b. Determine the nature of each critical point.

3.6.8 a. Write the Taylor polynomial of $f\begin{pmatrix} x \\ y \end{pmatrix} = \sqrt{1 - x + y^2}$ to degree 3 at the origin.

b. Show that $g\begin{pmatrix} x \\ y \end{pmatrix} = \sqrt{1 - x + y^2} + x/2$ has a critical point at the origin. What kind of critical point?

3.7 Constrained critical points and Lagrange multipliers

The shortest path between two points is a straight line. But what is the shortest path if you are restricted to paths that lie on a sphere (for example, because you are flying from New York to Paris)? This example is intuitively clear but actually quite difficult to address.

In this section we will look at problems in the same spirit, but easier. We will be interested in extrema of a function f when f is restricted to some manifold $X \subset \mathbb{R}^n$. In the case of the set $X \subset \mathbb{R}^8$ describing the position of four linked rods in the plane (example 3.1.8), we might imagine that each of the four joints connecting the rods at the vertices $\mathbf{x}_i$ is connected to the origin by a rubber band, and that the vertex $\mathbf{x}_i$ has a "potential" $|\vec{\mathbf{x}}_i|^2$. Then what is the equilibrium position, where the link realizes the minimum of the potential energy? Of course, all four vertices try to be at the origin, but they can't. Where will they go?

In this section we provide tools to answer this sort of question.

Another example occurs in section 2.9: the norm

$$\sup_{|\vec{\mathbf{x}}|=1} |A\vec{\mathbf{x}}|$$

of a matrix A answers the question, what is $\sup |A\vec{\mathbf{x}}|$ when we require that $\vec{\mathbf{x}}$ have length 1?

Finding constrained critical points using derivatives

A characterization of extrema in terms of derivatives should say that in some sense the derivative vanishes at an extremum. But when we take a function defined on $\mathbb{R}^n$ and consider its *restriction* to a manifold of $\mathbb{R}^n$, we cannot assert that an extremum of the restricted function is a point at which the derivative of the function vanishes. The derivative of the function may vanish at points not in the manifold (the shortest "unrestricted" path from New York to Paris would require tunneling under the Atlantic Ocean). And only very seldom will a constrained maximum be an unconstrained maximum (the tallest child in kindergarten is unlikely to be the tallest child in the entire elementary school). So only very seldom will the derivative of the function vanish at a critical point of the restricted function.

What we can say is that at an extremum of the function restricted to a manifold, the derivative of the function is 0 *when evaluated on any vector that is tangent to the manifold*. In other words, the derivative vanishes on the tangent space to the manifold.

Theorem 3.7.1. *If $X \subset \mathbb{R}^n$ is a manifold, $U \subset \mathbb{R}^n$ is open, $f : U \to \mathbb{R}$ is a C^1 function, and $\mathbf{c} \in (X \cap U)$ is a local extremum of f restricted to X, then*

$$T_{\mathbf{c}}X \subset \ker\,[\mathbf{D}f(\mathbf{c})]. \qquad 3.7.1$$

Formula 3.7.1 says that if $\vec{\mathbf{x}}$ is tangent to X at $\mathbf{c}$, then

$$[\mathbf{D}f(\mathbf{c})]\vec{\mathbf{x}} = 0.$$

We prove theorem 3.7.1 after giving some examples.

Definition 3.7.2 (Constrained critical point). A point $\mathbf{c}$ such that $T_{\mathbf{c}}X \subset \ker\,[\mathbf{D}f(\mathbf{c})]$ is called a *critical point of f constrained to X*.

Recall (theorem 3.2.4) that when a manifold X is described by $\mathbf{F}(\mathbf{z}) = \mathbf{0}$, where $\mathbf{F} : U \subset \mathbb{R}^n \to \mathbb{R}^{n-k}$, then

$$T_{\mathbf{z}}X = \ker\,[\mathbf{DF}(\mathbf{z})]. \qquad 3.7.2$$

Thus theorem 3.7.1 says that at a critical point $\mathbf{c}$ of f *restricted* to X,

$$\ker\,[\mathbf{DF}(\mathbf{c})] \subset \ker\,[\mathbf{D}f(\mathbf{c})]. \qquad 3.7.3$$

A space with more constraints is smaller than a space with fewer constraints: more people belong to the set of musicians than belong to the set of red-headed, left-handed cello players with last name beginning with W.

Note that both derivatives in formula 3.7.3 have the same width, as they must for that equation to make sense; $[\mathbf{DF}(\mathbf{c})]$ is a $(n-k) \times n$ matrix, and $[\mathbf{D}f(\mathbf{c})]$ is a $1 \times n$ matrix, so both can be evaluated on a vector in $\mathbb{R}^n$. It also makes sense that the kernel of the taller matrix should be a subset of the kernel of the shorter matrix. Saying that $\vec{\mathbf{v}} \in \ker\,[\mathbf{DF}(\mathbf{c})]$ means that

$$\begin{bmatrix} D_1F_1(\mathbf{c}) & \cdots & D_nF_1(\mathbf{c}) \\ \vdots & \vdots & \vdots \\ D_1F_{n-k}(\mathbf{c}) & \cdots & D_nF_{n-k}(\mathbf{c}) \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}; \qquad 3.7.4$$

$n-k$ equations need to be satisfied. Saying that $\vec{\mathbf{v}} \in \ker\,[\mathbf{D}f(\mathbf{c})]$ means that only one equation needs to be satisfied:

$$[D_1f(\mathbf{c}) \cdots D_nf(\mathbf{c})] \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = 0. \qquad 3.7.5$$

Theorem 3.7.1 says that any vector that satisfies equation 3.7.4 also satisfies equation 3.7.5.

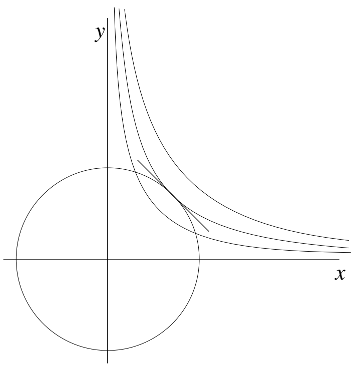

FIGURE 3.7.1.

The unit circle and several level curves of the function xy. The level curve $xy = 1/2$, which realizes the maximum value of xy restricted to the circle, is tangent to the circle at the point $\binom{1/\sqrt{2}}{1/\sqrt{2}}$, where the maximum value is realized.

Example 3.7.3 (Constrained critical point: a simple example). Suppose we wish to maximize the function $f\begin{pmatrix} x \\ y \end{pmatrix} = xy$ on the first quadrant of the circle $x^2 + y^2 = 1$, which we will denote by X. As shown in figure 3.7.1, some level sets of that function do not intersect the circle, and some intersect it in two points, but one, $xy = 1/2$, intersects it at the point $\mathbf{c} = \binom{1/\sqrt{2}}{1/\sqrt{2}}$. To show that $\mathbf{c}$ is the critical point of f constrained to X, we need to show that $T_{\mathbf{c}}X \subset \ker\,[\mathbf{D}f(\mathbf{c})]$.

Since $F\begin{pmatrix} x \\ y \end{pmatrix} = x^2 + y^2 - 1$ is the function defining the circle, we have

$$T_{\mathbf{c}}X = \ker\,[\mathbf{D}F(\mathbf{c})] = \ker\,[2c_1,\, 2c_2] = \ker\,\left[\frac{2}{\sqrt{2}},\, \frac{2}{\sqrt{2}}\right] \qquad 3.7.6$$

$$\ker\,[\mathbf{D}f(\mathbf{c})] = \ker\,[c_2,\, c_1] = \ker\,\left[\frac{1}{\sqrt{2}},\, \frac{1}{\sqrt{2}}\right]. \qquad 3.7.7$$