



9. Functions of several variables

These lecture notes present my interpretation of Ruth Lawrence's lecture notes (in Hebrew)

1

9.1 Definition

In the previous chapter we studied *paths* (מסילות), which are functions $\mathbb{R} \rightarrow \mathbb{R}^n$. We saw a path in \mathbb{R}^n can be represented by a vector of n real-valued functions. In this chapter we consider functions $\mathbb{R}^n \rightarrow \mathbb{R}$, i.e., functions whose input is an ordered set of n numbers and whose output is a single real number. In the next chapter we will generalize both topics and consider functions that take a vector with n components and return a vector with m components.

■ **Example 9.1** Consider the function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by

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

We may as well write

$$f: (s, t) \mapsto s^2 + t \sin s,$$

but since the number of arguments can be larger, it is more systematic to use the vector notation,

$$f(\mathbf{x}) = f(x_1, x_2) = x_1^2 + x_2 \sin x_1.$$

■

¹Image of Joseph-Louis Lagrange, 1736–1813

■ **Example 9.2** Like for the case $n = 1$, the domain of a function may be a subset of \mathbb{R}^n . For example, let

$$D = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 < 1\} \subset \mathbb{R}^3,$$

or in vector notation,

$$D = \{\underline{\mathbf{x}} \in \mathbb{R}^3 \mid |\underline{\mathbf{x}}| < 1\} \subset \mathbb{R}^3.$$

Define $g : D \rightarrow \mathbb{R}$ by

$$g(\underline{\mathbf{x}}) = \ln(1 - |\underline{\mathbf{x}}|^2) \quad \text{or} \quad g(x, y, z) = \ln(1 - x^2 - y^2 - z^2).$$

Then, for example, $(1/2, 1/2, 1/2) \in D$ and

$$g(1/2, 1/2, 1/2) = \ln \frac{1}{4}.$$

■

Comment 9.1 It is very common to denote the arguments of a bi-variate function by (x, y) and of a tri-variate function by (x, y, z) . We will often do so, but remember that there is nothing special about these choices.

■ **Example 9.3 Important example!** Recall that if V and W are vector spaces, we denote by $L(V, W)$ the space of **linear functions** $V \rightarrow W$. Let $\underline{\mathbf{a}} \in \mathbb{R}^n$. Then the function,

$$f : \mathbb{R}^n \rightarrow \mathbb{R} \quad \text{defined by} \quad f : \underline{\mathbf{x}} \mapsto \underline{\mathbf{a}} \cdot \underline{\mathbf{x}}$$

is linear, i.e., belongs to $L(\mathbb{R}^n, \mathbb{R})$, in the sense that for every $\underline{\mathbf{x}}, \underline{\mathbf{y}} \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$,

$$f(\alpha \underline{\mathbf{x}} + \beta \underline{\mathbf{y}}) = \alpha f(\underline{\mathbf{x}}) + \beta f(\underline{\mathbf{y}}).$$

For example, taking $n = 3$ and $\underline{\mathbf{a}} = (1, 2, 3)$,

$$f(x, y, z) = (1, 2, 3) \cdot (x, y, z) = x + 2y + 3z$$

is a linear function. In fact, *all* linear functions in $L(\mathbb{R}^n, \mathbb{R})$ are of this form. They are characterized by a vector $\underline{\mathbf{a}}$ that (scalar)-multiplies their argument. ■

■ **Example 9.4** The function

$$f(x, y) = \tan^{-1} \left(\frac{y}{x} \right)$$

can be defined everywhere where $x \neq 0$, i.e., its maximal domain is

$$\{(x, y) \in \mathbb{R}^2 \mid x \neq 0\}.$$

■

■ **Example 9.5** The maximal domain of the function

$$f(x, y) = \ln(x - y) \quad \text{is} \quad \{(x, y) \in \mathbb{R}^2 \mid x > y\}.$$

■

■ **Example 9.6** The maximal domain of the function

$$f(x, y) = \ln|x - y| \quad \text{is} \quad \{(x, y) \in \mathbb{R}^2 \mid x \neq y\}.$$

■

■ **Example 9.7** The maximal domain of the function

$$f(x, y) = \cos^{-1}y + x \quad \text{is} \quad \{(x, y) \in \mathbb{R}^2 \mid -1 \leq y \leq 1\}.$$

■

■ **Example 9.8** The maximal domain of the function

$$f(x, y) = \ln(\sin(x^2 + y^2)) \quad \text{is} \quad \bigcup_{n=0}^{\infty} \{(x, y) \in \mathbb{R}^2 \mid 2n\pi < x^2 + y^2 < (2n+1)\pi\}.$$

■

9.2 The graph of a function $\mathbb{R}^n \rightarrow \mathbb{R}$

9.2.1 Definition

Recall that for every two sets A and B , the graph $\text{Graph}(f)$ of a function $f : A \rightarrow B$ is a subset of the Cartesian product $A \times B$, with the condition that

$$\forall a \in A \quad \exists ! b \in B \quad \text{such that} \quad (a, b) \in \text{Graph}(f).$$

The value returned by f , $f(a)$, is the unique $b \in B$ that pairs with a in the graph set. In other words,

$$\text{Graph}(f) = \{(a, b) \in A \times B \mid b = f(a)\}.$$

Thus, if $D \subset \mathbb{R}^n$ and $f : D \rightarrow \mathbb{R}$, the graph of f is a subset of $D \times \mathbb{R} \subset \mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1}$,

$$\text{Graph}(f) = \{(x_1, x_2, \dots, x_n, z) \mid (x_1, \dots, x_n) \in D, z = f(x_1, \dots, x_n)\}.$$

For the particular case of $n = 2$,

$$\text{Graph}(f) = \{(x, y, z) \mid (x, y) \in D, z = f(x, y)\},$$

which is a surface in \mathbb{R}^3 .

■ **Example 9.9** The graph of the function $f(x, y) = x^2 + y^2$ whose domain of definition is the whole plane is a **paraboloid**,

$$\text{Graph}(f) = \{(x, y, z) \in \mathbb{R}^3 \mid z = x^2 + y^2\}.$$

■ **Example 9.10** The graph of any function of the form

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

where $h: \mathbb{R} \rightarrow \mathbb{R}$, is a **surface of revolution**.

9.2.2 Slices of graphs

Consider a function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, whose graph is

$$\text{Graph}(f) = \{(x, y, z) \mid z = f(x, y)\} \subset \mathbb{R}^3.$$

We obtain **slices** (חתיכים) of the graph by intersecting it with planes. For example, the intersection of this graph with the plane

$$“\{x = x_0\}” = \{(x, y, z) \in \mathbb{R}^3 \mid x = x_0\}$$

is

$$\{(x_0, y, z) \mid z = f(x_0, y)\},$$

which is the graph of a function of one variable (z as function of y for fixed $x = x_0$). Similarly, the intersection of the graph of f with the plane $y = y_0$ is

$$\{(x, y_0, z) \mid z = f(x, y_0)\},$$

which is also the graph of a function of one variable (z as function of x for fixed $y = y_0$). Thus, the fact that f is a function implies that both

$$y \mapsto f(x_0, y) \quad \text{and} \quad x \mapsto f(x, y_0)$$

are also functions (of one variable).

Finally, the intersection of the graph of f with the plane

$$“\{z = z_0\}” = \{(x, y, z) \in \mathbb{R}^3 \mid z = z_0\}$$

is the set

$$\{(x, y, z_0) \mid z_0 = f(x, y)\}.$$

This set is called a **contour line** (קו גובה) of f . It is a subset of space parallel to the xy -plane; it could be empty, it could be a closed curve, or a more complicated domain (even the whole of \mathbb{R}^2). The important observation is that in general a contour line is not the graph of a function.

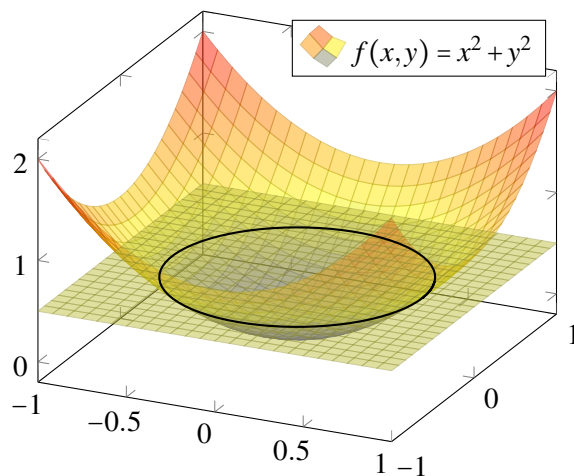
■ **Example 9.11** Consider the function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$,

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

Its graph is

$$\text{Graph}(f) = \{(x, y, z) \mid z = x^2 + y^2\}.$$

(It is a **paraboloid**.)



The intersection of its graph with the plane $x = a$ is

$$\{(a, y, z) \mid z = a + y^2\},$$

which is a parabola on a plane parallel to the yz -plane. The intersection of its graph with the plane $z = R$ is

$$\{(x, y, R) \mid R = x^2 + y^2\},$$

which is empty for $R < 0$, a point for $R = 0$ and a circle in a plane parallel to the xy -plane for $R > 0$.

■

■ **Example 9.12** Consider a linear function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$,

$$f(x, y) = ax + by.$$

Its graph is a plane in \mathbb{R}^3 ,

$$\text{Graph}(f) = \{(x, y, z) \mid z = ax + by\}.$$

All its slices are lines, e.g.,

$$\{(x, y, z_0) \mid z_0 = ax + by\}.$$

■

9.3 Continuity

Let $f : D \rightarrow \mathbb{R}$, where $D \subset \mathbb{R}^n$. Loosely speaking, f is continuous at a point $\mathbf{a} = (a_1, \dots, a_n)$ if small deviations of $\mathbf{x} = (x_1, \dots, x_n)$ about \mathbf{a} imply small changes in f . More formally, f is continuous at \mathbf{a} if for every $\varepsilon > 0$ there exists a neighborhood of \mathbf{a} , such that for every \mathbf{x} is that neighborhood,

$$|f(\mathbf{x}) - f(\mathbf{a})| < \varepsilon.$$

All the functions that we will meet in this chapter will be continuous in their domains of definition, but be aware that there are many non-continuous functions.

■ **Example 9.13** The function

$$f(\mathbf{x}) = x_1^2 + x_2^2$$

is continuous at $\mathbf{a} = (1, 2)$, because for every $\varepsilon > 0$ we can find a neighborhood of $(1, 2)$, say,

$$D = \{\mathbf{x} \in \mathbb{R}^2 \mid (x_1 - 1)^2 + (x_2 - 2)^2 < \delta\},$$

such that

$$\text{for all } \mathbf{x} \in D \quad |f(\mathbf{x}) - 5| < \varepsilon.$$

■

9.4 Directional derivatives

Consider a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ in the vicinity of a point $\mathbf{a} = (a_1, a_2)$. Like for functions in \mathbb{R} , we often ask at what rate does f change when we slightly modify its argument. The difference is that here, it has two arguments that can be modified independently.

One possibility is to keep, say, the second argument, x_2 fixed at a_2 , and evaluate f as we vary the first argument, x_1 , from the value a_1 . By fixing the value $x_2 = a_2$, we are in fact considering a function of one variable,

$$x \mapsto f(x, a_2).$$

We could ask then about the rate of change of f as we vary x_1 near a_1 with $x_2 = a_2$ fixed,

$$\lim_{x_1 \rightarrow a_1} \frac{f(x_1, a_2) - f(a_1, a_2)}{x_1 - a_1}.$$

If this limit exists, we call it the **partial derivative** (נגזרת חלקית) of f along its first argument (or in the x -direction) at the point \mathbf{a} . We may denote it by the following alternative notations,

$$D_1 f(\mathbf{a}) \quad \frac{\partial f}{\partial x_1}(\mathbf{a}) \quad \text{or} \quad \frac{\partial f}{\partial x}(\mathbf{a}).$$

Equivalently,

$$\frac{\partial f}{\partial x_1}(\mathbf{a}) = \lim_{h \rightarrow 0} \frac{f(a_1 + h, a_2) - f(a_1, a_2)}{h}.$$

Similarly, if the limit

$$\lim_{x_2 \rightarrow a_2} \frac{f(a_1, x_2) - f(a_1, a_2)}{x_2 - a_2} = \lim_{h \rightarrow 0} \frac{f(a_1, a_2 + h) - f(a_1, a_2)}{h}.$$

exists, we call it the partial derivative of f along the second coordinate (or in the y -direction) and denote it by

$$D_2 f(\mathbf{a}) \quad \frac{\partial f}{\partial x_2}(\mathbf{a}) \quad \text{or} \quad \frac{\partial f}{\partial y}(\mathbf{a})$$

More generally, if $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $\mathbf{a} \in \mathbb{R}^n$, then (assuming that the limit exists)

$$D_j f(\mathbf{a}) = \frac{\partial f}{\partial x_j}(\mathbf{a}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{a} + h \hat{\mathbf{e}}_j) - f(\mathbf{a})}{h},$$

where $\hat{\mathbf{e}}_j$ is the j -th unit vector in \mathbb{R}^n .

Comment 9.2 The symbol ∂ for partial derivatives is due to the Marquis de Condorcet in 1770 (a French philosopher, mathematician, and political scientist).

Partial derivatives quantify the rate at which a function changes when moving away from a point in very specific directions (along the unit vectors $\hat{\mathbf{e}}_j$). More generally, we could evaluate the rate of change of a function along any unit vector $\hat{\mathbf{v}}$ in \mathbb{R}^n ,

$$D_{\hat{\mathbf{v}}} f(\mathbf{a}) = \frac{\partial f}{\partial \hat{\mathbf{v}}}(\mathbf{a}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{a} + h \hat{\mathbf{v}}) - f(\mathbf{a})}{h}.$$

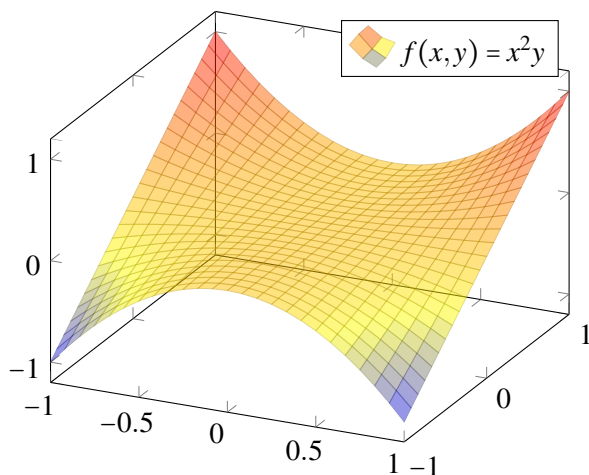
Such a derivative is called a **directional derivative** (נגזרת כיוונית). It is the rate of change of f at \mathbf{a} along the direction $\hat{\mathbf{v}}$. Note that $D_{\hat{\mathbf{v}}} f(\mathbf{a})$ is the derivative of the function

$$t \mapsto f(\mathbf{a} + t \hat{\mathbf{v}})$$

at the point $t = 0$. Partial derivatives are particular cases of directional derivatives with the choice $\hat{\mathbf{v}} = \hat{\mathbf{e}}_j$.

■ **Example 9.14** Consider the function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$,

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



The partial derivative of f in the x -direction at a point (x_0, y_0) is the derivative of the function

$$x \mapsto x^2 y_0$$

at $x = x_0$ namely,

$$D_1 f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0) = 2x_0 y_0.$$

The partial derivative of f in the y -direction at a point (x_0, y_0) is the derivative of the function

$$y \mapsto x_0^2 y$$

at $y = y_0$, namely,

$$D_2 f(x_0, y_0) = \frac{\partial f}{\partial y}(x_0, y_0) = x_0^2.$$

Take now any unit vector $\hat{\mathbf{v}} = (\cos \theta, \sin \theta)$. The directional derivative of f in this direction at a point (x_0, y_0) is the derivative of the function

$$t \mapsto f(x_0 + t \cos \theta, y_0 + t \sin \theta) = (x_0 + t \cos \theta)^2 (y_0 + t \sin \theta)$$

at $t = 0$, i.e.,

$$D_{\hat{\mathbf{v}}} f(x_0, y_0) = 2x_0 y_0 \cos \theta + x_0^2 \sin \theta.$$

This means that

$$D_{\hat{\mathbf{v}}} f(x_0, y_0) = \cos \theta \frac{\partial f}{\partial x}(x_0, y_0) + \sin \theta \frac{\partial f}{\partial y}(x_0, y_0).$$

We will soon see that this is a general result. The derivative of f along any direction can be inferred from its partial derivatives (this is not a trivial statement). ■

■ **Example 9.15** Consider the function

$$f(x, y) = \sqrt{x^2 + y^2},$$

whose graph is a cone. Its partial derivatives at (x, y) are

$$\frac{\partial f}{\partial x}(x, y) = \frac{x}{\sqrt{x^2 + y^2}} \quad \text{and} \quad \frac{\partial f}{\partial y}(x, y) = \frac{y}{\sqrt{x^2 + y^2}},$$

which holds everywhere except for the origin where the function is continuous but not differentiable.

Taking an arbitrary direction, $\hat{v} = (\cos \theta, \sin \theta)$

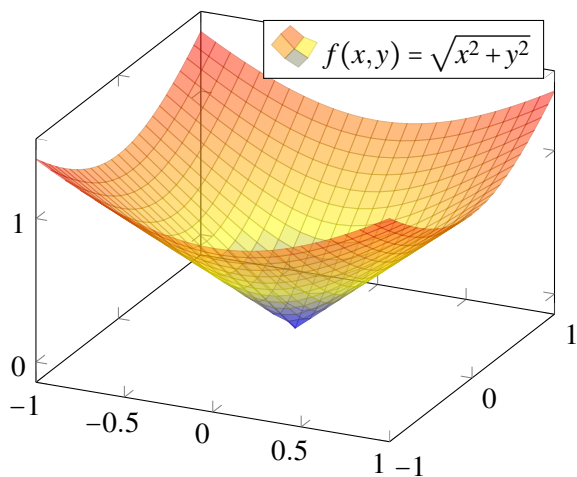
$$D_{\hat{v}}f(x, y) = F'(0),$$

where

$$F(t) = f(x + t \cos \theta, y + t \sin \theta) = \sqrt{(x + t \cos \theta)^2 + (y + t \sin \theta)^2}.$$

Thus,

$$D_{\hat{v}}f(x, y) = \frac{x \cos \theta + y \sin \theta}{\sqrt{x^2 + y^2}} = \cos \theta \frac{\partial f}{\partial x}(x, y) + \sin \theta \frac{\partial f}{\partial y}(x, y).$$



■

9.5 Differentiability

We have defined so far directional derivatives of functions $\mathbb{R}^n \rightarrow \mathbb{R}$, but we haven't yet defined what is a differentiable function, nor we defined what is *the* derivative of a multivariate function. Naively, one could define $f: \mathbb{R}^n \rightarrow \mathbb{R}$ to be differentiable

if all its partial derivatives exist. This requirement turns out not to be sufficiently stringent.

The differentiability of a function $\mathbb{R}^n \rightarrow \mathbb{R}$ generalizes the particular of case $n = 1$. For $n = 1$ there are several equivalent definitions of differentiability: $f: \mathbb{R} \rightarrow \mathbb{R}$ is differentiable at a if

$$\lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h} \text{ exists.}$$

Let's try to generalize this definition. Since f takes for input a vector, it would be differentiable at \mathbf{a} if

$$\lim_{\mathbf{h} \rightarrow 0} \frac{f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a})}{\mathbf{h}} \text{ exists?}$$

The numerator is well-defined; the limit $\mathbf{h} \rightarrow 0$ means that every component of \mathbf{h} tends to zero; but division by \mathbf{h} is not defined. The fraction we're trying to evaluate the limit of is

$$\frac{f(a_1 + h_1, \dots, a_n + h_n) - f(a_1, \dots, a_n)}{(h_1, \dots, h_n)},$$

and this is not a well-formed expression.

The differentiability of a function $f: \mathbb{R} \rightarrow \mathbb{R}$ can also be defined in an alternative way: f is differentiable at a if when x is close to a , $\Delta f = f(x) - f(a)$ is approximately a linear function of $\Delta x = x - a$. That is, there exists a number L , such that

$$f(x) - f(a) \approx L(x - a),$$

in the sense that

$$f(x) - f(a) = L(x - a) + o(|x - a|),$$

i.e.,

$$\lim_{x \rightarrow a} \frac{f(x) - f(a) - L(x - a)}{x - a} = 0.$$

If this is the case, we call the number L the derivative of f at a , and write $f'(a) = L$.

The latter characterization of the derivative turns out to be the one we can generalize to multivariate functions. Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$. Recall that linear functions $\mathbb{R}^n \rightarrow \mathbb{R}$ can be represented by a scalar multiplication by a constant vector. Thus, f is differentiable at \mathbf{a} if when \mathbf{x} is close to \mathbf{a} , i.e., when $|\mathbf{x} - \mathbf{a}|$ is small, $\Delta f = f(\mathbf{x}) - f(\mathbf{a})$ is approximately a linear function of $\Delta \mathbf{x} = \mathbf{x} - \mathbf{a}$.

Definition 9.1 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is differentiable at \mathbf{a} if there exists a vector \mathbf{L} such that

$$f(\mathbf{x}) - f(\mathbf{a}) = \mathbf{L} \cdot (\mathbf{x} - \mathbf{a}) + o(|\mathbf{x} - \mathbf{a}|).$$

In other words,

$$\lim_{|\mathbf{x} - \mathbf{a}| \rightarrow 0} \frac{f(\mathbf{x}) - f(\mathbf{a}) - \mathbf{L} \cdot (\mathbf{x} - \mathbf{a})}{|\mathbf{x} - \mathbf{a}|} = 0.$$

We call the vector $\underline{\mathbf{L}}$ the derivative of f at $\underline{\mathbf{a}}$, or the **gradient** (גרדיינט) of f at $\underline{\mathbf{a}}$, and denote

$$\underline{\mathbf{D}}f(\underline{\mathbf{a}}) = \underline{\mathbf{L}},$$

or

$$\underline{\nabla}f(\underline{\mathbf{a}}) = \underline{\mathbf{L}}.$$

Comment 9.3 Note that the derivative of a multivariate function at a point is a vector. If we turn now the point $\underline{\mathbf{a}}$ into a variable, the gradient $\underline{\nabla}f$ is a function that takes a vector \mathbb{R}^n (a point in the domain of f) and returns a vector in \mathbb{R}^n (the value of the gradient at that point), namely,

$$\text{For } f : \mathbb{R}^n \rightarrow \mathbb{R} \quad \underline{\nabla}f : \mathbb{R}^n \rightarrow \mathbb{R}^n$$

Comment 9.4 The limit $\underline{\mathbf{x}} \rightarrow \underline{\mathbf{a}}$ requires some clarification. It is required to exist regardless of *how* $\underline{\mathbf{x}}$ approaches $\underline{\mathbf{a}}$. For any sequence $\underline{\mathbf{x}}_k$ satisfying $|\underline{\mathbf{x}}_k - \underline{\mathbf{a}}| \rightarrow 0$ we want the above ratio to tend to zero.

Comment 9.5 The fact that $f(\underline{\mathbf{x}}) - f(\underline{\mathbf{a}})$ is approximated by a linear function of $(\underline{\mathbf{x}} - \underline{\mathbf{a}})$ means that

$$f(\underline{\mathbf{a}} + \Delta\underline{\mathbf{x}}) - f(\underline{\mathbf{a}}) = (\underline{\nabla}f(\underline{\mathbf{a}}))_1 \Delta x_1 + \cdots + (\underline{\nabla}f(\underline{\mathbf{a}}))_n \Delta x_n + o(|\Delta\underline{\mathbf{x}}|).$$

The immediate question is what is the relation between the derivative (or gradient) of a multivariate function and its partial derivatives. The answer is the following:

Proposition 9.1 Suppose that $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is differentiable at $\underline{\mathbf{a}}$ with $\underline{\nabla}f(\underline{\mathbf{a}})$. Then all its partial derivatives exist and

$$\frac{\partial f}{\partial x_j}(\underline{\mathbf{a}}) = (\underline{\nabla}f(\underline{\mathbf{a}}))_j.$$

In other words,

$$\underline{\nabla}f(\underline{\mathbf{a}}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\underline{\mathbf{a}}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\underline{\mathbf{a}}) \end{pmatrix}.$$

Proof. By definition,

$$\lim_{|\underline{\mathbf{x}} - \underline{\mathbf{a}}| \rightarrow 0} \frac{f(\underline{\mathbf{x}}) - f(\underline{\mathbf{a}}) - \underline{\nabla}f(\underline{\mathbf{a}}) \cdot (\underline{\mathbf{x}} - \underline{\mathbf{a}})}{|\underline{\mathbf{x}} - \underline{\mathbf{a}}|} = 0,$$

regardless of how $\underline{\mathbf{x}}$ tends to $\underline{\mathbf{a}}$. Set now $\underline{\mathbf{x}} = \underline{\mathbf{a}} + h\hat{\mathbf{e}}_j$ and let $h \rightarrow 0$. Since

$$|\underline{\mathbf{x}} - \underline{\mathbf{a}}| = |h| \rightarrow 0,$$

it follows that

$$\lim_{h \rightarrow 0} \frac{f(\underline{\mathbf{a}} + h\hat{\mathbf{e}}_j) - f(\underline{\mathbf{a}}) - h(\nabla f(\underline{\mathbf{a}}))_j}{h} = 0.$$

By definition, this means that the j -th partial derivative of f at $\underline{\mathbf{a}}$ exists and is equal to $(\nabla f(\underline{\mathbf{a}}))_j$. ■

■ **Example 9.16** Consider the function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$,

$$f(x, y) = \sqrt{1+x} \sin y.$$

Take $\underline{\mathbf{a}} = (0, 0)$. Then,

$$\frac{\partial f}{\partial x}(0, 0) = \left. \frac{\sin y}{2\sqrt{1+x}} \right|_{(0,0)} = 0 \quad \text{and} \quad \frac{\partial f}{\partial y}(0, 0) = \left. \sqrt{1+x} \cos y \right|_{(0,0)} = 1.$$

By Proposition 9.1, if f is differentiable at $(0, 0)$ then its gradient at $(0, 0)$ is a vector whose entries are the partial derivatives of f at that point,

$$\nabla f(0, 0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

To determine whether f is differentiable at $(0, 0)$ we have to check whether

$$\lim_{|(x,y)| \rightarrow 0} \frac{f(x, y) - f(0, 0) - \nabla f(0, 0) \cdot (x, y)}{|(x, y)|} = 0,$$

i.e., whether

$$\lim_{|(x,y)| \rightarrow 0} \frac{\sqrt{1+x} \sin y - y}{\sqrt{x^2 + y^2}} = 0.$$

You can use, say, Taylor's expansions to verify that this is indeed the case. ■

Comment 9.6 Most functions that you will encounter are differentiable. In particular, sums, products and compositions of differentiable functions are differentiable.

■ **Example 9.17** Consider again the function $r: \mathbb{R}^n \rightarrow \mathbb{R}$,

$$r(\underline{\mathbf{x}}) = |\underline{\mathbf{x}}|.$$

Then,

$$\nabla f(\underline{\mathbf{x}}) = \begin{pmatrix} x_1/|\underline{\mathbf{x}}| \\ \vdots \\ x_n/|\underline{\mathbf{x}}| \end{pmatrix} = \hat{\mathbf{x}}.$$

Thus, the gradient of the function that measures the length of a vector is the unit vector of that vector. (We will discuss the meaning of the gradient more in depth below.) ■

It remains to establish the relation between the gradient and directional derivatives.

Proposition 9.2 Suppose that $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is differentiable at \mathbf{a} and let $\hat{\mathbf{v}}$ be a unit vector. Then

$$\frac{\partial f}{\partial \hat{\mathbf{v}}}(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \hat{\mathbf{v}}.$$

That is,

$$\frac{\partial f}{\partial \hat{\mathbf{v}}}(\mathbf{a}) = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(\mathbf{a}) \hat{v}_j.$$

Proof. By definition, since f is differentiable at \mathbf{a} ,

$$\lim_{\|\mathbf{x}-\mathbf{a}\| \rightarrow 0} \frac{f(\mathbf{x}) - f(\mathbf{a}) - \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a})}{\|\mathbf{x} - \mathbf{a}\|} = 0.$$

Set now $\mathbf{x} = \mathbf{a} + h\hat{\mathbf{v}}$ and let $h \rightarrow 0$. Then,

$$\|\mathbf{x} - \mathbf{a}\| = |h| \rightarrow 0,$$

hence

$$\lim_{h \rightarrow 0} \frac{f(\mathbf{a} + h\hat{\mathbf{v}}) - f(\mathbf{a}) - h\nabla f(\mathbf{a}) \cdot \hat{\mathbf{v}}}{|h|} = 0.$$

This precisely means that the directional derivative of f at \mathbf{a} exists and is equal to $\nabla f(\mathbf{a}) \cdot \hat{\mathbf{v}}$. ■

Remains the following question: is it possible that a function has all its directional derivatives at a point, and yet fails to be differentiable at that point? The following example shows that the answer is positive, proving that differentiability is more stringent than the mere existence of directional derivatives.

■ **Example 9.18** Consider the function

$$f(x, y) = (x^2 y)^{1/3}$$

For every direction $\hat{\mathbf{v}} = (\cos \theta, \sin \theta)$,

$$\frac{\partial f}{\partial \hat{\mathbf{v}}}(0, 0) = \lim_{h \rightarrow 0} \frac{f(h \cos \theta, h \sin \theta) - f(0, 0)}{h} = \cos^{2/3} \theta \sin^{1/3} \theta,$$

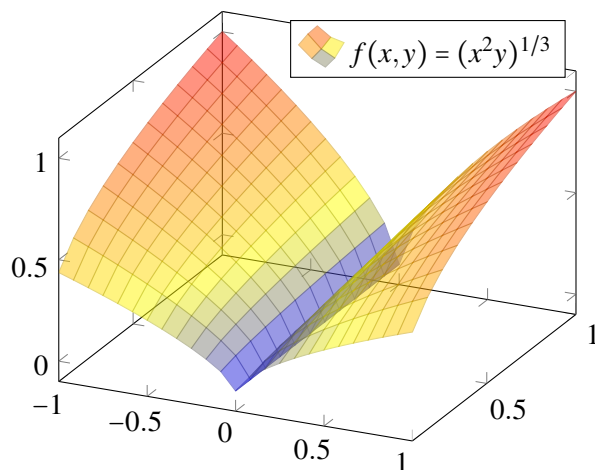
so that all the directional derivatives exist. In particular, setting $\theta = 0, \pi/2$,

$$\frac{\partial f}{\partial x}(0, 0) = 0 \quad \text{and} \quad \frac{\partial f}{\partial y}(0, 0) = 0.$$

Is f differentiable at $(0, 0)$? If it were, then by Proposition 9.2

$$\frac{\partial f}{\partial \hat{\mathbf{v}}}(0, 0) = \cos \theta \frac{\partial f}{\partial x}(0, 0) + \sin \theta \frac{\partial f}{\partial y}(0, 0),$$

which is not the case. Hence, we deduce that f is not differentiable at $(0, 0)$.



■

Comment 9.7 If the partial derivatives are continuous in a neighborhood of a point, then f is differentiable at that point.

9.6 Composition of multivariate functions with paths

Consider a function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ (representing, for example, temperature as a function of position). Let $\underline{x}: \mathbb{R} \rightarrow \mathbb{R}^n$ be a path in \mathbb{R}^n (representing, for example, the position of a fly as a function of time). If we compose these two functions,

$$f \circ \underline{x},$$

we get a function $\mathbb{R} \rightarrow \mathbb{R}$ (the temperature measured by the fly as function of time). Indeed,

$$t \mapsto \underline{x}(t) \mapsto f(\underline{x}(t)) \quad \text{is a mapping} \quad \mathbb{R} \rightarrow \mathbb{R}^n \rightarrow \mathbb{R}.$$

In component notation,

$$f \circ \underline{x}(t) = f(\underline{x}(t)) = f(x_1(t), x_2(t), \dots, x_n(t))$$

(recall that a path can be identified with n real-valued functions).

■ **Example 9.19** The trajectory of a fly in \mathbb{R}^2 is

$$\underline{x}(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} t \\ t^2 \end{pmatrix},$$

and the temperature as function of position on the plane is

$$f(x, y) = y \sin x + \cos x.$$

Then the temperature measured by the fly as function of time is

$$f \circ \underline{\mathbf{x}}(t) = f(\underline{\mathbf{x}}(t)) = t^2 \sin t + \cos t.$$

■

The question is the following: suppose that we know the derivative of $\underline{\mathbf{x}}$ at t (i.e., we know the velocity of the path) and we know the derivative of f at $\underline{\mathbf{x}}(t)$ (i.e., we know how the temperature changes in response to small changes in position at the current location). Can we deduce the derivative of $f \circ \underline{\mathbf{x}}$ at t (i.e., can we deduce the rate of change of the temperature measured by the fly)?

Let's try to guess the answer. For univariate functions the derivative of a composition is the product of the derivatives (the **chain rule**), so we would guess,

$$\frac{d(f \circ \underline{\mathbf{x}})}{dt}(t) = f'(\underline{\mathbf{x}}(t)) \dot{\underline{\mathbf{x}}}(t).$$

This expression is meaningless. The derivative of f is a vector and so is $\dot{\underline{\mathbf{x}}}(t)$. The left-hand side is a scalar, hence an educated guess would be:

Proposition 9.3 — Chain rule. Let $\underline{\mathbf{x}}: \mathbb{R} \rightarrow \mathbb{R}^n$ and $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be differentiable at t_0 and at $\underline{\mathbf{x}}(t_0)$, respectively. Then $f \circ \underline{\mathbf{x}}$ is differentiable at t_0 and

$$\frac{d(f \circ \underline{\mathbf{x}})}{dt}(t_0) = \nabla f(\underline{\mathbf{x}}(t_0)) \cdot \dot{\underline{\mathbf{x}}}(t_0).$$

■ **Example 9.20** Let's examine the above example, in which

$$\nabla f(\underline{\mathbf{x}}) = \begin{pmatrix} y \cos x - \sin x \\ \sin x \end{pmatrix} \quad \text{and} \quad \dot{\underline{\mathbf{x}}}(t) = \begin{pmatrix} 1 \\ 2t \end{pmatrix},$$

so that

$$\nabla f(\underline{\mathbf{x}}(t)) \cdot \dot{\underline{\mathbf{x}}}(t) = \begin{pmatrix} t^2 \cos t - \sin t \\ \sin t \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2t \end{pmatrix} = t^2 \cos t - \sin t + 2t \sin t.$$

On the other hand,

$$\frac{d(f \circ \underline{\mathbf{x}})}{dt}(t) = t^2 \cos t + 2t \sin t - \sin t,$$

i.e., Proposition 9.3 seems correct. ■

Proof. Since $\underline{\mathbf{x}}$ is differentiable at t_0 it follows that

$$\underline{\mathbf{x}}(t) = \underline{\mathbf{x}}(t_0) + \dot{\underline{\mathbf{x}}}(t_0)(t - t_0) + o(|t - t_0|),$$

which we may also write as

$$\Delta \mathbf{x} = \dot{\mathbf{x}}(t_0) \Delta t + o(|\Delta t|).$$

Since f is differentiable at $\mathbf{x}(t_0)$ it follows that

$$f(\mathbf{x}(t)) = f(\mathbf{x}(t_0)) + \nabla f(\mathbf{x}(t_0)) \cdot \Delta \mathbf{x} + o(|\Delta \mathbf{x}|).$$

Putting things together,

$$f(\mathbf{x}(t)) = f(\mathbf{x}(t_0)) + \nabla f(\mathbf{x}(t_0)) \cdot \dot{\mathbf{x}}(t_0) \Delta t + o(|\Delta t|),$$

which by definition, implies the desired result. ■

9.7 Differentiability and implicit functions

Recall that a function $x \mapsto f(x)$ may be defined implicitly via a relation of the form

$$G(x, f(x)) = 0,$$

where $G: \mathbb{R}^2 \rightarrow \mathbb{R}$. Another way to state it is that f is defined via its graph

$$\text{Graph}(f) = \{(x, y) \mid G(x, y) = 0\}.$$

Suppose that $(x_0, y_0) \in \text{Graph}(f)$, i.e., $y_0 = f(x_0)$. Consider now the directional derivatives of G along $\hat{\mathbf{v}} = (\cos \theta, \sin \theta)$ at (x_0, y_0) ,

$$\frac{\partial G}{\partial \hat{\mathbf{v}}}(x_0, y_0) = \hat{\mathbf{v}} \cdot \nabla G(x_0, y_0) = \cos \theta \frac{\partial G}{\partial x}(x_0, y_0) + \sin \theta \frac{\partial G}{\partial y}(x_0, y_0).$$

The line tangent to the graph of f at (x_0, y_0) is along the direction in which the directional derivative of G vanishes, i.e.,

$$\frac{\partial G}{\partial \hat{\mathbf{v}}}(x_0, y_0) = \cos \theta \left(\frac{\partial G}{\partial x}(x_0, y_0) + \tan \theta \frac{\partial G}{\partial y}(x_0, y_0) \right) = 0.$$

Thus,

$$f'(x_0) = \tan \theta = - \frac{\frac{\partial G}{\partial x}(x_0, y_0)}{\frac{\partial G}{\partial y}(x_0, y_0)}.$$

■ **Example 9.21** Let $G(x, y) = x^2 + y^2 - 16$ and $(x_0, y_0) = (2, \sqrt{12})$. Then,

$$\frac{\partial G}{\partial x}(2, \sqrt{12}) = 4 \quad \text{and} \quad \frac{\partial G}{\partial y}(2, \sqrt{12}) = 2\sqrt{12},$$

so that

$$f'(2) = - \frac{4}{2\sqrt{12}} = - \frac{1}{\sqrt{3}}.$$
■

9.8 Interpretation of the gradient

We saw that for every function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and unit vector $\hat{\mathbf{v}}$,

$$\frac{\partial f}{\partial \hat{\mathbf{v}}}(\mathbf{a}) = \hat{\mathbf{v}} \cdot \nabla f(\mathbf{a}).$$

$\nabla f(\mathbf{a})$ is a vector in \mathbb{R}^n . What is its direction? What is its magnitude.

If θ is the angle between the gradient of f at \mathbf{a} and $\hat{\mathbf{v}}$, then

$$\frac{\partial f}{\partial \hat{\mathbf{v}}}(\mathbf{a}) = |\nabla f(\mathbf{a})| \cos \theta.$$

The directional derivative is maximal when $\theta = 0$, which means that the gradient points to the direction in which the function changes the fastest. The magnitude of the gradient is this maximal directional derivative.

Likewise,

$$\frac{\partial f}{\partial \hat{\mathbf{v}}}(\mathbf{a}) = 0 \quad \text{when} \quad \nabla f(\mathbf{a}) \perp \hat{\mathbf{v}},$$

i.e., $\nabla f(\mathbf{a})$ is perpendicular to the contour line of f at \mathbf{a} .

■ **Example 9.22** Consider once again the distance function

$$f(\mathbf{x}) = |\mathbf{x}|.$$

We saw that

$$\nabla f(\mathbf{x}) = \hat{\mathbf{x}},$$

i.e., the gradient points along the radius vector (the direction in which the radius vector changes the fastest) and its magnitude is one, as

$$\frac{\partial f}{\partial \hat{\mathbf{x}}} = \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\hat{\mathbf{x}}) - f(\mathbf{x})}{h} = \lim_{h \rightarrow 0} \frac{|\mathbf{x} + h\hat{\mathbf{x}}| - |\mathbf{x}|}{h} = 1.$$

■

Definition 9.2 Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$. A point $\mathbf{a} \in \mathbb{R}^n$ at which

$$\nabla f(\mathbf{a}) = 0$$

is called a **stationary point**, or a **critical point** (נקודה קריטית).

What is the meaning of a point being stationary? That in any direction the pointwise rate of change of the function is zero. Like for univariate functions, stationary points of multivariate functions can be classified:

1. **Local maxima:** For every $\hat{\mathbf{v}} \in \mathbb{R}^n$, the function

$$g(t) = f(\mathbf{x}_0 + t\hat{\mathbf{v}})$$

has a local maximum at $t = 0$. Example, $f(x, y) = -x^2 - y^2$.

2. **Local minima:** For every $\hat{\mathbf{v}} \in \mathbb{R}^n$, the function

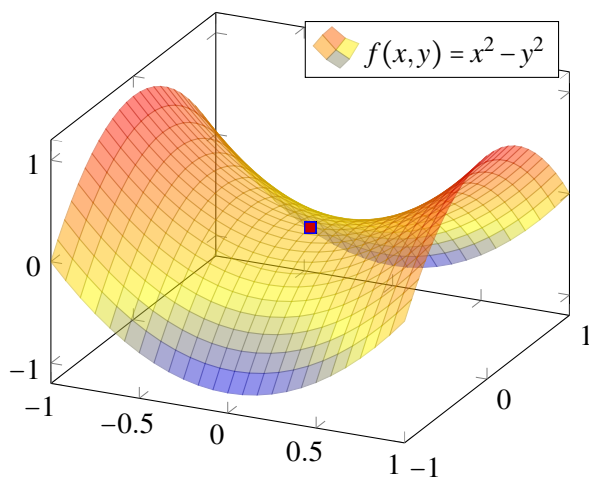
$$g(t) = f(\underline{\mathbf{x}}_0 + t\hat{\mathbf{v}})$$

has a local minimum at $t = 0$. Example, $f(x, y) = x^2 + y^2$.

3. **Saddle points** (נקודות אוכף): A stationary point that's neither a local minimum nor a local maximum. Typically,

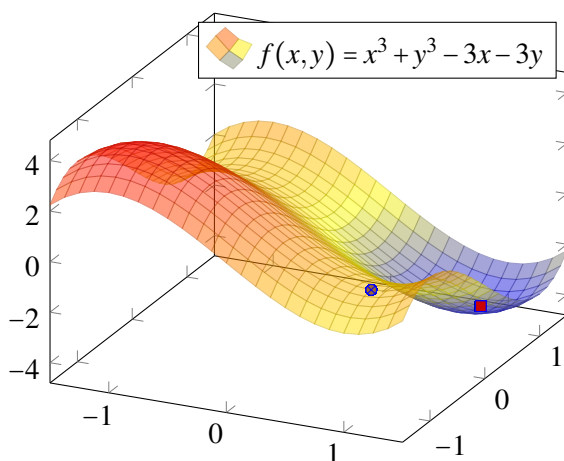
$$g(t) = f(\underline{\mathbf{x}}_0 + t\hat{\mathbf{v}})$$

may have a local minimum, a local maximum or an inflection point at $t = 0$, depending on the direction $\hat{\mathbf{v}}$. Example, $f(x, y) = x^2 - y^2$.



■ **Example 9.23** Consider the function

$$f(x, y) = x^3 + y^3 - 3x - 3y.$$



Its gradient is

$$\nabla f(x, y) = \begin{pmatrix} 3x^2 - 3 \\ 3y^2 - 3 \end{pmatrix},$$

hence its stationary points are $(\pm 1, \pm 1)$.

Take first the point $\underline{\mathbf{a}} = (1, 1)$. For $\hat{\mathbf{v}} = (\cos \theta, \sin \theta)$,

$$f(\underline{\mathbf{a}} + t\hat{\mathbf{v}}) = (1 + t \cos \theta)^3 + (1 + t \sin \theta)^3 - 3(1 + t \cos \theta) - 3(1 + t \sin \theta).$$

Expanding, we find

$$f(\underline{\mathbf{a}} + t\hat{\mathbf{v}}) = -4 + 3t^2 \cos^2 \theta + 3t^2 \sin^2 \theta + o(t^2) = -4 + 3t^2 + o(t^2),$$

hence in every direction $\hat{\mathbf{v}}$, $f(\underline{\mathbf{a}} + t\hat{\mathbf{v}})$ has a local minimum at $t = 0$. That is, $\underline{\mathbf{a}}$ is a local minimum of f .

Take next the point $\underline{\mathbf{b}} = (1, -1)$. For $\hat{\mathbf{v}} = (\cos \theta, \sin \theta)$,

$$f(\underline{\mathbf{b}} + t\hat{\mathbf{v}}) = (1 + t \cos \theta)^3 + (-1 + t \sin \theta)^3 - 3(1 + t \cos \theta) - 3(-1 + t \sin \theta).$$

Expanding, we find,

$$f(\underline{\mathbf{b}} + t\hat{\mathbf{v}}) = 3t^2 \cos^2 \theta - 3t^2 \sin^2 \theta + o(t^2),$$

For $\theta = 0$, $f(\underline{\mathbf{b}} + t\hat{\mathbf{v}})$ has a local minimum at $t = 0$, whereas for $\theta = \pi/2$, $f(\underline{\mathbf{b}} + t\hat{\mathbf{v}})$ has a local maximum at $t = 0$. Thus, $\underline{\mathbf{b}}$ is a saddle point of f . ■

9.9 Higher derivatives and multivariate Taylor theorem

Let $f: \mathbb{R}^2 \rightarrow \mathbb{R}$. If f is differentiable in a certain domain, then it has partial derivatives,

$$\frac{\partial f}{\partial x} \quad \text{and} \quad \frac{\partial f}{\partial y},$$

which both are also functions $\mathbb{R}^2 \rightarrow \mathbb{R}$ (recall the the gradient is a function $\mathbb{R}^2 \rightarrow \mathbb{R}^2$). If the partial derivatives are differentiable, then they have their own partial derivatives, which we denote by

$$\begin{aligned} \frac{\partial}{\partial x} \frac{\partial f}{\partial x} &= \frac{\partial^2 f}{\partial x^2} & \frac{\partial}{\partial y} \frac{\partial f}{\partial x} &= \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial}{\partial y} \frac{\partial f}{\partial y} &= \frac{\partial^2 f}{\partial y^2} & \frac{\partial}{\partial x} \frac{\partial f}{\partial y} &= \frac{\partial^2 f}{\partial y \partial x}. \end{aligned}$$

More generally, for functions $f: \mathbb{R}^n \rightarrow \mathbb{R}$, we have an n -by- n matrix of partial second derivatives (called the **Hessian**),

$$\frac{\partial^2 f}{\partial x_i \partial x_j} \quad i, j = 1, \dots, n.$$

Each such partial second derivative is a function $\mathbb{R}^n \rightarrow \mathbb{R}$, which may be differentiated along any direction. A function $\mathbb{R}^n \rightarrow \mathbb{R}$ that can be differentiated infinitely many times along any combination of directions is called **smooth** (הקלח).

Theorem 9.4 — Clairaut. If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ has continuous second partial derivatives then

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

That is, the matrix of second derivatives is symmetric.

Comment 9.8 This is not a trivial statement. One has to show that

$$\lim_{h \rightarrow 0} \frac{\frac{\partial f}{\partial x_i}(\mathbf{x} + h\hat{\mathbf{e}}_j) - \frac{\partial f}{\partial x_i}(\mathbf{x})}{h} = \lim_{h \rightarrow 0} \frac{\frac{\partial f}{\partial x_j}(\mathbf{x} + h\hat{\mathbf{e}}_i) - \frac{\partial f}{\partial x_j}(\mathbf{x})}{h}.$$

■ **Example 9.24** Consider the function,

$$f(x, y) = yx^3 + xy^4 - 3x - 3y.$$

Then,

$$\frac{\partial f}{\partial x} = 3x^2y + y^4 - 3 \quad \frac{\partial f}{\partial y} = x^3 + 4y^3x - 3,$$

and

$$\frac{\partial^2 f}{\partial x^2} = 6xy \quad \frac{\partial^2 f}{\partial y \partial x} = 3x^2 + 4y^3 \quad \frac{\partial^2 f}{\partial y^2} = 12y^2x \quad \frac{\partial^2 f}{\partial x \partial y} = 3x^2 + 4y^3.$$

We may calculate higher derivatives. For example,

$$\frac{\partial^3 f}{\partial x \partial y \partial x} = 6x.$$

■

Suppose we know $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ and its derivatives at a point (x_0, y_0) . What can we say about its values in the vicinity of that point, i.e., at a point $(x_0 + \Delta x, y_0 + \Delta y)$? It turns out that the concept of a Taylor polynomial can be generalized to multivariate function.

Definition 9.3 Let $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ be n -times differentiable at $\mathbf{x}_0 = (x_0, y_0)$. Then, its Taylor polynomial of degree n about the point \mathbf{x}_0 is

$$\begin{aligned} P_{f,n,\mathbf{x}_0}(\mathbf{x}) &= f(\mathbf{x}_0) \\ &+ \frac{\partial f}{\partial x}(\mathbf{x}_0)(x-x_0) + \frac{\partial f}{\partial y}(\mathbf{x}_0)(y-y_0) \\ &+ \frac{1}{2} \left(\frac{\partial^2 f}{\partial x^2}(\mathbf{x}_0)(x-x_0)^2 + 2 \frac{\partial^2 f}{\partial x \partial y}(\mathbf{x}_0)(x-x_0)(y-y_0) + \frac{\partial^2 f}{\partial y^2}(\mathbf{x}_0)(y-y_0)^2 \right) \\ &+ \cdots + \sum_{k=0}^n \binom{n}{k} \frac{\partial^n f}{\partial x^k \partial y^{n-k}}(\mathbf{x}_0)(x-x_0)^k (y-y_0)^{n-k}. \end{aligned}$$

Theorem 9.5 Let $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ be n -times differentiable at $\mathbf{x}_0 = (x_0, y_0)$. Then,

$$\lim_{|\mathbf{x}-\mathbf{x}_0| \rightarrow 0} \frac{f(\mathbf{x}) - P_{f,n,\mathbf{x}_0}(\mathbf{x})}{|\mathbf{x}-\mathbf{x}_0|^n} = 0,$$

or using the equivalent notation,

$$f(\mathbf{x}) = P_{f,n,\mathbf{x}_0}(\mathbf{x}) + o(|\mathbf{x}-\mathbf{x}_0|^n).$$

■ **Example 9.25** Calculate the Taylor polynomial of degree 3 of the function

$$f(x, y) = x^3 \ln y + xy^4$$

Then,

$$\begin{aligned} \frac{\partial f}{\partial x}(x, y) &= 3x^2 \ln y + y^4 & \frac{\partial f}{\partial y}(x, y) &= \frac{x^3}{y} + 4xy^3, \\ \frac{\partial^2 f}{\partial x^2}(x, y) &= 6x \ln y & \frac{\partial^2 f}{\partial x \partial y}(x, y) &= \frac{3x^2}{y} + 4y^3 & \frac{\partial^2 f}{\partial y^2}(x, y) &= -\frac{x^3}{y^2} + 12xy^2, \end{aligned}$$

and

$$\frac{\partial^3 f}{\partial x^3}(x, y) = 6 \ln y \quad \frac{\partial^3 f}{\partial x^2 \partial y}(x, y) = \frac{6x}{y} \quad \frac{\partial^3 f}{\partial x \partial y^2}(x, y) = -\frac{3x^2}{y^2} + 12y^2 \quad \frac{\partial^3 f}{\partial y^3}(x, y) = \frac{2x^3}{y^3} + 24xy.$$

At the point $(1, 1)$,

$$\begin{aligned} P_{f,3,(1,1)}(1 + \Delta x, 1 + \Delta y) &= 1 + (\Delta x + 5\Delta y) + \frac{1}{2} (14\Delta x \Delta y + 11\Delta y^2) \\ &+ \frac{1}{6} (18\Delta x^2 \Delta y + 27\Delta x \Delta y^2 + 26\Delta y^3). \end{aligned}$$

■

9.10 Classification of stationary points

Taylor's theorem for functions of two variables can be used to classify stationary points. If \mathbf{a} is a stationary point of $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ and f is twice differentiable at that point, then by Taylor's theorem

$$f(\mathbf{a} + \Delta \mathbf{x}) = f(\mathbf{a}) + \frac{1}{2} \left(\frac{\partial^2 f}{\partial x^2}(\mathbf{a}) \Delta x^2 + 2 \frac{\partial^2 f}{\partial x \partial y}(\mathbf{a}) \Delta x \Delta y + \frac{\partial^2 f}{\partial y^2}(\mathbf{a}) \Delta y^2 \right) + o(|\Delta \mathbf{x}|^2).$$

Near stationary points, $f - f(\mathbf{a})$ is dominated by the quadratic terms (unless they vanish, in which case we have to look at higher-order terms). To shorten notations let's write,

$$\frac{\partial^2 f}{\partial x^2}(\mathbf{a}) = A \quad \frac{\partial^2 f}{\partial x \partial y}(\mathbf{a}) = B \quad \text{and} \quad \frac{\partial^2 f}{\partial y^2}(\mathbf{a}) = C,$$

i.e.,

$$f(\mathbf{a} + \Delta \mathbf{x}) = f(\mathbf{a}) + \frac{1}{2} \underbrace{(A \Delta x^2 + 2B \Delta x \Delta y + C \Delta y^2)}_{M(\Delta \mathbf{x})} + o(|\Delta \mathbf{x}|^2).$$

Consider the quadratic term $M(\Delta \mathbf{x})$:

1. If it is **positive** for all $\Delta \mathbf{x} \neq \mathbf{0}$, then \mathbf{a} is a local **minimum**.
2. If it is **negative** for all $\Delta \mathbf{x} \neq \mathbf{0}$, then \mathbf{a} is a local **maximum**.
3. If it **changes sign** for different $\Delta \mathbf{x}$, then \mathbf{a} is a **saddle point**.

We now characterize under what conditions on A, B, C each case occurs:

1. $M(\Delta \mathbf{x}) > 0$ for all $\Delta \mathbf{x} \neq \mathbf{0}$ if $A, C > 0$, i.e., if

$$\frac{\partial^2 f}{\partial x^2}(\mathbf{a}), \frac{\partial^2 f}{\partial y^2}(\mathbf{a}) > 0.$$

This is still not sufficient. Writing,

$$M(\mathbf{x}) = \left(\sqrt{A} \Delta x + \frac{B}{\sqrt{A}} \Delta y \right)^2 + \left(C - \frac{B^2}{A} \right) \Delta y^2,$$

we obtain that $AC > B^2$, i.e.,

$$\frac{\partial^2 f}{\partial x^2}(\mathbf{a}) \frac{\partial^2 f}{\partial y^2}(\mathbf{a}) > \left(\frac{\partial^2 f}{\partial x \partial y}(\mathbf{a}) \right)^2.$$

2. $M(\Delta \mathbf{x}) < 0$ for all $\Delta \mathbf{x} \neq \mathbf{0}$ if $A, C < 0$, i.e., if

$$\frac{\partial^2 f}{\partial x^2}(\mathbf{a}), \frac{\partial^2 f}{\partial y^2}(\mathbf{a}) < 0.$$

This is still not sufficient. Writing,

$$M(\mathbf{x}) = -(|A|\Delta x^2 - 2B\Delta x\Delta y + |C|\Delta y^2) = -\left(\sqrt{|A|}\Delta x - \frac{B}{\sqrt{|A|}}\Delta y\right)^2 - \left(\frac{B^2}{|A|} - |C|\right)\Delta y^2,$$

we obtain once again that $AC > B^2$, i.e.,

$$\frac{\partial^2 f}{\partial x^2}(\mathbf{a}) \frac{\partial^2 f}{\partial y^2}(\mathbf{a}) > \left(\frac{\partial^2 f}{\partial x \partial y}(\mathbf{a})\right)^2.$$

3. A saddle point occurs if

$$\frac{\partial^2 f}{\partial x^2}(\mathbf{a}) \frac{\partial^2 f}{\partial y^2}(\mathbf{a}) < \left(\frac{\partial^2 f}{\partial x \partial y}(\mathbf{a})\right)^2.$$

We can see it by setting $\Delta x = 1$, in which case

$$M(1, \Delta y) = A + 2B\Delta y + C\Delta y^2,$$

which changes sign if the discriminant is positive.

To conclude:

Type	Conditions		
Local minimum	$\frac{\partial^2 f}{\partial x^2} > 0$	$\frac{\partial^2 f}{\partial y^2} > 0$	$\frac{\partial^2 f}{\partial x^2} \cdot \frac{\partial^2 f}{\partial y^2} > \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2$
Local maximum	$\frac{\partial^2 f}{\partial x^2} < 0$	$\frac{\partial^2 f}{\partial y^2} < 0$	$\frac{\partial^2 f}{\partial x^2} \cdot \frac{\partial^2 f}{\partial y^2} > \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2$
Saddle	$\frac{\partial^2 f}{\partial x^2} \cdot \frac{\partial^2 f}{\partial y^2} < \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2$		

■ **Example 9.26** Analyze the stationary points $(\pm 1/\sqrt{3}, 0)$ of the function

$$f(x, y) = x^3 - x - y^2.$$

■

9.11 Extremal values in a compact domain

Definition 9.4 A domain $D \subset \mathbb{R}^n$ is called **bounded** if there exists a number R such that

$$\forall \underline{\mathbf{x}} \in D \quad |\underline{\mathbf{x}}| < R.$$

Comment 9.9 In other words, a domain is bounded if it can be enclosed in a ball.

Definition 9.5 A domain $D \subset \mathbb{R}^n$ is called **open** if any point $\underline{\mathbf{x}} \in D$ has a neighborhood

$$\{\underline{\mathbf{y}} \in \mathbb{R}^n \mid |\underline{\mathbf{y}} - \underline{\mathbf{x}}| < \delta\}$$

contained in D . It is called **closed** if its complement is open, i.e., if any point $\underline{\mathbf{x}} \notin D$ has a neighborhood

$$\{\underline{\mathbf{y}} \in \mathbb{R}^n \mid |\underline{\mathbf{y}} - \underline{\mathbf{x}}| < \delta\}$$

disjoint from D . This means that if $\underline{\mathbf{x}} \in \mathbb{R}^n$ is a point that has the property that every ball around $\underline{\mathbf{x}}$ intersects D , then $\underline{\mathbf{x}} \in D$.

In this section we discuss extremal points of continuous functions defined on a bounded and closed domain $D \subset \mathbb{R}^n$. Such domains are called **compact** (קומפקט). Why are closedness and boundedness important? Recall that every for univariate function, a function continuous may fail to have extrema if its domain is not closed or unbounded. Like for univariate functions we have the following results:

Theorem 9.6 Let $f: D \rightarrow \mathbb{R}$ be continuous with $D \subset \mathbb{R}^n$ compact. Then f assumes in D both minima and maxima, i.e., where exist $\underline{\mathbf{a}}, \underline{\mathbf{b}} \in D$ such that

$$\forall \underline{\mathbf{x}} \in D \quad f(\underline{\mathbf{a}}) \leq f(\underline{\mathbf{x}}) \leq f(\underline{\mathbf{b}}).$$

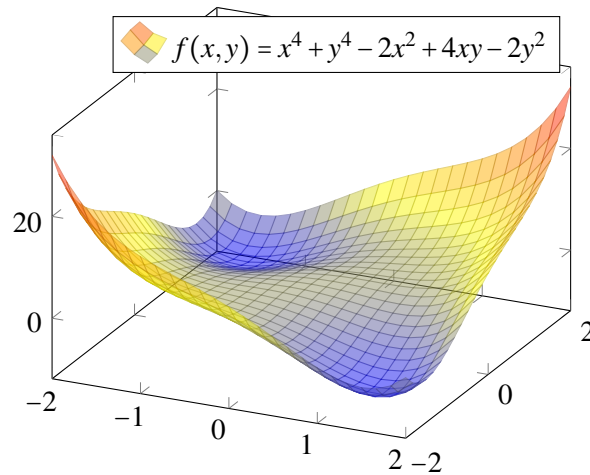
Furthermore, if f assumes a local extremum at an internal point of D and f is differentiable at that point then its gradient at that point vanishes.

■ **Example 9.27** Consider the function

$$f(x, y) = x^4 + y^4 - 2x^2 + 4xy - 2y^2$$

in the domain

$$D = \{(x, y) \mid -2 \leq x, y \leq 2\}.$$



We start by looking for stationary points inside the domain. The gradient is

$$\nabla f(x, y) = \begin{pmatrix} 4x^3 - 4x + 4y \\ 4y^3 + 4x - 4y \end{pmatrix},$$

which vanishes for

$$y = x - x^3 \quad \text{and} \quad x = y - y^3,$$

i.e.,

$$y = y - y^3 - (y - y^3)^3,$$

from gives $y^3 = y^3(1 - y^2)^3$, i.e., $y = 0$ or $y = \pm\sqrt{2}$, i.e., the stationary points are

$$\mathbf{a} = (0, 0) \quad \mathbf{b} = (-\sqrt{2}, \sqrt{2}) \quad \text{and} \quad \mathbf{c} = (\sqrt{2}, -\sqrt{2}).$$

To determine the types of the stationary points we calculate the Hessian,

$$\frac{\partial^2 f}{\partial x^2}(x, y) = 12x^2 - 4 \quad \frac{\partial^2 f}{\partial x \partial y}(x, y) = 4 \quad \text{and} \quad \frac{\partial^2 f}{\partial y^2}(x, y) = 12y^2 - 4.$$

I.e.,

$$H(\mathbf{a}) = \begin{pmatrix} 0 & 4 \\ 4 & 0 \end{pmatrix} \quad H(\mathbf{b}) = H(\mathbf{c}) = \begin{pmatrix} 20 & 4 \\ 4 & 20 \end{pmatrix},$$

from which we deduce that \mathbf{a} is a saddle point and \mathbf{b} and \mathbf{c} are local minima, with

$$f(\mathbf{b}) = f(\mathbf{c}) = -8.$$

We now turn to consider the boundaries, which comprise of four segments. Since f is symmetric,

$$f(x, y) = f(y, x) = f(-x, -y),$$

it suffices to check one segment, say, the segment $\{-2\} \times [-2, 2]$, where f takes the form

$$y \mapsto 16 + y^4 - 8 - 8y - 2y^2.$$

This is a univariate function whose derivative is

$$y \mapsto 4y^3 - 8 - 4y.$$

The derivative vanishes at $y \approx 1.5214$. The second derivative is

$$y \mapsto 12y^2 - 2.$$

Since it is positive at that point, the point is a local minimum. The function at that point equals approximately 20.9.

Finally, we need to verify the values of f at the four corners:

$$f(2, 2) = f(-2, -2) = 32 \quad \text{and} \quad f(2, -2) = f(-2, 2) = 0.$$

To conclude f assumes its minimal value, (-8) at $(-\sqrt{2}, \sqrt{2})$ and $(\sqrt{2}, -\sqrt{2})$, and its maximal value, 32, at $(2, 2)$ and $(-2, -2)$. ■

9.12 Extrema in the presence of constraints

9.12.1 Problem statement

The general problem: Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and let also

$$g_1, g_2, \dots, g_m: \mathbb{R}^n \rightarrow \mathbb{R}.$$

We are looking for a point $\underline{\mathbf{x}} \in \mathbb{R}^n$ that minimizes or maximizes f under the constraint that

$$g_1(\underline{\mathbf{x}}) = g_2(\underline{\mathbf{x}}) = \dots = g_m(\underline{\mathbf{x}}) = 0.$$

That is, the set of interest is

$$\{\underline{\mathbf{x}} \in \mathbb{R}^n \mid g_1(\underline{\mathbf{x}}) = g_2(\underline{\mathbf{x}}) = \dots = g_m(\underline{\mathbf{x}}) = 0\}.$$

Without the constraints, we know that the extremal point must be a critical point of $\underline{\mathbf{x}}$. In the presence of constraints, it may be that the extremal point is not stationary since the directions in which it changes violate the constraints.

■ **Example 9.28** Find the extremal points of

$$f(x, y) = 2x^2 + 4y^2$$

under the constraint

$$g(x, y) = x^2 + y^2 - 1 = 0.$$

Note that f by itself does not have a maximum in \mathbb{R}^n since it grows unbounded as $|\underline{\mathbf{x}}| \rightarrow \infty$. However, under the constraint that $|\underline{\mathbf{x}}| = 1$ it has a maximum. ■

9.12.2 A single constraint

Let's start with a single constraint. We are looking for extremal points of $f: \mathbb{R}^n \rightarrow \mathbb{R}$ under the constraint

$$g(\underline{\mathbf{x}}) = 0.$$

One approach could be the following: since the constraint is of the form

$$g(x_1, \dots, x_n) = 0,$$

invert it to get a relation,

$$x_n = h(x_1, \dots, x_{n-1}),$$

substitute it in f to get a new function

$$F(x_1, \dots, x_{n-1}) = f(x_1, \dots, x_{n-1}, h(x_1, \dots, x_{n-1})),$$

which we then minimize without having to deal with constraints. The problem with this approach is that it is often impractical to invert the constraint. Moreover, the constraint may fail to define an implicit function (like in Example 9.28, where the constraint is an ellipse).

The constrained optimization problem can also be approached as follows: the set

$$D = \{\underline{\mathbf{x}} \in \mathbb{R}^n \mid g(\underline{\mathbf{x}}) = 0\}$$

over which we extremize f is in general an $(n-1)$ -dimensional hyper-surface (a curve in Example 9.28). The gradient of g in D points to the direction in which g changes the fastest. The plane perpendicular to this direction spans the directions along which g remains constant. f has an extremal point in D if the gradient of f is perpendicular to the hyperplane along which g remains constant. In other words, f has an extremal point at $\underline{\mathbf{x}} \in D$ if the gradient of f is parallel to the gradient of g , i.e., there exists a scalar λ such that

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

A equivalent explanation is the following: let $\underline{\mathbf{x}}(t)$ be a general path in \mathbb{R}^n satisfying

$$\underline{\mathbf{x}}(0) = \underline{\mathbf{x}}_0.$$

The paths along which g does not change satisfy

$$g(\underline{\mathbf{x}}(t)) = 0.$$

For all such path,

$$(g \circ \underline{\mathbf{x}})'(0) = \underline{\nabla} g(\underline{\mathbf{x}}(t)) \cdot \underline{\mathbf{x}}'(0) = 0.$$

$\underline{\mathbf{x}}_0$ is a solution to our problem if for all such paths,

$$f(\underline{\mathbf{x}}(t)) \quad \text{has an extremum at } t = 0,$$

i.e.,

$$(f \circ \underline{\mathbf{x}})'(0) = \nabla f(\underline{\mathbf{x}}_0) \cdot \underline{\mathbf{x}}'(0) = 0.$$

This condition holds if $\nabla f(\underline{\mathbf{x}}_0)$ and $\nabla g(\underline{\mathbf{x}}_0)$ are parallel.

Comment 9.10 Note that $\nabla f(\underline{\mathbf{x}}_0)$ and $\nabla g(\underline{\mathbf{x}}_0)$ are parallel if and only if there exists a number λ such that $\underline{\mathbf{x}}_0$ is a stationary point of the function

$$\Phi(\underline{\mathbf{x}}) = f(\underline{\mathbf{x}}) - \lambda g(\underline{\mathbf{x}}).$$

In this context the number λ is called a **Lagrange multiplier** (כופל לגרנז).

■ **Example 9.29** Let's return to Example 9.28. Then, $\underline{\mathbf{x}} = (x, y)$ is an extremal point if there exists a constant λ such that $\underline{\mathbf{x}}$ is a critical point of

$$\Phi(x, y) = (2x^2 + 4y^2) - \lambda (x^2 + y^2 - 1)$$

Note,

$$\nabla \Phi(x, y) = \begin{pmatrix} 4x - 2\lambda x \\ 8y - 2\lambda y \end{pmatrix},$$

i.e., either $x = 0$ (in which case $y = \pm 1$) and $\lambda = 4$ or $y = 0$ (in which case $x = \pm 1$) and $\lambda = 2$. To find which point is a minimum/maximum we have to substitute in f . ■

9.12.3 Multiple constraints

For simplicity, let's assume that there are just two constraints,

$$D = \{\underline{\mathbf{x}} \in \mathbb{R}^n \mid g_1(\underline{\mathbf{x}}) = g_2(\underline{\mathbf{x}}) = 0\}.$$

let $\underline{\mathbf{x}}(t)$ be a general path in \mathbb{R}^n satisfying

$$\underline{\mathbf{x}}(0) = \underline{\mathbf{x}}_0.$$

The paths along which g_1 and g_2 do not change satisfy

$$g_1(\underline{\mathbf{x}}(t)) = g_2(\underline{\mathbf{x}}(t)) = 0.$$

For all such path,

$$(g_1 \circ \underline{\mathbf{x}})'(0) = \nabla g_1(\underline{\mathbf{x}}(t)) \cdot \underline{\mathbf{x}}'(0) = 0$$

$$(g_2 \circ \underline{\mathbf{x}})'(0) = \nabla g_2(\underline{\mathbf{x}}(t)) \cdot \underline{\mathbf{x}}'(0) = 0.$$

$\underline{\mathbf{x}}_0$ is a solution to our problem if for all such paths,

$$f(\underline{\mathbf{x}}(t)) \quad \text{has an extremum at } t = 0,$$

i.e.,

$$(f \circ \underline{\mathbf{x}})'(0) = \nabla f(\underline{\mathbf{x}}_0) \cdot \underline{\mathbf{x}}'(0) = 0.$$

This condition holds if $\nabla f(\underline{\mathbf{x}}_0)$ is any linear combination of $\nabla g_1(\underline{\mathbf{x}}_0)$ and $\nabla g_2(\underline{\mathbf{x}}_0)$, i.e., there exists two constants, λ and μ such that

$$\nabla f(\underline{\mathbf{x}}_0) = \lambda \nabla g_1(\underline{\mathbf{x}}_0) + \mu \nabla g_2(\underline{\mathbf{x}}_0),$$

or equivalently, $\underline{\mathbf{x}}_0$ is a stationary point of the function

$$\Phi(\underline{\mathbf{x}}) = f(\underline{\mathbf{x}}) - \lambda g_1(\underline{\mathbf{x}}) - \mu g_2(\underline{\mathbf{x}}).$$

■ **Example 9.30** Find a minimum point of

$$f(x, y, z) = x^2 + 2y^2 + 3z^2$$

under the constraints

$$g_1(x, y, z) = x + y + z - 1 = 0$$

$$g_2(x, y, z) = x^2 + y^2 + z^2 - 1 = 0.$$

For the function,

$$\Phi(x, y, z) = (x^2 + 2y^2 + 3z^2) - \lambda(x + y + z - 1) - \mu(x^2 + y^2 + z^2 - 1).$$

Then,

$$\nabla \Phi(x, y, z) = \begin{pmatrix} 2x - \lambda - 2\mu x \\ 4y - \lambda - 2\mu y \\ 6z - \lambda - 2\mu z \end{pmatrix}.$$

The condition that $\underline{\mathbf{x}}$ be a stationary point of Φ imposes three conditions on five unknowns. The constraints provide the two missing conditions. We may first get rid of $z = 1 - x - y$, i.e.,

$$2(1 - \mu)x = \lambda \quad 2(2 - \mu)y = \lambda \quad 2(3 - \mu)(1 - x - y) = \lambda$$

and

$$x^2 + y^2 + (1 - x - y)^2 = 1.$$

We can then get rid of λ ,

$$(1 - \mu)x = (2 - \mu)y \quad (3 - \mu)(1 - x - y) = (1 - \mu)x \quad \text{and} \quad x^2 + y^2 + (1 - x - y)^2 = 1.$$

Finally, we may eliminate μ ,

$$\mu = \frac{x - 2y}{x - y},$$

and remain with two equations for two unknowns. ■