BIT
2002, Vol. 43, No. 2, pp. 001–004
0006-3835/00/4004-0001 $15.00

A NEW ZERO-FINDER FOR TIKHONOV REGULARIZATION *

LOTHAR REICHEL[1,†] and ANDRIY SHYSHKOV[2,‡]

[1] *Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA. email: reichel@math.kent.edu*

[2] *Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA. email: ashyshko@math.kent.edu*

Abstract.

Tikhonov regularization with the regularization parameter determined by the discrepancy principle requires the computation of a zero of a rational function. We describe a cubically convergent zero-finder for this purpose.

AMS subject classification: 65F22, 65H05, 65R32.

Key words: ill-posed problem, Tikhonov regularization, discrepancy principle, zero-finder.

1 Introduction

Consider the problem of determining an approximate solution of the least-squares problem

$$\min_{x\in\mathbb{R}^n} \|Ax - b\| \tag{1.1}$$

with a right-hand side vector $b \in \mathbb{R}^m$, which has been contaminated by an error $e \in \mathbb{R}^m$, and a matrix $A \in \mathbb{R}^{m\times n}$ of ill-determined rank, i.e., A has many singular values of different size close to the origin. In particular, A is severely ill-conditioned and may be singular. We allow $m \geq n$, and do not require b to be in the range of A. Least-squares problems of this kind arise from the discretization of linear ill-posed problems, such as Fredholm integral equations of the first kind with a smooth kernel; see, e.g., Engl et al. [5], Hansen [8], and Morozov [11] for discussions on ill-posed problems, their discretization, and approximate solution.

In engineering applications that give rise to least-squares problems of this kind, the contamination e of b typically stems from measurement error. Let $\tilde{b}$ denote the unknown error-free vector associated with b, i.e.,

$$b = \tilde{b} + e. \tag{1.2}$$

*Received XXXXXXXXX. Revised XXXXXXXXX. Communicated by XXXXXXXX.
†Research supported in part by an OBR Research Challenge Grant.
‡Research supported in part by an OBR Research Challenge Grant.

We assume that $\tilde{b}$ is in the range of A and that an estimate of the error norm

$$\epsilon = \|e\|$$

is available. Throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm or the associated induced matrix norm.

Let $A^\dagger$ denote the Moore-Penrose pseudoinverse of A. We are interested in computing an approximation of $\tilde{x} = A^\dagger \tilde{b}$, the minimal-norm solution of the error-free least-squares problem associated with (1.1). Note that due to the error in b and the ill-conditioning of A, the minimal-norm least-squares solution $A^\dagger b$ of (1.1) typically is severely contaminated by propagated error and therefore, generally, is not a meaningful approximation of $\tilde{x}$.

Tikhonov regularization is a popular method for computing an approximation of $\tilde{x}$. This method replaces (1.1) by a penalized least-squares problem of the form

$$\min_{x\in\mathbb{R}^n} \left\{ \|Ax - b\|^2 + \frac{1}{\beta}\|Lx\|^2 \right\}, \tag{1.3}$$

where $\beta > 0$ is a regularization parameter and $L \in \mathbb{R}^{p\times n}$, $p \leq n$, a regularization operator. We will comment on the use of β instead of $\lambda = 1/\beta$ as regularization parameter in Section 2. Common choices of L include the identity matrix and approximations of differential operators. For ease of discussion, we let L be the identity in the present paper, however, our method also is applicable to other choices of L. We denote the solution of (1.3) by x_β.

The regularization parameter β is determined by the discrepancy principle, i.e., we seek to compute a positive value $\beta = \beta(\epsilon)$, so that the associated solution x_β of (1.3) satisfies

$$\|Ax_\beta - b\| = \eta\epsilon, \tag{1.4}$$

where $\eta > 1$ is a user-specified constant independent of ϵ. Generally, the more accurate our estimate ϵ of the norm of the error e, the closer to unity we choose η. It is known that $\lim_{\epsilon\searrow 0} x_\beta = \tilde{x}$; see, e.g., Engl et al. [5] or Groetsch [9] for proofs in Hilbert space settings.

The positive value of β that satisfies (1.4) is often determined by Newton's method applied to the function

$$\phi(\beta) = \|b - Ax_\beta\|^2 - \eta^2\epsilon^2; \tag{1.5}$$

see, e.g., Engl et al. [5, Section 9.4] or [2]. A different approach is described by Morozov [11, Section 26]. We discuss these schemes in Section 2.

Note that the function ϕ is decreasing and convex; see Proposition 2.1 below. Therefore Newton's method applied to ϕ gives quadratic and monotonic convergence if the initial iterate, β_0, is smaller than the desired zero. For every iterate β_j, Newton's method requires the evaluation of the function $\phi(\beta_j)$ and its derivative $\phi'(\beta_j)$. Let ϕ'' denote the second derivative of ϕ. In this paper, we point out that $\phi''(\beta_j)$ can be evaluated very inexpensively when the values $\phi(\beta_j)$ and $\phi'(\beta_j)$ are available. This observation makes it attractive to use a cubically convergent method for determining the positive zero of (1.5). We describe such methods and present comparisons to Newton's method.

This paper is organized as follows. Section 2 provides a few more details on Tikhonov regularization and also comments on a related problem that arises in spline approximation and trust-region computations. Cubically convergent zero-finders are described in Section 3, and a few computed examples can be found in Section 4. Section 5 contains concluding remarks.

2 Tikhonov regularization

The solution of (1.3) for any $\beta > 0$, with $L = I$, is given by

$$x_\beta = (A^T A + \beta^{-1} I)^{-1} A^T b. \tag{2.1}$$

The regularization parameter $\beta > 0$ determines how sensitive the solution of (1.3) is to perturbations in the right-hand side and how close x_β is to the solution $\tilde{x}$ of the unperturbed problem. The following proposition, essentially shown in [2], collects some properties of the function (1.5).

PROPOSITION 2.1. *Assume that $A^T b \neq 0$. Then the function (1.5) can be expressed as*

$$\phi(\beta) = b^T (\beta A A^T + I)^{-2} b - \eta^2 \epsilon^2, \tag{2.2}$$

which shows that ϕ and ϕ'' are decreasing and strictly convex for $\beta \geq 0$. Moreover, the equation

$$\phi(\beta) = \tau$$

has a unique solution β, such that $0 < \beta < \infty$, for any τ that satisfies

$$\|b_0\|^2 - \eta^2 \epsilon^2 < \tau < \|b\|^2 - \eta^2 \epsilon^2,$$

where b_0 denotes the orthogonal projection of b onto the null space of AA^T.

The representation (2.2) shows that the function (1.5) is rational. Its poles are negative reciprocal eigenvalues of AA^T. In particular, all poles are real and strictly negative.

Generally, $\|b_0\|$ is tiny or even vanishing. We will assume that

$$\|b_0\| < \eta\epsilon < \|b\|. \tag{2.3}$$

Then ϕ has a positive zero. We wish to determine this zero. The right-hand side inequality limits the amount of error in b, but is satisfied in typical applications.

It is not meaningful to compute the positive zero of (1.5) to very high accuracy since ϵ is only an estimate of the norm of the error. We therefore seek to determine a value β, such that

$$-(\eta^2 - 1)\epsilon^2 \leq \phi(\beta) \leq 0. \tag{2.4}$$

When a regularization parameter β much larger than the positive zero of ϕ is used, the associated Tikhonov solution x_β may be severely contaminated by propagated error due to the error e in b. In fact, it may be difficult to compute x_β due to the ill-conditioning of the linear system that has to be solved. The use of a too large value of β is referred to as *underregularization*. The upper bound in (2.4) secures that underregularization cannot take place, i.e., that the computed value of β is smaller or equal to the positive zero of ϕ.

We remark that the convexity of ϕ simplifies the design of iterative methods for determining a value of β that satisfies (2.4). If the regularization parameter β in (1.3) is replaced by $\lambda = 1/\beta$, then convexity of the analogue of the function ϕ so obtained is in general not guaranteed.

We describe the computations required when our zero-finder is applied to problems (1.1) with a matrix $A \in \mathbb{R}^{m \times n}$ of small enough size to allow factorization. On many PCs this limits n to be a few hundred and m to be a moderate

multiple (≥ 1) of n. Many discretized linear ill-posed problems that arise in applications satisfy these size-constraints. At the end of Section 3, we outline how the zero-finder can be applied to large-scale problems for which it is infeasible or undesirable to factor A.

Thus, assume that a factorization of the form

$$A = UBV^T \tag{2.5}$$

can be computed, where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal, and $B \in \mathbb{R}^{m\times n}$ is upper bidiagonal. The matrices U and V do not have to be formed explicitly; for instance, they may be products of Householder transformations or Givens rotations. Substituting (2.5) into (2.2) gives

$$\phi(\beta) = \hat{b}^T(\beta BB^T + I)^{-2}\hat{b} - \eta^2\epsilon^2 \tag{2.6}$$

with $\hat{b} = U^T b$. It follows that

$$\phi(\beta) = \|z_\beta\|^2 - \eta^2\epsilon^2, \tag{2.7}$$

where $z_\beta \in \mathbb{R}^m$ solves the equation

$$(\beta BB^T + I)z = \hat{b}.$$

These are the normal equations associated with the least-squares problem

$$\min_{z\in\mathbb{R}^m} \left\| \begin{bmatrix} \beta^{1/2}B^T \\ I \end{bmatrix} z - \begin{bmatrix} 0 \\ \hat{b} \end{bmatrix} \right\|. \tag{2.8}$$

We compute $\phi(\beta)$ for any positive value of β by first solving this least-square problem for z_β and then evaluating (2.7). When the matrix B is available, the least-squares problem (2.8) can be solved in $O(m)$ arithmetic floating point operations for each value of $\beta > 0$ by a scheme similar to the one described by Eldén in [4].

The first and second derivatives of ϕ at β are evaluated in the same manner as $\phi(\beta)$ as follows. Let z_β solve (2.8) and introduce the vector

$$w_\beta = (\beta BB^T + I)^{-1}z_\beta,$$

which we compute by solving a least-squares problem analogous to (2.8). Then

$$\phi'(\beta) = \frac{2}{\beta}z_\beta^T(w_\beta - z_\beta).$$

For the second derivative of ϕ at β, we obtain similarly

$$\phi''(\beta) = \frac{6}{\beta^2}(z_\beta - w_\beta)^T(z_\beta - w_\beta). \tag{2.9}$$

With the matrix B available, the evaluation of $\phi(\beta)$ and $\phi'(\beta)$ requires the solution of two structured least-squares problems, the computation of the vector $w_\beta - z_\beta$, and the evaluation of two inner products. The evaluation of $\phi''(\beta)$

only requires the additional computation of one inner product and a few scalar arithmetic operations; see (2.9). The ease with which $\phi''(\beta)$ can be computed makes it natural to try to use this quantity in a zero-finder for (1.5). This is discussed in the following section.

We remark that when β is large, both vectors z_β and w_β may be of large norm and loss of significant digits might occur when forming the difference $z_\beta - w_\beta$. In our numerical experiments this has not caused any problems. Nevertheless, explicit computation of the difference $z_\beta - w_\beta$ can be avoided by using the expression

$$z_\beta - w_\beta = (\beta BB^T + I)^{-1}\beta BB^T z_\beta.$$

Thus, we can determine $z_\beta - w_\beta$ by solving a least-squares problem of the form (2.8) with $\hat{b}$ replaced by $\beta BB^T z_\beta$. However, this requires more computational effort than computing z_β and w_β separately.

For future reference, we provide expressions for ϕ and its derivatives at the origin. These expressions can be determined by differentiating (2.6) and letting $\beta \searrow 0$. This yields

$$\phi(0) = \|\hat{b}\|^2 - \eta^2\epsilon^2, \quad \phi'(0) = -2\|B^T\hat{b}\|^2, \quad \phi''(0) = 6\|BB^T\hat{b}\|^2. \tag{2.10}$$

We will use our zero-finder with initial approximation $\beta_0 = 0$ of the positive zero of ϕ, and apply the above formulas to determine an improved approximation $\beta_1 > 0$. In order to avoid underregularization during the computation of the desired zero, we would like the zero-finder to yield a sequence of β-values that converge to the desired value from the left. Generally, positive values of β that are smaller than the desired solution are not explicitly known. Therefore we use the initial value $\beta_0 = 0$. This value is attractive also because the formulas (2.10) are simple to evaluate. We note that the initial value $\beta_0 = 0$ also is used for Newton's method for the same reason; see, e.g., [2].

We remark that Morozov [11, Section 26] discusses the zero-finding problem of the present paper and proposes that Newton's method be applied to the computation of the positive zero of

$$\hat{\phi}_s(\beta) = \left(b^T(\beta AA^T + I)^{-2}b\right)^s - (\eta\epsilon)^{2s} \tag{2.11}$$

for $s = -1/2$. This function is shown to be increasing and convex. Therefore, given an initial approximation, β_0, smaller than the desired solution, Newton's method determines a new approximation, β_1, larger than the smallest positive zero of $\hat{\phi}_s$. The computation of the next approximation, β_2, of this zero therefore requires the solution of an underregularized problem. The possible difficulties that underregularization may cause has been discussed above. It is the aim of the present paper to describe a cubically convergent zero-finder that generally avoids the solution underregularized problems.

For positive values of s in (2.11), the function $\beta \to \hat{\phi}_s(\beta)$ is decreasing and convex; in particular, $s = 1$ yields the function ϕ defined by (1.5). Numerical experiments with initial value $\beta_0 = 0$ do not display faster convergence of Newton's method or our cubically convergent zero-finder for s-values strictly between 0 and 1 than for $s = 1$. We therefore apply our zero-finder to the function (1.5) and compare it to Newton's method applied to this function.

Having determined a value of β that satisfies (2.4) with our cubically convergent zero-finder, we compute the approximate solution x_β of (1.1), given by

(2.1), by substituting the factorization (2.5) into (2.1) and solving a least-squares problem of a form related to (2.8).

We conclude this section with some comments on zero-finding problems related to the one discussed in the present paper. Trust-region methods for the minimization of a function $f : \mathbb{R}^n \to \mathbb{R}$ require the repeated solution of problems of the form

$$(H + \beta I)x = b, \qquad \|x\| = \rho, \tag{2.12}$$

where $\rho > 0$ is the trust-region radius, the symmetric matrix $H \in \mathbb{R}^{n\times n}$ is an approximation of the Hessian of f at $z \in \mathbb{R}^n$, the available approximation of the minimum, and $b \in \mathbb{R}^n$ is the negative gradient of f at z. The solution x of (2.12) is a correction of z; see, e.g., [3] for an introduction to trust-region methods. Thus, we need to compute both $\beta > 0$ and $x \in \mathbb{R}^n$ that satisfy (2.12). Let

$$\psi(\beta) = b^T(H + \beta I)^{-2}b.$$

The parameter β is computed by solving

$$\psi(\beta) = \rho^2. \tag{2.13}$$

Zero-finding problems of this form also arise in spline approximation; see, e.g., Reinsch [12]. Common solution methods include a rational zero-finder discussed, e.g., by Dennis and Schnabel [3] and a modification of Newton's method proposed by Reinsch [12]; see also Moré and Sorensen [10]. These methods require the evaluation of ψ and ψ' at approximations of the solution of (2.13). The values can be computed similarly as those of ϕ and ϕ'. Moreover, ψ'' can be evaluated inexpensively in the same manner as ϕ''. Whether the use of a cubically convergent zero-finder is warranted in this context depends on the quality of the available initial approximate solution. A preliminary comparison of a cubically convergent zero-finder for the solution of equations of the form (2.13) with the quadratically convergent method by Reinsch [12] showed the former to require only slightly fewer or the same number of iterations than the latter. Since the observed difference in performance was small, we omit a discussion on cubically convergent zero-finders for equation (2.13) in the present paper. A thorough study would be required to assess the benefits of using a cubically convergent method. Moreover, since standard trust-region subproblems are not very ill-conditioned, underregularization would generally be acceptable. Therefore other zero-finders than those discussed in the present paper also can be used for the solution of (2.13).

3 Cubically convergent zero-finders

This section describes cubically convergent zero-finders for functions $\varphi : \mathbb{R}_+ \to \mathbb{R}$ that are three times continuously differentiable and satisfy

$$\varphi'(\beta) < 0, \qquad \varphi''(\beta) > 0, \qquad \beta > 0, \tag{3.1}$$

as well as

$$\lim_{\beta\searrow 0} \varphi(\beta) > 0, \qquad \lim_{\beta\to\infty} \varphi(\beta) < 0. \tag{3.2}$$

Theorems 3.2 and 3.3 below also requires φ'' to be decreasing and convex. Our aim is to compute the positive zero of the function (1.5). This function satisfies all the conditions imposed on φ; cf. Proposition 2.1 and (2.3).

It follows from (3.1) and (3.2) that φ has a unique zero, denoted by $\tilde{\beta}$, in $\mathbb{R}_+$. We first describe the zero-finder used for determining $\tilde{\beta}$ in the numerical experiments of Section 4. Other zero-finders will be discussed below.

Let β_j be an available approximation of $\tilde{\beta}$ and define the function

$$\tau_j(\beta) = \alpha_j \sqrt{\beta - \mu_j} + \gamma_j, \tag{3.3}$$

whose coefficients α_j, μ_j, and γ_j are determined by the interpolation conditions

$$\tau_j(\beta_j) = \varphi(\beta_j), \qquad \tau_j'(\beta_j) = \varphi'(\beta_j), \qquad \tau_j''(\beta_j) = \varphi''(\beta_j). \tag{3.4}$$

Then

$$\alpha_j = 2\varphi'(\beta_j)\sqrt{\beta_j - \mu_j}, \tag{3.5}$$

$$\mu_j = \beta_j + \frac{\varphi'(\beta_j)}{2\varphi''(\beta_j)}, \tag{3.6}$$

$$\gamma_j = \varphi(\beta_j) - \alpha_j \sqrt{\beta_j - \mu_j}. \tag{3.7}$$

It follows from the properties (3.1) of φ that

$$\beta_j - \mu_j > 0, \qquad \alpha_j < 0, \qquad \gamma_j > \varphi(\beta_j). \tag{3.8}$$

Thus, the function τ_j is decreasing and convex for $\beta \geq \mu_j$. Further, τ_j has a unique real zero given by

$$\beta_{j+1} = \mu_j + \frac{\gamma_j^2}{\alpha_j^2}. \tag{3.9}$$

We use β_{j+1} as our new approximation of the zero $\tilde{\beta}$ of φ. The following theorem establishes cubic convergence of this zero-finder.

THEOREM 3.1. *Let the coefficients α_j, μ_j, and γ_j be defined by the equations (3.5)-(3.7), let β_{j+1} be given by (3.9), and let $\tilde{\beta}$ denote the zero of $\varphi(\beta)$. Then*

$$|\tilde{\beta} - \beta_{j+1}| = \mathcal{O}(|\tilde{\beta} - \beta_j|^3) \tag{3.10}$$

for β_j sufficiently close to $\tilde{\beta}$. Moreover, if $\beta_j < \tilde{\beta}$ then

$$\beta_{j+1} > \beta_j. \tag{3.11}$$

PROOF. We first show (3.11). It follows by application of (3.9), (3.7), and (3.8), in order, that

$$\beta_{j+1} - \beta_j = \mu_j + \frac{\gamma_j^2}{\alpha_j^2} - \beta_j = \frac{\gamma_j^2}{\alpha_j^2} - \frac{(\varphi(\beta_j) - \gamma_j)^2}{\alpha_j^2} = \frac{2\varphi(\beta_j)\gamma_j - \varphi^2(\beta_j)}{\alpha_j^2} > 0,$$

where the inequality follows from $\gamma_j - \varphi(\beta_j) = -\alpha_j\sqrt{\beta_j - \mu_j} > 0$ and $\varphi(\beta_j) > 0$.

To prove (3.10), assume that β_j is close to $\tilde{\beta}$. Then the interpolation conditions (3.4) yield

$$\varphi(\tilde{\beta}) - \tau_j(\tilde{\beta}) = \mathcal{O}(|\tilde{\beta} - \beta_j|^3).$$

Therefore,

$$\begin{aligned}
|\tilde{\beta}-\beta_{j+1}| &= \left|\tilde{\beta}-\mu_j-\frac{\gamma_j^2}{\alpha_j^2}\right| = \left|\frac{\alpha_j^2(\tilde{\beta}-\mu_j)-\gamma_j^2}{\alpha_j^2}\right| \\
&= \left|\frac{(\alpha_j\sqrt{\tilde{\beta}-\mu_j}-\gamma_j)(\alpha_j\sqrt{\tilde{\beta}-\mu_j}+\gamma_j)}{\alpha_j^2}\right| \\
&= \left|\frac{\alpha_j\sqrt{\tilde{\beta}-\mu_j}-\gamma_j}{\alpha_j^2}\right| \left|\tau_j(\tilde{\beta})-\varphi(\tilde{\beta})\right| \\
&= \left|\frac{\alpha_j\sqrt{\tilde{\beta}-\mu_j}-\gamma_j}{\alpha_j^2}\right| \mathcal{O}(|\tilde{\beta}-\beta_j|^3).
\end{aligned}$$

The theorem now follows from (3.5), (3.6), and the observation that there exist finite positive constants c_1 and c_2, such that $|\varphi'(\beta)| > c_1$ and $|\varphi''(\beta)| < c_2$ for $\beta \leq \tilde{\beta}$. □

Numerical experiments with this zero-finder, some of which are reported in Section 4, show fast and monotonic convergence. However, we cannot guarantee that $\beta_j < \tilde{\beta}$ for all j and therefore describe a cubically and monotonically convergent zero-finder that can be used if the zero-finder above determines an iterate β_{j+1} larger than $\tilde{\beta}$.

Thus, assume that for some $k \geq 1$, iterates β_j and β_{j+k}, such that

$$\beta_j < \tilde{\beta} < \beta_{j+k} \tag{3.12}$$

are available. Then $\varphi(\beta_j) > 0$ and $\varphi(\beta_{j+k}) < 0$. Introduce the function

$$\begin{aligned}
\psi_k(\beta) &= \frac{\varphi''_{j+k}-\varphi''_j}{6(\beta_{j+k}-\beta_j)}(\beta-\beta_{j+k})^3 + \frac{\varphi''_{j+k}}{2}(\beta-\beta_{j+k})^2 \\
&+ \varphi'_{j+k}(\beta-\beta_{j+k}) + \varphi_{j+k},
\end{aligned} \tag{3.13}$$

where we have used the notation

$$\varphi_\ell = \varphi(\beta_\ell), \qquad \varphi'_\ell = \varphi'(\beta_\ell), \qquad \varphi''_\ell = \varphi''(\beta_\ell).$$

The following theorem establishes a few properties of ψ_k.

THEOREM 3.2. *Assume that (3.12) holds and that φ satisfies (3.1) and (3.2). Moreover, let φ'' be decreasing and strictly convex in the open interval (β_j, β_{j+k}). Then the function ψ_k, given by (3.13), is decreasing and strictly convex in (β_j, β_{j+k}), and satisfies*

$$\varphi(\beta) < \psi_k(\beta) \tag{3.14}$$

and

$$\varphi(\beta_{j+k}) = \psi_k(\beta_{j+k}). \tag{3.15}$$

Let β_{j+k+1} denote the zero of ψ_k in the interval (β_j, β_{j+k}). Then

$$\tilde{\beta} < \beta_{j+k+1} < \beta_{j+k}, \qquad k \geq 1. \tag{3.16}$$

PROOF. We first note that (3.15) follows immediately from the definition of ψ_k. Consider the representation of φ in the interval (β_j, β_{j+k}),

$$\varphi(\beta) = \int_{\beta_{j+k}}^{\beta} \int_{\beta_{j+k}}^{\zeta} \varphi''(\xi) d\zeta d\xi + \varphi'_{j+k}(\beta - \beta_{j+k}) + \varphi_{j+k}. \tag{3.17}$$

We obtain a similar representation of $\psi_k(\beta)$ by replacing $\varphi''(\xi)$ in (3.17) by the function

$$\psi''_k(\xi) = \varphi''_{j+k} + \frac{\varphi''_{j+k} - \varphi''_j}{\beta_{j+k} - \beta_j}(\xi - \beta_{j+k}). \tag{3.18}$$

Since φ'' is decreasing and convex, and ψ''_k is linear and interpolates φ'' at β_j and β_{j+k}, it follows that $\psi''_k(\xi) > \varphi''(\xi)$ in the interval (β_j, β_{j+k}). The latter inequality, the representation (3.17), and the analogous representation of ψ_k yields (3.14).

It is easy to see that

$$\psi'_k(\beta) = \frac{\varphi''_{j+k} - \varphi''_j}{2(\beta_{j+k} - \beta_j)}(\beta - \beta_{j+k})^2 + \varphi''_{j+k}(\beta - \beta_{j+k}) + \varphi'_{j+k} \tag{3.19}$$

is negative in (β_j, β_{j+k}). It follows from (3.18) that ψ''_k is positive in this interval. Hence, ψ_k is strictly convex and decreasing in (β_j, β_{j+k}). Moreover, ψ_k changes sign in the interval $(\tilde{\beta}, \beta_{j+k})$. The latter is a consequence of (3.12), (3.14), and (3.15). Thus, ψ_k has a unique zero in $(\tilde{\beta}, \beta_{j+k})$. This shows (3.16). Repetition of the above argument for $k = 1, 2, \ldots$ yields $\beta_{j+1} > \beta_{j+2} > \beta_{j+3} > \cdots$. This completes the proof of the theorem. □

Assume that the zero-finder (3.3)-(3.9) yields an approximation β_{j+1} of $\tilde{\beta}$ which satisfies (3.12). We then determine the zeros β_{j+k+1} of ψ_k for increasing values of $k \geq 1$ until $\varphi(\beta_{j+k+1})$ is sufficiently close to zero. The next theorem discusses the limit of this sequence and the rate of convergence.

THEOREM 3.3. *Let, for each $k \geq 1$, ψ_k be defined by (3.13) and let β_{j+k+1} denote the zero of ψ_k in the interval (β_j, β_{j+k}), whose end points satisfy (3.12). Let φ satisfy the conditions of Theorem 3.2. Then the β_{j+k} converge monotonically to $\tilde{\beta}$ as $k \to \infty$. Moreover,*

$$|\tilde{\beta} - \beta_{j+k+1}| = \mathcal{O}(|\tilde{\beta} - \beta_{j+k}|^3), \tag{3.20}$$

for k sufficiently large, which shows that convergence is cubic.

PROOF. We first show monotonic convergence of the β_{j+k} as k increases. It follows from (3.16) that the β_{j+k} decrease monotonically and are bounded below as k increases. Therefore, there is a constant $\bar{\beta} \geq \tilde{\beta}$, such that $\bar{\beta} = \lim_{k\to\infty} \beta_{j+k}$. We will now show that $\bar{\beta} = \tilde{\beta}$.

Assume that $\bar{\beta} > \tilde{\beta}$. Since φ is convex and $\varphi(\tilde{\beta}) = 0$, there exist a constant $\sigma > 0$, such that $\varphi(\beta_{j+k}) < \varphi(\bar{\beta}) < -\sigma$ for all $k \geq 1$, which together with (3.15)

gives $\psi_k(\beta_{j+k}) < -\sigma$ for all $k \geq 1$. In addition, $\psi_k(\beta_{j+k+1}) = 0$ for all $k \geq 1$, and $\beta_{j+k} \to \beta_{j+k+1}$ as $k \to \infty$. This shows that the sequence of functions ψ_k, $k = 1, 2, \ldots$, is not uniformly continuous in the interval (β_j, β_{j+k}). However, the sequence ψ_k, $k = 1, 2, \ldots$, is uniformly continuous in $(\bar{\beta}, \beta_{j+k})$. This follows easily from (3.19) and the fact that $\beta_{j+k} - \beta_j > c$ for some constant $c > 0$ and all $k \geq 1$. For instance, we may choose $c = \bar{\beta} - \beta_j$. This contradiction shows that $\bar{\beta} = \tilde{\beta}$.

To prove (3.20), assume that β_{j+k} is close to $\tilde{\beta}$. It follows from (3.13) that

$$\psi_k(\beta_{j+k}) = \varphi(\beta_{j+k}), \quad \psi_k'(\beta_{j+k}) = \varphi'(\beta_{j+k}), \quad \psi_k''(\beta_{j+k}) = \varphi''(\beta_{j+k}).$$

Then, for β close to β_{j+k}, we have

$$\varphi(\beta) - \psi_k(\beta) = (\beta_{j+k} - \beta)^3 \frac{\varphi'''(\xi)}{6}, \tag{3.21}$$

for some ξ in the interval with end points β and β_{j+k}. Since β_{j+k+1} is a zero of ψ_k, we obtain from (3.21) and (3.16) that

$$\varphi(\beta_{j+k+1}) = (\beta_{j+k} - \beta_{j+k+1})^3 \frac{\varphi'''(\breve{\xi})}{6}, \tag{3.22}$$

for some $\breve{\xi} \in (\beta_{j+k+1}, \beta_{j+k})$. The property $\varphi(\tilde{\beta}) = 0$ yields

$$\varphi(\beta_{j+k+1}) = (\beta_{j+k+1} - \tilde{\beta})\varphi'(\tilde{\xi}) \tag{3.23}$$

for some $\tilde{\xi} \in (\tilde{\beta}, \beta_{j+k})$. Using the fact that φ' is bounded away from zero in a vicinity of $\tilde{\beta}$, as well as (3.22) and (3.23), gives

$$|\beta_{j+k+1} - \tilde{\beta}| = |\beta_{j+k+1} - \beta_k|^3 \frac{\varphi'''(\breve{\xi})}{6\varphi'(\tilde{\xi})}, \tag{3.24}$$

and by the monotonicity of the sequence β_{j+k}, cf. (3.16), we obtain

$$|\beta_{j+k+1} - \beta_{j+k}| \leq |\tilde{\beta} - \beta_{j+k}|, \qquad k \geq 1. \tag{3.25}$$

Finally, (3.24) and (3.25) yield $|\beta_{j+k+1} - \tilde{\beta}| \leq \theta |\beta_{j+k} - \tilde{\beta}|^3$ for all $k \geq 1$ and some constant $\theta > 0$ independent of k. This shows cubic convergence. □

We refer to Traub [14, pp. 67–75] for general results on zero-finders based on polynomial approximation. Note that ψ_k is a cubic polynomial with negative leading coefficient. If the zeros of ψ_k are real and distinct, then ψ_k decreases only in a neighborhood of the smallest and largest zeros. Further, ψ_k is convex only in a neighborhood of the smallest zero. Since by Theorem 3.2, ψ_k is convex and decreasing in the interval (β_j, β_{j+k}) and has exactly one zero there, this zero is the smallest zero of ψ_k.

The zeros $\beta_{j+k,\ell}$, $1 \leq \ell \leq 3$, of ψ_k are given by Cardano's formula

$$\begin{aligned}
\beta_{j+k,1} &= -\frac{1}{3}a_2 + (S + T), \\
\beta_{j+k,2} &= -\frac{1}{3}a_2 - \frac{1}{2}(S + T) + \frac{1}{2}i\sqrt{3}(S - T), \\
\beta_{j+k,3} &= -\frac{1}{3}a_2 - \frac{1}{2}(S + T) - \frac{1}{2}i\sqrt{3}(S - T),
\end{aligned}$$

where

$$S = \sqrt[3]{R + \sqrt{D}}, \qquad T = \sqrt[3]{R - \sqrt{D}}, \qquad D = Q^3 + R^2, \qquad i = \sqrt{-1},$$

with

$$Q = \frac{3a_1 - a_2^2}{9}, \qquad R = \frac{9a_1a_2 - 27a_0 - 2a_2^3}{54}$$

and

$$\begin{aligned} a_2 &= \frac{3\varphi''_{j+k}(\beta_{j+k} - \beta_j)}{\varphi''_{j+k} - \varphi''_j}, \\ a_1 &= \frac{6\varphi'_{j+k}(\beta_{j+k} - \beta_j)}{\varphi''_{j+k} - \varphi''_j}, \\ a_0 &= \frac{6\varphi_{j+k}(\beta_{j+k} - \beta_j)}{\varphi''_{j+k} - \varphi''_j}. \end{aligned}$$

If $D > 0$, then ψ_k has a pair of complex conjugate zeros, $\beta_{j+k,2}$ and $\beta_{j+k,3}$, which we do not need to compute. In this situation, we take $\beta_{j+k,1}$ as our next approximation, β_{j+k+1}, of $\tilde{\beta}$. Otherwise, all zeros are real, and we choose the smallest positive zero as our next approximation β_{j+k+1} of $\tilde{\beta}$.

There are many other cubically convergent zero-finders available. However, not all of them are well suited for the problem at hand. We illustrate this with Halley's method, which determines the correction

$$\Delta\beta_j = -\frac{\varphi(\beta_j)}{\varphi'(\beta_j) - \frac{\varphi''(\beta_j)\varphi(\beta_j)}{2\varphi'(\beta_j)}} \tag{3.26}$$

of the iterate β_j; see, e.g., Gander [6] for a discussion of this and other methods. When applied to determine the smallest positive zero of (1.5) with an initial iterate smaller than the desired zero, the denominator in (3.26) often became negative and convergence was not monotonic; moreover, complex arithmetic may be required. We therefore feel that Halley's method is not the method of choice for the present application.

We briefly comment on the situation when A is so large to make the computation of the factorization (2.5) undesirable or impossible. An approximate solution x_β of (1.3) with a β-value, such that (1.4) holds, then can be determined by partial Lanczos bidiagonalization of A. Application of ℓ steps of Lanczos bidiagonalization to A yields an $(\ell + 1) \times \ell$ lower bidiagonal matrix, which can be used similarly as the matrix B in (2.5). An algorithm is described in [2]. This algorithm uses Newton's method applied to a function related to ϕ to compute a suitable value of β. Newton's method in this algorithm may be replaced by the zero-finders of the present paper. For small problems, we may compute the singular value decomposition of A instead of the factorization (2.5). The zero-finders of this paper also can be used in this context.

4 Computed examples

This section compares the performance of the new cubically convergent zero-finder (3.3)-(3.9) to Newton's method. We use $\eta = 1.1$ and the initial iterate

$\beta_0 = 0$ in all examples. Our reasons for the latter choice is that $\beta_0 = 0$ is smaller than the positive zero of (1.5) and, moreover, ϕ and its first two derivatives easily can be evaluated; cf. (2.10). All computations are carried out in MATLAB with about 16 significant decimal digits.

δ	i	β_i	$\phi(\beta_i)$	$\|x_{\beta_i} - \tilde{x}\|/\|\tilde{x}\|$
$1 \cdot 10^{-2}$	10	$3.98 \cdot 10^{2}$	$-1.15 \cdot 10^{-4}$	$2.05 \cdot 10^{-1}$
$1 \cdot 10^{-3}$	14	$1.10 \cdot 10^{4}$	$-1.13 \cdot 10^{-6}$	$1.45 \cdot 10^{-1}$
$1 \cdot 10^{-4}$	17	$2.05 \cdot 10^{5}$	$-5.12 \cdot 10^{-10}$	$1.15 \cdot 10^{-1}$
$1 \cdot 10^{-5}$	22	$4.56 \cdot 10^{7}$	$-1.20 \cdot 10^{-10}$	$5.91 \cdot 10^{-2}$

Table 4.1: Example 4.1: Baart test problem using the zero finder (3.3)-(3.9). The table shows the number of iterations i, the value of regularization parameter β_i, the value of the function ϕ, defined by (1.5), at β_i, and the relative error of the computed approximate solution x_{β_i}.

δ	i	β_i	$\phi(\beta_i)$	$\|x_{\beta_i} - \tilde{x}\|/\|\tilde{x}\|$
$1 \cdot 10^{-2}$	15	$2.75 \cdot 10^{2}$	$-6.73 \cdot 10^{-5}$	$2.20 \cdot 10^{-1}$
$1 \cdot 10^{-3}$	21	$6.45 \cdot 10^{3}$	$-5.99 \cdot 10^{-7}$	$1.53 \cdot 10^{-1}$
$1 \cdot 10^{-4}$	27	$2.41 \cdot 10^{5}$	$-4.74 \cdot 10^{-9}$	$1.14 \cdot 10^{-1}$
$1 \cdot 10^{-5}$	34	$3.18 \cdot 10^{7}$	$-7.18 \cdot 10^{-11}$	$6.33 \cdot 10^{-2}$

Table 4.2: Example 4.1: Baart test problem using Newton's method. The table shows the number of iterations i, the value of regularization parameter β_i, the value of the function ϕ, defined by (1.5), at β_i, and the relative error of the computed approximate solution x_{β_i}.

Example 4.1. We consider the Fredholm integral equation of the first kind,

$$\int_0^{\pi/2} \kappa(\sigma,\tau)x(\tau)d\sigma = b(\tau), \qquad 0 \leq \tau \leq \pi, \tag{4.1}$$

where $\kappa(\sigma,\tau) = \exp(\sigma\cos(\tau))$ and $b(\tau) = 2\sinh(\tau)/\tau$. This equation has been discussed by Baart [1]. It has the solution $x(\tau) = \sin(\tau)$. We use the code baart from the MATLAB package Regularization Tools by Hansen [7] to discretize (4.1) by a Galerkin method with 200 orthonormal box functions as test and trial functions. This yields the matrix $A \in \mathbb{R}^{200\times 200}$ and the right-hand side vector $\tilde{b} \in \mathbb{R}^{200}$. The matrix is nonsymmetric and extremely ill-conditioned; its condition number $\kappa(A) = \|A\|\|A^\dagger\|$ is $\kappa(A) = 5.2 \cdot 10^{18}$. We generate error vectors $e \in \mathbb{R}^{200}$ with normally distributed zero-mean random entries, normalized to yield the relative error

$$\delta = \frac{\|e\|}{\|\tilde{b}\|}$$

for $1 \cdot 10^{-5} \leq \delta \leq 1 \cdot 10^{-2}$. The contaminated right-hand side of (1.1) is obtained from (1.2).

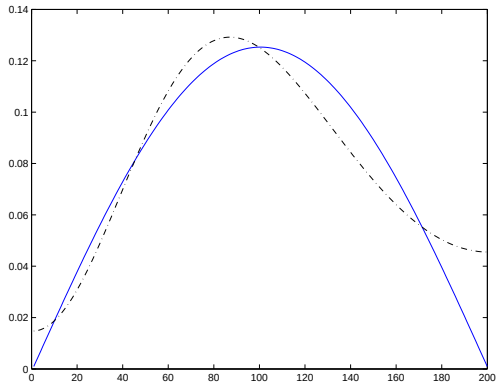

Figure 4.1: Example 4.1: Baart test problem. The exact solution $\tilde{x}$ (continuous blue graph) and the approximate solution $x_{\beta_{15}}$ (dash-dotted black graph) computed by Tikhonov regularization using the zero-finder (3.3)-(3.9) with $\delta = 1 \cdot 10^{-3}$.

We apply Tikhonov regularization to determine an approximate solution of (1.1). Table 4.1 reports the number of iterations required by the cubically convergent zero-finder defined by (3.3)-(3.9). This zero-finder gave monotonic convergence in all our computed examples, i.e., the computed iterates β_i satisfied $\beta_i < \beta_{i+1}$ for all $i \geq 0$. Therefore, there was no need to switch to the zero-finder described in Theorem 3.2.

Figure 4.1 displays the solution $x_{\beta_{15}}$ determined by our zero-finder for the relative error $\delta = 1 \cdot 10^{-3}$ (dash-dotted black graph). The figure also shows the desired solution $\tilde{x}$ of the (unknown) associated linear system of equations with error-free right-hand side $\tilde{b}$ (continuous blue graph).

Table 4.2 is analogous to Table 4.1 and displays results obtained with Newton's method. The zero-finder (3.3)-(3.9) is seen to reduce the number of iterations by about 1/3, compared with Newton's method, for all values of δ. Each iteration with the cubically convergent zero-finder (3.3)-(3.9) and with Newton's method requires the solution of two linear systems of equations with triangular matrices R and R^T, respectively, of order 200.

We remark that Tables 4.1 and 4.2 show the zero-finder (3.3)-(3.9) and Newton's method to determine different values of the β. This depends on that the computations with these schemes are terminated as soon as a β-value that satisfies (2.4) has been found. □

Example 4.2. Consider the Fredholm integral equation of the first kind,

$$\int_{-\pi/2}^{\pi/2} \kappa(\sigma,\tau)x(\tau)d\sigma = b(\tau), \qquad -\pi/2 \leq \tau \leq \pi/2, \tag{4.2}$$

with $\kappa(\sigma,\tau) = (\cos(\sigma)+\cos(\tau))^2(\sin(\xi)/\xi)^2$, $\xi = \pi(\sin(\sigma)+\sin(\tau))$, and the right-hand side chosen so that the solution x is the sum of two Gaussian functions. This integral equation is discussed by Shaw [13]. Using the MATLAB code shaw from [7], we discretize (4.2) by a quadrature rule with 200 nodes and obtain the

δ	i	β_i	$\phi(\beta_i)$	$\|x_{\beta_i} - \tilde{x}\|/\|\tilde{x}\|$
$1 \cdot 10^{-2}$	9	$9.09 \cdot 10^{1}$	$-2.64 \cdot 10^{-3}$	$1.51 \cdot 10^{-1}$
$1 \cdot 10^{-3}$	13	$3.92 \cdot 10^{3}$	$-4.09 \cdot 10^{-5}$	$5.65 \cdot 10^{-2}$
$1 \cdot 10^{-4}$	17	$5.83 \cdot 10^{4}$	$-1.12 \cdot 10^{-6}$	$4.60 \cdot 10^{-2}$
$1 \cdot 10^{-5}$	21	$2.32 \cdot 10^{6}$	$-1.09 \cdot 10^{-8}$	$3.27 \cdot 10^{-2}$

Table 4.3: Example 4.2: Shaw test problem using the zero-finder (3.3)-(3.9). The table shows the number of iterations i, the value of regularization parameter β_i, the value of the function ϕ, defined by (1.5), at β_i, and the relative error of the computed approximate solution x_{β_i}.

δ	i	β_i	$\phi(\beta_i)$	$\|x_{\beta_i} - \tilde{x}\|/\|\tilde{x}\|$
$1 \cdot 10^{-2}$	14	$8.95 \cdot 10^{1}$	$-2.14 \cdot 10^{-3}$	$1.52 \cdot 10^{-1}$
$1 \cdot 10^{-3}$	21	$5.20 \cdot 10^{3}$	$-1.25 \cdot 10^{-4}$	$5.35 \cdot 10^{-2}$
$1 \cdot 10^{-4}$	26	$4.76 \cdot 10^{4}$	$-3.61 \cdot 10^{-7}$	$4.63 \cdot 10^{-2}$
$1 \cdot 10^{-5}$	33	$2.46 \cdot 10^{6}$	$-1.24 \cdot 10^{-8}$	$3.26 \cdot 10^{-2}$

Table 4.4: Example 4.2: Shaw test problem using Newton's method. The table shows the number of iterations i, the value of regularization parameter β_i, the value of the function ϕ, defined by (1.5), at β_i, and the relative error of the computed approximate solution x_{β_i}.

matrix $A \in \mathbb{R}^{200\times 200}$ with condition number $\kappa(A) = 5.5 \cdot 10^{19}$ and the right-hand side $\tilde{b} \in \mathbb{R}^{200}$. The contaminated right-hand side b of (1.1) is constructed in the same way as in Example 4.1.

We determine approximate solutions of (1.1) for different relative errors in the right-hand side by Tikhonov regularization. Table 4.3 reports the number of iterations required by the cubically convergent zero-finder (3.3)-(3.9), which gave monotonic convergence.

Figure 4.2 displays the approximate solution $x_{\beta_{14}}$ determined by this zero-finder for the relative error $\delta = 1 \cdot 10^{-3}$ (dash-dotted black graph), as well as the desired solution $\tilde{x}$ of the (unknown) linear system of equations with error-free right-hand side (continuous blue graph).

Table 4.4 is analogous to Table 4.3 and shows results for Newton's method. The zero-finder (3.3)-(3.9) can be seen to reduce the number of iterations by about 1/3 for all values of δ. Similarly as in Example 4.1, the zero-finder (3.3)-(3.9) and Newton's method determine slightly different values of β, due to that the computations are terminated as soon as a value of β that satisfies (2.4) has been found. □

We remark that the performances of both our zero-finder and Newton's method depend on the choice of the initial value $\beta_0 = 0$ of the parameter β. For positive β_0, say $\beta_0 = 1/10$, both our and the Newton methods require fewer iterations. However, the use of a positive β_0 may require the solution of an underregularized problem. Since we would like to avoid this, we have not pursued a comparison of zero-finders for $\beta_0 > 0$.

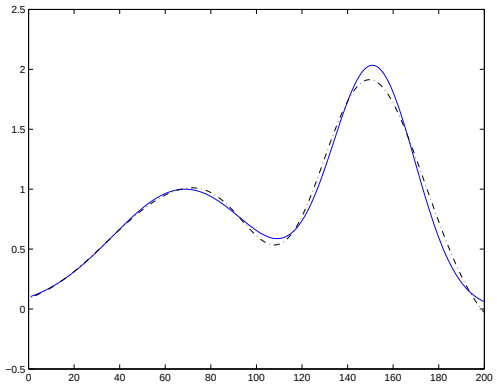

Figure 4.2: Example 4.2: Shaw test problem. The exact solution $\tilde{x}$ (continuous blue graph) and the approximate solution $x_{\beta_{14}}$ (dash-dotted graph) computed by Tikhonov regularization using the zero-finder (3.3)-(3.9) with $\delta = 1 \cdot 10^{-3}$.

5 Conclusion

This paper describes new cubically convergent zero-finders, such that each evaluation of a new iterate costs essentially the same as the computation of a new iterate by the quadratically convergent Newton's method. Numerical examples illustrate that the cubically convergent zero-finder (3.3)-(3.9) reduces the cost for the iterations required to determine a suitable value of the regularization parameter by about 1/3. In our computational experience this zero-finder always gave monotonic convergence. We therefore do not expect the cubically convergent zero-finder discussed in Theorems 3.2 and 3.3 to be of frequent use.

Acknowledgment. We would like to thank Per Christian Hansen for comments.

REFERENCES

1. M. L. Baart, *The use of auto-correlation for pseudo-rank determination in noisy ill-conditioned least-squares problems*, IMA J. Numer. Anal., 2 (1982), pp. 241–247.
2. D. Calvetti and L. Reichel, *Tikhonov regularization of large linear problems*, BIT, 43 (2003), pp. 263–283.
3. J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization*, SIAM, Philadelphia, 1996.
4. L. Eldén, *Algorithms for the regularization of ill-conditioned least squares problems*, BIT, 17 (1977), pp. 134–145.
5. H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
6. W. Gander, *On Halley's iteration method*, Amer. Math. Monthly, 92 (1985), pp. 131–134.

7. P. C. Hansen, *Regularization tools: A MATLAB package for analysis and solution of discrete ill-posed problems*, Numer. Algor., 6 (1994), pp. 1–35. Software is available in Netlib at http://www.netlib.org.
8. P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.
9. C. W. Groetsch, *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Pitman, Boston, 1984.
10. J. J. Moré and D. C. Sorensen, *Computing a trust-region step*, SIAM J. Sci. Stat. Comput., 4 (1983), pp. 553–572.
11. V. A. Morozov, *Methods for Solving Incorrectly Posed Problems*, Springer, New York, 1984.
12. C. H. Reinsch, *Smoothing by spline functions II*, Numer. Math., 16 (1971), pp. 451–454.
13. C. B. Shaw, Jr., *Improvements of the resolution of an instrument by numerical solution of an integral equation*, J. Math. Anal. Appl., 37 (1972), pp. 83–112.
14. J. F. Traub, *Iterative Methods for the Solution of Equations*, Prentice-Hall, Englewood Cliffs, 1964.