4. Univariate and Multivariate Differentiation (cont) 1

4.6. Taylor’s Theorem (cont) 1

4.7. Application of Taylor’s theorem: second order conditions for an unconstrained maximum. 2

4.8. Another application of Taylor 5

4.9. Terminology Review 6

4. Univariate and Multivariate Differentiation (cont)

4.6. Taylor’s Theorem (cont)

Taylor’s Theorem (continued): Why is the theorem so tremendously important? Because if you are only interested in the sign of \((f(\bar{x} + dx) - f(\bar{x}))\) and you have an n’th order Taylor expansion, then you know that for some neighborhood about \(\bar{x}\), the sign of your expansion will be the same as the sign of the true difference.
4.7. **Application of Taylor’s theorem: second order conditions for an unconstrained maximum.**

Going to be talking about necessary and sufficient conditions for an optimum of a differentiable function.

Terminology is that first order conditions are *necessary* while second order conditions are *sufficient*.

The terms necessary and sufficient conditions have a formal meaning:

- If an event $A$ cannot happen unless an event $B$ happens, then $B$ is said to be a *necessary condition* for $A$.

- If an event $B$ implies that an event $A$ will happen, then $B$ is said to be a *sufficient condition* for $A$.

For example, consider a differentiable function from $\mathbb{R}^1$ to $\mathbb{R}^1$.

- $f$ cannot attain an interior maximum at $\bar{x}$ *unless* $f'(\bar{x}) = 0$.
  - i.e., the maximum is $A$; the derivative condition is $B$.
  - Thus, the condition that the first derivative is zero is *necessary* for an interior maximum; called the first order conditions.
  - Emphasize strongly that this necessity business is delicate: derivative condition is only necessary provided that $f$ is differentiable *and* we’re talking interior maximum. Also, only talking LOCAL maximum.

- $f'(\bar{x}) = 0$ certainly doesn’t IMPLY that $f$ attains an interior maximum at $\bar{x}$
• If \( f''(\bar{x}) < 0 \), then the condition \( f'(\bar{x}) = 0 \) is both necessary and sufficient for an interior local maximum;

• Alternatively, if you know in advance that \( f \) is strictly concave, then the condition that \( f'(\bar{x}) \) is zero is necessary and sufficient for a strict global maximum.

Generalizing to functions defined on \( \mathbb{R}^n \), a simple application of Taylor’s theorem proves that if an appropriate local second order condition is satisfied, then the first order conditions are in fact necessary and sufficient for a strict local maximum.

**Theorem:** Consider a twice continuously differentiable function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) and a point \( \bar{x} \in \mathbb{R}^n \) such that the Hessian of \( f \) is negative definite at \( \bar{x} \). \( f \) attains a strict local maximum at \( \bar{x} \in \mathbb{R}^n \) iff \( \nabla f(\bar{x}) = 0 \).

**Proof of Necessity:** Suppose \( \nabla f(\bar{x}) \neq 0 \) and consider the first order Taylor expansion of \( f \) about \( \bar{x} \):

\[
f(\bar{x} + \Delta x) - f(\bar{x}) = \nabla f(\bar{x}) \Delta x + \text{a remainder term}.
\]  

Let \( \Delta x = \lambda \nabla \). Clearly, if \( \Delta x = \lambda \nabla \), \( \lambda > 0 \), then the first term in the Taylor’s expansion is positive. Applying Taylor’s theorem from last time, there exists \( \epsilon > 0 \), such that if \( 0 < |\lambda| < \epsilon \), then the absolute value of the first term of the Taylor expansion of \( f \) about \( \bar{x} \) is larger than the absolute value of the remainder term. Therefore, \( f(\bar{x} + \Delta x) > f(\bar{x}) \) and \( \bar{x} \) cannot be a maximum.

**Joint sufficiency:** This needs a little more work. Consider the second order Taylor expansion of \( f \) about \( \bar{x} \):

\[
f(\bar{x} + \Delta x) - f(\bar{x}) = \nabla f(\bar{x}) \Delta x + \frac{1}{2} \Delta x' \nabla^2 f(\bar{x}) \Delta x + \text{a remainder term}
\]
Note that since \( H_f(\bar{x}) \) is negative definite, the second term in the Taylor’s expansion is negative, for any \( \mathbf{d}x \neq 0 \). Now fix \( \mathbf{v} \) in the unit circle and define \( \epsilon(\mathbf{v}) \) as follows:

\[
\epsilon(\mathbf{v}) = \max\{\epsilon \in [0,1] : \forall \delta \leq \epsilon, (0.5(\delta \mathbf{v})' \cdot H_f(\bar{x}) \cdot (\delta \mathbf{v}) - \text{the remainder term for } \delta \mathbf{v}) \leq 0\}.
\]

Taylor’s theorem tells us that for all \( \mathbf{v} \), \( \epsilon(\mathbf{v}) > 0 \). Moreover, it follows (after a little work which we won’t do) that \( \epsilon(\cdot) \) is a continuous function of \( \mathbf{v} \).\(^1\) Now a continuous function on a compact set (the unit circle) attains a minimum on that set. Let \( \bar{\epsilon} \) denote the minimum of \( \epsilon(\cdot) \) on the unit circle and note that \( \bar{\epsilon} > 0 \). Next note that for all \( \mathbf{d}x \) such that \( ||\mathbf{d}x|| < \bar{\epsilon} \), the sum of the first two terms in the Taylor expansion (i.e., the Hessian term since the first term is zero) is larger in absolute magnitude than the remainder term. That is, for all \( 0 \neq \mathbf{d}x \) in the \( \bar{\epsilon} \) ball around \( \bar{x} \), \( f(\bar{x} + \mathbf{d}x) < f(\bar{x}) \). \( \square \)

Note that one has to be extremely careful about the wording of these necessary and sufficient conditions: The following statement is FALSE: A function \( f : \mathbb{R}^n \to \mathbb{R} \) attains a strict local maximum at \( \bar{x} \in \mathbb{R}^n \) iff \( \nabla f(\bar{x}) = 0 \) and \( H_f \) is negative definite at \( \bar{x} \). The “if” part of this statement is true, but the “only if” isn’t: you could have a strict local max at \( \bar{x} \) without \( f \) being negative definite at \( \bar{x} \), e.g. \(-x^4\) attains a global max at 0 but it isn’t negative definite at 0.

Notice that the theorem above only gives sufficient conditions for a local maximum. On the other hand, if \( H_f(\cdot) \) is globally negative definite (i.e., if \( f \) is strictly concave), then it can be shown that \( f \) attains a strict global maximum at \( \bar{x} \in \mathbb{R}^n \) iff \( \nabla f(\bar{x}) = 0 \).

\(^1\) Clearly \( \frac{1}{2} \mathbf{d}x' H_f(\bar{x}) \mathbf{d}x \) is a continuous function of \( \mathbf{d}x \). Moreover, by assumption, the third derivative of \( f(\cdot) \) is continuous, so that (here’s the little bit of work) the remainder term is also a continuous function of \( \mathbf{d}x \). The continuity of \( \epsilon(\cdot) \) now follows easily.
4.8. Another application of Taylor

A second important application of Taylor’s theorem is the following result, which is obtained by essentially duplicating the proof of sufficiency above.

**Theorem:** Given $f: \mathbb{R}^n \to \mathbb{R}$ thrice continuously differentiable and $\bar{x} \in \mathbb{R}^n$, if $Hf$ is positive (negative) definite at $\bar{x}$ then there exists $\epsilon > 0$ such that the tangent plane to $f$ at $\bar{x}$ lies below (above) the graph of the function on the $\epsilon$-ball around $\bar{x}$.

**Proof:** We’ll do the case when $Hf$ is negative definite at $\bar{x}$. Once again, for any $d\mathbf{x} \neq 0$,

$$f(\bar{x} + d\mathbf{x}) - f(\bar{x}) = \nabla f(\bar{x})d\mathbf{x} + \frac{1}{2}d\mathbf{x}'Hf(\bar{x} + \lambda d\mathbf{x})d\mathbf{x}, \text{ for some } \lambda \in [0, 1] \quad (2)$$

Now, since $Hf(\bar{x})$ is negative definite by assumption, and $Hf(\cdot)$ is continuous, then if $d\mathbf{x}$ is sufficiently small, then $Hf(\bar{x} + \lambda d\mathbf{x})$ will be negative definite also, by continuity. Hence the second term in the expression above will be negative. Subtracting $\nabla f(\bar{x})d\mathbf{x}$ from both sides, we obtain that for all $d\mathbf{x}$ with $||d\mathbf{x}|| < \epsilon$

$$\left(\frac{f(\bar{x} + d\mathbf{x})}{\text{the height of } f \text{ at } \bar{x} + d\mathbf{x}}\right) - \left(\frac{(f(\bar{x}) + \nabla f(\bar{x})d\mathbf{x})}{\text{the height of the tangent plane at } \bar{x} + d\mathbf{x}}\right) = \frac{1}{2}d\mathbf{x}'Hf(\bar{x} + \lambda d\mathbf{x})d\mathbf{x} < 0 \quad (3)$$

Notice that the theorem above isn’t necessarily true without restricting $d\mathbf{x}$ to lie in a neighborhood of zero. Think of a camel. Put a tangent plane against the smaller hump, and the whole camel isn’t underneath the plane. On the other hand, if $Hf(\cdot)$ is *globally* negative definite (i.e., if $f$ is concave), then it can be shown that the tangent plane lies *everywhere* above the graph of the function. To establish this, we need only slightly perturb the proof of the above theorem.
Theorem: Given $f : \mathbb{R}^n \to \mathbb{R}$ thrice continuously differentiable and concave (convex), then for all $x \in \mathbb{R}^n$, the tangent plane to $f$ at $x$ lies everywhere above (below) the graph of the function.

Proof: The proof is identical to the proof of the preceding result except that we omit the caveat about $dx$ being small. Since $Hf(\cdot)$ is everywhere negative semi definite, then we know that regardless of the size of $dx$, the matrix $Hf(\bar{x} + \lambda dx)$ is negative-semi definite, so that the term $\frac{1}{2}dx'Hf(\bar{x} + \lambda dx)dx$ in expression (2) above is non-positive. Hence the left hand side of expression (3) above is also non-positive. \qed

Finally, for a strictly concave (convex) function the tangent plane at $x$ is strictly above (below) the graph everywhere except at $x$. To prove this, replace semi-definiteness in the proof above with definiteness, and the weak inequalities with strict ones.

4.9. Terminology Review

Keep these straight.

- The derivative of $f$ at $\bar{x}$: this is a point (which may be a scalar, vector or matrix).
  - partial derivative of $f$ w.r.t variable $i$ at $\bar{x}$.
  - directional derivative of $f$ in the direction $h$ at $\bar{x}$.
  - crosspartial derivative of $f_i$ w.r.t variable $j$ at $\bar{x}$.
  - total derivative of $f$ w.r.t variable $i$ at $\bar{x}$. The total derivative is different from the partial derivative w.r.t. $i$ iff other variables change as the $i$'th variable changes. In this case, the change in the other variables determine a direction; e.g., if $q$ depends on $p$, then a change in $p$ induces a change in $(p,q)$-space in the direction $h = (1, q'(p))$. 

However, whenever $q'(p) \neq 0$, the total derivative is \textit{strictly larger} than the directional derivative in the direction $h$.

- The derivative of $f$: this is a function. It’s the generic term
  - functions from $\mathbb{R}^1$ to $\mathbb{R}^1$: derivative is usually just called the derivative.
  - functions from $\mathbb{R}^n$ to $\mathbb{R}^1$: derivative is usually called the gradient.
  - functions from $\mathbb{R}^n$ to $\mathbb{R}^m$: derivative is called the Jacobian matrix
  - the Hessian of $f$ is the Jacobian of the derivative of $f$: i.e., the matrix of 2nd partial derivatives.

- The differential of $f$ at $\bar{x}$: this is also a function, but a different one from the gradient. It’s the unique linear function whose coefficient vector is the gradient vector evaluated at $\bar{x}$. Again, the differential may be a linear mapping from $\mathbb{R}^1$ to $\mathbb{R}^1$, from $\mathbb{R}^n$ to $\mathbb{R}^1$, or from $\mathbb{R}^n$ to $\mathbb{R}^m$, depending on the domain and range of the original function $f$. 