Abstract. We present an $hp$-error analysis of the discontinuous Galerkin time-stepping method for Volterra integro-differential equations with weakly singular kernels. We derive new error bounds that are explicit in the time-steps, the degrees of the approximating polynomials, and the regularity properties of the exact solution. It is then shown that start-up singularities can be resolved at exponential rates of convergence by using geometrically graded time-steps. Our theoretical results are confirmed in a series of numerical tests.

Key words. Volterra integro-differential equation, discontinuous Galerkin time-stepping, geometrically refined time-steps, exponential convergence.

AMS subject classifications. 65R20, 65L05, 65L60

1. Introduction. We introduce and analyze the $hp$-version of the discontinuous Galerkin (DG) time-stepping method for the Volterra integro-differential equation (VIDE):

$$u'(t) + a(t)u(t) + \int_0^t k_\alpha(t-s)b(s)u(s)\,ds = f(t), \quad t \in [0, T];$$
$$u(0) = u_0 \in \mathbb{R}. \quad (1.1)$$

Here, $a$, $b$ and $f$ are real functions that are continuous on $[0, T]$. Moreover, we assume that there are constants $\mu^* \geq \mu_*>0$ such that

$$\mu_* \leq a(t) \leq \mu^*, \quad |b(t)| \leq \mu^*, \quad t \in [0, T]. \quad (1.2)$$

The convolution kernel $k_\alpha$ is the weakly singular function given by

$$k_\alpha(s) := s^{-\alpha} \quad \text{for } \alpha \in (0, 1). \quad (1.3)$$

For any initial datum $u_0 \in \mathbb{R}$, the VIDE (1.1) has a unique solution $u : [0, T] \to \mathbb{R}$ which is continuously differentiable; see, e.g., [5, 2] and the references cited therein. More precisely, smooth (analytic) data $a$, $b$ and $f$ in (1.1) lead to solutions $u$ that are smooth (analytic) away from $t = 0$, but their second derivatives are unbounded at $t = 0$ and behave like

$$|u''(t)| \leq C t^{-\alpha}, \quad t > 0,$$

see [5, 3, 4] and [2, Section 7.1]; compare also Theorem 4.1 below. This loss of regularity in $u$ at $t = 0$ has the consequence that, on uniform time-steps with length $k$,
approximations $U$ generated by standard DG or collocation methods only possess low convergence order, that is,
\[ \|u - U\|_{L^\infty(0,T)} \leq Ck^{1-\alpha}, \]
see [4, 2]. This problem can be overcome by using meshes that are suitably refined near $t = 0$. We will show that the $hp$-version of the DG time-stepping method with geometrically graded time-steps leads to exponential rates of convergence.

The discontinuous Galerkin method was first proposed in [11] as a non-standard finite element method for the numerical solution of neutron transport problems. Applied to initial-value ODEs, it can be viewed as an implicit single-step scheme that allows for arbitrary variation in the time-steps and the degrees of the approximating polynomials. It has been shown in [11] that, in spite of the underlying Galerkin structure, the discontinuous Galerkin time-stepping method corresponds to certain implicit schemes of Runge-Kutta type. Subsequently, several important issues concerning the a-priori and a-posteriori error analyses of these schemes have been addressed; see, e.g., [7, 9, 8, 1] and the references therein. DG time-stepping has also been applied successfully to partial differential equations, and, in the context of parabolic problems, a large body of literature exists; we refer here only to the recent monograph [18] and the references cited therein. An error analysis of the DG time-stepping method applied to a parabolic integro-differential equation was recently presented in [10].

All the works mentioned above are concerned with the $h$-version of the DG time-stepping method where convergence is achieved on successively refined time-steps using a fixed, typically low approximation order. This is in contrast to the so-called $p$- and $hp$-versions where approximating polynomials of high degree are employed. The $hp$-approach is particularly beneficial for piecewise analytic solutions as its judicious combination of $h$- and $p$-refinement results in exponential rates of convergence. The time discretization of linear parabolic problems by the $hp$-DG time-stepping method was recently analyzed in [15, 19] (see also [16] for extensions to problems whose spatial operators are not self-adjoint). In particular, it has been shown that temporal start-up singularities induced by incompatible initial data can be resolved at exponential rates of convergence. Furthermore, in [14], a complete $hp$-error analysis of the DG time-stepping method has been carried out for non-linear initial value problems in $\mathbb{R}^d$.

In the present work, we derive new $hp$-error bounds in $L^2(0,T)$ and $L^\infty(0,T)$ for the DG time-stepping method applied to the Volterra integro-differential equation (1.1). The $L^2$-framework will be particularly important in the extension of the present results to partial VIDEs. Our estimates are completely explicit in the time-steps, the polynomial degrees, and the regularity properties of the exact solution. While these estimates give optimal convergence rates in the time-steps, they also show that the DG method converges if the polynomial degrees are increased at fixed time-steps. In particular, we prove that the $p$-version DG approach gives spectral accuracy for solutions with smooth time dependence, i.e., the convergence rates are of arbitrarily high algebraic order. In order to resolve start-up singularities induced by the weakly singular kernel $k_\alpha$ in (1.3), we employ time-steps that are geometrically refined towards $t = 0$, combined with polynomial degrees that are linearly increasing. We show that this $hp$-version approach leads to exponential rates of convergence for analytic data $a$, $b$, and $f$, in spite of the unboundedness of the second derivative of $u$ near $t = 0$. We present a series of numerical experiments that confirm our theoretical results.

Finally, we observe that, since the main purpose of this paper is to obtain insight into the basic $hp$-error analysis of DG methods on geometrically graded time-steps
for partial VIDEs, there will be no loss of generality by using the model problem
given by (1.1)–(1.3). In a sequel we shall use this insight as the key to obtain an
analogous estimate for partial VIDEs; we will then also describe typical applications
of such VIDEs.

The outline of the paper is as follows. In Section 2, we introduce the DG time-
stepping method for the VIDE (1.1), and prove existence and uniqueness of approx-
imate solutions. In Section 3 we carry out a complete $hp$-error analysis of the DG
method. In Section 4 we show that, on the basis of precise regularity results, the
solutions of (1.1) can be approximated exponentially fast on time-steps that are geo-
metrically graded towards $t = 0$. Our theoretical results are verified in the numerical
tests in Section 5. Finally, we end our presentation in Section 6 with concluding
remarks pointing to future work and open problems.

Throughout, standard notations and conventions are used. For an interval $I$, we
write $L^p(I)$, $1 \leq p \leq \infty$, for the Lebesgue space of $p$-integrable functions, endowed
with the norm $\| \cdot \|_{L^p(I)}$. We write $W^{k,p}(I)$ for the Sobolev space of order $k \in \mathbb{N}_0$
equipped with the usual norm $\| \cdot \|_{W^{k,p}(I)}$. For a non-integer exponent $s \geq 0$, the space
$W^{s,p}(I)$ is defined by the $K$-method of interpolation. We set $H^s(I) = W^{s,2}(I)$. We
write $P^r(I)$ for the space of all polynomials of degree $\leq r$. We denote by $C$ generic
constants not necessarily identical at different places, but always independent of the
discretization parameters of interest (such as time-steps and polynomial degrees).

2. Discontinuous Galerkin time-stepping. In this section, we introduce the
discontinuous Galerkin time-stepping method for the numerical approximation of the
Volterra integro-differential equation (1.1). We then show the existence and unique-
ness of the approximate solutions.

2.1. Discontinuous Galerkin discretization. Let $\mathcal{M}$ be a partition of $(0, T)$
into intervals $\{I_m\}_{m=1}^M$ given by $I_m := (t_{m-1}, t_m)$, with nodes
$$0 =: t_0 < t_1 < \ldots < t_{M-1} < t_M := T.$$ The length of $I_m$ is $k_m := t_m - t_{m-1}$. As usual, we set $k := \max_{m=1}^M k_m$. The
partition $\mathcal{M}$ is called quasi-uniform if there is a constant $C > 0$ such that $k \leq Ck_m$
for all $1 \leq m \leq M$.

We assign to each interval $I_m$ a polynomial degree $r_m \geq 0$ and introduce the
degree vector $\underline{r} = \{r_m\}_{m=1}^M$. We define $\overline{r} := \max_{m=1}^M r_m$. The tuple $(\mathcal{M}, \underline{r})$ is called
an $hp$-discretization of $(0, T)$. If $r_m = r$ for all $1 \leq m \leq M$, we simply write $(\mathcal{M}, r)$.

Let $\varphi : (0, T) \to \mathbb{R}$ be a function that is piecewise continuous with respect to the
partition $\mathcal{M}$. At the nodes the left- and right-sided limits of $\varphi$ are defined by
$$\varphi^+_m = \lim_{s \to 0^+, s > 0} \varphi(t_m + s), \quad 0 \leq m \leq M - 1,$n $$\varphi^-_m = \lim_{s \to 0^-, s > 0} \varphi(t_m - s), \quad 1 \leq m \leq M.$$ The jumps across interior nodes are given by $[\varphi]_m = \varphi^+_m - \varphi^-_m$, $1 \leq m \leq M - 1$.

For a given $hp$-discretization $(\mathcal{M}, \underline{r})$ of $(0, T)$, we introduce the discrete space
$$\mathcal{V}(\mathcal{M}, \underline{r}) := \{ \varphi \in L^2(0, T) : \varphi|_{I_m} \in P^{r_m}(I_m), 1 \leq m \leq M \}. \quad (2.1)$$ Note that functions in $\mathcal{V}(\mathcal{M}, \underline{r})$ can be discontinuous across the nodes $\{t_m\}$.

We consider the following discontinuous Galerkin approximation of the Volterra
integro-differential equation in (1.1): find $U \in \mathcal{V}(\mathcal{M}, \underline{r})$ such that
$$B_{DG}(U, V) = F_{DG}(V) \quad (2.2)$$
for all \( V \in \mathcal{V}(\mathcal{M}, \mathcal{L}) \).

The forms \( B_{DG} \) and \( F_{DG} \) are given by

\[
B_{DG}(U, V) := \sum_{m=1}^{M} \int_{I_m} \left( U'(t) + a(t)U(t) \right)V(t) \, dt \\
+ \sum_{m=1}^{M} \int_{I_m} \left( \int_{t_m}^{t} k_\alpha(t-s)b(s)U(s) \, ds \right) V(t) \, dt \\
+ \sum_{m=1}^{M-1} \left[ U \right]_{m} V_{m}^{+} + U_{0}^{+} V_{0}^{+},
\]

\[
F_{DG}(V) := u_{0}V_{0}^{+} + \sum_{m=1}^{M} \int_{I_m} f(t)V(t) \, dt.
\]

Note that the exact solution \( u \) of problem (1.1) satisfies \( B_{DG}(u, V) = F_{DG}(V) \) for all \( V \in \mathcal{V}(\mathcal{M}, \mathcal{L}) \). Hence, we have the Galerkin orthogonality property:

\[
B_{DG}(u - U, V) = 0
\] (2.3)

for all \( V \in \mathcal{V}(\mathcal{M}, \mathcal{L}) \).

**Remark 2.1.** The discontinuous Galerkin discretization in (2.2) is a time-stepping scheme: if \( U \) is given on \( I_{n} \), \( 1 \leq n \leq m - 1 \), we find \( U|_{I_{n}} \in \mathcal{P}^{r_{m}}(I_{m}) \) by solving

\[
\int_{I_m} \left( U'(t) + a(t)U(t) \right)V(t) \, dt + \int_{I_m} \left( \int_{t_m}^{t} k_\alpha(t-s)b(s)U(s) \, ds \right) V(t) \, dt + U_{m-1}^{+} V_{m-1}^{+}
= U_{m-1}^{+} V_{m-1}^{+} + \int_{I_m} f(t)V(t) \, dt - \int_{I_m} \left( \int_{0}^{t_m} k_\alpha(t-s)b(s)U(s) \, ds \right) V(t) \, dt
\]

for all \( V \in \mathcal{P}^{r_{m}}(I_{m}) \). Here, we set \( U_{0}^{-} = u_{0} \).

### 2.2. Existence and uniqueness of discrete solutions

To show that the DG time-stepping method (2.2) defines a unique approximate solution \( U \in \mathcal{V}(\mathcal{M}, \mathcal{L}) \), we make use of the discrete Gronwall inequality from [10, Lemma 6.4].

**Lemma 2.2.** Let \( \mathcal{M} = \{I_{m}\}_{m=1}^{M} \) be a partition of \((0, T)\) with \( k = \max_{m=1}^{M} \{k_{m}\} \). Let \( \{a_{m}\}_{m=1}^{M} \) and \( \{b_{m}\}_{m=1}^{M} \) be sequences of numbers with \( 0 \leq b_{1} \leq b_{2} \leq \ldots \leq b_{M} \). Assume that there is a constant \( K \geq 0 \) such that

\[
a_{1} \leq b_{1}, \quad a_{m} \leq b_{m} + K \sum_{n=1}^{m} w_{m,n}(\alpha)a_{n}, \quad m = 2, \ldots, M,
\]

where \( w_{m,n}(\alpha) = \int_{I_n} (t_{m} - t)^{-\alpha} \, dt \). Assume further that \( \delta = \frac{Kk^{1-\alpha}}{1-\alpha} < 1 \). Then we have

\[
a_{m} \leq Cb_{m}, \quad m = 1, \ldots, M,
\]

with a constant \( C > 0 \) that solely depends on \( \delta, K, \alpha \) and \( T \).

Furthermore, we recall the following technical result from [10, Lemma 6.3].
Lemma 2.3. For \( f \in L^2(0, \tau) \) and \( \alpha \in (0, 1) \) there holds
\[
\int_0^\tau \left( \int_0^t (t-s)^{-\alpha} f(s) \, ds \right)^2 \, dt \leq \frac{\tau^{1-\alpha}}{(1-\alpha)} \int_0^\tau (\tau-t)^{-\alpha} \left( \int_0^t f(s)^2 \, ds \right) \, dt.
\]

We now address the existence and uniqueness of discrete solutions.

Proposition 2.4. Let \((\mathcal{M}, \mathcal{P})\) be an \(hp\)-discretization of \((0, T)\) with
\[
(\mu^*/\mu_*)^{2(Tk)^{(1-\alpha)}}/(1-\alpha)^2 < 1. \tag{2.4}
\]

Then the discrete problem (2.2) has a unique solution \( U \in \mathcal{V}(\mathcal{M}, \mathcal{P}) \).

Remark 2.5. Note that condition (2.4) is independent of the degree vector \( \mathcal{P} \).

Proof. We first show the uniqueness of DG solutions. To this end, let \( U \) and \( \bar{U} \) be two solutions of (2.2). The difference \( E = U - \bar{U} \) then satisfies
\[
\int_{I_m} \left( E' + aE \right) V \, dt + E_{m-1}^+ E_{m-1}^+ = E_{m-1}^+ E_{m-1}^+ - \int_{I_m} \left( \int_0^t k_\alpha(t-s)b(s)E(s) \, ds \right) V(t) \, dt,
\]
for any \( V \in \mathcal{P}^r(I_m) \), \( m = 1, \ldots, M \). Selecting \( V = E \) yields
\[
\frac{1}{2} (E_m^-)^2 + \frac{1}{2} (E_{m-1}^+)^2 + \int_{I_m} aE^2 \, dt = E_{m-1}^+ E_{m-1}^- - \int_{I_m} \left( \int_0^t k_\alpha(t-s)b(s)E(s) \, ds \right) E(t) \, dt.
\]

Since
\[
E_{m-1}^+ E_{m-1}^- \leq \frac{1}{2} (E_{m-1}^-)^2 + \frac{1}{2} (E_{m-1}^+)^2,
\]
we have
\[
\frac{1}{2} (E_m^-)^2 + \int_{I_m} aE^2 \, dt \leq \frac{1}{2} (E_{m-1}^-)^2 + \int_{I_m} \left( \int_0^t k_\alpha(t-s)|b(s)|E(s)| \, ds \right) |E(t)| \, dt.
\]

In view of \( E_0^- = 0 \), iterating the above estimate yields
\[
\frac{1}{2} (E_m^-)^2 + \int_0^{t_m} aE^2 \, dt \leq \int_0^{t_m} \left( \int_0^t k_\alpha(t-s)|b(s)|E(s)| \, ds \right) |E(t)| \, dt =: S_m, \tag{2.5}
\]
for \( 1 \leq m \leq M \). By invoking the bounds for \( a \) and \( b \) in (1.2), the Cauchy-Schwarz inequality and Lemma 2.3, the integral \( S_m \) in (2.5) can be bounded by
\[
S_m \leq \mu^* \mu_*^{-1/2} \left( \int_0^{t_m} \left( \int_0^t k_\alpha(t-s)|E(s)| \, ds \right)^2 \, dt \right)^{1/2} \left( \int_0^{t_m} aE^2 \, dt \right)^{1/2}
\leq \frac{1}{2} (\mu^*/\mu_*)_{1/2} \left( \int_0^{t_m} (t_m-t)^{-(1-\alpha)} \left( \int_0^t E(s)^2 \, ds \right) \, dt + \frac{1}{2} \int_0^{t_m} aE^2 \, dt \right)
\leq \frac{1}{2} (\mu^*/\mu_*)^{(1-\alpha)}(1-\alpha) \int_0^{t_m} (t_m-t)^{-(1-\alpha)} \left( \int_0^t a(s)E(s)^2 \, ds \right) \, dt + \frac{1}{2} \int_0^{t_m} aE^2 \, dt.
\]
Hence, we obtain
\[
\frac{1}{2} \int_0^{t_m} aE^2 \, dt \leq \frac{1}{2} \left( \mu^*/\mu_s \right)^2 \frac{T(1-\alpha)}{(1-\alpha)} \sum_{n=1}^{m} \left( \int_{I_n} (t_m - t)^{-\alpha} \, dt \right) \left( \int_0^{t_m} aE^2 \, ds \right) .
\]

Setting \( a_m = \int_0^{t_m} aE^2 \, dt \) and \( b_m = 0 \), the Gronwall inequality in Lemma 2.2 gives
\[
\int_0^{t_m} aE^2 \, dt = 0, \quad m = 1, \ldots, M,
\]
provided that (2.4) is satisfied. The boundedness of \( a \) thus shows that \( E \equiv 0 \) and \( U \equiv \bar{U} \).

As problem (2.2) is linear and finite dimensional, the existence of solutions follows from their uniqueness. This completes the proof. \( \square \)

3. Error analysis. In this section, we derive \( hp \)-version error bounds for the DG time-stepping method in (2.2).

3.1. Abstract error bounds. We start be showing abstract error bounds. To this end, for a continuous function \( u : [0, T] \to \mathbb{R} \), we define the interpolant \( \mathcal{I} u \in \mathcal{V}(M, r) \) by
\[
\begin{align*}
(\mathcal{I} u)^-_m &= u^-_m, \quad 1 \leq m \leq M, \\
\int_{I_m} \mathcal{I} u(t) V'(t) \, dt &= \int_{I_m} u(t) V'(t) \, dt, \quad V \in \mathcal{P}^r_m(I_m), \ 1 \leq m \leq M.
\end{align*}
\]

Remark 3.1. The same interpolant has been used in the \( h \)-version analysis in [10]; we also refer to [18] and the references cited therein in the context of parabolic problems. The \( hp \)-approximation properties of \( \mathcal{I} \) have been thoroughly investigated in [14, 15] and will be used in Section 3.2 below.

Let now \( u \) be the exact solution of (1.1) and \( U \in \mathcal{V}(M, r) \) the DG approximation in (2.2). We split the error \( e = u - U \) into \( e = \eta + \theta \) with \( \eta := u - \mathcal{I} u \) and \( \theta := \mathcal{I} u - U \). Using Galerkin orthogonality in (2.3) and the construction of \( \mathcal{I} u \), the function \( \theta \) satisfies
\[
\begin{align*}
\int_{I_m} \left( \theta' + a \theta \right) V \, dt + \theta^+_m V^+_m - \theta^-_{m-1} V^-_{m-1} &= \int_{I_m} a \eta V \, dt \\
- \int_{I_m} \left( \int_0^t k_\alpha(t-s)b(s)\eta(s) \, ds \right) V(t) \, dt - \int_{I_m} \left( \int_0^t k_\alpha(t-s)b(s)\theta(s) \, ds \right) V(t) \, dt,
\end{align*}
\]
for any \( V \in \mathcal{P}^r_m(I_m) \) and \( m = 1, \ldots, M \).

Our first result establishes an \( L^2 \)-control of \( \theta \) in terms of \( \eta \).

Lemma 3.2. Let \( (M, r) \) be an \( hp \)-discretization of \( (0, T) \) with
\[
\delta = 3(\mu^*/\mu_s)^2 \frac{(Tk)^{(1-\alpha)}}{(1-\alpha)^2} < 1.
\]

Then we have
\[
\frac{1}{2} \int_0^{t_m} a \theta^2 \, dt + \frac{1}{2} (\theta^-_m)^2 \leq C \int_0^{t_m} a \eta^2 \, dt, \quad m = 1, \ldots, M,
\]
with a constant $C > 0$ that solely depends on $\mu_*, \mu^*, \alpha, T,$ and $\delta$ in (3.4).

**Remark 3.3.** Note that assumption (3.4) is slightly stronger than that in (2.4) and thus implies the existence and uniqueness of discrete solutions.

**Proof.** We select $V = \theta$ in (3.3). This yields

$$
\frac{1}{2} (\theta_m^*)^2 + \frac{1}{2} (\theta_{m-1}^*)^2 + \int_{I_m} a\theta^2 \, dt = \theta_m^+ - \theta_{m-1}^+ - \int_{I_m} a\eta \theta \, dt
$$

$$
- \int_{I_m} \left( \int_0^t k_\alpha(t-s)b(s)\eta(s) \, ds \right) \theta(t) \, dt - \int_{I_m} \left( \int_0^t k_\alpha(t-s)b(s)\theta(s) \, ds \right) \theta(t) \, dt.
$$

Since

$$
\theta_m^+ \theta_{m-1}^- \leq \frac{1}{2} (\theta_m^*)^2 + \frac{1}{2} (\theta_{m-1}^*)^2,
$$

we obtain

$$
\frac{1}{2} (\theta_m^*)^2 + \int_{I_m} a\theta^2 \, dt \leq \frac{1}{2} (\theta_{m-1}^*)^2 + \int_{I_m} a\eta \theta \, dt
$$

$$
+ \int_{I_m} \left( \int_0^t k_\alpha(t-s)b(s)\eta(s) \, ds \right) |\theta(t)| \, dt
$$

$$
+ \int_{I_m} \left( \int_0^t k_\alpha(t-s)b(s)\theta(s) \, ds \right) |\theta(t)| \, dt.
$$

Iterating this estimate gives

$$
\frac{1}{2} (\theta_m^*)^2 + \int_{0}^{t_m} a\theta^2 \, dt \leq T_1 + T_2 + T_3,
$$

where

$$
T_1 = \int_{0}^{t_m} a|\eta\theta| \, dt,
$$

$$
T_2 = \int_{0}^{t_m} \left( \int_0^t k_\alpha(t-s)|b(s)\eta(s)| \, ds \right) |\theta(t)| \, dt,
$$

$$
T_3 = \int_{0}^{t_m} \left( \int_0^t k_\alpha(t-s)|b(s)\theta(s)| \, ds \right) |\theta(t)| \, dt.
$$

We estimate each of the above terms separately.

First, we note that

$$
T_1 \leq \frac{3}{2} \int_{0}^{t_m} a\eta^2 \, dt + \frac{1}{6} \int_{0}^{t_m} a\theta^2 \, dt.
$$

Next, using the bounds for $a$ and $b$ in (1.2), the Cauchy-Schwarz inequality, and Lemma 2.3, we have

$$
T_2 \leq \frac{\mu_*^{1/2}}{\mu^*} \left( \int_{0}^{t_m} \left( \int_0^t k_\alpha(t-s)|\eta(s)| \, ds \right)^2 \, dt \right)^{1/2} \left( \int_{0}^{t_m} a\theta^2 \, dt \right)^{1/2}
$$

$$
\leq \frac{3}{2} \left( \mu^*/\mu_* \right)^{1/2} \frac{T_m^{1-\alpha}}{(1-\alpha)} \int_0^{t_m} (t_m - t)^{-\alpha} \left( \int_0^t a(s)\eta(s)^2 \, ds \right) \, dt + \frac{1}{6} \int_{0}^{t_m} a\theta^2 \, dt
$$

$$
\leq \frac{3}{2} \left( \mu^*/\mu_* \right)^{1/2} \frac{T_m^{2(1-\alpha)}}{(1-\alpha)^2} \int_{0}^{t_m} a\eta^2 \, ds + \frac{1}{6} \int_{0}^{t_m} a\theta^2 \, dt.
$$
Analogously, we obtain
\[ T_3 \leq \frac{3}{2} \left( \mu^* / \mu_* \right)^2 \frac{T^{1-\alpha}}{(1-\alpha)} \sum_{n=1}^{m} \left( \int_{I_n} (t_m - t)^{-\alpha} \, dt \right) \left( \int_{0}^{t_m} a \theta^2 \, ds \right) + \frac{1}{6} \int_{0}^{t_m} a \theta^2 \, dt \]
\[ \leq \frac{3}{2} \left( \mu^* / \mu_* \right)^2 \frac{T^{1-\alpha}}{(1-\alpha)} \sum_{n=1}^{m} \left( \int_{I_n} (t_m - t)^{-\alpha} \, dt \right) \left( \int_{0}^{t_m} a \theta^2 \, ds \right) + \frac{1}{6} \int_{0}^{t_m} a \theta^2 \, dt. \]
Combining the above estimates results in
\[ \frac{1}{2} (\theta_m^-)^2 + \frac{1}{2} \int_{0}^{t_m} a \theta^2 \, dt \leq \max \left\{ \frac{3}{2} \left( \frac{3}{2} \left( \frac{\mu^* / \mu_*} {1-\alpha} \right)^2 \right) \right\} \int_{0}^{t_m} a \eta^2 \, dt \]
\[ + \frac{3}{2} \left( \mu^* / \mu_* \right)^2 \frac{T^{1-\alpha}}{(1-\alpha)} \sum_{n=1}^{m} \left( \int_{I_n} (t_m - t)^{-\alpha} \, dt \right) \left( \int_{0}^{t_m} a \theta^2 \, ds \right). \]

Setting
\[ a_m = \int_{0}^{t_m} a \theta^2 \, dt, \]
\[ b_m = \max \left\{ 3, 3 \left( \mu^* / \mu_* \right)^2 \frac{T^{2(1-\alpha)}}{(1-\alpha)^2} \right\} \int_{0}^{t_m} a \eta^2 \, dt, \]
the assertion follows from Lemma 2.2. \( \Box \)

Next, we bound the derivative of \( \theta \) as follows.

**Lemma 3.4.** We have
\[ \int_{t_m}^{t} |\theta'|^2 (t - t_{m-1}) \, dt \leq C k_m \int_{0}^{t_m} a (\theta^2 + \eta^2) \, dt, \quad m = 1, \ldots, M, \]
with a constant \( C > 0 \) that solely depends on \( \mu_*, \mu^*, \alpha, \) and \( T. \)

**Proof.** We choose \( V(t) = \theta'(t)(t - t_{m-1}) \) in (3.3) and obtain
\[ \int_{t_m}^{t} |\theta'|^2 (t - t_{m-1}) \, dt \leq T_1 + T_2 + T_3 + T_4, \]
where
\[ T_1 = \int_{t_m}^{t} a |\theta' (t - t_{m-1})| \, dt, \]
\[ T_2 = \int_{t_m}^{t} a |\eta' (t - t_{m-1})| \, dt, \]
\[ T_3 = \int_{t_m}^{t} \left( \int_{0}^{t} k_\alpha(t - s) |b(s) \eta(s)| \, ds \right) |\theta' (t - t_{m-1})| \, dt, \]
\[ T_4 = \int_{t_m}^{t} \left( \int_{0}^{t} k_\alpha(t - s) |b(s) \theta(s)| \, ds \right) |\theta' (t - t_{m-1})| \, dt. \]

Clearly, using the bounds for \( a \) in (1.2),
\[ T_1 \leq (\mu^*)^{1/2} \left( \int_{t_m}^{t} a \theta^2 \, dt \right)^{1/2} k_m \left( \int_{t_m}^{t} |\theta'|^2 (t - t_{m-1}) \, ds \right)^{1/2}, \]
\[ T_2 \leq (\mu^*)^{1/2} \left( \int_{t_m}^{t} a \eta^2 \, dt \right)^{1/2} k_m \left( \int_{t_m}^{t} |\theta'|^2 (t - t_{m-1}) \, ds \right)^{1/2}. \]
Furthermore, by Lemma 2.3,

\[ T_3 \leq \mu^* \left( \int_{t_m}^t \left( \int_0^t (t-s)^\alpha |\eta(s)| ds \right)^2 dt \right)^{\frac{1}{2}} \left( \int_{t_m}^t |\theta'|^2 (t-t_m-1) ds \right)^{\frac{1}{2}} \]

\[ \leq \mu^* \mu^*^{-1/2} \frac{T^{1-\alpha}}{1-\alpha} \left( \int_0^t a \eta^2 dt \right)^{\frac{1}{2}} \left( \int_{t_m}^t |\theta'|^2 (t-t_m-1) ds \right)^{\frac{1}{2}}. \]

Analogously, we obtain

\[ T_4 \leq \mu^* \mu^*^{-1/2} \frac{T^{1-\alpha}}{1-\alpha} \left( \int_0^t a \eta^2 dt \right)^{\frac{1}{2}} \left( \int_{t_m}^t |\theta'|^2 (t-t_m-1) ds \right)^{\frac{1}{2}}. \]

Combining these estimates results in

\[ \int_{t_m}^t |\theta'|^2 (t-t_m-1) dt \leq C k_m \int_0^{t_m} a (\eta^2 + \eta^2) dt. \]

This completes the proof. \( \square \)

To control the \( L^\infty \)-norm of \( \theta \) in terms of the interpolation error \( \eta \), we make use of the following inverse inequality from [14, Lemma 3.1]:

**Lemma 3.5.** On each interval \( I_m \) there holds

\[ \| \varphi \|_{L^\infty(I_m)}^2 \leq C \left( \log \left( \max \{r_m, 2\} \right) \int_{t_m}^t |\varphi'(t)|^2 (t-t_m-1) dt + \left( \varphi_m^- \right)^2 \right), \]

for any \( \varphi \in \mathcal{P}^r_m(I_m), r_m \geq 0 \). The constant \( C > 0 \) is independent of \( k_m \) and \( r_m \).

Furthermore, the estimate cannot be improved asymptotically as \( r_m \to \infty \).

The following result states an abstract error bound.

**Theorem 3.6.** Let \( (\mathcal{M}, T) \) be an \( hp \)-discretization of \( (0, T) \) satisfying (3.4). Then the error \( u - U \) between the exact solution \( u \) and the DG approximation \( U \) satisfies

\[ \| u - U \|_{L^2(0,T)} \leq C \| u - I u \|_{L^2(0,T)} \]

and

\[ \| u - U \|_{L^\infty(0,T)} \leq C \log \left( \max \{L, 2\} \right) \| u - I u \|_{L^\infty(0,T)}, \]

with a constant \( C > 0 \) that solely depends on \( \mu^* \), \( \mu^* \), \( \alpha \), \( T \), and \( \delta \) in (3.4).

**Proof.** As before, we split the error into \( u - U = \eta + \theta \). Lemma 3.2 and Lemma 3.4 yield

\[ \left( \varphi^- \right)^2 + \int_0^{t_m} a \eta^2 dt + \int_0^{t_m} |\theta'|^2 (t-t_m-1) dt \leq C \int_0^{t_m} a \eta^2 dt. \]

In view of the boundedness of \( a \) in (1.2), we obtain \( \| \theta \|_{L^2(0,T)} \leq C \| \eta \|_{L^2(0,T)} \). Furthermore, by Lemma 3.5,

\[ \| \theta \|_{L^\infty(I_m)}^2 \leq C \log \left( \max \{L, 2\} \right) \| \eta \|_{L^2(0,T)}^2 \leq C \log \left( \max \{L, 2\} \right) \| \eta \|_{L^\infty(0,T)}^2, \]

for \( 1 \leq m \leq M \). The error bounds follow from the triangle inequality. \( \square \)
3.2. Error bounds. In this section, we employ the $hp$-version approximation properties of the interpolant $\mathcal{I}$ to make explicit the error bounds in Theorem 3.6.

We first recall the following results from [15, Theorem 3.10] and [14, Corollary 3.10]. We denote by $\Gamma$ the Gamma function.

**Theorem 3.7.** Let $u|_{I_m} \in H^{s_m+1}(I_m)$ for $s_m \geq 0$. Then

$$
\|u - \mathcal{I}u\|_{L^2(I_m)}^2 \leq C \left( \frac{k_m}{2} \right)^{2t_m+2} \frac{1}{\max\{1, r_m^2\}} \frac{\Gamma(r_m + 1 - t_m)}{\Gamma(r_m + 1 + t_m)} \|u\|_{H^{t_m+1}(I_m)}^2,
$$

for any real $0 \leq t_m \leq \min\{r_m, s_m\}$. The constant $C > 0$ is independent of $k_m$, $r_m$, $t_m$, and $s_m$. Moreover, if $u|_{I_m} \in W^{s_m+1, \infty}(I_m)$ for $s_m \geq 0$, then

$$
\|u - \mathcal{I}u\|_{L^\infty(I_m)}^2 \leq C \left( \frac{k_m}{2} \right)^{2t_m+2} \frac{\Gamma(r_m + 1 - t_m)}{\Gamma(r_m + 1 + t_m)} \|u\|_{W^{t_m+1, \infty}(I_m)}^2,
$$

for any real $0 \leq t_m \leq \min\{r_m, s_m\}$.

From Theorem 3.6 and Theorem 3.7 we obtain the following $hp$-error estimates.

**Theorem 3.8.** Let $(\mathcal{M}, \mathcal{P})$ be an $hp$-discretization of $(0, T)$ satisfying (3.4), and let $U \in V(\mathcal{M}, \mathcal{P})$ be the DG approximation (2.2). Let the exact solution $u$ of (1.1) satisfy

$$
u|_{I_m} \in H^{s_m+1}(I_m), \quad s_m \geq 0, \quad m = 1, \ldots, M.
$$

Then we have the $L^2$-error bound

$$
\|u - U\|_{L^2(0,T)}^2 \leq C \sum_{m=1}^{M} \left( \left( \frac{k_m}{2} \right)^{2t_m+2} \frac{1}{\max\{1, r_m^2\}} \frac{\Gamma(r_m + 1 - t_m)}{\Gamma(r_m + 1 + t_m)} \|u\|_{H^{t_m+1}(I_m)}^2 \right)
$$

for any real $0 \leq t_m \leq \min\{s_m, r_m\}, 1 \leq m \leq M$. Moreover, if

$$
u|_{I_m} \in W^{s_m+1, \infty}(I_m), \quad s_m \geq 0, \quad m = 1, \ldots, M,
$$

then we have the $L^\infty$-error bound

$$
\|u - U\|_{L^\infty(0,T)}^2 \leq C \log(\max \{|2|, 2\}) \cdot \sum_{m=1}^{M} \left( \left( \frac{k_m}{2} \right)^{2t_m+2} \frac{\Gamma(r_m + 1 - t_m)}{\Gamma(r_m + 1 + t_m)} \|u\|_{W^{t_m+1, \infty}(I_m)}^2 \right),
$$

for any real $0 \leq t_m \leq \min\{s_m, r_m\}, 1 \leq m \leq M$.

The constants $C > 0$ solely depend on $\mu_*, \mu^*, T, \alpha$ and $\delta$ in (3.4).

We remark that the estimates in Theorem 3.8 are explicit in the time-steps $k_m$, the polynomial degrees $r_m$, and the regularity exponents $s_m$ of the exact solution. From the bounds in Theorem 3.8, the following convergence rates can be deduced for the $h$- and $p$-version of the DG time-stepping method.

**Corollary 3.9.** Let $(\mathcal{M}, \mathcal{P})$ be an $hp$-discretization of $(0, T)$ satisfying (3.4), with uniform polynomial degree $r \geq 0$. Let $u$ be the exact solution of (1.1) and $U$ the discontinuous Galerkin approximation (2.2). If $u \in H^{s+1}(0,T)$ for $s \geq 0$, we have the $L^2$-error bound

$$
\|u - U\|_{L^2(0,T)} \leq C \frac{k_m^{\min(s,r)+1}}{p^{s+1}} \|u\|_{H^{s+1}(0,T)}.
$$
Additionally, if $u \in W^{s+1, \infty}(0, T)$ for $s \geq 0$, we have the $L^\infty$-error bound

$$
\|u - U\|_{L^\infty(0, T)} \leq C \log \left( \max \{r, 2\} \right)^{k_{\min}} \|u\|_{W^{s+1, \infty}(0, T)}.
$$

The constants $C > 0$ solely depend on $\mu_*, \mu^*, T, \alpha, \delta$ in (3.4), and the regularity exponent $s$.

Proof. This follows from Theorem 3.8 and Stirling’s formula; cf. [17].

The estimates in Corollary 3.9 show that the DG time-stepping method converges either as the time-steps are decreased ($k \to 0$) or as $r$ is increased ($r \to \infty$). Both estimates are optimal in $k$. However, while the $L^2$-estimate is also optimal in the polynomial degree $r$, the $L^\infty$-estimate is one power of $r$ short from being optimal; this is due to the slightly suboptimal $L^\infty$-approximation properties of the interpolant $I$ in Theorem 3.7; see also [14].

It can be seen from Corollary 3.9 that for solutions $u$ for which $s$ is large it is more advantageous to increase $r$ rather than to reduce $k$ at fixed, low $r$. Indeed, if $u$ is smooth on $[0, T]$, arbitrarily high algebraic convergence rates are possible if the polynomial degree $r$ is raised. This is referred to as spectral convergence. Moreover, the $p$-version of the DG time-stepping method converges exponentially if the solution $u$ is analytic on $[0, T]$. To see this, we first recall the following result.

Lemma 3.10. On each interval $I_m$ there holds

$$
\|u - Iu\|_{L^2(I_m)} \leq C \inf_{q \in P^{r_m}(I_m)} \|u - q\|_{H^1(I_m)},
$$

$$
\|u - Iu\|_{L^\infty(I_m)} \leq Cr_m \inf_{q \in P^{r_m}(I_m)} \|u - q\|_{W^{1, \infty}(I_m)},
$$

with a constant $C > 0$ independent of $I_m$, $r_m$ and $u$.

Proof. The first estimate follows from [16, Lemma 3.6] and a scaling argument. The second estimate follows similarly from [14, Lemma 3.8].

Theorem 3.11. Let $(\mathcal{M}, r)$ be an $hp$-discretization of $(0, T)$ satisfying (3.4), with polynomial degree $r \geq 0$. Let the exact solution $u$ of (1.1) be analytic on $[0, T]$. For the DG approximation (2.2), we then have the error bound

$$
\|u - U\|_{L^p(0, T)} \leq C \exp(-br), \quad p = 2 \text{ or } p = \infty,
$$

with constants $C, b > 0$ that are independent of $r$.

Proof. The assertion follows from Theorem 3.6, the results in Lemma 3.10 and standard approximation theory for analytic functions.

4. Exponential convergence for analytic data. The exponential convergence result in Theorem 3.11 is valid for solutions that are analytic in $[0, T]$. However, this regularity assumption is unrealistic since, as discussed previously, solutions of (1.1) with analytic data have strong start-up singularities, due to the presence of the weakly singular kernel $k_\alpha$, and are only analytic away from $t = 0$. In this section we show that, in spite of this singular behavior, the $hp$-version of the DG method with geometrically graded time-steps near $t = 0$ yields exponential rates of convergence.

4.1. Analyticity of solutions. Let $A(0, T)$ denote the space of the functions which are analytic on $[0, T]$. A function $g$ in $A(0, T)$ can be characterized by analyticity constants $C_g, d_g > 0$ and the growth conditions (see [17, pp. 78-79] for details)

$$
|g^{(s)}(t)| \leq C_g d_g^s \Gamma(s + 1), \quad t \in [0, T], \quad s \geq 0.
$$
We assume the data $a$, $b$, and $f$ to satisfy
\begin{align}
  a, b &\in A(0, T), \\
  f(t) = f_1(t) + t^\beta f_2(t), &\quad f_i \in A(0, T), \; i = 1, 2, \; \beta > 0, \; \beta \not\in \mathbb{N}. 
\end{align}

The following result describes the analyticity properties of the exact solution $u$.

**Theorem 4.1.** Assume (4.1)-(4.2) and let \( \theta = \min\{2 - \alpha, 1 + \beta\} \). Then there exist constants $C, d > 0$ depending only on the analyticity constants of $a$, $b$, $f_1$, and $f_2$, such that the solution $u$ of (1.1) satisfies
\[ |u^{(s)}(t)| \leq C d^s \Gamma(s + 1)t^{\theta - s}, \quad t \in (0, T], \; s \in \mathbb{N}. \]

**Proof.** This regularity result slightly generalizes earlier results in [4]; see also [12, 3] and [2]. We give a brief sketch of the proof; additional details can be found in [2, Section 7.1].

The initial-value problem for the given Volterra integro-differential equation (1.1) is equivalent to the second-kind Volterra integral equation
\begin{equation}
  u(t) = g(t) + \int_0^t h_\alpha(t, s)b(s)u(s) \, ds, \quad t \in [0, T],
\end{equation}
with
\[ g(t) := u_0 + \int_0^t (f_1(s) + s^\beta f_2(s)) \, ds, \]
\[ h_\alpha(t, s) = -a(s) - \int_s^t k_\alpha(v - s)dv. \]
In particular, if $a(t) = a > 0$, $b(t) = \lambda > 0$, $f_i(t) = f_i = \text{const}$ for $t \in [0, T]$, then we have
\[ g(t) = u_0 + f_1 t + \frac{f_2}{1 + \beta} t^{1 + \beta}, \]
\[ h_\alpha(t, s) = -a - \frac{\lambda}{1 - \alpha} (t - s)^{1 - \alpha}. \]

The resolvent kernel $R_\alpha(t, s)$ associated with the kernel
\[ K_\alpha(t, s) := h_\alpha(t - s)b(s) \quad (t, s) \in D := \{(t, s) : 0 \leq s \leq t \leq T\}, \]
has the form
\[ R_\alpha(t, s) = (t - s)^{1 - \alpha} Q_\alpha(t, s), \quad (t, s) \in D. \]
Here,
\[ Q_\alpha(t, s) := \sum_{n=1}^{\infty} (t - s)^{(a-1)/2} \Phi_n(t, s; \alpha), \]
where the series is uniformly convergent on $D$ for all $\alpha \in (0, 1)$. If the given data $a$ and $b$ are in $A(0, T)$ then we have $\Phi_n(\cdot, \cdot; \alpha) \in A(D)$ for all $\alpha \in (0, 1)$. Here, $A(D)$ denotes the space of the functions that are analytic on $D$. 
Since the (unique) solution of the Volterra integral equation (4.3) is given by

$$u(t) = g(t) + \int_0^t R_\alpha(t,s)g(s)\,ds, \quad t \in [0,T],$$

(4.4)

the regularity properties of the nonhomogeneous term $g$ imply the asserted bounds for $u^{(k)}(t)$ on $(0,T)$. 

4.2. Exponential convergence for analytic data. In this section, we show that, under the analyticity assumption in (4.1)-(4.2), the $hp$-version of the DG time-stepping method leads to exponential rates of convergence.

We start with the following definition.

**Definition 4.2.** The basic geometric partition $\hat{M}_{n,\sigma} = \{I_m\}_{m=1}^{n+1}$ of $\hat{J} = (0,1)$ with grading factor $\sigma \in (0,1)$ and $n$ levels of refinement is given by

$$t_0 = 0, \quad t_m = \sigma^{n-m+1}, \quad 1 \leq m \leq n+1.$$

Away from $t = 0$, i.e., for $2 \leq m \leq n+1$, the intervals $I_m \in \hat{M}_{n,\sigma}$ satisfy

$$k_m = t_m - t_{m-1} = \lambda t_{m-1}, \quad \lambda := \sigma^{-1}(1-\sigma).$$

(4.5)

**Definition 4.3.** A geometric partition $M_{n,\sigma}$ of $(0,T)$ with grading factor $\sigma \in (0,1)$ and $n$ levels of refinement is obtained by first quasi-uniformly partitioning $(0,T)$ into intervals $\{J_k\}_{k=1}^K$. The first interval $J_1 = (0,t_1)$ near $t = 0$ is then further subdivided into $n+1$ subintervals $\{I_m\}_{m=1}^{n+1}$, by linearly mapping to basic geometric mesh $\hat{M}_{n,\sigma}$ in Definition 4.2 onto $J_1$.

An illustration of a geometric partition $M_{n,\sigma}$ is given in Figure 4.1. We point out that the coarse intervals $\{J_k\}_{k=3}^K$ will be kept fixed; convergence will be achieved there by increasing the polynomial degrees.

**Lemma 4.4.** Assume (4.1)-(4.2) and set $\theta = \min\{2-\alpha,1+\beta\}$. Let $M_{n,\sigma}$ be a geometric mesh of $(0,T)$ with $\{J_k\}_{k=1}^K$ denoting the underlying quasi-uniform partition of $(0,T)$ and $\{I_m\}_{m=1}^{n+1}$ the geometric refinement of $J_1$. Then the solution $u$ of (1.1) satisfies

$$\|u\|_{W^{1,\infty}(I_1)}^2 \leq C,$$

and

$$\|u\|_{W^{s+1,\infty}(I_m)}^2 \leq C d^{s}\Gamma(2s+1)\sigma^{2(n-m+2)(\theta-s-1)}, \quad 2 \leq m \leq n+1,$$

$$\|u\|_{W^{s+1,\infty}(I_k)}^2 \leq C d^{s}\Gamma(2s+1), \quad 2 \leq k \leq K,$$

for $s \geq 0$. The constants $C,d > 0$ are independent of $m, n$ and $s$.

**Remark 4.5.** We point out that the constants $C$ and $d$ in Lemma 4.4 depend on the underlying quasi-uniform partition $\{J_k\}_{k=1}^K$ of $M_{n,\sigma}$. 
Next, an element we denote by

on the first element the underlying quasi-uniform partition of the geometric re

with slope $\mu > 0$ if $r_m = \lfloor \mu m \rfloor$

on the geometrically refined elements $\{I_m\}_{m=1}^{n+1}$ and if $r_k = \lfloor \mu(n+1) \rfloor$ on the coarse elements $J_k$, $2 \leq k \leq K$, away from $t = 0$.

Our next result establishes exponential rates of convergence under the analyticity assumptions in (4.1) and (4.2).

**Theorem 4.7.** Assume (4.1)-(4.2). Let $\mathcal{M}_{n,\sigma}$ be a geometric partition of $(0,T)$ satisfying (3.4). Then there exists a slope $\mu_0 > 0$ solely depending on $\sigma$, $\alpha$, $\beta$ and the constants $C$ and $d$ in Lemma 4.4 such that for all linear polynomial degree vectors $\mathbf{c}$ with slope $\mu \geq \mu_0$ the DG approximation $U \in \mathcal{V}(\mathcal{M}_{n,\sigma}, \mathbb{L})$ satisfies the error estimate

$$\|u - U\|_{L^p(0,T)} \leq C \exp\left(-b N^{\frac{1}{2}}\right), \quad p = 2 \text{ or } p = \infty,$$

with constants $C, b > 0$ that are independent of $N = \dim(\mathcal{V}(\mathcal{M}_{n,\sigma}, \mathbb{L}))$.

**Proof.** We first note that

$$\|u - U\|_{L^2(0,T)} \leq \sqrt{T} \|u - U\|_{L^\infty(0,T)}.$$

In view of this inequality, we only need to prove the bound for the $L^\infty$-error. To do so, we denote by $\{J_k\}_{k=1}^K$ underlying quasi-uniform partition of $\mathcal{M}_{n,\sigma}$ and by $\{I_m\}_{m=1}^{n+1}$ the geometric refinement of the first-time step $J_1$ near $t = 0$. From Theorem 3.8 and Lemma 10.1, we find

$$\|u - U\|_{L^\infty(0,T)} \leq C \log \left( \max \{ \lfloor \mu(n+1) \rfloor, 2 \} \right) \max \left\{ \max_{m=1}^{n+1} e_m, \max_{k=2}^K e_k \right\},$$

with

$$e_m = \left( \frac{k_m}{2} \right)^{2t_m + 2} \frac{\Gamma(r_m + 1 - t_m)}{\Gamma(r_m + 1 + t_m)} \|u\|_{W^2(r_m + 1, \infty)}^2, \quad 1 \leq m \leq n + 1,$$

$$e_k = \inf_{q \in P^k_{\text{DG}}(I_k)} \|u - q\|_{W^1(0, \infty)}^2, \quad 2 \leq k \leq K,$$

and $0 \leq t_m \leq \min(s_m, r_m)$. Due to Theorem 4.1, $u$ is analytic away from $t = 0$ and, hence, the regularity exponents $s_m$ can be chosen arbitrarily large for $m = 2, \ldots, n+1$.

We first bound the errors $\{e_m\}$ on the geometrically refined elements $\{I_m\}_{m=1}^{n+1}$. On the first element $I_1$ near $t = 0$, we select $s_1 = t_1 = 0$ and have from Lemma 4.4

$$e_1 \leq C k_1^2 = C \sigma^{2n}.$$

Next, fix an element $I_m$, $2 \leq m \leq n + 1$, away from $t = 0$. From Lemma 4.4 and the definition of $\lambda$ in (4.5), we obtain

$$e_m \leq C \left( \frac{\lambda^{n-m+2}}{2} \right)^{2t_m + 2} \frac{\Gamma(r_m + 1 - t_m)}{\Gamma(r_m + 1 + t_m)} \left( \sigma^{n-m+2} \right)^{2(\theta - t_m - 1)} d^{2t_m} \Gamma(2t_m + 1)$$

$$= C \sigma^{(n-m+2)/2} \left( \frac{\lambda d}{2} \right)^{2t_m} \frac{\Gamma(r_m + 1 - t_m)}{\Gamma(r_m + 1 + t_m)} \Gamma(2t_m + 1).$$
Taking $t_m = \gamma_m r_m$ with $\gamma_m \in (0, 1)$, Stirling’s formula leads to
\[
e_m \leq C\sigma^{(n-m+2)/2} r_m^{1/2} \left( (\lambda d)^2 \gamma_m \left( \frac{(1 - \gamma_m)^{1-\gamma_m}}{(1 + \gamma_m)^{1+\gamma_m}} \right) \right)^{r_m}.
\]
The function $f_{\lambda,d}(\gamma) = (\lambda d)^2 \gamma \left( \frac{(1 - \gamma)^{1-\gamma}}{(1 + \gamma)^{1+\gamma}} \right)$ satisfies
\[
0 < \inf_{0 < \gamma < 1} f_{\lambda,d}(\gamma) =: f_{\lambda,d}(\gamma_{\min}) < 1 \quad \text{with } \gamma_{\min} = \frac{1}{\sqrt{1 + \lambda^2 d^2}}.
\]
Set $f_{\min} = f_{\min}(\lambda, d) = f_{\lambda,d}(\gamma_{\min})$ and select $\gamma_m = \gamma_{\min}$ for $2 \leq m \leq n + 1$. Hence, for $r_m = \lfloor \mu m \rfloor$, we have
\[
e_m \leq C\sigma^{(n-m+2)/2} r_m^{1/2} f_{\min}^r m \leq C\sigma^{(n-m+2)/2} (\mu m)^{1/2} f_{\min}^r m
\]
\[
\leq C\sigma^{2\theta n} \left( \mu(n + 1) \right)^{1/2} \left( \sigma^{(m+2)/2} f_{\min}^r m \right).
\]
Let
\[
\mu \geq \max \left\{ \frac{2\theta \log(\sigma)}{\log(f_{\min})}, 1 \right\}. \tag{4.6}
\]
Then, $f_{\min}^r m \leq \sigma^{2\theta m}$ and, consequently,
\[
e_m \leq C\sigma^{2\theta n} \left( \mu(n + 1) \right)^{1/2} \left( \sigma^{4\theta} \right) \leq C\sigma^{2\theta n} \left( \mu(n + 1) \right)^{1/2}, \quad m \geq 2.
\]
Thus, we obtain for $1 \leq m \leq n + 1$ the bound
\[
e_m \leq C \max \left\{ \sigma^{2n}, \sigma^{2\theta n} \left( \mu(n + 1) \right)^{1/2} \right\}. \tag{4.7}
\]

Further, from standard approximation properties for analytic functions, we can bound the errors $\{e_k\}$ on the elements $\{J_k\}_{k=2}^K$ away from $t = 0$ as follows:
\[
e_k \leq C e^{-b r_k} = C e^{-b \lfloor \mu(n + 1) \rfloor}, \quad 2 \leq k \leq K, \tag{4.8}
\]
with constants $C$ and $b$ that solely depend on the constants $C$ and $d$ in Lemma 4.4. Combining the estimates in (4.7) and (4.8) yields
\[
\|u - U\|_{L^\infty(0,T)}^2 \leq C \log \left( \max \{\mu(n + 1), 2\} \right) \max \left\{ \sigma^{2n}, \sigma^{2\theta n} \left( \mu(n + 1) \right)^{1/2}, e^{-b \lfloor \mu(n + 1) \rfloor} \right\}.
\]
Since we have
\[
\log \left( \max \{\mu(n + 1), 2\} \right) \max \left\{ \sigma^{2n}, \sigma^{2\theta n} \left( \mu(n + 1) \right)^{1/2}, (e^{-b})^{\lfloor \mu(n + 1) \rfloor} \right\} \leq C \exp(-bn),
\]
as $n \to \infty$, and $N = \dim(V_\nu) \leq C n^2$, the $L^\infty$-error bound follows. \[\]

**Remark 4.8.** From a practical point of view, it may be more convenient to use a fixed polynomial degree $r$ on a geometric partition $\mathcal{M}_{\mu, \sigma}$. In this case, exponential convergence results for all $\sigma \in (0, 1)$ provided that $r$ is proportional to the number of refinements, i.e., $r = \lfloor \mu(n + 1) \rfloor$ with the slope parameter $\mu$. Indeed, we see from the proof of Theorem 4.7, that
\[
\|u - U\|_{L^\infty(0,T)} \leq C \max(\sigma^{2n}, r^{1/2} f_{\min}) \leq C \exp(-br) \leq C \exp(-bN^{1/2}).
\]
Note that condition (4.6) on the slope is not necessary in this case.
5. Numerical experiments. In this section, we present a set of numerical experiments that confirm our theoretical error bounds. Throughout, we consider problem (1.1)–(1.3) with $T = 1$ and $a(t) = 1$, $b(t) = \exp(t)$, $u_0 = 0$.

We choose the right-hand side $f$ such that the solution $u$ of (1.1) is given by

$$u(t) = t^{2-\alpha} \exp(-t). \tag{5.1}$$

Notice that this solution is analytic away from $t = 0$ and that, for $\alpha \in (0, 1)$, the second derivative $u''$ is unbounded near $t = 0$. Thus, the solution (5.1) is ideally suited to test the performance of the $hp$-version DG method.

5.1. Smooth solution. We start by considering the case $\alpha = -1$ so that $u$ in (5.1) is analytic on $[0, 1]$.

In Figure 5.1, we show the errors in $L^\infty(0,1)$ that have been obtained for the $h$-version DG method on a sequence $\{M_i\}_{i=1}^9$ of equidistant time partitions with fixed polynomial degree $r = 1, \ldots, 5$. The partition $M_i$ consists of $2^i$ intervals of length $2^{-i}$. Hence, the straight error curves correspond to algebraic convergence in the time step $k$, for each polynomial degree. To illustrate this, we compute in Table 5.1 the numerical rates of convergence $\{\kappa_i\}$ given by

$$\kappa_i = \log\left(\frac{e(M_i)}{e(M_{i-1})}\right) / \log(0.5),$$

with $e(M_i)$ denoting the error on the partition $M_i$ measured in the $L^\infty$-norm. The convergence rates of order $r + 1$ are clearly visible, which confirms the $h$-version result in Corollary 3.9 for a smooth solution.

Next, let us consider the $p$-version of the DG time-stepping method. To that end, we increase the polynomial degree from $r = 1$ to $r = 50$ for fixed partitions with time-step length $k = 1$, $k = 0.5$, $k = 0.25$ and $k = 0.1$, respectively. The performance of the $p$-version method is displayed in Figure 5.2. For each of the fixed time partitions the results show that exponential rates of convergence are achieved, in agreement with
Discontinuous Galerkin time-stepping for Volterra integro-differential equations

<table>
<thead>
<tr>
<th>degree $r$</th>
<th>$\mathcal{M}_i$</th>
<th>error</th>
<th>order $\kappa_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>1.03e-05</td>
<td>1.9982</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>2.57e-06</td>
<td>1.9992</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>6.41e-07</td>
<td>1.9996</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>4.69e-08</td>
<td>2.9762</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>5.91e-09</td>
<td>2.9881</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>7.42e-10</td>
<td>2.9941</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>1.05e-10</td>
<td>3.9852</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>6.62e-12</td>
<td>3.9925</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>4.15e-13</td>
<td>3.9963</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>3.64e-12</td>
<td>4.9761</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>1.15e-13</td>
<td>4.9882</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>3.59e-15</td>
<td>4.9940</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>2.03e-11</td>
<td>5.9170</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3.27e-13</td>
<td>5.9585</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>5.17e-15</td>
<td>5.9793</td>
</tr>
</tbody>
</table>

Table 5.1

$h$-version: solution with $\alpha = -1$.

...the theoretical findings in Theorem 3.11 (remember that for $\alpha = -1$ the solution $u$ is analytic in $[0, 1]$). As expected, the smaller the underlying fixed time-step the smaller the errors that are actually obtained.

<table>
<thead>
<tr>
<th>Degree $r$</th>
<th>$\mathcal{M}_i$</th>
<th>Error $|u-u_h|_{L^\infty(0,1)}$</th>
<th>Order $\kappa$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>1.03e-05</td>
<td>1.9982</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>2.57e-06</td>
<td>1.9992</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>6.41e-07</td>
<td>1.9996</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>4.69e-08</td>
<td>2.9762</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>5.91e-09</td>
<td>2.9881</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>7.42e-10</td>
<td>2.9941</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>1.05e-10</td>
<td>3.9852</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>6.62e-12</td>
<td>3.9925</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>4.15e-13</td>
<td>3.9963</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>3.64e-12</td>
<td>4.9761</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>1.15e-13</td>
<td>4.9882</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>3.59e-15</td>
<td>4.9940</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>2.03e-11</td>
<td>5.9170</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3.27e-13</td>
<td>5.9585</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>5.17e-15</td>
<td>5.9793</td>
</tr>
</tbody>
</table>

Table 5.1

$p$-version: solution with $\alpha = -1$.

Fig. 5.2. $p$-version: solution with $\alpha = -1$.

5.2. Nonsmooth solution. Next, we consider the case where $\alpha = 0.5$ so that the solution $u$ in (5.1) has a singularity at $t = 0$. In fact, we have that $u \in W^{1.5,\infty}(0,1)$ while the second derivative of $u$ is unbounded near $t = 0$. In Figure 5.3, we show the performance of the $h$-version DG method on the uniform partitions $\mathcal{M}_i$ from Section 5.1. The optimal order $r + 1$ is not obtained anymore, due to the loss of smoothness of $u$ near the origin. Instead, the same asymptotic rate of convergence is...
observed for all polynomial degrees \( r \geq 1 \). This rate is computed in Table 5.2. It is of the order of 1.5 for all \( r \geq 1 \), thereby confirming the sharpness of the \( h \)-version result in Corollary 3.9.

<table>
<thead>
<tr>
<th>degree ( r )</th>
<th>( i )</th>
<th>error</th>
<th>order ( \kappa_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>1.3563e-04</td>
<td>1.4738</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>4.8388e-05</td>
<td>1.4870</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>1.7185e-05</td>
<td>1.4935</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>1.9960e-05</td>
<td>1.5169</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>7.0142e-06</td>
<td>1.5088</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>2.4723e-06</td>
<td>1.5044</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>6.5853e-06</td>
<td>1.5048</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>2.3244e-06</td>
<td>1.5024</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>8.2115e-07</td>
<td>1.5011</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>2.9812e-06</td>
<td>1.5023</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1.0532e-06</td>
<td>1.5011</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>3.7221e-07</td>
<td>1.5006</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>1.5981e-06</td>
<td>1.5014</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>5.6473e-07</td>
<td>1.5007</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>1.9961e-07</td>
<td>1.5004</td>
</tr>
</tbody>
</table>

Table 5.2

\( h \)-version: solution with \( \alpha = 0.5 \).

Since for \( \alpha = 0.5 \) the solution \( u \) in (5.1) has a singularity at \( t = 0 \), the \( p \)-version of the DG method can only be expected to yield algebraic rates of convergence, in contrast to the test in Section 5.1. Algebraic convergence behavior is indeed observed in Figure 5.4, where we increase the polynomial degree \( r \) on the same time partitions as above. The numerical convergence rates are shown in Table 5.3. In the context of
the $p$-version DG methods, these rates are defined as

$$\kappa_r = -\log \left( \frac{e(r)}{e(r-1)} \right) / \log \left( \frac{r}{r-1} \right),$$

where $e(r)$ denotes the $L^\infty$-error that is obtained for order $r$ (on a fixed partition of $(0,1)$). We note that Corollary 3.9 ensures at least the order 0.5. However, rates of order 3 are observed in Table 5.3. This indicates that the estimate in Corollary 3.9 is slightly suboptimal in the polynomial degree, as remarked in the discussion after Corollary 3.9. In fact, we observe twice the rate that would correspond to the regularity exponent 1.5 of the exact solution. This doubling phenomena is well-known in $p$-version finite element methods for second-order boundary-value problems; see [17] and the references therein. In our context, a theoretical explanation of this observation remains an open problem.

Next, we consider the performance of the $hp$-version time-stepping method on the basic geometric partitions $\mathcal{M}_{n,n} = \{I_m\}_{m=1}^{n+1}$ of $(0,1)$ introduced in Definition 4.2. In
addition, we use linearly increasing polynomial degrees as described in Definition 4.6: on time-step $I_m$ we set $r_m = [\mu m]$, with a slope $\mu > 0$. In Figure 5.5, we display the errors against the square root of the number of degrees of freedom (dofs) in the underlying discretization space, for $\mu = 1$ and various values of the grading factor $\sigma$. The straight curves indicate exponential convergence for each grading factor $\sigma$, as predicted by Theorem 4.7. It can further be seen that the grading $\sigma = 0.3$ gives the best results; for example, they are several orders of magnitude better than those for $\sigma = 0.5$. This is in contrast to the case of elliptic boundary-value problems, where the optimal choice of the grading is known to be given by $\sigma \approx 0.15$, independently of the strength of the singularity; see [17] and the references therein. In Figure 5.6, we show the convergence curves for $\sigma = 0.3$ and several values of the slope parameter $\mu$. The exponential convergence rates are less sensitive to variations in this parameter and good results are obtained for $\mu = 1$.

Finally, we test the performance of the $hp$-version DG method for the prob-

![Figure 5.5](image1)

**Fig. 5.5.** $hp$-version: solution with $\alpha = 0.5$.

![Figure 5.6](image2)

**Fig. 5.6.** $hp$-version: solution with $\alpha = 0.5$. 
lem (5.1) with $\alpha = 0.99$. In view of the above discussions, we set $\sigma = 0.3$ and $\mu = 1$.

In Table 5.4 it can be seen that, with this particular choice, the $hp$-version gives an $L^\infty$-error smaller than $1e-6$ with less than 44 degrees of freedom. To obtain the same error with the $h$-version approach on the meshes $\mathcal{M}_r$ from above and with $r = 2$, more than 10,000 degrees of freedom are needed. This clearly underlines the suitability of $hp$-version approaches for the numerical approximation of the VIDE (1.1).

<table>
<thead>
<tr>
<th>dofs</th>
<th>error in $L^\infty(0,1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1.2685e-02</td>
</tr>
<tr>
<td>9</td>
<td>1.1820e-03</td>
</tr>
<tr>
<td>14</td>
<td>6.9907e-05</td>
</tr>
<tr>
<td>20</td>
<td>1.7007e-05</td>
</tr>
<tr>
<td>27</td>
<td>9.3099e-06</td>
</tr>
<tr>
<td>35</td>
<td>3.1848e-06</td>
</tr>
<tr>
<td>44</td>
<td>9.8316e-07</td>
</tr>
</tbody>
</table>

Table 5.4: $hp$-version: solution with $\alpha = 0.99$.

6. Concluding remarks. We conclude the paper by pointing out some extensions and future work.

In applications it often happens that at least one of the functions $f_1$ and $f_2$ in (4.2) is only piecewise analytic on $[0,T]$. According to the proof of Theorem 4.1 (cf. (4.4)) the corresponding solution $u$ of (1.1) inherits this property: it is piecewise analytic on $[0,T]$, with its second derivative unbounded at $t = 0$. If the points in $[0,T]$ at which analyticity is lost are denoted by $\tau_1, \ldots, \tau_l$, it will be necessary to geometrically grade the time-steps individually near each point $\tau_i, 1 \leq i \leq l$, in order to obtain exponential convergence.

We mention in passing a related VIDE for which the above observation is relevant. Let (1.1) be replaced by

$$u'(t) + a(t)u(t) + \int_{t-\tau}^t k_\alpha(t-s)b(s)u(s)\,ds = f(t), \quad t \in [0,T],$$

$$u(t) = \phi(t), \quad t \leq 0,$$

with delay $\tau > 0$. It is well known (see, e.g., [2, Section 7.1]) that, regardless of the smoothness of the given functions, the solution $u$ of (6.1) exhibits lower regularity at the so-called primary discontinuity points $\{\kappa \tau\}_{\kappa \in \mathbb{N}_0}$ induced by the delay $\tau$. If $\phi$, $a$, $b$, $f_1$, $f_2$ are analytic on $[0,T]$, then $u$ will be analytic on each interval $(\kappa \tau, (\kappa + 1)\tau]$ but only piecewise analytic on $[0,T]$.

As we mentioned in Section 1, we shall study the exponential convergence of the $hp$-version of the DG method for time-stepping in a (spatially semi-discretized) parabolic partial VIDE (see the book [6]) in a forthcoming paper. Assume that such a partial VIDE has the form

$$u_t + Lu + \int_0^t k_\alpha(t-s)Bu(s)\,ds = f, \quad t \in [0,T], \quad x \in \Omega \subset \mathbb{R}^d,$$  

where $-L$ denotes a strongly elliptic (spatial) partial differential operator and where $B$ is given, for example, by $B = \Delta$, or by the scalar factor $b(s,x)$. If $L_h (= L_h(t))$ and
\(B_h (= B_h(s))\) denote discrete representations of \(L\) and \(B\) corresponding to a spatial discretization of (6.2) with respect to a mesh \(\Omega_h\) of \(\Omega\), then (6.2) is approximated by a system of ordinary VIDEs analogous to (1.1) in which the roles of \(a(t)\) and \(b(s)\) are now assumed by the matrices \(L_h(t)\) and \(B_h(s)\). This suggests that our “scalar” convergence analysis can in principle be extended to these systems of VIDEs. The analysis hinges on appropriate regularity results for the solution of (6.2).

The situation becomes rather more difficult if we have \(L = 0\) in (6.2) (see, e.g., [13]): we note that the convergence properties of the \(hp\)-DG method for (1.1) with \(a(t) \equiv 0\) are not covered by our analysis and remain open.

Acknowledgments. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors also gratefully acknowledge the suggestion by one of the referees that led to a more general version of Theorem 4.7.

REFERENCES