Square-Root Free Orthogonalization Algorithms

Mohammed A. Hasan
Department of Electrical & Computer Engineering
University of Minnesota Duluth
E.mail:mhasan@d.umn.edu

Abstract

This paper is concerned with the derivation and analysis of higher order algorithms of polynomial type for computing an orthonormal basis of a subspace. These algorithms are derived from unconstrained optimization of certain cost functions. The proposed methods are efficient and do not require square root computation. Based on these, algorithms for orthonormalization with respect to a positive definite matrix and principal and minor subspace methods are developed. Numerical experiments illustrate the theoretical results.

Keywords: Orthonormalization, Gram-Schmidt Process, higher order fixed point polynomials, global stability

1 Introduction

Orthonormalization has many applications in scientific computation. It is a useful stabilizing techniques in many numerical methods such as the subspace power method. In independent component analysis, it is used as prewhitening technique. Generally any matrix X with linearly independent columns can be decomposed as $X = QR$, where Q has orthonormal column vectors and where R is an upper triangular square matrix.

Consider the matrix $X = [\,x_1 \quad x_2 \quad \cdots \quad x_p\,]\,\mathbb{R}^{n\times p}$. If X is a full rank matrix, then one can verify that the matrix $Y = X(X^TX)^{\frac{-1}{2}}$ is orthogonal, i.e., $Y^TY = I$ or equivalently, for each two different columns y_i, and y_j of the matrix Y: $y_i^T y_j = \delta_{ij}$, where δ_{ij} is the Kronecker delta function. Clearly, X and Y span the same subspace, however $||Y|| = 1$ regardless of the norm of X. Here $||Y||$ denotes the Euclidean norm of Y. Note that if X is not full rank, then $Y = X\{(X^TX)^+\}^{\frac{1}{2}}$ satisfies the relation $(Y^TY)^2 = Y^TY$, i.e., Y^TY is a projection. Here $(.)^+$ denotes the Moore-Penrose generalized inverse of (.). Typically, orthogonalization is accomplished by means of some variant of the Gram-Schmidt process. In this work, a new type of algorithms is proposed, which involves fast implementation of square root methods without performing any square root operation.

The Gram-Schmidt process is related to the QR-factorization [1] which is used in solving linear equations. The QR factorization allows fast computation of the determinant, and least square solutions. One of the main motivation of this work is to develop dynamical systems for computing Y without the costly computation of matrix square root. Another motivation of using orthonormalization is that it is a stabilizing tool in many numerical methods [2]. For example it is used in the context of Krylov-type eigenvalue solvers such as Arnoldi and related methods [3].

In this work, new higher order orthogonalization algorithms have been proposed. These algorithms are based on unconstrained optimization of certain cost functions, where a class of methods are derived from polynomial representation of analytic functions having zeros at 1 and -1.

The following notation will be used throughout. The symbols $\mathbb{R}$, denotes the set of real numbers. The derivative of x with respect to time is written as x'. The identity matrix of appropriate dimension is expressed with the symbol I. Finally, the derivative of a Lyapunov function $V(x)$ with respect to time is denoted by $\dot{V}$.

2 Quadratically Convergent Algorithm

Many merit functions have been proposed to derive orthonormalization algorithms [4]. In this section, analysis of some of these algorithms will be given.

First, consider the following unconstrained optimization problem:

$$\text{Optimize } F(x) = tr\{\frac{1}{2}x^Tx - \frac{1}{4}(x^Tx)^2\}, \tag{1}$$

where $x \in \mathbb{R}^{n\times p}$. The gradient of F is

$$\nabla F(x) = x - xx^Tx.$$

If $x \in \mathbb{R}^{n\times 1}$, then the Hessian of F can be shown to be

$$\nabla^2 F(x) = I - x^TxI - 2xx^T.$$

The set of equilibrium points of the above system consists of all points for which $x - xx^Tx = 0$. Clearly if x is full rank, then $\nabla F(x) = x - xx^Tx = 0$ if and only if $x^Tx = I$. If x is not full rank, then x^Tx and xx^T are orthogonal projections, i.e., $(x^Tx)^2 = x^Tx$ and $(xx^T)^2 = xx^T$. Also, if $x^Tx = I$, then $\nabla^2F(x) = I - x^TxI - 2xx^T = -2xx^T$ which is negative semidefinite. In general, the Hessian matrix is

$$\nabla^2 F(x) = I \otimes I - x^Tx \otimes I - I \otimes xx^T - Kx \otimes x^T, \tag{2a}$$

where K is a symmetric permutation matrix. Hence at any full rank equilibrium point $\hat{x}$ where $\hat{x}^T\hat{x} = I$, it follows that

$$\begin{aligned}\nabla^2 F(\hat{x}) &= I \otimes I - \hat{x}^T\hat{x} \otimes I - I \otimes \hat{x}\hat{x}^T - K\hat{x} \otimes \hat{x}^T \\ &= I \otimes I - I \otimes I - I \otimes \hat{x}\hat{x}^T - K\hat{x} \otimes \hat{x}^T \\ &= -I \otimes \hat{x}\hat{x}^T - K\hat{x} \otimes \hat{x}^T \\ &= -(I \otimes \hat{x})(I \otimes I + K)(I \otimes \hat{x}^T),\end{aligned} \tag{2b}$$

for some symmetric permutation matrix K [4]. Since K is symmetric and $K^2 = I$, then each eigenvalue of K is 1 or -1. Thus $\nabla^2F(x)$ is negative semidefinite, i.e., each nonzero solution of $\nabla F(x) = 0$ satisfies the first and second order of optimality conditions.

Based on the gradient of the above cost function, we obtain the following dynamical system:

$$x' = x - xx^Tx. \tag{3}$$

The exact solution of the system (3) when $x \in \mathbb{R}^{n\times 1}$ is

$$x(t) = \frac{e^tx_0}{\sqrt{x_0^Tx_0e^{2t} - x_0^Tx_0 + 1}}, \tag{4}$$

where $x_0 = x(0)$. This shows that if $x_0 \neq 0$, then $\lim_{t\to\infty} x(t) = \frac{x_0}{\sqrt{x_0^T x_0}}$, and that if $x_0^T x_0 = 1$, then $x(t)^T x(t) = 1$ for all $t \geq 0$. Generally, $\text{rank}(x(t)) = \text{rank}(x(0))$ for all $t \geq 0$.

The systems (3) can be shown to be globally stable [6,7] using the function $V(x) = \frac{1}{4}tr\{(x^T x - I)^2\}$. It follows that the time derivative of V along the trajectory of (3) is

$$\dot{V} = -\frac{1}{2}tr\{(x^T x - I)^2(x^T x)\} \leq 0. \tag{5}$$

It follows that the system (3) is globally stable [7]. We also note that $\dot{V} = 0$ if and only if $(x^T x - I)x^T = 0$. Moreover, the set $\Omega_1 = \{x \in \mathbb{R}^{n\times p} : x^T x = I\}$ is an invariant set for the system (3). Here $n, p \in \mathbb{N}$ and $p \leq n$. This implies that if $x_0 \in \Omega_1$, then $x(t) \in \Omega_1$ for each $t \geq 0$. Thus, in the limit, $(x^T x)^2 = x^T x$ which means that as $t \to \infty$, $x(t)^T x(t)$ converges to an identity matrix provided that x_0 is full rank.

Using Euler method for solving the ordinary differential equation (3) with a stepsize $\alpha = \frac{1}{2}$, the discretized version of (3) leads to the following iterative equation:

$$x_{k+1} = \frac{x_k}{2}\{3I - x_k^T x_k\}, \tag{6}$$

where x_0 is the initial matrix. The analysis of convergence can be established using the iteration function

$$f(z) = \frac{3}{2}z - \frac{1}{2}z^3, \tag{7}$$

where $z_0 = \sigma$ is a singular value of x_0. Clearly, $f(z) = z$ if and only if $z = 0, 1, -1$.

The quadratic convergence of (7) is obtained from the equations:

$$f(z) - 1 = (z-1)^2(-\frac{1}{2}z - 1),$$

$$f(z) + 1 = (z+1)^2(-\frac{1}{2}z + 1).$$

To insure convergence, the initial condition x_0 should be scaled so that its singular values are in the interval $(-\sqrt{5}, \sqrt{5}) = (-2.236067, 2.236067)$. This means that if $z_0 \in (-\sqrt{5}, \sqrt{5})$, then

$$z_{k+1} = \frac{3}{2}z_k - \frac{1}{2}z_k^3, \tag{8}$$

converges to 1 or -1 depending on whether $z_0 > 0$ or $z_0 < 0$, respectively. A proper scaling factor is $||A||_1 = \max_i\{\sum_{j=1}^n |a_{ij}|\}$, or $||A||_\infty = \max_j\{\sum_{j=1}^m |a_{ij}|\}$ [8]. From the inequalities $\sigma_1 = ||A||_2 \leq ||A||_1$, $\sigma_1 = ||A||_2 \leq ||A||_\infty$, or $\sigma_1 = ||A||_2 \leq \sqrt{||A||_1||A||_\infty}$, one may use $x_0 = \frac{A}{||A||_1}$, $x_0 = \frac{A}{||A||_\infty}$, or $x_0 = \frac{A}{\sqrt{||A||_1||A||_\infty}}$ as an initial matrix for the system (6). The singular values of each of these matrices are in the interval [0 1].

3 Higher-Order Polynomial Type Fixed Point Functions

Motivated by the quadratically convergent algorithm (6), higher order convergent fixed point functions by optimizing some merit functions. By considering unconstrained optimization problems of the form

$$\text{Optimize } F(x) = tr\{\frac{1}{2}x^T x - \frac{1}{2r+2}(x^T x)^{r+1}\}, \tag{9}$$

where the integer $r \geq 2$ and $x \in \mathbb{R}^{n\times p}$. The dynamical system based on the gradient of F is

$$x' = x - x(x^T x)^r, \ x(0) = x_0. \tag{10}$$

The analysis of convergence can be established using a fixed point function $z = g_1(z)$ where g_1 is obtained from a discretized version of (10):

$$g_1(z) = (1+\alpha)z - \alpha z^{2r+1}, \ z \in \mathbb{R}$$

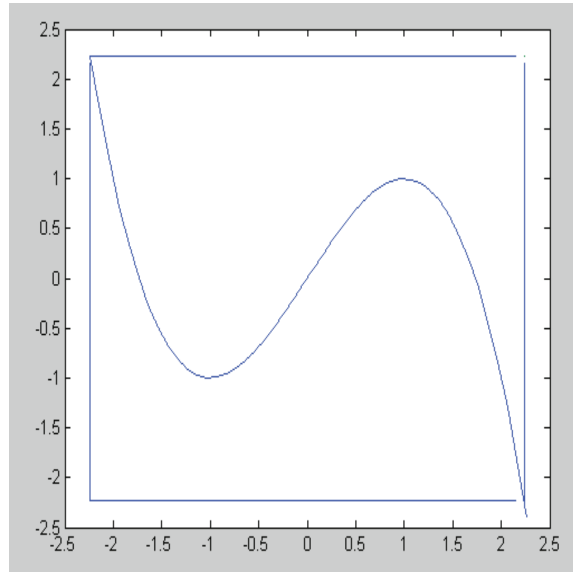

Figure 1: This figure shows that $g_1(z)$ increasing in the interval $(-1, 1)$ and decreasing on the intervals $(-\sqrt{5}, -1)$ and $(1, \sqrt{5})$.

with initial value $z_0 = \sigma$ where σ is a singular value of x_0.

We are interested in the nonzero real fixed points of g_3 which are $-1, 1$. The values of α for which $g_1'(\pm 1) = 0$ are obtained by solving the equation

$$1 + \alpha - (2r+1)\alpha = 0. \tag{11}$$

Therefore,

$$\alpha = \frac{1}{2r}.$$

Hence the fixed point function

$$g_1(z) = \frac{1+2r}{2r}z - \frac{1}{2r}z^{2r+1}, \tag{12}$$

is quadratically convergent near $z = \pm 1$. Note that $-z \leq g_1(z) \leq z$ if $z \in (-(4r+1)^{\frac{1}{2r}}, (4r+1)^{\frac{1}{2r}})$. Thus the set of initial values for which (12) converges becomes smaller as $r \to \infty$.

Methods of higher order convergence may be obtained by combining different methods of using different values of the positive integer r. For example it is possible to design a method of order 3 of the form

$$g_2(z) = \frac{1+\alpha}{2}z - \frac{\alpha}{2}z^3 + \frac{1+\beta}{2}z - \frac{\beta}{2}z^5,$$

for some α and β. The parameters α and β are determined by solving the equations

$$g_2'(\pm 1) = 0 = \frac{2+\alpha+\beta}{2} - \frac{3\alpha}{2}z^2 - \frac{5\beta}{2}z^4|_{z=\pm 1},$$

$$g_2''(\pm 1) = 0 = -\frac{6\alpha}{2}z - \frac{20\beta}{2}z^3|_{z=\pm 1}.$$

These imply that

$$\begin{array}{c}\alpha + 2\beta = 1 \\ 3\alpha + 10\beta = 0\end{array},$$

and hence $\alpha = \frac{5}{2}$ and $\beta = \frac{-3}{4}$. This leads to the following iterative equation:

$$g_2(z) = \frac{z}{8}\{15 - 10z^2 + 3z^4\}. \tag{13}$$

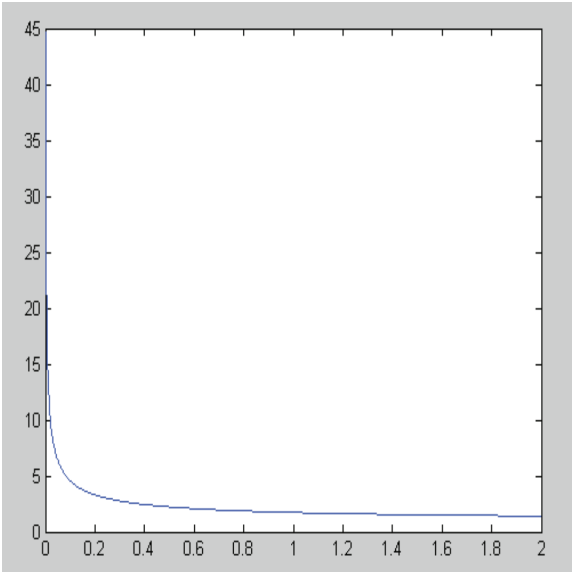

Figure 2: This figure provides the radius of intervals, given in the vertical axis, as a function of α, given in the horizontal axis, for which the iteration $z_{k+1} = (1+\alpha)z_k - \alpha z_k^3$ converges. For example, $\alpha = 1$ corresponds to $\sqrt{2}$ which implies that (8) converges from any initial value in the interval $(-\sqrt{2}, \sqrt{2})$. This figure also shows that the smaller α the larger the interval of convergence.

In matrix form, the iteration (13) may be expressed as

$$x_{k+1} = \frac{x_k}{8}\{15I - 10x_k^T x_k + 3(x_k^T x_k)^2\}. \tag{14}$$

The initial condition x_0 should be chosen so that its singular values are in the interval $(-1.5275, 1.5275)$. Thus (14) converges provided that the singular values of $x_0 \in (-1.5275, 1.5275)$. These bounds may be derived from the inequality $-z \leq g_2(z) = \frac{z}{8}\{15 - 10z^2 + 3z^4\} \leq z$. Figure 3 shows that if $z \in (-1.5275, 1.5275)$, then $f(z) \in (-1.5275, 1.5275)$.

Another example of deriving a method of order 4 involves determining α, β, γ such that

$$g_3(z) = \frac{1+\alpha}{3}z - \frac{\alpha}{3}z^3 + \frac{1+\beta}{3}z - \frac{\beta}{3}z^5 + \frac{1+\gamma}{3}z - \frac{\gamma}{3}z^7.$$

The points $z = \pm 1$ are fixed points of F for any parameters α, β and γ. To derive a fourth order fixed point function, the following equations must be satisfied:

$$g_3'(\pm 1) = 0 = \frac{3+\alpha+\beta+\gamma}{3} - \frac{3\alpha}{3}z^2 - \frac{5\beta}{3}z^4 - \frac{7\gamma}{3}z^6|_{z=\pm 1}$$
$$g_3''(\pm 1) = 0 = -\frac{6\alpha}{3}z - \frac{20\beta}{3}z^3 - \frac{42\gamma}{3}z^5|_{z=\pm 1}$$
$$g_3'''(\pm 1) = 0 = -\frac{6\alpha}{3} - \frac{60\beta}{3}z^2 - \frac{210\gamma}{3}z^4|_{z=\pm 1}.$$

This leads to the following set of equations

$$\begin{aligned} 2\alpha + 4\beta + 6\gamma &= 3, \\ 6\alpha - 20\beta - 42\gamma &= 0, \\ 6\alpha + 60\beta + 210\gamma &= 0. \end{aligned} \tag{15}$$

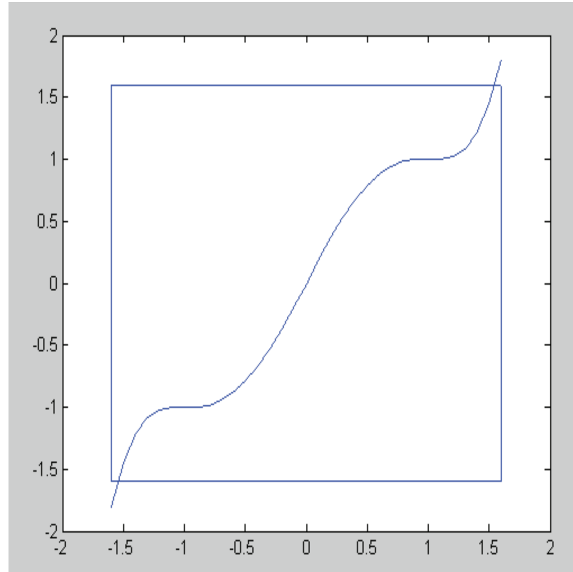

Figure 3: This figure shows that $f(z) = \frac{z}{8}\{15 - 10z^2 + 3z^4$ increasing in the interval $(-1, 1)$ and decreasing on the intervals $(-1.5275, -1)$ and $(1, 1.5275)$.

The solution of these equations is
$(\alpha, \beta, \gamma) = (6.5625, -3.9375, 0.9375)$. Thus

$$g_3(z) = \frac{z}{16}(35 - 35z^2 + 21z^4 - 5z^6). \tag{16}$$

This fixed point iteration converges provided that $z_0 \in (-1.73205080756888, 1.73205080756888)$

The fixed point iteration (16) may be expressed in matrix form as

$$G_3(x) = \frac{x}{16}(35 - 35x^T x + 21(x^T x)^2 - 5(x^T x)^3). \tag{17}$$

This converges provided that the largest singular value of the initial matrix x_0 is in the interval (0,1.73205080756888).

4 Orthonormalization with Respect to a Matrix

Let B be a positive definite matrix, and let x_0 be a full rank matrix. The above analysis may be extended to develop dynamical systems that converge to a matrix $\bar{x}$ such that $\bar{x}^T B \bar{x} = I$. This may be accomplished by replacing the quantity $x^T x$ with $x^T Bx$ in each of the orthogonalization algorithms. Thus we obtain the following:

$$x_{k+1} = \frac{3}{2}x_k - \frac{1}{2}x_k(x_k^T Bx_k). \tag{18}$$

The initial condition x_0 should be chosen so that the singular values of $B^{\frac{1}{2}}x_0$ are in the interval $(0, \sqrt{5})$.

Similarly, the iteration functions (14) and (17) may be modified as

$$x_{k+1} = \frac{x_k}{8}\{15I - 10x_k^T Bx_k + 3(x_k^T Bx_k)^2\}. \tag{19}$$

$$x_{k+1} = \frac{x_k}{16}(35 - 35x_k^T Bx_k + 21(x_k^T Bx_k)^2 - 5(x_k^T Bx_k)^3). \tag{20}$$

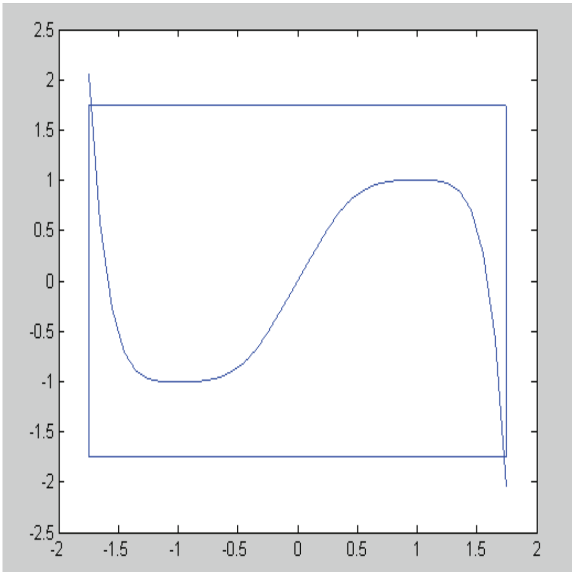

Figure 4: The graph of $g_3(z) = \frac{z}{16}(35 - 35z^2 + 21z^4 - 5z^6)$ in the interval (-1.73205080756888,1.73205080756888)

Convergence of (19) and (20) is insured if the initial condition x_0 is chosen so that the singular values of $B^{\frac{1}{2}}x_0$ are in the intervals $(0, 1.5275)$, and $(0, 1.73205080756888)$, respectively.

5 Application to MSA and PSA

Principal and/or minor subspace analysis (PSA/MSA) of a positive definite matrix are fundamental tasks for many signal processing applications. From the observation that $Ax(x^TAx)^{-1} \approx x^{-T}$ provided that x is a good approximation of an eigenvector of A, it is possible to modify the iterations (6), (14), and (17) so that

$$x_{k+1} = \frac{3}{2}Ax_k(x_k^TAx_k)^{-1}x_k^Tx_k - \frac{1}{2}x_kx_k^Tx_k, \qquad (21a)$$

$$x_{k+1} = \frac{3}{2}x_k - \frac{1}{2}Ax_k(x_k^TAx_k)^{-1}(x_k^Tx_k)^2, \qquad (21b)$$

$$x_{k+1} = \frac{Ax_k(x_k^TAx_k)^{-1}x_k^Tx_k}{8}\{15I - 10x_k^Tx_k + 3(x_k^Tx_k)^2\}. \quad (22)$$

$$x_{k+1} = \frac{Ax_k(x_k^TAx_k)^{-1}x_k^Tx_k}{16} \times \{35 - 35x_k^Tx_k + 21(x_k^Tx_k)^2 - 5(x_k^Tx_k)^3\}. \qquad (23)$$

It can be verified that the iterations (21a), (22), (23) converge to principal subspaces of A while the iteration (21b) converges to minor subspaces of A.

6 Simulation Results

In order to confirm the validity and performance of the proposed algorithms, a simulation example is given below. The iteration (6) is applied to the 6×3 matrix x_0 which is given by:

```
x0 =0.9602    1.0210    1.1673
    1.2967    0.5765    1.6790
    1.0132    0.3442    0.7447
    1.2916    1.0366    1.4550
    0.9513    1.4546    1.5331
    0.6148    0.9578    1.1575
```

Using the Matlab function *rand*, the matrix x_0 is generated by adding two random matrices of size 6×3. The nonzero singular values of x_0 are $4.6812, 0.8306, 0.3496$. Since some of the singular values are greater than $\sqrt{5}$, the matrix x_0 needs to be scaled so that the singular values of the scaled matrix are smaller than $\sqrt{5}$. Thus the matrix x_0 is divided by $||x_0||_1$ and the resulting matrix is used as initial matrix for the iteration (6). After 20 iterations, Algorithm (6) converges to the matrix

```
x=0.2979    0.4226    0.0341
  0.2094   -0.4667    0.8247
  0.7598   -0.0560   -0.1711
  0.4891    0.2545    0.1225
 -0.0621    0.6588    0.2891
 -0.2168    0.3189    0.4369
```

The accuracy of x is measured by the error $||x^Tx - I||_2 = 2.4195(10)^{-16}$.

7 Conclusion

Polynomial type higher order methods for orthonormalization are developed from converting continuous differential equations into discrete ones by choosing optimal stepsizes. These methods are shown to be equivalent to methods for computing the polar decomposition of matrices. Convergence analysis and the sets of initial conditions for which these methods converge are given. The proposed iterations may be generalized in many directions. For example, the proposed methods can be used to extend and analyze the behavior of these algorithms for the complex case. Additionally, more simulations need to be conducted to examine the convergence behavior of the iterations stated in Sections 4 and 5.

References

[1] G. H. Golub and C. F. van Loan, Matrix Computations. Baltimore, MD: The Johns Hopkins Univ. Press, 1989.

[2] D. R. Fokkema, G. L. G. Sleijpen, and H. A. van der Vorst. Jacobi-Davidson style QR and QZ algorithms for the partial reduction of matrix pencils. SIAM J. Sci. Comput., 20:94-125, 1998.

[3] H. Simon, Analysis of the symmetric Lanczos algorithm with reorthogonalization methods. Linear Algebra Appl., 61:101-132, 1984.

[4] Hasan, M.A., "Orthonormalization Learning Algorithms," proceedings of the International Joint Conference on Neural Networks, 2007 (IJCNN 2007), 12-17 Aug. 2007 Page(s):1894-1899.

[5] J. R. Magnus and H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, 2nd ed. New York: Wiley, 1991.

[6] Hassan K. Khalil, Nonlinear Systems, 3rd edition. Prentice Hall, 2002.

[7] J. J. E. Slotine and W. Li. Applied nonlinear Control. Prentice Hall, 1991.

[8] J. Stoer and R. Bulirsch, Introduction to Numerical Analysis, Springer-Verlag, New York 1980.