ANALYSIS OF A NONLOCAL-IN-TIME PARABOLIC EQUATION

QIANG DU, JIANG YANG AND ZHI ZHOU*

Department of Applied Physics and Applied Mathematics
Columbia University
New York, NY, 10027, USA

(Communicated by Yuan Lou)

Abstract. In this paper, we consider an initial boundary value problem for nonlocal-in-time parabolic equations involving a nonlocal in time derivative. We first show the uniqueness and existence of the weak solution of the nonlocal-in-time parabolic equation, and also the spatial smoothing properties. Moreover, we develop a new framework to study the local limit of the nonlocal model as the horizon parameter δ approaches 0. Exploiting the spatial smoothing properties, we develop a semi-discrete scheme using standard Galerkin finite element method for the spatial discretization, and derive error estimates dependent on data smoothness. Finally, extensive numerical results are presented to illustrate our theoretical findings.

1. Introduction.

1.1. Mathematical model. The aim of this paper is to study the following initial boundary value problem associated with a nonlocal-in-time parabolic equation (NPE) for $u = u(x, t)$:

$$
\mathcal{G}_\delta u(x, t) - \Delta u(x, t) = f(x, t), \quad \text{in } \Omega \times [0, T],
$$

$$
u(x, t) = 0, \quad \text{in } \partial\Omega \times [0, T],
$$

$$
u(x, t) = g(x, t), \quad \text{in } \Omega \times [-\delta, 0),
$$

where $\Omega$ is a bounded convex polygonal domain in $\mathbb{R}^d$ ($d = 1, 2, 3$) with a boundary $\partial \Omega$, $g$ and $f$ are given functions and $T > 0$ is a positive terminal time. Here, associated with a given nonnegative, symmetric kernel function $\rho_\delta(s) = \rho_\delta(|s|)$, the nonlocal operator $\mathcal{G}_\delta$ is defined by

$$
\mathcal{G}_\delta v(t) = \lim_{\epsilon \to 0} \int_{-\epsilon}^{\epsilon} (v(t) - v(t - s)) s \rho_\delta(s) \, ds.
$$

whenever the limit exists in $L^2(0, T)$ for any function $v \in L^2(-\delta, T)$. Throughout the paper, we assume that $\rho_\delta$ has a compact support contained in $[-\delta, \delta]$, and a

2010 Mathematics Subject Classification. Primary: 45K05; Secondary: 35E15, 65M60.

Key words and phrases. Nonlocal-in-time parabolic equation, nonlocal initial value problem, well-posedness, smoothing property, localization, Galerkin finite element method, error estimates.

This research is supported in part by the AFOSR MURI center for Material Failure Prediction through peridynamics and the ARO MURI Grant W911NF-15-1-0562.

* Corresponding author: Zhi Zhou (zhizhou0125@gmail.com).
normalized second moment, i.e.,
\[ \int_{0}^{\delta} s^2 \rho_\delta(s) \, ds = \frac{1}{2} \int_{-\delta}^{\delta} s^2 \rho_\delta(s) \, ds = 1. \]

In case \( s \rho_\delta(s) \notin L^1(0, \delta) \), we assume further that the possible singularity of \( s \rho_\delta(s) \) appears only at the origin, to ensure that the definition (1.2) makes sense. The positive parameter \( \delta \) appearing in (1.2) represents the range of nonlocal interactions or memory spans, and is called the nonlocal horizon parameter. For suitably defined kernels, as \( \delta \to 0 \), nonlocal or memory effect diminishes, so that the zero-horizon limit of the nonlocal operator \( G_\delta u \) corresponds to the standard first order derivative \( \frac{d}{dt} u \), and hence the model (1.1) becomes the standard diffusion model in the local limit (see Section 4 for details). Meanwhile, in the extreme case that \( \delta = \infty \), the initial data \( v(t) := v(0) \) for all \( t \in (-\infty, 0) \), and the kernel function is of fractional type without normalization, i.e., \( \rho_\infty(s) = \frac{\alpha}{\Gamma(1+\alpha)} s^{-\alpha-2} \) with \( \alpha \in (0, 1) \), the nonlocal operator reproduces the Caputo fractional derivative for \( t > 0 \) (e.g., [18, p. 91] and [2]) provided that \( v \in C(-\infty, T] \) and \( v \) is piece-wise differentiable with \( v' \in L^1(0, T) \cap C(0, T] \)

\[
\Gamma(1-\alpha) G_\infty v(t) = \alpha \lim_{\epsilon \to 0} \int_{-\infty}^{t-\epsilon} \frac{v(t) - v(s)}{(t-s)^{\alpha+1}} \, ds = \lim_{\epsilon \to 0} \left( -(t-s)^{-\alpha} v(s) \right|_{-\infty}^{t-\epsilon} + \alpha \int_{-\infty}^{t-\epsilon} \frac{v(t)}{(t-s)^{\alpha+1}} \, ds + \int_{0}^{t-\epsilon} v'(s)(t-s)^{-\alpha} \, ds \right)
\]

\[ = \int_{0}^{t} v'(s)(t-s)^{-\alpha} \, ds = \Gamma(1-\alpha) \partial_t^\alpha v(t) \]

Consequently, the nonlocal parabolic model (1.1), with initial data \( g(x, t) := g(x) \) for all \( t < 0 \), becomes the time-fractional diffusion model

\[
\partial_t^\alpha u(x, t) - \Delta u(x, t) = f(x, t), \quad \text{in } \Omega \quad T > t > 0, \\
u(x, t) = 0, \quad \text{on } \partial \Omega \quad T > t > 0, \\
u(x, 0) = g(x), \quad \text{in } \Omega. 
\]

We refer to [2] for more details. The model (1.4) has been used to describe sub-diffusion dynamics [24]. A comprehensive survey on fractional order differential equations arising in many applications such as dynamical systems in control theory, electrical circuits with fractance, generalized voltage divider, viscoelasticity, electrochemistry, and model of neurons in biology, can be found in [4] [11] [12] [24].

In comparison with the classical diffusion (the \( \delta \to 0 \) local limit) and fractional diffusion equations (the \( \delta = \infty \) limit), the nonlocal-in-time parabolic equation (1.1) provides an interesting intermediate model to study the ‘finite history dependence’ with a given \( \delta \in (0, \infty) \).

1.2. Relevant studies and our contributions. Our work is motivated by recent interests in nonlocal models. For instance, the nonlocal peridynamics (PD) theory, formulated in integro-differential equations, has attracted much attention as a possible alternative to classical partial differential equation formulation of solid mechanics [20] [8] [19] [27] [28]. The nonlocality in PD has been largely described by spatial nonlocal interactions but it is natural, from a modeling perspective, to consider nonlocal or memory effect as well. The theoretical study on (1.1) involving the nonlocal-in-time operator \( G_\delta \) remains limited, although there have been some
works concerning the related nonlocal gradient operator in $\mathbb{R}^d$ in the form
\[
\mathcal{G}_\delta(u)(x) = \int_{B_\delta(x)} (u(y) - u(x)) \frac{y - x}{|y - x|^2} \rho_\delta(|y - x|) \, dy,
\] (1.5)
where the nonnegative radial function $\rho_\delta = \rho_\delta(|x|) \in L^1(\mathbb{R}^d)$ is compactly supported in the ball $B_\delta(0)$ with the horizon parameter $\delta$. In Section 2, similar to \cite{7, 22}, we present some basic properties of the nonlocal calculus related to $\mathcal{G}_\delta$ and its adjoint operator, and discuss their localization theory in $L^2$-sense.

To the best of our knowledge, this is the first study of the nonlocal-in-time parabolic model in the form of (1.1). The focus of this work is to establish the well-posedness and spatial smoothing properties associated with (1.1). As a direct application of the regularity results, we provide some preliminary discussion on the semi-discrete Galerkin finite element approximation. We are especially interested in the case of nonsmooth data, which is a typical case for related inverse problems, see, e.g., \cite{17, 31} for its local parabolic counterpart. Specifically, the main results are summarized below.

First, in Section 3 using Banach fixed point theorem, in Theorem 3.1, we establish the well-posedness of the initial value problem: to find $u(t)$ satisfying
\[
\mathcal{G}_\delta u(t) + \lambda u(t) = f(t), \quad t \in (0, T],
\]
\[
u(t) = g(t), \quad t \in (-\delta, 0),
\] (1.6)
after almost everywhere, with $\lambda \geq \lambda_0 > 0$, and $f, g$ are in some suitable function spaces. As an example, if $g \in L^\infty(-\delta, 0)$ and $f \in L^\infty(0, T)$, then the problem (1.6) has a unique solution, in the function space $V_\delta$ defined by (2.4), such that (Lemma 3.1)
\[
\|u\|_{L^\infty(0,T)} \leq c(\|g\|_{L^\infty(-\delta,0)} + \lambda^{-1}\|f\|_{L^\infty(0,T)}).
\]
A similar argument yields an estimate when $f \in L^2(0, T)$. The estimate for the short-time asymptotic behavior of the solution is given by Theorem 3.2 and Remark 3.5. Furthermore, exploiting the localization theory established in Section 2, a new framework is developed via a duality argument to show that the solution of the nonlocal problem indeed converges to that of its local counterpart \cite{3, 35} in the sense of $L^2(0, T)$; c.f. Theorem 3.3 and Corollary 3.2. All these results can be easily extended to the case of infinite interval by letting $T \to \infty$. Although our discussion focuses on positive $\lambda$, analogous estimates can be derived for $\lambda = 0$ in a finite interval, i.e., $T < \infty$, using the same idea (cf. Corollary 3.1 and Remark 3.6). However, without the decay property, the estimate for $\lambda = 0$ is more delicate in case of an infinite interval. A reasonable estimate in $L^\infty$-norm, provided that $g \in L^\infty(-\delta, 0)$ and $f \in L^p(0, \infty)$ (cf. Remark 3.4), requires more studies on Duhamel’s principle, which is beyond the scope of the paper.

The extension to the initial boundary value problem (1.1) is made next in Section 4 using the spectrum decomposition based on the results for (1.6) established in Section 3. We show in Theorem 4.1 that the problem (1.1) with a variables-separable initial data $g(x, t) \in L^\infty(-\delta, 0; H^r(\Omega))$ with $r \geq 0$ and trivial source data ($f \equiv 0$) has a unique weak solution which satisfies the a priori estimate
\[
\|u\|_{H^r(\Omega)} \leq c\tau^{-(\frac{1}{r}-\frac{1}{2})} \|g\|_{L^\infty(-\delta,0;H^r(\Omega))}, \quad \text{with } l \in [r, r+2].
\] (1.7)
Moreover, the above estimate can be slightly improved in case of non-integrable fractional kernels, i.e. $\rho_\delta = (\alpha + 1)s^{\alpha-2}\delta^{-\alpha-1}$ with $-1 < \alpha < 0$. In the latter case,
it holds for $r \geq 0$ that (cf. Theorem 4.3)
\[
\|u(t)\|_{H^r(\Omega)} \leq c\delta^{-\alpha-1}(l-r)/2t^{\alpha(l-r)/2}\|g\|_{L^\infty(-\delta,0;H^r(\Omega))}, \quad l \in [r,r+2].
\] (1.8)

This estimate (1.8) is very similar to the short time behavior of the fractional diffusion; cf. [25, Theorem 2.1], while the appearance of $\delta^{-\alpha-1}$ is due to the normalization of the kernel. Furthermore, Theorem 4.4 shows that the second-order regularity pick-up still holds even if the variables-separable initial data is replaced by more general data, for integrable kernels, i.e., $sp_\delta(s) \in L^1(0, \delta)$, and the estimate is uniform in $t$, but with a $\delta$-dependent constant. Meanwhile, we obtain the well-posedness and smoothing property for the inhomogeneous problem, by allowing less regular source term, cf. Theorem 4.2.

While the model (1.1) offers restricted spatial smoothing at a short time, we observe that, by repeating the argument, $u = u(t) \in H^{s+2k}(\Omega)$ for all $t > (k-1)\delta$, cf. Remark 4.1. This interesting and $\delta$-dependent property is in sharp contrast with well-known facts on standard local parabolic equations [29] and some fractional diffusion models [25, 14], since the solution of standard diffusion model has an infinite smoothing property while the subdiffusion has a limited smoothing property with order 2 for any $t > 0$. This reflects the distinct feature of the nonlocal model (1.1) that it has the same smoothing property as the one of anomalous subdiffusion at a short time, but exhibits the behavior of standard diffusion when $t$ approaches infinity. The limited smoothing is also present for the semi-discrete Galerkin finite element approximation; see Section 6. In addition, we also study the localization of the nonlocal model (1.1) in the sense of $L^2(0,T;L^2(\Omega))$, with nonsmooth initial data $g \in L^\infty(-\delta,0;L^2(\Omega))$ and weak source data $f \in L^2(0,T;H^{-1}(\Omega))$, which is fully confirmed by the numerical experiments.

As a direct application of the smoothing property, we then analyze in Section 5 a spatially semi-discrete scheme for the nonlocal model (1.1) and provide some preliminary results. In particular, we develop the semi-discrete finite element approximation using continuous piecewise linear functions, and derive the error estimates in terms of the data regularity. Numerical tests on various one-dimensional examples, including both regular and weak data, are given in Section 6 which illustrate and complement our theoretical study.

Throughout the paper, we use the notation $c$ and $C$, with or without a subscript, to denote a generic positive constant, which may change at different occurrences, but is always independent of the horizon parameter $\delta$, the mesh size $h$, the solution $u$, and the data $g$ and $f$.

2. Preliminaries. In this section, we collect useful facts on the nonlocal calculus and the localization theory of the nonlocal operator $G_\delta$ given by (1.2). Some similar results for the operator $G_\delta$ given in (1.5) have been studied in [22, 23]. For completeness, we recall some useful results and outline their proofs for the simple one dimensional scalar version here.

2.1. Nonlocal calculus. Nonlocal vector calculus has been first introduced in [7] as a framework to analyze nonlocal models. It involves some basic nonlocal operators and associated integral identities similar to those in classical, local calculus. The first result needed for our study is a formula of nonlocal integration by parts. The proof is identical to that of [23, Theorem 1.4] and hence omitted here.
Lemma 2.1. Suppose that \( u \in L^2(-\delta, T) \), and \( \varphi \in C_0^\infty(0, T) \) with zero extension out of the interval \((0, T)\). Further, we suppose that
\[
\int_0^T \int_{-\delta}^t |u(t) - u(y)|(t-y)\rho_\delta(t-y)\,dy\,dt < \infty. \tag{2.1}
\]
Then it holds that
\[
\int_0^T \mathcal{G}_\delta u(t)\varphi(t)\,dt = -\int_0^T u(t)(\mathcal{G}_\delta \varphi)(t)\,dt
\]
with
\[
\mathcal{G}_\delta \varphi(t) = \int_{t+\delta}^{t-\delta} (\varphi(y) - \varphi(t))(y-t)\rho_\delta(y-t)\,dy. \tag{2.2}
\]

The next lemma gives an upper bound for the nonlocal operator for smooth functions in Sobolev spaces, see similar discussions in \([5, 23, 22]\).

Lemma 2.2. Let the operator \( \mathcal{G}_\delta \) be defined by \((1.2)\) with \( s^2\rho_\delta(s) \in L^1(0, \delta) \). Then it holds that
\[
\|\mathcal{G}_\delta(\varphi)\|_{L^2(0, T)} \leq C\|\varphi'\|_{L^2(-\delta, T)}, \quad \varphi \in H^1(-\delta, T).
\]

Proof. It follows the definition of \( \mathcal{G}_\delta \), Hölder’s inequality and \([5, \text{Theorem 1}]\) that
\[
\|\mathcal{G}_\delta(\varphi)\|_{L^2(0, T)}^2 \\
\leq c \int_{-\delta}^T \left( \int_{-\delta}^t |\varphi(t) - \varphi(y)||y-t|\rho_\delta(|y-t|)\,dy \right)^2\,dt
\]
\[
\leq c \int_{-\delta}^T \left( \int_{-\delta}^t |\varphi(y) - \varphi(t)|\rho_\delta(|y-t|)\,dy \right) \left( \int_{-\delta}^t |y-t|^2\rho_\delta(|y-t|)\,dy \right)\,dt
\]
\[
\leq c \int_{-\delta}^T \left( \int_{-\delta}^t |\varphi(y) - \varphi(t)|^2\rho_\delta(|y-t|)\,dy \right) \,dx \leq c\|\varphi'\|_{L^2(-\delta, T)}^2.
\]
This completes the proof of first inequality, and the second one follows analogously. \(\square\)

A similar argument for the nonlocal operator \( \mathcal{G}_\delta^* \) leads to the following.

Corollary 2.1. Let the operator \( \mathcal{G}_\delta^* \) be defined by \((2.2)\) with \( s^2\rho_\delta(s) \in L^1(0, \delta) \). Then for \( \varphi \in H^1(0, T+\delta) \), it holds that
\[
\|\mathcal{G}_\delta^*(\varphi)\|_{L^2(0, T+\delta)} \leq C\|\varphi'\|_{L^2(0, T+\delta)}.
\]

As a result of previous discussion, a distributional nonlocal operator \( \tilde{\mathcal{G}}_\delta \) may be defined on \( L^2(-\delta, T) \) by
\[
\langle \tilde{\mathcal{G}}_\delta(u), \varphi \rangle = -\int_{-\delta}^T u(t)(\mathcal{G}_\delta^* \varphi)(t)\,dt, \quad \text{for all } \varphi \in C_0^\infty(0, T). \tag{2.3}
\]
This definition is a nonlocal analogue of the notion of local distributional derivative, and \( \tilde{\mathcal{G}}_\delta \) indeed maps onto a distribution because
\[
|\langle \tilde{\mathcal{G}}_\delta(u), \varphi \rangle| \leq C\|\varphi\|_{H^1(0, T+\delta)}\|u\|_{L^2(-\delta, T)}.
\]
A consequence of Lemma 2.1 is that the function \( \mathcal{G}_\delta(u) \in L^2(0, T) \) is exactly the distribution \( \tilde{\mathcal{G}}_\delta(u) \) for all \( u \in L^2(-\delta, T) \) satisfying \((2.1)\), as noted in \([22]\). Next, we
may show that \( \overline{G}_\delta(u)(t) \in L^2(-\delta, 0) \) implies \( G_\delta(u)(t) = \overline{G}_\delta(u)(t) \) a.e. in \((0, T)\). To this end, let us now define a related nonlocal function space

\[
V_\delta = \{ u \in L^2(-\delta, T) : G_\delta u \in L^2(0, T) \},
\]
equipped with the norm

\[
\|u\|_{V_\delta}^2 = \|u\|_{L^2(-\delta, T)}^2 + \|G_\delta u\|_{L^2(0, T)}^2.
\]

The next lemma shows the density of smooth functions in \( V_\delta \), a result missing from earlier works \([22]\).

**Lemma 2.3.** Suppose that \( u \in V_\delta \). Then there exists a sequence of smooth functions \( \{u_n\} \subset C^\infty(-\delta, T) \) such that \( u_n \rightarrow u \) in \( L^2(-\delta, T) \) and \( G_\delta u_n = G_\delta u \rightarrow G_\delta u \) in \( L^2(0, T) \).

**Proof.** Let \( J_\epsilon \) be the standard mollifier \([1]\) pp. 36. For any \( a, b \) s.t. \( 0 < a < b < T, \epsilon < \min(a, T - b) \) and \( \varphi \in C^\infty(a, b) \) we obtain by Lemma 2.1

\[
\int_0^T G_\delta(J_\epsilon * u)(t) \varphi(t) \, dt = - \int_{-\delta}^T (J_\epsilon * u)(t) G_\delta \varphi(t) \, dt = - \int_{-\delta}^T u(t)(J_\epsilon * (G_\delta \varphi))(t) \, dt.
\]

Since \( \varphi \) is smooth and compactly supported in \((a, b)\) and \( \epsilon < \min(a, T - b) \), then we have \( G_\delta(J_\epsilon * \varphi) = J_\epsilon * (G_\delta \varphi) \) for \( t \in (-\delta, T) \). Next, using the fact that \( J_\epsilon * \varphi \in C_c^\infty(0, T) \) and the definition of \( G_\delta \) in (2.3), we have

\[
- \int_{-\delta}^T u(t)(J_\epsilon * \varphi)(t) \, dt = \int_{-\delta}^T (J_\epsilon * G_\delta(u))(t)(J_\epsilon * \varphi)(t) \, dt = \int_{-\delta}^T (J_\epsilon * \tilde{G}_\delta(u))(t)(J_\epsilon * \varphi)(t) \, dt.
\]

Therefore, \( G_\delta(J_\epsilon * u) = J_\epsilon * (\tilde{G}_\delta u) \) in \((a, b)\). Consequently, using the properties of mollification (cf. \([1]\) Theorem 2.29) we derive that

\[
\lim_{\epsilon \to 0} (\|J_\epsilon * u - u\|_{L^2(-\delta, \delta)} + \|G_\delta(J_\epsilon * u) - G_\delta u\|_{L^2(0, \delta)}) = 0.
\]

Without loss of generality, we assume \( T > \delta \). Let \( a \in (0, \delta) \) and consider the finite covering of \((-\delta, T)\) such that \( U_1 = (-\delta, a) \) and \( U_2 = (a, 2T) \). Let \( \{\eta_k\}_{k=1}^\infty \) be a \( C^\infty \) unity of partition subordinate to the above covering. Using the shifted point \( x^\sigma = x + \sigma \) with small \( \sigma \in (0, (T - a - \delta)/2) \) and shifted function \( T_\sigma f(x) = f(x^\sigma) \), we have for \( \epsilon < \sigma/2 \) and \( u_1 = \eta_1 u \) with zero extension that

\[
\|J_\epsilon * (T_\sigma u_1) - u_1\|_{L^2(-\delta, T)} + \|G_\delta(J_\epsilon * (T_\sigma u_1)) - \tilde{G}_\delta(u_1)\|_{L^2(0, T)} \leq \|J_\epsilon * (T_\sigma u_1) - (T_\sigma u_1)\|_{L^2(-\delta, \sigma, \sigma/2)} + \|G_\delta(J_\epsilon * (T_\sigma u_1)) - \tilde{G}_\delta(u_1)\|_{L^2(0, T)} + (\|T_\sigma u_1\|_{L^2(-\delta, T)} + \|\tilde{G}_\delta(T_\sigma u_1) - \tilde{G}_\delta(u_1)\|_{L^2(0, T)}) =: I_1(\sigma, \epsilon) + I_2(\sigma).
\]

Meanwhile, we observe that for fixed \( \sigma > 0, \)

\[
\lim_{\epsilon \to 0} I_1(\sigma, \epsilon) = \lim_{\epsilon \to 0} (\|J_\epsilon * u_1 - u_1\|_{L^2(-\delta + \sigma, \sigma, 3\sigma/2)} + \|G_\delta(J_\epsilon * (u_1)) - \tilde{G}_\delta(u_1)\|_{L^2(0, T)}) = 0,
\]

using the previous argument, and \( \lim_{\sigma \to 0} I_2(\sigma) = 0 \) due to the continuity of translation in \( L^2 \)-norm. Therefore, for arbitrary \( \zeta > 0 \), the triangle inequality argument
implies the existence of \((\sigma_1, \epsilon_1)\) such that \(\phi_1 = J_{\epsilon_1} * (T_{\sigma_1}(u_1)) \in C^\infty[-\delta, T]\) and
\[
\|\phi_1 - u_1\|_{L^2(-\delta, T)} + \|G_\delta \phi_1 - \tilde{G}_\delta(u_1)\|_{L^2(0,T)} \leq \zeta/2.
\]

Further, for \(U_2\), we use the shifted operation \(T_{-\sigma}\) with a small \(\sigma \in (0, a/2)\). Then for \(\epsilon < \sigma/2\) and \(u_2 = \eta_2 u\) with zero extension, we have
\[
\|J_\epsilon \ast (T_{-\sigma}u_2) - u_2\|_{L^2(-\delta, T)} + \|G_\delta(J_\epsilon \ast (T_{-\sigma}u_2)) - \tilde{G}_\delta(u_2)\|_{L^2(0, T)} \\
\leq (\|J_\epsilon \ast (T_{-\sigma}u_2) - (T_{-\sigma}u_2)\|_{L^2(0, T)}) + \|G_\delta(J_\epsilon \ast (T_{-\sigma}u_2)) - \tilde{G}_\delta(T_{-\sigma}u_2)\|_{L^2(0, T)} \\
+ \|G_\delta(T_{-\sigma}u_2) - \tilde{G}_\delta(T_{-\sigma}u_2)\|_{L^2(0, T)} =: I_1(\sigma, \epsilon) + I_2(\sigma).
\]
Hence, we have the existence of \((\sigma_2, \epsilon_2)\) such that \(\phi_2 = J_{\epsilon_2} \ast (T_{-\sigma_2}u_2) \in C^\infty[-\delta, T]\) and
\[
\|\phi_2 - u_2\|_{L^2(-\delta, T)} + \|G_\delta \phi_2 - \tilde{G}_\delta(u_2)\|_{L^2(0,T)} \leq \zeta/2.
\]
Consequently, by defining \(\phi = \phi_1 + \phi_2 \in C^\infty[-\delta, T]\), we deduce that
\[
\|\phi - u\|_{L^2(-\delta, T)} + \|G_\delta \phi - \tilde{G}_\delta(u)\|_{L^2(0, T)} \\
\leq \sum_{k=1}^2 (\|\phi_k - u_k\|_{L^2(-\delta, T)} + \|G_\delta \phi_k - \tilde{G}_\delta(\eta_k u)\|_{L^2(0, T)}) \leq \zeta.
\]
This completes the proof. \(\square\)

**Lemma 2.4.** Suppose \(u \in V_\delta\). Then \(G_\delta u \in \mathcal{L}^2(0, T)\) and \(\tilde{G}_\delta u = \tilde{G}_\delta(u)\), where \(G_\delta\) is defined by (1.2).

**Proof.** By Lemma [2.3], there exists a sequence of smooth functions \(\{u_n\} \subset C^\infty[-\delta, T]\) such that \(u_n \to u\) and \(G_\delta u_n \to \tilde{G}_\delta(u)\) in \(\mathcal{L}^2(0, T)\). Now, with reference to the definition (1.2), we introduce the operator \(G_\delta^\prime\) for small \(\epsilon > 0\) as
\[
G_\delta^\prime(u) := \int_{\delta}^\delta (u(t) - u(t - s))s\rho_\delta(s) \, ds.
\]

First, using the density of \(C^\infty_c(0, T)\) in \(\mathcal{L}^2(0, T)\) and the fact that
\[
\lim_{\epsilon \to 0} (G_\delta^\prime(u), \varphi) = (G_\delta u, \varphi), \quad \text{for all } \varphi \in C^\infty_c(0, T),
\]
we have that \(\|G_\delta^\prime(u)\|_{\mathcal{L}^2(0, T)}\) is bounded uniformly in \(\epsilon\), by the uniform boundedness principle, and hence \(\tilde{G}_\delta u\) is a weak limit of \(G_\delta^\prime u\) in \(\mathcal{L}^2(0, T)\). Then it suffices to show that \(\lim_{n \to 0} \|G_\delta^\prime u_n\|_{\mathcal{L}^2(0, T)} = \|\tilde{G}_\delta u\|_{\mathcal{L}^2(0, T)}\). To this end, we choose a sequence of smooth functions \(\{u_n\}_{n=1}^\infty \subset C^\infty[-\delta, T]\) such that
\[
\lim_{n \to 0} \left(\|u_n - u\|_{\mathcal{L}^2(-\delta, T)} + \|G_\delta u_n - \tilde{G}_\delta u\|_{\mathcal{L}^2(0, T)}\right) = 0,
\]
using the construction in the proof of Lemma [2.3]. In particular, we choose a sequence of \(\{(\rho_n, \sigma_n)\}\) s.t.
\[
u_n = J_{\rho_n} \ast (T_{\sigma_n}(\eta_1 u)) + J_{\rho_n} \ast (T_{-\sigma_n}(\eta_2 u)).
\]
Then using the continuity of the translation operator and the property that \(\|J_{\rho} \ast \varphi\|_{\mathcal{L}^2(0, T)} \leq \|\varphi\|_{\mathcal{L}^2(0, T)}\), we derive that
\[
\|G_\delta u_n\|_{\mathcal{L}^2(0, T)} \leq C \left(\|G_\delta(\eta_1 u)\|_{\mathcal{L}^2(0, T)} + \|G_\delta(\eta_2 u)\|_{\mathcal{L}^2(0, T)}\right) \\
\leq C(\|G_\delta u\|_{\mathcal{L}^2(0, T)} + \|u\|_{\mathcal{L}^2(-\delta, T)}),
\]
where the last inequality is due to $\eta_1, \eta_2 \in C^{\infty}[-\delta, T]$. Therefore, $\|G_\delta u_n\|_{L^2(\Omega)}$ is bounded uniformly in $\epsilon$ and $n$, and hence

$$
\lim_{\epsilon \to 0} \|G_\delta u\|_{L^2(0, T)} = \lim_{\epsilon \to 0} \lim_{n \to \infty} \|G_\delta u_n\|_{L^2(0, T)} = \lim_{n \to \infty} \lim_{\epsilon \to 0} \|G_\delta u_n\|_{L^2(0, T)} = \lim_{n \to \infty} \|G_\delta u_n\|_{L^2(0, T)} = \|\tilde{G}_\delta u\|_{L^2(0, T)}.
$$

By the definition (1.2), $G_\delta(u)$ exists and $G_\delta(u) = \tilde{G}_\delta(u)$.

Similarly, we may define the space $V_\delta^*$ by

$$
V_\delta^* = \{ v \in L^2(0, T + \delta) \text{ s.t. } \tilde{G}_\delta^*(v) \in L^2(0, T) \},
$$

with the operator $\tilde{G}_\delta^*$ on $L^2(0, T + \delta)$ by

$$
(\varphi, \tilde{G}_\delta^* v) = \int_0^{T+\delta} G_\delta(\varphi)(t)v(t) \, dt, \quad \text{for all } \varphi \in C_c(0, T).
$$

Then we have the following result of integration by parts.

**Corollary 2.2.** For $u \in V_\delta \cap \{ u = 0 \text{ in } (-\delta, 0) \}$ and $v \in V_\delta^* \cap \{ u = 0 \text{ in } (T, T + \delta) \}$, it holds that $(G_\delta u, v) = -(u, G_\delta^* v)$.

**Proof.** By Lemma 2.3, we may find $\phi \in C^{\infty}[-\delta, T]$ such that

$$
\|\phi - u\|_{L^2(-\delta, T)} + \|G_\delta \phi - G_\delta u\|_{L^2(0, T)} \leq \epsilon.
$$

Then we derive that

$$
-(u, G_\delta^* v) = -(\phi, G_\delta^* v) + (\phi - u, G_\delta^* v) = (G_\delta \phi, v) + (\phi - u, G_\delta^* v) = (G_\delta u, v) + (G_\delta \phi - G_\delta u, v) + (\phi - u, G_\delta^* v).
$$

Since

$$
|(G_\delta \phi - G_\delta u, v) + (\phi - u, G_\delta^* v)| \leq (\|G_\delta^* v\|_{L^2(0, T)} + \|v\|_{L^2(0, T)})\epsilon,
$$

the desired result follows. \qed

2.2. Localization. Next, we state results on localization theory, which plays a crucial role in studying local limits of nonlocal models. They are again similar to the more general cases presented in [22].

**Lemma 2.5.** Let $\varphi \in H^1(-\delta_0, T)$ for some $\delta_0 > 0$ and the nonlocal operator $G_\delta$ be given in [1.2]. Then it holds that

$$
\lim_{\delta \to 0} \|G_\delta(\varphi) - \varphi'\|_{L^2(0, T)} = 0.
$$

A localization result for the adjoint nonlocal operator $G_\delta^*$ may be derived via a similar argument.

**Corollary 2.3.** Let $\varphi \in H^1(0, T + \delta_0)$ for some $\delta_0 > 0$, and $G_\delta^*$ be the nonlocal operator given in [2.2]. Then it holds that

$$
\lim_{\delta \to 0} \|G_\delta^*(\varphi) - \varphi'\|_{L^2(0, T)} = 0.
$$

**Remark 2.1.** Further generalizations of the localization theory of the nonlocal gradient $G_\delta$ in the strong topology of Sobolev spaces $W^{1,p} (p \neq 2)$, or the space of functions of bounded variation, have been discussed in [23] and [22].

3. Nonlocal initial value problem. To prepare for the study on equation (1.1), we consider the space-independent nonlocal initial value problem (1.6) in this section. Our discussions are divided into the case of integrable kernels and that of more general kernels.
3.1. Well-posedness for integrable kernels. We start with the case of \( s \rho(s) \in L^1(0, \delta) \), and, hereinafter, we denote \( b_\delta = \int_0^\delta s \rho(s) \, ds \). The problem \( (1.6) \) can be seen as an standard integral equation with a nonlocal constrained initial value condition, whose well-posedness follows from the fixed point theorem, as stated in the following lemma.

**Lemma 3.1.** Suppose that \( g \in L^\infty(-\delta, 0) \) and \( f \in L^\infty(0, T) \), and \( s \rho_\delta(s) \in L^1(0, \delta) \). Then the initial value problem \( (1.6) \) with \( \lambda > 0 \) has a unique solution \( u \in L^\infty(0, T) \) which satisfies

\[
\|u\|_{L^\infty(0, T)} \leq \|g\|_{L^\infty(0, T)} + \lambda^{-1}\|f\|_{L^\infty(0, T)}.
\]

**Proof.** We note the fact that \( u \in L^\infty(0, T) \) is a solution of \( (1.6) \) if and only if it is a fixed point of the operator \( K \) such that for \( t \in (0, T) \)

\[
(K \varphi)(t) = \frac{1}{b_\delta + \lambda} \left( \int_0^\delta (t - s) s \rho_\delta(s) \, ds + f(t) \right),
\]

(3.1)

with a nonlocal constraint \( \varphi = g \) for \( t \in (-\delta, 0) \). Now we define a function set \( W_\delta \) in \( L^\infty(0, T) \) for some \( T > 0 \),

\[
W_\delta = \{ u \in L^\infty(0, T) : u = g \text{ in } (-\delta, 0) \text{ and } \|u\|_{L^\infty(0, T)} \leq \|g\|_{L^\infty(-\delta, 0)} + \lambda^{-1}\|f\|_{L^\infty(0, T)} \}.
\]

We observe that

\[
|K \varphi(t)| \leq \frac{1}{b_\delta + \lambda} \left( \int_0^\delta s \rho_\delta(s) \, ds \|\varphi\|_{L^\infty(0, T)} + \|f\|_{L^\infty(0, T)} \right)
\]

\[
\leq \frac{1}{b_\delta + \lambda} \left( b_\delta \|g\|_{L^\infty(-\delta, 0)} + \lambda^{-1} b_\delta \|f\|_{L^\infty(0, T)} + \|f\|_{L^\infty(0, T)} \right)
\]

\[
\leq \|g\|_{L^\infty(-\delta, 0)} + \lambda^{-1}\|f\|_{L^\infty(0, T)}.
\]

Further, we have for \( \varphi, \psi \in W_\delta \) and \( t \in (0, T) \),

\[
|K \varphi(t) - K \psi(t)| \leq b_\delta (b_\delta + \lambda)^{-1} \|\varphi - \psi\|_{L^\infty(0, T)},
\]

with \( b_\delta (b_\delta + \lambda)^{-1} < 1 \). So \( K \) is a contraction from \( W_\delta \) to \( W_\delta \). Then applying the fixed point theorem, there exists a unique solution \( u \in L^\infty(0, T) \) for \( (1.6) \) which satisfies the a priori estimate. \( \square \)

The preceding estimate is suitable for large \( \lambda \) and may be exploited to pick up the regularity of \( (1.1) \). However, in the case that \( \lambda \) is close to 0, it is beneficial to develop a sharper a priori estimate. To this end, it suffices to consider the trivial initial data and a nontrivial right hand side. Now we define

\[
W_\delta = \{ u \in L^\infty(0, c_0 \delta) : u = 0 \text{ in } (-\delta, 0) \text{ and } \|u\|_{L^\infty(0, c_0 \delta)} \leq 2\delta \|f\|_{L^\infty(0, c_0 \delta)} \}
\]

with \( c_0 \) such that

\[
2\delta \int_{c_0 \delta}^\delta s \rho_\delta(s) \, ds \geq 1.
\]

The existence of such a constant is assured by the fact that

\[
2\delta \int_0^\delta s \rho_\delta(s) \, ds \geq 2 \int_0^\delta s^2 \rho_\delta(s) \, ds = 2 > 1.
\]
Then it holds for $\varphi \in W_\delta$ and $t \in (0, c_0 \delta)$ that
\[
|K\varphi(t)| = \frac{1}{b_\delta + \lambda} \left( \int_0^t s \rho_\delta(s) \, ds \|\varphi\|_{L^\infty(0, c_0 \delta)} + \|f\|_{L^\infty(0,T)} \right)
\leq \frac{1}{b_\delta + \lambda} \left( 2\delta \int_0^{c_0 \delta} s \rho_\delta(s) \, ds \|f\|_{L^\infty(0, c_0 \delta)} + \|f\|_{L^\infty(0, c_0 \delta)} \right)
\leq \frac{1}{b_\delta + \lambda} \left( 2\delta \int_0^{c_0 \delta} s \rho_\delta(s) \, ds + 1 \right) \|f\|_{L^\infty(0, c_0 \delta)} \leq 2\delta \|f\|_{L^\infty(0, c_0 \delta)}.
\]

The contraction property follows as before, and hence there exists a unique solution $u \in W_\delta$. Then repeating this argument, we may conclude the following result.

**Proposition 3.1.** Suppose that $g \in L^\infty(-\delta,0)$ and $f \in L^\infty(0,T)$, and $s \rho_\delta(s) \in L^1(0,\delta)$. Then the initial value problem (1.6) with $\lambda \geq 0$ has a unique solution $u \in L^\infty(0,T)$ satisfying
\[
\|u\|_{L^\infty(0,t)} \leq c \left( \|g\|_{L^\infty(-\delta,0)} + \max(\delta, t) \|f\|_{L^\infty(0,T)} \right).
\]

An estimate in $L^2$-norm follows from Lemma 3.1 directly, i.e., for $g \in L^\infty(-\delta,0)$ and $f \in L^\infty(0,T)$, the unique solution satisfies
\[
\|u\|_{L^2(0,T)} \leq T^{1/2} \left( \|g\|_{L^\infty(-\delta,0)} + \lambda^{-1} \|f\|_{L^\infty(0,T)} \right).
\]

A sharper estimate (for large $T$) in the $L^2$-norm is given below.

**Lemma 3.2.** Suppose that $g \in L^\infty(-\delta,0)$, $f \in L^2(0,T)$ and $s \rho(s) \in L^1(0,\delta)$. Then there exists an unique solution satisfying
\[
\|u\|_{L^2(0,T)} \leq \lambda^{-1/2} \|g\|_{L^\infty(-\delta,T)} + \lambda^{-1} \|f\|_{L^2(0,T)}.
\]

**Proof.** Let the operator $K$ be defined by (3.1) and
\[
W_\delta = \{ u \in L^2(0,T) : u = g \text{ in } (-\delta,0) \text{ and } \|u\|_{L^2(0,T)} \leq \lambda^{-1/2} \|g\|_{L^\infty(-\delta,0)} + \lambda^{-1} \|f\|_{L^2(0,T)} \}.
\]

By repeating the argument in the proof of Lemma 3.1 we have for any $\varphi \in W_\delta$,
\[
\|K\varphi(t)\|_{L^2(0,T)} \leq \frac{1}{b_\delta + \lambda} \left( \int_0^\delta s \rho_\delta(s) \|\varphi(t-s)\|_{L^2(0,T)} \, ds + \|f\|_{L^2(0,T)} \right)
\leq \frac{1}{b_\delta + \lambda} \left( b_\delta^{1/2} \left( \int_0^\delta s \rho_\delta(s) \|\varphi(t-s)\|_{L^2(0,T)}^2 \, ds \right)^{1/2} + \|f\|_{L^2(0,T)} \right)
\leq \lambda^{-1/2} \|g\|_{L^\infty(-\delta,T)} + \lambda^{-1} \|f\|_{L^2(0,T)}.
\]

This shows the existence and uniqueness of the solution, and hence the a priori estimate.

**Remark 3.1.** As in Proposition 3.1 we may provide a sharper estimate in case of small $\lambda$. In particular, for $g \in L^\infty(-\delta,0)$, $f \in L^\infty(0,T)$, and $s \rho(s) \in L^1(0,\delta)$, the initial value problem with $\lambda \geq 0$ has a unique solution $u \in L^2(0,T)$ such that
\[
\|u\|_{L^2(0,T)} \leq cT^{1/2} \left( \|g\|_{L^\infty(-\delta,0)} + \max(\delta, T) \|f\|_{L^\infty(0,T)} \right).
\]
Remark 3.2. In case of the integrable kernel associated with trivial right hand side, we can slightly strengthen the result, for $g \in L^2(-\delta,0)$. To this end, we let

$$W_\delta = \{ u = g \text{ in } (-\delta,0) \text{ and } \| u \|_{L^2(0,T)} \leq b_\delta^{1/2} \lambda^{-1/2} \| g \|_{L^2(-\delta,0)} \}.$$

and use the following estimate for $\varphi \in W_\delta$

$$\| K \varphi(t) \|_{L^2(0,T)} \leq \frac{1}{b_\delta + \lambda} \int_0^\delta s \rho_\delta(s) \| \varphi(t-s) \|_{L^2(0,T)} \, ds$$

$$\leq \frac{b_\delta^{1/2} \lambda^{-1/2}}{b_\delta + \lambda} \left( \int_0^\delta s \rho_\delta(s) \| \varphi(t-s) \|^2_{L^2(0,T)} \, ds \right)^{1/2}$$

$$\leq \frac{b_\delta^{1/2} \lambda^{-1/2}}{b_\delta + \lambda} \left( b_\delta + b_\delta^2 \lambda^{-1} \right)^{1/2} \| g \|_{L^2(-\delta,0)} \leq b_\delta^{1/2} \lambda^{-1/2} \| g \|_{L^2(-\delta,0)}.$$

Then we obtain the well-posedness of the initial value problem as well as the following a priori estimate

$$\| u \|_{L^2(0,T)} \leq b_\delta^{1/2} \lambda^{-1/2} \| g \|_{L^2(-\delta,0)} + \lambda^{-1} \| f \|_{L^2(0,T)}.$$

3.2. Well-posedness in case of non-integrable kernels. Next, we turn to the case of $s \rho_\delta(s) \notin L^1(\Omega)$, which is more delicate due to the weak singularity of the kernel. We thus consider the limit of a sequence of initial value problems for $\{ u_\epsilon \}_{\epsilon > 0}$ given by

$$\mathcal{G}_\delta^\epsilon u_\epsilon(t) + \lambda u_\epsilon(t) = f(t), \quad t \in [0,T],$$

$$u_\epsilon(t) = g(t), \quad t \in [-\delta,0). \quad (3.2)$$

with the truncated nonlocal operator defined by (2.9). With the singularity removed, argument given in the previous subsection becomes applicable which then leads to the well-posedness of (3.2), i.e., there is a unique solution $u_\epsilon$ of (3.2) satisfying the uniform bound

$$\| u_\epsilon \|_{L^2(0,T)} \leq \lambda^{-1/2} \| g \|_{L^\infty(-\delta,0)} + \lambda^{-1} \| f \|_{L^2(0,T)}, \quad (3.3)$$

and hence for $\lambda > 0$,

$$\| \mathcal{G}_\delta^\epsilon u_\epsilon \|_{L^2(0,T)} + \| u_\epsilon \|_{L^2(0,T)} \leq C,$$

with $C > 0$ independent of $\epsilon$. Therefore $u_\epsilon$ weakly converges (up to a subsequence) to $u$ in $L^2(0,T)$, and $u$ satisfies the estimate

$$\| u \|_{L^2(0,T)} \leq \liminf_{\epsilon \to 0} \| u_\epsilon \|_{L^2(0,T)} \leq \left( \lambda^{-1/2} \| g \|_{L^\infty(-\delta,0)} + \lambda^{-1} \| f \|_{L^2(0,T)} \right), \quad (3.4)$$

The next lemma shows that $u$ is indeed a solution of (1.6).

Lemma 3.3. Let $u$ be an $L^2(0,T)$ weak limit of the sequence of functions $\{ u_\epsilon \}$ defined by (3.2). Then $u$ is the unique solution of (1.6).

Proof. From the definition of $u_\epsilon$ and properties of weak convergence, the initial condition is satisfied automatically. It suffices to show that $u$ satisfies the governing equation in (1.6). Taking $\varphi \in C_c^\infty(0,T)$ with zero extension out of the interval $(0,T)$ and multiplying $\varphi$ on both sides of (3.2), we have the following after integrating from 0 to $T$ and interchanging the order of integration,

$$(f, \varphi) = (\mathcal{G}_\delta^\epsilon u_\epsilon, \varphi) + (\lambda u_\epsilon, \varphi)$$

$$= - \int_{-\delta}^T u_\epsilon(t) \int_t^{t+\epsilon} (\varphi(y) - \varphi(t))(y-t) \rho_\delta(|y-t|) \, dy \, dt + \lambda (u_\epsilon, \varphi).$$
Using the proof of Lemma 2.2, we note the fact that
\[ \int_{-\delta}^{T} \left( \int_{t}^{t+\epsilon} (\varphi(y)-\varphi(t))(y-t)\rho_{\delta}|y-t| \, dy \right)^{2} \, dt \leq c \int_{-\delta}^{T} \left( \int_{t}^{t+\epsilon} |y-t|^{2}\rho_{\delta}|y-t| \, dy \right)^{2} \, dt \|\varphi'\|_{L^{2}(0,T)}^{2}. \]

Taking \( \epsilon \to 0 \) and applying (2.3), we derive that
\[ (f,\varphi) = -\int_{-\delta}^{T} u(t)G_{\delta}\varphi(t) \, dt + \lambda(u,\varphi) = \langle G_{\delta}u, \varphi \rangle + \lambda(u,\varphi). \]

From the observation \( G_{\delta}u = f - \lambda u \in L^{2}(0,T) \) and Lemma 2.4, we get that \( G_{\delta}u = G_{\delta}u \in L^{2}(0,T) \) and hence \( u \) satisfies (1.6) almost everywhere. Finally, the uniqueness follows from the comparison principle. \( \square \)

To sum up, we state the main result in this section, which gives the well-posedness of the initial value problem (1.6) and the a priori estimates.

**Theorem 3.1.** Let \( g \in L^{\infty}(-\delta,0) \) and \( f \in L^{2}(0,T) \), then there is a unique solution \( u \in V_{\delta} \) of problem (1.1) with \( \lambda > 0 \) and it holds that
\[ \|u\|_{L^{2}(0,T)} \leq c(\lambda^{-1/2}\|g\|_{L^{\infty}(-\delta,0)} + \lambda^{-1}\|f\|_{L^{2}(0,T)}). \]

Moreover, if \( f \in L^{\infty}(0,T) \), the solution \( u \in L^{\infty}(0,T) \) satisfies
\[ \|u(t)\|_{L^{\infty}(0,T)} \leq c \left( \|g\|_{L^{\infty}(-\delta,0)} + \lambda^{-1}\|f\|_{L^{\infty}(0,T)} \right). \]

**Proof.** The first estimate is a consequence of (3.4) and Lemma 3.3 while the second assertion follows from the similar argument (by replacing weak convergence with weak* convergence), and the uniform upper bound given in Lemma 3.1. \( \square \)

**Remark 3.3.** The preceding argument also works for unbounded intervals, i.e., when the governing equation of (1.6) is posed in \((0,\infty)\). Suppose that \( g \in L^{\infty}(-\delta,0) \) and \( f \in L^{2}(0,\infty) \), then there is a unique solution \( u \in V_{\delta} \) of the problem (1.1) and it holds that
\[ \|u\|_{L^{2}(0,\infty)} \leq c \left( \lambda^{-1/2}\|g\|_{L^{\infty}(-\delta,0)} + \lambda^{-1}\|f\|_{L^{2}(0,\infty)} \right). \]

In view of Proposition 3.1 and 3.1, we may derive the a priori estimate for (1.6) in case of \( \lambda = 0 \), for arbitrary kernels. This is given in the following theorem.

**Corollary 3.1.** Let \( g \in L^{\infty}(-\delta,0) \) and \( f \in L^{\infty}(0,T) \), then there is a unique solution \( u \in V_{\delta} \) of problem (1.1) with \( \lambda = 0 \), and it holds that
\[ \|u\|_{L^{2}(0,T)} \leq c T^{1/2}\|g\|_{L^{\infty}(-\delta,0)} + \max(\delta,T)\|f\|_{L^{\infty}(0,T)}. \]

Further, the solution \( u \in L^{\infty}(0,T) \) satisfies for all \( t > 0 \)
\[ \|u(t)\|_{L^{\infty}(0,T)} \leq c \left( \|g\|_{L^{\infty}(-\delta,0)} + \max(\delta,t)\|f\|_{L^{\infty}(0,T)} \right). \]

**Remark 3.4.** Corollary 3.1 provides a uniform upper bound in case that \( \lambda = 0 \) in a finite interval, i.e., \( T < \infty \). However, the estimate for the infinite interval is more delicate due to the lack of decay property. Analogously to the local counterpart, in which case \( \|u\|_{L^{\infty}(0,\infty)} \leq |u(0)| + \|f\|_{L^{1}(0,\infty)} \), we expect a reasonable a priori bound in \( L^{\infty}(0,\infty) \) provided that \( g \in L^{\infty}(-\delta,0) \) and \( f \in L^{p}(0,\infty) \). This requires further studies on Duhambel’s principle and is beyond the scope of this paper.
3.3. Short-time behavior of the solution. The next theorem gives the short-time asymptotic behavior of the solution of problem (3.8) that is useful to establish regularity pick-up in the Section 4. We are particularly interested in the case of non-integrable fractional kernels, i.e., \( \rho(s) = (\alpha + 1)s^{\alpha - 2}\delta^{-\alpha - 1} \) with \( \alpha \in (-1, 0) \), which is given in the following theorem.

**Theorem 3.2.** Let \( g \in L^\infty(-\delta, 0) \) and \( f \equiv 0 \), and suppose that the kernel satisfies \( \rho(s) = (\alpha + 1)s^{\alpha - 2}\delta^{-\alpha - 1} \) with \( \alpha \in (-1, 0) \). Then there is a unique solution \( u \in V_\delta \) of problem (1.1) and it holds that

\[
|u(t)| \leq \frac{\|g\|_{L^\infty(-\delta, 0)}}{1 + c_\alpha \delta^{1+\alpha} \lambda^{-\alpha}},
\]

where the constant \( c_\alpha \) only depends on \( \alpha \).

**Proof.** The unique existence has been given in \( 3.1 \). To prove the estimate, we first consider the case \( g \geq 0 \). This implies \( u(t) \geq 0 \) for any \( t \in [0, T] \) by comparison principle. We then define

\[
v(t) = \begin{cases} \|g\|_{L^\infty}, & t \in (-\delta, 0], \\ \|g\|_{L^\infty}(1 + c_\alpha \delta^{1+\alpha} \lambda^{-\alpha})^{-1}, & t \in (0, T], \end{cases}
\]

which is the solution of the following initial value problem

\[
G_\delta v(t) + \lambda v(t) = f_v(t), \quad t \in (0, T],
\]

\[
v(t) = \|g\|_{L^\infty}, \quad t \in [-\delta, 0].
\]  \( (3.5) \)

Next, we show that \( f_v \in L^\infty(0, T) \) and hence the problem is well-posed. In fact, due to the nonincreasing property of \( v \), we have \( G_\delta v(t) \geq 0 \), and hence for \( t \in [\delta, T] \)

\[
-G_\delta v(t) = c_\alpha \delta^{1+\alpha} \lambda \|g\|_{L^\infty} \int_0^\delta \frac{s \rho(s)(t^{-\alpha} - (t - s)^{-\alpha})}{(1 + c_\alpha \delta^{1+\alpha} \lambda^{-\alpha})(1 + c_\alpha \delta^{1+\alpha} \lambda(t-s)^{-\alpha})} \, ds \\
\leq c_\alpha (1 + \alpha) \lambda \|g\|_{L^\infty} \int_0^\delta s^{\alpha-1}(t^{-\alpha} - (t - s)^{-\alpha}) \, ds \\
\leq c_\alpha (1 + \alpha) \lambda \|g\|_{L^\infty} B(\alpha + 1, -\alpha),
\]

where \( B(\alpha+1, -\alpha) \) denotes the Beta function such that \( B(\alpha+1, -\alpha) = \frac{\Gamma(\alpha+1)\Gamma(-\alpha)}{\Gamma(\alpha+1-\alpha)} = \frac{\pi}{\sin(-\alpha\pi)} \). Further, for \( t \in (0, \delta) \) we may derive that

\[
-G_\delta v(t) \leq c_\alpha \delta^{1+\alpha} \lambda \|g\|_{L^\infty} \left( \int_0^t s \rho(s)(t^{-\alpha} - (t - s)^{-\alpha}) \, ds + \int_t^\delta s \rho(s)t^{-\alpha} \, ds \right) \\
\leq c_\alpha (1 + \alpha) \lambda \|g\|_{L^\infty} B(\alpha + 1, -\alpha).
\]  \( (3.6) \)

Hence \( f_v = G_\delta v(t) + \lambda v \in L^\infty(0, T) \). We now show that \( f_v \geq 0 \). Combining \( (3.6) \) and \( (3.7) \) we achieve that for all \( t > 0 \),

\[
f_v(t) = G_\delta v(t) + \lambda v(t) \geq \frac{\lambda \|g\|_{L^\infty}}{1 + c_\alpha \delta^{1+\alpha} \lambda^{-\alpha}} (1 - c_\alpha (1 + \alpha) B(\alpha + 1, -\alpha)).
\]
Then by choosing
\[ c_\alpha = \frac{1}{(1 + \alpha)B(\alpha + 1, -\alpha)} = \frac{\sin(-\alpha \pi)}{(1 + \alpha)\pi}, \]
we obtain that \( f_v(t) \geq 0 \) and hence the desired result.

Finally, for a general initial data \( g \in L^\infty(0, T) \), we consider the splitting \( g = g^+ - g^- \) where \( g^+ \geq 0 \) and \( g^- \leq 0 \) are in \( L^\infty(0, T) \). Let \( u^+ \) and \( u^- \) be solutions of problem \((1.6)\) with initial data \( g^+ \) and \( g^- \) respectively. Then it holds that
\[ |u(t)| = u^+ + (-u^-) \leq \frac{\|g^+\|_{L^\infty(0, T)} + \|g^-\|_{L^\infty(0, T)}}{1 + c_\alpha \delta^{\frac{1}{1+\alpha} - \frac{\lambda}{\lambda t}}} = \frac{\|g\|_{L^\infty(0, T)}}{1 + c_\alpha \delta^{\frac{1}{1+\alpha} - \frac{\lambda}{\lambda t}}}, \]
which completes the proof of the lemma. \( \Box \)

**Remark 3.5.** For a general kernel, the solution of \((1.6)\) satisfies the bound
\[ |u(t)| \leq (1 + \lambda t)^{-1}\|g\|_{L^\infty(-\delta, 0)}. \]
The basic idea is similar to the one of proving Theorem 3.2. Define \( u_m(t) \) as
\[ u(t) = \begin{cases} \|g\|_{L^\infty(-\delta, 0)}, & t \in [-\delta, 0] \\ (1 + \lambda t)^{-1}\|g\|_{L^\infty(-\delta, 0)}, & t \in (0, T] \end{cases} \]
We may verify that \( G_\delta u_m + \lambda u_m \geq 0 \) for \( t \in (0, \delta) \), in view of the fact
\[ G_\delta u_m + \lambda u_m \geq \frac{\lambda \|g\|_{L^\infty}}{1 + \lambda t} \left( -\int_0^t s^2 \rho_\delta(s) \, ds - \int_t^\delta s t \rho_\delta(s) \, ds + 1 \right) = \frac{\lambda \|g\|_{L^\infty}}{1 + \lambda t} \int_t^\delta s(s-t) \rho_\delta(s) \, ds \geq 0, \]
where the fact that \( \int_0^\delta s^2 \rho_\delta(s) \, ds = 1 \) is used.

### 3.4 Local limits of nonlocal models

In this part, we develop a new framework to study the local limit of the nonlocal problem \((1.6)\) as \( \delta \to 0 \), i.e., \( \|u-u_\delta\|_{L^2(0, T)} \to 0 \), where \( u_\delta \) is the solution of \((1.6)\) and \( u \) is the solution of the following local initial value problem
\[ u'(t) + \lambda u(t) = f(t), \quad t \in (0, T], \]
\[ u(0) = g(0). \]
Here we assume that \( g \in H^1(-\delta_0, 0) \) with some \( \delta_0 > 0 \) so that its trace at \( t = 0 \) is well defined. Now we give the main theorem of this part.

**Theorem 3.3.** Let \( u_\delta \) and \( u \) be solutions of problems \((1.6)\) and \((3.8)\), respectively, with \( g \equiv 0 \) and \( f \in L^2(0, T) \). Then it holds that
\[ \lim_{\delta \to 0} \|u_\delta - u\|_{L^2(0, T)} = 0. \]
**Proof.** By Theorem 3.1 we know that \( \|u_\delta\|_{L^2(0, T)} \) is bounded uniformly in \( \delta \). Then there exists a weakly convergent subsequence such that \( u_\delta \rightharpoonup w \in L^2(0, T) \). Next, we show that \( w = u \). In fact, for all \( \varphi \in C_c^\infty(0, T) \), it holds that
\[ (G_\delta(u_\delta), \varphi) + (\lambda u_\delta, \varphi) = (f, \varphi) \quad \text{for all } \delta. \]
Using Lemma 2.1 and Corollary 2.3 we derive that
\[ (G_\delta(u_\delta), \varphi) = -\int_0^T u_\delta(t) (G_\delta^*(\varphi))(t) \, dt \to -(w, \varphi') = \langle w', \varphi \rangle. \]
Therefore, by multiplying $\varphi \in C_c^\infty(0, T)$ on the equation of (1.6), integrating from 0 to $T$, and letting $\delta \to 0$, we derive that

$$(f, \varphi) = \lim_{\delta \to 0} ((G_\delta(u_\delta), \varphi) + (\lambda u_\delta, \varphi)) = \langle w', \varphi \rangle + (\lambda w, \varphi).$$

Hence $w' + \lambda w = f \in L^2(0, T)$. Next we shall show that $w$ satisfies the initial condition $w(0) = 0$. Let $C_c^\infty[0, T]$ denote the space consists of functions in $C^\infty[0, T]$ which can be extended to $C^\infty[0, \infty)$ by zero extension. For any $\varphi \in C_c^\infty(0, T)$, we may find a sequence $\varphi_n$ in $C_c^\infty(0, T)$ such that $\lim_{n \to 0} ||\varphi_n - \varphi||_{L^2(0,T)} = 0$. Then using the uniform boundedness of $\|G_\delta u_\delta\|_{L^2(0,T)}$, we have

$$\lim_{n \to \infty} \lim_{\delta \to 0} (G_\delta(u_\delta), \varphi_n) = \lim_{\delta \to 0} \lim_{n \to \infty} (G_\delta(u_\delta), \varphi_n).$$

The using (3.9), the left hand side can be written by

$$\lim_{n \to \infty} \lim_{\delta \to 0} (G_\delta(u_\delta), \varphi_n) = \lim_{\delta \to 0} (w', \varphi_n) = (w', \varphi)$$

On the other hand, Lemma 2.1 and Corollary 2.3 yield that

$$\lim_{\delta \to 0} \lim_{n \to \infty} (G_\delta(u_\delta), \varphi_n) = \lim_{\delta \to 0} (G_\delta(u_\delta), \varphi) = \lim_{\delta \to 0} (u_\delta, G_\delta^* \varphi) = (w, \varphi').$$

Consequently, we have $\langle w', \varphi \rangle = (w, \varphi')$ for $\varphi \in C_c^\infty(0, T)$, and hence $w(0) = 0$. Then $w = u$ follows immediately from well-posedness of the local initial value problem (3.8).

Finally, we shall show that the convergence is indeed strong in $L^2(0, T)$. To this end, we define a sequence of auxiliary problems: for a fixed $\delta$, to find $v_\delta \in V_\delta^*$ such that

$$-(G_\delta^* v_\delta)(t) + \lambda v_\delta(t) = u_\delta(t), \quad t \in [0, T],$$
$$v_\delta(t) = 0, \quad t \in (T, T + \delta).$$

(3.10)

The argument in Section 3.1 shows uniqueness and existence of the solution $v_\delta \in V_\delta^*$ which satisfies

$$\|v_\delta\|_{L^2(0,T)} + \|G_\delta^* v_\delta\|_{L^2(0,T)} \leq c \|u_\delta\|_{L^2(0,T)} \leq c \|f\|_{L^2(0,T)}.$$ 

Then up to a subsequence, $v_\delta \to v$ in $L^2(0, T)$ and

$$(u, \varphi) = \lim_{\delta \to 0} (u_\delta, \varphi) = \lim_{\delta \to 0} (-(G_\delta^* v_\delta, \varphi) + (\lambda v_\delta, \varphi)) = -\langle v', \varphi \rangle + (\lambda v, \varphi).$$

Hence $v$ satisfies $-v' + \lambda v = u \in L^2(0, T)$ and the preceding discussion shows that $v(T) = 0$. Further, multiplying $u_\delta$ on the equation (3.10) and integrating from 0 to $T$, and using Corollary 2.2 and the equation (1.6), we achieve that

$$\|u_\delta\|_{L^2(0,T)}^2 = (G_\delta u_\delta, v_\delta) + \lambda (u_\delta, v_\delta) = (f, v_\delta).$$

On the other hand, we have

$$\|u\|_{L^2(0,T)}^2 = (u, v') + \lambda (v, u) = (u', v) + \lambda (v, u) = (f, v).$$

Then the weak convergence of $\{v_\delta\}$ yields that

$$\lim_{\delta \to 0} \|u_\delta\|_{L^2(0,T)}^2 = \lim_{\delta \to 0} (f, v_\delta) = (f, v) = \|u\|_{L^2(0,T)}^2.$$

As a result, the strong convergence is implied and the theorem is proved. \hfill \square

**Corollary 3.2.** Let $u_\delta$ and $u$ be solutions of problems (1.6) and (3.8) respectively, with $g \in H^1(-\delta_0, 0)$ for some $\delta_0 > 0$. Then it holds that

$$\lim_{\delta \to 0} \|u_\delta - u\|_{L^2(0,T)} = 0.$$
we denote by \( \dot{\cdot} \) and consider the error equation for small \( \delta \) with
\begin{align*}
\{4.\text{ Nonlocal-in-time parabolic equations.}\}
\end{align*}
Then by the a priori estimate in Lemma 3.1, we may prove that 
\( \|r\|_{L^2(0,T)} \leq C \) where the constant \( C > 0 \) independent of \( \delta \). Then we consider the problem
\( \ddot{\mathbf{G}}_\delta w(t) + \lambda w(t) = \ddot{f}(t) := f(t) + \mathbf{G}_\delta u_\delta(t) + \lambda u_\delta(t), \quad t \in (0,T), \)
with the homogeneous initial data. Since the \( L^2 \)-norm of \( \ddot{f} \) is bounded uniformly in \( \delta \) and \( \|\mathbf{G}_\delta u_\delta - u_\delta\|_{L^2(0,T)} \to 0 \) by Lemma 2.5, we have \( w_\delta \to w \) in \( L^2(0,T) \), where \( w \) is the solution of local initial value problem with right hand side \( f + u_\delta + \lambda u_\delta \). Then \( u_\delta = w_\delta + u_\delta \), the solution of the nonlocal problem \( (1.6) \), converges to \( u = w + u_\delta \), the solution of the local problem \( (3.8) \), in sense of \( L^2(0,T) \).

**Remark 3.6.** Analogously, the localization theory in a finite interval can be established in the case that \( \lambda = 0 \), in view of a priori estimates in Corollary 3.1.

**Remark 3.7.** It is possible to derive an optimal rate of the \( \delta \)-convergence, in case of very regular data, e.g. \( g \equiv 0 \) and \( f \in C^1_1[0,T] \). Here the space \( C^1_1[0,T] \) is defined by the norm \( \|f\|_{C^1_1[0,T]} = \|\ddot{f}\|_{C^0(-\infty,T)} \), where \( \ddot{f} \) denotes the zero extension of \( f \). Then using the solution representation \( u = \int_0^t e^{-\lambda(t-s)}f(s)\,ds \) we have the following estimate
\( \|u\|_{C^1_1[0,T]} \leq C\|f\|_{C^1_1[0,T]} \).
The convergence analysis follows from the Taylor’s expansion. We let \( e = u_\delta - \tilde{u} \) and consider the error equation for small \( \delta \)
\( \mathbf{G}_\delta e + \lambda e = -\mathbf{G}_\delta \tilde{u} + \partial_\delta \tilde{u} = O(\delta). \)
Then by the a priori estimate in Lemma 3.1, we may prove that \( \|e\|_{L^2(0,T)} \leq c\delta \). However, the convergence rate for irregular or incompatible data is still an open question.

4. **Nonlocal-in-time parabolic equations.** In this section, we turn to the non-local-in-time parabolic equation \( (1.1) \) and study its well-posedness and smoothing properties. We first introduce some standard notation for convenience. For \( r \geq -1 \), we denote by \( H^r(\Omega) \subset H^{-1}(\Omega) \) the Hilbert space induced by the norm:
\( \|v\|_{H^r(\Omega)}^2 = \sum_{j=1}^{\infty} \lambda_j^r \langle v, \varphi_j \rangle^2, \)
with \( \{\lambda_j\}_{j=1}^{\infty} \) and \( \{\varphi_j\}_{j=1}^{\infty} \) being respectively the eigenvalues and the \( L^2 \)-orthonormal eigenfunctions of \( -\Delta \) on the domain \( \Omega \) with a homogeneous Dirichlet boundary condition. Further, \( \|v\|_{H^0(\Omega)} = \|v\|_{L^2(\Omega)} = (v, v)^{1/2} \) is the norm in \( L^2(\Omega) \) and \( \|v\|_{H^{-1}(\Omega)} = \|v\|_{H^{-1}(\Omega)} \) is the norm in \( H^{-1}(\Omega) \). Besides, it is easy to verify that \( \|v\|_{H^r(\Omega)} = \|\nabla v\|_{L^2(\Omega)} \) is also the norm in \( H^r_\partial(\Omega) \) and \( \|v\|_{H^r(\Omega)} \) is equivalent to the norm in \( H^2(\Omega) \cap H^r_\partial(\Omega) \) [29] Lemma 3.1. Motivated by the fact that \( H^r(\Omega), r \geq -1 \) form a Hilbert scale of interpolation spaces, we denote \( \|\cdot\|_{H^r(\Omega)} \) to be the norm on the interpolation scale between \( H^0_\partial(\Omega) \) and \( L^2(\Omega) \) when \( r \in [0,1] \) and \( \|\cdot\|_{H^r(\Omega)} \) to be the norm on the interpolation scale between \( L^2(\Omega) \) and \( H^{-1}(\Omega) \) when \( r \) is in \([-1,0] \). Then \( \|\cdot\|_{H^r(\Omega)} \) and \( \|\cdot\|_{H^r(\Omega)} \) are equivalent for \( r \in [-1,1] \) by interpolation.

For a Banach space \( B \), we define the space involving time
\( L^r(0,T;B) = \{ u(t) \in B \text{ for a.e. } t \in (0,T) \text{ and } \|u\|_{L^r(0,T;B)} < \infty \}, \)
for any \( r \geq 1 \), and the norm \( \| \cdot \|_{L^r(0,T;B)} \) is defined by

\[
\|u\|_{L^r(0,T;B)} = \left\{ \begin{array}{ll}
\left( \int_0^T \|u(t)\|_B^r \, dt \right)^{1/r}, & r \in [1, \infty), \\
\esssup_{t \in (0,T)} \|u(t)\|_B, & r = \infty.
\end{array} \right.
\]

Next we consider the well-posedness of mode (1.1). It is of particular interest to consider the case of a weak data, which is widely used in inverse and control problems. Similarly to the standard parabolic counterpart, we may define the weak solution of the nonlocal-in-time parabolic equation (1.1), using the preceding notations.

**Definition 4.1.** We say a function \( u(t) \in \dot{H}^1(\Omega) \) with \( \mathcal{G}_\delta u(t) \in \dot{H}^{-1}(\Omega) \) is a weak solution of the nonlocal-in-time parabolic initial boundary value problem (1.1) provided

1. \( \langle \mathcal{G}_\delta u, v \rangle + (\nabla u, \nabla v) = \langle f, v \rangle \), for all \( v \in \dot{H}^1(\Omega) \) and almost all \( t \in [0,T] \), and
2. \( u = g \) for \( t \in (-\delta,0) \).

**4.1. Solution theory of nonlocal-in-time parabolic equations.** Now we state our main results in the section, which gives the existence and uniqueness of the weak solution defined above.

**Theorem 4.1.** Let \( g(x,t) = p(t)q(x) \in L^\infty(-\delta,0;\dot{H}^r(\Omega)) \) with \( r \geq 0 \) and \( f \equiv 0 \). Then the problem (1.1) has a unique weak solution \( u \in C([0,T];\dot{H}^r(\Omega)) \cap C((0,T];\dot{H}^{r+2}(\Omega)) \) such that \( \mathcal{G}_\delta u(t) \in \dot{H}^r(\Omega) \) and

\[
\|u(t)\|_{\dot{H}^r(\Omega)} \leq c t^{-(r-\delta)/2} \|g\|_{L^\infty(-\delta,0;\dot{H}^r(\Omega))}, \quad \text{for all } t \in [r,r+2]
\]

Further, \( u \in L^2(0,T;\dot{H}^{-1}(\Omega)) \) and \( u \) satisfies

\[
\|u\|_{L^2(0,T;\dot{H}^{-1}(\Omega))} + \|\mathcal{G}_\delta u\|_{L^2(0,T;\dot{H}^{-1}(\Omega))} \leq c \|g\|_{L^\infty(-\delta,0;\dot{H}^r(\Omega))},
\]

where the constant \( c \) is independent of horizon parameter \( \delta \).

**Proof.** First, we give the solution representation of (1.1). Let \( g_n(x) = \langle g(x,t), \varphi_n \rangle = p(t) c_n \) where \( c_n = \langle g, \varphi_n \rangle \) is a time-independent constant. Define \( u_n \) as the solution of the initial value problem

\[
\mathcal{G}_\delta u_n(t) + \lambda_n u_n(t) = 0, \quad t \in [0,T],
\]

\[
u_n(t) = g_n(t), \quad t \in (-\delta,0).
\]

Theorem 3.1 ensures a unique solution that \( \|u_n\|_{L^2(0,T)} \leq c \lambda_n^{-1/2} \|g_n\|_{L^\infty(-\delta,0)} \). Then we define the function

\[
u(x,t) = \sum_{n=1}^{\infty} u_n(t) \varphi_n(x), \quad (4.1)
\]

Applying Theorem 3.1, we have the following estimate

\[
\|u(t)\|_{\dot{H}^r(\Omega)}^2 = \sum_{n=1}^{\infty} \lambda_n^r |u_n(t)|^2 \leq c \|p\|_{L^\infty(-\delta,0)}^2 \sum_{n=1}^{\infty} \lambda_n^r |\langle g, \varphi_n \rangle|^2 = c \|g\|_{L^\infty(-\delta,0;\dot{H}^r(\Omega))}^2.
\]
Moreover, by Remark 3.3 we derive that
\[ ||u(t)||_{H^{r+2}(\Omega)}^2 = \sum_{n=1}^{\infty} \lambda_n^{r+2} |u_n(t)|^2 \leq c ||p||_{L^\infty}^2 \sum_{n=1}^{\infty} \frac{\lambda_n^2}{(1 + \lambda_n t)^2} \lambda_n^r |\langle q, \varphi_n \rangle|^2 \]
\[ = ct^{-2} ||p||_{L^\infty(-\delta, 0; L^2(\Omega))}^2 ||q||_{H^r(\Omega)} = ct^{-2} ||q||_{L^\infty(-\delta, 0; H^r(\Omega))}. \]

Then the intermediate cases can be obtained by interpolation. Further, \( u \in C([0, T], H^r(\Omega)) \) is the direct result of the Lebesgue’s dominated convergence theorem and the fact that the series in \( 4.1 \) converges in \( H^r(\Omega) \) uniformly in \( t \in [0, T] \). Besides, we note that \( h > \delta \) and \( H_\varphi \) is evident. To show that \( u \) is a weak solution of \( 1.1 \), let \( u_N = \sum_{n=1}^{N} \lambda_n u_n(t) \varphi_n(t) \), then \( u_N \) satisfies
\[ \langle G_{\delta} u_N, v \rangle + \langle \nabla u_N, \nabla v \rangle = 0, \quad \text{for all } v \in V^N := \text{span}\{\varphi_1, ..., \varphi_N\}, \]
with the initial condition \( u_N(t) = \sum_{n=1}^{N} \lambda_n u_n(t) \varphi_n(t) \). We have immediately that
\[ ||u_N||_{L^2(0, T; H^1(\Omega))} + ||G_{\delta} u_N||_{L^2(0, T; H^{-1}(\Omega))} \leq C \]
for some constant \( C \) uniform in \( N \). Hence, \( u_N \to u \) in \( L^2(0, T; H^1(\Omega)) \) and \( G_{\delta} u_N \to G_{\delta} u \) in \( L^2(0, T; H^{-1}(\Omega)) \) up to a subsequence by the duality argument. Then for \( v \in C^1([0, T]; H^1(\Omega)) \) having the form \( v(t) = \sum_{n=1}^{N} v_n(t) \varphi_n(t) \) with \( \{v_n\} \) smooth enough, we have
\[ \int_0^T \langle G_{\delta} u_N, v \rangle + \langle \nabla u_N, \nabla v \rangle dt = \int_0^T \langle f, v \rangle dt. \]

By passing to weak limit we obtain that
\[ \int_0^T \langle G_{\delta} u, v \rangle + \langle \nabla u, \nabla v \rangle dt = \int_0^T \langle f, v \rangle dt. \]

Then this equation is valid for all \( v \in L^2(0, T; H^1(\Omega)) \) by density argument. Thus, for all \( v \in H^1(\Omega) \)
\[ \langle G_{\delta} u, v \rangle + \langle \nabla u, \nabla v \rangle = \langle f, v \rangle \quad \text{a.e. } t \in [0, T]. \]
As a result, \( u \) is a weak solution of \( 1.1 \). Consequently, \( G_{\delta} u = \Delta u \), which together with the a priori estimate on \( u \) yields the bound of \( G_{\delta} u \).

Finally, we prove the uniqueness of the weak solution to \( 1.1 \) within the class given in Definition 4.1. To this end, we set \( g \equiv 0 \) and \( f \equiv 0 \) and prove that the problem \( 1.1 \) has only a trivial solution. Let \( u \) be the weak solution. Then by the definition of \( \{\lambda_n, \varphi_n\}_{n=1}^{\infty} \), one observe that \( u_n = (u, \varphi_n) \) must satisfy \( G_{\delta} u_n(t) + \lambda_n u_n(t) = 0 \), and the initial condition \( u_n(t) = 0 \) for all \( t \in (-\delta, 0) \). Due to the existence and uniqueness of the initial value problem given in Theorem 3.1, we derive that \( u_n = 0, n = 1, 2, ..., \) Then by the completeness of \( \{\varphi_n\} \), we obtain that \( u \equiv 0 \) in \( (0, T) \times \Omega \).}

**Remark 4.1.** In view of Theorem 4.1, for initial value \( g = p(t) q(x) \in L^\infty(-\delta, 0; H^r(\Omega)) \), the solution \( u(t) \) of model \( 1.1 \) belongs to \( H^{r+2}(\Omega) \) for all \( t > 0 \). Then for \( t > \delta \) and \( t_0 \in (\delta, t) \), the similar argument works for the problem
\[ G_{\delta} y(x, t) - \Delta y(x, t) = 0, \quad \text{in } \Omega \times [0, T - t_0] \]
with homogeneous Dirichlet boundary condition, and the initial condition \( y(t) = u(t + t_0) \) for all \( t \in (-t_0, 0) \). Then we have \( y(t) \in \dot{H}^{r+4}(\Omega) \) for any \( t > \delta \) in view of the fact that

\[
\|y\|_{\dot{H}^{r+4}(\Omega)}^2 = \sum_{n=1}^{\infty} \lambda_n^{r+4} |y_n(t)|^2 \leq c(t - t_0)^{-2} \sum_{n=1}^{\infty} \lambda_n^{r+2} \sup_{t \in (t_0 - \delta, t_0)} |y_n(t)|^2
\]

\[
\leq c(t - t_0)^{-2}(t_0 - \delta)^{-2} \sum_{n=1}^{\infty} \lambda_n^r \sup_{t \in (-\delta, 0)} |g_n(t)|^2
\]

\[
= c(t - t_0)^{-2}(t_0 - \delta)^{-2} \|g\|_{L^\infty(-\delta, 0; \dot{H}^r(\Omega))}^2.
\]

Consequently \( u(t) \in \dot{H}^{r+4}(\Omega) \) for all \( t > \delta \). Then repeating the discussion, we may deduce that \( u(t) \in \dot{H}^{r+2k}(\Omega) \) for all \( t > (k - 1)\delta \). This is in sharp contrast to the its counterpart estimates for the fractional diffusion [25] and the standard diffusion [29, Chapter 3]. Recall that the solution of heat equation is infinitely differentiable, irrespective of the smoothness of the initial data, while the fractional diffusion only exhibits the spatial smoothing property of order 2 even for \( t \gg 0 \). Hence the nonlocal-in-time parabolic model (1.1) has the similar smoothing property as that of fractional diffusion for short time, i.e., \( t < \delta \), while for \( t \gg 0 \) it exhibits the similar solution behavior as that of standard diffusion.

The next theorem gives the well-posedness for inhomogeneous source data.

**Theorem 4.2.** Let \( g(x, t) \equiv 0 \) and \( f \in L^2(0, T; \dot{H}^r(\Omega)) \), with \( r \geq -1 \). Then the problem (1.1) has a unique weak solution \( u \in L^2(0, T; \dot{H}^{r+2}(\Omega)) \) such that

\[
\|u\|_{L^2(0, T; \dot{H}^{r+2}(\Omega))} + \|\mathcal{G}_\delta u\|_{L^2(0, T; \dot{H}^r(\Omega))} \leq c\|f\|_{L^2(0, T; \dot{H}^r(\Omega))},
\]

where the constant \( c \) is independent of horizon parameter \( \delta \).

**Proof.** Let \( u \) be the function defined by (4.1) where \( u_n(t) \) is the solution of the initial value problem

\[
\mathcal{G}_\delta u_n(t) + \lambda_n u_n(t) = f_n(t) := \langle f, \varphi_n \rangle, \quad t \in [0, T],
\]

\[
u_n(t) = 0, \quad t \in [-\delta, 0).
\]

By Theorem 3.1 the unique solution \( u_n(t) \) satisfies \( \|u_n\|_{L^2(0, T; \dot{H}^{r+2}(\Omega))} \leq c\lambda_n^{-1} \|f_n\|_{L^2(0, T; \dot{H}^r(\Omega))} \), and hence

\[
\|u\|_{L^2(0, T; \dot{H}^{r+2}(\Omega))}^2 = \sum_{n=1}^{\infty} \lambda_n^{r+2} \int_0^T |u_n(t)|^2 \, dt
\]

\[
\leq c \sum_{n=1}^{\infty} \int_0^T |f_n(t)|^2 \, dt = c\|f\|_{L^2(0, T; \dot{H}^r(\Omega))}^2.
\]

Using the similar argument as that of Theorem 4.1, we may show that \( u \) is the unique weak solution of the model (1.1). The estimate for \( \mathcal{G}_\delta u \) is direct result of triangle inequality, say,

\[
\|\mathcal{G}_\delta u\|_{L^2(0, T; \dot{H}^r(\Omega))} \leq \|\Delta u\|_{L^2(0, T; \dot{H}^r(\Omega))} + \|f\|_{L^2(0, T; \dot{H}^r(\Omega))} \leq c\|f\|_{L^2(0, T; \dot{H}^r(\Omega))}.
\]

This completes the proof. \( \square \)
4.2. Smoothing properties for particular types of kernels. We now consider two special types of kernels, the non-integrable fractional kernels and the integrable kernels, that are of particular interest. All the proofs use similar techniques as ones used in the preceding section, and are thus either omitted or given in Appendix.

As we mentioned in Section 1 the case of non-integrable fractional kernels, i.e., \( s \rho_\delta(s) = (\alpha + 1)s^{\alpha - 1}\delta - \alpha - 1 \) with \( \alpha \in (-1, 0) \), relates closely to the fractional diffusion. We can improve the estimate in Theorem 4.1 in this case, and characterize the short time behavior similar to that of fractional diffusion.

**Theorem 4.3.** Let \( g(x, t) = p(t)q(x) \in L^\infty(-\delta, 0; \dot{H}^r(\Omega)) \) and \( f \equiv 0 \), with \( r \geq 0 \). Let \( \rho_\delta = (\alpha + 1)s^{\alpha - 2}\delta - \alpha - 1 \) with \( \alpha \in (-1, 0) \). Then the solution \( u \) of problem (1.1) satisfies

\[
\|u(t)\|_{H^r(\Omega)} \leq c\delta^{\alpha - 1}(l-r)^2\|p\|_{L^\infty(-\delta, 0; \dot{H}^r(\Omega))}, \quad l \in [r, r + 2].
\]

**Proof.** The existence of the weak solution is given by Theorem 4.1. Now we deduce the a priori bound. By Theorem 3.2 we have for \( l \in [r, r + 2] \)

\[
\|u(t)\|_{H^r(\Omega)}^2 \leq \sum_{n=1}^{\infty} \lambda_n^l |u_n(t)|^2 \leq \sum_{n=1}^{\infty} \frac{\lambda_n^l}{(1 + c_\alpha\delta^{\alpha + 1}\lambda_n^{l-\alpha})^2} \|p(t)\|_{L^\infty(-\delta, 0)}^2 |q, \varphi_n| ^2
\]

\[
= c\delta^{\alpha - 1}l^\alpha \sum_{n=1}^{\infty} \frac{(\delta^{\alpha + 1}\lambda_n^{l-\alpha})^{l-r}}{(1 + c_\alpha\delta^{\alpha + 1}\lambda_n^{l-\alpha})^2} \lambda_n^l \|p(t)\|_{L^\infty(-\delta, 0)}^2 |q, \varphi_n| ^2
\]

\[
\leq c\delta^{\alpha - 1}l^\alpha \|p(t)\|_{L^\infty(-\delta, 0)}^2 \sum_{n=1}^{\infty} \lambda_n^l |q, \varphi_n| ^2
\]

\[
\leq c\delta^{\alpha - 1}(l-r)\|p\|_{L^\infty(-\delta, 0)}^2 |q|_{L^\infty(-\delta, 0; \dot{H}^r(\Omega))}. \]

\( \square \)

**Remark 4.2.** Let \( r = 0 \) and the kernel \( \rho_\delta \) be of the type \( s \rho_\delta(s) = (\alpha + 1)s^{\alpha - 1}\delta - \alpha - 1 \), \( \alpha \in (-1, 0) \). Then Theorem 4.3 provides a short time behavior

\[
\|u(t)\|_{H^2(\Omega)} \leq c\delta^{\alpha - 1}t^\alpha |g|_{L^\infty(-\delta, 0; L^2(\Omega))}
\]

for the solution of model (1.1). Recall that this is very close to the short time estimate for the fractional diffusion (cf. 23 Theorem 2.1), while the appearance of \( \delta^{\alpha - 1} \) is due to the normalization of the kernel.

Next, we consider the case of integrable kernels, i.e., \( s \rho_\delta(s) \in L^1(0, \delta) \), for which we may improve the smoothing property estimates given in the Theorem 4.1 even if the initial data is not variables-separable.

**Theorem 4.4.** Let \( g \in L^\infty(-\delta, 0; \dot{H}^r(\Omega)) \) with \( r \geq -1 \) and \( f \equiv 0 \), and the kernel \( \rho_\delta \) with \( s \rho_\delta(s) \in L^1(0, \delta) \). Then the problem (1.1) has a unique weak solution \( u \in L^\infty(0, T; \dot{H}^{r+2}(\Omega)) \) such that

\[
\|u\|_{L^\infty(0, T; \dot{H}^{r+2}(\Omega))} + \|\mathcal{G}_\delta u\|_{L^\infty(0, T; \dot{H}^r(\Omega))} \leq C\delta^{-1} |g|_{L^\infty(-\delta, 0; \dot{H}^r(\Omega))},
\]

where \( C \) denotes a constant independent of \( T \).

**Proof.** Without loss of generality, we consider the case that \( r = 0 \). The proof is based on the Banach fixed point theory. It is easy to observe that a function \( u \in L^2(0, T; H^2(\Omega)) \) is a solution of (1.1) if and only if it satisfies

\[
u(t) = (b_\delta + A)^{-1} \int_0^\delta s \rho_\delta(s) u(t - s) \, ds \quad \text{with} \quad u(t) = g(t), \quad t \in (-\delta, 0), \quad (4.2)
\]
where \( b_\delta = \int_0^\delta s \rho_\delta(s) \, ds \). We first show that the problem (4.2) has a unique solution in the set

\[ W_\delta = \{ u \in L^\infty(0,T;L^2(\Omega)) : u = g \text{ in } (-\delta,0), \|u\|_{L^\infty(0,T;L^2(\Omega))} \leq \|g\|_{L^\infty(-\delta,0;L^2(\Omega))} \} . \]

To this end, we set \( A = -\Delta \) for convenience and define the operator \( K \) as

\[ K\varphi(t) = (b_\delta + A)^{-1} \left( \int_0^\delta \varphi(t-s)s\rho_\delta(s) \, ds + f(t) \right) , \tag{4.3} \]

with \( b_\delta = \int_0^\delta s \rho_\delta(s) \, ds \). Then we have for \( \varphi \in W_\delta \) and \( t \in (0,T] \)

\[ \|K\varphi(t)\|_{L^2(\Omega)} \leq \|(b_\delta + A)^{-1}\| \int_0^\delta s \rho_\delta(s) \, ds \|\varphi\|_{L^\infty(0,T;L^2(\Omega))} \leq \|g\|_{L^\infty(-\delta,0;L^2(\Omega))} . \]

Hence the operator \( K \) maps \( W_\delta \) to \( W_\delta \). Further, for \( \varphi, \psi \in W_\delta \) and \( t \in (0,T] \), we derive that

\[ \|T\varphi(t) - T\psi(t)\|_{L^2(\Omega)} \leq b_\delta \|(b_\delta + A)^{-1}\| \|\varphi - \psi\|_{L^\infty(0,T;L^2(\Omega))} . \]

Then the contraction property follows directly from the fact \( b_\delta \|(b_\delta + A)^{-1}\| < 1 \). Hence there exists a unique solution in \( W_\delta \). Finally, we turn to the result of the lemma. For \( t \in (0,T) \), we note that

\[ \|u(t)\|_{H^2(\Omega)} = \|A(b_\delta + A)^{-1}\| \int_0^\delta s \rho_\delta(s) \|u(t) - s\|_{L^2(\Omega)} \, ds \leq b_\delta \|g\|_{L^\infty(-\delta,0;L^2(\Omega))} , \]

where the constant \( b_\delta \) is proportional to \( \delta^{-1} \). Then the estimate for \( \|G_\delta u\|_{L^\infty(0,T;L^2(\Omega))} \) is a direct consequence of the triangle inequality.

**Corollary 4.1.** Let \( g \in L^\infty(-\delta,0;\dot{H}^r(\Omega)) \) with \( r \geq -1 \) and \( f \equiv 0 \), and the kernel \( \rho_\delta \) satisfy \( s \rho_\delta(s) \in L^1(0,\delta) \). Then the problem (1.1) has a unique weak solution \( u \in L^2(0,T;H^{r+2}(\Omega)) \) such that

\[ \|u\|_{L^2(0,T;H^{r+2}(\Omega))} + \|G_\delta u\|_{L^2(0,T;H^r(\Omega))} \leq C\delta^{-1} T^{1/2} \|g\|_{L^\infty(-\delta,0;\dot{H}^r(\Omega))} . \]

### 4.3. Local limits of nonlocal models

In this section, we discuss the zero-horizon limit of the nonlocal model (1.1). Specifically, let \( u_\delta \) be the solution of (1.1), and \( u \) be the solution of the standard heat equation

\[ \begin{align*}
  u_t - \Delta u &= f & \text{in } \Omega, & T \geq t > 0, \\
  u &= 0 & \text{on } \partial\Omega, & T \geq t > 0, \\
  u(0) &= g(0) & \text{in } \Omega, .
\end{align*} \tag{4.4} \]

The next theorem shows that \( u_\delta \) strongly converges to \( u \) in \( L^2(0,T;L^2(\Omega)) \).

**Theorem 4.5.** Let \( u \) and \( u_\delta \) be the solutions of (4.4) and (1.1), with \( g = p(t)q(x) \in H^l(-\delta_0,0;\dot{H}^r(\Omega)) \), \( r \geq 0 \), \( \delta_0 > 0 \), and \( f \in L^2(0,T;H^l(\Omega)) \), \( l \geq -1 \). Then \( \|u - u_\delta\|_{L^2(0,T;L^2(\Omega))} \to 0 \) as \( \delta \to 0 \).

**Proof.** Let \( g_n = \langle g, \varphi_n \rangle \), \( f_n = \langle f, \varphi_n \rangle \), and \( u_\delta^n(t) \) be the solution of the initial value problem

\[ \begin{align*}
  G_\delta u_\delta^n(t) + \lambda u_\delta^n(t) &= f_n(t) & t \in [0,T] , \\
  u_\delta^n(t) &= g_n(t) & t \in (-\delta,0) .
\end{align*} \]
From Theorem 4.1 and 4.2, \( u(t) = \sum_{n=1}^{\infty} u_n(t) \phi_n(x) \) be the weak solution of (1.1). Similarly, \( u(t) = \sum_{n=1}^{\infty} u_n(t) \phi_n(x) \) be the weak solution of (4.4), where \( u_n(t) \) is the solution of the problem

\[
\begin{align*}
\frac{du}{dt} + \lambda u(t) &= f(t), \quad t \in (0, T], \\
\lambda u(0) &= g_0(0).
\end{align*}
\]

Using spectral decomposition and Corollary 3.2 we have

\[
\lim_{\delta \to 0} \| u - u_\delta \|_{L^2(0,T;L^2(\Omega))}^2 \leq \lim_{\delta \to 0} \sum_{n=1}^{\infty} |u_n(t) - u_n(t)|^2 dt \\
= \lim_{\delta \to 0} \sum_{n=1}^{\infty} \int_0^T |u_n(t) - u_n(t)|^2 dt \\
= \sum_{n=1}^{\infty} \left( \lim_{\delta \to 0} \int_0^T |u_n(t) - u_n(t)|^2 dt \right) = 0,
\]

This completes the proof of the theorem.

\( \square \)

**Remark 4.3.** Theorem 4.5 states the convergence of nonlocal-in-time parabolic model to the standard diffusion in \( L^2(0,T;L^2(\Omega)) \) as \( \delta \) approaches zero. However, the convergence rate still awaits further theoretical studies. Numerically, it highly depends on the smoothness of the initial data, i.e., it exhibits first order convergence for smooth data, but the order deteriorates for nonsmooth initial data; cf. Fig. 1 in Section 6.

5. **Semi-discrete finite element approximation.** The smoothing properties obtained in preceding section can be used to analyze a semi-discrete finite element scheme. Using the notations from [29], we let \( \{ T_h \} (0 < h < 1) \) be a family of shape regular and quasi-uniform partitions of the domain \( \Omega \) into \( d \)-simplices, called finite elements, with \( h \) denoting the maximum diameter. An approximate solution \( u_h \) is sought in the finite element space \( X_h \) of continuous piecewise linear elements over the triangulation \( \mathcal{T}_h \)

\[
X_h = \{ \chi \in H_0^1(\Omega) : \chi \text{ is a linear function over } \tau \quad \forall \tau \in \mathcal{T}_h \}.
\]

The semi-discrete Galerkin FEM for problem (1.1) reads: find \( u_h(t) \in X_h \) such that

\[
(\mathcal{G}_\delta u_h(t), \chi) + a(u_h, \chi) = (f, \chi), \quad \forall \chi \in X_h, \quad t > 0,
\]

with \( u_h(t) = P_h g(t) \) with \( t \in (-\delta, 0) \), and \( a(u, w) = (\nabla u, \nabla w) \) for \( u, w \in H_0^1(\Omega) \).

Upon introducing the discrete Laplace operator \( \Delta_h : X_h \to X_h \) via

\[
-(\Delta_h \psi, \chi) = (\nabla \psi, \nabla \chi) \quad \forall \psi, \chi \in X_h,
\]

the spatially semi-discrete problem can be written as

\[
\mathcal{G}_\delta u_h(t) - \Delta_h u_h(t) = P_h f(t) \quad \text{for } t \geq 0 \quad \text{with} \quad u_h(t) = P_h g(t), \quad t \in (-\delta, 0)
\]

(5.1)

where the discrete source term \( f_h = P_h f \).

Now we define the discrete norm \( \| \cdot \|_{H^p(\Omega)} \) on \( X_h \) for any \( p \in \mathbb{R} \)

\[
\| \psi \|_{H^p(\Omega)}^2 = \sum_{j=1}^{N} (\lambda_j^p)^2 (\psi, \phi_j^p)^2 \quad \psi \in X_h.
\]
Analogously, we introduce the associated spaces $\| \| \cdot \|_{L_r^r(0,T;\bar{H}^r(\Omega))}$, $r \in [1, \infty]$, on the space $X_h$.

The following estimate is the discrete analogue of Theorem 4.2. It is essential for the analysis of the lumped mass method in Section 4.

**Lemma 5.1.** Let $u_h$ be the solution of (5.1) with $g_h \equiv 0$. Then for arbitrary $p > -1$,

$$
\int_0^T \|G_{\delta}u_h(t)\|^2_{H^p(\Omega)} + \|u_h(t)\|^2_{H^{r+2}(\Omega)} dt \leq c \int_0^T \|f_h(t)\|^2_{\bar{H}^{r+1}(\Omega)} dt.
$$

In our analysis we the $L^2$-orthogonal projection $P_h : L^2(\Omega) \to X_h$ and the Ritz projection $R_h : H^1_0(\Omega) \to X_h$ are used. They are, respectively, defined by

$$(P_h \psi, \chi) = (\psi, \chi), \quad \forall \chi \in X_h,$$

$$(\nabla R_h \psi, \nabla \chi) = (\nabla \psi, \nabla \chi), \quad \forall \chi \in X_h.$$  

We need some properties of the Ritz projection $R_h$ and the $L^2$-projection $P_h$.

**Lemma 5.2.** Let the mesh be quasi-uniform. Then the operator $R_h$ satisfies:

$$\|R_h \psi - \psi\|_{L^2(\Omega)} + h \|\nabla(R_h \psi - \psi)\|_{L^2(\Omega)} \leq c h^2 \|\psi\|_{H^q(\Omega)} \quad \text{for} \ \psi \in \bar{H}^q(\Omega), \ q \in [1,2].$$

Further, for $s \in [0,1]$ we have

$$
\|(I - P_h) \psi\|_{H^s(\Omega)} \leq c h^{2-s} \|\psi\|_{H^2(\Omega)}, \quad \text{for all} \ \psi \in H^2(\Omega) \cap H^1_0(\Omega),
$$

$$
\|(I - P_h) \psi\|_{H^r(\Omega)} \leq c h^{1-s} \|\psi\|_{H^1(\Omega)}, \quad \text{for all} \ \psi \in H^1_0(\Omega).
$$

In addition, by duality $P_h$ is stable on $\bar{H}^s(\Omega)$ for $s \in [-1,0]$.

Now, we establish error estimates in $L^2(0,T;\bar{H}^p(\Omega))$-norm with $p = 0, 1$.

**Theorem 5.1.** Suppose $g = p(t)q(x) \in L^\infty(-\delta,0;\bar{H}^1(\Omega))$ and $f \in L^2(0,T;L^2(\Omega))$. Let $u$ and $u_h$ be the solutions of (1.1) and (5.1), respectively. Then

$$
\|u - u_h\|_{L^2(0,T;L^2(\Omega))} + h \|\nabla(u - u_h)\|_{L^2(0,T;L^2(\Omega))} \leq c h^2 \left( \|g\|_{L^\infty(-\delta,0;\bar{H}^1(\Omega))} + \|f\|_{L^2(0,T;L^2(\Omega))} \right),
$$

where the constant $c$ is independent of the horizon $\delta$.

**Proof.** We employ the $L^2$-projection $P_h u$, and split the error $u - u_h$ into two terms as:

$$
u_h - u = (u_h - P_h u) + (P_h u - u) := \tilde{\theta} + \bar{\theta}.
$$

Obviously, $P_h G_{\delta} \tilde{\theta} = G_{\delta} P_h (P_h u - u) = 0$ and using the identity $\Delta_h R_h = P_h \Delta$, we get the following equation for $\tilde{\theta}$:

$$
G_{\delta} \tilde{\theta}(t) - \Delta_h \tilde{\theta}(t) = - \Delta_h (R_h u - P_h u)(t), \quad t > 0, \quad \tilde{\theta}(0) = 0.
$$

By Theorems 4.1 and 4.2 and Lemma 5.2

$$
\|\tilde{\theta}\|_{L^2(0,T;L^2(\Omega))} + h \|\nabla \tilde{\theta}\|_{L^2(0,T;L^2(\Omega))} \leq C h^2 \|u\|_{L^2(0,T;H^2(\Omega))},
$$

$$
\leq C h^2 \left( \|g\|_{L^\infty(-\delta,0;\bar{H}^1(\Omega))} + \|f\|_{L^2(0,T;L^2(\Omega))} \right).
$$

}
By (5.3), Theorems 4.1, 4.2 and Lemmas 5.1, 5.2 we have for \( p = 0, 1 \):

\[
\int_0^T \| \tilde{v}(t) \|_p^2 dt \leq C \int_0^T \| \Delta_h (R_h u - P_h u)(t) \|_{H^{p-2}(\Omega)}^2 dt \\
\leq C \int_0^T \| (R_h u - P_h u)(t) \|_{H^p(\Omega)}^2 dt \leq Ch^{4-2p} \| u(t) \|_{L^2(0,T;H^{2+\gamma}(\Omega))}^2 \\
\leq Ch^{4-2p} \left( \| g \|_{L^\infty(-\delta,0;H^{\gamma}(\Omega))}^2 + \| f(t) \|_{L^2(0,T;L^2(\Omega))}^2 \right).
\]

Combing the preceding two estimates yields the desired assertion. 

The argument in the proof of Theorem 5.1 together with Theorem 4.2 deduces the following estimate in case of nonsmooth data.

**Corollary 5.1.** Suppose \( g = p(t)q(x) \in L^\infty(-\delta,0;L^2(\Omega)) \) and \( f \in L^2(0,T;H^{-1}(\Omega)) \). Let \( u \) and \( u_h \) be the solutions of (1.1) and (5.1), respectively. Then

\[
\| u_h - u \|_{L^2(0,T;L^2(\Omega))} \leq c\delta \| g \|_{L^\infty(-\delta,0;L^2(\Omega))} + \| f \|_{L^2(0,T;H^{-1}(\Omega))},
\]

where the constant \( C \) is independent of the horizon \( \delta \).

In case of integrable kernel, i.e., \( sp_\delta(s) \in L^1(-\delta,0) \), the assumption in Theorem 5.1 can be weakened, in view of Theorem 4.1 and the similar error estimate can be achieved with a \( \delta \)-dependent constant.

**Corollary 5.2.** Suppose \( g \in L^\infty(-\delta,0;H^{r}(\Omega)) \), \( f \in L^2(0,T;H^{r}(\Omega)) \) with \( r \in [-1,0] \), and \( sp_\delta(s) \in L^1(-\delta,0) \). Let \( u \) and \( u_h \) be the solutions of (1.1) and (5.1), respectively. Then

\[
\| u_h - u \|_{L^2(0,T;L^2(\Omega))} + h\| \nabla (u_h - u) \|_{L^2(0,T;L^2(\Omega))} \\
\leq c\delta h^{2+r} \left( \| g \|_{L^\infty(-\delta,0;H^{\gamma}(\Omega))} + \| f \|_{L^2(0,T;H^{-\gamma}(\Omega))} \right),
\]

where the constant \( c \) is independent of the horizon parameter \( \delta \).

Then the localization property follows directly from Theorem 5.1 and 4.5.

**Theorem 5.2.** Suppose that \( g = p(t)q(x) \in H^{1}(-\delta_0,0;L^2(\Omega)) \) with some \( \delta_0 > 0 \) and \( f \in L^2(0,T;H^{-1}(\Omega)) \). Let \( u_h^\delta \) be the solution of the nonlocal diffusion model (1.1) with \( \delta \in (0,\delta_0) \) and \( u \) be the solution of standard heat equation (4.4). Then it holds that

\[
\lim_{\delta \to 0} \lim_{h \to 0} \| u_h^\delta - u \|_{L^2(\Omega)} = \lim_{h \to 0} \lim_{\delta \to 0} \| u_h^\delta - u \|_{L^2(\Omega)} = 0
\]

In other words, the above theorem establishes the asymptotical compatibility of the semi-discrete Galerkin finite element approximation in the spirit of [30].

**Remark 5.1.** It is interesting to explore the pointwise-in-time error i.e., \( \|(u - u_h)(t)\|_{H^p(\Omega)} \), especially for nonsmooth data. However, these estimate generally utilizes Duhamel’s principle [14], Laplace transform [16] or energy estimate [29, Theorem 3.1, 3.2], all of which remain to be explored for the nonlocal model (1.1).

6. **Numerical results.** Here we present some one-dimensional numerical tests to substantiate and complement the theoretical findings in Section 4 and 5.
6.1. **Test problems.** We consider the problem (1.1) on the unit interval $\Omega = (0, 1)$, and perform numerical tests for three different groups of data:

(a) Smooth initial data: $g(x, t) = (-x^2 + x)e^t \in L^\infty(-\delta, 0; \dot{H}^2(\Omega))$ and $f \equiv 0$.

(b) Nonsmooth initial data: $g(x, t) \equiv \delta_\frac{1}{2}(x)$, a Dirac-delta function concentrated at $x = 0.5$, and $f \equiv 0$.

(c) Nonsmooth source data: $f(x, t) = (\chi_{[1/2, 1]}(t) + 1)\delta_{1/2}(x)$ and $g \equiv 0$.

In our computation, we divide the unit interval $(0, 1)$ into $N$ equally spaced subintervals, with a mesh size $h = 1/N$. The finite element space $X_h$ consists of continuous piecewise linear functions. We examine separately the spatial and temporal convergence rates at the terminal time $t = 1$. For the case of nonsmooth initial data, we are especially interested in the errors for $t$ close to zero, and thus we also present the errors at $t = 0.1, 0.01$ and $0.001$. To compute a numerical solution we have used also a time-stepping by discretizing the time interval, initial data, and discretize the nonlocal operator by

$$G_h(u)(t_n) = \int_0^\delta s\rho(s)(u(t_n) - u(t_n - s)) \, ds$$

$$= \sum_{k=0}^{m-1} \int_{k\tau}^{(k+1)\tau} s\rho(s)(u(t_n) - u(t_n - s)) \, ds$$

$$\approx \sum_{k=0}^{m-1} \int_{k\tau}^{(k+1)\tau} s\rho(s) \left( u(t_n) - \tau^{-1}(s - (k+1)\tau)u(t_{n-k}) \right)$$

$$+ (s - k\tau)(u(t_{n-k} - 1))) \, ds$$

(6.1)

It can be easily verified that the truncation error is second order provided that the function is smooth enough. Moreover, one may establish the unconditional stability. Indeed, even though it is not presented here, our experiments present a pointwisely first-order convergence no matter how singular the initial data is. Whenever needed, we use very fine meshes in both space and time to compute a reference solution. Unless otherwise specified, we set $\tau = 10^{-4}$, so that the error incurred by temporal discretization is negligible.

6.2. **Example (a): Smooth initial data.** In our experiments, we consider fractional kernels of the type $\rho_\alpha(s) = (\alpha + 1)s^{\alpha-2}\delta^{\alpha-1}$ with $\alpha > -1$. We measure the accuracy of the approximation $u_h(t)$ by the error $\|u - u_h\|_{L^2(0,T;\dot{H}^\alpha(\Omega))}$ for $p = 0, 1$, cf. Table [1]. In the table, rate refers to the ratio of the errors when the mesh size $h$ halves, and the numbers in the bracket denote theoretically predicted convergence rates. The numerical results show $O(h^2)$ and $O(h)$ convergence rates for the $L^2(\Omega)$- and $\dot{H}^1(\Omega)$-norms of the error, respectively, which fully confirm the theoretical estimates given in Theorem [5,4]. When $\delta$ approaches 0, the solution of the nonlocal model converges to a solution of the local heat equation in $L^2(0,T;\dot{H}^1(\Omega))$ (cf., Figure 1(a)), as stated in Theorem [5,2]. One may observe a first order convergence, i.e. $O(\delta)$, which theoretically remains to be explored.

6.3. **Example (b): Weak initial data.** The results of example (b) are presented in Tables [2,4] in which case $g(x) \in \dot{H}^{-1/2-\epsilon}(\Omega)$ for $t \in (-\delta, 0)$. The analysis for this very weak data is beyond the scope of the paper, but the numerical solutions presented here illustrate the limited smoothing of the nonlocal model (1.1). Since the solution is no longer be in $L^2(0,T;\dot{H}^1(\Omega))$ in case of $s\rho(s) \notin L^1(0, \delta)$, here we...
Table 1. $\|u - u_h\|_{L^2(0,T;\dot{H}^p(\Omega))}$ for example (a) with $\delta = 1$, $h = 1/N$, $\tau = 10^{-5}$, $T = 1$.

| $\gamma$ | norm $\backslash$ $N$ | 10 | 20 | 40 | 80 | 160 | rate |
|---------|------------------|----|----|----|----|-----|------|
| -0.8    | $L^2$            | 4.58e-4 | 1.15e-4 | 2.87e-5 | 7.18e-6 | 1.79e-6 | $\approx$ 2.00 (2.00) |
|         | $H^1$            | 9.42e-3 | 4.71e-3 | 2.35e-3 | 1.18e-3 | 5.53e-4 | $\approx$ 1.05 (1.00) |
| -0.2    | $L^2$            | 2.99e-4 | 7.51e-5 | 1.88e-5 | 4.69e-6 | 1.17e-6 | $\approx$ 2.00 (2.00) |
|         | $H^1$            | 5.57e-3 | 2.78e-3 | 1.39e-3 | 6.94e-4 | 3.26e-4 | $\approx$ 1.05 (1.00) |
| 0.2     | $L^2$            | 2.54e-4 | 6.36e-5 | 1.59e-5 | 3.97e-6 | 1.17e-6 | $\approx$ 2.00 (2.00) |
|         | $H^1$            | 4.59e-3 | 2.29e-3 | 1.14e-3 | 5.72e-4 | 3.26e-4 | $\approx$ 1.07 (1.00) |
| 0.8     | $L^2$            | 2.22e-4 | 5.58e-5 | 1.40e-5 | 3.49e-6 | 9.90e-7 | $\approx$ 2.00 (2.00) |
|         | $H^1$            | 3.98e-3 | 1.99e-3 | 9.94e-4 | 4.97e-4 | 2.34e-4 | $\approx$ 1.06 (1.00) |

Figure 1. Plot of $\|u^h_\delta - u^0_\delta\|_{L^2(0,T;L^2(\Omega))}$ for example (a), (b) and (c), with $\tau = 10^{-4}$, $h = 10^{-3}$. The plot shows the pointwise error, i.e., $\|(u - u_h)(t)\|_{\dot{H}^p(\Omega)}$ with $p = 0, 1$. In Tables 2 and 3, the mesh size is chosen to be $h = 1/(10k + 1)$, and thus the Dirac-$\delta$ function is not aligned with the grid. We observe that the $L^2(D)$- and $\dot{H}^1(\Omega)$-norms of error is of order $O(h^{3/2})$ and $O(h^{1/2})$, respectively, in case that $t < \delta$. However, for large $t$, i.e., $t > \delta$, the solution $u = u(t)$ could be smoother, cf. Remark 4.1. Hence, one may observe a higher-order convergence, see Table 4.1. This confirms the theoretical
findings discussed in Section 4. In Table 4, we report the results for the case that the \( \delta \)-function is supported at a grid point. Even if \( t \leq \delta \), one may observe a convergence with orders \( O(h^2) \) and \( O(h) \) respectively for errors measured in \( L^2(\Omega) \)- and \( H^1(\Omega) \)-norms, i.e., the method exhibits a superconvergence of one-half order higher, which theoretically remains to be established.

Figure 1(b) shows the convergence of the nonlocal models to the local one as \( \delta \to 0 \). Numerical results show that the rate of \( \delta \)-convergence deteriorates for nonsmooth data, which reflects the strong dependence on data regularity and remains to be studied further.

| Table 2. \( \| (u-u_h)(t) \|_{H^p(\Omega)} \) for example (b) with \( \delta = 1, h = 1/m, \tau = 10^{-4}, t = 0.5 \). |
| \( \gamma \) | norm \( \backslash m \) | 11 | 21 | 41 | 81 | 161 | 321 | rate |
|---|---|---|---|---|---|---|---|---|
| -0.8 | \( L^2 \) | 2.40e-3 | 8.04e-4 | 2.72e-4 | 9.29e-5 | 3.21e-5 | 1.10e-5 | \( \approx 1.55 \) |
| | \( H^1 \) | 7.40e-2 | 5.10e-2 | 3.54e-2 | 2.45e-2 | 1.66e-2 | 1.09e-2 | \( \approx 0.55 \) |
| -0.2 | \( L^2 \) | 6.04e-3 | 2.24e-3 | 8.08e-4 | 2.88e-4 | 1.02e-4 | 3.52e-5 | \( \approx 1.48 \) |
| | \( H^1 \) | 2.22e-1 | 1.59e-1 | 1.12e-1 | 7.86e-2 | 5.38e-2 | 3.53e-2 | \( \approx 0.53 \) |
| 0.2 | \( L^2 \) | 7.67e-3 | 2.88e-3 | 1.05e-3 | 3.77e-4 | 1.33e-4 | 4.63e-5 | \( \approx 1.47 \) |
| | \( H^1 \) | 2.88e-1 | 2.07e-1 | 1.47e-1 | 1.03e-1 | 7.08e-2 | 4.66e-2 | \( \approx 0.53 \) |
| 0.8 | \( L^2 \) | 9.35e-3 | 3.56e-3 | 1.30e-3 | 4.70e-4 | 1.66e-4 | 5.78e-5 | \( \approx 1.47 \) |
| | \( H^1 \) | 3.57e-1 | 2.58e-1 | 1.84e-1 | 1.29e-1 | 8.50e-2 | 5.81e-2 | \( \approx 0.52 \) |

| Table 3. \( \| (u-u_h)(t) \|_{H^p(\Omega)} \) for example (b) with \( \delta = 1, h = 1/m, \tau = 10^{-4}, t = 1.2 \). |
| \( \gamma \) | norm \( \backslash m \) | 11 | 21 | 41 | 81 | 161 | 321 | rate |
|---|---|---|---|---|---|---|---|---|
| -0.8 | \( L^2 \) | 1.36e-4 | 3.75e-5 | 9.86e-6 | 2.53e-6 | 6.37e-7 | 1.58e-7 | \( \approx 1.95 \) |
| | \( H^1 \) | 1.56e-3 | 8.09e-4 | 4.08e-4 | 2.02e-4 | 9.77e-5 | 4.47e-5 | \( \approx 1.03 \) |
| -0.2 | \( L^2 \) | 4.11e-4 | 1.13e-4 | 2.98e-5 | 7.63e-5 | 1.93e-4 | 4.77e-5 | \( \approx 1.95 \) |
| | \( H^1 \) | 5.98e-3 | 3.10e-3 | 1.57e-3 | 7.78e-4 | 3.76e-4 | 1.72e-4 | \( \approx 1.02 \) |
| 0.2 | \( L^2 \) | 5.22e-4 | 1.44e-4 | 3.78e-5 | 9.68e-6 | 2.44e-6 | 6.06e-7 | \( \approx 1.95 \) |
| | \( H^1 \) | 8.01e-3 | 4.15e-3 | 2.10e-3 | 1.04e-4 | 5.03e-4 | 2.30e-4 | \( \approx 1.02 \) |
| 0.8 | \( L^2 \) | 6.32e-4 | 1.74e-4 | 4.58e-5 | 1.17e-5 | 2.96e-6 | 7.34e-7 | \( \approx 1.95 \) |
| | \( H^1 \) | 1.01e-2 | 5.24e-3 | 2.65e-3 | 1.31e-4 | 6.34e-4 | 2.90e-4 | \( \approx 1.02 \) |

| Table 4. \( \| (u-u_h)(t) \|_{H^p(\Omega)} \) for example (b) with \( \delta = 1, h = 1/m, \tau = 10^{-4}, t = 0.5 \). |
| \( \gamma \) | norm \( \backslash m \) | 10 | 20 | 40 | 80 | 160 | 320 | rate |
|---|---|---|---|---|---|---|---|---|
| -0.8 | \( L^2 \) | 2.88e-4 | 7.21e-5 | 1.80e-5 | 4.51e-6 | 1.12e-6 | 2.77e-7 | \( \approx 2.00 \) |
| | \( H^1 \) | 7.88e-3 | 3.95e-3 | 1.97e-3 | 9.87e-4 | 4.64e-4 | 2.10e-4 | \( \approx 1.05 \) |
| 0.8 | \( L^2 \) | 6.27e-5 | 1.57e-5 | 3.94e-6 | 9.85e-7 | 2.45e-7 | 6.07e-8 | \( \approx 2.00 \) |
| | \( H^1 \) | 1.43e-3 | 7.12e-4 | 3.56e-4 | 1.78e-4 | 8.37e-5 | 3.78e-5 | \( \approx 1.05 \) |
6.4. Example (c): Nonsmooth source data. In this part, we present numerical results for example (c), i.e., \( g \equiv 0 \) and \( f \in L^\infty(0,T;H^s(\Omega)) \) with \( s \in (-1,-1/2) \), in Tables 5 and 6. The results are similar to ones for example (b), i.e., if the delta-function is not supported at a grid node, we observe that the method converges with the order \( O(h^{1/2}) \) in \( H^1(\Omega) \)-norm, while the convergence rate in the \( L^2(\Omega) \)-norm is \( O(h^{3/2}) \). The numerical results fully confirm our theoretical predictions. However, in case that the delta-function is aligned with the grid, we observe a superconvergence with one-half higher order. Figure 1(c) plots the error \( \|u^\delta_h - u_h\|_{L^2(0,T;L^2(\Omega))} \) as \( \delta \to 0 \), where \( u^\delta_h \) and \( u_h \) denote the solutions of nonlocal and local parabolic equations, respectively. One may observe a first order convergence, which confirms the theoretical finding in Theorem 4.5.

| \( \gamma \) | \( \text{norm} \backslash m \) | 21 | 41 | 81 | 161 | 321 | rate |
|---|---|---|---|---|---|---|---|
| -0.5 | \( L^2 \) | 4.67e-3 | 1.73e-3 | 6.23e-4 | 2.21e-4 | 7.68e-5 | \( \approx 1.47 \) \( (1.50) \) |
| | \( H^1 \) | 3.41e-1 | 2.43e-1 | 1.71e-1 | 1.17e-1 | 7.71e-2 | \( \approx 0.52 \) \( (0.50) \) |
| 0.5 | \( L^2 \) | 4.68e-3 | 1.73e-3 | 6.23e-4 | 2.21e-4 | 7.68e-5 | \( \approx 1.47 \) \( (1.50) \) |
| | \( H^1 \) | 3.41e-1 | 2.43e-1 | 1.71e-1 | 1.17e-1 | 7.72e-2 | \( \approx 0.52 \) \( (0.50) \) |

| \( \gamma \) | \( \text{norm} \backslash m \) | 10 | 20 | 40 | 80 | 160 | 320 | rate |
|---|---|---|---|---|---|---|---|---|
| -0.5 | \( L^2 \) | 2.51e-4 | 6.29e-5 | 1.57e-5 | 3.93e-6 | 9.79e-7 | 2.42e-7 | \( \approx 2.00 \) |
| | \( H^1 \) | 9.33e-3 | 4.68e-3 | 2.34e-3 | 1.17e-3 | 5.51e-4 | 2.49e-4 | \( \approx 1.07 \) |
| 0.5 | \( L^2 \) | 2.06e-4 | 5.18e-5 | 1.29e-5 | 3.23e-6 | 8.06e-7 | 1.99e-7 | \( \approx 2.02 \) |
| | \( H^1 \) | 7.12e-3 | 3.57e-3 | 1.79e-3 | 8.93e-4 | 4.20e-4 | 1.90e-4 | \( \approx 1.08 \) |

7. Conclusion. A nonlocal-in-time parabolic model is studied in this paper. Its well-posedness and smoothing properties are established, based on similar results derived first for the space-independent nonlocal-in-time models. These mathematical theory allow us to further discuss the zero-horizon limit of the nonlocal model as the horizon parameter \( \delta \) approaches zero, by establishing a novel duality argument. A semi-discrete scheme using piecewise linear finite element spatial approximation is also studied. Extensive numerical experiments are reported to confirm the theoretical analysis.

One distinct property of the nonlocal-in-time model is its limited smoothing property as we discussed in Section 4. In particular, one may observe a restricted spatial smoothing property at a small time, e.g., \( t < \delta \); cf. Theorem 4.1 and 4.3. However, we may repeat the argument to deduce a higher spatial regularity, say \( u(t) \in H^{r+2k}(\Omega) \) for all \( t > (k-1)\delta \); cf. Remark 4.1. Recall the fact that the solution of standard diffusion model is infinitely differentiable for all \( t > 0 \) [29] while the subdiffusion has a limited smoothing property with order 2 even if \( t \) approaches infinity [25]. Therefore the nonlocal model has the same smoothing property as the one of anomalous diffusion at a short time, but exhibits the behavior of standard diffusion when \( t \) approaches infinity. This illustrates our point that
(1.1) is in fact an intermediate case between standard and anomalous diffusion. The limited smoothing is numerically confirmed in Section 6 by evaluating the convergence order of the semi-discrete Galerkin finite element approximation.

Our analysis can be extended to problems with more general boundary conditions and/or spatially varying coefficients. As an example, one may replace the negative Laplacian with a generalized self-joint and positive definite second order elliptic operator, say $A = -\nabla \cdot \alpha(x) \nabla + q(x)$, where $\alpha(x)$ is a symmetric and positive definite $d \times d$ matrix-valued measurable function on the domain with smooth entries, and $q$ is an $L^\infty$-function. Then all the analytical studies in Section 4 and 5 can be repeated directly. Moreover, the argument can be extended to other types of boundary conditions, e.g., pure Neumann boundary condition whose spectrum contains a zero eigenvalue. In this case, we may treat nonzero eigenvalues analogously but applying Corollary 3.1 to the zero one so that all discussions in Section 4 and 5 remain valid.

The current work represents a first step towards the nonlocal-in-time parabolic model (1.1), and highly motivates further studies. First, the temporal regularity awaits rigorous theoretical analysis (e.g., see [10] and [2, 25] respectively for its standard and fractional counterparts). This will help to derive a sharp estimate for the time-stepping (6.1). The truncation error of (6.1) is second-order provided a smooth solution but it only exhibits a point-wisely first-order convergence for a general initial data empirically due to its lack of temporal smoothing property. This often happens to time-dependent problems with a initial layer, e.g., [15, 16, 20, 21]. Moreover, it will be beneficial to study the Duhamel’s principle of the model (1.1) which not only will provide reasonable a priori estimates for (1.6) in case that $\lambda = 0$ in the infinite interval (cf. Remark 3.4), but also is a crucial tool to analyze the point-wise error estimate, i.e., $\| (u - u_h)(t) \|_{\dot{H}^p(\Omega)}$ for the semi-discrete scheme.

Finally, it is interesting to explore the rate of $\delta$-convergence, which is of first-order in case of smooth data but deteriorates for very weak data, as presented in Section 6.

REFERENCES

[1] R. A. Adams and J. J. Fournier, *Sobolev Spaces*, vol. 140, Academic press, 2003.
[2] M. Allen, L. Caffarelli and A. Vasseur, A parabolic problem with a fractional time derivative, *Archive for Rational Mechanics and Analysis*, 221 (2016), 603–630.
[3] E. Askari, F. Bobaru, R. Lehoucq, M. Parks, S. Silling and O. Weckner, Peridynamics for multiscale materials modeling, in *Journal of Physics: Conference Series*, IOP Publishing, 125 (2008), 012078.
[4] B. Berkowitz, J. Klafter, R. Metzler and H. Scher, Physical pictures of transport in heterogeneous media: Advection-dispersion, random-walk, and fractional derivative formulations, *Water Resources Research*, 38 (2002), 1–9.
[5] J. Bourgain, H. Brezis and P. Mironescu, Another look at Sobolev spaces, *IOS Press, Amsterdam*, (2001), 439–455.
[6] O. Defterli, M. D’Elia, Q. Du, M. Gunzburger, R. Lehoucq and M. M. Meerschaert, Fractional diffusion on bounded domains, *Fractional Calculus and Applied Analysis*, 18 (2015), 342–360.
[7] Q. Du, M. Gunzburger, R. Lehoucq and K. Zhou, A nonlocal vector calculus, nonlocal volume-constrained problems, and nonlocal balance laws, *Mathematical Models and Methods in Applied Sciences*, 23 (2013), 493–540.
[8] Q. Du, M. Gunzburger, R. B. Lehoucq and K. Zhou, Analysis and approximation of nonlocal diffusion problems with volume constraints *SIAM Review*, 54 (2012), 667–696.
[9] Q. Du, Y. Tao, X. Tian and J. Yang, Robust a posteriori stress analysis for quadrature collocation approximations of nonlocal models via nonlocal gradients, *Comput. Methods Appl. Mech. Engrg.*, 310 (2016), 605–627.
[10] L. C. Evans, *Partial Differential Equations*, vol. 19 of Graduate Studies in Mathematics, American Mathematics Society, 2010.

[11] G. Gilboa and S. Osher, Nonlocal operators with applications to image processing, *Multiscale Modeling & Simulation*, 7 (2008), 1005–1028.

[12] M. Giona, S. Cerbelli and H. E. Roman, Fractional diffusion equation and relaxation in complex viscoelastic materials, *Physica A: Statistical Mechanics and its Applications*, 191 (1992), 449–453.

[13] B. Jin, R. Lazarov, J. Pasciak and Z. Zhou, Error analysis of semidiscrete finite element methods for inhomogeneous time-fractional diffusion, *IMA Journal of Numerical Analysis*, 35 (2014), 561–582.

[14] B. Jin, R. Lazarov and Z. Zhou, Error estimates for a semidiscrete finite element method for fractional order parabolic equations, *SIAM Journal on Numerical Analysis*, 51 (2013), 445–466.

[15] B. Jin, R. Lazarov and Z. Zhou, An analysis of the L1 scheme for the subdiffusion equation with nonsmooth data, *IMA Journal of Numerical Analysis*, 36 (2016), 197–221.

[16] B. Jin, R. Lazarov and Z. Zhou, Two fully discrete schemes for fractional diffusion and diffusion-wave equations with nonsmooth data, *SIAM Journal on Scientific Computing*, 38 (2016), A146–A170.

[17] Y. L. Keung and J. Zou, Numerical identifications of parameters in parabolic systems, *Inverse Problems*, 14 (1998), 83–100.

[18] A. Kilbas, H. Srivastava and J. Trujillo, *Theory and Applications of Fractional Differential Equations*, Elsevier, Amsterdam, 2006.

[19] B. Kilic and E. Madenci, Coupling of peridynamic theory and the finite element method, *Journal of Mechanics of Materials and Structures*, 5 (2010), 707–733.

[20] C. Lubich, I. Sloan and V. Thomée, Nonsmooth data error estimates for approximations of an evolution equation with a positive-type memory term, *Mathematics of Computation of the American Mathematical Society*, 65 (1996), 1–17.

[21] W. McLean and K. Mustapha, Time-stepping error bounds for fractional diffusion problems with non-smooth initial data, *Journal of Computational Physics*, 293 (2015), 201–217.

[22] T. Mengesha and Q. Du, Characterization of function spaces of vector fields and an application in nonlinear peridynamics, *Nonlinear Analysis A: Theory, Methods and Applications*, 140 (2016), 82–111.

[23] T. Mengesha and D. Spector, Localization of nonlocal gradients in various topologies, *Calculus of Variations and Partial Differential Equations*, 52 (2015), 253–279.

[24] R. Metzler and J. Klafter, The random walk’s guide to anomalous diffusion: A fractional dynamics approach, *Physics Reports*, 339 (2000), 1–77.

[25] K. Sakamoto and M. Yamamoto, Initial value/boundary value problems for fractional diffusion-wave equations and applications to some inverse problems, *Journal of Mathematical Analysis and Applications*, 382 (2011), 426–447.

[26] S. Silling, Reformulation of elasticity theory for discontinuities and long-range forces, *Journal of the Mechanics and Physics of Solids*, 48 (2000), 175–209.

[27] S. Silling and R. Lehoucq, Peridynamic theory of solid mechanics, *Advances in Applied Mechanics*, 44 (2010), 73–166.

[28] S. Silling, O. Weckner, E. Askari and F. Bobaru, Crack nucleation in a peridynamic solid, *International Journal of Fracture*, 162 (2010), 219–227.

[29] V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, vol. 25 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 2006.

[30] X. Tian and Q. Du, Asymptotically compatible schemes and applications to robust discretization of nonlocal models, *SIAM Journal on Numerical Analysis*, 52 (2014), 1641–1665.

[31] J. Xie and J. Zou, Numerical reconstruction of heat fluxes, *SIAM Journal on Numerical Analysis*, 43 (2005), 1504–1535.

Received March 2016; revised July 2016.

E-mail address: qd2125@columbia.edu
E-mail address: jyanghkbu@gmail.com
E-mail address: zhizhou0125@gmail.com