INTRODUCTION TO MAXIMUM PRINCIPLES

This is basically a set of class notes for 2/6/2001. I have incorporated suggestions that people made in class, so the material will move along more systematically than it did when I was trying to pull the next step out of the class.

0. Local Maxima on the Line. Everybody has learned, in some elementary calculus class or other, that if a function $f(x)$ is defined for x in some open interval $a < x < b$ of the real line, and if $x_{\max} \in (a, b)$ is a point such that for some subinterval $\alpha < x_{\max} < \beta$ it is true that $f(x_{\max}) \geq f(x)$ for every (other) $\alpha < x < \beta$ (which is what is meant by $x_{\max}$ being a (**weak**) **local maximum**), if $f'(x)$ exists for $\alpha < x < \beta$ and $f''(x_{\max})$ exists, then $f'(x_{\max}) = 0$ and $f''(x_{\max}) \leq 0$ must hold. There are various proofs of this, some using Taylor's theorem, but an elementary argument is the following: the limit defining $f'(x_{\max})$ is

$$f'(x_{\max}) = \lim_{\Delta x \to 0} \frac{f(x_{\max} + \Delta x) - f(x_{\max})}{\Delta x} \ .$$

If the limit is taken with $\Delta x > 0$ then the numerator is negative (or zero) and the denominator is positive, so the quotient is negative (or zero) and the limit can't be positive. If the limit is taken with $\Delta x < 0$ then the numerator is negative (or zero) and the denominator is negative, so the quotient is positive (or zero) and the limit can't be negative. The only value that the limit $f'(x_{\max})$ can take is therefore zero: first derivatives are zero at extrema, as everybody knows. Now if $f''(x_{\max}) > 0$ were true, we could reach a contradiction to the assumption that there was a local maximum at $x_{\max}$, as follows. Since we already know $f'(x_{\max}) = 0$ we would have

$$f''(x_{\max}) = \lim_{\Delta x \to 0} \frac{f'(x_{\max} + \Delta x) - f'(x_{\max})}{\Delta x} = \lim_{\Delta x \to 0} \frac{f'(x_{\max} + \Delta x)}{\Delta x} > 0 \ .$$

Since that limit was supposed to be positive, the values of $f'(x_{\max} + \Delta x)$ would be positive for $\Delta x > 0$ and negative for $\Delta x < 0$ provided Δx was sufficiently small; in other words, there would be an interval $A < x < B$ with $x_{\max}$ in its center, such that $f'(x) > 0$ if $x_{\max} < x < B$ and $f'(x) < 0$ if $A < x < x_{\max}$. But now, applying the MVTDC,[(1)] we see that (for example) for each x with $x_{\max} < x < B$ there would exist some "sample point" c_x with $x_{\max} < c_x < x < B$ for which

$$\frac{f(x) - f(x_{\max})}{x - x_{\max}} = f'(c_x) > 0 \ .$$

The r. h. s. is stricly positive because it is a value of $f'(\cdot)$ at a point for which $x_{\max} < c_x < x < B$. On the l. h. s., however, the denominator is positive, and therefore the numerator must be positive, so we must have $f(x) - f(x_{\max}) > 0$, or $f(x) > f(x_{\max})$. But that says: moving slightly to the right from $x_{\max}$ makes $f(x)$ take larger values. That contradicts the definition of a local maximum! Thus $f''(x_{\max}) > 0$ is untenable: we must have $f''(x_{\max}) \leq 0$. (One could make a similar argument from the left side. One could also pull out of these considerations the fact that $f'(x^*) = 0$ and $f''(x^*) > 0$ imply that x^* is a local **minimum** of the function $f(x)$, under fairly minimal hypotheses.)

1. The Maximum Principle on the Line. This is very little more than a negative way to state what we knew from freshman calculus.

The Maximum Principle on the Line: Suppose $f(x)$ is a function defined and continuous on a closed interval $[x_0, x_1] \subseteq \mathbb{R}$ and that $f(x)$ has the following two properties:

(1) $f'(x)$ exists at every point $x_0 < x < x_1$;

(2) $f''(x)$ exists and $f''(x) \geq 0$ at every point $x_0 < x < x_1$.

Then $\max\{f(x) : x_0 \leq x \leq x_1\} = \max\{f(x_0),\ f(x_1)\}$; in words, $f(x)$ takes its maximum value on $[x_0, x_1]$ at one of the "boundary points" x_0 or x_1.

(1) The Mean Value Theorem of the Differential Calculus. One can get awfully tired of typing out that long phrase.

Proof. Suppose the statement were false. Because it is a continuous function on a closed interval, $f(x)$ has to take an *absolute* maximum somewhere on the interval, and if we make the assumption $\max\{f(x) : x_0 \leq x \leq x_1\} > \max\{f(x_0),\ f(x_1)\}$ then the maximum is attained at some point x_2 with $x_0 < x_2 < x_1$. Applying the reasoning of §0 above, we see that $f''(x_2) \leq 0$ must hold. Unfortunately, this is not quite enough to contradict the assumption that $f''(x) \geq 0$ at every point $x_0 < x < x_1$. So how do we strengthen the argument?

Consider functions of the form $f_\epsilon(x) = f(x) + \epsilon(x - x_2)^2$, where $\epsilon > 0$. Such functions satisfy the same hypotheses about continuity, differentiability, etc., that $f(x)$ does, but they have the property $f_\epsilon''(x) = f''(x) + 2\epsilon > 0$ at every point $x_0 < x < x_1$. Under the assumption we're making about $f(x)$, we can choose a value of $\epsilon > 0$ such that $\max\{f_\epsilon(x) : x_0 \leq x \leq x_1\} > \max\{f_\epsilon(x_0),\ f_\epsilon(x_1)\}$. The reason is that the value of an f_ϵ at x_2 is the same as that of f, so if $\epsilon > 0$ satisfies the inequality

$$f(x_2) > f(x_1) + \epsilon(x_1 - x_2)^2 \quad \text{or equivalently}$$
$$\frac{f(x_2) - f(x_1)}{(x_2 - x_1)^2} > \epsilon$$

and similarly also

$$\frac{f(x_2) - f(x_0)}{(x_0 - x_2)^2} > \epsilon\ ,$$

then $f_\epsilon(x_2) = f(x_2) > \max\{f_\epsilon(x_0),\ f_\epsilon(x_1)\}$. This does not imply that the maximum value of $f_\epsilon(x)$ is attained at x_2, but because f_ϵ has to attain an absolute maximum *somewhere* on $[x_0, x_1]$ and we know that it takes a value at x_2 that is greater than either endpoint value, that maximum point must occur in the "interior," say at x_ϵ with $x_0 < x_\epsilon < x_1$ (there may be a maximum at more than one point, of course). But now we have reached a contradiction: $f_\epsilon''(x_\epsilon) \geq 2\epsilon > 0$ and by the freshman calculus calculus result of §0 above, we *can't* be at a maximum point of f_ϵ. The contradiction proves the maximum principle.

2. The Maximum Principle in Higher Dimensions. This discussion will mainly take place in the plane $\mathbb{R}^2$, because the step from dimension 1 to dimension 2 is the big step: the arguments that work in $\mathbb{R}^2$ will work, almost without alteration, in $\mathbb{R}^n$. The first question is: what is the condition(2) on $u(x, y)$ that generalizes the condition $f''(x) \geq 0$? One could require $u_{xx} \geq 0$ and $u_{yy} \geq 0$, but while that would work that is actually too strong: it suffices, for example, that the sum of those second partials be nonnegative. Here's the argument.

The Maximum Principle for the Laplace Operator in $\mathbb{R}^2$: Suppose $u(x, y)$ is a function defined and continuous on a closed and bounded set $D \subseteq \mathbb{R}^2$ and that $u(x, y)$ has the following two properties:

(1) u_x and u_y exist and are continuous at every interior point $(x, y) \in D$;

(2) $u_{xx}(x)$ and u_{yy} exist and $\nabla^2 u \geq 0$ at every interior point of D.

Then $\max\{u(x, y) : (x, y) \in D\} = \max\{u(x, y) : (x, y) \in \partial D\}$; in words, $u(x, y)$ takes its maximum value on D at one (or more) of the "boundary points" of D.

Proof. Basically, we assume the contrary and fight the ensuing contradiction back to what we already know about the 1-dimensional case. Suppose, therefore, that the hypotheses placed on $u(x, y)$ above are satisfied but that somehow $\max\{u(x, y) : x \in D\} > \max\{u(x, y) \in \partial D\}$. Then, as before, there is some *interior* point $(x_2, y_2) \in D$ at which an absolute maximum is attained. We consider functions of the form $u_\epsilon(x, y) + \epsilon \cdot [(x - x_2)^2 + (y - y_2)^2]$ on D; the Laplacian of such a function is

$$\nabla^2 u_\epsilon(x, y) = \nabla^2\{u(x, y) + \epsilon \cdot [(x - x_2)^2 + (y - y_2)^2]\} = \nabla^2 u(x, y) + 4\epsilon \geq 4\epsilon > 0\ .$$

The "fudge term" $\epsilon \cdot [(x - x_2)^2 + (y - y_2)^2]$ is basically ϵ-times the square of the distance from (x_2, y_2) to the "moveable point" (x, y). Because D is a bounded set, the distances from (x_2, y_2) to points $(x, y) \in \partial D$

(2) Actually, the answer is "there are many generalizations." As we shall see (or as Strauss proves on pp. 41–42), there is a version for the heat equation (or heat inequality), and there is a version for every elliptic operator of 2nd order with no 0-th order term.

cannot be larger than a certain $\delta > 0$, so the $\epsilon \cdot [(x-x_2)^2 + (y-y_2)^2]$ term cannot be larger than $\epsilon \cdot \delta^2$. If we take $\epsilon > 0$ small enough that the inequality

$$\frac{u(x_2, y_2) - \max\{u(x,y) \in \partial D\}}{\delta^2} > \epsilon$$

is satisfied, then—because $u_\epsilon(x_2, y_2) = u(x_2, y_2)$—the value of u_ϵ at (x_2, y_2) is larger than its maximum value on the boundary, and so u_ϵ must attain its maximum value on D at some point (a, b) in the interior of D. Well, now we're in trouble as usual. On any small interval of the horizontal line $y = b$ that contains (a, b) and lies entirely in D—and such intervals exist, because (a, b) is an interior point of D—the value $u_\epsilon(a, b)$ is an absolute maximum of the function of x given by $x \mapsto u(x, b)$. This cannot happen if $(u_\epsilon)_{xx}(a, b) > 0$, so $(u_\epsilon)_{xx}(a, b) \le 0$. By the same argument applied to vertical lines $x = a$, $(u_\epsilon)_{yy}(a, b) \le 0$. But now we have

$$0 < 4\epsilon \le \nabla^2 u_\epsilon(a, b) = (u_\epsilon)_{xx}(a, b) + (u_\epsilon)_{yy}(a, b) \le 0 + 0 = 0$$

and we have reached a contradiction, starting with the assumption that $u(x, y)$ took a larger value in the interior of D than its maximum value on ∂D. Hence that assumption is untenable, and the maximum principle must hold.

The only special property of the Laplacian used in the argument given above is that it is an elliptic 2nd-order operator with no 0-th-order term. For constant-coefficient operators one can see this by using Strauss's discussion in §1.6, pp. 29–30, which shows that a change of coördinates will convert such an elliptic 2nd-order operator into an operator of the form $Lu = \nabla^2 u + a_1 u_x + a_2 u_y$. The argument just given works just as well for such an operator, because at a (putative) absolute maximum point (a, b) interior to D the first partial derivatives $(u_\epsilon)_x(a, b) = 0$ and $(u_\epsilon)_y(a, b) = 0$ anyway. The result also holds for nonconstant-coefficient elliptic 2nd-order operators with no 0-th-order term, but one has to go back and reëxamine the construction given above. The interested reader is invited to do this.

3. The Maximum Principle for the Heat Operator. This is essentially the same thing that Strauss does on pp. 41–42, but done for the heat equation in the plane, so that the modifications necessary in higher dimensions can be examined. The analogue of a rectangle for the heat equation in the plane is a "box" or "product set" of the form

$$V = D \times [0, T] = \{(x, y, t) : (x, y) \in D,\ 0 \le t \le T\}$$

where $D \subseteq \mathbb{R}^2$ is a closed and bounded subset of the plane (the construction for dimensions ≥ 3 is similar). The "**distinguished boundary**" or **Šilov boundary** of V is the set $\check{V} = (D \times \{0\}) \cup (\partial D \times [0, T])$ consisting of all the points (x, y, t) such that either $t = 0$ or $(x, y) \in \partial D$ (or both); we want to prove

Maximum Principle for the Heat Operator: Suppose $u(x, y, t)$ is a function defined and continuous on a set V as described above, and that $u(x, y, t)$ has the following two properties:

(1) u_x, u_y and u_t exist and are continuous at every interior point $(x, y, t) \in V$, and u_t exists at least one-sidedly at points (x, y, T) where (x, y) is an interior point of D;

(2) $u_{xx}(x)$ and u_{yy} exist and $k\nabla^2 u - \dfrac{\partial u}{\partial t} \ge 0$ at every interior point of V and at points (x, y, T) where (x, y) is an interior point of D.

Then $\max\{u(x, y, t) : (x, y, t) \in V\} = \max\{u(x, y, t) : (x, y, t) \in \check{V}\}$; in words, $u(x, y, t)$ takes its maximum value on V at one of the "distinguished boundary points" of V.

Proof: This is going to look very familiar. Suppose the statement fails for some such function $u(x, y, t)$. Then the absolute maximum value of u on V is attained at some point $(x_2, y_2, t) \in V$ which is either of the form (x_2, y_2, t_2) with (x_2, y_2) in the interior of D and $0 < t_2 < T$, or of the form (x_2, y_2, T), where (x_2, y_2) is an interior point of D. We replace $u(x, y, t)$ by $u_\epsilon(x, y, t) = u(x, y, t) + \epsilon[(x-x_2)^2 + (y-y_2)^2]$ with $\epsilon > 0$ taken sufficiently small that $u_\epsilon(x_2, y_2, t_2) = u(x_2, y_2, t_2) > \max\{u_\epsilon(x, y, t) : (x, y, t) \in \check{V}\}$, just as we did for the Laplace operator. For u_ϵ we have $k\nabla^2 u_\epsilon - \dfrac{\partial u_\epsilon}{\partial t} = k\nabla^2 u + 4\epsilon - \dfrac{\partial u}{\partial t} \geq 4\epsilon > 0$. $u_\epsilon(x, y, t)$ must still take an absolute maximum at some point $(a, b, t_0) \in V$ which is either in the interior of the set V or has the form (a, b, T) for some point (a, b) in the interior of D. In the first case, we are in the interior of the set $V \subseteq \mathbb{R}^3$, and therefore $(u_\epsilon)_t(a, b, t_0) = 0$ and the inequality $k\nabla^2 u_\epsilon - \dfrac{\partial u_\epsilon}{\partial t} \geq 4\epsilon$ becomes $k\nabla^2 u(a, b, t_0) \geq 4\epsilon > 0$. The function $(x, y) \mapsto u_\epsilon(x, y, t_0)$, defined in D, then has an absolute maximum at the interior point (a, b)—but at this point its Laplacian $\nabla^2 u(a, b, t_0) \geq 4\epsilon/k > 0$, and as we have already seen, the Laplacian of a function cannot be strictly positive at a maximum point. In the second case, we are at a point (a, b, T) where (a, b) is an interior point of D. Since the value $u_\epsilon(a, b, T)$ is an absolute maximum, $u_\epsilon(a, b, T - \Delta t) \leq u_\epsilon(a, b, T)$ for any $\Delta t > 0$, and consequently the one-sided derivative

$$\frac{\partial u_\epsilon(a, b, T)}{\partial t} = \lim_{\Delta t \to 0^+} \frac{u_\epsilon(a, b, T - \Delta t) - u_\epsilon(a, b, T)}{-\Delta t} \geq 0$$

since each difference quotient has a nonpositive numerator and a negative denominator. In this situation the inequality $k\nabla^2 u_\epsilon - \dfrac{\partial u_\epsilon}{\partial t} \geq 4\epsilon$ becomes $k\nabla^2 u_\epsilon(a, b, T) \geq \dfrac{\partial u_\epsilon(a, b, T)}{\partial t} + 4\epsilon \geq 4\epsilon > 0$, so again the function $(x, y) \mapsto u_\epsilon(x, y, T)$, defined in D, has an absolute maximum at the interior point (a, b)—at which its Laplacian is strictly positive. This final contradiction shows that the absolute maximum of u on V must have occurred on $\check{V}$.