APPROXIMATION OF ILL-POSED VOLTERRA PROBLEMS VIA PREDICTOR-CORRECTOR REGULARIZATION METHODS *

PATRICIA K. LAMM†

Abstract. First-kind Volterra problems arise in numerous applications, from inverse problems in mathematical biology to inverse heat conduction problems. Unfortunately, such problems are also ill-posed due to lack of continuous dependence of solutions on data. Consequently, numerical methods to solve first-kind Volterra equations are only effective when regularizing features are built into the algorithms or used to control stepsize. Classical methods often combine numerical discretization with Tikhonov regularization, but in doing so the underlying Volterra (or causal) nature of the original problem is often destroyed. Instead, a "predictor-corrector" type of numerical method is proposed which combines at each step "local regularization" ideas with the use of small intervals of future data. The result is a regularized numerical method which retains much of the causal nature of the Volterra problem and may be solved in fast sequential steps, often improving upon the performance of classical algorithms such as those based on standard Tikhonov regularization.

In this paper, the discretized local regularization method is described and proofs are given of convergence of the method, with rate of convergence being "best possible" with regard to the amount of error in the data. Further, by linking the regularization parameter of the stabilizing method (i.e., the length of the "future interval" in the future-sequential method) to the approximation stepsize, great simplification of the resulting numerical algorithm is obtained. Relevant numerical examples are included.

Key words. First-kind Volterra equations, inverse problems, regularization, collocation method

AMS subject classifications. 65J20, 45

1. Introduction. We consider here the solution of a first-kind Volterra integral equation of the form

$$\int_0^t k(t-s)u(s)\,ds = f(t), \quad t \in (0,1], \tag{1.1}$$

where we assume that given data f and convolution kernel $k = k(t)$ are such that a unique solution $\overline{u}$ of (1.1) exists in $L_2(0,1)$; very general conditions guaranteeing such existence/uniqueness may be found, for example, in [6]. Although we only consider scalar equations here, there is no difficulty in extending our ideas to a vector version of (1.1).

First-kind Volterra equations of this type occur naturally in many applications, a classic example being the inverse heat conduction problem (IHCP) or sideways heat equation, the problem of determining from *internal* temperature or temperature-flux measurements the unknown heat (or heat flux) source which is being applied at the *surface* of a solid [1]. Unfortunately, first-kind integral equations of the form (1.1) are well-known to be ill-posed in the sense that solutions are unstable with respect to $L_2(0,1)$ (or $L_\infty(0,1)$) perturbations f^ε in the data f. In fact, the Volterra integral equation associated with the IHCP is so severely ill-posed that closeness of $\dfrac{d^j}{dt^j} f^\varepsilon$ to $\dfrac{d^j}{dt^j} f$ for $j = 0, \ldots, p$, still does not guarantee the corresponding closeness of u^ε to $\overline{u}$, a fact which holds for any $p > 0$.

* This research was supported in part by the U. S. Air Force Office of Scientific Research under contract AFOSR-89-0419 and by the Clare Boothe Luce Foundation, NY, NY.

† Department of Mathematics, Michigan State University, East Lansing, MI 48824–1027 (lamm@math.msu.edu)

The extreme instability of the IHCP is due to the fact that the kernel k for this problem belongs to $C^\infty[0,1]$ and satisfies $k^{(n)}(0) = 0$ for all $n = 1, 2, \ldots$. The situation improves somewhat if the kernel does not degenerate so badly at $t = 0$, although the underlying problem is still unstable. We shall say that k is "ν-smoothing" if there exists an integer $\nu > 0$ such that $k \in H^\nu(0,1)$ and $k(0) = k'(0) = \ldots = k^{(\nu-2)}(0) = 0$, $k^{(\nu-1)}(0) \neq 0$. It is known that for such finitely-smoothing problems the degree of instability associated with solving (1.1) increases as the degree ν of smoothing increases (see, for example, [4, 10]). So long as ν is finite, the resulting instability is never as severe as it is for problems such as the IHCP, but in fact even with ν small (e.g., problems with $\nu = 1$ or 2 arise in mathematical biology [3, 6, 15] and many other fields), the problem is sufficiently unstable as to require regularization techniques in order to compute reasonably accurate solutions.

Stabilizing the solution of the integral equation (1.1) in the presence of noisy data may be handled via a classical technique such as Tikhonov regularization [7], where (for example) one seeks to determine the solution u_α^ε of the problem

$$\min_{u\in L_2(0,1)} \int_0^1 \left| \int_0^t k(t-s)u(s)\,ds - f^\varepsilon(t) \right|^2 dt + \alpha \int_0^1 |u(t)|^2 dt$$

where α is a "regularization parameter" and $f^\varepsilon = f(t) + \delta(t)$ is given in $L_2(0,1)$ with $\|\delta(\cdot)\| < \varepsilon$ (here $\|\cdot\|$ denotes the $L_2(0,1)$ norm). From the standard theory of Tikhonov regularization, one obtains a criterion for selecting $\alpha = \alpha(\varepsilon)$ which guarantees that $\alpha(\varepsilon) \to 0$ and $\|u^\varepsilon_{\alpha(\varepsilon)} - \overline{u}\| \to 0$ as $\varepsilon \to 0$, where $\overline{u}$ denotes the solution of the original unperturbed problem.

However, a distinct drawback of Tikhonov regularization for a first-kind Volterra problem lies in the fact that the normal equation (i.e., necessary condition for the minimization problem) associated with Tikhonov regularization is non-Volterra, or noncausal, in structure, even though the original equation (1.1) has Volterra, or causal, structure. As a consequence, while numerical methods to solve equation (1.1) can be devised which involve simple, lower-triangular matrices and fast, sequential techniques, numerical methods to solve the equations associated with Tikhonov regularization generally involve full matrices and more expensive non-sequential solution techniques. Even in the case where a transformation is made to restore to the regularized problem the lower-triangular structure of the original problem (e.g., [5]), the transformation still necessitates the use of *all* of the data at every time-step of the algorithm, a property which again illustrates how the causal nature of the original problem has been lost in the process.

In [8] a general "future-sequential" regularization theory is described which enables the stable solution of (1.1) but which also has the advantage of being structure-preserving in the sense that most of the basic causal properties associated with the underlying Volterra equation are preserved. Although this method is a generalization of an extremely simple and effective discretization method used by J. V. Beck for the stable solution of the IHCP, in its most general form our ideas depart significantly from those of Beck. Nevertheless, the convergence theory described in the next section does apply to the method used by Beck (for 1-smoothing problems) and thus we provide the first proof of convergence of Beck's method in this case.

In order to describe our more general approach here, we first let $T > 1$ be fixed and make the assumption that equation (1.1) actually holds on the interval $(0, T]$, with $\overline{u}$, k, f, and f^ε defined on this extended interval as well; as is discussed in [8], this assumption is made to simplify the presentation and may actually be avoided via slight changes in the theory of that reference. Let $\Delta_R > 0$ be a fixed constant with $1 + \Delta_R \leq T$, and define the regularization parameter for the method by Δ_r, where $0 < \Delta_r \leq \Delta_R$.

The idea behind future-sequential regularization is this. We know that the "true" solution $\overline{u}$ of (1.1) must satisfy the advanced-in-time equation

$$\int_0^{t+\rho} k(t+\rho-s)u(s)\,ds = f(t+\rho), \quad t \in (0,1], \;\; \rho \in [0,\Delta_r].$$

In order to regularize the solution of this problem in the presence of perturbed data, we fix t for the moment and assume that a reasonable solution has already been obtained on $[0,t]$; we then seek to find a solution on the interval $[t, t+\Delta_r]$, asking that the solution be "regular" in some sense on that interval. For example, one may seek a solution which is actually constant-valued on that interval, the value of the constant being given by that $\overline{c}_t \in \mathbb{R}$ for which the integral is matched locally to data in a least-squares sense over the interval $[t, t+\Delta_r]$; i.e., $\overline{c}_t$ is defined by

$$\overline{c}_t = \arg\min_{c\in\mathbf{R}} \int_0^{\Delta_r} \left| \int_0^t k(t+\rho-s)u(s)\,ds + c\int_t^{t+\rho} k(t+\rho-s)\,ds - f(t+\rho)\right|^2 d\rho. \tag{1.2}$$

But in fact we do not *a priori* know the solution on $[0,t]$ for any $t \in (0,1]$, so we instead use the above idea to motivate our general approach. That is, we now let $t \in (0,1]$ be arbitrary and seek a function u such that $u(t) = \overline{c}_t$ for *every* $t \in (0,1]$, with $\overline{c}_t$ given by (1.2). The normal equation associated with the t-dependent minimization problem in (1.2) is given as follows, where we have identified $u(t)$ with $\overline{c}_t$ and reversed an order of integration:

$$\begin{aligned}\int_0^t \int_0^{\Delta_r} k(t+\rho-s)\left(\int_0^\rho k(\rho-\tau)\,d\tau\right) d\rho\, u(s)\,ds + u(t)\int_0^{\Delta_r}\left(\int_0^\rho k(\rho-\tau)\,d\tau\right)^2 d\rho \\ = \int_0^{\Delta_r} f(t+\rho)\left(\int_0^\rho k(\rho-\tau)\,d\tau\right) d\rho,\end{aligned} \tag{1.3}$$

a second-kind equation which may be generalized as follows,

$$\int_0^t \tilde{k}(t-s;\Delta_r)u(s)\,ds + \alpha(\Delta_r)u(t) = \tilde{f}(t;\Delta_r), \tag{1.4}$$

where

$$\begin{aligned}\tilde{k}(t;\Delta_r) &= \int_0^{\Delta_r} k(t+\rho)\,d\eta_{\Delta_r}(\rho) \\ \alpha(\Delta_r) &= \int_0^{\Delta_r}\int_0^\rho k(\rho-\tau)d\tau\,d\eta_{\Delta_r}(\rho) \\ \tilde{f}(t;\Delta_r) &= \int_0^{\Delta_r} f(t+\rho)\,d\eta_{\Delta_r}(\rho)\end{aligned} \tag{1.5}$$

and η_{Δ_r} is a positive Borel-Stieltjes measure on the Borel subsets of $\mathbb{R}$, with $\int_0^{\Delta_r}(\cdot)\,d\eta_{\Delta_r}(\rho)$ a Stieltjes integral. In this paper the primary focus will be on two special cases of η_{Δ_r}:

(i) We consider the case of η_{Δ_r} a density, defined via

$$\int_0^{\Delta_r} \phi(\rho)\, d\eta_{\Delta_r}(\rho) = \int_0^{\Delta_r} \phi(\rho)\, \omega(\rho)\, d\rho, \tag{1.6}$$

for $\phi \in C[0,\Delta_r]$ and $\omega \in L_\infty[0,\Delta_r]$, $\omega(\rho) \geq \omega_{\text{min}} > 0$ for all $\rho \in [0,\Delta_r]$. We note that for the particular choice $\omega(\rho) = \int_0^\rho k(\rho-\tau)\, d\tau$, approximating equation (1.4) reduces to (1.3), the equation generated by the localized least-squares fitting described above.

(ii) We also consider the example of η_{Δ_r} a discrete measure, defined by

$$\eta_{\Delta_r} = \sum_{i=1}^{K} s_i\, \delta_{\tau_i \Delta_r}, \tag{1.7}$$

where $K = K(\Delta_r) \geq 2$ is an integer, $s_i > 0$ for $i = 1,\ldots,K$, and $0 \equiv \tau_1 < \tau_2 < \ldots < \tau_K \equiv 1$; here we use the notation $\delta_{\hat{t}}$ to denote the usual delta-function centered at $\hat{t} \in [0,\Delta_r]$. Depending on the choice of s_i and τ_i, the discrete measure may correspond to a localized discrete least-squares criterion (matching to K *discrete* present and future values of data f in the interval $[t, t+\Delta_r]$) for every $t \in [0,1]$. This particular formulation will be especially useful when we look at a numerical method for solving (1.4) and coordinate the numerical grid for the computations with the values of the parameters τ_i.

In [8, 10], general conditions on k, $\overline{u}$, η_{Δ_r}, and ν are given which guarantee the existence and uniqueness of a solution $u(\cdot;\Delta_r)$ of (1.4) and ensure the convergence of $u(\cdot;\Delta_r)$ to the solution $\overline{u}$ of (1.1) as $\Delta_r \to 0$ (in the case of noise-free data). Further, because (1.4) is a *second*-kind Volterra equation, its solution is stable with respect to perturbations in the data f. Therefore a statement of regularized approximation is obtained as follows: defining $u(\cdot;\Delta_r, f^\varepsilon)$ to be the solution of (1.4) associated with an ε-perturbation f^ε of f, there exists a choice of $\Delta_r = \Delta_r(\varepsilon)$ such that $\Delta_r(\varepsilon) \to 0$ and $u(\cdot;\Delta_r(\varepsilon), f^\varepsilon)$ converges in $L_2(0,1)$ to the solution $\overline{u}(\cdot;f)$ of (1.1) as the noise level $\varepsilon \to 0$ [8, 10]. Again these results are obtained under prescribed conditions on k, $\overline{u}$, η_{Δ_r}, and ν.

The convergence/stabilization theory described above for future-sequential regularization has been developed in [8, 10] in an infinite-dimensional setting. The central purpose of this paper is to provide a similar analysis of (1.4) in a finite-dimensional setting, where the resulting numerical algorithm exhibits an interesting "predictor-corrector" type of regularization. In the next section we set up the finite-dimensional algorithm and prove convergence of the algorithm for η_{Δ_r} given by (1.6) or (1.7). Finally, in §3 we describe the way in which the numerical algorithm becomes a predictor-corrector regularization algorithm, presenting there our numerical findings.

2. Finite-Dimensional Approximation. Finite-dimensional approximations are always regularizing in the sense that an unstable infinite-dimensional inverse problem is transformed into a stable finite-dimensional inverse problem, with degree of stability worsening as the dimension of the finite-dimensional approximation space increases. Thus we may always attempt to regularize the first-kind integral equation (1.1) via finite-dimensional approximation. When such an approximation converges, it does so as the approximation stepsize and the level ε of noise in the perturbed data f^ε decreases, so that the discretization parameter or approximation stepsize (defined below to be Δt) plays the role of a "regularization parameter". Generally, we must let $\Delta t \to 0$ slowly as $\varepsilon \to 0$ and it is this "holding-back" of Δt which performs the regularization. Unfortunately, in many situations the size of $\Delta t = \Delta t(\varepsilon)$ which is necessitated by the level ε of noise is unacceptably large, with the result that one is forced to accept crude approximations from finite-dimensional spaces of very low dimension in order to avoid the highly oscillatory approximations which result from data or computational error. In contrast, discretization of (1.4) leads to a method which, from all numerical evidence and from a related conditioning analysis [11], allows for the stable approximation of $\overline{u}$ using much smaller stepsize than does a method based on direct discretization of (1.1).

In this section we examine the convergence properties of collocation-based approximations of solutions $u(\cdot;\Delta_r)$ of (1.4), showing that a selection of $\Delta_r = \Delta_r(\varepsilon)$ and $\Delta t = \Delta t(\varepsilon)$ may be made ensuring that approximations $u_{\Delta t(\varepsilon)}(\cdot;\Delta_r(\varepsilon))$ converge to $\overline{u}$ as $\varepsilon \to 0$. The theory we develop is limited to integral equations with 1-smoothing kernels, although (as numerical results later show) convergence appears to obtain for general ν-smoothing problems as well, and even to infinitely smoothing problems like the IHCP [1]. Since 1-smoothing first-kind integral equations become well-posed second-kind equations after a single differentiation of the equation (provided that the data is smooth), the discretized analog of the first-kind equation may be analyzed similarly via differencing instead of differentiation; this is the approach taken here, following [2]. However, we do not make the stringent assumption that the perturbed data f^ε is smooth enough to allow differentiation, so we depart significantly from [2] in our treatment of f^ε and additionally in the presence of a regularization parameter Δ_r which enters into every term in (1.4).

Throughout this section we will make the following assumptions regarding k, $\overline{u}$ and perturbed data f^ε:

> **Hypothesis 2.1:** Let $\overline{u}$, $k \in W^{1,\infty}(0,T)$. In addition, let k be 1-smoothing and such that $\alpha(\Delta_r)$ defined by (1.5) satisfies $\alpha(\Delta_r) > 0$ for all $\Delta_r \in (0, \Delta_R]$ (conditions which are satisfied if $k > 0$ on $[0,\Delta_R]$). Further, let $f^\varepsilon(t) = f(t) + \delta(t)$, where $\delta \in L_\infty(0,T)$ satisfies $\|\delta\|_\infty \leq \varepsilon$ for some fixed $\varepsilon > 0$ and where $\int_0^{\Delta_r} \delta(t+\rho)\, d\eta_{\Delta_r}(\rho)$ is well defined for all $t \in (0,1)$ and all $\Delta_r \in (0,\Delta_R]$; here $\|\cdot\|_\infty$ denotes the $L_\infty(0,T)$ norm.

In order to implement a discretized approximation of the solution $u(\cdot;\Delta_r;f^\varepsilon)$ of (1.4), we let $N = 1,2,\ldots$ be fixed and define a gridsize $\Delta t = 1/N$ with equally-spaced gridpoints on $[0,1]$, $t_j \equiv j\Delta t$, $j = 0,1,\ldots,N$. It will also be convenient to extend the gridpoints past the interval $[0,1]$ by defining t_j in the same way for $j = N+1, N+2, \ldots$. We define piecewise-constant approximation spaces of the form $\mathcal{S}_{\Delta t}^{-1} = \mathrm{span}\{\phi_j\}_{j=0}^{N-1}$, where $\phi_j(t) = 1$ on the interval $(t_j, t_{j+1}]$, $\phi_j(t) = 0$ otherwise on $[0,1]$. We then seek $u_{\Delta t}(\cdot;\Delta_r;f^\varepsilon) \in \mathcal{S}_{\Delta t}^{-1}$ which exactly matches, or collocates, the regularized equation (1.4) (with f^ε in place of f) to N collocation points t_j, for $j = 1,2,\ldots,N$. Writing

$$u_{\Delta t}(t;\Delta_r;f^\varepsilon) = \sum_{j=0}^{N-1} \alpha_j \phi_j(t), \quad t \in [0,1],$$

the collocation procedure leads to linear equations in the basis coefficients α_j which are especially simple if we assume that the "future" interval Δ_r is exactly proportional to gridsize; we henceforth make this assumption and define

$$\Delta_r = (r-1)\Delta t,$$

where $r \geq 2$ is a fixed integer.

If equation (1.4) is to be exactly satisfied at t_{j+1} for $j = 0, \ldots, N-1$, the equation in α_j then becomes

$$\alpha(\Delta_r)\,\alpha_j + \int_0^{t_{j+1}} \tilde{k}(t_{j+1}-s;\Delta_r)\left(\sum_{\mu=0}^{j} \alpha_\mu \phi_\mu(s)\right) ds$$
$$= \int_0^{\Delta_r} f^\varepsilon(t_{j+1}+\rho)\, d\eta_{\Delta_r}(\rho), \quad j = 0, \ldots, N-1, \tag{2.1}$$

a lower-triangular linear system in $(\alpha_0, \ldots, \alpha_{N-1})^\top$ for which the governing matrix has nonzero diagonal (for Δt sufficiently small), guaranteeing existence and uniqueness of a solution $u_{\Delta t}(\cdot;\Delta_r;f^\varepsilon) \in \mathcal{S}_{\Delta t}^{-1}$ of (2.1) for all small Δt. We note that piecewise-polynomial spaces more general than $\mathcal{S}_{\Delta t}^{-1}$ may be used in this collocation procedure, however piecewise-constant functions are ideally suited as approximations for a method which regularizes by temporarily holding solutions constant over the interval $[t, t+\Delta_r]$ (as was discussed in motivating the future-sequential regularization method in §1).

We consider the question of convergence for the above approximation scheme, beginning with the case of η_{Δ_r} satisfying (1.6) and then turning to the case of discrete η_{Δ_r} given by (1.7). We note that under the conditions of the theorem, we obtain an $\mathcal{O}(\varepsilon^{1/2})$ rate of convergence as the level ε of error in data goes to zero, which is the best possible rate of convergence with respect to ε that one can expect in the case of a 1-smoothing kernel k, $\overline{u} \in H^1(0,1)$, and for error in data measured in an $L_2(0,1)$ sense, conditions which hold in the theorems of this section (see, e.g., [13, 14] for a discussion of "best possible" convergence rates for ill-posed problems, and [10] for a discussion of 1-smoothing Volterra problems).

THEOREM 2.1. *Assume k, $\overline{u}$, and f^ϵ satisfy Hypothesis 2.1 for $\varepsilon > 0$, and let $\Delta_r = (r-1)\Delta t$, where $r \geq 2$ is a fixed integer and $\Delta t > 0$. Suppose further that η_{Δ_r} is defined via*

$$\int_0^{\Delta_r} \phi(\rho)\, d\eta_{\Delta_r}(\rho) = \int_0^{\Delta_r} \phi(\rho)\,\omega(\rho)\, d\rho, \quad \phi \in C[0,\Delta_r],$$

where $\omega \in L_\infty[0,\Delta_r]$ and $\omega(\rho) \geq \omega_{\min} > 0$ for all $\rho \in [0,\Delta_r]$ and all $\Delta_r \in [0,\Delta_R]$, and that $u_{\Delta t}(\cdot;\Delta_r,f^\varepsilon)$ denotes the solution of (2.1) associated with η_{Δ_r} and perturbed data f^ε.

Then if $\Delta t = c\sqrt{\varepsilon}$, for any $c > 0$, convergence of approximations $u_{\Delta t}(t;\Delta_r,f^\varepsilon)$ to the true solution $\overline{u}(t)$ occurs at collocation points as $\varepsilon \to 0$, with best possible rate with regard to ε; i.e.,

$$|u_{\Delta t}(t_j;\Delta_r,f^\varepsilon) - \overline{u}(t_j)| \leq \varepsilon^{1/2} C(r) + \mathcal{O}(\varepsilon), \quad \text{for } j = 1, \ldots, N(\varepsilon),$$

as $\varepsilon \to 0$, where $C(r)$ is a positive constant and $N(\varepsilon) = 1/\Delta t(\varepsilon)$.

Proof. Subtracting (2.1), evaluated at t_j, $j = 1, \ldots, N$, from the same equation evaluated at t_{j+1} yields the differenced equation

$$\alpha(\Delta_r)\,\alpha_j + \int_{t_j}^{t_{j+1}} \tilde{k}(t_{j+1} - s; \Delta_r)\alpha_j \phi_j(s)\,ds \tag{2.2}$$
$$= \int_0^{\Delta_r} [f^{\varepsilon}(t_{j+1} + \rho) - f^{\varepsilon}(t_j + \rho)]\; d\eta_{\Delta_r}(\rho)$$
$$- \sum_{\mu=0}^{j-1} \int_{t_\mu}^{t_{\mu+1}} \Big(\tilde{k}(t_{j+1} - s; \Delta_r) - \tilde{k}(t_j - s; \Delta_r)\Big)\alpha_\mu \phi_\mu(s)\,ds + \alpha(\Delta_r)\,\alpha_{j-1},$$

valid for $j = 1, \ldots, N-1$. But for $\Delta_r \in (0, \Delta_R]$, the true solution $\overline{u}$ of equation (1.1) satisfies

$$\int_0^{\Delta_r} \left(\int_0^{t+\rho} k(t + \rho - s)\overline{u}(s)\,ds \right) d\eta_{\Delta_r}(\rho) = \int_0^{\Delta_r} f(t + \rho)\,d\eta_{\Delta_r}(\rho), \quad t \in (0, 1],$$

or, splitting the inner integral at t in the first term and switching the order of integration,

$$\int_0^t \tilde{k}(t - s; \Delta_r)\overline{u}(s)\,ds + \int_0^{\Delta_r} \left(\int_0^{\rho} k(\rho - s)\overline{u}(t + s)\,ds \right) d\eta_{\Delta_r}(\rho)$$
$$= \int_0^{\Delta_r} f(t + \rho)\,d\eta_{\Delta_r}(\rho), \quad t \in (0, 1]. \tag{2.3}$$

We evaluate (2.3) at $t = t_j$ for $j = 1, \ldots, N$, and subtract this from the same equation evaluated at t_{j+1}; subtracting the resulting differenced equation in $\overline{u}$ from (2.2), one obtains

$$\alpha(\Delta_r)[\overline{u}(t_{j+1}) - \alpha_j] + \int_{t_j}^{t_{j+1}} \tilde{k}(t_{j+1} - s; \Delta_r)[\overline{u}(s) - \alpha_j \phi_j(s)]\,ds$$
$$= \; - \int_0^{\Delta_r} [\delta(t_{j+1} + \rho) - \delta(t_j + \rho)]\; d\eta_{\Delta_r}(\rho)$$
$$- \sum_{\mu=0}^{j-1} \int_{t_\mu}^{t_{\mu+1}} \Big(\tilde{k}(t_{j+1} - s; \Delta_r) - \tilde{k}(t_j - s; \Delta_r)\Big) [\overline{u}(s) - \alpha_\mu \phi_\mu(s)]\,ds$$
$$+ \alpha(\Delta_r)\,[\overline{u}(t_j) - \alpha_{j-1}] \tag{2.4}$$
$$- \left\{ \int_0^{\Delta_r} \int_0^{\rho} k(\rho - s)[\overline{u}(t_{j+1}+s) - \overline{u}(t_{j+1})]\,ds\,d\eta_{\Delta_r}(\rho) \right.$$
$$\left. - \int_0^{\Delta_r} \int_0^{\rho} k(\rho - s)[\overline{u}(t_j+s) - \overline{u}(t_j)]\,ds\,d\eta_{\Delta_r}(\rho) \right\}.$$

Using a Taylor expansion, we write $\overline{u}(t) - \alpha_j \phi_j(t) = \Delta t \left[\beta_j + \dfrac{(t - t_{j+1})}{\Delta t}\,\overline{u}'(z_j(t)) \right]$ for $t \in (t_j, t_{j+1}]$, $j = 0, 1, \ldots, N-1$, where $\beta_j = \dfrac{\overline{u}(t_{j+1}) - \alpha_j}{\Delta t}$ and $z_j(t) \in (t, t_{j+1})$. Then (2.4) becomes

$$\beta_j \alpha(\Delta_r) + \int_{t_j}^{t_{j+1}} \tilde{k}(t_{j+1} - s; \Delta_r) \left[\beta_j + \frac{(s - t_{j+1})}{\Delta t}\,\overline{u}'(z_j(s)) \right] ds$$

$$
\begin{aligned}
&= -\frac{1}{\Delta t}\int_0^{\Delta_r} [\delta(t_{j+1}+\rho)-\delta(t_j+\rho)]\, d\eta_{\Delta_r}(\rho) \\
&\quad - \sum_{\mu=0}^{j-1}\int_{t_\mu}^{t_{\mu+1}} \left[\tilde{k}(t_{j+1}-s;\Delta_r)-\tilde{k}(t_j-s;\Delta_r)\right]\left[\beta_\mu + \frac{(s-t_{\mu+1})}{\Delta t}\,\overline{u}'(z_\mu(s))\right] ds \\
&\quad + \beta_{j-1}\alpha(\Delta_r) \qquad\qquad (2.5)\\
&\quad - \frac{1}{\Delta t}\left\{\int_0^{\Delta_r}\int_0^{\rho} k(\rho-s)[\overline{u}(t_{j+1}+s)-\overline{u}(t_{j+1})]\, ds\, d\eta_{\Delta_r}(\rho)\right.\\
&\qquad\qquad \left. - \int_0^{\Delta_r}\int_0^{\rho} k(\rho-s)[\overline{u}(t_j+s)-\overline{u}(t_j)]\, ds\, d\eta_{\Delta_r}(\rho)\right\}.
\end{aligned}
$$

But, making the change of variable $s \to s/\Delta t$ in $\alpha(\Delta_r) \equiv \int_0^{\Delta_r}\int_0^{\rho} k(\rho-s)\, ds\, d\eta_{\Delta_r}(\rho)$, and making the changes of variable in (2.5) of $s \to (s-t_j)/\Delta t$ and $s \to (s-t_\mu)/\Delta t$ (for appropriate j and μ), we have

$$
\begin{aligned}
\beta_j \;&=\; W(r,\Delta t)\beta_{j-1} - \Delta t\sum_{\mu=0}^{j-1} V_{j,\mu}(r,\Delta t)\beta_\mu \\
&\quad - \frac{1}{\Delta t^2}E_j(\varepsilon,r,\Delta t) - Z_j(r,\Delta t), \qquad j=1,\ldots,N-1,
\end{aligned} \tag{2.6}
$$

where, for appropriate ξ-variables,

$$
W(r,\Delta t) = \frac{\int_0^{\Delta_r}\int_0^{\rho/\Delta t} k(\rho-s\Delta t)\, ds\, d\eta_{\Delta_r}(\rho)}{d(r,\Delta t)}, \tag{2.7}
$$

$$
V_{j,\mu}(r,\Delta t) = \frac{\int_0^1\int_0^{\Delta_r} k'(\xi_{j,\mu,\Delta t}(s,\rho))\, d\eta_{\Delta_r}(\rho)\, ds}{d(r,\Delta t)}, \tag{2.8}
$$

$$
E_j(\varepsilon,r,\Delta t) = \frac{\int_0^{\Delta_r}[\delta(t_{j+1}+\rho)-\delta(t_j+\rho)]\, d\eta_{\Delta_r}(\rho)}{d(r,\Delta t)}, \tag{2.9}
$$

$$
Z_j(r,\Delta t) = \frac{r_j(r,\Delta t)}{d(r,\Delta t)}, \tag{2.10}
$$

for $j=1,\ldots,N-1,\ \mu=0,\ldots,j-1$, and

$$
\begin{aligned}
d(r,\Delta t) &= \int_0^{\Delta_r}\int_0^{\rho/\Delta t} k(\rho-s\Delta t)\, ds\, d\eta_{\Delta_r}(\rho) + \int_0^1\int_0^{\Delta_r} k((1-s)\Delta t+\rho)\, d\eta_{\Delta_r}(\rho)\, ds,\\
r_j(r,\Delta t) &= \int_0^{\Delta_r}\int_0^{\rho/\Delta t} k(\rho-s\Delta t)\overline{u}'(\xi_{j,\Delta t}(s))\, ds\, d\eta_{\Delta_r}(\rho)\\
&\quad - \int_0^{\Delta_r}\int_0^{\rho/\Delta t} k(\rho-s\Delta t)\overline{u}'(\xi_j(s))\, ds\, d\eta_{\Delta_r}(\rho)\\
&\quad + \Delta t\sum_{\mu=0}^{j-1}\int_0^1\int_0^{\Delta_r} k'(\xi_{j,\mu,\Delta t}(s,\rho))\, d\eta_{\Delta_r}(\rho)(s-1)\overline{u}'(z_\mu(t_\mu+s\Delta t))\, ds\\
&\quad + \int_0^1\int_0^{\Delta_r} k((1-s)\Delta t+\rho)\, d\eta_{\Delta_r}(\rho)(s-1)\overline{u}'(z_j(t_j+s\Delta t))\, ds.
\end{aligned}
$$

If we repeat the above steps for the collocation equation at t_1, a similar equation is obtained for β_0,

$$\beta_0 = -\frac{1}{\Delta t^2} E_0(\varepsilon, r, \Delta t) - Z_0(r, \Delta t), \tag{2.11}$$

where

$$E_0(\varepsilon, r, \Delta t) = \frac{\int_0^{\Delta_r} \delta(t_1+\rho)\, d\eta_{\Delta_r}(\rho)}{d(r,\Delta t)}, \tag{2.12}$$

$$Z_0(r, \Delta t) = \frac{r_0(r,\Delta t)}{d(r,\Delta t)}, \tag{2.13}$$

and where r_0 is defined for appropriate $\xi_{1,\Delta t}(\cdot)$ by

$$\begin{aligned} r_0(r,\Delta t) &= \int_0^1 \int_0^{\Delta_r} k((1-s)\Delta t + \rho)\, d\eta_{\Delta_r}(\rho)(s-1)\overline{u}'(z_0(s\Delta t))\, ds \\ &\quad + \int_0^{\Delta_r} \int_0^{\rho/\Delta t} k(\rho - s\Delta t)\overline{u}'(\xi_{1,\Delta t}(s))\, s\, ds\, d\eta_{\Delta_r}(\rho). \end{aligned}$$

We claim that, if certain uniform (in Δt) bounds can be obtained, namely,

$$|W(r,\Delta t)| \le w(r), \tag{2.14}$$

$$|V_{j,\mu}(r,\Delta t)| \le v(r), \quad j = 1,\dots,N-1;\ \mu = 0,\dots,j-1, \tag{2.15}$$

$$|E_j(\varepsilon, r,\Delta t)| \le e(r)\varepsilon, \quad j = 0,\dots,N-1, \tag{2.16}$$

$$|Z_j(r,\Delta t)| \le z(r), \quad j = 0,\dots,N-1, \tag{2.17}$$

where $v(r)$, $z(r)$, and $e(r)$ are nonnegative and $0 \le w(r) < 1$, then the results of the theorem are true. Indeed, an induction argument may be used in this case to argue that $|\beta_j| \le B_j$ for $j = 0,\dots,N-1$, where

$$\begin{aligned} B_0 &= \frac{1}{\Delta t^2} e(r)\varepsilon + z(r), \\ B_j &= w(r)B_{j-1} + \Delta t\, v(r) \sum_{\mu=0}^{j-1} B_\mu + \frac{1}{\Delta t^2} e(r)\varepsilon + z(r), \quad j = 1,\dots,N-1, \end{aligned} \tag{2.18}$$

and where each coefficient above is bounded in the limit as $\varepsilon \to 0$ provided $\Delta t = \Delta t(\varepsilon)$ is selected such that $\Delta t(\varepsilon) \to 0$ and $e(r)\varepsilon/\Delta t(\varepsilon)^2$ remains bounded as $\varepsilon \to 0$. Simplifying (2.18) one obtains a second order difference equation in B_j,

$$\begin{aligned} B_0 &= \frac{1}{\Delta t^2} e(r)\varepsilon + z(r), \\ B_1 &= (w(r) + \Delta t\, v(r))B_0 + \frac{1}{\Delta t^2} e(r)\varepsilon + z(r), \\ B_j &= (1 + w(r) + \Delta t\, v(r))B_{j-1} - w(r)B_{j-2}, \quad j = 2,\dots,N-1, \end{aligned}$$

so that the theory of difference equations may be used to obtain

$$|\beta_j| \le C_1(\varepsilon, r, \Delta t)\, [\tau_1(r,\Delta t)]^j \ + \ C_2(\varepsilon, r, \Delta t)\, [\tau_2(r,\Delta t)]^j,$$

for $j = 0, \ldots, N-1$, where

$$\begin{aligned} \tau_1(r,\Delta t) &= 1 + \Delta t \frac{v(r)}{1-w(r)} + \mathcal{O}(\Delta t^2), \\ \tau_2(r,\Delta t) &= w(r)\left(1 - \Delta t \frac{v(r)}{1-w(r)}\right) + \mathcal{O}(\Delta t^2), \\ C_1(\varepsilon, r, \Delta t) &= \frac{z(r) + e(r)\varepsilon/\Delta t^2}{1-w(r)} + \mathcal{O}(\Delta t), \\ C_2(\varepsilon, r, \Delta t) &= -w(r)\frac{z(r) + e(r)\varepsilon/\Delta t^2}{1-w(r)} + \mathcal{O}(\Delta t). \end{aligned} \tag{2.19}$$

Thus

$$|\beta_j| \le 2\, C_1(\varepsilon, r, \Delta t) \exp\left(\frac{2v(r)}{1-w(r)}\right), \quad j = 0, \ldots, N-1,$$

for Δt sufficiently small. Using the definition of β_j we thus have the error estimate

$$|\alpha_j - \overline{u}(t_{j+1})| \le 2\, \Delta t\, C_1(\varepsilon, r, \Delta t) \exp\left(\frac{2v(r)}{1-w(r)}\right), \quad j = 0, \ldots, N-1, \tag{2.20}$$

for Δt sufficiently small, where $\alpha_j = u_{\Delta t}(t_{j+1}; \Delta_r, f^\varepsilon)$. Thus, if we can show that the bounds in (2.14)–(2.17) hold for η_{Δ_r} given by the statement of the theorem, the convergence estimate (2.20) then holds when $\Delta t = c\sqrt{\varepsilon}$.

To verify (2.14)–(2.17), we substitute the representation of η_{Δ_r} (as a density) into (2.7) and make a change of integration variable to obtain

$$W(r,\Delta t) = \frac{\int_0^{r-1}\int_0^\rho k((\rho - s)\Delta t)\, ds\ \omega(\rho\Delta t)\, d\rho}{\int_0^{r-1}\int_0^\rho k((\rho - s)\Delta t)\, ds\ \omega(\rho\Delta t)\, d\rho + \int_0^1\int_0^{r-1} k((\rho + 1 - s)\Delta t)\omega(\rho\Delta t)\, d\rho\, ds}$$

so that $W(r, \Delta t) \le w(r)$ where

$$w(r) = \frac{\|k\|\, \|\omega\|(r-1)}{2k_{\min}\omega_{\min} + \|k\|\, \|\omega\|(r-1)}$$

(for simplicity in what follows, we will use $\|\cdot\| \equiv \|\cdot\|_\infty$ throughout). Making similar changes of integration variable in quantities in (2.8)–(2.10) and (2.12)–(2.13), we calculate

$$\begin{aligned} v(r) &= \frac{2\|k'\|\, \|\omega\|}{k_{\min}\omega_{\min}(r+1)}, \\ e(r) &= \frac{4\|\omega\|}{k_{\min}\omega_{\min}(r+1)}, \\ z(r) &= \frac{\|\overline{u}'\|\, \|\omega\|}{k_{\min}\omega_{\min}(r+1)}(3\|k\|(r-1)^2 + \|k'\|), \end{aligned}$$

where $k_{\min} \equiv \inf\{\, k(t) : t \in [0,T]\,\} > 0$. Since $0 < w(r) < 1$ for $r \ge 1$, we have shown that the needed bounds in (2.14)–(2.17) hold in the case of η_{Δ_r} a density, and thus the proof of the theorem is complete. $\square$

In the the proof of the last theorem we obtained sufficient conditions for convergence of the approximation method, and we state these here as a corollary.

COROLLARY 2.2. *Assume $\overline{u}$, k, and f^ε satisfy Hypothesis 2.1, for $\varepsilon > 0$. Let $\Delta_r = (r-1)\Delta t$ for some fixed integer $r \geq 2$ and assume that Δt-independent bounds exist for the quantities defined in (2.7)–(2.10), (2.12)–(2.13), namely*

$$\begin{aligned} |W(r,\Delta t)| &\leq w(r) \\ |V_{j,\mu}(r,\Delta t)| &\leq v(r), \quad j = 1,\ldots,N-1;\ \mu = 0,\ldots,j-1 \\ |E_j(\varepsilon,r,\Delta t)| &\leq e(r)\varepsilon, \quad j = 0,\ldots,N-1, \\ |Z_j(r,\Delta t)| &\leq z(r), \quad j = 0,\ldots,N-1, \end{aligned}$$

where $v(r)$, $z(r)$, and $e(r)$ are nonnegative and $0 \leq w(r) < 1$. Then, if $\Delta t = \Delta t(\varepsilon)$ is selected such that $\Delta t(\varepsilon) \to 0$ and such that $e(r)\varepsilon/\Delta t(\varepsilon)^2$ remains bounded as $\varepsilon \to 0$, then there exists $C_1 = C_1(\varepsilon, r, \Delta t)$ (given by (2.19)) such that the solution $u_{\Delta t}(\cdot;\Delta_r;f^\varepsilon)$ of the discretized future-sequential equations associated with perturbed data f^ε satisfies

$$|u_{\Delta t}(t_j;\Delta_r,f^\varepsilon) - \overline{u}(t_j)| \leq 2\,\Delta t(\varepsilon)\ C_1(\varepsilon,r,\Delta t(\varepsilon)) \exp\left(\frac{2v(r)}{1-w(r)}\right), \quad j = 1,\ldots,N,$$

for Δt sufficiently small. In particular, the choice $\Delta t = c\sqrt{\varepsilon}$ for any $c > 0$ yields

$$|u_{\Delta t}(t_j;\Delta_r,f^\varepsilon) - \overline{u}(t_j)| \leq \varepsilon^{1/2}C(r) + \mathcal{O}(\varepsilon), \quad \text{for } j = 1,\ldots,N(\varepsilon),$$

where $N(\varepsilon) = 1/\Delta t(\varepsilon)$ and $C(r)$ is a positive constant.

We turn now to a consideration of the discrete measure η_{Δ_r} given by (1.7), and additionally allow the possibility of $r = 1$ (defining in this case $K(1) = 1$ and $\eta_{\Delta_r} = s_1\,\delta_0$ for some $s_1 > 0$); thus $r = 1$ implies $\Delta_r = 0$ (no future interval) and that (1.4) in this case is equivalent to collocation of the original problem (1.1).

THEOREM 2.3. *Assume $\overline{u}$, k, and f^ε satisfy Hypothesis 2.1 for $\varepsilon > 0$, and let $\Delta_r = (r-1)\Delta t$, where $r \geq 1$ is a fixed integer and $\Delta t > 0$. Assume further that η_{Δ_r} is given by*

$$\eta_{\Delta_r} = \sum_{i=1}^{K(r)} s_i \delta_{\tau_i \Delta_r}$$

where $K(r)$ is an integer ($K(r) \geq 2$ for $r \geq 2$, $s_i = s_i(r) > 0$ and $0 \equiv \tau_1 < \tau_2 < \ldots < \tau_K \equiv 1$, $\tau_i = \tau_i(r)$, while for $r = 1$ we assume $K(1) = 1$, $s_1 > 0$, and $\tau_1 = 0$). Then if $\Delta t = c\sqrt{\varepsilon}$ for any constant $c > 0$, we have convergence of approximations $u_{\Delta t}(t;\Delta_r,f^\varepsilon)$ to the true solution $\overline{u}(t)$ at collocation points as $\varepsilon \to 0$, with best possible rate with respect to ε; i.e.,

$$|u_{\Delta t}(t_j;\Delta_r,f^\varepsilon) - \overline{u}(t_j)| \leq \varepsilon^{1/2}C(r) + \mathcal{O}(\varepsilon), \quad j = 1,2,\ldots,N(\varepsilon),$$

as $\varepsilon \to 0$, where $C(r)$ is a positive constant and $N(\varepsilon) = 1/\Delta t(\varepsilon)$.

Proof. Substitution of $\int_0^{\Delta_r} \phi(\rho)\,d\eta_{\Delta_r}(\rho) \equiv \sum_{i=1}^{K(r)} s_i\phi(\tau_i\Delta_r)$ (for ϕ continuous on $[0,\Delta_r]$) into (2.7)–(2.10) and (2.12)–(2.13) yields the needed bounds in (2.14)–(2.17), namely,

$$w(r) = \frac{\lambda(r)}{\lambda(r) + k_{\min}/\|k\|},$$

$$
\begin{aligned}
(2.21)\quad v(r) &= \sigma_0(r)\|k'\|/k_{\min}, \\
e(r) &= 2\sigma_0(r)/k_{\min}, \\
z(r) &= \|\overline{u}'\| \left(\frac{1}{2}(\|k\|+\|k'\|)\sigma_0(r) + 2\|k\|\sigma_1(r) + \|k\|\sigma_2(r)/2)\right)/k_{\min},
\end{aligned}
$$

and where the quantities $\lambda(r)$ and $\sigma_\ell(r)$ are defined by

$$
\lambda(r) = \frac{(r-1)\sum_{i=1}^{K(r)} s_i\tau_i}{\sum_{i=1}^{K(r)} s_i}, \qquad \sigma_\ell(r) = \frac{\sum_{i=1}^{K(r)} s_i[\tau_i(r-1)]^\ell}{\sum_{i=1}^{K(r)} s_i \,+\, (r-1)\sum_{i=1}^{K(r)} s_i\tau_i}.
$$

Then $\lambda \in [0, r-1]$ for all $r \geq 1$, $\lambda(1) = 0$, while $\sigma_0(r)$ satisfies $\sigma_0(r) = 1/(1+\lambda(r))$, $\sigma_0(r) \in [\frac{1}{r}, 1]$ for all $r \geq 1$, $\sigma_0(1) = 1$. Thus $0 \leq w(r) < 1$ for all $r \geq 1$, and $w(r) = 0$ at $r = 1$.

The selection of $\Delta t = c\sqrt{\varepsilon}$ for $c > 0$ guarantees that the conditions of Corollary 2.2 hold for $r \geq 2$. Further, using $r = 1$ we obtain $\eta_{\Delta_r} = s_1\delta_0$ so that proof of convergence in this case is merely a proof of convergence of collocation for the original first-kind problem (in the absence of future-sequential regularization). Repeating the arguments from the proof of Corollary 2.2 in the case of the original first-kind problem, we find that the bounds for the quantities in (2.7)–(2.10) and (2.12)–(2.13) are the same as those given by (2.21) using there the value of $r = 1$. Thus the results of the theorem are also true for the case of $r = 1$. □

Remark 2.1. We have found that collocation-based discretizations of (1.4), with typical choices of η_{Δ_r} converge at a "best possible" asymptotic rate $\mathcal{O}(C(r)\varepsilon^{1/2})$ as $\varepsilon \to 0$, provided that Δt is chosen appropriately, i.e., $\Delta t = c\,\varepsilon^{1/2}$, with $c > 0$ an arbitrary constant. The fact that the last theorem is valid in the case of $r = 1$ gives that standard collocation of equation (1.1) also yields an approximation which converges at the best possible asymptotic rate, for the same choice of Δt. However, as numerical examples in the next section indicate, the same choice of Δt which is suitable in practice for $r > 1$ is often associated with highly oscillatory, inaccurate approximations in the case of $r = 1$. Thus it is evident that the way in which the leading constant $C(r)$ depends on r in our convergence estimates is relevant. But $C(r)$ is only an upper bound in these estimates and, indeed, it is generally quite difficult to determine the way in which the actual convergence rate depends on r.

One relevant aspect of the question of dependence of approximations on r is addressed in [11], where an analysis of the condition number of the matrix $A_{r,\Delta t}$ (as it depends on r) is performed; here $A_{r,\Delta t}$ denotes the governing matrix in the linear system associated with (2.1), where η_{Δ_r} is the discrete measure used in the next section (we note that this particular choice of η_{Δ_r} is one of the most effective and simple in practice, and additionally satisfies the conditions of Theorem 2.3). The results in [11] apply to general ν-smoothing kernels k and show (under conditions on k and ν) that stability of the linear system sharply improves as r increases, for the same choice of Δt. Indeed, it is shown that

$$
\mathrm{cond}_\infty(A_{r,\Delta t}) < \mathrm{cond}_\infty(A_{1,\Delta t}), \quad \text{for } r = 2, 3, \ldots
$$

(where $\mathrm{cond}_\infty(A_{r,\Delta t})$ denotes the usual $\mathbb{R}^N$ infinity-norm condition number of $A_{r,\Delta t}$), and that

$$
\mathrm{cond}_\infty(A_{r,\Delta t}) \leq \eta(r)^{2\left(\frac{1}{\Delta t}-1\right)}\mathrm{cond}_\infty(A_{1,\Delta t}), \quad \text{for } r = 3, 4, \ldots,
$$

where $\eta(r) < 1$ is a decreasing function of r for $r = 3, 4, \ldots$. We note that the precise choice of r to use in practice, for a given value of ε, remains a difficult question.

3. A "Predictor-Corrector" Regularization Algorithm. Here we shall discuss implementation of the method described in §2, and illustrate the way in which a "predictor-corrector" type of numerical regularization scheme obtains.

For each example in this section, $\Delta_r = (r-1)\Delta t$ (for integer $r \geq 1$) and η_{Δ_r} denotes the discrete measure given by $\eta_{\Delta_r} = \sum_{i=1}^{r} s_i \delta_{\tau_i \Delta_r}$, where, for $i = 1, \dots, r$,

$$s_i = \frac{\int_0^{i\Delta t} k(i\Delta t - s)\, ds}{\int_0^{\Delta t} k(\Delta t - s)\, ds},$$

and $\tau_i = \dfrac{(i-1)}{(r-1)}$ in the case of $r \geq 2$; $\tau_1 = 0$ in the (unregularized) case of $r = 1$. For this choice of η_{Δ_r}, it can be shown that the collocation-based approximation described in the last section is equivalent to the following "predictor-corrector" regularization algorithm: That is, let $u_{\Delta t}(t; \Delta_r; f^\varepsilon) = \sum_{j=0}^{N-1} \alpha_j \phi_j(t)$, as before, and solve sequentially for α_0, $\alpha_1, \dots$, such that each α_j is the optimal value one would use if forced to retain α_j as the value of the *present* basis coefficient, as well as of $r-1$ *future* coefficients while performing data-fitting to 1 present and $r-1$ future data points. For example, in the first (**"predictor"**) step we select α_0 minimizing J_0, the least-squares criterion associated with discrete data points on the interval $[0, t_r]$, i.e.,

$$\begin{aligned} J_0(\alpha_0) \equiv & \left| \int_0^{t_1} k(t_1 - s)\alpha_0 \phi_0(s)\, ds - f(t_1) \right|^2 \\ & + \left| \int_0^{t_2} k(t_2 - s)\alpha_0 \left(\phi_0(s) + \phi_1(s) \right) ds - f(t_2) \right|^2 + \dots \\ & + \left| \int_0^{t_r} k(t_r - s)\alpha_0 \left(\phi_0(s) + \dots + \phi_{r-1}(s) \right) ds - f(t_r) \right|^2. \end{aligned}$$

After computing the optimal α_0, we hold this value fixed on $[0, t_1]$ but (now, in the **"corrector"** step) we do not retain the value of α_0 on the remaining subinterval $(t_1, t_r]$; instead we continue to the second step of the process where α_1 is selected in a similar manner. That is, we now choose α_1 minimizing J_1, the least squares functional associated with the interval $[t_1, t_{r+1}]$, i.e.,

$$\begin{aligned} J_1(\alpha_1) \equiv & \left| \int_0^{t_2} k(t_2 - s) \left[\alpha_0 \phi_0(s) + \alpha_1 \phi_1(s) \right] ds - f(t_2) \right|^2 \\ & + \left| \int_0^{t_3} k(t_3 - s) \left[\alpha_0 \phi_0(s) + \alpha_1 \left(\phi_1(s) + \phi_2(s) \right) \right] ds - f(t_3) \right|^2 + \dots \\ & + \left| \int_0^{t_{r+1}} k(t_{r+1} - s) \left[\alpha_0 \phi_0(s) + \alpha_1 \left(\phi_1(s) + \dots + \phi_r(s) \right) \right] ds - f(t_{r+1}) \right|^2, \end{aligned}$$

and we then retain this optimal value on the interval $[t_1, t_2]$ only. And so on, until all α_j are selected. This is exactly the method used with much success by Beck for the IHCP. As can be seen, the coordination of the length Δ_r of the future interval with the stepsize Δt (through $\Delta_r = (r-1)\Delta t$) leads to a very simple routine. In addition, the local regularization is apparent in the process, accomplished via the temporary "thickening" of the basis element $\phi_j(\cdot)$ in the determination of α_j.

Making other choices of η_{Δ_r} leads to other predictor-corrector regularization methods for Volterra equations, each of which is based on predicting a new regularized value on a small (future) time interval, and then correcting the prediction via a shortening of the interval over which the new value is to be assumed. One such variation of this approach is a sequential Tikhonov regularization method, where a Tikhonov problem is solved on each subinterval of length Δ_r, retaining the optimal value only on the initial subinterval of length Δt; the details of this method may be found in [12].

For the following test examples, the kernel k and true solution $\overline{u}$ were selected *a priori*, and the unperturbed data function f was then constructed from the (exact) integration of $f(t) = \int_0^t k(t-s)\overline{u}(s)\,ds$; varying amounts of random noise were then added to f, with the noise uniformly distributed in the interval $[-s\|f\|_\infty, s\|f\|_\infty]$ for some positive number s (representing the amount of relative error in the resulting perturbed data). A noise level of roughly 1% is evidently quite common for some applications, including the IHCP; thus the noise levels of 5% used below are extreme for such applications and useful in verifying the effectiveness of the regularization method. For more reasonable levels of 1% or so, one can expect to use a much smaller regularization parameter r than is exhibited below.

For the first two examples, the kernel k satisfies the conditions of the last section (i.e., k is 1-smoothing), while the third example shows that the method applies to more general k; the last example is for $\overline{u}$ more general than that needed for the theory of the last section, i.e., for $\overline{u}$ piecewise continuous. In all cases, $N = 20$.

Example 3.1. For this example, we use $k(t) = 1+t^2$ and $\overline{u}(t) = t^2 + \cos(t)$. Random noise is added to f at the level of 5% relative error ($s = .05$). Our findings are shown in Figure 3.1, with standard collocation on equation (1.1) given by the results for $r = 1$, and regularized approximation given for $r = 2, \ldots, 8$.

Example 3.2. Here we use $k(t) = 1 + 2t^4$ and the more oscillatory $\overline{u}(t) = 11 + 5\sin(8t)$. Random noise is added to f at the level of 3% relative error, and the results given in Figure 3.2. We note that this example clearly shows that r too large leads to "over-regularization" of the solution via excessive smoothing. This phenomenon is a common feature in all regularization methods, including Tikhonov regularization.

Example 3.3. For this example, we use a kernel which does not satisfy the assumptions of the last section, namely, $k(t) = t^5 + t^2$ (such a kernel is 3-smoothing in the language of §1, meaning that the resulting integral equation is more ill-posed than the two previous examples). We let $\overline{u}(t) = \cos(4t)$ and add random noise to f at the level of both 5% and 1%. Our findings are shown in Figures 3.3 and 3.4; it is not surprising that, for this example, the additional ill-posedness of (1.1) in this case requires a larger value of r (as high as $r = 14$) in order to effectively implement regularization in the presence of 5% noise.

Example 3.4. Finally, we consider a situation in which $\overline{u}$ is piecewise continuous and thus does not satisfy the assumptions of the theory in the last section; we show in Figure 3.5 our numerical findings for the example of

$$\overline{u} = \begin{cases} 1 + 2.5\,t, & 0 \le t < .2, \\ 2 - 2.5\,t, & .2 \le t < .4, \\ t - .5, & t \ge .4, \end{cases}$$

and $k(t) \equiv 1$ (the results are similar in our tests of other 1-smoothing k). The first

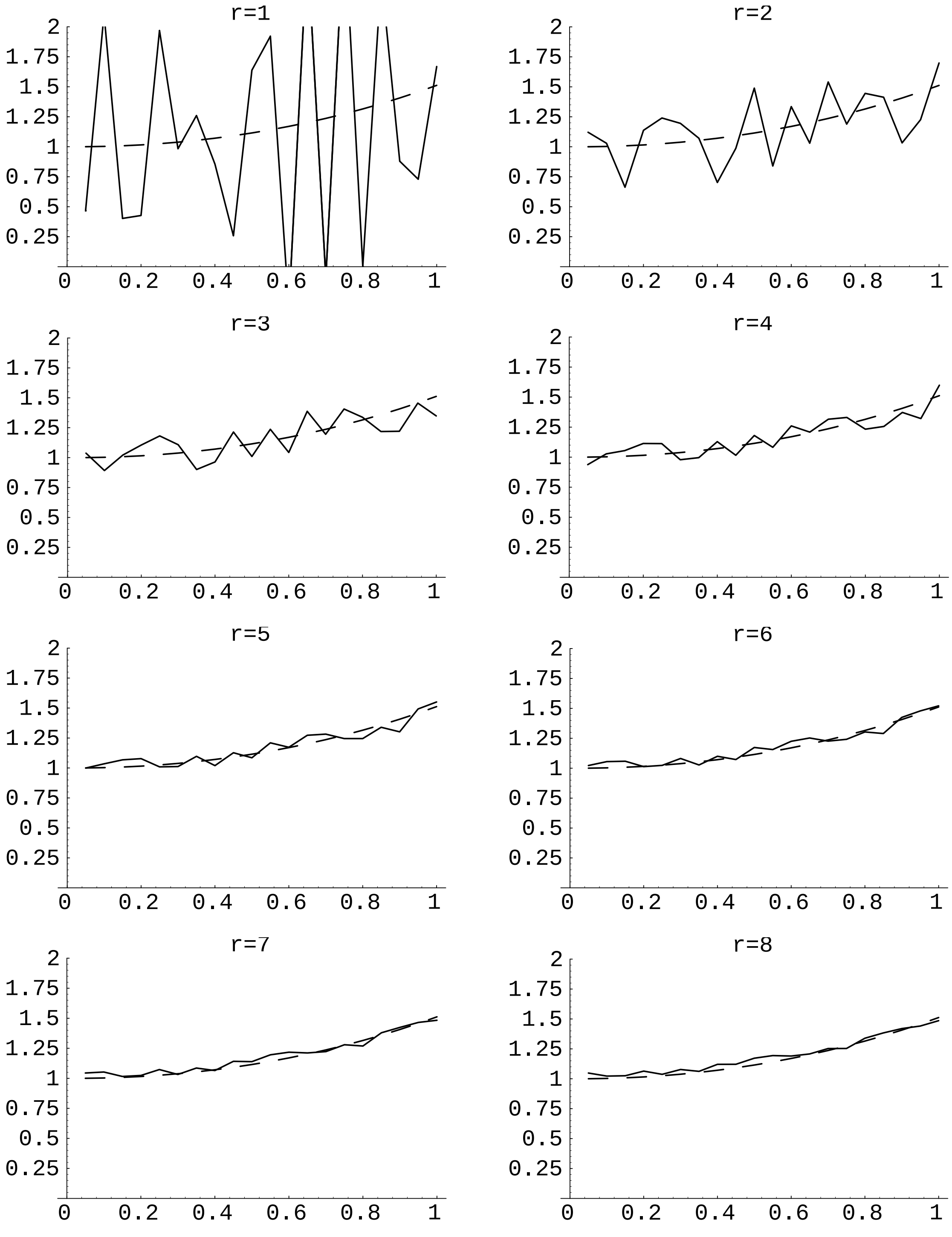

FIG. 3.1. *Results from Example 3.1.*

column of graphs in that figure represents the results for a 5% level of perturbation in the data and $r = 1, 2, 3$, while in the second column we display our results for the same values of r but with 1% level of noise in the data. We note that our usual manner of displaying output in previous figures (i.e., using piecewise linear curves for approximations) suggests continuity that should not be assumed for this problem; on the other hand, replacing piecewise-linear curves by discrete points (plotted at collocation locations) gives way to graphs that are very difficult to read, especially in the highly oscillatory $r = 1$ case. We have therefore chosen to display the results in Figure 3.5 via piecewise-linear curves as before, but with an artificially-generated break in the approximate solution curve at the true location of the discontinuity ($t = .4$). This break is for display purposes only and is not intended to imply that the numerical method introduced in this paper is able to resolve discontinuities. Indeed, like many regularization methods, this method also fails to accurately predict *locations* of discontinuity; however, as can be seen from the graphs in Figure 3.5, the method does apparently still improve upon collocation alone in the determination of the approximate shape of a desired solution.

4. Conclusion. To summarize the major results of this paper, we have examined the convergence properties of a collocation-based regularization method for the stable numerical solution of first-kind Volterra equations of convolution type. For standard implementations of the method, we find that convergence obtains at a "best possible" rate with respect to error in the data when the method is applied to 1-smoothing Volterra problems . In addition, our numerical investigations indicate how regularization via this approach allows for the use of much smaller approximation stepsize than may be used for the stable approximate solution of the original equation, and how a "predictor-corrector" form of numerical regularization obtains.

In related work [11], we have examined the conditioning of the linear system associated with the discretization in Section 4, showing that stability of the system improves as the value of the regularization parameter (i.e., the length of the "future interval" Δ_r, or the size of r) increases. Work is also in progress on extensions of these general regularization ideas to non-Volterra problems, allowing for a local regularization method for general integral equations of the first kind. In addition, in joint work with L. Eldén [12], we consider a sequential Tikhonov regularization method that is also of "predictor-corrector" type and in which Tikhonov regularization is performed sequentially over small future intervals of length Δ_r. This work is considered in a more abstract setting in [9].

Acknowledgments. A special case of one of the results in this paper (i.e., Theorem 2.3, for a particular choice of discrete measure η_{Δ_r}) was jointly worked out with K. A. Murphy, Department of Mathematics, University of North Carolina at Chapel Hill.

The author would also like to thank the referees for comments and suggestions which improved both the content and exposition of this paper.

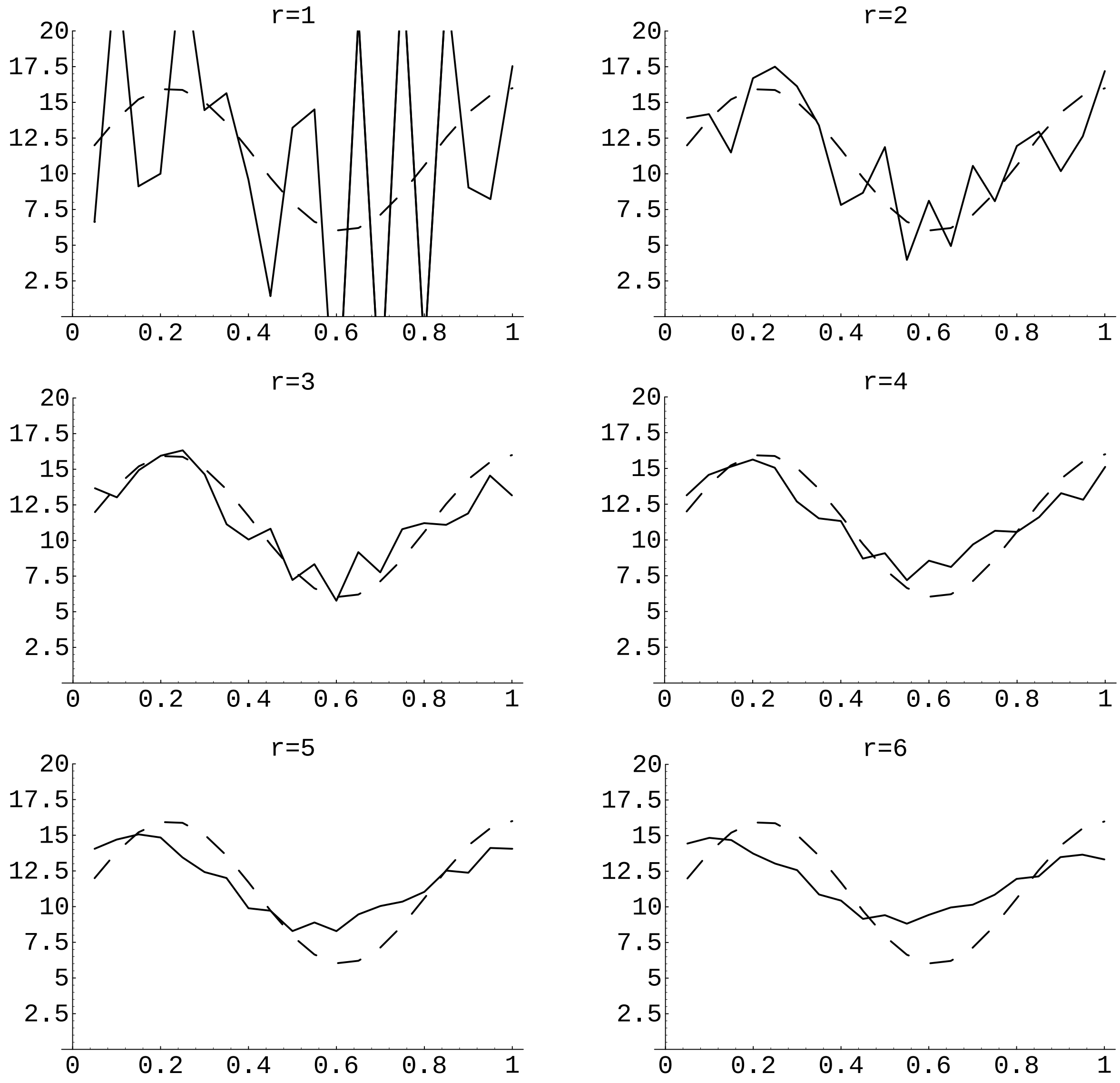

FIG. 3.2. *Findings for Example 3.2.*

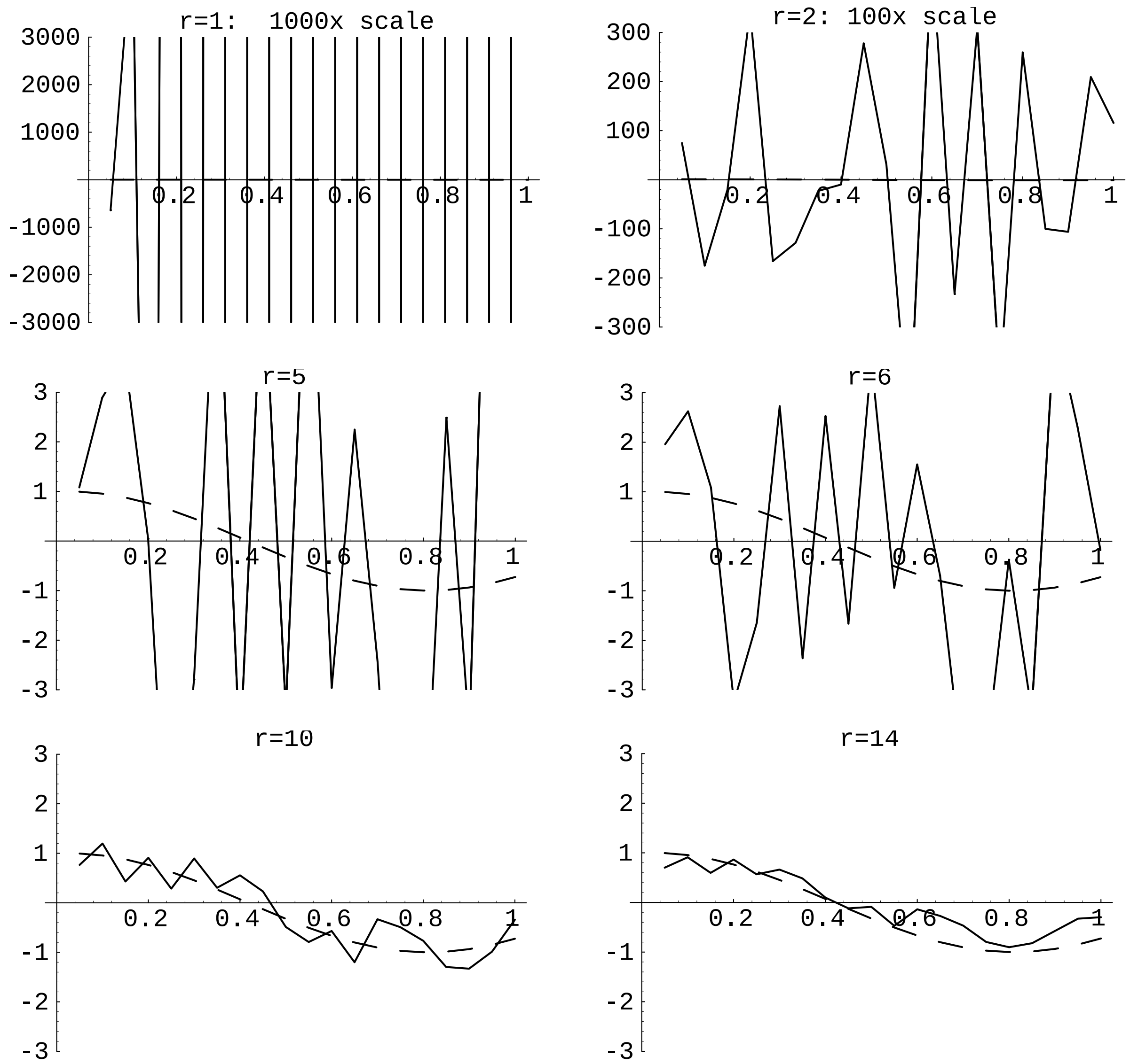

FIG. 3.3. *Results from Example 3.3 with 5% Noise.*

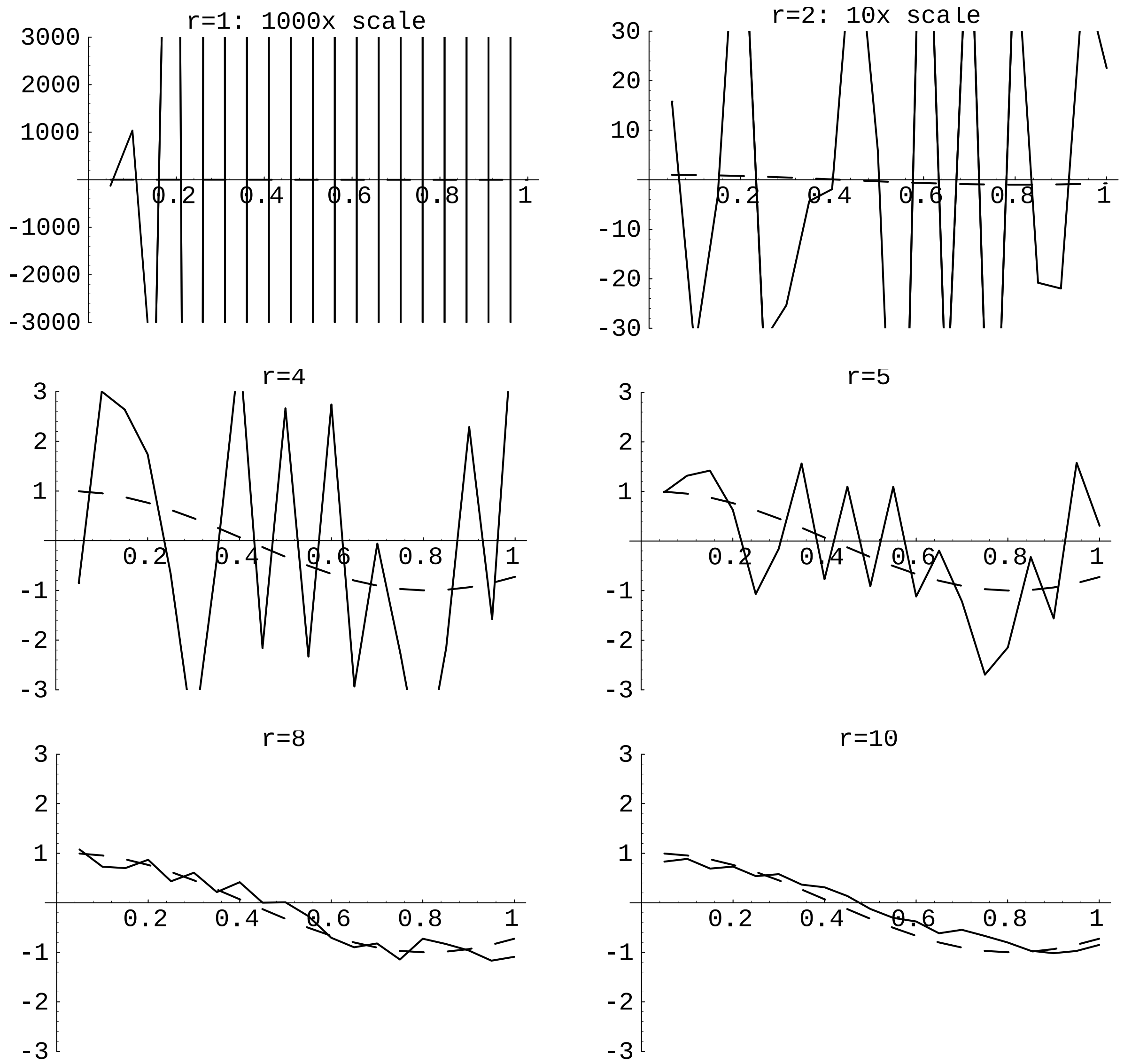

FIG. 3.4. *Results from Example 3.3 with 1% Noise.*

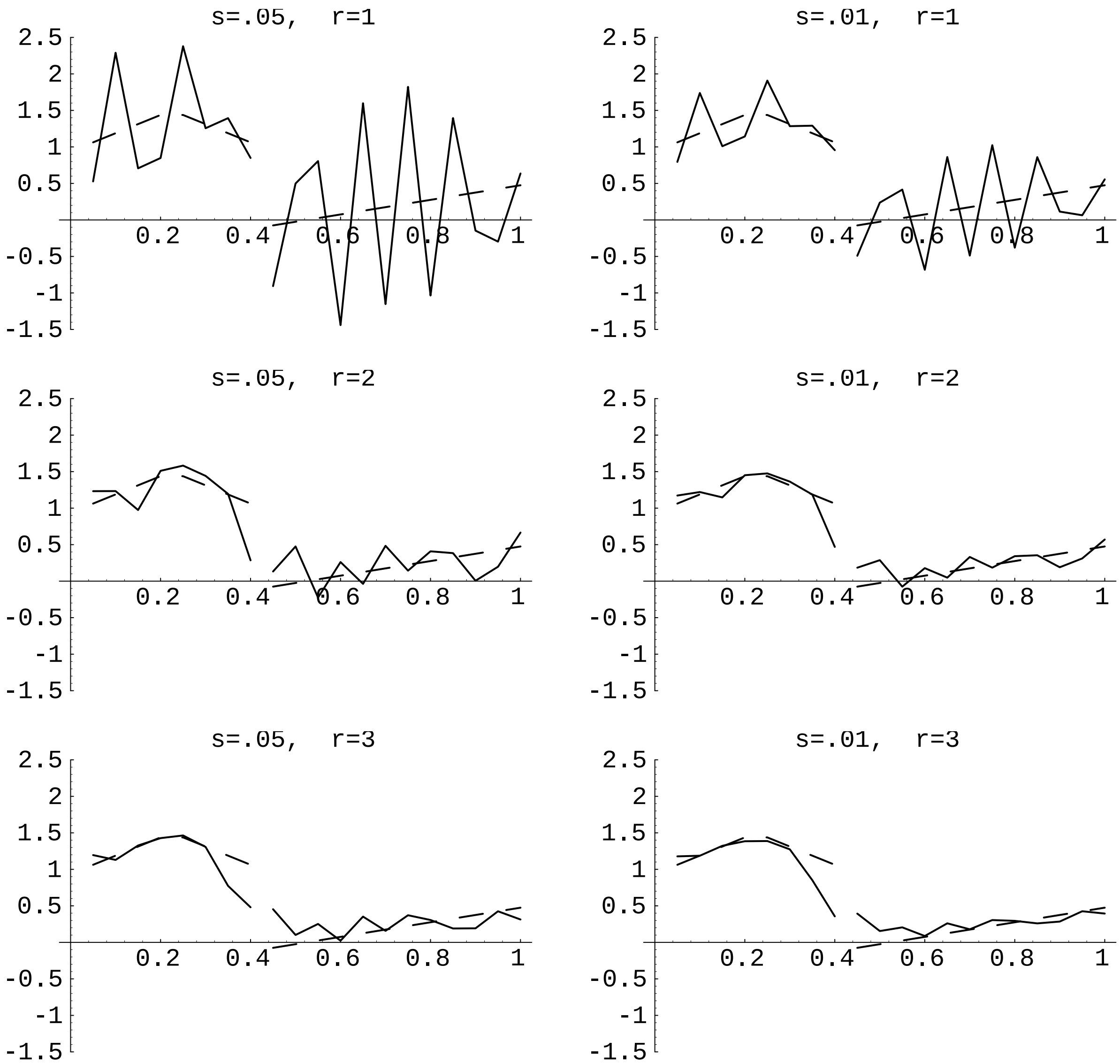

FIG. 4.1. *Results from Example 3.4 with 5% Noise (graphs on the left above) and with 1% Noise (graphs on the right above).*

REFERENCES

[1] J. V. Beck, B. Blackwell, and J. C. R. St. Clair, *Inverse Heat Conduction*, Wiley-Interscience, 1985.

[2] H. Brunner, *Discretization of Volterra integral equations of the first kind*, Numer. Math., 30 (1978), pp. 117–136.

[3] T. A. Burton, *Volterra Integral and Differential Equations*, Academic Press, 1983.

[4] F. R. de Hoog, *Review of Fredholm equations of the first kind*, in The application and numerical solution of integral equations, R. S. A. *et al*, ed., Sijthoof & Noordhoff, The Netherlands, 1980, pp. 119–134.

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

[6] G. Gripenberg, S. O. Londen, and O. Staffens, *Volterra Integral and Functional Equations*, Cambridge University Press, Cambridge, 1990.

[7] C. W. Groetsch, *The Theory of Tikhonov regularization for Fredholm equations of the first kind*, Pitman, Boston, 1984.

[8] P. K. Lamm, *Future-sequential regularization methods for ill-posed Volterra equations*, J. Math. Analysis Appl., to appear.

[9] ———, *On the local regularization of inverse problems of Volterra type*. To appear in Special Symposium on Parameter Identification, *Conference Proceedings, 15th ASME Biennial Conference*, Boston, September 1995.

[10] ———, *Regularized inversion of finitely smoothing Volterra operators*. Submitted, 1993.

[11] ———, *On the condition number associated with numerical solution of locally-regularized first-kind integral equations*. Submitted, 1994.

[12] P. K. Lamm and L. Eldén, *Numerical solution of first-kind Volterra equations by sequential Tikhonov regularization*, SIAM J. Numerical Analysis, to appear.

[13] 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, 1983, pp. 189–203.

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

[15] G. M. Wing, *A Primer of Integral Equations of the First Kind*, SIAM, 1991.