CLASS NOTES for CHAPTER 4, Nonlinear Programming

1 Space Curves and Tangent Lines

Recall that space curves are defined by a set of parametric equations,

$$\mathbf{r}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}$$

In Calc III, we might have written this a little differently,

$$\vec{r}(t) =< x(t), y(t), z(t) >$$

but here we want to use n dimensions rather than two or three dimensions. The derivative and antiderivative of $\mathbf{r}$ with respect to t is done componentwise,

$$\mathbf{r}(t) = \begin{bmatrix} x_1'(t) \\ x_2'(t) \\ \vdots \\ x_n'(t) \end{bmatrix}, \qquad \mathbf{R}(t) = \begin{bmatrix} \int x_1(t)\,dt \\ \int x_2(t)\,dt \\ \vdots \\ \int x_n(t)\,dt \end{bmatrix}$$

And the local linear approximation to $\mathbf{r}(t)$ is also done componentwise. The tangent line (in n dimensions) can be written easily- the derivative at $t = a$ is the direction of the curve, so the tangent line is given by:

$$\mathbf{L}(t) = \begin{bmatrix} x_1(a) \\ x_2(a) \\ \vdots \\ x_n(a) \end{bmatrix} + t \begin{bmatrix} x_1'(a) \\ x_2'(a) \\ \vdots \\ x_n'(a) \end{bmatrix}$$

In Class Exercise: Use Maple to plot the curve $\mathbf{r}(t) = [\cos(t), \sin(t), t]^T$ and its tangent line at $t = \frac{\pi}{2}$.

2 Gradients and Tangent Planes

Let $f : \mathbb{R}^n \to \mathbb{R}$. In this case, we can write:

$$y = f(x_1, x_2, x_3, \ldots, x_n)$$

Note that any function that we wish to optimize must be of this form- It would not make sense to find the maximum of a function like a space curve; n dimensional coordinates are not well-ordered like the real line- so the following statement would be meaningless: $(3, 5) > (1, 2)$. This type of function is sometimes called a surface even though we can't graph it if the domain has a higher dimension than 2.

The gradient of f, ∇f, is defined by its partial derivatives. We will denote $\frac{\partial f}{\partial x_i}$ as f_{x_i}, and so:

$$\nabla f(x_1, x_2, \ldots, x_n) = \begin{bmatrix} f_{x_1}(x_1, x_2, \ldots, x_n) \\ f_{x_2}(x_1, x_2, \ldots, x_n) \\ \vdots \\ f_{x_n}(x_1, x_2, \ldots, x_n) \end{bmatrix}$$

NOTE: When writing the gradient, we will not distinguish between it being a row vector or a column vector. For some ideas, it may be more clear notationally to write it as a column, but for computations, it is a row vector. Normally, it will be clear from the context whether it is a row or column.

The rate of change of f at $\mathbf{x} = \mathbf{a}$ in the direction of the unit vector $\mathbf{u}$ is defined as the directional derivative of f:

$$D_{\mathbf{u}} f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \mathbf{u}$$

If we recall the useful relationship between any two vectors $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^n$,

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|_2 \|\mathbf{v}\|_2 \cos(\theta)$$

then we see that the direction for which the rate of change of f is a maximum is in the direction of $\nabla f(\mathbf{a})$.

In Class Example: Find the maximum rate of change of f at the given point and the direction in which it occurs: $f(x, y) = x + y/z$ at $(4, 3)$.

The linearization of f at $\mathbf{x} = \mathbf{a}$ is the tangent plane,

$$L(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a})$$

which is also the first order Taylor series approximation to f at $\mathbf{a}$.

2.1 Surfaces and Contours

In maximizing or minimizing a function, we'll consider equations like: $F(\mathbf{x}) = k$, where k is a real number. From our earlier discussion on the Implicit Function Theorem, we know that this equation could implicitly define one of the variables in terms of the others. In Calculus III, for example, we saw

- $F(x, y) = k$ defines a level curve in the plane.
- $F(x, y, z) = k$ defines a level surface in three dimensions.

Now let $\mathbf{r}(t)$ be a curve on the surface defined by $F(\mathbf{x}) = k$. Then we could write F as dependent on t, and differentiate with respect to t:

$$\nabla F(\mathbf{r}(t)) \cdot \mathbf{r}'(t) = 0$$

which implies that the gradient is orthogonal to any curve on the surface. In particular, we can compute the tangent line or plane at the point (a, b) or (a, b, c):

- $F_x(a, b) \cdot (x - a) + F_y(a, b) \cdot (y - b) = 0$
- $F_x(a, b, c) \cdot (x - a) + F_y(a, b, c) \cdot (y - b) + F_z(a, b, c) \cdot (z - c) = 0$

If $\mathbf{x} \in \mathbb{R}^n$, this can be written compactly as $\nabla F(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) = 0$

In Class Example: Compute the tangent line and show that the gradient is orthogonal to it, if $x^2 + y^2 = 1$ at $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$. In this case, the gradient of $F(x, y) = x^2 + y^2$ is $[2x, 2y]$, and at the given point, the gradient is $[\sqrt{2}, \sqrt{2}]$. Notice that this vector points directly opposite from the center of the circle. On the other hand, the equation of the tangent line is:

$$\sqrt{2}\left(x - \frac{1}{\sqrt{2}}\right) + \sqrt{2}\left(y - \frac{1}{\sqrt{2}}\right) = 0$$

or in the more familiar form, $y = -x + \sqrt{2}$. The "slope" of the gradient is 1, while the slope of the tangent line is -1. In two dimensions such as this, a line in the direction of the gradient would have slope $\dfrac{F_y}{F_x}$, while the slope of the tangent line would be $-\dfrac{F_x}{F_y}$.

The important point in this section: Given a level surface of the form $F(\mathbf{x}) = k$, the gradient ∇F at $\mathbf{x} = \mathbf{a}$ is orthogonal. Furthermore, the gradient points in the direction where F increases the fastest (for $y = F(\mathbf{x})$.

3 Derivatives we don't see in Calculus III

3.1 The Hessian Matrix

Let $f : \mathbb{R}^n \to \mathbb{R}$, so that we have something in the form of $y = f(x_1, x_2, \ldots, x_n)$. We have already seen that the derivative of f takes the form of the gradient. We now examine the second derivative of f, and we'll see that the second derivative forms what is called the Hessian matrix.

The first derivative is in the form of the gradient, where we have a vector of functions. Let $f_{x_i}(x_1, x_2, \ldots, x_n) = g_i(x_1, x_2, \ldots, x_n)$. Now each function g is also a function from $\mathbb{R}^n$ to $\mathbb{R}$, and so the derivative of g is also a gradient. We could therefore write the second derivative of f in matrix form to obtain what is called the Hessian of f:

$$Hf = \begin{bmatrix} \nabla g_1 \\ \nabla g_2 \\ \vdots \\ \nabla g_n \end{bmatrix} = \begin{bmatrix} f_{x_1x_1} & f_{x_1x_2} & \cdots & f_{x_1x_n} \\ f_{x_2x_1} & f_{x_2x_2} & \cdots & f_{x_2x_n} \\ \vdots & \vdots & & \vdots \\ f_{x_nx_1} & f_{x_nx_2} & \cdots & f_{x_nx_n} \end{bmatrix}$$

In Class Example: Let $f(x, y) = x^3 + 2xy - y^2$. Compute the gradient and Hessian of f.

$$\nabla f = \begin{bmatrix} 3x^2 + 2y \\ 2x - 2y \end{bmatrix}, \quad Hf = \begin{bmatrix} 6x + 2 & 2 \\ 2 & -2 \end{bmatrix}$$

You might note that, if the second partials of f are continuous, then Hf will always be a symmetric matrix (this is a consequence of *Clairaut's Theorem*).

In Class Example: Let $f(x, y) = 2 + x^2 + x + xy + y + y^2$. Show that f can be written as:

$$f(\mathbf{a}) + \nabla f(\mathbf{a})(\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T Hf(\mathbf{a})(\mathbf{x} - \mathbf{a})$$

where $\mathbf{a} = (0, 0)$.

First, we compute ∇f and Hf:

$$\nabla f = \begin{bmatrix} 2x + 1 + y \\ x + 1 + 2y \end{bmatrix}, \quad Hf = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Substituting $(x, y) = (0, 0)$, the gradient is $[1, 1]$, and now compute:

$$2 + [1, 1] \cdot \begin{bmatrix} x - 0 \\ y - 0 \end{bmatrix} + \frac{1}{2}[x - 0, y - 0] \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x - 0 \\ y - 0 \end{bmatrix}$$

$$2 + x + y + \frac{1}{2}[x, y] \begin{bmatrix} 2x + y \\ x + 2y \end{bmatrix} = 2 + x + y + x^2 + xy + y^2$$

The Second Order Taylor Approximation

If $f : \mathbb{R}^n \to \mathbb{R}$, then the Taylor approximation to f at $\mathbf{a}$ to second order is:

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a})(\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T Hf(\mathbf{a})(\mathbf{x} - \mathbf{a})$$

3.2 Connecting the Hessian to Calculus III

In Calc III, when we were finding local extrema, we had the "Second Derivatives Test". To generalize this result requires a little linear algebra, but the end result is that if $\mathbf{x} = \mathbf{a}$ is a critical point of f, and $Hf(\mathbf{a})$ has all positive eigenvalues, then f has a local minimum at $\mathbf{a}$. If $Hf(\mathbf{a})$ has all negative eigenvalues, then f has a local maximum at $\mathbf{a}$. If the eigenvalues are mixed in sign, there is a saddle point.

3.3 The Jacobian Matrix

Let $f : \mathbb{R}^n \to \mathbb{R}^m$. The we could write f using m coordinate functions, each depending on n variables,

$$f(x_1, \ldots, x_n) = \begin{bmatrix} f_1(x_1, \ldots, x_n) \\ f_2(x_1, \ldots, x_n) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{bmatrix}$$

Definition: The derivative of f takes the form of a matrix called the Jacobian of f, constructed by taking the partial derivatives:

$$Jf = \begin{bmatrix} (f_1)_{x_1} & (f_1)_{x_2} & \cdots & (f_1)_{x_n} \\ (f_2)_{x_1} & (f_2)_{x_2} & \cdots & (f_2)_{x_n} \\ \vdots & \vdots & & \vdots \\ (f_m)_{x_1} & (f_m)_{x_2} & \cdots & (f_m)_{x_n} \end{bmatrix}$$

In Class Example: Compute the Jacobian of:

$$f(x,y,z) = \begin{bmatrix} x^2+yz+z \\ y^2+xy+3 \\ \sin(xyz) \end{bmatrix}$$

$$Jf = \begin{bmatrix} 2x & z & y+1 \\ y & 2y+x & 0 \\ yz\cos(xyz) & xz\cos(xyz) & xy\cos(xyz) \end{bmatrix}$$

If $f : \mathbb{R}^n \to \mathbb{R}^m$, then the linearization of f at $\mathbf{x} = \mathbf{a}$ is given by:

$$L(\mathbf{x}) = f(\mathbf{a}) + Jf(\mathbf{a})(\mathbf{x} - \mathbf{a})$$

4 The Extreme Value Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous on a closed and bounded domain D. Then f attains a global maximum and global minimum on D. Furthermore, this maximum and minimum value is obtained either at the critical points of f (where the gradient is zero or not defined), or on the boundary of D.

In Class Example: Find the maximum value of $f(x,y) = x^2 - 2xy + 2y$ where the domain is the rectangle $D = \{(x,y)|x \in [0,3], y \in [0,2]\}$

The gradient is $[2x-2y, -2x+2]$. Setting these to zero, we see the only solution is $(1,1)$. The value of the function here is $f(1,1) = 1$.

Since the boundary is made up of line segments, it is easy to analyze f along them:

- For $y = 0, 0 \le x \le 3$, $f(x,0) = x^2$, and the max of 9 occurs at $x = 3$
- For $x = 0, 0 \le y \le 2$, $f(0,y) = 2y$, and the max of 4 occurs at $y = 2$.
- For $x = 3, 0 \le y \le 2$, $f(3,y) = 9 - 4y$, and the max of 5 occurs at $y = 0$.
- For $y = 2, 0 \le x \le 3$, $f(x,2) = x^2 - 4x + 4 = (x-2)^2$, and has a max of 4 at $x = 0$.

Now we see that the global max is 9 and occurs at the boundary point $(3,0)$.

5 Newton's Method Revisited

We saw that for a problem of the form $f(x) = 0$, we could solve for x iteratively by using an initial guess, x_0, and iterate:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

We also saw that there could be problems with Newton's Method:

- If x^* is the desired point, then $f'(x^*) \neq 0$.
- We're not sure which root we're converging to.
- We can only guarantee convergence if x_0 is sufficiently close to x^*.

We got the formula for Newton's Method by linearizing the function f at x_k, then we set $y = 0$ and solved for x_{k+1}:

$$y - f(x_k) = f'(x_k)(x_{k+1} - x_k) \Rightarrow -f(x_k) = f'(x_k)(x_{k+1} - x_k)$$

5.1 Scalar Valued Functions

If $f : \mathbb{R}^n \to \mathbb{R}$, then Newton's Method is not well defined. For example, suppose $f(x, y) = x^2 + y^2 - 1$. Then the zeros to this function are all points on the unit circle- Generally, the solution set is a level surface (or level curve).

5.2 Vector Valued Functions

Let $g : \mathbb{R}^n \to \mathbb{R}^n$. The derivative of g is then its Jacobian matrix $(n \times n)$, and the linearization of g about $\mathbf{x}_k$ is given by:

$$\mathbf{y} - g(\mathbf{x}_k) = Jg(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k)$$

Set $\mathbf{y} = 0$ and solve for $\mathbf{x}_{k+1}$, and we get:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (Jg(\mathbf{x}_k))^{-1} g(\mathbf{x}_k)$$

Let's do a simple example just to see what is at work here. Let $g(x, y) = [x^2 - y^2, 2x + y - 1]$. Then the Jacobian matrix at a point (x, y) is given by:

$$Jg(x, y) = \begin{bmatrix} 2x & -2y \\ 2 & 1 \end{bmatrix}$$

And since this is a 2×2 matrix, its inverse can be computed directly:

$$(Jg)^{-1}(x, y) = \frac{1}{2x + 4y} \begin{bmatrix} 1 & 2y \\ -2 & 2x \end{bmatrix}$$

Let the initial guess be $(1, 1)$. The next iterate is then:

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \frac{1}{6} \begin{bmatrix} 1 & 2 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/3 \end{bmatrix}$$

You can double check that this is indeed a solution to $g(x) = 0$, but this was because we had a lucky first guess. Other initial conditions may take longer for the algorithm to converge.

5.3 Newton's Method in Optimization

In one dimension, we look for points where $f'(x) = 0$. Numerically, we can use Newton's Method to find the x by letting $g(x) = f'(x)$:

$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

If $f : \mathbb{R}^n \to \mathbb{R}$, then we can think of $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$. To find the zeros of the gradient, we can use Newton's Method in a similar way. Let $g(\mathbf{x}) = \nabla f$. Then, from our previous computation,

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (Jg(\mathbf{x}_k))^{-1} g(\mathbf{x}_k) = \mathbf{x}_k - (Hf(\mathbf{x}_k))^{-1} \nabla f(\mathbf{x}_k)$$

Note that this last product is an $n \times n$ inverse matrix of the Hessian times the gradient as a column vector.

If the domain of f is unbounded, we may or may not have an optimal value- recall that we are simply finding where the gradient is zero, which may correspond to a local or global maximum, local or global minimum, or a saddle point.

EXAMPLE: Find the maximum of $f(x, y) = 12 + 10y - 2x^2 - 8xy - y^4$. Since this is a function of two variables, we can look at the contour plot to predict what will happen, and this is given in Figure 1. After looking at the function, it is clear that there is a global maximum, since the value of the function will tend to $-\infty$ as (x, y) become unbounded.

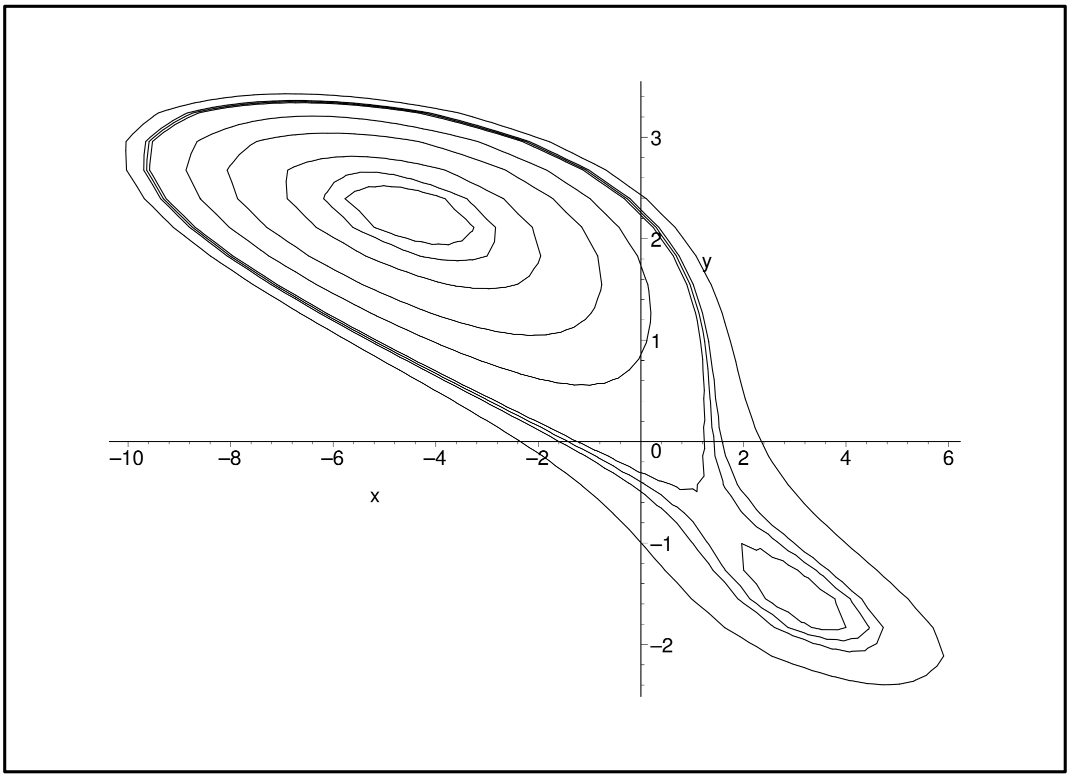

Figure 1: Contours of $f(x,y) = 12 + 10y - 2x^2 - 8xy - y^4$. We see that the contours are encircling two points and there might be a saddle in between.

We compute the gradient and Hessian of f:

$$\nabla f = \begin{bmatrix} -4x - 8y & 10 - 8x - 4y^3 \end{bmatrix}$$

$$Hf = \begin{bmatrix} -4 & -8 \\ -8 & -12y^2 \end{bmatrix}$$

We can look at the gradient directly- Solutions to where the gradient is zero will lie along the line $y = -\frac{1}{2}x$, from which we would have to solve: $10 - 8x + \frac{1}{2}x^3 = 0$, and we get three solutions for x, which are approximately:

$$x \approx -4.159, \quad x \approx 3.084, \quad x \approx 1.434$$

Using the Matlab script below, we can numerically solve for these using Newton's Method:

```
x=[-8;1];
for j=1:30
    gradf=[-4*x(1)-8*x(2); 10-8*x(1)-4*x(2)^3];
    Hf=[-4 8; -8 -12*x(2)^2];
    iHf=inv(Hf);
    x=x-iHf*gradf;
end
```

Running this code will give a final value of approximately $(-4.519, 2.2597)$. We can put additional things into the code so that we can track the values of $\mathbf{x}$, ∇f and Hf. We could also use a stopping criteria rather than running the algorithm 30 times (30 was chosen arbitrarily). At this point, the functional value was approximately 49.37.

Starting Newton's Method at $(3, -1)$ gave a different ending value of approximately $(3.084, -1.542)$, and the value of f was 9.484.

Starting Newton's Method at $(1, -1)$ gave a different ending value of approximately $(3.084, -1.542)$. Starting Newton's Method at $(1.5, -0.5)$ gave us $(-4.519, 2.259)$.

Starting very close to the third point we found by other methods, Newton's method did not want to converge to it- it is in fact a saddle point.

We see some typical behavior in this example:

- If Newton's method starts close to the optimum, it will reach that point fairly quickly.
- The solution given by Newton's method may be only a local maximum (or minimum).
- Different initial conditions may give different answers.

In practice, it is common to run an optimization algorithm with many randomly chosen initial values to probe the surface.

6 Lagrange Multipliers

In the last section, we considered optimizing a nonlinear function over an unbounded domain. In this section, we place constraints on the domain- in particular, we assume the constraints are of a particular form: $g(\mathbf{x}) = k$.

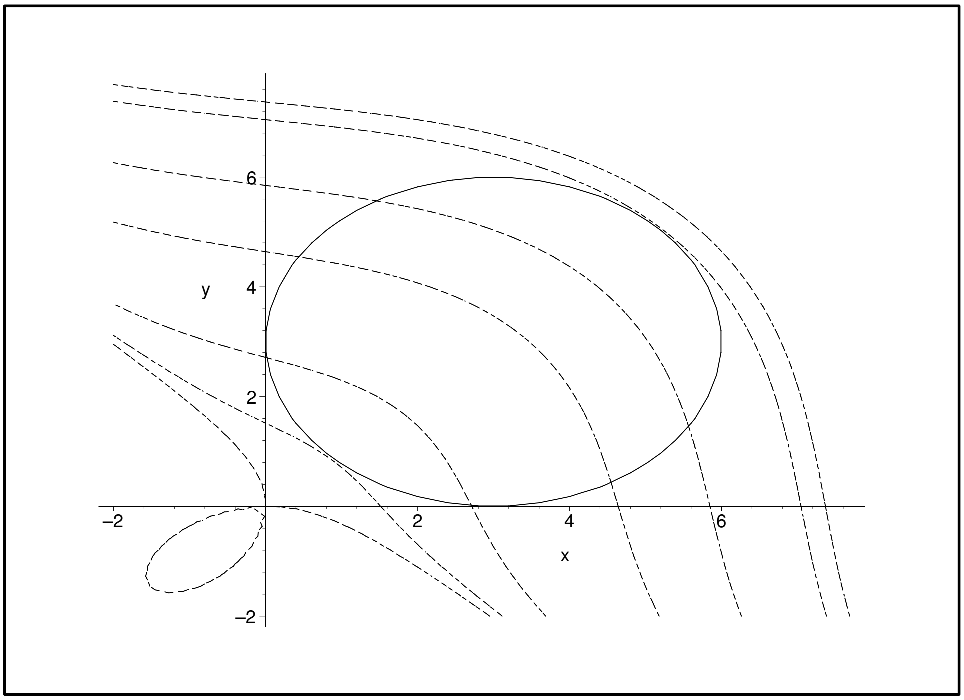

Figure 2: Some contours associated with $f(x, y) = x^3 + y^3 + 3xy$ (these are $0, 3.5, 20, 100, 200, 350, 400$) which are shown as dotted curves (the higher the value, the higher the curve for this function). The constraint $(x-3)^2 + (y-3)^2 = 9$ is shown as the solid curve.

Before looking at the case in $\mathbb{R}^n$, let's consider the problem where we are optimizing a function $f(x, y)$ under constraint that $g(x, y) = k$. In this case, plots of the form $f(x, y) = c$ are level curves of f, and the plot of $g(x, y) = k$ is a curve.

For example, consider $f(x, y) = x^3 + y^3 - 3xy$ subject to the restriction that the points lie on a circle of radius 3 centered at $(3, 3)$. In Figure 2, we show a plot of some contours of f together with the circle. We'll recall that the gradient of f, in this case

$$\nabla f(x, y) = [3x^2 + 3y, \quad 3y^2 + 3x]$$

is orthogonal to the level curve at a given point on the level curve. Similarly,

$$\nabla g(x, y) = [2(x-3), \quad 2(y-3)]$$

is orthogonal to the circle. Now, consider f restricted to the curve where we'll assume that $g(x, y) = k$ has been parametrized by $\vec{r}(t) = [x(t), y(t)]$. In the example, this would mean

$$x(t) = 3\cos(t) + 3, \quad y(t) = 3\sin(t) + 3$$

Note that:

$$\vec{r}'(t) = [-3\sin(t), \quad 3\cos(t)]$$

and this is the *tangent* vector to the curve. Now back to the general case, we can write:

$$h(t) = f(\vec{r}(t)) = f(x(t), y(t))$$

and if h has an extreme point at t^*, then $h'(t^*) = 0$:

$$0 = h'(t^*) = \nabla f(x(t^*), y(t^*)) \cdot \vec{r}'(t)$$

Thus, at an extreme point, the gradient of f is orthogonal to the tangent vector on the constraint, which is in turn orthogonal to ∇g.

Method of Lagrange Multipliers

If $y = f(\mathbf{x})$ has an extreme point on the surface defined by $g(\mathbf{x}) = k$, then candidates for where the extrema occur are where ∇f and ∇g are parallel. That is,

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

The constant(s) λ are called the Lagrange multipliers.

Let us solve our previous example: Find the optimal values of $f(x, y) = x^3 + y^3 + 3xy$, such that $(x - 3)^2 + (y - 3)^2 = 9$.

$$\nabla f(x, y) = [3x^2 + 3y, \quad 3y^2 + 3x]$$

$$\nabla g(x, y) = [2(x - 3), \quad 2(y - 3)]$$

so that we solve the following system of equations:

$$\begin{aligned} 3x^2 + 3y &= \lambda 2(x - 3) \\ 3y^2 + 3x &= \lambda 2(y - 3) \\ (x - 3)^2 + (y - 3)^2 &= 9 \end{aligned}$$

These types of systems are generally difficult to solve by hand, so we'll use a computer algebra system or Matlab to help us. In this example, we have:

$$x = 3 \pm \frac{3}{\sqrt{2}}, \quad y = 3 \pm \frac{3}{\sqrt{2}}, \quad \lambda = 21 \pm \frac{33}{\sqrt{2}}$$

Note where these points occur in our Figure. The extrema are found by substituting these 2 values back into f. Doing this, we find that the maximum occurs at:

$$x = 3 + \frac{3}{\sqrt{2}}, \quad y = 3 + \frac{3}{\sqrt{2}}, \quad \lambda = 21 + \frac{33}{\sqrt{2}}$$

and the maximum value is approximately 347.32, which we expected from the contour plot. The minimum is approximately 3.67, which is also expected from the contour plot.

Of course, in this example, we could actually convert f to a function of one variable by explicitly performing the parameterization. In this case, we have:

$$f(t) = 27(\cos(t) + 1)^3 + 27(\sin(t) + 1)^3 + 27(\cos(t) + 1)(\sin(t) + 1)$$

and we can use a numerical solver (like Newton's method) to solve for where $f'(t) = 0$. In this case, we get that

$$t \approx .7854, \quad 3.926$$

where we get a maximum value for f for the first value of t, and a minimum for f at the second value of t.

This method can be employed for several constraints, $g_1(\mathbf{x}) = k_1, g_2(\mathbf{x}) = k_2, \ldots, g_m(\mathbf{x}) = k_m$. In this case, the gradient of f is parallel to a *linear combination* of the gradients of g_i, and we have the system of equations:

$$\begin{array}{cl} \nabla f & = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2 + \ldots + \lambda_m \nabla g_m \\ g_1 & = k_1 \\ \vdots & \vdots \\ g_m & = k_m \end{array}$$

Note that this is a system of $n+m$ equations in $n+m$ unknowns (if $\mathbf{x} \in \mathbb{R}^n$). In general, we could translate this system into:

$$G(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m) = 0$$

and use Newton's method to solve it numerically.

Example: Maximize $\frac{1}{3}x_1^3 + \frac{1}{3}x_2^3 + \frac{1}{2}x_3^3$ subject to the conditions that $x_1 + x_2 + x_3 = 0$ and $x_1^2 + x_2^2 + x_3^2 = 3$.

The system of equations we need to solve is given by:

$$\begin{aligned}
x_1^2 &= \lambda_1 + 2\lambda_2 x_1 \\
x_2^2 &= \lambda_1 + 2\lambda_2 x_2 \\
x_3^2 &= \lambda_1 + 2\lambda_2 x_3 \\
x_1 + x_2 + x_3 &= 0 \\
x_1^2 + x_2^2 + x_3^2 &= 3
\end{aligned}$$

In this example, we have:

$$G = \begin{bmatrix} x_1^2 - 2\lambda_2 x_1 - \lambda_1 \\ x_2^2 - 2\lambda_2 x_2 - \lambda_1 \\ x_3^2 - 2\lambda_2 x_3 - \lambda_1 \\ x_1 + x_2 + x_3 \\ x_1^2 + x_2^2 + x_3^2 - 3 \end{bmatrix}$$

and the Jacobian of G:

$$JG = \begin{bmatrix} 2x_1 - 2\lambda_2 & 0 & 0 & -1 & -2x_1 \\ 0 & 2x_2 - \lambda_2 & 0 & -1 & -2x_2 \\ 0 & 0 & 2x_3 - 2\lambda_2 & -1 & -2x_3 \\ 1 & 1 & 1 & 0 & 0 \\ 2x_1 & 2x_2 & 2x_3 & 0 & 0 \end{bmatrix}$$

Performing Newton's method gives a maximum where $\lambda_1 = 1, \lambda_2 = \frac{\sqrt{2}}{4}$, and triples of the form $(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, \sqrt{2})$, so any two variables can be $-\frac{1}{\sqrt{2}}$, and the third is $\sqrt{2}$.