Section 4.3. Optimization of Functions (Partial)

Note. We now consider $f : \mathbb{R}^n \to \mathbb{R}$ and attempt to find optima (local maxima/minima) of f. We describe some analytic and numerical techniques, but offer (like Gentle) no proofs/justifications.

Note. For $f : \mathbb{R}^n \to \mathbb{R}$, a *stationary point* is a point at which

$$\frac{\partial f}{\partial x} = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right]$$

is zero. We can consider the Hessian at the stationary point to see if it is a local maximum, local minimum, or saddle point. We assume that f is sufficiently continuous so that $\partial^2 f/\partial x_i \partial x_j = \partial^2 f/\partial x_j \partial x_i$ so that the Hessian of f is symmetric. Then, by Theorem 3.8.14, the Hessian is positive definite if and only if all of its eigenvalues are positive. So:

(1) If (but not only if) the stationary point is a local minimum then the Hessian is nonnegative definite.

(2) If the Hessian is positive definite then the stationary point is a local minimum.

(3) If (but not only if) the stationary point is a local minimum then the Hessian is nonpositive definite.

(4) If the Hessian is negative definite then the stationary point is a local maximum.

(5) If the Hessian has both positive and negative eigenvalues then the stationary point is a saddle point.

Note. Let $F : \mathbb{R}^n \to \mathbb{R}$. *Newton's Method* is an iterative numerical technique that estimates a stationary point of $f(x)$ (that is, a point where $\nabla f(x) = 0$). We start with an initial guess $x^{(0)}$ as an estimate of a solution and then revise it for $k = 1, 2, \ldots$ by solving the linear system

$$\nabla^2 f(x^{(k)}) p^{(k+1)} = -\nabla f(x^{(k)})$$

and updating the estimated solution as

$$x^{(k+1)} = x^{(k)} + p^{(k+1)}.$$

The process is then iterated.

Note. This looks somewhat different from the Newton's Method you encountered in Calculus 1 (see, for example, my online Calculus 1 notes on 4.7. Newton's Method). There, you were likely looking for zeros of a function f. Here you are looking for zeros of the *derivative* ∇f (which is why in Calculus 1 you only used the function and its derivative, but here you use the derivative ∇f and the second derivative $\nabla^2 f$). You may also be aware that the stationary point to which Newton's Method is attracted likely depends on the initial guess $x^{(0)}$ and that Newton's Method can produce unexpected behavior (like cycling around instead of approaching a stationary point or acting in a chaotic way).

Note. In a *least squares fit of a linear model*, $y = X\beta + \varepsilon$, where y is an n-vector (of "output values" or dependent values), X is an $n \times m$ matrix, and β is an m-vector (of "input values" or independent values), we replace β with variable vector b and define the *residual vector* $r = y - Xb$ (notice that this is a difference of an "observed value" y and a "predicted value" Xb). We want to minimize the Euclidean norm of the residual vector $f(b) = \langle r, r\rangle = r^T r = (y - Xb)^T(y - Xb)$. Notice that this is the sum of squares (since we use the Euclidean norm) of the difference between observed and predicted values. So we differentiate $f(b)$ to get

$$\begin{aligned}
\frac{\partial}{\partial b}[f(b)] &= \frac{\partial}{\partial b}\left[(y - Xb)^T(y - Xb)\right] \\
&= \frac{\partial}{\partial b}[y^T y - y^T Xb - b^T X^T y + b^T X^T Xb] \\
&= 0 - (y^T X)^T - X^T y + (X^T X + (X^T X)^T)b \text{ by the differentiation} \\
&\qquad \text{properties given in Section 4.2} \\
&= -2X^T y + 2X^T Xb.
\end{aligned}$$

Setting $\partial f/\partial b = 0$ implies $X^T Xb = X^T y$. So a solution b of the system $X^T Xb = X^T y$ is a stationary point of $f(b)$. Now $\partial^2 f/\partial b^2 = 2X^T X$. Gentle states that $X^T X \succeq 0$ (presumably based on some hypothesis on the linear system $y = X\beta + \varepsilon$ that guarantees a unique solution to $X^T Xb = X^T y$) and then concludes that, since the Hessian is positive definite, stationary point b corresponds to a minimum of $f(b)$ and so b minimizes the Euclidean norm of the residual vector.

Note. For a geometric argument of the least squares technique, see my online Linear Algebra (MATH 2010) notes on 6.5 The Method of Least Squares. The desired vector (denoted $\overline{r}$ in these notes) results from a projection of a vector onto a subspace; this projection then minimizes the residual vector. This technique yields the same result as Gentle's approach.

Note. We omit the rest of this section since we do not need it for the rest of the course.

Revised: 4/25/2020