Sequential predictor-corrector methods for the variable regularization of Volterra inverse problems

Patricia K Lamm† and Thomas L Scofield

† Department of Mathematics, Michigan State University, E. Lansing, MI 48824–1027
‡ Department of Mathematics, The University of Michigan – Flint, Flint, MI 48502-1950

Abstract. We analyze the convergence of a class of discrete predictor-corrector methods for the sequential regularization of first-kind Volterra integral equations. In contrast to classical methods such as Tikhonov regularization, this class of methods preserves the Volterra (causal) structure of the original problem. The result is a discretized regularization method for which the number of arithmetic operations is $\mathcal{O}(N^2)$ (where N is the dimension of the approximating space) in contrast to standard Tikhonov regularization which requires $\mathcal{O}(N^3)$ operations.

In addition, the method considered here is defined using functional regularization parameters so that the possibility for more or less smoothing at different points in the domain of the solution is allowed. We establish a convergence theory for these methods and present relevant numerical examples, illustrating how one functional regularization parameter may be adaptively selected as part of the sequential regularization process. This work generalizes earlier results by the first author to the case of a penalized predictor-corrector formulation, functional regularization parameters, and nonconvolution Volterra equations.

Revised October 1999.

PACS numbers: 02.60, 02.30

Submitted to: *Inverse Problems*

1. Introduction

We consider the following inverse problem. Given a suitable function f, find $\bar{u}$ satisfying the first-kind Volterra integral equation

$$\mathcal{A}u(t) = f(t), \tag{1.1}$$

for a.e. $t \in [0,1]$, where $\mathcal{A}$ is the bounded linear operator on $L^2(0,1)$ defined by

$$\mathcal{A}u(t) := \int_0^t k(t,s)u(s)\,ds, \qquad \text{a.e. } t \in [0,1]. \tag{1.2}$$

Problems based on (1.1) are ill-posed due to lack of continuous dependence on data $f \in L^2(0,1)$, with the severity of ill-posedness related to properties of the kernel k. For example, if $k \in C^1([0,1]\times[0,1])$ satisfies $k(t,t) \neq 0$ for all $t \in [0,1]$, it is well known that under this condition differentiation of (1.1) with respect to t (for sufficiently smooth f) leads to a well-posed second-kind Volterra equation with solutions depending continuously on the (new) data $f' \in L^2(0,1)$. We will say that the operator $\mathcal{A}$ is "one-smoothing" in this case. But even if the "true" data function f is smooth, the usual situation is that we only have available a nonsmooth perturbation of f. Thus, in the case of problems with one-smoothing operators $\mathcal{A}$, the "degree" of ill-posedness of (1.1) is that associated with first order differentiation of noisy data.

More generally, if the kernel k is such that $\partial^\nu k/\partial t^\nu$ is continuous with $\left(\partial^{\nu-1}k/\partial t^{\nu-1}\right)(t,t) \neq 0$, $\left(\partial^\ell k/\partial t^\ell\right)(t,t) = 0$, for $\ell = 0,\ldots,\nu-2$ ($\nu \geq 2$ integer) and $0 \leq t \leq 1$, then it takes ν differentiations of equation (1.1) (for sufficiently smooth f) to obtain a well-posed second-kind equation, with solutions depending continuously on $f^{(\nu)}$. We will say that the operator $\mathcal{A}$ is "ν-smoothing" in this case, and "infinitely-smoothing" in the case of smooth k with $\left(\partial^\ell k/\partial t^\ell\right)(t,t) = 0$, for $0 \leq t \leq 1$ and all $\ell = 0,1,2,\ldots$. Of course, not all equations of the form (1.1) fall into one of these classes of problems; however, these terms will be useful in discussing below the severity of ill-posedness associated with particular Volterra equations.

Equations of the form (1.1) arise in a number of applications. For example, the inverse heat conduction problem (or sideways heat equation) [2] is based on such a model with infinitely-smoothing operator $\mathcal{A}$, while the differentiation problem [12] is associated with a one-smoothing operator $\mathcal{A}$. In both of these examples the operator $\mathcal{A}$ has a convolution kernel; however numerous nonconvolution kernels may also be found in applications (see, e.g., a recent example from capillary viscometry in [26]).

A regularization method must be used to solve (1.1), and certainly the classical method of Tikhonov regularization is a simple and effective approach. However, a disadvantage of this method when applied to Volterra problems is that it replaces the original "causal" problem (1.1) with a "full-domain" regularized problem, $(\mathcal{A}^\star\mathcal{A} + \alpha I)u = \mathcal{A}^\star f$, where $\mathcal{A}^\star$ is the adjoint operator associated with $\mathcal{A}$ and $\alpha > 0$ is the Tikhonov regularization parameter. By a "full-domain" problem we mean that, instead of using only values of f on the interval $[0,t]$ to recover $\bar{u}$ on the same interval, Tikhonov regularization requires that data values from $[t,1]$ also be used, destroying the causal nature of the original problem and leading to inefficient solution of Volterra problems. This is true for all regularization methods based on the computation of (suitably defined) $g_\alpha(\mathcal{A}^\star\mathcal{A})$ [12] for the reason that, although the original operator $\mathcal{A}$ in (1.2) is non-anticipatory (causal), the adjoint $\mathcal{A}^\star$ is an anticipatory operator.

In addition to destroying the causal nature of Volterra problems, classical methods such as Tikhonov regularization tend to oversmooth solutions. Other regularization approaches, such as the technique of bounded variation regularization [1, 5, 8, 15, 16, 35] and the idea of regularization for curve representations [27], have been developed to handle the problem of oversmoothing. Although quite promising, such methods do not retain the causal nature of the Volterra problem and, in addition, require either the formulation of a nondifferentiable or nonlinear optimization problem.

The goal of this paper is to establish convergence results for a discrete regularization method for the solution of (1.1), a method which retains the causal nature of the original problem and also has the potential for avoiding excessive oversmoothing. This approach falls into a broad class of "local regularization" methods for Volterra equations [17]. In addition to retaining both the causal and linear structure of Volterra problems, it also has the advantage of being formulated as a differentiable optimization technique in local regions of the solution. Numerical implementation of this local regularization method leads to a sequential algorithm which exhibits certain "predictor-corrector" characteristics. Indeed, at each step in the sequential algorithm, solutions are held rigid for a short time into the future, yielding a locally-regularized "prediction" of the desired solution. Then, in a "correction" step, the local solution is truncated in order to avoid oversmoothing and to improve accuracy. The result is a method that is easily implemented numerically and which, due to its sequential nature, has the capability of providing very fast solutions. In fact, as we will see in Section 2.1, the method we consider requires only $\mathcal{O}(N^2)$ arithmetic operations while standard Tikhonov regularization requires $\mathcal{O}(N^3)$ operations. In the case of k a convolution kernel the local regularization method is still more efficient, although the difference in cost (to highest order in N) is less dramatic; in this case the local method takes $N^2/2$ multiplications while standard Tikhonov regularization requires $9N^2/2$ multiplications.

The method we present here is a generalization of previous work [18], where now we consider *functional* regularization parameters, a sequence of penalized local regularization problems (with functional penalty parameter $\mu = \mu(t)$), and an extension to the case of nonconvolution kernels k. We note that the extension of [18] to the case of functional regularization parameters required nontrivial theoretical changes. In this paper we additionally formulate a sequential discrepancy principle for the adaptive selection of the penalty parameter μ.

It is worth noting that, although computational costs tend to be smaller with a method which preserves the original Volterra structure of the problem, there is generally additional cost in terms of the assumptions which must be made in order to prove convergence of the method. Classical (non-causal) methods based on the operator $\mathcal{A}^\star\mathcal{A}$ are generally associated with well-developed convergence theories for even infinitely-smoothing problems because such theories may be advanced using the special spectral properties of $\mathcal{A}^\star\mathcal{A}$. The same cannot be said, in general, of "Volterra-preserving" methods because they are based on the operator $\mathcal{A}$ alone and do not make use of the non-causal operator $\mathcal{A}^\star\mathcal{A}$. Thus, theoretical results for such methods are generally limited by the assumption that the underlying equation is only moderately ill-posed [17]. It is worth noting that this is often only a theoretical limitation, as a given Volterra-preserving method may work quite well in practice for even severely ill-posed problems.

Among methods which retain the Volterra/causal nature of the original problem we mention the following. Lavrent′ev's classical method, or the small parameter method, is associated with a well-developed convergence theory for one-smoothing operators $\mathcal{A}$ [7] (see [9, 24, 25, 37] to name only a few of the references in this area). The method works quite well if $\bar{u}(0)$ is known precisely, but suffers from boundary layer effects (requiring solution methods for stiff singularly perturbed equations) if $\bar{u}(0)$ is not exactly known [17]. A related method has been developed to include general ν-smoothing problems [37], but in this case precise knowledge is required of $\bar{u}$ and higher order derivatives of $\bar{u}$ at $t = 0$ in order to avoid the boundary layer effects.

Other Volterra-preserving methods include Lavrent′ev's m-times iterated method [29, 30], Richardson iteration [32, 38, 39], and certain implicit iterative methods [28, 30, 31, 32]. The regularized convergence theory for these methods, in the case of noisy data, appears to be limited to only very moderately ill-posed problems (such as the classical Abel integral equation, which is generally considered "half-smoothing"); we are not aware of successful application of these methods (in practice) to more severely ill-posed problems. See [17] for an expanded discussion of these methods and underlying theoretical assumptions, and for additional references on these and other methods (e.g., mollification methods).

The method that is the focus of this paper is no different from those mentioned above in that our theoretical convergence proofs are limited to only moderately ill-posed problems. We present a convergence theory for the case of one-smoothing problems, and make the assumption throughout that $k \in C^1$, $k(t,t) \neq 0$ for $t \in [0,1]$. (Without loss of generality we will assume $k(t,t) = 1$, $t \in [0,1]$.) However, despite the fact that the theoretical development presented here is based on such an assumption, our method is a generalization of a numerical technique developed by J. V. Beck which has been used successfully for over thirty years for the severely ill-posed (infinitely-smoothing) inverse heat conduction problem [2]. In Section 2.4 we also illustrate the effectiveness of the method when applied to a two-smoothing example. Indeed, practical application seems to indicate that the method applies to a wide variety ill-posed Volterra problems, with both finitely- and infinitely-smoothing operators $\mathcal{A}$.

The paper is organized as follows. In Section 2 we describe implementation and convergence results for a predictor-corrector regularization algorithm that is a special case of the more general class of discrete local regularization methods to be considered in this paper. In this section we also illustrate, via numerical examples, the effectiveness of the method and of a strategy for adaptively selecting the penalty parameter. In

Section 3 we formulate the hypotheses and more general structure in which convergence is to be examined, stating convergence results in Section 4. Finally, proofs of these results are presented in Section 5.

2. Discrete predictor-corrector regularization methods

2.1. A sequential collocation-based discretization

We will motivate the discrete local regularization method to be considered in this paper by first examining a collocation-based discretization of (1.1). To this end, let $N = 1, 2, \ldots$ be fixed and divide $[0, 1]$ into N subintervals $[t_{i-1}, t_i]$, $i = 1, \ldots N$, each of width $h = 1/N$. We seek constants c_i, $i = 1, \ldots, N$, so that the step function

$$u_h(t) := \sum_{i=1}^{N} c_i \chi_i(t), \quad t \in [0, 1], \tag{2.1}$$

satisfies (1.1) at the collocation points $t = t_j$, $j = 1, \ldots, N$. That is,

$$\mathcal{A}\left(\sum_{i=1}^{j} c_i \chi_i\right)(t_j) = f(t_j), \quad j = 1, \ldots, N. \tag{2.2}$$

In the above, χ_i is the usual characteristic function on the interval $(t_{i-1}, t_i]$ for $i = 2, \ldots, N$, while χ_1 is the characteristic function on the interval $[0, t_1]$. Because the operator $\mathcal{A}$ is of Volterra type, equation (2.2) is a triangular system of equations for which the solution is determined sequentially provided the diagonal entries are nonzero (guaranteed under reasonable assumptions on the kernel k).

It is useful at this point to mention an equivalent formulation of the same procedure, which we state as follows. Assuming $c_1, \ldots, c_{j-1}$ have already been found, determine the h-dependent constant c_j satisfying

$$c_j = \arg\min_{c \in \mathbb{R}} J_j(c), \tag{2.3}$$

$$J_j(c) := \left(\mathcal{A}\left(\sum_{i=1}^{j-1} c_i \chi_i + c\chi_j\right)(t_j) - f(t_j)\right)^2. \tag{2.4}$$

Although the procedure (2.2), equivalently (2.3)–(2.4), for determining an approximate solution of (1.1) is a well-posed problem (because it is finite-dimensional), it is not well-conditioned and can lead to poor approximations. The idea for a regularized improvement of this simple algorithm may be traced back to a numerical method developed by J. V. Beck in the 1960's for the inverse heat conduction problem. This particular approach was generalized in [6, 18, 19, 20, 22] and examined in those references from the point of view of stability and convergence. (Other relevant treatments of Beck's method may be found, for example, in [33, 34].) Here we extend these ideas even further by considering a similar method but now with a functional "local regularization parameter" r. For example, given $h = 1/N$, define an h-dependent regularization function $r = r(t)$ by

$$
\begin{aligned}
r(t) &:= \sum_{i=1}^{N} r_i\, \chi_i(t), \quad t \in [0,1], && (2.5)\\
r_i &:= \gamma_i h, \text{ for integer } \gamma_i \geq 0, \quad i = 1, \ldots, N. && (2.6)
\end{aligned}
$$

The idea behind the new method is to seek u_h of the form (2.1) as before but instead to determine the coefficients in (2.1) in the following manner. Assuming $c_1, \ldots, c_{j-1}$ have already been found, the j^{th} step in the process is to determine c_j such that

$$
\begin{aligned}
c_j &= \arg\min_{c\in\mathbb{R}} J_{j,r}(c), && (2.7)\\
J_{j,r}(c) &:= \sum_{s=0}^{\gamma_j} \left(\mathcal{A}\left(\sum_{i=1}^{j-1} c_i \chi_i + c \sum_{\ell=j}^{j+s} \chi_\ell \right)(t_{j+s}) - f(t_{j+s}) \right)^2. && (2.8)
\end{aligned}
$$

Thus the constant c_j determined via (2.7)–(2.8) is the best constant-valued solution (in a least-squares sense) over the interval $[t_{j-1}, t_j + r(t_j)] = [t_{j-1}, t_j + \gamma_j h]$. This process of temporarily holding the solution rigid over a small future interval leads to a regularized "prediction" of the optimal solution $u_h(\cdot) = u_h(\cdot; r)$ on the interval $[t_{j-1}, t_j + r(t_j)]$. We "correct" this over-regularized solution by only retaining this solution on the interval $[t_{j-1}, t_j]$ (i.e., the predicted value of u_h on $[t_j, t_j + r(t_j)]$ is not retained) at the j^{th} step of the process. We note that in the case of $\gamma_j = 0$, $j = 1, \ldots, N$, we have $r(t) = 0$, $t \in [0,1]$, and the algorithm (2.7)–(2.8) reduces to the discrete algorithm (2.3)–(2.4) for the original (unregularized) problem.

We can generalize these ideas even further by considering a penalized version of the process described in (2.7)–(2.8). Suppose, for example, an h-dependent function $\mu = \mu(t)$ is given by

$$
\mu(t) := \sum_{i=1}^{N} \mu_i\, \chi_i(t), \; t \in [0,1], \qquad \mu_i \geq 0, \; i = 1, \ldots, N. \tag{2.9}
$$

We now find u_h of the form (2.1) where the coefficients c_j in this expression are determined as follows. Assuming that $c_1, \ldots, c_{j-1}$ have already been found, the idea is to determine c_j such that

$$
\begin{aligned}
c_j &= \arg\min_{c\in\mathbb{R}} J_{j,r,\mu}(c), && (2.10)\\
J_{j,r,\mu}(c) &:= J_{j,r}(c) + \mu_j\, c^2, && (2.11)
\end{aligned}
$$

with $J_{j,r}$ given by (2.8) for $j = 1, \ldots, N$. Thus the parameter $\mu_j = \mu(t_j) > 0$ serves to penalize large values of the constant being determined in the j^{th} step of the numerical process. Obviously, in the case of $\mu_j = 0$, $j = 1, \ldots, N$, the process reduces to (2.7)–(2.8).

The j^{th} coefficient c_j found by each of the above algorithms may be written explicitly. Indeed, making the definition of the h-dependent quantity Δ_{nm},

$$
\Delta_{nm} := \mathcal{A}\chi_m(t_n) = \int_0^{t_1} k(t_n, t_{m-1} + s)\, ds,
$$

for $1 \le m \le n \le N$, it follows that $J_{j,r,\mu}(c)$ may be written as

$$J_{j,r,\mu}(c) = \|\mathbf{b}_j c - \mathbf{d}_j\|_j^2 + \mu_j\, c^2 \tag{2.12}$$

where $\|\cdot\|_j$ denotes the usual Euclidean norm in $\mathbb{R}^{\gamma_j+1}$. Here the h-dependent quantities $\mathbf{b}_j$ and $\mathbf{d}_j$ are given by $\mathbf{b}_j := \left(b_1^{(j)}, \ldots, b_{\gamma_j+1}^{(j)}\right)^\top$, with

$$b_\ell^{(j)} := \sum_{i=1}^{\ell} \Delta_{j+\ell-1,\, j+i-1}\,, \quad \ell = 1, \ldots, \gamma_j + 1, \tag{2.13}$$

and the ℓ^{th} entry in $\mathbf{d}_j \in \mathbb{R}^{\gamma_j+1}$ is given by

$$(\mathbf{d}_j)_\ell = f(t_{j+\ell-1}) - \sum_{i=1}^{j-1} c_i\, \Delta_{j+\ell-1,\, i}, \;\; \ell = 1, \ldots, \gamma_j + 1. \tag{2.14}$$

Thus the scalar c_j which solves the penalized algorithm (2.10)–(2.11) is given by

$$c_j = \left(\|\mathbf{b}_j\|_j^2 + \mu_j\right)^{-1} \mathbf{b}_j^\top \mathbf{d}_j, \tag{2.15}$$

where $\|\mathbf{b}_j\|_j^2 + \mu_j \ge \|\mathbf{b}_j\|_j^2 > 0$ under reasonable assumptions on the kernel k (see, e.g., the assumptions in Section 3.2). By making specific choices of the parameters μ_j and γ_j in (2.15) (where $\gamma_j + 1$ is the vector dimension of $\mathbf{b}_j$, $\mathbf{d}_j$ in (2.15)), one may also recover the solutions c_j to the remaining two algorithms considered above. In particular, the choices $\gamma_j = 0$ and $\mu_j = 0$ in (2.15) prescribe the solution c_j of the original collocation algorithm (2.3)–(2.4) while the choices $\gamma_j \ge 0$, $\mu_j = 0$, determine c_j as the solution of the unpenalized predictor-corrector algorithm (2.7)–(2.8).

The operation count for the algorithm in the case of nonconvolution kernel k is as follows. The biggest expense is the computation of $\mathbf{d}_j$, but the cost is lowered by noting that, for $j = 2, \ldots, N$, the ℓ^{th} entry $(\mathbf{d}_j)_\ell$ of $\mathbf{d}_j$ may be written $(\mathbf{d}_j)_\ell = (\mathbf{d}_{j-1})_{\ell+1} - c_{j-1}\Delta_{\ell+j-1,j-1}$, for $\ell = 1, \ldots, \gamma_j + 1$. The computation of $(\mathbf{d}_j)_\ell$ requires no multiplications for $j = 1$, while for $j = 2, \ldots, N$, one multiplication is required for each ℓ. Thus the worst-case cost of computing all $\mathbf{d}_j$'s is $\sum_{j=2}^{N}(\gamma_{\max} + N - j + 1) = N^2/2 + (\gamma_{\max} - 1/2)N - \gamma_{\max}$, where $\gamma_{\max} = \max_{1 \le i \le N} \gamma_i$.

The computation of c_j in (2.15) requires $2(\gamma_j + 1) + 1$ additional multiplications for each N, bringing the total algorithm count to $N^2/2 + (3\gamma_{\max} + 5/2)N - \gamma_{\max}$ multiplications. Since $\gamma_{\max}$ is generally taken to be much smaller than N in practice, this estimate compares quite favorably to standard Tikhonov regularization which requires $\mathcal{O}(N^3)$ multiplications in the nonconvolution case [22].

In the case of a convolution kernel, $k(t, s) = \kappa(t - s)$, it can be shown that the algorithm presented here requires $N^2/2 + (2\gamma_{\max} + 3/2)N + 1$ multiplications, which compares favorably to the multiplication count of $4.5N^2$ (to highest order) for standard Tikhonov regularization as applied to the convolution case [11]. Our local regularization method is less expensive (to highest order) than standard Tikhonov regularization for the convolution problem provided $\gamma_{\max} \le 2N - 1$ (again, in numerical examples it is seen that an appropriate value of $\gamma_{\max}$ is generally much less than N, even for severely ill-posed problems).

2.2. Convergence of the discrete predictor-corrector regularization method

Since the local regularization algorithm (2.10)–(2.11) involves solving optimization problems over small future intervals, the theory we develop will require that we either seek a regularized approximation to $\bar{u}$ on an interval of the form $[0, 1-\varepsilon]$, for $\varepsilon > 0$ small, or else slightly extend the domain of definition of the original problem. We take the latter route here and make the following standing hypothesis.

Let $T > 1$ and assume $k \in \mathcal{C}^1([0,T] \times [0,T])$ with $k(t,t) = 1$ for $0 \le t \le T$.

In Sections 3–5, we will generalize the local regularization algorithm given above and develop an associated convergence theory. Because this generalization is somewhat technical, it is worth stating here the results of these sections as they apply to the more practical algorithm given in Section 2.1 above. The result demonstrates convergence of the regularized approximation scheme in the case where true data f is used, as well as in the more usual case where only a perturbation f^δ of f is available. Convergence of approximations will be in the following sense.

Definition 2.1 *We say that $u_h(\cdot)$ converges to $u(\cdot)$* uniformly at collocation points *as $h \to 0$ if for each $\epsilon > 0$ there exists $H = H(\epsilon) > 0$ for which $|u_h(t_l) - u(t_l)| < \epsilon$ for each $t_l = lh$, $l = 1, \ldots, 1/h$, whenever $0 < h \le H(\epsilon)$.*

Because convergence occurs as $h \to 0$, we will need sequences (r_h) and (μ_h) of regularization parameters, selected satisfying certain conditions in the limit as $h \to 0$.

Theorem 2.1 *Assume $f : [0,T] \to \mathbb{R}$ is a bounded Borel measurable function for which the unique solution $\bar{u}$ of (1.1) corresponding to f is in $C^1[0,T]$. For each $h = 1/N$, $N = 1, 2, \ldots$, let the regularization parameters r_h and μ_h be given by (2.5) and (2.9), respectively, i.e.,*

$$r_h(t) := \sum_{i=1}^{N} r_{h,i}\, \chi_i(t), \qquad \mu_h(t) := \sum_{i=1}^{N} \mu_{h,i}\, \chi_i(t), \quad t \in [0,1],$$

where we assume

$$r_{h,i} = \gamma(t_i) h, \qquad \mu_{h,i} = \ell(t_i) h^2, \quad i = 1, \ldots N,$$

for all h sufficiently small. Here γ is piecewise continuous and integer-valued, with $\gamma(t^+) \ge \gamma(t^-) - 1$, $t \in [0,1]$, and $\ell : [0,1] \mapsto (0,\infty)$ bounded. Then the solution $u_h = \sum_{i=1}^{N} c_j \chi_j(\cdot)$ of (2.10)–(2.11) (where r_h and μ_h are used in place of r and μ) converges to $\bar{u}$ uniformly at collocation points t_j, $j = 1, \ldots, N$, as $h \to 0$.

In addition, let $\delta > 0$ and f^δ be bounded Borel measurable with $\|f - f^\delta\|_\infty \le \delta$. If $h = h(\delta)$ is selected so that

$$\delta / h^2(\delta) \le M, \quad h(\delta) \to 0, \ \text{ as } \delta \to 0,$$

then $u_h^\delta = \sum_{i=1}^{N} c_j \chi_j$ of (2.10)–(2.11) (defined additionally using f^δ in place of f) converges to $\bar{u}(\cdot)$ uniformly at collocation points t_j, $j = 1, \ldots, N(\delta)$, as $\delta \to 0$. This convergence is at the best possible rate with respect to δ, that is,

$$|u_h^\delta(t_j) - \bar{u}(t_j)| \le K \delta^{1/2} + \mathcal{O}(\delta), \quad j = 1, \ldots, N(\delta)$$

as $\delta \to 0$, where $K > 0$ is independent of h and δ.

The implications of the theorem (the proof of which follows immediately from Theorem 4.2) are that the penalty regularization parameter $\mu_h(\cdot)$ may be t-varying provided it is not too large (relative to h^2) and, in fact, μ_h may be zero. In addition, the local regularization parameter r_h may also vary with t, provided it too is not excessively large (relative to h) and provided it does not decrease too rapidly as t increases. We note that there are no limitations on increases in r_h.

2.3. Sequential selection of the penalty parameter μ

As is true with all regularization methods, proper selection of the regularization parameter(s) is an important issue. For simplicity we will assume that $h = 1/N$ is given and that the regularization function r is fixed and given by (2.5)–(2.6). We note that a principle for the sequential selection of r has been considered in numerical examples for the inverse heat conduction problem in [3], but we will not address selection of this parameter here. Our main interest in this section concerns the selection of the regularization parameter $\mu = \mu(t)$ of the form (2.9) in the regularization algorithm (2.10)–(2.11), in the case of perturbed data f^δ. We will give an explicit formula for $\mu_j = \mu(t_j)$ at the j^{th} step in the sequential process.

Let $j \geq 1$. Then if $c_1^\delta, \ldots, c_{j-1}^\delta$ have already been found, we determine the (h-dependent constant) c_j^δ from (2.10), (2.12), where now the perturbed data f^δ will be used in place of f. That is,

$$c_j^\delta = \arg\min_{c\in\mathbb{R}} J_{j,r,\mu}^\delta(c), \tag{2.16}$$

$$J_{j,r,\mu}^\delta(c) := \left\|\mathbf{b}_j c - \mathbf{d}_j^\delta\right\|_j^2 + \mu_j\, c^2, \tag{2.17}$$

with $\mathbf{b}_j \in \mathbb{R}^{\gamma_j+1}$ defined via (2.13) and the ℓ^{th} entry in $\mathbf{d}_j^\delta \in \mathbb{R}^{\gamma_j+1}$ given by

$$\left(\mathbf{d}_j^\delta\right)_\ell = f^\delta(t_{j+\ell-1}) - \sum_{i=1}^{j-1} c_i^\delta\, \Delta_{j+\ell-1,i}\,,$$

for $\ell = 1, \ldots, \gamma_j + 1$. For a given value of $\mu_j = \mu(t_j) \geq 0$, the solution $c_j^\delta = c_j^\delta(\mu_j)$ of (2.16)–(2.17) is then given by

$$c_j^\delta(\mu_j) = \left(\|\mathbf{b}_j\|_j^2 + \mu_j\right)^{-1}\left(\mathbf{b}_j^\top \mathbf{d}_j^\delta\right), \tag{2.18}$$

where $\|\mathbf{b}_j\|_j^2 + \mu_j \geq \|\mathbf{b}_j\|_j^2 > 0$ when $\Delta_{jj} \neq 0$ (which occurs under the standing assumptions on the kernel k given in Section 2.2).

In order to determine an appropriate value of $\mu_j = \mu(t_j)$ at the j^{th} step, we apply a Morozov discrepancy principle. To this end we let $C_j \geq 1$ be fixed and assume that we know δ_j for which

$$\|\mathbf{d}_j^\delta - \mathbf{d}_j\|_j \leq \delta_j \tag{2.19}$$

(where $\mathbf{d}_j$ is given by (2.14)), where we assume that the signal-to-noise assumption,

$$C_j \delta_j^2 < \|\mathbf{d}_j^\delta\|_j^2, \tag{2.20}$$

is satisfied at the j^{th} step. Then a discrete Morozov discrepancy principle determines the selection of μ_j at this step via

$$F_j(\mu_j) = C_j \delta_j^2, \tag{2.21}$$

where F_j represents the j^{th} discrete discrepancy function. That is, for $\nu \geq 0$,

$$F_j(\nu) = \|\mathbf{b}_j c_j^\delta(\nu) - \mathbf{d}_j^\delta\|_j^2 = \left\| \left(\|\mathbf{b}\|_j^2 + \nu\right)^{-1} \mathbf{b}_j\, \mathbf{b}_j^\top \mathbf{d}_j^\delta - \mathbf{d}_j^\delta \right\|_j^2. \tag{2.22}$$

The unique μ_j determined by this process is given by the following theorem.

Theorem 2.2 *Let $h = 1/N > 0$ and let r be given by (2.5)–(2.6). For $j \geq 1$, assume that $c_1^\delta, \ldots, c_{j-1}^\delta$ have already been determined. Then if δ_j satisfies (2.19)–(2.20) for fixed $C_j \geq 1$, an application of the discrete Morozov discrepancy principle (2.21) determines a unique μ_j at the j^{th} step given by*

$$\mu_j = \begin{cases} 0, & \text{if } C_j\delta_j^2 \leq D_j, \\ \sigma_j \left(\sigma_j + \left|\mathbf{b}_j^\top \mathbf{d}_j^\delta\right|\right) \left(\|\mathbf{d}_j^\delta\|_j^2 - C_j\delta_j^2\right)^{-1}, & \text{if } C_j\delta_j^2 > D_j. \end{cases} \tag{2.23}$$

Here $D_j \geq 0$ and $\sigma_j \in \mathbb{R}$, $\sigma_j > 0$, are given respectively by

$$D_j := \|\mathbf{b}_j\|_j^{-2} \left(\|\mathbf{b}_j\|_j^2 \, \|\mathbf{d}_j^\delta\|_j^2 - \left(\mathbf{b}_j^\top \mathbf{d}_j^\delta\right)^2\right), \tag{2.24}$$

$$\sigma_j \; := \; \left(\left(\mathbf{b}_j^\top \mathbf{d}_j^\delta\right)^2 - \|\mathbf{b}_j\|_j^2 \left(\|\mathbf{d}_j^\delta\|_j^2 - C_j\delta_j^2\right)\right)^{1/2}$$

in the case of $C_j\delta_j^2 > D_j$. Using this value of μ_j, the solution c_j^δ at the j^{th} step is then given by (2.18).

Proof: It is not difficult to show that $F_j(0) = D_j$ and that $F_j' > 0$ on $(0, \infty)$. Thus there is a unique $\mu_j > 0$ satisfying the j^{th} discrete discrepancy equation (2.21) for all $C_j\delta_j^2 \in (D_j, \|\mathbf{d}_j^\delta\|_j^2)$.

Let $C_j\delta_j^2 > D_j$. We note from (2.24) that this condition is equivalent to $\left(\mathbf{b}_j^\top \mathbf{d}_j^\delta\right)^2 - \|\mathbf{b}\|_j^2 \left(\|\mathbf{d}_j^\delta\|_j^2 - C_j\delta_j^2\right) > 0$, from which it follows that σ_j is real-valued and positive. In addition, it is easy to see from the definition of σ_j that $|\mathbf{b}_j^\top \mathbf{d}_j^\delta| > \sigma_j$ so that $|\mathbf{b}_j^\top \mathbf{d}_j^\delta| \neq 0$.

Rewriting $F_j(\nu)$ in (2.22), we have

$$F_j(\nu) = \tau^{-2}(\nu) \left(\tau^2(\nu)\|\mathbf{d}_j^\delta\|_j^2 - 2\,\tau(\nu)\, \left(\mathbf{b}_j^\top \mathbf{d}_j^\delta\right)^2 + \|\mathbf{b}_j\|_j^2 \left(\mathbf{b}_j^\top \mathbf{d}_j^\delta\right)^2\right) \tag{2.25}$$

where $\tau(\nu) = \|\mathbf{b}\|_j^2 + \nu > 0$ for $\nu \geq 0$ since $\Delta_{jj} \neq 0$. We seek $\mu_j > 0$ which uniquely solves $F_j(\nu) = C_j\delta_j^2$, or, equivalently, ν satisfying

$$0 = \tau^2(\nu) \left(\|\mathbf{d}_j^\delta\|_j^2 - C_j\delta_j^2\right) - 2\,\tau(\nu)\, \left(\mathbf{b}_j^\top \mathbf{d}_j^\delta\right)^2 + \|\mathbf{b}_j\|_j^2 \left(\mathbf{b}_j^\top \mathbf{d}_j^\delta\right)^2.$$

Solving this equation for $\tau(\nu)$ we obtain explicit values of $\mu_j = \tau(\nu) - \|\mathbf{b}_j\|_j^2$, i.e.,

$$\mu_j = \left(\sigma_j^2 \pm \sqrt{\left(\mathbf{b}_j^\top \mathbf{d}_j^\delta\right)^2 \sigma_j^2}\right) \left(\|\mathbf{d}_j^\delta\|_j^2 - C_j\delta_j^2\right)^{-1}. \tag{2.26}$$

But $\sigma_j < |\mathbf{b}_j^\top \mathbf{d}_j^\delta|$ implies $\sqrt{\left(\mathbf{b}_j^\top \mathbf{d}_j^\delta\right)^2 \sigma_j^2} > \sigma_j^2$, so that there is only one nonnegative μ_j in (2.26) above. This is the unique μ_j found by a discrete Morozov discrepancy principle at the j^{th} sequential step. The remainder of the theorem follows easily. $\square$

2.4. Numerical implementation

We consider an example in which the true solution $\bar{u}$ has a discontinuous derivative. In Figures 1–4, this solution is represented by a dashed curve. Approximate solutions are computed using $N = 40$ ($h = 1/40$) and in these figures are represented by solid curves joining midpoints of piecewise constant approximations by line segments. The operator $\mathcal{A}$ is given by (1.2) where the kernel k is given by $k(t,s) = t - s$, for $0 \leq s \leq t \leq 1$. The data f^δ used in the regularization process is a (uniformly distributed) random perturbation of $f = \mathcal{A}\bar{u}$, where f^δ differs from f with approximately 3% relative error.

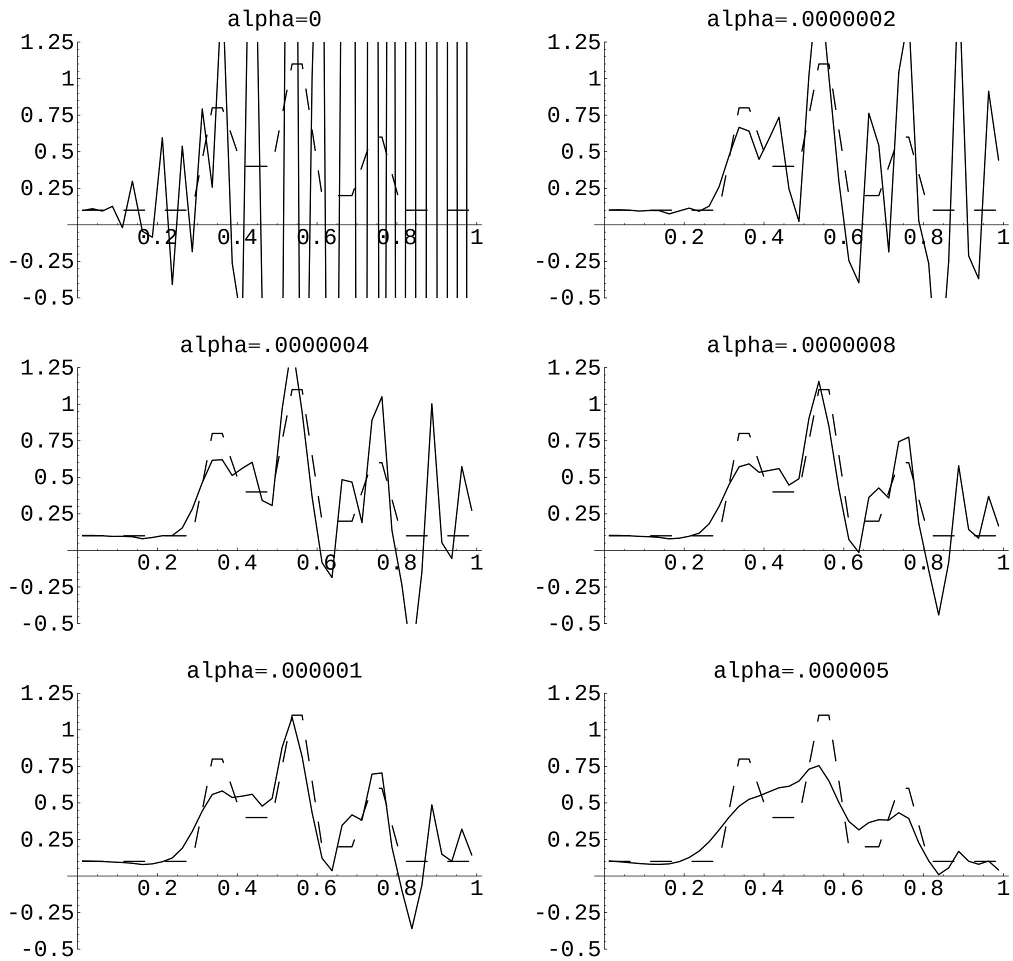

Figure 1. Results from Tikhonov regularization for various α values

As a baseline for comparison, we show in Figure 1 the results of standard Tikhonov regularization as applied to this example, using various choices of the Tikhonov parameter α. We show results for the same example in Figure 2 where now the local regularization ("predictor-corrector") ideas of this paper are used to find approximate solutions. In each graph in this figure, a constant value of $r \equiv 2h$ is used, while different

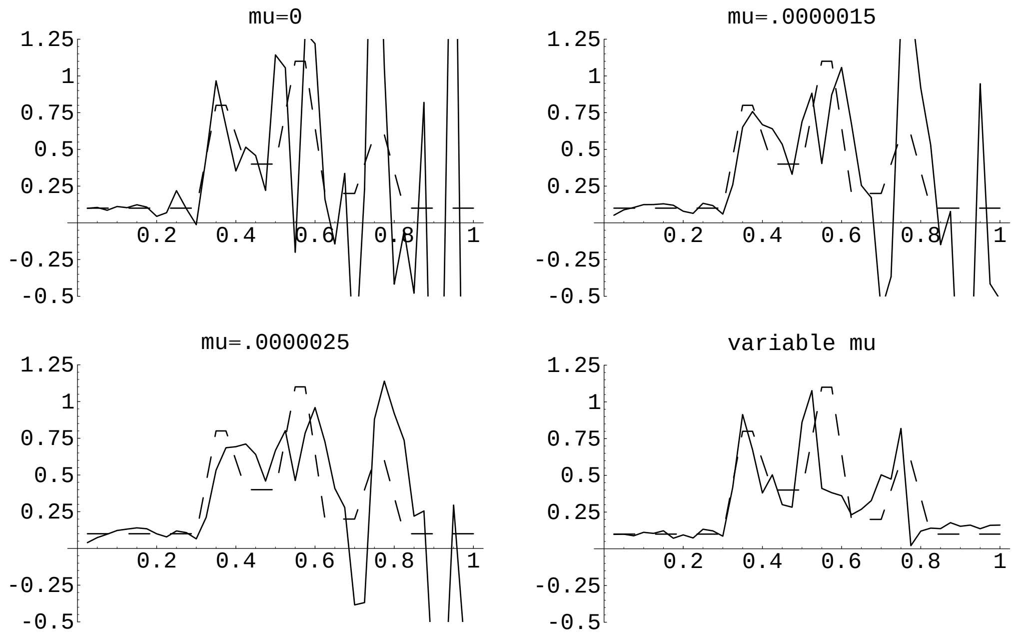

Figure 2. Results using the "local regularization" method with various choices of $\mu(t)$ (`mu`).

values of the penalty parameter μ are selected. In the first three graphs in Figure 2, μ is constant-valued (taking the values $\mu \equiv 0.0$, $1.5{\times}10^{-6}$, and $2.5{\times}10^{-6}$, respectively); in the final graph in that figure, an *a priori* selection of a functional parameter $\mu = \mu(t)$ is made, with values of μ in this case varying from 10^{-8} at the beginning of the interval to 10^{-4} at the end of the interval. (See Figure 3 for an even better choice of variable μ for this problem.) Further improvements in the results are obtained if r is also allowed to vary with t, however the advantages of variable r have been illustrated in numerical examples elsewhere (see [23] for an example similar to that considered here and [36] for other numerical results). For this reason, we keep r constant and focus here instead on results obtained through the use of a variable penalty parameter μ.

It is worth making a comparison between the first graph in Figure 1 (standard Tikhonov regularization with $\alpha = 0$, i.e., the solution to the discrete equations (2.3)–(2.4)) and the first graph in Figure 2 (local regularization with $\mu \equiv 0$). The latter graph shows an improved approximate solution, but this is because the choice of $r \equiv 2h$ offers some regularization even when $\mu \equiv 0$.

In Figure 3 we illustrate an application of a sequential discrepancy principle to select $\mu(t)$. As given in Theorem 2.2, we have an explicit representation for $\mu_j \equiv \mu(t_j)$ given an estimate of δ_j at the j^{th} step in the sequential process. It is our experience that useful results require a fairly reasonable estimate of *both* the data error component and the

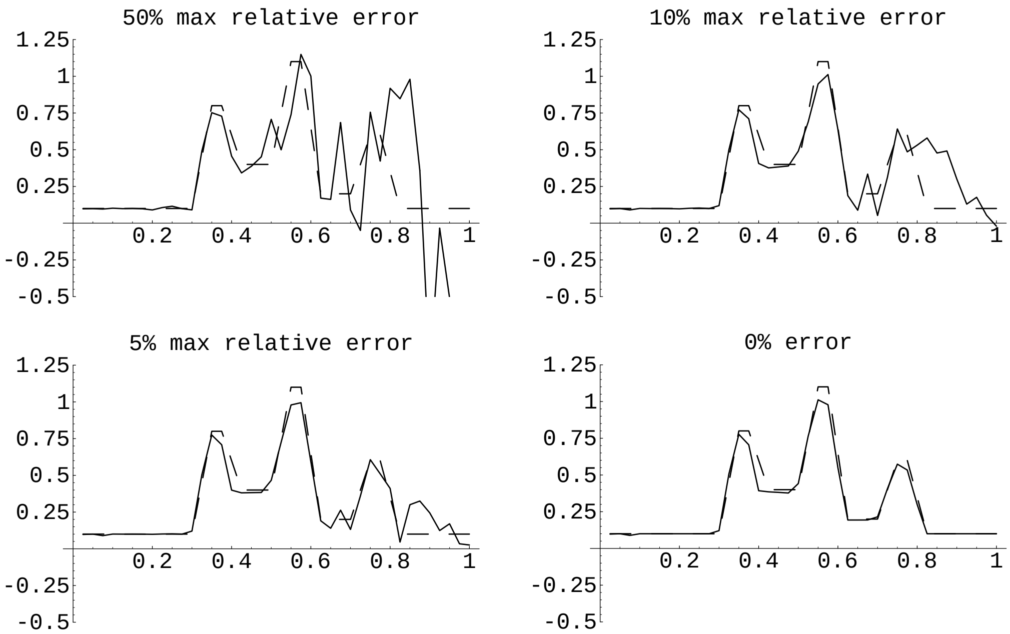

Figure 3. Results obtained using the sequential computation of $\mu(t)$ (as prescribed by Theorem 2.2). In the above graphs, the maximum relative errors refer to the computation of the δ_j needed in Theorem 2.2.

propagated error component (the latter being more difficult to estimate) which comprises δ_j. We note that this is in contrast to initial findings for the method of sequential Tikhonov regularization (in which a local, reduced-dimension Tikhonov regularization problem is solved at the j^{th} step). Indeed, numerical tests for this particular method seem to indicate that one need only provide an estimate of the average data error (ignoring the effect of propagated error) in order to sequentially determine a variable Tikhonov-like parameter which works well in practice [21].

For the results in Figure 3 we use $\delta_j = \bar{\delta}_j(1+\nu_j)$ where $\bar{\delta}_j$ is the *exact* error (exact data error, plus exact propagated error) at the j^{th} step in the sequential process, and ν_j is a uniformly distributed random variable scaled to obtain 50%, 10%, 5%, and 0%, respectively, maximum relative error in δ_j. In each example we use $C_j = 1$, $j = 1, \ldots, N$, in the formula (2.23) for μ_j, $j = 1, \ldots, N$. In Figure 4, we repeat the graph of the approximate solution found using 0% relative error in δ_j. In the second graph in Figure 4, we graph the μ that was determined by the sequential discrepancy principle. In the third graph in this figure, we rescale the y-axis for μ so that the detail on the first half of the interval can be clearly seen. It is interesting to note that decreases in values of predicted μ correspond to locations of larger/steeper values of the true solution (at which points less regularization is required). In addition, the sequentially-determined

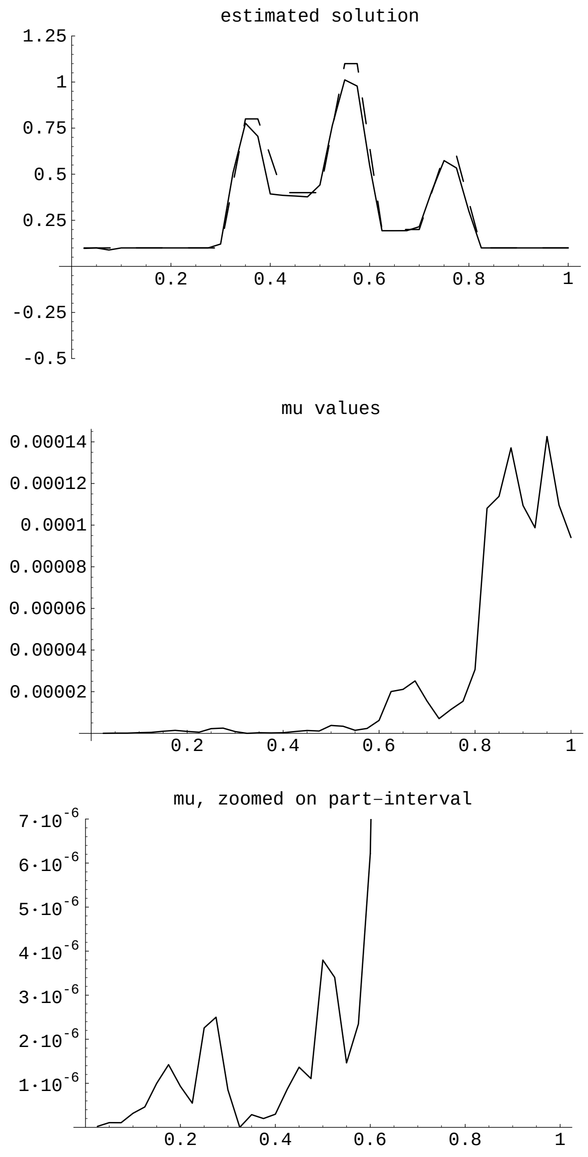

Figure 4. Sequential selection of $\mu(t)$ (`mu`)

μ increases greatly toward the end of the interval, when propagated error is having the largest effect.

3. Generalized discrete predictor-corrector methods

3.1. An equivalent representation of the predictor-corrector algorithm

Assuming we are given r, μ of the form (2.5)–(2.6) and (2.9) respectively, it is useful to view the penalized predictor-corrector algorithm (2.10)–(2.11) in a slightly different context. Recall that (2.15) gives an explicit solution of this algorithm at the j^{th} step.

Rewriting (2.15) we have

$$\sum_{\ell=1}^{\gamma_j+1} b_\ell^{(j)} \left(\sum_{i=1}^{j-1} c_i \, \Delta_{j+\ell-1,i} \right) + \left(\|\mathbf{b}_j\|_j^2 + \mu_j \right) c_j = \sum_{\ell=1}^{\gamma_j+1} b_\ell^{(j)} f(t_{j+\ell-1}) \tag{3.1}$$

where, using (2.1),

$$\sum_{i=1}^{j-1} c_i \Delta_{j+\ell-1,i} = \int_0^{t_{j-1}} k(t_{j+\ell-1}, s) u_h(s) \, ds$$

and

$$\begin{aligned} c_j \|\mathbf{b}_j\|_j^2 &= c_j \sum_{\ell=1}^{\gamma_j+1} b_\ell^{(j)} \Delta_{j+\ell-1,j} + c_j \sum_{\ell=1}^{\gamma_j+1} b_\ell^{(j)} \sum_{i=2}^{\ell} \Delta_{j+\ell-1,j+i-1} \\ &= \sum_{\ell=1}^{\gamma_j+1} b_\ell^{(j)} \int_{t_{j-1}}^{t_j} k(t_{j+\ell-1}, s) u_h(s) \, ds + u_h(t_j) \sum_{\ell=1}^{\gamma_j+1} b_\ell^{(j)} \int_0^{t_{\ell-1}} k(t_{j+\ell-1}, s+t_j) \, ds. \end{aligned}$$

Thus (3.1) becomes

$$\begin{aligned} \int_0^{t_j} \left(\sum_{\ell=1}^{\gamma_j+1} b_\ell^{(j)} k(t_j + t_{\ell-1}, s) \right) u_h(s) \, ds + \left[\sum_{\ell=1}^{\gamma_j+1} b_\ell^{(j)} \int_0^{t_{\ell-1}} k(t_j + t_{\ell-1}, s+t_j) ds + \mu_j \right] u_h(t_j) \\ = \sum_{\ell=1}^{\gamma_j+1} b_\ell^{(j)} f(t_j + t_{\ell-1}), \end{aligned}$$

or

$$\int_0^{t_j} \tilde{k}(t_j, s; r, h) u_h(s) \, ds + [\alpha(t_j; r, h) + \mu(t_j)] u_h(t_j) = \tilde{f}(t_j; r, h), \tag{3.2}$$

for $j = 1, \ldots, N$, where for $t \in [0,1]$,

$$\tilde{k}(t, s; r, h) := \int_0^{r(t)} k(t+\rho, s) \, d\eta(\rho; t), \tag{3.3}$$

$$\alpha(t; r, h) \quad := \int_0^{r(t)} \int_0^{\rho} k(t+\rho, s+t) \, ds \, d\eta(\rho; t), \tag{3.4}$$

$$\tilde{f}(t; r, h) \quad := \int_0^{r(t)} f(t+\rho) \, d\eta(\rho; t). \tag{3.5}$$

Here, for each $t \in [0,1]$, $\eta(\cdot; t)$ is an (r,h)-dependent Borel measure on $[0, r(t)]$ defined via

$$\int_0^{r(t)} g(\rho) \, d\eta(\rho; t) := \sum_{\ell=1}^{K(t)} s_\ell(t) \, g(t_{\ell-1}), \tag{3.6}$$

for g a Borel function on $[0, r(t)]$, where the h-dependent functions K and s_ℓ are given by

$$K(t) := r(t)/h + 1, \tag{3.7}$$

$$s_\ell(t) := \int_0^{t_\ell} k(t + t_{\ell-1}, t + (s-h)) \, ds, \quad \ell = 1, \ldots, K(t). \tag{3.8}$$

The equivalence of (3.1) and (3.2) results from the fact that $s_\ell(t_j) = b_\ell^{(j)}$, for $j = 1, \ldots, N$, and thus $\int_0^{r(t_j)} g(\rho)\, d\eta(\rho; t_j) = \sum_{\ell=1}^{\gamma_j+1} b_\ell^{(j)} g(t_{\ell-1})$, $j = 1, \ldots, N$. We summarize our findings in the following lemma.

Lemma 3.1 *Let r and μ be given by (2.5)–(2.6) and (2.9), respectively. The problem of finding c_j solving the penalized predictor-corrector algorithm (2.10)–(2.11) for $j = 1, \ldots, N$, is equivalent to the problem of seeking u_h of the form (2.1) which solves the Volterra equation*

$$\int_0^t \tilde{k}(t, s; r, h) u(s)\, ds + [\alpha(t; r, h) + \mu(t)] u(t) = \tilde{f}(t; r, h), \tag{3.9}$$

precisely at collocation points $t = t_j$, $j = 1, \ldots, N$. The quantities $\tilde{k}$, α, and $\tilde{f}$ are defined in (3.3), (3.4), and (3.5), respectively, and η is given by (3.6).

Under our standing assumptions on k, the coefficient of $u(t)$ in (3.9) is nonzero and relevant quantities in that equation are square-integrable; thus the above lemma gives that the predictor-corrector algorithm is a collocation-based discretization of a well-posed second-kind Volterra equation. This is in contrast to the unregularized algorithm (2.3)–(2.4), which results from a collocation-based discretization of the original ill-posed first-kind Volterra equation (1.1).

The selection of η above can be generalized, as can the choices of r and μ. We do this in the next section, and make rigorous the assumptions needed in the more general framework. Theoretical convergence arguments will also be constructed in this setting, with convergence results given in Sections 4–5.

3.2. Definitions and hypotheses

For a generalization of the method presented in Section 2, we let $T > 1$ and k be given as in Section 2.2, let $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ denote a normed linear space of functions defined on $[0, T]$, and let the subspace $\mathcal{F}_D$ of $\mathcal{F}$ denote the admissible space of data functions, where it is assumed that all $g \in \mathcal{F}_D$ are Borel functions. We assume that the data f defined in (1.1) belongs to $\mathcal{F}_D$ and is such that (1.1) has a unique solution $\bar{u} \in C^1[0, T]$. The perturbation f^δ of f will be assumed to be such that $f^\delta \in \mathcal{F}_D$, where f^δ is close to f in an appropriate sense.

For a functional local regularization parameter r of more general form than that considered in the last section, we assume $r \in \Gamma$, where

$$\Gamma := \{r : [0,1] \to \mathbb{R} \,:\, r \text{ piecewise continuous}, \min_{t\in[0,1]} r(t) > 0,\ \max_{t\in[0,1]} t + r(t) \le T\}.$$

For $r \in \Gamma$, we will use the notation $r_{min} := \min_{t\in[0,1]} r(t) > 0$ and $\|r\|_\infty := \max_{t\in[0,1]} r(t) \le T$. Corresponding to $r \in \Gamma$ and $h > 0$ we make the following definition of a family $\mathcal{N} = \mathcal{N}(r, h)$ of measures which are compatible with r and h in a specific sense.

Definition 3.1 *Given $r \in \Gamma$ and $h > 0$, we say that the one-parameter family $\mathcal{N} = \{\eta(\cdot;t), t \in [0,1]\}$ is an* (r,h)-suitable family of measures *if, for each $t \in [0,1]$, $\eta(\cdot;t)$ is a finite, positive Borel measure defined on $[0, r(t)]$ satisfying*

$$\int_0^{r(t)} \rho \, d\eta(\rho;t) > 0, \quad t \in [0,1], \tag{3.10}$$

and for which

$$\int_0^{r(t)} f(t+\rho) \, d\eta(\rho;t) \text{ is well-defined for all } f \in \mathcal{F}_D, \ t \in [0,1]. \tag{3.11}$$

We note that, in general, the family $\mathcal{N}$ depends on both r and h, as well as the selection of $\mathcal{F}$. We give two examples of (r,h)-suitable $\mathcal{N}$ below. In each case we will define $\eta(\cdot,t)$ on $[0, \|r\|_\infty]$ for $t \in [0,1]$, where it will be understood that $\int_0^{r(t)} g(\rho)\, d\eta(\rho;t) := \int_0^{\|r\|_\infty} g(\rho)\chi_{[0,r(t)]}(\rho)\, d\eta(\rho;t)$, $t \in [0,1]$, where $\chi_{[0,r(t)]}$ is the characteristic function on $[0, r(t)]$.

Our first example is a generalization of the family of measures defined via (3.6)–(3.8).

Example 3.1 Let $h = 1/N$ and $r \in \Gamma$ be given, and suppose $\mathcal{F}_D \subseteq \mathcal{F} := \{f : [0,T] \to \mathbb{R} : \|f\|_{\mathcal{F}} := \sup_{t\in[0,T]} |f(t)| < \infty\}$. Define $\mathcal{N} = \{\eta(\cdot,t), t \in [0,1]\}$ where for each $t \in [0,1]$,

$$\int_0^{\|r\|_\infty} g(\rho)\, d\eta(\rho;t) := \sum_{\ell=1}^{K} s_\ell(t) g(\tau_\ell)$$

for bounded Borel-measurable g on $[0, \|r\|_\infty]$, where the (r,h)-dependent parameters K, s_ℓ, τ_ℓ, satisfy $0 < K < \infty$, K integer; $0 < s_\ell(t) \le \|s_\ell\|_\infty < \infty$, $t \in [0,1]$, $\ell = 1, \ldots, K$; and $0 \le \tau_1 < \tau_2 < \ldots < \tau_K \le \|r\|_\infty$, with $\tau_\ell \in (0, r_{min}]$ for some ℓ. It then follows that $\mathcal{N}$ is an (r,h)-suitable family of measures.

Example 3.2 Let h, r, $\mathcal{F}_D$, $\mathcal{F}$, and g be as in Example 3.1, and assume that the (r,h)-dependent function ω satisfies $0 < \underline{\omega} \le \omega(\rho,t) \le \bar{\omega} < \infty$, a.a. $(\rho,t) \in [0, \|r\|_\infty] \times [0,T]$. Then if $\mathcal{N} = \{\eta(\cdot;t), t \in [0,1]\}$, where for each $t \in [0,1]$,

$$\int_0^{\|r\|_\infty} g(\rho) d\eta(\rho;t) := \int_0^{\|r\|_\infty} g(\rho)\omega(\rho,t) d\rho, \tag{3.12}$$

it follows that $\mathcal{N}$ is an (r,h)-suitable family of measures.

Given noisy data f^δ, the discrete regularization problem $\mathcal{P}_h^\delta$ (which generalizes the problem described in Section 2 and in Lemma 3.1) is as follows:

Definition 3.2 *Let $f^\delta \in \mathcal{F}_D$ and $\mu : [0,1] \to [0,\infty)$ be specified. Given $h = 1/N$ and $r \in \Gamma$, let $\mathcal{N} = \{\eta(\cdot;t), t \in [0,1]\}$ denote an (r,h)-suitable family of measures. We define the* discrete regularization problem, *denoted by $\mathcal{P}_h^\delta \equiv \mathcal{P}_h^\delta(r;\mu;\mathcal{N})$, to be the problem of* *determining* $u_h^\delta = u_h^\delta(\cdot; r; \mu; \mathcal{N})$*, a step-function of the form (2.1), which satisfies the regularization equation*

$$\int_0^t \tilde{k}(t, s; r, h) u(s)\, ds + [\alpha(t; r, h) + \mu(t)] u(t) = \tilde{f}^\delta(t; r, h), \tag{3.13}$$

exactly at collocation points $t = t_j$, $j = 1, \ldots, N$*. We will also use the notation* $\mathcal{P}_h \equiv \mathcal{P}_h(r; \mu; \mathcal{N})$ *to designate the same problem as above, but with* f^δ *replaced by the true data* f*.*

In the above definition, $\tilde{k}$ and α are given by (3.3) and (3.4), respectively, while $\tilde{f}^\delta$ is defined by (3.5) with f^δ used in place of f. Well-posedness of the discrete regularization problem $\mathcal{P}_h^\delta$ is guaranteed by the following theorem.

Theorem 3.1 *Let* $f^\delta \in \mathcal{F}_D$ *and* $\mu : [0,1] \to [0, \infty)$ *be specified. Given* $h = 1/N$ *and* $r \in \Gamma$*, let* $\mathcal{N} = \{\eta(\cdot; t), t \in [0,1]\}$ *denote an* (r,h)*-suitable family of measures. Then if* $\|r\|_\infty$ *and* h *are sufficiently small, there is a unique solution* $u_h^\delta = u_h^\delta(\cdot; r; \mu; \mathcal{N})$ *of the discrete regularization problem* $\mathcal{P}_h^\delta = \mathcal{P}_h^\delta(r; \mu; \mathcal{N})$*.*

Proof: Substituting (2.1) into (3.13) and evaluating at t_j, we have

$$\sum_{i=1}^{j} c_i \int_{t_{i-1}}^{t_i} \tilde{k}(t_j, s; r, h)\, ds + [\alpha(t_j; r, h) + \mu(t_j)] c_j = \tilde{f}^\delta(t_j; r, h), \quad j = 1, \ldots N, \tag{3.14}$$

a lower-triangular linear system in the vector $(c_1, \ldots, c_N)^\top$, with diagonal elements in the governing matrix given by

$$\begin{aligned}
&\int_{t_{j-1}}^{t_j} \tilde{k}(t_j, s; r, h)\, ds + \alpha(t_j; r, h) + \mu(t_j) \\
&\qquad = \int_{t_{j-1}}^{t_j} \int_0^{r(t_j)} k(t_j + \rho, s)\, d\eta(\rho; t_j)\, ds + \int_0^{r(t_j)} \int_0^{\rho} k(t_j + \rho, s + t_j)\, ds\, d\eta(\rho; t_j) + \mu(t_j),
\end{aligned}$$

for $j = 1, \ldots, N$. But we have assumed that $k \in C^1$ has been normalized so that $k(t,t) = 1$ for $t \in [0,1]$, so it follows that for h and $\|r\|_\infty$ sufficiently small, the integrands in the first two terms on the right above are positive, bounded below by some $\underline{k} = \underline{k}(r,h) > 0$. Thus, for $j = 1, \ldots, N$, $\int_{t_{j-1}}^{t_j} \int_0^{r(t_j)} k(t_j + \rho, s)\, d\eta(\rho; t_j)\, ds \geq 0$, and $\int_0^{r(t_j)} \int_0^{\rho} k(t_j + \rho, s + t_j)\, ds\, d\eta(\rho; t_j) \geq \underline{k} \int_0^{r(t_j)} \rho\, d\eta(\rho; t_j) > 0$, where we have used the assumption (3.10) on $\eta(\cdot, t)$. Therefore, the lower-triangular matrix system determined by (3.14) has a unique solution $(c_1, \ldots, c_N)^\top$. □

In the next section we focus on the problem of convergence for discrete approximations of equation (3.13). Before doing so, it is worth noting that if no discretization is performed, then equation (3.13) may alternatively be used to define a *continuous* regularization method. This idea is pursued in [23] and there one may find conditions guaranteeing well-posedness of the continuous regularization problem associated with (3.13), along with convergence results depending on choices of the functional regularization parameters $r = r(\cdot; \delta)$ and $\mu(\cdot; \delta)$ as $\delta \to 0$. The theory in [23] serves to generalize the continuous regularization ideas in [19] to nonconvolution kernels and, more importantly, to the case of a variable regularization parameter r.

4. Convergence results

Throughout this section we will assume $f \in \mathcal{F}_D$, k, and $\bar{u}$ satisfy the assumptions made at the beginning of Section 3.2. Let $h = 1/N$ for $N = 1, 2, \ldots$. We are interested here in the limiting behavior as $h \to 0$ of solutions u_h of the discrete regularization problem $\mathcal{P}_h$, given

- a sequence $(r_h) \subset \Gamma$ of functional local regularization parameters;
- a sequence $(\mathcal{N}_h)$, where each $\mathcal{N}_h = \{\eta_h(\cdot, t), t \in [0,1]\}$ is an (r_h, h)-suitable family of measures; and,
- a sequence (μ_h) of functional penalty parameters, with $\mu_h(t) \geq 0$, $t \in [0,1]$.

In Theorem 5.1 and Corollary 5.1, we will state convergence results under fairly general conditions on the above quantities. We will also examine convergence of u_h^δ (the solution of problem $\mathcal{P}_h^\delta$) to $\bar{u}$, given $f^\delta \in F_D$, $|f(t) - f^\delta(t)| \leq \delta$, $t \in [0,T]$, and under conditions relating h to δ as $\delta \to 0$.

Before turning to the main convergence theorem (for which the statement of results becomes somewhat technical), we first describe a couple of useful special cases of these findings. The first special case requires one of the following conditions on local regularization parameters $r_h(\cdot)$, given a positive constant M_r:

(1a) The parameters $r_h \in \Gamma$ are constant functions given by

$$r_h(t) = C_h h, \quad t \in [0,1],$$

where $0 < C_h \leq M_r < \infty$ for all h sufficiently small;

(1b) The parameters $r_h \in \Gamma$ are functions satisfying

$$r_h(t) \leq M_r h, \quad t \in [0,1],$$

for all h sufficiently small, where $M_r < 1$.

When condition (1a) holds we will need the following condition on the families $\mathcal{N}_h$ of measures:

(2a) For some $\varepsilon > 0$ and each $h > 0$, there exists a finite, positive (t-independent) Borel measure $\bar{\eta}_h$ on $[0, \varepsilon]$ for which (3.10) and (3.11) hold, and such that, for $m = 0, 1$ and $\eta_h(\cdot; t) \in \mathcal{N}_h$

$$\int_0^{r_h(t)} \rho^m \, d\eta_h(\rho; t) = (1 + \bar{w}_m(t; h)) \int_0^{r_h(t)} \rho^m \, d\bar{\eta}_h(\rho), \quad t \in [0,1],$$

where $\|\bar{w}_m(\cdot; h)\|_\infty = \mathcal{O}(h)$ as $h \to 0$.

Finally, we require the following condition on the penalty regularization parameters μ_h:

(3) For some $M_\mu > 0$, the penalty parameters μ_h satisfy

$$0 \leq \mu_h(t) \leq h^2 M_\mu \int_0^{r_h(t)} d\eta_h(\rho; t), \quad t \in [0,1],$$

for all h sufficiently small.

We note that condition (3) relates the size of μ_h to h, r_h, and η_h, and allows for the possibility of $\mu_h \equiv 0$ for all $h = 1/N$, $N = 1, 2, \ldots$. More general conditions on μ_h are allowed, as can be seen in Section 5.

Remark 4.1 Condition (2a) on η_h requires that the measures be approximately t-independent (in some sense) for all h small. Although technical, this condition is satisfied by the families of measures most commonly used in practice. Indeed, the discrete measure defined via (3.6)–(3.8) (which is the measure associated with the predictor-corrector algorithms of Section 2) can be seen to satisfy this condition provided we make the natural assumption (which is more general than (1a)) that $r_h \in \Gamma$ is of the form $r_h(t) = \gamma(t)h$, where γ is a fixed nonnegative integer-valued function on $[0, 1]$. (We note that we may equivalently assume, as in Section 2, that $r_h \in \Gamma$ is of the form $r_h(t) = \sum_{i=1}^{N} \gamma(t_i)h\, \chi_i(t)$ since only $r_h(t_j)$, $j = 1, \ldots, N$, will be required in constructing the solution of $\mathcal{P}_h^\delta$.) In this case, using a Taylor expansion on k, the functions $s_\ell = s_\ell(t)$, $\ell = 1, \ldots K = \|\gamma\|_\infty + 1$, in (3.8) satisfy

$$s_\ell(t) = \int_0^{t_\ell} \left[k(t,t) + t_{\ell-1} D_1 k(\zeta_1, \zeta_2) + (s - h) D_2 k(\zeta_1, \zeta_2)\right] ds,$$

where $\zeta_i = \zeta_i(t, s, h, \ell)$, so that $s_\ell(t) = \ell h(1 + \hat{s}_\ell(t, h))$, $\ell = 1, \ldots, K$, with $\|\hat{s}_\ell(\cdot, h)\|_\infty = \mathcal{O}(h)$ as $h \to 0$. Thus, the "approximate t-independence" of the quantity $s_\ell(\cdot)$ in η_h is sufficient to argue that condition (2a) above holds for the associated family of measures. It is also not difficult to see how other families of measures (such as those in Examples 3.1–3.2) may be constructed in order to easily satisfy condition (2a) above.

The next theorem follows immediately from Theorem 5.1 and Corollary 5.1, both of which are proven in the next section.

Theorem 4.1 *For $h = 1/N$, $N = 1, 2, \ldots$, assume the parameters $r_h \in \Gamma$ satisfy either conditions (1a) (in which case the (r_h, h)-suitable families $\mathcal{N}_h$ of measures satisfy condition (2a)) or (1b), and that the parameters μ_h satisfy condition (3). Then the solution $u_h(\cdot) = u_h(\cdot; r_h; \eta_h; \mu_h)$ of the discretization problem $\mathcal{P}_h$ converges to $\bar{u}(\cdot)$ uniformly at collocation points t_j, $j = 1, \ldots, N$, as $h \to 0$.*

If, in addition, $f^\delta \in \mathcal{F}_D$ satisfies $|f(t) - f^\delta(t)| \leq \delta$, $t \in [0, T]$, and $h = h(\delta)$ is selected so that

$$\delta / h^2(\delta) \leq M$$

and $h(\delta) \to 0$ as $\delta \to 0$, then the solution $u_h^\delta(\cdot) = u_h^\delta(\cdot; r_h; \eta_h; \mu_h)$ of discretization problem $\mathcal{P}_h^\delta$ converges to $\bar{u}(\cdot)$ uniformly at collocation points t_j, $j = 1, \ldots, N(\delta)$, as $\delta \to 0$. This convergence is at the best possible rate with respect to δ, that is,

$$|u_h^\delta(t_j) - \bar{u}(t_j)| \leq K\delta^{1/2} + \mathcal{O}(\delta), \quad j = 1, \ldots, N(\delta)$$

as $\delta \to 0$, where $K > 0$ is independent of h and δ.

We note that when r_h is specified to satisfy condition (1a) in Theorem 4.1, this theorem generalizes the results in [18] to a penalized (i.e., $\mu \neq 0$) predictor-corrector method, to nonconvolution kernels k, and to the case of more general measures.

The above result is fairly limited for truly variable r_h in that condition (1b) implies that $r_h(t) < h$ for all $t \in [0,1]$. In practical numerical calculations, we are interested in using r_h as in Section 2.1 and Remark 4.1, e.g., r_h of the form $r_h(t) = \gamma(t)h$ where $\gamma(t)$ is integer-valued for $t \in [0,1]$; such a choice allows us to coordinate the length $r_h(t_j)$ of the j^{th} future regularization interval with the discretization stepsize. But condition (1b) requires $0 < \gamma(t) < 1$, $t \in [0,1]$, thus ruling out integer-valued γ. However, in fact we can allow integer-valued γ under an additional condition on the measures η_h, a property which is satisfied by measures of practical interest in computations. We give a definition prescribing this "p-condition" below, and note that the condition is of greatest interest when $p \geq 1$ is an integer.

Definition 4.1 *Let $h = 1/N$, $N = 1, 2, \ldots$. Suppose we are given a sequence $(r_h) \subset \Gamma$ with $|r_h(t)| \leq M_r h$, $t \in [0,1]$, for some $M_r > 0$, and assume that for each h we have an associated (r_h, h)-suitable family $\mathcal{N}_h = \{\eta_h(\cdot; t), t \in [0,1]\}$ of measures. For $p > 0$, we say that $(\mathcal{N}_h)$* satisfies a p-condition with respect to (r_h) *if there is $b_1 \in (0, 1/p)$, $C_1 > 0$, so that, for all h sufficiently small,*

$$\frac{\int_0^{r_h(t)} \rho \, d\eta_h(\rho; t)}{\int_0^{r_h(t)} d\eta_h(\rho; t)} = b_1 r_h(t) + c_1(t; h), \quad t \in [0,1], \tag{4.1}$$

where

$$|c_1(t; h)| \leq C_1 h^2, \quad t \in [0,1].$$

We consider in Examples 4.1 and 4.2 below some measures standardly used in computations, and demonstrate that these classes of measures satisfy a p-condition with $p = 1$ for reasonable choices of (r_h). But before giving these examples, we state a theorem which shows how we may relax conditions on (r_h) when the p-condition is satisfied by (η_h).

Theorem 4.2 *Suppose that the sequence $(r_h) \subset \Gamma$ is given, with*

$$r_h(t) = \gamma(t)h, \quad 0 < \gamma_{min} \leq \gamma(t) \leq M_r, \tag{4.2}$$

for some γ_{min}, M_r, and all $t \in [0,1]$. Suppose further that there is a corresponding sequence $(\mathcal{N}_h)$ of families of measures, each (r_h, h)-suitable, which satisfies a p-condition with respect to (r_h) for some $p > 0$. Then if

$$\gamma(t^+) \geq \gamma(t^-) - p, \quad t \in [0,1], \tag{4.3}$$

and if penalty parameters μ_h satisfy condition (3), then the conclusion of the first part of Theorem 4.1 holds.

If, in addition, $f^\delta \in \mathcal{F}_D$ satisfies $|f(t) - f^\delta(t)| \leq \delta$, $t \in [0,T]$, and $h = h(\delta)$ is selected so that

$$\delta / h^2(\delta) \leq M$$

and $h(\delta) \to 0$ as $\delta \to 0$, then the conclusions of the second part of Theorem 4.1 hold.

The significance of this theorem (the proof of which appears in Section 5) is clearly seen when $p \geq 1$ is an integer. In this case condition (4.3) allows γ to be integer valued with "decreasing jumps" of at most p, and no limit on "increasing jumps" (other than the fact that γ must remain finite on $[0,1]$). In other words, as t increases, one may increase the level of regularization quickly, but one must decrease it more deliberately.

Below we give some examples of standard measures satisfying p-conditions for $p = 1$, and a final example where the p-condition may be satisfied using arbitrary $p \geq 2$.

Example 4.1 Let $(r_h) \subset \Gamma$ be given satisfying (4.2), with piecewise continuous, integer-valued γ, and, for each (r_h, h), let $\mathcal{N}_h = \{\eta_h(\cdot; t), t \in [0,1]\}$ where η_h is given by (3.6)–(3.8) (which is the standard measure associated with the predictor-corrector scheme described in Section 2, and a special case of Example 3.1). From Remark 4.1 we have that $s_\ell(t) = \ell h(1 + \mathcal{O}(h))$, $t \in [0,1]$, for $\ell = 1, \ldots, K$. Thus,

$$\begin{aligned}\frac{\int_0^{r_h(t)} \rho \, d\eta_h(\rho; t)}{\int_0^{r_h(t)} d\eta_h(\rho; t)} &= \left(\sum_{\ell=1}^{\gamma(t)+1} s_\ell((\ell-1)h)\right)\left(\sum_{\ell=1}^{\gamma(t)+1} s_\ell\right)^{-1} \\ &= h\,(1 + \mathcal{O}(h))\left(\sum_{\ell=1}^{\gamma(t)+1} \ell(\ell-1)\right)\left(\sum_{\ell=1}^{\gamma(t)+1} \ell\right)^{-1} \\ &= (2/3)\, h\gamma(t) + \mathcal{O}(h^2).\end{aligned}$$

It follows that $(\mathcal{N}_h)$ satisfies (4.1) with $b_1 = 2/3$ and thus with $p = 1$.

Example 4.2 Let $\mathcal{F}_D$ and g be as in Example 3.1, let $(r_h) \subset \Gamma$ with $\|r_h\|_\infty \leq M_r h$, and suppose $\mathcal{N}_h = \{\eta_h(\cdot; t), t \in [0,1]\}$, where η_h is defined (similar to Example 3.2) for each $h = 1/N$, $N = 1, 2, \ldots$, by

$$\int_0^{\|r_h\|_\infty} g(\rho) d\eta_h(\rho; t) := \int_0^{\|r_h\|_\infty} g(\rho)\omega_h(\rho, t)\, d\rho. \tag{4.4}$$

Here we assume $\omega_h(\rho, t) = \hat{\omega}(\rho)(1 + \omega_0(\rho, t; h))$ for fixed $\hat{\omega} \in C^1[0, T]$, $0 < \underline{\hat{\omega}} \leq \hat{\omega}(\rho)$, $\rho \in [0, T]$, and for $\|\omega_0(\cdot, \cdot; h)\|_\infty = \mathcal{O}(h)$. Then $(\mathcal{N}_h)$ satisfies a p-condition with $p = 1$. Indeed, for all $t \in [0,1]$, a Taylor expansion of $\hat{\omega}(\rho)$ about $\rho = 0$ gives

$$\begin{aligned}\frac{\int_0^{r_h(t)} \rho\, \omega_h(\rho, t)\, d\rho}{\int_0^{r_h(t)} \omega_h(\rho, t)\, d\rho} &= \frac{\hat{\omega}(0) \int_0^{r_h(t)} \rho\, d\rho}{\hat{\omega}(0) \int_0^{r_h(t)} d\rho}\,(1 + \mathcal{O}(h)) \\ &= (1/2)\, r_h(t) + \mathcal{O}(h^2)\end{aligned}$$

as $h \to 0$. Thus (4.1) is satisfied with $b_1 = 1/2$, and $(\mathcal{N}_h)$ satisfies a p-condition with $p = 1$.

Example 4.3 Let $p \geq 2$ be arbitrary. Following the ideas of the last example (with the same $\mathcal{F}_D$, g, and (r_h)), we may construct families $\mathcal{N}_h$ of measures satisfying a p-condition for this prescribed p if we define η_h via (4.4) using instead an unbounded $\hat{\omega}$, here given by $\hat{\omega}(\rho) = \rho^{-m/(m+1)}$, $\rho \in (0, 1]$, for $m > p - 2$. Then it's not difficult to show that (4.1) holds with $b_1 = 1/(m+2) < 1/p$.

5. Proofs of convergence

The proofs of Theorems 4.1 and 4.2 follow from the results below and will be presented at the end of this section. Our main convergence theorem, a rather technical result, is given first.

Theorem 5.1 *Suppose $f \in \mathcal{F}_D$, k, and $\bar{u}$ satisfy the conditions at the beginning of Section 3.2, and let ε_r, M_r, and M_μ be fixed positive scalars. For each $N = 1, 2, \ldots$, let $h = 1/N$ and suppose we are given $r_h \in \Gamma$ and an (r_h, h)-suitable family $\mathcal{N}_h = \{\eta_h(\cdot; t), t \in [0,1]\}$ of measures for which*

$$r_h(t) \leq M_r\, h, \quad t \in [0,1], \tag{5.1}$$

and

$$\frac{\int_0^{r_h(t_{j+1})} \rho\, d\eta_h(\rho; t_{j+1})}{h \int_0^{r_h(t_{j+1})} d\eta_h(\rho; t_{j+1})} - \frac{\int_0^{r_h(t_j)} \rho\, d\eta_h(\rho; t_j)}{h \int_0^{r_h(t_j)} d\eta_h(\rho; t_j)} + 1 \geq \varepsilon_r, \quad j = 0, \ldots, N-1, \tag{5.2}$$

for all N sufficiently large. Then if $\mu_h : [0,1] \mapsto [0, \infty)$ is selected satisfying

$$\frac{1}{h}\left(\frac{\mu_h(0)}{\int_0^{r_h(0)} d\eta_h(\rho; 0)}\right) \leq M_\mu, \tag{5.3}$$

$$\frac{1}{h^2}\left|\frac{\mu_h(t_{j+1})}{\int_0^{r_h(t_{j+1})} d\eta_h(\rho; t_{j+1})} - \frac{\mu_h(t_j)}{\int_0^{r_h(t_j)} d\eta_h(\rho; t_j)}\right| \leq M_\mu, \quad j = 0, \ldots, N-1, \tag{5.4}$$

for all h sufficiently small, and if $\bar{u}(0) = 0$, the solution $u_h(\cdot) = u_h(\cdot; r_h; \mu_h; \mathcal{N}_h)$ of the discretization problem $\mathcal{P}_h$ converges to $\bar{u}(\cdot)$ uniformly at collocation points $t = t_j$, $j = 1, \ldots, N$, as $h \to 0$.

If, in addition, $f^\delta \in \mathcal{F}_D$ satisfies $|f(t) - f^\delta(t)| \leq \delta$, $t \in [0, T]$, and $h = h(\delta)$ is selected so that

$$\delta / h^2(\delta) \leq M$$

and $h(\delta) \to 0$ as $\delta \to 0$, then the solution $u_h^\delta(\cdot) = u_h^\delta(\cdot; r_h; \mu_h; \mathcal{N}_h)$ of discretization problem $\mathcal{P}_h^\delta$ converges to $\bar{u}(\cdot)$ uniformly at collocation points t_j, $j = 1, \ldots, N(\delta)$, as $\delta \to 0$. This convergence is at the best possible rate with respect to δ, that is,

$$|u_h^\delta(t_j) - \bar{u}(t_j)| \leq C\delta^{1/2} + \mathcal{O}(\delta), \quad j = 1, \ldots, N(\delta)$$

as $\delta \to 0$, where $C > 0$ is independent of h and δ.

Proof: Let $N = 1, 2, \ldots$, and $h = 1/N$. Let $d(t) := f^\delta(t) - f(t)$, $t \in [0,1]$. Since the sequences (r_h) and (μ_h) are indexed by the discretization parameter h, we will simplify notation throughout the proof by writing $\alpha(t; h) := \alpha(t; r_h, h)$, $\tilde{f}^\delta(t; h) := \tilde{f}^\delta(t; r_h, h)$, $t \in [0,1]$, and $\tilde{k}(t, s; h) := \tilde{k}(t, s; r_h, h)$, for $0 \leq s \leq t \leq 1$. In addition, it will be useful to define, for $t \in [0,1]$,

$$\begin{aligned} \nu_h(t) \quad &:= (\alpha(t; h) + \mu_h(t)) / a_0(t; h), \\ a_m(t; h) &:= \int_0^{r_h(t)} \rho^m\, d\eta_h(\rho, t), \end{aligned} \tag{5.5}$$

for $m = 0, 1$. Clearly $a_0(t;h) \geq a_1(t;h)/\|r_h\|_\infty > 0$ for all $t \in [0,1]$, $h = 1/N$, $N = 1, 2, \ldots$. We will also make the definitions

$$\tilde{k}_0(t,s;h) := \tilde{k}(t,s;h)/a_0(t;h), \quad f_0^\delta(t) := f^\delta(t)/a_0(t;h)$$

with similar definitions for f_0, d_0, and $\bar{u}_0$.

Using these definitions, (3.14) may be written, after a division by $a_0(t_j;h)$ as

$$\sum_{i=1}^{j} c_i \int_{t_{i-1}}^{t_i} \tilde{k}_0(t_j,s;h)\,ds + \nu_h(t_j)c_j = \int_0^{r_h(t_j)} f_0^\delta(t_j+\rho)\,d\eta_h(\rho;t_j), \tag{5.6}$$

for $j = 1, \ldots, N$.

We will use a differencing technique similar to that used in [18] to analyze convergence. To this end, we replace j in equation (5.6) by $j+1$ (for $j = 0, \ldots, N-1$) and subtracting (5.6) from the resulting equation we obtain

$$\begin{aligned}
\nu_h(t_{j+1})c_{j+1} &+ \int_{t_j}^{t_{j+1}} \tilde{k}_0(t_{j+1},s;h)c_{j+1}\chi_{j+1}(s)\,ds \\
&= \int_0^{r_h(t_{j+1})} [f_0(t_{j+1}+\rho) + d_0(t_{j+1}+\rho)]\; d\eta_h(\rho;t_{j+1}) \\
&\quad - \int_0^{r_h(t_j)} [f_0(t_j+\rho) + d_0(t_j+\rho)]\; d\eta_h(\rho;t_j) + \nu_h(t_j)c_j \\
&\quad - \sum_{i=1}^{j} \int_{t_{i-1}}^{t_i} \left[\tilde{k}_0(t_{j+1},s;h) - \tilde{k}_0(t_j,s;h)\right] c_i\chi_i(s)\,ds,
\end{aligned} \tag{5.7}$$

for $j = 1, \ldots, N-1$.

For given $h = 1/N$ and r_h, the true solution $\bar{u}$ of (1.1) satisfies

$$\begin{aligned}
&\int_0^t \tilde{k}(t,s;h)\bar{u}(s)\,ds + \int_0^{r_h(t)} \int_0^\rho k(t+\rho, s+t)\,[\bar{u}(s+t) - \bar{u}(t)]\; ds\,d\eta_h(\rho;t) \\
&\qquad + [\alpha(t;h) + \mu_h(t)]\bar{u}(t) = \int_0^{r_h(t)} f(t+\rho)\,d\eta_h(\rho;t) + \mu_h(t)\bar{u}(t),
\end{aligned} \tag{5.8}$$

for all $t \in [0,1]$. We evaluate (5.8) at $t = t_j$ and divide through by $a_0(t_j;h)$, and then subtract the resulting equation from the one arising from evaluation at $t = t_{j+1}$ and division by $a_0(t_{j+1};h)$. Then, subtracting (5.7) from the resulting equation in $\bar{u}$, the result is (for $j = 1, \ldots, N-1$)

$$\begin{aligned}
\nu_h(t_{j+1})\,[\bar{u}(t_{j+1}) - c_{j+1}] &+ \int_{t_j}^{t_{j+1}} \tilde{k}_0(t_{j+1},s;h)[\bar{u}(s) - c_{j+1}\chi_{j+1}(s)]\,ds \\
&= \nu_h(t_j)[\bar{u}(t_j) - c_j] + \mu_h(t_{j+1})\bar{u}_0(t_{j+1}) - \mu_h(t_j)\bar{u}_0(t_j) \\
&\quad - \int_0^{r_h(t_{j+1})} d_0(t_{j+1}+\rho)\,d\eta_h(\rho;t_{j+1}) + \int_0^{r_h(t_j)} d_0(t_j+\rho)\,d\eta_h(\rho;t_j) \\
&\quad - \sum_{i=1}^{j} \int_{t_{i-1}}^{t_i} \left[\tilde{k}_0(t_{j+1},s;h) - \tilde{k}_0(t_j,s;h)\right] [\bar{u}(s) - c_i\chi_i(s)]\; ds
\end{aligned}$$

$$
\begin{aligned}
&- \int_0^{r_h(t_{j+1})} \int_0^{\rho} k(t_{j+1}+\rho, s+t_{j+1}) \frac{\bar{u}(s+t_{j+1}) - \bar{u}(t_{j+1})}{a_0(t_{j+1}; h)}\, ds\, d\eta_h(\rho; t_{j+1}) \\
&+ \int_0^{r_h(t_j)} \int_0^{\rho} k(t_j+\rho, s+t_j) \frac{\bar{u}(s+t_j) - \bar{u}(t_j)}{a_0(t_j; h)}\, ds\, d\eta_h(\rho; t_j). \qquad (5.9)
\end{aligned}
$$

Using a Taylor expansion we can write, for $t \in (t_j, t_{j+1}]$ and $j = 0, \ldots, N-1$,

$$
\bar{u}(t) - c_{j+1}\chi_{j+1}(t) = h\left[\beta_{j+1} + h^{-1}(t - t_{j+1})\bar{u}'(z_{j+1}(t))\right], \qquad (5.10)
$$

for some $z_{j+1}(t)$ between t and t_{j+1}, and where $\beta_{j+1} := h^{-1}(\bar{u}(t_{j+1}) - c_{j+1})$. Using this expansion and dividing through by h, (5.9) becomes

$$
\beta_{j+1} = W_j(h)\beta_j - h\sum_{i=1}^{j} V_{j,i}(h)\beta_i - \frac{1}{h^2}E_j(\delta, h) - Z_j(h), \qquad (5.11)
$$

for $j = 1, \ldots, N-1$, where

$$
\begin{aligned}
D_j(h) &:= \nu_h(t_{j+1}) + \int_{t_j}^{t_{j+1}} \tilde{k}_0(t_{j+1}, s; h)\, ds, && (5.12)\\
W_j(h) &:= \nu_h(t_j)/D_j(h), && (5.13)\\
E_j(\delta, h) &:= \frac{h}{D_j(h)}\left[\int_0^{r_h(t_{j+1})} d_0(t_{j+1}+\rho)\, d\eta_h(\rho; t_{j+1}) - \int_0^{r_h(t_j)} d_0(t_j+\rho)\, d\eta_h(\rho; t_j)\right], && (5.14)\\
V_{j,i}(h) &:= (hD_j(h))^{-1}\int_{t_{i-1}}^{t_i}\left[\tilde{k}_0(t_{j+1}, s; h) - \tilde{k}_0(t_j, s; h)\right] ds, && (5.15)\\
Z_j(h) &:= (hD_j(h))^{-1}R_j(h), && (5.16)\\
R_j(h) &:= \int_0^{r_h(t_{j+1})} \int_0^{\rho} k(t_{j+1} + \rho, s + t_{j+1}) \frac{\bar{u}(s + t_{j+1}) - \bar{u}(t_{j+1})}{a_0(t_{j+1}; h)}\, ds\, d\eta_h(\rho; t_{j+1}) && (5.17)\\
&\quad - \int_0^{r_h(t_j)} \int_0^{\rho} k(t_j + \rho, s + t_j) \frac{\bar{u}(t_j + s) - \bar{u}(t_j)}{a_0(t_j; h)}\, ds\, d\eta_h(\rho; t_j)\\
&\quad + \sum_{i=1}^{j} \int_{t_{i-1}}^{t_i} \left[\tilde{k}_0(t_{j+1}, s; h) - \tilde{k}_0(t_j, s; h)\right] (s - t_i)\bar{u}'(z_i(s))\, ds\\
&\quad + \int_{t_j}^{t_{j+1}} \tilde{k}_0(t_{j+1}, s; h)(s - t_{j+1})\bar{u}'(z_{j+1}(s))\, ds - \mu_h(t_{j+1})\bar{u}_0(t_{j+1}) + \mu_h(t_j)\bar{u}_0(t_j),
\end{aligned}
$$

for $i = 1, \ldots, j,\;\; j = 1, \ldots, N-1$.

Similarly, if we evaluate (5.6) at $j = 1$ and (5.8) at $t = t_1$ (dividing through the resulting equation by $a_0(t_1; h)$) and then subtract the two equations, we get

$$
\beta_1 = -\frac{1}{h^2}E_0(\delta, h) - Z_0(h), \qquad (5.18)
$$

where $D_0(h)$ and $Z_0(h)$ are defined by (5.12) and (5.16), respectively (using $j = 0$ in each) and where

$$
\begin{aligned}
E_0(\delta, h) &:= hD_0^{-1}(h)\int_0^{r_h(t_1)} d_0(t_1 + \rho)\, d\eta_h(\rho, t_1), && (5.19)\\
R_0(h) &:= \int_0^{r_h(t_1)} \int_0^{\rho} k(t_1 + \rho, s + t_1)\left[\bar{u}(s + t_1) - \bar{u}(t_1)\right] a_0^{-1}(t_1; h)\, ds\, d\eta_h(\rho, t_1)
\end{aligned}
$$

$$+\int_0^{t_1} \tilde{k}_0(t_1,s;h)(s-t_1)\bar{u}'(z_1(s))\,ds - \mu_h(t_1)\bar{u}_0(t_1). \tag{5.20}$$

Now suppose we can show that there are positive constants w, v, ϵ, and z, all independent of h for which

$$W_j(h) \le w, \quad V_{j,i}(h) \le v, \tag{5.21}$$

for $i = 1,\dots,j$, $j = 1,\dots,N-1$, and $w \in (0,1)$, and

$$E_j(\delta,h) \le \epsilon\,\delta, \quad Z_j(h) \le z, \tag{5.22}$$

for $j = 0,\dots,N-1$. Then applying the arguments found in [18], it follows that $|\beta_j| \le B$, for $j = 1,\dots,N$, where B is independent of N and h; the bound is obtained using the assumption of a uniform bound on $\delta/h^2(\delta)$. Thus, using (5.10), $\left|\bar{u}(t_j) - u_h^\delta(t_j)\right| = |\bar{u}(t_j) - c_j| \le B\,h(\delta)$ so that $\left|\bar{u}(t_j) - u_h^\delta(t_j)\right| \to 0$ as $\delta \to 0$, for $j = 1,\dots,N$.

It remains only to show that the bounds in (5.21)–(5.22) hold. We first show $D_j(h) > 0$ for all h sufficiently small and obtain estimates on $D_j^{-1}(h)$ for $j = 0,\dots,N-1$ and h sufficiently small. To this end we note that the quantity $a_0(t_{j+1};h)D_j(h)$ is the $(j+1)^{\text{st}}$ diagonal entry in the matrix system in (3.14) for $j = 0,\dots,N-1$. Thus from the proof of Theorem 3.1, $D_j(h) > 0$ for $j = 0,\dots,N-1$ and all h sufficiently small. In addition, after a change of integration variable,

$$\begin{aligned}\int_{t_j}^{t_{j+1}} \tilde{k}(t_{j+1},s;h)\,ds &= \int_0^{t_1}\int_0^{r_h(t_{j+1})} k(t_{j+1}+\rho, t_j+s)\,d\eta_h(\rho;t_{j+1})\,ds \\ &= \int_0^{t_1}\int_0^{r_h(t_{j+1})} \left[k(t_j,t_j) + (\rho+h)D_1k(\xi_j,\zeta_j) + sD_2k(\xi_j,\zeta_j)\right]\,d\eta_h(\rho;t_{j+1})\,ds\end{aligned}$$

for suitable $\xi_j = \xi_j(\rho,s)$, $\zeta_j = \zeta_j(\rho,s)$, $j = 0,\dots,N-1$. Thus

$$\int_{t_j}^{t_{j+1}} \tilde{k}(t_{j+1},s;h)\,ds = h\,a_0(t_{j+1};h)\left[1 + g(t_{j+1};h)\right], \tag{5.23}$$

where for $j = 0,\dots,N-1$, $|g(t_{j+1};h)| \le \|k\|_{1,\infty} h(M_r+2)$. Thus for all h sufficiently small, $\int_{t_j}^{t_{j+1}} \tilde{k}(t_{j+1},s;h)\,ds \ge h\,a_0(t_{j+1};h)/2 > 0$, for $j = 0,\dots,N-1$, and

$$D_j^{-1}(h) \le 2/h, \quad j = 0,\dots,N-1. \tag{5.24}$$

Using this estimate on $D_j^{-1}(h)$, we return to the computation of the bounds in (5.21)–(5.22) and see that

$$|E_j(\delta,h)| \le 2h^{-1}\left(h\,\delta\,a_0^{-1}(t_{j+1};h)\int_0^{r_h(t_{j+1})} d\eta_h(\rho;t_{j+1}) + h\,\delta\,a_0^{-1}(t_j;h)\int_0^{r_h(t_j)} d\eta_h(\rho;t_j)\right),$$

so that $|E_j(\delta,h)| \le 4\delta$ for $j = 1,\dots,N-1$, and likewise the same bound is obtained for $|E_0(\delta,h)|$. Thus we obtain the needed bound for E_j in (5.22) with $\epsilon = 4$.

In considering the bound for $V_{j,i}(h)$ in (5.21), we note that

$$\begin{aligned}\tilde{k}(t_j,s;h) &= \int_0^{r_h(t_j)} \left[k(t_j,s) + \rho\,D_1k(t_j+\xi_{s,t_j}(\rho),s)\right]\,d\eta_h(\rho;t_j) \\ &= a_0(t_j;h)k(t_j,s) + \int_0^{r_h(t_j)} \rho\,D_1k(t_j+\xi_{s,t_j}(\rho),s)\,d\eta_h(\rho;t_j), \end{aligned} \tag{5.25}$$

for $j = 1, \ldots, N$, and suitable $\xi_{s,t_j}(\rho)$. Thus, for $i = 1, \ldots, j,\ \ j = 1, \ldots, N-1$,

$$\int_{t_{i-1}}^{t_i} \left| \tilde{k}_0(t_{j+1}, s; h) - \tilde{k}_0(t_j, s; h) \right| \, ds \leq \|k\|_{1,\infty}(1 + 2M_r)h^2.$$

From this estimate and (5.24), it follows that $|V_{j,i}(h)| \leq 2\|k\|_{1,\infty}(1 + 2M_r)$, $i = 1, \ldots, j$, $j = 1, \ldots, N-1$, so that the bound for $V_{j,i}$ in (5.21) is established.

Turning to $Z_j(h)$, we see from (5.16) and (5.24) that we need only show that each of the terms in $R_j(h)$ is $\mathcal{O}(h^2)$, $j = 0, \ldots, N-1$, as $h \to 0$. We have that the first term in $R_j(h)$, $j = 1, \ldots, N-1$, satisfies

$$\begin{aligned} a_0^{-1}(t_{j+1}; h) & \left| \int_0^{r_h(t_{j+1})} \int_0^{\rho} k(t_{j+1} + \rho, s + t_{j+1})(\bar{u}(s + t_{j+1}) - \bar{u}(t_{j+1})) \, ds \, d\eta_h(\rho; t_{j+1}) \right| \\ & \leq \|k\|_\infty \|\bar{u}\|_{1,\infty} \, a_0^{-1}(t_{j+1}; h) \, r_h(t_{j+1}) \int_0^{r_h(t_{j+1})} \rho \, d\eta_h(\rho; t_{j+1}) \\ & \leq \|k\|_\infty \|\bar{u}\|_{1,\infty} M_r^2 \, h^2. \end{aligned}$$

The first term of $R_0(h)$ is bounded similarly, as is the second term in the expression for $R_j(h)$, $j = 1, \ldots, N-1$.

The above estimates for bounding $V_{j,i}(h)$ may be used to show that

$$\left| \sum_{i=1}^{j} \int_{t_{i-1}}^{t_i} \left[\tilde{k}_0(t_{j+1}, s; h) - \tilde{k}_0(t_j, s; h) \right] (s - t_i)\bar{u}'(z_i(s)) \, ds \right| \leq \|\bar{u}\|_{1,\infty} \|k\|_{1,\infty}(1 + 2M_r)Nh^3$$

so that the summation term in $R_j(h)$ is $\mathcal{O}(h^2)$, $j = 1, \ldots, N-1$. The fourth term in $R_j(h)$ for these same values of j is handled similarly, as is the second term in the expression for $R_0(h)$. Using the conditions (5.3)–(5.4) on $\mu_h(t_j)$, the remaining terms from $R_j(h)$, $j = 1, \ldots, N-1$, are estimated as follows. First we note that for $j = 1, \ldots, N-1$,

$$\begin{aligned} \mu_h(t_j)a_0^{-1}(t_j; h) &= \sum_{i=0}^{j-1} \left\{ \left|\mu_h(t_{i+1})a_0^{-1}(t_{i+1}; h)\right| - \left|\mu_h(t_i)a_0^{-1}(t_i; h)\right| \right\} + \left|\mu_h(t_0)a_0^{-1}(t_0; h)\right| \\ &\leq \sum_{i=0}^{N-1} \left|\mu_h(t_{i+1})a_0^{-1}(t_{i+1}; h) - \mu_h(t_i)a_0^{-1}(t_i; h)\right| + \left|\mu_h(t_0)a_0^{-1}(t_0; h)\right| \\ &\leq Nh^2 M_\mu + hM_\mu \end{aligned} \tag{5.26}$$

so that

$$|\mu_h(t_{j+1})\bar{u}_0(t_{j+1}) - \mu_h(t_j)\bar{u}_0(t_j)| \leq \|\bar{u}\|_\infty M_\mu h^2 + 2M_\mu h^2 \|\bar{u}\|_{1,\infty}, \tag{5.27}$$

where we have added and subtracted a term of the form $\bar{u}(t_{j+1})\mu_h(t_j)a_0^{-1}(t_j; h)$. Since $\bar{u}(0) = 0$, the final term in the expression for $R_0(h)$ can be written as

$$a_0^{-1}(t_1; h) \, |\mu_h(t_1)\bar{u}(t_1)| = |\mu_h(t_1)\bar{u}_0(t_1) - \mu_h(0)\bar{u}_0(0)| \tag{5.28}$$

and is handled similarly. Thus the bound in (5.22) is obtained.

Finally, we have that $W_j(h) = (1 + K_j(h))^{-1}$, for $j = 1, \ldots, N-1$, where

$$K_j(h) := \nu_h^{-1}(t_j)\left(\nu_h(t_{j+1}) - \nu_h(t_j) + a_0^{-1}(t_{j+1};h)\int_{t_j}^{t_{j+1}} \tilde{k}(t_{j+1}, s; h)\, ds\right)$$
$$= \nu_h^{-1}(t_j)\left(\frac{\alpha(t_{j+1};h) + \mu_h(t_{j+1})}{a_0(t_{j+1};h)} - \frac{\alpha(t_j;h) + \mu_h(t_j)}{a_0(t_j;h)} + h\left[1 + g(t_{j+1};h)\right]\right),$$

where we have used (5.23). Thus

$$K_j(h) = h\,\nu_h^{-1}(t_j)\left(\frac{\alpha(t_{j+1};h)}{h\,a_0(t_{j+1};h)} - \frac{\alpha(t_j;h)}{h\,a_0(t_j;h)} + \hat{g}(t_{j+1};h) + 1 + g(t_{j+1};h)\right)$$

where $|\hat{g}(t_{j+1};h)| := h^{-1}|\mu_h(t_{j+1})a_0^{-1}(t_{j+1};h) - \mu_h(t_j)a_0^{-1}(t_j;h)| \le M_\mu h$, for $j = 1, \ldots, N-1$. In addition,

$$\alpha(t_j;h) = \int_0^{r_h(t_j)}\int_0^{\rho}\left[1 + \rho D_1 k(\xi_j, \zeta_j) + s D_2 k(\xi_j, \zeta_j)\right]\, ds\, d\eta_h(\rho; t_j),$$

for $\xi_j = \xi_j(\rho, s)$, $\zeta_j = \zeta_j(\rho, s)$, so that $\alpha(t_j;h)(h\,a_0(t_j;h))^{-1} = a_1(t_j;h)(h\,a_0(t_j;h))^{-1} + \tilde{g}(t_j;h)$, where $|\tilde{g}(t_j,h)| \le \frac{3}{2}\|k\|_{1,\infty}M_r^2\, h$, for $j = 1, \ldots, N-1$. Thus, for $j = 1, \ldots, N-1$,

$$\begin{aligned}
K_j(h) &= h\nu_h^{-1}(t_j)\left[a_1(t_{j+1};h)(h\,a_0(t_{j+1};h))^{-1} - a_1(t_j;h)(h\,a_0(t_j;h))^{-1} + 1 + \mathcal{O}(h)\right] \\
&\ge h\nu_h^{-1}(t_j)(\varepsilon_r/2) \\
&\ge h\left[\|k\|_\infty r_h(t_j) + \mu_h(t_j)a_0^{-1}(t_j;h)\right]^{-1}(\varepsilon_r/2),
\end{aligned}$$

for all h sufficiently small, where we have used assumption (5.2). Then, using (5.1) and (5.26), we have that $K_j(h) \ge K$ for $j = 1, \ldots, N-1$ (for all h sufficiently small), where

$$K := \varepsilon_r(2\|k\|_\infty M_r + 4M_\mu)^{-1} > 0. \tag{5.29}$$

Thus $W_j(h) = 1/(1 + K_j(h)) \le w$ for $j = 1, \ldots, N-1$, where $w = 1/(1+K) \in (0,1)$. Thus the bound in (5.21) holds and the proof of the theorem is complete. □

We may relax the condition $\bar{u}(0) = 0$ in the statement of Theorem 5.1 under stricter restrictions on μ than those given in (5.3) and (5.4). The proof of the following corollary is identical to the proof of Theorem 5.1 except for fairly obvious changes in the estimates in equations (5.26)–(5.28), and (5.29).

Corollary 5.1 *Suppose $f \in \mathcal{F}_D$, k, and $\bar{u}$ satisfy the conditions at the beginning of Section 3.2, and let ε_r, M_r, and M_μ be fixed positive scalars. For each $N = 1, 2, \ldots$, let $h = 1/N$ and suppose we are given $r_h \in \Gamma$ and an (r_h, h)-suitable family $\mathcal{N}_h = \{\eta_h(\cdot, t), t \in [0,1]\}$ of measures for which (5.1) and (5.2) are satisfied for all N sufficiently large. Then if $\mu_h : [0,1] \mapsto [0,\infty)$ is selected satisfying*

$$\frac{1}{h^2}\left(\frac{\mu_h(t_j)}{\int_0^{r_h(t_j)} d\eta_h(\rho, t_j)}\right) \le M_\mu, \quad j = 0, \ldots, N, \tag{5.30}$$

for all h sufficiently small, the conclusions of Theorem 5.1 still hold.

We conclude with the proofs of Theorem 4.1 and 4.2 from Section 3.

Proof of Theorem 4.1: The proof follows from Corollary 5.1. The only estimate that is not immediate is that condition (1b) in the statement of Theorem 4.1 implies the inequality in (5.2). Indeed

$$\frac{\int_0^{r_h(t_{j+1})} \rho \, d\eta_h(\rho; t_{j+1})}{h \int_0^{r_h(t_{j+1})} d\eta_h(\rho; t_{j+1})} - \frac{\int_0^{r_h(t_j)} \rho \, d\eta_h(\rho; t_j)}{h \int_0^{r_h(t_j)} d\eta_h(\rho; t_j)} + 1 \geq 0 - M_r + 1$$

so that $\varepsilon_r = -M_r + 1 > 0$ under the assumptions of the theorem. □

Proof of Theorem 4.2: The proof follows from Corollary 5.1 provided we show that condition (5.2) holds. In fact, under the conditions of Theorem 4.2,

$$\begin{aligned}\frac{\int_0^{r_h(t_{j+1})} \rho \, d\eta_h(\rho; t_{j+1})}{h \int_0^{r_h(t_{j+1})} d\eta_h(\rho; t_{j+1})} - \frac{\int_0^{r_h(t_j)} \rho \, d\eta_h(\rho; t_j)}{h \int_0^{r_h(t_j)} d\eta_h(\rho; t_j)} + 1 &= b_1 \left(\gamma(t_{j+1}) - \gamma(t_j) + 1/b_1 + \hat{c}_2(h)/b_1\right) \\ &\geq b_1[\gamma(t_{j+1}) - \gamma(t_j) + \hat{p}]\end{aligned}$$

where $\hat{c}_2(h) = \mathcal{O}(h)$, and the inequality is valid for all h sufficiently small, using $\hat{p} = 1/b_1 - (1/b_1 - p)/2 > p$. Thus if γ satisfies (4.3), condition (5.2) holds with $\varepsilon_r > 0$. □

6. Conclusions

We have considered a local regularization method for the solution of ill-posed Volterra problems, focusing in particular on discrete realizations of the method. We have provided theoretical results guaranteeing convergence of the discretized method, and have examined the role played by functional regularization parameters r_h and μ_h. Further, we have developed a sequential discrepancy principle to select the penalty parameter $\mu_h = \mu_h(t)$, presenting numerical examples to illustrate the effectiveness of this adaptive procedure.

We should mention that several parts of the analysis and practical use of this sequential discrepancy principle have not been presented here. Current study involves the development of a convergence theory guaranteeing that this selection of μ_h leads to a convergent approximation method as $h \to 0$ and the level of the noise decreases to zero. In addition, we are investigating models of propagated error in order to make best use of the results in Theorem 2.2. Finally, the results in [20] give hope that we may be able to extend our theoretical results to ν-smoothing problems (most likely for small integer ν), under additional assumptions on the problems.

Acknowledgments

This work was supported in part (for the first author) by the National Science Foundation under contract number NSF DMS 9704899. The authors wish to thank the referees for comments which simplified some of the assumptions in our theory and improved the exposition of the paper.

References

[1] Acar R and Vogel C 1994 Analysis of bounded variation penalty methods for ill-posed problems *Inverse Problems* **10** 1217–1229

[2] Beck J V, Blackwell B and St Clair C R Jr 1985 *Inverse Heat Conduction* (New York: Wiley-Interscience)

[3] Blanc G, Beck J V and Raynaud M 1997 Solution of the inverse heat conduction problem with a time-variable number of future temperatures *Numerical Heat Transfer, Part B* **32** 437–451

[4] Burton T A 1983 *Volterra Integral and Differential Equations* (New York: Academic Press)

[5] Chambolle A and Lions P-L 1997 Image recovery via total variation minimization and related problems *Numer Math.* **72** 167–188

[6] Cinzori A C and Lamm P K 1998 Future polynomial regularization of first-kind Volterra operators *SIAM J. Numerical Analysis*, to appear

[7] Corduneanu C 1991 *Integral Equations and Applications* (Cambridge: Cambridge University Press)

[8] Dobson D C and Vogel C R 1997 Convergence of an iterative method for total variation denoising *SIAM J. Numerical Analysis* **34** 1779–1971

[9] Denisov A M 1975 The approximate solution of a Volterra equation of the first kind *Ž. Vyčisl. Mat. i Mat. Fiz.* **15** 1053–1056, 1091 (English transl.: *USSR Comput. Math. Math. Phys.* **15** 237–239

[10] Edwards R E 1979 *Fourier Series (a Modern Introduction)* vol 1 2nd edn (Berlin: Springer-Verlag)

[11] Eldén, L 1984 An efficient algorithm for the regularization of ill-conditioned least squares problems with triangular Toeplitz matrix *SIAM J. Sci. Stat. Comput.* **5** 229–236

[12] Engl H W, Hanke M, and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht, The Netherlands: Kluwer Academic Publishers)

[13] Gripenberg G, Londen S O, and Saffens O 1990 *Volterra Integral and Functional Equations* (Cambridge: Cambridge University Press)

[14] Groetsch C W 1984 *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind* (Boston: Pitman)

[15] Ito K and Kunisch K An active set strategy based on the augmented Lagrangian formulation for image restoration *R.A.I.R.O. Math. Mod. and Numer. Anal* to appear

[16] Ito K and Kunisch K 1996 Augmented Lagrangian formulation of nonsmooth convex optimization in Hilbert spaces, in Control of Partial Differential Equations and Applications *Proc. IFIP TC7/WG-7.2 Int. Conf. (1994 Laredo, Spain) Lect. Notes Pur Appl. Math.* **174** E Casas, ed (New York: Dekker) 107–117

[17] Lamm P K 1999 A survey of regularization methods for first-kind Volterra equations, to appear in *Surveys on Solution Methods for Inverse Problems*, D Colton, et al, eds (Springer: Vienna)

[18] Lamm P K 1996 Approximation of ill-posed Volterra problems via predictor–corrector regularization methods *SIAM J. Appl. Math.* **56** 524–541

[19] Lamm P K 1995 Future-sequential regularization methods for ill-posed Volterra equations: applications to the inverse heat conduction problem *J. Math. Anal. Appl.* **195** 469–494

[20] Lamm P K 1997 Regularized inversion of finitely smoothing Volterra operators: predictor–corrector regularization methods *Inverse Problems* **13** 375–402

[21] Lamm P K 1998 Variable-smoothing regularization methods for inverse problems, to appear in *Conf. Proceedings of 6th Mediterranean Conference on Control and Systems*, June 1998

[22] Lamm P K and Eldén L 1997 Numerical solution of first-kind Volterra equations by sequential Tikhonov regularization *SIAM J. Numer. Anal.* **34** 1432–1450

[23] Lamm P K and Scofield T L 1999 Local regularization methods for the stabilization of ill-posed Volterra problems, preprint

[24] Lavrent′ev M M 1967 *Some improperly posed problems of mathematical physics* (Berlin: Springer-Verlag)

[25] Magnickiĭ N A 1978 The approximate solution of certain Volterra integral equations of the first kind

Vestnik Moskov. Univ. Ser. XV Vyčisl. Mat. Kibernet. **1978** 91–96 (English transl.: *Moscow Univ. Comput. Math. Cybernetics* **1978** 74–78

[26] Munoz J C and Yeow Y L 1996 Applications of maximum entropy method in capillary viscometry *Rheo. Acta* **35** 76–82

[27] Neubauer A and Scherzer O 1998 Regularization for curve representations: uniform convergence for discontinuous solutions of ill-posed problems *SIAM J. Appl. Math.* **58** 1891–1900

[28] Plato R 1995 *Iterative and parametric methods for linear ill-posed problems* (TU Berlin: Habiliationsschrift Fachbereich Mathematik)

[29] Plato R 1997 Lavrentiev's method for linear Volterra integral equations of the first kind, with applications to the non-destructive testing of optical-fibre preforms in *Inverse problems in medical imaging and nondestructive testing (Oberwolfach, 1996, 196–211)* Vienna: Springer

[30] Plato R 1996 On the discrepancy principle for iterative and parametric methods to solve linear ill-posed equations *Numer. Math.* **75** 99–120

[31] Plato R 1997 Resolvent estimates for Abel integral operators and the regularization of associated first kind integral equations *J. Integral Equations Appl.* **9** 253–278

[32] Plato R and Hämarik U 1996 On pseudo-optimal parameter choices and stopping rules for regularization methods in Banach spaces *Numer. Funct. Anal. Optim.* **17** 181–195

[33] Reinhardt H-J 1994 Analysis of the discrete Beck method of solving ill-posed parabolic equations *Inverse Problems* **10** 1345–1360

[34] Reinhardt H-J and Hào D N 1994 Sequential approximation to nonlinear inverse heat conduction problems *Math. Comput. Modelling* **20** 189–200

[35] Rudin L I, Osher S, and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica D* **60** 259–268

[36] Scofield, T L 1998 Sequential predictor-corrector methods for variable regularization of ill-posed Volterra problems, Ph.D. Thesis, Department of Mathematics, Michigan State University, E. Lansing, MI 48824–1027

[37] Sergeev V O 1971 Regularization of Volterra equations of the first kind *Dokl. Akad. Nauk SSSR* **197** 531–534 (English transl.: *Soviet Math. Dokl .* **12** 501–505)

[38] Vasin V V 1996 Monotone iterative processes for nonlinear operator equations and their applications to Volterra equations *J. Inverse Ill-Posed Probl.* **4** 331–340

[39] Vasin V V 1996 Monotone iterative processes for operator equations in partially ordered spaces *Dokl. Akad. Nauk* **349** 7–9