NUMERICAL SOLUTION OF FIRST-KIND VOLTERRA EQUATIONS BY SEQUENTIAL TIKHONOV REGULARIZATION

PATRICIA K. LAMM* AND LARS ELDÉN†

Abstract. We consider the problem of finding regularized solutions to ill-posed Volterra integral equations. The method we consider is a sequential form of Tikhonov regularization that is particularly suited to problems of Volterra type. We prove that when this sequential regularization method is coupled with several standard discretizations of the integral equation (collocation, rectangular and midpoint quadrature), one obtains convergence of the method at an optimal rate with respect to noise in the data. In addition we describe a fast algorithm for the implementation of sequential Tikhonov regularization, and show that for small values of the regularization parameter, the method is only slightly more expensive computationally than the numerical solution of the original unregularized integral equation. Finally, numerical results are presented to show that the performance of sequential Tikhonov regularization nearly matches that of standard Tikhonov regularization in practice, but at considerable savings in cost.

Key words. First-kind Volterra equations, inverse problems, regularization, sequential numerical methods, fast algorithm

AMS subject classifications. 65J20, 65F20, 45

1. Introduction. Consider the problem of finding f solving

$$(1.1) \qquad\qquad Kf = g$$

where $Kf(t) = \int_0^t k(t - \tau)f(\tau)\, d\tau$, $0 \le t \le 1$, is a first-kind Volterra integral operator with convolution kernel $k \in L_2(0,1)$, $k(t) > 0$ on $(0,1]$, and data $g \in L_2(0,1)$. In particular, we are interested in the case where we only know a perturbation of the "ideal" data g, the perturbation given by $g^\delta(t) = g(t) + d(t)$, where $\|d\| < \delta$ for some $\delta > 0$, and where $\|\cdot\|$ is typically the L_2 or L_∞ norm. We are then forced to consider the related problem

$$(1.2) \qquad\qquad Kf^\delta = g^\delta$$

and it is for this problem that we seek an approximate solution which reasonably estimates the solution f of (1.1).

Because problem (1.1) is ill-posed, some form of regularization method must be utilized in order to obtain a useful approximation to the desired solution. In this paper we consider a variation of classical Tikhonov regularization which is *sequential* in nature and, as such, is particularly suited for the solution of a first-kind Volterra integral equation (and other causal problems). This approach was mentioned in Chapter 5 of [1] for a particular Volterra problem, the sideways heat equation (this problem is discussed in some more detail in §6), although it was not explored further in that reference because it was said to be too costly to implement. Here we improve upon the initial idea in [1] by developing an extremely efficient algorithm for implementation of the method. In addition, we go further and make an analysis of convergence of the algorithm, obtaining theoretical results which are valid for zeroth order regularization of a particular class of Volterra problems (i.e., the class of "one-smoothing" Volterra

*Department of Mathematics, Michigan State University, East Lansing, MI 48824-1027, lamm@math.msu.edu

†Department of Mathematics, Linköping University, S-581 83 Linköping, Sweden, laeld@math.liu.se

problems, which appear in numerous applications but do not include the sideways heat equation).

The plan of the paper is as follows. In the following section we introduce sequential Tikhonov regularization in more detail. Then we discuss the relation to standard Tikhonov regularization in §3, and establish there some notation and estimates useful in the convergence proofs in §4. In §4, we prove that when the sequential Tikhonov algorithm is coupled with several standard discretizations (collocation-based methods, rectangular quadrature, or midpoint quadrature) of the original integral equation, the method converges at an optimal rate with respect to the level δ of error in the data.

A fast algorithm for the implementation of sequential Tikhonov regularization is presented in §5. It is shown that for large problems the cost of the proposed method (in terms of the number of operations) is comparable to that of solving a triangular system of equations, and thus the new method is considerably more efficient than standard Tikhonov regularization. In §6 some numerical examples are given, which indicate that sequential regularization gives about as accurate solutions as standard Tikhonov regularization.

2. Sequential Tikhonov Regularization. Consider the solution of the system of equations

$$(2.1) \qquad\qquad \mathbf{Kf} = \mathbf{g},$$

which is obtained by discretizing the perturbed equation (1.2). We assume that the particular discretization method used generates a lower-triangular, Toeplitz matrix $\mathbf{K} \in \mathbb{R}^{n \times n}$, given by

$$(2.2) \qquad\qquad \mathbf{K} = \begin{pmatrix} k_1 & 0 & \dots & 0 \\ k_2 & k_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ k_n & k_{n-1} & \dots & k_1 \end{pmatrix},$$

and that $k_i > 0$ for all i.

Discrete Tikhonov regularization stabilizes the solution of (2.1) via the solution of minimization problem

$$(2.3) \qquad\qquad \min_{\mathbf{f} \in \mathbb{R}^n} \left\{ \|\mathbf{Kf} - \mathbf{g}\|^2 + \mu \|\mathbf{Lf}\|^2 \right\},$$

where $\mu > 0$ is a given parameter, and $\mathbf{L}$ is a lower-triangular Toeplitz matrix (typically the $n \times n$ identity, or a discretized differentiation operator). We propose here a sequential variation of discrete Tikhonov regularization which is particularly useful for the stable solution of problems of the form (2.1), (2.2). Assuming that $f_1, f_2, \dots, f_{i-1}$ have already been found, this method determines f_i by first finding the vector $\beta \in \mathbb{R}^r$, $r < n$, which solves the reduced-dimension Tikhonov problem

$$(2.4) \qquad\qquad \min_{\beta \in \mathbb{R}^r} \left\{ \|\mathbf{K}_r \beta - \mathbf{h}^{(i)}\|^2 + \mu \|\mathbf{L}_r \beta\|^2 \right\},$$

where $\mathbf{K}_r$ and $\mathbf{L}_r$ are the leading $r \times r$ blocks of $\mathbf{K}$ and $\mathbf{L}$, respectively, and where $\mathbf{h}^{(1)} = (g_1, \dots, g_r)^\top$ and $\mathbf{h}^{(i)} = (h_1^{(i)}, h_2^{(i)}, \dots, h_r^{(i)})^\top$ for $i \geq 2$, with

$$(2.5) \qquad\qquad h_p^{(i)} = g_{i+p-1} - \sum_{j=1}^{i-1} k_{i+p-j} f_j, \qquad p = 1, \dots, r.$$

(Here r is typically much smaller than n.) We then take f_i to be the first component of the vector β and discard all remaining components of β. It is not difficult to modify this algorithm for the case where p components of the solution are retained at each step, $1 \le p < r$; however, this variation will not be a focus of attention here.

It is important to note that this sequential approach is not simply a matter of performing a decomposition of the original matrix system into smaller subproblems, with standard Tikhonov regularization then applied individually to each smaller problem. In contrast, the method performs Tikhonov regularization sequentially on small overlapping subproblems, updating the definition of each subproblem as new information about the solution is obtained.

Because $r-1$ "future" values $g_{i+1}, \ldots, g_{i+r-1}$ of data are used to compute f_i (for $r \ge 1$ fixed), it is clear that one can only expect to stably compute $f_1, f_2, \ldots, f_{n-r+1}$ using this algorithm, unless k and g^δ are available past the interval $[0, 1]$ and one assumes that (2.1) is the discrete approximation of (1.2) on this extended interval as well. In order to simplify the theoretical arguments which follow, we will assume in §2 and 3 that k and g^δ are defined on $[0, T]$, for some $T > 0$ sufficiently large, and that (1.2) holds on $[0, T]$ as well.

3. Comparison with Tikhonov Regularization. Standard Tikhonov regularization applied to (2.1) can be viewed as a discretized version of Tikhonov regularization applied to the original problem (1.2); in the case of $\mathbf{L} = \mathbf{I}$, the associated infinite-dimensional Tikhonov problem becomes an integral equation in f_μ^δ of the form

$$(3.1) \qquad (K^\star K + \mu I) f_\mu^\delta(t) = K^\star g^\delta(t), \quad t \in [0, 1],$$

where $K^\star u(t) = \int_t^1 k(s - t) u(s) \, ds$ for $u \in L_2(0, 1)$. Thus f_μ^δ is the solution of a *well-posed* second-kind integral equation, and, as such, depends continuously on data g^δ. One disadvantage of this approach however is that (3.1) is non-Volterra in nature, while the original equation (1.2) is Volterra in form.

In this section we will show how the discrete sequential Tikhonov regularization also can be viewed as a discretization of a well-posed second-kind integral equation; in the sequential case however the second-kind equation is still Volterra in form provided that r is small.

The discretization of (1.2) given by (2.1), (2.2), typically arises from a standard numerical method such as rectangular quadrature, midpoint quadrature, or collocation over piecewise constant functions. The details of these methods will be considered further in the next section, but for the time being we shall make rather general assumptions about the entries k_i in $\mathbf{K}$ and g_i in $\mathbf{g}$, conditions which remain valid for any of these three methods. That is, we assume that $k_i = k_i(k)$ is defined using the kernel k in such a way that (1) $k \mapsto k_i(k)$ is a linear map for each $i = 1, \ldots, n$, and, (2) for some fixed $\Delta t > 0$,

$$k_i(\mathcal{S}_{\Delta t} k) = k_{i+1}(k),$$

where $\mathcal{S}_{\Delta t}$ is the shift operator defined by $(\mathcal{S}_{\Delta t} k)(t) = k(t + \Delta t)$ for $t \in [0, 1]$. Typically Δt is the discretization stepsize of the underlying numerical method, $\Delta t = \mathcal{O}(1/n)$. We make similar assumptions concerning the dependence of the discretized vector $\mathbf{g} = (g_1, \ldots, g_n)^\top$ in (2.1) on $g^\delta \in L_2(0, T)$; that is, we assume that the map $g^\delta \mapsto g_i(g^\delta)$ is linear and that $g_i(\mathcal{S}_{\Delta t} g^\delta) = g_{i+1}(g^\delta)$, for $i = 1, \ldots, n$.

From (2.4) it is seen that, in the sequential Tikhonov algorithm, one uses previously computed $f_1, \ldots, f_{i-1}$ to determine the solution $\beta \in \mathbb{R}^r$ of (2.4), or equivalently,

β solving the normal equations

$$(3.2) \qquad (\mathbf{K}_r^\top \mathbf{K}_r + \mu \mathbf{L}_r^\top \mathbf{L}_r)\beta = \mathbf{K}_r^\top \mathbf{h}^{(i)};$$

one then sets f_i equal to the first component of β. Using Cramer's Rule,

$$f_i = \frac{\det\left(\mathbf{K}_r^\top \mathbf{h}^{(i)} \,\middle|\, \mathbf{t}_2, \ldots, \mathbf{t}_r\right)}{\det\left(\mathbf{t}_1 \,\middle|\, \mathbf{t}_2, \ldots, \mathbf{t}_r\right)},$$

where $\mathbf{t}_i$ denotes the ith column of $\mathbf{K}_r^\top \mathbf{K}_r + \mu \mathbf{L}_r^\top \mathbf{L}_r$. For an arbitrary vector $\alpha = (\alpha_1, \ldots, \alpha_r)^\top \in \mathbb{R}^r$, properties of the determinant may be used to expand along the first column of $(\mathbf{K}_r^\top \alpha \,|\, \mathbf{t}_2, \ldots, \mathbf{t}_r)$ so that

$$(3.3) \qquad \det\left(\mathbf{K}_r^\top \alpha \,\middle|\, \mathbf{t}_2, \ldots, \mathbf{t}_r\right) = \sum_{\ell=1}^{r} c_\ell \, \alpha_\ell$$

for constants $c_i = c_i(\mu; \mathbf{K}_r, \mathbf{L}_r)$. In fact, the upper-triangular form of $\mathbf{K}_r^\top$ gives α_1 appearing only in the first component of $\mathbf{K}_r^\top \alpha$, so

$$(3.4) \qquad c_1 = k_1 \det(\mathbf{e_1} \,|\, \mathbf{t}_2, \ldots, \mathbf{t}_r)$$

where $\mathbf{e}_1 = (1, 0, \ldots, 0)^\top$. Using (3.3) and the fact that

$$\mathbf{t}_1 = \mathbf{K}_r^\top \begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_r \end{pmatrix} + \mu \mathbf{L}_r^\top \ell_1,$$

for ℓ_1 the first column of $\mathbf{L}_r$, we have

$$(3.5) \qquad \det(\mathbf{K}_r^\top \mathbf{K}_r + \mu \mathbf{L}_r^\top \mathbf{L}_r) = \sum_{\ell=1}^{r} c_\ell k_\ell + \mu \gamma_r(\mathbf{K}_r, \mathbf{L}_r)$$

where we have made the definition

$$\gamma_r(\mathbf{K}_r, \mathbf{L}_r) = \det\left(\mathbf{L}_r^\top \ell_1 \,\middle|\, \mathbf{t}_2, \ldots, \mathbf{t}_r\right),$$

and where γ_r also depends on μ through its dependence on $\mathbf{t}_i$, $i = 2, \ldots, r$. Therefore,

$$(3.6) \qquad \begin{aligned} f_i &= \frac{\displaystyle\sum_{\ell=1}^{r} c_\ell h_\ell^{(i)}}{\displaystyle\sum_{\ell=1}^{r} c_\ell k_\ell + \mu \gamma_r(\mathbf{K}_r, \mathbf{L}_r)} \\[2em] &= \frac{\displaystyle\sum_{\ell=1}^{r} c_\ell \left(g_{i+\ell-1} - \sum_{j=1}^{i-1} k_{i+\ell-j} f_j\right)}{\displaystyle\sum_{\ell=1}^{r} c_\ell k_\ell + \mu \gamma_r(\mathbf{K}_r, \mathbf{L}_r)} \end{aligned}$$

for $i = 1, \ldots, n$, so that f_i satisfies

$$(3.7) \quad \sum_{j=1}^{i} \left(\sum_{\ell=1}^{r} c_\ell\, k_{i+\ell-j} \right) f_j + \mu \gamma_r(\mathbf{K}_r, \mathbf{L}_r) f_i = \sum_{\ell=1}^{r} c_\ell\, g_{i+\ell-1}, \quad i = 1, \ldots, n.$$

For the particular choice of $\mathbf{L}_r = \mathbf{I}_r$, the $r \times r$ identity, block matrix multiplication gives

$$\mathbf{K}_r^\top \mathbf{K}_r + \mu \mathbf{I}_r = \left(\begin{array}{c|c} \begin{array}{c} k_1 \\ \hline 0 \\ \vdots \\ 0 \end{array} & \begin{array}{c} k_2 \ldots k_r \\ \hline \\ \mathbf{K}_{r-1}^\top \\ \\ \end{array} \end{array} \right) \left(\begin{array}{c|c} \begin{array}{c} k_1 \\ \hline k_2 \\ \vdots \\ k_r \end{array} & \begin{array}{c} 0 \ldots 0 \\ \hline \\ \mathbf{K}_{r-1} \\ \\ \end{array} \end{array} \right) + \mu \left(\begin{array}{c|c} \begin{array}{c} 1 \\ \hline 0 \\ \vdots \\ 0 \end{array} & \begin{array}{c} 0 \ldots 0 \\ \hline \\ \mathbf{I}_{r-1} \\ \\ \end{array} \end{array} \right)$$

$$= \left(\begin{array}{c|c} k_1^2 + \ldots + k_r^2 + \mu & (k_2 \ldots k_r) \mathbf{K}_{r-1} \\ \hline \mathbf{K}_{r-1}^\top \begin{pmatrix} k_2 \\ \vdots \\ k_r \end{pmatrix} & \mathbf{K}_{r-1}^\top \mathbf{K}_{r-1} + \mu \mathbf{I}_{r-1} \end{array} \right),$$

where $\mathbf{I}_{r-1}$ denotes the $(r-1) \times (r-1)$ identity, and $\mathbf{K}_{r-1}$ is the $(r-1)$-square Toeplitz matrix formed by deleting the first row and column of $\mathbf{K}_r$. Since $\mathbf{L}_r^\top \ell_1 = \mathbf{e}_1$ in this case, it follows that

$$(3.8) \quad \begin{aligned} \gamma_r(\mathbf{K}_r, \mathbf{L}_r) &= \det\left(\mathbf{e}_1 \mid \mathbf{t}_2, \ldots, \mathbf{t}_r \right) \\ &= \det\left(\mathbf{K}_{r-1}^\top \mathbf{K}_{r-1} + \mu \mathbf{I}_{r-1} \right), \end{aligned}$$

or, for $\mu > 0$, that $\gamma_r(\mathbf{K}_r, \mathbf{L}_r) \neq 0$. Further, (3.4) gives that $c_1 = k_1 \gamma_r$ is also nonzero; dividing both sides of equation(3.7) by c_1 we obtain

$$(3.9) \quad \sum_{j=1}^{i} \left(\sum_{\ell=1}^{r} s_\ell\, k_{i+\ell-j} \right) f_j + \frac{\mu}{k_1} f_i = \sum_{\ell=1}^{r} s_\ell\, g_{i+\ell-1}, \quad i = 1 \ldots, n,$$

where $s_\ell \equiv c_\ell / c_1$ for $\ell = 1, \ldots, r$. That is, the f_i satisfy

$$(3.10) \quad \left\{ \begin{pmatrix} \tilde{k}_1 & 0 & \ldots & 0 & 0 \\ \tilde{k}_2 & \tilde{k}_1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \tilde{k}_n & \tilde{k}_{n-1} & \ldots & \tilde{k}_2 & \tilde{k}_1 \end{pmatrix} + \frac{\mu}{k_1} \mathbf{I}_r \right\} \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix} = \begin{pmatrix} \tilde{g}_1 \\ \tilde{g}_2 \\ \vdots \\ \tilde{g}_n \end{pmatrix}$$

where

$$(3.11) \quad \begin{aligned} \tilde{k}_i &\equiv \sum_{\ell=1}^{r} s_\ell\, k_{i+\ell-1}(k) \\ &= k_i \left(\sum_{\ell=1}^{r} s_\ell\, \mathcal{S}_{[(\ell-1)\Delta t]} k \right), \end{aligned}$$

using the assumptions on k_i; a similar expression holds for $\tilde{g}_i \equiv \sum_{\ell=1}^{r} s_\ell \, g_{i+\ell-1}(g^\delta)$. Therefore, making the following definitions for $t \in [0, 1]$,

$$(3.12) \qquad\qquad \tilde{k}(t) = \sum_{\ell=1}^{r} s_\ell \, k(t + (\ell - 1)\Delta t)$$

$$(3.13) \qquad\qquad \tilde{g}^\delta(t) = \sum_{\ell=1}^{r} s_\ell \, g^\delta(t + (\ell - 1)\Delta t),$$

it follows from the assumptions on k_i, g_i that (3.10) corresponds to an n-dimensional discretization of the second-kind Volterra integral equation

$$(3.14) \qquad\qquad \int_0^t \tilde{k}(t - s) f(s) \, ds + \frac{\mu}{k_1} f(t) = \tilde{g}^\delta(t), \quad t \in [0, 1].$$

Thus, analogous to the case of standard Tikhonov regularization, the discrete sequential Tikhonov algorithm corresponds to a direct discretization of a stable second-kind integral equation, provided the parameter μ is positive. In addition, this equation retains the Volterra structure with a new kernel $\tilde{k}$ and data $\tilde{g}^\delta$, each of which is constructed using a small (r-dependent) amount of future information at each time t.

4. Convergence of the Sequential Algorithm. The results which follow require that k and f belong to $C^1[0, T]$, where f is the solution of (1.2) on $[0, T]$ using precise data g. The goal of this section is to show that the vector $\mathbf{f} = (f_1, \ldots, f_n)^\top$ defined by the discrete sequential Tikhonov regularization algorithm is a suitable approximation for $(f(x_1), \ldots, f(x_n))^\top$ for appropriate choices of $x_1, \ldots, x_n \in [0, 1]$ (the choice depends on the discretization method) and for n sufficiently large. We shall consider the cases where the unregularized approximation (2.1) is constructed using either collocation over piecewise constant functions, rectangular quadrature, or midpoint quadrature. Throughout we assume that the data g^δ in (1.2) satisfies $g^\delta = g(t) + d(t)$ with $|d(t)| \leq \delta$ for all $t \in [0, T]$ and $\delta > 0$.

We first present the arguments for convergence of approximations based on collocation over spaces of piecewise constant functions. To this end we let $n = 1, 2, \ldots$, be fixed, let $\Delta t \equiv 1/n$ and $t_i \equiv i\Delta t$, for $i = 0, 1, \ldots, n+r-1$, for $r \geq 1$ a fixed integer. We designate the space of piecewise-constant functions on $[0, 1]$ by $X_{\Delta t} \equiv \text{span}\{\chi_i\}_{i=1}^n$, where χ_i is the characteristic function defined by $\chi_i(t) = 1$, for $t_{i-1} < t \leq t_i$, and $\chi_i(t) = 0$ otherwise; $\chi_1(0) = 1$. We then seek $f_{\Delta t}^\delta \in X_{\Delta t}$, $f_{\Delta t}^\delta(t) = \sum_{i=1}^n f_i \chi_i(t)$, solving the collocation equations

$$(4.1) \qquad\qquad K f_{\Delta t}^\delta(t_j) = g(t_j) + d(t_j), \quad j = 1, 2, \ldots, n,$$

where

$$K f_{\Delta t}^\delta(t_j) = \sum_{\nu=1}^{j} f_\nu \int_{t_{\nu-1}}^{t_\nu} k(t_j - s) \, ds$$

$$= \sum_{\nu=1}^{j} f_\nu \int_0^{t_1} k(t_{j-\nu+1} - s) \, ds.$$

Therefore when (2.1), (2.2), corresponds to a discretization of (1.2) based on collocation over piecewise constants, the entries k_i in $\mathbf{K}$ are given by

$$(4.2) \qquad\qquad k_i \equiv \int_0^{t_1} k(t_i - s) \, ds$$

and $\mathbf{g} = (g(t_1) + d(t_1), \ldots, g(t_n) + d(t_n))^\top$.

Using the construction in the previous section, the sequential Tikhonov regularization method defines the regularized f_j recursively via (3.10), or by

$$(4.3) \qquad \frac{\mu}{k_1} f_j + \sum_{\nu=1}^{j} f_\nu \tilde{k}_{j-\nu+1} = \tilde{g}(t_j) + \tilde{d}(t_j), \quad j = 1, \ldots, n,$$

where $\tilde{g}(t) = \sum_{\ell=1}^{r} s_\ell g(t + (\ell - 1)\Delta t)$ with $\tilde{d}(t)$ defined similarly, and where $\tilde{k}_i = \sum_{\ell=1}^{r} s_\ell k_{i+\ell-1}$. Thus the sequential Tikhonov algorithm for collocation defines $\tilde{k}_i = \int_0^{t_1} \tilde{k}(t_i - s) \, ds$ (using the definition of $\tilde{k}(t)$ in (3.12)) and f_j via

$$(4.4) \qquad \frac{\mu}{k_1} f_j + \sum_{\nu=1}^{j} f_\nu \int_{t_{\nu-1}}^{t_\nu} \tilde{k}(t_j - s) \, ds = \tilde{g}(t_j) + \tilde{d}(t_j), \quad j = 1, \ldots, n.$$

THEOREM 4.1. *Let $r = 1, 2, \ldots$ be a fixed integer and let $k, f \in C^1[0, T]$ where f is the solution of (1.1) on $[0, T]$ using precise data g. In addition, assume that $k(t) \geq k_{\min} > 0$ on $[0, T]$ and $f(0) = 0$, and that for $\delta > 0$, the perturbed data g^δ in (1.2) satisfies $g^\delta(t) = g(t) + d(t)$, $t \in [0, T]$, with $|d(t)| \leq \delta$ on $[0, T]$.*

Then if $\mu = \mu(\Delta t)$ is selected such that $\mu = \hat{\mu} \Delta t^2$, with $\hat{\mu} > 0$ constant, and if $\Delta t = \Delta t(\delta)$ satisfies $\Delta t(\delta) = \tau \sqrt{\delta}$, with $\tau > 0$ constant, it follows that as $\delta \to 0$, we have $\Delta t(\delta) \to 0$, $\mu(\Delta t) \to 0$, and

$$|f_j - f(t_j)| \leq \delta^{1/2} \overline{C}(r) + \mathcal{O}(\delta) \to 0, \qquad \text{for } j = 1, \ldots, n(\delta),$$

where $\mathbf{f} = (f_1, \ldots, f_n)^\top$ is the solution of (4.4), the sequential Tikhonov regularization problem for approximations based on collocation over piecewise contants, with perturbed data g^δ. Here $n(\delta) = 1/\Delta t(\delta)$ and $\overline{C}(r)$ is a fixed positive constant.

Proof. The proof, which is lengthy, is accomplished in the following steps.

- *Step 1:* Let

$$(4.5) \qquad \sigma_j = [f(t_j) - f_j]/\Delta t \qquad j = 1, \ldots, n.$$

Then the $\{\sigma_j\}$ satisfy a system of difference equations of the form

$$(4.6) \qquad \sigma_1 = -d_1 - R_1$$

$$(4.7) \qquad \sigma_{j+1} = W\sigma_j - \Delta t \sum_{\nu=1}^{j} V_{j,\nu}\sigma_\nu - d_{j+1} - R_{j+1},$$
$$\text{for } j = 1, \ldots, n - 1,$$

where W, $V_{j,\nu}$, d_j, and R_j, $j = 1, \ldots, n$, are suitable functions of Δt, δ, r, and μ (given in (4.17) below).

- *Step 2:* For the choice $\mu = \hat{\mu} \Delta t^2$ (where $\hat{\mu} > 0$ is constant), the functions W, $V_{j,\nu}$, d_j, and R_j are bounded as follows

$$|W(\Delta t, r, \mu)| \leq w(r),$$
$$|V_{j,\nu}(\Delta t, r, \mu)| \leq v(r), \quad \nu = 1, \ldots, j; \; j = 1, \ldots, n - 1$$
$$(4.8) \qquad |d_j(\Delta t, r, \mu)| \leq e(r)\frac{\delta}{\Delta t^2}, \quad j = 1, \ldots, n$$
$$|R_j(\Delta t, r, \mu)| \leq z(r), \quad j = 1, \ldots, n,$$

where $v(r)$, $e(r)$, and $z(r)$ are nonnegative and $0 < w(r) < 1$ for all $r = 1, 2, \ldots$.

- *Step 3:* Using the bounds in (4.8),

$$(4.9) \qquad |\sigma_j| \le 2\,C_1(\delta, \Delta t, r)\exp\left(\frac{2v(r)}{1-w(r)}\right), \quad j = 1, \ldots, n-1,$$

for suitable $C_1(\delta, \Delta t, r)$ and for Δt sufficiently small, so that using (4.5), the error estimate

$$|f_j - f(t_j)| \le 2\,\Delta t\,C_1(\delta, \Delta t, r)\exp\left(\frac{2v(r)}{1-w(r)}\right), \quad j = 1, \ldots, n-1,$$

follows for Δt sufficiently small. The results of the theorem then obtain.

Proof of Step 1:

We first write an equation similar to (4.4) for f_{j+1}, subtracting (4.4) from this equation to obtain

$$
\begin{aligned}
&\alpha f_{j+1} + f_{j+1}\int_{t_j}^{t_{j+1}} \tilde{k}(t_{j+1}-s)\,ds \\
(4.10) \qquad &= [\tilde{g}(t_{j+1}) - \tilde{g}(t_j)] + [\tilde{d}(t_{j+1}) - \tilde{d}(t_j)] \\
&\quad - \sum_{\nu=1}^{j} f_\nu \int_{t_{\nu-1}}^{t_\nu} \left[\tilde{k}(t_{j+1}-s) - \tilde{k}(t_j - s)\right]ds + \alpha f_j,
\end{aligned}
$$

for $j = 1, \ldots, n-1$, where we have defined $\alpha = \mu/k_1$.

Similarly, the solution f of (1.1) satisfies

$$
\begin{aligned}
&\sum_{\ell=1}^{r} s_\ell \int_0^{t_j+(\ell-1)\Delta t} k(t_j+(\ell-1)\Delta t-s)f(s)\,ds + \alpha f(t_j) \\
&= \sum_{\ell=1}^{r} s_\ell g(t_j+(\ell-1)\Delta t) + \alpha f(t_j), \quad j = 1, \ldots, n,
\end{aligned}
$$

or,

$$
\begin{aligned}
(4.11) \qquad &\int_0^{t_j} \tilde{k}(t_j - s)f(s)\,ds + \alpha f(t_j) \\
&= \tilde{g}(t_j) + \alpha f(t_j) \\
&\quad - \sum_{\ell=1}^{r} s_\ell \int_0^{(\ell-1)\Delta t} k((\ell-1)\Delta t - s)f(t_j + s)\,ds, \quad j = 1, \ldots, n.
\end{aligned}
$$

A similar equation holds at t_{j+1}, so subtracting (4.11) from this new equation, and then subtracting (4.10) from the difference, yields

$$
\begin{aligned}
&\alpha[f(t_{j+1}) - f_{j+1}] + \int_{t_j}^{t_{j+1}} \tilde{k}(t_{j+1}-s)[f(s) - f_{j+1}]\,ds \\
(4.12) \qquad &= -[\tilde{d}(t_{j+1}) - \tilde{d}(t_j)] + \alpha[f(t_{j+1}) - f(t_j)] + \alpha[f(t_j) - f_j]
\end{aligned}
$$

$$-\sum_{\nu=1}^{j}\int_{t_{\nu-1}}^{t_\nu}\left[\tilde{k}(t_{j+1}-s)-\tilde{k}(t_j-s)\right][f(s)-f_\nu]\,ds$$

$$-\sum_{\ell=1}^{r}s_\ell\int_0^{(\ell-1)\Delta t}k((\ell-1)\Delta t-s)\left[f(t_{j+1}+s)-f(t_j+s)\right]\,ds,$$

for $j=1,\ldots,n-1$. Now let σ_j be given by (4.5). Then

$$f(t)-f_j=\Delta t\left[\sigma_j+\left(\frac{t-t_j}{\Delta t}\right)f'(\zeta_j(t))\right],\quad t\in(t_{j-1},t_j],$$

where $\zeta_j(t)\in(t,t_j)$, and (4.12) becomes

$$\alpha\sigma_{j+1}+\int_{t_j}^{t_{j+1}}\tilde{k}(t_{j+1}-s)\left[\sigma_{j+1}+\left(\frac{s-t_{j+1}}{\Delta t}\right)f'(\zeta_{j+1}(s))\right]\,ds$$

$$=-\frac{1}{\Delta t}[\tilde{d}(t_{j+1})-\tilde{d}(t_j)]+\alpha\sigma_j+\alpha\frac{[f(t_{j+1})-f(t_j)]}{\Delta t}$$

$$-\sum_{\nu=1}^{j}\int_{t_{\nu-1}}^{t_\nu}\left[\tilde{k}(t_{j+1}-s)-\tilde{k}(t_j-s)\right]\left[\sigma_\nu+\left(\frac{s-t_\nu}{\Delta t}\right)f'(\zeta_\nu(s))\right]\,ds$$

$$-\frac{1}{\Delta t}\sum_{\ell=1}^{r}s_\ell\int_0^{(\ell-1)\Delta t}k((\ell-1)\Delta t-s)\left[f(t_{j+1}+s)-f(t_j+s)\right]\,ds,$$

or, for ξ_j and $\hat{\xi}_j$ suitably defined and $j=1,\ldots,n-1$,

(4.13)
$$\left[\alpha+\int_{t_j}^{t_{j+1}}\tilde{k}(t_{j+1}-s)\,ds\right]\sigma_{j+1}$$

$$=\alpha\sigma_j-\Delta t\sum_{\nu=1}^{j}\sigma_\nu\int_{t_{\nu-1}}^{t_\nu}\tilde{k}'(\xi_j(s))\,ds$$

$$-\frac{1}{\Delta t}[\tilde{d}(t_{j+1})-\tilde{d}(t_j)]-r_{j+1}$$

where

$$r_{j+1}(\Delta t,r,\mu)=-\alpha f'(\hat{\xi}_j)$$

$$+\int_{t_j}^{t_{j+1}}\tilde{k}(t_{j+1}-s)\left(\frac{s-t_{j+1}}{\Delta t}\right)f'(\zeta_{j+1}(s))\,ds$$

$$+\Delta t\sum_{\nu=1}^{j}\int_{t_{\nu-1}}^{t_\nu}\tilde{k}'(\xi_j(s))\left(\frac{s-t_\nu}{\Delta t}\right)f'(\zeta_\nu(s))\,ds$$

$$+\sum_{\ell=1}^{r}s_\ell\int_0^{(\ell-1)\Delta t}k((\ell-1)\Delta t-s)f'(\hat{\xi}_{j,\ell}(s))\,ds.$$

An equation in σ_1 is obtained by subtracting (4.4) from (4.11) for $j=1$ and using the definition of σ_1 in (4.5), i.e.,

(4.14)
$$\left[\alpha+\int_0^{t_1}\tilde{k}(t_1-s)\,ds\right]\sigma_1=-\frac{\tilde{d}(t_1)}{\Delta t}-r_1$$

where

$$r_1(\Delta t, r, \mu) = \int_0^{t_1} \tilde{k}(t_1 - s) \left(\frac{s - t_1}{\Delta t} \right) f'(\zeta_1(s)) \, ds$$
$$- \alpha \frac{f(t_1)}{\Delta t} + \frac{1}{\Delta t} \sum_{\ell=1}^{r} s_\ell \int_0^{(\ell-1)\Delta t} k((\ell-1)\Delta t - s) f(t_1 + s) \, ds.$$

Let $D(\Delta t, r, \mu) = \alpha + \int_0^{\Delta t} \tilde{k}(\Delta t - s) \, ds$ denote the coefficient of σ_1 in (4.14),

$$D(\Delta t, r, \mu) = \frac{\mu}{k_1} + \sum_{\ell=1}^{r} s_\ell \int_0^{\Delta t} k(\Delta t + (\ell-1)\Delta t - s) \, ds.$$

Then, for $r = 1$,

$$D(\Delta t, 1, \mu) = \frac{\mu + k_1^2}{k_1} > 0,$$

which holds for all $\Delta t > 0$ and $\mu > 0$; on the other hand, for general $r = 2, 3, \ldots$,

$$D(\Delta t, r, \mu) = \frac{1}{c_1} \left[\gamma_r \mu + \sum_{\ell=1}^{r} c_\ell k_\ell \right]$$

(4.15)
$$= \frac{1}{c_1} \det(\mathbf{K}_r^\top \mathbf{K}_r + \mu \mathbf{I}_r)$$

where we have used (3.5) and the relationships $c_1 = \gamma_r k_1$ and $s_\ell = c_\ell/c_1$ from the previous section. But (3.8) gives

$$D(\Delta t, r, \mu) = \frac{1}{k_1} \frac{\det(\mathbf{K}_r^\top \mathbf{K}_r + \mu \mathbf{I}_r)}{\det(\mathbf{K}_{r-1}^\top \mathbf{K}_{r-1} + \mu \mathbf{I}_{r-1})}$$

or

$$D(\Delta t, r, \mu) = \frac{1}{k_1} \frac{\prod_{i=1}^{r}(\lambda_{i,r} + \mu)}{\prod_{i=1}^{r-1}(\lambda_{i,r-1} + \mu)}$$

where $0 < \lambda_{1,r} \leq \ldots \leq \lambda_{r,r}$ denote the eigenvalues of $\mathbf{K}_r^\top \mathbf{K}_r$. Since $\mathbf{K}_{r-1}^\top \mathbf{K}_{r-1}$ is a principal submatrix of $\mathbf{K}_r^\top \mathbf{K}_r$, a standard eigenvalue interlacing theorem (see, for example, [6], p. 294) may be used to claim that

$$\lambda_{i,r} \leq \lambda_{i,r-1} \leq \lambda_{i+1,r}, \quad i = 1, \ldots, r - 1,$$

which yields the estimate

(4.16)
$$D(\Delta t, r, \mu) \geq \frac{1}{k_1}(\lambda_{1,r} + \mu),$$

in particular, $D(\Delta t, r, \mu) > 0$ for all $\Delta t > 0$, $\mu > 0$, and $r = 1, 2, \ldots$. We may therefore divide both sides of equations (4.13) and (4.14) by $D(\Delta t, r, \mu)$ to obtain equations (4.6) and (4.7), where

$$W(\Delta t, r, \mu) = \frac{\alpha}{D(\Delta t, r, \mu)}$$

$$V_{j,\nu}(\Delta t, r, \mu) = \frac{\int_{t_{\nu-1}}^{t_\nu} \tilde{k}'(\xi_j(s)) \, ds}{D(\Delta t, r, \mu)}, \quad \nu = 1, \ldots, j; \; j = 1, \ldots, n-1$$

(4.17) $\quad d_1(\delta; \Delta t, r, \mu) = \dfrac{1}{\Delta t} \dfrac{\tilde{d}(t_1)}{D(\Delta t, r, \mu)}$

$$d_{j+1}(\delta; \Delta t, r, \mu) = \frac{1}{\Delta t} \frac{[\tilde{d}(t_{j+1}) - \tilde{d}(t_j)]}{D(\Delta t, r, \mu)}, \quad j = 1, \ldots, n-1$$

$$R_{j+1}(\Delta t, r, \mu) = \frac{r_{j+1}(\Delta t, r, \mu)}{D(\Delta t, r, \mu)}, \quad j = 0, \ldots, n-1.$$

Thus the proof of Step 1 is complete.

Proof of Step 2:

An observation which aids in obtaining bounds for the quantities in (4.17)) is that for any $\rho = (\rho_1, \ldots, \rho_r)^\top$ and $r = 2, 3, \ldots,$

$$\frac{1}{D(\Delta t, r, \mu)} \sum_{\ell=1}^r s_\ell \rho_\ell = \frac{\dfrac{1}{c_1} \displaystyle\sum_{\ell=1}^r c_\ell \rho_\ell}{\dfrac{1}{c_1} \det(\mathbf{K}_r^\top \mathbf{K}_r + \mu \mathbf{I}_r)}$$

$$= \frac{\det(\mathbf{K}_r^\top \rho \,|\, \mathbf{t}_2, \ldots, \mathbf{t}_r)}{\det(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_r)}$$

$$= h_1$$

where we have used the definition of c_ℓ and $\mathbf{t}_\ell$ from the previous section, and where h_1 is the first component of the solution $\mathbf{h}$ of $(\mathbf{K}_r^\top \mathbf{K}_r + \mu \mathbf{I}_r)\mathbf{h} = \mathbf{K}_r^\top \rho$. Thus,

$$\frac{1}{D(\Delta t, r, \mu)} \sum_{\ell=1}^r s_\ell \rho_\ell \le \|\mathbf{h}\|$$

$$\le \|(\mathbf{K}_r^\top \mathbf{K}_r + \mu \mathbf{I}_r)^{-1}\| \, \|\mathbf{K}_r^\top\| \, \|\rho\|$$

$$\le \frac{\Delta t}{\mu} r^2 \|k\|_\infty \rho_{\max}$$

where $\rho_{\max} \equiv \max_{1 \le j \le r} |\rho_j|$ and $\|\cdot\|_\infty$ denotes the usual $C[0, T]$ norm, and where we have used standard relationships between the usual Euclidean norm and the max-norm on $\mathbb{R}^r$. In the case of $r = 1$ we obtain the same estimate since

$$\left| \frac{s_1 \rho_1}{D(\Delta t, 1, \mu)} \right| = \frac{k_1 |\rho_1|}{k_1^2 + \mu}$$

$$\le \frac{\Delta t}{\mu} \|k\|_\infty \rho_{\max}.$$

It follows from (4.16) that

$$\frac{\alpha}{D(\Delta t, r, \mu)} \le \frac{\mu}{\lambda_{1,r} + \mu},$$

for $r = 2, 3, \ldots,$ while for $r = 1$ the same bound also obtains (in that case, $\mathbf{K}_r^\top \mathbf{K}_r$ is the scalar k_1^2).

Using these observations we may bound the quantities in (4.17) as follows:

$$0 < W(\Delta t, r, \mu) \le \frac{1}{\frac{\lambda_{1,r}}{\mu} + 1}$$

$$|V_{j,\nu}(\Delta t, r, \mu)| \le \frac{1}{D(\Delta t, r, \mu)} \sum_{\ell=1}^{r} s_\ell \int_{t_{\nu-1}}^{t_\nu} |k'(\xi_{j,\ell}(s))| \, ds$$

$$\le \frac{\Delta t^2}{\mu} r^2 \|k\|_\infty \|k'\|_\infty$$

$$|d_{j+1}(\delta; \Delta t, r, \mu)| \le \frac{2}{\Delta t} \frac{\sum_{\ell=1}^{r} s_\ell \delta}{D(\Delta t, r, \mu)}$$

$$\le \frac{2\delta}{\mu} r^2 \|k\|_\infty, \qquad j = 0, 1, \ldots, n-1,$$

for suitably defined $\xi_{j,\ell}(s)$ and for all $r = 1, 2, \ldots$, $\mu > 0$, and $\Delta t > 0$. Further,

$$|R_{j+1}(\Delta t, r, \mu)| \le \frac{\alpha}{D(\Delta t, r, \mu)} \|f'\|_\infty$$

$$+ \frac{1}{D(\Delta t, r, \mu)} \sum_{\ell=1}^{r} s_\ell \left[\int_{t_j}^{t_{j+1}} k(t_{j+1} + (\ell-1)\Delta t - s) \left(\frac{s - t_{j+1}}{\Delta t} \right) f'(\zeta_{j+1}(s)) \, ds \right]$$

$$+ \frac{\Delta t}{D(\Delta t, r, \mu)} \sum_{\ell=1}^{r} s_\ell \left[\sum_{\nu=1}^{j} \int_{t_{\nu-1}}^{t_\nu} k'(\xi_{j,\ell}(s)) \left(\frac{s - t_\nu}{\Delta t} \right) f'(\zeta_\nu(s)) \, ds \right]$$

$$+ \frac{1}{D(\Delta t, r, \mu)} \sum_{\ell=1}^{r} s_\ell \left[\int_0^{(\ell-1)\Delta t} k((\ell-1)\Delta t - s) f'(\hat{\xi}_{j,\ell}(s)) \, ds \right]$$

$$\le \frac{\|f'\|_\infty}{\frac{\lambda_{1,r}}{\mu} + 1} + \frac{\Delta t^2}{\mu} r^2 \|k\|_\infty \|f'\|_\infty \left[r\|k_\infty\| + \|k'\|_\infty \right],$$

for $j = 1, \ldots N - 1$ and $r = 1, 2, \ldots$. Under the additional assumption that $f(0) = 0$, it follows that

$$R_1(\Delta t, r, \mu) = - \frac{\alpha}{D(\Delta t, r, \mu)} \frac{f(t_1)}{\Delta t}$$

$$+ \frac{1}{D(\Delta t, r, \mu)} \sum_{\ell-1}^{r} s_\ell \left[\int_0^{t_1} k(t_1 + (\ell-1)\Delta t - s) \left(\frac{s - t_1}{\Delta t} \right) f'(\zeta_1(s)) \, ds \right]$$

$$+ \frac{1}{\Delta t} \frac{1}{D(\Delta t, r, \mu)} \sum_{\ell=1}^{r} s_\ell \left[\int_0^{(\ell-1)\Delta t} k((\ell - s)\Delta t - s) f(t_1 + s) \, ds \right]$$

$$\le \frac{\|f'\|_\infty}{\frac{\lambda_{1,r}}{\mu} + 1} + \frac{\Delta t^2}{\mu} r^4 \|k\|_\infty^2 \|f'\|_\infty.$$

for $r = 1, 2, \ldots$.

We have yet to use the last assumed property of the kernel k, namely, $k(t) \ge k_{\min} > 0$ for $t \in [0, T]$. For this case, $k_i \equiv \int_0^{\Delta t} k(t_i - s) \, ds = \kappa_i \Delta t$ for some $\kappa_i \in$

$[k_{\min}, \|k\|_\infty]$ independent of Δt. Thus,

$$
\begin{aligned}
\lambda_{1,r} &= \min_{x \in \mathbb{R}^r} \frac{x^\top \mathbf{K}_r^\top \mathbf{K}_r x}{x^\top x} \\
&= \Delta t^2 \min_{x \in \mathbb{R}^r} \frac{x^\top \hat{\mathbf{K}}_r^\top \hat{\mathbf{K}}_r x}{x^\top x} \\
&= \Delta t^2 \rho(r)
\end{aligned}
$$

where $\rho(r) > 0$ is the smallest eigenvalue of $\hat{\mathbf{K}}_r^\top \hat{\mathbf{K}}_r$,

$$
\hat{\mathbf{K}}_r = \begin{pmatrix} \kappa_1 & 0 & \dots & 0 \\ \kappa_2 & \kappa_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \kappa_r & \kappa_{r-1} & \dots & \kappa_1 \end{pmatrix};
$$

it follows that $\rho(r)$ is independent of Δt. The choice of $\mu = \mu(\Delta t)$ given by $\mu = \hat{\mu}\Delta t^2$ ensures that $\lambda_{1,r}/\mu = \rho(r)/\hat{\mu} > 0$ for all $\Delta t > 0$. Thus, we obtain bounds for W, $V_{j,\nu}$, d_j and R_j as given by (4.8), and the proof of Step 2 is complete.

Proof of Step 3:

Using the results of Steps 1 and 2, a simple induction argument may be applied to (4.6),(4.7), to show that $|\sigma_j| \le B_j$ for $j = 1, \dots, n$, where

$$
B_1 = e(r)\frac{\delta}{\Delta t^2} + z(r)
$$

$$
B_{j+1} = w(r)B_j + \Delta t\, v(r) \sum_{\nu=1}^{j} B_\nu + e(r)\frac{\delta}{\Delta t^2} + z(r), \quad j = 1, \dots, n-1,
$$

and where each coefficient in this system of difference equations is bounded in the limit provided $\Delta t = \Delta t(\delta)$ is selected such that $\delta/\Delta t(\delta)^2$ remains bounded as $\delta \to 0$. It follows (see, for example, [5]) from the theory of finite difference equations that

$$
|\sigma_j| \le C_1(\delta, \Delta t, r)\, [\tau_1(\Delta t, r)]^j \;+\; C_2(\delta, \Delta t, r)\, [\tau_2(\Delta t, r)]^j, \quad \text{for } j = 1, \dots, n,
$$

where

$$
\tau_1(\Delta t, r) = 1 + \Delta t\frac{v(r)}{1 - w(r)} + \mathcal{O}(\Delta t^2)
$$

$$
\tau_2(\Delta t, r) = w(r)\left(1 - \Delta t\frac{v(r)}{1 - w(r)}\right) + \mathcal{O}(\Delta t^2)
$$

$$
C_1(\delta, \Delta t, r) = \frac{z(r) + e(r)\delta/\Delta t^2}{1 - w(r)} + \mathcal{O}(\Delta t)
$$

$$
C_2(\delta, \Delta t, r) = -w(r)\frac{z(r) + e(r)\delta/\Delta t^2}{1 - w(r)} + \mathcal{O}(\Delta t).
$$

Thus, we have that (4.9) holds and the remainder of the proof of Step 3 (and the theorem) follows. □

Remark: We note that under the assumption that $k(0) \ne 0$ for the kernel k of (1.1), it follows that the convergence rate of $\mathcal{O}(\delta^{1/2})$ we obtain in the last theorem is optimal with regard to the level δ of noise in the data [5, 8, 9].

We obtain similar results for sequential Tikhonov regularization where the discretization in (2.1), (2.2), is based on rectangular quadrature or midpoint quadrature. For rectangular quadrature the sequential Tikhonov iteration is given by (4.3) where $\tilde{k}_k$ is defined by (3.11) and k_i is given in this case by

$$k_i = \Delta t\, k(t_i), \tag{4.18}$$

while for midpoint quadrature, the approximation leads to (4.3) with

$$k_i = \Delta t\, k(t_{i-1/2}), \tag{4.19}$$

where $t_{i-1/2} = (i - \frac{1}{2})\Delta t$ for $i = 1, \ldots, n + r - 1$.

THEOREM 4.2. *Assume that the conditions of Theorem 4.1 hold. Then if $\mu = \mu(\Delta t)$ is selected such that $\mu = \hat{\mu}\Delta t^2$, with $\hat{\mu} > 0$ constant, and if $\Delta t = \Delta t(\delta)$ satisfies $\Delta t(\delta) = \tau\sqrt{\delta}$, with $\tau > 0$ constant, it follows that $\Delta t(\delta) \to 0$, $\mu(\Delta t) \to 0$, and*

$$|f_j - f(t_{j-1})| \leq \delta^{1/2}\overline{C}(r) + \mathcal{O}(\delta) \to 0, \qquad \text{for } j = 1, \ldots, n(\delta),$$

as $\delta \to 0$, where $\mathbf{f} = (f_1, \ldots, f_n)^\top$ is the solution of the sequential Tikhonov regularization problem for approximations based on rectangular quadrature and perturbed data g^δ (i.e., $\mathbf{f}$ solves (4.3) with $\tilde{k}_i = \sum_{\ell=1}^{r} s_\ell k_{i+\ell-1}$, where k_i is given by (4.18)). Here $n(\delta) = 1/\Delta t(\delta)$ and $\overline{C}(r)$ is a fixed positive constant.

Proof. The proof is a slight alteration of the proof of Theorem 4.1, due to the fact that here the equations satisfied by the solution f of (1.1) are written

$$\alpha f(t_{j-1}) + \sum_{\nu=1}^{j} \tilde{k}_{j-\nu+1} f(t_{\nu-1}) = \tilde{g}(t_j) + \alpha f(t_{j-1})$$
$$- \sum_{\ell=1}^{r} s_\ell \int_0^{(\ell-1)\Delta t} k((\ell-1)\Delta t - s) f(t_j + s)\, ds$$
$$- \sum_{\nu=1}^{j} \left[\int_{t_{\nu-1}}^{t_\nu} \tilde{k}(t_j - s) f(s)\, ds - \tilde{k}_{j-\nu+1} f(t_{\nu-1}) \right],$$

for $j = 1, \ldots, n$ (in place of the equations given in (4.11) for collocation). Thus in this case, (4.12) becomes

$$[\alpha + \tilde{k}_1][f(t_j) - f_{j+1}]$$
$$= -[\tilde{d}(t_{j+1}) - \tilde{d}(t_j)] + \alpha[f(t_j) - f(t_{j-1})]$$
$$+ \alpha[f(t_{j-1}) - f_j] - \sum_{\nu=1}^{j} \left[\tilde{k}_{j-\nu+2} - \tilde{k}_{j-\nu+1} \right] [f(t_{\nu-1}) - f_\nu]$$
$$- \sum_{\ell=1}^{r} s_\ell \int_0^{(\ell-1)\Delta t} k((\ell-1)\Delta t - s) [f(t_{j+1} + s) - f(t_j + s)]\, ds$$
$$- \left[\int_{t_j}^{t_{j+1}} \tilde{k}(t_{j+1} - s) f(s)\, ds - \tilde{k}_1 f(t_j) \right]$$
$$- \sum_{\nu=1}^{j} \left\{ \int_{t_{\nu-1}}^{t_\nu} \left[\tilde{k}(t_{j+1} - s) - \tilde{k}(t_j - s) \right] f(s)\, ds - \left[\tilde{k}_{j-\nu+2} - \tilde{k}_{j-\nu+1} \right] f(t_{\nu-1}) \right\}.$$

Letting $\sigma_{j+1} = [f(t_j) - f_{j+1}]/\Delta t$, we derive equations that are simple modifications of (4.13). Equations for σ_1 are similar, and the remaining analysis for convergence of the rectangular quadrature approximation follows in a standard way from the proof of Theorem 4.1, where we need only add the result that rectangular quadrature is an $\mathcal{O}(\Delta t)$ approximation of the integrals which appear above.   □

The proof of a similar result for midpoint quadrature is an easy modification of the last proof. Although midpoint quadrature is more accurate than rectangular quadrature, the presence of perturbed data g^δ in the equations for sequential Tikhonov regularization means that the order of convergence with respect to δ is unchanged from that given in the last two theorems.

THEOREM 4.3. *Assume that the conditions of Theorem 4.1 hold. Then if $\mu = \mu(\Delta t)$ is selected such that $\mu = \hat{\mu}\Delta t^2$, with $\hat{\mu} > 0$ constant, and if $\Delta t = \Delta t(\delta)$ satisfies $\Delta t(\delta) = \tau\sqrt{\delta}$, with $\tau > 0$ constant, it follows that $\Delta t(\delta) \to 0$, $\mu(\Delta t) \to 0$, and*

$$|f_j - f(t_{j-1/2})| \le \delta^{1/2}\overline{C}(r) + \mathcal{O}(\delta) \to 0, \qquad \text{for } j = 1, \ldots, n(\delta),$$

as $\delta \to 0$, where $\mathbf{f} = (f_1, \ldots, f_n)^\top$ is the solution of the sequential Tikhonov regularization problem for approximations based on midpoint quadrature and using perturbed data g^δ (i.e., $\mathbf{f}$ solves (4.3) with $\tilde{k}_i = \sum_{\ell=1}^r s_\ell k_{i+\ell-1}$, where k_i is given by (4.19)). Here $n(\delta) = 1/\Delta t(\delta)$ and $\overline{C}(r)$ is a fixed positive constant.

5. Implementation. For algorithmic purposes it is more useful to write the Tikhonov problem (2.4) in the form

$$(5.1) \qquad \min_\beta \left\| \begin{pmatrix} \mathbf{K}_r \\ \mu\mathbf{L}_r \end{pmatrix} \beta - \begin{pmatrix} \mathbf{h}^{(i)} \\ 0 \end{pmatrix} \right\|.$$

The algorithm described in §2 can be written as follows.

Sequential regularization algorithm, preliminary version.
    1. **for** $i := 1$ **to** $n - r + 1$
        (a) Solve (5.1) for β.
        (b) Put $f_i := \beta_1$.

We will now discuss how to implement this algorithm efficiently. We assume that $i-1$ iterations have been completed and that we shall next perform iteration i.

Consider first the computation of the right hand side $\mathbf{h}^{(i)}$. Using the definition (2.5) we immediately see that the first $r-1$ components of $\mathbf{h}^{(i)}$ can be obtained easily by updating quantities from the previous step. Thus we can compute $\mathbf{h}^{(i)}$ from

$$(5.2) \qquad h_p^{(i)} = h_{p+1}^{(i-1)} - k_p f_{i-1}, \qquad p = 1, \ldots, r-1,$$

$$(5.3) \qquad h_r^{(i)} = g_{i+r-1} - \sum_{j=1}^{i-1} k_{i+r-j} f_j.$$

The operation count for this step is $r + i - 2$ multiplications.

A least squares problem (5.1) is solved in each step of the algorithm. For reasons of numerical stability it not recommended [4, p. 230] to solve this using the normal equations (3.2). Instead a method based on an orthogonal decomposition of the matrix should be used. The matrices $\mathbf{K}_r$ and $\mathbf{L}_r$ in (5.1) both have lower triangular Toeplitz structure, and therefore (5.1) can be efficiently and accurately solved using

the method in [2]. In this algorithm an orthogonal $\mathbf{Q}$ matrix is determined, which transforms the matrix in (5.1) to the form,

$$(5.4) \qquad \mathbf{Q}^\top \begin{pmatrix} \mathbf{K}_r \\ \mu \mathbf{L}_r \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{K}}_r \\ 0 \end{pmatrix},$$

where $\hat{\mathbf{K}}_r$ is lower triangular but no longer has Toeplitz structure. The matrix $\mathbf{Q}$ is represented as a product of Givens rotations, and is never formed explicitly. The operation count for this transformation is $2r^2$ multiplications [2]. Note that this computation needs to be done only once, as a preprocessing step.

After multiplying the right hand side by $\mathbf{Q}^\top$,

$$(5.5) \qquad \mathbf{Q}^\top \begin{pmatrix} \mathbf{h}^{(i)} \\ 0 \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{h}}^{(i)} \\ \tilde{\mathbf{h}}^{(i)} \end{pmatrix},$$

the first component of the solution is obtained from

$$(5.6) \qquad f_i := \beta_1 := \frac{\hat{h}_1^{(i)}}{\hat{k}_1},$$

where $\hat{k}_1$ is the element in position $(1,1)$ in $\hat{\mathbf{K}}_r$. The transformation (5.5) requires approximately $2r^2$ multiplications [2].

Below we present the overall algorithm.

Sequential regularization algorithm.
1. (a) Compute the decomposition (5.4) and store the matrix $\mathbf{Q}$ as a product of Givens rotations.
   (b) Initialize $\mathbf{h}^{(0)} = (g_1, \ldots, g_r)^\top$.
2. **for** $i := 1$ **to** $n - r + 1$
   (a) Compute $\mathbf{h}^{(i)}$ from (5.2) and (5.3).
   (b) Perform the multiplication (5.5).
   (c) Compute f_i from (5.6).

Assuming that $r \ll n$, we can disregard the cost for Step 1. Similarly, in the operation count for Step 2(a) we omit terms that are $\mathcal{O}(n + nr + r^2)$. We obtain the following simplified operation count for the algorithm,

$$\frac{n^2}{2} + 2nr^2,$$

multiplications, where the first term comes from the forward substitution (5.2) and (5.3), and the second from (5.5). This should be compared to the cost of solving a triangular system of dimension n, which is $n^2/2$ multiplications, and the cost for standard Tikhonov regularization (using the fast algorithm in [2]), which is $4.5n^2$ multiplications[1]. Thus, the sequential algorithm is faster than standard regularization for $r < \sqrt{2n}$.

In some cases it may be desired to retain more than one component of the vector β in each iteration of the algorithm. It is easy to modify the above algorithm to the

---

[1]The algorithm in [2] takes advantage of the Toeplitz structure. If this is not done, then the operation count is $\mathcal{O}(n^3)$.

case when $p > 1$ components of the solution are retained. For example, in the case $p = r/2$ and n is large, we get the operation count $n^2/2 + \mathcal{O}(nr)$ multiplications.

One potentially important special case is when $\mathbf{K}$ and $\mathbf{L}$ are *band matrices.* For example, this may happen when the sideways heat equation[2] is solved over a very long time interval. Here it seems appropriate not to let measurements from long ago influence the present. In such cases one can define the kernel function of the Volterra equation (1.2) to be equal to zero for $t > t_0$, for some value of t_0. This leads to a matrix $\mathbf{K}$ with band structure, i.e., the elements of $\mathbf{K}$ satisfy $k_j = 0$, for $j > w$.

Since in most cases, the matrix $\mathbf{L}$ is a discretization of the identity or a differentiation operator, it is usually a band matrix with a narrow band.

It is trivial to modify the sequential regularization algorithm for this case. If $r \leq w$, then only the forward substitution (5.2) and (5.3) is affected, and it is easy to see that the operation count (in the case when only one unknown is computed in each step) becomes

$$nw + 2nr^2,$$

multiplications. If r is considerably larger than w, then more savings can be done in Steps 1 and 2(b), by taking advantage of band structure also here.

Of course, when $n \gg w$ the sequential method is much faster than standard Tikhonov regularization.

6. Numerical Examples. In this section we give a few numerical examples that illustrate sequential regularization. The purpose is not to investigate in detail the numerical properties but rather to demonstrate the usefulness of the proposed method. More extensive numerical experiments will be reported elsewhere.

Example I.   Consider the Volterra equation of the first kind

$$g_m(t) = Kf(t) = \int_0^t k_\kappa(t - \tau)f(\tau)d\tau, \qquad 0 \leq t \leq 1,$$

where the kernel function is defined by

$$(6.1) \qquad\qquad k_\kappa(t) = \frac{1}{2t^{3/2}\sqrt{\kappa\pi}} \exp\left[\frac{-1}{4\kappa t}\right].$$

This corresponds to the sideways heat equation (IHCP) in a quarter plane, where temperature measurements are done at $x = 1$ and $f(t)$ is the unknown temperature at $x = 0$. The parameter κ determines the heat conduction properties of the material: the smaller κ, the more ill-conditioned the discretized system. For information about this problem, see e.g. the recent surveys [1], [7], [3], and the references therein. This application is very important, and we choose to consider the example, in spite of the fact that since $k_\kappa(0) = 0$, the Volterra equation is considerably more ill-conditioned than is assumed in the theory in §4.

A solution function $f(t)$ was constructed, and the integral equation was discretized using midpoint quadrature, giving a lower triangular system of equations of dimension 256. The right hand side was formed by multiplying the vector $\mathbf{f}$ by the matrix $\mathbf{K}$, and then adding a random perturbation, normally distributed with mean zero and standard deviation d.

---

[2]Or Inverse Heat Conduction Problem (IHCP). For recent surveys of numerical procedures for the sideways heat equation, see [1], [7], [3].
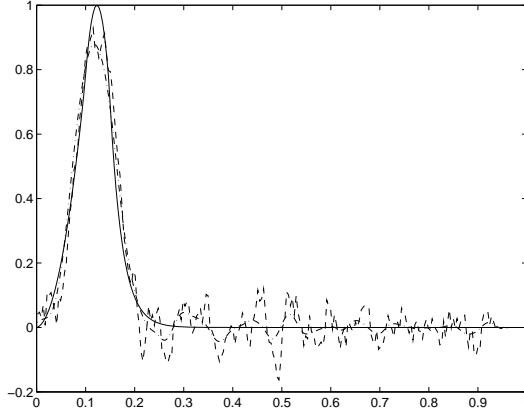
Fig. 6.1. *Numerical results for Example I, using sequential and standard regularization, with* $\kappa = 1$, $d = 10^{-3}$, $\mu = 0.25 \cdot 10^{-4}$, *and* $r = 18$.

In all graphs, the solid curve is the exact solution, and the dashed curve is the approximation using sequential regularization with $\mathbf{L}_r$ equal to the identity matrix. For comparison we also plot the result of standard Tikhonov regularization (2.3) (dashed-dotted curve), where we have used $\mathbf{L}$ equal to the identity matrix and chosen the same value of the regularization parameter as in sequential regularization.

In the examples given, we have chosen r as small as possible, such that the results using standard and sequential regularization agree well visually. Therefore, in a couple of the graphs it is difficult to distinguish between the two approximations.

In the first test case, we chose $d = 10^{-3}$ and $\kappa = 1$, and the results are shown in Figure 6.1. It is seen that a relatively large value of r is needed to make sequential regularization perform almost as well as standard regularization. However, note that with $r = 18$ sequential regularization is slightly more efficient than standard regularization.

Figure 6.2 shows that the value of r can be chosen considerably smaller, if κ is larger. Note that the standard deviation of the random perturbation is the same as in the previous example. If the noise level is lower, 10^{-4} say, then it is possible to get good results with r as small as 4.

Example II.   Consider the Volterra equation with kernel function $\hat{k}(t) = k_\kappa(t) + \theta$,

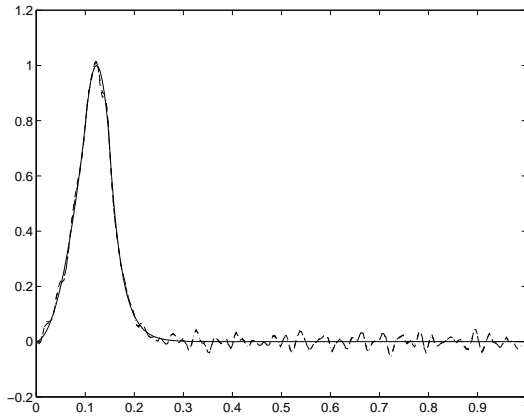FIG. 6.2. *Example I with $\kappa = 2$, $d = 10^{-3}$, $\mu = 0.49 \cdot 10^{-4}$, and $r = 8$.*

where $\kappa = 2$, $\theta = 10^{-3}$, and k_κ is defined by (6.1). Since $\hat{k}(t) > 0$, the assumptions of the previous sections are satisfied. The problem was discretized, and a solution and a right hand side were constructed exactly as in Example I. The matrix **K** has a condition number of the order 10^{18} so the problem cannot be solved without some kind of regularization.

The results are shown in Figure 6.3. Notice that here a very small value of r can be used, which makes sequential regularization much more efficent than standard regularization.

Our examples indicate that when the noise level is not too high and the problem is not too ill-conditioned, then r can be chosen so small that the sequential algorithm is considerably faster than and about as accurate as standard Tikhonov regularization (using the fast algorithm in [2]). For ill-conditioned problems with $k(t) > 0$ it may be possible to choose r as small as 4. More research is needed to find a recipe for choosing r, depending on the kernel function and the noise level.

REFERENCES

[1] J. V. BECK, B. BLACKWELL, AND C. R. ST. CLAIR, JR., *Inverse Heat Conduction. Ill-Posed Problems*, Wiley, New York, NY, 1985.
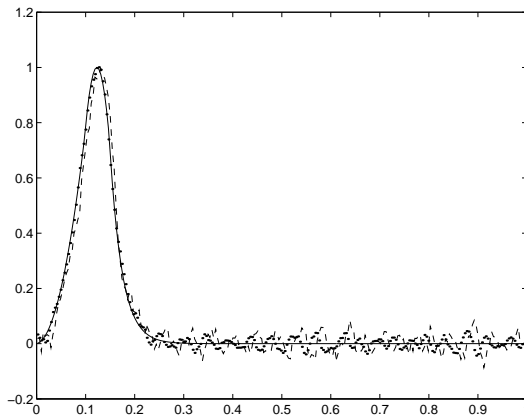
FIG. 6.3. *Numerical results for Example II, with $d = 10^{-3}$, $\mu = 0.2 \cdot 10^{-3}$ and $r = 4$.*

[2] L. ELDÉN, *An efficient algorithm for the regularization of ill-conditioned least squares problems with triangular Toeplitz matrix*, SIAM J. Sci. Statist. Comput., 5 (1984) pp. 229–236.

[3] ———, *Numerical solution of the sideways heat equation*, in Inverse Problems in Diffusion Processes, H. Engl and W. Rundell, eds., SIAM, Philadelphia, PA, 1995, pp. 130–150.

[4] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, Second ed., The Johns Hopkins University Press, Baltimore, MD, 1989.

[5] P. K. LAMM, *Approximation of ill-posed Volterra problems via predictor-corrector regularization methods*, SIAM J. Applied Math., 56 (1996), pp. 524–541.

[6] P. LANCASTER AND M. TISMENETSKY, *The Theory of Matrices*, Second ed., Academic Press, London, 1985.

[7] D. A. MURIO, *The Mollification Method and the Numerical Solution of Ill-Posed Problems*, J. Wiley & Sons, New York, NY, 1993.

[8] F. NATTERER, *On the order of regularization methods*, in Improperly Posed Problems and their Numerical Treatment, G. Hämmerlin and K. H. Hoffman, eds., Birkhäuser, Basel, pp. 189–203, 1983.

[9] ———, *Error bounds for Tikhonov regularization in Hilbert scales*, Applicable Analysis, 18 (1984), pp. 29–37.