Available online at www.sciencedirect.com
ScienceDirect
Linear Algebra and its Applications 424 (2007) 240–281
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

The higher-order derivatives of spectral functions ☆

Hristo S. Sendov

Department of Mathematics and Statistics, 50 Stone Road East, University of Guelph, Guelph, Ont., Canada N1G 2W1

Received 14 July 2006; accepted 20 December 2006
Available online 19 January 2007
Submitted by R.A. Brualdi

Dedicated to Professor Roger Horn on the occasion of his 65th birthday

Abstract

We are interested in higher-order derivatives of functions of the eigenvalues of real symmetric matrices with respect to the matrix argument. We describe a formula for the kth derivative of such functions in two general cases.

The first case concerns the derivatives of the composition of an arbitrary (not necessarily symmetric) k-times differentiable function with the eigenvalues of symmetric matrices at a symmetric matrix with distinct eigenvalues.

The second case describes the derivatives of the composition of a k-times differentiable *separable* symmetric function with the eigenvalues of symmetric matrices at an arbitrary symmetric matrix. We show that the formula significantly simplifies when the separable symmetric function is k-times continuously differentiable.

As an application of the developed techniques, we re-derive the formula for the Hessian of a general *spectral* function at an arbitrary symmetric matrix. The new tools lead to a shorter, cleaner derivation than the original one.

To make the exposition as self contained as possible, we have included the necessary background results and definitions. Proofs of the intermediate technical results are collected in the appendices.

AMS classification: 49R50; 47A75

Keywords: Spectral function; Differentiable; Twice differentiable; Higher-order derivative; Eigenvalue optimization; Symmetric function; Perturbation theory; Tensor analysis; Hadamard product

☆ Research supported by NSERC.
E-mail address: hssendov@uoguelph.ca

0024-3795/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2006.12.013

1. Introduction

We say that a real-valued function F of a real symmetric matrix argument is *spectral* if

$$F(UXU^{\mathrm{T}}) = F(X)$$

for every real symmetric matrix X in its domain and every orthogonal matrix U. That is, $F(X) = F(Y)$ if X and Y are symmetric and similar. The restriction of F to the subspace of diagonal matrices defines a function $f(x) = F(\mathrm{Diag}\, x)$ on a vector argument $x \in \mathbb{R}^n$. The function $f : \mathbb{R}^n \to \mathbb{R}$ is *symmetric*, that is, has the property

$$f(x) = f(Px) \quad \text{for any permutation matrix } P \text{ and any } x \text{ in the domain of } f,$$

and in addition, $F(X) = (f \circ \lambda)(X)$, in which the *eigenvalue map*

$$\lambda(X) = (\lambda_1(X), \ldots, \lambda_n(X))$$

is the vector of eigenvalues of X arranged in non-increasing order.

What smoothness properties of the symmetric function f are inherited by F? The eigenvalue map $\lambda(X)$ is continuous but not always differentiable with respect to X. Even in domains where $\lambda(X)$ is differentiable, it is difficult to organize the differentiation process so that one arrives at an elegant formula for the higher-order derivatives of $(f \circ \lambda)(X)$.

An important subclass of spectral functions is obtained when $f(x) = g(x_1) + \cdots + g(x_n)$ for some function g of one real variable. We call such symmetric functions *separable*; their corresponding spectral functions are called *separable spectral functions*.

In [13] there is an explicit formulae for the gradient of the spectral function F in terms of the derivatives of the symmetric function f:

$$\nabla(f \circ \lambda)(X) = V(\mathrm{Diag}\, \nabla f(\lambda(X)))V^{\mathrm{T}}, \tag{1}$$

where V is any orthogonal matrix such that $X = V(\mathrm{Diag}\, \lambda(X))V^{\mathrm{T}}$ is the *ordered* spectral decomposition of X. In [17] a formula for the Hessian of F was given, whose structure appeared quite different from the one for the gradient. Calculating the third and higher-order derivatives of F becomes unmanageable without an appropriate language for describing them.

In this work we generalize the work in [13,17] by proving, in two general cases, the following formula for the kth derivative of a spectral function

$$\nabla^k(f \circ \lambda)(X) = V\left(\sum_{\sigma \in P^k} \mathrm{Diag}^{\sigma} \mathscr{A}_{\sigma}(\lambda(X))\right)V^{\mathrm{T}}, \tag{2}$$

where again $X = V(\mathrm{Diag}\, \lambda(X))V^{\mathrm{T}}$. The sum is taken over all permutations on k elements, which are a convenient tool for enumerating the maps $\mathscr{A}_{\sigma}(x)$. The precise meanings of the operators Diag^{σ} and the conjugation by the orthogonal matrix V are explained in the next section; see (6) and (9) respectively. The maps $\mathscr{A}_{\sigma}(x)$ depend only on the partial derivatives of $f(x)$ up to order k, and do not depend on the eigenvalues; they reveal how the higher-order derivatives depend on the eigenvalue map $\lambda(X)$. Formula (2) depends on the eigenvalues only through the compositions $\mathscr{A}_{\sigma}(\lambda(X))$ and conjugation by the orthogonal matrix V.

We show that (2) is valid (a) when f is a k-times (continuously) differentiable function, not necessarily symmetric, and X is a matrix with distinct eigenvalues, and (b) when f is a k-times (continuously) differentiable separable symmetric function and X is an arbitrary symmetric matrix. We give a recipe for computing the maps $\mathscr{A}_{\sigma}(x)$ in these two cases.