Filomat 36:8 (2022), 2739–2748
https://doi.org/10.2298/FIL2208739F

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

A Hybrid Regularization Model for Linear Inverse Problems

Ximing Fang[a]

[a] *School of Mathematics and Statistics, Zhaoqing University, Zhaoqing, 526000, China*

Abstract. For the ill-posed linear inverse problem, we propose a hybrid regularization model, which possesses the characters of Tikhonov regularization and TV regularization to some extent. Through transformation, the hybrid regularization is reformulated as an equivalent minimization problem. To solve the minimization problem, we present two modified iterative shrinkage-thresholding algorithms (MISTA) based on the fast iterative shrinkage-thresholding algorithm (FISTA) and the iterative shrinkage-thresholding algorithm (ISTA). The numerical experiments are performed to show the effectiveness and superiority of the regularization model and the presented algorithms.

1. Introduction

In this paper, we consider the numerical method for solving the linear inverse problem

$$\mathbf{b} = H\mathbf{u} + \mathbf{n}, \tag{1}$$

where $H \in R^{n\times n}$ is a nonsingular matrix, $\mathbf{b}$ is the observed result, $\mathbf{u}$ is the true solution to be recovered and $\mathbf{n}$ is a noise vector. The linear inverse problem arises in a wide range of applications, such as astrophysics, signal and image processing, see [2, 10, 15, 23] and the references therein.

It is well known that if the condition number of H is very large, the recovering problem of $\mathbf{u}$ is ill-posed. In order to obtain a satisfactory numerical solution, some regularization approaches are usually required. The basic idea of regularization is to replace the original ill-posed problem by an approximate well-posed problem, and for detailed materials, see [4, 7, 8, 10–12, 15–17, 20–22]. One popular regularization approach is to transform (1) into a least squares problem, that is,

$$\min_{\mathbf{u}} \|H\mathbf{u} - \mathbf{b}\|_2^2 + \lambda\|L\mathbf{u}\|_l^l, \tag{2}$$

where $l = 1, 2$, L is the approximating matrix of the first order derivative operator, $\lambda(> 0)$ is a regularization parameter, which trades off the matching term $\|H\mathbf{u} - \mathbf{b}\|_2^2$ and the regularization term $\|L\mathbf{u}\|_l^l$. For $l = 1$, the model (2) is called the total variation (TV) regularization and is usually applied to solve the discontinuous problems. For $l = 2$, the model (2) is called Tikhonov regularization and is usually applied to solve the

2020 *Mathematics Subject Classification*. 65F22; 65F10
Keywords. The linear inverse problem; Iterative algorithm; Regularization.
Received: 12 January 2019; Accepted: 10 August 2021
Communicated by Yimin Wei
Research supported by Zhaoqing University Research Program: No.611-612279, Guangdong Basic and Applied Basic Research Foundation No. 2022A1515011081, Zhaoqing Science and Technology Innovation Guidance Project No. 2022040315016 and the Innovative Research Team Project of Zhaoqing University.
Email address: fangxm504@163.com (Ximing Fang)

continuous and smooth problems. Both regularization models have some drawbacks, which are mainly determined by the models themselves. For example, Tikhonov regularization is not good at judging the discontinuous points and solving the constant functions, and the numerical solutions obtained by the TV regularization often suffer the staircase effects and loss fine details. For the two regularization methods, readers can refer to [3, 5, 6, 9, 14, 15, 18, 19, 22–25] and the references therein.

In this paper, by introducing two positive parameters α, β and a parameter vector $\mathbf{w}$, we consider the hybrid regularization model for the linear inverse problem (1), that is,

$$\min_{\mathbf{u},\mathbf{w}} \|H\mathbf{u} - \mathbf{b}\|_2^2 + \alpha\|L\mathbf{u} - \mathbf{w}\|_2^2 + \beta\|\mathbf{w}\|_1. \tag{3}$$

In applications, we usually have no information for the true solution(s) in advance, such as the smoothness and the continuity. We expect that the hybrid regularization (3) possesses some good characters of both Tikhonov regularization and the TV regularization by selecting proper parameters. In order to solve the hybrid regularization (3), we first reformulate it as an equivalent minimization model, then present two modified iterative shrinkage-thresholding algorithms (MISTA) based on the fast iterative shrinkage-thresholding algorithm (FISTA) and the iterative shrinkage-thresholding algorithm (ISTA). For the hybrid regularization and the MISTA, we provide some numerical experiments to show the effectiveness and practicality.

The outline of this paper is as follows. In Section 2, we discuss the regularization model (3) and reformulate it as an equivalent minimization problem. In Section 3, we present the MISTA for solving (3). In Section 4, we illustrate some numerical experiments for the hybrid regularization model and the MISTA. Finally, we end this paper by some conclusions in Section 5.

2. Reformulation

In this section, we discuss the transformation problem of the hybrid regularization model (3) for the linear inverse problem (1).

If we denote the objective function in (3) by $F(\mathbf{u}, \mathbf{w})$, then

$$\min_{\mathbf{u},\mathbf{w}} F(\mathbf{u}, \mathbf{w}) \approx \begin{cases} \min_{\mathbf{u}} \|H\mathbf{u} - \mathbf{b}\|_2^2 + \alpha\|L\mathbf{u}\|_2^2, & \text{when } \|\mathbf{w}\|_1 \to 0, \\ \min_{\mathbf{u}} \|H\mathbf{u} - \mathbf{b}\|_2^2 + \beta\|L\mathbf{u}\|_1, & \text{when } \|L\mathbf{u} - \mathbf{w}\|_2 \to 0, \end{cases}$$

which are Tikhonov regularization and the TV regularization, respectively. So the regularization model (3) implies the characters of both regularization models to some extent.

It is well known that both the ISTA and the FISTA can solve the following minimization problem

$$\min \bar{f}(\mathbf{w}) + \bar{g}(\mathbf{w}). \tag{4}$$

Here, $\bar{g}(\mathbf{w})$ is a continuous convex function, and $\bar{f}(\mathbf{w})$ is a smooth convex function of type $C^{1,1}$, i.e., continuously differentiable with Lipschitz constant $C_{\bar{f}} > 0$, that is,

$$\|\nabla\bar{f}(\mathbf{w}_1) + \nabla\bar{f}(\mathbf{w}_2)\| \le C_{\bar{f}}\|\mathbf{w}_1 - \mathbf{w}_2\|. \tag{5}$$

For the difference and the relationship between the ISTA and the FISTA, see [1] and the references therein.

According to the matrix theories, for the hybrid regularization model (3), we have

$$\begin{aligned}
&\min_{\mathbf{u},\mathbf{w}} \|H\mathbf{u}-\mathbf{b}\|_2^2+\alpha\|L\mathbf{u}-\mathbf{w}\|_2^2+\beta\|\mathbf{w}\|_1\\
&=\min_{\mathbf{w}}\Big\{\min_{\mathbf{u}}\{\|H\mathbf{u}-\mathbf{b}\|_2^2+\alpha\|L\mathbf{u}-\mathbf{w}\|_2^2\}+\beta\|\mathbf{w}\|_1\Big\}\\
&=\min_{\mathbf{w}}\Big\{\min_{\mathbf{u}}\Big\{\Big\|\begin{pmatrix} H\\ \sqrt{\alpha}L\end{pmatrix}\mathbf{u}-\begin{pmatrix}\mathbf{b}\\ \sqrt{\alpha}\mathbf{w}\end{pmatrix}\Big\|_2^2\Big\}+\beta\|\mathbf{w}\|_1\Big\}\\
&=\min_{\mathbf{w}}\Big\|\Big(\begin{pmatrix} H\\ \sqrt{\alpha}L\end{pmatrix}(H^{\mathrm{T}}H+\alpha L^{\mathrm{T}}L)^{-1}\begin{pmatrix} H\\ \sqrt{\alpha}L\end{pmatrix}^{\mathrm{T}}-I\Big)\begin{pmatrix}\mathbf{b}\\ \sqrt{\alpha}\mathbf{w}\end{pmatrix}\Big\|_2^2+\beta\|\mathbf{w}\|_1\\
&=\min_{\mathbf{w}}\|\sqrt{\alpha}A_{12}\mathbf{w}+A_{11}\mathbf{b}\|_2^2+\beta\|\mathbf{w}\|_1\\
&\triangleq\min_{\mathbf{w}}\bar{f}(\mathbf{w})+\bar{g}(\mathbf{w})\\
&\triangleq\min_{\mathbf{w}}\bar{F}(\mathbf{w}),
\end{aligned}\tag{6}$$

where $A:=\begin{pmatrix} H\\ \sqrt{\alpha}L\end{pmatrix}(H^{\mathrm{T}}H+\alpha L^{\mathrm{T}}L)^{-1}\begin{pmatrix} H\\ \sqrt{\alpha}L\end{pmatrix}^{\mathrm{T}}-I$, $A_{11}=A(:,1:n)$, $A_{12}=A(:,n+1:2n)$, and

$$L=\begin{pmatrix} 1 & 0 & \cdots & 0\\ -1 & 1 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & 1\\ 0 & 0 & \cdots & -1\end{pmatrix}_{(n+1)\times n},\tag{7}$$

where L is the approximation matrix of the first order derivative operator with the zero boundary conditions. Here, we introduce $\bar{f}(\mathbf{w})$ and $\bar{g}(\mathbf{w})$ for convenience, and the hybrid regularization model (3) is reformulated as (4).

For model (6), if we denote

$$\bar{Q}_C(\mathbf{w},\mathbf{v}):=\bar{f}(\mathbf{v})+\langle\mathbf{w}-\mathbf{v},\nabla\bar{f}(\mathbf{v})\rangle+\frac{C}{2}\|\mathbf{w}-\mathbf{v}\|^2+\bar{g}(\mathbf{w})\tag{8}$$

and

$$\begin{aligned}
\bar{\mathcal{P}}_C(\mathbf{v}):&=\operatorname{argmin}\{\bar{Q}_C(\mathbf{w},\mathbf{v}):\mathbf{w}\in R^n\}\\
&=\operatorname{argmin}\{\frac{C}{2}\|\mathbf{w}-(\mathbf{v}-\frac{1}{C}\nabla\bar{f}(\mathbf{v}))\|^2+\bar{g}(\mathbf{w}):\mathbf{w}\in R^n\}\\
&=\operatorname{argmin}\{\frac{C}{2}\|\mathbf{w}-(\mathbf{v}-\frac{2}{C}\sqrt{\alpha}A_{12}^{\mathrm{T}}(\sqrt{\alpha}A_{12}\mathbf{v}+A_{11}\mathbf{b}))\|^2+\beta\|\mathbf{w}\|_1:\mathbf{w}\in R^n\},
\end{aligned}\tag{9}$$

where $C>0$ is a positive number, and $\mathbf{v}$ is a given vector, then both the ISTA and the FISTA can be applied to solve (6) for $\mathbf{w}$ based on (8) and (9) (see [1]). We show the basic step of the ISTA below.

$$\begin{aligned}
\mathbf{w}_k&=\bar{\mathcal{P}}_C(\mathbf{w}_{k-1})\\
&=\bar{\mathcal{T}}_{\beta/C}(\mathbf{w}_{k-1}-\frac{2}{C}\sqrt{\alpha}A_{12}^{\mathrm{T}}(\sqrt{\alpha}A_{12}\mathbf{w}_{k-1}+A_{11}\mathbf{b}))\\
&=\max\{|\mathbf{w}_{k-1}-\frac{2}{C}\sqrt{\alpha}A_{12}^{\mathrm{T}}(\sqrt{\alpha}A_{12}\mathbf{w}_{k-1}+A_{11}\mathbf{b})|-\frac{\beta}{C},0\}\cdot\operatorname{sign}(\mathbf{w}_{k-1}-\frac{2}{C}\sqrt{\alpha}A_{12}^{\mathrm{T}}(\sqrt{\alpha}A_{12}\mathbf{w}_{k-1}+A_{11}\mathbf{b})),
\end{aligned}\tag{10}$$

where $\bar{\mathcal{T}}_\gamma:R^n\to R^n$ is the shrinkage operator defined by

$$\bar{\mathcal{T}}_\gamma(x)_i=(|x_i|-\gamma)_+\operatorname{sign}(x_i)$$

Algorithm 1 MISTA for (6)

1 Take $C_0 > 0$, $\eta > 1$ and $\mathbf{w}_0 \in R^n$. Set $\mathbf{v}_1 = \mathbf{w}_0, t_1 = 1$ and $k = 1$.
2 Repeat
3 Find the smallest nonnegative integer i_k such that $\bar{C} = \eta^{i_k} C_0$ satisfies

$$\bar{F}(\bar{\mathcal{P}}_{\bar{C}}(\mathbf{v}_k)) \le \bar{Q}_{\bar{C}}(\bar{\mathcal{P}}_{\bar{C}}(\mathbf{v}_k), \mathbf{v}_k).$$

4 Set $C_k = \eta^{i_k} C_0$ and compute

$$\mathbf{w}_k = \bar{\mathcal{P}}_{C_k}(\mathbf{v}_k),$$

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2},$$

$$\mathbf{v}_{k+1} = \mathbf{w}_k + \frac{t_k - 1}{t_{k+1}}(\mathbf{w}_k - \mathbf{w}_{k-1}).$$

5 If $\bar{F}(\mathbf{v}_{k+1}) > \bar{F}(\mathbf{w}_k)$, set $\mathbf{v}_{k+1} = \mathbf{w}_k$.
6 Set $C_0 = \max\{C_0, M(C_k)\}$ with $M(C_k) \le C_k$.
7 Until the stopping criterion is satisfied.

Algorithm 2 MISTA for (6)

1 Take $C_0 > 0$, $\eta > 1$ and $\mathbf{w}_0 \in R^n$. Set $k = 1$.
2 Repeat
3 Find the smallest nonnegative integer i_k such that with $\bar{C} = \eta^{i_k} C_{k-1}$

$$\bar{F}(\bar{\mathcal{P}}_{\bar{C}}(\mathbf{w}_{k-1})) \le \bar{Q}_{\bar{C}}(\bar{\mathcal{P}}_{\bar{C}}(\mathbf{w}_{k-1}), \mathbf{w}_{k-1}).$$

4 Set $C_k = \eta^{i_k} C_{k-1}$ and compute

$$\mathbf{w}_k = \bar{\mathcal{P}}_{C_k}(\mathbf{w}_{k-1}).$$

5 Set $C_0 = \max\{C_0, M(C_k)\}$ with $M(C_k) \le C_k$.
6 Until the stopping criterion is satisfied.

and the positive constant C will be changed in the iteration processes.

Once we obtain the solution $\mathbf{w}$ in (6), then we can obtain $\mathbf{u}$ in (3) by

$$\begin{aligned} \mathbf{u} &= (H^{\mathrm{T}}H + \alpha L^{\mathrm{T}}L)^{-1} \begin{pmatrix} H \\ \sqrt{\alpha}L \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \mathbf{b} \\ \sqrt{\alpha}\mathbf{w} \end{pmatrix} \\ &= (H^{\mathrm{T}}H + \alpha L^{\mathrm{T}}L)^{-1}(H^{\mathrm{T}}\mathbf{b} + \alpha L^{\mathrm{T}}\mathbf{w}). \end{aligned} \tag{11}$$

We can find that $\mathbf{u}$ is the function of $\mathbf{w}$, α and β.

3. MISTA

In this section, we mainly discuss the solving method of (6) for $\mathbf{w}$. Based on (8), (9), (10) and the FISTA and the ISTA, we have two modified iterative shrinkage-thresholding algorithms (MISTA), that is Algorithm 1 and Algorithm 2 as follows.

In Step 6 of Algorithm 1, $M(C_k)$ is a function of C_k, which has many forms, such as, $M(C_k) = C$ with $C(\ge C_0)$ being a positive constant, $M(C_k) = \tau C_k$ with $\tau(< 1)$ being a positive constant and $M(C_k) = \sqrt{C_k}$ if $C_k \ge 1$. We set $C_0 = \max\{C_0, M(C_k)\}$ in Step 5 so that the new constant C_0 will not less than the original C_0.

The main difference between Algorithm 1 and the FISTA with backtracking is that Algorithm 1 has Step 5 and Step 6. The FISTA with backtracking does not have Step 6, and the process of searching for C_k is

based on C_{k-1}. However, for Algorithm 1, C_k is based on the new C_0, which is less than C_{k-1}. So, the two algorithms have different changing ways of C_k. For the FISTA with backtracking, C_k will increase till a larger value $C_k \geq C_{\bar{f}}$ appears, then the value C_k will be fixed in the latter iteration processes (see Lemma 2.1 in [1]). So, the iteration process of the FISTA can be regarded as the searching process of $C_{\bar{f}}$. However, for Algorithm 1, the C_k does not always increase but can restart from a lower value due to Step 6. In a word, if we denote the parameter C_ks generated by the FISTA and Algorithm 1 by C_k(F) and C_k(M), respectively, then the inequality $C_k(\mathrm{M}) \leq C_k(\mathrm{F})$ holds. So from (8), the inequality

$$\min_{\mathbf{w}} \bar{Q}_{C_k(\mathrm{M})}(\mathbf{w}, \mathbf{v}_k) \leq \min_{\mathbf{w}} \bar{Q}_{C_k(\mathrm{F})}(\mathbf{w}, \mathbf{v}_k) \tag{12}$$

holds for a given vector $\mathbf{v}_k$. In other words, when we search for the $\mathbf{w}_k = \bar{\mathcal{P}}_{C_k}(\mathbf{v}_k)$ in Step 3, Algorithm 1 owns a smaller upper boundary than the FISTA with backtracking. In this sense, we say that Algorithm 1 is better than the FISTA with backtracking. Meanwhile, from Step 3, we know that since the C_0 in Algorithm 1 is lower than C_{k-1}(F) in the FISTA with backtracking, the searching process of C_k(M) in Algorithm 1 may take more time.

For the convergence of Algorithm 1, from (8), (9), Step 3 and Step 5, we have

$$\begin{aligned}
\bar{F}(\mathbf{w}^*) \leq \bar{F}(\mathbf{w}_k) &= \bar{F}(\bar{\mathcal{P}}_{\bar{C}}(\mathbf{v}_k)) \\
&\leq \bar{Q}_{\bar{C}}(\bar{\mathcal{P}}_{\bar{C}}(\mathbf{v}_k), \mathbf{v}_k) \\
&= \bar{f}(\mathbf{v}_k) + \langle \bar{\mathcal{P}}_{\bar{C}}(\mathbf{v}_k) - \mathbf{v}_k, \nabla \bar{f}(\mathbf{v}_k) \rangle + \frac{\bar{C}}{2} \|\bar{\mathcal{P}}_{\bar{C}}(\mathbf{v}_k) - \mathbf{v}_k\|^2 + \bar{g}(\bar{\mathcal{P}}_{\bar{C}}(\mathbf{v}_k)) \\
&\leq \bar{f}(\mathbf{v}_k) + \bar{g}(\mathbf{v}_k) \\
&= \bar{F}(\mathbf{v}_k) \leq \bar{F}(\mathbf{w}_{k-1}).
\end{aligned}$$

Thus, $\{\bar{F}(\mathbf{w}_k)\}_{k=1}^{+\infty}$ is a decreasing sequence. Similarly, we know that $\{\bar{F}(\mathbf{w}_k)\}_{k=1}^{+\infty}$ generated by Algorithm 2 is also a decreasing sequence.

4. Numerical examples

In this section, we present several experiments to show the performance of the hybrid regularization model (3) and the MISTA, that is Algorithm 1.

For one dimensional problem, the coefficient matrix H in (6) is obtained by discretizing the function

$$H(x, t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2\sigma^2}} \quad \text{with} \quad \sigma = 0.05$$

at the grid point $(x_i, t_j) = (\frac{i}{n}, \frac{j}{n})$ for $i, j = 1, 2, \ldots, n$. For two dimensional problem, the coefficient matrix H in (6) is obtained by discretizing the function

$$H(x, y, s, t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-s)^2+(y-t)^2}{2\sigma^2}} \quad \text{with} \quad \sigma = 0.05,$$

where $(x, y, s, t) \in [0, 1]^4$. The true solution is generated by discretizing

$$U_{\mathrm{tr}}(x, y) = \begin{cases} 1, & x \in [0.1, 0.4], y \in [0.1, 0.4], \\ \sin(4\pi(x - 0.5)(y - 0.5)), & (x - 0.75)^2 + (y - 0.25)^2 \leq 0.15^2, \\ 1.5x + 0.3, & x \in [0.1, 0.4], y \in [0.6, 0.9], \\ u_4(x, y), & x \in [0.6, 0, 9], y \in [0.6, 0.9], \\ 0, & \text{elsewhere} \end{cases} \tag{13}$$

at the grid point $(x_i, y_j) = (\frac{i}{n}, \frac{j}{n})$ for $i, j = 1, 2, \ldots, n$ and reordering the resulted matrix column by column to get the $n^2 \times 1$ solution vector $\mathbf{U}_{\text{tr}}$, where

$$u_4(x, y) = 2000(x - 0.6)(0.9 - x)(y - 0.6)(0.9 - y).$$

In Example 1 and Example 3, The observation data $\mathbf{b}$ is generated by

$$\mathbf{b} = \begin{cases} H\mathbf{u}_{tr} + \delta(\text{rand}(n, 1) - 0.5), & \text{for one dementional problem,} \\ H\mathbf{U}_{tr} + \delta(\text{rand}(n^2, 1) - 0.5), & \text{for two dementional problem,} \end{cases}$$

where δ is the noise level and "rand" is a function in Matlab.

The elapsed time is denoted by CPU (unit: second) and the number of iteration steps is denoted by IT. The absolute error and relative error of the recovered solutions are denoted by Er_{ab} and Er_{re}, respectively. Here, we consider the 2-norm $\|.\|_2$ in Er_{ab} and Er_{re}. It is well known that for a regularization model, the proper regularization parameters is very important, and for the parameter selecting methods, readers can refer to [13, 23] and the references therein. We select the best parameter values α and β in our experiments by comparing many numerical results.

Example 1 (The comparison of different regularization models)

In this example, we compare the hybrid regularization model (6) with (2). We use the $\text{MISTA}_{\text{sqrt}}$ to solve the hybrid model (6) with (11), that is Algorithms 1 with $M(C_k) = \sqrt{C_k}$ in Step 6 with $\eta = 1.5$ and $C_0 = 3$. For Tikhonov regularization model and the TV regularization model, we apply the directly method and the lagged diffusivity fixed point algorithm (LDFPA) to solve them, respectively. For the two dimensional problems, we consider the matrix L in (6) with the zero boundary conditions, that is,

$$L = \begin{pmatrix} D_x \\ D_y \end{pmatrix} = \begin{pmatrix} L \otimes I_n \\ I_n \otimes L \end{pmatrix},$$

where the matrix L in the bracket is given by (7), the symbol $\otimes$ denotes the Kronecker tensor product of two matrices, I_n denotes the identity matrix of order n.

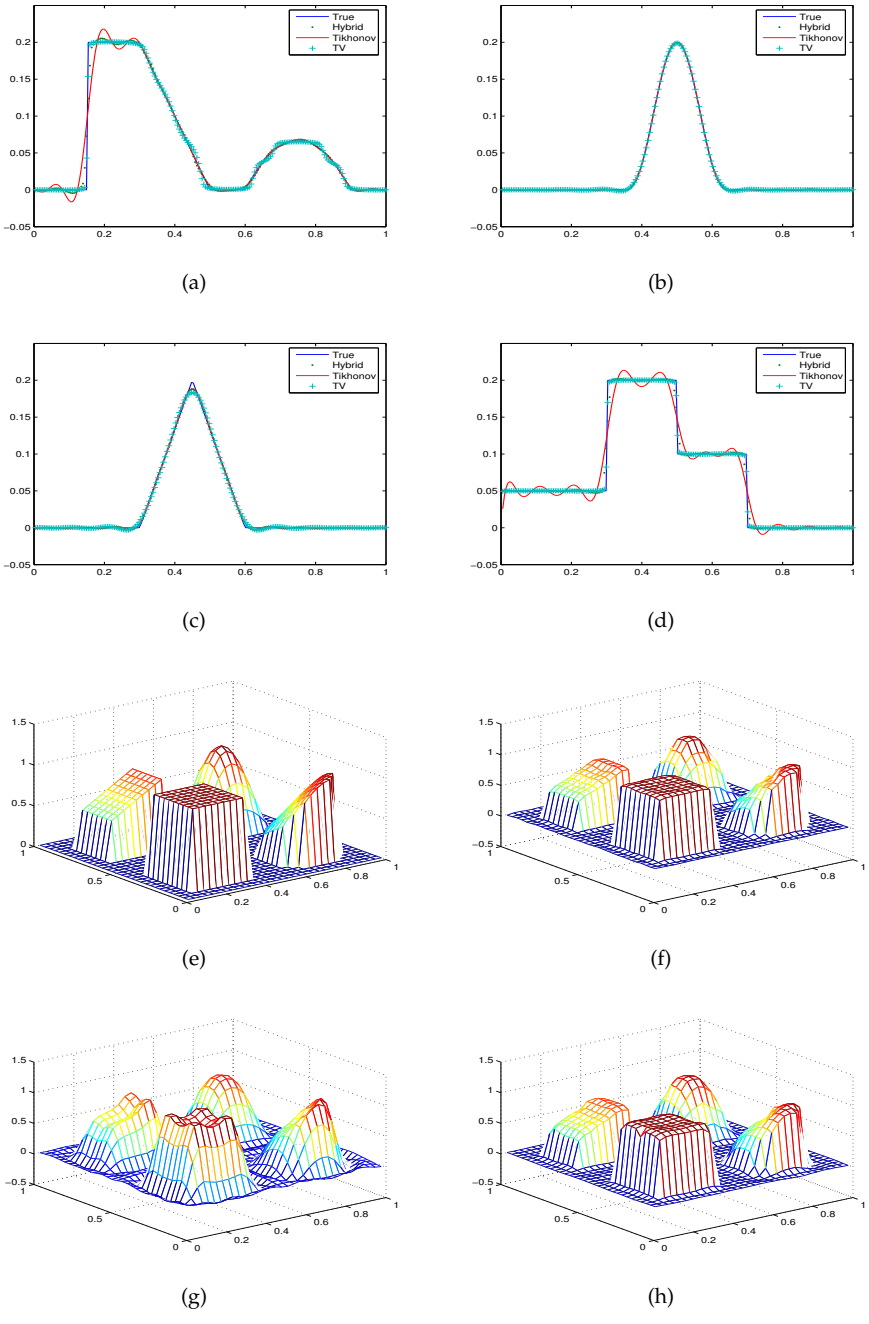
The regularization term $\alpha\|L\mathbf{u}\|_1$ in (2) is approximated by a smoothing function with a parameter ϵ, and the concrete forms are

$$\alpha\|Lu\|_1 \approx \begin{cases} \Sigma_{i=1}^{n+1} \sqrt{(L_i\mathbf{u})^2 + \epsilon}, \\ \Sigma_{i=1}^{n^2+n} \sqrt{(D_{xi}\mathbf{u})^2 + \epsilon} + \Sigma_{i=1}^{n^2+n} \sqrt{(D_{y_i}\mathbf{u})^2 + \epsilon}, \end{cases}$$

for one dimensional and two dimensional problems, respectively, where the subscript i denotes the i-th row of a matrix. The noise levels are set to $\delta = 1e-2$ and $\delta = 1e-1$, respectively. The initial iteration vectors are set to be $\mathbf{w}_0 = 0.1\text{ones}(n, 1)$ with $n = 300$ and $\mathbf{w}_0 = 0.1\text{ones}(2n^2 + 2n, 1)$ with $n = 32$, respectively. The number of iteration steps is 100. We obtain Figure 1, Table 1 and Table 2 as follows.

Table 1: The parameters for one dimensional problem

	one dimension											
	Ex_1			Ex_2			Ex_3			Ex_4		
	α	β	ϵ	α	β	ϵ	α	β	ϵ	α	β	ϵ
Hybrid	3	1e-2		5	1		2	0.06		2	2e-3	
Tikhonov	3			5			2			2		
TV		3	1e-3		5	1e+1		2	1e+1		2	1e-3



(a) (b)

(c) (d)

(e) (f)

(g) (h)

Figure 1: The comparison of Tikhonov, TV and hybrid regularization.

Table 2: The parameters, Errors and CPU for two dimensional problem

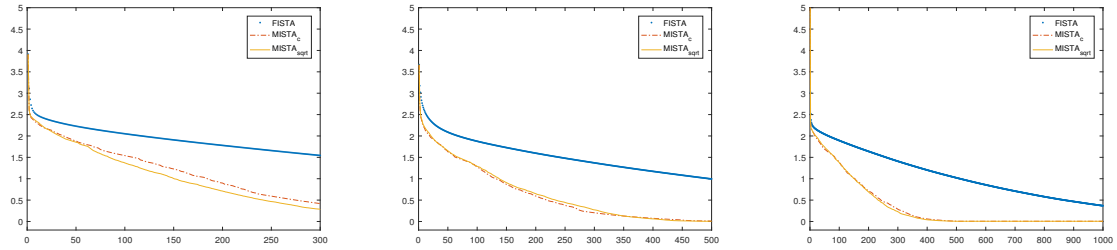
	two dimension				
	α	β	ϵ	Er_{ab}	Er_{re}
hybrid	1	8e-2		0.3344	0.0262
Tikhonov	6			2.1358	0.1674
TV		1	1e-3	0.5082	0.0398

The first four sub-figures in Figure 1 are for one dimensional problem. The four lines in each sub-figure represent the true solution and the numerical solutions generated by different regularization models. The other four sub-figures are for two dimensional problem, they are the true solution (e), the hybrid regularization solution (f), Tikhonov regularization solution (g) and the TV regularization solution (h) in turn. The parameter ϵ in the TV regularization model is selected the best in many numerical experiments. From Figure 1, we can find that the hybrid regularization model (6) possesses the characters of Tikhonov regularization and the TV regularization to some extent.

Example 2 (**The comparison between the MISTA and the FISTA**)
In this example, we compare the numerical results of the MISTA and the FISTA for solving (3), that is (6) for $\bar{F}(\mathbf{w}_k)$. Since the regularization parameters α, β are not required, we set $\alpha = 5$, $\beta = 0.01$, $n = 300$ for convenience, then the Lipschitz constant of $\bar{f}(\mathbf{w})$ is that $C_{\bar{f}} = 10$. The true solution $\mathbf{u}_{\text{tr}}$ is obtained by discretizing the function

$$u_{\text{tr}}(t) = \begin{cases} 0.2, & t \in (0.15, 0.3], \\ 0.5 - t, & t \in (0.3, 0.5], \\ 3(t - 0.6)(0.9 - t), & t \in [0.6, 0.9], \\ 0, & \text{otherwise} \end{cases}$$

at the grid points $t_i = \frac{i}{n}$, $i = 1, 2, \ldots, n$. Set $\mathbf{b} = \text{ones}(n, 1)$ in (3), $n = 300$, $\eta = 1.2$, $C_0 = 2$ and $\mathbf{w}_0 = \text{randn}(n + 1, 1)$ in Algorithm 1, then we obtain Figure 2 and Table 3 as follows.



(a) (b) (c)

Figure 2: The function $\bar{F}(\mathbf{w}_k)$

.

MISTA_c and $\text{MISTA}_{\text{sqrt}}$ denote the two cases of Algorithm 1 with $C_0 = C = 2$ and $C_0 = \max\{C_0, \sqrt{C_k}\}$ in step 6, respectively. In Figure 2, the horizontal axis denotes IT and the vertical axis denotes $\bar{F}(\mathbf{w}_k)$. Figure 2 shows that both the MISTA_c and the $\text{MISTA}_{\text{sqrt}}$ have the smaller value $\bar{F}(\mathbf{w}_k)$ compared with the FISTA when the iteration steps are same. Meanwhile, we can see that the iteration sequence $\{\bar{F}(\mathbf{w}_k)\}_{k=1}^{+\infty}$ is decreasing, and although the MISTA_c has a larger range of C_0 compared with the $\text{MISTA}_{\text{sqrt}}$ since $C_0 \leq \max\{C_0, \sqrt{C_k}\}$ in step 6, the values $\bar{F}(\mathbf{w}_k)$ generated by the MISTA_c are similar to that generated by the $\text{MISTA}_{\text{sqrt}}$. Table 3

Table 3: The comparison between the MISTA and the FISTA

	IT=300		IT=500		IT=1000	
	CPU	$\min \bar{F}(\mathbf{w}_k)$	CPU	$\min \bar{F}(\mathbf{w}_k)$	CPU	$\min \bar{F}(\mathbf{w}_k)$
FISTA	1.188660	1.5420	1.875522	0.9928	3.657459	0.3670
MISTA_{c}	6.408545	0.4276	9.545819	0.0099	14.272323	0.0052
$\text{MISTA}_{\text{sqrt}}$	5.850016	0.3185	9.065983	0.0052	13.771557	0.0052

shows the minimum values of $\bar{F}(\mathbf{w}_k)$ and the elapsed time CPU, which correspond to Figure 2. From Table 3, we can see that the MISTA has more efficient that the FISTA although the latter has a shorter single step running time. Besides, We point out that since the initial iteration vector $\mathbf{w}_0$ is arbitrary in our experiments, the numerical results are different each time, however, the basic situation is similar.

Example 3 (**The numerical results of the MISTA and the FISTA**)
In this example, we compare the recovered results of the MISTA and the FISTA for solving $\mathbf{u}$ in (6) under the same iteration steps, that is, IT = 100. We consider one dimensional problem and set $\delta = 1e-1$, $\mathbf{w}_0 = 0.1\text{ones}(n+1,1)$ with $n = 201$. The numerical results are shown in Figure 3 as follows.

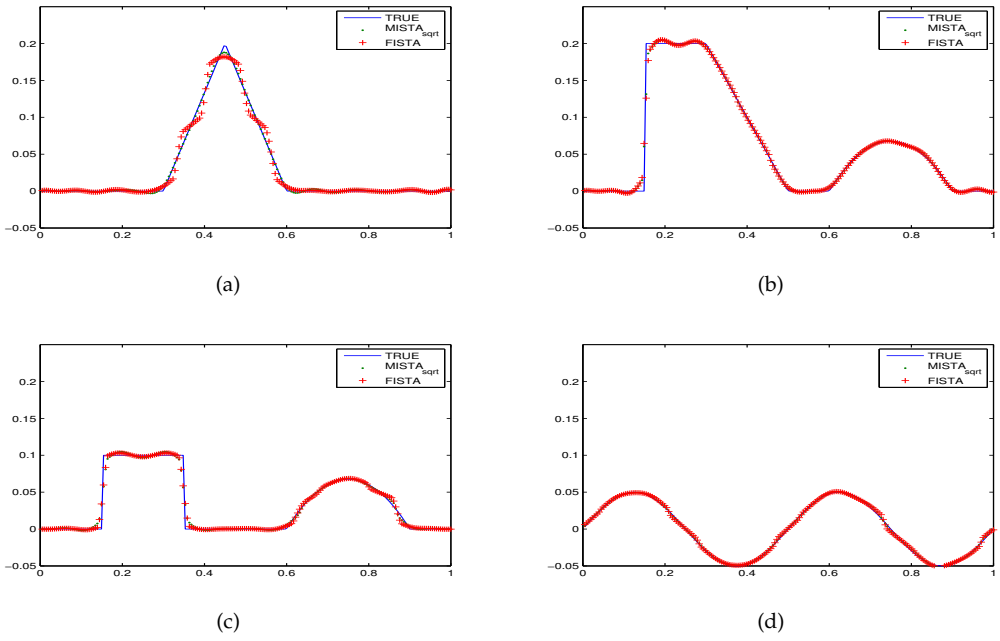


Figure 3: The numerical results of the MISTA and the FISTA.

From Figure 3, We can find that the numerical solution obtained by the MISTA is better than that obtained by the FISTA sometimes. The parameter α, β, η, C_0 in the four figures are as follows: 20, 0.1, 1.2, 2; 10, 0.1, 1.2, 12; 40, 0.1, 1.2, 40; 20, 0.1, 1.2, 10, respectively.

5. Conclusions

In this paper, a hybrid regularization model is proposed for the ill-posed linear inverse problem. Then we reformulate the regularization model as a concrete minimum problem and present the MISTA to solve it. Compared with Tikhonov regularization and the TV regularization, the hybrid regularization model

shows its superiority. In addition, the numerical experiments illustrate the high-efficiency of the MISTA compared with the FISTA.

References

[1] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM Journal on Imaging Sciences 2 (2009) 183-202.
[2] J. M. Bioucas-Dias, M. A. Figueiredo, A new TWIST: two-step iterative shrinkage thresholding algorithms for image restoration, IEEE Transactions on Image Processing 16 (2007) 2992-3004.
[3] A. Beck, M. Teboulle, Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems, IEEE Transactions on Image Processing 18 (2009) 2419-2434.
[4] P. L. Combettes, V. R. Wajs, Signal recovery by proximal forward-backward splitting, Multiscale Modeling and Simulation 4 (2005) 1168-1200.
[5] A. Chambolle, An algorithm for total variation minimization and application: special issue on mathematics and image analysis, Journal of Mathematical Imaging and Vision 20 (2004) 89-97.
[6] S. H. Chan, R. Khoshabeh, K. B. Gibson, P. E. Gill and T. Q. Nguyen, An augmented Lagrangian method for total variation video restoration, IEEE Transactions on Image Processing 20 (2011) 3097-3111.
[7] I. Daubechies, M. Defrise and C. D. Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Communications on pure and applied mathematics 57 (2004) 1413-1457.
[8] M. Elad, B. Matalon and M. Zibulevsky, Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization, Applied and Computational Harmonic Analysis 23 (2007) 346-367.
[9] A. E. Hamidi, M. Ménard, M. Lugiez and C. Ghannam, Weighted and extended total variation for image restoration and decomposition, Pattern Recognition 43 (2010) 1564-1576.
[10] M. A. Figueiredo, R. D. Nowak, An EM algorithm for wavelet-based image restoration, IEEE Transactions on Image Processing 12 (2003) 906-916.
[11] M. A. Figueiredo, R. D. Nowak and S. J. Wright, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems, IEEE Journal of Selected Topics in Signal Processing 1(2007) 586-597.
[12] W. W. Hager, D. T. Phan and H. C. Zhang, Gradient-based methods for sparse recovery, SIAM Journal on Imaging Sciences 4 (2011) 146-165.
[13] P. C. Hansen, Discrete Inverse Problems: Insight and Algorithms, SIAM, Philadelphia, 2010.
[14] Y. Liu, J. H. Ma, Y. Fan and Z. R. Liang, Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction, Inverse Problems 57 (2012) 7923-7956.
[15] F. R. Lin and S. W. Yang, A weighted H_1 seminorm regularization method for Fredholm integral equations of the first kind, International Journal of Computer Mathematics 91 (2014) 1012-1029.
[16] V. A. Morozov, Methods for Solving Incorrectly Posed Problems, Springer-Verlag, New York, 1984.
[17] K. Maleknejad, S. Sohrabi, Numerical solution of Fredholm integral equations of the first kind by using Legendre wavelets, Applied Mathematics and Computation 186 (2007) 836-843.
[18] J. P. Oliveira, J. M. Bioucas-Dias and M. A. Figueiredo, Adaptive total variation image deblurring: a majorization-minimization approach, Signal Processing 89 (2009) 1683-1693.
[19] K. Papafitsoros, C. B. Schönlieb, A combined first and second order variational approach for image reconstruction, Journal of Mathematical Imaging and Vision 48 (2014) 308-338.
[20] Z. W. Qin, D. Goldfarb, Structured sparsity via alternating direction methods, Journal of Machine Learning Research 13 (2012) 1435-1468.
[21] L.I. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D: Nonlinear Phenomena 60 (1992) 259-268.
[22] A. N. Tikhonov, Regularization of incorrectly posed problems, Soviet Mathematics Doklady 4 (1964) 1624-1627.
[23] C. R. Vogel, Computational Methods for Inverse Problems, SIAM, Philadelphia, 2002.
[24] Y. L. Wang, J. F. Yang, W. T. Yin and Y. Zhang, A new alternating minimization algorithm for total variation image reconstruction, SIAM Journal on Imaging Sciences 1 (2008) 248-272.
[25] J. F. Yang, Y. Zhang and W. T. Yin, An efficient TVL_1 algorithm for deblurring multichannel images corrupted by impulsive noise, SIAM Journal on Scientific Computing 31 (2009) 2842-2865.