Appendix D

Optimization and Lagrange Multipliers

Suppose that $x^* \in \mathbb{R}$ is a local minimum of a twice-differentiable objective function $f(x)$, $f : \mathbb{R} \to \mathbb{R}$, in the sense that for all x near x^*, we have $f(x) \geq f(x^*)$. We can then expect that the slope of $f(x)$ at x^* is zero, i.e.,

$$\frac{\partial f}{\partial x}(x^*) = 0,$$

and also that

$$\frac{\partial^2 f}{\partial x^2}(x^*) \geq 0.$$

If f is multi-dimensional, i.e., $f : \mathbb{R}^n \to \mathbb{R}$, and all partial derivatives of f exist up to second-order, then a necessary condition for $x^* \in \mathbb{R}^n$ to be a local minimum is that its gradient be zero:

$$\nabla f(x^*) = \left[\begin{array}{ccc} \frac{\partial f}{\partial x_1}(x^*) & \cdots & \frac{\partial f}{\partial x_n}(x^*) \end{array} \right]^{\mathrm{T}} = 0.$$

For example, consider the linear equation $Ax = b$, where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ $(m > n)$ are given. Because there are more constraints (m) than variables (n), in general a solution to $Ax = b$ will not exist. Suppose we seek the x that best approximates a solution, in the sense of satisfying

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2}\|Ax - b\|^2 = \frac{1}{2}(Ax - b)^{\mathrm{T}}(Ax - b) = \frac{1}{2}x^{\mathrm{T}}A^{\mathrm{T}}Ax - 2b^{\mathrm{T}}Ax + b^{\mathrm{T}}b.$$

The first-order necessary condition is given by

$$A^{\mathrm{T}}Ax - A^{\mathrm{T}}b = 0. \tag{D.1}$$

If rank $A = n$ then $A^{\mathrm{T}}A \in \mathbb{R}^{n\times n}$ is invertible, and the solution to (D.1) is

$$x^* = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b.$$

Now suppose that we wish to find, among all $x \in \mathbb{R}^n$ that satisfy $g(x) = 0$ for some differentiable $g : \mathbb{R}^n \to \mathbb{R}^m$ (typically $m \leq n$ to ensure that there exists an infinity of solutions to $g(x) = 0$), the x^* that minimizes the objective function $f(x)$. Suppose that x^* is a local minimum of f that is also a regular point of the surface parametrized implicitly by $g(x) = 0$, i.e., x^* satisfies $g(x^*) = 0$ and

$$\operatorname{rank} \frac{\partial g}{\partial x}(x^*) = m.$$

Then, from the fundamental theorem of linear algebra, it can be shown that there exists some $\lambda^* \in \mathbb{R}^m$ (called the **Lagrange multiplier**) that satisfies

$$\nabla f(x^*) + \frac{\partial g}{\partial x}^{\mathrm{T}}(x^*)\lambda^* = 0 \tag{D.2}$$

Equation (D.2) together with $g(x^*) = 0$ constitute the first-order necessary conditions for x^* to be a feasible local mininum of $f(x)$. Note that these two equations represent $n + m$ equations in the $n + m$ unknowns x and λ.

As an example, consider the quadratic objective function $f(x)$ such that

$$\min_{x\in\mathbb{R}^n} f(x) = \frac{1}{2}x^{\mathrm{T}}Qx + c^{\mathrm{T}}x,$$

subject to the linear constraint $Ax = b$, where $Q \in \mathbb{R}^n$ is symmetric positive-definite (that is, $x^{\mathrm{T}}Qx > 0$ for all $x \in \mathbb{R}^n$) and the matrix $A \in \mathbb{R}^{m\times n}$, $m \leq n$, is of maximal rank m. The first-order necessary conditions for this equality-constrained optimization problem are

$$\begin{aligned} Qx + A^{\mathrm{T}}\lambda &= -c, \\ Ax &= b. \end{aligned}$$

Since A is of maximal rank and Q is invertible, the solutions to the first-order necessary conditions can be obtained, after some manipulation, as

$$\begin{aligned} x &= Gb + (I - GA)Q^{-1}c, \\ \lambda &= Bb + BAQ^{-1}c, \end{aligned}$$

where $G \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{m\times m}$ are defined as

$$G = Q^{-1}A^{\mathrm{T}}B, \qquad B = (AQ^{-1}A^{\mathrm{T}})^{-1}.$$