ARTICLE IN PRESS

ISA Transactions xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

ISA Transactions

journal homepage: www.elsevier.com/locate/isatrans

Research article

Matrix function optimization under weighted boundary constraints and its applications in network control

Pei Tang[a,b,1], Guoqi Li[a,b,*,1], Chen Ma[a,b], Ran Wang[c], Gaoxi Xiao[d], Luping Shi[a,b,**]

[a] Department of Precision Instrument, Center for Brain Inspired Computing Research, Tsinghua University, Beijing, 100084, China
[b] Beijing Innovation Center for Future Chip, Tsinghua University, Beijing, China
[c] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[d] School of EEE, Nanyang Technological University, Singapore

ARTICLE INFO

Keywords:
Matrix function optimization
Matrix variable
Weighted orthornormal constraint
Weighted trace constraint
Network control

ABSTRACT

The matrix function optimization under weighted boundary constraints on the matrix variables is investigated in this work. An "index-notation-arrangement based chain rule" (*I-Chain rule*) is introduced to obtain the gradient of a matrix function. By doing this, we propose the weighted trace-constraint-based projected gradient method (WTPGM) and weighted orthornormal-constraint-based projected gradient method (WOPGM) to locate a point of minimum of an objective/cost function of matrix variables iteratively subject to weighted trace constraint and weighted orthonormal constraint, respectively. New techniques are implemented to establish the convergence property of both algorithms. In addition, compared with the existing scheme termed "orthornormal-constraint-based projected gradient method" (OPGM) that requires the gradient has to be represented by the multiplication of a symmetrical matrix and the matrix variable itself, such a condition has been relaxed in WOPGM. Simulation results show the effectiveness of our methods not only in network control but also in other learning problems. We believe that the results reveal interesting physical insights in the field of network control and allow extensive applications of matrix function optimization problems in science and engineering.

1. Introduction

It is well known that the derivative is a fundamental tool in many science and engineering problems [1,2]. For a scalar function of a real variable, the derivative measures the sensitivity of function change with respect to such a variable, which has meaningful physical insights. For example, the derivative of the position of a moving object with respect to time is the object's velocity, and it measures how quickly the object position changes when time involves. However, finding the derivative of a function with respect to a real variable is not enough when one wants to describe a more complicated problem in which a function is determined by a set of variables. In such a case, the study of the derivative of a function with respect to a vector becomes necessary. Vector derivatives that take in vector variables are extremely important, where they arise throughout fluid mechanics [3], electricity and magnetism [4], elasticity [5], and many other areas of theoretical and applied physics [6]. Vector derivatives can be combined in different ways, such as divergence [7] and curl [8] operators, producing sets of identities that are also very important in physics.

A vector is a special form of a matrix in which all elements are organized in a line, and a matrix can always be stacked to a vector form. However, in various practical statistics and engineering problems, stacking a matrix into a vector will lose the physical meaning within each column. For example, in the problem of control of complex networks [9], we need to design an input matrix to achieve the control objective. The number of columns of the input matrix is the number of external control sources available, and stacking the input matrix into a vector makes the network become uncontrollable. In these cases, taking the derivative of a function with respect to a matrix variable becomes essential. To this end, we need to collect various partial derivatives of a single function with respect to many variables, and/or of a multivariate function with respect to a single variable, and obtain the total differential information. Thus, operations in finding a local maximum/minimum of a multivariate function solving differential equations can

* Corresponding author. Department of Precision Instrument, Center for Brain Inspired Computing Research, Tsinghua University, Beijing, 100084, China.
** Corresponding author. Department of Precision Instrument, Center for Brain Inspired Computing Research, Tsinghua University, Beijing, 100084, China.
E-mail addresses: tangp14@mails.tsinghua.edu.cn (P. Tang), liguoqi@mail.tsinghua.edu.cn (G. Li), macheng@mail.tsinghua.edu.cn (C. Ma), wangran@nuaa.edu.cn (R. Wang), EGXXiao@ntu.edu.sg (G. Xiao), lpshi@mail.tsinghua.edu.cn (L. Shi).
[1] G.Li and P. Tang contribute equally to this work.

https://doi.org/10.1016/j.isatra.2018.06.010
Received 4 May 2017; Received in revised form 27 May 2018; Accepted 23 June 2018
0019-0578/

Please cite this article as: Tang, P., ISA Transactions (2018), https://doi.org/10.1016/j.isatra.2018.06.010

be significantly simplified via various gradient descent methods [10]. In this paper, optimization problems where vector variables and matrix variables are involved are termed by *vector function optimization* and *matrix function optimization* problems, respectively. The notation “vector” and “matrix” used here is commonly used in statistics and engineering, while the tensor index notation [11,12] is preferred in physics.

Motivated by an irresistible longing to understand the above issues, we moved from *vector function optimization* problems to *matrix function optimization* problems whose variables are matrices in our most recently work [13]. To accomplish this issue, we hope to know the gradient information of a matrix function. However, it is generally hard to obtain such information when matrices-by-matrices derivatives are involved. It should be noticed that the derivative of a vector function with respect to another matrix is a high-order tensor. For example, for a scalar cost function $E(B)$ where $B \in \mathbb{R}^{n\times m}$, it is a function of another matrix $Q \in \mathbb{R}^{p\times q}$. The derivative of cost function $E(B)$ with respect to B captures how the matrix variable B affects the cost function. To this end, we explore how the value of the element B_{kl} of the matrix B affects the value of element Q_{ij} of the matrix Q. Particularly, we need to differentiate $\frac{\partial Q_{ij}}{\partial B_{kl}}$ for all i, j and k, l. Such a differentiation results in a fourth order tensor [14,15]. In short it is an $m \times n$ matrix, and each of its entries is a $p \times q$ matrix.

Although vector derivative has been well established, matrix derivative is difficult. Currently there is no unified framework that can completely solve this problem [16]. Existing schemes mainly use two basic ways to deal with this issue, one is the Vec operator and Kronecker products arrangement [17,18], and the other is the index notation arrangement [19]. However, for implementations, there are a lot of intricacy and tedious calculation. The main difficulty here is keeping track of where things are put since a matrix variable may depend on numerous intermediate matrix variables. This situation becomes worse when $E(.)$ has a more complicated form. We find that index notation arrangement relatively simplifies the presentation and manipulation of differential geometry when doing matrix differentiation. Thus, we proposed an index-notation-arrangement based chain rule (*I-Chain rule*) in Ref. [13]. By obtaining the gradient of a matrix function using *I-Chain rule*, two iterative algorithms, namely, *trace-constraint-based projected gradient method* (TPGM) and *orthornormal-constraint-based projected gradient method* (OPGM) were presented to solve the matrix function optimization problems. Projection and normalization operators were utilized to establish the convergence of TPGM and OPGM. This work provided a unified framework which reveals important physical insights and deepens our understanding of various matrix function optimization problems, and inspires wide applications in science and engineering.

However, in the work [13], to guarantee the convergence of TPGM/OPGM, it is required that the gradient can be represented by the multiplication of a symmetrical matrix and the matrix variable itself. That is to say, the gradient of the cost function should be represented in the form of $\nabla E(B) = F(B) \cdot B$ where $F(B) \in \mathbb{R}^{N\times N}$ is symmetrical and $B_k \in \mathbb{R}^{N\times M}$. Although such an assumption holds in various applications, it indeed does not hold in some cases. For example, consider the case that $E(B) = \mathrm{tr}((L - BX)^T(L - BX))$ where tr(.) denotes the matrix trace function. In this case, the gradient $\nabla E(B)$ cannot be represented by $\nabla E(B_k) = F(B_k) \cdot B_k$. Therefore we investigate how to develop an algorithm and ensure its convergence in this work. We also find that the boundary constraints in Ref. [13] can be further relaxed. More particularly, by introducing an real symmetry positive definite weight matrix G, the trace constraint and the orthornormal constraint can be relaxed to the weighted trace constraint and weighted orthornormal constraint, respectively. The boundary constraints in Ref. [13] become special cases of this work in which G is an identity matrix.

The main problem we faced is how to deal with the non-symmetrical of $\nabla E(B)B^T$ for a formulated matrix objective function under more relaxed constraints. It brought this question up: a non symmetrical matrix diagonalized as its eigenvalues may not be all real values and therefore existing techniques cannot guarantee the convergence of the TPGM/OPGM algorithm in this case. To this end, we propose the weighted trace-constraint-based projected gradient method (WTPGM) and the weighted orthornormal-constraint-based projected gradient method (WOPGM) to locate a point of minimum of an objective/cost function of matrix variables iteratively subject to weighted trace boundary constraint condition and weighted orthornormal constraint condition, respectively. Our main idea is to replace $\nabla E(B)$ by $\nabla E(B) B^TGB$, which can be represented by $F(B) \cdot B$ such that $F(B) = \nabla E(B)B^TG$. The key technique in guaranteeing the convergence of WTPGM is to obtain the orthonormal basis of GB in the iteration process. While for WOPGM, the essential issue is to the establishment of the value condition of λ_k in the iteration process. Introducing the parameter λ_k is similar to the idea of Levenberg-Marquardt stabilization, also known as the damped least-squares (DLS) [20,21] method, which is generally used to solve non-linear least squares problems. In the next section, we shall prove that $E(B_k)$ is convergent to $E(B^*)$ as $k \to \infty$, with B^* having orthonormal columns, provided that the step length η is sufficiently small. Thus, both the assumption on the gradient of the cost function and requirement on the boundary constraint have been relaxed in this paper. This means that we are able to extend our method regarding the optimization of matrix functions to more extensive applications in science and engineering. Simulation results show the effectiveness of our framework.

To show the effectiveness of our method, various case studies including two in the area of network control are illustrated. In the first case study, we focus on how to identify nodes to which the external control sources are connected so as to minimize a pre-defined energy cost function of a control strategy. Different from the work in Ref. [13], a positive definite diagonal weight matrix G reflecting the restriction on each external control source is considered. The matrix function optimization model built in this work allows us to investigate how G can affect the control cost. By applying WTPGM and WOPGM, we uncover that the control cost is related to the condition number of G. This interesting observation may lead to heuristic algorithm design for the minimum cost control of large scale of real life complex networks, which deserves great attention for the future research. In the second case study, we consider controlling directed networks by only evolving the connection strengths on a fixed network structure. In this case, the topology matrix A becomes a matrix variable of the control cost function while the input matrix B is fixed. By this example, we also show that the proposed WTPGM and WOPGM are applicable when G becomes an identity matrix, which suggests that WTPGM and WOPGM are more general than TPGM and OPGM, respectively. In addition, we uncover that, when the control sources are evenly allocated, the system can be considered as a few identical subsystems and the control cost attains its minimum. This is meaningful when one want to explore how network topology evolution affects the cost of controlling these networks.

There are some literature considering optimization problems where matrix variables are involved under specific constraints [13,22,22–38]. However, the cost functions in these works are in relatively simple and specific forms [24,25]. For example, they are usually simple trace functions such as $\mathrm{tr}(X^TAX)$ where X are the matrix variable and A is a given symmetrical matrix [37,38]. Regarding the constraints, applications of trace and orthornormal constraints can be found in many practical problems such as in machine learning problems [26,27], image processing [28,29], signal processing [22,30,31], modularity detection [32,33] and complex networks [34,36]. Existing schemes translate each of the above applications into some particular models that are manageable, so they fail to deal with the problems in a general way.

The remaining part of the paper is organized as follows. In Section 2, we illustrate how a matrix function optimization problem is