International Journal of Pure and Applied Mathematics
Volume 58 No. 4 2010, 473-492

NAIVE VS. REGULARIZED. NUMERICAL STUDY

Alexandra Smirnova[1] §, MaryGeorge L. Whitney[2]
[1,2]Department of Mathematics and Statistics
Georgia State University
Atlanta, GA 30303, USA
e-mail: asmirnova@gsu.edu
e-mail: mwhitney1@gsu.edu

Abstract: In computational mathematics and applications, there are numerous examples of problems that are unstable with respect to noise in the input data. Classical numerical algorithms, when used for such problems, turn out to be divergent. Hence, in order to solve the problem in a stable fashion, one has to combine a numerical method with special regularization techniques that would take advantage of an *a priori* information available in each particular case. In this paper, numerical analysis of Tikhonov's (variational) regularization for a first kind integral equation is given. The regularization parameter is computed by the discrepancy principle of Morozov.

AMS Subject Classification: 47A52, 65F22
Key Words: Tikhonov regularization, Morozov discrepancy principle, ill-posed problems, integral equations

1. Introduction

Hadamard's [6] initial concept of a well posed problem reflected the idea that any mathematical model of a physical phenomena must have the properties of uniqueness, existence, and stability of the solution. If at least one of these properties is not satisfied, the problem is ill-posed (unstable). Among classi-

Received: January 15, 2010
§Correspondence author

cal ill-posed problems are stable numerical differentiation of noisy data, stable inversion of ill-conditioned matrices, parameter identification in partial differential equations, stable solution of first-kind integral equations, and many other examples.

Let X and Y be normed spaces and the problem

$$Kx = f,\ K : X \to Y \tag{1.1}$$

be ill-posed. Suppose the right-hand side f is given by its δ-approximation f_δ such that $\|f - f_\delta\| \le \delta$. It is tempting to seek an approximate solution to (1.1) in the set $Q_\delta := \{x \in X : \|Kx - f_\delta\| \le \delta\}$. However, in the ill-posed case, this is not a good idea. Since x_δ does not depend continuously on f_δ, the fact that $\|Kx - f_\delta\| \le \delta$ does not guarantee that x_δ is close to the solution we are looking for.

Indeed, assume for example, that $K : \mathbb{R}^n \to \mathbb{R}^n$ is a nonsingular $n \times n$ matrix. One has

$$||x - x_\delta|| \le ||K^{-1}||\,||f - f_\delta|| \quad \text{and} \quad \frac{1}{\|x\|} \le \frac{\|K\|}{\|f\|}.$$

Therefore

$$\begin{aligned} \frac{\|x - x_\delta\|}{\|x\|} &\le \frac{\|K^{-1}\|\,\|f - f_\delta\|}{\|x\|} = \|K^{-1}\|\,\frac{\|f - f_\delta\|}{\|f\|}\,\|K\| \\ &\le \underbrace{\|K\|\,\|K^{-1}\|}_{\text{cond}(K)}\,\frac{\|f - f_\delta\|}{\|f\|}. \end{aligned}$$

One can see from above that a large condition number can result in a considerable change in the solution even if the relative error in the right-hand side is small (unstable problem). In fact, cond(K) measures how close K is to a singular matrix. The following theorem helps to explain.

Theorem 1.1. (see [4]) *Let $K \in \mathbb{R}^{n\times n}$ be nonsingular. Then for any singular matrix $B \in \mathbb{R}^{n\times n}$, $\frac{1}{\text{cond(K)}} \le \frac{\|K-B\|}{\|K\|}$.*

So, as K gets close to singular, cond(K) approaches infinity. In the two dimensional case represented by a 2×2 system

$$\begin{bmatrix} \vec{k_1} & \vec{k_2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{f} \ \Leftrightarrow\ \vec{k_1}x_1 + \vec{k_2}x_2 = \vec{f} \Leftrightarrow \vec{k_1}x_1 = \vec{f} - \vec{k_2}x_2$$

the solution is equivalent to finding the intersection of the one dimensional subspace $\{x\vec{k_1} : x \in \mathbb{R}\}$ with the affine subspace $\{\vec{f} - \vec{k_2}x : x \in \mathbb{R}\}$. Such intersections are shown in Figure 1. If the subspaces are nearly parallel then the condition number is large since the matrix is almost singular, and a slight

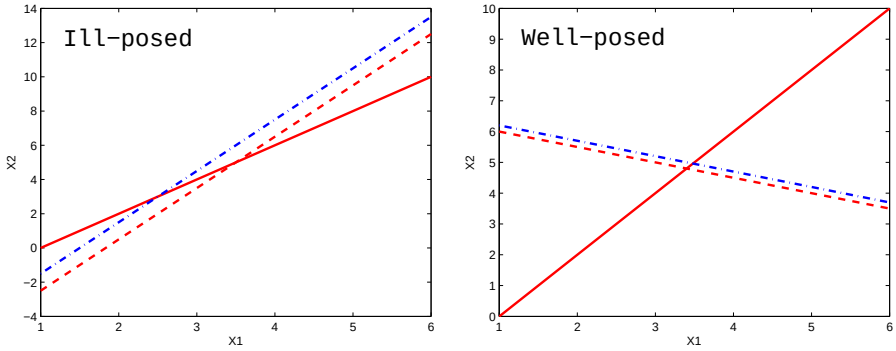

Figure 1: Ill-posed and well-posed linear systems

perturbation in the system may result in a considerable change of the solution. On the other hand, if the subspaces are almost perpendicular, a small perturbation or slight shift means one still gets a good solution. *Thus an arbitrary element $x_\delta \in Q_\delta$ cannot be used as an approximate solution to (1.1)*

Now let us turn our attention to linear equations (1.1) with compact operators K acting from infinite dimensional spaces.

Theorem 1.2. (see [9]) *Linear equations $Kx = f$ with compact operators $K : X \to Y$, where X and Y are normed spaces and $\dim X = \infty$ are always ill-posed.*

Proof. To show that $Kx = f$ is ill-posed, we will prove that K^{-1} is unbounded. Assume the converse: K^{-1} exists and is bounded, then $K^{-1}K = I$ should be compact since a superposition of a bounded and compact operators is a compact operator. Contradiction, since a unit ball in infinite dimensional space is not a compact set. □

The first kind integral equation is a classical example of an ill-posed problem with a compact operator:

$$Kx := \int_a^b k(t,s)x(s)ds = f(t), \quad K : X \to Y, \quad t \in (c,d). \tag{1.2}$$

This is the consequence of the following theorem:

Theorem 1.3. (see [9])

(a) Let $k(t,s) \in L^2\Big((c,d) \times (a,b)\Big)$. The operator $K : L^2(a,b) \to L^2(c,d)$,

	$\|x(t_i) - x_i\|/\|x(t_i)\|$				
t	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$x(t_i) = \exp(\frac{-t_i}{9})\cos(\frac{2\pi t_i}{3})$
1.00	4.45	3.19	2.34×10^4	7.57×10^6	-0.45
3.25	0.23	1.06	3.47×10^5	2.33×10^6	0.60
5.50	0.73	4.92	9.19×10^5	1.87×10^7	0.27
7.75	0.96	3.03	1.60×10^4	1.05×10^6	-0.37
10.00	1.48	4.70	3.41	4.80×10^3	-0.16

Table 1: Comparative results

defined by

$$Kx := \int_a^b k(t,s)x(s)ds, \quad t \in (c,d), \quad x \in L^2(a,b),$$

is compact from $L^2(a,b)$ *into* $L^2(c,d)$.

(b) Let $k(t,s)$ *be continuous on* $[c,d]\times[a,b]$. *Then K defined by the integral above is also compact as an operator from* $C[a,b]$ *into* $C[c,d]$.

The 'naive' way to solve (1.2) numerically is to discretize the integral by using some quadrature formula and to find x from the resulting linear system. Our next example illustrates what can happen if we actually implement this method.

2. 'Naive' Discretization

Consider the integral equation

$$Kx := \int_1^{10} \frac{s}{25}\sin(st)x(s)ds = f(t), \quad K : L^2(1,10) \to L^2(1,10),$$

where $1 \le t \le 10$. With the help of *Symbolic Toolbox* in *MatLab*, we use the following model solution to generate the exact right-hand side $f(t)$ of the equation:

$$x(s) = \exp\frac{-s}{9}\cos\left(\frac{2\pi s}{3}\right).$$

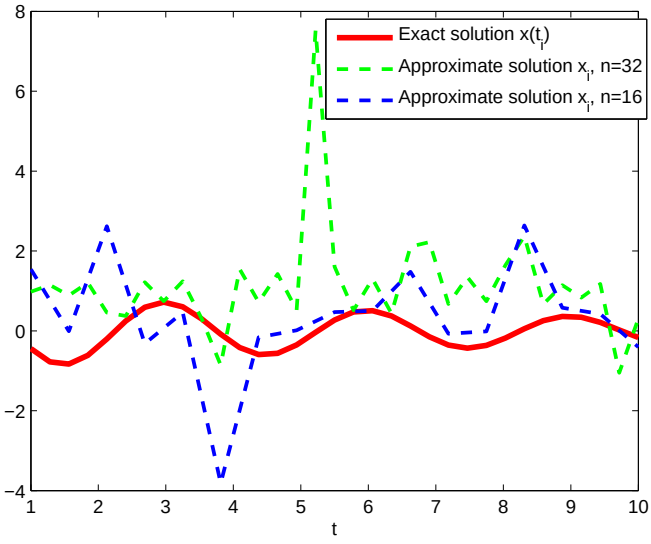

Figure 2: Exact and approximate solutions

The integral is approximated by the trapezoidal rule

$$(K_n x)_i; = \frac{h}{25}\left[\frac{1}{2}\sin t_i x_0 + \frac{1}{2}\Big(10\sin(10t_i)\Big)x_n + \sum_{j=1}^{n-1}(1+jh)\sin((1+jh)t_i)x_j\right],$$

$i = 0, 1, \ldots n$, where $h = \frac{9}{n}$ and $t_i = 1+ih$. For numerical simulations conducted in *MatLab*, the vector $[x_0\ x_1\ \ldots x_n]^T$ represents a computed solution given the discretized right-hand side $[f_0\ f_1\ \ldots f_n]^T$. We take $n = 16, 32, 64$, and 128 as the number of partitions for both s and t to compare approximate solutions corresponding to different values of n. Table 1 contains the relative errors $|x(t_i) - x_i|/|x(t_i)|$ and the corresponding values of the exact solution $x(t_i)$ for five different test points. In Figure 2, one can see the exact solution along with the two approximate solutions for $n = 16$ and $n = 32$. Approximate solutions for higher values of n are so inaccurate that they cannot be displayed in the picture. Once again, the inequality $||Kx_\delta - f_\delta|| \leq \delta$ does not imply that $||x - x_\delta|| = O(\delta)$.

Notice, that 'naive' algorithm does not work even though our example is noise-free ($f_\delta = f$) and the only error is due to numerical method:

$$||Kx_\delta - f_\delta|| \leq ||Kx_\delta - K_n x_\delta|| + ||K_n x_\delta - f|| = O(h^2) + O(10^{-16}) := \delta_n.$$

As the number of partitions n increases, the error δ_n gets smaller, but the matrix K_n becomes more and more ill-conditioned. *As the result, the finer discretization, the worse the computed solution.*

3. The Tikhonov (Variational) Regularization

When the right-hand side of (1.1) is known approximately, the perturbed equation $Kx = f_\delta$ may not be solvable. A common approach in that case is to solve the problem in the sense of least squares, i.e., to minimize $||Kx - f_\delta||^2$ with respect to $x \in X$. Since the minimizer is a solution to the corresponding normal equation, the above minimization problem is also ill-posed. Hence one should penalize either the functional $||Kx - f_\delta||^2$ or the equation $K^*Kx = K^*f_\delta$ in such a way that the generating operator is no longer compact. Both ideas lead to the following regularized optimization problem:

$$\min_{x\in X} J_\alpha(x) := \min\{\|Kx - f\|^2 + \alpha\|x\|^2 : \ x \in X\},$$

where J_α is called the *Tikhonov functional*.

Theorem 3.1. (see [9]) *Let $K : X \to Y$ be a linear bounded operator between Hilbert spaces and $\alpha > 0$. Then J_α has a unique minimum $x_\alpha \in X$. This minimum is the unique solution to the normal equation*

$$\alpha x_\alpha + K^*Kx_\alpha = K^*f. \tag{3.1}$$

From (3.1), it follows that $x_\alpha = (\alpha I + K^*K)^{-1}K^*f := R_\alpha f,\ R_\alpha : Y \to X$, where

$$R_\alpha := (\alpha I + K^*K)^{-1}K^*.$$

Suppose now that f is given by its $\delta -$ approximation, i.e., $\|f_\delta - f\| \le \delta$ and one solves the equation $Kx = f_\delta$. Then one gets

$$x_{\alpha,\delta} = (\alpha I + K^*K)^{-1}K^*f_\delta := R_\alpha f_\delta,$$

and

$$\begin{aligned} ||x_{\alpha,\delta} - x|| &= ||R_\alpha f_\delta - R_\alpha f + R_\alpha f - x|| \le ||R_\alpha||\,||f - f_\delta|| + ||R_\alpha f - x|| \\ &\le ||R_\alpha||\delta + ||R_\alpha Kx - x||. \end{aligned}$$

By polar decomposition $K = U(K^*K)^{1/2}$, where U is a partial isometry:

$$\|Uq\| = \|q\| \text{ for all } \ q \in N(U)^\perp, \quad N(U) = \{q : Uq = 0\}.$$

Therefore,

$$\|R_\alpha\| = \|(\alpha I + K^*K)^{-1}(K^*K)^{1/2}U\| \le \|(\alpha I + K^*K)^{-1}(K^*K)^{1/2}\| := \|\psi(A)\|$$

with $A := K^*K$. Since A is self-adjoint, by spectral theorem one derives

$$\|\psi(A)\| = \sup_{\lambda\in\sigma(A)} |f(\lambda)| = \sup_{\lambda\in\sigma(A)} \frac{\sqrt{\lambda}}{\lambda+\alpha} = \frac{1}{2\sqrt{\alpha}}.$$

Here $\sigma(A) = [0, \|K\|^2]$. Thus, $\delta\|R_\alpha\| \le \frac{\delta}{2\sqrt{\alpha}}$.

In order to analyze $\|R_\alpha Kx - x\|$, we make an additional assumption:

$$x = K^*z \in K^*(Y),\ z \in Y.$$

Then

$$\begin{aligned}
\|R_\alpha Kx - x\| &= \|[K^*K + \alpha I]^{-1}K^*Kx - x\| \\
&= \|[K^*K + \alpha I]^{-1}K^*Kx - [K^*K + \alpha I]^{-1}[K^*K + \alpha I]x\| \\
&= \|[K^*K + \alpha I]^{-1}(K^*Kx - K^*Kx - \alpha x)\| \\
&= \|[K^*K + \alpha I]^{-1}(-\alpha I x)\| = \|[K^*K + \alpha I]^{-1}(-\alpha x)\| \\
&= \alpha\|[K^*K + \alpha I]^{-1}K^*z\| \le \alpha\|[K^*K + \alpha I]^{-1}K^*\|\,\|z\| \\
&\le \frac{\alpha\|z\|}{2\sqrt{\alpha}} = \frac{\|z\|\sqrt{\alpha}}{2}.
\end{aligned}$$

This inequality proves that the operators $R_\alpha : Y \to X,\ R_\alpha = (\alpha I + K^*K)^{-1}K^*$ form a regularization strategy with $\lim_{\alpha\to 0}\|R_\alpha Kx - x\| \le \lim_{\alpha\to 0}\frac{\|z\|\sqrt{\alpha}}{2} = 0$. It is called the *Tikhonov regularization* method [15], [16]. Combining the estimates for $\delta\|R_\alpha\|$ and $\|R_\alpha Kx - x\|$, one concludes

$$\|x_{\alpha,\delta} - x\| \le \frac{\delta}{2\sqrt{\alpha}} + \frac{\|z\|\sqrt{\alpha}}{2} := E(\alpha).$$

Formally, one can now minimize $E(\alpha)$ in order to find the optimal value of the regularization parameter. However, $\|z\|$ is unknown. In the next section we discuss a posteriori choice of α by the so called discrepancy principle, which does not use $\|z\|$.

4. The Discrepancy Principle of Morozov

In this section, we consider the determination of $\alpha(\delta)$ from the discrepancy principle (DP) [10], [11], [12] of Morozov. The discrepancy principle suggests computing $\alpha = \alpha(\delta) > 0$ in such a way that the corresponding Tikhonov solution:

$$\alpha x_{\alpha,\delta} + K^*Kx = K^*f_\delta,$$

that is, the minimum of the functional

$$J_{\alpha,\delta}(x) := \|Kx - f_\delta\|^2 + \alpha\|x\|^2,$$

satisfies the equation

$$\|Kx_{\alpha,\delta} - f_\delta\| = \delta.$$

This choice of α guarantees that, on one hand, the discrepancy is equal to δ and, on the other hand, α is not too small. Uniqueness and existence of the solution to $\|Kx_{\alpha,\delta} - f_\delta\| = \delta$ are justified by the following theorem.

Theorem 4.1. (see [9]) *Let X and Y be Hilbert spaces and $K : X \to Y$ be linear, compact, and one-to-one with a dense range in Y. Also, let $x = K^*z \in K^*(Y)$ be the exact solution to $Kx = f$ and $\|f - f_\delta\| \le \delta < \|f_\delta\|$. If $\|Kx_{\alpha,\delta} - f_\delta\| = \delta$, then $\|x_{\alpha,\delta} - x\| \le 2\sqrt{\delta\|z\|}$.*

Proof. Since $x_{\alpha,\delta}$ minimizes the Tikhonov functional $J_{\alpha,\delta}$, one can conclude that

$$\alpha\|x_{\alpha,\delta}\|^2 + \delta^2 = J_{\alpha,\delta}(x_{\alpha,\delta}) \le J_{\alpha,\delta}(x) = \alpha\|x\|^2 + \|f - f_\delta\|^2 \le \alpha\|x\|^2 + \delta^2,$$

and hence $\|x_{\alpha,\delta}\| \le \|x\|$ for all $\delta > 0$. This gives us the following important estimate:

$$\begin{aligned} \|x_{\alpha,\delta} - x\|^2 &= \|x_{\alpha,\delta}\|^2 - 2\Re e\,\langle x_{\alpha,\delta}, x\rangle + \|x\|^2 \\ &\le 2[\|x\|^2 - \Re e\,\langle x_{\alpha,\delta}, x\rangle] = 2\Re e\,\langle x - x_{\alpha,\delta}, x\rangle\,. \end{aligned}$$

Let $x = K^*z, z \in Y$. Then

$$\begin{aligned} \|x_{\alpha,\delta} - x\|^2 &\le 2\Re e\,\langle x - x_{\alpha,\delta}, K^*z\rangle = 2\Re e\,\langle f - Kx_{\alpha,\delta}, z\rangle \\ &\le 2\Re e\,\langle f - f_\delta, z\rangle + 2\Re e\,\langle f_\delta - Kx_{\alpha,\delta}, z\rangle \\ &\le 2\delta\|z\| + 2\delta\|z\| = 4\delta\|z\|. \end{aligned}$$

Therefore, $\|x_{\alpha,\delta} - x\| \le 2\sqrt{\delta\|z\|}$, as was to be shown. □

The condition $\|f_\delta\| > \delta$ is reasonable, since otherwise the right-hand side is less than δ, and one can take $x_\delta = 0$. This also shows that the determination of α does not have to satisfy the equation $\|Kx_{\alpha,\delta} - f_\delta\| = \delta$ exactly. One can use the following bounds that will guarantee the same level of accuracy.

$$c_1\delta \le \|Kx_{\alpha,\delta} - f_\delta\| \le c_2\delta.$$

In order to solve the equation $\|Kx_{\alpha,\delta} - f_\delta\| = \delta$, we introduce

$$\varphi(\alpha) := \|Kx_{\alpha,\delta} - f_\delta\|^2 - \delta^2. \tag{4.1}$$

Then the equation $\varphi(\alpha) = 0$ is equivalent to $\|Kx_{\alpha,\delta} - f_\delta\| = \delta$, and it can be solved numerically for example, by Newton's method:

$$\alpha_{j+1} = \alpha_j - \frac{\varphi(\alpha_j)}{\varphi'(\alpha_j)},\ j = 0, 1, 2, ...$$

The derivative $\varphi'(\alpha)$ can be calculated as follows [9]:

$$\begin{aligned}\varphi'(\alpha) &= \left[\langle Kx_{\alpha,\delta} - f_\delta, Kx_{\alpha,\delta} - f_\delta\rangle - \delta^2\right]'_\alpha \\ &= \left\langle K\frac{dx_{\alpha,\delta}}{d\alpha}, Kx_{\alpha,\delta} - f_\delta\right\rangle + \left\langle Kx_{\alpha,\delta} - f_\delta, K\frac{dx_{\alpha,\delta}}{d\alpha}\right\rangle \\ &= 2\Re e\left\langle K\frac{dx_{\alpha,\delta}}{d\alpha}, Kx_{\alpha,\delta} - f_\delta\right\rangle.\end{aligned}$$

One can compute $\frac{dx_{\alpha,\delta}}{d\alpha}$ by differentiating the identity

$$\alpha x_{\alpha,\delta} + K^*Kx_{\alpha,\delta} = K^*f_\delta$$

implicitly. Since the right-hand side does not depend on α, one has

$$x_{\alpha,\delta} + \alpha\frac{dx_{\alpha,\delta}}{d\alpha} + K^*K\frac{dx_{\alpha,\delta}}{d\alpha} = 0.$$

Solving for $\frac{dx_{\alpha,\delta}}{d\alpha}$, one gets

$$\frac{dx_{\alpha,\delta}}{d\alpha} = -[\alpha I + K^*K]^{-1}x_{\alpha,\delta} = -[\alpha I + K^*K]^{-1}[\alpha I + K^*K]^{-1}K^*f_\delta.$$

Substituting this into $\varphi'(\alpha)$, one derives

$$\varphi'(\alpha) = 2\Re e\left\langle -K[\alpha I + K^*K]^{-1}x_{\alpha,\delta}, Kx_{\alpha,\delta} - f_\delta\right\rangle$$

$$= 2\Re e\left\langle -K[\alpha I + K^*K]^{-1}[\alpha I + K^*K]^{-1}K^*f_\delta, K[\alpha I + K^*K]^{-1}K^*f_\delta - f_\delta\right\rangle.$$

5. Numerical Simulations

In this section, we will show that as opposed to the 'naive' discretization used in Section 2, the Tikhonov regularization combined with Morozov's discrepancy principle, results in a stable numerical solution to an ill-posed integral equation.

To illustrate the efficiency of Tikhonov-Morozov algorithm we consider the same equation as in Section 2:

$$Kx := \int_a^b k(t,s)x(s)ds = f(t), \quad K : L^2(a,b) \to L^2(c,d),$$

with $k(t,s) = \frac{s}{25}\sin(st)$, $[a,b] = [c,d] = [1,10]$, and $f(t)$ is the exact right-hand side. The values of the partition number n are 16, 32, 64, 128 and 256 while t

$\|x - x_{\alpha,\delta}\|,\ \delta_n = ch^2 + \sigma$			
$n\backslash\sigma$	0.1	0.05	0.001
64	1.5866	1.3273	1.0320
128	0.4744	0.3318	0.1557
256	0.1645	0.1027	0.0255

Table 2: Comparison of errors

σ	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$
0.1	n/a	n/a	1.186437	0.017617	0.005637
0.05	n/a	n/a	0.170580	0.011570	0.003773
0.001	n/a	n/a	0.068301	0.005483	0.001180

Table 3: Values of α by discrepancy principle

is varying from 1 to 10 for our tables. Note that

$$K^*x := \int_c^d \overline{k(s,t)}x(s)ds, \text{ where } K : L^2(c,d) \to L^2(a,b),$$

and $\overline{k(s,t)} = \overline{\frac{t}{25}\sin(ts)} = \frac{t}{25}\sin(ts)$.

We discretized the operator K by again using the trapezoidal rule to approximate the integral. For the discrete analog of the Tikhonov solution, one gets

$$x_{\alpha,\delta} = (\alpha I_n + K_n^* K_n)^{-1} K_n^* f_\delta.$$

In order to implement the discrepancy principle, one has to solve the following nonlinear equation:

$$\varphi(\alpha) := \|K_n x_{\alpha,\delta} - f_\delta\|^2 - \delta^2 = 0.$$

Introduce the notation

$$G_n := (\alpha I_n + K_n^* K_n)^{-1}.$$

Substituting $x_{\alpha,\delta} = G_n K_n^* f_\delta$ in the expression for $\varphi(\alpha)$, one concludes

$$\varphi(\alpha) := \|K_n G_n K_n^* f_\delta - f_\delta\|^2 - \delta^2 = \|(K_n G_n K_n^* - I_n) f_\delta\|^2 - \delta^2 = 0.$$

We solve $\varphi(\alpha) = 0$ by Newton's method:

$$\alpha_{j+1} = \alpha_j - \frac{\varphi(\alpha_j)}{\varphi'(\alpha_j)}, \quad j = 0, 1, \ldots .$$

	$\|\|x - x_{\alpha,\delta}\|\|$				
t	$n = 16, 32$	$n = 64$	$n = 128$	$n = 256$	$x = \exp(\frac{-t}{9})\cos(\frac{2\pi}{3}t)$
1.00	n/a	1.0380	1.0014	0.7116	-0.4474
3.25	n/a	0.7948	0.2415	0.0459	0.6035
5.50	n/a	0.6988	0.1562	0.0652	0.2714
7.75	n/a	0.6193	0.0950	0.0034	-0.3661
10.00	n/a	0.4484	0.0458	0.0019	-0.1646

Table 4: Comparative results for $\sigma = 0.001$

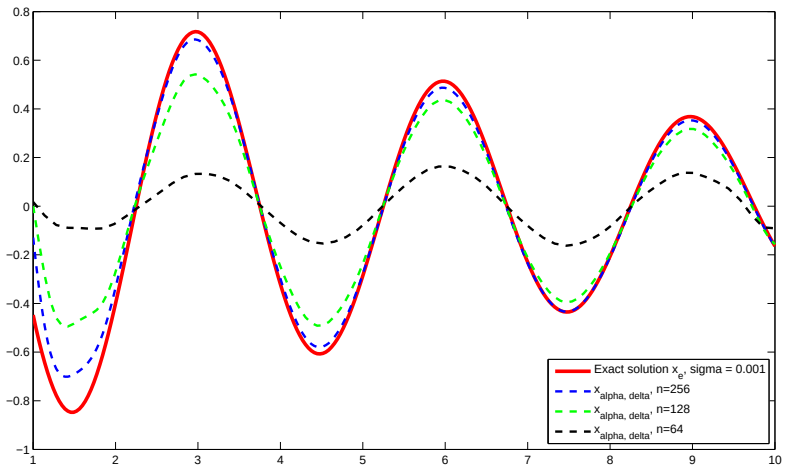

Figure 3: Exact and approximate solutions for $\sigma = 0.001$

As it has been shown in the previous section,

$$\varphi'(\alpha) = 2\left\langle K_n \frac{dx_{\alpha,\delta}}{d\alpha},\ K_n x_{\alpha,\delta} - f_\delta \right\rangle.$$

Recall, that the operator in our example is real-valued. Besides,

$$\frac{dx_{\alpha,\delta}}{d\alpha} = -G_n x_{\alpha,\delta} = -G_n^2 K_n^* f_\delta.$$

Hence,

$$\varphi'(\alpha) = -2\left\langle K_n G_n^2 K_n^* f_\delta, K_n G_n K_n^* f_\delta - f_\delta \right\rangle$$
$$= -2\left\langle K_n G_n^2 K_n^* f_\delta, (K_n G_n K_n^* - I_n) f_\delta \right\rangle.$$

	$\|\|x - x_{\alpha,\delta}\|\|$				
t	$n = 16, 32$	$n = 64$	$n = 128$	$n = 256$	$x = \exp(\frac{-t}{9})\cos(\frac{2\pi}{3}t)$
1.00	n/a	1.0147	1.0661	0.9153	-0.4474
3.25	n/a	0.9021	0.4116	0.1928	0.6035
5.50	n/a	0.8643	0.2558	0.0891	0.2714
7.75	n/a	0.8090	0.1970	0.0583	-0.3661
10.00	n/a	0.6796	0.0526	0.0010	-0.1646

Table 5: Comparative results for $\sigma = 0.05$

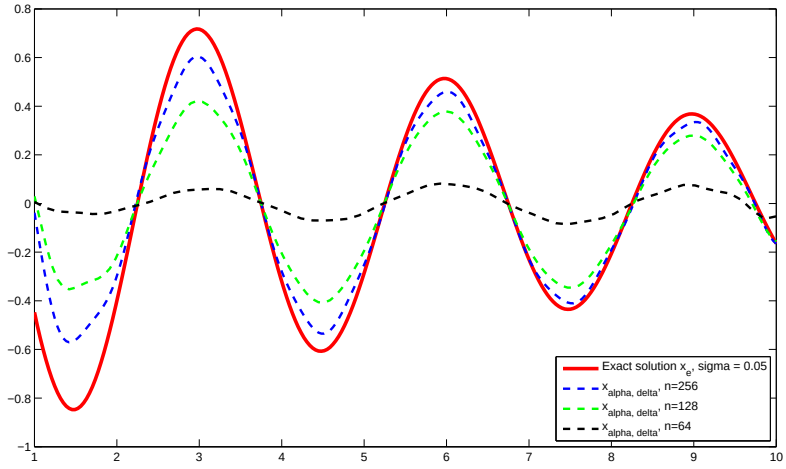

Figure 4: Exact and approximate solutions for $\sigma = 0.05$

We terminate Newton's method when

$$|\alpha_{j+1} - \alpha_j| < 10^{-6} \text{and } |\varphi(\alpha_j)| < 10^{-6},$$

which is comparable to the *MatLab* solver *fsolve*.

	$\|\|x - x_{\alpha,\delta}\|\|$				
t	$n = 16, 32$	$n = 64$	$n = 128$	$n = 256$	$x = \exp(\frac{-t}{9})\cos(\frac{2\pi}{3}t)$
1.00	n/a	1.0018	1.0769	0.9764	-0.4474
3.25	n/a	0.9839	0.5178	0.2759	0.6035
5.50	n/a	0.9798	0.3389	0.1037	0.2714
7.75	n/a	0.9680	0.2775	0.0906	-0.3661
10.00	n/a	0.9342	0.0822	0.0623	-0.1646

Table 6: Comparative results for $\sigma = 0.1$

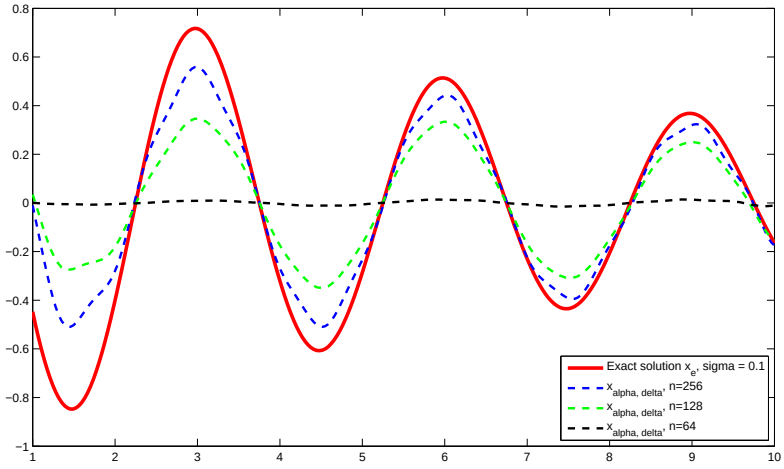

Figure 5: Exact and approximate solutions for $\sigma = 0.1$

6. Sources of Noise

To generate f_δ, we add a noise term to the exact function f. That noise term includes an input value of σ, and is equal to $\frac{\sigma\sqrt{2}}{3}\sin(10\pi t)$. Therefore, f_δ is:

$$f_\delta = f + \frac{\sigma\sqrt{2}}{3}\sin(10\pi t).$$

In our numerical simulations, the error in the data comes from two sources. First, it is the error due to inaccurate measurements, which we assume to be $\frac{\sigma\sqrt{2}}{3}\sin(10\pi t)$. Second, it is the discretization error that is generated when the

integral is replaced with the trapezoidal formula.

For large values of n, the discretization error can be ignored. However, we want to compare our results to those obtained by 'naive' numerical method used in Section 2. There we used $n = 16, 32, 64, 128$ and 256, since for larger values of n, the matrix K_n becomes singular to the *MatLab* precision. Hence, in this experiment, we take the same values of n (16, 32, 64, 128 and 256), and therefore the discretization error must be thoroughly investigated. The system we actually solve is $K_n x = f_n$, where f_n is a discrete analog of the exact right-hand side generated by the function $x(s) = \exp\left(-\frac{s}{9}\right)\cos\left(\frac{2\pi s}{3}\right)$. One has

$$\|f_n - f_\delta\| \le \|f_n - f\| + \|f - f_\delta\|.$$

The error for the composite trapezoidal rule is

$$f_n - f = K_n x - Kx = -\frac{g''(\xi)(b-a)}{12}h^2,$$

where $\xi \in [a, b]$, and $g(s)$ is the function which appears under the integral. In our case,

$$g(s) = k(t,s)x(s) \approx \frac{s}{25}\sin(st)\exp\left(\frac{-s}{9}\right)\cos\left(\frac{2\pi s}{3}\right)\ [a,b] = [1,10].$$

Hence

$$\begin{aligned}
\|f_n - f\| &= \left\| -\frac{g''(\xi)(b-a)}{12}h^2 \right\| \\
&= \left(\int_a^b \left(\frac{g''(\xi)(b-a)}{12}h^2 \right)^2 dt \right)^{\frac{1}{2}} \le \left(\int_a^b \left(\frac{M(b-a)}{12}h^2 \right)^2 dt \right)^{\frac{1}{2}} \\
&= \frac{M(b-a)}{12}h^2 \left(\int_a^b dt \right)^{\frac{1}{2}} = \frac{M(b-a)}{12}h^2(b-a)^{\frac{1}{2}} := ch^2.
\end{aligned}$$

Here $M = \max_{\xi\in[a,b]} |g''(\xi)| \le 17$ and $[a, b] = [1, 10]$, which gives us $c = 4.25$. One can also see that

$$\|f_e - f_\delta\|^2 = \int_1^{10} \sigma^2 \frac{2}{9}\sin^2(10\pi t)dt = \sigma^2 \frac{2}{9}\int_1^{10} \frac{1-\cos(20\pi t)}{2}dt = \sigma^2.$$

Therefore, $\|f_n - f_\delta\| \le ch^2 + \sigma := \delta_n$. Note that δ_n changes as n changes since h, the step size, is dependent on n, the number of partitions.

We apply the discrepancy principle with the values of $\sigma = 0.1$, $\sigma = 0.05$, and $\sigma = 0.001$. As α_0, the initial approximation, we take 0.25, because $\varphi(0) < 0$ and $\varphi(\frac{1}{2}) > 0$. Table 2 shows a comparison of the norms and confirms that as n gets larger we have convergence to the exact solution x.

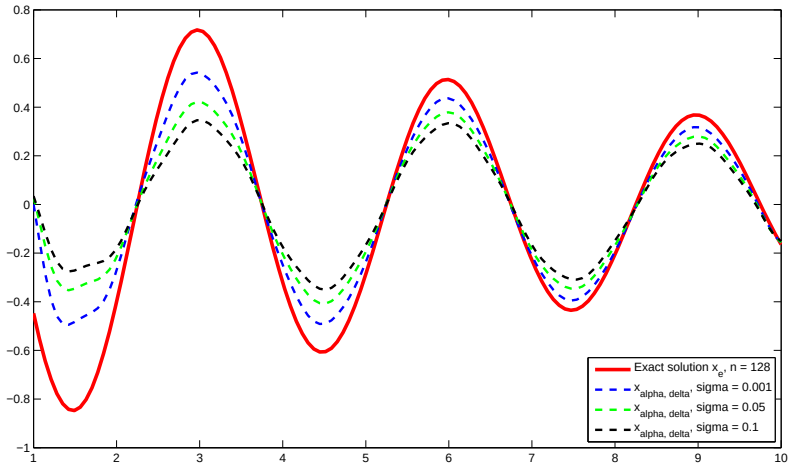

Figure 6: Exact and approximate solutions for $n = 128$

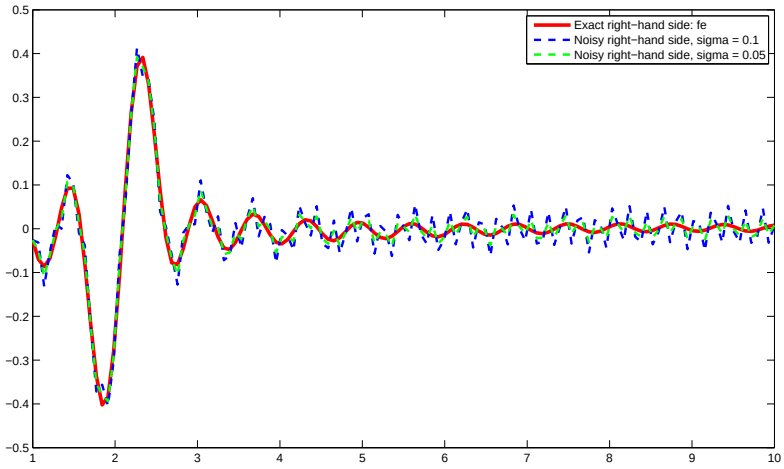

Figure 7: Exact and approximate values of right-hand side

The second Table 3, shows α versus σ for different values of n. Note that $\sigma = 0$ is not included in the table since we would only have the discretization error of the trapezoidal rule (as done in Table 1). One can see in Table 3, as

	$\|x(t_i) - x_i\|/\|x(t_i)\|$					
t	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	x
1.00	4.45	3.19	2.33×10^4	7.57×10^6	1.92×10^7	-0.4474
3.25	0.23	1.06	3.47×10^5	2.33×10^6	2.52×10^7	0.6035
5.50	0.73	4.92	9.19×10^5	1.87×10^7	2.75×10^7	0.2714
7.75	0.96	3.03	1.60×10^4	1.05×10^6	4.38×10^6	-0.3661
10.00	1.48	4.70	3.41	4.80×10^3	3.34×10^5	-0.1646

Table 7: Results for the 'naive' discretization

the noise gets smaller, α also gets smaller, as to be expected. After evaluating α by Morozov's discrepancy principle, we can now find $x_{\alpha,\delta}$ and compare it to our exact value of $x(t) = \exp(\frac{-t}{9})\sin(\frac{2\pi}{3}t)$ for $\sigma = 0.1,\ 0.05,\ 0.001$, and $n = 64, 128, 256$. Table 4 shows how $x_{\alpha,\delta}$ differs from x for $\sigma = 0.001$, our smallest value for σ, Figure 3 is the corresponding picture. As one can see, as n gets larger, the approximation converges towards the exact solution as guaranteed by the convergence theorem. Table 5 and Figure 4 show a similar result for $\sigma = 0.05$, as does Table 6 and Figure 5 for $\sigma = 0.1$, our largest σ. As expected, when σ is small we get a much better approximation of our exact solution. Figure 6 compares the exact and approximate solutions for various σ and $n = 128$. Again when $n = 256$ and $\sigma = 0.001$ the approximation is the best.

7. Summary

For our summary, we are using $\sigma = 0.05$ as part of the noise in the Tikhonov regularization with Morozov's discrepancy since it is the middle of the three that we choose to work with. We start by comparing the two tables, Table 7 for our 'naive' discretization and Table 8 for the Tikhonov-Morozov algorithm. Note that for the 'naive' discretization, $\sigma = 0$, and the only noise is due to finite-dimensional approximatation for the integral. As one can see, the numerical solutions obtained by the Tikhonov-Morozov algorithm are much better than those done by the 'naive' discretization. The errors for each n show that while the 'naive' discretization oscillates wildly and gets worse as n increases, the Tikhonov-Morozov algorithm actually gets better as n gets larger. This is confirmed by the general theory.

	$x_e - x_{\alpha,\delta}$				
t	$n = 16, 32$	$n = 64$	$n = 128$	$n = 256$	$x = \exp(\frac{-t}{9})\cos(\frac{2\pi}{3}t)$
1.00	n/a	1.0147	1.0661	0.9153	-0.4474
3.25	n/a	0.9021	0.4116	0.1928	0.6035
5.50	n/a	0.8643	0.2558	0.0891	0.2714
7.75	n/a	0.8090	0.1970	0.0583	-0.3661
10.00	n/a	0.6796	0.0526	0.0010	-0.1646

Table 8: Comparative results for $\sigma = 0.05$

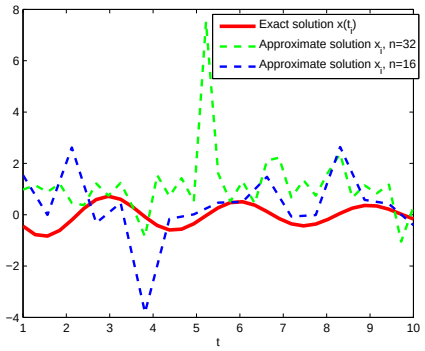

Figure 8: Exact and approximate solutions

We now move to the figures. Again Figure 8 refers to the 'naive' discretization while Figure 9 refers to the Tikhonov-Morozov algorithm. As foretold in the tables, the 'naive' discretization varies wildly while the Tikhonov-Morozov algorithm generates a smooth line that converges toward the exact value as n gets larger. All of the values of n could not even be shown in the 'naive' discretization due to the fact that at $n = 64$, $n = 128$ and $n = 256$ the matrix is nearly singular and does not return a better result but a much worse one. For the Tikhonov-Morozov algorithm, the situation is different. When $n = 256$, the computed solutions are most accurate. Note that even thought we are showing $\sigma = 0.05$ as our representative for the Tikhonov-Morozov scheme, the one with $\sigma = 0.1$, our largest value for σ, is still much better approximation than the naive discretization, which uses $\sigma = 0$.

There are definite advantages to using the Tikhonov regularization with Morozov's discrepancy principle. The first and most obvious is that it returns

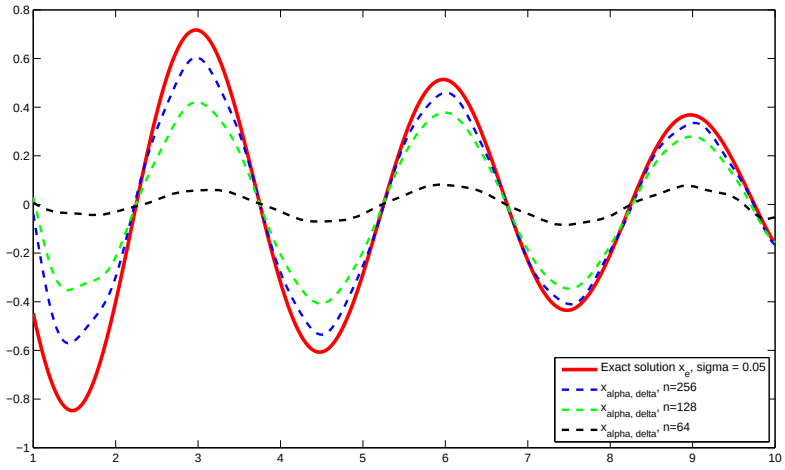

Figure 9: Exact and approximate solution for $\sigma = 0.05$

a much better approximate solution than the 'naive' discretization does. While using the same trapezoidal rule for the approximation of the integral, the relatively simple regularization by the Tikhonov method along with Morozov's discrepancy principle produces little to no discernible error as the partitions get sufficiently large, meaning you get a much closer approximation as the number of partitions gets larger for any amount of noise. In other words, the accuracy of the regularized solutions gets better as n approaches infinity, which is not the case for the 'naive' discretization.

The biggest disadvantage of Tikhonov regularization with Morozov's discrepancy principle is the repeated matrix manipulation done to compute the solution. Since the approximate solution uses the inverse of $(\alpha_j I + K^*K)$ one has to make sure that it exists for every j by starting Newton's iterations with an overestimate of α rather than underestimate. Another disadvantage of Tikhonov regularization is the over-smoothing effect. In order to reconstruct nonsmooth or discontinuous solutions, one has to use a different penalty term.

References

[1] A. Bakushinsky, A.Goncharsky, *Ill-Posed Problem Theory and Application*, Kluver, Dordrecht (1994).

[2] J. Demmel, *Applied Numerical Linear Algebra*, Society for Industrial and Applied Mathematics (SIAM), Philidephia (1997).

[3] H. Engl, M. Hanke, A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht (1996).

[4] James F. Epperson, *An Introduction to Numerical Methods and Analysis*, John Wiley and Sons, Inc., New York (2002).

[5] C.W. Groetsch, *The theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Pitman, Boston (1984).

[6] J. Hadamard, *Lectures on the Cauchy Problem in Linear Partial Differential Equations*, Yale University Press, New Haven (1923).

[7] M. Hanke, Accelerated Landweber iterations for the solution of ill-posed equations, *Numer. Math.*, **60** (1991) 341-373.

[8] B. Hofmann, *Regularization of Applied Inverse and Ill-Posed Problems*, Teubner, Leipzig (1986).

[9] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems*, Springer-Verlag, New York, NY (1996).

[10] V.A. Morozov, Choice of a parameter for the solution of functional equations by the regularization method, *Sov. Math. Doklady*, **8** (1967), 1000-1003.

[11] V.A. Morozov, The error principle in the solution of operational equations by the regularization method, *USSR Comput. Math. Math. Phys.*, **8** (1968), 63-87.

[12] V.A. Morozov, *Methods for Solving Incorrectly Posed Problems*, Springer-Verlag, Berlin (1984).

[13] V.A. Morozov, The principle of discrepancy in the solution of inconsistent equations by Tikhonov's regularization method, *Zhurnal Vychislitel'noy Matematiky i Matematicheskoy Fisiki*, **13** (1973), 5.

[14] D.L. Phillips, A technique for the numerical solution of certain integral equation of the first kind, *J. Assoc. Comput. Machinery*, **9**, No. 1 (1962), 84-97.

[15] A.N. Tikhonov, The solution of ill-posed problems, *Doklady Akad. Nauk SSSR*, **151**, (1963), 3.

[16] A.N. Tikhonov, A. Leonov, A. Yagola, *Nonlinear Ill-Posed Problems*, Chapman and Hall, London (1998).

[17] G. Vainikko, A. Veretennikov, *Iterative Procedures in Ill-Posed Problems*, Moscow, Nauka (1986).

[18] V.V. Vasin, A.L. Ageev, *Ill-Posed Problems with a priori Information*, VNU, Utrecht (1995).