Gradient Calculation for Nonlinear Linear Squares Problems with Complex Numbers

Qiang Ning
Department of Electrical and Computer Engineering
Beckman Institute for Advanced Science and Techonology
University of Illinois at Urbana-Champaign

Fall 2013

Abstract

When handling non-linear least squares problems, we usually need to calculate the gradient of the sum of squares, which has been well defined if the squares are dependent on real variables. However, sometimes these squares are dependent on complex variables, in which case more careful consideration should be involved.

I. Non-linear Least Squares Problem

The general non-linear squares problem is trying to find the optimality,

$$\alpha^\star = arg \min_{\alpha \in \mathbb{F}^n} f(r(\alpha)), \tag{1}$$

where $f = \|r(\alpha)\|_2^2$, and $r(\alpha) : \mathbb{F}^n \to \mathbb{F}^m$ with $m \geq n$.

If $\mathbb{F} = \mathbb{R}$, i.e., $r \in \mathbb{R}^m, \alpha \in \mathbb{R}^n$, then

$$\bigtriangledown f = 2J^T r, \tag{2}$$

where J is the Jacobian matrix of $r(\alpha)$ and defined as

$$J = \begin{pmatrix} \frac{\partial r_1}{\partial \alpha_1} & \frac{\partial r_1}{\partial \alpha_2} & \cdots & \frac{\partial r_1}{\partial \alpha_n} \\ \frac{\partial r_2}{\partial \alpha_1} & \frac{\partial r_2}{\partial \alpha_2} & \cdots & \frac{\partial r_2}{\partial \alpha_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial r_m}{\partial \alpha_1} & \frac{\partial r_m}{\partial \alpha_2} & \cdots & \frac{\partial r_m}{\partial \alpha_n} \end{pmatrix}. \tag{3}$$

However, Eqn. (2) does not always hold when $r(\alpha) \in \mathbb{C}^m$.

II. GRADIENT

In this section we will talk about the case where $r(\alpha) : \mathbb{R}^n \to \mathbb{C}^m$.

A. $n = m = 1$

When $n = m = 1$, $r(\alpha)$ is simply a complex scalar function with one single real variable. Assuming $r(\alpha) = u(\alpha) + iv(\alpha)$, where $u(\alpha)$ and $v(\alpha)$ are real functions, we have

$$
\begin{aligned}
f &= \|r(\alpha)\|_2^2 = u^2 + v^2, && (4)\\
\frac{\partial f}{\partial \alpha} &= 2u\frac{\partial u}{\partial \alpha} + 2v\frac{\partial v}{\partial \alpha}. && (5)
\end{aligned}
$$

Considering

$$
\frac{\partial r}{\partial \alpha} = \frac{\partial u}{\partial \alpha} + i\frac{\partial v}{\partial \alpha}, \tag{6}
$$

we have

$$
\frac{\partial f}{\partial \alpha} = r^*\frac{\partial r}{\partial \alpha} + r(\frac{\partial r}{\partial \alpha})^*. \tag{7}
$$

where "$*$" denotes complex conjugate.

B. $n = 1, m > 1$

When $m > 1$, $r(\alpha)$ is a vector function:

$$
r(\alpha) = \begin{pmatrix} r_1(\alpha) \\ r_2(\alpha) \\ \vdots \\ r_m(\alpha) \end{pmatrix}. \tag{8}
$$

Then we have:

$$
f = \|r(\alpha)\|_2^2 = \|r_1(\alpha)\|_2^2 + \|r_2(\alpha)\|_2^2 + \cdots + \|r_m(\alpha)\|_2^2, \tag{9}
$$

$$
\begin{aligned}
\frac{\partial f}{\partial \alpha} &= r_1^*\frac{\partial r_1}{\partial \alpha} + r_1(\frac{\partial r_1}{\partial \alpha})^* + \cdots + r_m^*\frac{\partial r_m}{\partial \alpha} + r_m(\frac{\partial r_m}{\partial \alpha})^* \\
&= \begin{bmatrix} r_1^*, \cdots, r_m^* \end{bmatrix} \begin{bmatrix} \frac{\partial r_1}{\partial \alpha} \\ \vdots \\ \frac{\partial r_m}{\partial \alpha} \end{bmatrix} + \begin{bmatrix} (\frac{\partial r_1}{\partial \alpha})^*, \cdots, (\frac{\partial r_m}{\partial \alpha})^* \end{bmatrix} \begin{bmatrix} r_1 \\ \vdots \\ r_m \end{bmatrix} \\
&= (J_\alpha^H r)^H + J_\alpha^H r, && (10)
\end{aligned}
$$

where "H" denotes conjugate transpose and J_α is the Jacobian matrix:

$$J_\alpha = \begin{pmatrix} \frac{\partial r_1}{\partial \alpha} \\ \frac{\partial r_2}{\partial \alpha} \\ \vdots \\ \frac{\partial r_m}{\partial \alpha} \end{pmatrix}. \tag{11}$$

C. $n > 1, m > 1$

$$\bigtriangledown f = \begin{pmatrix} \frac{\partial f}{\partial \alpha_1} \\ \vdots \\ \frac{\partial f}{\partial \alpha_n}. \end{pmatrix} \tag{12}$$

According to Eqn. (10), we have,

$$\frac{\partial f}{\partial \alpha_j} = (J_j^H r)^H + J_j^H r,\ j = 1, 2, ..., n, \tag{13}$$

where J_j is the j-th column of the Jacobian matrix as shown in Eqn. (3).

$$\begin{aligned} \bigtriangledown f &= \begin{pmatrix} (J_1^H r)^H \\ \vdots \\ (J_n^H r)^H \end{pmatrix} + \begin{pmatrix} J_1^H r \\ \vdots \\ J_n^H r \end{pmatrix} \\ &= (J^H r)^* + J^H r. \end{aligned} \tag{14}$$

Obviously this result is consistent with the result of $\mathbb{F} = \mathbb{R}$ as shown in Eqn. (2).

III. Another Way of Calculating the Gradient

We have already known that the difficulty is brought by the fact that $r(\alpha) \in \mathbb{C}^m$. We can directly use the result as shown in Eqn. (14). Nevertheless, we can also decompose r into real part and imaginary part:

$$\begin{aligned} r(\alpha) &= u(\alpha) + iv(\alpha) \\ &= \begin{pmatrix} u_1(\alpha) \\ \vdots \\ u_m(\alpha) \end{pmatrix} + i \begin{pmatrix} v_1(\alpha) \\ \vdots \\ v_m(\alpha) \end{pmatrix}. \end{aligned} \tag{15}$$

If we redefine our non-linear least squares problem as following, we can overcome the inconvenience caused by the complexity of r.

$$\alpha^\star = arg \min_{\alpha \in \mathbb{F}^n} f(\hat{r}(\alpha)), \tag{16}$$

$$\hat{r}(\alpha) = \begin{pmatrix} u(\alpha) \\ v(\alpha) \end{pmatrix}_{2m\times 1}, \tag{17}$$

$$f = \|\hat{r}(\alpha)\|_2^2 \tag{18}$$

Denoting the original Jacobian matrix as described in Eqn. (3) as J_r, and the new Jacobian matrix as $J_{\hat{r}}$, we can see:

$$J_{\hat{r}} = \begin{pmatrix} \frac{\partial u_1}{\partial \alpha_1} & \cdots & \frac{\partial u_1}{\partial \alpha_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial u_m}{\partial \alpha_1} & \cdots & \frac{\partial u_m}{\partial \alpha_n} \\ \frac{\partial v_1}{\partial \alpha_1} & \cdots & \frac{\partial v_1}{\partial \alpha_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial v_m}{\partial \alpha_1} & \cdots & \frac{\partial v_m}{\partial \alpha_n} \end{pmatrix} = \begin{pmatrix} J_u \\ J_v \end{pmatrix}. \tag{19}$$

In this way, we can calculate the gradient as:

$$\bigtriangledown f = 2J_{\hat{r}}^T \hat{r}. \tag{20}$$

A useful fact is that:

$$J_u = real(J_r), J_v = imag(J_r). \tag{21}$$