Schur functions for approximation problems

Nadezda Sukhorukova, Julien Ugon and David Yost

Abstract In this paper we propose a new approach to least squares approximation problems. This approach is based on partitioning and Schur function. The nature of this approach is combinatorial, while most existing approaches are based on algebra and algebraic geometry. This problem has several practical applications. One of them is curve clustering. We use this application to illustrate the results.

1 Introduction

In this paper we formulate a specific least least squares approximation problem and provide a signal processing application where this problem is used. The main technical difficulty for this problem is to solve linear systems with the same system matrix and different right-hand sides. One simple approach that can be proposed here is to invert the system matrix and multiply the updated right-hand side by this inverse at each iteration. In general, it is not very efficient to solve linear systems through computing matrix inverses, but in this particular application it is very beneficial. One technical difficulty here is to know in advance whether the system matrix

Nadezda Sukhorukova
Faculty of Science, Engineering and Technology, Swinburne University of Technology, PO Box 218, Hawthorn, Victoria, Australia and Centre for Informatics and Applied Optimization, Federation University Australia e-mail: nsukhorukova@swin.edu.au

Julien Ugon
Faculty of Science, Engineering and Built Environment, Deakin University, 221 Burwood Highway Burwood Victoria 3125, Australia and Centre for Informatics and Applied Optimization, Federation University Australia e-mail: julien.ugon@deakin.edu.au

David Yost
Centre for Informatics and Applied Optimization, Federation University Australia e-mail: d.yost@federation.edu.au

is invertible or not. Similar problems appear in Chebyshev (uniform) approximation problems as well.

In this paper we suggest a new approach for dealing with this kind of systems. This approach in based on Schur functions, a well-established techniques that is used to describe partitioning [2]. The very nature of these functions is combinatorial. Based on our previous experience [4], the characterisation of the necessary and sufficient optimality conditions for multivariate Chebyshev approximation is also combinatorial and therefore Schur function is a very natural tool to work with these problems.

This paper is organised as follows. In section 2 we introduce a signal processing application that relies on approximation and optimisation. In section 3 we provide a mathematical formulation to the signal processing problem and discuss how it can be simplified. In section 4 we introduce an innovative approach for solving the problem. This approach is based on Schur functions. Finally, in section 5 we provide future research directions.

2 Signal clustering

In signal processing, there is a need for constructing signal prototypes. Signal prototypes are summary curves that may replace the whole group of signal segments, where the signals are believed to be similar to each other. Signal prototypes may be used for characterising the structure of the signal segments and also for reducing the amount of information to be stored.

Any signal group prototype should be an accurate approximation for each member of the group. On the top of this, it is desirable that the process of recomputing group prototypes, when new group members are available, is not computationally expensive.

In this paper we suggest a k-means and least square approximation based model. Similar models are proposed in [3]. This is a convex optimisation problem. There are several advantages of this model. First of all, it provides an accurate approximation to the group of signals. Second, this problem can be obtained as a solution to a linear system and can be solved efficiently. Finally, the proposed approach allows one to compute prototype updates without recomputing from scratch.

3 Mathematical formulation

3.1 Prototype construction

Assume that there is a group of l signals $S_1(t),\dots,S_l(t)$, whose values are measured at discrete time moments

$$t_1, \ldots, t_N,\ t_i \in [a,b],\ i = 1, \ldots, N.$$

We suggest to construct the prototype as a polynomial $P_n(\mathbf{X},t) = \sum_{i=0}^{n} x_i t^i$ of degree n, whose least squares deviation from each member of the group on $[a,b]$ is minimal. That is, one has to solve the following optimisation problem:

$$\text{minimise } F(\mathbf{X}) = \sum_{i=1}^{N} \sum_{j=1}^{l} (S_j(t_i) - P_n(\mathbf{X},t_i))^2, \tag{1}$$

where $\mathbf{X} = (x_0, \ldots, x_n) \in \mathbb{R}^{n+1}$, $x_k,\ k = 0, \ldots, n$ are the polynomial parameter and also the decision variables. Each signal is a column vector

$$\mathbf{S}^j = (S_j(t_1), \ldots, S_j(t_N))^T,\ j = 1, \ldots, l.$$

Problem (1) can be formulated in the following matrix form:

$$\text{minimise } F(\mathbf{X}) = \|\mathbf{Y} - \mathbf{B}\mathbf{X}\|, \tag{2}$$

where

$\mathbf{X} = (x_0, \ldots, x_n) \in \mathbb{R}^{n+1}$, are the decision variables (same as in (1));
vector

$$\mathbf{Y} = \begin{pmatrix} \mathbf{S}^1 \\ \mathbf{S}^2 \\ \vdots \\ \mathbf{S}^l \end{pmatrix} \in \mathbb{R}^{(n+1)l}$$

matrix $\mathbf{B}$ contains repeated matrix blocks, namely,

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_0 \\ \mathbf{B}_0 \\ \mathbf{B}_0 \\ \vdots \\ \mathbf{B}_0 \end{pmatrix},$$

where

$$\mathbf{B}_0 = \begin{pmatrix} 1 & t_1 & t_1^2 & \ldots & t_1^n \\ 1 & t_2 & t_2^2 & \ldots & t_2^n \\ \vdots & \vdots & \ddots & \ldots & \vdots \\ 1 & t_N & t_N^2 & \ldots & t_N^n \end{pmatrix}.$$

This least squares problem can be solved using a system of normal equations:

$$\mathbf{B}^T\mathbf{B}\mathbf{X} = \mathbf{B}^T\mathbf{Y}. \tag{3}$$

Taking into account the structure of the system matrix in (3), the problem can be significantly simplified:

$$l\mathbf{B}_0^T\mathbf{B}_0\mathbf{X} = \mathbf{B}_0^T\sum_{k=1}^{l}\mathbf{S}^k. \tag{4}$$

Therefore, instead of solving (3), one can solve

$$\mathbf{B}_0^T\mathbf{B}_0\mathbf{X} = \mathbf{B}_0^T\frac{\sum_{k=1}^{l}\mathbf{S}^k}{l} = \mathbf{B}_0^T\mathbf{S}, \tag{5}$$

where $\mathbf{S}$ is the average of all l signals of the group (centroid).

3.2 Prototype update

Suppose that a signal group prototype has been constructed. Assume now that we need to update our group of signals: some new signals have to be included, while some others are to be excluded. To update the prototype, one needs to update the centroid and solve (5) with the updated right-hand side, while the system matrix $\mathbf{B}_0^T\mathbf{B}$ remains the same.

If only few signals are moving in and out of the group, then the updated centroid can be calculated without recomputing from scratch. Assume that l_a signals are moving in the group (signals $S_a^1(t),\ldots S_a^{l_a}$), while l_r are moving out (signals $S_r^1(t),\ldots S_r^{l_r}$), then the centroid can be recalculated as follows:

$$S_{new}(t) = \frac{lS_{old}(t) + \sum_{k=1}^{l_a}S_a^k(t) + \sum_{k=1}^{l_r}S_r^k(t)}{l - l_r + l_a}.$$

Since the same system has to be solved repeatedly with different right-hand sides, one approach is to invert matrix $\mathbf{B}_0^T\mathbf{B}_0$, which is an $(n+1)\times(n+1)$ matrix. In most cases, n is much smaller than N or l and therefore this approach is quite attractive, if we can guarantee that matrix $\mathbf{B}_0^T\mathbf{B}_0$ is invertible. In the next section we discuss the verification of this property.

4 Schur functions and matrix inverse

4.1 Vandermonde and generalised Vandermonde matrices

Consider matrix $\mathbf{B}_0^T\mathbf{B}_0$. In general, matrix $\mathbf{B}_0$ can be defined as follows:

$$\mathbf{B}_0 = \begin{pmatrix} g_1(t_1) & g_2(t_1) & g_3(t_1) & \ldots & g_{n+1}(t_1) \\ g_1(t_2) & g_2(t_2) & g_3(t_2) & \ldots & g_{n+1}(t_2) \\ \vdots & \vdots & \ddots & \ldots & \vdots \\ g_1(t_N) & g_2(t_N) & g_3(t_N) & \ldots & g_{n+1}(t_N) \end{pmatrix},$$

where $g_i,\ i = 1,\dots,n+1$ are basis functions. In section 3 we were discussing polynomial approximation and therefore, the components of matrix $\mathbf{B}_0^T\mathbf{B}_0$ are monomials that are evaluated at different time-moments. Recall that $n+1 << N$. Matrix $\mathbf{B}_0^T\mathbf{B}_0$ is invertible if and only if matrix $\mathbf{B}_0$ has exactly $n+1$ linearly independent rows. This is always the case when functions $g_i,\ i = 1,\dots,n+1$ form a Chebyshev system (for example, monomials $g_i = t^{i-1}$, some systems of trigonometric functions). This is not always the case when, for example, some of the monomials are "missing" from the system. This situation is illustrated in the following example.

Example 1. Consider the system of two monomials on the segment $[-1,1]$:

$$g_1(t) = 1,\ g_2(t) = t^2,\ t_1 \neq t_2,$$

the monomial t is "missing". Take time-moments $t_1,t_2 \in [-1,1]$. The determinant

$$\begin{vmatrix} 1 & t_1^2 \\ 1 & t_2^2 \end{vmatrix} = 0 \Leftrightarrow t_1 = -t_2.$$

Therefore, these functions do not form a Chebyshev system, since the corresponding determinant is zero when, for example, $t_1 = -t_2 = 1$ and there is only one linear independent row.

Recall that in the case of classical polynomial approximation (all monomials are included into the set of basis functions), the corresponding determinant is non-zero as it is the determinant of a Vandermonde matrix. We now need to introduce so called generalised Vandermonde matrices.

Definition 1. Generalised Vandermonde matrices have the following structure:

$$G = \begin{pmatrix} t_1^{m_1} & t_2^{m_1} & \dots & t_{n+1}^{m_1} \\ t_1^{m_2} & t_2^{m_2} & \dots & t_{n+1}^{m_2} \\ \vdots & \vdots & \ddots & \vdots \\ t_1^{m_n} & t_2^{m_n} & \dots & t_{n+1}^{m_n} \end{pmatrix}.$$

Denote

$$m_1 = \lambda_1 + n - 1,\ m_2 = \lambda_2 + n - 2, \dots, m_n = \lambda_n + n - n = \lambda_n. \tag{6}$$

Define the following function

$$s_\lambda(t_1,\dots,t_{n+1}) = \frac{\det(G)}{\det(V)}, \tag{7}$$

where G is the matrix with one or more missing monomials and V is the Vandermonde matrix. Vandermonde matrices correspond to

$$\lambda_1 = \lambda_2 = \dots = \lambda_n = 0.$$

$s_\lambda(t_1, \dots, t_{n+1})$ is called Schur function, named after Issai Schur. Schur polynomials are certain symmetric polynomials of n variables. This polynomials are used in representation and partitioning. A good introduction to Schur polynomials can be found in [2]. Therefore,

$$\det(G) = s_\lambda(t_1, \dots, t_n)\det(V)$$

and hence one needs to study the behaivour of Schur functions. Therefore, the following theorem holds.

Theorem 1. *Matrix* $\mathbf{B}_0^T\mathbf{B}_0$ *is non-singular if and only if the corresponding Schur function (7) is non-zero.*

In particular, if $t_i > 0,\ i = 1, \dots, n+1$, then the system is Chebyshev. Note that this statement can be proven using a logarithmic transformation [1]. We believe, however, that our approach is also applicable to more general settings.

There are many studies on Schur polynomials and many efficient ways for computing them. This approach can be used, for example, if one needs to know if $\mathbf{B}_0^T\mathbf{B}_0$ is invertible. If the matrix is invertible, one can develop a very fast and efficient algorithm for curve cluster prototype updates. If the matrix is singular, one can use the singular-value decomposition for constructing the prototype updates. This decomposition can be computed once, since $\mathbf{B}_0^T\mathbf{B}_0$ remains unchanged when the cluster membership is updated.

5 Discussions and future research directions

There are many studies on how to compute Schur functions. We are particularly interested in the extension of this approach to Chebyshev (uniform) approximation and multivariate approximation. This is a very promising approach for dealing with this type of problems, since, as our previous studies suggested [4] the corresponding optimality conditions are very combinatorial in their nature and therefore, Schur functions are a very natural tool for study this kind of systems.

We are also planning to conduct a thorough numerical study of the signal processing application we are discussing in this paper.

Acknowledgements This paper was inspired by the discussions during a recent MATRIX program “Approximation Optimisation and Algebraic Geometry'' that took place in February 2018. We are thankful to the MATRIX organisers, support team and participants for a terrific research atmosphere and productive discussions.

This study was supported by the Australian Research Council Discovery Project DP18010060 “Solving hard Chebyshev approximation problems through nonsmooth analysis”.

References

1. Samuel Karlin and William Studden, *Tchebycheff systems, with applications in analysis and statistics*, Interscience Publishers New York, 1966 (English).
2. I. G. Macdonald, *Symmetric functions and hall polynomials*, Clarendon Press Oxford University Press, Oxford New York, 1995.
3. H. Späth, *Cluster analysis algorithms for data reduction and classification of objects*, Ellis Horwood Limited, Chichester, 1980.
4. Nadezda Sukhorukova, Julien Ugon, and David Yost, *Chebyshev multivariate polynomial approximation: alternance interpretation*, 2016 MATRIX Annals (David R. Wood, Jan de Gier, Cheryl Praeger, and Terence Tao, eds.), MATRIX Book Series, vol. 1, Springer, 2018, pp. 177–182.