Maxima, Minima and other Stationary Points

Let f be a function of two or more variables. Then a *necessary* condition for $\mathbf{x_0}$ to be a minimum (or maximum) of f is

$$\nabla f(\mathbf{x_0}) = \mathbf{0}.$$

We see this as follows. If $\mathbf{h}$ is small,

$$f(\mathbf{x_0} + \mathbf{h}) \approx f(\mathbf{x_0}) + \nabla f(\mathbf{x_0}).\mathbf{h}.$$

Now suppose that $\mathbf{x_0}$ is a minimum of f. Then

$$f(\mathbf{x_0}) \leq f(\mathbf{x_0} + \mathbf{h})$$

for all $\mathbf{h}$. So

$$\nabla f(\mathbf{x_0}).\mathbf{h} \geq 0$$

for all small $\mathbf{h}$.

Now put

$$\mathbf{h} = t\nabla f(\mathbf{x_0}).$$

Then

$$\nabla f(\mathbf{x_0}).\mathbf{h} = t\|\nabla f(\mathbf{x_0}\|^2 \geq 0$$

for all small t. The only way this can be true for both $t > 0$ and $t < 0$ is if

$$\nabla f(\mathbf{x_0}) = \mathbf{0}. \tag{1}$$

A point $\mathbf{x_0}$ where (1) holds is called a *stationary* point or a *critical* point.

Second Derivative Test for Type of Stationary Point

Let $f = f(x, y)$ with continuous first and second partial derivatives, and let (x_0, y_0) be a stationary point of f, that is

$$\nabla f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0\mathbf{i} + \frac{\partial f}{\partial y}(x_0, y_0)\mathbf{j} = \mathbf{0}.$$

Then write

$$A = \frac{\partial^2 f}{\partial x^2}(x_0, y_0), \quad B = \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0),$$

$$C = \frac{\partial^2 f}{\partial y^2}(x_0, y_0).$$

Then if

$$AC - B^2 > 0, A > 0,$$

(x_0, y_0) is a *local minimum* of f. If

$$AC - B^2 > 0, A < 0,$$

(x_0, y_0) is a *local maximum* of f. If

$$AC - B^2 < 0,$$

(x_0, y_0) is a *saddle point* of f. If

$$AC - B^2 = 0$$

then we *don't know.*

Lagrange Multipliers

Suppose that f and g are both functions of n variables, with $f(\mathbf{x})$ and $g(\mathbf{x})$ defined for all vectors $\mathbf{x}$. Then a necessary condition for $\mathbf{x_0}$ to be a local maximum or local minimum of f restricted to the set

$$g(\mathbf{x}) = c,$$

for any constant c, is that

$$\nabla f(\mathbf{x_0}) = \lambda \nabla g(\mathbf{x_0})$$

for some λ. The reason is that we need

$$\nabla f(\mathbf{x_0}).\mathbf{h} = 0$$

for all $\mathbf{h}$ such that $g(\mathbf{x_0} + \mathbf{h}) \approx g(\mathbf{x_0}) = c$, that is, for all $\mathbf{h}$ such that

$$g(\mathbf{x_0}) + \nabla g(\mathbf{x_0}).\mathbf{h} = g(\mathbf{x_0}),$$

that is, for all $\mathbf{h}$ such that

$$\nabla g(\mathbf{x_0}).\mathbf{h} = 0.$$