Review of Optimization Theory

Consider a real-valued function f(x) of real variable x. The function local minima points satisfy to (*necessary conditions*)

$$\frac{df(x)}{dx} = 0, \qquad \frac{d^2 f(x)}{dx^2} \geq 0 \tag{1}$$

and the *sufficient conditions* are

$$\frac{df(x)}{dx} = 0, \qquad \frac{d^2 f(x)}{dx^2} > 0$$

These conditions are easily derived from the Taylor series expansion up to 2nd order term (e.g. see [3][4]). Maxima are obtained from minima via $f \to -f$.
There maybe several local minima. Global minimum (the minimum) is found by inspecting all the local minima is choosing the smallest one. If f(x) is strictly convex,

$$f\left(ax_1 + (1-a)x_2\right) < af\left(x_1\right) + (1-a)\, f\left(x_2\right) \tag{2}$$

then there exists only one minimum, which is also the global minimum. Example of a convex function: $f(x) = a_1 x^2 + a_2 x + a_3, \quad a_1 > 0$.

For a convex function, the 1st part of (1) is a sufficient condition of the minimum.

In practice, a real-valued function f(z) of a complex variable z is of interest,

$$z = x + jy, \quad x = \mathrm{Re}\{z\}, \quad y = \mathrm{Im}\{z\}. \tag{3}$$

If the derivative df(z)/dz exists, a minimum can be found as in (1). In many cases, however, this derivative does not exist. A simple example of such a function is $f(z)=|z|^2$, which is of large practical interest. Obviously, the minimum of $|z|^2$ is at z=0, but we cannot apply (1) as dz^*/dz does not exist (Rieman-Cauchy conditions are not satisfied).

This difficulty can be resolved in 2 ways.

1). Express f(z) as f(x, y) and use (1) for x and y independently,

$$\frac{df(x,y)}{dx}=0, \quad \frac{df(x,y)}{dy}=0$$
$$\frac{d^2 f(x,y)}{dx^2}>0, \quad \frac{d^2 f(x,y)}{dy^2}>0 \tag{4}$$

This results in a complicated (lengthy) analysis.

2). An elegant solution to this problem is to express

$$f(z)=f(z,z^*) \tag{5}$$

where z and z* are <u>considered</u> to be independent variables (constant with respect to each other). Under this assumption, for example, $dz^*/dz=0$ and $d|z|^2/dz=z^*$. The minima of f(z, z*) are found using

$$\frac{df(z,z^*)}{dz}=0 \quad or \quad \frac{df(z,z^*)}{dz^*}=0 \tag{6}$$

where we assumed that f(z, z*) is a convex function (which is the case for all our applications as f(z) are various quadratic forms).

When f is a function of complex- valued vector w, the stationary points are found from

$$\frac{df}{d\mathbf{w}} = \left[\frac{df}{dw_1}, \quad \frac{df}{dw_2}, \quad .. \quad \frac{df}{dw_n} \right] = \mathbf{0} \tag{7}$$

or, equivalently,

$$\frac{df}{d\mathbf{w}^+} = \left[\frac{df}{dw_1^*}, \quad \frac{df}{dw_2^*}, \quad .. \quad \frac{df}{dw_n^*} \right]^T = \mathbf{0} \tag{8}$$

i.e. (6) is applied element-wise.

Example: consider $f(w, w^+) = |w|^2 = w^2 w$,

$$\frac{df}{dw^+} = \frac{d(w^+ w)}{dw^+} = w = 0$$

i.e. , the minimum is achieved at w=0.

Differentiation with respect to a vector

Consider a scalar function f of a vector argument $\mathbf{w}$, $f(\mathbf{w})$. Its derivative with respect to $\mathbf{w}$ (gradient) is the following vector

$$\frac{df}{d\mathbf{w}} = \left[\frac{df}{dw_1}, \quad \frac{df}{dw_2}, \quad .. \quad \frac{df}{dw_n} \right] \tag{9}$$

i.e. it is a row vector. The derivative of $f(\mathbf{w})$ with respect to a row vector $\mathbf{w}^T$ is the following column vector,

$$\nabla f = \frac{df}{d\mathbf{w}^T} = \left[\frac{df}{dw_1}, \quad \frac{df}{dw_2}, \quad .. \quad \frac{df}{dw_n} \right]^T \tag{10}$$

Derivative of a vector $\mathbf{u}(\mathbf{w})$ with respect to $\mathbf{w}$ is the following matrix $\mathbf{D}$,

$$\frac{d\mathbf{u}}{d\mathbf{w}} = \left[\frac{d\mathbf{u}}{dw_1}, \quad \frac{d\mathbf{u}}{dw_2}, \quad .. \quad \frac{d\mathbf{u}}{dw_n} \right] = \mathbf{D} \rightarrow d_{ij} = \frac{du_i}{dw_j} \tag{10}$$

where d_{ij} denotes *ij*-th entry of $\mathbf{D}$.

Constrained optimization

Suppose we wish to find an extremum (maximum or minimum) of a function $f(\mathbf{x})$ subject to a constrain of the form $h(\mathbf{x}) = 0$ (all functions and arguments are assumed to be real here, for simplicity),

$$\min_{\mathbf{x}} f(\mathbf{x}), \text{ subject to } h(\mathbf{x}) = 0 \tag{10}$$

Introducing the Lagrangian,

$$L = f(\mathbf{x}) + \lambda h(\mathbf{x}) \tag{11}$$

where λ is the Lagrange multiplier, the *necessary condition* for an extremum is

$$\nabla L = \frac{\partial L}{\partial \mathbf{x}} = \mathbf{0} \tag{12}$$

which gives the solution $\mathbf{x}_0(\lambda)$ and λ is found from the constrain:

$$h(\mathbf{x}_0(\lambda)) = 0 \tag{13}$$

The *sufficient conditions* are 1st order condition in (12) plus the following 2nd order condition: if the Hessian $\mathbf{H} = \nabla^2 L = \partial^2 L / \partial\mathbf{x}\partial\mathbf{x}^T$ is positive definite on the tangent plane of $h(\mathbf{x}_0)$, $\mathbf{y}^T\mathbf{H}\mathbf{y} > 0 \forall \mathbf{y} : \mathbf{y}^T \nabla h(\mathbf{x}_0) = 0$, then $\mathbf{x}_0$ is a minimum; if $\mathbf{y}^T\mathbf{H}\mathbf{y} < 0 \forall \mathbf{y} : \mathbf{y}^T \nabla h(\mathbf{x}_0) = 0$, then it is a maximum. This gives only local extrema and global ones can only be found by inspection of all local ones, in general.

For *convex* problems [5], global extremum is unique and can be easily found (unlike the generic, non-linear case): if $f(\mathbf{x})$ is strictly convex, $\nabla^2 f(\mathbf{x}) > 0$, and $h(\mathbf{x})$ is affine ($= \mathbf{a}^T\mathbf{x} + c$), then $\mathbf{x}_0$ is the unique, global minimum; if $f(\mathbf{x})$ is strictly concave and $h(\mathbf{x})$ is affine, then $\mathbf{x}_0$ is the unique, global maximum.

Multiple constraints of the form $h_i(\mathbf{x}) = 0$ can be handled via the substitution $\lambda h(\mathbf{x}) \to \sum_i \lambda_i h_i(\mathbf{x})$ in (11).

Example (from high school): among all rectangles of given perimeter p, find one that has the maximum area.

$f(\mathbf{x}) = x_1 x_2,\ h(\mathbf{x}) = x_1 + x_2 - p$

$L = x_1 x_2 + \lambda(x_1 + x_2 - p)$

$\nabla L = \mathbf{0} \to x_1^* = x_1^* = -\lambda$

$\lambda = -p/2 \to x_1^* = x_1^* = p/2$

i.e. the extremum area is for the square. To see that this is indeed the maximum,

$$\mathbf{H} = \nabla^2 L = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix};\ \mathbf{y} : \mathbf{y}^T \nabla h(\mathbf{x}^*) = 0 \to \mathbf{y} = [a, -a] \forall a \neq 0$$

$$\mathbf{y}^T \mathbf{H} \mathbf{y} = -2a^2 < 0$$

i.e. indeed the maximum. Note that, in this case, $f(\mathbf{x})$ is not convex, so that the problem itself is not convex, yet the global maximum is easily found (since there is only one point satisfying the sufficient conditions).

Home exercise: among all rectangles inscribed in a given circle, find the one with the maximum area.

More detailed review of matrix and vector differentiation techniques can be found in [3, Appendix E] and [1, Appendix 7].

More detailed discussion of the optimization theory relevant to signal processing applications (including smart antennas) can be found in [3, Section 18] and [2, Appendix C], [4], or by taking the Convex optimization course (based on [5]).

References

[1] H.L. Van Trees, Optimum Array Processing, Wiley, 2002. Appendix 7.

[2] D.H. Johnson, D.E. Dudgeon, Array Signal Processing, Prentice Hall, 1993. Appendix C

[3] T.K. Moon, W.C. Stirling, Mathematical Methods and Algorithms for Signal Processing, Prentice Hall, 2000.

[4] D. G. Luenberger, Linear and Nonlinear Programming, Springer, 2005.

[5] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.