SIAM J. NUMER. ANAL.
Vol. 58, No. 2, pp. 1295–1318
© 2020 Society for Industrial and Applied Mathematics

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

APPROXIMATING THE SINGULAR VALUE EXPANSION OF A COMPACT OPERATOR*

DANIEL K. CRANE[†], MARK S. GOCKENBACH[†], AND MATTHEW J. ROBERTS[†]

Abstract. The singular values and singular vectors of a compact operator T can be estimated by discretizing T (in a variety of ways) and then computing the singular value decomposition of a suitably scaled Galerkin matrix. In general, the singular values and singular vectors converge at the same rate, which is governed by the error (in the operator norm) in approximating T by the discretized operator. However, when the discretization is accomplished by projection (variational approximation), the computed singular values converge at an increased rate; the typical case is that the errors in the singular values are asymptotically equal to the square of the errors in the singular vectors (this statement must be modified if the approximations to the left and right singular vectors converge at different rates). Moreover, in the case of variational approximation, the error in the singular vectors can be compared with the optimal approximation error, with the two being asymptotically equal in the typical case.

Key words. singular value expansion, convergence, Galerkin discretization

AMS subject classifications. 65J22, 47A52

DOI. 10.1137/18M1226002

1. Introduction.

The *singular value expansion* (SVE) of a compact linear operator $T : X \to Y$, where X and Y are separable Hilbert spaces, enables a straightforward analysis of several related problems: computing the generalized inverse $T^\dagger$ of T, understanding the inverse problem $Tx = y$ (given $y \in Y$, estimate $x \in X$), regularizing the inverse problem using Tikhonov regularization or another scheme, and so forth. Although the SVE is used mostly for analysis, it is employed in certain computational schemes, most notably truncated SVE regularization. In this paper, we analyze general schemes for approximating the singular values and vectors of a compact operator.

The SVE and its analysis have a long history; in the context of integral equations, this history dates from 1907 [15]. (For even earlier work, the reader can consult Stewart's brief history of the singular value decomposition (SVD) [17].) In addition to deriving the basic properties of the SVE, much of the work focused on clarifying the sense in which the kernel of the integral operator could be represented in terms of the singular values and vectors (for example, [14], [16]) and characterizing the singular values, including the rate at which they converge to zero (for example, [19], [10], [16]). As we discuss below, there is little work in the literature about computing numerical estimates of the singular values and vectors of a compact operator.

Throughout this paper, X and Y denote separable Hilbert spaces, and $T : X \to Y$ denotes a compact linear operator with SVE

$$T = \sum_{k=1}^{\infty} \sigma_k \psi_k \otimes \phi_k.$$

Thus, $\{\phi_k\}$ and $\{\psi_k\}$ are orthonormal sequences in X and Y, respectively, and $\{\sigma_k\}$

*Received by the editors November 19, 2018; accepted for publication (in revised form) January 14, 2020; published electronically April 22, 2020.
https://doi.org/10.1137/18M1226002

[†]Department of Mathematical Sciences, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931-1295 (dkcrane@mtu.edu, msgocken@mtu.edu, majrober@mtu.edu).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

is a sequence of positive numbers decreasing monotonically to zero. (If T has finite rank, then the SVE contains only finitely many terms. For simplicity of exposition, we will assume the typical case that T has infinite rank.)

Let $\{T_h : X \to Y \,|\, h > 0\}$ be a family of compact linear operators with the property that $T_h \to T$ in the operator norm as $h \to 0$; more specifically, we assume that

$$\|T_h - T\|_{\mathcal{L}(X,Y)} \le \epsilon_h \to 0 \text{ as } h \to 0.$$

Each operator T_h has an SVE

$$T_h = \sum_k \sigma_{h,k} \psi_{h,k} \otimes \phi_{h,k},$$

where the sum contains finitely many terms if T_h has finite rank and infinitely many terms otherwise. This notation for T_h, its SVE, and ϵ_h will be used throughout the paper. We wish to analyze how the singular values and singular vectors of T_h converge to those of T.

We will frequently use the fact that the singular values and singular vectors of T are related to the nonzero eigenvalues and corresponding eigenvectors of T^*T and TT^*. Specifically, for each $k \in \mathbb{Z}^+$, we have

$$\begin{aligned} T\phi_k &= \sigma_k \psi_k, \\ T^*T\phi_k &= \sigma_k^2 \phi_k, \\ TT^*\psi_k &= \sigma_k^2 \psi_k. \end{aligned}$$

It follows that the subspace of right singular vectors associated with the singular value σ_k can be defined as

$$E_k = \{\phi \in X \,:\, T^*T\phi = \sigma_k^2 \phi\}. \tag{1}$$

Similarly,

$$F_k = \{\psi \in Y \,:\, TT^*\psi = \sigma_k^2 \psi\} \tag{2}$$

is the subspace of left singular vectors associated with the singular value σ_k. We assume that the singular values are enumerated according to multiplicity. That is, if $\dim(E_k) = d > 1$, then the value σ_k appears d times in the list $\sigma_1, \sigma_2, \sigma_3, \ldots$.

By definition, the singular values of T and T_h satisfy

$$\sigma_1 \ge \sigma_2 \ge \sigma_3 \ge \cdots$$

and

$$\sigma_{h,1} \ge \sigma_{h,2} \ge \sigma_{h,3} \ge \cdots .$$

There is therefore no ambiguity in asserting that the singular values of T_h converge to those of T; it simply means that, for each $k \in \mathbb{Z}^+$, $\sigma_{h,k} \to \sigma_k$ as $h \to 0$, where it is understood that, for a given k, $\sigma_{h,k}$ is defined for all h sufficiently small.

The convergence of the singular vectors is a more subtle question. If the singular space E_k has dimension $d > 1$, then convergence of the singular values implies that there will be d singular values of T_h converging to σ_k,

$$\sigma_{h,k_i} \to \sigma_k \text{ as } h \to 0 \text{ for } i = 1, 2, \ldots, d,$$

where $\sigma_{k_i} = \sigma_k$ for all $i = 1, 2, \ldots, d$. However, there is no reason to expect that σ_{h,k_i}, $i = 1, 2, \ldots, d$, all have the same value; more typically, the singular space associated

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

with each σ_{h,k_i} is one-dimensional. Since the right singular vectors associated with σ_k need only form an orthonormal basis for E_k, it need not be the case that ϕ_{h,k_i} converge to any particular ϕ_{k_j}. For this reason, we define

$$E_{h,k} = \mathrm{sp}\{\phi \in X \,:\, \exists\, \ell \in \mathbb{Z}^+,\, \sigma_{h,\ell} \to \sigma_k \text{ as } h \to 0 \text{ and } T_h^* T_h \phi = \sigma_{h,\ell}^2 \phi\} \tag{3}$$

and

$$F_{h,k} = \mathrm{sp}\{\psi \in Y \,:\, \exists\, \ell \in \mathbb{Z}^+,\, \sigma_{h,\ell} \to \sigma_k \text{ as } h \to 0 \text{ and } T_h T_h^* \psi = \sigma_{h,\ell}^2 \psi\}. \tag{4}$$

We then require that $E_{h,k}$ converges to E_k in the sense that the *gap* between $E_{h,k}$ and E_k converges to zero, and similarly for $F_{h,k}$ and F_k.

DEFINITION 1. *Given subspaces U and V of a Hilbert space H, we define*

$$\delta(U,V) = \sup_{\substack{u \in U \\ \|u\|_H = 1}} \inf_{v \in V} \|v - u\|_H.$$

We then define the gap *between U and V by*

$$\hat{\delta}(U,V) = \max\{\delta(U,V), \delta(V,U)\}.$$

In the case that V is closed (the only case discussed in this paper), we could equivalently define $\delta(U,V)$ as

$$\delta(U,V) = \sup_{\substack{u \in U \\ \|u\|_H = 1}} \|P_V u - u\|_H,$$

where P_V denotes the (orthogonal) projection onto V. It is clear that $0 \le \delta(U,V) \le 1$. In general, $\delta(U,V)$ and $\delta(V,U)$ may differ, but it is known (see, for example, [11, section 2]) that the two are equal when both are strictly less than 1. Moreover, this also holds if U and V are finite-dimensional subspaces with the same dimension.

LEMMA 2. *Let U and V be k-dimensional subspaces of a Hilbert space H, where k is a positive integer. Then $\delta(U,V) = \delta(V,U)$.*

Proof. Without loss of generality, let us assume that $\delta(U,V) \le \delta(V,U)$. As noted above, $\delta(V,U) < 1$ implies that $\delta(U,V) = \delta(V,U)$. It suffices, therefore, to show that the assumption $\delta(U,V) < \delta(V,U) = 1$ produces a contradiction.

Since V is finite-dimensional,

$$\delta(V,U) = \max_{\substack{v \in V \\ \|v\|_H = 1}} \|P_U v - v\|_H.$$

Therefore, the assumption that $\delta(V,U) = 1$ implies that there exists $\hat{v} \in V$ such that $\|\hat{v}\|_H = 1$ and $\|P_U \hat{v} - \hat{v}\|_H = 1$. This is possible only if $P_U \hat{v} = 0$, that is, if $\hat{v} \in U^\perp$. On the other hand, the assumption that $\delta(U,V) < 1$ implies that $\|P_V u - u\|_H < 1$ for all $u \in U$ and hence that the null space of $P = P_V|_U$ (P_V restricted to U) is trivial. Since $\dim(U) = \dim(V) = k$, the fundamental theorem of linear algebra implies that P maps U onto V; thus, there exists $\hat{u} \in U$ such that $P\hat{u} = \hat{v}$. But then

$$\begin{aligned}
\left\| P\left(\frac{\hat{u}}{\|\hat{u}\|_H}\right) - \frac{\hat{u}}{\|\hat{u}\|_H} \right\|_H < 1 &\Rightarrow \|P\hat{u} - \hat{u}\|_H < \|\hat{u}\|_H \\
&\Rightarrow \|\hat{v} - \hat{u}\|_H < \|\hat{u}\|_H \\
&\Rightarrow \|\hat{v}\|_H^2 + \|\hat{u}\|_H^2 < \|\hat{u}\|_H^2,
\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where we used $\delta(U,V) < 1$ in the first step and $\hat{v} \in U^\perp$ in the last step. Since $\|\hat{v}\|_H = 1$, the last inequality is impossible, and the proof is complete. ∎

Convergence of the singular values and vectors of T_h to those of T follows from the theory of Babuška and Osborn [3]; the following result is Theorem 9.1 of [4], specialized to the self-adjoint case.

THEOREM 3. *Let X be a Hilbert space; let $A : X \to X$ and $A_h : X \to X$, $h > 0$, be compact self-adjoint linear operators; and assume that $A_h \to A$ in the operator norm as $h \to 0$. Then, for any compact subset K of $\rho(A)$ (the resolvent of A), there exists $h_0 > 0$ such that for all $h \in (0, h_0)$, $K \subset \rho(A_h)$. If λ is a nonzero eigenvalue of A with multiplicity equal to m, then there are m eigenvalues $\lambda_{h,1}, \lambda_{h,2}, \ldots, \lambda_{h,m}$ of A_h, repeated according to their multiplicities, such that each $\lambda_{h,i} \to \lambda$ as $h \to 0$. Moreover, the gap between the direct sum of the eigenspaces of A_h corresponding to $\lambda_{h,1}, \lambda_{h,2}, \ldots, \lambda_{h,m}$ and the eigenspace of A corresponding to λ tends to zero as $h \to 0$.*

We can use Theorem 3 to demonstrate the convergence of the singular values and singular vectors of T_h to those of T by applying it to T^*T and TT^*.

LEMMA 4. *There exists a constant $C > 0$ such that*

$$\|T_h^*T_h - T^*T\|_{\mathcal{L}(X,X)} \leq C\epsilon_h \text{ and } \|T_hT_h^* - TT^*\|_{\mathcal{L}(Y,Y)} \leq C\epsilon_h \ \forall\, h > 0.$$

Proof. We have

$$\begin{aligned}\|T_h^*T_h - T^*T\|_{\mathcal{L}(X,X)} &\leq \|T_h^*T_h - T^*T_h\|_{\mathcal{L}(X,X)} + \|T^*T_h - T^*T\|_{\mathcal{L}(X,X)} \\ &\leq (\|T_h\|_{\mathcal{L}(X,Y)} + \|T\|_{\mathcal{L}(X,Y)})\|T_h - T\|_{\mathcal{L}(X,Y)},\end{aligned}$$

and the first result follows (note that $\{\|T_h\|_X\}$ is bounded because $\{T_h\}$ converges as $h \to 0$; also, $\|T^*\|_{\mathcal{L}(Y,X)} = \|T\|_{\mathcal{L}(X,Y)}$). The proof of the second is similar. ∎

THEOREM 5. *For each $k \in \mathbb{Z}^+$,*

$$\begin{aligned}\sigma_{h,k} &\to \sigma_k \text{ as } h \to 0,\\ \hat{\delta}(E_{h,k}, E_k) &\to 0 \text{ as } h \to 0,\\ \hat{\delta}(F_{h,k}, F_k) &\to 0 \text{ as } h \to 0,\end{aligned}$$

where E_k, F_k, $E_{h,k}$, and $F_{h,k}$ are defined by (1–4).

Proof. The convergence of $\sigma_{h,k}$ to σ_k and $\hat{\delta}(E_{h,k}, E_k)$ to zero follows from applying Theorem 3 to $\{T_h^*T_h\}$ and T^*T, while the convergence of $\hat{\delta}(F_{h,k}, F_k)$ to zero follows from applying Theorem 3 to $\{T_hT_h^*\}$ and TT^*. ∎

We will not use Theorems 3 and 5 in the rest of the paper, preferring to give a direct analysis that also provides rates of convergence.

We will use the following well-known max-min characterizations of the singular values:

$$\begin{aligned}\sigma_k &= \max_{\substack{S \subset X\\ \dim(S)=k}} \min_{\substack{x \in S\\ x \neq 0}} \frac{\|Tx\|_Y}{\|x\|_X},\\ \sigma_{h,k} &= \max_{\substack{S \subset X\\ \dim(S)=k}} \min_{\substack{x \in S\\ x \neq 0}} \frac{\|T_hx\|_Y}{\|x\|_X}\end{aligned} \tag{5}$$

(see [8, Theorem 8.6.1] or [12, Chapter 28, Theorem 4]).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

A particular discretization (discretization by projection or variational approximation) is of special interest: We choose families $\{X_h\}$ and $\{Y_h\}$ of finite-dimensional subspaces of X and Y, respectively, having the property that $P_{X_h} \to I_X$ and $P_{Y_h} \to I_Y$ pointwise as $h \to 0$ (where I_X and I_Y denote the identity operators on X and Y, respectively), and define $T_h = P_{Y_h} T P_{X_h}$.

It is straightforward to use the max-min formulas to prove the following results.

THEOREM 6.

1. *For each $k \in \mathbb{Z}^+$, $|\sigma_{h,k} - \sigma_k| \le \|T_h - T\|_{\mathcal{L}(X,Y)}$ for all $h > 0$. (This holds regardless of the form of T_h.)*
2. *If $T_h = P_{Y_h} T P_{X_h}$, then, for each $k \in \mathbb{Z}^+$, $\sigma_{h,k} \le \sigma_k$ for all $h > 0$.*

Proof.

1. See [7, Chapter IV, Corollary 1.6].
2. Immediate from (5). ∎

Note that, in Theorem 6, we define $\sigma_{h,k} = 0$ for $k > \dim(X_h)$.

Relatively little work has been done on approximating the SVE of a compact operator. Hansen [9] analyzed the rate of convergence of the *method of moments* (equivalent to variational approximation) for the case that T is an integral operator of the first kind. Specifically, assume that

$$(Tx)(s) = \int_{I_t} k(s,t)x(t)\,dt, \ (T_h x)(s) = \int_{I_t} k_h(s,t)x(t)\,dt, \ s \in I_s,$$

where k_h is the kernel produced by variational approximation (namely, the projection of k onto the tensor product space $Y_h \otimes X_h$ in $L^2(I_s \times I_t)$). Hansen's analysis is based on

$$\tilde{\epsilon}_h = \|k_h - k\|_{L^2(I_s \times I_t)}$$

(an upper bound for $\|T_h - T\|_{\mathcal{L}(X,Y)}$), and his Theorem 4 implies that

$$\sigma_k - \sigma_{h,k} \le \frac{\tilde{\epsilon}_h^2}{\sigma_{h,k}} \,\forall\, k = 1, 2, \ldots, N_h = \min\{\dim(X_h), \dim(Y_h)\}.$$

As we show below (Theorem 16), a more precise estimate can be formulated in terms of the optimal approximation errors for ϕ_k and ψ_k, which are $\|(I - P_{X_h})\phi_k\|_X$ and $\|(I - P_{Y_h})\psi_k\|_Y$, respectively, and ϵ_h (rather than $\tilde{\epsilon}_h$). (This statement assumes that the singular spaces corresponding to σ_k are one-dimensional; see Theorem 16 for the general statement.)

Regarding the singular vectors, Hansen's Theorem 5 implies (under the assumption that the singular spaces are one-dimensional) that

$$\max\{\|\phi_{h,k} - \phi_k\|_X, \|\psi_{h,k} - \psi_k\|_Y\} \le \frac{\sqrt{2}\sqrt{\tilde{\epsilon}_h}}{\sqrt{\sigma_k - \sigma_{k+1}}} \,\forall\, k = 1, 2, \ldots, N_h.$$

Our Theorem 8 below improves this to an $O(\epsilon_h)$ upper bound for an arbitrary approximation T_h of T, while Theorem 11 improves the estimate for the case of variational approximation and shows when the error in the singular vectors is asymptotically optimal.

Given the paucity of results about approximation of the SVE, it is natural to look to the literature on eigenvalues and eigenvectors, which is extensive (see, for example, [3] or [5]). We could approximate the singular values and right singular vectors of T

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

by approximating the eigenvalues and eigenvectors of T^*T as, for example, described in Babuška and Osborn [2]. They analyze Galerkin methods for solving self-adjoint eigenvalue problems posed in variational form; the eigenvalue problem $T^*T\phi = \sigma^2\phi$ would be posed in variational form as

$$\text{(6)} \qquad \text{find } \phi \in X,\ \lambda \in \mathbb{R},\ \text{such that } \langle \phi, v\rangle_X = \lambda \langle T\phi, Tv\rangle_Y \ \ \forall\, v \in X$$

(with $\sigma^2 = \lambda^{-1}$). The Galerkin method discretizes this variational problem as

$$\text{(7)} \qquad \text{find } \phi \in X_h,\ \lambda \in \mathbb{R},\ \text{such that } \langle \phi, v\rangle_X = \lambda \langle T\phi, Tv\rangle_Y \ \ \forall\, v \in X_h.$$

It is straightforward to show that (7) is equivalent to $T_h^*T_h\phi = \sigma^2\phi$ with $T_h = TP_{X_h}$ (again, with $\sigma^2 = \lambda^{-1}$).

Given any $u \in E_k$ (that is, any right singular vector corresponding to the singular value σ_k) with $\|u\|_X = 1$, and not assuming that E_k is one-dimensional, the optimal approximation error for u is $\|(I - P_{X_h})u\|_X$, and we obviously have

$$\|(I - P_{X_h})u\|_X \le \|(I - P_{E_{h,k}})u\|_X.$$

Theorem 4.4 of [2] depends on

$$\hat{\epsilon}_h = \|T_h^*T_h - T^*T\|_{\mathcal{L}(X,X)}$$

and implies that

$$\text{(8)} \qquad \|(I - P_{E_{h,k}})u\|_X \le (1 + d_k\hat{\epsilon}_h)\|(I - P_{X_h})u\|_X,$$

where d_k is a constant proportional to the reciprocal of the gap between σ_k^2 and the closest distinct eigenvalue of T^*T. (Babuška and Osborn [2] also prove a version of (8) with $\hat{\epsilon}_h$ replaced by $\|T_h^*T_h - T^*T\|^2_{\mathcal{L}(X_1,X_1)}$, where X_1 is the completion of $\mathcal{N}(T)^\perp$ under the inner product defined by $\langle u, v\rangle_{X_1} = \langle Tu, Tv\rangle_Y$. For some problems, this provides a more precise bound because $\|T_h^*T_h - T^*T\|^2_{\mathcal{L}(X_1,X_1)}$ can be asymptotically smaller than $\hat{\epsilon}_h$ for some problems. We will not discuss this and similar refined bounds any further.) The bound (8) is comparable to the result in our Theorem 11:

$$\text{(9)} \qquad \|(I - P_{E_{h,k}})u\|_X \le (1 + \sqrt{2}q_k\epsilon_h)\|(I - P_{X_h})u\|_X + \sqrt{2}q_k\epsilon_h\|(I - P_{Y_h})v\|_Y.$$

Our constant $\sqrt{2}q_k$ is proportional to the reciprocal of the gap between σ_k and the closest distinct singular value of T, which implies that $\sqrt{2}q_k$ is asymptotically smaller than d_k as $k \to \infty$. On the other hand, while $\hat{\epsilon}_h = O(\epsilon_h)$ as $h \to 0$, it could be the case that $\hat{\epsilon}_h$ is asymptotically smaller than ϵ_h. Therefore, the two constants, $d_k\hat{\epsilon}_h$ and $\sqrt{2}q_k\epsilon_h$, are not directly comparable. Nevertheless, (8) implies that

$$\|(I - P_{E_{h,k}})u\|_X \sim \|(I - P_{X_h})u\|_X \text{ as } h \to 0,$$

and (9) yields the same result *if* $\epsilon_h\|(I - P_{Y_h})v\|_Y = o(\|(I - P_{X_h})u\|_X)$. We will see by example that this is not always the case and thus that optimal approximability of the right singular vectors can be lost due to poor approximability of the corresponding left singular vectors.

With respect to the singular values, Theorem 4.2 of [2] implies that there exists a right singular vector u corresponding to σ_k such that

$$\frac{\sigma_k^2 - \sigma_{h,k}^2}{\sigma_{h,k}^2} \le (1 + d_k\hat{\epsilon}_h)\|(I - P_{X_h})u\|_X^2.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Since

$$\frac{\sigma_k^2 - \sigma_{h,k}^2}{\sigma_{h,k}^2} = \frac{\sigma_k - \sigma_{h,k}}{\sigma_{h,k}} \cdot \frac{\sigma_k + \sigma_{h,k}}{\sigma_{h,k}}$$

and (since $\sigma_{h,k} \leq \sigma_k$)

$$\frac{\sigma_{h,k}}{\sigma_k + \sigma_{h,k}} \leq \frac{\sigma_{h,k}}{\sigma_{h,k} + \sigma_{h,k}} = \frac{1}{2},$$

we obtain

$$\frac{\sigma_k - \sigma_{h,k}}{\sigma_{h,k}} \leq \frac{1}{2}\left(1 + d_k \hat{\epsilon}_h\right) \|(I - P_{X_h})u\|_X^2. \tag{10}$$

Our Theorem 16 implies that if we estimate the singular values by computing the SVE of $T_h = P_{Y_h} T P_{X_h}$ directly (rather than solving an eigenvalue problem), we obtain

$$\begin{aligned}\frac{\sigma_k - \sigma_{h,\ell}}{\sigma_k} \leq \frac{1}{2}&\left(\|(I - P_{X_h})u\|_X^2 + \|(I - P_{Y_h})v\|_Y^2\right)\\ &+ C_k \epsilon_h \left(\|(I - P_{X_h})u\|_X + \|(I - P_{Y_h})v\|_Y\right)^2 \ \ \forall\, h > 0 \text{ sufficiently small,}\end{aligned}$$

where $u \in X$ and $v \in Y$ satisfy $\|u\|_X = 1$, $\|v\|_Y = 1$, and $Tu = \sigma_k v$. If $\|(I - P_{X_h})u\|_X$ and $\|(I - P_{Y_h})v\|_Y$ go to zero at the same rate, this suggests that the error in the singular values (computed directly) is twice the error in the same quantities computed by the eigenvalue approach. However, this ignores the limitations of finite precision arithmetic, as discussed below.

It is natural to compute the singular values and singular vectors of T directly, rather than compute the eigenvalues and eigenvectors of T^*T, for several reasons. First, given that $\sigma_k \to 0$ as $k \to \infty$, we can expect to compute σ_k directly as long as it is larger than machine epsilon ϵ_{mach}. By the eigenvalue approach, we expect to be able to compute σ_k^2 as long as it remains above ϵ_{mach}, that is, as long as σ_k is larger than $\sqrt{\epsilon_{mach}}$. Thus, the eigenvalue approach reduces the range of the singular values than can be estimated. Second, both approaches require the computation of a Galerkin matrix. The eigenvalue approach requires $\hat{A}$ defined by

$$\hat{A}_{ij} = \langle x_i, T^*Tx_j \rangle_X = \langle Tx_i, Tx_j \rangle_X ,$$

while the direct approach requires A defined by

$$A_{ij} = \langle y_i, Tx_j \rangle_X ,$$

where $\{x_1, x_2, \ldots, x_n\}$ and $\{y_1, y_2, \ldots, y_m\}$ are the bases for X_h and Y_h, respectively. In the common case that T is an integral operator, $\hat{A}_{ij}$ is defined by a triple integral, while A_{ij} is defined by a double integral. Therefore, the direct approach may be more efficient. Finally, the eigenvalue approach gives no direct estimate of the left singular vectors.

In the remainder of the paper, we will analyze the errors in both the singular values and the singular vectors in both the generic case (T_h is assumed only to converge to T in the operator norm) and the case of variational approximation ($T_h = P_{Y_h} T P_{X_h}$ and the related cases of $T_h = TP_{X_h}$ and $T_h = P_{Y_h} T$). Specifically, section 2 contains error estimates for the computed singular vectors in the generic case. In section 3, we present an analysis of the convergence of both the singular vectors and the singular values in the case of variational approximation. Numerical examples are presented in section 4, and we present some conclusions in section 5. The appendix shows how to compute the singular values and singular vectors of $T_h = P_{Y_h} T P_{X_h}$ from the SVD of a scaled Galerkin matrix.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

2. Convergence of singular vectors: The general case. We have already seen that

$$|\sigma_{h,k} - \sigma_k| \le \|T_h - T\|_{L(X,Y)} \le \epsilon_h \text{ for all } h > 0.$$

We now show that $\hat{\delta}(E_{h,k}, E_k)$ and $\hat{\delta}(F_{h,k}, F_k)$ also converge to zero like $O(\epsilon_h)$. We begin by establishing more notation.

We have already defined E_k and F_k in (1) and (2); let I_k be the corresponding index set, so that

$$E_k = \mathrm{sp}\{\phi_\ell : \ell \in I_k\},\ F_k = \mathrm{sp}\{\psi_\ell : \ell \in I_k\}.$$

Let $\sigma_{h,\ell}$, $\phi_{h,\ell}$, $\psi_{h,\ell}$, $\ell \in \mathcal{I}_h$, be the singular values and singular vectors of T_h, where $\mathcal{I}_h = \{1, 2, \ldots, N_h\}$ or $I_h = \mathbb{Z}^+$, according as T_h has finite rank or not, and define

$$I_{h,k} = \{\ell \in \mathcal{I}_h : \sigma_{h,\ell} \to \sigma_k \text{ as } h \to 0\}.$$

(Note that $I_{h,k} = I_k$ for all $h > 0$ sufficiently small.) Then $E_{h,k} = \mathrm{sp}\{\phi_{h,\ell} : \ell \in I_{h,k}\}$. Extend $\{\phi_{h,\ell} : \ell \in \mathcal{I}_h\}$ to a complete orthonormal sequence $\{\phi_{h,\ell} : \ell \in \mathcal{J}_h\}$ for X ($\mathcal{I}_h \subset \mathcal{J}_h$), and define $\sigma_{h,\ell} = 0$ for $\ell \in \mathcal{J}_h \setminus \mathcal{I}_h$.

We will write gap_k for the gap between σ_k and the nearest distinct singular value of T. If $k > 1$ and $\sigma_{k-1} > \sigma_k = \sigma_{k+1} = \cdots = \sigma_{k+n_k-1} > \sigma_{k+n_k}$, then

$$\mathrm{gap}_k = \min\{\sigma_{k-1} - \sigma_k, \sigma_k - \sigma_{k+n_k}\},$$

while $\mathrm{gap}_1 = \sigma_1 - \sigma_t$, where σ_t is the largest singular value not equal to σ_1.

LEMMA 7. *Assume that $\|T_h - T\|_{\mathcal{L}(X,Y)} \le \epsilon_h$ for $h > 0$. Then, for each $k \in \mathbb{Z}^+$,*

$$\max\left\{\frac{1}{|\sigma_k - \sigma_{h,\ell}|} : \ell \in \mathcal{J}_h \setminus I_{h,k}\right\} \le \frac{\sqrt{2}}{\mathrm{gap}_k}\ \forall\, h > 0 \textit{ sufficiently small.}$$

Proof. Let $k \in \mathbb{Z}^+$ be given, and note that $\sigma_{h,\ell} \to \sigma_\ell$ as $h \to 0$ for all $\ell \in \mathbb{Z}^+$. It follows that, for $h > 0$ sufficiently small, $I_{h,k} = I_k$ and

$$|\sigma_{h,\ell} - \sigma_k| \ge \frac{\mathrm{gap}_k}{\sqrt{2}}\ \forall\, \ell \in \mathcal{J}_h \setminus I_{h,k}.$$

The result follows. ∎

In Lemma 7, we could obviously replace $\sqrt{2}$ by any constant strictly greater than 1. We will write $q_k = \sqrt{2}/\mathrm{gap}_k$.

THEOREM 8. *Assume that $\|T_h - T\|_{\mathcal{L}(X,Y)} \le \epsilon_h$ for $h > 0$. Let $k \in \mathbb{Z}^+$, and let E_k and $E_{h,k}$ be defined by (1) and (3), respectively. Then, for all $u \in E_k$ satisfying $\|u\|_X = 1$ and for all $h > 0$ sufficiently small, $\|(I - P_{E_{h,k}})u\|_X \le 2q_k\epsilon_h$.*

Proof. Given $u \in E_k$, we define $v = \sigma_k^{-1}Tu$; then $T^*v = \sigma_k u$. We have

$$u = \sum_{\ell \in \mathcal{J}_h} \langle \phi_{h,\ell}, u\rangle_X\, \phi_{h,\ell},$$

$$P_{E_{h,k}}u = \sum_{\ell \in I_{h,k}} \langle \phi_{h,\ell}, u\rangle_X\, \phi_{h,\ell}$$

and hence

$$u - P_{E_{h,k}}u = \sum_{\ell \in \mathcal{J}_h \setminus I_{h,k}} \langle \phi_{h,\ell}, u\rangle_X\, \phi_{h,\ell}.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

For each $\ell \in \mathcal{J}_h \setminus I_{h,k}$, we have

$$\begin{aligned}
\sigma_k |\langle \phi_{h,\ell}, u\rangle_X| &= |\langle \phi_{h,\ell}, \sigma_k u\rangle_X| = |\langle \phi_{h,\ell}, T^* v\rangle_X| = |\langle T\phi_{h,\ell}, v\rangle_Y|,\\
\sigma_{h,\ell} |\langle \phi_{h,\ell}, u\rangle_X| &= |\langle \sigma_{h,\ell}\phi_{h,\ell}, u\rangle_X| = |\langle T_h^*\psi_{h,\ell}, u\rangle_X| = |\langle \psi_{h,\ell}, T_h u\rangle_Y|,\\
\sigma_k |\langle \psi_{h,\ell}, v\rangle_Y| &= |\langle \psi_{h,\ell}, \sigma_k v\rangle_Y| = |\langle \psi_{h,\ell}, Tu\rangle_Y|,\\
\sigma_{h,\ell} |\langle \psi_{h,\ell}, v\rangle_Y| &= |\langle \sigma_{h,\ell}\psi_{h,\ell}, v\rangle_Y| = |\langle T_h\phi_{h,\ell}, v\rangle_Y|.
\end{aligned}$$

Subtracting the second and fourth equations from the first and third, respectively, and then adding yields

$$\begin{aligned}
&|\sigma_k - \sigma_{h,\ell}| \left(|\langle \phi_{h,\ell}, u\rangle_X| + |\langle \psi_{h,\ell}, v\rangle_Y|\right)\\
&= \left| |\langle T\phi_{h,\ell}, v\rangle_Y| - |\langle T_h\phi_{h,\ell}, v\rangle_Y| + |\langle \psi_{h,\ell}, Tu\rangle_Y| - |\langle \psi_{h,\ell}, T_h u\rangle_Y| \right|\\
&\le \left| |\langle T\phi_{h,\ell}, v\rangle_Y| - |\langle T_h\phi_{h,\ell}, v\rangle_Y| \right| + \left| |\langle \psi_{h,\ell}, Tu\rangle_Y| - |\langle \psi_{h,\ell}, T_h u\rangle_Y| \right|\\
&\le \left|\langle (T-T_h)\phi_{h,\ell}, v\rangle_Y\right| + \left|\langle \psi_{h,\ell}, (T-T_h)u\rangle_Y\right|.
\end{aligned}$$

This implies that

$$|\langle \phi_{h,\ell}, u\rangle_X| \le \frac{\left|\langle (T-T_h)\phi_{h,\ell}, v\rangle_Y\right| + \left|\langle \psi_{h,\ell}, (T-T_h)u\rangle_Y\right|}{|\sigma_k - \sigma_{h,\ell}|}. \tag{11}$$

Therefore,

$$\begin{aligned}
\|u - P_{E_{h,k}}u\|_X^2 &= \sum_{\ell \in \mathcal{J}_h \setminus I_{h,k}} \langle \phi_{h,\ell}, u\rangle_X^2\\
&\le \sum_{\ell \in \mathcal{J}_h \setminus I_{h,k}} \left(\frac{\left|\langle (T-T_h)\phi_{h,\ell}, v\rangle_Y\right| + \left|\langle \psi_{h,\ell}, (T-T_h)u\rangle_Y\right|}{|\sigma_k - \sigma_{h,\ell}|}\right)^2\\
&\le 2q_k^2 \sum_{\ell \in \mathcal{J}_h \setminus I_{h,k}} \left(\left|\langle \phi_{h,\ell}, (T-T_h)^* v\rangle_X\right|^2 + \left|\langle \psi_{h,\ell}, (T-T_h)u\rangle_Y\right|^2\right)\\
&\le 2q_k^2 \left(\|(T-T_h)^* v\|_X^2 + \|(T-T_h)u\|_Y^2\right) \le 4q_k^2\epsilon_h^2
\end{aligned}$$

(since $\|T_h - T\|_{\mathcal{L}(X,Y)} = \|T_h^* - T^*\|_{\mathcal{L}(Y,X)} \le \epsilon_h$ and $\|u\|_X = \|v\|_Y = 1$). Therefore, $\|u - P_{E_{h,k}}u\|_X \le 2q_k\epsilon_h$, as desired. ◻

The previous theorem implies that $\delta(E_k, E_{h,k}) \le 2q_k\epsilon_h$ for all $h > 0$ sufficiently small. Since Lemma 2 shows that $\delta(E_{h,k}, E_k) = \delta(E_k, E_{h,k})$ for all h sufficiently small (h must be small enough that $\dim(E_{h,k}) = \dim(E_k)$), we obtain the desired bound on the gap between E_k and $E_{h,k}$.

COROLLARY 9. *For each $k \in \mathbb{Z}^+$,*

$$\hat{\delta}(E_{h,k}, E_k) \le 2q_k\epsilon_h \ \forall\, h > 0 \text{ sufficiently small.}$$

The same analysis, applied to T^* and T_h^*, shows that

$$\hat{\delta}(F_{h,k}, F_k) \le 2q_k\epsilon_h \ \forall\, h > 0 \text{ sufficiently small,}$$

where F_k and $F_{h,k}$ are defined by (2) and (4), respectively.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

3. Accelerated convergence. The case of variational approximation deserves special attention because it leads to increased rates of convergence. Given the families of finite-dimensional subspaces $\{X_h\}$ and $\{Y_h\}$ of X and Y, respectively, we define $T_h : X \to Y$ by $T_h = P_{Y_h} T P_{X_h}$ and

$$\epsilon_h = \|P_{Y_h} T P_{X_h} - T\|_{\mathcal{L}(X,Y)} \ \forall\, h > 0.$$

Under our assumptions on $\{X_h\}$ and $\{Y_h\}$ (namely, that $P_{X_h} \to I_X$ and $P_{Y_h} \to I_Y$ pointwise), it is guaranteed that $\epsilon_h \to 0$ as $h \to 0$.

THEOREM 10. *The operator $P_{Y_h} T P_{X_h}$ converges in the operator norm to T as $h \to 0$.*

Proof. Consider

$$\begin{aligned}\|P_{Y_h} T P_{X_h} - T\|_{\mathcal{L}(X,Y)} &\le \|P_{Y_h} T P_{X_h} - P_{Y_h} T\|_{\mathcal{L}(X,Y)} + \|P_{Y_h} T - T\|_{\mathcal{L}(X,Y)} \\ &\le \|T(P_{X_h} - I_X)\|_{\mathcal{L}(X,Y)} + \|(P_{Y_h} - I_Y)T\|_{\mathcal{L}(X,Y)}.\end{aligned}$$

The second norm goes to zero by a standard result (see [1] or [6, Theorem 7]), while the first goes to zero by Theorem 16 of [6]. ∎

The results of section 2 apply and show that the singular values and corresponding singular spaces of T_h satisfy

$$\begin{aligned}|\sigma_{h,k} - \sigma_k| &\le \epsilon_h, \\ \hat{\delta}(E_{h,k}, E_k) &\le 2q_k\epsilon_h, \\ \hat{\delta}(F_{h,k}, F_k) &\le 2q_k\epsilon_h,\end{aligned}$$

where the above inequalities hold for all $h > 0$ sufficiently small. We will now show that, for $T_h = P_{Y_h} T P_{X_h}$, better rates of convergence are obtained. Recall that $\{\phi_{h,\ell} : \ell \in I_h\}$ was extended to a complete orthonormal sequence $\{\phi_{h,\ell} : \ell \in \mathcal{J}_h\}$ for X. We now assume that this is done in such a way that $\{\phi_{h,\ell} : \ell \in \tilde{I}_h\}$ ($I_h \subset \tilde{I}_h \subset \mathcal{J}_h$) is an orthonormal basis for X_h ($\tilde{I}_h = I_h$ if $\mathcal{N}(T_h) \cap X_h$ is trivial). Similarly, we extend $\{\psi_{h,\ell} : \ell \in I_h\}$ to a complete orthonormal sequence $\{\psi_{h,\ell} : \ell \in \mathcal{K}_h\}$ for Y and assume that $\{\psi_{h,\ell} : \ell \in \hat{I}_h\}$ is an orthonormal basis for Y_h. These definitions imply that $\{\phi_{h,\ell} : \ell \in \mathcal{J}_h \setminus \tilde{I}_h\}$ is a complete orthonormal sequence for $X_h^\perp$ and $\{\psi_{h,\ell} : \ell \in \mathcal{K}_h \setminus \hat{I}_h\}$ is a complete orthonormal sequence for $Y_h^\perp$.

3.1. Singular vectors.

THEOREM 11. *Suppose $u \in E_k$ and $v \in F_k$ satisfy $Tu = \sigma_k v$ and $T_h = P_{Y_h} T P_{X_h}$. Then, for all $h > 0$ sufficiently small,*

$$\|(I - P_{E_{h,k}})u\|_X \le (1 + \sqrt{2}q_k\epsilon_h)\|(I - P_{X_h})u\|_X + \sqrt{2}q_k\epsilon_h\|(I - P_{Y_h})v\|_Y, \tag{12}$$

$$\|(I - P_{F_{h,k}})v\|_Y \le (1 + \sqrt{2}q_k\epsilon_h)\|(I - P_{Y_h})v\|_Y + \sqrt{2}q_k\epsilon_h\|(I - P_{X_h})u\|_X. \tag{13}$$

Proof. As in the proof of Theorem 8,

$$(I - P_{E_{h,k}})u = \sum_{\ell \in \mathcal{J}_h \setminus I_{h,k}} \langle \phi_{h,\ell}, u \rangle_X \, \phi_{h,\ell},$$

which yields

$$\|(I - P_{E_{h,k}})u\|_X^2 = \sum_{\ell \in \mathcal{J}_h \setminus \tilde{I}_h} \langle \phi_{h,\ell}, u \rangle_X^2 + \sum_{\ell \in \tilde{I}_h \setminus I_{h,k}} \langle \phi_{h,\ell}, u \rangle_X^2 \,.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

The first term on the right is exactly $\|(I-P_{X_h})u\|_X^2$. To estimate the second term, we use the following inequality from the proof of Theorem 8:

$$|\langle \phi_{h,\ell}, u\rangle_X| \le \frac{1}{|\sigma_k-\sigma_{h,\ell}|}\left(\left|\langle (T-T_h)\phi_{h,\ell}, v\rangle_Y\right| + \left|\langle \psi_{h,\ell}, (T-T_h)u\rangle_Y\right|\right).$$

We argue as follows:

$$\begin{aligned}\sum_{\ell\in\tilde{I}_h\setminus I_{h,k}} \langle \phi_{h,\ell}, u\rangle_X^2 &\le \sum_{\ell\in\tilde{I}_h\setminus I_{h,k}} \left(\frac{\left|\langle (T-T_h)\phi_{h,\ell}, v\rangle_Y\right| + \left|\langle \psi_{h,\ell}, (T-T_h)u\rangle_Y\right|}{|\sigma_k-\sigma_{h,\ell}|}\right)^2 \\ &\le 2q_k^2 \sum_{\ell\in\tilde{I}_h\setminus I_{h,k}} \left(\left|\langle (T-T_h)\phi_{h,\ell}, v\rangle_Y\right|^2 + \left|\langle \psi_{h,\ell}, (T-T_h)u\rangle_Y\right|^2\right).\end{aligned}$$

For any $x\in X_h$,

$$(T-T_h)x = Tx - P_{Y_h}TP_{X_h}x = Tx - P_{Y_h}Tx = (I-P_{Y_h})Tx \in Y_h^{\perp},$$

and therefore

$$\langle (T-T_h)\phi_{h,\ell}, v\rangle_Y = \langle (T-T_h)\phi_{h,\ell}, (I-P_{Y_h})v\rangle_Y.$$

Similarly, for $y\in Y_h$,

$$(T-T_h)^*y = T^*y - P_{X_h}T^*P_{Y_h}y = T^*y - P_{X_h}T^*y = (I-P_{X_h})T^*y \in X_h^{\perp},$$

which yields

$$\langle \psi_{h,\ell}, (T-T_h)u\rangle_Y = \langle (T-T_h)^*\psi_{h,\ell}, u\rangle_X = \langle (T-T_h)^*\psi_{h,\ell}, (I-P_{X_h})u\rangle_X.$$

Therefore,

$$\begin{aligned}&\sum_{\ell\in\tilde{I}_h\setminus I_{h,k}} \left(\left|\langle (T-T_h)\phi_{h,\ell}, v\rangle_Y\right|^2 + \left|\langle \psi_{h,\ell}, (T-T_h)u\rangle_Y\right|^2\right) \\ &= \sum_{\ell\in\tilde{I}_h\setminus I_{h,k}} \left(\left|\langle (T-T_h)\phi_{h,\ell}, (I-P_{Y_h})v\rangle_Y\right|^2 + \left|\langle (T-T_h)^*\psi_{h,\ell}, (I-P_{X_h})u\rangle_X\right|^2\right) \\ &= \sum_{\ell\in\tilde{I}_h\setminus I_{h,k}} \left(\left|\langle \phi_{h,\ell}, (T-T_h)^*(I-P_{Y_h})v\rangle_X\right|^2 + \left|\langle \psi_{h,\ell}, (T-T_h)(I-P_{X_h})u\rangle_Y\right|^2\right) \\ &\le \left(\|(T-T_h)^*(I-P_{Y_h})v\|_X^2 + \|(T-T_h)(I-P_{X_h})u\|_Y^2\right) \\ &\le \epsilon_h^2\left(\|(I-P_{Y_h})v\|_Y^2 + \|(I-P_{X_h})u\|_X^2\right).\end{aligned}$$

The conclusion is

$$\|(I-P_{E_{h,k}})u\|_X^2 \le \|(I-P_{X_h})u\|_X^2 + 2q_k^2\epsilon_h^2\left(\|(I-P_{Y_h})v\|_Y^2 + \|(I-P_{X_h})u\|_X^2\right).$$

Inequality (12) follows immediately.

The proof of (13) is exactly analogous. ∎

Inequality (12) suggests that the approximation error for a given right singular vector u is affected by the optimal approximation error for that singular vector and

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

also the optimal approximation error for the corresponding left singular vector. As long as

$$\epsilon_h \|(I - P_{Y_h})v\|_Y = o\left(\|(I - P_{X_h})u\|_X\right) \text{ as } h \to 0,$$

it follows that

$$\|(I - P_{E_{h,k}})u\|_X \sim \|(I - P_{X_h})u\|_X \text{ as } h \to 0;$$

that is, the error in the approximation to u is asymptotically optimal. However, it is not difficult to construct an example in which

$$\|(I - P_{X_h})u\|_X = o\left(\epsilon_h \|(I - P_{Y_h})v\|_Y\right) \text{ as } h \to 0,$$

in which case the error in the approximation to u is suboptimal. Examples are presented below.

The situation with respect to optimal approximation of the left singular vectors is exactly analogous.

The bound (12) is not directly comparable to the result (8) of Babuška and Osborn [2]. When we approximate both the right and the left singular vector simultaneously, it is not guaranteed that we obtain an optimal rate of convergence for both, although Theorem 11 shows that at least one of the left or right singular vectors will exhibit an optimal rate of convergence. The examples given below verify that suboptimal convergence is observed in some cases. If, for some reason, it were desired to approximate just the singular values and either the right singular vectors or the left singular vectors, we could (at least in principle) proceed with $T_h = TP_{X_h}$ or $T_h = P_{Y_h}T$.

Theorem 12.

1. *Let the family $\{X_h\}$ of subspaces of X be given and, for each $h > 0$, define $Y_h = T(X_h)$, $T_h = TP_{X_h}$. Let $u \in E_k$ be given and define $v = \sigma_k^{-1}Tu$. Then, for all $h > 0$ sufficiently small,*

$$\|(I - P_{E_{h,k}})u\|_X \le (1 + \sqrt{2}q_k\epsilon_h)\|(I - P_{X_h})u\|_X, \tag{14}$$

$$\|(I - P_{F_{h,k}})v\|_Y \le \|(I - P_{Y_h})v\|_Y + \sqrt{2}q_k\epsilon_h\|(I - P_{X_h})u\|_X. \tag{15}$$

2. *Let the family $\{Y_h\}$ of subspaces of Y be given and, for each $h > 0$, define $X_h = T^*(Y_h)$, $T_h = P_{Y_h}T$. Let $v \in F_k$ be given and define $u = \sigma_k^{-1}T^*v$. Then, for all $h > 0$ sufficiently small,*

$$\|(I - P_{E_{h,k}})u\|_X \le \|(I - P_{X_h})u\|_X + \sqrt{2}q_k\epsilon_h\|(I - P_{Y_h})v\|_Y, \tag{16}$$

$$\|(I - P_{F_{h,k}})v\|_Y \le (1 + \sqrt{2}q_k\epsilon_h)\|(I - P_{Y_h})v\|_Y. \tag{17}$$

Proof. The proofs of (14)–(17) are similar to the proofs of (12–13) and will not be given in detail. The difference between (12) and (14) is that, for $x \in X_h$ and $T = TP_{X_h}$, $(T - T_h)x = 0$ (rather than just $(T - T_h)x \in X_h^\perp$, as in the case of $T_h = P_{Y_h}TP_{X_h}$). The difference between (13) and (17) is similar. ∎

Therefore, if we use $T_h = TP_{X_h}$, we are guaranteed that the error in the approximations of the right singular vectors is optimal and of the same quality as obtained in (8). Similarly. with $T_h = P_{Y_h}T$, we are guaranteed optimal approximation of the left singular vectors. In the appendix, we show how to compute the SVE of $P_{Y_h}TP_{X_h}$ by computing the SVD of a suitably scaled Galerkin matrix. The numerical computation of the SVE of $T_h = TP_{X_h}$ or $T_h = P_{Y_h}T$ is problematic; we also comment on this in the appendix.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

3.2. Singular values. We continue to discuss the case of $T_h = P_{Y_h} T P_{X_h}$. We need a bound on $\|T_h(I - P_{E_{h,k}})u\|_Y$ for $u \in E_k$. We will use the following technical results.

LEMMA 13. *If $\{s_k\}$ is a strictly decreasing sequence of positive real numbers that converges to zero, then*

$$\frac{s_{k-1}}{s_{k-1} - s_k} < \frac{2s_k}{\min\{s_{k-1} - s_k, s_k - s_{k+1}\}} \ \forall\, k > 1.$$

Proof. Given $k > 1$, we consider two cases.

1. If $s_{k-1} - s_k \le s_k - s_{k+1}$, then it follows that $s_{k-1} - s_k < s_k$ and hence that $s_{k-1} < 2s_k$. Therefore,

$$\frac{s_{k-1}}{s_{k-1} - s_k} = \frac{s_{k-1}}{\min\{s_{k-1} - s_k, s_k - s_{k+1}\}} < \frac{2s_k}{\min\{s_{k-1} - s_k, s_k - s_{k+1}\}}.$$

   Thus, the result holds in this case.

2. If $s_k - s_{k+1} < s_{k-1} - s_k$, define

$$\theta_1 = \frac{s_k}{s_{k-1}},\ \theta_2 = \frac{s_{k+1}}{s_{k-1}},$$

   and note that $\theta_1 \in (0, 1)$, $\theta_2 \in (0, \theta_1)$. We must show that

$$\begin{aligned} &\frac{s_{k-1}}{s_{k-1} - s_k} < \frac{2s_k}{\min\{s_{k-1} - s_k, s_k - s_{k+1}\}} = \frac{2s_k}{s_k - s_{k+1}} \\ \Leftrightarrow\ &\frac{s_{k-1}(s_k - s_{k+1})}{s_k(s_{k-1} - s_k)} < 2 \ \Leftrightarrow\ \frac{\theta_1 - \theta_2}{\theta_1(1 - \theta_1)} < 2. \end{aligned}$$

   Note that $s_k - s_{k+1} < s_{k-1} - s_k$ is equivalent to $\theta_1 - \theta_2 < 1 - \theta_1$. It is now an easy exercise to prove that $(\theta_1 - \theta_2)/(\theta_1(1 - \theta_1)) < 2$ for θ_1, θ_2 satisfying $0 < \theta_2 < \theta_1 < 1$ and $\theta_1 - \theta_2 < 1 - \theta_1$. This completes the proof. ∎

LEMMA 14. *For a given $k \in \mathbb{Z}^+$, there exists $h_0 > 0$ such that*

$$\frac{\sigma_{h,\ell}}{|\sigma_k - \sigma_{h,\ell}|} \le \frac{2\sigma_k}{\mathrm{gap}_k} = \sqrt{2} q_k \sigma_k \ \forall\, \ell \notin I_{h,k} \ \forall\, h \in (0, h_0).$$

Proof. We will assume that $\sigma_k = \sigma_{k+1} = \cdots = \sigma_{k+n_k-1} > \sigma_{k+n_k}$ and either $k = 1$ or $\sigma_{k-1} > \sigma_k$. If we prove the result in this case, it obviously follows for any other value of k. Suppose first that $\ell \ge k + n_k$. Then $\sigma_{h,\ell} \le \sigma_\ell \le \sigma_{k+n_k} < \sigma_k$. Therefore,

$$\begin{aligned} \frac{\sigma_{h,\ell}(\sigma_k - \sigma_{k+n_k})}{\sigma_k(\sigma_k - \sigma_{h,\ell})} \le \frac{\sigma_{k+n_k}(\sigma_k - \sigma_{k+n_k})}{\sigma_k(\sigma_k - \sigma_{k+n_k})} < 1 &\Rightarrow \frac{\sigma_{h,\ell}\mathrm{gap}_k}{\sigma_k(\sigma_k - \sigma_{h,\ell})} < 1 \\ &\Rightarrow \frac{\sigma_{h,\ell}}{\sigma_k - \sigma_{h,\ell}} < \frac{\sigma_k}{\mathrm{gap}_k}. \end{aligned}$$

This proves the desired result in the case $\ell \ge k + n_k$. If $\ell < k$, then there exists $h_0' > 0$ such that $\sigma_{h,k-1} > \sigma_k$ for all $h \in (0, h_0')$. For such h,

$$\frac{\sigma_{h,\ell}}{|\sigma_k - \sigma_{h,\ell}|} = \frac{\sigma_{h,\ell}}{\sigma_{h,\ell} - \sigma_k} \le \frac{\sigma_{h,k-1}}{\sigma_{h,k-1} - \sigma_k}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

(using the fact that $s/(s-\sigma_k)$ increases as s decreases toward σ_k). Since

$$\frac{\sigma_{h,k-1}}{\sigma_{h,k-1}-\sigma_k} \to \frac{\sigma_{k-1}}{\sigma_{k-1}-\sigma_k} \text{ as } h \to 0$$

and

$$\frac{\sigma_{k-1}}{\sigma_{k-1}-\sigma_k} < \frac{2\sigma_k}{\mathrm{gap}_k}$$

by Lemma 13, it follows that there exists $h_0 \in (0, h_0')$ such that

$$\frac{\sigma_{h,k-1}}{\sigma_{h,k-1}-\sigma_k} \leq \frac{2\sigma_k}{\mathrm{gap}_k} \;\forall\, h \in (0, h_0).$$

Thus, the desired results holds in the case that $\ell < k$, and the proof is complete. ◻

We can now prove the desired bound on $\|T_h(I-P_{E_{h,k}})u\|_Y$ for $u \in E_k$.

THEOREM 15. *Let $k \in \mathbb{Z}^+$ be given, suppose $T_h = P_{Y_h} T P_{X_h}$, and let $u \in E_k$, $v \in F_k$ satisfy $Tu = \sigma_k v$. Then, for all $h > 0$ sufficiently small,*

$$\|T_h(u - P_{E_{h,k}}u)\|_Y \leq 2\sigma_k q_k \epsilon_h \left(\|(I-P_{Y_h})v\|_Y + \|(I-P_{X_h})u\|_X\right).$$

Proof. We will assume that $\sigma_k = \sigma_{k+1} = \cdots = \sigma_{k+n_k-1} > \sigma_{k+n_k}$ and either $k=1$ or $\sigma_{k-1} > \sigma_k$. As in the proof of Theorem 8, we have

$$(I - P_{E_{h,k}})u = \sum_{\ell \in \mathcal{J}_h \setminus I_{h,k}} \langle \phi_{h,\ell}, u\rangle_X \, \phi_{h,\ell},$$

which yields

$$\begin{aligned} T_h(u - P_{E_{h,k}}u) &= \sum_{\ell \in \mathcal{J}_h \setminus I_{h,k}} \sigma_{h,\ell} \langle \phi_{h,\ell}, u\rangle_X \, \psi_{h,\ell} \\ \Rightarrow \|T_h(u - P_{E_{h,k}}u)\|_Y^2 &= \sum_{\ell \in \mathcal{J}_h \setminus I_{h,k}} \sigma_{h,\ell}^2 \langle \phi_{h,\ell}, u\rangle_X^2 . \end{aligned}$$

Now we use the upper bound (11) and Lemma 14 to obtain

$$\begin{aligned} &\|T_h(I-P_{E_{h,k}})u)\|_Y^2 \\ &\leq \sum_{\ell \in \mathcal{J}_h \setminus I_{h,k}} \frac{\sigma_{h,\ell}^2}{|\sigma_k - \sigma_{h,\ell}|^2} \left(\left|\langle (T-T_h)\phi_{h,\ell}, v\rangle_Y\right| + \left|\langle \psi_{h,\ell}, (T-T_h)u\rangle_Y\right|\right)^2 \\ &\leq 2\sigma_k^2 q_k^2 \sum_{\ell \in \mathcal{J}_h \setminus I_{h,k}} \left(\left|\langle (T-T_h)\phi_{h,\ell}, v\rangle_Y\right| + \left|\langle \psi_{h,\ell}, (T-T_h)u\rangle_Y\right|\right)^2, \end{aligned}$$

where the last inequality holds for all $h > 0$ sufficiently small. Proceeding as in the proof of Theorem 11, we obtain

$$\|T_h(I-P_{E_{h,k}})u)\|_Y^2 \leq 4\sigma_k^2 q_k^2 \epsilon_h^2 \left(\|(I-P_{Y_h})v\|_Y^2 + \|(I-P_{X_h})u\|_X^2\right).$$

The desired result follows. ◻

We can now give our main result on the convergence of the singular values.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

THEOREM 16. *Let $k \in \mathbb{Z}^+$ be given, and suppose $T_h = P_{Y_h} T P_{X_h}$. For each $\ell \in I_k$, there exist $u \in E_k$, $v \in F_k$, and a constant $C_k > 0$ such that*

$$0 \leq \frac{\sigma_k - \sigma_{h,\ell}}{\sigma_k} \leq \frac{e_X^2}{2} + \frac{e_Y^2}{2} + C_k \epsilon_h \left(e_X + e_Y\right)^2 \;\; \forall\, h > 0 \text{ sufficiently small},$$

where $e_X = \|(I - P_{X_h})u\|_X$ and $e_Y = \|(I - P_{Y_h})v\|_Y$.

Proof. Let $\ell \in I_k$ be given, let $h > 0$ be sufficiently small that $\ell \in I_{h,k}$, and choose $\phi_{h,\ell} \in E_{h,k}$, $\psi_{h,\ell} \in F_{h,k}$ such that

$$\|\phi_{h,\ell}\|_X = \|\psi_{h,\ell}\|_Y = 1 \text{ and } T_h \phi_{h,\ell} = \sigma_{h,\ell} \psi_{h,\ell}.$$

We note that $\sigma_{h,\ell} \leq \sigma_k$ by Theorem 6. We must derive an upper bound on $\sigma_k - \sigma_{h,\ell}$. For $h > 0$ sufficiently small, $\delta(E_{h,k}, E_k) = \delta(E_k, E_{h,k}) < 1$, which implies that $P_{E_{h,k}}$ defines a bijection from E_k onto $E_{h,k}$ (see the proof of Lemma 2). Thus, there exists $u \in E_k$ such that $\|u\|_X = 1$ and $P_{E_{h,k}} u = \alpha \phi_{h,\ell}$ for some $\alpha \in (0, 1]$. Define $v \in F_k$ by $Tu = \sigma_k v$ and note that $\|v\|_Y = 1$. As in the statement of the theorem, we will write

$$e_X = \|(I - P_{X_h})u\|_X, \; e_Y = \|(I - P_{Y_h})v\|_Y.$$

We have

$$\begin{aligned}
\sigma_k &= \langle v, Tu \rangle_Y \\
&= \left\langle P_{F_{h,k}} v + (I - P_{F_{h,k}})v, T(P_{E_{h,k}} u + (I - P_{E_{h,k}})u) \right\rangle_Y \\
&= \left\langle P_{F_{h,k}} v, T P_{E_{h,k}} u \right\rangle_Y + \left\langle (I - P_{F_{h,k}})v, T P_{E_{h,k}} u \right\rangle_Y \\
&\quad + \left\langle P_{F_{h,k}} v, T(I - P_{E_{h,k}})u \right\rangle_Y + \left\langle (I - P_{F_{h,k}})v, T(I - P_{E_{h,k}})u \right\rangle_Y .
\end{aligned}$$

We now consider each of these four inner products. For the first, we have

$$\left\langle P_{F_{h,k}} v, T P_{E_{h,k}} u \right\rangle_Y = \alpha \left\langle P_{F_{h,k}} v, T_h \phi_{h,\ell} \right\rangle_Y = \alpha \sigma_{h,\ell} \left\langle P_{F_{h,k}} v, \psi_{h,\ell} \right\rangle_Y .$$

By the Pythagorean theorem and Taylor's theorem,

$$\begin{aligned}
\alpha = \|P_{E_{h,k}} u\|_X &= \sqrt{1 - \|(I - P_{E_{h,k}})u\|_X^2} \\
&= 1 - \frac{\|(I - P_{E_{h,k}})u\|_X^2}{2} + O(\|(I - P_{E_{h,k}})u\|_X^4) \\
&\leq 1 - \frac{e_X^2}{2} + O(\|(I - P_{E_{h,k}})u\|_X^4) \\
&= 1 - \frac{e_X^2}{2} + O(e_X^4),
\end{aligned}$$

where the bound from Theorem 11 is used in the last step. Similarly,

$$\left\langle P_{F_{h,k}} v, \psi_{h,\ell} \right\rangle_Y \leq \|P_{F_{h,k}} v\|_Y \leq 1 - \frac{e_Y^2}{2} + O(e_Y^4).$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

It follows that

$$\begin{aligned}
\left\langle P_{F_{h,k}}v, TP_{E_{h,k}}u\right\rangle_Y &\leq \sigma_{h,\ell}\left(1-\frac{e_X^2}{2}+O(e_X^4)\right)\left(1-\frac{e_Y^2}{2}+O(e_Y^4)\right)\\
&= \sigma_{h,\ell}-\left(\frac{e_X^2}{2}+\frac{e_Y^2}{2}+O(e_X^4+e_Y^4)\right)\sigma_{h,\ell}\\
&= \sigma_{h,\ell}-\left(\frac{e_X^2}{2}+\frac{e_Y^2}{2}+O(e_X^4+e_Y^4)\right)\sigma_k\\
&\quad +\left(\frac{e_X^2}{2}+\frac{e_Y^2}{2}+O(e_X^4+e_Y^4)\right)(\sigma_k-\sigma_{h,\ell})\\
&= \sigma_{h,\ell}-\left(\frac{e_X^2}{2}+\frac{e_Y^2}{2}\right)\sigma_k+O(\sigma_k\epsilon_h(e_X+e_Y)^2)
\end{aligned}$$

(using the fact that $\sigma_k-\sigma_{h,\ell}\leq\epsilon_h$ and $e_X, e_Y = O(\epsilon_h)$ by Theorems 6 and 8).

To bound the second inner product, notice that $\left\langle (I-P_{F_{h,k}})v, T_hP_{E_{h,k}}u\right\rangle_Y = 0$ because $(I-P_{F_{h,k}})v\in F_{h,k}^{\perp}$ and $T_hP_{E_{h,k}}u\in F_{h,k}$. Thus,

$$\begin{aligned}
\left\langle (I-P_{F_{h,k}})v, TP_{E_{h,k}}u\right\rangle_Y &= \left\langle (I-P_{F_{h,k}})v, TP_{E_{h,k}}u-T_hP_{E_{h,k}}u\right\rangle_Y\\
&= \left\langle (I-P_{F_{h,k}})v, (I-P_{Y_h})TP_{E_{h,k}}u\right\rangle_Y\\
&= \left\langle (I-P_{Y_h})v, (I-P_{Y_h})TP_{E_{h,k}}u\right\rangle_Y\\
&\leq e_Y\|(I-P_{Y_h})TP_{E_{h,k}}u\|_Y\\
&\leq e_Y\left(\|(I-P_{Y_h})Tu\|_Y+\|(I-P_{Y_h})T(I-P_{E_{h,k}})u\|_Y\right)\\
&\leq e_Y\left(\sigma_k e_Y+\epsilon_h\|(I-P_{E_{h,k}})u\|_X\right)\\
&\leq \sigma_k e_Y\left(e_Y+\frac{\epsilon_h}{\sigma_k}\|(I-P_{E_{h,k}})u\|_X\right)\\
&\leq \sigma_k e_Y\left(e_Y+\epsilon_h q_k\|(I-P_{E_{h,k}})u\|_X\right)\\
&= \sigma_k e_Y^2+O\left(\sigma_k\epsilon_h(e_X+e_Y)^2\right)
\end{aligned}$$

(using Theorem 11 and the fact that $\|T-P_{Y_h}T\|_{\mathcal{L}(X,Y)}\leq\|T-T_h\|_{\mathcal{L}(X,Y)}=\epsilon_h$).

We use similar reasoning to bound the third inner product as follows:

$$\left\langle P_{F_{h,k}}v, T(I-P_{E_{h,k}})u\right\rangle_Y\leq\sigma_k e_X^2+O\left(\sigma_k\epsilon_h(e_X+e_Y)^2\right).$$

Finally, for the fourth inner product, we use Theorem 15 to obtain

$$\begin{aligned}
&\left\langle (I-P_{F_{h,k}})v, T(I-P_{E_{h,k}})u\right\rangle_Y\\
&= \left\langle (I-P_{F_{h,k}})v, T_h(I-P_{E_{h,k}})u\right\rangle_Y+\left\langle (I-P_{F_{h,k}})v, (T-T_h)(I-P_{E_{h,k}})u\right\rangle_Y\\
&\leq \|(I-P_{F_{h,k}})v\|_Y\|T_h(I-P_{E_{h,k}})u\|_Y\\
&\qquad +\|(I-P_{F_{h,k}})v\|_Y\|T-T_h\|_{\mathcal{L}(X,Y)}\|(I-P_{E_{h,k}})u\|_X\\
&\leq 2\sigma_k q_k\epsilon_h\|(I-P_{F_{h,k}})v\|_Y\,(e_Y+e_X)+\sigma_k\left(\frac{\epsilon_h}{\sigma_k}\right)\|(I-P_{F_{h,k}})v\|_Y\|(I-P_{E_{h,k}})u\|_X\\
&= O\left(\sigma_k\epsilon_h(e_X+e_Y)^2\right).
\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

We thus obtain

$$\sigma_k \le \sigma_{h,\ell} - \sigma_k \left(\frac{e_X^2}{2} + \frac{e_Y^2}{2} \right) + \sigma_k e_X^2 + \sigma_k e_Y^2 + O\left(\sigma_k \epsilon_h (e_X + e_Y)^2 \right)$$
$$= \sigma_{h,\ell} + \sigma_k \left(\frac{e_X^2}{2} + \frac{e_Y^2}{2} \right) + O\left(\sigma_k \epsilon_h (e_X + e_Y)^2 \right)$$

and hence

$$0 \le \frac{\sigma_k - \sigma_{h,\ell}}{\sigma_k} \le \frac{e_X^2}{2} + \frac{e_Y^2}{2} + O\left(\epsilon_h (e_X + e_Y)^2 \right),$$

as desired. ∎

Since $e_X = \|(I - P_{X_h})u\|_X$ and $e_Y = \|(I - P_{Y_h})v\|_X$ are both $O(\epsilon_h)$ by Theorem 8, Theorem 16 implies that the relative error in each computed singular value is $O(\epsilon_h^2)$ when $T_h = P_{Y_h} T P_{X_h}$ (as opposed to $O(\epsilon_h)$ for a general approximation T_h of T).

4. Numerical experiments. The first example demonstrates the convergence guaranteed by Theorems 6 and 8.

Example 1. Consider the first-kind integral operator $T : L^2(0,1) \to L^2(0,1)$ defined by

$$(Tx)(s) = \int_0^1 k(s,t)x(t)\,dt, \ 0 < s < 1,$$

where $k(s,t) = se^{st}$. We will use the techniques described in this paper, with X_h and Y_h chosen to be finite element spaces, to estimate the SVE of T. Since the kernel is smooth, Chebyshev approximation allows for a sequence of approximations that converge exponentially quickly (see [13] and [18]). The finite element approximations described here do not lead to a competitive algorithm, but they serve to illustrate the convergence theorems of this paper.

To apply the above results, we define $X_h = Y_h$ to be the space of continuous piecewise polynomial functions relative to the uniform mesh on $[0,1]$ with $1/h$ elements. We denote the nodes by $t_0, t_1, \ldots, t_n$ (where $t_j = j/n$ for each j) and use the standard nodal basis $\{x_0, x_1, \ldots, x_n\}$ (defined by $x_i(t_j) = \delta_{ij}$). Note that $n = d/h$ for piecewise polynomials of degree d. We discretize the integral operator by interpolating the kernel k onto the tensor-product finite element space

$$Y_h \otimes X_h = \mathrm{sp}\{z_{k\ell} : 0 \le k, \ell \le n\},$$

where $z_{k\ell}$ is defined by $z_{k\ell}(s,t) = x_k(s)x_\ell(t)$. We then define $T_h : X_h \to Y_h$ by

$$(T_h x)(s) = \int_0^1 k_h(s,t)x(t)\,dt, \ 0 < s < 1,$$

where k_h is the interpolated kernel:

$$k_h(s,t) = \sum_{k=0}^{n} \sum_{\ell=0}^{n} k(t_k, t_\ell) x_k(s) x_\ell(t).$$

Standard finite element approximation results can be used to show that, for piecewise polynomials of degree d,

$$\|k_h - k\|_{L^2((0,1)\times(0,1))} \le Ch^{d+1}.$$

It follows immediately that $\|T_h - T\|_{\mathcal{L}(L^2(0,1), L^2(0,1))} \le Ch^{d+1}$, and therefore we expect $\sigma_{h,k}$, $E_{h,k}$, and $F_{h,k}$ to converge with an error of $O(h^{d+1})$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 1
The computed singular values $\sigma_{h,1}, \sigma_{h,2}, \sigma_{h,3}$ for Example 1. Richardson extrapolation is used to estimate p such that $|\sigma_{h,k} - \sigma_k| = O(h^p)$ appears to hold; the estimated exponents p_{est} appear in parentheses.

h	$\sigma_{h,1}(p_{\mathrm{est}})$	$\sigma_{h,2}(p_{\mathrm{est}})$	$\sigma_{h,3}(p_{\mathrm{est}})$
1/8	$8.95937 \cdot 10^{-1}$	$4.25077 \cdot 10^{-2}$	$1.19714 \cdot 10^{-3}$
1/16	$8.93464 \cdot 10^{-1}$	$4.26331 \cdot 10^{-2}$	$1.22087 \cdot 10^{-3}$
1/32	$8.92847 \cdot 10^{-1}(2.00)$	$4.26627 \cdot 10^{-2}(2.09)$	$1.22608 \cdot 10^{-3}(2.19)$
1/64	$8.92693 \cdot 10^{-1}(2.00)$	$4.26700 \cdot 10^{-2}(2.02)$	$1.22734 \cdot 10^{-3}(2.05)$
1/128	$8.92655 \cdot 10^{-1}(2.00)$	$4.26718 \cdot 10^{-2}(2.00)$	$1.22765 \cdot 10^{-3}(2.01)$
1/256	$8.92645 \cdot 10^{-1}(2.00)$	$4.26722 \cdot 10^{-2}(2.00)$	$1.22773 \cdot 10^{-3}(2.00)$
1/512	$8.92643 \cdot 10^{-1}(2.00)$	$4.26723 \cdot 10^{-2}(2.00)$	$1.22775 \cdot 10^{-3}(2.00)$
1/1024	$8.92642 \cdot 10^{-1}(2.00)$	$4.26724 \cdot 10^{-2}(2.00)$	$1.22775 \cdot 10^{-3}(2.00)$

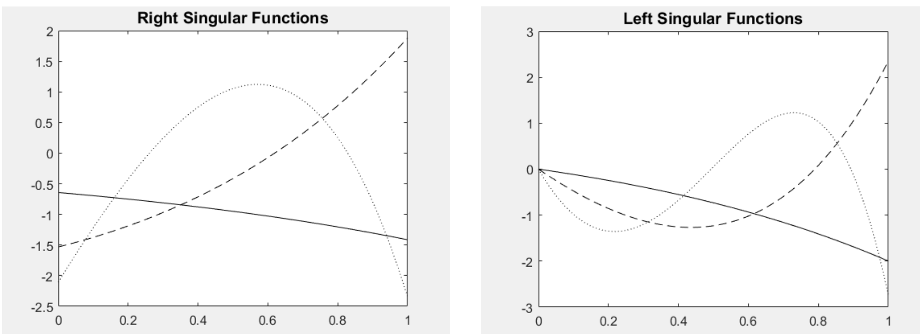

FIG. 1. *The computed singular functions $\phi_{h,1}, \phi_{h,2}, \phi_{h,3}$ (left) and $\psi_{h,1}, \psi_{h,2}, \psi_{h,3}$ (right). The solid curves represent $\phi_{h,1}$ and $\psi_{h,1}$, the dashed curves $\phi_{h,2}$ and $\psi_{h,2}$, and the dotted curves $\phi_{h,3}$ and $\psi_{h,3}$.*

It is easy to show that the Galerkin matrix A is given by $A = MCM$, where M is the Gram matrix for the basis $\{x_0, x_1, \ldots, x_n\}$ and C is defined by $C_{k\ell} = k(t_k, t_\ell)$. It follows that it is simple to implement this particular discretization.

Table 1 shows the computed values of $\sigma_{h,1}$, $\sigma_{h,2}$, $\sigma_{h,3}$ for $h = 2^{-3}, 2^{-4}, \ldots, 2^{-10}$, using piecewise linear functions ($d = 1$). The exact values are unknown, but we use Richardson extrapolation to estimate an exponent p such that the error appears to converge to zero like $O(h^p)$. As expected, the results suggest that the errors are $O(h^2)$. Although we do not show the results, the same method suggests that the singular functions $\{\phi_{h,k}\}$ and $\{\psi_{h,k}\}$ converge at the same rate. (The singular spaces all appear to be one-dimensional.) Figure 1 shows the first three right and left singular functions.

The next experiment illustrates the accelerated convergence guaranteed by Theorem 16.

Example 2. We now repeat Example 1 but using $T_h = P_{Y_h} T P_{X_h}$ to approximate T. We are required to compute the Galerkin matrix A defined by

$$A_{k\ell} = \langle y_k, T_h x_\ell \rangle_{L^2(0,1)} = \langle y_k, T x_\ell \rangle_{L^2(0,1)}$$
$$= \int_0^1 \int_0^1 k(s,t) x_\ell(t) y_k(s)\, dt\, ds.$$

We use a tensor-product Gauss quadrature rule to compute the entries of A to high accuracy. For an integral operator such as T, it is straightforward to show that $T_h =$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 2

The computed singular values $\sigma_{h,1}, \sigma_{h,2}, \sigma_{h,3}$ for Example 2 (piecewise linear functions). Richardson extrapolation is used to estimate p such that $|\sigma_{h,k} - \sigma_k| = O(h^p)$ appears to hold; the estimated exponents p_{est} appear in parentheses.

h	$\sigma_{h,1}(p_{\text{est}})$	$\sigma_{h,2}(p_{\text{est}})$	$\sigma_{h,3}(p_{\text{est}})$
1/8	$8.92640 \cdot 10^{-1}$	$4.26673 \cdot 10^{-2}$	$1.22572 \cdot 10^{-3}$
1/16	$8.92642 \cdot 10^{-1}$	$4.26720 \cdot 10^{-2}$	$1.22763 \cdot 10^{-3}$
1/32	$8.92642 \cdot 10^{-1}(4.00)$	$4.26723 \cdot 10^{-2}(4.00)$	$1.22775 \cdot 10^{-3}(3.99)$
1/64	$8.92642 \cdot 10^{-1}(4.00)$	$4.26724 \cdot 10^{-2}(4.00)$	$1.22776 \cdot 10^{-3}(4.00)$
1/128	$8.92642 \cdot 10^{-1}(4.00)$	$4.26724 \cdot 10^{-2}(4.00)$	$1.22776 \cdot 10^{-3}(4.00)$
1/256	$8.92642 \cdot 10^{-1}(4.00)$	$4.26724 \cdot 10^{-2}(4.00)$	$1.22776 \cdot 10^{-3}(4.00)$
1/512	$8.92642 \cdot 10^{-1}(4.00)$	$4.26724 \cdot 10^{-2}(4.00)$	$1.22776 \cdot 10^{-3}(4.00)$
1/1024	$8.92642 \cdot 10^{-1}(4.00)$	$4.26724 \cdot 10^{-2}(4.00)$	$1.22776 \cdot 10^{-3}(4.00)$

TABLE 3

The computed singular values $\sigma_{h,1}, \sigma_{h,2}, \sigma_{h,3}$ for Example 2 (piecewise quadratic functions). Richardson extrapolation is used to estimate p such that $|\sigma_{h,k} - \sigma_k| = O(h^p)$ appears to hold; the estimated exponents p_{est} appear in parentheses. With piecewise quadratic approximations, the errors quickly reach the level of machine epsilon, and so the estimates of p deteriorate as the mesh is refined.

h	$\sigma_{h,1}(p_{\text{est}})$	$\sigma_{h,2}(p_{\text{est}})$	$\sigma_{h,3}(p_{\text{est}})$
1/8	$8.92642 \cdot 10^{-1}$	$4.26724 \cdot 10^{-2}$	$1.22775 \cdot 10^{-3}$
1/16	$8.92642 \cdot 10^{-1}$	$4.26724 \cdot 10^{-2}$	$1.22776 \cdot 10^{-3}$
1/32	$8.92642 \cdot 10^{-1}(5.93)$	$4.26724 \cdot 10^{-2}(5.91)$	$1.22776 \cdot 10^{-3}(5.86)$
1/64	$8.92642 \cdot 10^{-1}(5.98)$	$4.26724 \cdot 10^{-2}(5.95)$	$1.22776 \cdot 10^{-3}(5.93)$
1/128	$8.92642 \cdot 10^{-1}(6.27)$	$4.26724 \cdot 10^{-2}(5.98)$	$1.22776 \cdot 10^{-3}(5.96)$
1/256	$8.92642 \cdot 10^{-1}(1.91)$	$4.26724 \cdot 10^{-2}(7.38)$	$1.22776 \cdot 10^{-3}(6.00)$
1/512	$8.92642 \cdot 10^{-1}(-4.15)$	$4.26724 \cdot 10^{-2}(-2.39)$	$1.22776 \cdot 10^{-3}(4.79)$

$P_{Y_h} T P_{X_h}$ is the integral operator defined by the kernel $\hat{k}_h$, where $\hat{k}_h$ is the projection (in the L^2 inner product) of the true kernel k onto the tensor-product space $Y_h \otimes X_h$ (see [9]). It follows that $\|\hat{k}_h - k\|_{L^2((0,1)\times(0,1))}$ is no greater than $\|k_h - k\|_{L^2((0,1)\times(0,1))}$ (where k_h is the interpolated kernel used in Example 1), and numerical evidence suggests that the asymptotic rate is the same: $\|\hat{k}_h - k\|_{L^2((0,1)\times(0,1))} = O(h^{d+1})$ if piecewise polynomials of degree d are used. By Theorem 16, then, we expect the computed singular values to converge at a rate of $O((h^{d+1})^2) = O(h^{2d+2})$.

Table 2 shows the computed values of $\sigma_{h,1}, \sigma_{h,2}, \sigma_{h,3}$ for $h = 2^{-3}, 2^{-4}, \ldots, 2^{-10}$, using piecewise linear functions ($d = 1$). As expected, the errors are consistent with $O(h^4)$ convergence. Table 3 shows the analogous results for piecewise quadratic polynomials ($d = 2$). In this case, the results are consistent with $O(h^6)$ convergence, as predicted by Theorem 16. (With piecewise quadratic functions, the singular values converge quickly enough that the errors reach the level of round-off error in our computations. This is reflected in the fact that the estimated exponent is not as close to $2d + 2$ as in the linear case and even becomes negative in some cases.)

With both piecewise linear and piecewise quadratic functions, the computed singular functions are consistent with $O(h^{d+1})$ convergence; the increased rate of convergence applies only to the singular values.

We also note that the first nine (that is, the nine largest) singular values of T are approximately

$$8.926 \cdot 10^{-1},\ 4.267 \cdot 10^{-2},\ 1.228 \cdot 10^{-3},\ 2.434 \cdot 10^{-5},\ 3.685 \cdot 10^{-7},$$
$$4.507 \cdot 10^{-9},\ 4.621 \cdot 10^{-11},\ 4.076 \cdot 10^{-13},\ 3.15 \cdot 10^{-15}.$$

Using piecewise linear functions and a uniform mesh with 256 elements, we are able

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

to estimate these values accurately (error approximately 1% in σ_9 and much less than 1% for $\sigma_1, \ldots, \sigma_8$) by computing the SVE of $T_h = P_{Y_h} T P_{X_h}$. Using the eigenvalue approach and the same mesh, it is possible only to estimate the first five singular values, with the error in $\sigma_5 \approx 3.685 \cdot 10^{-7}$ already about 0.6%.

Finally, as we have noted, Theorem 11 does not guarantee an optimal rate of convergence for the estimates of the singular vectors, at least not in all scenarios. The following example, in which we use different finite element spaces for X_h and Y_h, suggests that Theorem 11 correctly predicts the observed rate of convergence (optimal or suboptimal).

Example 3. Let $T : L^2(0,1) \to L^2(0,1)$ be defined by

$$(Tx)(s) = \int_0^1 k(s,t)x(t)\,dt,\ 0 < s < 1,$$

where k is the discontinuous kernel defined as follows:

$$k(s,t) = \begin{cases} s^2 - t, & s \le t, \\ se^{st}, & t < s. \end{cases}$$

We use continuous piecewise polynomials of degree $p-1$ and $q-1$ for X_h and Y_h, respectively, yielding the following rates of convergence:

$$\begin{aligned} \text{Right singular vectors: } & \|(I - P_{X_h})u\|_X = O(h^p); \\ \text{Left singular vectors: } & \|(I - P_{Y_h})v\|_Y = O(h^q). \end{aligned}$$

Since k is discontinuous, $\epsilon_h = O(h)$. Theorems 11 and 16 suggest that we should observe

$$\begin{aligned} \|(I - P_{E_{h,k}})u\|_X &= O\left(h^r\right),\ r = \min\{p, q+1\}, \\ \|(I - P_{F_{h,k}})v\|_Y &= O\left(h^s\right),\ s = \min\{q, p+1\}, \\ \frac{\sigma_k - \sigma_{h,k}}{\sigma_k} &= O\left(h^t\right),\ t = 2\min\{r, s\} = 2\min\{p, q\}. \end{aligned}$$

Table 4 presents the results of our numerical experiments for different values of p and q; the results are fully consistent with the predictions of Theorems 11 and 16. For illustration, we use $k = 1$; specifically, we estimate $u = \phi_1$ and $v = \psi_1$ on uniform meshes with 16, 32, and 64 elements. We then use Richardson extrapolation to estimate r, s, and t so that

$$\|(I - P_{E_{h,1}})u\|_X = O(h^r),\ \|(I - P_{F_{h,1}})v\|_Y = O(h^s),\ \text{and}\ \frac{\sigma_k - \sigma_{h,k}}{\sigma_k} = O(h^t).$$

Although we do not show the results here, we observe the same behavior for various values of k, except that, as k increases, finer meshes are needed to observe predicted rates of convergence.

5. Conclusions. The singular values and singular vectors of a compact operator can be estimated using a variety of discretizations. In the generic case, where the operator T is approximated by a family $\{T_h : h > 0\}$ of discretized operators, the errors in both singular values and singular vectors go to zero at least as fast as does $\|T_h - T\|_{\mathcal{L}(X,Y)}$. With the special choice of $T_h = P_{Y_h} T P_{X_h}$, the error in the computed singular vectors can be expressed in terms of the optimal approximation errors, but

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Table 4

Observed rates of convergence of the first right and left singular vectors for different discretizations. In each case, the observed rate of convergence agrees with the prediction of Theorem 11. *Suboptimal rates of convergence are indicated in boldface. The last column shows the rate of convergence for the first singular value; the results agree with Theorem* 16.

		$\|(I-P_{E_{h,1}})u\|_X$	$\|(I-P_{F_{h,1}})v\|_Y$	$\frac{\sigma_1-\sigma_{h,1}}{\sigma_1}$
p	q	observed r	observed s	observed t
2	2	2.0213	2.0057	4.0217
2	3	2.0201	2.9772	4.0450
3	2	2.8997	2.0061	4.0117
2	4	2.0201	**3.1194**	4.0395
4	2	**3.0198**	2.0061	4.0113
4	3	3.9531	2.9552	5.9103
5	2	**3.0194**	2.0061	4.0113
3	5	2.8427	**3.7315**	5.6847
5	3	**3.8940**	2.9552	5.9096
6	2	**3.0195**	2.0061	4.0113

the approximability of both left and right singular vectors affects the error in either the computed left singular vectors or the computed right singular vectors. In many cases, the error in the computed singular vector is asymptotically optimal, but a poor degree of approximability in the left singular vectors, for example, can cause the error in the computed right singular vectors to be suboptimal. This is suggested by the bounds in Theorem 11 and confirmed by the numerical examples in section 4. Still considering the case of $T_h = P_{Y_h}TP_{X_h}$, Theorem 16 shows that the computed singular values converge at an increased rate. The typical case is that the error goes to zero like the square of the optimal approximation error for the singular vectors, but once again the approximability of both left and right singular vectors must be taken into account.

Although the approximation of the singular values and singular vectors of T is related to the approximation of the eigenvalues and eigenvectors of T^*T, the fact that both left and right singular vectors (which may have different degrees of approximability) are involved means that the situation cannot be understood by simply transferring the results of the eigenvalue theory found in [5], [2], [3], and other references) to the singular value problem. Moreover, as discussed in section 1, there are advantages to computing the singular values and singular vectors directly rather than by formulating a related eigenvalue problem, particularly the fact that we can approximate smaller singular values accurately.

Appendix. We now show how to compute the SVE of $T_h = P_{Y_h}TP_{X_h}$. It is clear that the right singular vectors of T_h belong to X_h and the left singular vectors to Y_h. Therefore, it suffices to show how to compute the SVE of an operator of the form $\hat{T} : \hat{X} \to \hat{Y}$, where $\hat{X}$ and $\hat{Y}$ are finite-dimensional subspaces of X and Y, respectively.

The following lemma will be useful.

Lemma 17. *Let $\{x_1, x_2, \ldots, x_n\}$ be a basis for a finite-dimensional inner product space $\hat{X}$, and let $M \in \mathbb{R}^{n\times n}$ be the Gram matrix for this basis. Then $\{\phi_1, \phi_2, \ldots, \phi_n\}$ is an orthonormal basis for $\hat{X}$ if and only if*

$$\phi_j = \sum_{i=1}^{n} U_{ij}x_i,\ j = 1, 2, \ldots, n, \tag{18}$$

where $M^{1/2}U$ is an orthogonal matrix.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Proof. Straightforward. ∎

We can now state the desired theorem.

THEOREM 18. *Let $\hat{X}$ and $\hat{Y}$ be finite-dimensional subspaces of X and Y, respectively, and let $\{x_1, x_2, \ldots, x_n\}$ and $\{y_1, y_2, \ldots, y_m\}$ be bases for $\hat{X}$ and $\hat{Y}$, respectively. Suppose $\hat{T} : \hat{X} \to \hat{Y}$ is linear, and let $A \in \mathbb{R}^{m\times n}$ be the Galerkin matrix defined by*

$$A_{k\ell} = \left\langle y_k, \hat{T}x_\ell \right\rangle_Y .$$

Let $H^{-1/2}AM^{-1/2} = U\Sigma V^T$ be an SVD of the matrix $\hat{A} = H^{-1/2}AM^{-1/2}$, where M and H are the Gram matrices for $\{x_1, x_2, \ldots, x_n\}$ and $\{y_1, y_2, \ldots, y_m\}$, respectively, and define $\hat{V} = M^{-1/2}V$ and $\hat{U} = H^{-1/2}U$. Then an SVE of $\hat{T}$ is given by

$$\hat{T} = \sum_{k=1}^{r} \hat{\sigma}_k \hat{\psi}_k \otimes \hat{\phi}_k, \tag{19}$$

where $\hat{\sigma}_1, \hat{\sigma}_2, \ldots, \hat{\sigma}_r$ are the nonzero singular values of $\hat{A}$ and

$$\begin{aligned} \hat{\phi}_\ell &= \sum_{k=1}^{n} \hat{V}_{k\ell} x_k,\ \ \ell = 1, 2, \ldots, n, \\ \hat{\psi}_\ell &= \sum_{k=1}^{m} \hat{U}_{k\ell} y_k,\ \ \ell = 1, 2, \ldots, m. \end{aligned} \tag{20}$$

Proof. By Lemma 17, (20) defines orthonormal bases

$$\{\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_n\},\ \ \{\hat{\psi}_1, \hat{\psi}_2, \ldots, \hat{\psi}_m\}$$

of $\hat{X}$, $\hat{Y}$, respectively. It remains only to verify that (19) holds, which can be done by showing directly that

$$\hat{T}x = \left(\sum_{k=1}^{r} \hat{\sigma}_k \hat{\psi}_k \otimes \hat{\phi}_k \right) x = \sum_{k=1}^{r} \hat{\sigma}_k \left\langle \hat{\phi}_k, x \right\rangle_X \hat{\psi}_k \ \forall\, x \in \hat{X}. \qquad ∎$$

As discussed above, it might be desirable to compute the SVE of $\hat{T} : \hat{X} \to \hat{Y}$, where $\hat{X}$ is a given finite-dimensional subspace of X, $\hat{Y} = T(\hat{X})$, and $\hat{T}x = Tx$ for all $x \in \hat{X}$. We suppose first that $\{x_1, x_2, \ldots, x_n\}$ is a basis for $\hat{X}$ and that $\{y_1, y_2, \ldots, y_n\}$, where $y_i = Tx_i$ for $i = 1, 2, \ldots, n$, is linearly independent and hence is a basis for $\hat{Y} = T(X_h)$. In this case, the Galerkin matrix A is defined by

$$A_{ij} = \langle y_i, Tx_j \rangle_Y = \langle Tx_i, Tx_j \rangle_Y$$

and coincides with the Gram matrix for the basis $\{y_1, y_2, \ldots, y_n\}$. Since A is likely to be dense (even if X_h is a finite element space), the square root $A^{1/2}$ required by Theorem 18 is expensive to compute when n is large. For this reason, we use the Cholesky factorization instead (as, indeed, we could have done in Theorem 18). Let $A = LL^T$ and $M = NN^T$ be Cholesky factorizations of A and M, respectively (so that L and N are lower triangular). We then define $\hat{A} = L^{-1}AN^{-T} = L^TN^{-T}$ and

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

compute an SVD of $\hat{A}$: $L^T N^{-T} = U\Sigma V^T$. Defining $\hat{V} = N^{-T}V$ and $\hat{U} = L^{-T}U$, it is easy to show that

$$\hat{T} = \sum_{k=1}^{r} \hat{\sigma}_k \hat{\psi}_k \otimes \hat{\phi}_k,$$

where $\hat{\sigma}_1, \hat{\sigma}_2, \ldots, \hat{\sigma}_r$ are the nonzero singular values of $\hat{A}$ and

$$\hat{\phi}_\ell = \sum_{k=1}^{n} \hat{V}_{k\ell} x_k,\ \ \ell = 1, 2, \ldots, n,$$
$$\hat{\psi}_\ell = \sum_{k=1}^{n} \hat{U}_{k\ell} y_k,\ \ \ell = 1, 2, \ldots, n.$$

The above scheme is likely to be effective if $\mathcal{N}(T) \cap \hat{X}$ is trivial and the singular vectors of T go to zero slowly enough to allow a sufficiently fine discretization of X for accurate approximation of the left singular vectors while maintaining the positive definiteness of A. For many operators T, though, the Galerkin matrix A will be singular for a reasonable discretization $\hat{X}$ of X. In this case, a more careful algorithm is needed.

Acknowledgments. The authors would like to acknowledge the helpful advice of two anonymous referees whose suggestions resulted in a significantly improved paper. We also thank Professor Nick Trefethen for a helpful conversation.

REFERENCES

[1] P. M. Anselone, *Collectively Compact Operator Approximation Theory and Applications to Integral Equations*, Prentice Hall, Englewood Cliffs, NJ, 1971.

[2] I. Babuška and J. Osborn, *Finite element Galerkin approximation of the eigenvalues and eigenvectors of selfadjoint problems*, Math. Comp., 52 (1989), pp. 275–297.

[3] I. Babuška and J. Osborn, *Eigenvalue problems*, in Handbook of Numerical Analysis. Volume 2: Finite Element Methods: Part 1, P. G. Ciarlet and J. L. Lions, eds., Elsevier, Amsterdam, 1991, pp. 641–787.

[4] D. Boffi, *Finite element approximation of eigenvalue problems*, Acta Numer., 19 (2010), pp. 1–120.

[5] F. Chatelin, *Spectral Approximation of Linear Operators*, SIAM, Philadelphia, 2011.

[6] M. S. Gockenbach and M. J. Roberts, *Approximating the generalized singular value expansion*, SIAM J. Numer. Anal., 56 (2018), pp. 2776–2795.

[7] I. Gohberg, S. Goldberg, and M. A. Kaashoek, *Classes of Linear Operators*, Vol. 1, Birkhauser Verlag, Basel, 1990.

[8] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 4th ed., Johns Hopkins University Press, Baltimore, MD, 2013.

[9] P. C. Hansen, *Computation of the Singular Value Expansion*, Computing, 40 (1988), pp. 185–199.

[10] E. Hille and J. D. Tamarkin, *On the characteristic values of linear integral equations*, Acta Math., 57 (1931), pp. 1–76.

[11] T. Kato, *Perturbation theory for nullity, deficiency and other quantities of linear operators*, J. Anal. Math., 6 (1958), pp. 261–322.

[12] P. D. Lax, *Functional Analysis*, John Wiley & Sons, New York, 2002.

[13] G. Little and J. B. Reade, *Eigenvalues of analytic kernels*, SIAM J. Math. Anal., 15 (1984), pp. 133–136.

[14] J. Mercer, *Functions of positive and negative type, and their connection the theory of integral equations*, Philos. Trans. Roy. Soc. A, 209.441-458 (1909), pp. 415–446.

[15] E. Schmidt, *Zur Theorie der linearen und nichtlinearen Integralgleichungen.* I. *Teil: Entwicklung willkiirlichen funktionen nach System vorgeschriebener*, Math. Ann., 63 (1907), pp. 433–476.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[16] F. Smithies, *The eigen-values and singular values of integral equations*, Proc. Lond. Math. Soc., 43 (1937), pp. 255–279.

[17] G. W. Stewart, *On the early history of the singular value decomposition*, SIAM Rev., 35 (1993), pp. 551–566.

[18] A. Townsend and L. N. Trefethen, *Continuous analogues of matrix factorizations*, Proc. Roy. Soc. A, 471.2173 (2015), 20140585.

[19] H. Weyl, *Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung)*, Math. Ann., 71 (1912), pp. 441–479.

Downloaded 05/07/20 to 144.173.6.94. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.