3.6.5 a. Find the critical points of the function $f\begin{pmatrix} x \\ y \\ z \end{pmatrix} = xy + yz - xz + xyz$.

b. Determine the nature of each critical point.

3.6.6 Find all the critical points of the following functions:

a. $\sin x \cos y$ b. $xy + \dfrac{8}{x} + \dfrac{1}{y}$

*c. $\sin x + \sin y + \sin(x+y)$

3.6.7 a. Find the critical points of the function $f\begin{pmatrix} x \\ y \end{pmatrix} = (x^2 + y^2)e^{x^2 - y^2}$.

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 is it?

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 quite difficult to address.

Here we will look at easier problems in the same spirit. We will be interested in extrema of a function f when f is either defined on a manifold $X \subset \mathbb{R}^n$ or restricted to it. 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 case the function "sum of the $|\vec{\mathbf{x}}_i|^2$" is defined on the ambient space, but there are important functions that are not, such as the curvature of a surface.

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

Finding shortest paths goes under the name of the *calculus of variations*. The set of paths from New York to Paris is an infinite-dimensional manifold. We will be restricting ourselves to finite-dimensional problems. But the tools we develop apply quite well to the infinite-dimensional setting.

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

Recall that in Section 3.2 we defined the derivative of a function defined on a manifold. Thus we can make the obvious generalization of Definition 3.6.2.

The derivative $[\mathbf{D}f(\mathbf{x})]$ of f is only defined on the tangent space $T_{\mathbf{x}}X$, so saying that it is 0 is saying that it vanishes *on tangent vectors* to X.

Definition 3.7.1 (Critical point of function defined on manifold). Let $X \subset \mathbb{R}^n$ be a manifold, and $f : X \to \mathbb{R}$ be a C^1 function. Then a critical point of f is a point $\mathbf{x} \in X$ where $[\mathbf{D}f(\mathbf{x})] = 0$.

An important special case is when f is defined not just on X but on an open neighborhood $U \subset \mathbb{R}^n$ of X; in that case we are looking for critical points of the restriction $f|_X$ of f to X.

Analyzing critical points of functions $f : X \to \mathbb{R}$ isn't quite the focus of this section; we are really concerned with functions g defined on a neighborhood of a manifold X and studying critical points of the restriction of g to X.

Traditionally a critical point of g restricted to X is called a *constrained critical point*.

Theorem 3.7.2. *Let $X \subset \mathbb{R}^n$ be a manifold, $f : X \to \mathbb{R}$ a C^1 function, and $\mathbf{c} \in X$ a local extremum of f. Then $\mathbf{c}$ is a critical point of f.*

Proof. Let $\gamma : V \to X$ be a parametrization of a neighborhood of $\mathbf{c} \in X$, with $\gamma(\mathbf{x}_0) = \mathbf{c}$. Then $\mathbf{c}$ is an extremum of f precisely if $\mathbf{x}_0$ is an extremum of $f \circ \gamma$. By Theorem 3.6.3, $[\mathbf{D}f \circ \gamma(\mathbf{x}_0)] = [0]$, so, by Proposition 3.2.11,

$$[\mathbf{D}(f \circ \gamma)(\mathbf{x}_0)] = [\mathbf{D}f(\gamma(\mathbf{x}_0))][\mathbf{D}\gamma(\mathbf{x}_0)] = [\mathbf{D}f(\mathbf{c})][\mathbf{D}\gamma(\mathbf{x}_0)] = [0].$$

By Proposition 3.2.7, the image of $[\mathbf{D}\gamma(\mathbf{x}_0)]$ is $T_{\mathbf{c}}X$, so $[\mathbf{D}f(\mathbf{c})]$ vanishes on $T_{\mathbf{c}}X$. The proof is illustrated by Figure 3.7.1.

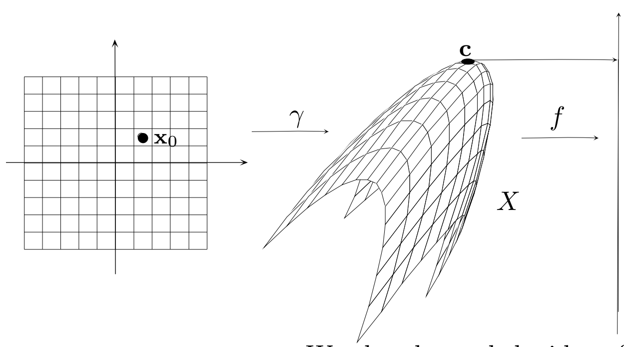

FIGURE 3.7.1 The parametrization γ takes a point in $\mathbb{R}^2$ to the manifold X; f takes it to $\mathbb{R}$. An extremum of the composition $f \circ \gamma$ corresponds to an extremum of f. $\square$

We already used the idea of the proof in Example 2.9.9, where we found the maximum of $|A\mathbf{x}|$ restricted to the unit circle, for $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

In Examples 3.7.3 and 3.7.4 we know a critical point to begin with, and we show that equation 3.7.1 is satisfied:

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

Theorem 3.7.5 will show how to find critical points of a function restricted to a manifold (rather than defined on the manifold, as in Definition 3.7.1), when the manifold is known by an equation $\mathbf{F}(\mathbf{z}) = \mathbf{0}$.

Examples 3.7.3 and 3.7.4 illustrate constrained critical points. They show how to check that a maximum or minimum is indeed a critical point satisfying Definition 3.7.1.

Suppose a manifold X is defined by the equation $\mathbf{F}(\mathbf{z}) = \mathbf{0}$, where $\mathbf{F} : U \subset \mathbb{R}^n \to \mathbb{R}^{n-k}$ has onto derivative $[\mathbf{DF}(\mathbf{x})]$ for all $\mathbf{x} \in U \cap X$, and suppose $f : U \to \mathbb{R}$ is a C^1 function. Then Definition 3.7.1 says that $\mathbf{c}$ is a critical point of f restricted to X if

$$T_{\mathbf{c}}X \underbrace{=}_{\text{Thm. 3.2.4}} \ker\,[\mathbf{DF}(\mathbf{c})] \underbrace{\subset}_{\text{Def. 3.7.1}} \ker\,[\mathbf{D}f(\mathbf{c})]. \qquad 3.7.1$$

Note that both derivatives in formula 3.7.1 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