Multivariable Calculus – Lecture #7 Notes

The central theme of this lecture is the existence of extrema (maxima and minima) for continuous functions defined on a closed and bounded set (The Extreme Value Theorem), and how to find where these extrema occur. This provides methods for optimization. We will define stationary points and test them using a Second Derivative Test, and we'll also exploit the geometry of the gradient vector to create a method for solving constrained optimization problems (The Method of Lagrange Multipliers).

Calculus Theorems for which the proofs will have to wait

There are only a few theorems in basic Calculus for which "the proof is beyond the scope of this text." These include the Monotone Convergence Theorem, the Intermediate Value Theorem, and the Extreme Value Theorem. The first is a statement about the "completeness" of the real numbers, and the last are statements about continuous functions. Specifically, the Intermediate Value Theorem states that the continuous image of a "connected" set will also be connected, and the Extreme Value Theorem states that the continuous image of a "compact" set (closed and bounded in our context) will also be compact.

Extreme Value Theorem – Suppose a real-valued function $f(x_1,\cdots,x_n)$ is continuous on a closed and bounded domain $D\subset\mathbf{R}^n$. Then this function must attain its maximum and minimum values somewhere within this domain. [Note: You can think of a closed set as one for which any convergent sequence in this set has its limit in the set. In particular, this means that the boundary of the set must be included.]

Note: By considering closed subsets of this domain, the Extreme Value Theorem also guarantees the existence (and attainment) of relative maxima and minimum within any closed and bounded set.

Recall from one-variable Calculus that when seeking maxima and minima of a continuous function defined on a closed interval you would seek stationary points (points where the derivative was zero), points where the derivative was not defined (e.g. cusps and points where there was a vertical tangent line), and you would also check the endpoints. These were the only places where relative (or absolute) maxima and minima could possibly occur. You may also recall that a stationary point could yield a relative maximum or relative minimum, but it could also yield a point of inflection. The Second Derivative Test was a handy tool for helping to distinguish these possibilities.

The situation for a function $f(x_1,\cdots,x_n)$ of several variables is fundamentally the same. Potential maxima or minima will occur at:

(a) Stationary points (points where $\overrightarrow{\nabla f}=\mathbf{0}$, i.e. where all of the partial derivatives of f vanish); or
(b) Non-differentiable points (points where the given function is not differentiable); or
(c) On the boundary of the domain.

We generally refer to stationary points and non-differentiable points as ***critical points***. These will usually occur in the interior of the given domain, but they could potentially occur on the boundary. The usual routine for finding extrema of a function is to (1) seek interior critical points, and (2) examining the boundary for candidates where extrema might occur. We will soon develop good methods for doing the latter.

Quadratic approximation and the Second Derivative Test (for functions of two variables)

We are already familiar with linear approximation of a function $f(x,y)$ in the vicinity of a point (x_0,y_0):

$$f(x,y)\cong L(x,y)=f(x_0,y_0)+f_x(x_0,y_0)(x-x_0)+f_y(x_0,y_0)(y-y_0)$$

Since this is based on the idea of approximating the graph of $f(x,y)$ by a tangent plane, it will not be especially helpful in distinguishing whether a stationary point yields a local maximum or a local minimum or something like a saddle point. The tangent planes will be horizontal in all of these cases. However, if we use second derivative information we may be able to get a better grasp of the shape of the graph and possibly be able to distinguish these cases.

To do this, let's suppose we want to approximate the graph of $f(x,y)$ by a quadratic function of two variables. To this end, we would want to find the appropriate choices of the coefficients A,B,C,D,E,F so that:

$$f(x,y) \cong Q(x,y) = A + B(x-x_0) + C(y-y_0) + D(x-x_0)^2 + E(x-x_0)(y-y_0) + F(y-y_0)^2$$

To get the best fit, we would want to ensure that at (x_0, y_0) we match the values, slopes, concavity (and other quantities relating to shape) for the actual function $f(x,y)$ and its approximating quadratic function $Q(x,y)$. That is, we must require that:

$Q(x_0,y_0) = f(x_0,y_0)$ to match values

$Q_x(x_0,y_0) = f_x(x_0,y_0)$ and $Q_y(x_0,y_0) = f_y(x_0,y_0)$ to match slopes

$H_Q(x_0,y_0) = H_f(x_0,y_0)$ to match 2nd derivative information

Here H_Q and H_f are the Hessian matrices of 2$^{\text{nd}}$ derivatives.

A quick calculation gives that:

$Q(x_0,y_0) = \boxed{A = f(x_0,y_0)}$, $Q_x(x_0,y_0) = \boxed{B = f_x(x_0,y_0)}$, and $Q_y(x_0,y_0) = \boxed{C = f_y(x_0,y_0)}$ which we expected from the expression for linear approximation.

We also calculate that $H_Q(x_0,y_0) = \begin{bmatrix} 2D & E \\ E & 2F \end{bmatrix} = H_f(x_0,y_0) = \begin{bmatrix} f_{xx}(x_0,y_0) & f_{xy}(x_0,y_0) \\ f_{yx}(x_0,y_0) & f_{yy}(x_0,y_0) \end{bmatrix}$, so $2D = f_{xx}(x_0,y_0)$, $E = f_{xy}(x_0,y_0) = f_{yx}(x_0,y_0)$, and $2F = f_{yy}(x_0,y_0)$.

Therefore $\boxed{D = \tfrac{1}{2} f_{xx}(x_0,y_0)}$, $\boxed{E = f_{xy}(x_0,y_0)}$, and $\boxed{F = \tfrac{1}{2} f_{yy}(x_0,y_0)}$ and the best quadratic approximation is:

$$\boxed{\begin{aligned} f(x,y) \cong Q(x,y) &= f(x_0,y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0) \\ &+ \tfrac{1}{2} f_{xx}(x_0,y_0)(x-x_0)^2 + f_{xy}(x_0,y_0)(x-x_0)(y-y_0) + \tfrac{1}{2} f_{yy}(x_0,y_0)(y-y_0)^2 \end{aligned}}$$

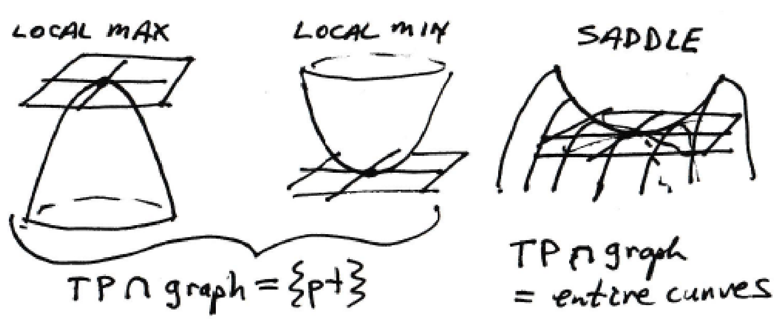
Note that the quadratic terms simply improve upon the linear approximation.

Second Derivative Test

If (x_0,y_0) is a stationary point, then $f_x(x_0,y_0) = 0$ and $f_y(x_0,y_0) = 0$ and the tangent plane will be horizontal, i.e. the plane will have the simple equation $z = f(x_0,y_0)$. The approximating quadratic will be:

$$z = f(x_0,y_0) + \tfrac{1}{2} f_{xx}(x_0,y_0)(x-x_0)^2 + f_{xy}(x_0,y_0)(x-x_0)(y-y_0) + \tfrac{1}{2} f_{yy}(x_0,y_0)(y-y_0)^2 .$$

For a local maximum or a local minimum, the graph and the tangent plane will have only a single intersection point, but for a saddle point the intersection will consist of entire curves.



If we use the quadratic approximation as a substitute for the actual function for points (x,y) near (x_0,y_0) and equate values to determine the intersection with the horizontal tangent plane, we get the equation:

$z = f(x_0,y_0) = f(x_0,y_0) + \frac{1}{2} f_{xx}(x_0,y_0)(x-x_0)^2 + f_{xy}(x_0,y_0)(x-x_0)(y-y_0) + \frac{1}{2} f_{yy}(x_0,y_0)(y-y_0)^2$ or, more simply, $\frac{1}{2} f_{xx}(x_0,y_0)(x-x_0)^2 + f_{xy}(x_0,y_0)(x-x_0)(y-y_0) + \frac{1}{2} f_{yy}(x_0,y_0)(y-y_0)^2 = 0$. If, for simplicity, we write $h = x - x_0$ and $k = y - y_0$, and if we temporarily suppress the point at which things are evaluated, we have: $\frac{1}{2} f_{xx}h^2 + \left[f_{xy}h \right] k + \left[\frac{1}{2} f_{yy} \right] k^2 = 0$. We can solve for $k = k(h)$ using the quadratic formula. This yields:

$$k = \frac{-f_{xy}h \pm \sqrt{f_{xy}{}^2 h^2 - f_{xx}f_{yy}h^2}}{f_{yy}} = \frac{-f_{xy}h \pm |h| \sqrt{f_{xy}{}^2 - f_{xx}f_{yy}}}{f_{yy}}$$

We don't really care so much what the solutions are other than to determine whether there actually are any solutions other than the one point where $h = k = 0$, i.e. the stationary point (x_0, y_0). This is completely determined by the sign of the discriminant $f_{xy}{}^2 - f_{xx}f_{yy}$. If it's positive there will be additional solutions (saddle point). If it's negative there will be no additional solutions (local minimum or local maximum). In the borderline case where the discriminant is equal to zero, it would be best not to draw any conclusions since the analysis is based only on the quadratic approximation, and that may be inadequate in this case. It's easier to work with the negative of the discriminant since we observe that $\det(H_f) = \det \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} = f_{xx}f_{yy} - f_{xy}{}^2$.

Putting these observations together and using concavity in the cross-sections to distinguish a local maximum from a local minimum, we can state the following:

Second Derivative Test: Suppose (x_0, y_0) is a stationary point of a function $f(x,y)$ and that this function is at least twice differentiable. Then

(1) If $\det(H_f(x_0,y_0)) > 0$, then (x_0, y_0) will give either a local maximum or a local minimum.

(a) If, in addition, $f_{xx}(x_0,y_0) > 0$, then the stationary point will yield a local minimum.

(b) If, in addition, $f_{xx}(x_0,y_0) < 0$, then the stationary point will yield a local maximum.

(2) If $\det(H_f(x_0,y_0)) < 0$, then (x_0, y_0) will give a saddle point (neither maximum nor minimum).

(3) If $\det(H_f(x_0,y_0)) = 0$, then the test is inconclusive.

Note: For functions of more than two variables, there's a different version of the Second Derivative Test that utilizes the idea of quadratic forms that is expressed in terms of the eigenvalues of the (symmetric) Hessian matrix. It makes use of the Spectral Theorem from Linear Algebra.

Example 1: Find the stationary points for the function $f(x,y) = x^3 - 3x + y^3 - 12y$ and test each to determine which yield local maxima, local minima, or saddle points.

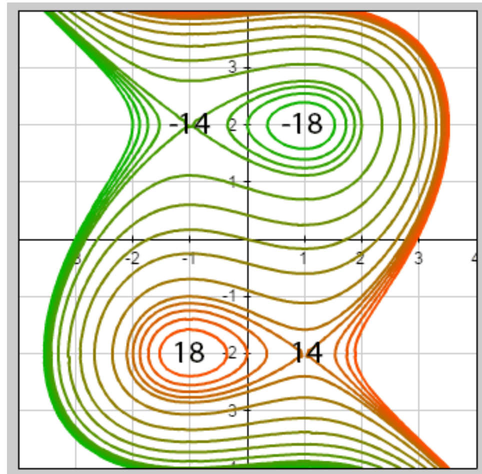
Solution: Setting the 1st partial derivatives equal to zero gives $f_x = 3x^2 - 3 = 3(x^2 - 1) = 0$ and $f_y = 3y^2 - 12 = 3(y^2 - 4) = 0$, and these imply that $x = \pm 1$ and $y = \pm 2$. We therefore have four stationary points: $(1,2)$, $(1,-2)$, $(-1,2)$, and $(-1,-2)$. The Hessian matrix of 2nd partial derivatives is $H_f = \begin{bmatrix} 6x & 0 \\ 0 & 6y \end{bmatrix}$. We then test each of the stationary points:

$H_f(1,2) = \begin{bmatrix} 6 & 0 \\ 0 & 12 \end{bmatrix}$, so $\det H_f(1,2) = 72 > 0$, so the stationary point gives either a local maximum or a local minimum. Furthermore, $f_{xx}(1,2) = 6 > 0$, so the graph is concave up in cross-section and the stationary point must therefore yield a local minimum.

$H_f(1,-2)=\begin{bmatrix}6 & 0\\ 0 & -12\end{bmatrix}$, so $\det H_f(1,-2)=-72<0$, so the stationary point gives a saddle point

$H_f(-1,2)=\begin{bmatrix}-6 & 0\\ 0 & 12\end{bmatrix}$, so $\det H_f(-1,2)=-72<0$, so the stationary point gives a saddle point

$H_f(-1,-2)=\begin{bmatrix}-6 & 0\\ 0 & -12\end{bmatrix}$, so $\det H_f(-1,-2)=72>0$, so the stationary point gives either a local maximum or a local minimum. Furthermore, $f_{xx}(-1,-2)=-6<0$, so the graph is concave down in cross-section and the stationary point must therefore yield a local maximum.



As was the case in one-variable Calculus where plotting critical points and their local qualities allowed us to "interpolate" the rest of the graph, we can also do something analogous for a function of two variables. In this example, if we evaluate to get $f(1,2)=-18$, $f(1,-2)=14$, $f(-1,2)=-14$, and $f(-1,-2)=18$ and use the above analysis to determine what the level sets must look like in the vicinity of each stationary point, this helps to explain the level sets and thus an understanding of the graph.

Example 2: Find the point on the plane with equation $2x-y+z=5$ that is closest to the origin.

Solution: This can be done without Calculus using some basic geometry and parametrization of a normal line to the plane passing through the origin. Specifically, we observe that the normal vector to the plane is $\mathbf{n}=\langle 2,-1,1\rangle$ and the normal line can be parameterized by $\left\{\begin{matrix} x=2t\\ y=-t\\ z=t\end{matrix}\right\}_{t\in\mathbf{R}}$. This line pierces the given plane when $2(2t)-(-t)+(t)=6t=5$, so $t=\frac{5}{6}$, and therefore the closest point is $(x,y,z)=(\frac{5}{3},-\frac{5}{6},\frac{5}{6})$.

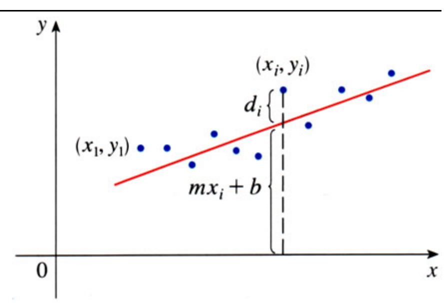
We can view this as an optimization problem by finding which point on the plane minimizes the distance to the origin. For the sake of algebraic simplicity, it's easier to minimize the square of the distance.

For any point (x,y,z) on the plane, the square of the distance is given by $D^2=x^2+y^2+z^2$. It's worth noting that this is minimized at the origin, but this is not a point on the plane. So we have to somehow incorporate the ***constraint*** that $2x-y+z=5$. One way to do this is to simply absorb this into the function to be minimized via substitution. For example, we can solve for $z=5-2x+y$ and substitute to get $D^2=x^2+y^2+(5-2x+y)^2=F(x,y)$. This function of two variable now is unconstrained, i.e. we can freely vary x and y. We must therefore seek its stationary points. We have:

$$\left\{\begin{matrix} F_x=2x+2(5-2x+y)(-2)=10x-4y-20=0\\ F_y=2y+2(5-2x+y)(1)=-4x+4y+10=0\end{matrix}\right\} \Rightarrow \left\{\begin{matrix} 10x-4y=20\\ -4x+4y=-10\end{matrix}\right\} \Rightarrow x=\tfrac{5}{3},\ y=-\tfrac{5}{6}$$

Substitution then gives $z=5-2(\frac{5}{3})+(-\frac{5}{6})=\frac{5}{6}$, so $(x,y,z)=(\frac{5}{3},-\frac{5}{6},\frac{5}{6})$. You can check this using the 2nd derivative test if you wish, but since there is a closest point and we have found just this one candidate for this closest point, this must be it.

Method of Least Squares



A situation frequently encountered is that of fitting a curve (or surface or other relation) to a data set. There are many ways to approach this problem, but one way is to assume a particular model (such as a straight line relationship) and find the line that best fits the data. Suppose we have a data set that consists of ordered pairs $\{(x_i, y_i\}_{i=1}^{i=N}$. If we plot these N points, we get a *scatterplot*. If this scatterplot suggests a pattern that might be explained via a straight line relationship $y = mx + b$, then we have to find what values of the slope m and intercept b yield a line that best fits this data. If we view $y = mx + b$ as the theoretical model, then we can for each data point measure the difference between the theoretical $\hat{y}_i = mx_i + b$ and the actual measure value y_i. That is, we consider $\hat{y}_i - y_i = mx_i + b - y_i$. The Method of Least Squares is based on the idea that we can get a best fit by minimizing the sum of the squares of these deviations. This notion is analogous with the idea of line-of-sight distance between points where the square of the distance is the sum of the squares of the differences in their respective coordinates.

So, we want to minimize $F(m,b) = \sum_{i=1}^{N} (mx_i + b - y_i)^2$. This is an unconstrained optimization problem that is solved by finding the stationary points of $F(m,b)$. These will occur where $\frac{\partial F}{\partial m} = 0$ and $\frac{\partial F}{\partial b} = 0$.

Using standard rules for differentiation as well as the distributive law, we calculate that:

$$\frac{\partial F}{\partial m} = \sum_{i=1}^{N} 2(mx_i + b - y_i)(x_i) = 2\sum_{i=1}^{N} (mx_i^2 + bx_i - x_i y_i) = 2\left(\left(\sum_{i=1}^{N} x_i^2\right)m + \left(\sum_{i=1}^{N} x_i\right)b - \sum_{i=1}^{N} x_i y_i\right) = 0$$

$$\frac{\partial F}{\partial b} = \sum_{i=1}^{N} 2(mx_i + b - y_i)(1) = 2\sum_{i=1}^{N} (mx_i + b - y_i) = 2\left(\left(\sum_{i=1}^{N} x_i\right)m + Nb - \sum_{i=1}^{N} y_i\right) = 0$$

These two equations yield the linear system (in m and b): $\left\{\begin{array}{l} \left(\sum_{i=1}^{N} x_i^2\right)m + \left(\sum_{i=1}^{N} x_i\right)b = \sum_{i=1}^{N} x_i y_i \\ \left(\sum_{i=1}^{N} x_i\right)m + \qquad Nb \quad = \sum_{i=1}^{N} y_i \end{array}\right\}$.

These can also be expressed in matrix form as: $\begin{bmatrix} \sum_{i=1}^{N} x_i^2 & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & N \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} x_i y_i \\ \sum_{i=1}^{N} y_i \end{bmatrix}$.

Example 3: Find the line that best fits the data $\{(2,1),(3,1),(4,2),(5,3)\}$.

Solution: It is perhaps worth noting that the Method of Least Squares should really be used for much larger data sets, but this example will illustrate the practical aspects of the method. It's generally best to make a table with all relevant given and derived information:

	x_i	y_i	x_i^2	$x_i y_i$
	2	1	4	2
	3	1	9	3
	4	2	16	8
	5	3	25	15
Σ	14	7	54	28

The associated matrix is then $\begin{bmatrix} 54 & 14 \\ 14 & 4 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 28 \\ 7 \end{bmatrix}$ which yields $\begin{bmatrix} m \\ b \end{bmatrix} = \frac{1}{20}\begin{bmatrix} 4 & -14 \\ -14 & 54 \end{bmatrix}\begin{bmatrix} 28 \\ 7 \end{bmatrix} = \frac{1}{20}\begin{bmatrix} 14 \\ -14 \end{bmatrix} = \begin{bmatrix} 0.7 \\ -0.7 \end{bmatrix}$.

So the line of best fit has equation $\boxed{y = .7x - .7}$.

It's worth noting that the Method of Least Squares can also be used to fit data to other simple models. For example, if we felt that a quadratic model of the form $y = ax^2 + bx + c$ might better explain the data $\{(x_i, y_i)\}_{i=1}^{i=N}$, we could look at $F(a,b,c) = \sum_{i=1}^{N} (ax_i^2 + bx_i + c - y_i)^2$ and again find the stationary points by setting all the first partial derivatives equal to 0. After a little calculus and algebra, this would yield the matrix system:

$$\begin{bmatrix} \sum_{i=1}^{N} x_i^4 & \sum_{i=1}^{N} x_i^3 & \sum_{i=1}^{N} x_i^2 \\ \sum_{i=1}^{N} x_i^3 & \sum_{i=1}^{N} x_i^2 & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i^2 & \sum_{i=1}^{N} x_i & N \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} x_i^2 y_i \\ \sum_{i=1}^{N} x_i y_i \\ \sum_{i=1}^{N} y_i \end{bmatrix}$$

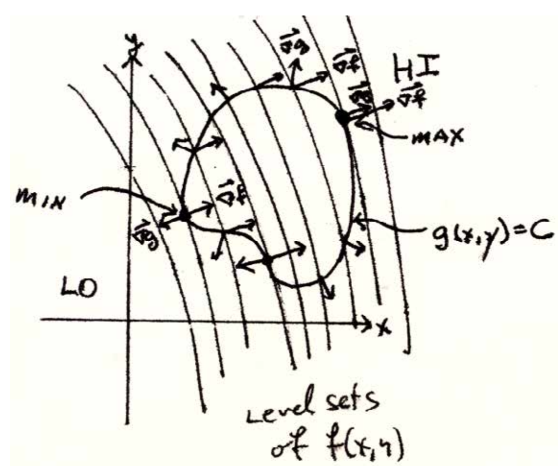
We would then solve this to get a, b, and c to determine the best-fitting parabola with equation $y = ax^2 + bx + c$.

Constrained optimization and the Method of Lagrange Multipliers

We often have to find the extrema of a differentiable function of several variables in a situation where there are constraints on the variables. For example, in Example 2 we were seeking the minimum of $f(x,y,z) = x^2 + y^2 + z^2$ subject to the constraint that $2x - y + z = 5$. As an alternative to solving for one variable in terms of the others and substituting this into the function to reduce it to an unconstrained function of fewer variables, we can instead use the **Method of Lagrange Multipliers**. Perhaps an even greater benefit of this method is that it produces relations that must be true for any optimal solution. The method is based on geometric facts about gradient vectors.

Let's look at two situations to motivate the method:

(A) Suppose we want to find the extrema of a function $f(x,y)$ subject to a single constraint of the form $g(x,y) = c$ (constant). For simplicity, let's also assume that this constraint represents a smooth curve.



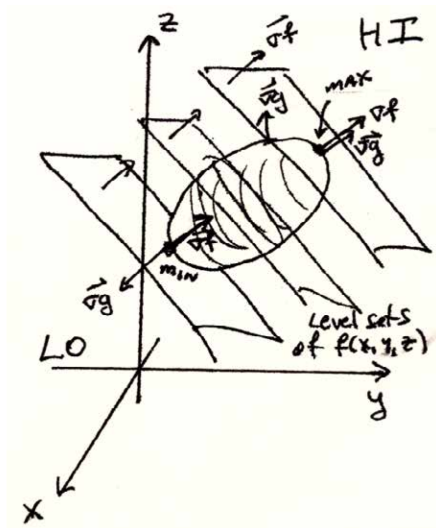
The diagram shows the relationship between the level sets of the function $f(x,y)$ and the constraint curve, especially in the vicinity of the extreme values. If you imagine moving along the constraint as you approach the maximum value and then head toward lower values, you glance off a level set of $f(x,y)$ tangentially. Note that the gradient $\overrightarrow{\nabla f}$ (when nonzero) is always perpendicular to the level sets of f. The constraint, when expressed in the form $g(x,y) = c$ is a level set of g, so $\overrightarrow{\nabla g}$ must be perpendicular to the constraint curve. We observe from the diagram that at the point where the maximum occurs, $\overrightarrow{\nabla f}$ and $\overrightarrow{\nabla g}$ must be parallel. Therefore $\overrightarrow{\nabla f} = \lambda \overrightarrow{\nabla g}$ at the maximum point for some scalar λ (called a *Lagrange multiplier*). The same argument also applies at a point where a minimum occurs. As the diagram indicates, it's also possible that this condition may hold at a point that's neither a maximum nor a minimum (very much like a point of inflection). If we take this parallelism condition together with the constraint, these comprise the **Method of Lagrange Multipliers**.

$$\begin{Bmatrix} \overrightarrow{\nabla f} = \lambda \overrightarrow{\nabla g} \\ g(x,y) = c \end{Bmatrix}$$

Note that the first condition provides two equations $f_x = \lambda g_x$ and $f_y = \lambda g_y$ (from the two components), so there is a total of three equations in the three variables $\{x, y, \lambda\}$, so it may be possible to solve these equations.

(B) Suppose we want to find the extrema of a function $f(x, y, z)$ subject to a single constraint of the form $g(x, y, z) = c$ (constant). For simplicity, let's also assume that this constraint represents a smooth surface.

The diagram shows the relationship between the level surfaces of the function $f(x, y, z)$ and the constraint surface, especially in the vicinity of the extreme values. If you imagine moving along the constraint approaching the maximum value and then heading toward lower values, you glance off a level surface of $f(x, y, z)$ tangentially. The gradient $\overrightarrow{\nabla f}$ (when nonzero) is always perpendicular to these level surfaces. The constraint, when expressed in the form $g(x, y, z) = c$ is a level set of g, so $\overrightarrow{\nabla g}$ must be perpendicular to this constraint surface. We again observe from the diagram that at the point where the maximum occurs, $\overrightarrow{\nabla f}$ and $\overrightarrow{\nabla g}$ must be parallel. Therefore $\overrightarrow{\nabla f} = \lambda \overrightarrow{\nabla g}$ at the maximum point for some scalar λ. The same argument also applies at a point where a minimum occurs. It's also possible that this condition may hold at a point that's neither a maximum nor a minimum. One again, if we take this parallelism condition together with the constraint, the **Method of Lagrange Multipliers** can be expressed as:



$$\begin{Bmatrix} \overrightarrow{\nabla f} = \lambda \overrightarrow{\nabla g} \\ g(x, y, z) = c \end{Bmatrix}$$

Note that the first condition now provides three equations $f_x = \lambda g_x$, $f_y = \lambda g_y$, and $f_z = \lambda g_z$ (from the three components), so there is a total of four equations in the three variables $\{x, y, z, \lambda\}$, so it may be possible to solve these equations.

These observations can also be proven using the Chain Rule. Suppose we parameterize a differentiable curve that satisfies the constraint at every point, and then evaluate the given function at all points of this curve. In the first instance, we can diagram this as a composition:

$$t \to (x(t), y(t)) \to f(x(t), y(t))$$

The Chain Rule then gives that $\frac{d}{dt}\left[f(x(t), y(t))\right] = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = \overrightarrow{\nabla f} \cdot \mathbf{v} = 0$ at any stationary point for this function of t. Since the velocity vector is tangent to the constraint curve, this means that $\overrightarrow{\nabla f}$ must be perpendicular to the constraint and therefore parallel to $\overrightarrow{\nabla g}$ (since this gradient is everywhere perpendicular to the $g(x, y) = c$ constraint). Therefore $\overrightarrow{\nabla f} = \lambda \overrightarrow{\nabla g}$ in addition to the constraint $g(x, y) = c$.

In the case of a function of three variables, the argument is basically the same:

$$t \to (x(t), y(t), z(t)) \to f(x(t), y(t), z(t))$$

The Chain Rule then gives that $\frac{d}{dt}\left[f(x(t), y(t), z(t))\right] = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} = \overrightarrow{\nabla f} \cdot \mathbf{v} = 0$ at any stationary point for this function of t. Since the velocity vector is tangent to the constraint curve, this means that $\overrightarrow{\nabla f}$ must be perpendicular to the constraint and therefore parallel to $\overrightarrow{\nabla g}$ (since this gradient is everywhere perpendicular to the $g(x, y, z) = c$ constraint). Therefore $\overrightarrow{\nabla f} = \lambda \overrightarrow{\nabla g}$ in addition to the constraint $g(x, y, z) = c$.

The same argument also works for a function of n variables with one constraint. We will also soon address what happens in the presence of multiple constraints.

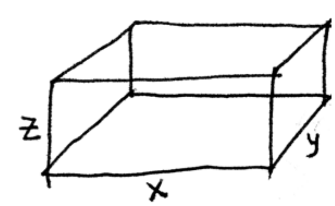
Example 4: Let's redo the previous problem of finding the point on the plane $2x-y+z=5$ closest to the origin. The (square of the) distance from the origin to any point is $f(x,y,z)=x^2+y^2+z^2$, and the plane represents a constraint $g(x,y,z)=2x-y+z=5$. The Lagrange Multiplier condition yields the three equations:

$$\begin{Bmatrix} f_x=\lambda g_x \\ f_y=\lambda g_y \\ f_z=\lambda g_z \end{Bmatrix} \Rightarrow \begin{Bmatrix} 2x=\lambda\cdot 2 \\ 2y=\lambda\cdot(-1) \\ 2z=\lambda\cdot 1 \end{Bmatrix}$$

We don't necessarily care what λ is, so dividing respective sides of these equations will eliminate λ. Dividing the 1st and 2nd equations yields $\frac{x}{y}=-2$ or $\boxed{x=-2y}$. Dividing the 3rd and 2nd equations yields $\frac{z}{y}=-1$ or $\boxed{z=-y}$. If we substitute these into the constraint we get $-4y-y-y=-6y=5$, so $y=-\frac{5}{6}$. Therefore $x=\frac{5}{3}$ and $z=\frac{5}{6}$. So the only candidate is the point $(x,y,z)=(\frac{5}{3},-\frac{5}{6},\frac{5}{6})$. It should be emphasized that there is no 2nd Derivative Test available with the Method of Lagrange Multipliers (though it is possible to formulate a somewhat complicated alternative). Consequently, you will have to use other methods (such as evaluating the function and comparing values). In this case we simply observe that there is a closest point and we have produced only one possible candidate, so this must be that closest point.

Example 5: Prove that among all rectangles with a fixed perimeter the one that encloses the maximum area is a square.
Solution: If we let x and y be the lengths of adjacent edges of the square and let P be the (constant) perimeter, then the area is $A(x,y)=xy$ and the perimeter is $g(x,y)=2x+2y=P$. The Method of Lagrange Multipliers gives: $\begin{Bmatrix} y=2\lambda \\ x=2\lambda \end{Bmatrix} \Rightarrow \boxed{x=y}$. So the rectangle must be a square.

Example 6: Suppose we wish to construct an open-top rectangular box that must contain a volume of 96 cubic units. If the bottom of the box costs twice as much per square unit as the sides, find the dimensions of the least expensive box that can be constructed.
Solution: If we denote the lengths of the bottom edges by x and y and let the height be z, and if we let the price per square unit of the sides be p (and therefore the price per square unit of the bottom will be $2p$), we can express the total cost as $C(x,y,z)=3p(xy)+p(2xz+2yz)=p(3xy+2xz+2yz)$. The volume is constrained as $V(x,y,z)=xyz=96$. The Lagrange Multiplier optimality condition will be given by $\overrightarrow{\nabla C}=\lambda\overrightarrow{\nabla V}$ which yields the three equations: $\begin{Bmatrix} p(3y+2z)=yz \\ p(3x+2z)=xz \\ p(2x+2y)=xy \end{Bmatrix}$.



If we divide respective sides of the first and second equations we get

$$\frac{3y+2z}{3x+2z}=\frac{y}{x} \Rightarrow 3xy+2xz=3xy+2yz \Rightarrow 2xz=2yz \Rightarrow \boxed{x=y}.$$

If we divide respective sides of the first and third equations we get

$$\frac{3y+2z}{2x+2y}=\frac{z}{x} \Rightarrow 3xy+2xz=2xz+2yz \Rightarrow 3xy=2yz \Rightarrow \boxed{z=\tfrac{3}{2}x}.$$

It's worth noting that these determine the *relative* dimensions of the box regardless of the constrained volume. We substitute these into the given constraint to get $\frac{3}{2}x^3=96 \Rightarrow x^3=64 \Rightarrow \boxed{x=y=4,\ z=6}$.
So the box will have a square 4 by 4 foot base and a height of 6 units.

Notes by Robert Winters and Renée Chipman

Revised January 7, 2017