CHAPTER 3

Inverse and Implicit functions

1. Inverse Functions and Coordinate Changes

Let $U \subseteq \mathbb{R}^d$ be a domain.

THEOREM 1.1 (Inverse function theorem). *If $\varphi : U \to \mathbb{R}^d$ is differentiable at a and $D\varphi_a$ is invertible, then there exists a domains U', V' such that $a \in U' \subseteq U$, $\varphi(a) \in V'$ and $\varphi : U' \to V'$ is bijective. Further, the inverse function $\psi : V' \to U'$ is differentiable.*

The proof requires compactness and is beyond the scope of this course.

REMARK 1.2. The condition $D\varphi_a$ is necessary: If φ has a differentiable inverse in a neighbourhood of a, then $D\varphi_a$ must be invertible. (Proof: Chain rule.)

This is often used to ensure the existence (and differentiability of local coordinates).

DEFINITION 1.3. A function $\varphi : U \to V$ is called a (differentiable) coordinate change if φ is differentiable and bijective and $D\varphi$ is invertible at every point.

Practically, let φ be a coordinate change function, and set $(u, v) = \varphi(x, y)$. Let $\psi = \varphi^{-1}$, and we write $(x, y) = \psi(u, v)$. Given a function $f : U \to \mathbb{R}$, we treat it as a function of x and y. Now using ψ, we treat (x, y) as functions of (u, v).

Thus we can treat f as a function of u and v, and it is often useful to compute $\partial_u f$ etc. in terms of $\partial_x f$ and $\partial_y f$ and the coordinate change functions. By the chain rule:

$$\partial_u f = \partial_x f \partial_u x + \partial_y f \partial_u y,$$

and we compute $\partial_u x$, $\partial_u y$ etc. either by directly finding the inverse function and expressing x, y in terms of u, v; or implicitly using the chain rule:

$$I = D\psi_\varphi \, D\varphi = \begin{pmatrix} \partial_u x & \partial_v x \\ \partial_u y & \partial_v y \end{pmatrix} \begin{pmatrix} \partial_x u & \partial_y u \\ \partial_x v & \partial_y v \end{pmatrix} \implies \begin{pmatrix} \partial_u x & \partial_v x \\ \partial_u y & \partial_v y \end{pmatrix} = \begin{pmatrix} \partial_x u & \partial_y u \\ \partial_x v & \partial_y v \end{pmatrix}^{-1}.$$

EXAMPLE 1.4. Let $u(x, y) = x^2 - y^2$ and $v(x, y) = 2xy$. Let $\varphi(x, y) = (u, v)$. For any $a \neq 0 \in \mathbb{R}^2$, there exists a small neighbourhood of a in which φ has a differentiable inverse.

The above tells us that *locally* x, y can be expressed as functions of u, v. This might not be true globally. In the above case we can explicitly solve and find x, y:

$$x = \Big(\frac{\sqrt{u^2 + v^2} + u}{2}\Big)^{1/2} \quad \text{and} \quad y = \Big(\frac{\sqrt{u^2 + v^2} - u}{2}\Big)^{1/2} \tag{1.1}$$

is one solution. (Negating both, gives another solution.)

Regardless, even without using the formulae, we can implicitly differentiate and find $\partial_u x$. Consequently,

$$\begin{pmatrix} \partial_u x & \partial_v x \\ \partial_u y & \partial_v y \end{pmatrix} = \begin{pmatrix} \partial_x u & \partial_y u \\ \partial_x v & \partial_y v \end{pmatrix}^{-1} = \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix}^{-1} = \frac{1}{2(x^2 + y^2)} \begin{pmatrix} x & y \\ -y & x \end{pmatrix}.$$

It is instructive to differentiate (1.1) directly and double check that the answers match.

Polar coordinates is another example, and has been done extensively your homework.

2. Implicit functions

Let $U \subseteq \mathbb{R}^{d+1}$ be a domain and $f : U \to \mathbb{R}$ be a differentiable function. If $x \in \mathbb{R}^d$ and $y \in \mathbb{R}$, we'll concatenate the two vectors and write $(x, y) \in \mathbb{R}^{d+1}$.

THEOREM 2.1 (Implicit function theorem). *Suppose $c = f(a, b)$ and $\partial_y f(a, b) \neq 0$. Then, there exists a domain $U' \ni a$ and differentiable function $g : U' \to \mathbb{R}$ such that $g(a) = b$ and $f(x, g(x)) = c$ for all $x \in U'$. Further, there exists a domain $V' \ni b$ such that $\{(x, y) \mid x \in U', y \in V', f(x, y) = c\} = \{(x, g(x)) \mid x \in U'\}$. (In other words, for all $x \in U'$ the equation $f(x, y) = c$ has a unique solution in V' and is given by $y = g(x)$.)*

REMARK 2.2. To see why $\partial_y f \neq 0$ is needed, let $f(x, y) = \alpha x + \beta y$ and consider the equation $f(x, y) = c$. To express y as a function of x we need $\beta \neq 0$ which in this case is equivalent to $\partial_y f \neq 0$.

REMARK 2.3. If $d = 1$, one expects $f(x, y) = c$ to some curve in $\mathbb{R}^2$. To write this curve in the form $y = g(x)$ using a differentiable function g, one needs the curve to never be vertical. Since ∇f is perpendicular to the curve, this translates to ∇f never being horizontal, or equivalently $\partial_y f \neq 0$ as assumed in the theorem.

REMARK 2.4. For simplicity we chose y to be the last coordinate above. It could have been any other, just as long as the corresponding partial was non-zero. Namely if $\partial_i f(a) \neq 0$, then one can locally solve the equation $f(x) = f(a)$ (uniquely) for the variable x_i and express it as a differentiable function of the remaining variables.

EXAMPLE 2.5. $f(x, y) = x^2 + y^2$ with $c = 1$.

PROOF OF THE IMPLICIT FUNCTION THEOREM. Let $\varphi(x, y) = (x, f(x, y))$, and observe $D\varphi_{(a,b)} \neq 0$. By the inverse function theorem φ has a unique local inverse ψ. Note ψ must be of the form $\psi(x, y) = (x, g(x, y))$. Also $\varphi \circ \psi = \text{Id}$ implies $(x, y) = \varphi(x, g(x, y)) = (x, f(x, g(x, y)))$. Hence $y = g(x, c)$ uniquely solves $f(x, y) = c$ in a small neighborhood of (a, b). ☐

Instead of $y \in \mathbb{R}$ above, we could have been fancier and allowed $y \in \mathbb{R}^n$. In this case f needs to be an $\mathbb{R}^n$ valued function, and we need to replace $\partial_y f \neq 0$ with the assumption that the $n \times n$ minor in Df (corresponding to the coordinate positions of y) is invertible. This is the general version of the implicit function theorem.

THEOREM 2.6 (Implicit function theorem, general case). *Let $U \subseteq \mathbb{R}^{d+n}$ be a domain, $f : \mathbb{R}^{d+n} \to \mathbb{R}^n$ be a differentiable function, $a \in U$ and M be the $n \times n$ matrix obtained by taking the i_1^{th}, i_2^{th}, $\dots i_n^{th}$ columns from Df_a. If M is invertible, then one can locally solve the equation $f(x) = f(a)$ (uniquely) for the variables x_{i_1}, ..., x_{i_n} and express them as a differentiable function of the remaining d variables.*

To avoid too many technicalities, we only state a more precise version of the above in the special case where the $n \times n$ matrix obtained by taking the last n columns of Df is invertible. Let's use the notation $(x, y) \in \mathbb{R}^{d+n}$ when $x \in \mathbb{R}^d$ and $y \in \mathbb{R}^n$. Now the precise statement is almost identical to Theorem 2.1:

THEOREM 2.7 (Implicit function theorem, precise statement in a special case). *Suppose $c = f(a, b)$ and the $n \times n$ matrix obtained by taking the last n columns of $Df_{a,b}$ is invertible. Then, there exists a domain $U' \subseteq \mathbb{R}^d$ containing a and differentiable function $g : U' \to \mathbb{R}^n$ such that $g(a) = b$ and $f(x, g(x)) = c$ for all $x \in U'$. Further, there exists a domain $V' \subseteq \mathbb{R}^n$ containing b such that $\{(x, y) \mid x \in U', y \in V', f(x, y) = c\} = \{(x, g(x)) \mid x \in U'\}$. (In other words, for all $x \in U'$ the equation $f(x, y) = c$ has a unique solution in V' and is given by $y = g(x)$.)*

EXAMPLE 2.8. Consider the equations

$$(x-1)^2 + y^2 + z^2 = 5 \quad \text{and} \quad (x+1)^2 + y^2 + z^2 = 5$$

for which $x = 0$, $y = 0$, $z = 2$ is one solution. For all other solutions close enough to this point, determine which of variables x, y, z can be expressed as differentiable functions of the others.

SOLUTION. Let $a = (0, 0, 1)$ and

$$F(x, y, z) = \begin{pmatrix} (x-1)^2 + y^2 + z^2 \\ (x+1)^2 + y^2 + z^2 \end{pmatrix}$$

Observe

$$DF_a = \begin{pmatrix} -2 & 0 & 4 \\ 2 & 0 & 4 \end{pmatrix},$$

and the 2×2 minor using the first and last column is invertible. By the implicit function theorem this means that in a small neighbourhood of a, x and z can be (uniquely) expressed in terms of y. □

REMARK 2.9. In the above example, one can of course solve explicitly and obtain

$$x = 0 \quad \text{and} \quad z = \sqrt{4 - y^2},$$

but in general we won't get so lucky.

3. Tangent planes and spaces

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be differentiable, and consider the implicitly defined curve $\Gamma = \{(x, y) \in \mathbb{R}^2 \mid f(x, y) = c\}$. (Note this is some level set of f.) Pick $(a, b) \in \Gamma$, and suppose $\partial_y f(a, b) \neq 0$. By the implicit function theorem, we know that the y-coordinate of this curve can locally be expressed as a differentiable function of x. In this case the tangent line through (a, b) has slope $\frac{dy}{dx}$.

Directly differentiating $f(x, y) = c$ with respect to x (and treating y as a function of x) gives

$$\partial_x f + \partial_y f \frac{dy}{dx} = 0 \iff \frac{dy}{dx} = \frac{-\partial_x f(a, b)}{\partial_y f(a, b)}.$$

Further, note that the normal vector at the point (a, b) has direction $(-\frac{dy}{dx}, 1)$. Substituting for $\frac{dy}{dx}$ using the above shows that the normal vector is parallel to ∇f.

REMARK 3.1. Geometrically, this means that ∇f is *perpendicular* to level sets of f. This is the direction along which f is changing "the most". (Consequently, the directional derivative of f along directions tangent to level sets is 0.)

The same is true in higher dimensions, which we study next. Consider the surface $z = f(x, y)$, and a point (x_0, y_0, z_0) on this surface. Projecting it to the x-z plane, this becomes the curve $z = f(x, y_0)$ which has slope $\partial_x f$. Projecting it onto the y-z plane, this becomes the curve with slope $\partial_y f$. The *tangent plane* at the point (x_0, y_0, z_0) is defined to be the unique plane passing through (x_0, y_0, z_0) which projects to a line with slope $\partial_x f(x_0, y_0)$ in the x-z plane and projects to a line with slope $\partial_y f(x_0, y_0)$ in the y-z plane. Explicitly, the equation for the tangent plane is

$$z - z_0 = (x - x_0)\partial_x f(x_0, y_0) + (y - y_0)\partial_y f(x_0, y_0).$$

REMARK 3.2. Consider a curve Γ in $\mathbb{R}^2$ and $a \in \Gamma$. The usual scenario is that Γ "touches" the tangent line at a and the continues (briefly) on the same side of the tangent line. The exception is of course inflection points, where Γ passes through its tangent line. In a generic curve, inflection points are usually isolated and this doesn't happen too often.

In $2D$ however, the picture is quite different. A surface will "touch" and locally stay on the same side of the tangent plane if the Hessian is either positive definite or negative definite. If the Hessian has both a strictly positive and a strictly negative eigenvalue, then the curve will necessarily "pass through" the tangent plane at the point of contact. Further, it is possible to construct surfaces where this happens at *every single point.* One such example is the surface $z = x^2 - y^2$.

DEFINITION 3.3. The *tangent space* to the surface $z = f(x, y)$ at the point (x_0, y_0, z_0) is defined to be the subspace

$$T = \left\{ (x, y, z) \in \mathbb{R}^3 \;\middle|\; z = x\partial_x f(x_0, y_0) + y\partial_x f(x_0, y_0) = Df_{(x_0, y_0)} \begin{pmatrix} x \\ y \end{pmatrix} \right\}$$

Elements of the tangent space are said to be *tangent vectors* at the point (x_0, y_0, z_0).

REMARK 3.4. The tangent space is *parallel* to the tangent plane, but shifted so that is passes through the origin (and hence is also a vector subspace).

REMARK 3.5. Clearly the vector $(\partial_x f, \partial_y f, -1)$ is *normal* to the tangent space of the surface $z = f(x, y)$.

Now let $g : \mathbb{R}^3 \to \mathbb{R}$ be differentiable, $c \in \mathbb{R}$ and consider the *implicitly defined* surface $\Sigma = \{(x, y, z) \mid g(x, y, z) = c\}$. Note again, this is a level set of g. Suppose (x_0, y_0, z_0) is a point on this surface and $\partial_z g(x_0, y_0, z_0) \neq 0$. Then using the implicit function theorem, the z-coordinate of this surface can locally be expressed as a differentiable function of x and y (say $z = f(x, y)$). In terms of f we know how to compute the tangent plane and space of Σ. Our aim is to write this directly in terms of g.

PROPOSITION 3.6. *Let* $a = (x_0, y_0, z_0) \in \Sigma$.

- *The* tangent space *at* a *(denoted by* $T\Sigma_a$*) is exactly* $\ker(Dg_a)$.
- *The* tangent plane *at* a *is* $\{x \in \mathbb{R}^3 \mid Dg_a(x - a) = 0\}$.

Recall elements of the tangent space are called *tangent vectors*. If $v \in T\Sigma_a$ then $Dg_a(v) = 0$, and hence the directional derivative of g in the direction v must be 0. Note further ∇g is *normal* to the surface Σ. Both these statements were made earlier, but not explored in detail as we didn't have the implicit function theorem at our disposal.

PROOF OF PROPOSITION 3.6. Substituting $z = f(x, y)$ in $g(x, y, z) = c$ and differentiating with respect to x and y gives

$$\partial_x g + \partial_z g \partial_x f = 0 \quad \text{and} \quad \partial_y g + \partial_z g \partial_y f = 0$$

Thus the tangent plane to the surface $g(x, y, z) = c$ at the point (x_0, y_0, z_0) is given by

$$z - z_0 = Df_{(x_0,y_0)} \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} \iff Dg_{(x_0,y_0,z_0)} \begin{pmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{pmatrix} = 0$$

The tangent space is given by

$$T = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \;\middle|\; \begin{pmatrix} x \\ y \\ z \end{pmatrix} \cdot \nabla g_{(x_0,y_0,z_0)} = 0 \right\}. \qquad \square$$

These generalizes in higher dimensions. Without being too precise about the definitions, here is the bottom line:

PROPOSITION 3.7. *Let* $g : \mathbb{R}^{n+d} \to \mathbb{R}^n$ *be a differentiable function,* $c \in \mathbb{R}^n$ *and let* $M = \{x \in \mathbb{R}^{n+d} \mid g(x) = c\}$. *Suppose the implicit function theorem applies at all points in* M. *Then* M *is a* d*-dimensional "surface" (called a* d*-dimensional manifold). At any point* $a \in M$*, the tangent space is exactly* $\ker Dg_a$. *Consequently,* $D_v g(a) = 0$ *for all tangent vectors* v*, and* $\nabla g_1, \dots \nabla g_n$ *are* n *linearly independent vectors that are orthogonal to the tangent space.*

4. Parametric curves.

DEFINITION 4.1. Let $\Gamma \subseteq \mathbb{R}^d$ be a (differentiable) closed curve. We say γ is a (differentiable) parametrization of Γ if $\gamma : [a, b] \to \Gamma$ is differentiable, $D\gamma \neq 0$, $\gamma : [a, b) \to \Gamma$ is bijective, $\gamma(b) = \gamma(a)$ and $\gamma'(a) = \gamma'(b)$. A curve with a parametrization is called a parametric curve.

EXAMPLE 4.2. The curve $x^2 + y^2 = 1$ can be parametrized by $\gamma(t) = (\cos t, \sin t)$ for $t \in [0, 2\pi]$

REMARK 4.3. A curve can have many parametrizations. For example, $\delta(t) = (\cos t, \sin(-t))$ also parametrizes the unit circle, but runs clockwise instead of counter clockwise. Choosing a parametrization requires choosing the direction of traversal through the curve.

REMARK 4.4. If γ is a curve with endpoints, then we require $\{\gamma(a), \gamma(b)\}$ to be the endpoints of the curve (instead of $\gamma(b) = \gamma(a)$).

REMARK 4.5. If γ is an open curve, then we only require γ to be defined (and bijective) on (a, b).

REMARK 4.6. While curves can not self-intersect, we usually allow parametric curves to self-intersect. This is done by replacing the requirement that γ is injective with the requirement that if for $x, y \in (a, b)$ we have $\gamma(x) = \gamma(y)$ then $D\gamma_x$ and $D\gamma_y$ are linearly independent. Sometimes, one also allows parametric curves loop back on themselves (e.g. $\gamma(t) = (\cos(t), \sin(t))$ for $t \in \mathbb{R}$.

DEFINITION 4.7. If γ represents a differentiable parametric curve, we define $\gamma' = D\gamma$.

REMARK 4.8. For any t, $\gamma'(t)$ is a vector in $\mathbb{R}^d$. Think of $\gamma(t)$ representing the position of a particle, and γ' to represent the velocity.

PROPOSITION 4.9. *Let* Γ *be a curve and* γ *be a parametrization,* $a = \gamma(t_0) \in \Gamma$. *Then*

$$T\Gamma_a = \operatorname{span}\{\gamma'(t_0)\}.$$

Consequently, tangent line through a *is* $\{\gamma(t_0) + t\gamma'(t_0) \mid t \in \mathbb{R}\}$.

If we think of $\gamma(t)$ as the position of a particle at time t, then the above says that the tangent space is spanned by the *velocity* of the particle. That is, the velocity of the particle is always tangent to the curve it traces out. However, the acceleration of the particle (defined to be γ'') *need not* be tangent to the curve! In fact if the magnitude of the velocity $|\gamma'|$ is constant, then the acceleration will be *perpendicular* to the curve!

PROOF OF PROPOSITION 4.9. We only do the proof in 3D. Write $\Gamma = \{f = 0\}$ where $f : \mathbb{R}^3 \to \mathbb{R}^2$ is a differentiable function such that $\operatorname{rank}(Df_a) = 2$. In this case $\Gamma = S^{(1)} \cap S^{(2)}$ where $S^{(i)}$ is the surface $\{f_i = 0\}$. Since $\Gamma \subseteq S^{(i)}$, $f_i \circ \gamma = 0$ and hence (by the chain rule) $\gamma'(t) \in \ker(Df_i(a))$. By dimension counting this forces

$$T\Gamma_a = TS^{(1)}_a \cap TS^{(2)}_a = \ker(Df_1(a)) \cap \ker(Df_2(a)) = \operatorname{span}\{\gamma'(t)\}. \qquad \square$$

5. Curves, surfaces, and manifolds

In the previous sections we talked about tangents to curves and surfaces. However, we haven't ever precisely defined what a curve or surface is. For the curious, the definitions are here. The main result (i.e. that ∇g is orthogonal to level sets, and that $\ker(Dg)$ is the tangent space) is still true in arbitrary dimensions.

DEFINITION 5.1. We say $\Gamma \subseteq \mathbb{R}^n$ is a (differentiable) *curve* if for every $a \in \Gamma$ there exists domains $U \subseteq \mathbb{R}^n$, $V \subseteq \mathbb{R}$ and a differentiable function $\varphi : V \to U$ such that $D\varphi \neq 0$ in V and $U \cap \Gamma = \varphi(V)$.

REMARK 5.2. Many authors insist $V = (0,1)$ or $V = \mathbb{R}$. This is equivalent to what we have.

EXAMPLE 5.3. If $f : \mathbb{R} \to \mathbb{R}^n$ is a differentiable function, then the graph $\Gamma \subseteq \mathbb{R}^{n+1}$ defined by $\Gamma = \{(x, f(x)) \mid x \in \mathbb{R}\}$ is a differentiable curve.

PROPOSITION 5.4. *Let $f : \mathbb{R}^{n+1} \to \mathbb{R}^n$ is differentiable, $c \in \mathbb{R}^n$ and $\Gamma = \{x \in \mathbb{R}^{n+1} \mid f(x) = c\}$ be the level set of f. If at every point in Γ, the matrix Df has rank n then Γ is a curve.*

PROOF. Let $a \in \Gamma$. Since $\operatorname{rank}(Df_a) = d$, there must be d linearly independent columns. For simplicity assume these are the first d ones. The implicit function theorem applies and guarantees that the equation $f(x) = c$ can be solved for $x_1, \dots, x_n$, and each x_i can be expressed as a differentiable function of x_{n+1} (close to a). That is, there exist open sets $U' \subseteq \mathbb{R}^n$, $V' \subseteq \mathbb{R}$ and a differentiable function g such that $a \in U' \times V'$ and $\Gamma \cap (U' \times V') = \{(g(x_{n+1}), x_{n+1}) \mid x_{n+1} \in V'\}$. □

Surfaces in higher dimensions are defined similarly.

DEFINITION 5.5. We say $\Sigma \subseteq \mathbb{R}^n$ is a (differentiable) *surface* if for every $a \in \Sigma$ there exists domains $U \subseteq \mathbb{R}^n$, $V \subseteq \mathbb{R}^2$ and a differentiable function $\varphi : V \to U$ such that $\operatorname{rank}(D\varphi) = 2$ at every point in V and $U \cap \Sigma = \varphi(V)$.

The difference from a curve is that now $V \subseteq \mathbb{R}^2$ and not $\mathbb{R}$.

DEFINITION 5.6. We say $M \subseteq \mathbb{R}^n$ is a d-dimensional (differentiable) *manifold* if for every $a \in M$ there exists domains $U \subseteq \mathbb{R}^n$, $V \subseteq \mathbb{R}^d$ and a differentiable function $\varphi : V \to U$ such that $\operatorname{rank}(D\varphi) = d$ at every point in V and $U \cap M = \varphi(V)$.

REMARK 5.7. For $d = 1$ this is just a curve, and for $d = 2$ this is a surface.

REMARK 5.8. If $d = 1$ and Γ is a connected, then there exists an interval U and an injective differentiable function $\gamma : U \to \mathbb{R}^n$ such that $D\gamma \neq 0$ on U and $\gamma(U) = \Gamma$. If $d > 1$ this is no longer true: even though near every point the surface is a differentiable image of a rectangle, the entire surface need not be one.

As before d-dimensional manifolds can be obtained as level sets of functions $f : \mathbb{R}^{n+d} \to \mathbb{R}^d$ provided we have $\operatorname{rank}(Df) = d$ on the entire level set.

PROPOSITION 5.9. *Let $f : \mathbb{R}^{n+d} \to \mathbb{R}^n$ is differentiable, $c \in \mathbb{R}^n$ and $\Gamma = \{x \in \mathbb{R}^{n+1} \mid f(x) = c\}$ be the level set of f. If at every point in Γ, the matrix Df has rank d then Γ is a d-dimensional manifold.*

The results from the previous section about tangent spaces of implicitly defined manifolds generalize naturally in this context.

DEFINITION 5.10. Let $U \subseteq \mathbb{R}^d$, $f : U \to R$ be a differentiable function, and $M = \{(x, f(x)) \in \mathbb{R}^{d+1} \mid x \in U\}$ be the graph of f. (Note M is a d-dimensional manifold in $\mathbb{R}^{d+1}$.) Let $(a, f(a)) \in M$.

- The *tangent "plane"* at the point $(a, f(a))$ is defined by
$$\{(x, y) \in \mathbb{R}^{d+1} \mid y = f(a) + Df_a(x - a)\}$$
- The *tangent space* at the point $(a, f(a))$ (denoted by $TM_{(a,f(a))}$) is the subspace defined by
$$TM_{(a,f(a))} = \{(x, y) \in \mathbb{R}^{d+1} \mid y = Df_a x\}.$$

REMARK 5.11. When $d = 2$ the tangent plane is really a plane. For $d = 1$ it is a line (the tangent line), and for other values it is a d-dimensional hyper-plane.

PROPOSITION 5.12. *Suppose $f : \mathbb{R}^{n+d} \to \mathbb{R}^n$ is differentiable, and the level set $\Gamma = \{x \mid f(x) = c\}$ is a d-dimensional manifold. Suppose further that Df_a has rank n for all $a \in \Gamma$. Then the tangent space at a is precisely the kernel of Df_a, and the vectors $\nabla f_1, \dots \nabla f_n$ are n linearly independent vectors that are normal to the tangent space.*

6. Constrained optimization.

Consider an implicitly defined surface $S = \{g = c\}$, for some $g : \mathbb{R}^3 \to \mathbb{R}$. Our aim is to maximise or minimise a function f *on this surface.*

DEFINITION 6.1. We say a function f attains a local maximum at a on the surface S, if there exists $\varepsilon > 0$ such that $|x - a| < \varepsilon$ and $x \in S$ imply $f(a) \geqslant f(x)$.

REMARK 6.2. This is sometimes called constrained local maximum, or local maximum subject to the constraint $g = c$.

PROPOSITION 6.3. *Suppose $\nabla g \neq 0$ at every point on S. If f attains a constrained local maximum (or minimum) at a on the surface S, then $\exists \lambda \in \mathbb{R}$ such that $\nabla f(a) = \lambda \nabla g(a)$.*

INTUITION. If $\nabla f(a) \neq 0$, then $S' \stackrel{\text{def}}{=} \{f = f(a)\}$ is a surface. If f attains a constrained maximum at a then S' must be tangent to S at the point a. This forces $\nabla f(a)$ and $\nabla g(a)$ to be parallel. □

PROPOSITION 6.4 (Multiple constraints). *Let $f, g_1, \dots, g_n : \mathbb{R}^d \to \mathbb{R}$ be $: \mathbb{R}^d \to \mathbb{R}$ be differentiable. If f attains a local maximum (or minimum) at a subject to the constraints $g_1 = c_1$, $g_2 = c_2$, $\dots g_n = c_n$, and $\nabla g_1(a), \dots, \nabla g_n(a)$ are linearly independent, then $\exists \lambda_1, \dots \lambda_n \in \mathbb{R}$ such that $\nabla f(a) = \sum_1^n \lambda_i \nabla g_i(a)$.*

PROOF IN A SPECIAL CASE. Let's assume $n = 1$ and $d = 3$. The proof in this case generalizes without much difficulty to the general case, but the notation is much more cumbersome. For notational convenience we will also write $g = g_1$.

By assumption $\nabla g(a) \neq 0$ and hence $\partial_i g(a) \neq 0$ for some i. For notational simplicity, we assume $i = 3$. Now, by the implicit function theorem there exists a function h defined in a small neighborhood of (a_1, a_2) such that for all x close to a, we have $g(x) = c$ if and only if $x_3 = h(x_1, x_2)$. In particular, this means that close to a, all points on the surface $\{g = c\}$ are of the form $(x_1, x_2, h(x_1, x_2))$. Thus, f attains a constrained local extremum at a if and only if the function $(x_1, x_2) \mapsto f(x_1, x_2, h(x_1, x_2))$ attains an *unconstrained* local extremum at (a_1, a_2).

Define
$$\varphi(x_1, x_2) \stackrel{\text{def}}{=} f(x_1, x_2, h(x_1, x_2)) .$$
By the above argument we know that φ attains an *unconstrained* local extremum at (a_1, a_2). Now we know that unconstrained local extrema are only attained at

critical points. Thus to finish the proof, all we need to do is compute the critical points of φ and check that it yields $\nabla f(a) = \lambda \nabla g(a)$ for some $\lambda \in \mathbb{R}$.

To do this, define

$$\psi(x_1, x_2) = (x_1, x_2, h(x_1, x_2)),$$

and observe $\varphi = f \circ \psi$. By the chain rule, we know $D\varphi = Df_\psi D\psi$. Transposing this, and using the fact that $D\varphi_{(a_1,a_2)} = 0$, gives

$$(D\varphi_a)^T = (D\psi_{(a_1,a_2)})^T \nabla f(a) = 0 \implies \nabla f(a) \in \ker\big((D\psi_{(a_1,a_2)})^T\big). \tag{6.1}$$

Moreover, by construction of h, we know $g \circ \psi(x_1, x_2) = c$, and hence $D(g \circ \psi) = 0$. Transposing this and evaluating at (a_1, a_2) gives

$$(D\psi_{(a_1,a_2)})^T \nabla g(a) = 0 \implies \nabla g(a) \in \ker\big(D\psi^T_{(a_1,a_2)}\big). \tag{6.2}$$

Observe now that

$$(D\psi)^T = \begin{pmatrix} 1 & 0 & \partial_1 h \\ 0 & 1 & \partial_2 h \end{pmatrix},$$

and hence $\operatorname{rank}(D\psi)^T = 2$. Thus by the rank nullity theorem, $\dim\ker((D\psi)^T) = 1$. Since $\nabla g(a) \neq 0$, by assumption and $\nabla g(a) \in \ker(D\psi_{(a_1,a_2)})$ (from (6.2)), we must have $\ker((D\psi_{(a_1,a_2)})^T) = \operatorname{span}\{\nabla g(a)\}$. But since $\nabla f(a) \in \ker((D\psi_{(a_1,a_2)})^T)$ (by (6.1)), this implies there exists $\lambda \in \mathbb{R}$ such that $\nabla f(a) = \lambda \nabla g(a)$. □

REMARK 6.5 (Proof when $n \neq 1$). The main difference in this case is to note that $\nabla g_1, \ldots, \nabla g_n$ form a basis of $\ker((D\psi)^T)$, and use this to conclude the existence of $\lambda_1, \ldots, \lambda_n$.

To explicitly find constrained local maxima in $\mathbb{R}^d$ with n constraints we do the following:

- Simultaneously solve the system of equations

$$\begin{aligned} \nabla f(x) &= \lambda_1 \nabla g_1(x) + \cdots \lambda_n \nabla g_n(x) \\ g_1(x) &= c_1, \\ &\cdots \\ g_n(x) &= c_n. \end{aligned}$$

- The unknowns are the d-coordinates of x, and the Lagrange multipliers $\lambda_1, \ldots, \lambda_n$. This is $n + d$ variables.
- The first equation above is a vector equation where both sides have d coordinates. The remaining are scalar equations. So the above system is a system of $n + d$ equations with $n + d$ variables.
- The typical situation will yield a finite number of solutions.
- There is a test involving the *bordered Hessian* for whether these points are constrained local minima / maxima or neither. These are quite complicated, and are usually more trouble than they are worth, so one usually uses some ad-hoc method to decide whether the solution you found is a local maximum or not.

EXAMPLE 6.6. Find necessary conditions for $f(x, y) = y$ to attain a local maxima/minima of subject to the constraint $y = g(x)$.

Of course, from one variable calculus, we know that the local maxima / minima must occur at points where $g' = 0$. Let's revisit it using the constrained optimization technique above.

SOLUTION. Note our constraint is of the form $y - g(x) = 0$. So at a local maximum we must have

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \nabla f = \lambda \nabla (y - g(x)) = \begin{pmatrix} -g'(x) \\ 1 \end{pmatrix} \quad \text{and} \quad y = g(x).$$

This forces $\lambda = 1$ and hence $g'(x) = 0$, as expected. □

EXAMPLE 6.7. Maximise xy subject to the constraint $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$.

SOLUTION. At a local maximum,

$$\begin{pmatrix} y \\ x \end{pmatrix} = \nabla(xy) = \lambda \nabla\Big(\frac{x^2}{a^2} + \frac{y^2}{b^2}\Big) = \lambda \begin{pmatrix} 2x/a^2 \\ 2y/b^2. \end{pmatrix}$$

which forces $y^2 = x^2 b^2 / a^2$. Substituting this in the constraint gives $x = \pm a/\sqrt{2}$ and $y = \pm b/\sqrt{2}$. This gives four possibilities for xy to attain a maximum. Directly checking shows that the points $(a/\sqrt{2}, b/\sqrt{2})$ and $(-a/\sqrt{2}, -b/\sqrt{2})$ both correspond to a local maximum, and the maximum value is $ab/2$. □

PROPOSITION 6.8 (Cauchy-Schwartz). *If $x, y \in \mathbb{R}^n$ then $|x \cdot y| \leqslant |x||y|$.*

PROOF. Maximise $x \cdot y$ subject to the constraint $|x| = a$ and $|y| = b$. □

PROPOSITION 6.9 (Inequality of the means). *If $x_i \geqslant 0$, then*

$$\frac{1}{n} \sum_1^n x_i \geqslant \Big(\prod_1^n x_i\Big)^{1/n}.$$

PROPOSITION 6.10 (Young's inequality). *If $p, q > 1$ and $1/p + 1/q = 1$ then*

$$|xy| \leqslant \frac{|x|^p}{p} + \frac{|y|^q}{q}.$$