## $3.\ \text{What}'\text{s difficult to apply AD}$

1. svd

(1) Complex derivative

Natutal way:

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

When $f$ is a real value function, it's easy to get that $df$ is real sa well, that is to say:

$$df = \frac{1}{2}(u'_x - iu'_y)dz + c.c$$

(2) Complex gradient

How define complex gradient for a function? Denote a complex linear function:

$$\nabla_z f(z) = R(z) + iI(z)$$

When $f(z) = u(z)$, we hope that the complex gradient is the direction of the fastest increase of $u$, which is to say:

$$\nabla u(z) = u'_x + iu'_y$$

Now we consider analytic functions:

$$\nabla f(z) = \nabla u(z) + i\nabla v(z)$$

$$= u'_x + iu'_y + i(v'_x + iv'_y)$$

$$= u'_x + iu'_y + i(-u'_y + iu'_x) = 0$$

So we can define $\nabla f(z)$ on $\mathbb{H}(\mathbb{C})$ as

$$\nabla f(z) = k\frac{\partial f}{\partial z^*}$$

And consider $\nabla u(z)$ we get $k = 2$.

Continue this to all complex function, we get:

$$\nabla f(z) = 2\frac{\partial f}{\partial z^*}$$

Specifically, when $f(z)$ is a real value function, we have:

$$\nabla f(z) = 2\left(\frac{\partial f}{\partial z}\right)^*$$

(3) Complex derivative of on $\mathbb{C}^n$
When $f : \mathbb{C}^n \to \mathbb{C}$, we can define its complex derivative on $\mathbb{C}^n$ as:

$$\frac{\partial f}{\partial Z} = \left(\frac{\partial f}{\partial z_i}\right)_{1\le i\le n}$$

$$\nabla f(Z) = 2\left(\frac{\partial f}{\partial z_i^*}\right)_{1\le i\le n}$$

Here we have clear that for

$$Z = (z_{ij})_{1\le i\le n, 1\le j\le m}$$

We define $\frac{\partial f}{\partial Z}$ as:

$$\left(\frac{\partial f}{\partial z_{ij}}\right)_{ij}$$

But not

$$\left(\frac{\partial f}{\partial z_{ji}}\right)_{ij}$$

This also means that if $Z$ is a colunm (row) vector, $\frac{\partial f}{\partial Z}$ is also a row (colunm) vector.

(4) Introduction to AD for svd
Denote

$$\bar{A} = \frac{\partial Loss}{\partial A}$$

Assume that $L(A)$ is a real gauge loss function. Gauge means for a svd composition

$$A = USV^\dagger$$

$$Loss(A) = Loss(U, S, V)$$

has nothing with the choose of $U, V$.
And assume that $A$ is a matrix has no zero or same eigenvalue.
Then we have:

$$dLoss = \mathrm{Tr}(\bar{A}^T dA + c.c)$$

$$\Longrightarrow \nabla f(A) = 2(\bar{A})^* = 2\left(A_s + A_J + A_K + A_O\right)^*$$

$$\begin{aligned} A_s^* &= U(\bar{S})^* V^\dagger \\ A_J^* &= U(J^* + J^T)SV^\dagger \\ A_K^* &= US(K^* + K^T)V^\dagger \\ A_O^* &= \frac{1}{2}US^{-1}(O - O^\dagger)V^\dagger \end{aligned}$$

$$\begin{aligned} J &= F \circ (U^T \bar{U}) \\ K &= F \circ (V^T \bar{V}) \\ O &= I \circ (V^T \bar{V}) \end{aligned}$$

$$F = \frac{1}{s_j^2 - s_i^2}\chi_{i \neq j}$$

In formulas above, $\circ$ is:

$$(a_{ij})_{n\times n} \circ (b_{ij})_{n\times n} = (a_{ij}b_{ij})_{n\times n}$$

$I$ is identity matrix.

(5) Theoretical validation of method effectiveness.

(a) Gauge freedom

Arbitrary given

$$\mathrm{diag}\{e^{i\theta_1}, .., e^{i\theta_m}\},\ V = [v_1, .., v_m]$$

$$\Longrightarrow V\mathrm{diag} = [.., e^{i\theta_m} v_m]$$

$$\Longrightarrow \|\mathrm{Bary}(v_m e^{i\theta_m}) - A\|_2^2 = \|\mathrm{Bary}(v_m) - A\|_2^2 = Loss$$

(b) Different eigenvalues

In practice, input green function values take some noise and therefore we can assume that all submatrices of $L$ without zero eigenvalues are non-singular with probability 1.

(c) non-zero eigenvalues

Denote size of $L$ is $N \times N$ and size of $L_{sub}$ is $(N - m) \times m$.

For equation:

$$L_{sub} w = G_{sub}$$

In practice, G_{sub} has noise so if $m \leq N/2$ and not full column rank, this equation has no solution with probability 1.
So if $m \leq N/2$, we can think that $L_{sub}$ is full column and therefore has no zero eigenvalue.