (2) are regarded as the vertical and horizontal axes separately, see Fig. 1. The conflicts between these two terms shown in Fig. 1 can be proven reasonable by the frequent use of L-curve method for choosing proper regularization parameters, e.g., Refs. [53, 54]. The curve shape of objective function versus variable $\boldsymbol{x}$ is plotted as in Fig. 2 according to its convexity. We can link these two figures together by plotting a series of contour lines: $L(\boldsymbol{x})+\lambda P(\boldsymbol{x}) = C$ in Fig. 1. The searching process of the minimizer in Fig. 2: $\boldsymbol{x}_1 \to \boldsymbol{x}_2 \to \cdots \to \boldsymbol{x}^*$ corresponds to the update process of coordinates: $(P(\boldsymbol{x}_1), L(\boldsymbol{x}_1)) \to (P(\boldsymbol{x}_2), L(\boldsymbol{x}_2)) \to \cdots \to (P(\boldsymbol{x}^*), L(\boldsymbol{x}^*))$ in Fig. 1. The point of tangency $(P(\boldsymbol{x}^*), F(\boldsymbol{x}^*))$ of curve l with slope $-\lambda_1$ in Fig. 1 corresponds to the minimizer $\boldsymbol{x}^*$ of objective $E(\boldsymbol{x})$ in Fig. 2. If we change the regularization parameter λ, the point of tangency with slope $-\lambda$ and its preimage $\boldsymbol{x}^*$ will change correspondingly. To be specific, if the parameter λ increases, the penalty term will decrease and the fidelity term will increase. If λ decreases, the opposite result will be received. These conclusions are in accordance with the analysis in above text.
- For the convex function $E(\boldsymbol{x})$, if the optimization methods for minimization problem described by Eq. (2) converges, it converges to the global optimal $\boldsymbol{x}^*$ which lies in the Pareto optimal set of multi-objective problem: $\min_{\boldsymbol{x}} \{L(\boldsymbol{x}), P(\boldsymbol{x})\}$. Otherwise, there must be an Pareto optimal solution $\boldsymbol{x}'$ in Pareto optimal set which dominates $\boldsymbol{x}^*$, i.e., $L(x') \leqslant L(x^*), P(x') < P(x^*)$ or $L(x') < L(x^*), P(x') \leqslant P(x^*)$. Therefore, $E(\boldsymbol{x}') < E(\boldsymbol{x}^*)$, $\boldsymbol{x}^*$ is not the optimal solution, which deduces a contradiction. If the Pareto optimal set of multi-objective problem can be guaranteed by some optimization methods, the decision maker would prefer to select the ideal one from the Pareto optimal set rather than obtain a unique solution with a settled parameter in advance. New optimization approaches and Pareto dominance strategies are continuously put forward for this purpose. Therefore, multi-objective optimization has a hopeful future in coping with ill-posed problems.
- For nonlinear ill-posed problems, optimization approaches for regularization models such as Newton-type methods, have to overcome a series of extra difficulties like gradient. While multi-objective optimization should consider nothing about these problems but an increase of computational work by taking advantages of its global optimization techniques.

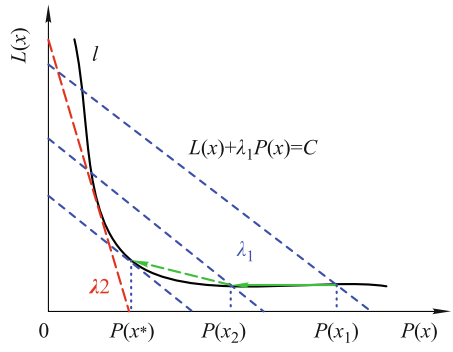

Fig. 1 $E(\boldsymbol{x})$ as the linear programming of $L(\boldsymbol{x})$ and $P(\boldsymbol{x})$ via λ

Through the above analysis on convex function, we can conclude some characters of regularization problems which are suitable for introducing multi-objective optimization. Firstly, the minimizer of objective $E(\boldsymbol{x})$ lies in the regions where fidelity term and penalty term are conflict, multi-objective optimization is just right for handling conflict objectives. Secondly, the choice of regularization parameter directly influences the minimizer of Eq. (2), but the magnitude of the influence is not easy to control. By contract, the Pareto optimal set of multi-optimization is more conducive for decision maker to choose from. Thirdly, optimization methods for better approximation of Pareto optimal set are put forward successively, which makes the use of multi-objective optimization in ill-posed problems possible. What's more, multi-objective optimization can be used for nonlinear ill-posed problem just with an increase of computational work compared with its use for linear ill-posed problem.

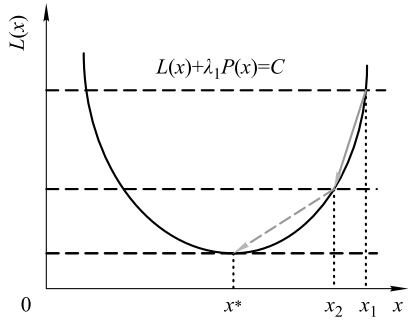

Fig. 2 The searching demo for the minimizer of objective function

4.3 Establishment of multi-objective framework for ill-posed problem

Now that the feasibility of multi-objective framework has been affirmed fully, we can simply reformulate the regularization problem Eq. (2) as

$$\min_{\boldsymbol{x}} \{L(\boldsymbol{x}), P(\boldsymbol{x})\}, \tag{28}$$

the advantage of this reformulation is that it eases the burden of users for determining the regularization parameter before solving the problem described in Eq. (2). To speak further, the users are able to exact more knowledge of the problem and choose their preferred solution from the Pareto optimal set of Eq. (28).

However, it is unadvisable for us to study multi-objective optimization trapped within regularization models. In fact, the concrete forms of fidelity term and penalty term are not mentioned in analyzing their contradictory relationship. That is to say, we might as well get rid of thinking bondage from regularization model and directly confront the ill-posed problem described in Eq. (1). We can construct multiple objectives from the original ill-posed problem Eq. (1), according to the application requirements and even use some of them at the same time. In general, the multi-objective problem can be modelled as

$$\min_{\boldsymbol{x}} \{\boldsymbol{H}(\boldsymbol{x})\}, \tag{29}$$

where $\boldsymbol{H}(\boldsymbol{x})$ is a vector of no less than two objectives. It should include a fit-to-data objective which reproduces the major information of the actual solution and a elaborately designed objective which collects our desirous details. Certainly, various objectives can be designed to pursue different types of details such as texture and smoothness.

The most popular mean square error (MSE) and relative error can be used as fit-to-data objective, these error measurements with other space norms can also be used as fit-to-date objective. Generally speaking, the construction of multiple objectives should be conflict to each other, and objectives should be convenient for practical operation. For example, two unusual conflict objectives for restraining noise and preserving details separately are proposed in Ref. [290] for better performance of change detection.

However, the multiple objectives are usually not easy to tackle because of the large dimension of the decision vector. It seems impossible for global search schemes to search for global optimal in such a large decision space within an acceptable time cost. Therefore, some schemes such as clustering process [291] and sparse representation are used to reduce the dimension of decision space.

5 A simple case study on multi-objective sparse reconstruction

In this section, we mainly investigate the performance of multi-objective optimization together with EA used in signal recovery. For simplicity, we use the sparse constraint and the square error as two conflict objectives as follows:

$$\min_{\boldsymbol{x}} \boldsymbol{H}(\boldsymbol{x}) = \left\{\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2, \|\boldsymbol{x}\|_0\right\}. \tag{30}$$

5.1 Effects of dimension of observation matrix $\boldsymbol{A}$ on signal recovery by EMO

As we know, how much information can be observed from signal is determined by the dimension of observation matrix $\boldsymbol{A}$, that is, of course, the rows number of Matrix $\boldsymbol{A}$. In the condition of less than or equal to the number of columns, the more rows $\boldsymbol{A}$ has, the more information of signal will be recorded. Nevertheless, the reservation of the signal does not

need to be observed completely because of the sparsity of signal. Therefore, we intend to investigate the effects of rows number of matrix $\boldsymbol{A}$ on signal recovery.

The experiment is set as follows: the length of signal is fixed at 2 000, the sparsity ratio ($S-ratio$) of signal is fixed at 0.05, 0.15, 0.30 separately. The fidelity term $\|\boldsymbol{Ax}-\boldsymbol{y}\|_2^2$ and relative error $\|\boldsymbol{x}-\boldsymbol{x}_0\|^2/\|\boldsymbol{x}_0\|^2$ (RE) versus the sparsity ratio of recovery signal $\boldsymbol{x}$ are shown in Fig. 3.

From the first column denoted by $S-ratio=0.05$, we can find that the optimal solutions in the knee region of the Pareto front have almost the same sparsity ratio as the exact signal $\boldsymbol{x}_0$. What's more, the value of fidelity term $\|\boldsymbol{Ax}-\boldsymbol{y}\|_2^2$ of solutions in the knee region is getting closer to the noise level $\|\boldsymbol{Ax}_0-\boldsymbol{y}\|_2^2=\Delta$ as rows number of matrix $\boldsymbol{A}$ increases. Therefore, the optimal solutions in knee region have not only the similar sparsity to the exact signal but also the acceptable fidelity error referring to the noise level. These conclusions can also be obtained by analyzing details of column $S-ratio=0.15$ and $S-ratio=0.25$. In addition, the relative error of Pareto optimal solution increases with the increasing sparsity ratio of $\boldsymbol{x}_0$ and decreases with the increasing rows number of matrix $\boldsymbol{A}$, seeing the lower left corner subfigure of Fig. 3 exhibits perfect performance of recovery. Furtherly, from each row which has the same scale matrix $\boldsymbol{A}$ of Fig. 3, we can simply find that the knee region moves toward the corresponding area of the exact signal, and thus the solution in knee region will maintain the similar sparsity ratio as the sparsity ratio of the exact signal changes from 0.05 to 0.15, then to 0.25, But these details are not enough to draw a conclusion. We are intend to anatomize the effect of signal sparsity ratio on its recovery.

5.2 Effects of sparsity ratio of signals on its recovery by EMO

In the previous experiment, we have aware of that the knee region of Pareto front will follow suit of the exact signal in

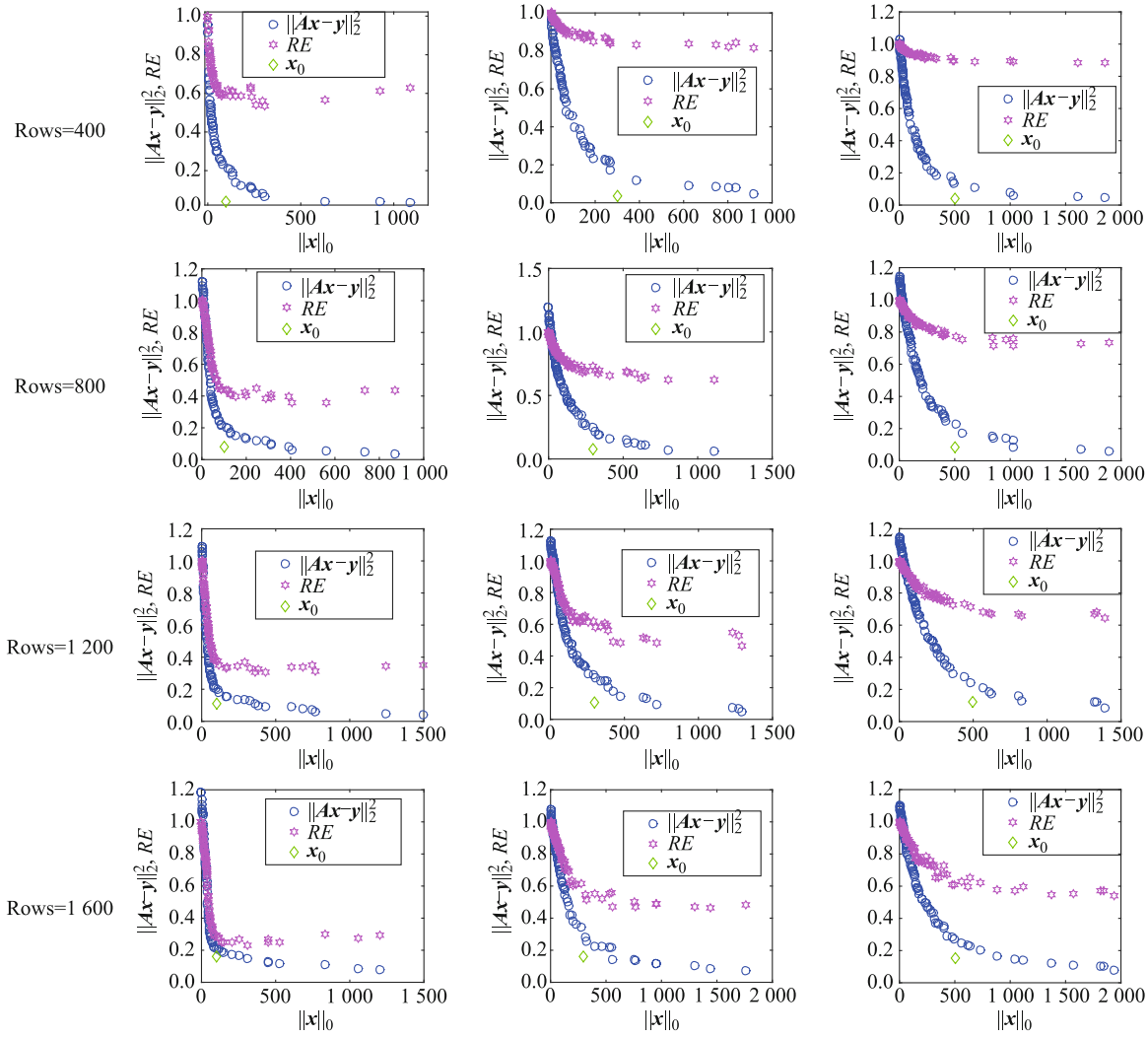

Fig. 3 Under different sparsity ratio settings, the Pareto front of bi-objective problem described in Eq. (30) changes as the rows number of matrix $\boldsymbol{A}$ increases. For the left, middle and right column, $S-ratio$ are 0.05, 0.15 and 0.25, respectively

changing sparsity ratio. We are going to explore deeply by further experiments. Going on with the experiments of last three rows shown in Fig. 3, we continue to increase the sparsity ratio of signal $\boldsymbol{x}_0$ and see the changes of the knee region of Pareto front. By combining the subfigures with the same rows number in Figs. 3 and 4 together, we can observe that optimal solutions in the knee region of Pareto front can keep pace well with the exact signal $\boldsymbol{x}_0$ in sparsity ratio when the sparsity of $\boldsymbol{x}_0$ is strong, e.g., subfigures of the same row in Fig. 3. However, as the sparsity ratio of $\boldsymbol{x}_0$ continues to increase, the solutions in knee region are unable to well approximate the exact signal $\boldsymbol{x}_0$ any longer because they are far away from $\boldsymbol{x}_0$, e.g., subfigures of the same row in Fig. 4. Nevertheless, although we can not approximate the exact signal by solutions in the knee region of Pareto front, we can still use other solutions in the Pareto front to achieve good approximation because there are still some other solutions in Pareto front close to $\boldsymbol{x}_0$, e.g., Fig. 4.

From the above discussion, we can conclude that the exact signal can be well approximated by solutions in the knee region of Pareto front if it is sparse enough. If the signal is not sparse enough, we can still choose a good approximation of it from the Pareto front as long as we have priori knowledge about the sparsity ratio of the signal.

5.3 Average estimation error of EMO

We still use EMO to reconstruct the signal repeatedly and compare its average estimation error (estimation of average relative error) with these of other advanced sparse reconstruction methods read as Orthogonal Matching Pursuit (OMP) [215], Homotopy Methods [230], L1LS [229], PFP [292], Alternating Direction Method (ADM) [212], SPARSA [293], and FISTA [81], respectively.

The experiments are set as follows: the length of exact signal is fixed at 2 000 and the trial is repeated ten times for each experiment. For the first experiment, the sparsity ratio of exact signal is fixed at 0.10, the average estimation errors of eight algorithms are investigated by varying rows number of matrix $\boldsymbol{A}$ from 500 to 1 900 at a step of 100. For the second experiment, the dimension of matrix $\boldsymbol{A}$ is fixed at $1\,600 * 2\,000$, the average estimation errors of eight algorithms are investigated by varying signal sparsity ratio from 0.1 to 0.6 at a step of 0.05. The average estimation errors of eight algorithms in the above two experiments are shown in Figs. 5(a) and 5(b).

As shown in Fig. 5(a), when the rows number of matrix $\boldsymbol{A}$ is small, the average estimation errors of EMO and other seven algorithms will reduce more or less as the rows number

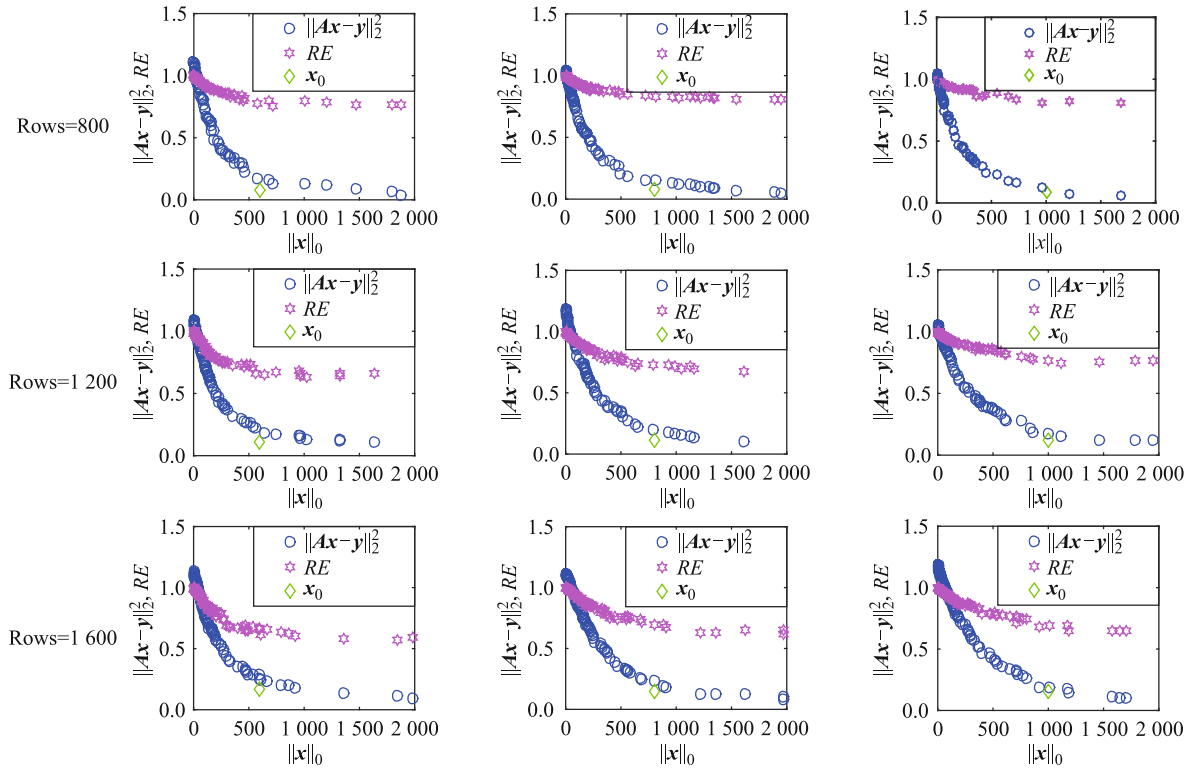

Fig. 4 Under different rows number settings of matrix $\boldsymbol{A}$, the Pareto front of bi-objective problem described in Eq. (30) changes as sparsity ratio of $\boldsymbol{x}_0$ increases. For the left, middle and right column, *S-ratio* are 0.05, 0.15 and 0.25, respectively

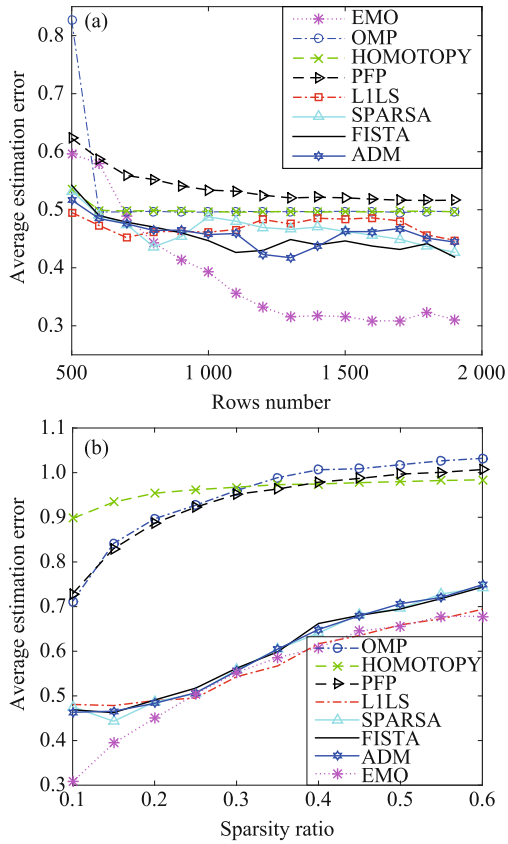

Fig. 5 Average estimation error of EMO compared with these of other seven algorithms versus (a) varying rows number of matrix $\boldsymbol{A}$, (b) varying sparsity ratio of signal $\boldsymbol{x}_0$

increases, this is because the increased scale of matrix $\boldsymbol{A}$ is able to record some extra information of the exact signal. As the rows number continues to increase, the average estimation errors of Homotopy and OMP algorithms are the earliest to stop to reduce, closely followed by these of other algorithms except for EMO, the average estimation error of EMO does not stop to reduce untill the rows number increases to 1 300. What's more, the average estimation error of the EMO gets to less than these of other algorithms when rows number exceeds 800. The average estimation error of EMO is much less than these of other algorithms when all of them are stabilized at their stage.

As shown in Fig. 5(b), The average estimation errors of OMP, Homotopy and PFP algorithms are obviously lager than these of other five algorithms. For small sparsity ratio of signal, the EMO outperforms other four algorithms in accuracy. As the sparsity ratio increases, the average estimation errors of all the eight algorithms will increase at the same time. Nevertheless, the EMO algorithm can match the L1LS algorithm which is the optimal in other seven algorithms.

To summarize, The EMO outperforms the other seven sparse reconstruction algorithms in reconstruction accuracy against the varying rows number of observation matrix and varying sparsity ratio of signal on the whole.

6 Concluding remarks

In this paper, we illustrated a purpose-driven modeling process of regularization, during which we explained the construction of regularization matrix and the effects of regularization parameter. Then, we reviewed the most widely used regularization models such as TSVD, iterative regularization, variational regularization and so on. Meanwhile, we explained the regularizing effects of these models and summarized their corresponding optimization methods. In the latter part of this paper, we established a multi-objective framework for ill-posed problems after some in-depth analysis on regularization model. A case study on signal recovery showed favorable results and demonstrated the high accuracy of EMO compared with some advanced algorithms.

There are many merits of using EMO framework for ill-posed problems. Firstly, the choice of regularization parameter in regularization model is avoided. Secondly, the EMO framework can obtain a set of nondominated solutions in a single run, which are conducive to decision-making. Finally, EMO can well handle complex problems [234, 294, 295]. Recently, researchers have realized the effectiveness of EMO in signal processing [234], image processsing [296] and machine learning [291, 297–300]. In these applications, the objectives were constructed to meet the demand and purpose of problem, the algorithms were also designed to generate diverse optional objective combinations. In addition, the EMO together with designed algorithms exhibited superiority in some aspects. For example, in Ref. [296], the selected bands are supposed to have less numbers and contain much information simultaneously in band selection for hyperspectral images. Therefore, the number of selected bands and information entropy are designed as two conflict objectives. The designed EMO algorithm can obtain a series of band subsets with different numbers. Researchers have also proved that EMO can overcome the difficulty that limits the use of greedy algorithm in some NP-hard problems [294, 295].

We hope to explore more practical use of MOO in ill-posed problems because of its advantages. There are some guidelines for applying MOO to a specific ill-posed problem. Firstly, multiple objectives should be designed as contradictory as possible to model a MOP according to the demands of problem. Secondly, efficient MOO algorithms should be designed to optimize the established MOP. EAs are encouraged

to apply to MOO, because they can well handle complex features of problems such as discontinuity, non-convexity, non-convex feasible region and so on. Cooperating with EAs, there are several EMO frameworks to choose from, specifically, domination based framework, indicator based framework and decomposition based framework. We should select proper EMO framework and design appropriate schemes for diversity and convergence according to characteristics of the problem. Finally, priori information for a specific problem can be introduced in the process of EMO such as the initialization, the design of genetic operators, the selection of the final solutions. Following the above guidelines, we are going to exploit effective algorithms for further use of EMO in image processing, signal processing and machine learning.

Acknowledgements This work was supported in part by the National Nature Science Foundation of China (Grant No. 61422209), in part by the National Program for Support of Top-Notch Young Professionals of China.

References

1. Hadamard J. Sur les Problemes aux Derivees Partielles et Leur Signification Physique. Princeton University Bulletin, 1902, 13: 49–52
2. Kabanikhin S I. Inverse and Ill-Posed Problems: Theory and Applications. Berlin: Water De Gruyter, 2011
3. Zhang B Y, Xu D H, Liu T W. Stabilized algorithms for ill-posed problems in signal processing. In: Proceedings of the IEEE International Conferences on Info-tech and Info-net. 2001, 1: 375–380
4. Scherzer O. Handbook of Mathematical Methods in Imaging. Springer Science & Business Media, 2011
5. Groetsch C W. Inverse problems in the mathematical sciences. Mathematics of Computation, 1993, 63(5): 799–811
6. Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena, 1992, 60(1): 259–268
7. Tikhonov A N. Solution of incorrectly formulated problems and the regularization method. Soviet Math, 1963, 4: 1035–1038
8. Tikhonov A N, Arsenin V I. Solutions of Ill-posed Problems. Washington, DC: V. H. Winston & Sons, 1977
9. Landweber L. An iteration formula for Fredholm integral equations of the first kind. American Journal of Mathematics, 1951, 73(3): 615–624
10. Hestenes M R, Stiefel E. Methods of conjugate gradients for solving linear systems. Journal of Research of the National Bureau of Standards, 1952, 49(6): 409–436
11. Vogel C R. Computational Methods for Inverse Problems. Society for Industrial and Applied Mathematics, 2002, 23
12. Hansen P C. The truncated SVD as a method for regularization. Bit Numerical Mathematics, 1987, 27(4): 534–553
13. Honerkamp J, Weese J. Tikhonovs regularization method for ill-posed problems. Continuum Mechanics and Thermodynamics, 1990, 2(1): 17–30
14. Zhang X Q, Burger M, Bresson X, Osher S. Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. SIAM Journal on Imaging Sciences, 2010, 3(3): 253–276
15. Deb K. Multi-Objective Optimization Using Evolutionary Algorithms. New York: John Wiley & Sons, 2001, 16
16. Fonseca C M, Fleming P J. An overview of evolutionary algorithms in multiobjective optimization. Evolutionary Computation, 1995, 3(1): 1–16
17. Coello C A C, Van Veldhuizen D A, Lamont G B. Evolutionary Algorithms for Solving Multi-objective Problems. New York: Kluwer Academic, 2002
18. Tan K C, Khor E F, Lee T H. Multiobjective Evolutionary Algorithms and Applications. Springer Science & Business Media, 2005
19. Knowles J, Corne D, Deb K. Multiobjective Problem Solving from Nature: from Concepts to Applications. Springer Science & Business Media, 2008
20. Raquel C, Yao X. Dynamic multi-objective optimization: a survey of the state-of-the-art. In: Yang S X, Yao X, eds. Evolutionary Computation for Dynamic Optimization Problems. Springer Berlin Heidelberg, 2013, 85–106
21. Lücken C V, Barán B, Brizuela C. A survey on multi-objective evolutionary algorithms for many-objective problems. Computational Optimization and Applications, 2014, 58(3): 707–756
22. Hwang C L, Masud A S M. Multiple Objective Decision Making-Methods and Applications. Springer Science & Business Media, 1979, 164
23. Girosi F, Jones M B, Poggio T. Regularization theory and neural networks architectures. Neural Computation, 1995, 7(2): 219–269
24. Belge M, Kilmer M E, Miller E L. Efficient determination of multiple regularization parameters in a generalized L-curve framework. Inverse Problems, 2002, 18(4): 1161
25. Hansen P C. Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion. American Mathematical Monthly, 1997, 4(5): 491
26. Eriksson P, Jiménez C, Buehler S A. Qpack, a general tool for instrument simulation and retrieval work. Journal of Quantitative Spectroscopy and Radiative Transfer, 2005, 91(1): 47–64
27. Giusti E. Minimal Surfaces and Functions of Bounded Variation. Springer Science & Business Media, 1984, 80
28. Catté F, Coll T. Image selective smoothing and edge detection by nonlinear diffusion. SIAM Journal on Numerical Analysis, 1992, 29(1): 182–193
29. Björck A. Numerical Methods for Least Squares Problems. Society for Industrial and Applied Mathematics, 1996
30. Groetsch C W. The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind. Pitman Advanced Publishing Program, 1984
31. Hanson R J. A numerical method for solving Fredholm integral equations of the first kind using singular values. SIAM Journal on Numerical Analysis, 1971, 8(3): 616–622
32. Stewart G W. Rank degeneracy. SIAM Journal on Scientific and Statistical Computing, 1984, 5(2): 403–413
33. Hansen P C, Sekii T, Shibahashi H. The modified truncated SVD method for regularization in general form. SIAM Journal on Scien-

tific and Statistical Computing, 1992, 13(5): 1142–1150

34. Van Loan C F. Generalizing the singular value decomposition. SIAM Journal on Numerical Analysis, 1976, 13(1): 76–83
35. Hansen P C. Regularization, GSVD and truncated GSVD. BIT Numerical Mathematics, 1989, 29(3): 491–504
36. Paige C C. Computing the generalized singular value decomposition. SIAM Journal on Scientific and Statistical Computing, 1986, 7(4): 1126–1146
37. Morigi S, Reichel L, Sgallari F. A truncated projected SVD method for linear discrete ill-posed problems. Numerical Algorithms, 2006, 43(3): 197–213
38. Fernando K V, Hammarling S. A product induced singular value decomposition (ΠSVD) for two matrices and balanced realization. In: Proceedings of SIAM Conference on Linear Algebra in Signals, Systems and Control. 1988, 128–140
39. Zha H Y. The restricted singular value decomposition of matrix triplets. SIAM Journal on Matrix Analysis and Applications, 1991, 12(1): 172–194
40. De Moor B, Golub G H. The restricted singular value decomposition: properties and applications. SIAM Journal on Matrix Analysis and Applications, 1991, 12(3): 401–425
41. De Moor B, Zha H Y. A tree of generalizations of the ordinary singular value decomposition. Linear Algebra and Its Applications, 1991, 147: 469–500
42. De Moor B. Generalizations of the OSVD: structure, properties and applications. In: Vaccaro R J, ed. SVD & Signal Processing, II: Algorithms, Analysis & Applications. 1991, 83–98
43. Noschese S, Reichel L. A modified TSVD method for discrete ill-posed problems. Numerical Linear Algebra with Applications, (in press)
44. Dykes L, Noschese S, Reichel L. Rescaling the GSVD with application to ill-posed problems. Numerical Algorithms, 2015, 68(3): 531–545
45. Edo L, Franco W, Martinsson P G, Rokhlin V, Tygert M. Randomized algorithms for the low-rank approximation of matrices. Proceedings of the National Academy of Sciences, 2007, 104(51): 20167–20172
46. Woolfe F, Liberty E, Rokhlin V, Tygert M. A fast randomized algorithm for the approximation of matrices. Applied & Computational Harmonic Analysis, 2008, 25(3): 335–366
47. Sifuentes J, Gimbutas Z, Greengard L. Randomized methods for rank-deficient linear systems. Electronic Transactions on Numerical Analysis, 2015, 44: 177–188
48. Liu Y G, Lei Y J, Li C G, Xu W Z, Pu Y F. A random algorithm for low-rank decomposition of large-scale matrices with missing entries. IEEE Transactions on Image Processing, 2015, 24(11): 4502–4511
49. Sekii T. Two-dimensional inversion for solar internal rotation. Publications of the Astronomical Society of Japan, 1991, 43: 381–411
50. Scales J A. Uncertainties in seismic inverse calculations. In: Jacobsen B H, Mosegaard K, Sibani P, eds. Inverse Methods. Berlin: Springer-Verlag, 1996, 79–97
51. Lawless J F, Wang P. A simulation study of ridge and other regression estimators. Communications in Statistics-Theory and Methods, 1976, 5(4): 307–323
52. Dempster A P, Schatzoff M, Wermuth N. A simulation study of alternatives to ordinary least squares. Journal of the American Statistical Association, 1977, 72(357): 77–91
53. Hansen P C, O'Leary D P. The use of the L-curve in the regularization of discrete ill-posed problems. SIAM Journal on Scientific Computing, 1993, 14(6): 1487–1503
54. Hansen P C. Analysis of discrete ill-posed problems by means of the L-curve. SIAM Review, 1992, 34(4): 561–580
55. Xu P L. Truncated SVD methods for discrete linear ill-posed problems. Geophysical Journal International, 1998, 135(2): 505–514
56. Wu Z M, Bian S F, Xiang C B, Tong Y D. A new method for TSVD regularization truncated parameter selection. Mathematical Problems in Engineering, 2013
57. Chicco D, Masseroli M. A discrete optimization approach for SVD best truncation choice based on ROC curves. In: Proceedings of the 13th IEEE International Conference on Bioinformatics and Bioengineering. 2013: 1–4
58. Golub G H, Heath M, Wahba G. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics, 1979, 21(2): 215–223
59. Jbilou K, Reichel L, Sadok H. Vector extrapolation enhanced TSVD for linear discrete ill-posed problems. Numerical Algorithms, 2009, 51(2): 195–208
60. Bouhamidi A, Jbilou K, Reichel L, Sadok H, Wang Z. Vector extrapolation applied to truncated singular value decomposition and truncated iteration. Journal of Engineering Mathematics, 2015, 93(1): 99–112
61. Vogel C R. Computational Methods for Inverse Problems. Society for Industrial and Applied Mathematics, 2002
62. Doicu A, Trautmann T, Schreier F. Numerical Regularization for Atmospheric Inverse Problems. Springer Science & Business Media, 2010
63. Bakushinsky A B, Goncharsky A V. Iterative Methods for the Solution of Incorrect Problems. Moscow: Nauka, 1989
64. Rieder A. Keine Probleme mit Inversen Problemen: Eine Einführung in ihre stabile Lösung. Berlin: Springer-Verlag, 2013
65. Nemirovskiy A S, Polyak B T. Iterative methods for solving linear ill-posed problems under precise information. Engineering Cybernetics, 1984, 22(4): 50–56
66. Brakhage H. On ill-posed problems and the method of conjugate gradients. Inverse and Ill-posed Problems, 1987, 4: 165–175
67. Hanke M. Accelerated Landweber iterations for the solution of ill-posed equations. Numerische Mathematik, 1991, 60(1): 341–373
68. Barzilai J, Borwein J M. Two-point step size gradient methods. IMA Journal of Numerical Analysis, 1988, 8(1): 141–148
69. Axelsson O. Iterative Solution Methods. Cambridge: Cambridge University Press, 1996
70. Van der Sluis A, Van der Vorst H A. The rate of convergence of conjugate gradients. Numerische Mathematik, 1986, 48(5): 543–560
71. Scales J A, Gersztenkorn A. Robust methods in inverse theory. Inverse Problems, 1988, 4(4): 1071–1091
72. Björck Å, Eldén L. Methods in numerical algebra for ill-posed problems. Technical Report LiTH-MAT-R-33-1979. 1979
73. Trefethen L N, Bau D. Numerical Linear Algebra. Society for Industrial and Applied Mathematics, 1997
74. Calvetti D, Lewis B, Reichel L. On the regularizing properties of the

GMRES method. Numerische Mathematik, 2002, 91(4): 605–625

75. Calvetti D, Lewis B, Reichel L. Alternating Krylov subspace image restoration methods. Journal of Computational and Applied Mathematics, 2012, 236(8): 2049–2062
76. Brianzi P, Favati P, Menchi O, Romani F. A framework for studying the regularizing properties of Krylov subspace methods. Inverse Problems, 2006, 22(3): 1007–1021
77. Sonneveld P, Van Gijzen M B. IDR(s): a family of simple and fast algorithms for solving large nonsymmetric systems of linear equations. SIAM Journal on Scientific Computing, 2008, 31(2): 1035–1062
78. Fong D C L, Saunders M. LSMR: an iterative algorithm for sparse least-squares problems. SIAM Journal on Scientific Computing, 2011, 33(5): 2950–2971
79. Zhao C, Huang T Z, Zhao X L, Deng L J. Two new efficient iterative regularization methods for image restoration problems. Abstract & Applied Analysis, 2013
80. Perez A, Gonzalez R C. An iterative thresholding algorithm for image segmentation. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1987, 9(6): 742–751
81. Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences, 2009, 2(1): 183–202
82. Bioucas-Dias J M, Figueiredo M A T. Two-step algorithms for linear inverse problems with non-quadratic regularization. In: Proceedings of the IEEE International Conference on Image Processing. 2007, 105–108
83. Bioucas-Dias J M, Figueiredo M A T. A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration. IEEE Transactions on Image Processing. 2007, 16(12): 2992–3004
84. Bayram I, Selesnick I W. A subband adaptive iterative shrinkage/thresholding algorithm. IEEE Transactions on Signal Processing, 2010, 58(3): 1131–1143
85. Yamagishi M, Yamada I. Over-relaxation of the fast iterative shrinkage-thresholding algorithm with variable stepsize. Inverse Problems, 2011, 27(10): 105008–105022
86. Bhotto M Z A, Ahmad M O, Swamy M N S. An improved fast iterative shrinkage thresholding algorithm for image deblurring. SIAM Journal on Imaging Sciences, 2015, 8(3): 1640–1657
87. Zhang Y D, Dong Z C, Phillips P, Wang S H, Ji G L, Yang J Q. Exponential wavelet iterative shrinkage thresholding algorithm for compressed sensing magnetic resonance imaging. Information Sciences, 2015, 322: 115–132
88. Zhang Y D, Wang S H, Ji G L, Dong Z C. Exponential wavelet iterative shrinkage thresholding algorithm with random shift for compressed sensing magnetic resonance imaging. IEEJ Transactions on Electrical and Electronic Engineering, 2015, 10(1): 116–117
89. Wu G M, Luo S Q. Adaptive fixed-point iterative shrinkage/thresholding algorithm for MR imaging reconstruction using compressed sensing. Magnetic Resonance Imaging, 2014, 32(4): 372–378
90. Fang E X, Wang J J, Hu D F, Zhang J Y, Zou W, Zhou Y. Adaptive monotone fast iterative shrinkage thresholding algorithm for fluorescence molecular tomography. IET Science Measurement Technology, 2015, 9(5): 587–595
91. Zuo W M, Meng D Y, Zhang L, Feng X C, Zhang D. A generalized iterated shrinkage algorithm for non-convex sparse coding. In: Proceedings of the IEEE International Conference on Computer Vision. 2013, 217–224
92. Krishnan D, Fergus R. Fast image deconvolution using hyper-laplacian priors. In: Bengio Y, Schuurmans D, Lafferty J D, et al., eds. Advances in Neural Information Processing Systems 22. 2009, 1033–1041
93. Chartrand R, Yin W. Iteratively reweighted algorithms for compressive sensing. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. 2008, 3869–3872
94. She Y Y. An iterative algorithm for fitting nonconvex penalized generalized linear models with grouped predictors. Computational Statistics & Data Analysis, 2012, 56(10): 2976–2990
95. Gong P H, Zhang C S, Lu Z S, Huang J Z, Ye J P. A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. In: Proceedings of International Conference on Machine Learning. 2013, 37–45
96. Bredies K, Lorenz D A. Linear convergence of iterative soft-thresholding. Journal of Fourier Analysis and Applications, 2008, 14(5–6): 813–837
97. Kowalski M. Thresholding rules and iterative shrinkage/thresholding algorithm: a convergence study. In: Proceedings of the IEEE International Conference on Image Processing. 2014, 4151–4155
98. Chambolle A, Dossal C. On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm". Journal of Optimization Theory & Applications, 2015, 166(3): 968–982
99. Johnstone P R, Moulin P. Local and global convergence of an inertial version of forward-backward splitting. Advances in Neural Information Processing Systems, 2014, 1970–1978
100. Morozov V A. On the solution of functional equations by the method of regularization. Soviet Mathematics Doklady, 1966, 7(11): 414–417
101. Vainikko G M. The discrepancy principle for a class of regularization methods. USSR Computational Mathematics and Mathematical Physics, 1982, 22(3): 1–19
102. Vainikko G M. The critical level of discrepancy in regularization methods. USSR Computational Mathematics and Mathematical Physics, 1983, 23(6): 1–9
103. Plato R. On the discrepancy principle for iterative and parametric methods to solve linear ill-posed equations. Numerische Mathematik, 1996, 75(1): 99–120
104. Borges L S, Bazán F S V, Cunha M C C. Automatic stopping rule for iterative methods in discrete ill-posed problems. Computational & Applied Mathematics, 2015, 34(3): 1175–1197
105. Dziwoki G, Izydorczyk J. Stopping criteria analysis of the OMP algorithm for sparse channels estimation. In: Proceedings of the International Conference on Computer Networks. 2015, 250–259
106. Favati P, Lotti G, Menchi O, Romani F. Stopping rules for iterative methods in nonnegatively constrained deconvolution. Applied Numerical Mathematics, 2014, 75: 154–166
107. Engl H W, Hanke M, Neubauer A. Regularization of Inverse Problems. Springer Science & Business Media, 1996
108. Amster P. Iterative Methods. Universitext, 2014, 53–82
109. Waseem M. On some iterative methods for solving system of nonlinear equations. Dissertation for the Doctoral Degree. Islamabad:

COMSATS Institute of Information Technology, 2012

110. Burger M, Osher S. A guide to the TV zoo. In: Burger M, Mennucci A C G, Osher S, et al., eds. Level Set and PDE Based Reconstruction Methods in Imaging. Springer International Publishing, 2013, 1–70
111. Tikhonov A N. Regularization of incorrectly posed problems. Soviet Mathematics Doklady, 1963, 4(1): 1624–1627
112. Nikolova M. Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares. Multiscale Modeling & Simulation, 2005, 4(3): 960–991
113. Burger M, Osher S. Convergence rates of convex variational regularization. Inverse Problems, 2004, 20(5): 1411–1421
114. Hofmann B, Kaltenbacher B, Pöschl C, Scherzer O. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. Inverse Problems, 2007, 23(3): 987–1010
115. Resmerita E. Regularization of ill-posed problems in Banach spaces:convergence rates. Inverse Problems, 2005, 21(4): 1303–1314
116. Resmerita E, Scherzer O. Error estimates for non-quadratic regularization and the relation to enhancement. Inverse Problems, 2006, 22(3): 801–814
117. Engl H W. Discrepancy principles for Tikhonov regularization of ill-posed problems leading to optimal convergence rates. Journal of Optimization Theory and Applications, 1987, 52(2): 209–215
118. Gfrerer H. An a posteriori parameter choice for ordinary and iterated Tikhonov regularization of ill-posed problems leading to optimal convergence rates. Mathematics of Computation, 1987, 49(180): 507–522
119. Natterer F. Error bounds for Tikhonov regularization in Hilbert scales. Applicable Analysis, 1984, 18(1–2): 29–37
120. Neubauer A. An a posteriori parameter choice for Tikhonov regularization in the presence of modeling error. Applied Numerical Mathematics, 1988, 4(6): 507–519
121. Engl H W, Kunisch K, Neubauer A. Convergence rates for Tikhonov regularisation of non-linear ill-posed problems. Inverse Problems, 1989, 5(4): 523–540
122. Scherzer O, Engl H W, Kunisch K. Optimal a posteriori parameter choice for Tikhonov regularization for solving nonlinear ill-posed problems. SIAM Journal on Numerical Analysis, 1993, 30(6): 1796–1838
123. Varah J M. On the numerical solution of ill-conditioned linear systems with applications to ill-posed problems. SIAM Journal on Numerical Analysis, 1973, 10(2): 257–267
124. Vinod H D, Ullah A. Recent Advances in Regression Methods. Danbury: Marcel Dekker Incorporated, 1981
125. O'Sullivan F. A statistical perspective on ill-posed inverse problems. Statistical Science, 1986, 1(4): 502–518
126. Grafarend E W, Schaffrin B. Ausgleichungsrechnung in linearen modellen. BI Wissenschaftsverlag Mannheim, 1993
127. Rodgers C D. Inverse Methods for Atmospheric Sounding: Theory and Practice. Singapore: World Scientific, 2000
128. Ceccherini S. Analytical determination of the regularization parameter in the retrieval of atmospheric vertical profiles. Optics Letters, 2005, 30(19): 2554–2556
129. Mallows C L. Some comments on C_p. Technometrics, 1973, 15(4): 661–675
130. Rice J. Choice of smoothing parameter in deconvolution problems. Contemporary Mathematics, 1986, 59: 137–151
131. Hanke M, Raus T. A general heuristic for choosing the regularization parameter in ill-posed problems. SIAM Journal on Scientific Computing, 1996, 17(4): 956–972
132. Wu L M. A parameter choice method for Tikhonov regularization. Electronic Transactions on Numerical Analysis, 2003, 16: 107–128
133. Gao W, Yu K P. A new method for determining the Tikhonov regularization parameter of load identification. In: Proceedings of the International Symposium on Precision Engineering Measurement and Instrumentation. 2015
134. Ito K, Jin B, Takeuchi T. Multi-parameter Tikhonov regularization-an augmented approach. Chinese Annals of Mathematics, Series B, 2014, 35(03): 383–398
135. Jin B, Lorenz D A. Heuristic parameter-choice rules for convex variational regularization based on error estimates. SIAM Journal on Numerical Analysis, 2010, 48(3): 1208–1229
136. Pazos F, Bhaya A. Adaptive choice of the Tikhonov regularization parameter to solve ill-posed linear algebraic equations via Liapunov optimizing control. Journal of Computational and Applied Mathematics, 2015, 279: 123–132
137. Hämarik U, Palm R, Raus T. A family of rules for parameter choice in Tikhonov regularization of ill-posed problems with inexact noise level. Journal of Computational and Applied Mathematics, 2012, 236(8): 2146–2157
138. Reichel L, Rodriguez G. Old and new parameter choice rules for discrete ill-posed problems. Numerical Algorithms, 2013, 63(1): 65–87
139. Kryanev A V. An iterative method for solving incorrectly posed problems. USSR Computational Mathematics and Mathematical Physics, 1974, 14(1): 24–35
140. King J T, Chillingworth D. Approximation of generalized inverses by iterated regularization. Numerical Functional Analysis & Optimization, 1979, 1(5): 499–513
141. Fakeev A G. A class of iterative processes for solving degenerate systems of linear algebraic equations. USSR Computational Mathematics and Mathematical Physics, 1981, 21(3): 15–22
142. Brill M, Schock E. Iterative solution of ill-posed problems: a survey. In: Proceedings of the 4th International Mathematical Geophysics Seminar. 1987
143. Hanke M, Groetsch C W. Nonstationary iterated Tikhonov regularization. Journal of Optimization Theory and Applications, 1998, 98(1): 37–53
144. Lampe J, Reichel L, Voss H. Large-scale Tikhonov regularization via reduction by orthogonal projection. Linear Algebra and Its Applications, 2012, 436(8): 2845–2865
145. Reichel L, Yu X B. Tikhonov regularization via flexible Arnoldi reduction. BIT Numerical Mathematics, 2015, 55(4): 1145–1168
146. Huang G, Reichel L, Yin F. Projected nonstationary iterated Tikhonov regularization. BIT Numerical Mathematics, 2016, 56(2): 467–487
147. Ambrosio L, Fusco N, Pallara D. Functions of Bounded Variation and Free Discontinuity Problems. Oxford: Oxford University Press, 2000
148. Acar R, Vogel C R. Analysis of bounded variation penalty methods for ill-posed problems. Inverse Problems, 1997, 10(6): 1217–1229
149. Hunt B R. The application of constrained least squares estimation to

image restoration by digital computer. IEEE Transactions on Computers, 1973, 100(9): 805–812

150. Demoment G. Image reconstruction and restoration: overview of common estimation structures and problems. IEEE Transactions on Acoustics, Speech and Signal Processing, 1989, 37(12): 2024–2036
151. Katsaggelos A K. Iterative image restoration algorithms. Optical Engineering, 1989, 28(7): 735–748
152. Katsaggelos A K, Biemond J, Schafer R W, Mersereau R M. A regularized iterative image restoration algorithm. IEEE Transactions on Signal Processing, 1991, 39(4): 914–929
153. Babacan S D, Molina R, Katsaggelos A K. Parameter estimation in TV image restoration using variational distribution approximation. IEEE Transactions on Image Processing, 2008, 17(3): 326–339
154. Wen Y W, Chan R H. Parameter selection for total-variation-based image restoration using discrepancy principle. IEEE Transactions on Image Processing, 2012, 21(4): 1770–1781
155. Chen A, Huo B M, Wen C W. Adaptive regularization for color image restoration using discrepancy principle. In: Proceedings of the IEEE International Conference on Signal processing, Comminications and Computing. 2013, 1–6
156. Lin Y, Wohlberg B, Guo H. UPRE method for total variation parameter selection. Signal Processing, 2010, 90(8): 2546–2551
157. Stein C M. Estimation of the mean of a multivariate normal distribution. Annals of Statistics, 1981, 9(6): 1135–1151
158. Ramani S, Blu T, Unser M. Monte-Carlo SURE: a black-box optimization of regularization parameters for general denoising algorithms. IEEE Transactions on Image Processing, 2008, 17(9): 1540–1554
159. Palsson F, Sveinsson J R, Ulfarsson M O, Benediktsson J A. SAR image denoising using total variation based regularization with sure-based optimization of regularization parameter. In: Proceedings of the IEEE International Conference on Geoscience and Remote Sensing Symposium. 2012, 2160–2163
160. Liao H Y, Li F, Ng M K. Selection of regularization parameter in total variation image restoration. Journal of the Optical Society of America A, 2009, 26(11): 2311–2320
161. Bertalmío M, Caselles V, Rougé B, Solé A. TV based image restoration with local constraints. Journal of Scientific Computing, 2003, 19(1–3): 95–122
162. Almansa A, Ballester C, Caselles V, Haro G. A TV based restoration model with local constraints. Journal of Scientific Computing, 2008, 34(3): 209–236
163. Vogel C R, Oman M E. Iterative methods for total variation denoising. SIAM Journal on Scientific Computing, 1997, 17(1): 227–238
164. Chan T F, Golub G H, Mulet P. A nonlinear primal-dual method for total variation-based image restoration. Lecture Notes in Control & Information Sciences, 1995, 20(6): 1964–1977
165. Chambolle A. An algorithm for total variation minimization and applications. Journal of Mathematical Imaging & Vision, 2004, 20(1–2): 89–97
166. Huang Y M, Ng M K, Wen Y W. A fast total variation minimization method for image restoration. SIAM Journal on Multiscale Modeling & Simulation, 2008, 7(2): 774–795
167. Bresson X, Chan T F. Fast dual minimization of the vectorial total variation norm and applications to color image processing. Inverse Problems & Imaging, 2008, 2(4): 455–484
168. Ng M K, Qi L Q, Yang Y F, Huang Y M. On semismooth Newton's methods for total variation minimization. Journal of Mathematical Imaging & Vision, 2007, 27(3): 265–276
169. Zhu M Q, Chan T F. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. UCLA CAM Report. 2008, 8–34
170. Zhu M Q, Wright S J, Chan T F. Duality-based algorithms for total-variation-regularized image restoration. Computational Optimization and Applications, 2010, 47(3): 377–400
171. Krishnan D, Lin P, Yip A M. A primal-dual active-set method for non-negativity constrained total variation deblurring problems. IEEE Transactions on Image Processing, 2007, 16(11): 2766–2777
172. Krishnan D, Pham Q V, Yip A M. A primal-dual active-set algorithm for bilaterally constrained total variation deblurring and piecewise constant Mumford-Shah segmentation problems. Advances in Computational Mathematics, 2009, 31(1–3): 237–266
173. Osher S, Burger M, Goldfarb D, Xu J J, Yin W T. An iterative regularization method for total variation-based image restoration. Multiscale Modeling & Simulation, 2005, 4(2): 460–489
174. Goldstein T, Osher S. The split Bregman method for l_1-regularized problems. SIAM Journal on Imaging Sciences, 2009, 2(2): 323–343
175. Glowinski R, Le Tallec P. Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics. Society for Industrial and Applied Mathematics, 1989
176. Wu C C, Tai X C. Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models. SIAM Journal on Imaging Sciences, 2010, 3(3): 300–339
177. Darbon J, Sigelle M. Image restoration with discrete constrained total variation part I: fast and exact optimization. Journal of Mathematical Imaging & Vision, 2006, 26(3): 261–276
178. Duan Y P, Tai X C. Domain decomposition methods with graph cuts algorithms for total variation minimization. Advances in Computational Mathematics, 2012, 36(2): 175–199
179. Fu H Y, Ng M K, Nikolova M, Barlow J L. Efficient minimization methods of mixed l_2-l_1 and l_1-l_1 norms for image restoration. SIAM Journal on Scientific Computing, 2005, 27(6): 1881–1902
180. Goldfarb D, Yin W T. Second-order cone programming methods for total variation-based image restoration. SIAM Journal on Scientific Computing, 2005, 27(2): 622–645
181. Oliveira J P, Bioucas-Dias J M, Figueiredo M A T. Adaptive total variation image deblurring: a majorization-minimization approach. Signal Processing, 2009, 89(9): 1683–1693
182. Bioucas-Dias J M, Figueiredo M A T, Oliveira J P. Total variation-based image deconvolution: a majorization-minimization approach, In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. 2006, 861–864
183. Chan T F, Esedoglu S. Aspects of total variation regularized l_1 function approximation. SIAM Journal on Applied Mathematics, 2004, 65(5): 1817–1837
184. He L, Burger M, Osher S. Iterative total variation regularization with non-quadratic fidelity. Journal of Mathematical Imaging & Vision, 2006, 26(1–2): 167–184

185. Jonsson E, Huang S C, Chan T F. Total variation regularization in positron emission tomography. CAM Report. 1998
186. Panin V Y, Zeng G L, Gullberg G T. Total variation regulated EM algorithm. IEEE Transactions on Nuclear Science, 1999, 46(6): 2202–2210
187. Le T, Chartrand R, Asaki T J. A variational approach to reconstructing images corrupted by Poisson noise. Journal of Mathematical Imaging & Vision, 2007, 27(3): 257–263
188. Rudin L, Lions P L, Osher S. Multiplicative denoising and deblurring: theory and algorithms. In: Osher S, Paragios N, eds. Geometric Level Set Methods in Imaging, Vision, and Graphics. New York: Springer, 2003, 103–119
189. Huang Y M, Ng M K, Wen Y W. A new total variation method for multiplicative noise removal. SIAM Journal on Imaging Sciences, 2009, 2(1): 20–40
190. Bonesky T, Kazimierski K S, Maass P, Schöpfer F, Schuster T. Minimization of Tikhonov functionals in Banach spaces. Abstract & Applied Analysis, 2008, 2008(1): 1563–1569
191. Meyer Y. Oscillating Patterns in Image Processing and Nonlinear Evolution Equations, University Lecture Series. Rhode Island: American Mathematical Society, 2002
192. Blomgren P, Chen T F. Color TV: total variation methods for restoration of vector valued images. IEEE Transactions on Image Processing, 1970, 7(3): 304–309
193. Setzer S, Steidl G, Popilka B, Burgeth B. Variational methods for denoising matrix fields. In: Laidlaw D, Weickert J, eds. Visualization and Processing of Tensor Fields. Berlin: Springer Berlin Heidelberg, 2009, 341–360
194. Esedoglu S, Osher S. Decomposition of images by the anisotropic Rudin-Osher-Fatemi model. Communications on Pure and Applied Mathematics, 2004, 57(12): 1609–1626
195. Shi Y Y, Chang Q S. Efficient algorithm for isotropic and anisotropic total variation deblurring and denoising. Journal of Applied Mathematics, 2013
196. Marquina A, Osher S. Explicit algorithms for a new time dependent model based on level set motion for nonlinear deblurring and noise removal. SIAM Journal on Scientific Computing, 2000, 22(2): 387–405
197. Chan T F, Marquina A, Mulet P. High-order total variation-based image restoration. SIAM Journal on Scientific Computing, 2000, 22(2): 503–516
198. Gilboa G, Osher S. Nonlocal operators with applications to image processing. SIAM Journal on Multiscale Modeling & Simulation, 2008, 7(3): 1005–1028
199. Kindermann S, Osher S, Jones P W. Deblurring and denoising of images by nonlocal functionals. SIAM Journal on Multiscale Modeling & Simulation, 2005, 4(4): 1091–1115
200. Hu Y, Jacob M. Higher degree total variation (HDTV) regularization for image recovery. IEEE Transactions on Image Processing, 2012, 21(5): 2559–2571
201. Yang J S, Yu H Y, Jiang M, Wang G. High-order total variation minimization for interior SPECT. Inverse Problems, 2012, 28(1): 15001–15024
202. Liu X W, Huang L H. A new nonlocal total variation regularization algorithm for image denoising. Mathematics and Computers in Simulation, 2014, 97: 224–233
203. Ren Z M, He C J, Zhang Q F. Fractional order total variation regularization for image super-resolution. Signal Processing, 2013, 93(9): 2408–2421
204. Oh S, Woo H, Yun S, Kang M. Non-convex hybrid total variation for image denoising. Journal of Visual Communication & Image Representation, 2013, 24(3): 332–344
205. Donoho D L. Compressed sensing. IEEE Transactions on Information Theory, 2006, 52(4): 1289–1306
206. Candè E J, Wakin M B. An introduction to compressive sampling. IEEE Signal Processing Magazine, 2008, 25(2): 21–30
207. Tsaig Y, Donoho D L. Extensions of compressed sensing. Signal Processing, 2006, 86(3): 549–571
208. Candès E J, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory, 2006, 52(2): 489–509
209. Candès E J, Tao T. Near-optimal signal recovery from random projections: Universal encoding strategies?. IEEE Transactions on Information Theory, 2006, 52(12): 5406–5425
210. Donoho D L, Elad M. Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization. Proceedings of National Academy of Sciences, 2003, 100(5): 2197–2202
211. Wright J, Ma Y. Dense error correction via l_1-minimization. IEEE Transactions on Information Theory, 2010, 56(7): 3540–3560
212. Yang J F, Zhang Y. Alternating direction algorithms for ℓ_1-problems in compressive sensing. SIAM Journal on Scientific Computing, 2011, 33(1): 250–278.
213. Natarajan B K. Sparse approximate solutions to linear systems. SIAM Journal on Computing, 1995, 24(2): 227–234
214. Mallat S G, Zhang Z. Matching pursuits with time-frequency dictionaries. IEEE Transactions on Signal Processing, 1993, 41(12): 3397–3415
215. Tropp J, Gilbert A C. Signal recovery from random measurements via orthogonal matching pursuit. IEEE Transactions on Information Theory, 2007, 53(12): 4655–4666
216. Blumensath T, Davies M E. Iterative thresholding for sparse approximations. Journal of Fourier Analysis and Applications, 2008, 14(5–6): 629–654
217. Gorodnitsky I F, Rao B D. Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm. IEEE Transactions on Signal Processing, 1997, 45(3): 600–616
218. Bao C L, Ji H, Quan Y H, Shen Z W. l_0 norm based dictionary learning by proximal methods with global convergence. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014, 3858–3865
219. Foucart S, Lai M J. Sparsest solutions of underdetermined linear systems via l_q-minimization for $0 < q \leqslant 1$. Applied and Computational Harmonic Analysis, 2009, 26(3): 395–407
220. Cai T T, Wang L, Xu G. Shifting inequality and recovery of sparse signals. IEEE Transactions on Signal Processing, 2010, 58(3): 1300–1308
221. Cai T T, Wang L, Xu G. New bounds for restricted isometry constants. IEEE Transactions on Information Theory, 2010, 56(9): 4388–4394
222. Chen S S, Donoho D L, Saunders M A. Atomic decomposition by

basis pursuit. SIAM Journal on Scientific Computing, 1998, 20(1): 33–61

223. Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. The Annals of Statistics, 2004, 32(2): 407–499
224. Figueiredo M A T, Nowak R D. An EM algorithm for wavelet-based image restoration. IEEE Transactions on Image Processing, 2002, 12(8): 906–916
225. Starck J L, Mai K N, Murtagh F. Wavelets and curvelets for image deconvolution: a combined approach. Signal Processing, 2003, 83(10): 2279–2283
226. Herrholz E, Teschke G. Compressive sensing principles and iterative sparse recovery for inverse and ill-posed problems. Inverse Problems, 2010, 26(12): 125012–125035
227. Jin B, Lorenz D, Schiffler S. Elastic-net regularization: error estimates and active set methods. Inverse Problems, 2009, 25(11): 1595–1610
228. Figueiredo M A T, Nowak R D, Wright S J. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. IEEE Journal of Selected Topics in Signal Processing, 2007, 1(4): 586–597
229. Kim S J, Koh K, Lustig M, Boyd S, Gorinevsky D. An interior-point method for large-scale l_1-regularized least squares. IEEE Journal of Selected Topics in Signal Processing, 2007, 1(4): 606–617
230. Donoho D L, Tsaig Y. Fast solution of l_1-norm minimization problems when the solution may be sparse. IEEE Transactions on Information Theory, 2008, 54(11): 4789–4812
231. Combettes P L, Wajs E R. Signal recovery by proximal forward-backward splitting. SIAM Journal on Multiscale Modeling & Simulation, 2005, 4(4): 1168–1200
232. Becker S, Bobin J, Candés E J. NESTA: a fast and accurate first-order method for sparse recovery. SIAM Journal on Imaging Sciences, 2011, 4(1): 1–39
233. Osborne M R, Presnell B, Turlach B A. A new approach to variable selection in least squares problems. IMA Journal of Numerical Analysis, 1999, 20(3): 389–403
234. Li L, Yao X, Stolkin R, Gong M G, He S. An evolutionary multiobjective approach to sparse reconstruction. IEEE Transactions on Evolutionary Computation, 2014, 18(6): 827–845
235. Chartrand R. Exact reconstruction of sparse signals via nonconvex minimization. IEEE Signal Processing Letters, 2007, 14(10): 707–710
236. Candes E J, Tao T. Decoding by linear programming. IEEE Transactions on Information Theory, 2005, 51(12): 4203–4215
237. Saab R, Chartrand R, Yilmaz Ö. Stable sparse approximations via nonconvex optimization. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. 2008, 3885–3888
238. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B, 1996, 58(1): 267–288
239. Zhang C H. Nearly unbiased variable selection under minimax concave penalty. The Annals of Statistics, 2010, 38(2): 894–942
240. Fan J Q, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association, 2001, 96(456): 1348–1360
241. Nikolova M, Ng M K, Zhang S, Ching W K. Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization. SIAM Journal on Imaging Sciences, 2008, 1(1): 2–25
242. Frank L E, Friedman J H. A statistical view of some chemometrics regression tools. Technometrics, 1993, 35(2): 109–135
243. Fu W J. Penalized regressions: the bridge versus the lasso. Journal of Computational and Graphical Statistics, 1998, 7(3): 397–416
244. Lyu Q, Lin Z C, She Y Y, Zhang C. A comparison of typical l_p minimization algorithms. Neurocomputing, 2013, 119: 413–424
245. Candes E J, Wakin M B, Boyd S P. Enhancing sparsity by reweighted l_1 minimization. Journal of Fourier Analysis and Applications, 2008, 14(5–6): 877–905
246. Rao B D, Kreutz-Delgado K. An affine scaling methodology for best basis selection. IEEE Transactions on Signal Processing, 1999, 47(1): 187–200
247. She Y Y. Thresholding-based iterative selection procedures for model selection and shrinkage. Electronic Journal of Statistics, 2009, 3: 384–415
248. Xu Z B, Zhang H, Wang Y, Chang X Y, Liang Y. $L_{1/2}$ regularization. Science China Information Sciences, 2010, 53(6): 1159–1169
249. Xu Z B, Guo H L, Wang Y, Zhang H. Representative of $L_{1/2}$ regularization among l_q $(0 < q \leqslant 1)$ regularizations: an experimental study based on phase diagram. Acta Automatica Sinica, 2012, 38(7): 1225–1228
250. Candes E J, Plan Y. Matrix completion with noise. Proceedings of the IEEE, 2009, 98(6): 925–936
251. Cai J F, Candes E J, Shen Z. A singular value thresholding algorithm for matrix completion. SIAM Journal on Optimization, 2010, 20(4): 1956–1982
252. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations & Trends in Machine Learning, 2011, 3(1): 1–122
253. Qian J J, Yang J, Zhang F L, Lin Z C. Robust low-rank regularized regression for face recognition with occlusion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2014, 21–26
254. Liu Y J, Sun D, Toh K C. An implementable proximal point algorithmic framework for nuclear norm minimization. Mathematical Programming, 2012, 133(1–2): 399–436
255. Yang J F, Yuan X M. Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization. Mathematics of Computation, 2013, 82(281): 301–329
256. Li T, Wang W W, Xu L, Feng X C. Image denoising using low-rank dictionary and sparse representation. In: Proceedings of the 10th IEEE International Conference on Computational Intelligence and Security. 2014, 228–232
257. Waters A E, Sankaranarayanan A C, Baraniuk R G. SpaRCS: recovering low-rank and sparse matrices from compressive measurements. In: Proceedings of the Neural Information Processing Systems Conference. 2011, 1089–1097
258. Li Q, Lu Z B, Lu Q B, Li H Q, Li W P. Noise reduction for hyperspectral images based on structural sparse and low-rank matrix decomposition. In: Proceedings of the IEEE International on Geoscience and Remote Sensing Symposium. 2013, 1075–1078

259. Zhou T Y, Tao D C. Godec: randomized low-rank & sparse matrix decomposition in noisy case. In: Proceedings of the 28th International Conference on Machine Learning. 2011, 33–40
260. Zhang H Y, He W, Zhang L P, Shen H F, Yuan Q Q. Hyperspectral image restoration using low-rank matrix recovery. IEEE Transactions on Geoscience & Remote Sensing, 2014, 52(8): 4729–4743
261. Zhang Z, Xu Y, Yang J, Li X L, Zhang D. A survey of sparse representation: algorithms and applications. IEEE Access, 2015, 3: 490–530
262. Burger M, Franek M, SchÖnlieb C B. Regularized regression and density estimation based on optimal transport. Applied Mathematics Research eXpress, 2012, 2012(2): 209–253
263. Osher S, Solè A, Vese L. Image decomposition and restoration using total variation minimization and the H^1 norm. Multiscale Modeling & Simulation, 2003, 1(3): 349–370
264. Barbara K. Iterative regularization methods for nonlinear ill-posed problems. Algebraic Curves & Finite Fields Cryptography & Other Applications, 2008, 6
265. Miettinen K. Nonlinear Multiobjective Optimization. Springer Science & Business Media, 2012
266. Marler R T, Arora J S. Survey of multi-objective optimization methods for engineering. Structural and Multidisciplinary Optimization, 2004, 26(6): 369–395
267. Gong M G, Jiao L C, Yang D D, Ma W P. Research on evolutionary multi-objective optimization algorithms. Journal of Software, 2009, 20(20): 271–289
268. Fonseca C M, Fleming P J. Genetic algorithm for multiobjective optimization: formulation, discussion and generation. In: Proceedings of the International Conference on Genetic Algorithms. 1993, 416–423
269. Srinivas N, Deb K. Multiobjective optimization using non-dominated sorting in genetic algorithms. Evolutionary Computation, 1994, 2(3): 221–248
270. Horn J, Nafpliotis N, Goldberg D E. A niched Pareto genetic algorithm for multiobjective optimization. In: Proceedings of the 1st IEEE Conference on Evolutionary Computation. 1994, 1: 82–87
271. Zitzler E, Thiele L. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. IEEE Transactions on Evolutionary Computation, 1999, 3(4): 257–271
272. Zitzler E, Laumanns M, Thiele L. SPEA2: improving the strength Pareto evolutionary algorithm. Eurogen, 2001, 3242(103): 95–100
273. Kim M, Hiroyasu T, Miki M, Watanabe S. SPEA2+: improving the performance of the strength Pareto evolutionary algorithm 2. In: Proceedings of the International Conference on Parallel Problem Solving from Nature. 2004, 742–751
274. Knowles J D, Corne D W. Approximating the non-dominated front using the Pareto archived evolution strategy. Evolutionary Computation, 2000, 8(2): 149–172
275. Corne D W, Knowles J D, Oates M J. The Pareto-envelope based selection algorithm for multi-objective optimization. In: Proceedings of the Internatioal Conference on Parallel Problem Solving from Nature. 2000, 869–878
276. Corne D W, Jerram N R, Knowles J D, Oates M J. PESA-II: region-based selection in evolutionary multi-objective optimization. In: Proceedings of the Genetic and Evolutionary Computation Conference. 2001, 283–290
277. Deb K, Agrawal S, Pratap A, Meyarivan T. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. Lecture Notes in Computer Science, 2000, 1917: 849–858
278. Zhang Q F, Li H. MOEA/D: a multiobjective evolutionary algorithm based on decomposition. IEEE Transactions on Evolutionary Computation, 2007, 11(6): 712–731
279. Ishibuchi H, Sakane Y, Tsukamoto N, Nojima Y. Simultaneous use of different scalarizing functions in MOEA/D. In: Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation. 2010, 519–526
280. Wang L P, Zhang Q F, Zhou A M, Gong M G, Jiao L C. Constrained subproblems in decomposition based multiobjective evolutionary algorithm. IEEE Transactions on Evolutionary Computation, 2016, 20(3): 475–480
281. Li K, Fialho A, Kwong S, Zhang Q F. Adaptive operator selection with bandits for a multiobjective evolutionary algorithm based on decomposition. IEEE Transactions on Evolutionary Computation, 2014, 18(1): 114–130
282. Ke L J, Zhang Q F, Battiti R. Hybridization of decomposition and local search for multiobjective optimization. IEEE Transactions on Cybernetics, 2014, 44(10): 1808–1820
283. Cai X Y, Wei O. A hybrid of decomposition and domination based evolutionary algorithm for multi-objective software next release problem. In: Proceedings of the 10th IEEE International Conference on Control and Automation. 2013, 412–417
284. Deb K, Jain H. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints. IEEE Transactions on Evolutionary Computation , 2014, 18(4): 577–601
285. Yuan Y, Xu H, Wang B. An improved NSGA-III procedure for evolutionary many-objective optimization. In: Proceedings of ACM Annual Conference on Genetic & Evolutionary Computation. 2014, 661–668
286. Seada H, Deb K. U-NSGA-III: a unified evolutionary optimization procedure for single, multiple, and many objectives: proof-of-principle results. In: Proceedings of the International Conference on Evolutionary Multi-Criterion Optimization. 2015, 34–49
287. Zhu Z X, Xiao J, Li J Q, Zhang Q F. Global path planning of wheeled robots using multi-objective memetic algorithms. Integrated Computer-Aided Engineering, 2015, 22(4): 387–404
288. Zhu Z X, Jia S, He S, Sun Y W, Ji Z, Shen L L. Three-dimensional Gabor feature extraction for hyperspectral imagery classification using a memetic framework. Information Sciences, 2015, 298: 274–287
289. Zhu Z X, Xiao J, He S, Ji Z, Sun Y W. A multi-objective memetic algorithm based on locality-sensitive hashing for one-to-many-to-one dynamic pickup-and-delivery problem. Information Sciences, 2015, 329: 73–89
290. Li H, Gong M G, Wang Q, Liu J, Su L Z. A multiobjective fuzzy clustering method for change detection in synthetic aperture radar images. Applied Soft Computing, 2016, 46: 767–777
291. Jin Y, Sendhoff B. Pareto based approach to machine learning: an overview and case studies. IEEE Transactions on Systems, Man, and Cybernetics, Part C, 2008, 38(3): 397–415
292. Plumbley M D. Recovery of sparse representations by polytope faces pursuit. In: Proceedings of the 6th International Conference on In-

dependent Component Analysis and Blind Source Separation. 2006, 206–213

293. Wright S J, Nowak R D, Figueiredo M A T. Sparse reconstruction by separable approximation. IEEE Transactions on Signal Processing, 2009, 57(7): 2479–2493
294. Yang Y, Yao X, Zhou Z H. On the approximation ability of evolutionary optimization with application to minimum set cover. Artificial Intelligence, 2012, 180(2): 20–33
295. Qian C, Yu Y, Zhou Z H. An analysis on recombination in multi-objective evolutionary optimization. Artificial Intelligence, 2013, 204(1): 99–119
296. Gong M G, Zhang M Y, Yuan Y. Unsupervised band selection based on evolutionary multiobjective optimization for hyperspectral images. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(1): 544–557
297. Qian C, Yu Y, Zhou Z H. Pareto ensemble pruning. In: Proceedings of AAAI Conference on Artificial Intelligence. 2015, 2935–2941
298. Qian C, Yu Y, Zhou Z H. On constrained Boolean Pareto optimization. In: Proceedings of the 24th International Joint Conference on Artificial Intelligence. 2015, 389–395
299. Qian C, Yu Y, Zhou Z H. Subset selection by Pareto optimization. In: Proceedings of the Neural Information Processing Systems Conference. 2015, 1765–1773
300. Gong M G, Liu J, Li H, Cai Q, Su L Z. A multiobjective sparse feature learning model for deep neural networks. IEEE Transactions on Neural Networks and Learning Systems, 2015, 26(12): 3263–3277

Maoguo Gong received the BS degree in electronic engineering and PhD degree in electronic science and technology from Xidian University, China in 2003 and 2009, respectively. He is currently a full professor with Xidian University. His research interests are in the area of computational intelligence with applications to optimization, learning, data mining and image understanding. He is a senior member of IEEE and Chinese Computer Federation. He is the awardee of the NSFC Excellent Young Scholars Program in 2014.

Xiangming Jiang received the BS degree in mathematics and applied mathematics from Xidian University, China in 2015. He is currently pursuing the MS degree in pattern recognition and intelligent systems at the School of Electronic Engineering, Xidian University. His research interests include image understanding and inverse problem.

Hao Li received the BS degree in electronic engineering from Xidian University, China in 2013. He is currently pursuing the PhD degree in pattern recognition and intelligent systems at the School of Electronic Engineering, Xidian University. His research interests include computational intelligence and image understanding.