

BACKWARD PROBLEM FOR A TIME-SPACE FRACTIONAL DIFFUSION EQUATION

JUNXIONG JIA

School of Mathematics and Statistics
Xi'an Jiaotong University
Xi'an, 710049, China

JIGEN PENG*

Department of Mathematics and Information Science
Guangzhou University
Guang Zhou, 510006, China

JINGHUAI GAO

School of Electronic and Information Engineering
Xi'an Jiaotong University
Xi'an, 710049, China

YUJIAO LI

Department of Bioengineering
Xi'an Jiaotong University
Xi'an, 710049, China

(Communicated by the associate editor name)

ABSTRACT. In this paper, a backward problem for a time-space fractional diffusion process has been considered. For this problem, we propose to construct the initial function by minimizing data residual error in Fourier space domain with variable total variation (TV) regularizing term which can protect the edges as TV regularizing term and reduce staircasing effect. The well-posedness of this optimization problem is obtained under a very general setting. Actually, we rewrite the time-space fractional diffusion equation as an abstract fractional differential equation and deduce our results by using fractional operator semigroup theory, hence, our theoretical results can be applied to other backward problems for the differential equations with more general fractional operator. Then a modified Bregman iterative algorithm has been proposed to approximate the minimizer. The new features of this algorithm is that the regularizing term altered in each step and we need not to solve the complex Euler-Lagrange equation of variable TV regularizing term (just need to solve a simple Euler-Lagrange equation). The convergence of this algorithm and the strategy of choosing parameters are also obtained. Numerical implementations are provided to support our theoretical analysis to show the flexibility of our minimization model.

2010 *Mathematics Subject Classification.* Primary: 35R30, 35R11; Secondary: 65L09.

Key words and phrases. Backward time-space fractional diffusion, Fractional operator semigroup, Bregmann iterative method, Variable TV regularization.

* Corresponding author: Jigen Peng.

1. Introduction. Diffusion phenomenon is ubiquitous in our physical world. From the point of view of probability theory, we can derive diffusion equations by applying central limit theorem to the random walk problem. If we assume the distribution of particle jump is Gaussian, we will obtain normal diffusion equations

$$\begin{cases} \partial_t v(t, x) - \Delta v(t, x) = 0 \\ v(x, 0) = u(x). \end{cases} \quad (1)$$

If we assume the particle jump satisfy Lévy distribution, by continuous time random walk (CTRW) model, we will derive time-space fractional diffusion equation (FDE) as follows

$$\begin{cases} \partial_t^\alpha v(t, x) + (-\Delta)^\beta v(t, x) = 0 & (x, t) \in \mathbb{R}^2 \times (0, \infty) \\ v(x, 0) = u(x) & x \in \mathbb{R}^2 \end{cases} \quad (2)$$

with $\alpha \in (0, 1]$, $\beta \in (1/2, 1]$. Here the time derivative is in Djrbashian-Caputo sense defined as follows

$$\partial_t^\alpha v(t) = (g_{1-\alpha} * \partial_t v)(t) := \int_0^t g_{1-\alpha}(t-s)v(s)ds, \quad (3)$$

where

$$g_\gamma(t) := \begin{cases} \frac{1}{\Gamma(\gamma)} t^{\gamma-1}, & t > 0, \\ 0, & t \leq 0, \end{cases} \quad (4)$$

with $\Gamma(\gamma)$ is the Gamma function. Denote the Fourier transform of function v as $\mathcal{F}(v)$ or \hat{v} , the inverse fourier transform as $\mathcal{F}^{-1}v$ or \check{v} . Then the space fractional derivative $(-\Delta)^\beta$ can be defined by Fourier transform as $(-\Delta)^\beta v = \mathcal{F}^{-1}(|\xi|^{2\beta}\hat{v})$. Usually, we call this type fractional derivative operator as symmetric Riesz-Feller space fractional derivative operator.

Fractional time-space diffusion equation (2) attracts lots of researchers attention. From the physical point of view, there are two long papers [36, 15] provide a good summary. From the stochastic point of view, there is a good book [28] which gives rigorous mathematical deductions. From the functional analysis point of view, Peng, Li [21] propose fractional semigroup theory; Li, Chen [27] propose α -resolvent operator to provide a general theory for the fractional abstract Cauchy problem which can be applied to FDE (2) and some more general FDEs. B. Baeumer et al. [4, 3, 1, 2] propose the concept of stochastic solutions for fractional evolution equations and study FDEs by using stochastic methods combined with operator semigroup theory.

In this paper, we focus on the backward problem for equation (2). As mentioned in a recent tutorial [6], the mathematical theory of inverse problems for FDEs is still in its infancy. However, there are already some pioneering work in this direction. Cheng et al. [18] establish the uniqueness in an inverse problem for a one-dimensional fractional diffusion equation. Sakamoto and Yamamoto [23] establish the unique existence of weak solutions and the asymptotic behavior as time t goes to ∞ . They also prove the stability in the backward problem in time and the uniqueness in determining an initial value. Liu and Yamamoto [20] study a backward problem for a time-fractional diffusion equation. Zhang and Xu [41] investigate an inverse source problem for fractional diffusion equation. They obtain the uniqueness of the inverse problem by analytic continuation and Laplace transform. Zheng

and Wei [14] study backward problem of space fractional diffusion equations. They show that the problem is severely ill-posed and propose a regularization method.

Recently, Wang and Liu [25] propose to use more general anomalous diffusion models to describe the blurring effect, and the backward problems give the mathematical formulation for the de-blurring process in image restoration. By using total variation regularization term, the discontinuity of the initial data can be recovered. In Wang and Liu's paper, they use time-fractional diffusion models, here we intend to use a more general time-space fractional diffusion model (2). Recovering initial data with discontinuities not only can find applications for de-blurring process but also appeared for recovering initial temperature when we know temperature field diffuse according to some fractional equations. Apart from the usual smooth assumptions, the initial temperature may also exhibit discontinuities as shown in Figure 10.4 in [35].

In the following, we assume the initial data $u(x) \in \mathbb{R}^2$ with compact support in a bounded convex open subset Ω of the plane with Lipschitz continuous boundary $\partial\Omega$, which is a reasonable assumption in many applications. Denote $g^\delta(x)$ in Ω to be the measurement data, our backward problem is to approximate $v(0, x) = u(x)$ from $g^\delta(x)$. For some known error level $\delta > 0$, the noisy data of the exact gray level $g(x) := v(T, x)$ satisfying

$$\|g^\delta(\cdot) - g(\cdot)\|_{L^2(\Omega)} \leq \delta. \quad (5)$$

In many applications of the backward diffusion problem, the initial distribution $u(x)$ is in general not smooth. Because $v(T, x)$ generated from the Cauchy problem (2), $v(t, x)$ need not have compact support as $u(x)$. Here, we only use the measurement data $v(T, x)$ in Ω , i.e., the values $v(T, x)$ outside of Ω have nothing to do with our reconstruction process, we can define $v(T, x)$ for $x \in \Omega$ such that

$$\|g^\delta(\cdot) - g(\cdot)\|_{L^2(\mathbb{R}^2 \setminus \Omega)} = 0. \quad (6)$$

Taking Fourier transform with respect to x in (2), we obtain

$$\partial_t^\alpha \hat{v}(t, \xi) = -|\xi|^{2\beta} \hat{v}(t, \xi). \quad (7)$$

By using the Laplace transform with respect to t in (7), we can establish the relation between $u(x)$ and $v(T, x)$ in frequency domain as

$$\hat{g}(\xi) = \hat{v}(T, \xi) = \hat{S}(\xi) \hat{u}(\xi), \quad \hat{S}(\xi) := E_{\alpha, 1}(-|\xi|^{2\beta} T^\alpha), \quad (8)$$

where $E_{\alpha, 1}(\cdot)$ is the Mittag-Leffler function defined as

$$E_{\alpha, \gamma}(z) := \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\alpha k + \gamma)}, \quad z \in \mathbb{C}, \quad \alpha > 0, \quad \gamma > 0, \quad (9)$$

which can be seen as a generalization of exponential function e^z . Denote $Su = \mathcal{F}^{-1}(\hat{S}(\xi) \hat{u}(\xi))$, then we have

$$g(x) = v(T, x) = (Su)(x). \quad (10)$$

Intuitively, the operator S is a convolution operator with kernel $\mathcal{F}^{-1}(\hat{S}(\xi))$. In part 3 of Section 2, we will define the operator S (formula (60)) by the solution operator of an abstract fractional evolution equation.

As mentioned by the previous works [6, 25], recovering $u(x)$ from the noisy measurement of exact $v(T, x)$ base on relation (8) in the frequency domain is ill posed due to the rapid decay of the forward process. Usually, there are two conventional

methods, namely, Tikhonov regularization and truncated Fourier transform regularization, to overcome this difficulty in the frequency domain. In 2013, Wang and Liu [25] proposed to use total variation (TV) regularization for time fractional diffusion model.

TV regularization method successfully recover the edges of the initial data and is robust for the noise. However, TV regularization method suffers from staircasing effect, namely the transformation of smooth regions into piecewise constant regions. A linear combination of the Tikhonov regularization and TV regularization with an adaptive weight is a natural way to reduce this effect. However, linear combination of two regularization terms could be seen as a compromise between Tikhonov regularization and TV regularization, which may not reflect the specific properties of functions on local regions. Hence, Blomgren et al. [31] suggest letting the exponent in the regularization term depend on the data. Li et al. [11] study the variable exponent TV regularization when exponent $1 < \bar{p}(x) \leq 2$. Harjulehto et al. [32] discuss the variable TV regularization allowing $\bar{p}(x) = 1$ for some x by using techniques developed in [33]. Boltt et al. [9] focus on the following variation model

$$\min_u J(u) = \int_{\Omega} |\nabla u|^{\bar{p}(x)} dx + \frac{\lambda}{2} \int_{\Omega} |u - g|^2 dx, \quad (11)$$

where $\lambda > 0$ is a positive constant, g is the noisy image, $\bar{p}(\cdot)$ defined as

$$\bar{p}(x) = p(u) = P_N(|\nabla(G_{\bar{\delta}} * u)(x)|^2) \quad (12)$$

with $G_{\bar{\delta}} : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a symmetric mollifier centered at 0 and belongs to $C^2 \cap W^{3,2}$. $P_N : \mathbb{R}^+ \rightarrow [1, 2]$ is a non-increasing C^2 function with $P_N(N) = 1$ with $N > 0$ is a positive real number. For example, P_N can be taken as follows:

$$P_N(s) = \begin{cases} 2 - \frac{10s^3}{N^3} + \frac{15s^4}{N^4} - \frac{6s^5}{N^5} & \text{if } s \leq N \\ 1 & \text{if } s > N \end{cases}. \quad (13)$$

They discuss how parameter choices affect recover results and prove the existence and uniqueness of minimizers. Recently, in [22], the author studies image decomposition problems by using variable TV regularization combined with variable Besov space.

We attempt to use the variable total variation regularization term to penalize our fractional backward diffusion problem. More specifically, we intend to use the following model

$$u^{\delta} = \operatorname{argmin}_{u \in \mathcal{K}} \int_{\Omega} |\nabla u|^{\bar{p}(x)} dx \text{ such that } \|Su - g^{\delta}\|_{L^2(\mathbb{R}^2)}^2 \leq \delta, \quad (14)$$

where $\delta > 0$ is error level, \mathcal{K} is some suitable admissible set of the approximate solution, $\bar{p}(x)$ defined as in (12).

By applying the Lagrangin formulation, the variable TV restoration model (14) can be transformed into the following unconstrained minimization problem:

$$u^{\delta} = \operatorname{argmin}_{u \in \mathcal{K}} \int_{\Omega} |\nabla u|^{\bar{p}(x)} dx + \frac{\lambda}{2} \|Su - g^{\delta}\|_{L^2(\mathbb{R}^2)}^2, \quad (15)$$

where λ is a positive parameter that controls the tradeoff between a good fit to the measurement data and the regularized solution. For each $\delta > 0$, there exists some λ such that (14) and (15) are equivalent.

Comparing (11) and our model (15), the forward operator is more complex than the identity operator. In order to solve (11), we can simply take

$$\bar{p}(x) = p(g) = P_N(|\nabla(G_{\bar{\delta}} * g)(x)|^2) \quad (16)$$

where g is the noisy image, and the exponent will not change during our computation, by doing this the Euler-Lagrange equation will be simpler than

$$\bar{p}(x) = p(u) = P_N(|\nabla(G_{\bar{\delta}} * u)(x)|^2). \quad (17)$$

For clarity, we list the Euler-Lagrange equations for (15) with (16) as follows

$$0 = -\nabla \cdot \left(\frac{\nabla u}{|\nabla u|} P_N(|\nabla(G_{\bar{\delta}} * g)|^2) |\nabla u|^{P_N(|\nabla(G_{\bar{\delta}} * g)|^2)-1} \right) + \lambda S^*(Su - g^\delta), \quad (18)$$

and for (15) with (17) as follows

$$\begin{aligned} 0 = & -G_{\bar{\delta}} * \nabla \cdot \left(|\nabla u|^{P_N(|\nabla(G_{\bar{\delta}} * u)|^2)} P'_N(|\nabla(G_{\bar{\delta}} * u)|^2) \cdot 2\nabla(G_{\bar{\delta}} * u) \right) \\ & - \nabla \cdot \left(|\nabla u|^{P_N(|\nabla(G_{\bar{\delta}} * u)|^2)} \frac{\nabla u}{|\nabla u|} \right) + \lambda S^*(Su - g^\delta). \end{aligned} \quad (19)$$

For model (11), because the edges will not change so much during the computation, the reduction (16) which is taken in [9] is suitable. However, for our model (15), because the edges will change dramatically during the evolution process, we must iterate the value of $\bar{p}(x)$ during our computation. To make this clear, we consider an image as the initial data and the solution $v(1, x)$ of the fractional diffusion equation (2) shown in Figure 1. The solution $v(1, x)$ shown on the right hand

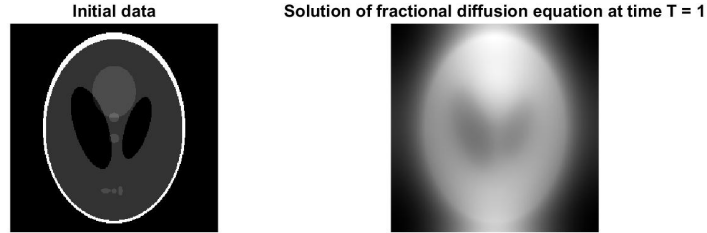


FIGURE 1. Left: Initial data; Right: The solution of the fractional diffusion equation (2) at time $T = 1$ with $\alpha = 0.6$, $\beta = 0.9$.

side of Figure 1 is calculated by Fourier transform and formula (8) with $\alpha = 0.6$, $\beta = 0.9$ and $T = 1$. The left image in Figure 2 is the boundary of the initial data detected by Canny algorithm in the Matlab toolbox. The right image in Figure 2 is the boundary of $v(1, x)$ also detected by Canny algorithm. From these figures, it is clear that the boundary of the initial data will change dramatically during the fractional evolution process as claimed in the beginning of this paragraph.

In summary, theories about existence, uniqueness and stability will be proved in a very general setting, then restricted to backward problem for equation (2) we propose a modified Bregman iterative algorithm to solve problem (15). In the following, we will describe the key point of our proof. In order to prove existence, uniqueness and stability of problem (15), we first generalize the theory constructed in [34] to our variable total variation regularization model, during the proof we propose a concept named CBV-coercive. After building the general theory, we need

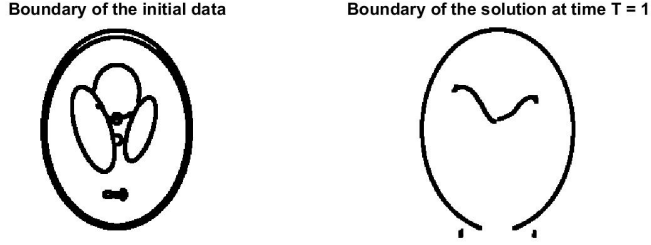


FIGURE 2. Left: Boundaries of the initial data; Right: Boundaries of the solution of the fractional diffusion equation (2) at time $T = 1$ with $\alpha = 0.6$, $\beta = 0.9$.

to verify the operator S appeared in (10) satisfy the conditions in the general theory. One of the new ingredients of this paper is that we propose an abstract fractional evolution equation (58), then prove the solution operator of this abstract evolution equation satisfy the required conditions. Through the abstract formulation, we obtain the existence, uniqueness and stability when the forward problem is (2) and in addition, all the theoretical results can be applied to more general systems. More specifically, our results is valid for system (58) in Section 2 with the spatial derivative operator $-A$ generate a C_0 -semigroup and satisfy condition (67).

The second new ingredients of this paper is that we propose a modified Bregman iterative algorithm to solve problem (15). To the best of our knowledge, researchers use Euler-Lagrange equations directly or construct an evolution process based on Euler-Lagrange equations to solve image restoration problems with variable TV regularizing term. In the traditional image restoration problem, the edges will not change dramatically for the forward operator is the identity operator, hence, we can assume (16) which highly reduce the computational task. In our setting, the edges will change during the diffusion process as shown in Figure 2, so we must iterate the exponent \bar{p} in our algorithm. Bregman iteration [37] is an efficient methods used to solve TV regularization based image restoration. In the framework of Bregman iteration methods, we obtain \bar{p}_{n+1} by using the value of recovered image u_n , so during every iteration we can use the Euler-Lagrange equations as in the case of (16). Hence, on one hand we allow the exponent \bar{p} update during each iteration. On the other hand, a simple Euler-Lagrange equation can be used to reduce the computational load. However, in our modified Bregman iterative algorithm, the regularization term changes its form at each iteration, so we need more techniques to provide theoretical analysis of our algorithm. In Section 3, we prove the convergence and provide a practical stopping criterion based on the detailed analysis.

The organization of this paper is as follows. In Section 2, we propose the concept of CBV-coercive and build a general theory then use the general theory to a general linear model with variable TV regularizing term. By using operator semigroup and fractional operator semigroup theory, we prove the backward problems for an abstract fractional differential equation satisfying the conditions in our general theory. In Section 3, we propose modified Bregman iterative algorithm, then provide detailed theoretical analysis. Finally, the numerical implementations are given in Section 4 to support our theoretical results and to show the validity of the proposed algorithm.

2. Existence, Uniqueness and Stability. In this section, we will prove existence, uniqueness and stability of our minimization problem (15) in a general setting. Here we need to clarify some notations used through all the following parts of this paper.

- d stands for dimension; $\Omega \subset \mathbb{R}^d$ is a bounded domain with Lipschitz boundary;
- C^m will stand for functions with continuous derivatives up to order m ; C_c^m stands for compactly supported function with continuous derivatives up to order m ;
- $W^{m,p}$ is the usual Sobolev space with weak derivatives of order up to m belongs to L^p ; For simplicity, we denote $H^m := W^{m,2}$ when $p = 2$; H_0^m will stand for the closure of C_c^∞ in H^m ;
- For a subset $\Omega \subset \mathbb{R}^d$, χ_Ω stands for indicator function which equal to 1 in Ω and equal to 0 outside of Ω ;
- $|\Omega|$ stands for the Lebesgue measure of $\Omega \subset \mathbb{R}^d$;
- If S is a bounded linear operator, we will denote $\|S\|$ as the operator norm of S ;
- BV in this paper stands for functions of bounded variation, the norm defined as

$$\|u\|_{BV} := \|u\|_{L^1(\Omega)} + \|u\|_{BV}$$

where $\|u\|_{BV} := \sup_{\sigma \in V} \int_{\Omega} (-u \operatorname{div} \sigma) dx$ with $V := \{\sigma \in C_c^1(\Omega; \mathbb{R}^d) : |\sigma(x)| \leq 1 \text{ for all } x \in \Omega\}$.

2.1. General Theory. In this subsection, we build a general theory for the following unconstrained minimization problem

$$\min_{u \in L^q(\Omega)} T(u). \quad (20)$$

In order to use compactness properties of function spaces for unconstrained minimization problems, we introduce the following property: define T to be CBV-coercive if

$$T(u) \rightarrow +\infty \quad \text{whenever} \quad J(u) \rightarrow +\infty, \quad (21)$$

where $J(\cdot)$ satisfies $\|u\|_{BV} \leq CJ(u)$.

Theorem 2.1. *Suppose J is defined as in (21) and T is CBV-coercive. If $1 \leq q < \frac{d}{d-1}$ and T is lower semi-continuous, then problem (20) has a solution. If in addition $q = \frac{d}{d-1}$, dimension $d \geq 2$, and T is weakly lower semi-continuous, then a solution also exists. In either case, the solution is unique if T is strictly convex.*

Proof. Let u_n be a minimizing sequence for T ; in other words,

$$\lim_{n \rightarrow \infty} T(u_n) = \inf_{u \in L^q(\Omega)} T(u) := T_{\min}. \quad (22)$$

Since T is CBV-coercive, the $\{u_n\}$ are BV-bounded. By Theorem 2.5 in [34], there exists a subsequence u_{n_k} which converges to some $\bar{u} \in L^q(\Omega)$. Convergence is weak if $q = \frac{d}{d-1}$. By the (weak) lower semi-continuity of T ,

$$T(\bar{u}) \leq \liminf_{k \rightarrow \infty} T(u_{n_k}) = T_{\min}. \quad (23)$$

Uniqueness of minimizers follows immediately from strict convexity. \square

Next, we consider a sequence of perturbed problems

$$\min_{u \in L^q(\Omega)} T_n(u). \quad (24)$$

Theorem 2.2. *Assume J is defined as in (21), $1 \leq q < \frac{d}{d-1}$ and that T and each of the T_n s are CBV-coercive, lower semi-continuous, and have a unique minimizer. Assume in addition:*

1. *Uniform CBV-coercivity: for any sequence $v_n \in L^q(\Omega)$,*

$$\lim_{n \rightarrow \infty} T_n(v_n) = \infty \quad \text{whenever} \quad \lim_{n \rightarrow \infty} J(v_n) = \infty. \quad (25)$$

2. *Consistency: $T_n \rightarrow T$ uniformly on J -bounded sets, i.e. given $B > 0$ and $\epsilon > 0$, there exists N such that*

$$|T_n(u) - T(u)| < \epsilon \quad \text{whenever} \quad n \geq N, \quad J(u) \leq B. \quad (26)$$

Then problem (20) is stable with respect to the perturbations (24), i.e. if \bar{u} minimizes T and u_n minimizes T_n , then

$$\|u_n - \bar{u}\|_{L^q(\Omega)} \rightarrow 0.$$

If $q = \frac{d}{d-1}$, $d \geq 2$, and one can replace the lower semi-continuity assumption on T and each T_n by weak lower semi-continuity, then convergence is weak:

$$u_n - \bar{u} \rightharpoonup 0.$$

Proof. Note that $T_n(u_n) \leq T_n(\bar{u})$, by assumption (2), we have

$$\liminf_{n \rightarrow \infty} T_n(u_n) \leq \limsup_{n \rightarrow \infty} T_n(u_n) \leq T(\bar{u}) < \infty$$

and hence by assumption (1), the u_n s are J -bounded. Remember the properties of J , the u_n s are BV-bounded. Now suppose our results does not hold. By Theorem 2.5 in [34], there exists a subsequence u_{n_k} which converges in $L^q(\Omega)$ (weak $L^q(\Omega)$ if $q = \frac{d}{d-1}$) to some $\hat{u} \neq \bar{u}$. By the (weak) lower semi-continuity of T ,

$$\begin{aligned} T(\hat{u}) &\leq \liminf_{k \rightarrow \infty} T(u_{n_k}) \\ &= \lim_{k \rightarrow \infty} (T(u_{n_k}) - T_{n_k}(u_{n_k})) + \liminf_{k \rightarrow \infty} T_{n_k}(u_{n_k}) \\ &\leq T(\bar{u}). \end{aligned}$$

But this contradicts the uniqueness of the minimizer \bar{u} of T . \square

2.2. Variable TV Regularization for General Linear Problems. In this subsection, we consider the following special form of T :

$$T(u) = F(u) + \frac{\lambda}{2} \|Su - g\|_{L^2(\Omega)}^2, \quad (27)$$

where

$$F(u) = \max_{\nu(x) \geq 0, |\sigma(x)| \leq 1} \int_{\Omega} (\nabla \cdot (\bar{p}\nu^{\bar{p}-1}\sigma)u - (\bar{p}-1)\nu^{\bar{p}}) dx \quad (28)$$

and ν ranges over the set of $C^1(\bar{\Omega})$ functions with positive minimum, σ ranges over functions in $C_c^1(\Omega)$ with $|\sigma(x)| \leq 1$, \bar{p} defined as in (12), S is a linear bounded operator from $L^q(\Omega)$ to $L^2(\Omega)$, g is a function in $L^2(\Omega)$, $\lambda > 0$ is a positive real number. As demonstrated in [9], if u, \bar{p} are C^1 , $F(u)$ defined above is equivalent to

$$\int_{\Omega} |\nabla u|^{\bar{p}} dx. \quad (29)$$

So instead of (15) in the previous section, in this subsection we consider T defined in (27). For this particular T , we define

$$J(u) = \|u\|_{L^2(\Omega)} + F(u) + \frac{1}{4}|\Omega|, \quad (30)$$

which obviously satisfies $\|u\|_{BV} \leq CJ(u)$.

Theorem 2.3. *Assume $d = 2$ and that $1 \leq q \leq \frac{d}{d-1}$, S is a linear bounded operator from $L^q(\Omega)$ to $L^2(\Omega)$ and g is a function in $L^2(\Omega)$. In addition, we assume that*

$$S\chi_\Omega \neq 0. \quad (31)$$

Then T defined in (27) is CBV-coercive with J defined in (30), and the functional T has a minimizer.

Proof. The lower semi-continuous or weakly lower semi-continuous of $F(u)$ follows from Theorem 9 and Theorem 12 in [9]. Hence the (weakly) lower semi-continuous of $J(u)$ obviously hold. If we can prove T in (27) is CBV-coercive, by Theorem 2.1, the functional T has a minimizer. So the main task is to prove T is CBV-coercive. Decompose u as follows:

$$u = \nu + w \quad (32)$$

where

$$w = \frac{\int_\Omega u dx}{|\Omega|} \chi_\Omega \quad \int_\Omega \nu dx = 0. \quad (33)$$

Using Poincaré inequality and Hölder's inequality, there exists a positive constant C such that for any q such that $1 \leq q \leq \frac{d}{d-1} = r$,

$$\begin{aligned} \|\nu\|_{L^q(\Omega)} &\leq |\Omega|^{\frac{1}{q} - \frac{1}{r}} \|\nu\|_{L^r(\Omega)} \\ &\leq (|\Omega| + 1)^{1 - \frac{1}{r}} C \|\nu\|_{BV} \\ &\leq C_1 \left(F(\nu) + \frac{1}{4}|\Omega| \right) \end{aligned} \quad (34)$$

where $C_1 := (|\Omega| + 1)^{\frac{1}{d}} C$. In the last inequality of (34), we used (35) in [9]. Using (34) and the decomposition (32), we have

$$\begin{aligned} J(u) &= \|u\|_{L^2(\Omega)} + F(u) + \frac{1}{4}|\Omega| \\ &\leq \|w\|_{L^2(\Omega)} + (C_1 + 1)(F(\nu) + \frac{1}{4}|\Omega|). \end{aligned} \quad (35)$$

From the assumption (31), there exists $C_2 > 0$ such that

$$\|Sw\|_{L^2(\Omega)} = C_2 \|w\|_{L^2(\Omega)}. \quad (36)$$

From the definition of T and the decomposition (32), we obtain

$$\begin{aligned} T(u) &= F(u) + \frac{\lambda}{2} \|(S\nu - g) + Sw\|_{L^2}^2 \\ &\geq F(u) + \frac{\lambda}{2} (\|S\nu - g\|_{L^2} - \|Sw\|_{L^2})^2 \\ &\geq F(u) + \frac{\lambda}{2} \|Sw\|_{L^2} (\|Sw\|_{L^2} - 2\|S\nu - g\|_{L^2}). \end{aligned} \quad (37)$$

By (34), we obtain

$$\|S\nu - g\|_{L^2} \leq C_1\|S\|F(u) + \frac{1}{4}C_1\|S\|\|\Omega\| + \|g\|_{L^2}. \quad (38)$$

Combining (36), (37) and (38), we have

$$\begin{aligned} T(u) \geq & F(u) + \frac{\lambda}{2}C_2\|w\|_{L^2}(C_2\|w\|_{L^2} \\ & - 2(C_1\|S\|F(u) + \frac{1}{4}C_1\|S\|\|\Omega\| + \|g\|_{L^2})). \end{aligned} \quad (39)$$

From the definition of $T(u)$ in (27), we obtain that

$$F(u) \leq T(u). \quad (40)$$

Case 1: $C_2\|w\|_{L^2} - 2(C_1\|S\|F(u) + \frac{1}{4}C_1\|S\|\|\Omega\| + \|g\|_{L^2}) \geq 1$. From (39), we obtain

$$\|w\|_{L^2} \leq \frac{2}{\lambda C_2}T(u). \quad (41)$$

Hence, considering (40), we finally obtain

$$J(u) \leq \left(\frac{2}{\lambda C_2} + (C_1 + 1) \right) T(u) + \frac{C_1 + 1}{4}\|\Omega\|. \quad (42)$$

Case 2: $C_2\|w\|_{L^2} - 2(C_1\|S\|F(u) + \frac{1}{4}C_1\|S\|\|\Omega\| + \|g\|_{L^2}) < 1$. Obviously, we have

$$\|w\|_{L^2} \leq \frac{1 + 2(C_1\|S\|F(u) + \frac{1}{4}C_1\|S\|\|\Omega\| + \|g\|_{L^2})}{C_2} \quad (43)$$

Combining (40) with the above inequality, we obtain

$$J(u) \leq \left(\frac{2C_1\|S\|}{C_2} + C_1 + 1 \right) T(u) + \frac{1 + \frac{1}{4}C_1\|S\|\|\Omega\| + \|g\|_{L^2}}{C_2}. \quad (44)$$

Considering case 1 and case 2, (42) and (44) yields the CBV-coercivity of T . \square

Remark 1. In the above theorem, if we assume the solution lies in a closed convex subset \mathcal{K} of L^q with $1 \leq q \leq \frac{d}{d-1}$, T can be written as

$$T(u) = F(u) + \frac{\lambda}{2}\|Su - g\|_{L^2(\Omega)}^2 + \frac{1}{\chi_{\mathcal{K}}}. \quad (45)$$

Because the above T is strictly convex, using same procedure as in the proof of theorem 2.3, we obtain the following theorem.

Theorem 2.4. Assume $d = 2$ and that $1 \leq q \leq \frac{d}{d-1}$, \mathcal{K} is a closed convex subset of $L^q(\Omega)$, S is a linear bounded operator from $L^q(\Omega)$ to $L^2(\Omega)$ and g is a function in $L^2(\Omega)$. In addition, we assume that

$$S\chi_{\Omega} \neq 0. \quad (46)$$

Then T defined in (27) is CBV-coercive with J defined in (30), and the functional T has a unique constrained minimizer over \mathcal{K} .

Next, we addresses the stability of minimizers to functionals of (27). Consider perturbed functionals

$$T_n(u) = F(u) + \frac{\lambda}{2}\|S_n u - g_n\|_{L^2(\Omega)}^2, \quad (47)$$

Theorem 2.5. Assume $1 \leq q \leq \frac{d}{d-1}$, $\lim_{n \rightarrow \infty} \|g_n - g\|_{L^2(\Omega)} = 0$, the S_n s are each bounded linear and converge pointwise to S , and for each n ,

$$\|S_n \chi_\Omega\|_{L^2} \geq \gamma > 0. \quad (48)$$

Also assume each S_n has a unique minimizer u_n and that S has a unique minimizer \bar{u} . Then for $1 \leq q < \frac{d}{d-1}$, we have

$$\lim_{n \rightarrow \infty} \|u_n - \bar{u}\|_{L^q(\Omega)} = 0, \quad (49)$$

for $q = \frac{d}{d-1}$, the convergence is weak

$$u_n \rightharpoonup \bar{u}. \quad (50)$$

Proof. It suffices to show that conditions (1) and (2) of Theorem 2.2 hold. For condition (1), put $u_n = \nu_n + w_n$ as in (32) and (33), and repeat the proof of theorem 2.2. Since $\|S_n \chi_\Omega\|_{L^2} \geq \gamma \sqrt{|\Omega|} \|w_n\|_{L^2(\Omega)}$, letting M be an upper bound on S and each S_n , m is an upper bound on $\|g\|_{L^2}$ and each $\|g_n\|_{L^2}$, one obtains

$$\begin{aligned} T_n(u_n) \geq & F(u_n) + \frac{\lambda}{2} \gamma \sqrt{|\Omega|} \|w_n\|_{L^2} (\gamma \sqrt{|\Omega|} \|w_n\|_{L^2} \\ & - 2(C_1 M F(u_n) + \frac{1}{4} C_1 \|S_n\| \text{Vol}(\Omega) + m)). \end{aligned} \quad (51)$$

This yields uniform CBV-coercivity by same argument as in the proof of Theorem 2.2. Since

$$\begin{aligned} |T_n(u) - T(u)| \leq & \frac{\lambda}{2} (\|S_n u - S u\|_{L^2} + \|g_n - g\|_{L^2}) (\|S_n\| + \|S\|) \|u\|_{L^2} \\ & + \|g_n\|_{L^2} + \|g\|_{L^2}, \end{aligned}$$

condition (2) is obviously satisfied. \square

2.3. Variable TV Regularization for Fractional Backward Diffusion. In this subsection, we firstly construct an abstract fractional evolution equation based on the following time-space fractional diffusion system for homogeneous media:

$$\begin{cases} \partial_t^\alpha v(t, x) + (-\Delta)^\beta v(t, x) = 0 & (x, t) \in \mathbb{R}^2 \times (0, \infty) \\ v(x, 0) = u(x) & x \in \mathbb{R}^2 \end{cases}. \quad (52)$$

with $0 < \alpha \leq 1$, $\beta \in (1/2, 1]$. u is the initial data in \mathbb{R}^2 with compact support in a bounded convex open subset Ω of the plane with Lipschitz continuous boundary $\partial\Omega$.

Define an operator $A = -\Delta$ with domain

$$D(A) = \{u \in H_0^1(\mathbb{R}^2) : \Delta u \in L^2(\mathbb{R}^2)\}. \quad (53)$$

Let $X = L^2(\mathbb{R}^2)$, then A is a bounded linear operator defined on X . By Theorem 2.4.1 in [7], we know that A is m-accretive and $-A$ generates a uniformly exponentially stable and contractive C_0 -semigroup, denoted as $T(t)$. From the proof of Theorem 2.4.1 in [7], we know that there exists a constant $c > 0$ such that

$$\|T(t)\| \leq e^{-ct}, \quad (54)$$

where $\|\cdot\|$ is the operator norm. Considering (54), for $\frac{1}{2} < \beta \leq 1$, we can define the following bounded linear operator

$$A^{-\beta} x = \frac{1}{\Gamma(\beta)} \int_0^\infty t^{\beta-1} T(t) x dt \quad \forall x \in X. \quad (55)$$

Then as illustrated in [7], [26] or [29], we can define the operator A^β as the inverse operator of $A^{-\beta}$. Hence, we have

$$A^\beta A^{-\beta}x = x \quad \forall x \in X \quad \text{and} \quad D(A^\beta) = R(A^{-\beta}). \quad (56)$$

For any $x \in D(A)$, the above positive power of operator A has the following Balakrishnan representations [26]

$$A^\beta x = \frac{\sin(\pi\alpha)}{\pi} \int_0^\infty t^{\alpha-1} (tI + A)^{-1} A x dt \quad \forall x \in D(A), \quad (57)$$

which may provide more intuitive ideas to the readers. With these preparations, we can recast system (52) into the following abstract ordinary differential equations on Banach space $X = L^2(\mathbb{R}^2)$ as follows

$$\begin{cases} \partial_t^\alpha v(t) + A^\beta v(t) = 0 & t \in (0, \infty) \\ v(0) = u \end{cases}. \quad (58)$$

Remark 2. Usually the fractional Laplacian operator defined as

$$(-\Delta)^\beta v := \mathcal{F}^{-1}(|\xi|^{2\beta} \hat{v}).$$

So in order to obtain a meaningful abstract form (58), we need to state the equivalence of A^β defined in (57) and the usual definition by Fourier transform. By Proposition 8.3.3 in [26], we know that if $v \in D(A^\beta)$ then $A^\beta v = \mathcal{F}^{-1}(|\xi|^{2\beta} \hat{v})$. That is to say if $v \in H_0^1(\mathbb{R}^2) \cap H^{2\beta}(\mathbb{R}^2)$, the abstract form is equivalent to the usual definition by Fourier transform.

Remark 3. The operator A can be defined more generally as follow

$$Au := -\nabla \cdot (a(x)\nabla u) + c(x)u \quad \text{if } u \in D(A),$$

where $a(\cdot), c(\cdot)$ are functions in C^1 , $D(A)$ defined as in (53) and in addition, we suppose the operator A defined above satisfies the strong elliptic condition. It is well known that this operator generate a contractive C_0 -semigroup [24] and we can define fractional operator as in (55), (56) and (57), so the abstract fractional evolution equation (58) incorporate a natural generalization of fractional Laplace operator. In the following part of this article, we only present the proof of the Laplace case. That is because once we define the general operator mentioned in this remark appropriately, it satisfies all the properties of the operator semigroup which we used and the proof will be almost same.

Now we can prove the main theorem in this subsection.

Theorem 2.6. *Let \mathcal{K} be a closed convex subset of $L^2(\Omega)$, g^δ is the measurement data in time T . Then the optimization problem*

$$\operatorname{argmin}_{u \in \mathcal{K}} F(u) + \frac{\lambda}{2} \|Su - g^\delta\|_{L^2(\mathbb{R}^2)} \quad (59)$$

has a unique minimizer over \mathcal{K} for any fixed $\lambda > 0$.

Proof. We attempt to use Theorem 2.4 to obtain the result, so we need to verify that the operator S is bounded from L^2 to L^2 , and $S\chi_\Omega \neq 0$.

Step 1. Bounded of operator S . Taking the measure μ in [19] to be $\delta_\beta(\cdot)$, using Theorem 3.7 in [19], we find that $-A^\beta$ generates a bounded C_0 -semigroup denoted as $S_\beta(t)$. For simplicity, we denote S_β as $S_\beta(T)$ for short. Using Corollary 2.10 in [8], we know that $-A^\beta$ generate an α -order fractional semigroup proposed in

[21]. We denote the α -order fractional semigroup generated by $-A^\beta$ as $S_{\alpha,\beta}(t)$, particularly for $t = T$, denote $S_{\alpha,\beta}$ as $S_{\alpha,\beta}(T)$ for short. Instead of the semigroup property, this α -order fractional semigroup satisfies the following equality

$$\begin{aligned} \int_0^{t+s} \frac{S_{\alpha,\beta}(\tau)}{(t+s-\tau)^\alpha} d\tau - \int_0^t \frac{S_{\alpha,\beta}(\tau)}{(t+s-\tau)^\alpha} d\tau - \int_0^s \frac{S_{\alpha,\beta}(\tau)}{(t+s-\tau)^\alpha} d\tau \\ = \alpha \int_0^t \int_0^s \frac{S_{\alpha,\beta}(\tau_1)S_{\alpha,\beta}(\tau_2)}{(t+s-\tau_1-\tau_2)^{1+\alpha}} d\tau_1 d\tau_2 \end{aligned}$$

in the strong operator topology. Noting our definition of $v(x, 1)$ for $x \notin \Omega$ in Section 1, we have the following expression for operator S :

$$(Su)(x) = \begin{cases} (S_{\alpha,\beta}u)(x), & x \in \Omega \\ g^\delta(x), & x \notin \Omega. \end{cases} \quad (60)$$

Now the meaning of the operator S is not restricted to the one used in (10). In order to obtain our results, we need to introduce the following lemma (restated in our setting) proved in Section 3 of [8].

Lemma 2.7. *Let $S_\beta(t)$ is a bounded C_0 -semigroup, $S_{\alpha,\beta}(t)$ is an α -order fractional semigroup with $\alpha \in (0, 1)$. Then the following representation holds*

$$S_{\alpha,\beta}(t) = \int_0^\infty \phi_{t,\alpha}(s) S_\beta(s) ds, \quad t > 0, \quad (61)$$

where $\phi_{t,\alpha}(s) = t^{-\alpha} \Phi_\alpha(st^{-\alpha})$, $\Phi_\alpha(z)$ is the function of Wright type defined as

$$\Phi_\alpha(z) := \sum_{n=0}^\infty \frac{(-z)^n}{n! \Gamma(-\alpha n + 1 - \alpha)}, \quad 0 < \alpha < 1, \quad (62)$$

and (61) holds in the strong sense.

For functions of Wright type, there are many usefully literatures [10, 12, 13]. Here we only recall that $\Phi_\alpha(t)$ is a probability density function satisfies:

$$\Phi_\alpha(t) \geq 0, \quad t > 0; \quad \int_0^\infty \Phi_\alpha(t) dt = 1. \quad (63)$$

Considering Lemma 2.7 and the above properties (63), for every $u \in X$, we have

$$\begin{aligned} \|Su\|_{L^2(\Omega)} &= \|S_{\alpha,\beta}u\|_{L^2(\Omega)} \leq \int_0^\infty \phi_{T,\alpha}(s) \|S_\beta(s)u\|_{L^2} ds \\ &\leq C \int_0^\infty \phi_{T,\alpha}(s) ds \|u\|_{L^2} \\ &= C \|u\|_{L^2(\Omega)}, \end{aligned} \quad (64)$$

where we used the fact that $S_\beta(s)$ is a bounded C_0 -semigroup.

Step 2. Operator S does not annihilate constant functions. Assume $\|S_{\alpha,\beta}\chi_\Omega\|_{L^2(\Omega)} = 0$, that is

$$\int_0^\infty \phi_{T,\alpha}(s) S_\beta(s) \chi_\Omega ds = 0 \quad a.e. \quad (65)$$

By the definition of $\phi_{T,\alpha}(s)$, we know that

$$\int_0^\infty T^{-\alpha} \Phi_\alpha(T^{-\alpha}s) S_\beta(s) \chi_\Omega ds = 0 \quad a.e. \quad (66)$$

Since $\Phi_\alpha(s)$ is a probability density, if $S_\beta(s)\chi_\Omega \geq 0$ a.e., we can conclude that $S_\beta(s)\chi_\Omega = 0$ a.e. for almost all $s > 0$. By the strong continuity of C_0 -semigroup, we will obtain $\lim_{s \rightarrow 0} \|S_\beta(s)\chi_\Omega - \chi_\Omega\|_{L^2(\mathbb{R}^2)} = 0$, that is to say $\|\chi_\Omega\|_{L^2(\mathbb{R}^2)} = 0$ which is a contradiction. So $\|S\chi_\Omega\|_{L^2(\Omega)} = \|S_{\alpha,\beta}\chi_\Omega\|_{L^2(\Omega)} > 0$ which means $S\chi_\Omega \neq 0$ in $L^2(\Omega)$.

Now we verify the positive condition $S_\beta(s)\chi_\Omega \geq 0$ a.e.. Since $\chi_\Omega \geq 0$, we just need to verify that $S_\beta(s)$ is a positive preserving operator semigroup. For more properties of positive preserving operator semigroup, we refer to [30, 40]. Here, for the reader's convenience, we give the following lemma (Theorem 4.6.14 in [30]) which is useful for our proof.

Lemma 2.8. *Let $\{T(t)\}_{t \geq 0}$ be a strongly continuous contraction semigroup on the space $L^p(\mathbb{R}^d, \mathbb{R})$, $1 < p < \infty$, with generator $(-A, D(A))$. The semigroup $\{T(t)\}_{t \geq 0}$ is positive preserving if and only if*

$$\int_{\mathbb{R}^d} (-Au)(u^+)^{p-1} dx \leq 0 \quad (67)$$

holds for all $u \in D(A)$.

In our setting, $A = -\Delta$, $p = 2$, (67) can be verified as follows

$$\int_{\mathbb{R}^d} (\Delta u)u^+ dx = - \int_{\mathbb{R}^d} \nabla u \cdot \nabla u^+ dx \quad (68)$$

$$= - \int_{\mathbb{R}^d} |\nabla u^+|^2 dx \leq 0, \quad (69)$$

where we used

$$\nabla u^+ = \begin{cases} \nabla u, & \text{a.e. on } \{u > 0\} \\ 0, & \text{a.e. on } \{u \leq 0\}. \end{cases}$$

Hence, the semigroup stated in (54) is positive preserving. Then from the proof of Theorem 3.7 in [19], there exists a probability measures $\{\mu(t)\}_{t \geq 0}$ such that

$$S_\beta(t) = \int_0^\infty T(s)\mu(t)ds. \quad (70)$$

Considering (70), the C_0 -semigroup $S_\beta(t)$ is obviously positive preserving that is $S_\beta(t)\chi_\Omega \geq 0$ a.e.. \square

At last, let us consider the stability of the minimizer of (59) with respect to the perturbations on the operator S and g^δ for any fixed $\lambda > 0$. We state this result as follows.

Theorem 2.9. *Let the hypothesis in Theorem 2.6 be satisfied. Define two functionals*

$$\begin{aligned} T_*(u) &:= F(u) + \frac{\lambda}{2} \|Su - g\|_{L^2(\mathbb{R}^2)}^2, \\ T_n(u) &:= F(u) + \frac{\lambda}{2} \|S_n u - g_n\|_{L^2(\mathbb{R}^2)}^2, \quad n = 1, 2, \dots \end{aligned}$$

and denote by

$$u_* := \operatorname{argmin}_{u \in \mathcal{K}} T_*(u), \quad u_n := \operatorname{argmin}_{u \in \mathcal{K}} T_n(u).$$

If $\lim_{n \rightarrow \infty} \|g_n - g\|_{L^2(\mathbb{R}^2)} = 0$, S_n converge pointwise to S as $n \rightarrow \infty$, then it follows that $u_n \rightharpoonup u$ as $n \rightarrow \infty$ in the weak topology of $L^2(\Omega)$.

Proof. We will use Theorem 2.5 to obtain the above result. Here we just need to verify condition (48). From the proof of Theorem 2.6, we know that there exists a constant $c > 0$ such that

$$\|S\chi_\Omega\|_{L^2} \geq c > 0. \quad (71)$$

Since here we just concern the case with large enough n and S_n converge to S pointwise, we can find a positive constant $N > 0$ such that if $n \geq N$ then we have

$$\|S_n\chi_\Omega - S\chi_\Omega\|_{L^2} \leq \frac{c}{2}. \quad (72)$$

From (71) and (72), we obtain

$$\|S_n\chi_\Omega\|_{L^2} \geq \frac{c}{2} > 0 \quad (73)$$

for every $n \geq N$. \square

Remark 4. For the general operator A mentioned in remark 3, in order to obtain the same results as in the Laplace case, we may need to assume the function $c(\cdot)$ has a small positive bound to verify the conditions mentioned in lemma 2.8. In summary, the proof of Theorem 2.6 and Theorem 2.9 just require $-A$ generate a contraction C_0 -semigroup and satisfies condition (67). In addition, by some modifications, we may obtain similar results for space distributed order fractional diffusion equations studied in [19, 5]. Because the modifications is not trivial, it may be more appropriate to report in another paper.

3. Numerical Approach. The well-posedness of our optimization problem (59) is obtained in the previous section. In this section, we propose modified Bregman iterative method to solve (59) efficiently. The Bregman distance associated with a convex functional $F(\cdot)$ between points u and v is defined as

$$D_F^p(u, v) := F(u) - F(v) - \langle p, u - v \rangle, \quad (74)$$

where $p \in \partial F(v) = \{w : F(u) - F(v) \geq \langle w, u - v \rangle, \forall u\}$ is the sub-gradient of $F(\cdot)$ at the point v . Then our optimization problem (59) becomes

$$u_* := \operatorname{argmin}_{u \in \mathcal{K}} F(u_*) + \langle p, u - u_* \rangle + D_F^p(u, u_*) + \frac{\lambda}{2} \|Su - g^\delta\|_{L^2(\mathbb{R}^2)}^2 \quad (75)$$

with $p \in \partial F(u_*)$. Since the forward problem can be solved efficiently in the frequency domain, by Parseval identity, we have the following equivalent form

$$u_* := \operatorname{argmin}_{u \in \mathcal{K}} F(u_*) + \langle p, u - u_* \rangle + D_F^p(u, u_*) + \frac{\lambda}{2} \|\hat{S}\hat{u} - \hat{g}^\delta\|_{L^2(\mathbb{R}^2)}^2 \quad (76)$$

where \hat{S} defined as in (8).

Instead of solving (14), Osher et al. [37] proposed Bregman iterative regularization to solve (76) approximately by using the following iterative formula:

$$u^{m+1} = \operatorname{argmin}_u D_F^p(u, u^m) + \frac{\lambda}{2} \|\hat{S}\hat{u} - \hat{g}^\delta\|_{L^2(\mathbb{R}^2)}^2 \quad (77)$$

for $m = 0, 1, 2, \dots$, beginning with $p^0 = u^0 = 0$. In order to simplify the computation, we introduce

$$F_m(u) = \int_{\Omega} |\nabla u|^{\bar{p}^m(x)} dx, \quad (78)$$

where $\bar{p}^m(x) = P_N(|\nabla(G_{\bar{\delta}} * u^m)(x)|^2)$. Instead of $F(u)$ (\bar{p} dependent on u) by $F_m(u)$ defined above, the Euler-Lagrange equation will be significantly simplified for each step as illustrated in (18) and (19). Through this simplification, we can still capture the change of edges during evolution process by just solving a simple Euler-Lagrange equation.

Using the definition of Bregman distance (74), problem (77) will becomes

$$u^{m+1} = \operatorname{argmin}_u F_m(u) - F_m(u^m) - \langle p^m, u - u^m \rangle + \frac{\lambda}{2} \|\hat{S}\hat{u} - \hat{g}^\delta\|_{L^2(\mathbb{R}^2)}^2. \quad (79)$$

From the results in Section 2, it is easy to find that u_1 is well defined. By similar arguments as deriving (3.4), (3.5) and (3.6) in [25], we can also deduce that $\{u^m : m \in \mathbb{N}\}$ is well defined. Now let us firstly provide a recursive procedure which can solve (77) numerically in Algorithm 1. Then, let us prove the properties of

Algorithm 1 Recursive Procedure

set

$$\begin{aligned} (u^0, p^0) &= (0, 0) \\ F_0(u) &= \int_{\Omega} |\nabla u|^{P_N(|\nabla(G_{\bar{\delta}} * g^\delta)(x)|^2)} dx \\ u^1 &= \operatorname{argmin}_u F_0(u) + \frac{\lambda}{2} \|\hat{S}\hat{u} - \hat{g}^\delta\|_{L^2(\mathbb{R}^2)}^2 \\ \hat{f}^1 &= \hat{g}^\delta - \hat{S}\hat{u}^1 \end{aligned}$$

repeat

$$\begin{aligned} u^{m+1} &= \operatorname{argmin}_u F_m(u) + \frac{\lambda}{2} \|\hat{S}\hat{u} - \hat{g}^\delta - \hat{f}^m\|_{L^2(\mathbb{R}^2)}^2 \\ \hat{f}^{m+1} &= \hat{f}^m + \hat{g}^\delta - \hat{S}\hat{u}^{m+1} \end{aligned}$$

until Some stopping condition is satisfied

$\{u^m : m \in \mathbb{N}\}$ appeared in Algorithm 1 and the stopping criterion for iteration based on (79). Since the regularizing term changed during iteration in our case, the proof is more complex than the TV regularization case.

Theorem 3.1. *For any fixed $\lambda > 0$, the data fitting error from the iteration is non-increasing, i.e.,*

$$\|\hat{S}\hat{u}^{m+1} - \hat{g}^\delta\|_{L^2(\mathbb{R}^2)} \leq \|\hat{S}\hat{u}^m - \hat{g}^\delta\|_{L^2(\mathbb{R}^2)}, \quad m = 1, 2, \dots \quad (80)$$

Moreover, it follows that

$$\|\hat{S}\hat{u}^M - \hat{g}^\delta\|_{L^2(\mathbb{R}^2)}^2 \leq \frac{2}{\lambda M} (\|\nabla u_*\|_{L^2} + |\Omega|) + \delta^2, \quad (81)$$

where u_* is the minimizer of functional $\|\hat{S}\hat{u} - \hat{g}^\delta\|_{L^2(\mathbb{R}^2)}^2$, $\delta > 0$ is the known error level.

Proof. Obviously, we have

$$\begin{aligned}
 \frac{\lambda}{2} \|\hat{S}\hat{u}^{m+1} - \hat{g}^\delta\|_{L^2}^2 &\leq D_{F_m}^{p^m}(u^{m+1}, u^m) + \frac{\lambda}{2} \|\hat{S}\hat{u}^{m+1} - \hat{g}^\delta\|_{L^2}^2 \\
 &\leq D_{F_m}^{p^m}(u^m, u^m) + \frac{\lambda}{2} \|\hat{S}\hat{u}^m - \hat{g}^\delta\|_{L^2}^2 \\
 &= \frac{\lambda}{2} \|\hat{S}\hat{u}^m - \hat{g}^\delta\|_{L^2}^2.
 \end{aligned} \tag{82}$$

By direct computations, we can obtain

$$\begin{aligned}
 &D_{F_m}^{p^m}(u, u^m) - D_{F_{m-1}}^{p^{m-1}}(u, u_{m-1}) + D_{F_{m-1}}^{p^{m-1}}(u^m, u^{m-1}) \\
 &= F_m(u) - F_m(u^m) + \langle p^m, u^m - u \rangle - F_{m-1}(u) + F_{m-1}(u^{m-1}) \\
 &\quad - \langle u_{m-1} - u, p_{m-1} \rangle + F_{m-1}(u_m) - F_{m-1}(u_{m-1}) \\
 &\quad + \langle u_{m-1} - u_m, p^{m-1} \rangle \\
 &= F_m(u) - F_m(u^m) - F_{m-1}(u) + F_{m-1}(u^m) + \langle u^m - u, p^m - p^{m-1} \rangle.
 \end{aligned} \tag{83}$$

Summing up the following estimates into (83)

$$\begin{aligned}
 \langle u^m - u, p^m - p^{m-1} \rangle &= \frac{\lambda}{2} \langle u^m - u, -\mathcal{F}^{-1}(\partial_{\hat{u}} \|\hat{S}\hat{u} - \hat{g}^\delta\|_{L^2}^2|_{\hat{u}=\hat{u}^m}) \rangle \\
 &= \frac{\lambda}{2} \langle \hat{u}^m - \hat{u}, -\partial_{\hat{u}} \|\hat{S}\hat{u} - \hat{g}^\delta\|_{L^2}^2|_{\hat{u}=\hat{u}^m} \rangle \\
 &\leq \frac{\lambda}{2} (\|\hat{S}\hat{u} - \hat{g}^\delta\|_{L^2}^2 - \|\hat{S}\hat{u}^m - \hat{g}^\delta\|_{L^2}^2),
 \end{aligned} \tag{84}$$

then we have

$$\begin{aligned}
 &D_{F_m}^{p^m}(u, u^m) - D_{F_{m-1}}^{p^{m-1}}(u, u_{m-1}) + D_{F_{m-1}}^{p^{m-1}}(u^m, u^{m-1}) \\
 &\leq F_m(u) - F_m(u^m) - F_{m-1}(u) + F_{m-1}(u^m) \\
 &\quad + \frac{\lambda}{2} (\|\hat{S}\hat{u} - \hat{g}^\delta\|_{L^2}^2 - \|\hat{S}\hat{u}^m - \hat{g}^\delta\|_{L^2}^2).
 \end{aligned} \tag{85}$$

Take $u = u_*$ in (85) and rewrite it as

$$\begin{aligned}
 D_{F_m}^{p^m}(u_*, u^m) + \frac{\lambda}{2} \|\hat{S}\hat{u}^m - \hat{g}^\delta\|_{L^2}^2 &\leq F_m(u) - F_m(u^m) - F_{m-1}(u) \\
 &\quad + F_{m-1}(u^m) + D_{F_{m-1}}^{p^{m-1}}(u_*, u^{m-1}) \\
 &\quad - D_{F_{m-1}}^{p^{m-1}}(u^m, u^{m-1}) + \frac{\lambda}{2} \delta^2.
 \end{aligned} \tag{86}$$

Taking summation for $m = 1, \dots, M$ yields

$$D_{F_M}^{p^M}(u_*, u^M) + \frac{\lambda}{2} \sum_{m=1}^M \|\hat{S}\hat{u}^m - \hat{g}^\delta\|_{L^2}^2 \leq F_M(u_*) + \frac{\lambda}{2} M \delta^2, \tag{87}$$

where we used $u^0 = p^0 = 0$. Considering $D_{F_M}^{p^M}(u_*, u^M) \geq 0$, we finally arrive at

$$\|\hat{S}\hat{u}^M - \hat{g}^\delta\|_{L^2}^2 \leq \frac{2}{\lambda M} (\|\nabla u_*\|_{L^2} + |\Omega|) + \delta^2. \tag{88}$$

□

Theorem 3.2. *Assume that u is the exact initial distribution. The optimal strategy for regularizing parameters (λ, M) is that*

$$\frac{2}{\lambda M}(\|\nabla u_*\|_{L^2} + |\Omega|) = \delta^2. \quad (89)$$

For such a strategy, we have the optimization convergence rate

$$\|\hat{S}(\hat{u}^M - \hat{u})\|_{L^2(\mathbb{R}^2)} \leq (\sqrt{2} + 1)\delta \quad \text{as } \delta \rightarrow 0. \quad (90)$$

Proof. The results can be obtained by using Theorem 3.1 and the following estimate

$$\|\hat{S}(\hat{u}^M - \hat{u})\|_{L^2} \leq \|\hat{S}\hat{u}^M - \hat{g}^\delta\|_{L^2} + \|\hat{S}\hat{u} - \hat{g}^\delta\|_{L^2} \leq \|\hat{S}\hat{u}^M - \hat{g}^\delta\|_{L^2} + \delta.$$

□

Since $\|\nabla u_*\|_{L^2}$ is unknown, the strategy (89) can not be implemented numerically for choosing (λ, M) . An implementable scheme is that we fix λ , then choosing the iteration stopping value M such that

$$\|\hat{S}\hat{u}^M - \hat{g}^\delta\|_{L^2(\mathbb{R}^2)} \leq \tau\delta \quad (91)$$

is satisfied first time for some specified $\tau > 1$.

For our problem, we must consider how to evaluate $\bar{p}(x)$ during each iteration. In order to simplify the computation and reduce the number of parameters needed to be specify, here we use a surrogate for $\bar{p}(x) = P_N(|\nabla(G_{\bar{\delta}} * u)(x)|^2)$. The essential idea of $\bar{p}(x)$ is that its value is 1 near the edges and its value is 2 away from the edges. So we can use some simple algorithm to detect the edges firstly, then based on the estimated edges build our exponent $\bar{p}(x)$. Denote u^m to be the input of the $(m+1)$ s iterate, we need to specify $\bar{p}(x)$ from u^m . Using some simple edge detection algorithm, we will obtain the estimated edges denoted as $E(u^m)$ which is an image with value 1 on the detected edges and with value 0 away from the edges. Then, we define

$$\bar{p}^m(x) = 2I - G_{\bar{\delta}} * E(u^m) \quad (92)$$

where I is the matrix which has the same dimension as u^m and each element of I is equal to 1, $G_{\bar{\delta}}$ defined as in (12) is a smoothing kernel.

Now we are ready to consider the algorithm for solving

$$u^{m+1} = \operatorname{argmin}_u F_m(u) + \frac{\lambda}{2} \|\hat{S}\hat{u} - \hat{g}^m\|_{L^2}^2 \quad (93)$$

with $\hat{g}^m = \hat{g}^\delta + \hat{f}^m$ for $m = 0, 1, \dots$. Firstly, we introduce a new function w to represent the gradient term ∇u in optimization problem (93), which generates an equivalent constrained convex optimization problem:

$$\min_{u, w} \left\{ \int_{\Omega} |w|^{\bar{p}^m(x)} dx + \frac{\lambda}{2} \|\hat{S}\hat{u} - \hat{g}^m\|_{L^2}^2 \right\} \quad \text{such that } w(x) = \nabla u(x). \quad (94)$$

Secondly, we split the domain Ω into three parts, for some small constant $\epsilon > 0$,

$$\begin{aligned} \Omega_1 &:= \{x \in \Omega : 1 \leq \bar{p}^m(x) < 1 + \epsilon\}, \\ \Omega_2 &:= \{x \in \Omega : 1 + \epsilon \leq \bar{p}^m(x) \leq 2 - \epsilon\}, \\ \Omega_3 &:= \{x \in \Omega : 2 - \epsilon < \bar{p}^m(x) \leq 2\}. \end{aligned} \quad (95)$$

Approximately, we can take $\bar{p}(x) = 1$ on Ω_1 and $\bar{p}(x) = 2$ on Ω_3 . Hence, we can rewrite (94) as follows

$$\min_{u,w} \left\{ \int_{\Omega_1} |w| dx + \int_{\Omega_2} |w|^{\bar{p}^m(x)} dx + \int_{\Omega_3} |w|^2 dx + \frac{\lambda}{2} \|\hat{S}\hat{u} - \hat{g}^m\|_{L^2}^2 \right\} \quad (96)$$

such that $w(x) = \nabla u(x)$.

Based on the domain decomposition, we define $w_1 := w|_{\Omega_1}$, $w_2 := w|_{\Omega_2}$ and $w_3 := w|_{\Omega_3}$.

Thirdly, by using the splitting technique, we construct an iterative procedure of alternately solving a pair of easy subproblems. The first three subproblems can be called 'w-subproblem' for fixed $u = u^*$:

$$\operatorname{argmin}_{w_1} \left\{ \int_{\Omega_1} |w_1| dx + \frac{\tilde{\lambda}}{2} \|w_1 - \nabla u^*\|_{L^2(\Omega_1)}^2 \right\}, \quad (97)$$

$$\operatorname{argmin}_{w_2} \left\{ \int_{\Omega_2} |w_2|^{\bar{p}^m(x)} dx + \frac{\tilde{\lambda}}{2} \|w_2 - \nabla u^*\|_{L^2(\Omega_2)}^2 \right\}, \quad (98)$$

$$\operatorname{argmin}_{w_3} \left\{ \int_{\Omega_3} |w_3|^2 dx + \frac{\tilde{\lambda}}{2} \|w_3 - \nabla u^*\|_{L^2(\Omega_3)}^2 \right\}. \quad (99)$$

The last subproblem is the 'u-subproblem' for fixed $w = w^*$:

$$\operatorname{argmin}_u \left\{ \frac{\lambda}{2} \|\hat{S}\hat{u} - \hat{g}^m\|_{L^2(\Omega)}^2 + \frac{\tilde{\lambda}}{2} \|w^* - \nabla u\|_{L^2(\Omega)}^2 \right\}. \quad (100)$$

Using the definition of Frechet derivatives and standard computations, we can easily obtain the minimizer of the subproblems (97), (99) and (100). Because the deduction is standard, we omit the details and just give the results as follows:

$$w_1[u^*](x) = \begin{cases} 0 & x \notin \Omega_1 \\ 0 & x \in \Omega_1 \text{ and } |\nabla u^*(x)| \leq \frac{1}{\tilde{\lambda}} \\ \left(|\nabla u^*(x)| - \frac{1}{\tilde{\lambda}} \right) \frac{\nabla u^*(x)}{|\nabla u^*(x)|} & x \in \Omega_1 \text{ and } |\nabla u^*(x)| > \frac{1}{\tilde{\lambda}} \end{cases}, \quad (101)$$

$$w_3[u^*](x) = \begin{cases} 0 & x \notin \Omega_3 \\ \frac{\nabla u^*(x)}{\tilde{\lambda}+2} & x \in \Omega_3, \end{cases} \quad (102)$$

$$\hat{u}[w^*, g^m](\xi) = \frac{\lambda \hat{S}(\xi) \hat{g}^m(\xi) - i \tilde{\lambda} \xi \cdot \hat{w}^*}{\lambda \hat{S}^2(\xi) + \tilde{\lambda} |\xi|^2}. \quad (103)$$

For subproblem (98), by a simple calculation, we can obtain the Euler-Lagrange equation

$$0 = -\nabla \cdot \left(\frac{w_2}{|w_2|} \bar{p}(x) |w_2|^{\bar{p}(x)-1} \right) + \tilde{\lambda} (w_2 - \nabla u^*) \quad (104)$$

Denote

$$J_{sub2}(w_2) := \int_{\Omega_2} |w_2|^{\bar{p}^m(x)} dx + \frac{\tilde{\lambda}}{2} \|w_2 - \nabla u^*\|_{L^2(\Omega_2)}^2.$$

Taking \tilde{g}_{ij}^m to be a standard finite difference approximation of the right hand side of (104) at $x_{i,j}$ and t_m , we get an Euler-like updating scheme

$$w_{2i,j}^{m+1}[u^*] = w_{2i,j}^m[u^*] - \Delta t \tilde{g}_{i,j}^m. \quad (105)$$

Here we use an adaptive step size scheme. The new value $w_{2i,j}^{m+1}[u^*]$ is accepted for each step in which the cost is improved, $J_{sub2}(w_{2i,j}^{m+1}[u^*]) < J_{sub2}(w_{2i,j}^m[u^*])$, and the step Δt is increased by a factor $\Delta t \rightarrow (1+s)\Delta t$, $s > 0$. For each unsuccessful step where $J_{sub2}(w_{2i,j}^{m+1}[u^*]) \geq J_{sub2}(w_{2i,j}^m[u^*])$, the trial step is not used, and the step size is decreased, $\Delta t \rightarrow (1-s)\Delta t$.

In order to solve optimization problem (94) by solving subproblems from (97) to (100), we need to solve subproblems from (97) to (100) iteratively with many times to obtain an accurate solution. However, as mentioned in [39], we actually only need to solve these subproblems with few iterations. Hence, we may not need to solve subproblem (98) with very high accuracy. That is to say we can run the iterative procedure (105) with few steps.

At last, we state the discrete version of gradient operator and frequency operation. For a function f , the discrete version of ∇f is $(\nabla f)_{i,j} := ((\nabla f)_{i,j}^1, (\nabla f)_{i,j}^2)$ with

$$(\nabla f)_{i,j}^1 = \begin{cases} \frac{f_{i+1,j} - f_{i,j}}{\delta x_1} & 1 \leq i < N \\ 0 & i = N, \end{cases} \quad (\nabla f)_{i,j}^2 = \begin{cases} \frac{f_{i,j+1} - f_{i,j}}{\delta x_2} & 1 \leq j < N \\ 0 & j = N. \end{cases}$$

For simplicity, we assume that $\Omega := [-L, L] \times [-L, L]$ is a square, which yields that $\delta x_1 = \delta x_2 := \delta x = \frac{2L}{N}$. Then from Shannon-Nyquist sampling principle, the maximum frequency from the spatial grids is $[-\Omega_0, \Omega_0]$ with

$$\Omega_0 = \frac{2\pi}{\delta x} = \frac{\pi N}{L}.$$

We can compute the discrete Fourier transform in $[-\Omega_0, \Omega_0] \times [-\Omega_0, \Omega_0]$ with uniform frequency distribution $\{\xi_{m,n} : m, n = 1, 2, \dots, N\}$.

Under these considerations, the iterative scheme for solving the optimization problem for the backward time-space fractional diffusion model can be implemented by the Bregman iterative algorithm with some modifications. For the details, see Algorithm 2.

4. Numerical Examples. In this section, we consider two typical examples. In these two examples, we will compare our results with TV regularizing and Tikhonov regularizing model. Here we first list the two models as follows

$$u_{TV} = \operatorname{argmin}_u \left\{ \|u\|_{TV} + \frac{\lambda}{2} \|\hat{S}\hat{u} - \hat{g}^\delta\|_{L^2(\mathbb{R}^2)}^2 \right\}, \quad (106)$$

$$u_{Tik} = \operatorname{argmin}_u \left\{ \|\nabla u\|_{L^2}^2 + \frac{\lambda}{2} \|\hat{S}\hat{u} - \hat{g}^\delta\|_{L^2(\mathbb{R}^2)}^2 \right\}, \quad (107)$$

where λ is the regularization parameter, g^δ is the measured data with noise. For the TV regularization model, we refer to [37] which described clearly how to solve TV regularization model. For Tikhonov regularization model, it can be solved just by a small modification of algorithm stated in [37]. More explicitly, we just need to change the Euler-Lagrange equation of problem (106) by the Euler-Lagrange equation of problem (107).

In our numerical examples, we choose

$$G_{\tilde{\delta}}(x) = \frac{1}{\tilde{\delta}} \exp\left(-\frac{1}{4\tilde{\delta}^2}|x|^2\right), \quad (108)$$

Algorithm 2 Modified Bregman Iteration Algorithm

Input: $g^\delta, \delta, \lambda, \tilde{\lambda}, \tau, m_{\max}, k_{\max}, \ell_{\max}, G_{\tilde{\delta}}, E(\cdot)$ is some edge detector, $\Delta t, \epsilon, s, \text{tol}$
 set $u^0 = p^0 = 0, g^0 = g^\delta, m = 0$
while $\|\hat{S}\hat{u}^m - \hat{g}^\delta\|_{L^2} \geq \tau\delta$ and $m \leq m_{\max}$ **do**
 $u^{m,0} = u^m$ $\bar{p}^m = 2I - G_{\tilde{\delta}} * E(u^m)$
 for $k = 0$ to k_{\max} **do**
 $w_1^{k+1} \leftarrow w_1[u^{m,k}]$
 while $\ell \leq \ell_{\max}$ and $\|w_2^{\ell+1} - w_2^\ell\|_{L^2} > \text{tol}$ **do**
 $\tilde{g}^\ell = -\nabla \cdot \left(\frac{w_2^\ell}{|w_2^\ell|} \bar{p}^m |w_2^\ell|^{\bar{p}^m - 1} \right) + \tilde{\lambda}(w_2^\ell - \nabla u^{m,k})$
 if $J_{\text{sub}2}(w_2^\ell) > J_{\text{sub}2}(w_2^\ell - \Delta t \tilde{g}^\ell)$ **then**
 $w_2^{\ell+1} \leftarrow w_2^\ell - \Delta t \tilde{g}^\ell, \Delta t \leftarrow (1+s)\Delta t$
 else
 $\Delta t \leftarrow (1-s)\Delta t$
 end if
 end while
 $w_3^{k+1} \leftarrow w_2[u^{m,k}], w^{k+1} \leftarrow w_1^{k+1} + w_2^{k+1} + w_3^{k+1}$
 $u^{m,k+1} \leftarrow u[w^{k+1}, \hat{g}^m]$
 end for
 $u^{m+1} \leftarrow u^{m,k+1}, \hat{g}^{m+1} \leftarrow \hat{g}^m + (\hat{g}^\delta - \hat{S}\hat{u}^{m+1})$
end while
Output: u^{m+1}

as the smoothing kernel $G_{\tilde{\delta}}$ appeared in Algorithm 2. In the following, we take parameters in Algorithm 2 as follows

$$\begin{aligned} \tilde{\delta} &= 0.4, \quad \tau = 1.01, \quad m_{\max} = 500, \quad k_{\max} = 2, \quad \ell_{\max} = 5, \quad \text{tol} = 10^{-6}, \\ s &= 0.1, \quad \Delta t = 0.1, \quad \epsilon = 0.1, \\ E(\cdot) &\text{ to be the Canny edge detection algorithm in the Matlab toolbox.} \end{aligned}$$

For the noise, we take $\delta = 0.0005$ and $\delta = 0.005$ respectively. Because the value of λ can determine the convergence rate of our algorithm, we take different λ for different noise level. If we take λ too big, $\|Su_1 - \hat{g}^\delta\|_{L^2}$ may less than $\tau\delta$ when the first iteration finished. In this case, we may incorporate more noise in our result u_1 . If we take λ too small, the iteration will converge too slow to obtain our final results, e.g. exceed 500 steps. For $\delta = 0.0005$ and $\delta = 0.005$, we take $\lambda = 10^{11}$ and $\lambda = 10^9$ respectively in our numerical experiments.

In order to avoid the error in solving the forward fractional differential equation, we solve (2) to obtain the solution at time T using the Laplace transform

$$\hat{v}(T, \xi) = E_{\alpha,1}(-|\xi|^{2\beta} T^\alpha) \hat{u}(\xi),$$

where the Mittag-Leffler function $E_{\alpha,1}(\cdot)$ is numerically calculated up to desired accuracy by standard algorithm provided by Podlubny [17].

Denote $x = (x_1, x_2) \in \mathbb{R}^2$ and $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$. We generate the final measurement data with noise by

$$g^\delta := v^\delta(T, x) = \mathcal{F}^{-1}(\hat{v}(T, \xi)) + \delta \cdot \text{randn}(x) \cdot \max(\mathcal{F}^{-1}(\hat{v}(T, \xi))),$$

where randn is the pseudo-random number generating from the standard normal distribution. Here we add noise as in [38] where the δ stands for the noise level is $100 \times \delta$, e.g. when $\delta = 0.05$ the noise level is 5%.

In our discretization, we discrete $\Omega = [-10, 10]^2$, the support of $u(x)$, by uniform grids $(x_1(i), x_2(j)) \in [-10, 10]^2$ with $i, j = 1, \dots, 256$.

We use relative error (RelErr) to quantitatively compare our solution with those based on TV regularization and Tikhonov regularization. For given finite dimensional vectors g and its noisy form g^δ representing the image, the above RelErr has the representation

$$\text{RelErr}(g^\delta, g) := \frac{\|g^\delta - g\|_{L^2(\Omega)}}{\|g\|_{L^2(\Omega)}} \times 100\%. \quad (109)$$

Example 1. We consider $u(x) = e^{-|x|^2}$, $T = 1$. In this case, the exact solution has the following form

$$\hat{v}(T, \xi) = \pi E_{\alpha, 1}(-|\xi|^{2\beta} T^\alpha) \hat{u}(\xi) e^{-|\xi|^2}. \quad (110)$$

We take $\alpha = 0.6$, $\beta = 1$ and $T = 1$ to see the difference between the three different models. In Table 1, relative error defined in (109) for three different methods are presented. Because the noise added by random algorithms, we run the three different algorithms 100 times and the data are the averages.

TABLE 1. The values of RelErr of three methods for Example 1

RelErr	TV model	Tikhonov model	Variable TV model
$\sigma = 0.0005$	3.8283%	0.3857%	0.3696%
$\sigma = 0.005$	8.8646%	0.6559%	0.6597%

From Table 1, we could clearly know that TV model's performance is much weaker than the Tikhonov model and our variable TV model's performance is comparable to the Tikhonov model. Because the differences for recovered functions obtained by different methods can not see clearly from the figures of the recovered function, we will not provide the comparison figures for the recovered function and only provide the original data, recovered data with $\delta = 0.0005$ and $\delta = 0.005$ in Figure 3 which show that the recovered data have no visual difference with the original data.

Example 2. Consider a phantom model generated by standard function phantom.m in Matlab with default parameters. We use the gray level (piecewise constant) of this image as the values of $u(x)$, see Figure 4. In this example, we take $\alpha = 0.6$ and $\beta = 0.9$ and $N = 256$. In the following, we provide Table 2 to present the performance of the three different models. As in Example 1, we also run the three different algorithms 100 times and the data in Table 2 are the averages. Table 2 demonstrate that the TV model's performance is better than Tikhonov model when the initial data is a piecewise constant function. Our variable TV model as expected perform comparable to the TV model. Hence, Example 1 and Example 2 reflect that our model can change the value of \bar{p} and the algorithm proposed in Section 3 can solve our variable TV regularization model effectively. Based on same considerations as stated in Example 1, we will not present the three different figures of the recovered function and only present the recovered functions of our variable

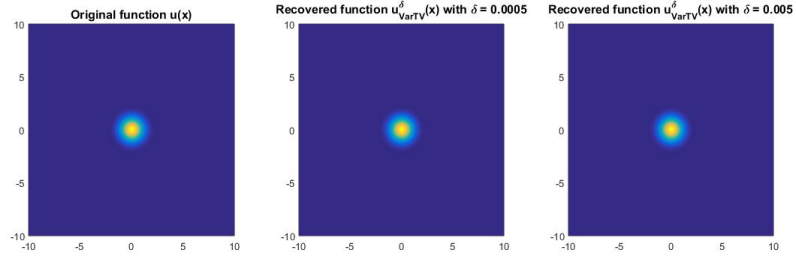


FIGURE 3. Left: Original function; Middle: Recovered function by variable TV model with $\delta = 0.0005$; Right: Recovered function by variable TV model with $\delta = 0.005$ for Example 1.



FIGURE 4. Initial function for Example 2.

TV model in Figure 5 which show that the recovered functions are much similar to the original data.

TABLE 2. The values of RelErr of three methods for Example 2

RelErr	TV model	Tikhonov model	Variable TV model
$\sigma = 0.0005$	13.0053%	13.7772%	13.0666%
$\sigma = 0.005$	22.7222%	25.2101%	22.7810%

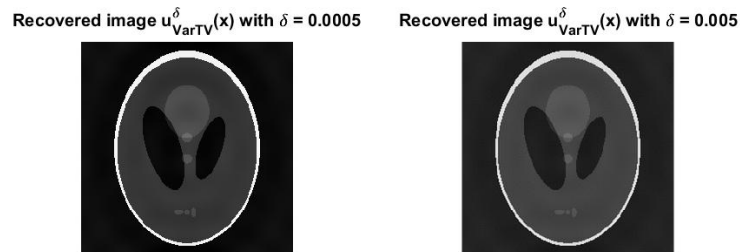


FIGURE 5. Left: Recovered function by variable TV model with $\delta = 0.0005$ for Example 2; Right: Recovered function by the variable TV model with $\delta = 0.005$ for Example 2.

Now, we provide a simple verification of our theoretical results Theorem 3.2. Here, we take $\delta = 0.0005$. From Theorem 3.2, we know that

$$\lambda M = \frac{2}{\delta^2} (\|\nabla u_*\|_{L^2} + |\Omega|) = C$$

for some unknown constant C . We take $\lambda = 10^{11}, \frac{1}{4} \times 10^{11}, \frac{1}{16} \times 10^{11}$ respectively, then run our program by taking $\tau = 1.01$ and using our stop criterion (91) to obtain the iterative step M . For clarity, the values of M , RelErr with different choices of λ are shown in Table 3. From Table 3, we find that $M = 9, 34, 150$ when

TABLE 3. The values of RelErr with different parameters λ of Variable TV model for Example 2

	$\lambda = 10^{11}$	$\lambda = \frac{1}{4} \times 10^{11}$	$\lambda = \frac{1}{16} \times 10^{11}$
$\sigma = 0.0005$	M = 9	M = 34	M = 150
	RelErr = 13.0792%	RelErr = 13.0342%	RelErr = 13.0275%

$\lambda = 10^{11}, \frac{1}{4} \times 10^{11}, \frac{1}{16} \times 10^{11}$ respectively, which indicate that the numerical results are almost the same as predicted by the theoretical results. Although we choose three different values of λ and obtain three different maximum iteration numbers, the values of RelErr are almost the same, which demonstrate that the estimated results are not sensitive to different values of parameter λ . The results shown in Table 3 are obtained by running our algorithm once, and each time the results will be a little different for the noise is added randomly.

After theoretical justifications, we want to clarify an interesting phenomena which reveals some essential different properties of the inverse problems for integer-order differential equations and fractional-order differential equations.

Discontinuity for normal and anomalous diffusion. In this part, we also use the phantom model generated by standard function phantom.m in Matlab with default parameters (same as in Example 2) as our initial data then take $T = 1$, $\sigma = 0.0005$, $\beta = 1$ and the time derivative $\alpha = 0.5, \dots, 1$. In order to provide a clear explanation, we take 100 points between $[0.5, 1]$ for α . Then we use our variable TV regularizing model to recover the true initial data and plot the RelErr value for each α in Figure 6. Because we used the same model, the degree of ill-posedness intuitively can be represented by the RelErr value. Small RelErr value indicate that our model can provide a good result, hence, the degree of ill-posedness is weak. In contradict, large RelErr value indicate that the degree of ill-posedness is strong. From Figure 6, we clearly find that even for $\alpha = 0.99$ the degree of ill-posedness is much weaker than the integer-order equation. This implies that for $\alpha < 1$ the degree of ill-posedness varies continuously, however, for $\alpha = 1$ the degree of ill-posedness is much higher than any value of $\alpha < 1$. The degree of ill-posedness may not change continuously at the point 1. This observation may be explained by the properties of Mittag-Leffler function $E_{\alpha,1}(z)$. For $\alpha = 1$, it is an exponential function, however, for any value $\alpha < 1$ the Mittag-Leffler function behaves like polynomial functions for large z (Theorem 1.3 in [16]). This property also be observed in [42] which propose a fractional extension of instantaneous frequency attribute to detect thin layers of sandstone formations. They use fractional order of 0.99 and illustrate only 0.01 smaller than integer-order 1 will bring very different results.

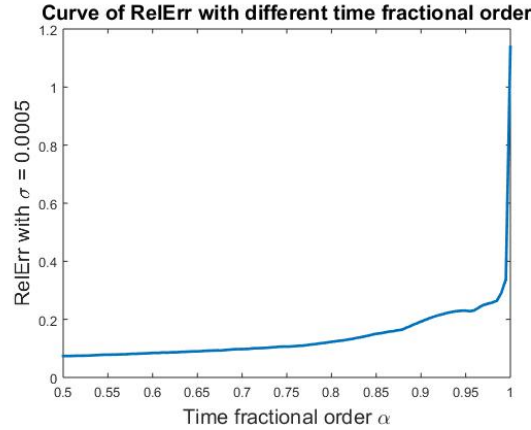


FIGURE 6. The curve of the relative error of the recovered data for different values of parameter α .

From the above two typical examples, it is obviously that our algorithm behaves like L^2 based Tikhonov regularization model when the initial function is smooth and behaves like TV regularization model when the initial function is piecewise constant. Hence, our model has more flexibility compared with Tikhonov regularization model and TV regularization model.

5. Conclusion. In this work, backward diffusion problem of a time-space fractional diffusion equation are discussed under the regularization framework. Variable total variation (TV) regularizing term are employed to overcome the deficiencies of Tikhonov and TV regularization methods. The existence, uniqueness and stability of the solution for the minimization problem with variable TV regularizing term are obtained by using the fractional operator semigroup theory. Meanwhile, a new modified Bregman iterative algorithm are proposed to approximate the minimizer. Concerning this new algorithm, both the convergence and the strategy of choosing parameters are illustrated theoretically. In the end, numerical tests are provided, which demonstrate the good performance and flexibility of the proposed method.

Finally, we point out some future directions along the line of this work. The first is concerned with the inversion of initial data and fractional-order simultaneously. In some problems, we can estimate the fractional-order firstly, then solve the backward problem. However, sometimes, we need to estimate the initial data and the fractional-order simultaneously, which require to construct more sophisticated algorithms, e.g., alternate iterative type algorithms for fractional-order and initial data. The second is concerned with the generalization of the isotropic fractional diffusion equations employed here to anisotropic equations. For anisotropic equations, Fourier transform may not be used directly and new techniques need to be developed.

Acknowledgments. The author thanks referees for many valuable comments and suggestions. This work was partially supported by the NSFC under grant Nos. 11501439, 11771347, 91730306, 41390454, 61372017, and partially supported by the Major projects of the NSFC under grant Nos. 41390450 and 41390454, and partially

supported by the postdoctoral science foundation project of China under grant no. 2017T100733.

REFERENCES

- [1] B. Baeumer, M. Kovács and H. Sankaranarayanan, Higher order Grünwald approximations of fractional derivatives and fractional powers of operators, *T. Am. Math. Soc.*, **367** (2015), 813–834.
- [2] B. Baeumer, M. Kovács and M. M. Meerschaert, R. Schilling and P. Straka, Reflected spectrally negative stable processes and their governing equations, *T. Am. Math. Soc.*, **368** (2016), 227–248.
- [3] B. Baeumer, M. M. Meerschaert and E. Nane, SpaceCtime duality for fractional diffusion, *J. Appl. Probab.*, **46** (2009), 1100–1115.
- [4] B. Baeumer, S. Kurita and M. M. Meerschaert, Inhomogeneous fractional diffusion equations, *Fract. Calc. Appl. Ana.*, **8** (2005), 371–386.
- [5] B. Jin, R. Lazarov, D. Sheen and Z. Zhou, Error estimates for approximations of distributed order time fractional diffusion with nonsmooth data, *Fract. Calc. Appl. Ana.*, **19** (2016), 69–93.
- [6] B. Jin and W. Rundell, A tutorial on inverse problems for anomalous diffusion processes, *Inverse Probl.*, **31** (2015), 035003.
- [7] C. M. Carracedo and M. S. Alix, *The Theory of Fractional Powers of Operators*, North-Holland Publishing Co., Amsterdam, 2001.
- [8] E. G. Bajlekova, *Fractional Evolution Equations in Banach Spaces*, Ph.D thesis, Eindhoven University of Technology in Eindhoven, 2001.
- [9] E. M. Bollt, R. Chartrand, S. Esedoglu, P. Schultz and K. R. Vixie, Graduated adaptive image denoising: local compromise between total variation and isotropic diffusion, *Adv. Comput. Math.*, **31** (2009), 61–85.
- [10] E. M. Wright, The generalized Bessel function of order greater than one, *Q. J. Math.*, **1** (1940), 36–48.
- [11] F. Li, Z. Li and L. Pi, Variable exponent functionals in image restoration, *Appl. Math. Comput.*, **216** (2010), 870–882.
- [12] F. Mainardi, Fractional relaxation-oscillation and fractional diffusion-wave phenomena, *Chaos Soliton Frac.*, **7** (1996), 1461–1477.
- [13] F. Mainardi and M. Tomirotti, On a special function arising in the time fractional diffusion-wave equation, *Transform Methods and Special Functions, Sofia'94, Proceedings of International Workshop*, (1994), 171–183.
- [14] G. H. Zheng and T. Wei, Two regularization methods for solving a Riesz-Feller space-fractional backward diffusion problem, *Inverse Probl.*, **26** (2010), 115017.
- [15] G. M. Zaslavsky, Chaos, fractional kinetics, and anomalous transport, *Phys. Rep.*, **371** (2002), 461–580.
- [16] I. Podlubny, *Fractional Differential Equations: An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of Their Solution and Some of Their Applications*, Academic Press, Inc., San Diego, CA, 1999.
- [17] I. Podlubny, Matlab program for computing Mittag-Leffler function $E_{\alpha,\beta}(\cdot)$, <http://www.mathworks.com/matlabcentral/fileexchange/8738-mittag-leffler-function>.
- [18] J. Cheng, J. Nakagawa, M. Yamamoto and T. Yamazaki, Uniqueness in an inverse problem for a one-dimensional fractional diffusion equation, *Inverse Probl.*, **25** (2009), 115002.
- [19] J. Jia, J. Peng and K. Li, Well-posedness of abstract distributed-order fractional-order fractional diffusion equations, *Commun. Pur. Appl. Anal.*, **13** (2014), 605–621.
- [20] J. Liu and M. Yamamoto, A backward problem for the time-fractional diffusion equation, *Appl. Anal.*, **89** (2010), 1769–1788.
- [21] J. Peng and K. Li, A novel characteristic of solution operator for the fractional abstract Cauchy problem, *J. Math. Anal. Appl.*, **385** (2012), 786–796.
- [22] J. Tiirola, Image decompositions using spaces of variable smoothness and integrability, *SIAM J. Imaging Sci.*, **7** (2014), 1558–1587.
- [23] K. Sakamoto and M. Yamamoto, Initial value/boundary value problems for fractional diffusion-wave equations and applications to some inverse problems, *J. Math. Anal. Appl.*, **382** (2011), 426–447.

- [24] L. C. Evans, *Partial Differential Equations*, American Mathematical Society, Providence, 1998.
- [25] L. Wang and J. Liu, Total variation regularization for a backward time-fractional diffusion problem, *Inverse Probl.*, **29** (2013), 115013.
- [26] M. Haase, *The Functional Calculus for Sectorial Operators*, Birkhäuser Verlag, Basel, 2006.
- [27] M. Li, C. Chen and F. B. Li, On fractional powers of generators of fractional resolvent families, *J. Funct. Anal.*, **259** (2010), 2702–2726.
- [28] M. M. Meerschaert and A. Sikorskii, *Stochastic Models for Fractional Calculus*, Walter de Gruyter & Co, Berlin/Boston, 2011.
- [29] M. X. Wang, *Operator Semigroup and Evolutionary Equations (in Chinese)*, Science Press, China, 2006.
- [30] N. Jacob, *Pseudo Differential Operators and Markov Processes: Fourier Analysis and Semigroups*, Imperial College Press, London, 2001.
- [31] P. Blomgren, T. F. Chan, P. Mulet, L. Vese and W. L. Wan, *Variational PDE Models and Methods for Image Processing*, Chapman & Hall/CRC, Boca Raton, 2000.
- [32] P. Harjulehto, P. Hästö, V. Latvala and O. Toivanen, Critical variable exponent functionals in image restoration, *Appl. Math. Lett.*, **26** (2013), 56–60.
- [33] P. Harjulehto, P. Hästö and V. Latvala, Minimizers of the variable exponent, non-uniformly convex Dirichlet energy, *J. Math. Pure. Appl.*, **89** (2008), 174–197.
- [34] R. Acar and C. R. Vogel, Analysis of bounded variation penalty methods for ill-posed problems, *Inverse Probl.*, **10** (1994), 1217–1229.
- [35] R. Klages, G. Radons and I. M. Sokolov, *Anomalous Transport: Foundations and Applications*, Wiley-VCH Verlag GmbH & Co. KGaA, Darmstadt, 2008.
- [36] R. Metzler and J. Klafter, The random walk's guide to anomalous diffusion: a fractional dynamics approach, *Phys. Rep.*, **339** (2000), 1–77.
- [37] S. Osher, M. Burger, D. Goldfarb, J. Xu and W. Yin, An iterative regularization method for total variation-based image restoration, *Multiscale Model. Sim.*, **4** (2005), 460–489.
- [38] T. Bui-Thanh and O. Ghattas, A scalable algorithm for MAP estimators in Bayesian inverse problems with Besov priors, *Inverse Probl. Imag.*, **9** (2015), 27–53.
- [39] T. Goldstein and S. Osher, The split Bregman method for L1-regularized problems, *SIAM J. Imaging Sci.*, **2** (2009), 323–343.
- [40] W. Arendt, C. J. Batty, M. Hieber and F. Neubrander, *Vector-Valued Laplace Transforms and Cauchy Problems*, Birkhäuser Verlag, Basel, 2001.
- [41] Y. Zhang and X. Xu, Inverse source problem for a fractional diffusion equation, *Inverse Probl.*, **27** (2011), 035010.
- [42] Z. Wang, J. Gao, Q. Zhou, K. Li and J. Peng, A new extension of seismic instantaneous frequency using a fractional time derivative, *J. Appl. Geophys.*, **98** (2013), 176–181.

Received March 2016; revised December 2017.

E-mail address: jjx323@xjtu.edu.cn

E-mail address: jgpeng@xjtu.edu.cn

E-mail address: jhgao@xjtu.edu.cn

E-mail address: liyujiao.323@stu.xjtu.edu.cn