Institute for Computational Mathematics
Hong Kong Baptist University

ICM Research Report
10-01

System Reduction with Optimal Replacement Variables

Alex Solomonoff[1*], Wai Sun Don[2]

[1] *Camberville Research Institute, Somerville, Mass., USA*
[2] *Department of Mathematics, Hong Kong Baptist University, Hong Kong, China*

Abstract. In this exploratory study, we present a new method of approximating a large system of ODEs by one with fewer equations, while attempting to preserve the essential dynamics of a reduced set of variables of interest. The method has the following key elements:

1. Put a (simple, ad-hoc) probability distribution on the phase space of the ODE;
2. Assert that a small set of *replacement variables* are to be unknown linear combinations of the not-of-interest variables, and let the variables of the reduced system consist of the variables-of-interest together with the replacement variables;
3. Find the linear combinations that minimize the difference between the dynamics of the original system and the reduced system.

We describe this approach in detail for linear systems of ODEs. Numerical techniques and issues for carrying out the required minimization are presented. Examples of systems of linear ODEs and variable-coefficient linear PDEs are used to demonstrate the method. We show that the resulting approximate reduced system of ODEs gives good approximations to the original system. Finally, many possible directions for further work are outlined.

AMS subject classifications: 65M06, 65M30, 65M70

Key words: System reduction, Optimal Replacement Variables, Resolved Variables, Optimal Prediction

1 Introduction

The framework of the problem studied in this paper is a system of ordinary differential equations (ODEs)

$$z_t = f(z), \quad t > 0, \tag{1.1}$$

*Corresponding author. *Email addresses:* alex.solomonoff@yahoo.com (A. Solomonoff), wsdon@math.hkbu.edu.hk (W.-S. Don)

where $z = (x,y) \in \mathbb{R}^{m+n}$ is divided into a set of *resolved* variables $x \in \mathbb{R}^m$ that one wants to observe or calculate, and a set of *unresolved* variables $y \in \mathbb{R}^n$ that one doesn't need to observe, but which the dynamics of the resolved variables x depend on. Furthermore, it may be that $m \ll n$ or even n infinite, in which case it will not be computationally feasible to solve the full system of equations.

The goal of this study is to approximate (model) the dynamics of x in a computationally efficient way, with useful accuracy, without actually including the full set of unresolved quantities y in the modeling.

This is desirable in many situations: For example, many partial differential equations, when discretized into a system of ODEs, require millions of degrees of freedom to adequately approximate the dynamics. *FIXME: citation?* However, most of these DOF are usually of no interest. Examples of such PDEs include weather simulations and many simulations of fluid or aerodynamic flows. In the flow-around-an-aircraft example, an engineer would be mainly interested in bulk features such as the total lift and drag, or average vorticity as a function of time. A flow field detailed enough to actually resolve all of the dynamics would not be needed in many situations. Calculating solutions to these equations can require quite large amounts of computing resources, and a system reduction method such as the one studied here has the potential to reduce these resource required, or allow the fast solution of more complex problems.

One approach for system reduction has been developed by Chorin et al. [1, 2] and Gottlieb et al. [3], which has been called the *t-system* or *Optimal Prediction*. An overview of other approaches to the problem can be found in [4].

Much previous work, including much work by Chorin and associates, has revolved around approximate systems having only x as the state variables and eliminating any trace of y. We suspect that better approximations will require the influence of some lower-dimensional replacement of the unresolved variables y. These *replacement variables* (RV) do not have to include all the information in y but just carry enough information to model the influence of unresolved y on the resolved x.

Table of Contents

and issues in the ORV method.

5. We outline some directions and questions for future research in section 6.

2 Setup and Framework

In this work, we only consider the case of a linear system of ODEs, but the general method described here can be applied to nonlinear systems. Consider a linear system of ODEs,

$$z_t = Fz = \begin{pmatrix} F_0 & F_1 \\ F_2 & F_3 \end{pmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, t > 0. \tag{2.1}$$

with an appropriate initial conditions $z(0)$.

We shall assume that

1. a probability distribution $\rho(z)$ exists for z, which approximates the dynamics of the system (2.1) in the sense that for any set $S \in \mathbb{R}^{n+m}$, the likelihood of $z \in S$ is $\rho(S)$, and

2. ρ is a reasonable approximation at all times $t > 0$.

The probability distribution $\rho(z)$ is an important part of the framework, and is an input to the process – it is supplied by the user. It has to be selected using domain-specific knowledge and intuition, and will often be ad-hoc and imprecise. Therefore, it is a crucial requirement for the success of the optimal replacement variables method (ORV) that the method has to work well with such an imprecise, inaccurate probability distribution. This approximate or utility distribution is used only for computing the expectation of physical quantities one want to minimize. It is unlikely that samples of $z(t)$ obey this distribution, one only hope that that minimizing expected errors with respect to it, will result in algorithms with good results and/or improved performance.

Mathematically, ρ will be taken to be Gaussian distribution, that is,

$$z \sim \mathcal{N}(\mu, S), \tag{2.2}$$

with

$$\mu = \begin{bmatrix} \mu_0 \\ \mu_1 \end{bmatrix}, \quad S = \begin{pmatrix} S_0 & S_1 \\ S_2 & S_3 \end{pmatrix}. \tag{2.3}$$

It will be assumed to be zero-mean, i.e., $\mu = 0$.

We shall construct a small set of variables

$$u = Ry, \quad R \in \mathbb{R}^{k \times n}, \quad k \ll n, \tag{2.4}$$

such that the dynamics of

$$w = (x, u), \tag{2.5}$$

mimic the dynamics of $z = (x, y)$ as closely as possible.

For the measure of closeness, we pick the error e as

$$e = E\big(\|w_t - E(w_t|w)\|^2\big), \tag{2.6}$$

where the expectations are taken with respect to the Gaussian density (2.2).

The variables u which minimize the closeness metric (2.6) we call *optimal replacement variables* (ORV) and we also use this as the name for the full system reduction method.

There are many possible versions of the closeness metric. Two of them are

$$\begin{aligned} e &= E\big(\|x_t - E(x_t|w)\|^2\big), \\ e &= E\big(\|z_t - E(z_t|w)\|^2\big). \end{aligned}$$

The first metric has the drawback of not requiring the replacement variables to reproduce their own dynamics with any accuracy, and this seems like it would not work – to approximate x_t well, only Ry needs to be accurate, but if the dynamics of y are all wrong, neither y or Ry will be accurate. The second metric goes too far in the other direction – it required that the dynamics of both x and y be preserved, even though most of the information in y is not needed. So the closeness measure (2.6) is a plausible compromise between these two extremes. Whether it is the best measure of closeness or what the best measure of closeness might be, seems to be an open question.

2.1 The Expected Error

We want to compute the best R,

$$R_* = \arg\min_R e(R). \tag{2.7}$$

We need to simplify the error (2.6) down to an expression that can be minimized numerically or analytically.

First, what is $E(z|w)$? Note that

$$w = \begin{pmatrix} I & 0 \\ 0 & R \end{pmatrix} z = Pz. \tag{2.8}$$

In general (see [7])

$$(z|Pz) \sim \mathcal{N}(\mu', S'), \tag{2.9}$$

with

$$\mu' = E(z|Pz) = \mu + SP^T ZP(z-\mu) = (I-KP)\mu + KPz, \tag{2.10}$$

where

$$Z = (PSP^T)^{-1} \qquad \text{and} \qquad K = SP^T Z. \tag{2.11}$$

This gives

$$z - E(z|Pz) = \big(I-KP\big)(z-\mu), \tag{2.12}$$

and the covariance matrix is

$$S' = E(zz^T|Pz) - \mu'\mu'^T = S - SP^TZPS. \tag{2.13}$$

It should be noted that S' does not depend on z, but $E(zz^T|Pz)$ does.

The replacement dynamics w will be

$$w_t = PF\ \mathrm{E}(z|w) \approx PFz, \tag{2.14}$$

and so

$$e \;=\; E\Big(\|PF\big(z - E(z|Pz)\big)\|^2\Big) = E\Big(\|PF\big(I-KP\big)(z-\mu)\|^2\Big). \tag{2.15}$$

Define $A = PF(I-KP)$, one has

$$e \;=\; E\big(\|A(z-\mu)\|^2\big) = \mathrm{tr}\big(PF(S-SP^TZPS)F^TP^T\big). \tag{2.16}$$

Hence, in term of the dynamics of the replacement system, the reduced system of ODEs becomes

$$w_t \;=\; PF\mathrm{E}(z|Pz) = PFKw \;=\; F_*w, \tag{2.17}$$

with $F_* = PFK$, assuming that $\mu = 0$. If not, there is an extra affine term.

3 The Error Equation

This section describes the unfolded and simplified equation for the expected error, and its gradient.

Consider three different parts of $e(R)$,

$$e(R) = e_0 + e_1 - e_2 = e_0 + \mathrm{tr}(PFSF^TP^T) - \mathrm{tr}(PFSP^TZPSF^TP^T) \tag{3.1}$$

where e_0 contains all the terms from e_1 and e_2 that do not depend on R. The trace depends only on the diagonal submatrices of the matrix inside the trace. Define

$$A = FSF^T = \begin{pmatrix} A_0 & A_1 \\ A_1^T & A_3 \end{pmatrix} \qquad \text{and} \qquad B = FS = \begin{pmatrix} B_0 & B_1 \\ B_2 & B_3 \end{pmatrix} \tag{3.2}$$

Then, one has

- $e_0 = \mathrm{tr}(A_0)$.
- $e_1 = \mathrm{tr}(RA_3R^T)$, where $A = FSF^T$.
- $e_2 = \mathrm{tr}(PBP^TZPB^TP^T)$.

By letting

$$H = BP^T ZPB = \begin{pmatrix} H_0 & H_1 \\ H_1^T & H_3 \end{pmatrix}, \tag{3.3}$$

one has

$$e_2 = \mathrm{tr}(H_0) + \mathrm{tr}(RH_3R^T), \tag{3.4}$$

where

$$\begin{aligned} H_0 &= B_0 Z_0 B_0^T + B_1 R^T Z_1^T B_0^T + B_0 Z_1 R B_1^T + B_1 R^T Z_3 R B_1^T, \\ H_3 &= B_2 Z_0 B_2^T + B_3 R^T Z_1^T B_2^T + B_2 Z_1 R B_3^T + B_3 R^T Z_3 R B_3^T. \end{aligned} \tag{3.5}$$

3.0.1 Unfolding Z

In general, the inverse of a 2×2-partitioned symmetric matrix [8]

$$V = \begin{pmatrix} V_0 & V_1 \\ V_1^T & V_3 \end{pmatrix}, \tag{3.6}$$

is

$$V^{-1} = \begin{pmatrix} C_0 + C_1 C_3^{-1} C_1^T & -C_1 C_3^{-1} \\ -C_3^{-1} C_1^T & C_3^{-1} \end{pmatrix}, \tag{3.7}$$

where the "half-inverse" matrix C is defined as

$$C = C(V) = \begin{pmatrix} V_0^{-1} & V_0^{-1} V_1 \\ V_1^T V_0^{-1} & V_3 - V_1^T V_0^{-1} V_1 \end{pmatrix}, \tag{3.8}$$

assuming that all the required submatrix inverses exist.

Letting $C = C(S)$ and $Q = (RC_3R^T)^{-1}$, we have

$$Z = \begin{pmatrix} C_0 + C_1 R^T Q R C_1^T & -C_1 R^T Q \\ -QRC_1^T & Q \end{pmatrix}, \tag{3.9}$$

If one define $X = R^T QR$, and now one can use these expressions for Z to unfold H_0 and H_3:

$$\begin{aligned} H_0 &= B_0 C_0 B_0^T + (B_0 C_1 - B_1) X (C_1^T B_0^T - B_1^T) \\ H_3 &= B_2 C_0 B_2^T + (B_2 C_1 - B_3) X (C_1^T B_2^T - B_3^T) \end{aligned}$$

Since the term $B_0 C_0 B_0^T$ from H_0 is independent on R, the term can be moved into e_0 as

$$e_0 = \mathrm{tr}(A_0) - \mathrm{tr}(B_2 C_0 B_0^T) = \mathrm{tr}(F_1 C_3 F_1^T). \tag{3.10}$$

By defining

$$\begin{aligned} D_0 &= B_2 C_0 B_2^T, \\ D_1 &= B_0 C_1 - B_1 = -F_1 C_3, \\ D_2 &= B_2 C_1 - B_3 = -F_3 C_3 \end{aligned} \tag{3.11}$$

the second error term e_2 can be expressed as

$$e_2 = \mathrm{tr}(R D_0 R^T - D_1 X D_1^T - R D_2 X D_2^T R^T).$$

Taking the liberty of assuming that $Q = I$ (which will be discussed/justified later), one has

$$e_2 = \mathrm{tr}\big(R(D_0 - D_1^T D_1) R^T - R D_2 X D_2^T R^T\big),$$

and

$$E = A_3 + D_0 - D_1^T D_1 = F_3 C_3 F_3^T - C_3 F_1^T F_1 C_3.$$

Note that E is symmetric.

The error function $e(R)$ becomes

$$e(R) = e_0 + \mathrm{tr}(R E R_t) - \mathrm{tr}(R D_2 X D_2^T R^T).$$

It is important to note that if U is any $k \times k$ orthogonal matrix, then $e(R) = e(UR)$. That is, the rows can be permuted, for example, without changing the error function which will be significant later.

The Gradient

When minimizing $e(R)$, the gradient of e will usually be needed. So let

$$R' = \frac{\partial R}{\partial R_{ij}} = e_i e_j^T,$$

for an element (i,j) of R.

Differentiating (3.12), one has

$$\begin{aligned} e' &= \mathrm{tr}(R' E R^T + R E R'^T - R' D_2 R^T R D_2^T R^T - R D_2 R'^T R D^T R^T \\ &\qquad - R D_2 R^T R' D_2^T R^T - R D_2 R^T R D_2 R'^T) \\ &= 2\mathrm{tr}\Big(R'\big(E R^T - D_2 R^T R D_2^T R^T - D_2^T R^T R D_2 R^T\big)\Big) \\ &= 2\big(E R^T - D_2 R^T R D_2^T R^T - D_2^T R^T R D_2 R^T\big). \end{aligned}$$

since for any matrix G, $\mathrm{tr}(R' G^t) = G$.

3.1 Expected Errors and Norms

The probabilistic framework can be used to calculate expected values of many quantities, such as:

- $E\|z\|^2 = \mathrm{tr}(S)$.
- $E\|Fz\|^2 = E\|z_t\|^2 = \mathrm{tr}(FSF^T) = \mathrm{tr}(A_0) + \mathrm{tr}(A_3)$.
- $E\|x\|^2 = \mathrm{tr}(S_0)$.
- The error of the Galerkin approximation:

$$E\|(F_0x + F_1y) - F_0x)\|^2 = E\|F_1y\|^2 = \mathrm{tr}(F_1S_3F_1^T).$$

- Error of replacing x_t by $E(x_t|x)$ is

$$E\|x_t - E(x_t|x)\|^2 = \mathrm{tr}(F_1C_3F_1^T).$$

- $E\|Pz_t\|^2 = \mathrm{tr}(A_0) + \mathrm{tr}(RA_3R^T)$.
- The x-part of the ORV error:

$$E\|x_t - E(x_t|w)\|^2 = \mathrm{tr}(F_1C_3F_1^T) - \mathrm{tr}(RD_1^TD_1R^T).$$

The last two properties in the list above depend on R.

These expressions are useful for normalizing the ORV error. It is more meaningful to compare the error of an ORV scheme to, for example the Galerkin approximation error or the expected value of w_t, than to simply look at the size of the numerical error obtained from the ORV method. This is particularly important when looking at the ORV error as parameters change.

This normalization is trickier than it might first look. The ORV-ed system will have errors in both x and Ry, so it is difficult to compare it with, say $E\|x_t\|$, which only involves x. This is especially a problem with looking at the behavior of the ORV system as the number of replacement variables changes. It is also difficult to compare the ORV error with $E\|z_t\|$ since the unresolved variables might have substantial dynamics that the ORV system does not (and does not need to) resolve.

Some plausible ways of normalizing the ORV error are

$$\frac{E\|x_t - E(x_t|w)\|}{E\|x_t\|}, \tag{3.12}$$

and

$$\frac{E\|w_t - E(w_t|w)\|}{E\|Pz_t\|}. \tag{3.13}$$

The first compares x errors only, which is not what ORV is minimizing although it is what we actually want to be small. The second compares w error from ORV with a quantity that depends on R in some random way, and which ORV has not tried to maximize. Both of these measures of ORV error are slightly problematic. But, we have not found a better way to normalized errors.

3.2 Orthogonality Constraints

The dynamics of the replacement-variables system clearly depend only on the subspace spanned by R, not on R itself. Unfortunately, our error criterion $e(R)$ does depend on R. For example if R is multiplied by 2, then the u-part of the ORV error is also multiplied by 2. To deal with this in the minimization R needs to be normalized or constrained. An obvious way to do this is to require that

$$RR^T = I, \tag{3.14}$$

or alternatively, require

$$RGR^T = I, \tag{3.15}$$

for some symmetric positive definite matrix G.

If we allow ourselves to use a $G \neq I$, we might pick a convenient G. For example, the error term e_2 has R inside a matrix inverse, as well as terms of high degree in R. This greatly complicates the algebra, and will make any algebraic manipulation of $e(R)$ more difficult. This can be bypassed by changing the orthogonality condition from $RR^T = I$ to

$$RC_3R^T = I. \tag{3.16}$$

This gives $Q = I$ and $X = R^TR$. We will use (3.16) for the development of ORV method in the following discussion.

4 Minimizing the Expected Error

In this section, we will discuss numerical methods for minimizing the expected error of the ORV method.

4.1 Lagrange Multipliers

The orthogonality condition (3.16) can be enforced using Lagrange multipliers with a constraint expression

$$l = \mathrm{tr}\big(L(RC_3R^T - I)\big), \tag{4.1}$$

where L is an $k \times k$ matrix.

Combining the error from (3.12) and the Lagrange multiplier term yields

$$e_l(R,L) = e_0 + \mathrm{tr}\big(RER^T - RD_2R^TRD_2^TR^T\big) + \mathrm{tr}\big(L(RC_3R^T - I)\big), \tag{4.2}$$

and the minimizer of $e(R)$ will satisfy the combined equation

$$0 = \nabla_R e_l = ER^T + DR^T RD^T R^T + D^T R^T RDR^T + C_3 R^T L. \tag{4.3}$$

4.2 Minimization Using the Self-Consistent Field Iteration

How can equation (4.3) be solved? It looks somewhat like a symmetric eigenvalue problem – Specifically, if

$$F(R) = E + DRR^T D^T + D^T RR^T D,$$

then (4.3) is

$$F(R)R^T - C_3 R^T L = 0,$$

which, except for the matrix depending on R, is a *symmetric eigenvalue* problem. This suggests the iteration

$$F(R_i)R_{i+1}^T = C_3 R_{i+1} L_{i+1},$$

which, if it converges, will converge to a minimizer of (4.2).

It turns out that minimization problems similar to (4.3) occur in electronic structure calculations, see [5]. The iteration (4.4) is used in that community, where it goes by the name of the *self-consistent field* algorithm (SCF) . They find that it often either converges slowly or fails altogether to converge. Enhancements have been developed to make it more efficient and reliable, but minimization of this kind of expression remains an active area of research. It is used in the electronic structure community in spite of its poor performance because the problems are extremely high-dimensional and using methods involving higher derivatives is impractical.

We have tried the iteration (4.4) and like the electronic structure community, found that it often converges slowly or not at all. However it worked well enough to carry out a few experiments, which did give encouraging results.

4.3 Unconstrained Minimization

The SCF iteration automatically handles the orthogonality constraints. A general minimization scheme such as Conjugate Gradient does not enforce constraints, although they could be introduced through the Lagrange multipliers.

Another way of dealing with the constraints is to create a new, lower-dimensional variable that incorporates the constraints automatically. This can be done in the following way: Let

$$R = \begin{pmatrix} R_1 & R_2 \end{pmatrix} = P^{-1} \begin{pmatrix} I & Y \end{pmatrix},$$

where the I is a $k \times k$ identity matrix. Then the new variable Y is

$$Y = Y(R) = R_1^{-1} R_2.$$

R can be reconstructed from Y using the orthogonality constraint as follows:

$$RC_3R^T = I = P^{-1} \begin{pmatrix} I & Y \end{pmatrix} C_3 \begin{bmatrix} I \\ Y^T \end{bmatrix} P^{-1},$$
$$M(Y) = PP^T = \begin{pmatrix} I & Y \end{pmatrix} C_3 \begin{bmatrix} I \\ Y^T \end{bmatrix} P^{-1},$$

and

$$P = \text{chol}\big(M(Y)\big),$$

where $\text{chol}(M)$ is any matrix P satisfying $PP^T = M$. Hence,

$$R = R(Y) = \text{chol}\big(M(Y)\big)^{-1} \begin{pmatrix} I & Y \end{pmatrix}, \tag{4.4}$$

and

$$X = R^T R = \begin{bmatrix} I \\ Y^T \end{bmatrix} M^{-1} \begin{pmatrix} I & Y \end{pmatrix}. \tag{4.5}$$

Note that the $R\big(Y(R)\big)$ will generally differ from R by an orthogonal transformation. This is desirable – the orthogonal transformation does not change the objective function.

How does this use of Y affect the gradient of $e(R)$? First, it can be shown that

$$\frac{\partial X}{\partial Y} = \text{symm}\Big(A(Y)Y'B(Y)\Big), \tag{4.6}$$

where

$$A(Y) = \begin{bmatrix} I \\ Y^T \end{bmatrix} Q^{-1} \qquad \text{and} \qquad B(Y) = \begin{pmatrix} 0_{k\times(n-n)} & I_{n-k} \end{pmatrix} \big(I - C_3 X\big). \tag{4.7}$$

Then it can be shown that

$$\frac{\partial e}{\partial Y} = 2B(Y)\Big(E - D_2 X D_2^T - D_2^T X D_2\Big) A(Y). \tag{4.8}$$

The function $e(R(Y))$ and equation (4.8) can be used in a standard (unconstrained) minimization scheme such as conjugate gradient. *FIXME: reference?*

4.3.1 Pivoting

The matrix $M(Y)$ is not guaranteed to be well-conditioned. In experiments, it appears that even a small amount of ill-conditioning of $M(Y)$ can substantially slow down the convergence of conjugate gradient minimization.

How can the $R \leftrightarrow Y$ transformation be modified to improve its conditioning? Suppose

$$R = R(Y) = P^{-1} \begin{pmatrix} I & Y \end{pmatrix} Q, \tag{4.9}$$

where

$$Q = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} \qquad \text{and} \qquad W = Q^{-1} = \begin{pmatrix} W_1 & W_2 \end{pmatrix}, \tag{4.10}$$

are constant matrices to be described below. Then by defining

$$RQ^{-1} = RW = \begin{pmatrix} RW_1 & RW_2 \end{pmatrix} = (RW_1) \begin{pmatrix} I & (RW_1)^{-1}(RW_2) \end{pmatrix},$$

one has

$$Y = (RW_1)^{-1}(RW_2). \tag{4.11}$$

The inverse transformation is derived as follows:

$$I = RC_3R^T = P^{-1} \begin{pmatrix} I & Y \end{pmatrix} QCQ^T \begin{bmatrix} I \\ Y^T \end{bmatrix} P^{-T},$$

and

$$PP^T = M(Y) = \begin{pmatrix} I & Y \end{pmatrix} QCQ^T \begin{bmatrix} I \\ Y^T \end{bmatrix},$$

and

$$R(Y) = \text{chol}(M(Y))^{-1} \begin{pmatrix} I & Y \end{pmatrix} Q. \tag{4.12}$$

As before, $\text{chol}(M)$ is any matrix P satisfying $PP^T = M$.

Now matrices Q and W must be picked so that for a given R_*, $\text{cond}(R_* W_1) = 1$. So

$$W_1 = R \div U \tag{4.13}$$

for some orthogonal matrix U. This is the only requirement. U and W_2 can be chosen for convenience.

CG with Pivoting

Pivoting would be added to a conjugate gradient minimization scheme as follows:

1. Choose an initial guess R_0.
2. Compute pivoting matrices W and Q based on R_0.

3. Do several/many iterations of CG minimization, giving an intermediate solution R_1.
4. compute new pivoting matrices W and Q based on R_1
5. if not yet converged, go to 3

In experiments, this substantially improved the convergence of the conjugate gradient minimization, and also seemed to improve the accuracy of finding the minimum.

5 Numerical Results

In this section, we will present some preliminary results from applying ORV method to two test problems that illustrate the potential and issues in the implementation of the ORV method. All calculations present below were done using version 3.0.1 of Octave, a free Matlab-like program, [9].

Random system of ODEs

The first test equation is the linear system of ODEs

$$z_t = Fz, \qquad z = (x,y). \tag{5.1}$$

with a random matrix F. Its entries were IID 0-1 Gaussian random numbers between zero and one. It was stabilized by computing its eigenvalues and adding the multiple of the identity that just makes the matrix stable.

A covariance matrix S was computed by cheating [†] – Eq. (5.1) was integrated in time, starting with an IID 0-1 random vector, and regularly-spaced-in-time samples used to compute an observed covariance matrix $S = \frac{1}{n}\sum_i z_i z_i^T$. This was found to be ill-conditioned, and was therefore backed off: The diagonal of S was first backed off to a multiple of the identity, then the full matrix backed off to the diagonal:

$$S \leftarrow (1-\epsilon)S + \epsilon\big((1-\epsilon)\mathrm{diag}(S) + \epsilon\frac{1}{n}\mathrm{tr}(S)I\big).$$

Scalar Variable-Coefficient PDEs

The second test problem is cosine expansion of the scalar variable-coefficient heat equation

$$u_t = \big(2+\cos(x)\big)u_{xx} \qquad \text{with} \qquad u(x,t) = \sum_{k=0}^{\infty} u_k(t)\cos(kx). \tag{5.2}$$

This gives the tridiagonal system of ODEs for each coefficient $u_k(t)$

$$\frac{d}{dt}u_k = -\frac{1}{2}(k-1)^2 u_{k-1} - 2k^2 u_k - \frac{1}{2}(k+1)^2 u_{k+1}. \tag{5.3}$$

[†] A wise person once said "If you can't win by cheating, then you can't win, so if you cheat and still lose, you know to give up."

This is an infinite matrix, which is truncated at a fixed number $m+n$ of equations.

For the covariance matrix we used an exponentially-decaying diagonal matrix

$$S_{ij} = \alpha^i \delta_{ij} \tag{5.4}$$

where the parameter $\alpha \in [0,1)$ is to be chosen to suit the experimenter's intuition. This is quite an ad-hoc choice, which is appropriate because often (usually?) a more correct covariance matrix will not be known.

5.1 Performance Results

It is not particularly meaningful to solve the ODEs $w_t = F_* w$, and $z_t = Fz$, look at the difference between the two solutions and say "look, they're small!" Small compared to what? To meaningfully evaluate the ORV system, its performance needs to be compared to other approximations of the original ODE.

Some reasonable approximations for comparison are Galerkin approximation, Random-R approximation, $m+k$ approximation, $\mathrm{E}(x_t|x)$ approximation.

- The Galerkin approximation is
$$x_t = F_0 x. \tag{5.5}$$
where F_0 is as defined in Eq. (2.1).

- Random-R approximation means picking a replacement matrix R at random and constructing a replacement-variable system using that R.

- The Expected-x_t approximation is
$$x_t = \mathrm{E}(x_t|x) = (F_0 + F_1 S_1^T S_0^{-1})x. \tag{5.6}$$
This uses the probability distribution ρ to improve the accuracy of the approximate dynamics over Galerkin.

- The $m+k$ approximation is for cases like the heat equation where there is an obvious reasonable way of defining the variables and an obvious ordering of them. In the case of a PDE with slowly-varying coefficients (such as our heat equation) the cosine modes of the solution are such variables. In this case a reasonable choice is to simply pick the k next modes in the sequence and let those be the additional degrees-of-freedom in the approximate dynamics.

5.2 Stability of the ORV System

The matrix F of the original system of ODEs will usually be a stable matrix. If the ORV matrix F_* is not also stable, then the ORV scheme will be of limited use. Figure 1 tests this in the random-matrix test case. The values at the extreme right of the figure are for the

original matrix F. The top curve is the spectral radius of F_*; $\rho(F_*)$, which is related to the degree of stiffness in F_*. Note that $\rho(F_*) < \rho(F)$ for all k, meaning that the ORV matrices are never any stiffer than the original matrix. The bottom curve is $\log(\rho(\exp(F_*)))$, which should be ≤ 0 for a stable matrix. It can be seen that F_* is stable for most values of k, but for $k = 5,6$ it is not quite stable. This tiny degree of instability would probably not have any negative effects unless time integration was carried out to quite large t.

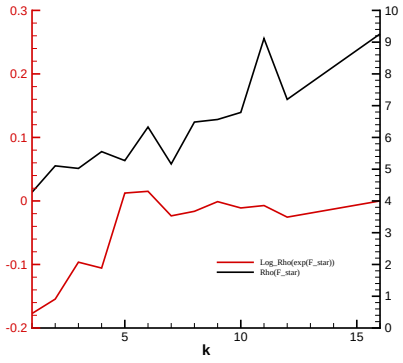

Figure 1: Spectral characteristics of F_* vs. k, $m = 4$, $n = 16$.

5.3 Accuracy of the ORV scheme

Figure 2 compares the accuracy of the ORV scheme for the random matrix problem to two obvious competitors – Galerkin and picking R at random. ORV scheme is clearly superior to both of these.

Figure 3 shows the error of the ORV scheme for the random matrix problem, compared with the two different normalizations of the the expected error from equations (3.12) and (3.13). As k increases, the actual error decreases in a slow but exponential way. We also see that the numerical error and the two expected errors are all roughly similar.

Figure 4 shows the accuracy of the ORV version of the heat equation for several different values of k. The ORV system is compared the Galerkin approximation and also to the $m+k$ approximation.

The heat equation is a case where a reasonable set of additional variables exists. This makes the $m+k$ scheme a strong competitor to the ORV scheme. In more general cases obvious replacement variables would not exist. The figure also shows that as k increases, the accuracy of the ORV scheme increases rapidly. The ORV scheme is, in fact, about the same accuracy as the $k+m$ competing scheme. It is better in one case, $k = 6$. The ORV scheme is far better than the Galerkin scheme (no additional variables).

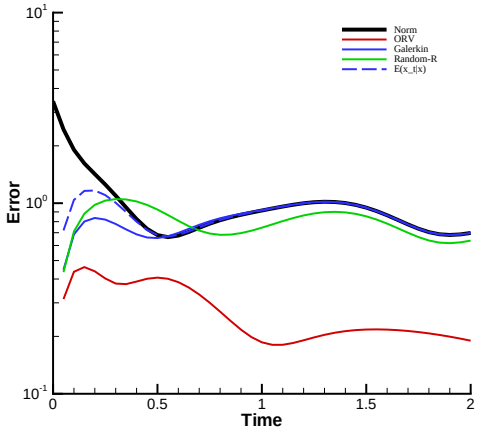

Figure 2: Errors in time for the random matrix problem with the ORV, Galerkin and random-R methods. $(m,n,k) = (8,40,12$.

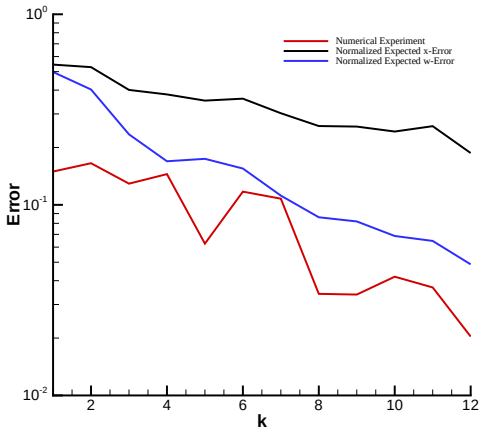

Figure 3: Expected and actual error vs. k for F_*.

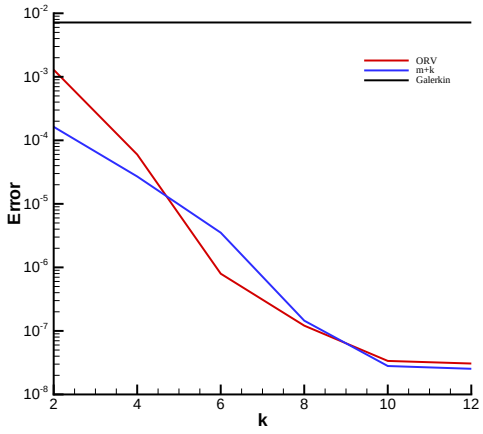

Figure 4: Accuracy of the ORV scheme versus number of replacement variables for the heat equation.

This can be viewed both as encouraging and discouraging. If we assume that the $k+m$ scheme is close to optimal, then this experiment shows that the ORV scheme has done as well as it is possible for it to do. But the discouraging viewpoint is that ORV was not superior to an obvious and simple competitor. We prefer the encouraging viewpoint.

5.4 Multiple Starts and Multiple Non-global Minima

The conjugate gradient minimization (CG) finds a local minimum, and if the function being minimized has multiple local minima, CG is not guaranteed to find the best one. Does our ORV error function have multiple minima? (Answer: sometimes) And if so, what should be done about it? (Answer: That's an open question)

The traditional, simple way of finding a global minimum amid (possibly) many local minima is with multiple random starts – make a random initial guess at the solution, use your favorite local minimizer starting there, save the result. Repeat this several (many?) times using different random initial guesses. If the solutions are all the same, you probably have a single global minimum. If some or all of them are different, you have multiple minima, and you use the best one.

We did this experiment on our ORV error function. We computed local minima of the ORV error function 200 times, counted the different solutions and looked at some statistics of each solution. The objective function value was normalized against the expected-x_t error of equation (3.12). The results are shown in tables 1, 2 and 3.

First, the random ODE problem has only a few solutions, and the best one attracts almost the entire state space. One or two random starts would be enough to get the global minimum in this situation.

num	obj func	counts	cond(hess)	$\|RDR^t\|_F$	$\|R\|_f$	nn-dist
1	1.141×10^{-2}	4	1.403×10^{2}	1.211×10^{2}	3.93355	10.17232
2	5.417×10^{-2}	16	1.392×10^{2}	1.419×10^{2}	4.12778	16.10386
3	5.493×10^{-2}	4	1.441×10^{2}	1.676×10^{2}	4.33858	8.17578
4	5.493×10^{-2}	9	1.505×10^{2}	1.980×10^{2}	4.58933	8.06855
5	5.493×10^{-2}	10	1.618×10^{2}	2.329×10^{2}	4.88402	11.62701
6	5.493×10^{-2}	5	1.770×10^{2}	2.719×10^{2}	5.22763	8.06377
7	5.493×10^{-2}	5	1.964×10^{2}	3.150×10^{2}	5.62522	8.06386
8	5.493×10^{-2}	1	3.448×10^{2}	4.131×10^{2}	6.59003	46.52897
9	0.20361	7	1.420×10^{2}	1.540×10^{2}	4.26404	8.11335
10	0.20376	6	1.428×10^{2}	1.780×10^{2}	4.48230	24.97835
18	0.20751	8	1.618×10^{2}	2.516×10^{2}	5.15148	17.28774
26	0.20761	2	1.767×10^{2}	3.195×10^{2}	5.83747	16.09629
34	0.52970	3	1.771×10^{2}	2.823×10^{2}	5.42198	6.31713
42	0.53165	2	1.771×10^{2}	3.052×10^{2}	5.72282	6.42484
50	0.55186	1	2.888×10^{2}	3.377×10^{2}	6.02601	8.34724
58	0.87336	1	1.557×10^{2}	2.476×10^{2}	5.15070	7.70220
66	0.99997	1	6.007×10^{2}	5.019×10^{2}	8.09102	37.74330

Table 1: $u_t = \big(2+\cos(x)\big)u_{xx}$,$\alpha_{init} = 0.9$,$\alpha_{prob} = 0.8$, 20 unresolved variables, 4 resolved, 4 replacement. Only some solutions are shown.

obj func	counts	cond(hess)	$\|RDR^t\|_F$	$\|R\|_f$	nn-dist
6.518×10^{-3}	37	1.337×10^{2}	1.211×10^{2}	10.00759	98.76203
3.310×10^{-2}	48	1.501×10^{2}	1.419×10^{2}	11.00082	1.692×10^{2}
3.634×10^{-2}	28	1.722×10^{2}	1.676×10^{2}	12.80773	59.50623
3.648×10^{-2}	5	2.160×10^{2}	1.980×10^{2}	15.25168	2.653×10^{2}
3.648×10^{-2}	3	4.758×10^{2}	2.329×10^{2}	18.46988	1.547×10^{2}
3.648×10^{-2}	1	1.364×10^{3}	2.719×10^{2}	22.62486	1.537×10^{2}
0.14753	21	1.722×10^{2}	1.780×10^{2}	13.36889	1.000×10^{11}
0.14753	12	1.504×10^{2}	1.541×10^{2}	12.26345	81.63501
0.14754	15	3.259×10^{2}	2.069×10^{2}	15.72647	45.25228
0.14755	5	8.935×10^{2}	2.404×10^{2}	18.86379	59.92150
0.14755	1	5.561×10^{3}	3.206×10^{2}	28.17489	1.033×10^{3}
0.16044	3	1.729×10^{2}	1.930×10^{2}	15.49814	35.34451
0.16045	4	3.860×10^{2}	2.197×10^{2}	16.55432	35.44029
0.16045	1	2.816×10^{3}	2.881×10^{2}	23.52398	1.051×10^{2}
0.16104	1	1.094×10^{3}	2.669×10^{2}	20.63488	4.268×10^{2}
0.16104	1	2.873×10^{3}	3.016×10^{2}	24.42192	6.444×10^{2}
0.45091	7	3.128×10^{2}	1.841×10^{2}	13.92223	44.82425
0.45092	1	7.447×10^{2}	2.122×10^{2}	16.19956	69.51190
0.45105	1	3.203×10^{2}	1.986×10^{2}	15.80370	1.596×10^{2}
0.45105	3	7.451×10^{2}	2.247×10^{2}	16.84083	96.04916
0.45105	2	1.896×10^{3}	2.559×10^{2}	19.80403	94.54004

Table 2: $u_t = \big(2+\cos(x)\big)u_{xx}$,$\alpha_{init} = 0.9$,$\alpha_{prob} = 0.6$, 20 unresolved variables, 4 resolved, 4 replacement.

obj func	counts	cond(hess)	$\|RDR^t\|_F$	$\|R\|_f$	nn-dist
0.55702	182	1.064×10^{2}	6.35782	3.76880	1.000×10^{11}
0.55753	16	6.342×10^{2}	9.01687	4.81273	20.78350
1.56383	1	-0.93490	6.42993	3.62001	10.69016
1.74289	1	-1.79985	4.94075	3.46182	9.64476

Table 3: $u_t = Fx$,F random, 4 resolved variables, 16 unresolved, 4 replacement.

The situation is completely different in the heat equation. Depending on parameter values, there can be a dozen or a hundred local minima, and the global minimum does not attract a very big section of the phase space. In this case, many random starts would be required to find the global minimum.

In the heat equation case, we observe that

1. The best and good solutions occur at least somewhat more often than the bad solutions.

2. The best and good solutions are somewhat correlated with small values of the condition number of the Hessian matrix at the minimum, the norm $\|R\|_f$ and the quantity $\|RD_2R^T\|_f$. We don't know of any way of taking advantage of these facts, though, or why they are so.

3. There are many sets of several (completely different) solutions with the same objective function value. We also do not know why this is so.

6 Discussion, Conclusions and Further Work

This paper presents a method of system reduction, and shows on a couple of test cases that it is able to reduce the dimension of the system and still reproduce the dynamics of the original system with a useful degree of accuracy.

The work of this paper is preliminary and can be extended in many directions. These fall into several categories:

6.1 Orthogonality of R and Variations on the Error Function

Does it matter to the performance of the ORV system what kind of orthogonality is used? If $k=1$ it makes no difference. Conjecture: if k is small, it makes only a small difference.

What about $RC_3R^T=\alpha I$? Then how would α be chosen?

Another direction for study is a matrix exponential version of the closeness metric.

6.2 Algorithms for Minimizing the Error Function

Is $RC_3R^T=I$ really necessary? Can an effective CG minimization use some other constraint? It simplifies the algebra of the gradient greatly, but is it essential? How is speed of CG convergence affected by choice of orthogonality?

How can multiple local minima be dealt with? Is there a way to recognize and avoid them? Can the objective function be reformulated to have only a few, or only one minimum?

What about other minimization procedures, such as the Grassman CG algorithm proposed in [6]. Does this have any effect on the multiple local minimum issue?

6.3 Optimizing Side Conditions, such as Sparseness of R and of F_*

The R that ORV computes is a full matrix. If the original F was very sparse, then F_* might have just as many non-zeroes as the original F did, in spite of being a smaller system of ODEs. How can be arrange for R or F_* to be a sparse matrix? (while still getting good system accuracy)

Another potentially-important question is how can one do the constrained minimization of requiring that F_* be stable or not too stiff, while also minimizing the error of the dynamics?

6.4 Scaling ORV up to Large Systems of ODEs

In this paper we have only considered systems of ODEs with small-to-moderate numbers of equations. To be most useful it needs to be usable on large systems. Does the ORV system give good performance in such cases? Is it computationally feasible to compute R? Are there issues that occur in large systems that do not occur in small ones?

6.5 Applying ORV to Nonlinear Systems

The ORV general framework can be applied to nonlinear ODEs without change — replace $z=(x,y)$ by $w=(x,Ry)$ and minimize $E\Big(\|w_t-E(w_t|w)\|^2\Big)$ over R, but now $z_t=f(z)$ and $f(z)$ is a nonlinear (perhaps quadratic) function. A Gaussian constant-in-time density could still be assumed for z, but now higher moments of the density (3rd and 4th in the quadratic case) will appear in the expression for the expected error.

An easier version of the nonlinear problem is where only the dynamics of x are nonlinear:

$$\begin{aligned} x_t &= f_0(x)+F_1 y \\ y_t &= f_2(x)+F_3 y. \end{aligned} \tag{6.1}$$

In this case the linear ORV program appears to go though almost unchanged.

6.6 Extending the ORV Framework

1. Non-Gaussian probability distributions, such as mixtures of Gaussian. In many cases, a Gaussian probability distribution would depend on a few parameters, such as the decay α in our heat equation example. These are generally known at best approximately, and one idea is to let the probability distribution be a mixture of Gaussian with different parameters.

2. Nonlinear replacement variables. This might be done by augmenting the set of unresolved variables with a set of nonlinear functions of them, such as

$$r(y)=\{y_i y_j\}_{i,j=1,\ldots,n} \tag{6.2}$$

and then computing replacement variables of the form $u=Ry+Nr(y)$. Since the size of $r(y)$ grows rapidly with n, we might want to boil down the unresolved variables first with a 2-stage ORV process. That is, first compute $v=Ry$, and then compute $u=Nr(v)$ or $u=[v,Nr(v)]$.

6.7 Miscellaneous Linear Algebra

There are many questions that should be studied as well, such as

1. Do the ORV equations change or simplify in any interesting or useful way if F or S are diagonal or block-diagonal?
2. If they commute?
3. If a generalized singular value decomposition of F and S together is considered?
4. What if F or S is close to block-diagonal?

Acknowledgments

The authors dedicate this article to Prof. David Gottlieb. The area of this work is one that David felt was interesting and important both theoretically and practically, before his passing.

The first author (Solomonoff) acknowledges the support of this work under the FRG grant FRG08-09-II-12 and appreciates the hospitality of the Department of Mathematics at Hong Kong Baptist University during his visit. The second author (Don) would like to thank the support provided by the FRG grant FRG08-09-II-12 from Hong Kong Baptist University and funding from Hong Kong Research Grants Council.

A Properties of Trace (tr) and (symm()

These are some identities that play an important role in the preceding results.

- $\mathrm{tr}(AA^T)=\|A\|_F^2=\sum_{i,j}A_{ij}^2$.
- $\mathrm{tr}(AB)=\mathrm{tr}(BA)$.
- $\mathrm{tr}(A)=\mathrm{tr}(A^T)$.
- $\mathrm{tr}\left(\frac{dA}{dA_{ij}}B^T\right)=B_{ij}$.
- $\mathrm{symm}(A)=A+A^T$. (definition).

- $\mathrm{symm}(\alpha A+\beta B)=\alpha\,\mathrm{symm}(A)+\beta\,\mathrm{symm}(B)$.
- $\mathrm{tr}(A\,\mathrm{symm}(B))=\mathrm{tr}(\mathrm{symm}(A)B)$.
- $A\,\mathrm{symm}(B)A^T=\mathrm{symm}(ABA^T)$.
- $\mathrm{symm}(A)=\mathrm{symm}(A^T)$.

References

[1] A. Chorin and P. Stinis, *Problem Reduction, Renormalization, and Memory*, Comm. App. Math. and Comp. Sci. **1** No.1, 2005

[2] A. J. Chorin and O.H. Hald, *Stochastic Tools in Mathematics and Science*, Springer 2006

[3] A. Chertock, D. Gottlieb and A. Solomonoff, *Modified Optimal Prediction and its Application to a Particle-Method Problem*, J. Sci. Computing **37**, pp. 189–201, 2008

[4] D. Givon, R. Kupferman and A. Stuart, *Extracting Macroscopic Dynamics: Model Problems and Algorithms*, Nonlinearity **17** No. 6, R55-R127, 2004. MR 2097022

[5] C. Yang, J. C. Meza and L.-W. Wang, *A Trust Region Direct Constrained Minimization Algorithm for the Kohn-Sham Equation*, SIAM J. Sci. Comput. **29** No. 5, pp. 1854–1875, 2007

[6] A. Edelman, T.A. Arias and S. T. Smith, *The Geometry of Algorithms with Orthogonality Constraints*, SIAM J. Matrix Anal. Appl. **20** No. 2, pp. 303–353.

[7] S. R. Searle, *Linear Models*, Wiley and Sons, 1997

[8] T. Fortmann, *A Matrix Inversion Identity*, IEEE Trans. Automatic Control **15** No. 5, 1970

[9] John W. Eaton, *GNU Octave Manual*, Network Theory Limited, 2002, isbn=0-9541617-2-6, also `www.octave.org`.