Inverse Problems

PAPER • OPEN ACCESS

A conjugate-gradient-type rational Krylov subspace method for ill-posed problems

To cite this article: Volker Grimm 2020 *Inverse Problems* **36** 015008

OPEN ACCESS
IOP Publishing Inverse Problems

Inverse Problems **36** (2020) 015008 (19pp) https://doi.org/10.1088/1361-6420/ab5819

A conjugate-gradient-type rational Krylov subspace method for ill-posed problems

Volker Grimm

Institute for Applied and Numerical Mathematics, Karlsruhe Institute of Technology, D–76128 Karlsruhe, Germany

E-mail: volker.grimm@kit.edu

Received 23 May 2019, revised 30 October 2019
Accepted for publication 15 November 2019
Published 19 December 2019

Abstract

Conjugated gradients on the normal equation (CGNE) is a popular method to regularise linear inverse problems. The idea of the method can be summarised as minimising the residuum over a suitable Krylov subspace. It is shown that using the same idea for the shift-and-invert rational Krylov subspace yields an order-optimal regularisation scheme.

Keywords: order-optimal regularisation scheme, rational Krylov subspace method, discrepancy principle

(Some figures may appear in colour only in the online journal)

1. Introduction

We consider the solution of the linear system

$$Tx = y^\delta \tag{1}$$

where the operator T acts continuously between the Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$. The linear system is assumed to be ill-posed, that is, the range $\mathcal{R}(T)$ is not closed in $\mathcal{Y}$. y^δ is a perturbation of the exact data y, such that $\|y^\delta - y\| \leqslant \delta$. y^δ is also called the *noisy data* and δ the *noise level*. For exact data, we assume that y is in the range of T, $y \in \mathcal{R}(T)$, which guarantees that there exists a unique $x^+ \in \mathcal{N}(T)^\perp$ such that $Tx^+ = y$. $\mathcal{N}(T)^\perp$ designates the orthogonal complement of the null space $\mathcal{N}(T)$ of T. x^+ can also be characterised as the unique $x^+ \in \mathcal{N}(T)^\perp$ that solves the *normal equation*

$$T^*Tx = T^*y. \tag{2}$$

In fact, the normal equation possesses a unique solution $x^+ \in \mathcal{N}(T)^\perp$ for every $y \in \mathcal{D}(T^+) := \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$, where $\mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$ designates the direct orthogonal sum of $\mathcal{R}(T)$ and its orthogonal complement $\mathcal{R}(T)^\perp = \mathcal{N}(T^*)$. The linear unbounded map $T^+ : \mathcal{D}(T^+) \to \mathcal{N}(T)^\perp$, $y \mapsto x^+$, is the *Moore–Penrose (generalised) inverse* and x^+ is the *minimum-norm* solution.

In order to reconstruct the solution x^+ of the unperturbed problem $Tx^+ = y$ as good as possible subject to a given noise level δ, special procedures, called *regularisation schemes*, have to be used. Let $\{R_m\}_{m\in\mathbb{N}_0}$ be a family of linear or nonlinear operators from $\mathcal{Y}$ to $\mathcal{X}$ with $R_m 0 = 0$. If there exists a mapping $m : \mathbb{R}^+ \times \mathcal{Y} \to \mathbb{N}_0$ such that

$$\lim_{\delta\to 0} \sup\{\|R_{m(\delta,y^\delta)} y^\delta - x^+\| \mid y^\delta \in \mathcal{Y}, \|y^\delta - Tx^+\| \leqslant \delta\} = 0$$

for any $x^+ \in \mathcal{N}(T)^\perp$, then the pair $(R_m, m(\delta, y^\delta))$ is called a (convergent) regularisation scheme for T. The mapping m is called *parameter choice* or *stopping rule*. We will always use the *discrepancy principle* as our stopping rule, which is due to Morozov [23]. The *discrepancy principle* reads: Choose a fixed $\tau > 1$ and set:

$$m(\delta, y^\delta) := \min\{m \in \mathbb{N}_0 \mid \|y^\delta - Tx_m^\delta\| \leqslant \tau\delta\}, \tag{3}$$

where $x_m^\delta := R_m y^\delta$. The discrepancy principle leads to convergent regularisation schemes (see [6]). Regularisation schemes might converge arbitrarily slow unless the (unperturbed) data x^+ satisfies some smoothness assumptions. Convergence rates can be given when x^+ is in the *source set* $\mathcal{X}_{\mu,\rho} := \{x \in \mathcal{X} \mid x = (T^*T)^\mu w,\ \|w\| \leqslant \rho\}$, $\mu > 0$. Regularisation schemes $(R_m, m(\delta, y^\delta))$ that attain the highest possible convergence speed are called of *optimal order* in $\mathcal{X}_{\mu,\rho}$ if

$$\sup\{\|R_{m(\delta,y^\delta)} y^\delta - x^+\| \mid x^+ \in \mathcal{X}_{\mu,\rho}, \|y^\delta - Tx^+\| \leqslant \delta\} \leqslant C_\mu \delta^{\frac{2\mu}{2\mu+1}} \rho^{\frac{1}{2\mu+1}},$$

where C_μ neither depends on δ nor on ρ.

One of the most popular iterative regularisation schemes is conjugated gradients on the normal equation (CGNE) that can be stated briefly as

$$x_m^\delta =: R_m y^\delta, \qquad x_m^\delta = \mathrm{argmin}_{x\in\mathcal{K}_m} \|y^\delta - Tx\|, \qquad m = 1, 2, \dots, \tag{4}$$

where $\mathcal{K}_m$ is the (polynomial) Krylov subspace

$$\mathcal{K}_m = \mathcal{K}_m(T^*T, T^*y^\delta) = \mathrm{span}\{T^*y^\delta, (T^*T)T^*y^\delta, \dots, (T^*T)^{m-1}T^*y^\delta\}.$$

An efficient algorithm is available to compute these approximations (see [15]). CGNE with the discrepancy principle as a stopping rule is an order-optimal regularisation scheme for all $\mu > 0$ (see theorem 7.12 in [6, 24]). And, due to its definition, CGNE is the fastest to satisfy the discrepancy principle with respect to all regularisation schemes that compute approximations in the Krylov subspace $\mathcal{K}_m$. The analysis of CGNE with respect to its regularisation properties is involved, since the operators R_m are nonlinear and not necessarily continuous (see theorem 7.6 in [5, 6]).

In this paper, we will define a method of the same type, but for the *shift-and-invert* or *resolvent* Krylov subspace

$$\begin{aligned} \mathcal{Q}_m &= \mathcal{K}_m\left((I + T^*T/\gamma)^{-1}, T^*y^\delta\right) \\ &= \mathrm{span}\left\{T^*y^\delta, (I + T^*T/\gamma)^{-1} T^*y^\delta, \cdots, (I + T^*T/\gamma)^{-m+1} T^*y^\delta\right\}, \end{aligned} \tag{5}$$

where $\gamma > 0$ is a fixed real number (e.g. [1, 11, 12, 20–22, 26, 29]). Due to the relation $(I + T^*T/\gamma)^{-1}T^*T = \gamma I - \gamma(I + T^*T/\gamma)^{-1}$, this rational Krylov subspace can also be written as

$$\mathcal{Q}_m = \mathcal{K}_m\left((I + T^*T/\gamma)^{-1}, T^*y^\delta\right) = \mathcal{K}_m\left((I + T^*T/\gamma)^{-1} T^*T, T^*y^\delta\right). \quad (6)$$

We define our method by

$$x_m^\delta =: R_m y^\delta, \qquad x_m^\delta = \operatorname{argmin}_{x \in \mathcal{Q}_m} \|y^\delta - Tx\|, \qquad m = 1, 2, \ldots, \quad (7)$$

combined with the discrepancy principle as its stopping rule. (The minimizer x_m^δ is uniquely defined, see lemma 2.3 below.) The subspace $\mathcal{Q}_m$ belongs to the class of rational Krylov subspaces which have been studied in recent years (e.g. references in [8, 13]). Since this method can be seen as solving the normal equation (2) approximatively in the shift-and-invert Krylov subspace $\mathcal{Q}_m$, the method will be called shift-and-invert on the normal equation (SINE). Several regularisation schemes have been proposed that compute approximations in the subspace $\mathcal{Q}_m$, see example 1.1. By definition, our method will be the fastest to stop with respect to the discrepancy principle. SINE is related (but not identical) to CGNE preconditioned by $(I + T^*T/\gamma)^{-1}$. Actually, SINE is not a preconditioning technique in the usual sense. Nevertheless, rational Krylov subspaces have been observed of being capable of accelerating the convergence (e.g. [10]). The analysis of SINE with respect to its regularisation properties shares the difficulties of the analysis of CGNE, the family of operators R_m is again nonlinear and not continuous in general, which can be seen by generalising the ideas of the proof of theorem 7.6 in [5, 6].

Example 1.1. Some regularisation schemes with approximations in the subspace $\mathcal{Q}_m$.

(i) Iterated Tikhonov–Phillips regularisation (see [3, 7, 17, 18]).
(ii) Applying the implicit Euler method, the implicit midpoint-rule, or the trapezoidal rule with fixed time-step to asymptotic regularisation (Showalter's regularisation) leads to approximations in $\mathcal{Q}_m$ (see [27]).
(iii) The method proposed by Riley in [28] applied to the normal equation.
(iv) The rational Arnoldi approach proposed in [2].

In section 2, we will show that x_m^δ in (7) can be computed efficiently and discuss some basic properties of the method. Convergence for unperturbed data is shown in section 3 before the SINE method is discussed with respect to its regularisation properties in section 4. In section 5, upper bounds on the number of iterations of SINE are discussed by comparing them to the known upper bounds on the number of iterations of CGNE. The findings are illustrated by an experiment in section 6.

Throughout we will use notations identical or closely related to the notations in [6]. In particular, the functional calculus described in section 2.3 of [6] is used without further note. Our proofs will follow closely or sometimes literally the corresponding proofs for CGNE in chapter 7 of [6].

2. Basic properties

We consider algorithm 1, where we choose $x_0^\delta = 0$ without loss of generality. If x_0^δ were not zero, the corresponding shift-and-invert Krylov subspace would be spanned with $r_0 = y^\delta - Tx_0^\delta$ instead of y^δ, which allows to use prior information on the solution. First, we aim for the following properties: algorithm 1 computes x_m^δ according to (7), as long as the algorithm does not break down. If algorithm 1 breaks down in step κ with $q_\kappa = 0$, we have $x^\delta_\kappa = T^+ y^\delta$ as well as $x^\delta_m = T^+ y^\delta$ for $m \geqslant \kappa$ in (7).

Algorithm 1. Shift-and-invert on the normal equation (SINE).

```
Choose $x^\delta_0$, set $r_0 = y^\delta - Tx^\delta_0$, $w_0 = T^* r_0$.
for $j = 0, 1, 2, \ldots$ do
    $q_j = Tw_j$
    $\delta_j = (q_j, q_j)$
    $\alpha_j = (r_j, q_j)/\delta_j$
    $x^\delta_{j+1} = x^\delta_j + \alpha_j w_j$
    $r_{j+1} = r_j - \alpha_j q_j$
    $s_j = T^* q_j$
    $t_{j+1} = (I + T^*T/\gamma)^{-1} T^* r_{j+1}$
    $\beta_j = (t_{j+1}, s_j)/\delta_j$
    $w_{j+1} = t_{j+1} - \beta_j w_j$
end for
```

An alternative way to compute the iterates of SINE is presented in algorithm 2. This variant is closer to the standard CG iteration and it might therefore help to compare SINE to known methods. All proofs and experiments refer to algorithm 1.

Algorithm 2. Variant of shift-and-invert on the normal equation (SINE).

```
Choose $x_0$, set $r_0 = y^\delta - Tx_0$, $t_0 = d_0 = w_0 = T^* r_0$,
for $j = 0, 1, 2, \ldots$ do
    $q_j = Tw_j$
    $\delta_j = (q_j, q_j)$
    $\alpha_j = (t_j, d_j)/\delta_j$
    $x_{j+1} = x_j + \alpha_j w_j$
    $r_{j+1} = r_j - \alpha_j q_j$
    $d_{j+1} = T^* r_{j+1}$
    $t_{j+1} = (I + T^*T/\gamma)^{-1} d_{j+1}$
    $\tilde\beta_j = (t_{j+1}, d_{j+1})/(t_j, d_j)$
    IF $j = 0$ , $\tilde\beta_j = \tilde\beta_j - (t_{j+1}, d_j)/(t_j, d_j)$
    $w_{j+1} = t_{j+1} + \tilde\beta_j w_j$
end for
```

Lemma 2.1. *As long as* $q_{m-1} \neq 0$

(i) $(r_m, q_j) = (T^* r_m, w_j) = 0,\ j = 0, \ldots, m-1$

(ii) $(q_m, q_j) = 0,\ j = 0, \ldots, m-1$

Proof. The proof is via induction on m. For $m = 1$, we have

$$(r_1, q_0) = (r_0, q_0) - \alpha_0 (q_0, q_0) = 0, \qquad \alpha_0 = \frac{(r_0, q_0)}{(q_0, q_0)}.$$

We further obtain

$$\begin{aligned}(q_1, q_0) &= (Tw_1, Tw_0) = (Tt_1 - \beta_0 Tw_0, Tw_0)\\ &= (Tt_1, Tw_0) - \beta_0 (Tw_0, Tw_0), \qquad \beta_0 = \frac{(t_1, s_0)}{(q_0, q_0)}\\ &= (Tt_1, q_0) - (t_1, T^* q_0) = 0,\end{aligned}$$

which concludes the proof of our statements for $m = 1$. Now we assume that the assertions are satisfied for m. Then, we have

$$\begin{aligned}(r_{m+1}, q_j) &= (r_m, q_j) - \alpha_m (q_m, q_j), \qquad \alpha_m = \frac{(r_m, q_m)}{(q_m, q_m)}\\ &= \begin{cases}(r_m, q_m) - \alpha_m (q_m, q_m) = 0, & j = m,\\ 0, & j < m\end{cases}.\end{aligned}$$

With respect to (ii), it follows

$$\begin{aligned}(q_{m+1}, q_m) &= (Tt_{m+1} - \beta_m Tw_m, q_m)\\ &= (Tt_{m+1}, q_m) - \beta_m (q_m, q_m), \qquad \beta_m = \frac{(t_{m+1}, T^* q_m)}{(q_m, q_m)}\\ &= 0\end{aligned}$$

and for $j < m$, we have

$$\begin{aligned}(q_{m+1}, q_j) &= (Tt_{m+1}, Tw_j) - \beta_m (q_m, q_j)\\ &= (Tt_{m+1}, Tw_j) = (T(I + T^*T/\gamma)^{-1} T^* r_{m+1}, Tw_j)\\ &= (T^* r_{m+1}, (I + T^*T/\gamma)^{-1} T^* q_j) = 0,\end{aligned} \tag{8}$$

since

$$(I + T^*T/\gamma)^{-1} T^* q_j = (I + T^*T/\gamma)^{-1} T^* T w_j \in \mathcal{Q}_{j+2}$$

due to (6) and the fact that assertion (i) is already proved for $m + 1$. □

The following lemma is a direct consequence of lemma 2.1.

Lemma 2.2. *As long as algorithm 1 does not break down with $q_m = 0$, we have*

$$\mathcal{Q}_{m+1} = \operatorname{span}\{w_0, \ldots, w_m\},$$

where $w_0, \ldots, w_m$ is a basis of $\mathcal{Q}_{m+1}$.

With these preparations, the statements we aimed for can be proved.

Lemma 2.3. *The iterates x_m^δ of algorithm 1 satisfy (7).*

Proof. Due to algorithm 1 with $x_0^\delta = 0$, we have $x_m^\delta \in \operatorname{span}\{w_0, \ldots, w_{m-1}\} = \mathcal{Q}_m$. Now, let $z_m \in \mathcal{Q}_m = \operatorname{span}\{w_0, \ldots, w_{m-1}\}$ such that $z_m \neq x_m^\delta \in \mathcal{Q}_m$. Hence, we can write

$$0 \neq z_m - x_m^\delta = \sum_{j=0}^{m-1} \xi_j w_j, \qquad \xi_j \in \mathbb{R},$$

and obtain

$$\begin{aligned}\|y^\delta - Tz_m\|^2 &= \|y^\delta - Tx_m^\delta\|^2 - 2\sum_{j=0}^{m-1}\xi_j(Tw_j, r_m) + \|T\sum_{j=0}^{m-1}\xi_j w_j\|^2\\ &> \|y^\delta - Tx_m^\delta\|^2 - 2\sum_{j=0}^{m-1}\xi_j(q_j, r_m) = \|y^\delta - Tx_m^\delta\|^2\end{aligned}$$

by lemma 2.1. The strict inequality, and therefore uniqueness of the minimizer, follows since $\mathcal{Q}_m \subseteq \mathcal{N}(T)^\perp$ and $\|T\sum_{j=0}^{m-1}\xi_j w_j\|^2 > 0$ as a consequence. □

Lemma 2.4. *If algorithm 1 breaks down in step κ with $q_\kappa = 0$, then $x_\kappa^\delta = x^+ = T^+ y^\delta$.*

Proof. We first show that $q_\kappa = 0$ means $T^* r_\kappa = 0$. First assume $\kappa = 0$. Then $0 = (r_0, q_0) = (r_0, TT^* r_0) = \|T^* r_0\|^2$, hence $T^* r_0 = 0$. Now let $\kappa > 0$ and assume $(r_\kappa, q_\kappa) = 0$. Then, with the help of statement (i) of lemma 2.1

$$\begin{aligned}0 &= (T^* r_\kappa, w_\kappa) = (T^* r_\kappa, t_\kappa - \beta_{\kappa-1} w_{\kappa-1}) = (T^* r_\kappa, t_\kappa)\\ &= ((I + T^*T/\gamma)t_\kappa, t_\kappa) = \|t_\kappa\|^2 + \frac{1}{\gamma}\|Tt_\kappa\|^2.\end{aligned}$$

Hence we have $0 = (I + T^*T/\gamma)t_\kappa = T^* r_\kappa$ in all cases. Since $r_\kappa = y^\delta - Tx_k^\delta$, this means

$$T^*Tx_\kappa^\delta = T^* y^\delta, \qquad x_\kappa^\delta \in \mathcal{Q}_\kappa \subseteq \mathcal{N}(T)^\perp,$$

which characterises the minimum-norm solution, that is $x_\kappa^\delta = x^+$ (see theorems 2.5 and 2.6 in [6]). □

If the algorithm stops with $q_\kappa = 0$, it follows from (7) that $x_m^\delta = x_\kappa^\delta$, $m \geqslant \kappa$.

Analogous to the description of the Krylov subspace $\mathcal{K}_m$ with the set Π_{m-1} of polynomials of degree less than m, the shift-and-invert Krylov subspace $\mathcal{Q}_m$ can be described with the help of rational functions as

$$\mathcal{Q}_m = \left\{ r(T^*T)T^* y^\delta \mid r \in \Pi_{m-1}/(1 + \cdot/\gamma)^{m-1} \right\}.$$

The functional calculus of section 2.3 in [6] applies to these rational functions. The iterates, residui etc of algorithm 1 can be identified with the corresponding rational functions (see [6, 14]). The following lemma describes some properties of the rational function $r_m(\lambda)$ that belongs to the residuum r_m, i.e. the function $r_m(\lambda)$ such that

$$r_m = y^\delta - Tx_m = r_m(TT^*)y^\delta \qquad \text{or} \qquad T^* r_m = r_m(T^*T)T^* y^\delta, \tag{9}$$

respectively.

Lemma 2.5. *As long as the stopping index κ has not been reached, we have*

$$r_m(\lambda) = \frac{p_m(\lambda)}{(1 + \lambda/\gamma)^{m-1}} \quad \text{with} \quad p_m(\lambda) = \prod_{j=1}^{m}\left(1 - \frac{\lambda}{\lambda_{j,m}}\right), \quad m \geqslant 1, \tag{10}$$

where $r_m(\lambda)$ is the rational function that describes the residuum r_m in algorithm 1. The values $\lambda_{j,m-1}$, $j = 1, \ldots, m-1$ of $r_{m-1}(\lambda)$ and the values $\lambda_{j,m}$, $j = 1, \ldots, m$ of $r_m(\lambda)$ are interlacing, real, and positive, that is

$$0 < \lambda_{1,m} < \lambda_{1,m-1} < \lambda_{2,m} < \cdots < \lambda_{m-1,m} < \lambda_{m,m-1} < \lambda_{m,m} \leqslant \|T\|^2.$$

Proof. Let $v_0, \cdots, v_{m-1}$ be an orthonormal basis such that

$$\mathcal{Q}_\ell = \operatorname{span}\{w_0, \cdots, w_{\ell-1}\} = \operatorname{span}\{v_0, \cdots, v_{\ell-1}\} \qquad \text{for} \quad \ell = 1, \dots, m.$$

Then we can represent the iterate x_ℓ^δ as $x_\ell^\delta = \sum_{j=0}^{\ell-1} z_{j,\ell} v_j$. Since $T^*y^\delta - T^*Tx_\ell^\delta \perp \mathcal{Q}_\ell$ is an equivalent condition to x_ℓ^δ being the minimizer in (7), we have

$$S_\ell = ((T^*Tv_i, v_j))_{j,i=0}^{\ell-1}, \qquad S_\ell z_\ell = \beta e_1, \qquad \beta = \|T^*y^\delta\|.$$

Since $\mathcal{Q}_\ell \subset \mathcal{N}(T)^\perp = \mathcal{N}(T^*T)^\perp$, S_ℓ is invertible, and therefore symmetric positive definite with $\|S_\ell\| \leqslant \|T\|^2$ and $z_\ell = \beta S_\ell^{-1} e_1$. Specifically,

$$x_m^\delta = \sum_{j=0}^{m-1} z_{j,m} v_j, \qquad \text{with} \qquad z_m = \beta S_m^{-1} e_1. \tag{11}$$

Furthermore, the $(\ell-1, \ell-1)$ submatrix of S_ℓ is $S_{\ell-1}$ for $\ell = 2, \dots, m$. Inductively, by the interlacing eigenvalue theorem (see theorem 3.6 in [30]), this leads to the result that the eigenvalues of S_ℓ are separated and interlaced with the eigenvalues of $S_{\ell-1}$. If we designate the eigenvalues of S_ℓ with $\lambda_{1,\ell} < \cdots < \lambda_{\ell,\ell}$ then we obtain the statement on the interlacing of the numbers in our theorem. Using the representation of the iterate in the rational Krylov subspace (see [9]), one obtains

$$x_m^\delta = \frac{q_{m-1}(T^*T)}{(1 + T^*T/\gamma)^{m-1}} T^*y^\delta = \sum_{j=0}^{m-1} \xi_{j,m} v_j, \quad \xi_m = \beta \frac{q_{m-1}(S_m)}{(1 + S_m/\gamma)^{m-1}} e_1. \tag{12}$$

By comparing (11) with (12), we obtain

$$\beta \frac{q_{m-1}(S_m)}{(1 + S_m/\gamma)^{m-1}} e_1 = \beta S_m^{-1} e_1 \qquad \text{and hence} \qquad \frac{q_{m-1}(S_m)}{(1 + S_m/\gamma)^{m-1}} = S_m^{-1},$$

since S_m comes from a Krylov process and the minimal polynomial of S_m with respect to e_1 has therefore degree m. This means that $r_m(\lambda) = 1 - \lambda q_{m-1}(\lambda)/(1 + \lambda/\gamma)^{m-1}$ has zeros $\lambda_{j,m}$, $j = 1, \dots, m$ by spectral decomposition of S_m. Together with the obvious value $r_m(0) = 1$, this shows the representation of $r_m(\lambda)$ as given in our lemma. □

The inner product introduced in the following lemma will be crucial for the proof of our main theorem. Also, the idea of algorithm 1 can be briefly stated as computing an orthogonal basis $w_0, \dots, w_{j-1}$ of the rational Krylov subspace $\mathcal{Q}_j$ with respect to the inner product $[\cdot, \cdot]$ in (13) when the vectors are identified with the corresponding rational functions. In the following, we will often not make a difference between the residuum r_m and the rational function $r_m(\lambda)$ and denote both by r_m.

Lemma 2.6. *The rational functions r_m generated by algorithm 1 are orthogonal to $\Pi_{m-1}/(1 + \cdot/\gamma)^{m-1}$ with respect to the inner product*

$$[\varphi, \psi] = \int_0^{\|T\|^2+} \varphi(\lambda)\psi(\lambda)\lambda\, d\|F_\lambda y^\delta\|^2, \tag{13}$$

where F_λ designates the spectral family of TT^*. Among all rational $\varphi \in \Pi_m/(1+\cdot/\gamma)^{m-1}$ with $\varphi(0) = 1$, r_m minimises the functional

$$\Phi[\varphi] = \int_0^{\|T\|^2+} \varphi^2(\lambda)\, d\|F_\lambda y^\delta\|^2. \tag{14}$$

Proof. We have

$$[\varphi, \psi] = \int_0^{\|T\|^2+} \varphi(\lambda)\psi(\lambda)\lambda\, d\|F_\lambda y^\delta\|^2 = (\varphi(T^*T)T^*y^\delta, \psi(T^*T)T^*y^\delta)$$

which gives the first assertion by lemma 2.1 (i). The second assertion follows by lemma 2.3. □

3. Convergence

The following theorem shows convergence of the iterates x_m in algorithm 1 to the minimum-norm solution $x^+ = T^+y$ for data $y \in \mathcal{D}(T^+)$ and our general choice $x_0 = 0$. For an initial guess $x_0 \neq 0$, it can be readily shown that the iterates converge to $T^+y + P_{\mathcal{N}(T)}x_0$, where $P_{\mathcal{N}(T)}$ is the orthogonal projector to the null space of T. The superscript δ has been dropped in this section in order to emphasize that data $y \in \mathcal{D}(T^+)$ without perturbation is considered.

Theorem 3.1. *The sequence of SINE iterates $\{x_m\}$ converges to T^+y for all $y \in \mathcal{D}(T^+)$.*

Proof. We basically follow the lines of the proof of theorem 7.9 in [6] or [25], respectively. If the iteration terminates after a finite number of steps then the corresponding iterate coincides with T^+y according to lemma 2.4. We therefore assume, that the iteration does not terminate. Then we have the sorting

$$0 < \lambda_{1,m} < \lambda_{2,m} < \cdots < \lambda_{m,m} \leqslant \|T\|^2$$

of the Ritz values according to lemma 2.5. From the representation of the residual rational function (10) in lemma 2.5, we obtain

$$|r_m'(0)| = -r_m'(0) = \sum_{j=1}^m \frac{1}{\lambda_{j,m}} + \frac{m-1}{\gamma}. \tag{15}$$

Since $r_m/(\lambda - \lambda_{1,m})$ is in the space $\Pi_{m-1}/(1+\cdot/\gamma)^{m-1}$, the orthogonality relation (13) yields

$$0 = \int_0^{\|T\|^2+} r_m(\lambda)\frac{r_m(\lambda)}{\lambda - \lambda_{1,m}}\lambda\, d\|F_\lambda y\|^2$$

which gives

$$\int_0^{\lambda_{1,m}} r_m^2(\lambda)\frac{\lambda}{\lambda_{1,m}-\lambda}\,d\|F_\lambda y\|^2 = \int_{\lambda_{1,m}}^{\|T\|^2+} r_m^2(\lambda)\frac{\lambda}{\lambda-\lambda_{1,m}}\,d\|F_\lambda y\|^2.$$

Since $\lambda/(\lambda-\lambda_{1,m}) \geqslant 1$ for $\lambda \geqslant \lambda_{1,m}$ we obtain

$$\int_0^{\lambda_{1,m}} r_m^2(\lambda)\frac{\lambda}{\lambda_{1,m}-\lambda}\,d\|F_\lambda y\|^2 \geqslant \int_{\lambda_{1,m}}^{\|T\|^2+} r_m^2(\lambda)\,d\|F_\lambda y\|^2.$$

And therefore,

$$\begin{aligned}\|y-Tx_m\|^2 &= \int_0^{\lambda_{1,m}} r_m^2(\lambda)\,d\|F_\lambda y\|^2 + \int_{\lambda_{1,m}}^{\|T\|^2+} r_m^2(\lambda)\,d\|F_\lambda y\|^2\\ &\leqslant \int_0^{\lambda_{1,m}} r_m^2(\lambda)\Big(1+\frac{\lambda}{\lambda_{1,m}-\lambda}\Big)\,d\|F_\lambda y\|^2.\end{aligned}$$

Defining

$$\varphi_m(\lambda) := r_m(\lambda)\Big(\frac{\lambda_{1,m}}{\lambda_{1,m}-\lambda}\Big)^{\frac{1}{2}}, \qquad 0 \leqslant \lambda \leqslant \lambda_{1,m},$$

we obtain the estimate

$$\|y-Tx_m\| \leqslant \|F_{\lambda_{1,m}}\varphi_m(TT^*)y\| \leqslant \max_{0\leqslant\lambda\leqslant\lambda_{1,m}} \sqrt{\lambda\varphi_m^2(\lambda)} \quad \|E_{\lambda_{1,m}}x^+\|, \tag{16}$$

where E_λ designates the spectral family of T^*T. For the last inequality, we additionally assumed $y \in \mathcal{R}(T)$, that is, $y = Tx^+$. For later use (e.g. lemma 4.1), we discuss the slightly more general function $\lambda^\nu\varphi_m^2(\lambda)$, $\nu > 0$. Standard calculations lead to

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}\lambda}\lambda^\nu\varphi_m^2(\lambda)\\ &= \lambda^{\nu-1}\varphi_m^2(\lambda)\cdot\left[\nu+\lambda\left(\frac{1}{\lambda_{1,m}-\lambda}-\sum_{j=1}^m\frac{2}{\lambda_{j,m}-\lambda}-2\cdot\frac{m-1}{\gamma}\cdot\frac{1}{1+\lambda/\gamma}\right)\right].\end{aligned}$$

Since $0^\nu\varphi_m^2(0) = 0 = \lambda_{1,m}^\nu\varphi_m^2(\lambda_{m,1})$, there is at least one $0 < \lambda^* < \lambda_{m,1}$, such that $(\lambda^\nu\varphi_m^2(\lambda))'(\lambda^*) = 0$ and such that the maximum is achieved at this point. Hence the equation

$$\nu = \lambda^*\left(\sum_{j=1}^m\frac{2}{\lambda_{j,m}-\lambda^*}-\frac{1}{\lambda_{1,m}-\lambda^*}+2\cdot\frac{m-1}{\gamma}\cdot\frac{1}{1+\lambda^*/\gamma}\right) \tag{17}$$

holds true. We need to distinguish two cases.

First case: $\gamma \geqslant \|T\|^2$. Then, we have $0 < \lambda^* < \lambda_{1,m} \leqslant \|T\|^2 \leqslant \gamma$. Since $\lambda^* < \gamma$, we have

$$2\cdot\frac{m-1}{\gamma}\cdot\frac{1}{1+\lambda^*/\gamma} \geqslant \frac{m-1}{\gamma}$$

and hence, by (17),

$$\nu \geqslant \lambda^* \left(\sum_{j=1}^{m} \frac{1}{\lambda_{j,m}} + \frac{m-1}{\gamma} \right) = \lambda^* \left(-r_m'(0)\right)$$

and therefore

$$\lambda^* \leqslant \frac{\nu}{-r_m'(0)} = \frac{\nu}{|r_m'(0)|}.$$

Second case: $\gamma < \|T\|^2$. We set

$$p = \frac{1}{2}\left(\frac{\|T\|^2}{\gamma} + 1\right) > 1.$$

We then have

$$2 \cdot \frac{m-1}{\gamma} \cdot \frac{1}{1+\lambda^*/\gamma} \geqslant \frac{1}{p} \cdot \frac{m-1}{\gamma}.$$

We can therefore conclude, by (17),

$$\nu \geqslant \lambda^* \cdot \frac{1}{p} \cdot \left(\sum_{j=1}^{m} \frac{1}{\lambda_{j,m}} + \frac{m-1}{\gamma} \right) = \lambda^* \cdot \frac{1}{p} \cdot \left(-r_m'(0)\right),$$

hence we have

$$\lambda^* \leqslant \frac{\nu p}{-r_m'(0)} = \frac{\nu p}{|r_m'(0)|} \qquad \text{with} \qquad p = \frac{1}{2}\left(\frac{\|T\|^2}{\gamma} + 1\right).$$

In both cases, we have

$$\lambda^* \leqslant c \cdot \frac{\nu}{|r_m'(0)|}, \qquad c = \max\{1, p\}.$$

Hence

$$\sup_{0 \leqslant \lambda \leqslant \lambda_{1,m}} \lambda^\nu \varphi_m^2(\lambda) \leqslant (\lambda^*)^\nu \varphi_m^2(\lambda^*) \leqslant (\lambda^*)^\nu \leqslant c^\nu \nu^\nu |r_m'(0)|^{-\nu}, \quad \nu > 0. \tag{18}$$

We now relax the assumption on y to $y \in \mathcal{D}(T^+) = \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$. Since $\mathcal{R}(T)^\perp = \mathcal{N}(T^*)$, algorithm 1 produces the same iterates for $y \in \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$ and $P_{\overline{R(T)}} y \in \mathcal{R}(T)$, respectively. Rewriting $P_{\overline{R(T)}} y = Tx^+$ with $x^+ = T^+ y$, we can apply (18) with $\nu = 1$ and obtain

$$\|P_{\overline{R(T)}} y - Tx_m\|^2 \leqslant \|F_{\lambda_{1,m}} \varphi_m(TT^*) Tx^+\|^2 \leqslant c |r_m'(0)|^{-1} \|E_{\lambda_{1,m}} x^+\|^2.$$

From here, a nearly literal copy of the corresponding part of the proof of theorem 7.9 in [6] will finish the proof. □

4. SINE is an order-optimal regularisation method

We assume that

$$y \in \mathcal{R}(T), \qquad \|y^\delta - y\| \leqslant \delta,$$

where the noise level $\delta > 0$ is known. Algorithm 1 is stopped with $m = m(\delta, y^\delta)$ according to the discrepancy principle (3). For the stopping index $m = m(\delta, y^\delta) \geqslant 1$,

$$\|y^\delta - Tx^\delta_{m(\delta,y^\delta)}\| \leqslant \tau\delta < \|y^\delta - Tx^\delta_{m(\delta,y^\delta)-1}\| \tag{19}$$

is satisfied (with an *a priori* chosen $\tau > 1$), for $m = 0$, only the first inequality holds true. The algorithm always terminates after a finite number of steps, which can be seen as follows. Lemma 4.1 also holds for $\mu = 0$ and $\rho = \|x^+\|$. Hence

$$\lim_{m\to\infty} \|y^\delta - Tx_m\| \leqslant \delta + \lim_{m\to\infty} c|r'_m(0)|^{-\frac{1}{2}}\|x^+\| = \delta,$$

since $|r'_m(0)|^{-\frac{1}{2}} \to 0$ (see (15)). The limit of the norm of the residuals exists, since the sequence is non-increasing due to lemma 2.3 or (7), respectively, and bounded from below by zero. Since $\tau\delta > \delta$, the discrepancy principle stops the algorithm after a finite number of steps. If the algorithm has a finite termination index κ, then $q_\kappa = 0$. According to lemma 2.4 we have $x^\delta_\kappa = T^+y^\delta$, in which case

$$\|y^\delta - Tx^\delta_\kappa\| = \|(I - P_{\overline{R(T)}})y^\delta\| = \|(I - P_{\overline{R(T)}})(y^\delta - y)\| \leqslant \delta$$

and therefore $m(\delta, y^\delta) \leqslant \kappa$.

The letter c designates a generic constant in the following lemmata and proofs.

Lemma 4.1. *Let $y = Tx^+$ with $x^+ \in \mathcal{X}_{\mu,\rho}$. Then for* $0 < m \leqslant \kappa$,

$$\|y^\delta - Tx^\delta_m\| \leqslant \delta + c|r'_m(0)|^{-\mu-\frac{1}{2}}\rho.$$

Proof. The bound (16) proved in theorem 3.1 reads

$$\|y^\delta - Tx^\delta_m\| \leqslant \|F_{\lambda_{1,m}}\varphi_m(TT^*)y^\delta\|.$$

As before φ_m is bounded by 1 in $[0, \lambda_{1,m}]$ and satisfies the equation (18) with $\nu = 2\mu + 1$

$$\lambda^{2\mu+1}\varphi_m^2(\lambda) \leqslant c^{2\mu+1}(2\mu+1)^{2\mu+1}|r'_m(0)|^{-2\mu-1}, \qquad 0 \leqslant \lambda \leqslant \lambda_{1,m}.$$

If we insert these estimates and use $y = T(T^*T)^\mu w$ with $\|w\| \leqslant \rho$, we obtain

$$\begin{aligned}\|y^\delta - Tx^\delta_m\| &\leqslant \|F_{\lambda_{1,m}}\varphi_m(TT^*)(y^\delta - y)\| + \|F_{\lambda_{1,m}}\varphi_m(TT^*)y\|\\ &\leqslant \delta + \|E_{\lambda_{1,m}}\varphi_m(T^*T)(T^*T)^{\mu+\frac{1}{2}}w\|\\ &\leqslant \delta + c^{\mu+\frac{1}{2}}(2\mu+1)^{\mu+\frac{1}{2}}|r'_m(0)|^{-\mu-\frac{1}{2}}\rho,\end{aligned}$$

which gives the assertion. □

The following lemma and its proof are a nearly literal copy of lemma 7.11 in [6], only that the functions representing the iteration are rational instead of polynomial.

Lemma 4.2. *Assume that* $y = Tx^+$ *with* $x^+ \in \mathcal{X}_{\mu,\rho}$. *Then for* $0 \leqslant m \leqslant \kappa$,

$$\|x_m^\delta - x^+\| \leqslant c(\rho^{\frac{1}{2\mu+1}} \delta_m^{\frac{2\mu}{2\mu+1}} + \sqrt{|r_m'(0)|}\delta_m),$$

where

$$\delta_m := \max\{\|y^\delta - Tx_m^\delta\|, \delta\}.$$

Proof. By the interpolation inequality (see [6], page 47) and $x_0^\delta = 0$,

$$\|x^+\| \leqslant \rho^{\frac{1}{2\mu+1}} \|y\|^{\frac{2\mu}{2\mu+1}} \leqslant \rho^{\frac{1}{2\mu+1}} (\|y^\delta\| + \|y - y^\delta\|)^{\frac{2\mu}{2\mu+1}} \leqslant c\rho^{\frac{1}{2\mu+1}} \delta_0^{\frac{2\mu}{2\mu+1}}.$$

We conclude that the assertion of the lemma is true for $m = 0$ by keeping in mind that $r_0' = 0$. Now let $0 < m \leqslant \kappa$. By assumption, we have

$$x^+ = T^+ y = (T^*T)^\mu w,$$

and we choose a positive ϵ such that

$$0 < \epsilon \leqslant |r_m'(0)|^{-1}, \tag{20}$$

which in particular implies that ϵ is smaller than or equal to $\lambda_{1,m}$, see (15). Next, we introduce

$$x_m^\delta = g_m(T^*T)T^*y^\delta, \quad g_m(\lambda) = \frac{q_{m-1}(\lambda)}{(1 + \lambda/\gamma)^{m-1}} \in \Pi_{m-1}/(1 + \cdot/\gamma)^{m-1}, \tag{21}$$

where g_m is the rational function that represents the mth SINE-iterate in $\mathcal{Q}_m$. We obtain

$$\begin{aligned}
\|x^+ - x_m^\delta\| &\leqslant \|E_\epsilon(x^+ - x_m^\delta)\| + \|(I - E_\epsilon)(x^+ - x_m^\delta)\| \\
&\leqslant \|E_\epsilon(x^+ - g_m(T^*T)T^*y)\| + \|E_\epsilon(g_m(T^*T)T^*y - x_m^\delta)\| \\
&\qquad + \epsilon^{-\frac{1}{2}}\|y - Tx_m^\delta\| \\
&\leqslant \|E_\epsilon r_m(T^*T)(T^*T)^\mu w\| + \|E_\epsilon g_m(T^*T)T^*(y - y^\delta)\| \\
&\qquad + \epsilon^{-\frac{1}{2}}\|y - Tx_m^\delta\| \\
&\leqslant \|\lambda^\mu r_m(\lambda)\|_{C[0,\epsilon]}\rho + \|\lambda^{\frac{1}{2}} g_m(\lambda)\|_{C[0,\epsilon]}\delta + \epsilon^{-\frac{1}{2}}(\|y^\delta - Tx_m^\delta\| + \delta).
\end{aligned}$$

From here, a literal copy of the proof of lemma 7.11 in [6] will do. The only difference is that r_m is a rational function (see (10)) instead of a polynomial. □

Finally, we can prove our main theorem.

Theorem 4.3. *If* $y \in \mathcal{R}(T)$ *and if* SINE *is stopped according to the discrepancy principle (19) with* $m(\delta, y^\delta)$, *then* SINE *is an order-optimal regularisation method, i.e. if* $T^+y \in \mathcal{X}_{\mu,\rho}$, *then*

$$\|T^+y - x_{m(\delta,y^\delta)}^\delta\| \leqslant c\rho^{\frac{1}{2\mu+1}} \delta^{\frac{2\mu}{2\mu+1}}.$$

Proof. By the definition of the stopping index $m(\delta, y^\delta)$ by the discrepancy principle, one obtains

$$\delta_{m(\delta,y^\delta)} = \max\{\|y^\delta - Tx^\delta_{m(\delta,y^\delta)}\|, \delta\} \leqslant \tau\delta.$$

With respect to lemma 4.2, it remains to estimate $|r'_{m(\delta,y^\delta)}(0)|$. For simplicity, we write m instead of $m(\delta, y^\delta)$ in the following and assume, without loss of generality, that $m \geqslant 2$. ($m = 0$ follows from lemma 4.2 with $r'_0 = 0$, $m = 1$ refers to the space $\mathcal{Q}_1 = \mathcal{K}_1$ and theorem 7.12 in [6] applies). By lemma 4.1, we conclude that

$$\tau\delta < \|y^\delta - Tx^\delta_{m-1}\| \leqslant \delta + c|r'_{m-1}(0)|^{-\mu-\frac{1}{2}}\rho.$$

Since $\tau > 1$, this implies that

$$|r'_{m-1}(0)| \leqslant c\left(\frac{\rho}{\delta}\right)^{\frac{2}{2\mu+1}}. \tag{22}$$

It remains to estimate

$$\pi_m := r'_{m-1}(0) - r'_m(0).$$

The rational function

$$u_m(\lambda) := \frac{r_{m-1}(\lambda) - r_m(\lambda)}{\lambda} \in \Pi_{m-1}/(1+\cdot/\gamma)^{m-1}$$

satisfies

$$[u_m, \lambda\varphi] = 0 \qquad \text{for every} \quad \varphi \in \Pi_{m-2}/(1+\cdot/\gamma)^{m-2}$$

due to (13) in lemma 2.6. Moreover, by definition of π_m and (15),

$$u_m(0) = \pi_m > \frac{1}{\gamma} > 0. \tag{23}$$

Substituting $u_m = \pi_m + \lambda\varphi$, then we have $\varphi \in \Pi_{m-2}/(1+\cdot/\gamma)^{m-1}$ and

$$[u_m, u_m] = \pi_m[u_m, 1] + [u_m, \lambda\varphi]. \tag{24}$$

We first show that

$$[u_m, \lambda\varphi] = [u_m, u_m - \pi_m] = -\frac{1}{\gamma}[r_{m-1}, \frac{r_{m-1}}{\lambda}] + \frac{1}{\gamma}[r_m, \frac{r_m}{\lambda}]. \tag{25}$$

Using (21), we obtain

$$u_m(\lambda) = \frac{r_{m-1}(\lambda) - r_m(\lambda)}{\lambda} = \frac{q_{m-1}(\lambda)}{(1+\lambda/\gamma)^{m-1}} - \frac{q_{m-2}(\lambda)}{(1+\lambda/\gamma)^{m-2}}$$

and

$$\pi_m = u_m(0) = q_{m-1}(0) - q_{m-2}(0).$$

Hence,

$$[u_m, u_m - \pi_m] = \left[\frac{r_{m-1} - r_m}{\lambda}, \lambda \frac{\tilde{q}_{m-2}}{(1+\lambda/\gamma)^{m-1}} - \lambda \frac{\tilde{q}_{m-3}}{(1+\lambda/\gamma)^{m-2}}\right],$$

with

$$\begin{aligned} \tilde{q}_{m-2}(\lambda) &= \frac{q_{m-1}(\lambda) - q_{m-1}(0)(1+\lambda/\gamma)^{m-1}}{\lambda} \in \Pi_{m-2}, \\ \tilde{q}_{m-3}(\lambda) &= \frac{q_{m-2}(\lambda) - q_{m-2}(0)(1+\lambda/\gamma)^{m-2}}{\lambda} \quad \begin{cases} \in \Pi_{m-3} & \text{for} \quad m \geqslant 3 \\ = 0 & \text{for} \quad m = 2 \end{cases}. \end{aligned}$$

Since

$$\left[r_m, \frac{\tilde{q}_{m-2}}{(1+\lambda/\gamma)^{m-1}} - \frac{\tilde{q}_{m-3}}{(1+\lambda/\gamma)^{m-2}}\right] = 0$$

and

$$\left[r_{m-1}, \frac{\tilde{q}_{m-3}}{(1+\lambda/\gamma)^{m-2}}\right] = 0$$

according to lemma 2.6, we have, again with lemma 2.6,

$$\begin{aligned} [u_m, u_m - \pi_m] &= \left[r_{m-1}, \frac{\tilde{q}_{m-2}}{(1+\lambda/\gamma)^{m-1}}\right] \\ &= \left[r_{m-1}, \frac{\tilde{q}_{m-2}}{(1+\lambda/\gamma)^{m-1}} - \frac{\tilde{q}_{m-2}}{(1+\lambda/\gamma)^{m-2}}\right] \\ &= \left[r_{m-1}, -\frac{\lambda}{\gamma} \cdot \frac{\tilde{q}_{m-2}}{(1+\lambda/\gamma)^{m-1}}\right] \\ &= \left[r_{m-1}, -\frac{\lambda}{\gamma} \cdot \frac{q_{m-1} - q_{m-1}(0)(1+\lambda/\gamma)^{m-1}}{\lambda\,(1+\lambda/\gamma)^{m-1}}\right] \\ &= -\frac{1}{\gamma} \cdot \left[r_{m-1}, \frac{q_{m-1}}{(1+\lambda/\gamma)^{m-1}} - q_{m-1}(0)\right] \\ &= -\frac{1}{\gamma} \cdot \left[r_{m-1}, \frac{q_{m-1}}{(1+\lambda/\gamma)^{m-1}}\right] \\ &= -\frac{1}{\gamma} \cdot \left[r_{m-1}, \frac{1-r_m}{\lambda}\right] = -\frac{1}{\gamma} \cdot \left[r_{m-1}, \frac{1}{\lambda}\right] + \frac{1}{\gamma} \cdot \left[r_{m-1}, \frac{r_m}{\lambda}\right]. \end{aligned}$$

By lemma 2.6 and (21), we obtain

$$[r_{m-1}, \frac{1}{\lambda}] = [r_{m-1}, -g_{m-1} + \frac{1}{\lambda}] = [r_{m-1}, \frac{1}{\lambda} r_{m-1}]$$

and, similarly,

$$[r_{m-1}, \frac{r_m}{\lambda}] = [1 - \lambda \frac{q_{m-2}}{(1+\lambda/\gamma)^{m-2}}, \frac{r_m}{\lambda}] = [1, \frac{r_m}{\lambda}] = [\frac{1}{\lambda}, r_m] = [\frac{r_m}{\lambda}, r_m]$$

which finally gives (25). Due to

$$[u_m, 1] = [r_{m-1}, \frac{1}{\lambda}] - [r_m, \frac{1}{\lambda}]$$

and by lemma 2.6, we obtain analogously

$$[r_j, \frac{1}{\lambda}] = [r_j, -g_j + \frac{1}{\lambda}] = [r_j, \frac{1}{\lambda} r_j], \qquad j = m-1, m,$$

and therefore

$$[u_m, 1] = [r_{m-1}, \frac{1}{\lambda} r_{m-1}] - [r_m, \frac{1}{\lambda} r_m]. \tag{26}$$

Hence, by setting (25) and (26) in (24), we obtain

$$[u_m, u_m] = \left(\pi_m - \frac{1}{\gamma}\right) [r_{m-1}, \frac{1}{\lambda} r_{m-1}] - \left(\pi_m - \frac{1}{\gamma}\right) [r_m, \frac{1}{\lambda} r_m]. \tag{27}$$

From here, the proof continues literally as the proof of theorem 7.12 in [6]. □

5. Upper bounds for the stopping index

The number $m(\delta, y^\delta)$ of necessary iterations to meet the discrepancy principle reflects the efficiency of the method. Due to construction, SINE will stop faster with respect to the discrepancy principle than any other method that relies on the shift-and-invert Krylov subspace. Here, we will additionally show that SINE stops earlier than CGNE (or at the same iterate as CGNE, in the worst case) under the same assumptions as in section 4. For this discussion, we designate the stopping index for SINE with parameter $\gamma > 0$ as $m^\gamma(\delta, y^\delta)$. When γ tends to infinity, SINE turns into CGNE. Therefore we designate the stopping index of CGNE with $m^\infty(\delta, y^\delta)$.

Theorem 5.1. *If* $y \in \mathcal{R}(T)$ *and* $\gamma > 0$ *then* $0 \leqslant m^\gamma(\delta, y^\delta) \leqslant m^\infty(\delta, y^\delta) < \infty$.

Proof. For $m \geqslant 1$, by

$$\min_{\substack{r \in \Pi_m/(1+\cdot/\gamma)^{m-1} \\ r(0)=1}} \int_0^{\|T\|^2+} r^2(\lambda)\, d\|F_\lambda y^\delta\|^2 \leqslant \min_{\substack{p \in \Pi_m \\ p(0)=1}} \int_0^{\|T\|^2+} p^2(\lambda)\, d\|F_\lambda y^\delta\|^2,$$

and hence by (9), we have

$$\min_{x \in \mathcal{Q}_m} \|y^\delta - Tx\| \leqslant \min_{x \in \mathcal{K}_m} \|y^\delta - Tx\|.$$

Due to the definition of CGNE (4) and SINE (7), it follows immediately that the norm of the residual of the mth-SINE iterate x_m^{SINE} is always smaller than or equal to the norm of the residual of the mth-CGNE iterate x_m^{CGNE}, i.e.

$$\|y^\delta - Tx_m^{\mathrm{SINE}}\| \leqslant \|y^\delta - Tx_m^{\mathrm{CGNE}}\|$$

holds for all $m \geqslant 0$, which proves our theorem. (The case $m = 0$ is trivial.) □

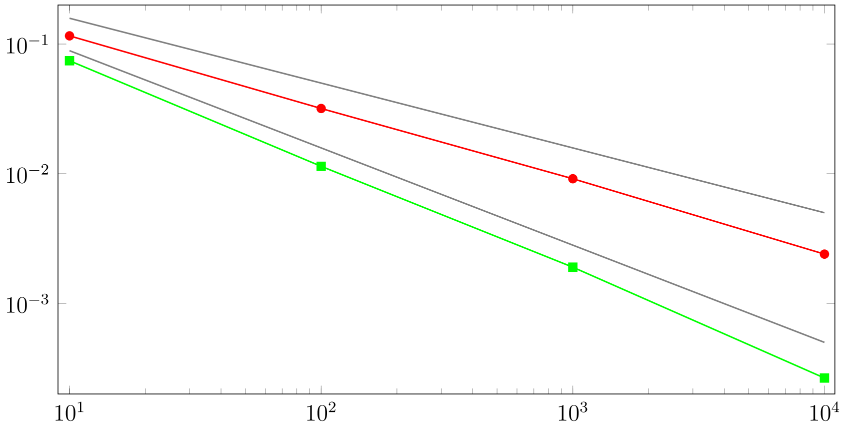

Figure 1. L_2-error versus inverse noise δ^{-1}.

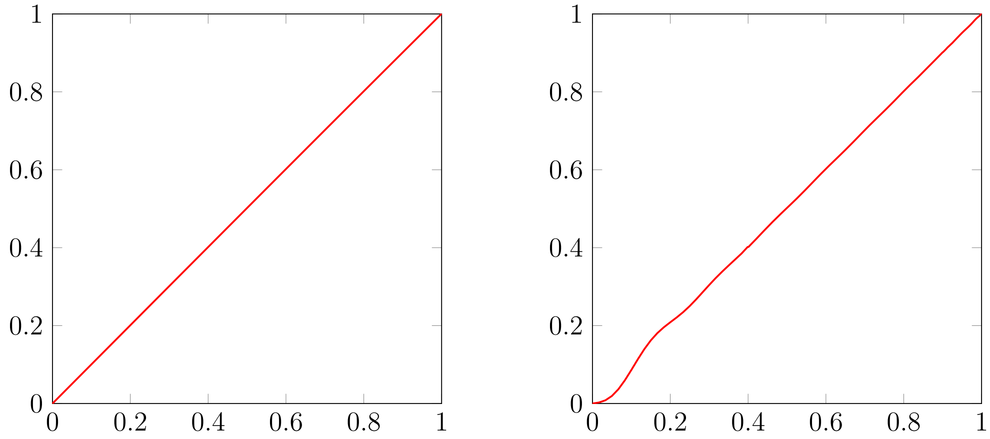

Figure 2. Left-hand side SINE-regularisation attained in 2nd step, right-hand side CGNE-regularisation attained in 19th step.

Theorem 5.1 basically shows that all upper bounds that are known for CGNE also apply to SINE, but SINE might be faster. We explicitly state some corollaries. As a corollary of theorem 7.13 in [6], which is due to [24], and theorem 5.1 we obtain the following statement.

Corollary 5.2. *If $y \in \mathcal{R}(T)$, $\gamma \in \mathbb{R}^+ \cup \{\infty\}$, and $T^+y \in \mathcal{X}_{\mu,\rho}$, then*

$$m^\gamma(\delta, y^\delta) \leqslant c\left(\frac{\rho}{\delta}\right)^{\frac{1}{2\mu+1}}$$

and this estimate is sharp in the sense that the exponent cannot be replaced by a smaller one and that the bound is supposed to hold true for all possible values of γ.

Theorems 7.14 and 7.15 in [6] also hold literally for SINE as simple corollaries of theorem 5.1

6. Illustration and discussion

We use the multiplication operator $T : L_2(0,1) \to L_2(0,1)$, $Tf(t) := tf(t)$, in order to illustrate the theoretical findings. The range of T is not closed. For example, it can be readily seen that any constant function apart from zero is in the closure $\overline{\mathcal{R}(T)}$ of the range $\mathcal{R}(T)$, but not in the range $\mathcal{R}(T)$ itself. We further have $T^* = T$ and $T^*Tf(t) = t^2 f(t)$. For the fractional powers of the operator T^*T, we obtain $(T^*T)^\mu f(t) = t^{2\mu} f(t)$, $\mu > 0$. We choose the exact solutions $x_1^+ = t$ and $x_2^+ = t^3$ with right-hand sides $y_1 = Tx_1^+ = t^2$ and $y_2 = Tx_2^+ = t^4$, respectively. Then $x_1^+ \in \mathcal{X}_{1/2,1}$ and $x_2^+ \in \mathcal{X}_{3/2,1}$. We use the perturbed right-hand sides $y_i^\delta = y_i + \delta, i = 1, 2$, with $\|y_i^\delta - y_i\| = \delta$. The linear systems with the perturbed right-hand sides y_i^δ, $i = 1, 2$, do not have a solution and a regularisation is necessary. We now use SINE to compute regularised solutions, where the iteration is stopped according to the discrepancy principle with $\tau = 1001/1000$. In figure 1, the L_2-norm of the error of the computed regularisation is plotted versus δ^{-1}. The red circle-marked line belongs to the error with respect to $x_1^+ = t$ and the green, square-marked line belongs to the error with respect to $x_2^+ = t^3$. As predicted by theorem 4.3, the convergence to the exact solution with decreasing perturbation δ is at least $\delta^{\frac{1}{2}}$ or $\delta^{\frac{3}{2}}$, respectively, which are indicated by the gray lines. The operator is simple enough such that all computations could be conducted exactly by using the computer algebra system Maple. Due to construction, SINE will stop faster with respect to the discrepancy principle than any other method computing regularisations in the shift-and-invert Krylov subspace $\mathcal{Q}_m$. That is, where these methods have been used successfully, SINE should also be a very good choice. That SINE might also be useful with respect to regularisation schemes that do not use the shift-and-invert Krylov subspace $\mathcal{Q}_m$, will be illustrated by another simple experiment where we compare SINE and CGNE. For $\gamma = \frac{1}{1000}$ and $x^+ = t$, $y^\delta = t^2 + \delta$, $\delta = \frac{1}{1000}$, SINE stops after two steps with

$$x_2 = -\frac{21}{5000}t^3 + \frac{1507}{1500}t = c_1 T^* y^\delta + c_2(1 + T^*T/\gamma)^{-1} T^* y^\delta \in \mathcal{Q}_2,$$

$c_1 = -21/5000$, $c_2 = 15\,070\,063/15\,000$, whereas CGNE produces a polynomial of degree 39 after 19 steps. Both methods have been stopped according to the discrepancy principle with $\tau = \frac{1001}{1000}$. In figure 2, it can be seen that the SINE regularisation on the left-hand side is qualitatively better than the CGNE regularisation on the right-hand side. The experiment also shows that the stopping index of SINE can be significantly smaller than the stopping index of CGNE. With the designations of section 5, we have

$$m^\gamma(\delta, y^\delta) = 2 < 19 = m^\infty(\delta, y^\delta).$$

Altogether, the theory and the experiment suggest that SINE is a valid order-optimal regularisation scheme. It is also hoped that the given analysis inspires further research on the regularisation properties of rational Krylov subspace methods. For example, it is immediately clear that theorem 5.1 carries over to rational Krylov subspaces with arbitrary negative real poles, when the method is defined analogous to (7). Even choosing negative poles at random can only improve on CGNE with respect to the stopping index. These more general rational Krylov subspace methods might also be seen as accelerations of the nonstationary iterated Tikhonov iteration (e.g. [16]) or of method (ii) in example 1.1 with varying step sizes. While there is only one polynomial Krylov subspace, rational Krylov subspaces inspire a wide range of methods that might be adapted to the needs at hand. As a possible application, rational Krylov subspaces have been successfully used to accelerate computations related to seismic imaging (e.g. [4, 19, 31, 32]), which is known to be an ill-posed inverse problem.

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project-ID 258734477—SFB 1173.

ORCID iDs

Volker Grimm https://orcid.org/0000-0002-6821-5388

References

[1] Botchev M A and Knizhnerman L A 2020 ART: adaptive residual-time restarting for Krylov subspace matrix exponential evaluations *J. Comput. Appl. Math.* **364** 112311

[2] Brezinski C, Novati P and Redivo-Zaglia M 2012 A rational Arnoldi approach for ill-conditioned linear systems *J. Comput. Appl. Math.* **236** 2063–77

[3] Buccini A, Donatelli M and Reichel L 2017 Iterated Tikhonov regularization with a general penalty term *Numer. Linear Algebr. Appl.* **24** e2089

[4] Druskin V, Remis R F, Zaslavsky M and Zimmerling J T 2018 Compressing large-scale wave propagation models via phase-preconditioned rational Krylov subspaces *Multiscale Model. Simul.* **16** 1486–518

[5] Eicke B, Louis A K and Plato R 1990 The instability of some gradient methods for ill-posed problems *Numer. Math.* **58** 129–34

[6] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (*Mathematics and its Applications* vol 375) (Dordrecht: Kluwer)

[7] Fakeev A G 1981 A class of iteration processes for solution of degenerate systems of linear algebraic equations *USSR Comput. Math. Math. Phys.* **21** 545–52

[8] Göckler T 2014 Rational Krylov subspace methods for φ-functions in exponential integrators *PhD Thesis* Karlsruhe Institute of Technology (KIT), Germany

[9] Göckler T and Grimm V 2013 Convergence analysis of an extended Krylov subspace method for the approximation of operator functions in exponential integrators *SIAM J. Numer. Anal.* **51** 2189–213

[10] Göckler T and Grimm V 2017 Acceleration of contour integration techniques by rational Krylov subspace methods *J. Comput. Appl. Math.* **316** 133–42

[11] Grimm V 2012 Resolvent Krylov subspace approximation to operator functions *BIT* **52** 639–59

[12] Grimm V and Göckler T 2017 Automatic smoothness detection of the resolvent Krylov subspace method for the approximation of C_0-semigroups *SIAM J. Numer. Anal.* **55** 1483–504

[13] Güttel S 2013 Rational Krylov approximation of matrix functions: numerical methods and optimal pole selection *GAMM-Mitt.* **36** 8–31

[14] Hanke M 1995 *Conjugate Gradient Type Methods for Ill-Posed Problems* (*Pitman Research Notes in Mathematics Series* vol 327) (Harlow: Longman Scientific & Technical)

[15] Hestenes M R and Stiefel E 1952 Methods of conjugate gradients for solving linear systems *J. Res. Nat. Bur. Stand.* **49** 409–36

[16] Jin Q and Stals L 2012 Nonstationary iterated Tikhonov regularization for ill-posed problems in Banach spaces *Inverse Problems* **28** 104011

[17] King J T and Chillingworth C 1979 Approximation of generalized inverses by iterated regularization *Numer. Funct. Anal. Optim.* **1** 499–513

[18] Krjanev A V 1974 An iterative method for solving incorrectly posed problems *USSR Comput. Math. Math. Phys.* **14** 25–35

[19] Liu W, Farquharson C G, Zhou J and Li X 2019 A rational Krylov subspace method for 3D modeling of grounded electrical source airborne time-domain electromagnetic data *J. Geophys. Eng.* **16** 451–62

[20] Liu Y and Gu Ch 2019 A shift and invert reorthogonalization Arnoldi algorithm for solving the chemical master equation *Appl. Math. Comput.* **349** 1–13

[21] Moret I and Novati P 2004 RD-rational approximations of the matrix exponential *BIT Numer. Math.* **44** 595–615

[22] Moret I and Novati P 2019 Krylov subspace methods for functions of fractional differential operators *Math. Comput.* **88** 293–312

[23] Morozov V A 1966 On the solution of functional equations by the method of regularization *Sov. Math. Dokl.* **7** 414–7

[24] Nemirovskiy A S 1986 The regularizing properties of the adjoint gradient method in ill-posed problems *USSR Comput. Math. Math. Phys.* **26** 7–16

[25] Nemirovskiy A S and Polyak B T 1984 Iterative methods for solving linear ill-posed problems under precise information. I *Eng. Cybern.* **22** 1–11

[26] Ramlau R and Reichel L 2019 Error estimates for Arnoldi–Tikhonov regularization for ill-posed operator equations *Inverse Problems* **35** 055002

[27] Rieder A 2005 Runge–Kutta integrators yield optimal regularization schemes *Inverse Problems* **21** 453–71

[28] Riley J D 1955 Solving systems of linear equations with a positive definite, symmetric, but possibly ill-conditioned matrix *Math. Tables Aids Comput.* **9** 96–101

[29] Ruhe A 1984 Rational Krylov sequence methods for eigenvalue computation *Linear Algebr. Appl.* **58** 391–405

[30] Stewart G W 2001 *Matrix Algorithms* vol II (Philadelphia, PA: SIAM) (Eigensystems)

[31] Zhou J, Liu W, Li X and Qi Z 2018 3D transient electromagnetic modeling using a shift-and-invert Krylov subspace method *J. Geophys. Eng.* **15** 1341–9

[32] Zimmerling J 2018 Model reduction of wave equations, theory and applications in forward modeling and imaging *PhD Thesis* Delft University of Technology, The Netherlands