Homotopy Methods for Solving the Optimal Projection Equations for the H_2 Reduced Order Model Problem

By Dragan Zigic, and Layne T. Watson, Emmanuel G. Collins and Dennis S. Bernstein

TR 91-13

HOMOTOPY METHODS FOR SOLVING THE OPTIMAL PROJECTION EQUATIONS FOR THE H_2 REDUCED ORDER MODEL PROBLEM

Dragan Žigić and Layne T. Watson
Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061

Emmanuel G. Collins, Jr. and Dennis S. Bernstein
Harris Corporation, Government Aerospace Systems Division
Melbourne, FL 32902

Abstract.

The optimal projection approach to solving the H_2 reduced order model problem produces two coupled, highly nonlinear matrix equations with rank conditions as constraints. Due to the resemblance of these equations to standard matrix Lyapunov equations, they are called modified Lyapunov equations. The algorithms proposed herein utilize probability-one homotopy theory as the main tool. It is shown that there is a family of systems (the homotopy) that make a continuous transformation from some initial system to the final system. With a carefully chosen initial problem a theorem guarantees that all the systems along the homotopy path will be asymptotically stable, controllable and observable. One method, which solves the equations in their original form, requires a decomposition of the projection matrix using the Drazin inverse of a matrix. It is shown that the appropriate inverse is a differentiable function. An effective algorithm for computing the derivative of the projection matrix that involves solving a set of Sylvester equations is given. Another class of methods considers the equations in a modified form, using a decomposition of the pseudogramians based on a contragredient transformation. Some freedom is left in making an exact match between the number of equations and the number of unknowns, thus effectively generating a family of methods.

1. Introduction.

In [12] Hyland and Bernstein considered the quadratic (H_2) reduced order model problem, which is to find a reduced order model for a given continuous time stationary linear system which minimizes a quadratic model error criterion. The necessary conditions for the optimal reduced order model are given in the form of two modified Lyapunov equations, matrix equations which resemble the (linear) matrix Lyapunov equations, but are highly nonlinear and mutually coupled.

Among many different approaches for finding reduced order models are component cost analysis [20]–[22], balancing [17]–[18], Hankel-norm approximation [14]–[15], aggregation [1], [16], nonminimal partial realization [10] and the optimal reduction method of Wilson [30]. Some other applications of the optimal projection approach include the H_2/H_∞ model reduction problem [8], the fixed order dynamic compensation problem [11] and the reduced order state estimation problem [3]. A homotopy based algorithm for solving the fixed order dynamic compensation problem is given by Richter and Collins [19].

The complete statement of the reduced order model problem is given in Section 2. Section 3 explains the basics of probability-one homotopy theory. Section 4 gives a way for constructing an initial problem. Section 5 gives an algorithm for computing a contragredient transformation. The method based on the Drazin inverse is presented in Section 6. Methods based on decompositions of pseudogramians are given in Section 7. Numerical results obtained by solving a number of model reduction problems are given in Section 8. Section 9 gives a conclusion.

2. Statement of the Problem.

Given the controllable and observable, time invariant, continuous time system

$$\dot{x}(t) = A\,x(t) + B\,u(t), \tag{1}$$
$$y(t) = C\,x(t), \tag{2}$$

where $A \in R^{n\times n}$, $B \in R^{n\times m}$, $C \in R^{l\times n}$, the goal is to find, for given $n_m < n$, a reduced order model

$$\dot{x}_m(t) = A_m\,x_m(t) + B_m\,u(t),$$
$$y_m(t) = C_m\,x_m(t),$$

where $A_m \in R^{n_m\times n_m}$, $B_m \in R^{n_m\times m}$, $C_m \in R^{l\times n_m}$, which minimizes the quadratic model-reduction criterion

$$J(A_m, B_m, C_m) \equiv \lim_{t\to\infty} E\left[(y - y_m)^t\,R\,(y - y_m)\right],$$

where the input $u(t)$ is white noise with positive definite intensity V and R is a positive definite weighting matrix.

It is assumed that A is asymptotically stable and diagonalizable, and a solution (A_m, B_m, C_m) is sought in the set

$$A_+ = \{(A_m, B_m, C_m) :\ A_m \text{ is stable, } (A_m, B_m) \text{ is controllable and } (A_m, C_m) \text{ is observable}\}.$$

DEFINITION 1. Given symmetric positive semidefinite matrices $\hat{Q}$, $\hat{P} \in R^{n\times n}$ such that rank $(\hat{Q})$ = rank $(\hat{P})$ = rank $(\hat{Q}\hat{P})$ = n_m, matrices $G, \Gamma \in R^{n_m\times n}$ and positive semisimple $M \in R^{n_m\times n_m}$ are called a (G, M, Γ)-factorization (projective factorization) of $\hat{Q}\hat{P}$ if

$$\hat{Q}\,\hat{P} = G^t\,M\,\Gamma,$$
$$\Gamma\,G^t = I_{n_m}.$$

Positive semisimple means similar to a symmetric positive definite matrix.

The following theorem from [12] gives necessary conditions for the optimal solution to the reduced order model problem.

THEOREM 2. *Suppose $(A_m, B_m, C_m) \in A_+$ solves the optimal model-reduction problem. Then there exist symmetric positive semidefinite matrices $\hat{Q}$, $\hat{P} \in R^{n\times n}$ such that for some projective factorization of $\hat{Q}\hat{P}$, A_m, B_m and C_m are given by*

$$A_m = \Gamma\,A\,G^t, \tag{3}$$
$$B_m = \Gamma\,B, \tag{4}$$
$$C_m = C\,G^t, \tag{5}$$

and such that, with $\tau \equiv G^t\Gamma$ the following conditions are satisfied:

$$0 = \tau[A\,\hat{Q} + \hat{Q}\,A^t + B\,V\,B^t], \tag{6}$$
$$0 = [A^t\,\hat{P} + \hat{P}\,A + C^t\,R\,C]\,\tau, \tag{7}$$
$$\text{rank } (\hat{Q}) = \text{rank } (\hat{P}) = \text{rank } (\hat{Q}\,\hat{P}) = n_m. \tag{8}$$

The equations (6)–(7) can be written in an equivalent form

$$A\hat{Q} + \hat{Q}A^t + \tau B V B^t + B V B^t \tau^t - \tau B V B^t \tau^t = 0,$$
$$A^t \hat{P} + \hat{P} A + \tau^t C^t R C + C^t R C \tau - \tau^t C^t R C \tau = 0.$$

The matrices $\hat{Q}$ and $\hat{P}$ are called the *controllability* and *observability pseudogramians*, respectively, since they are analogous to the Gramians G_c and G_o which satisfy the dual Lyapunov equations

$$A G_c + G_c A^t + B V B^t = 0,$$
$$A^t G_o + G_o A + C^t R C = 0.$$

τ is an oblique projection (idempotent) operator since $\tau^2 = \tau$. The projection matrix τ can be expressed as

$$\tau = (\hat{Q}\,\hat{P})\,(\hat{Q}\,\hat{P})^\sharp,$$

where $(\hat{Q}\,\hat{P})^\sharp$ is the Drazin inverse defined in Section 6.

3. Probability-One Homotopy Methods.

Homotopies are a traditional part of topology, and have found significant application in nonlinear functional analysis and differential geometry [28]. Homotopy methods are globally convergent, which distinguishes them from most iterative methods, which are only locally convergent. The general idea of homotopy methods is to make a continuous transformation from an initial problem, which can be solved trivially, to the target problem.

Following [26], the theoretical foundation of all probability-one globally convergent homotopy methods is given in the following differential geometry theorem:

DEFINITION 3. Let $U \subset R^m$ and $V \subset R^p$ be open sets, and let $\rho : U \times [0,1) \times V \to R^p$ be a C^2 map. ρ is said to be transversal to zero if the Jacobian matrix $D\rho$ has full rank on $\rho^{-1}(0)$.

THEOREM 4. *If $\rho(a,\lambda,x)$ is transversal to zero, then for almost all $a \in U$ the map*

$$\rho_a(\lambda, x) = \rho(a, \lambda, x)$$

is also transversal to zero; i.e, with probability one the Jacobian matrix $D\rho_a(\lambda,x)$ has full rank on $\rho_a^{-1}(0)$.

The recipe for constructing a globally convergent homotopy algorithm to solve the nonlinear system of equations

$$f(x) = 0,$$

where $f : R^p \to R^p$ is a C^2 map, is as follows: For an open set $U \subset R^m$ construct a C^2 homotopy map $\rho : U \times [0,1) \times R^p \to R^p$ such that

1) $\rho(a,\lambda,x)$ is transversal to zero,
2) $\rho_a(0,x) = \rho(a,0,x) = 0$ is trivial to solve and has a unique solution x_0,
3) $\rho_a(1,x) = f(x)$,
4) $\rho_a^{-1}(0)$ is bounded.

Then for almost all $a \in U$ there exists a zero curve γ of ρ_a, along which the Jacobian matrix $D\rho_a$ has rank n, emanating from $(0,x_0)$ and reaching a zero $\bar{x}$ of f at $\lambda = 1$. This zero curve γ does not intersect itself, is disjoint from any other zeros of ρ_a, and has finite arc length in every

compact subset of $[0,1) \times R^p$. Furthermore, if $Df(\bar{x})$ is nonsingular, then γ has finite arc length. The general idea of the algorithm is to follow the zero curve γ emanating from $(0, x_0)$ until a zero $\bar{x}$ of $f(x)$ is reached (at $\lambda = 1$).

The zero curve γ is tracked by the normal flow algorithm [29], a predictor-corrector scheme. In the predictor phase, the next point is produced using Hermite cubic interpolation. Starting at the predicted point, the corrector iteration involves computing (implicitly) the Moore-Penrose pseudo-inverse of the Jacobian matrix at each point. The most complex part of the homotopy algorithm is the computation of the tangent vectors to γ, which involves the computation of the kernel of the $p \times (p+1)$ Jacobian matrix $D\rho_a$. The kernel is found by computing a QR factorization of $D\rho_a$, and then using back substitution. This strategy is implemented in the mathematical software package HOMPACK [29], which was used for the curve tracking here.

Two different homotopy maps are used for solving the optimal projection equations. When the initial problem, $g(x;a) = 0$, can be solved, then the homotopy map is [27]

$$\rho_a(\lambda, x) = F(a, \lambda, x) \equiv \lambda\, f(x) + (1-\lambda) g(x;a), \tag{9}$$

where $f(x) = 0$ is the final problem, and a is a parameter vector used in defining the function g.

When the initial problem is not solved exactly, i.e., $g(x_0;b) \neq 0$, then the map is a Newton homotopy [23]

$$\rho_a(\lambda, x) = F(b, \lambda, x) - (1-\lambda) F(b, 0, x_0), \tag{10}$$

where $a = (b, x_0)$. For $\lambda = 0$, $\rho_a(0, x_0) = F(b, 0, x_0) - F(b, 0, x_0) = 0$, and for $\lambda = 1$, $\rho_a(1, x) = F(b, 1, x) = f(x) = 0$.

For the homotopies considered here, the theoretical verification of properties 1) and 4) is highly technical and has not been done.

4. Defining an Initial System.

While with homotopy algorithms in general an initial problem can be chosen practically at random, this problem has some special limitations. The reason is that Theorem 2 provides necessary conditions on a solution only under certain assumptions. In other words, every intermediate problem solution satisfies these equations only if the system is asymptotically stable, controllable and observable. While the absence of these features does not automatically mean that the intermediate problem solution will not satisfy the equations, it is clearly better to define a homotopy path in such a way that each problem along it corresponds to an asymptotically stable, controllable and observable system. Existence of a solution to the H_2 reduced order problem follows from [24]. Theorem 5 defines a class of initial systems such that these conditions are satisfied.

THEOREM 5. *For the given system* (1)–(2), *let* $A = X \Lambda X^{-1}$, *with* Λ *diagonal. Define* $D = X \Omega X^{-1}$ *for any diagonal matrix* $\Omega = diag\,(\omega_1, \ldots, \omega_n)$, *such that all* ω_i, *for* $i = 1, 2, \ldots, n$, *are in the open left half plane. Then for almost all such* D *every convex combination* $(A(\alpha), B, C)$ *of the systems* (D, B, C) *and* (A, B, C) *will be asymptotically stable, controllable and observable.*

Proof. (Stability.) Since

$$\begin{aligned} A(\alpha) &= \alpha\, A + (1-\alpha)\, D = \alpha\, X \Lambda X^{-1} + (1-\alpha)\, X \Omega X^{-1} \\ &= X\, [\alpha \Lambda + (1-\alpha)\, \Omega]\, X^{-1} = X\, \Lambda(\alpha)\, X^{-1}, \end{aligned}$$

and $\Lambda(\alpha)$ is diagonal with all diagonal elements in the open left half plane for $\alpha \in [0,1]$, the matrix $A(\alpha)$ is asymptotically stable for $\alpha \in [0,1]$.

(Controllability.) Let $B = X\,\hat{B}$. Consider the controllability matrices in the coordinate system obtained by the change of coordinates defined by the matrix X. In that coordinate system the controllability matrix is

$$\hat{B}_c(\alpha) = (\,\hat{B} \quad \Lambda(\alpha)\,\hat{B} \quad \ldots \quad \Lambda^{n-1}(\alpha)\,\hat{B}\,),$$

for $\alpha \in [0,1]$. For almost all choices of Ω the eigenvalues of the diagonal matrix $\Lambda(\alpha)$ will be distinct for $\alpha \in [0,1)$. That rank $\hat{B}_c(\alpha) = n$ follows from the results in [4]. A direct argument follows. Let $\lambda_1, \ldots, \lambda_n$ be the (distinct) eigenvalues of $\Lambda(\alpha)$ for some α, and

$$\hat{B} = \begin{pmatrix} \hat{b}_{11} & \hat{b}_{12} & \ldots & \hat{b}_{1m} \\ \hat{b}_{21} & \hat{b}_{22} & \ldots & \hat{b}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{b}_{n1} & \hat{b}_{n2} & \ldots & \hat{b}_{nm} \end{pmatrix}.$$

Since the system is controllable for $\alpha = 1$, it follows that each row of $\hat{B}$ has at least one nonzero element, because otherwise $\hat{B}_c(1)$ would have a zero row.

Reorder the rows and columns of $\hat{B}_c$ in the following way: for each row of $\hat{B}$ that has a nonzero element in the first column (suppose there are p_1 of them), exchange rows so that the selected p_1 rows are at the top. Next, exchange columns in such a way that columns $1, m+1, \ldots, (p_1-1)\,m+1$ become the first p_1 columns. This produces a submatrix in the upper left corner which can be expressed as

$$\hat{B}_1 = \begin{pmatrix} \hat{b}_1 & 0 & \ldots & 0 \\ 0 & \hat{b}_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \hat{b}_{p_1} \end{pmatrix} \begin{pmatrix} 1 & \lambda_1 & \ldots & \lambda_1^{p_1-1} \\ 1 & \lambda_2 & \ldots & \lambda_2^{p_1-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_{p_1} & \ldots & \lambda_{p_1}^{p_1-1} \end{pmatrix}.$$

Since $\hat{b}_1, \ldots, \hat{b}_{p_1}$ were chosen to be nonzero, and the λ_i are distinct, the Vandermonde matrix is nonsingular, and hence $\hat{B}_1$ is also nonsingular.

Repeating this procedure for the remaining rows gives a block upper triangular transformation of $\hat{B}_c(\alpha)$, with diagonal blocks $\hat{B}_1, \ldots, \hat{B}_r$, for some $r \leq n$. That means that rank $\hat{B}_c(\alpha) = n$ and the system is controllable.

(Observability.) The analogous construction for the observability matrices proves that the system is observable. Q. E. D.

While the random construction of the matrix D given in Theorem 5 is theoretically plausible, in practice it may not be wise. The reason is that the matrix X is complex in general, which for many choices of Ω leads to a complex matrix D, which is undesirable. Hence, it is better to directly construct a matrix D such that Ω satisfies the conditions given in Theorem 5.

One simple choice for D is

$$D \equiv -c_1\,I + \text{diag}\,\{\epsilon_1, \ldots, \epsilon_n\}, \tag{11}$$

where $c_1 > 0$ and ϵ_i are small random numbers that correspond to the parameter a in the theory. In this case Ω is a small perturbation of $-c_1\,I$.

Also, the matrix D can be defined as

$$D \equiv -c_1 I + c_2 A, \tag{12}$$

for $c_1, c_2 > 0$. In this case $\Omega = -c_1 I + c_2 \Lambda$.

The following strategy can be applied to find a good approximation to a solution of the initial system. Since the matrix D is asymptotically stable, the Lyapunov equation

$$D Q + Q D^t + B V B^t = 0 \tag{13}$$

has a unique solution Q. Let $Q = T \Sigma T^t$, where T is orthogonal and

$$\Sigma = \text{diag } \{\sigma_1, \ldots, \sigma_n\}.$$

Next, define

$$\Sigma_1 \equiv \text{diag } \{\sigma_1, \ldots, \sigma_{n_m}, 0, \ldots, 0\}, \qquad Q_0 \equiv T \Sigma_1 T^t.$$

If Q_0 is substituted for Q in (13), the equation will not be satisfied, but in general, if σ_i are sufficiently small, it will not be very different from zero. A similar procedure can be applied to compute P_0 that will 'almost' satisfy the equation

$$D^t P + P D + C^t R C = 0.$$

The point $x_0 = (Q_0, P_0)$ chosen in this way may lead to small values of $g(x_0)$. Also, this x_0 can be used as the initial guess for a quasi-Newton algorithm which may find a solution to the initial problem

$$\tau[D Q + Q D^t + B V B^t] = 0,$$
$$[D^t P + P D + C^t R C]\tau = 0.$$

5. Contragredient Transformation.

The following lemma from [12], which is a special case of a result in [7], gives an algorithm for simultaneous reduction of pseudogramians to diagonal forms using a contragredient transformation. The constructive proof given here is different from that in [12].

LEMMA 6. [12] *Let symmetric positive semidefinite Q, $P \in R^{n \times n}$ satisfy*

$$\text{rank } (Q) = \text{rank } (P) = \text{rank } (Q P) = n_m, \tag{14}$$

where $n_m \leq n$. Then, there exists a nonsingular $W \in R^{n \times n}$ (contragredient transformation) and positive definite diagonal Σ, $\Omega \in R^{n_m \times n_m}$ such that

$$Q = W \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} W^t, \quad P = W^{-t} \begin{pmatrix} \Omega & 0 \\ 0 & 0 \end{pmatrix} W^{-1}.$$

Proof. Since P is positive semidefinite and symmetric there exists orthogonal $V \in R^{n \times n}$ such that

$$P = V \begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix} V^t,$$

where $D_1 \in R^{n_m \times n_m}$ is diagonal and positive definite. Let

$$T_1 \equiv V \begin{pmatrix} D_1^{-1/2} & 0 \\ 0 & I \end{pmatrix}.$$

Then

$$T_1^t P T_1 = \begin{pmatrix} I_{n_m} & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad T_1^{-1} Q T_1^{-t} = \begin{pmatrix} Q_1 & Q_{12} \\ Q_{12}^t & Q_2 \end{pmatrix},$$

where $Q_1 \in R^{n_m \times n_m}$. Since rank $(Q\,P)$ = rank (P) implies that Q is one-to-one on the eigenspace of P corresponding to positive eigenvalues, and the quadratic form $x^t Q_1 x$ corresponds to the quadratic form $y^t Q\, y$ restricted to this eigenspace, and $y^t Q\, y = 0$ implies $Q\, y = 0$ for symmetric positive semidefinite Q, it follows that Q_1 is also symmetric positive definite. Therefore, there exists a positive definite diagonal $D_2 \in R^{n_m \times n_m}$ and orthogonal $U \in R^{n_m \times n_m}$ such that $Q_1 = U\, D_2\, U^t$. Let

$$T_2 \equiv \begin{pmatrix} U & 0 \\ Q_{12}^t\, U\, D_2^{-1} & -I \end{pmatrix}.$$

Then

$$\begin{aligned} T_2^t\, T_1^t\, P\, T_1\, T_2 &= \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}, \\ T_2^{-1}\, T_1^{-1}\, Q\, T_1^{-t}\, T_2^{-t} &= \begin{pmatrix} D_2 & 0 \\ 0 & Q_2 - Q_{12}^t\, U\, D_2^{-1}\, U^t\, Q_{12} \end{pmatrix} \\ &= \begin{pmatrix} D_2 & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

The last equality is a consequence of the rank conditions (14). If $W \equiv T_1\, T_2$ then

$$Q = W \begin{pmatrix} D_2 & 0 \\ 0 & 0 \end{pmatrix} W^t, \quad P = W^{-t} \begin{pmatrix} I_{n_m} & 0 \\ 0 & 0 \end{pmatrix} W^{-1}, \tag{15}$$

which completes the proof. Q. E. D.

REMARK 7. [12] *Let Q and P be as in Lemma 6. Then there exists a nonsingular $U \in R^{n \times n}$ and positive definite diagonal $\Lambda \in R^{n_m \times n_m}$ such that*

$$Q = U \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} U^t, \quad P = U^{-t} \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} U^{-1}.$$

Proof. The statement follows from (15) using $U \equiv W\, D_2^{1/4}$ Q. E. D.

The following lemma defines a projective factorization of the product of the pseudogramians and gives an effective way to compute it using a contragredient transformation. The proof here is slightly different from that of [12].

LEMMA 8. [12] *Let symmetric positive semidefinite $\hat{Q}$, $\hat{P} \in R^{n \times n}$ satisfy the rank conditions* (14). *Then, there exist G, $\Gamma \in R^{n_m \times n}$ and positive semisimple (*positive semisimple *means similar to a symmetric positive definite matrix) $M \in R^{n_m \times n_m}$ such that*

$$\hat{Q}\, \hat{P} = G^t\, M\, \Gamma, \tag{16}$$

$$\Gamma\, G^t = I_{n_m}. \tag{17}$$

Proof. Due to Remark 7 there exist nonsingular $W \in R^{n\times n}$ and positive definite diagonal $\Sigma \in R^{n_m \times n_m}$ such that

$$\hat{Q} = W \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} W^t, \quad \hat{P} = W^{-t} \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} W^{-1}. \tag{18}$$

The equations (18) can be expressed in the equivalent form

$$\hat{Q} = W_1 \Sigma W_1^t, \quad \hat{P} = U_1^t \Sigma U_1, \tag{19}$$

where

$$W = \overbrace{\begin{pmatrix} W_1 & W_2 \end{pmatrix}}^{n_m}, \quad W^{-1} = U = \begin{matrix} n_m \{ \\ \\ \end{matrix} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}. \tag{20}$$

From (20) with $G \equiv W_1^t$ and $\Gamma \equiv U_1$ follow (16)–(17). Q. E. D.

Matrices G, M and Γ from Lemma 8 are a (G, M, Γ)-factorization of $(\hat{Q}, \hat{P})$.

6. Homotopy Method Based on the Drazin Inverse.

One approach in designing a homotopy algorithm for solving the optimal projection equations is to use the decomposition of the projection matrix τ based on the Drazin inverse.

6.1. Theoretical Results.

Since the homotopy algorithm involves computation of derivatives at each step, it is essential that equations (6)–(7) be differentiable. The problem of proving the differentiability of the equations (6)–(7) with the rank conditions (8) reduces to the problem of proving the differentiability of the Drazin inverse.

DEFINITION 9. The index of $A \in R^{n\times n}$ is the smallest nonnegative integer k such that $\text{im}\,(A^k) = \text{im}\,(A^{k+1})$.

If $A \in R^{n\times n}$ has index k, then $R^n = \text{im}\,(A^k) \oplus \ker\,(A^k)$. Thus any $x \in R^n$ has the unique decomposition $x = u + v$, $u \in \text{im}\,(A^k)$, $v \in \ker\,(A^k)$. Observe further that A is invertible on $\text{im}\,(A^k)$.

DEFINITION 10. Let $A \in R^{n\times n}$ have index k, and $x = u + v$, $u \in \text{im}\,(A^k)$, $v \in \ker\,(A^k)$. The Drazin inverse $A^\sharp$ of A is defined by $A^\sharp x = A_1^{-1} u$, where A_1 is A restricted to the image of A^k.

If $k = 1$, the Drazin inverse $A^\sharp$ is called the group inverse, and in fact that is all that is needed here.

THEOREM 11. [5] *Let $A \in R^{n\times n}$ have index k. The Drazin inverse $A^\sharp$ of A is the unique matrix $A^\sharp$ such that:*

$$A^\sharp A\, A^\sharp = A^\sharp,$$

$$A^\sharp A = A\, A^\sharp,$$

$$A^{k+1}\, A^\sharp = A^k.$$

THEOREM 12. [5] *If $A \in R^{n\times n}$ has index k, then there exist nonsingular matrices T and C, and a nilpotent matrix N of index k such that*

$$A = T\begin{pmatrix} C & 0 \\ 0 & N \end{pmatrix} T^{-1} \qquad \text{and} \qquad A^{\sharp} = T\begin{pmatrix} C^{-1} & 0 \\ 0 & 0 \end{pmatrix} T^{-1}.$$

$A^{\sharp}$ is computed using the Hermite echelon form as described in Campbell and Meyer [5].

Hearon and Evans [9] give conditions for the differentiability of the Drazin inverse. That theorem is proven here in a simpler way and using different terminology.

THEOREM 13. *Let $A(t) \in C^s(I)$ for some interval I and*

$$\text{rank } (A(t)) = \text{rank } (A^2(t)) = r$$

for each $t \in I$. If, for each $t \in I$, $B(t)$ is the Drazin inverse of $A(t)$, then $B(t) \in C^s(I)$.

Proof. The rank condition implies A has either index 0 or 1. For index 0, $A^{\sharp} = A^{-1}$ is differentiable and there is nothing to prove. So assume A has index 1. Then for each $t \in I$, it has the Jordan decomposition

$$A = T\begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix} T^{-1},$$

where A_1 is square and nonsingular. Using the characteristic polynomial of A_1, a divisor of the characteristic polynomial of A, A_1^{-1} can be expressed as a polynomial in A_1 whose coefficients are polynomials in the elements of A:

$$A_1^{-1} = p(A_1).$$

Let

$$B_1 \equiv p(A) = T\,p\left(\begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix}\right) T^{-1} = T\begin{pmatrix} A_1^{-1} & 0 \\ 0 & p(0) \end{pmatrix} T^{-1}.$$

Then

$$B \equiv B_1\,A\,B_1 = T\begin{pmatrix} A_1^{-1} & 0 \\ 0 & 0 \end{pmatrix} T^{-1} = A^{\sharp}$$

is the Drazin inverse of A. Furthermore, since the elements of B are polynomial functions of the elements of A, $B \in C^k(I)$. Q. E. D.

The derivative τ' of τ is actually computed using the Sylvester equations. Since $\tau = (\hat{Q}\,\hat{P})\,(\hat{Q}\,\hat{P})^{\sharp}$, it follows that

$$\tau' = (\hat{Q}\,\hat{P})'(\hat{Q}\,\hat{P})^{\sharp} + (\hat{Q}\,\hat{P})\left[(\hat{Q}\,\hat{P})^{\sharp}\right]',$$

where everything is directly computable except $\left[(\hat{Q}\,\hat{P})^{\sharp}\right]'$. From (19),

$$\begin{aligned} \text{rank } (\hat{Q}\,\hat{P})^2 &= \text{rank } (W_1\,\Sigma\,W_1^t\,U_1^t\Sigma\,U_1)^2 \\ &= \text{rank } (W_1\,\Sigma^4\,U_1) = \text{rank } (\Sigma^4) = \text{rank } (\Sigma^2) \\ &= \text{rank } (W_1\,\Sigma\,W_1^t\,U_1^t\,\Sigma\,U_1) = \text{rank } (\hat{Q}\,\hat{P}) = n_m < n, \end{aligned}$$

and therefore $\hat{Q}\hat{P}$ has index 1. The following describes a procedure for the computation of the derivative of the Drazin inverse X of a matrix A of index 1.

For given A of index 1, X is the unique matrix that satisfies:

$$A\,X\,A = A, \tag{21}$$
$$X\,A\,X = X, \tag{22}$$
$$A\,X = X\,A. \tag{23}$$

Differentiating (21)–(23) yields:

$$A'\,X\,A + A\,X'\,A + A\,X\,A' = A', \tag{24}$$
$$X'\,A\,X + X\,A'\,X + X\,A\,X' = X', \tag{25}$$
$$A\,X' - X'\,A = X\,A' - A'\,X. \tag{26}$$

Substituting $A\,X'$ from (26) into (24), and summing up equations (24)–(26) gives the Sylvester equation

$$\begin{aligned}(A + X\,A - I)X' + X'(A\,X - A + A^2)\\ = X\,A' - A'\,X - X\,A'\,X + A' - X\,A'\,A - A\,X\,A',\end{aligned} \tag{27}$$

which has a unique solution with probability one due to the randomness in $A(\lambda)$.

Solving (27) for $X' = \left[(\hat{Q}\,\hat{P})^{\sharp}\right]'$ completes the computation of τ'.

6.2. Description of Algorithm.

The following is a description of the algorithm. The algorithm is based on the normal flow algorithm for dense Jacobian matrices described in [29], slightly modified here to handle the rank requirements of the solution $(\hat{Q}, \hat{P})$.

The algorithm starts at the point

$$(\lambda, x) = (0, x_0) = (0, \hat{Q}_0, \hat{P}_0)$$

with some $x_0 = (\hat{Q}_0, \hat{P}_0)$ chosen as explained in Section 4. Then it follows the zero curve γ of the homotopy map (9) until a point where $\lambda = 1$ is reached.

$F(a, \lambda, x)$ is represented by two equations:

$$A(\lambda)\,\hat{Q} + \hat{Q}\,A^t(\lambda) + \tau\,B\,V\,B^t + B\,V\,B^t\,\tau^t - \tau\,B\,V\,B^t\,\tau^t = 0,$$
$$A^t(\lambda)\,\hat{P} + \hat{P}\,A(\lambda) + \tau^t\,C^t\,R\,C + C^t\,R\,C\,\tau - \tau^t\,C^t\,R\,C\,\tau = 0,$$

where

$$A(\lambda) = \lambda\,A + (1 - \lambda)\,D.$$

Recall from (11) that D was a function of the parameter vector a.

Since the equations are symmetric, only the upper right triangles are considered, i.e., q_{ij} and p_{ij} are computed only for $j \geq i$. Therefore, the number of variables is $2\,[n(n+1)/2] = n(n+1)$.

The mathematical software package HOMPACK requires that the user provide routines to evaluate $\rho_a(\lambda, x)$ and the Jacobian matrix $D\rho_a$ at each step. While the former is relatively simple, the latter involves considerable computational effort.

The Jacobian matrix consists of $n(n+1)+1$ derivative vectors, which correspond to the partial derivatives with respect to q_{ij}, p_{ij}, and λ. The terms that do not include τ are simple to evaluate analytically. For example,

$$\left[\frac{\partial}{\partial q_{ij}}(\hat{Q}\,\hat{P})\right]_{kl} = \frac{\partial}{\partial q_{ij}}\sum_{m=1}^{n} q_{km}p_{ml} = \delta_{ik}\,p_{jl}. \tag{28}$$

On the other hand, the components of τ' have to be evaluated numerically. Each evaluation involves solving a Sylvester equation (27). Fortunately, since for different q_{ij} and p_{ij} only the right hand side of the equation changes, the whole process can be done efficiently. In order to compute $\partial/\partial q_{11}$ all the computations related to (27) are done completely. For all subsequent partial derivative evaluations only the right hand side of (27) is evaluated and submitted to the procedure that solves a Sylvester equation. The procedure used for that purpose [2] supports this approach very efficiently.

The derivatives of $(1-\lambda)F(a,0,x_0)$ with respect to q_{ij} and p_{ij} are zero, and the derivative with respect to λ is $-F(a,0,x_0)$.

When a final solution to the equations (6)–(7) is obtained, the computation of (A_m, B_m, C_m) is completed by applying the formulas (3)–(5), where G and Γ are obtained as explained in Lemma 8.

In summary, the whole algorithm is:

1) Define $D = -c\,I + \text{diag}\,\{\epsilon_1,\ldots,\epsilon_n\}$, with $c > 0$ and small random ϵ_i.
2) Choose Q_0 and P_0 that satisfy (8) as small perturbations of, respectively, $B\,V\,B^t/2\,c$ and $C^t\,R\,C/2\,c$.
3) Set $\lambda := 0$, $x := x_0$.
4) Compute Drazin inverse $(\hat{Q}\,\hat{P})^\sharp$.
5) Compute $\tau = (\hat{Q}\,\hat{P})\,(\hat{Q}\,\hat{P})^\sharp$.
6) Evaluate $\rho_a(\lambda, x)$.
7) (Evaluate $D\rho_a(\lambda,x)$.) For each p_{ij}, q_{ij} such that $j \geq i$, and λ, do Steps 8–11.
   8) Compute derivatives of terms that do not include τ using analytical formulas similar to (28).
   9) Compute $\left[(\hat{Q}\,\hat{P})^\sharp\right]'$ using equation (27).
   10) Complete computation of τ' as $\tau' := (\hat{Q}\,\hat{P})'(\hat{Q}\,\hat{P})^\sharp + (\hat{Q}\,\hat{P})\left[(\hat{Q}\,\hat{P})^\sharp\right]'$.
   11) Sum values obtained in steps 8 and 10 to the final value of the derivative vector.
12) Take a step along the curve and obtain $x_1 = (\hat{Q}_1, \hat{P}_1)$.
13) Compute a contragredient transformation (18) as if x_1 satisfied the rank conditions.
14) Use formulas (19) to compute $\bar{x}_1 = (\hat{Q}, \hat{P})$.
15) If $\lambda < 1$, then set $x := \bar{x}_1$, and go to Step 4.
16) If $\lambda \geq 1$, compute the solution $\bar{x}_1$ of (6)–(8) at $\lambda = 1$. Obtain G and Γ as explained in Lemma 8.
17) Compute the reduced order model using (3)–(5).

7. Methods Based on Decompositions of Pseudogramians.

Homotopy algorithms for solving optimal projection equations can be designed using decompositions of the pseudogramians based on contragredient transformations.

7.1. Descriptions of Methods.

The equations (6)–(7) can be considered in another, equivalent form. If (6) is multiplied by U_1 from the left, and (7) is multiplied by W_1 from the right, using (19)–(20), the following two equations are obtained:

$$U_1 A W_1 \Sigma W_1^t + \Sigma W_1^t A^t + U_1 B V B^t = 0, \tag{29}$$

$$A^t U_1^t \Sigma + U_1^t \Sigma U_1 A W_1 + C^t R C W_1 = 0. \tag{30}$$

The third equation

$$U_1 W_1 - I = 0 \tag{31}$$

determines the relationship between W_1 and U_1.

The matrix equations (29)–(31) contain $2\,n\,n_m + n_m^2$ scalar equations. On the other side, the only natural unknowns in (29)–(31), W_1, U_1 and diagonal Σ, contain $2\,n\,n_m + n_m$ variables. Hence, some additional techniques are necessary in order to make an exact match between the number of equations and the number of unknowns.

One approach is to consider Σ to be symmetric and all elements of Σ as unknowns. This is appropriate, since the equations (29)–(31) with a full symmetric Σ can be transformed into equations of the same form with a diagonal Σ by computing

$$\Sigma = T \bar{\Sigma} T^t, \quad \bar{W}_1 = W_1 T, \quad \bar{U}_1 = T^t U_1,$$

where $\bar{\Sigma}$ is diagonal and T is orthogonal.

Another approach is to consider the decomposition from the statement of Lemma 6, which leads to the equations

$$U_1 A W_1 \Sigma W_1^t + \Sigma W_1^t A^t + U_1 B V B^t = 0,$$

$$A^t U_1^t \Omega + U_1^t \Omega U_1 A W_1 + C^t R C W_1 = 0,$$

$$U_1 W_1 - I = 0,$$

which also have $2\,n\,n_m + n_m^2$ scalar equations. In this case the number of unknowns in W_1, U_1 and symmetric Σ and Ω is $2\,n\,n_m + n_m^2 + n_m$. An additional n_m equations can be obtained, for example, by requiring

$$\sigma_{ii} - \omega_{ii} = 0 \quad \text{for } i = 1, \ldots, n_m.$$

Alternatively, the number of unknowns can be reduced to $2\,n\,n_m + n_m^2$ if the diagonal elements of Ω are actually the diagonal elements of Σ.

7.2. Algorithm.

The following is a description of the algorithm for the method determined by the equations (29)–(31). The algorithm is based on the normal flow algorithm for dense Jacobian matrices described in [29]. Depending on the relative size of $F(a, 0, x_0)$ the algorithm may be modified. If $F(a, 0, x_0)$ is relatively large, computational experience shows that it is desirable (but not theoretically necessary) to enforce the symmetry of Σ along the homotopy path. This is done by observing that a symmetrized Σ corresponds to *some* homotopy map that *could* have been chosen initially. In effect, x_0 is changed in the homotopy map at each step along the homotopy zero curve γ. Obviously, in that case the homotopy map (10) must be used.

The algorithm is using the homotopy map (9) or (10), where $F(a,\lambda,x)$ is represented by three equations:

$$U_1\,A(\lambda)\,W_1\Sigma\,W_1^t + \Sigma\,W_1^t\,A^t(\lambda) + U_1\,B\,V\,B^t = 0, \tag{32}$$

$$A^t(\lambda)\,U_1^t\,\Sigma + U_1^t\,\Sigma\,U_1\,A(\lambda)\,W_1 + C^t\,R\,C\,W_1 = 0, \tag{33}$$

$$U_1\,W_1 - I = 0. \tag{34}$$

A detailed description of the algorithm for evaluation of the Jacobian matrix $D\rho_a(\lambda,x)$ is given in Appendix B. A program that implements this method is given in Appendix C.

In summary, the whole algorithm is:

1) Define D using formula (11) or (12).
2) Choose a starting point $x_0 = (Q_0, P_0)$ using one of strategies explained in Section 5.2. Compute $(W_1)_0$, $(U_1)_0$ and Σ_0 using a contragredient transformation.
3) Set $\lambda := 0$, $x := x_0 = ((W_1)_0, (U_1)_0, \Sigma_0)$.
4) Evaluate $\rho_a(\lambda,x)$ given by (9) or (10), and (32)–(34).
5) Evaluate $D\rho_a(\lambda,x)$.
6) Take a step along the curve and obtain $x_1 = (W_1, U_1, \Sigma)$, $\bar\lambda$.
7) Compute $\bar x_1 = (W_1, U_1, \bar\Sigma) = (W_1, U_1, (\Sigma + \Sigma^t)/2)$.
8) Change the homotopy to
$$F(a,\lambda,x) - (1-\lambda)v = 0,$$
where $v = F(a,\bar\lambda,\bar x_1)/(1-\bar\lambda)$.
9) If $\bar\lambda < 1$, then set $x := \bar x_1$, $\lambda := \bar\lambda$, and go to Step 4.
10) If $\bar\lambda \geq 1$, compute the solution $\bar x_1$ at $\bar\lambda = 1$. Compute the reduced order model by diagonalizing $\Sigma = T\,\bar\Sigma\,T^t$.

Note: if $F(a,0,x_0)$ is small, steps 7 and 8 can be omitted without a serious loss of efficiency.

8. Numerical Results.

Three examples are discussed here. The methods were tested on a number of additional examples, reported in [31]. The results were obtained using the method based on a contragredient transformation that has W_1, U_1 and Σ as unknowns, using the homotopy (9). For all examples $V = R = I$.

EXAMPLE 1 [13]. The system is given by

$$A = \begin{pmatrix} -0.05 & -0.99 \\ -0.99 & -5000.0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 100 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 100 \end{pmatrix}.$$

For the starting point

$$y_0 = \begin{pmatrix} 0 \\ 0.0099995 \\ 0.99995 \\ 0.0099995 \\ 0.99995 \\ 0.5 \end{pmatrix}$$

the homotopy algorithm converges to a solution corresponding to the model of order $n_m = 1$ given by

$$A_m = (-4998.078625), \quad B_m = (100.000194), \quad C_m = (100.000194).$$

This model yields the (minimum) cost $J = 96.078058$.

For the starting point

$$y_0 = \begin{pmatrix} 0 \\ 1 \\ 0.25 \\ 1.07 \\ -0.27 \\ 1 \end{pmatrix}$$

the solution found corresponds to the model of order $n_m = 1$ given by

$$A_m = (-0.485152), \quad B_m = (-0.0000011427), \quad C_m = (-0.000000073400),$$

which yields the (maximum) cost $J = 10100$. This example shows that the homotopy method can obtain different solutions.

EXAMPLE 2. This is a model of a synchronous machine connected to an infinite busbar [10]. The system is given by

$$A = \begin{pmatrix} -6.2036 & 15.054 & -9.8726 & -376.58 & 251.32 & -162.24 & 66.827 \\ 0.53 & -2.0176 & 1.4363 & 0 & 0 & 0 & 0 \\ 16.846 & 25.079 & -43.555 & 0 & 0 & 0 & 0 \\ 377.4 & -89.449 & -162.83 & 57.998 & -65.514 & 68.579 & 157.57 \\ 0 & 0 & 0 & 107.25 & -118.05 & 0 & 0 \\ 0.36992 & -0.14445 & -0.26303 & -0.64719 & 0.49947 & -0.21133 & 0 \\ 0 & 0 & 0 & 0 & 0 & 376.99 & 0 \end{pmatrix},$$

$$B = \begin{pmatrix} 89.353 & 0 \\ 376.99 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0.21133 \\ 0 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

A model of order $n_m = 3$ is

$$A_m = \begin{pmatrix} -0.0261157 & 9.349756 & -0.0528086 \\ -9.352068 & -0.912444 & 0.506220 \\ -0.0541716 & -0.506226 & -0.198770 \end{pmatrix},$$

$$B_m = \begin{pmatrix} -2.414471 & -0.571953 \\ 14.906052 & 0.0416151 \\ -14.944459 & 0.0512237 \end{pmatrix}, \quad C_m = \begin{pmatrix} -0.371712 & 0.00240265 & 0.0122915 \\ 2.453290 & 14.906110 & 14.944542 \end{pmatrix}.$$

This model yields the cost $J = 0.673079$.

A model of order $n_m = 4$ is

$$A_m = \begin{pmatrix} -37.55440 & -0.0546940 & 0.326197 & -0.0709427 \\ 0.0561170 & -0.0261155 & 9.349755 & -0.0528084 \\ 0.324384 & -9.352067 & -0.912440 & 0.506220 \\ 0.0705453 & -0.0541714 & -0.506226 & -0.198769 \end{pmatrix},$$

$$B_m = \begin{pmatrix} -2.666516 & -0.00624702 \\ -2.414464 & -0.571953 \\ 14.906036 & 0.0416151 \\ -14.944458 & 0.0512237 \end{pmatrix}, \quad C_m^t = \begin{pmatrix} 0.280123 & -2.651769 \\ -0.371711 & 2.453283 \\ 0.00240433 & 14.906094 \\ 0.0122916 & 14.944541 \end{pmatrix}.$$

This model yields the cost $J = 3.27495 \cdot 10^{-6}$.

EXAMPLE 3. This is a state space model of the transfer function between a torque activator and an approximately collocated torsional rate sensor for the ACES structure [6], located at NASA Marshall Space Flight Center, Huntsville, AL. The system in this example is of size $n = 17$, $m = 1$, $l = 1$. The nonzero elements of A are

$A(1,1) = A(2,2) = -0.031978272$, $A(1,2) = -A(2,1) = -78.54$,
$A(1,17) = 0.0097138566$, $A(2,17) = -0.0060463517$,
$A(3,3) = A(4,4) = -5.152212$, $A(3,4) = -A(4,3) = -51.457677$,
$A(3,17) = -0.021760771$, $A(4,17) = -0.0054538246$.
$A(5,5) = A(6,6) = -0.1351159$, $A(5,6) = -A(6,5) = -15.417859$,
$A(5,17) = -0.02179972$, $A(6,17) = -0.015063913$,
$A(7,7) = A(8,8) = -0.42811443$, $A(7,8) = -A(8,7) = -14.698408$,
$A(7,17) = -0.01042631$, $A(8,17) = -0.0088479697$,
$A(9,9) = A(10,10) = -0.064896745$, $A(9,10) = -A(10,9) = -12.077045$,
$A(9,17) = -0.030531575$, $A(10,17) = -0.030260987$,
$A(11,11) = A(12,12) = -0.048520356$, $A(11,12) = -A(12,11) = -8.9654448$,
$A(11,17) = -0.016843335$, $A(12,17) = -0.011449591$,
$A(13,13) = A(14,14) = -0.036781718$, $A(13,14) = -A(14,13) = -4.9057426$,
$A(13,17) = -0.1248007$, $A(14,17) = -0.0005136047$,
$A(15,15) = A(16,16) = -0.025112482$, $A(15,16) = -A(16,15) = -3.8432892$,
$A(15,17) = -0.035415526$, $A(16,17) = -0.028115589$,
$A(17,17) = -92.399784$.

The matrices B and C are

$$B = \begin{pmatrix} 1.8631111 \\ -1.1413786 \\ -1.2105758 \\ 0.31424169 \\ 0.013307797 \\ -0.211128913 \\ 0.19552894 \\ -0.037391511 \\ -0.01049736 \\ -0.011486242 \\ -0.029376402 \\ 0.0082391613 \\ -0.012609562 \\ -0.0022040505 \\ -0.030853234 \\ 0.0011671662 \\ 0 \end{pmatrix}, \quad C^t = \begin{pmatrix} -0.0097138566 \\ 0.0060463517 \\ 0.021760771 \\ -0.0054538246 \\ 0.02179972 \\ 0.015063913 \\ -0.01042631 \\ -0.0088479697 \\ 0.030531575 \\ 0.030260987 \\ 0.016843335 \\ 0.011449591 \\ 0.1248007 \\ -0.0005136047 \\ 0.035415526 \\ 0.028115589 \\ 184.79957 \end{pmatrix}.$$

A model of order $n_m = 8$ is given by

$$A = \begin{pmatrix} -70.1470 & 21.9187 & -2.7406 & 2.9917 & -0.3721 & 0.2286 & 0.0246 & 0.0836 \\ 54.1619 & -32.1862 & 4.6829 & 9.2995 & -0.4958 & 0.1804 & 0.0289 & 0.0934 \\ 3.5118 & -4.6512 & -0.2083 & -51.3962 & 0.1211 & -0.0130 & -0.0049 & -0.0157 \\ -22.2535 & 19.0459 & 51.8525 & -12.0437 & 1.0945 & -0.6389 & -0.0741 & -0.2438 \\ 1.2271 & -1.1976 & -0.2000 & 1.1602 & -0.1936 & 15.4491 & 0.0243 & 0.0807 \\ 0.5249 & -0.5415 & -0.0764 & 0.6934 & -15.4500 & -0.0147 & -0.0125 & 0.0414 \\ -0.0705 & 0.0708 & 0.0106 & -0.0770 & 0.0238 & 0.0122 & 0.0181 & -78.5742 \\ -0.2393 & 0.2397 & 0.0357 & -0.2610 & 0.0803 & 0.0415 & 78.5083 & -0.0823 \end{pmatrix},$$

$$B_m = \begin{pmatrix} -0.05753 \\ -0.06445 \\ 0.01043 \\ 0.16983 \\ -0.05959 \\ 0.02622 \\ 0.04591 \\ 0.15167 \end{pmatrix}, \quad C_m^t = \begin{pmatrix} -0.16432 \\ 0.16512 \\ 0.02442 \\ -0.18165 \\ 0.05966 \\ 0.02629 \\ -0.04472 \\ -0.15162 \end{pmatrix}.$$

This model yields the cost $J = 3.95223 \cdot 10^{-5}$.

9. Conclusion.

This paper has considered the use of probability-one homotopy methods to solve the optimal projection equations for the model reduction problem. Four different approaches were given for solving the equations. The first approach is based on solving the optimal projection equations in their original form. The three other approaches stem from a decomposition of the pseudogramians based on a contragredient transformation and proved to be more numerically robust than the first approach. The "best" algorithm was shown to be effective in finding the optimal reduced order models for several examples.

The number of variables associated with the first approach is of order n^2 (n is the dimension of the original model), while the number of variables for the latter approaches is of order $n\,n_m$ (n_m is the dimension of the reduced order model). Future research will involve the development of homotopy algorithms with fewer variables. It appears that by using a more rudimentary form of the optimal projection equations, it is possible to reduce the number of variables to be of order $n(m+l)$ (m and l are, respectively, the number of inputs and outputs). Future research will also consider the H_2/H_∞ reduced order model problem [8].

REFERENCES

[1] M. Aoki, "Control of large-scale dynamic systems by aggregation," *IEEE Trans. Autom. Contr.*, vol. AC-13, pp. 246–253, 1968.

[2] R. H. Bartels and G. W. Stewart, "Solution of the matrix equation $A\,X + X\,B = C$," *Comm. of the ACM*, vol. 15, pp. 820–826, 1972.

[3] D. S. Bernstein and D. C. Hyland, "The optimal projection equations for reduced-order state estimation," *IEEE Trans. Autom. Contr.*, vol. AC-30, pp. 583–585, 1985.

[4] W. L. Brogan, *Modern Control Theory*, Englewood Cliffs: Prentice Hall, 1985.

[5] S. L. Campbell and C. D. Meyer, Jr., *Generalized Inverses of Linear Transformations*, London, England: Pitman, 1979, ch. 7, pp. 120–127.

[6] E. G. Collins, Jr., D. J. Phillips and D. C. Hyland, "Robust decentralized control laws for the ACES structure," *Contr. Sys. Mag.*, pp. 62–70, April 1991.

[7] K. Glover, "All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds," *Int. J. Contr.*, vol. 39, pp. 1115–1193, 1984.

[8] W. M. Haddad and D. S. Bernstein, "Combined L_2/H_∞ model reduction," *Int. J. Contr.*, vol. 49, pp. 1523–1535, 1989.

[9] J. Z. Hearon and J. W. Evans, "Differentiable generalized inverses," *Jour. Res. NBS*, vol. 72B, No. 2, pp. 109–113, 1968.

[10] J. Hickin and N. K. Sinha, "Model reduction for linear multivariable systems," *IEEE Trans. Autom. Contr.*, vol. AC-25, pp. 1121–1127, 1980.

[11] D. C. Hyland and D. S. Bernstein, "The optimal projection equations for fixed-order dynamic compensation," *IEEE Trans. Autom. Contr.*, vol. AC-29, pp. 1034–1037, 1984.

[12] D. C. Hyland and D. S. Bernstein, "The optimal projection equations for model reduction and the relationships among the methods of Wilson, Skelton, and Moore," *IEEE Trans. Autom. Contr.*, vol. AC-30, pp. 1201–1211, 1985.

[13] P. T. Kabamba, "Balanced gains and their significance for L^2 model reduction," *IEEE Trans. Autom. Contr.*, vol. AC-30, pp. 690–693, 1985.

[14] S. Kung and D. W. Lin, "Optimal Hankel-norm model reductions: Multivariable systems," *IEEE Trans. Autom. Contr.*, vol. AC-26, pp. 832–852, 1981.

[15] S. Kung and D. W. Lin, "A state-space formulation for optimal Hankel-norm approximations," *IEEE Trans. Autom. Contr.*, vol. AC-26, pp. 942–946, 1981.

[16] C. P. Kwong, "Optimal chained aggregation for reduced-order modelling," *Int. J. Contr.*, vol. 35, pp. 964–982, 1982.

[17] B. C. Moore, "Principal component analysis in linear systems: Controllability, observability, and model reduction," *IEEE Trans. Autom. Contr.*, vol. AC-26, pp. 17–32, 1981.

[18] L. Pernebo and L. M. Silverman, "Model reduction via balanced state space representations," *IEEE Trans. Autom. Contr.*, vol. AC-27, pp. 382–387, 1982.

[19] S. Richter and E. G. Collins, "A homotopy algorithm for reduced order compensator design using the optimal projection equations," in *Proc. IEEE Conf. Dec. Contr.*, Tampa, FL, pp. 506–511, Dec 1989.

[20] R. E. Skelton, "Cost decomposition of linear systems with application to model reduction," *Int. J. Contr.*, vol. 32, pp. 1031–1055, 1980.

[21] R. E. Skelton and P. C. Hughes, "Modal cost analysis for linear matrix-second-order systems," *J. Dynamic Systems, Measurement and Control*, vol. 102, pp. 151–158, 1980.

[22] R. E. Skelton and A. Yousuff, "Component cost analysis of large scale systems," *Int. J. Contr.*, vol. 37, pp. 285–304, 1983.

[23] S. Smale, "A convergent process of price adjustment and global Newton methods," *J. Math. Econom.*, vol. 3, pp. 107–120, 1976.

[24] J. T. Spanos, M. H. Milman and D. L. Mingori, "Optimal model reduction and frequency-weighted extension," *AIAA paper number AIAA–90–3345–CP*, pp. 271–284, 1990.

[25] C. de Villemagne and R. E. Skelton, "Model reductions using a projection formulation," *Int. J. Contr.*, vol. 46, pp. 2141–2169, 1987.

[26] L. T. Watson, "Globally convergent homotopy methods: A tutorial," *Appl. Math. Comput.*, vol. 31BK, pp. 369–396, 1989.

[27] L. T. WATSON, "Globally convergent homotopy algorithms for nonlinear systems of equations," *Nonlinear Dynamics*, vol. 1, pp. 143–191, 1990.

[28] L. T. WATSON, "Numerical linear algebra aspects of globally convergent homotopy methods," *SIAM Rev.*, vol. 28, pp. 529–545, 1986.

[29] L. T. WATSON, S. C. BILLUPS AND A. P. MORGAN, "HOMPACK: A suite of codes for globally convergent homotopy algorithms," *ACM Trans. Math. Software*, vol. 13, pp. 281–310, 1987.

[30] D. A. WILSON, "Optimum solution of model-reduction problem," *Proc. IEE*, vol. 117, pp. 1161–1165, 1970.

[31] D. ŽIGIĆ, "Homotopy methods for solving the optimal projection equations for the reduced order model problem," , MS thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA, May 1991.