ECE 251A: Digital Signal Processing I
Complex Gradients: Wirtinger Calculus

Florian Meyer

Electrical and Computer Engineering Department
University of California San Diego

UC San Diego
JACOBS SCHOOL OF ENGINEERING

Real Functions and Gradients

For real valued parameter $x = [x_1, \ldots, x_n]^T \in R^n$ and scalar function $f(x)$,we have the gradient given by

$$\nabla_x f = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right]^T$$

and the second derivative or Hessian is given by

$$\mathbf{H} = [H_{i,j}], \text{ where } H_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

In optimization problems, the gradient is usually zero at the extrema (local minimum or local maximum) and the Hessian is (positive definite or negative definite).

For function $g(A)$, where A is a matrix, we have $\nabla_A(g)$ is a matrix with (ij)th entry equal to $\frac{\partial g}{\partial a_{ij}}$,

Some Useful Formulas

Consider $f(x) = y^T x = x^T y = \sum_{l=1}^{n} y_l x_l$. Then $\frac{\partial f}{\partial x_p} = y_p$. Hence

$$\nabla_x (y^T x) = \nabla_x (x^T y) = y$$

This leads to

$$\nabla_x (x^T A y) = Ay \text{ and } \nabla_x (y^T A x) = A^T y$$

For $f(x) = x^T A x$, it can be shown that

$$\nabla_x (x^T A x) = (A + A^T)x \text{ and } \mathbf{H} = \left[\frac{\partial^2 f}{\partial x_i \partial x_j} \right] = A + A^T$$

Some examples of matrix functions

$$\nabla_A (x^T A y) = x y^T$$
$$\nabla_A (x^T A x) = x x^T$$

Applications to the Least Squares Problem[1]

Objective function is

$$f(c) = \|y - Ac\|_2^2 = (y - Ac)^T(y - Ac) = y^T y + c^T A^T A c - c^T A^T y - y^T A c$$

$c \in R^n$, $y \in R^L, L \geqslant n$, and $A \in R^{L \times n}$.

Goal: Minimize $f(c)$ with respect to $c \in R^n$

$$\nabla_c f = 0 + (A^T A + A^T A)c - A^T y - A^T y = 2A^T A c - 2A^T y$$

Setting the gradient to zero leads to

$$A^T A c_{LS} = A^T y \text{ or } c_{LS} = (A^T A)^{-1} A^T y$$

if $A^T A$ is invertible.

The Hessian or second derivative is given by $2A^T A$, which is positive semi-definite and positive definite if $A^T A$ is invertible.

[1] $\|x\|_2^2 = \sum_l x_l^2 = x^T x$.

Functions of Complex Variables and Gradients

For complex valued parameter $z = [z_1, \ldots, z_n]^T \in C^n$ and scalar function $f(z)$, we have the gradient given by

$$\nabla_z f = \left[\frac{\partial f}{\partial z_1}, \frac{\partial f}{\partial z_2}, \ldots, \frac{\partial f}{\partial z_n} \right]^T$$

Scalar Case: For scalar $z = a + jb$, a scalar function $f(z) = u(a, b) + jv(a, b)$.

A function $f(z)$ is continuous at z_0, if $f(z_0)$ is defined, and $\lim_{z \to z_0} f(z) = f(z_0)$.

A function $f(z)$ is differentiable at z_0, if the limit

$$f'(z_0) = \lim_{\Delta z \to 0} \frac{f(z_0 + \Delta z) - f(z_0)}{\Delta z} = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

exists and the limit is called the derivative.

Derivative Continued

Cauchy-Riemann Conditions:

Let $f(z) = u(a, b) + jv(a, b)$ be defined and continuous in some neighborhood of a point $z = a + jb$ and differentiable at z itself. Then at that point, the first order partial derivative of u and v exists and satisfy the Cauchy-Riemann equations given below

$$\frac{\partial u}{\partial a} = \frac{\partial v}{\partial b} \text{ and } \frac{\partial u}{\partial b} = -\frac{\partial v}{\partial a}$$

Examples: $f(z) = z^l$, $f(z) = e^z$, $f(z) = \sum_l h[l]z^{-l}$, are all differentiable in their ROC.

$f(z) = z^*$

What about $f(z) = z^* = a - jb$.

$u(a,b) = a$, and $v(a,b) = -b$ and $\frac{\partial u}{\partial a} = 1$, $\frac{\partial v}{\partial b} = -1$, $\frac{\partial u}{\partial b} = 0$, $\frac{\partial v}{\partial a} = 0$.

Conclusion: Not differentiable

Alternate Verification

$$\lim_{\Delta z \to 0} \frac{f(z_0 + \Delta z) - f(z_0)}{\Delta z} = \lim_{\Delta z \to 0} \frac{(\Delta z)^*}{\Delta z} = \lim_{\Delta z \to 0} \frac{\Delta a - j\Delta b}{\Delta a + j\Delta b}$$

Least Squares Problem: Complex Case[2]

Objective function is

$$f(c) = \|y - Ac\|_2^2 = (y - Ac)^H(y - Ac) = y^H y + c^H A^H Ac - c^H A^H y - y^H Ac$$

$c \in C^n$, $y \in C^L, L \geqslant n$, and $A \in C^{L\times n}$.

Goal: Minimize $f(c)$ with respect to $c \in C^n$

$f(c)$ not differentiable with respect to c because of the presence of $C^H = [c_1^*, \ldots, c_n^*]$.

Optimization path based on taking the gradient is unclear.

[2] $\|x\|_2^2 = \sum_l |x_l|^2 = x^H x$.

Path Forward

$f(c)$ is a real and positive. So reduce problem to one involving real variables only.

Replace $c = c_R + jc_I$, $y = y_R + jy_I$, and $A = A_R + jA_I$.

The objective function is a real function of c_R, c_I and we can use real calculus.

The answer may not have a very nice form and be easy to interpret. Wirtinger calculus solves this issue of simplicity and interpretability.

Real Functions of Complex Variables and Gradients

For problems of interest to us, even though the variables are complex, the objective function being minimized is real.We can convert the problem involving complex variables into one of real variables and use ordinary calculus. Wirtinger calculus formalizes this approach with the additional benefits of simplicity and tractable expressions.

Let $x = a + jb$, where a and b are real vectors. We define the complex derivative as

$$\nabla_x(g) \triangleq \frac{1}{2}\left[\frac{\partial g}{\partial a} - j\frac{\partial g}{\partial b}\right] \quad \text{and} \quad \nabla_{x^*}(g) \triangleq \frac{1}{2}\left[\frac{\partial g}{\partial a} + j\frac{\partial g}{\partial b}\right].$$

$$\nabla_x(g) = 0 \leftrightarrow \frac{\partial g}{\partial a} = 0 \text{ and } \frac{\partial g}{\partial b} = 0$$

$$\nabla_{x^*}(g) = 0 \leftrightarrow \frac{\partial g}{\partial a} = 0 \text{ and } \frac{\partial g}{\partial b} = 0$$

Some Examples

Consider $f(x) = x^H y$.

$$\nabla_{x^*}(x^H y) = \frac{1}{2}\left[\frac{\partial(a^T y - jb^T y)}{\partial a} + j\frac{\partial(a^T y - jb^T y}{\partial b}\right] = \frac{1}{2}[y + j(-jy)] = y$$

This proves (B.16) in textbook. Similarly (B.17) is derived next. Since $y^H = y^{*T}$,

$$\nabla_{x^*}(y^H x) = \frac{1}{2}\left[\frac{\partial(y^{*T} a + jy^{*T} b)}{\partial a} + j\frac{\partial(y^{*T} a + jy^{*T} b)}{\partial b}\right] = \frac{1}{2}[y^* + j(jy^*)] = 0$$

Quadratic Functions

$$
\begin{aligned}
x^H Ax &= a^T Aa + b^T Ab - jb^T Aa + ja^T Ab \\
\frac{\partial x^H Ax}{\partial a} &= (A + A^T)a - jA^T b + jAb = Ax + A^T x^* \\
\frac{\partial x^H Ax}{\partial b} &= (A + A^T)b - jAa + jA^T a = -jA(a + jb) + jA^T(a - jb) \\
&= -jAx + jA^T x^*
\end{aligned}
$$

Using the above results, we have

$$
\begin{aligned}
\nabla_{x^*}(x^H Ax) &= \frac{1}{2}\left[\frac{\partial x^H Ax}{\partial a} + j\frac{\partial x^H Ax}{\partial b}\right] \\
&= \frac{1}{2}\left[Ax + A^T x^* + Ax - A^T x^*\right] = Ax \\
\nabla_{x}(x^H Ax) &= \frac{1}{2}\left[\frac{\partial x^H Ax}{\partial a} - j\frac{\partial x^H Ax}{\partial b}\right] \\
&= \frac{1}{2}\left[Ax + A^T x^* - Ax + A^T x^*\right] = A^T x^*
\end{aligned}
$$

Functions of Matrices

Let $A = B + jC$, where B and C are real matrices. Then

$$\begin{aligned}\nabla_A(x^H A y) &= \tfrac{1}{2}\left[\nabla_B \cdot (x^{*T} B y + j x^H C y) - j \nabla_C \cdot (x^H B y + j x^{*T} C y)\right] \\ &= \tfrac{1}{2}\left[x^* y^T - j(j x^* y^T)\right] \;=\; x^* y^T\end{aligned}$$

Replacing y with x in the above equation, we get $\nabla_A(x^H A x) = x^* x^T$

Least Squares Problem: Complex Case Revisited

Objective function is

$$f(c) = \|y - Ac\|_2^2 = (y - Ac)^H (y - Ac) = y^H y + c^H A^H A c - c^H A^H y - y^H A c$$

$c \in C^n$, $y \in C^L, L \geqslant n$, and $A \in C^{L \times n}$.

Goal: Minimize $f(c)$ with respect to $c \in C^n$

$$\begin{aligned} \nabla_{c^*} f(c) &= \nabla_{c^*} y^H y + \nabla_{c^*} c^H A^H A c - \nabla_{c^*} c^H A^H y - \nabla_{c^*} y^H A c \\ &= 0 + A^H A c - A^H y - 0 \end{aligned}$$

Setting the gradient to zero

$$A^H A c_{LS} = A^H y \;\text{ or }\; c_{LS} = (A^H A)^{-1} A^H y$$

Optimization with Equality Constraints

Real Case: Minimize $f(c)$ subject to $g(c) = 0$, where $g(c) \in R^p$.
An example of $g(c)$ is $Tc - d = 0$ or $Tc = d$.

Using Lagrange Multipliers

$$\min_{c,\lambda} f(c) + \lambda^T g(c)$$

where $\lambda \in R^p$ is the Lagrange multiplier of dimension p, equal to the number of constraints.
Optimize with respect to c and λ.

Complex case: Minimize $f(c)$ subject $g(c) = 0$, where $g(c) \in C^p$.
An example of $g(c)$ is $Tc - d = 0$ or $Tc = d$. (c, T, d are complex, but $f(c)$ is real)

Can show the equivalent Lagrange multiplier problem is

$$\min_{c,\lambda} f(c) + \lambda^H g(c) + \lambda^T g^*(c)$$

where $\lambda = \lambda_R + j\lambda_I$.

Optimize with respect to c and λ using Wirtinger Calculus

Derivation

Constraint $g(c) = 0$.

With $g(c) = g_R(c) + jg_I(c)$, we have two set of p real constraints $g_R(c) = 0$, and $g_I(c) = 0$.

This leads to the following Lagrange Multiplier problem in real variables

$$\min_{c,\lambda} \left[f(c) + 2\lambda_R^T g_R(c)) + 2\lambda_I^T g_I(c) \right]$$

Now the constraints can also be written as $2g_R(c) = g(c) + g^*(c)$ and $2g_I(c) = -j(g(c) - g^*(c))$

Substituting this we get

$$\begin{aligned} &\min_{c,\lambda} \left[f(c) + \lambda_R^T(g(c) + g^*(c)) - j\lambda_I^T(g(c) - g^*(c)) \right] \\ &= \min_{c,\lambda} \left[f(c) + \lambda^H g(c) + \lambda^T g^*(c) \right] \end{aligned}$$

where $\lambda = \lambda_R + j\lambda_I$.