Automatic Differentiation for Complex Valued SVD

Zhou-Quan Wan and Shi-Xin Zhang*
Institute for Advanced Study, Tsinghua University, Beijing 100084, China
(Dated: September 4, 2019)

In this note, we report the back propagation formula for complex valued singular value decompositions (SVD). This formula is an important ingredient for a complete automatic differentiation(AD) infrastructure in terms of complex numbers, and it is also the key to understand and utilize AD in tensor networks.

INTRODUCTION

Automatic differentiation(AD) evaluates derivatives or gradients of any functions specified by computer programs[1]. It is implemented by propagating derivatives of primitive operations via chain rules. Such approach is different from classical symbolic or numerical differentiations. Symbolic differentiation faces the difficulty of converting a complicated computer program into expressions, while numerical differentiation faces the difficulty of numerical errors in the discretization. Besides, both symbolic and numerical methods have problems in calculating higher order derivatives and are also slow at computing gradients with respect to lots of inputs variables, e.g. in the case for gradient-based optimization algorithms. AD solves all of these drawbacks and is emerging as a new programming paradigm which is extensively utilized in machine learning(ML) community and serves as the most important infrastructure for ML libraries.

AD find applications in various other fields as well. Specifically, AD has been applied to computational physics where the interplay between tensor networks and AD has been studied recently[2, 3]. In the context of quantum physics, complex number plays an important role which is often overlooked by ML community. Therefore, although there are many works on AD for real valued linear algebra[4–6], there are very few reports on AD for complex-valued linear algebra[7, 8]. And as mentioned by both [3] and [7], reverse mode AD formula for complex valued SVD is still missing. AD formula for complex valued SVD would allow us to directly utilize AD in tensor network context where tensors are complex valued and operations include SVD. AD knowledge on complex valued SVD is necessary to explore AD on tensor networks for complicated models as well as other scenarios where SVD lies within the forward functions for numerical stability or information extraction.

RESULTS

In this note, we give the formula and derivation of reverse mode AD for complex valued SVD.

For square matrix SVD $A = USV^\dagger$, with the final loss function L being real, the back propagation formula is given by the following:

$$\bar{A} = U\bar{S}V^\dagger + U(\mathcal{J} + \mathcal{J}^\dagger)SV^\dagger + US(\mathcal{K} + \mathcal{K}^\dagger)V^\dagger + \frac{1}{2}US^{-1}(\mathcal{L}^\dagger - \mathcal{L})V^\dagger, \tag{1}$$

where $\mathcal{J} = F \circ (U^\dagger \bar{U}), \mathcal{K} = F \circ (V^\dagger \bar{V}), \mathcal{L} = I \circ (V^\dagger \bar{V})$ and

$$F_{ij} = \begin{cases} \dfrac{1}{s_j^2 - s_i^2} & (i \neq j) \\ 0 & (i = j) \end{cases}$$

The main contribution of this work is the fourth term in (1) as

$$\bar{A}_d = \frac{1}{2}US^{-1}(\mathcal{L}^\dagger - \mathcal{L})V^\dagger. \tag{2}$$

This term is new and has no counterpart in real valued SVD.

DERIVATION

Basics on complex calculus and some notations:

The partial derivatives for function $f(z)$ where $z = x + iy$ is defined as:

$$\frac{\partial f}{\partial z} = \frac{1}{2}(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}), \quad \frac{\partial f}{\partial z^*} = \frac{1}{2}(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}). \tag{3}$$

Such form of derivative is justified as

$$df = \frac{\partial f}{\partial z}dz + \frac{\partial f}{\partial z^*}dz^*. \tag{4}$$

If f is a holomorphic function of z, then $\frac{\partial f}{\partial z^*} = 0$. Only for these holomorphic f, we have well-defined derivatives for complex function as

$$\frac{\mathrm{d}f}{\mathrm{d}z} = \frac{\partial f}{\partial z}. \tag{5}$$

In this note, we only focus on real valued functions depending on complex variables. This is the case as we can only optimize a real value; optimization problem makes no sense for complex valued object. Besides, for physics settings, all observables give real valued results. Note real valued f cannot be holomorphic unless f is a constant independent of z.

For optimization problems and gradient descent approach, what we really care about is gradients instead of derivatives. For real variable x and real function $f(x)$, it is coincident that $\nabla_x f = \frac{\partial f}{\partial x}$. However this relation doesn't hold for complex variable functions. Instead, for complex variable functions, the gradient is defined as

$$\nabla_z f(z) = 2\frac{\partial f}{\partial z^*} = \frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}. \tag{6}$$

If f is real valued function, we further have $\nabla_z f = 2(\frac{\partial f}{\partial z})^*$. See [9] for more details on the calculus of complex and matrix based functions.

In the following, we only focus on reverse mode AD only. We use L to represent the final real valued loss function, and use the following notation for derivatives and gradients:

$$\bar{A} = \frac{\partial L}{\partial A} \quad \bar{\mathcal{A}} = 2\frac{\partial L}{\partial A^*} = \nabla_A L. \tag{7}$$

$\circ$ stands for elements multiplication between two matrices. The following facts will be utilized in the proof below:

$$\mathrm{Tr}[A(C \circ B)] = \mathrm{Tr}[(C^T \circ A)B], \tag{8}$$

$$(A \circ B)^T = A^T \circ B^T. \tag{9}$$

I denotes the identity matrix while $\bar{I}$ represents a square matrix with diagonal zero and all off-diagonal elements 1.

Parallel proof with real case:

For the first three terms in (1), we follow the similar proof for real valued SVD with special attention on unique features for complex functions. The SVD is defined as:

$$A = USV^\dagger, \tag{10}$$

where A is the input matrix while U, S, V is the output matrix. S only has positive real elements in the main diagonal, and U, V are unitary matrix: $U^\dagger U = I, V^\dagger V = I$.

A direct differentiation on (10) gives:

$$dA = dUSV^\dagger + UdSV^\dagger + USdV^\dagger. \tag{11}$$

Let $dC = U^\dagger dU, dD = V^\dagger dV$, according to the unitary property of U and V, we have

$$dC = -dC^\dagger \quad dD = -dD^\dagger. \tag{12}$$

(12) indicates that the diagonal elements in dC and dD are **pure imiginary**. (not zero!) Plug dC and dD into (11), we have

$$dP = dCS + dS - SdD, \tag{13}$$

where $dP = U^\dagger dA V$.

We have several relations from real and imaginary diagonal part as well as off-diagonal part from (13), respectively. (Note that S and dS is real diagonal matrix and dC, dD have pure imaginary diagonal elements.)

$$dS = I \circ (\frac{dP + dP^\dagger}{2}), \tag{14}$$

$$I \circ (dC - dD) = (I \circ (\frac{dP - dP^\dagger}{2}))S^{-1}, \tag{15}$$

$$\bar{I} \circ (dPS + SdP^\dagger) = dCS^2 - S^2 dC, \tag{16}$$

$$\bar{I} \circ (SdP + dP^\dagger S) = dDS^2 - S^2 dD. \tag{17}$$

Our proof is based on the following relation given by chain rules:

$$dL = \mathrm{Tr}[\bar{U}^T dU + \bar{V}^T dV + \bar{S}^T dS + h.c.] = \mathrm{Tr}[\bar{A}^T dA + h.c.] \tag{18}$$

and our aim is to find the analytical formula for $\bar{A}$ in terms of U, S, V and $\bar{U}, \bar{S}, \bar{V}$. dL in (18) can be decomposed as

$$\begin{aligned} dL &= \mathrm{Tr}[\bar{S}^T dS + \bar{U}^T U(I \circ dC) + \bar{U}^T U(\bar{I} \circ dC)] + \bar{V}^T V(I \circ dD) + \bar{V}^T V(\bar{I} \circ dD) + h.c.] &(19)\\ &= \mathrm{Tr}[\bar{A}_s^T dA + \bar{A}_{ud}^T dA + \bar{A}_{uo}^T dA + \bar{A}_{vd}^T dA + \bar{A}_{vo}^T dA + h.c.]. &(20)\end{aligned}$$

For $\bar{A}_s$ part, we have:

$$\mathrm{Tr}[\bar{S}^T(I \circ dP) + h.c.] = \mathrm{Tr}[\bar{S}^T U^\dagger dA V + h.c.] = \mathrm{Tr}[V\bar{S}^T U^\dagger dA + h.c.]. \tag{21}$$

Namley,

$$\bar{A}_s = U^* \bar{S} V^T. \tag{22}$$

For $\bar{A}_{uo}$ part, note $\bar{I} \circ dC = F \circ (dPS + SdP^\dagger)$, we have:

$$\begin{aligned} &\mathrm{Tr}[\bar{U}^T U(\bar{I} \circ dC) + h.c.] = \mathrm{Tr}[\bar{U}^T U F \circ (dPS + SdP^\dagger) + h.c.] = &(23)\\ &= \mathrm{Tr}[(-F \circ \bar{U}^T U)(U^\dagger dA V S + S V^\dagger dA^\dagger U) + h.c.] = \mathrm{Tr}[V S J^T U^\dagger dA + V S J^* U^\dagger dA + h.c.], &(24)\end{aligned}$$

where $J = F \circ (U^T \bar{U})$. From this, we conclude that

$$\bar{A}_{uo} = U^*(J + J^\dagger) S V^T. \tag{25}$$

Similarly, we have:

$$\bar{A}_{vo} = U^* S (K + K^\dagger) V^T, \tag{26}$$

where $K = F \circ (V^T \bar{V})$.

The three terms (22), (25) and (26) as well as the proof above are very similar with real valued SVD. The difference here is that dC and dD have pure imaginary diagonals instead of zero diagonals. Consequently, there are two more differentiation terms for dA, corresponding to dA_{ud} and dA_{vd}, originating from the differentiation of diagonal part in dC and dD. The two terms are totally new in complex context, and we will explore how to compute them in the following sections.

Gauge freedom:

There are some extra freedom for the choice of U and V in SVD as we can see from the following relation:

$$A = USV^\dagger = U\Lambda\Lambda^\dagger S \Lambda\Lambda^\dagger V^\dagger = U' S V'^\dagger, \tag{27}$$

where $U' = U\Lambda, V' = V\Lambda$, Λ is a diagonal matrix with diagonal elements in the form of a phase $e^{i\lambda}$. Therefore, the loss function L must be gauge invariant, i.e. $L(U, S, V) = L(U\Lambda, S, V\Lambda)$.

We can fix the gauge freedom for U and V, say, all diagonal elements of U are positive real. In this gauge fix scheme, we define $U = u(A), V = v(A), S = s(A)$. And we can introduce $\Lambda(A)$ describing any other gauge fix sheme as $U' = u(A)\Lambda(A)$. For given input matrix A, we consider any gauge scheme where $\Lambda(A) = I$.

$$\begin{aligned} dL(U,S,V) &= \mathrm{Tr}[\bar{S}^T dS + \bar{U}^T dU + \bar{V}^T dV] \\ &= \mathrm{Tr}[\bar{S}^T dS + \bar{U}^T d(U\Lambda) + \bar{V}^T d(V\Lambda)]. \end{aligned} \tag{28}$$

Therefore, we have

$$\mathrm{Tr}[(\bar{U}^T U + \bar{V}^T V)d\Lambda] = 0. \tag{29}$$

$d\Lambda$ can be any diagonal matrix based on the function form of $\Lambda(A)$, we have obtained the important equation (29) from gauge invariance.

New differentiation term in complex case:

Now considering the remaining term $\bar{A}_{ud} + \bar{A}_{vd}$, we have:

$$\mathrm{Tr}[\bar{U}^T U(I \circ dC) + \bar{V}^T V(I \circ dD)] = \mathrm{Tr}[(\bar{U}U + \bar{V}V)(I \circ dC)] + \mathrm{Tr}[\bar{V}^T V(I \circ (dD - dC))]. \tag{30}$$

Note the first term on rhs is zero due to gauge invariance (29), the remaining one can be replaced by (15):

$$\mathrm{Tr}[(\bar{A}_{ud} + \bar{A}_{vd})^T dA] = \frac{1}{2}\mathrm{Tr}[\bar{V}^T V(I \circ (dP^\dagger - dP)S^{-1}) + h.c.] = \frac{1}{2}\mathrm{Tr}[V(L^* S^{-1} - S^{-1}L^T)U^\dagger dA + h.c.]. \tag{31}$$

where $L = I \circ (V^T \bar{V})$.

From this, we have:

$$\bar{A}_d \equiv \bar{A}_{ud} + \bar{A}_{vd} = \frac{1}{2}U^*(L^\dagger S^{-1} - S^{-1}L)V^T. \tag{32}$$

Now we have derived all contributions to $\bar{A}$, the sum of (22), (25), (26) and (32) gives the back propagation formula of derivatives for complex valued SVD.

In practice, our aim is to obtain gradients of L, since $\bar{\mathcal{A}} = 2\bar{A}^*$ for real L, we can easily transform the relation for derivatives to the one for gradients, and the final result is (1) [10]. Note how $\mathcal{J}, \mathcal{K}, \mathcal{L}$ changed accordingly from J, K, L.

DISCUSSIONS

The results can directly apply to SVD on rectangular matrix A, with two extra terms similar to the real SVD case, namely

$$\bar{\mathcal{A}} = \bar{\mathcal{A}}_{square} + (I - UU^\dagger)\bar{\mathcal{U}}S^{-1}V^\dagger + US^{-1}\bar{\mathcal{V}}^\dagger(1 - VV^\dagger), \tag{33}$$

where $\bar{\mathcal{A}}_{square}$ is defined as (1). In other words, for general complex valued SVD, we only need to add one extra term beyond real SVD case, which is just $\bar{\mathcal{A}}_d$ in (2).

It is worth noting that if loss function has no dependence on U or V, i.e. $L = L(U,S)$ or $L = L(V,S)$, the contribution $\bar{A}_d = 0$ following (29). In other words, to test the correctness of backprop for complex valued SVD, the loss function must depend on U and V at the time. Otherwise, the original formula for SVD without $\bar{A}_d$ still works. Therefore, one must devise a test loss function L, which is real valued, gauge invariant, and cannot decouple to two gauge invariant parts depending on U, S and S, V only. An example loss function might be something like $\mathcal{R}(U_{00}V_{00}^*)$.

The backprop formula for complex valued SVD has great potential for applications in various fields ranged from machine learning to computational physics. For example, AD for complexed value SVD may advance the development of relevant optimization algorithms on tensor networks[2].

Acknowledgement: We thank Jin-Guo Liu for helpful discussions on this topic.

* zsx16@mails.tsinghua.edu.cn

[1] Michael Bartholomew-Biggs, Steven Brown, Bruce Christianson, and Laurence Dixon, J. Comput. Appl. Math. 124, 171190 (2000).
[2] Hai-Jun Liao, Jin-Guo Liu, Lei Wang, and Tao Xiang, arXiv:1903.09650.
[3] Giacomo Torlai, Juan Carrasquilla, Matthew T. Fishman, Roger G. Melko, and Matthew P. A. Fisher, arXiv:1906.04654.
[4] Mike Giles, An extended collection of matrix derivative results for forward and reverse mode algorithmic differentiation, Tech. Rep. (2008).
[5] James Townsend, Differentiating the Singular Value Decomposition, Tech. Rep. (2016).
[6] Matthias Seeger, Asmus Hetzel, Zhenwen Dai, Eric Meissner, and Neil D. Lawrence, arXiv:1710.08717.
[7] Claudius Hubig, arXiv:1907.13422.
[8] Jin-Guo Liu, Linear Algebra Autodiff (complex valued) (2019).
[9] R. Hunger, An introduction to complex differentials and complex differentiability, Tech. Rep. (2007).
[10] See https://github.com/Zhouquan-Wan/SVD_autodiff for a possible implementation in Tensorflow.