Notes on Derivatives of Matrix Valued Matrix Functions

Hubertus van Dam

December 14, 2014

This document is concerned with derivatives of matrix valued matrix functions. I.e. expressions of the kind

$$\frac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}_{ij}} \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{N\times N}$ and $F : \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times N}$. These kinds of expressions arise e.g. in density matrix functional theory, grid free density functional theory [1], and continuum mechanics [4, 6, 3, 2].

In the field of continuum mechanics the emphasis seems to lie on deriving analytical expressions for matrix function derivatives. In that domain this approach is feasible as the matrices involved are rather small (typically 3x3). In density matrix functional theory the matrix function derivatives of interest deal with functions of the density matrix. The dimension of this matrix depends on the number of atoms in the system of interest. For many quantum mechanics applications today the dimensions range from hundreds to tens of thousands. In these applications analytical expressions are of little interest but effective numerical methods are essential.

The commonly used approach to derivatives of matrix functions is the Fréchet derivative. This is essentially a directional derivative of a matrix function given by

$$\lim_{h\to 0} \frac{||F(X+hA) - f(X) - Ah||_W}{||h||_V} \quad = \quad 0 \tag{2}$$

where A is the Fréchet derivative of f. This approach seems to rely on a change of the matrix in one particular direction (although this direction may be chosen arbitrarily). Also this approach can extended to higher order derivatives by recursively applying the same argument. Nevertheless the result is not so easily interpreted.

In this work we approach this problem in a way that is closer to regular function derivatives. Our approach is based on the consideration that Eq.1 is easily expressed for any matrix $\mathbf{X}$ in any basis if the matrix function is F(X) = X. In all cases the result is a fourth order identity matrix. However this particular

matrix function is an exceptionally simple one. In the general case the result is not so easy to formulate. However, in one particular representation of the matrix the result is equally easily formulated.

In the case we represent the matrix $\mathbf{X}$ in its eigenvalue representation, when we will refer to $\mathbf{X}$ as $\mathbf{\Lambda}$ to highlight that this is a diagonal matrix, the first order derivative is

$$\frac{\mathbf{F}_{kl}(\mathbf{\Lambda})}{\Lambda_{ij}} = (1-\delta_{kl})\delta_{ki}\delta_{lj} + \delta_{kl}\delta_{ki}\delta_{lj}\frac{F(\lambda_k)}{\lambda_k} \quad (3)$$

where $\lambda_k = \mathbf{\Lambda}_{kk}$. The first term stems from the fact that the off-diagonal elements have to be considered variables at the value zero, rather than zero constants.

This result is also easily extended to higher order derivatives

$$\frac{\partial^n \mathbf{F}(\mathbf{\Lambda})_{kl}}{\partial \mathbf{\Lambda}_{i_1 j_1} \dots \partial \mathbf{\Lambda}_{i_n j_n}} = \delta_{kl}(\prod_{p=1}^{n} \delta_{ki_p}\delta_{lj_p})F^{(n)}(\lambda_k) \quad (4)$$

Of interest for applications is the fact that the resulting tensors are sparse. The fourth order tensor that represents the first order derivative has only $O(N^2)$ non-zero elements out of a total of $O(N^4)$ tensor elements. In addition of these non-zero elements $N^2 - N$ of them are one and only N of them have a non-trivial value. Derivatives of orders higher than one are $2n+2$ order tensors with $O(N^{2n+2})$ elements of which only $O(N)$ are non-zero. This suggests that the diagonal representation of $\mathbf{X}$ is a particularly convenient basis to work with matrix function derivatives of $\mathbf{X}$ in.

In density matrix functional theory we write the energy expression for the energy of the electrons as

$$E = E_1 + E_C + E_X \quad (5)$$

where E_1 is the 1-electron energy which is the sum of the kinetic energy and the nuclear attraction energy (the positively charged nucleii are assumed stationary but they do attract the negatively charged and moving electrons). The term E_C represents the 2-electron repulsion and is referred to as the Coulomb energy. The term E_X is a negatively valued 2-electron interaction that stems from a purely quantum mechanical effect due to the anti-symmetric nature of the wavefunction. This phenomenon says that the wavefunction changes sign when two electrons are interchanged. The resulting energy term is correspondingly called the exchange term.

Given the 1-electron density matrix $\mathbf{D}$ (a symmetric non-negative matrix for which we have that all $\lambda_m \leq 1$ and that $\sum_m \lambda_m = N$ where N is the number of electrons the various terms are given by

$$E_1 = tr(hD) \quad (6)$$

$$E_C = \sum_{ijkl} D_{ij}(ij|kl)D_{kl} \quad (7)$$

$$E_X = \sum_{ijkl} F(D)_{ij}(ik|jl)F(D)_{kl} \quad (8)$$

Here the expressions $(ij|kl)$ are 2-electron integrals which are given by

$$(ij|kl) \quad = \quad \int\int \chi_i(r_1)\chi_j(r_1)\frac{1}{r_{12}}\chi_k(r_2)\chi_l(r_2)\mathrm{d}r_1\mathrm{d}r_2 \tag{9}$$

where the $\chi(r)$ are normalized basis functions. As you can see the exchange term contains a matrix valued matrix function. Typically $F(X) = X^p$ where $\frac{1}{2} \leq p \leq 1$ [5].

References

[1] Gerd Berghold, Jürg Hutter, and Michele Parrinello. Grid-free DFT implementation of local and gradient-corrected XC functionals. *Theoretical Chemistry Accounts*, 99:344–346, 1998.

[2] Morton E. Gurtin and Kathleen Spear. On the relationship between the logarithmic strain rate and the stretching tensor. *International Journal of Solids and Structures*, 19:437–444, 1983.

[3] Anne Hoger and Donald E. Carlson. On the derivative of the square root of a tensor and guo's rate theorems. *Journal of Elasticity*, 14:329–336, 1984.

[4] C.S. Jog. Derivatives of the stretch, rotation and exponential tensors in n-dimensional vector spaces. *Journal of Elasticity*, 82:175–192, 2006.

[5] N. N. Lathiotakis, S. Sharma, J. K. Dewhurst, F. G. Eich, M. A. L. Marques, and E. K. U. Gross. Density-matrix-power functional: performance for finite systems and the homogeneous electron gas. *Physical Review A*, 79:040501(R), 2009.

[6] Cristina Padovani. On the derivative of some tensor-valued functions. *Journal of Elasticity*, 58:257–268, 2000.