Overview of chapter 12. Just as curves in the plane are described (at least locally) as graphs of functions, **surfaces** in space are described as the set of point $(x, y, z) \in \mathbb{R}^3$ satisfying $z = f(x, y)$ with f a function of two variables.

Formal Derivatives. Most functions will be given by expressions, and the differential calculus of functions of several variables uses the same rules as single variable calculus. The rules are the same ones used by *Maple* — identify **one** independent variable, and treat all other variables as constants. The formulas of elementary calculus apply. Most attention will be devoted to functions of two variables, since that step introduces most of the new features.

Domain and range. Before getting into the calculus, there are some theoretical issues. The **domain** of a function $f(x, y)$ is the set of pairs $(x, y) \in \mathbb{R}^2$ for which you are allowed to use the formula. It should take into account rules like *do not divide by zero*, or *only take square roots of nonnegative numbers*, but some *arbitrary* restrictions may be added to the description of the function. It is almost always easy to find the domain of a function.

The **range** is the set of numerical values that can be attained by the function. For simple examples, you are able to find the range "by inspection", but the general problem of finding the range includes finding the maximum and minimum of the function, which is the goal of the chapter.

Limits and continuity. Applications of the derivative are based on being able to define derivatives as **limits**. As in the study of functions of one variable, a function is called **continuous** on its domain if the limit of the function as one approaches a point P of the domain is the value of f at P. Almost all functions given by expressions that we know how to write are easily proved to be continuous. This allows limits of functions at points outside the domain, but on the boundary of the domain, to be found by replacing the expression for the function by an equivalent one that can be defined at the desired point. In contrast to the single variable case, *this rarely works*.

Definitions are given, as in the single variable case, in terms of those annoying characters ϵ and δ. In some sense, the definitions are the same, so any proofs will look familiar. However, the examples are different: functions of two variables can fail to be continuous in more interesting ways.

Tangent planes. Just as a lot of single variable calculus dealt with tangent lines to curves, the concept of a **tangent plane** to a surface at a point P provides focus to the study of derivatives of functions of two variables. With few exceptions, if you *zoom in* on a surface near P, the surface looks like a plane. For the graph of $z = f(x, y)$, the plane found in this way can be calculated using derivatives of f. While formal derivatives are used to find tangent planes, the defining property asserts that the graph will be **approximated** by this tangent plane near P. Once more calculus leads to simple computation of useful things.

Total derivatives and gradients. Motivated by the way in which the derivative will be used, the derivatives we calculate are considered to be *partial* derivatives that are the components of a *vector* called the **gradient** of the function. Similar considerations lead to a related quantity called the **total differential**. The gradient should be preferred, since it can be used to express results in **geometric** terms that are *coordinate-free*.

The chain rule. When we think you are ready for it, we introduce a **chain rule**. Again the problem of finding dz/dt when $z = f(x, y)$ and x and y are functions of t contains most of the new features.

Lagrange multipliers. To find the maximum and minimum of a function of three variables on a surface in $\mathbb{R}^3$, special techniques are used. There is some difficulty proving that the method works, but the result is easy to state — and easy to use. If you want to find the extreme values of $f(x, y, z)$ on the surface given by $g(x, y, z) = 0$, you find the points on $g(x, y, z) = 0$ where the gradients of f and g are parallel. For smooth functions, the value of the function at *other* points can be shown *not* to be an extreme value. If you know for other reasons that the f has extreme values on the surface, then you have identified a small set (frequently a *finite* set) of values that need to be examined closely.

Exercise 11.9.27. This was started at end of the last lecture before exam, but not done in an efficient manner. The following treatment is slightly different from the method developed in the text, and leads to an *algorithm* for curvature that is not easily summarized in a formula. We are looking for a calculation of the scalar coefficients of **T** and **N** in

$$\mathbf{r}''(t) = \frac{d^2s}{dt^2}\,\mathbf{T} + \kappa\left(\frac{ds}{dt}\right)^2\mathbf{N},$$

when

$$\mathbf{r}(t) = (t - \sin t)\,\mathbf{i} + (1 - \cos t)\,\mathbf{j}.$$

In stating the problem, the word "components" is used to refer to these coefficients, which should suggest

$$\begin{aligned}\frac{d^2s}{dt^2} &= \operatorname{comp}_{\mathbf{T}}\mathbf{r}''(t)\\ &= \operatorname{comp}_{\mathbf{r}'(t)}\mathbf{r}''(t)\end{aligned}$$

since $\mathbf{T} \parallel \mathbf{r}'(t)$

$$= \frac{\mathbf{r}'(t) \cdot \mathbf{r}''(t)}{|\mathbf{r}'(t)|}$$

In this case,

$$\begin{aligned}
\mathbf{r}(t) &= (t - \sin t)\,\mathbf{i} + (1 - \cos t)\,\mathbf{j} \\
\mathbf{r}'(t) &= (1 - \cos t)\,\mathbf{i} + \sin t\,\mathbf{j} \\
\mathbf{r}''(t) &= \sin t\,\mathbf{i} + \cos t\,\mathbf{j} \\
\left|\mathbf{r}'(t)\right|^2 &= 2 - 2\cos(t) \\
\mathbf{r}'(t) \cdot \mathbf{r}''(t) &= \sin t
\end{aligned}$$

Thus,

$$\frac{d^2 s}{dt^2} = \sin t\big(2 - 2\cos(t)\big)^{-1/2}.$$

Also,

$$\begin{aligned}
\operatorname{proj}_{\mathbf{T}} \mathbf{r}''(t) &= \operatorname{proj}_{\mathbf{r}'(t)} \mathbf{r}''(t) \\
&= \frac{\mathbf{r}'(t) \cdot \mathbf{r}''(t)}{|\mathbf{r}'(t)|^2}\mathbf{r}'(t) \\
&= \frac{\sin t}{2 - 2\cos(t)}\big((1 - \cos t)\,\mathbf{i} + \sin t\,\mathbf{j}\big)
\end{aligned}$$

If this is subtracted from $\mathbf{r}''(t)$, one gets a vector perpendicular to $\mathbf{T}$ (a fact which can be used to derive these formulas easily). Simplifying this gives

$$\kappa\left(\frac{ds}{dt}\right)^2 \mathbf{N} = \frac{1}{2}\sin t\,\mathbf{i} - \frac{1}{2}(1-\cos t)\,\mathbf{j}.$$

This vector is *clearly* perpendicular to $\mathbf{T}$, and finding its length gives

$$\kappa\left(\frac{ds}{dt}\right)^2 = \frac{1}{2}\big(2 - 2\cos(t)\big)^{1/2}.$$

Definition of limit. It is useful at this point to review the definition of limit, extended to cover functions of several variables. For simplicity, we use only two variables.

We say that

$$\lim_{(x,y)\to(a,b)} f(x, y) = L \tag{$*$}$$

if we can force $f(x, y)$ to be arbitrarily close to L (within an ϵ given by *el Exigente*) by demanding that (x, y) be sufficiently close to (a, b) (at a distance no more than a δ that we choose), independent of direction.

Note that it is the whole sentence $(*)$ that is defined. The use of the equal sign in that definition is justified by . . . .

The main theorem. Limits are unique. That is, given the function f and the point (a, b), there is *at most one* choice of L satisfying $(*)$.

Proof. If not, let L_0 and L_1 be two different numbers that both satisfy $(*)$. The triangle inequality gives

$$|L_0 - L_1| \leq |L_0 - f(x, y)| + |f(x, y) - L_1| \quad (\dagger)$$

for any point (x, y) in the domain of f. If you persist in claiming that L_0 and L_1 both satisfy $(*)$, I calculate

$$d = |L_0 - L_1| .$$

To say that $L_0 \neq L_1$ is to say that $d > 0$ (strictly). Then I choose $\epsilon = d/2$. Then, if (x, y) is any point that you think is close enough to (a, b), then both terms on the right side of $(\dagger)$ are less than $d/2$, which contradicts $(\dagger)$.

Non-existence of limits. The idea of this proof is used to identify cases in which limits do not exist. If there are two different values v_0 and v_1 such that $f(x, y) = v_0$ for some points (x, y) arbitrarily close to (a, b) while $f(x, y) = v_1$ for other points arbitrarily close to (a, b), then it is not possible for $f(x, y)$ at *all* points (x, y) close to (a, b) to be within $|v_0 - v_1|/2$ of any single value.

Continuity. Almost every function that we know how to express has the property that, at points where the expression for the function can be evaluated, that value may be used for L in $(*)$. This property is called **continuity**. Continuity means that that the value of a function at a point can be approximated by evaluating the function at a nearby point. This is what we do every time we trust our calculators to do much more than verify that $1 + 1 = 2$, so continuity is an abstraction of the idea of *computable*. The rules for building functions that we use are easily seen to preserve continuity, although complicated functions (for which δ is small) may require extraordinary care to give reasonable accuracy.

Derivatives and tangent planes. The geometric version of the existence of a derivative of a function f at a point (a, b) is the existence of a tangent plane to the surface $z = f(x, y)$ at the point where $x = a$ and $y = b$. A tangent plane has an equation of the form $z = Ax + By + C$ for constants A, B and C, which we abbreviate $z = L(x, y)$ — L standing for **linear**. The definition giving the most efficient characterization of tangent planes is to require that, for all $\epsilon > 0$,

$$|f(x, y) - L(x, y)| < \epsilon\sqrt{(x-a)^2 + (y-b)^2}$$

for all (x, y) sufficiently close to (a, b), independent of direction. The expression on the right is chosen so that given any nonzero linear expression L, it must fail to bound L close to (a, b) for some ϵ. This allows the proof of the main theorem to be modified to show that there is at most one such L.

Using tangent lines to find tangent planes. Tangent lines to curves could have been defined in the same way. Hence, intersecting with the plane $y = b$ gives a curve whose equation is $z = f(x, b)$ and a tangent line to this curve at the point where $x = a$. Since tangent lines could have been characterized by the same ϵ-δ definition as tangent planes, these tangent lines *must* lie in the tangent plane. Finally, writing the equation of the tangent plane in the form

$$z = L(x, y) = A(x-a) + B(y-b) + C$$

we can use what we know about tangent lines to show that $C = f(a, b)$, A is the derivative of $f(x, b)$ with respect to x evaluated at $x = a$, and B is the derivative of $f(a, y)$ with respect to y evaluated at $y = b$.

Finally, some calculus. The coefficients in the equation of the tangent line have been identified as values of derivatives, so finding them leads us, after a long detour into theory, to finding derivatives. The point of the theory has been to show the importance of the operation of differentiating with respect to one variable while holding other variables constant. This is easy to do: you simply follow the rules from single-variable calculus. It only remains to describe the notation. If z has been defined by some expression in the variables x and y, then the derivative of z with respect to x, treating y as a constant, is denoted

$$\frac{\partial z}{\partial x} \text{ or } D_x z.$$

As usual, the result of finding this derivative is an expression that usually involves x and y. To use this result to find a tangent line, you will need to evaluate the expression at $(x, y) = (a, b)$. There is no nice notation for this.

Alternatively, you can think of the given expression as defining z as a function of x and y, which you typically write $z = f(x, y)$. This notation means that the function f requires two input variables and the value of the function is found by using the given expression with the first variable assigned to x and the second to y. One of the things that is typically done with functions is to evaluate them at arbitrary expressions. Thus, if

$$f(x, y) = x + y^2,$$

then

$$f(y, x) = y + x^2.$$

What this example shows is that one should never assume that the variables used to describe a function have any significance whatsoever. Unfortunately, most of the other notations used in the textbook violate this rule. The only notation that doesn't is one that uses f_1 to stand for the derivative of the function f with respect to its first variable. This has the advantage that one can write $f_1(a, b)$ for the result of evaluating this function at the point (a, b), i.e., the result of first differentiating the function and then evaluation the result at the base point.

Higher derivatives. Once one has a derivative, either as an expression or as a function, one can think of differentiating *that*. If one expects to do a lot of that sort of thing, an abbreviation is needed. Thus one writes

$$\frac{\partial^2 z}{\partial x^2} \text{ for } \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right); \ \frac{\partial^2 z}{\partial y \partial x} \text{ for } \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right);$$

and f_{12} for $(f_1)_2$.

Fortunately, you only need to be careful about the order of the variables in this expression when making the definition, since $f_{12} = f_{21}$ for functions that you will meet in practice.

The chain rule.There are other derivatives involving functions of several variables that can be found. Suppose that z is given in terms of x and y, and that x and y are each given in terms of t. You could (and *Maple* does) use this to find z in terms of t and then calculate $D_t z$ (this is the neatest notation for this discussion).

Alternatively, you could apply the rules of differentiation to the expressions that you are given. Whatever the expression for z tells you is the last step in its computation is the first differentiation formula to be applied. In this process, expressions equal to $D_t z$ are obtained that can contain x, y, t, $D_t x$ and $D_t y$. Since x and y are given in terms of t, their expressions are used when you need to expand $D_t x$ and $D_t y$ in terms of t. The idea behind the chain rule for functions of several variables is to delay the expansion of $D_t x$ and $D_t y$ as long as possible. This gives an expression for $D_t z$ in which t has yet to appear outside of a subscript.

The fundamental linearity of differentiation formulas. Let's take a close look at the formulas of elementary calculus:

$$D_t(x+y) = D_t x + D_t y \tag{S}$$

$$D_t(x \cdot y) = D_t x \cdot y + x \cdot D_t y \tag{P}$$

$$D_t\big(f(x)\big) = f'(x) \cdot D_t x \tag{C}$$

These rules suffice to differentiate all the functions met so far, when supplemented by special formulas for differentiating functions given by $f(x) =$ one the following expressions: a constant, x^n, e^x, $\ln x$, $\sin x$, or $\cos x$, $\arctan x$. A few more formulas are obtained to avoid deriving them from other formulas every time they are needed, but this short list of formulas is a good summary of elementary calculus. Of course, the course really deals with the understanding of functions that allows you to use these formulas and apply the results to things in the real world that can be modeled by this mathematical abstraction.

The thing to notice about formulas (S), (P) and (C) is that each *term* contains a *factor* that is a single application of D_t. This means that, in the setting at the start of this lecture,

$$D_t z = A \cdot D_t x + B \cdot D_t y, \qquad (*)$$

where A and B are expressions involving x and y.

Connection with partial derivatives. Formula $(*)$ holds independent of the dependence of x and y on t. The special cases used to define partial derivatives (or, at least, their values at particular points) are obtained by using the parameterizations: (1) $x = t$, $y = b$; or (2) $x = a$, $y = t$. This shows that $A = D_x z$ and $B = D_y z$. The usual statements of the chain rule are obtained by translating this statement into different notations.

A word about proving the chain rule. This discussion has emphasized how the chain rule appears in calculations, suppressing all aspects of proof. However, this is not completely lacking in mathematical rigor. The rules of elementary calculus were obtained with proofs based on a definitions in terms of limits. In fact, it would have been better to use the “good linear approximation” version introduced in connection with our discussion of the tangent plane. Inside these proofs are rules for obtaining a δ from a given ϵ, although we usually don’t look that closely.

For functions given by expressions that we recognize, our calculation is a proof that the derivatives exist. In fact, a close examination of what we have when we have reached the stage of formula $(*)$ shows that the surface defined by the given expression of z in terms of x and y has a tangent plane wherever the calculation is valid.

What this approach does *not* do is tell how to deal with a new function of several variables. What should be done, if such a function is ever met, is to show from the definition of the function that it has a tangent plane, and then prove a version of the chain rule we have already stated in the case where the graph of $z = f(x, y)$ has a tangent plane and the curve $x = g(t)$, $y = h(t)$ is differentiable.

Since we are not likely to meet any new functions of several variables in this course, this approach is more suitable for *Advanced Calculus*. However, in a sense, this was exactly what was done in deriving formulas (S) and (P)! Look closely at the expressions $x + y$ and $x \cdot y$. What are the partial derivatives? Does this agree with the statement given for the chain rule? (Of course, these questions are rhetorical.)

Gradients. There is one more thing to be seen in $(*)$. Whenever one has a sum of terms, each of which is a product of something of one type and something of another, it should be viewed as a dot product of vectors. We have already met the vector $\langle D_t x, D_t y\rangle$ as the velocity vector when $\mathbf{r}(t) = \langle x, y\rangle$ gives the position of a point at time t. This model suggests that it won't be long before we try to write $\mathbf{r}'(t)$ in the form $(ds/dt)\mathbf{T}$, but first we collect the other factors into a vector $\left\langle D_x z, D_y z\right\rangle$. When differentiating a function f instead of an expression z, this has the form $\langle f_1, f_2\rangle$. In this form, it is easy to imagine the generalization to functions of any number of variables. This vector is called the **gradient** of f and denoted ∇f. Gradients are very much a "function thing" since it emphasizes the domain of the function rather than the range — there is no good notation for the same object constructed from an expression.

Like all other derivatives, gradients will be evaluated at points of their domain when they appear in applications.

Directional derivatives. The chain rule can now be expressed as

$$D_t f\big(\mathbf{r}(t)\big) = \nabla f\big(\mathbf{r}(t)\big) \cdot \mathbf{r}'(t).$$

While f is an ordinary real valued function on some $\mathbb{R}^d$ (with $d = 2$ or $d = 3$ for most of the examples in this course, but a common description of those cases leads immediately to a vast generalization), ∇f is something else: for each point $P \in \mathbb{R}^d$, $\nabla f(P)$ is a vector in $\mathbb{R}^d$ based at P. One term frequently used for such a function $\mathbb{R}^d \to \mathbb{R}^d$ is **vector field**. The appearance of $\nabla f(P)$ in an inner product suggests that its principal interpretation will be in terms of expressions of the form $\nabla f(P) \cdot \mathbf{v}$. In particular, if $\mathbf{v}$ is a unit vector, this expression is called the **directional derivative** of f in the direction $\mathbf{v}$. The special case in which $\mathbf{v}$ lies along one of the coordinate axes has already been met under the name "partial derivative".

Now, as promised, we write $\mathbf{r}'(t) = (ds/dt)\mathbf{T}$ to express $D_t f\big(\mathbf{r}(t)\big)$ as the product of the *speed* with which $\mathbf{r}(t)$ is drawing the curve and the directional derivative of f in the *tangential direction* of the curve.

There are two important special cases: $\mathbf{v} \perp \nabla f$ is equivalent to the directional derivative being zero; the directional derivative takes its maximal value when $\mathbf{v} \parallel \nabla f$. Since

$$\nabla f \cdot (-\mathbf{v}) = -\nabla f \cdot \mathbf{v},$$

the minimum value of the directional derivative is in the *anti-parallel* direction.

An important application is that level curves of functions on $\mathbb{R}^2$, or level surfaces in $\mathbb{R}^3$, are perpendicular to the gradient of the function.

Changing coordinates. Since the last step of computing a gradient of f is to draw that vector field at points of the domain of f, the coordinates used in its computation have been pushed into the background. This suggests that it should be possible to perform these computations in other coordinate systems and get the same geometric answer. This is indeed true, with one important requirement — since inner products play an important role in the theory, only systems in which the coordinates are the components of a set of *mutually perpendicular unit vectors* can be used. While it is usually better to use a disjoint set of names for the elements of different coordinate systems, the restriction to orthonormal coordinates allows the same name to appear in different coordinate systems as long as it means the component with respect to the same vector.

Afterthoughts on an exercise. The comments on *Changing coordinates* also apply to implicit differentiation problems in several variables. When doing exercise 29 of section 12.5, there seemed to be one point in which $\partial z/\partial x$ had two different meanings. The textbook avoided this in equation (7) by introducing the derivatives of *functions* instead of variables. Another approach is to use different variables for variables in different spaces. The setting in that problem then becomes

$$\mathbb{R}^2(\xi, \eta) \to \mathbb{R}^3(x, y, z) \to \mathbb{R}(w), \qquad (*)$$

where the coordinates in the spaces are indicated after the name of the space. The new feature is to reserve x, y, and z for the orthogonal coordinates in $\mathbb{R}^3$. The first arrow in $(*)$ stands for the *assumed* ability to describe the surface as the graph of a function $z = g(x, y)$ giving a mapping of the xy-plane to the given surface. However, we indicated that the variables x and y are only available to represent rectangular coordinates in space, not coordinates on the surface. We can accommodate this pedantry by using the equivalents of x and

y in the Greek alphabet to represent the coordinates in the xy plane. The first arrow in $(*)$ is then the function $(\xi, \eta) \mapsto (\xi, \eta, g(\xi, \eta))$, while the second arrow is the given function $F(x, y, z)$ that is assumed constant on the surface.

It may be possible to blunder through simple problems, but the partial derivative notation can be misleading because it omits mention of the variables that are assumed constant.

Maxima and minima. Everyone's favorite problem of differential calculus is the determination of maxima and minima of a function on a region. An important theoretical result asserts that, if the function is *continuous* and the region *closed* and *bounded* in some $\mathbb{R}^n$, then these extreme values exist and are attained at points of the domain. Calculus improves this by saying that, if the function is *differentiable*, then the points at which the extreme values are taken on are **easy to find**. An important feature of the method is that the emphasis has shifted from the values of the function to points in the domain where the function takes those values.

To apply the methods of one-variable calculus to functions of several variables, we need the following

Secret Weapon. *If $A \subseteq B$ and if the maximum of a function f on B is taken on at a point $x \in A$, then $f(x)$ is also the maximum of f on A.*

Proof. Think about it!

To apply this, let B be the given region in $\mathbb{R}^n$ and let A be a curve lying in B. Composing the given function on B with the parameterization of A gives a real valued function of a real variable, and one-dimensional calculus applies.

You may protest that we don't know A, but that will be dealt with. What one-variable calculus gives us is a **necessary condition** for a point to be a maximum or a minimum. If we can find *any* curve A for which this condition fails, then the point is not an extreme value of the function on B and we can look somewhere else. It turns out that points on the boundary of B behave differently than interior points. Interior points are easier, so we do them first. However, it often turns out that all interior points can be excluded as possible locations of extreme vales, so we need to look at the boundary. In $\mathbb{R}^2$, this is almost good enough, since the boundary will frequently be a curve for which we know a parametric description. The general method turns out to be easier, even in this case, but we have a method that can find extrema.

The key special case. What then do we take as A? The simplest examples turn out to suffice: let A be a line segment parallel to one of the coordinate axes and lying in B. Every interior point has segments through it parallel to each coordinate axis. The coordinate on the axis can be taken as the parameter on the segment. The derivative with respect to this parameter is exactly the partial derivative of the given function with respect to the selected coordinate. The *secret weapon* then implies

Theorem. *If f takes on its maximum (or minimum) value on the set B at an interior point P, then all partial derivatives of f are zero at P.*

This result suffices to do the exercises.

Interior points that are not ruled out by this result are called **critical points** of the function. A boundary point should not be called a critical point if we can find a curve in the boundary such that the derivative of the function restricted to that curve is not zero. An appropriate criterion appears in section 12.8.

The second derivative test. A slight generalization of the same principle leads to a second derivative test for distinguishing maxima and minima. For this test, allow A to be any line segment containing an interior critical point P and compute the second derivative of the given function restricted to A. Here is how this is done in $\mathbb{R}^2$:

Let $P = (x_0, y_0)$ and choose A to be $x = x_0 + at$, $y = y_0 + bt$ for some real numbers a and b. If a and b are not both zero, this is a line with $t = 0$ corresponding to the point P. Put this into $f(x, y)$ and apply the chain rule:

$$
\begin{aligned}
\frac{d}{dt} f(x_0 + at, y_0 + bt) = \\
& f_1(x_0 + at, y_0 + bt) \cdot a \\
& + f_2(x_0 + at, y_0 + bt) \cdot b.
\end{aligned}
$$

Differentiating again gives sums of terms that are second partial derivatives of f times second degree terms in a and b. When we restrict this to $t = 0$ we must get an expression that does not change sign when evaluated at all choices of a and b. If P is a maximum point, the expression must not be positive; is p is a minimum point, it must not be negative.

If there are only a few critical points, it will be easier to list them all with the value of the function at each, than to perform a second derivative test. The chief interest in the test is theoretical. It demonstrates the existence of critical points that can be neither minima nor maxima because there are both positive and negative second derivatives. A simple example is $x^2 - y^2$.

Lagrange multipliers. We now characterize critical points on the boundary of a region. Let P be a point on the boundary of the set B.

Suppose first that a smooth curve can be drawn on the boundary of B *through* P. That is, there is a vector function $\mathbf{r}(t)$ whose domain contains a neighborhood of 0 and whose range lies entirely in the boundary of B, with $\mathbf{r}(0) = P$, and $\mathbf{r}'(0) \neq \mathbf{0}$. The last part is the *smoothness* condition. If B is a polygon, then any $\mathbf{r}(t)$ for which $\mathbf{r}(0)$ is a vertex of the polygon must have $\mathbf{r}'(0) = \mathbf{0}$ if the derivative exists. This follows from the fact that a nonzero $\mathbf{r}'$ has a well defined direction, but this direction changes when when one goes through a vertex.

Now, suppose that B is given in the form

$$\left\{ \mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) \leq 0 \right\}.$$

The boundary is given by $g(\mathbf{x}) = 0$. Hence, the composition $g(\mathbf{r}(t))$ is a constant function, so must have derivative zero. The chain rule, evaluated at $t = 0$, gives

$$\nabla g(P) \cdot \mathbf{r}'(0) = 0.$$

This says that $\mathbf{r}'(0)$ must be perpendicular to $\nabla g(P)$. If the boundary of B is smooth at P, then curves $\mathbf{r}(t)$ can be found for which $\mathbf{r}'(0)$ takes all values allowed by this condition.

Now, look at the function $f(\mathbf{x})$ that we are trying to maximize (or minimize). If we form $f(\mathbf{r}(t))$, this function of t must have a critical point at $t = 0$, which requires that

$$\nabla f(P) \cdot \mathbf{r}'(0) = 0.$$

In words, every vector perpendicular to $\nabla g(P)$ must also be perpendicular to $\nabla f(P)$. For points at which the boundary is smooth, this forces the existence of a value traditionally called λ for which $\nabla f(P) = \lambda \nabla g(P)$. In $\mathbb{R}^n$ this gives n equations in λ and the n coordinate functions. The condition, $g(P) = 0$ gives one more equation. We not have as many equations as we have variables, and we should be able to solve this system of equations. Unfortunately, the algebra for doing this is sometimes difficult. A systematic approach to this algebra is possible, but we don't have time to develop one. One consequence of this is that only simple examples will be considered.

This approach can be extended to deal with smooth points of an object of any dimension d in any $\mathbb{R}^n$, but few examples reduce to algebra that can be done easily.

There is one easy application of Lagrange multipliers that appears in many variants. There are two forms that reduce to the same algebra:

(1) Let x and y be nonnegative with $x + y = s$ (s constant). Find the maximum of xy.

(2) Let x and y be positive with $xy = p$ (p constant). Find the minimum of $x + y$.

Since our notation for gradients uses functions rather than expressions, introduce the functions $S(x, y) = x + y$ and $P(x, y) = xy$. Then the Lagrange multiplier method for either problem (1) or problem (2) says that extreme values occur when $\nabla S \parallel \nabla P$. Since $\nabla S = \langle 1, 1\rangle$ and $\nabla P = \langle y, x\rangle$, this says $x = y$. In (1), the given constraint then gives $x = y = s/2$ and $xy = s^2/4$. In (2), we get $x = y = \sqrt{p}$ and $x + y = 2\sqrt{p}$.

In each problem, the method selects a unique point.

The only other candidates for the location of extreme values of the function are the endpoints. In (1), the product is zero at both endpoints; in (2), there are no true endpoints, but $x + y \to \infty$ outside bounded parts of the curve. This explains how we knew that the interior point gave a maximum in (1) and a minimum in (2).

Quadric surfaces. The general form of an equation of second degree is mentioned in the text (and in Lab 2), but no techniques are given for studying the equation in that generality. It is the case that, apart from some special cases like cones and cylinders, the equations studied in detail in the text represent all cases. That is, there is a *rigid* change of coordinates reducing any equation of second degree to one of these types. (The word "rigid" refers to compositions of translations and rotations.) Finding such a change of coordinates would lead us into an advanced part of linear algebra, so we confine attention to the special cases.

If the equation has the form

$$Ax^2 + By^2 + Cz^2 + J = 0,$$

then the shape of surface depends on the signs of the coefficients.

The case $J = 0$ gives a cone (unless the remaining coefficients all have the same sign). Ignoring these surfaces, we may take scale the equation so that $J = -1$. The shape of the surface then depends on the signs of the remaining coefficients, and we emphasize this by writing $A = \pm 1/a^2$, and similarly for the other coefficients. The values of a, b and c then refer to scales on the different coordinate axes. These scale factors introduce only a simple stretching of the figure that does not affect the type of the surface. Finally, the order of the variables x, y and z can be permuted so that only the number of signs is significant. This leaves only four examples:

$$x^2 + y^2 + z^2 = 1; \tag{S}$$

$$x^2 + y^2 - z^2 = 1; \tag{$H1$}$$

$$x^2 - y^2 - z^2 = 1; \tag{$H2$}$$

$$-x^2 - y^2 - z^2 = 1. \tag{E}$$

We recognize (S) as a sphere, which becomes an ellipsoid when scales are introduced. Also (E) is empty, since the left side is negative and the right side is positive. The remaining two cases give two distinct types of hyperboloid. The analysis of equation (E) give the key to the general study of these surfaces: we can fix z and investigate the equation in x and y. This corresponds to taking sections parallel to the xy-plane. In case (S), only $-1 \leq z \leq 1$ leads to nonempty sections, and those sections are ellipses; in case $(H1)$, there are elliptical sections for all z, and none are small; in case $(H2)$, these sections exist for all z, but the sections are hyperbolas. Sections parallel to the yz-plane in case (H2) are ellipses, but are nonempty only for $|x| \geq 1$. The surface in case $(H2)$ falls into two parts, while the other cases are connected.

Be scaling coordinates, we produce equations that contain $x^2 + y^2$ or $y^2 + z^2$. These quantities give the distance from the origin in a coordinate plane, independent of direction, so the surface can be seen as a figure of rotation.

A similar analysis can be made for surfaces of the type

$$z = Ax^2 + By^2,$$

which cover the remaining cases.

This latter family of examples give the general appearance of surfaces which are the graphs of quadratic functions. The second derivative test that we mentioned briefly is based on showing that the graph of a general function near a critical point resembles the graph of the quadratic function with the same values of second derivatives.