Proceedings of the 35th
Conference on Decision and Control
Kobe, Japan • December 1996

L_2 Optimal Model Reduction

Wei-Yong Yan* James Lam†

* School of Electrical and Electronic Engineering
Nanyang Technological University
Nanyang Avenue, Singapore 639798

† Department of Mechanical Engineering
University of Hong Kong
Pokfulam Road, Hong Kong

Abstract

This paper deals with the problem of computing an L_2-optimal reduced-order model for a given stable multivariable linear system. By way of orthogonal projection, the problem is formulated as that of minimizing the L_2 model reduction cost over the Stiefel manifold so that the stability constraint on reduced-order models is automatically satisfied and thus totally avoided in the new problem formulation. The closed form formula for the gradient of the cost over the manifold is derived, from which a gradient flow is formed as an ordinary differential equation. A number of nice properties about such a flow are obtained. Among them are the decreasing property of the cost along the ODE solution and the convergence of the flow from any starting point in the manifold. Furthermore, an explicit iterative convergent algorithm is developed from the flow and inherits the properties that the iterates remain on the manifold starting from any orthogonal initial point and that the model-reduction cost is decreasing to minimums along the iterates.

1. Introduction

One of important optimality-based techniques for model reduction is to minimize the L_2 norm of the model mismatch between the original model and a reduced-order one. This minimization problem for a given stable plant over all stable models of fixed lower order has received a great deal of attention over the past three decades. Still, rigorous and convergent algorithms have remained to be found in the general multi-input multi-output case. So far, the most commonly taken approach to L_2-optimal model reduction problem is to work with first order necessary conditions for optimality, which were developed and simplified in one way or another by Meier and Luenberger [1], Wilson [2], Hyland and Bernstein [3], Halevi [4], Bryson and Carrier [5], Baratchart *et al.* [6], and more recently Spanos *et al.* [7]. Accordingly, they proposed their respective algorithms to seek a solution satisfying the conditions expressed in terms of nonlinear matrix equations. Many of the algorithms lack the proof of convergence and mathematical rigor, and some of them may even become divergent for certain initial conditions. Though Baratchart *et al.* [6] and Spanos *et al.* [7] established the convergence of their respective algorithms under certain conditions, the algorithms are only applicable to the single-input single-output case.

So far, it seems unclear whether the global minimum of the cost exists or not in the continuous-time multi-input multi-output case though the answer to this question in the discrete-time case was positive according to Baratchart [8]. This issue inevitably sheds some doubt on the theoretic basis of the above approach. Moreover, as pointed out by Spanos *et al.* [7], there are two technical difficulties associated with the approach; one is the stability constraint on reduced-order models and the other is the unboundedness of the level sets of the L_2 cost functional. It goes without saying that the first one is fundamentally intricate to accommodate and thus represents a major obstacle to the effectiveness of any algorithm based on that approach. We believe that this difficulty stems from direct parametrizations of all the reduced-order models in one form or another.

In this paper, we take a different approach to the L_2-optimal model reduction problem in the continuous-time case. The main idea is to treat the minimization problem over a subclass of stable reduced-order models parameterized by a projection matrix instead of the whole class of all the reduced-order models. It can be heuristically argued and numerically verified that the global minimum of the cost over such a subclass is very close if not identical to that over the entire class. In addition, the restriction to this subclass enables one to avoid the stability constraint entirely and leads to a more tractable minimization problem over the Stiefel manifold, which is compact. Our main purpose is to develop both continuous and iterative convergent algorithms which are rigorous and universally applicable.

The paper is briefly outlined as follows. In the next section, we modify the L_2-optimal model reduction problem as an unconstrained minimization problem over the Stiefel manifold. Section 3 centers on the development of the gradient flow of the model reduction cost and establishment of its associated properties including convergence by using differential manifold techniques. In Section 4, we turn to discuss recursive algorithms. The last section contains some conclusions.

2. Problem Formulation

Consider a linear time-invariant stable system $G(s)$ with the realization

$$\dot{x} = Ax + Bu \tag{2.1}$$
$$y = Cx \tag{2.2}$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times p}$, $C \in \mathbb{R}^{q\times n}$. An admissible reduced-order model $G_m(s)$ is defined to be of the form

$$\dot{x}_m = A_m x_m + B_m u \tag{2.3}$$
$$y_r = C_m x_m \tag{2.4}$$

where $A_m \in \mathbb{R}^{m\times m}$, $B \in \mathbb{R}^{m\times p}$, $C \in \mathbb{R}^{q\times m}$ with A a stable matrix. The mismatch between the full-order $G(s)$ and reduced-order $G_m(s)$ will be measured by the square of the L_2 norm of their difference $G_e(s)$, i.e.,

$$J(A_m, B_m, C_m) \triangleq \|G_e(s)\|_2^2$$

which is often termed the quadratic model-reduction cost.

The so-called L_2 or quadratically optimal model reduction problem is to minimize $J(A_m, B_m, C_m)$ over all the admissible reduced-order models $G_m(s)$. Note that one realization $(A_e,\ B_e,\ C_e)$ of the error model $G_e(s)$ is given by

$$(A_e,\ B_e,\ C_e) = \left(\begin{bmatrix} A & 0 \\ 0 & A_m \end{bmatrix},\ \begin{bmatrix} B \\ B_m \end{bmatrix},\ \begin{bmatrix} C & -C_m \end{bmatrix} \right)$$

Then it is a standard fact that the model-reduction cost can be expressed in terms of the controllability Gramian L_c and observability Gramian L_o of this realization. Namely, there holds

$$J(A_m, B_m, C_m) = \mathrm{trace}(C_e L_c C_e^T) = \mathrm{trace}(B_e^T L_o B_e) \tag{2.5}$$

with

$$A_e L_c + L_c A_e^T + B_e B_e^T = 0 \tag{2.6}$$
$$A_e^T L_o + L_o A_e + C_e^T C_e = 0 \tag{2.7}$$

It is known from [3] that any minimizing solution (A_m, B_m, C_m) must be of the form

$$(A_m, B_m, C_m) = (TAV,\ TB,\ CV) \tag{2.8}$$

where $V \in \mathbb{R}^{n\times m}$ and $T \in \mathbb{R}^{m\times n}$ satisfy

$$TV = I \tag{2.9}$$

Hence, the original model reduction problem amounts to minimizing $J(TAV,\ TB,\ CV)$ with respect to $(T,V) \in \mathbb{R}^{m\times n} \times \mathbb{R}^{n\times m}$ subject to the two constraints

(i) $TV = I$ (ii) TAV is stable

This is essentially a nonlinear optimization problem subject to both equality and inequality constraints as the stability constraint can be expressed in terms of inequalities by the Hurwitz criterion. Though it may be possible to use some constrained optimization techniques to find a local minimum, the computation involved could be formidable.

To formulate a more tractable problem without involving the stability constraint, we observe that T given by

$$T = V^\dagger = (V^T V)^{-1} V^T \tag{2.10}$$

satisfies the constraint (2.9) for any V of full column rank. It is therefore interesting to consider the following modified problem:

minimize $J(V) \triangleq J(V^\dagger AV,\ V^\dagger B,\ CV)$
subject to stability of $V^\dagger AV$

In fact, the above modification can be motivated from a geometric point of view. To see this, we decompose the state space $\mathbb{R}^n$ into

$$\mathbb{R}^n = \mathrm{range}\,(V) + \mathrm{range}\,(V)^\perp$$

That is, any state $x \in \mathbb{R}^n$ is expressed as

$$x = Vw + e$$

with $w \in \mathbb{R}^m$ and $e \in \mathrm{range}\,(V)^\perp$. Here, Vw is the orthogonal projection of the state onto the subspace $\mathrm{range}\,(V)$. By rewriting the state equation as

$$V\dot{w} - (AVw + Bu) = Ae - \dot{e}$$

and appealing to the fact that $\|Vz - s\|_2$ is minimized at $z = V^\dagger s$ for any given $s \in \mathbb{R}^n$, one sees that the best approximate to $\dot{w}$ given w is $V^\dagger(AVw + Bu)$ in the sense that e has the minimal effect. Removing e from the output equation naturally results in a reduced-order model

$$\dot{w} = V^\dagger AVw + V^\dagger Bu \tag{2.11}$$
$$y = CVw \tag{2.12}$$

As such, the above modified minimization problem may well be thought of as finding a dominant state subspace of dimension m, which is spanned by the columns of V. As a matter of fact, such a projection idea shares the same underlying principle with the method of aggregation [9, 10, 11].

Perhaps it is also interesting and relevant to note that the model-reduction cost can be expressed as

$$\begin{aligned} &J(TAV,\ TB,\ CV) \\ &= \|sC(sI - VTA)^{-1}(VT - I)(sI - A)^{-1}B\|_2^2 \end{aligned}$$

This implies that the cost only depends on the product VT and contains as one factor $VT - I$ whose norm is minimized at $T = V^\dagger$ as a function of T.

Another crucial implication of the above observation is that the modified problem can be virtually reduced to an unconstrained minimization problem on a Stiefel manifold, which will be dealt with in this paper. To see this, note that

$$J(V^\dagger AV,\ V^\dagger B,\ CV) = J(U^TAU,\ U^TB,\ CU) \quad (2.13)$$

with $U = V(V^TV)^{-1/2}$ due to

$$UU^T = V(V^TV)^{-1}V^T = VV^\dagger$$

Moreover, the stability of U^TAU is automatically guaranteed provided that $A + A^T$ is negative definite. We claim that such a property of the system realization can be assumed without loss of generality. In fact, since A is stable, for any negative definite matrix Q there exists a nonsingular square matrix T such that

$$ATT^T + TT^TA^T = 0$$

from which it is seen that the use of the so obtained T as a similarity transformation will result in a new realization with the required property. Besides, Lam and Hung [12] have shown that any balanced realization with distinct Hankel singular values automatically possesses the property as well.

In this way, it is obvious that with the above mentioned assumption on the system matrix A, the minimal model-reduction cost over the reduced-order model set

$$\begin{aligned} \{(A_m, B_m, C_m) = (V^\dagger AV,\ V^\dagger B,\ CV);& \\ V \in \mathbb{R}^{n\times m} \text{ and } V^\dagger AV \text{ is stable}\}& \end{aligned}$$

is exactly equal to that over

$$\begin{aligned} \{(A_m, B_m, C_m) = (U^TAU,\ U^TB,\ CU);& \\ U \in \mathbb{R}^{n\times m} \text{ and } U^TU = I\}& \end{aligned}$$

Since this latter set is a compact set, the minimum model-reduction cost over it indeed exists. Moreover, as will be numerically verified in the sequel, the minimum cost over the compact set is hardly different from that over all the admissible reduced-order models.

We now end this section by formally posing the following projection model reduction problem, which will be called the L_2-PMR problem.

L_2-Projection Model Reduction Problem: Given the realization system (2.1)-(2.2) with $A + A^T$ being negative definite, minimize

$$\mathfrak{J}(U) \triangleq J(U^TAU,\ U^TB,\ CU)$$

over the Stiefel manifold

$$St(m,n) = \{U \in \mathbb{R}^{n\times m} \mid U^TU = I\}$$

3. Gradient Flow on Manifold

In this section, we aim to solve the L_2-PMR problem posed last section using the gradient flow approach. Recall that an optimal solution to this problem exists. So the question is really how to find one. Also, recall that there is no loss of generality in assuming that $A + A^T$ is negative definite for the original realization (2.1)-(2.2), which will be our standing assumption throughout. In addition, we adopt the convention that $\|\cdot\|$ means the spectral norm of a matrix i.e. the maximum singular value while $\|\cdot\|_F$ means the Frobenius norm.

Let us first obtain a more explicit formula for $\mathfrak{J}(U)$. To do this, partition the solutions L_c and L_o to the Lyapunov equations (2.6) and (2.7) as

$$L_c = \begin{bmatrix} \Sigma_c & X \\ X^T & P \end{bmatrix} \quad \text{and} \quad L_o = \begin{bmatrix} \Sigma_o & Y \\ Y^T & Q \end{bmatrix}$$

As a result, the Lyapunov equations (2.6) and (2.7) become equivalent to

$$A\Sigma_c + \Sigma_cA^T + BB^T = 0 \quad (3.1)$$
$$AX + XU^TA^TU + BB^TU = 0 \quad (3.2)$$
$$U^TAUP + PU^TA^TU + U^TBB^TU = 0 \quad (3.3)$$
$$A^T\Sigma_o + \Sigma_oA + C^TC = 0 \quad (3.4)$$
$$A^TY + YU^TAU - C^TCU = 0 \quad (3.5)$$
$$U^TA^TUQ + QU^TAU + U^TC^TCU = 0 \quad (3.6)$$

and the cost $\mathfrak{J}(U)$ can be rewritten as

$$\begin{aligned} \mathfrak{J}(U) &= \text{trace}[C^TC(\Sigma_c + UPU^T - 2XU^T)] \\ &= \text{trace}[BB^T(\Sigma_o + UQU^T + 2YU^T)] \end{aligned}$$

Quite obviously, $\mathfrak{J}(U)$ is a smooth function on the manifold $St(m,n)$. From [13], its tangent space at a given $U \in St(m,n)$ is known to be

$$T_USt(m,n) = \{\Pi \in \mathbb{R}^{n\times m} \mid \Pi^TU + U^T\Pi = 0\}$$

By endowing $T_USt(m,n)$ with the inner product defined by

$$<\eta,\ \xi> \triangleq 2\text{trace}(\eta^T\xi), \quad \text{for } \eta,\ \xi \in T_USt(m,n)$$

$St(m,n)$ becomes a Riemannian manifold. Also, note that the derivative $D\mathfrak{J}_U$ of $\mathfrak{J}(U)$ at $U \in St(m,n)$ is a linear functional on the tangent space $T_USt(m,n)$, and that the gradient $\nabla\mathfrak{J}(U)$ of $\mathfrak{J}(U)$ at $U \in St(m,n)$ is a tangent vector in $T_USt(m,n)$ such that

$$D\mathfrak{J}_U(\Pi) = <\nabla\mathfrak{J}(U),\ \Pi>, \quad \forall\, \Pi \in T_USt(m,n)$$

The explicit expression of $\nabla\mathfrak{J}(U)$ is now given in the following lemma.

Lemma 3.1 *For any $U \in St(m,n)$, there holds*

$$\nabla \mathfrak{J}(U) = (I - UU^T)R$$

where

$$R \triangleq (-C^TC + A^TUY^T)X + (C^TCU + A^TUQ)P + (BB^T + AUX^T)Y + (BB^TU + AUP)Q \quad (3.7)$$

Proof: The proof is omitted. □

At this point, it is worth pointing out that the above gradient is different from the gradient of $\mathfrak{J}(U)$ as a usual function defined on $\mathbb{R}^{n\times m}$. This latter gradient is in $\mathbb{R}^{n\times m}$ but not necessarily in the tangent space $T_U St(m,n)$. From Appendix B, as a matter of fact, it is found equal to R, as defined in the above lemma.

As an immediate consequence of the above lemma, it follows from advanced calculus that any minimum point of $\mathfrak{J}(U)$ in $St(m,n)$ must satisfy

$$(I - UU^T)R = 0 \quad \text{and} \quad U^TU = I \qquad (3.8)$$

since any solution in $St(m,n)$ is a critical point of $\mathfrak{J}(U)$. So (3.8) expresses a first-order necessary condition for a minimum point. However, solving such an equation does not seem to be a sensible or effective way to go about finding a minimum point as it may be very difficult to solve and may have multiple solutions.

Remark 3.1 *It can be easily verified that U^TR is always a symmetric matrix for any $U \in St(m,n)$, which is instrumental in constructing iterative algorithms later on. In fact, there holds*

$$U^TR = Y^TAX + QU^TAUP + X^TA^TY + PU^TA^TUQ$$

Therefore, the first equation of (3.8) can be alternatively expressed as

$$R = UR^TU.$$

Now with the formula for $\nabla \mathfrak{J}(U)$ available, we can form the following gradient flow

$$\boxed{\dot{U} = (UU^T - I)R} \qquad (3.9)$$

as a basis for solving the L_2-PMR problem. Regarding this ordinary differential equation, it is natural to inquire questions such as whether a solution to the ODE always exists and lies on the manifold $St(m,n)$ on the whole time interval for any given initial value in $St(m,n)$, how the model-reduction cost evolves along a solution, and whether the solution can converge to a critical point of $\mathfrak{J}(U)$ on $St(m,n)$. The answers to these questions are crucial in order for the ODE to be able to serve as an continuous-time algorithm for computing an optimal solution to the L_2-PMR problem. We now address the raised issues by stating the following theorem, which summarizes the main features of the gradient flow.

Theorem 3.1 *Let the initial condition of (3.9) be given by*

$$U(0) = U_0 \in St(m,n)$$

Then,

1. *the ODE (3.9) has a unique solution $U(t)$ defined for all $t \geq 0$;*
2. *the solution $U(t)$ stays in $St(m,n)$ for all $t \geq 0$;*
3. *the cost $\mathfrak{J}(U)$ is non-increasing along $U(t)$ with*

$$\mathfrak{J}(U(s_2)) - \mathfrak{J}(U(s_1)) = -2\int_{s_1}^{s_2} \left\|(I - UU^T)R\right\|_F^2 dt, \quad \forall\, s_2 \geq s_1 \geq 0$$

4. *there holds*

$$\lim_{t\to\infty} \dot{U}(t) = \lim_{t\to\infty} (UU^TR - R) = 0$$

5. *the solution $U(t)$ converges to a connected component of the set of critical points of $\mathfrak{J}(U)$;*
6. *there exists a time sequence $\{s_k\}$ with*

$$s_k \geq 0 \quad \text{and} \quad \lim_{k\to\infty} s_k = \infty$$

such that the corresponding sequence $U(s_k)$ converges to a critical point of $\mathfrak{J}(U)$;

Proof: The first two statements follow from the compactness properties of the Stiefel manifold. In fact, it is straightforward to verify that the derivative of $U^T(t)U(t)$ is identically zero for all $t \geq 0$. Statement 3 is immediately obtained by noting that the derivative of $\mathfrak{J}(U(t))$ is equal to

$$\begin{aligned}\dot{\mathfrak{J}}(U(t)) &= <\nabla\mathfrak{J}(U),\, \dot{U}> \\ &= -2\text{trace}\left[R^T(I - UU^T)R\right] \\ &= -2\left\|(I - UU^T)R\right\|_F^2 \leq 0\end{aligned}$$

Statement 4 is due to the two facts — finiteness of the integral $\int_0^\infty \left\|(I - UU^T)R\right\|_F^2 dt$ and uniform continuity of $\dot{U}(t)$ on $[0,\ \infty)$. Finally, the last two statements are typical properties of a gradient flow on a Riemannian manifold. □

The above summarized properties of the gradient flow (3.9) give one confidence in finding a global minimum of $\mathfrak{J}(U)$ by integrating the differential equation, which can be done using any numerical ODE package, e.g. in Matlab. Since the model-reduction cost is getting smaller and smaller as the iteration goes on and no finite escape time will occur, one can keep on solving the ODE until a satisfactory suboptimal solution is reached. Finally, the last two statements suggest that a minimum point could be found from the solution history. In particular, it is guaranteed that if the cost has only isolated minimum points, the solution $U(t)$ is bound to converge to one of them.

Remark 3.2 *It should be pointed out that if the initial U_0 does not happen to be a critical point, then the cost $\mathfrak{J}(U)$ is actually strictly decreasing along the ODE solution $U(t)$, which is because of the uniqueness of solutions to an ODE.*

4. Iterative Gradient Flow

In this section, we will consider discretizing the gradient flow (3.9), which is necessary or desirable in order to take full advantage of digital computers as far as computation is concerned. In other words, we will seek iterative algorithms which can produce a sequence of iterates whose corresponding model reduction costs are decreasing to its minimum. Recall that the projection matrix U is required to be orthogonal. This restriction makes it difficult if not impossible to apply common discretizing techniques such as Runge-Kutta methods to derive an efficient iterative algorithm.

In what follows, a general form of iterative algorithm will first be suggested which automatically guarantees that all the iterates generated evolve on the manifold $St(m,n)$ for an arbitrary step size. Two schemes for selecting the step size will then be developed — one is constant and the other is varying and more effective.

We start by noting that the gradient flow can be rewritten as

$$\dot{U} = \Gamma U \tag{4.1}$$

because of Remark 3.1, where Γ is defined by

$$\Gamma = UR^T - RU^T \tag{4.2}$$

In addition, it is trivial but vital to observe that Γ is skew-symmetric. As a result, the matrix exponential $e^{t\Gamma}$ is orthogonal for any real scalar t. With this observation and the special structure of the gradient flow, it seems natural to propose the algorithm of the following form

$$\boxed{U_{k+1} = e^{t_k \Gamma_k} U_k} \tag{4.3}$$

where Γ_k is associated with U_k via (4.2) and (3.7), and t_k is the k-th step size to be determined. One nice thing about this algorithm is its ability to generate a sequence of orthogonal matrices from any starting orthogonal U_0 for any step size, and another is its simplicity in form in spite of the involved calculation of the matrix exponential. Of course, for such an algorithm to work, it remains to develop a mechanism for selecting the step size t_k so that the algorithm can converge to an orthogonal U at which the model reduction cost is minimum. As will be turned out, a certain constant step size can be chosen for this purpose.

Understandably, a workable step size should consistently reduce the model-reduction cost as the iteration goes on. With this in mind, we proceed by establishing the following auxiliary lemma before coming up with a scheme for choosing a constant step size.

Lemma 4.1 *Consider equations (3.1)-(3.6). Let $U \in St(m,n)$ be any differentiable function of t with the derivative U', and let R be defined by (3.7) accordingly. Then R and its derivative R' satisfy*

$$\|R(t)\|_F \leq \alpha_1 \tag{4.4}$$
$$\|R'(t)\|_F \leq \alpha_2 \|U'(t)\|_F \tag{4.5}$$

where

$$\alpha_1 \triangleq \frac{4\|B\|^2 \|C\|^2 (\alpha + \|A\|)}{\alpha^2}$$

$$\alpha_2 \triangleq \frac{2\|B\|^2 \|C\|^2 \left(3\alpha^2 + 10\|A\|\alpha + 8\|A\|^2\right)}{\alpha^3}$$

and α denotes the minimum eigenvalue of $-A - A^T$.

Proof: The proof is omitted. □

Theorem 4.1 *Consider the iterative algorithm (4.3) with $U_0 \in St(m,n)$ and*

$$0 < t_k < \frac{\sqrt{2}}{\alpha_1 + \sqrt{2}\alpha_2} \tag{4.6}$$

where α_1 and α_2 are defined as in Lemma 4.1. Then there holds

$$\mathfrak{J}(U_{k+1}) \leq \mathfrak{J}(U_k), \quad \forall\, k = 0,\, 1,\, 2,\, \cdots;$$

moreover, the equality holds if and only if U_k becomes a critical point of $\mathfrak{J}(U)$.

Proof: Set

$$U(t) = e^{t\Gamma_k} U_k$$

and let $R(t)$ be the corresponding R defined via (3.7). Then it is clear that $U(0) = U_k$ and $R(0) = R_k$. By the Taylor expansion, there exists some θ between 0 and t such that

$$\mathfrak{J}(U(t)) - \mathfrak{J}(U_k) = t\mathfrak{J}'(U(0)) + \frac{t^2}{2}\mathfrak{J}''(U(\theta))$$

It can be shown that

$$\begin{aligned}
\mathfrak{J}'(U(t)) &= 2\text{trace}\left[R^T(t)U'(t)\right] \\
&= 2\text{trace}\left[R^T(t)\Gamma_k U(t)\right] \\
\mathfrak{J}''(U(t)) &= 2\text{trace}\left[(R')^T(t)\Gamma_k U(t) + R^T(t)\Gamma_k^2 U(t)\right]
\end{aligned}$$

which imply that

$$\begin{aligned}
\mathfrak{J}'(U(0)) &= 2\text{trace}\left(R_k^T \Gamma_k U_k\right) = -\text{trace}\left(\Gamma_k^T \Gamma_k\right) \\
|\mathfrak{J}''(U(t))| &\leq 2\left(\|R'(t)\|_F \|\Gamma_k\|_F + \|R(t)\|_F \left\|\Gamma_k^2\right\|_F\right)
\end{aligned}$$

Furthermore, it follows by Lemma 4.1 that

$$\begin{aligned}
|\mathfrak{J}''(U(t))| &\leq 2\left(\alpha_2 \|U'(t)\|_F \|\Gamma_k\|_F + \alpha_1 \left\|\Gamma_k^2\right\|_F\right) \\
&\leq 2\left(\alpha_2 \|\Gamma_k\|_F^2 + \alpha_1 \|\Gamma_k\| \|\Gamma_k\|_F\right)
\end{aligned}$$

Consequently, there results

$$\begin{aligned}&\mathfrak{J}(U(t)) - \mathfrak{J}(U_k) \\ &\leq -t\left\|\Gamma_k\right\|_F^2 + t^2\left(\alpha_2\left\|\Gamma_k\right\|_F^2 + \alpha_1\left\|\Gamma_k\right\|\left\|\Gamma_k\right\|_F\right) \quad (4.7)\end{aligned}$$

As Γ_k is skew-symmetric, all its eigenvalues must be on the imaginary axis and thus the multiplicity of every nonzero singular value is at least 2, which implies that $\|\Gamma_k\| \leq \|\Gamma_k\|_F / \sqrt{2}$. Therefore, it is true that $\mathfrak{J}(U(t)) \leq \mathfrak{J}(U_k)$ for any t with

$$0 < t < \frac{\sqrt{2}}{\alpha_1 + \sqrt{2}\alpha_2}$$

and that the equality holds if and only if $\Gamma_k = 0$. □

Two important remarks are in order.

Remark 4.1 *Quite clearly, the model-reduction cost* $\mathfrak{J}(U_k)$ *is convergent as* $k \to \infty$.

Remark 4.2 *Given a constant step-size* $t_k = c$ *satisfying the inequality* (4.6), *note from* (4.7) *that*

$$\|\Gamma_k\|^2 \leq \frac{\mathfrak{J}(U_k) - \mathfrak{J}(U_{k+1})}{2c - c^2\left(\alpha_2 + \alpha_1/\sqrt{2}\right)}$$

As a result, it is true that

$$\lim_{k\to\infty} \Gamma_k = 0$$

which implies that U_k *generated by the algorithm will approach the critical points satisfying the first-order necessary conditions* (3.8) *for optimality as* $k \to \infty$.

5. Conclusions

The L_2 optimal model reduction problem has approximately been formulated as an unconstrained minimization problem over the Stiefel manifold. Using the differential techniques, we have derived explicit formulas for the gradient of the model reduction cost function over the manifold. Several convergent algorithms have been proposed. The first one is given in terms of an ordinary differential equation formed by the gradient flow, and concerning this algorithm a number of nice theoretical properties are obtained. For example, the cost is always decreasing along the solution to the ODE evolving on the Stiefel manifold until a minimum point is reached. Based on this gradient flow algorithm, an iterative algorithm in closed form has been generated, for which it has been shown that a fixed step size is adequate to ensure that the cost is decreasing to a minimum. All the proposed algorithms are well applicable to the multi-input multi-output case.

References

[1] L. Meier and D. G. Luenberger, "Approximation of linear constant systems," *IEEE Trans. Automat. Contr.*, vol. AC-12, pp. 585–588, 1967.

[2] D. A. Wilson, "Optimum solution of model-reduction problem," *Proc. IEE*, pp. 1161–1165, 1970.

[3] 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. Automat. Contr.*, vol. AC-30, pp. 1201–1211, 1985.

[4] Y. Halevi, "Frequency weighted model reduction via optimal projection," *Proc. IEEE Conf. on Decision and Control*, pp. 2906–2911, 1990.

[5] A. E. Bryson and A. Carrier, "Second-order algorithm for optimal model order reduction," *J. Guidance Control Dynam.*, pp. 887–892, 1990.

[6] L. Baratchart, M. Cardelli, and M. Olivi, "Identification and rational L_2 approximation: a gradient algorithm," *Automatica*, pp. 413–417, 1991.

[7] J. T. Spanos, M. H. Milman, and D. L. Mingori, "A new algorithm for L_2 optimal model reduction," *Automatica*, pp. 897–909, 1992.

[8] L. Baratchart, *R. F. Curtin (Ed.). Modelling, Robustness and Sensitivity Reduction in Control Systems*, ch. Recent and new results in rational L_2 approximation, pp. 119–126. Berlin: Springer-Verlag, 1987.

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

[10] M. Aoki, "Some approximation methods for estimation and control of large scale systems," *IEEE Trans. Automat. Contr.*, pp. 173–182, 1978.

[11] J. Hickin and N. K. Sinha, "Canonical forms for aggregated models," *Int. J. Contr.*, pp. 473–485, 1978.

[12] J. Lam and Y. S. Hung, "On the stability of projections of balanced realizations." To appear.

[13] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*. London: Springer-Verlag, 1994.

[14] A. Lepschy, G. A. Mian, G. Pinato, and U. Viaro, "Rational L_2 approximation: a non-gradient algorithm," *Proc. IEEE Conf. on Decision and Control*, pp. 2321–2323, 1991.