Optimal Filtering
I: Complex Gradient and Lagrange Multipliers

ECE416 Adaptive Algorithms

Prof. Fred L. Fontaine

Department of Electrical Engineering
The Cooper Union

Spring 2022

Gradient and Lagrange Multipliers

We shall frequently deal with optimization problems.

We first review the gradient and Lagrange multipliers in the real case, then extend these concepts to the complex case.

Real Case

Let $\mathbf{w} \in \mathbb{R}^M$, and $J(\mathbf{w})$ a real-valued function. Define:

$$\nabla = \begin{bmatrix} \frac{\partial}{\partial w_1} \\ \frac{\partial}{\partial w_2} \\ \vdots \\ \frac{\partial}{\partial w_M} \end{bmatrix}$$

Real Case

Omitting the tedious computations, we present some useful results:

$$J(\mathbf{w}) = \mathbf{p}^T\mathbf{w} \text{ or } \mathbf{w}^T\mathbf{p} \longrightarrow \nabla J = \mathbf{p}$$

$$J(\mathbf{w}) = \mathbf{w}^T A\mathbf{w} \longrightarrow \nabla J = \left(A + A^T\right)\mathbf{w}$$

As a special case, if A is *symmetric* $\left(A = A^T\right)$ then:

$$\nabla\left(\mathbf{w}^T A\mathbf{w}\right) = 2A\mathbf{w}$$

Lagrange Multipliers

Consider a *constrained optimization problem* where we want to minimize $J(\mathbf{w})$ subject to multiple constraints, expressed as:

$$g_1(\mathbf{w}) = 0, g_2(\mathbf{w}) = 0, \cdots, g_L(\mathbf{w}) = 0$$

We can combine these constraint functions into a vector:

$$\mathbf{g}(\mathbf{w}) = \begin{bmatrix} g_1(\mathbf{w}) \\ g_2(\mathbf{w}) \\ \vdots \\ g_L(\mathbf{w}) \end{bmatrix}$$

Lagrange Multipliers

Then the constrained optimization problem is:

$$\min_{\mathbf{w}} J(\mathbf{w}) \text{ subject to } \mathbf{g}(\mathbf{w}) = \mathbf{0}$$

Note a more challenging type of problem is when we have multiple *inequalities*, but here we will stay with the case of equalities.

Introduce $\boldsymbol{\lambda} = [\lambda_1 \; \lambda_2 \; \cdots \; \lambda_L]^T$ and consider the augmented cost function:

$$J_0(\boldsymbol{\lambda}, \mathbf{w}) = J(\mathbf{w}) + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{w})$$

Then the unconstrained problem is:

$$\min_{\boldsymbol{\lambda}, \mathbf{w}} J_0(\boldsymbol{\lambda}, \mathbf{w})$$

Lagrange Multipliers

Taking the gradient with respect to $\boldsymbol{\lambda}$ returns the constraints:

$$\nabla_\lambda J_0 = \mathbf{g}\left(\mathbf{w}\right) = 0$$

Taking the gradient with respect to $\mathbf{w}$ yields:

$$\nabla_w J\left(\mathbf{w}\right) + \sum \lambda_\ell \nabla_w g_\ell\left(\mathbf{w}\right) = \mathbf{0}$$

Complex Case

When $\mathbf{w}$ is complex valued, we could simply break it down to real and imaginary components, and instead of M complex variables, we have $2M$ real variables. However, in many cases it is convenient to maintain the complex formulation.

Recall J remains real valued (so it can be minimized), and the constraint functions g are also assumed real valued (which is fine considering in the end they are set to 0).

We want an algebraically convenient way to formulate *gradient of a real-valued function of a complex vector.* The approach is based on *Wirtinger calculus.*

Complex Gradient

Write:

$$\mathbf{w} = \mathbf{x} + j\mathbf{y}$$

where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$. Also:

$$\mathbf{w}^* = \mathbf{x} - j\mathbf{y}$$

Note that the mapping $(\mathbf{x}, \mathbf{y}) \to (\mathbf{w}, \mathbf{w}^*)$ is invertible:

$$\begin{aligned} \mathbf{x} &= \frac{1}{2}(\mathbf{w} + \mathbf{w}^*) \\ \mathbf{y} &= \frac{1}{2j}(\mathbf{w} - \mathbf{w}^*) \end{aligned}$$

The idea is to treat the components of $\mathbf{w}, \mathbf{w}^*$ as independent variables, and relate the gradients ∇_w, ∇_{w^*} to the gradients ∇_x, ∇_y.

Complex Gradient

Taking the scalar case first, applying the chain rule:

$$
\begin{aligned}
\frac{\partial}{\partial w} &= \frac{\partial}{\partial x}\cdot\frac{\partial x}{\partial w}+\frac{\partial}{\partial y}\cdot\frac{\partial y}{\partial w}=\frac{1}{2}\left(\frac{\partial}{\partial x}-j\frac{\partial}{\partial y}\right)\\
\frac{\partial}{\partial w^*} &= \frac{\partial}{\partial x}\cdot\frac{\partial x}{\partial w^*}+\frac{\partial}{\partial y}\cdot\frac{\partial y}{\partial w^*}=\frac{1}{2}\left(\frac{\partial}{\partial x}+j\frac{\partial}{\partial y}\right)
\end{aligned}
$$

Putting this in vector form:

$$
\frac{\partial}{\partial \mathbf{w}^*}=\begin{bmatrix}\frac{\partial}{\partial w_1^*}\\ \frac{\partial}{\partial w_2^*}\\ \vdots\\ \frac{\partial}{\partial w_M^*}\end{bmatrix}=\frac{1}{2}\left(\nabla_x+j\nabla_y\right)
$$

Complex Gradient

Now let us *define:*

$$\nabla_w = 2\frac{\partial}{\partial w^*} = \nabla_x + j\nabla_y = \begin{bmatrix} \frac{\partial}{\partial x_1} + j\frac{\partial}{\partial y_1} \\ \frac{\partial}{\partial x_2} + j\frac{\partial}{\partial y_2} \\ \vdots \\ \frac{\partial}{\partial x_M} + j\frac{\partial}{\partial y_M} \end{bmatrix}$$

Note that setting $\nabla_w J\left(\mathbf{w}\right) = \mathbf{0}$ is equivalent to setting all $\frac{\partial J}{\partial x_m} = 0$ and $\frac{\partial J}{\partial y_m} = 0$.

Digression: Analytic Functions

Let $f(z) = u(x,y) + jv(x,y)$ be a complex valued function of a complex variable. It is *analytic (holomorphic)* if it is complex differentiable:

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

In particular, the limit should be the same no matter how $\Delta z \longrightarrow 0$.

Take $\Delta z = \Delta x$ and then $\Delta z = j\Delta y$ we get:

$$f'(z) = \frac{\partial u}{\partial x} + j\frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} - j\frac{\partial u}{\partial y}$$

It turns out matching these two formulas is all we need!

Digression: Analytic Functions

Theorem: f is analytic iff:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \text{and} \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

These are called the *Cauchy-Riemann equations*.

They can be written in more compact form:

$$\frac{\partial f}{\partial z^*} = 0$$

Indeed, as we know, $f(z) = z^*$ is not analytic!

Special Cases

$J(\mathbf{w})$ with $\mathbf{w}$ complex and $J(\cdot)$ real:

$$J(\mathbf{w}) = 2\,\mathrm{Re}\left(\mathbf{p}^H\mathbf{w}\right) = \mathbf{p}^H\mathbf{w} + \mathbf{w}^H p \longrightarrow \nabla J = 2\mathbf{p}$$

and with A Hermitian:

$$J(\mathbf{w}) = \mathbf{w}^H A\mathbf{w} \longrightarrow \nabla J = 2A\mathbf{w}$$

Rayleigh Quotient Revisited

It can be shown our formulation of the complex gradient satisfies:

$$\nabla \left(\frac{f(\mathbf{w})}{g(\mathbf{w})} \right) = \frac{1}{g} \nabla f - \frac{1}{g^2} f \nabla g$$

where $f(\cdot), g(\cdot)$ are real-valued. Now take:

$$\eta(\mathbf{w}) = \frac{\mathbf{w}^H A \mathbf{w}}{\mathbf{w}^H \mathbf{w}}$$

with A Hermitian.

Rayleigh Quotient Revisited

$$\begin{aligned}
\nabla\eta &= \frac{1}{\mathbf{w}^H\mathbf{w}}\nabla\left(\mathbf{w}^H A\mathbf{w}\right) - \frac{\mathbf{w}^H A\mathbf{w}}{\left(\mathbf{w}^H\mathbf{w}\right)^2}\nabla\left(\mathbf{w}^H\mathbf{w}\right) \\
&= \frac{1}{\mathbf{w}^H\mathbf{w}}2A\mathbf{w} - \mathbf{2}\frac{\mathbf{w}^H A\mathbf{w}}{\left(\mathbf{w}^H\mathbf{w}\right)^2}\mathbf{w} \\
&= \frac{2}{\|\mathbf{w}\|^2}\left(A\mathbf{w}-\eta\left(\mathbf{w}\right)\mathbf{w}\right)
\end{aligned}$$

Setting $\nabla\eta\left(\mathbf{w}\right) = \mathbf{0}$ givens:

$$A\mathbf{w} = \eta\left(\mathbf{w}\right)\mathbf{w}$$

hence $\mathbf{w}$ is an eigenvector, and $\eta\left(\mathbf{w}\right)$ is the corresponding eigenvalue!

Complex Lagrange Multipliers

Let $\mathbf{w} \in \mathbb{C}^M$, and let $J(\mathbf{w})$ be a *real* cost function, $\mathbf{g}(\mathbf{w})$ a vector of L *complex* constraint functions. The constrained optimization problem is:

$$\min_{\mathbf{w}} J(\mathbf{w}) \text{ subject to } \mathbf{g}(\mathbf{w}) = \mathbf{0}$$

Introduce a *complex* vector of Lagrange multipliers $\boldsymbol{\lambda} = [\lambda_1 \cdots \lambda_L]^T$ and then:

$$\min_{\boldsymbol{\lambda},\mathbf{w}} J(\mathbf{w}) + \mathrm{Re}\left(\boldsymbol{\lambda}^H \mathbf{g}(\mathbf{w})\right) = \min_{\boldsymbol{\lambda},\mathbf{w}} J(\mathbf{w}) + \mathrm{Re}\left(\sum_{\ell=1}^{L} \lambda_\ell^* g_\ell(\mathbf{w})\right)$$

Then the constraints are returned via:

$$\nabla_{\boldsymbol{\lambda}} \left(J + \mathrm{Re}\left(\lambda^H g\right)\right) = \mathbf{g}(\mathbf{w}) = \mathbf{0}$$

OPTIMAL FILTERING
I: COMPLEX GRADIENT AND LAGRANGE MULTIPLIERS

ECE416 ADAPTIVE ALGORITHMS