Approximation of linear operators in the 2-norm *

A.C. Antoulas

Department of Electrical and Computer Engineering
Rice University
Houston, Texas 77251-1892, USA

e-mail: aca@rice.edu - fax: +1-713-524-5237

September 1, 1997

Abstract

The problems of approximating in the 2-induced norm, linear oparators which are (1) finite-dimensional, unstructured, and (2) infinite-dimensional structured (Hankel), have been solved. The solutions of these two problems exhibit striking similarities. These similarities suggest the search of a *unifying framework* for the approximation of linear operators in the 2-induced norm.

*Note submitted to LAA for the *Challenges in Matrix Theory* Issue.

1 Finite-dimensional operators

Given a matrix $A \in \mathbb{R}^{n\times n}$, there exist unitary matrices $U, V \in \mathbb{R}^{n\times n}$, such that

$$A = U\Sigma V^* \quad \text{where} \quad \Sigma = \text{diag}\,(\sigma_1, \; \cdots \; , \sigma_n) \in \mathbb{R}^{n\times n}, \;\; \sigma_1(A) \geq \cdots \geq \sigma_n(A) \geq 0 \tag{1.1}$$

This is the *singular value decomposition (SVD)* of the matrix A; the σ_i's are the *singular values* of A; $\sigma_1(A)$ is the 2-*induced norm* of A. The columns of U, V

$$U = (u_1 \;\; u_2 \;\; \cdots \;\; u_n), \;\; V = (v_1 \;\; v_2 \;\; \cdots \;\; v_n)$$

are the *left*, *right singular vectors* of A (these singular vectors are the eigenvectors of AA^*, A^*A). If in (1.1) $\sigma_r > 0$ while $\sigma_{r+1} = 0$, the rank of A is r. Given (1.1), A has a decomposition as a sum of r matrices of rank one, called the *dyadic decomposition*:

$$A = \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \cdots \sigma_r u_r v_r^* \tag{1.2}$$

1.1 Optimal approximation in the 2-induced norm

The singular value decomposition is the tool which leeds to the solution of the *optimal approximation* problem of A by matrices of lower rank, in the **2-induced norm**. Given $A \in \mathbb{R}^{n\times n}$, $\text{rank}\, A = r \leq n$, find $X \in \mathbb{R}^{n\times n}$, $\text{rank}\, X = k < r$, such that the 2-norm of the *error* matrix $E := A - X$ is minimized. The solution of this problem is given next.

Theorem 1.1 Schmidt-Mirsky. *With the notation introduced above*

$$\min_{\text{rank} X \leq k} \| A - X \|_2 = \sigma_{k+1}(A) \tag{1.3}$$

A (non-unique) minimizer X_* is obtained by truncating the dyadic decomposition (1.2) to contain the first k terms:

$$X_* = \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \cdots \sigma_k u_k v_k^* \tag{1.4}$$

1.2 Sub-optimal approximation in the 2-induced norm

In order to obtain a larger set of solutions, one may relax the optimality condition (1.3). Instead one may wish to obtain all matrices X of rank k satisfying

$$\sigma_{k+1}(A) < \| A - X \|_2 < \epsilon \tag{1.5}$$

for some given ϵ between the $(k+1)^{\text{st}}$ and the k^{th} singular values:

$$\sigma_{k+1}(A) < \epsilon < \sigma_k(A)$$

This is the **sub-optimal approximation** problem in the 2-induced norm.

In order to state the solution we need to construct a J-unitary matrix Θ. This is done as follows. There exist matrices Λ_i, $i = 1, 2$ such that:

$$I_n - \epsilon^{-2} AA^* = \Lambda_1 \Lambda_1^* - \Lambda_2 \Lambda_2^*, \;\; \text{rank}\,(\Lambda_1) + \text{rank}\,(\Lambda_2) = \text{rank}\,(A) \tag{1.6}$$

This relationship implies the existence of a J-unitary matrix Θ of size $2n$ such that

$$[I_n \quad \epsilon^{-1}T]\Theta = [\Lambda_2 \quad \Lambda_1], \text{ where} \tag{1.7}$$

$$\Theta J \Theta^* = J, \; J = \begin{pmatrix} I_n & 0 \\ 0 & -I_n \end{pmatrix} \tag{1.8}$$

$$\Theta = \begin{pmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{pmatrix} \in \mathbb{R}^{2n\times 2n} \tag{1.9}$$

Finally, let $E_i \in \mathbb{R}^{n\times n}$, $i = 1, 2$, and $\Delta \in \mathbb{R}^{n\times n}$ be related as follows

$$\begin{pmatrix} E_1 \\ E_2 \end{pmatrix} = \Theta \begin{pmatrix} \Delta \\ I_n \end{pmatrix} = \begin{pmatrix} \Theta_{11}\Delta + \Theta_{12} \\ \Theta_{21}\Delta + \Theta_{22} \end{pmatrix} \tag{1.10}$$

Theorem 1.2 [11]. *$\hat{X}$ satisfies (1.5) if, and only if, there exists a contraction Δ such that*

$$A - \hat{X} = E := E_1 E_2^{-1}, \; \| \Delta \|_2 < 1 \;\; \text{and} \;\; \operatorname{rank}(\Lambda_1 + \Lambda_2\Delta) = k \tag{1.11}$$

It should be remarked that for a particular choice of Λ_1 and Λ_2, the rank condition above can be converted to the block triangularity of the contraction Δ (see [11]).

2 Infinite-dimensional, structured operators

In this section we will consider structured linear maps of $\ell(\mathbb{Z}_+)$ (the space of real-valued sequences) onto itself. In particular we will turn our attention to *Hankel operators*:

$$\mathcal{H} : \; \ell(\mathbb{Z}_+) \longrightarrow \ell(\mathbb{Z}_+) \tag{2.1}$$

where in the canonical basis

$$\mathcal{H} = \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \cdots \\ \alpha_2 & \alpha_3 & \alpha_4 & \cdots \\ \alpha_3 & \alpha_4 & \alpha_5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{2.2}$$

We will assume that $\mathcal{H}$ has finite rank:

$$\operatorname{rank} \mathcal{H} = n < \infty$$

This is equivalent with the rationality of the (formal) power series $\sum_{t>0} \alpha_t z^{-t}$, i.e. the existence of polynomials π and χ such that

$$\sum_{t>0} \alpha_t z^{-t} = \frac{\pi(z)}{\chi(z)} =: G_{\mathcal{H}}(z) \text{ where } \deg \chi = n > \deg \pi \tag{2.3}$$

Furthermore, every strictly proper (degree of numerator less than degree of denominator) rational function G can be written as

$$G(z) = C(zI - A)^{-1}B, \; A \in \mathbb{R}^{n\times n}, \; B, C^T \in \mathbb{R}^n \tag{2.4}$$

In summary, realization theory of linear time-invariant systems establishes the equivalence of sequences whose associated Hankel matrix has finite rank and linear time-invariant, finite dimensional systems expressed in terms of their transfer functions G or in terms of the triple (C, A, B):

$$\mathcal{H}, \ \mathrm{rank}\,\mathcal{H} = n \ \stackrel{1-1}{\longleftrightarrow} \ G(z) = \frac{\pi(z)}{\chi(z)} = C(zI - A)^{-1}B, \ \deg \pi < \deg \chi = n \tag{2.5}$$

For details on realization theory of linear systems we refer e.g., to [3].

We will also assume that the roots of χ are inside the unit disc which is equivalent with the square summability of the sequence α_t

$$\sum_{t>0} \mid \alpha_t \mid^2 < \infty$$

2.1 Optimal approximation of $\mathcal{H}$ in the 2-norm

The problem is to approximate $\mathcal{H}$ by a Hankel matrix $\hat{\mathcal{H}}$ of lower rank in the 2-induced norm. By the equivalence (2.5), this corresponds to approximating the linear system described by the rational function $G = \frac{\pi}{\chi}$, by a system of lower complexity, i.e. $\hat{G} = \frac{\hat{\pi}}{\hat{\chi}}$, where $\deg \chi > \deg \hat{\chi}$. This problem is known in system theory as **linear system approximation in the Hankel norm**. For details see e.g. [8].

Since $\mathcal{H}$ is bounded and compact, it has a discrete set of non-zero singular values with an accumulation point at zero. Let the non-zero singular values of $\mathcal{H}$ be

$$\sigma_1(\mathcal{H}) \geq \sigma_2(\mathcal{H}) \geq \cdots \geq \sigma_n(\mathcal{H}) > 0 \tag{2.6}$$

According to the Schmidt-Mirsky theorem any, not necessarily structured, approximant $\mathcal{K}$ of $\mathcal{H}$ which has rank $k < n$ satisfies

$$\| \mathcal{H} - \mathcal{K} \|_2 \geq \sigma_{k+1}(\mathcal{H}) \tag{2.7}$$

The question arises: *does there exist an approximant of rank k which has Hankel structure and achieves the lower bound*? The answer is given by the result below.

Theorem 2.1 Adamjan, Arov, Krein, [1], [2]. *There exists a unique approximant $\hat{\mathcal{H}}$ of rank k, which has Hankel structure and attains the lower bound, namely*

$$\sigma_1(\mathcal{H} - \hat{\mathcal{H}}) = \sigma_{k+1}(\mathcal{H})$$

This result is proved in the above mentioned references using funcional analytic tools. A purely linear algebraic solution which gave rise to many subsequent developments, is given in [8].

2.2 Sub-optimal approximation of $\mathcal{H}$ in the 2-norm

As in the finite-dimesional operator case the problem of suboptimal solutions arises, namely, given ϵ, find all $\hat{\mathcal{H}}$ having Hankel structure and rank k, such that

$$\sigma_{k+1}(\mathcal{H}) \leq \sigma_1(\mathcal{H} - \hat{\mathcal{H}}) < \epsilon < \sigma_k(\mathcal{H}) \tag{2.8}$$

The solution to this problem (see pages 540-546 of [7], and [8]) is given in terms of the rational (transfer) function G attached to the Hankel matrix $\mathcal{H}$ by means of (2.5).

Recall that theorem 2.1 is valid for Hankel operators $\mathcal{H}$ whose associated rational function G has poles *inside* the unit disc; such rational functions are called *stable*. The solution discussed below however makes use of rational functions which have poles *outside* the unit disc; such rational functions will be called *anti-stable*. A rational function F with no poles on the unit circle, can be decomposed uniquely as a sum of an anti-stable and a stable rational function:

$$F = [F]_- + [F]_+$$

where $[F]_-$ has all its poles outside the unit disc and $[F]_+$ has all its poles inside the unit disc and is strictly proper.

Given $\mathcal{H}$ and the associated rational (transfer) function $G_{\mathcal{H}}(z) = C(zI - A)^{-1}B$, with (C, A, B) minimal, let

$$\sigma_k(\mathcal{H}) > \epsilon > \sigma_{k+1}(\mathcal{H})$$

The following rational matrix $\Theta(z)$ has dimension 2×2.

$$\Theta(z) = I_2 - (z - z_0)\Psi(z) \tag{2.9}$$

where z_0 is a fixed point on the unit circle, and

$$\Psi(z) = \begin{pmatrix} \epsilon^{-2}C(zI-A)^{-1}\Gamma^{-1}\mathcal{P}(I-z_0A^*)^{-1}C^* & \epsilon^{-1}C(zI-A)^{-1}\Gamma^{-1}(A-z_0I)^{-1}B \\ \epsilon^{-1}B^*(I-zA^*)^{-1}\Gamma^{-*}(I-z_0A^*)^{-1}C^* & B^*(I-zA^*)^{-1}\mathcal{Q}\Gamma^{-1}(A-z_0I)^{-1}B \end{pmatrix} \tag{2.10}$$

$\mathcal{P}$, $\mathcal{Q} > 0$ are the reachability, observability Grammians satisfying the Lyapunov equations

$$A\mathcal{P}A^* + BB^* = \mathcal{P},\ A^*\mathcal{Q}A + C^*C = \mathcal{Q} \tag{2.11}$$

and $\Gamma := I - \epsilon^{-2}\mathcal{P}\mathcal{Q}$. By construction Θ is J-unitary on the unit circle:

$$\Theta(e^{j\theta})J\Theta^*(e^{-j\theta}) = J$$

where J is the 2×2 signature matrix $J := \text{diag}\,(1, -1)$. Let E_i, $i = 1, 2$, and Δ be related as follows

$$\begin{pmatrix} E_1(\Delta) \\ E_2(\Delta) \end{pmatrix} := \Theta \begin{pmatrix} \Delta \\ I \end{pmatrix} = \begin{pmatrix} \Theta_{11}(z)\Delta(z) + \Theta_{12}(z) \\ \Theta_{21}(z)\Delta(z) + \Theta_{22}(z) \end{pmatrix} \tag{2.12}$$

We will call the rational function Δ a *contraction*, if its magnitude on the unit circle is less than unity: $\mid \Delta(e^{-j\theta}) \mid < 1$ for all $\theta \in [0, 2\pi]$. The following result, reminiscent of theorem 1.2, holds true.

Theorem 2.2 *$\hat{\mathcal{H}}$ satisfies (2.8) if, and only if, there exists an anti-stable contraction Δ, such that*

$$G_{\hat{\mathcal{H}}} = [F]_+ \text{ where } G_{\mathcal{H}} - F = E(\Delta) := E_1(\Delta)E_2^{-1}(\Delta) \tag{2.13}$$

By setting $\Delta = 0$ in the above formula, we obtain one sub-optimal approximant. It is namely, the Hankel matrix associated with the stable part of the rational function $G_{\mathcal{H}} - \Theta_{12}\Theta_{22}^{-1}$. It turns out that this Hankel matrix has rank k.

It should be mentioned that the optimal and sub-optimal approximation results presented in the last two sub-sections, hold equally for *block Hankel* matrices (the α_t are $p \times m$ blocks). We refer to [8] and [7] for details.

3 Problems

We have presented above two cases of optimal and sub-optimal approximation in the 2-induced norm, of linear operators of finite rank. The two cases were

- Unstructured operators in a finite-dimensional space
- Structured (Hankel) operators in an infinite-dimensional space

In the parametrization of all sub-optimal solutions as given in theorems 1.2, 2.2, **two** important ingredients manifest themselves **in both cases**:

- Linear fractions in terms of a J-unitary matrix Θ, see (1.10) and (2.12)
- All solutions are parametrized in terms of a contraction Δ

Both Θ and Δ are defined differently for the two cases (1.10) and (2.12). Thus the following problem arises:

Problem 1. Investigate the existence of a unifying framework for the approximation of operators in the induced 2-norm.

It is known that theorem 2.1 does not hold for *finite Hankel matrices*. This leads to

Problem 2. Investigate the problem of low rank approximation of structured finite matrices, and in particular Hankel matrices, in the induced 2-norm.

Remark. There are two abstract frameworks with relevance to the problems formulated above. The first one is the *Grassmannian/projective geometry* approach of Ball and Helton (see [4]). The second one is the *abstract band method* of Dym, Gohberg, Kaashoek, and Woerdeman (see the survey and references in [9]). The challenge problems call for dealing with a more general situation than what is discussed in [4] and [9], namely, the Takagi version of the generalized Nehari problem (where the approximating function is allowed a prescribed number of poles in the unit disk). While the Ball-Helton method has been worked out in particular cases for the Takagi version (see [5] and [6]) the abstract Hilbert module version as presented in [4] has not been so worked out. Moreover, the general band method presented in [9] has not been worked out at all for the Takagi version of the contractive completion problem. An elaboration of either the Grassmannian/projective geometry approach or of the abstract band method may thus yield a framework which handles the above challenge problems. ■

The problem of deriving an optimal or sub-optimal reduced order model of a system from input-output measurements is of importance for system identification and robust control. In [8] this problem is studied for the special case where the input to the system is an impulse and the output the impulse response; for the more general case of arbitrary input-output measurements the problem is studied in [12]. This leads us to the third problem.

Problem 3. Investigate the problem of obtaining reduced order models of a system from arbitrary measurements - finite or infinite - of the input and the output, which are optimal and/or suboptimal in the induced 2-norm.

References

[1] V.M. Adamjan, D.Z. Arov, and M.G. Krein, *Analytic properties of Schmidt pairs for a Hankel operator and the generalized Schur-Takagi problem*, Math. USSR Sbornik, **15**: 31-73 (1971).

[2] V.M. Adamjan, D.Z. Arov, and M.G. Krein, *Infinite block Hankel matrices and related extension problems*, American Math. Society Transactions, **111**: 133-156 (1978).

[3] A.C. Antoulas, T. Matsuo, and Y. Yamamoto, *Linear deterministic realization theory*, in *Mathematical System Theory*, A.C. Antoulas Ed., pp. 191-212, Springer Verlag (1991).

[4] J.A. Ball, *Nevanlinna-Pick interpolation: generalizations and applications*, in Surveys of Some Recent Results in Operator Theory, Volume I, J.B. Conway and B.B. Morrel Editors, Pitman Research Notes in Mathematics, **171**: 51-94.

[5] J.A. Ball and J.W. Helton, *Interpolation problems of Pick-Nevanlinna and Loewner types for meromorphic matrix functions: parametrization of the set of all solutions*, Integral Equations and Operator Theory, **9**: 155-203 (1986).

[6] J.A. Ball and I. Gohberg, *Shift invariant subspaces, factorization and interpolation for matrices I: The canonical case*, Linear Algebra and Its Applications, **74**: 87-150 (1986).

[7] J.A. Ball, I. Gohberg, and L. Rodman, *Interpolation of rational matrix functions*, Operator Theory: Advances and Applications, Birkhäuser (1990).

[8] K. Glover, *All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds*, International Journal of Control, **39**: 1115-1193 (1984).

[9] I. Gohberg, S. Goldberg, and M.A. Kaashoek, *Classes of Linear Operators*, Vol. II, **OT 63** Birkhäuser.

[10] G.W. Stewart and J. Sun, *Matrix perturbation theory*, Academic Press (1990).

[11] A.J. van der Veen, *A Schur method for low rank matrix approximation*, SIAM J. Matrix Analysis and Applications, January (1996).

[12] J.C. Willems, *From time series to linear system - Part III: Approximate modeling*, Automatica **23**: 87-115 (1987).