First order Taylor expansion - Part I

Tiep M. Hoang

I. Complex Partial Derivative

Let $\mathbf{z} = \mathbf{x} + j\mathbf{y}$ be a column vector with N complex-valued elements. Let $\mathbf{z}^*$ be the conjugate vector of $\mathbf{z}$, then $\mathbf{z}^* = \mathbf{x} - j\mathbf{y}$. Let $\mathbf{z}^\dagger$ be the conjugate transpose vector of $\mathbf{z}$, then $\mathbf{z}^\dagger = (\mathbf{z}^*)^\top$. We will write $f(\mathbf{z}, \mathbf{z}^*)$ to denote a function $f(\cdot)$ with respect to $\mathbf{z}$ and $\mathbf{z}^*$. For example, $\|\mathbf{z}\|^2 = \mathbf{z}^\dagger \mathbf{z} = (\mathbf{z}^*)^\top \mathbf{z}$ can be written as $f(\mathbf{z}, \mathbf{z}^*)$ instead of $f(\mathbf{z})$. This notation will allow us to remember that $f(\mathbf{z}, \mathbf{z}^*)$ is analytic both on $\{\mathbf{z} | \mathbf{z} \in \mathbb{C}^{N\times 1}\}$ and $\{\mathbf{z}^* | \mathbf{z}^* \in \mathbb{C}^{N\times 1}\}$.

The *complex partial derivatives* include the following:

$$\left.\frac{\partial f(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}}\right|_{\mathbf{z}^*=\text{const}}, \quad \left.\frac{\partial f(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}^*}\right|_{\mathbf{z}=\text{const}}, \quad \left.\frac{\partial f(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}^\top}\right|_{\mathbf{z}^*=\text{const}}, \quad \left.\frac{\partial f(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}^\dagger}\right|_{\mathbf{z}=\text{const}}$$

For simplicity, we define

$$\nabla_{\mathbf{z}} f(\mathbf{z}, \mathbf{z}^*) = \left.\frac{\partial f(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}}\right|_{\mathbf{z}^*=\text{const}} \text{ and } \nabla_{\mathbf{z}^*} f(\mathbf{z}, \mathbf{z}^*) = \left.\frac{\partial f(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}^*}\right|_{\mathbf{z}=\text{const}}. \tag{1}$$

Also, we define

$$D_{\mathbf{z}} f(\mathbf{z}, \mathbf{z}^*) = \left.\frac{\partial f(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}^\top}\right|_{\mathbf{z}^*=\text{const}} \text{ and } D_{\mathbf{z}^*} f(\mathbf{z}, \mathbf{z}^*) = \left.\frac{\partial f(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}^\dagger}\right|_{\mathbf{z}=\text{const}}.$$

The relationship between $\nabla_{(\cdot)} f(\mathbf{z}, \mathbf{z}^*)$ and $D_{(\cdot)} f(\mathbf{z}, \mathbf{z}^*)$ is as follows:

$$\nabla_{\mathbf{z}} f(\mathbf{z}, \mathbf{z}^*) = \Big(D_{\mathbf{z}} f(\mathbf{z}, \mathbf{z}^*)\Big)^\top \text{ and } \nabla_{\mathbf{z}^*} f(\mathbf{z}, \mathbf{z}^*) = \Big(D_{\mathbf{z}^*} f(\mathbf{z}, \mathbf{z}^*)\Big)^\top. \tag{2}$$

Following [1, Section 3.4], $\nabla_{\mathbf{z}} f(\mathbf{z}, \mathbf{z}^*)$ and $\nabla_{\mathbf{z}^*} f(\mathbf{z}, \mathbf{z}^*)$ are the *complex gradient vector* and the *complex conjugate gradient vector* of the function $f(\mathbf{z}, \mathbf{z}^*)$. On the other hand, $D_{\mathbf{z}} f(\mathbf{z}, \mathbf{z}^*)$ and $D_{\mathbf{z}^*} f(\mathbf{z}, \mathbf{z}^*)$ are the complex *cogradient* vector and the complex conjugate *cogradient* vector.

The relationship between the complex *differential* and the complex partial derivative is as follows [1, page 163]:

$$\begin{aligned} df(\mathbf{z}, \mathbf{z}^*) &= D_{\mathbf{z}} f(\mathbf{z}, \mathbf{z}^*) \times d\mathbf{z} + D_{\mathbf{z}^*} f(\mathbf{z}, \mathbf{z}^*) \times d\mathbf{z}^* \\ &= \frac{\partial f(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}^\top} d\mathbf{z} + \frac{\partial f(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}^\dagger} d\mathbf{z}^* \\ &= \Big(\nabla_{\mathbf{z}} f(\mathbf{z}, \mathbf{z}^*)\Big)^\top d\mathbf{z} + \Big(\nabla_{\mathbf{z}^*} f(\mathbf{z}, \mathbf{z}^*)\Big)^\top d\mathbf{z}^* \end{aligned} \tag{3}$$

TABLE I

$f(\mathbf{Z},\mathbf{Z}^*)$ AND ITS DIFFERENTIAL $df(\mathbf{Z},\mathbf{Z}^*)$

$f(\mathbf{Z},\mathbf{Z}^*)$	$df(\mathbf{Z},\mathbf{Z}^*)$	Complex gradient matrix $\nabla_{\mathbf{Z}} f(\mathbf{Z},\mathbf{Z}^*)$	Complex conjugate gradient matrix $\nabla_{\mathbf{Z}^*} f(\mathbf{Z},\mathbf{Z}^*)$
$\operatorname{tr}(\mathbf{A}\mathbf{Z})$	$\operatorname{tr}(\mathbf{A}d\mathbf{Z})$	$\mathbf{A}^\top$	$\mathbf{0}$
$\operatorname{tr}(\mathbf{A}\mathbf{Z}^\dagger)$	$\operatorname{tr}(\mathbf{A}^\top d\mathbf{Z}^*)$	$\mathbf{0}$	$\mathbf{A}$
$\operatorname{tr}(\mathbf{Z}\mathbf{A}\mathbf{Z}^\top)$	$\operatorname{tr}((\mathbf{A}+\mathbf{A}^\top)\mathbf{Z}^\top d\mathbf{Z})$	$\mathbf{Z}(\mathbf{A}^\top+\mathbf{A})$	$\mathbf{0}$
$\operatorname{tr}(\mathbf{Z}\mathbf{A}\mathbf{Z}^\dagger)$	$\operatorname{tr}(\mathbf{A}\mathbf{Z}^\dagger d\mathbf{Z}+\mathbf{A}^\top\mathbf{Z}^\top d\mathbf{Z}^*)$	$\mathbf{Z}^*\mathbf{A}^\top$	$\mathbf{Z}\mathbf{A}$

TABLE II

$f(\mathbf{z},\mathbf{z}^*)$ AND ITS DIFFERENTIAL $df(\mathbf{z},\mathbf{z}^*)$, GIVEN THAT $\mathbf{z}$ IS A COLUMN VECTOR

$f(\mathbf{z},\mathbf{z}^*)$	$df(\mathbf{z},\mathbf{z}^*) \approx f(\mathbf{z},\mathbf{z}^*) - f(\mathbf{z}_0,\mathbf{z}_0^*)$	Complex grad. vector at $\mathbf{z}_0$ $\nabla_{\mathbf{z}_0} f(\mathbf{z}_0,\mathbf{z}_0^*)$	Complex conj. grad. vector at $\mathbf{z}_0^*$ $\nabla_{\mathbf{z}_0^*} f(\mathbf{z}_0,\mathbf{z}_0^*)$
$\mathbf{A}\mathbf{z}$	$\mathbf{A}d\mathbf{z}$	$\mathbf{A}^\top$	$\mathbf{0}$
$\mathbf{A}\mathbf{z}^*$	$\mathbf{A}d\mathbf{z}^*$	$\mathbf{0}$	$\mathbf{A}^\top$
$\operatorname{tr}(\mathbf{z}^\top\mathbf{A}\mathbf{z})$	$\operatorname{tr}((\mathbf{A}+\mathbf{A}^\top)\mathbf{z}_0\ d\mathbf{z}^\top)$ $=\mathbf{z}_0^\top(\mathbf{A}+\mathbf{A}^\top)d\mathbf{z}$	$(\mathbf{A}^\top+\mathbf{A})\mathbf{z}_0$	$\mathbf{0}$
$\operatorname{tr}(\mathbf{z}^\dagger\mathbf{A}\mathbf{z})$	$\operatorname{tr}(\mathbf{A}\mathbf{z}_0\ d\mathbf{z}^\dagger+\mathbf{A}^\top\mathbf{z}_0^* d\mathbf{z}^\top)$ $=\mathbf{z}_0^\dagger\mathbf{A}d\mathbf{z}+\mathbf{z}_0^\top\mathbf{A}^\top d\mathbf{z}^*$	$\mathbf{A}^\top\mathbf{z}_0^*$	$\mathbf{A}\mathbf{z}_0$

Once we have the explicit expression of the complex differential, we can find $D_{\mathbf{z}} f(\mathbf{z},\mathbf{z}^*)$ and $D_{\mathbf{z}^*} f(\mathbf{z},\mathbf{z}^*)$. Then, using (2), we can find $\nabla_{\mathbf{z}} f(\mathbf{z},\mathbf{z}^*)$ and $\nabla_{\mathbf{z}^*} f(\mathbf{z},\mathbf{z}^*)$, which will be used to perform the Taylor series approximation.

In Table-I, we see that $\mathbf{Z}$ is a matrix but not just a vector. This table is extracted from Table 3.9 on page 170 of the book [1]. In general, Table-I holds true with $\mathbf{Z}$ being a (complex) matrix.

Interestingly, we can look up many widely-used gradient vectors in [2]. That would save a lot of time instead of deriving them from scratch.

II. First order Taylor expansion

In this section, we will consider $f(\mathbf{z},\mathbf{z}^*)$ instead of $f(\mathbf{Z},\mathbf{Z}^*)$. Moreover, we narrow down the topic to a **special** case, where $f(\mathbf{z},\mathbf{z}^*)$ is a **real** function.

A. Approximation

We consider $\mathbf{z},\mathbf{z}_0 \in \mathbb{C}^{N\times 1}$ so that $\mathbf{z}$ is close to $\mathbf{z}_0$. Using the first-order Taylor expansion, we can approximate $f(\mathbf{z},\mathbf{z}^*)$ as follows:

$$\begin{aligned} f(\mathbf{z},\mathbf{z}^*) &\approx f(\mathbf{z}_0,\mathbf{z}_0^*) + \left[\Big(\nabla_{\mathbf{z}_0} f(\mathbf{z}_0,\mathbf{z}_0^*)\Big)^\top,\ \Big(\nabla_{\mathbf{z}_0^*} f(\mathbf{z}_0,\mathbf{z}_0^*)\Big)^\top\right] \begin{bmatrix} (\mathbf{z}-\mathbf{z}_0) \\ (\mathbf{z}^*-\mathbf{z}_0^*) \end{bmatrix} \\ &= f(\mathbf{z}_0,\mathbf{z}_0^*) + \Big(\nabla_{\mathbf{z}_0} f(\mathbf{z}_0,\mathbf{z}_0^*)\Big)^\top (\mathbf{z}-\mathbf{z}_0) + \Big(\nabla_{\mathbf{z}_0^*} f(\mathbf{z}_0,\mathbf{z}_0^*)\Big)^\top (\mathbf{z}^*-\mathbf{z}_0^*). \end{aligned} \tag{4}$$

As aforementioned, the symbol "$\approx$" is for $\|\mathbf{z}-\mathbf{z}_0\| \to 0$ or at least $\|\mathbf{z}-\mathbf{z}_0\| \le \epsilon$.

B. Special case: $f(\mathbf{z},\mathbf{z}^*)$ is convex

Given that $f(\mathbf{z},\mathbf{z}^*)$ is convex, we can use the first-order Taylor expansion to find its **lower bound** as follows:

$$\begin{aligned} f(\mathbf{z},\mathbf{z}^*) &\ge f(\mathbf{z}_0,\mathbf{z}_0^*) + \left[\Big(\nabla_{\mathbf{z}_0} f(\mathbf{z}_0,\mathbf{z}_0^*)\Big)^\top,\ \Big(\nabla_{\mathbf{z}_0^*} f(\mathbf{z}_0,\mathbf{z}_0^*)\Big)^\top\right] \begin{bmatrix} (\mathbf{z}-\mathbf{z}_0) \\ (\mathbf{z}^*-\mathbf{z}_0^*) \end{bmatrix} \\ &= f(\mathbf{z}_0,\mathbf{z}_0^*) + \Big(\nabla_{\mathbf{z}_0} f(\mathbf{z}_0,\mathbf{z}_0^*)\Big)^\top (\mathbf{z}-\mathbf{z}_0) + \Big(\nabla_{\mathbf{z}_0^*} f(\mathbf{z}_0,\mathbf{z}_0^*)\Big)^\top (\mathbf{z}^*-\mathbf{z}_0^*), \end{aligned} \tag{5}$$

where both $\mathbf{z}$ and $\mathbf{z}_0$ are arbitrary vectors in $\mathbb{C}^{N\times 1}$. Noticeably, $\mathbf{z}$ is *NOT* necessarily close to $\mathbf{z}_0$, we still obtain (5) as long as $f(\mathbf{z},\mathbf{z}^*)$ is convex. The equality "=" holds for $\mathbf{z}=\mathbf{z}_0$. In general, when $\|\mathbf{z}-\mathbf{z}_0\| \to 0$, the right-hand side of (5) is a closely-tight lower bound of $f(\mathbf{z},\mathbf{z}^*)$.

Example 1. *Let us consider* $f(\mathbf{z},\mathbf{z}^*) = \operatorname{tr}\left(\mathbf{z}^\top \mathbf{A}\mathbf{z}\right)$*. Then,* $f(\mathbf{z},\mathbf{z}^*)$ *is convex if* $\mathbf{A} \succcurlyeq \mathbf{0}$ *(i.e.,* $\mathbf{A}$ *is positive semidefinte) and* $\mathbf{z} = \mathbf{z}^* \in \mathbb{R}^{N\times 1}$ *(i.e., the imaginary part is zero). From Table-II, we see that the complex gradient vector is* $\nabla_{\mathbf{z}_0} f(\mathbf{z}_0,\mathbf{z}_0^*) = (\mathbf{A}^\top + \mathbf{A})\mathbf{z}_0$*, while the complex conjugate gradient vector is* $\nabla_{\mathbf{z}_0^*} f(\mathbf{z}_0,\mathbf{z}_0^*) = \mathbf{0}$*. With* $f(\mathbf{z},\mathbf{z}^*)$ *being convex, we can use* (5) *to arrive at*

$$\begin{aligned}
\operatorname{tr}\left(\mathbf{z}^\top \mathbf{A}\mathbf{z}\right) &\geq \operatorname{tr}\left(\mathbf{z}_0^\top \mathbf{A}\mathbf{z}_0\right) + [\,\underbrace{(\mathbf{A}^\top + \mathbf{A})\mathbf{z}_0}_{=\nabla_{\mathbf{z}_0} f(\mathbf{z}_0,\mathbf{z}_0^*)}\,]^\top \underbrace{(\mathbf{z}-\mathbf{z}_0)}_{=d\mathbf{z}_0} + [\quad \underbrace{\mathbf{0}}_{=\nabla_{\mathbf{z}_0^*} f(\mathbf{z}_0,\mathbf{z}_0^*)} \quad]^\top \underbrace{(\mathbf{z}^*-\mathbf{z}_0^*)}_{=d\mathbf{z}_0^*} \\
&= \mathbf{z}_0^\top \mathbf{A}\mathbf{z}_0 + \mathbf{z}_0^\top (\mathbf{A}^\top + \mathbf{A})(\mathbf{z}-\mathbf{z}_0) \\
&= \mathbf{z}_0^\top (\mathbf{A}^\top + \mathbf{A})\mathbf{z} - \mathbf{z}_0^\top \mathbf{A}^\top \mathbf{z}_0 \qquad (6)
\end{aligned}$$

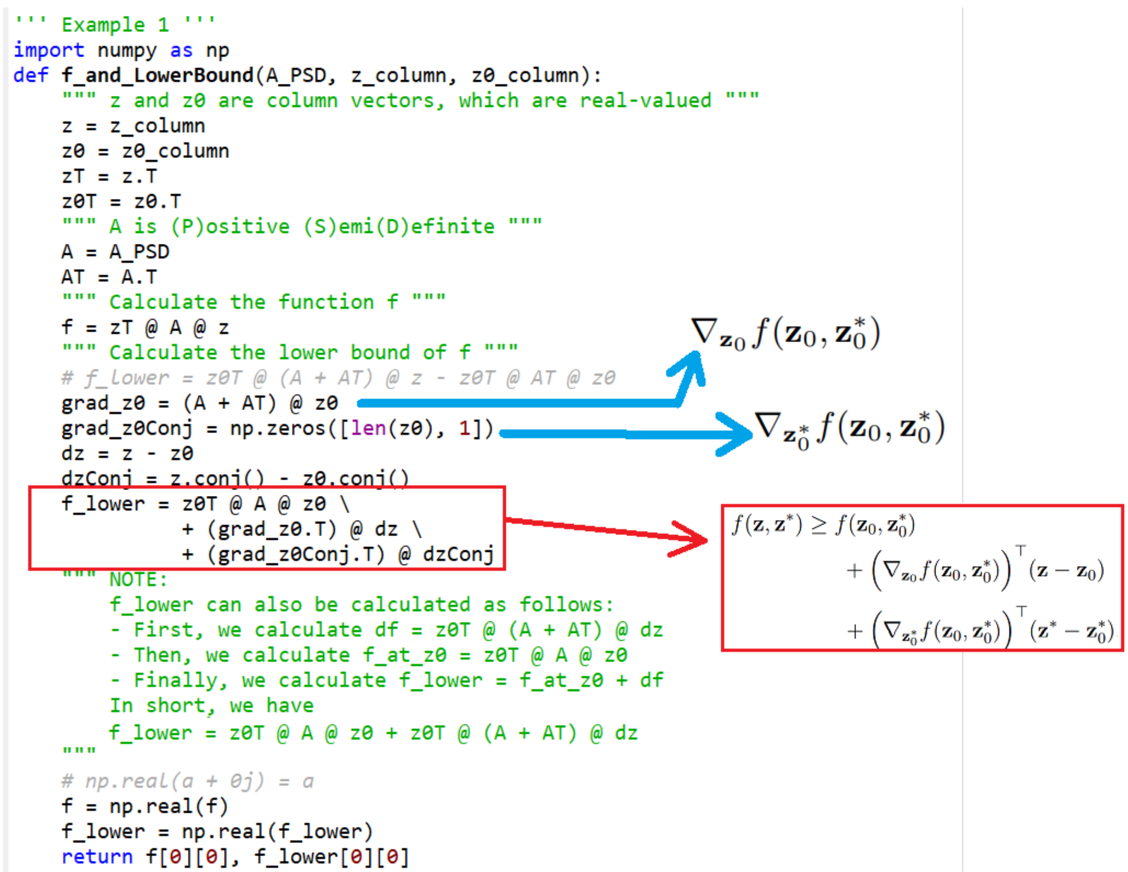

Fig. 1. A code snippet in the file “first-order-Taylor–ex1.py”.

Example 2. *Let us consider* $f(\mathbf{z}, \mathbf{z}^*) = \operatorname{tr}\left(\mathbf{z}^\dagger \mathbf{A}\mathbf{z}\right)$*. Then,* $f(\mathbf{z}, \mathbf{z}^*)$ *is convex if* $\mathbf{A} \succcurlyeq \mathbf{0}$*. From Table-II, we have* $\nabla_{\mathbf{z}_0} f(\mathbf{z}_0, \mathbf{z}_0^*) = \mathbf{A}^\top \mathbf{z}_0^*$ *and* $\nabla_{\mathbf{z}_0^*} f(\mathbf{z}_0, \mathbf{z}_0^*) = \mathbf{A}\mathbf{z}_0$*. Consequently, we apply the first-order Taylor expansion to the convex function* $f(\mathbf{z}, \mathbf{z}^*)$ *to obtain the following inequality:*

$$
\begin{aligned}
\operatorname{tr}\left(\mathbf{z}^\dagger \mathbf{A}\mathbf{z}\right) &\geq \operatorname{tr}\left(\mathbf{z}_0^\dagger \mathbf{A}\mathbf{z}_0\right) + [\underbrace{\quad \mathbf{A}^\top \mathbf{z}_0^* \quad}_{=\nabla_{\mathbf{z}_0} f(\mathbf{z}_0, \mathbf{z}_0^*)}]^\top \underbrace{(\mathbf{z} - \mathbf{z}_0)}_{=d\mathbf{z}_0} + [\underbrace{\quad \mathbf{A}\mathbf{z}_0 \quad}_{=\nabla_{\mathbf{z}_0^*} f(\mathbf{z}_0, \mathbf{z}_0^*)}]^\top \underbrace{(\mathbf{z}^* - \mathbf{z}_0^*)}_{=d\mathbf{z}_0^*} \\
&= \mathbf{z}_0^\dagger \mathbf{A}\mathbf{z}_0 + \mathbf{z}_0^\dagger \mathbf{A}(\mathbf{z} - \mathbf{z}_0) + \mathbf{z}_0^\top \mathbf{A}^\top (\mathbf{z}^* - \mathbf{z}_0^*) \\
&= \mathbf{z}_0^\dagger \mathbf{A}\mathbf{z} + \mathbf{z}_0^\top \mathbf{A}^\top \mathbf{z}^* - \mathbf{z}_0^\top \mathbf{A}^\top \mathbf{z}_0^*. \qquad (7)
\end{aligned}
$$

```python
''' Example 2 '''
import numpy as np
def f_and_LowerBound(A_PSD, z_column, z0_column):
    """ z and z0 are column vectors """
    z = z_column
    z0 = z0_column
    z0T = z0.T
    zH = z.conj().T
    z0H = z0.conj().T
    """ A is (P)ositive (S)emi(D)efinite """
    A = A_PSD
    AT = A.T
    """ Calculate the function f """
    f = zH @ A @ z
    """ Calculate the lower bound of f """
    grad_z0 = AT @ (z0.conj())
    grad_z0Conj = A @ z0
    dz = z - z0
    dzConj = z.conj() - z0.conj()
    f_lower = z0H @ A @ z0 \
            + (grad_z0.T) @ dz \
            + (grad_z0Conj.T) @ dzConj
    """ NOTE:
    f_lower can also be calculated as follows:
    - First, we calculate df = z0H @ A @ dz + z0T @ AT @ (dz.conj())
    - Then, we calculate f_at_z0 = z0H @ A @ z0
    - Finally, we have f_lower = f_at_z0 + df
    """
    f = np.real(f)  # np.real(a + 0j) = a
    f_lower = np.real(f_lower)  # np.real(a + 0j) = a
    return f[0][0], f_lower[0][0]
```

$\nabla_{\mathbf{z}_0} f(\mathbf{z}_0, \mathbf{z}_0^*)$

$\nabla_{\mathbf{z}_0^*} f(\mathbf{z}_0, \mathbf{z}_0^*)$

$$
f(\mathbf{z}, \mathbf{z}^*) \geq f(\mathbf{z}_0, \mathbf{z}_0^*) + \left(\nabla_{\mathbf{z}_0} f(\mathbf{z}_0, \mathbf{z}_0^*)\right)^\top (\mathbf{z} - \mathbf{z}_0) + \left(\nabla_{\mathbf{z}_0^*} f(\mathbf{z}_0, \mathbf{z}_0^*)\right)^\top (\mathbf{z}^* - \mathbf{z}_0^*)
$$

Fig. 2. A code snippet in the file “first-order-Taylor–ex2.py”.

References

[1] X.-D. Zhang, *Matrix Analysis and Applications*. Cambridge University Press, 2017.

[2] K. B. Petersen, M. S. Pedersen *et al.*, “The matrix cookbook,” *Technical University of Denmark*, vol. 7, no. 15, p. 510, 2008.