Chapter 5

Nonlinear optimisation

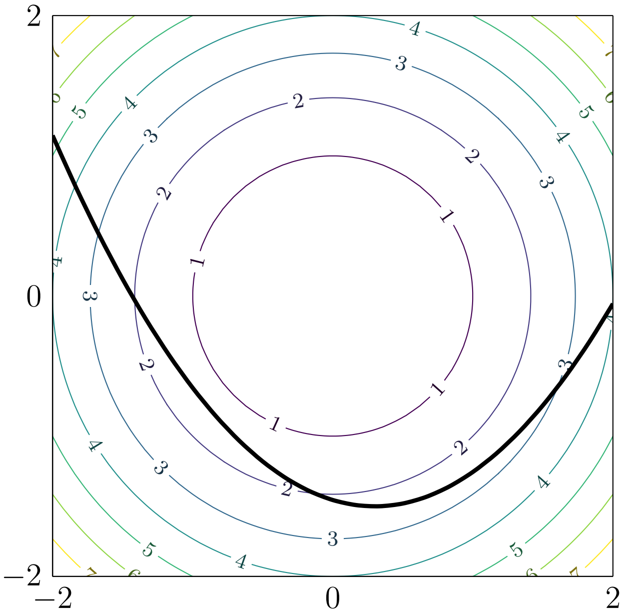
5.1 An introductory example

Consider the following optimisation problem:

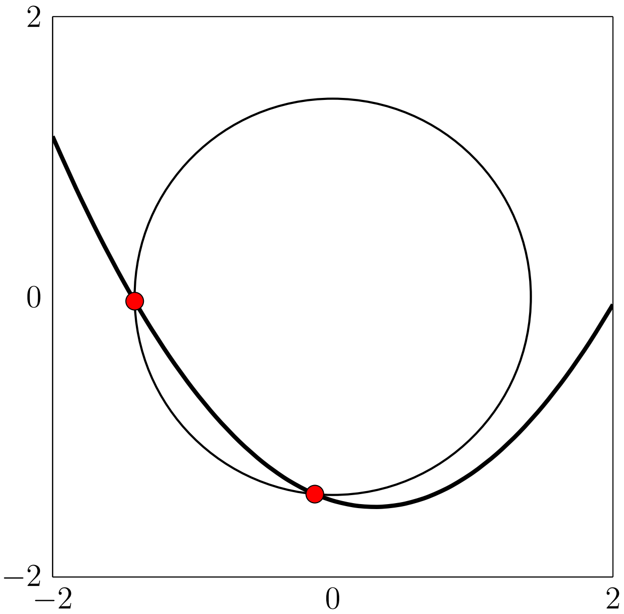
$$\begin{aligned} \text{minimise} \quad & f(\mathbf{x}) = x_1^2 + x_2^2 \\ \text{subject to} \quad & h(\mathbf{x}) = (x_1 - 0.3)^2 - 2x_2 = 3. \end{aligned}$$

The corresponding *unconstrained* problem of minimising $f(\mathbf{x})$ has an easy solution: as $x_1^2 + x_2^2$ is positive when $x_1, x_2 > 0$, the minimum of 0 is obtained when $x_1 = x_2 = 0$. But this is not a solution to the constrained problem, as $h(0,0) = 0.09 \neq 3$.

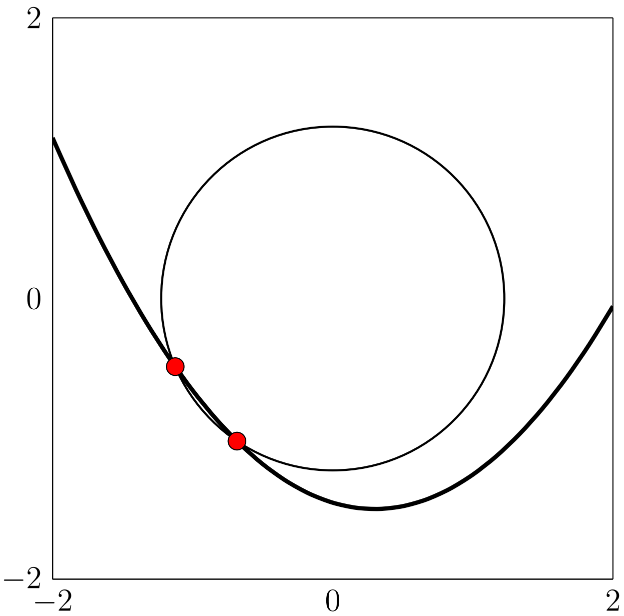
In this section, to illustrate the underlying concepts, we will investigate this problem from a geometric perspective. In the following sections we will see how these ideas generalise to higher dimensions. In the plot below, the concentric circles are level sets of $f(\mathbf{x})$. The curve $h(\mathbf{x}) = 3$ is shown in black; it forms the feasible set.



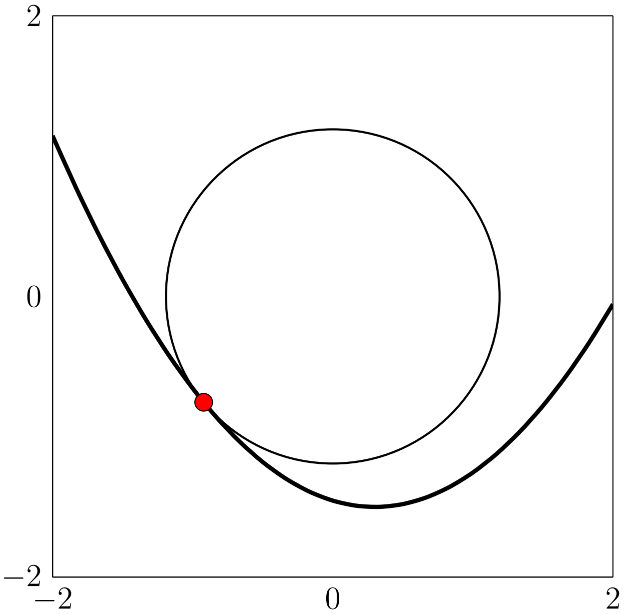
Consider the level set for $f(\mathbf{x}) = 2$, shown below. The two red dots mark the intersection of the curves defined by $f(\mathbf{x}) = 2$ and $h(\mathbf{x}) = 3$. Evaluating f at any point inside this circle will result in a value smaller than 2. Since a portion of the feasible set lies inside this circle, this immediately disqualifies the red points from being minimisers.



The same observations hold if we consider the level set given by $f(\mathbf{x}) = 1.5$ instead.



However, if we choose a level set that is tangent to the curve $h(\mathbf{x}) = 3$, as depicted below, then we find that there is no portion of the feasible set inside the circle. More formally, there is a neighbourhood of $\mathbf{x}$ whose intersection with the feasible set lies entirely on the side of the level set corresponding to higher values of f.



Thus, we reformulate the problem as follows: find a point $\mathbf{x}^*$ such that, at $\mathbf{x}^*$, the lines tangent to the curves given by $f(\mathbf{x}) = f(\mathbf{x}^*)$ and $h(\mathbf{x}) = 3$ coincide.

Let $\mathbf{x}^* = (a, b)$ be the supposed point. We then have $f(\mathbf{x}^*) = a^2 + b^2$, and so the corresponding level set has equation $x_1^2 + x_2^2 = a^2 + b^2$. One can show using elementary calculus that the line tangent to this circle at $\mathbf{x}^*$ has slope $-\frac{a}{b}$, and that the line tangent to the curve $h(\mathbf{x}) = 3$ at $\mathbf{x}^*$ has slope $a - 0.3$. Equating these results in

$$a - 0.3 = -\frac{a}{b} \iff b = \frac{a}{0.3 - a}.$$

As $\mathbf{x}^*$ is assumed feasible, we must have $h(\mathbf{x}^*) = 3$, giving

$$b = \frac{(a - 0.3)^2 - 3}{2}.$$

So we arrive at an equation in one variable,

$$\begin{aligned} &\frac{(a-0.3)^2-3}{2} = \frac{a}{0.3-a} \\ \iff\ & (a-0.3)^3 + 3(0.3-a) + 2a = 0 \\ \iff\ & a^3 - 0.9a^2 - 0.73a + 0.873 = 0. \end{aligned}$$

It is possible to solve this exactly using the cubic formula, but we will employ MATLAB's `roots` function instead:

```
>> roots([1 -0.9 -0.73 0.873])
ans =
  -0.921196686181078 + 0.000000000000000i
   0.910598343090540 + 0.344225231141957i
   0.910598343090540 - 0.344225231141957i
```

We are interested only in real solutions, so we take $a \approx -0.921197$, giving

$$b = \frac{(a-0.3)^2-3}{2} \approx -0.754339.$$

So the minimiser is $\mathbf{x}^* \approx (-0.921197, -0.754339)$ giving $f(\mathbf{x}^*) \approx 1.417631$.

It remains to be seen how to generalise this. We will see that this key idea works in any dimension, but in order to do this we have to understand two important spaces, connected to level sets of functions, namely the tangent space and normal space.

5.2 Tangent and normal spaces

Background: linear algebra

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be vectors in $\mathbb{R}^n$. A vector $\mathbf{v}$ is a **linear combination** of those vectors if there are real numbers $a_1, a_2, \ldots, a_k$ such that

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_k\mathbf{v}_k.$$

The set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is **linearly independent** if $a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_k\mathbf{v}_k = \mathbf{0}$ implies $a_i = 0$, for all i. Otherwise, they are called **linearly dependent**. For a set of vectors S, the **span** of S, denoted by $\mathrm{Sp}(S)$, is the set of all possible linear combinations of those vectors. Occasionally, set brackets will be omitted, so that $\mathrm{Sp}(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \mathrm{Sp}(\{\mathbf{v}_1, \ldots, \mathbf{v}_k\})$. In the special case $S = \varnothing$, we define $\mathrm{Sp}(\varnothing) = \{\mathbf{0}\}$. Also, note that the set $\{\mathbf{0}\}$ is *not* linearly independent.

A subset $\mathcal{V} \subset \mathbb{R}^n$ is a **vector subspace** of $\mathbb{R}^n$ if it is closed under vector addition and scalar multiplication; that is, if $\mathbf{u}, \mathbf{v} \in \mathcal{V}$ and $a \in \mathbb{R}$, then $\mathbf{u} + \mathbf{v} \in \mathcal{V}$ and $a\mathbf{u} \in \mathcal{V}$. Every span is a vector space, and conversely, every vector space can be written as the span of a finite set of vectors. If $\mathcal{V} = \mathrm{Sp}(B)$ and B is linearly independent, then B is called a **basis** of $\mathcal{V}$. In that case, the number of vectors in B is unique, and this number is called the **dimension** of $\mathcal{V}$, denoted by $\dim(\mathcal{V})$. The only subspace of $\mathbb{R}^n$ with dimension n is $\mathbb{R}^n$ itself.

The **orthogonal complement** of a vector subspace $\mathcal{V} \subset \mathbb{R}^n$ is the set of all vectors that are orthogonal to every vector in $\mathcal{V}$,

$$\mathcal{V}^\perp = \{\mathbf{v} : (\forall \mathbf{w} \in \mathcal{V})\ \mathbf{v}^T\mathbf{w} = 0\}.$$

For an $m \times n$ matrix $\mathbf{A}$, the **kernel** of $\mathbf{A}$, written $\mathrm{Ker}(\mathbf{A})$, is the set of vectors $\mathbf{v}$ in $\mathbb{R}^n$ which are solutions to $\mathbf{A}\mathbf{v} = \mathbf{0}$, and the **image** of $\mathbf{A}$, written $\mathrm{Im}(\mathbf{A})$, is the set of vectors in $\mathbb{R}^m$ that can be obtained by multiplying a vector in $\mathbb{R}^n$ by $\mathbf{A}$. That is,

$$\mathrm{Ker}(\mathbf{A}) = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{A}\mathbf{v} = \mathbf{0}\} \quad \text{and} \quad \mathrm{Im}(\mathbf{A}) = \{\mathbf{A}\mathbf{v} : \mathbf{v} \in \mathbb{R}^n\}.$$

Equivalently, $\mathrm{Im}(\mathbf{A})$ is the span of the columns of $\mathbf{A}$. Also, $\mathrm{Ker}(\mathbf{A})$ is a subspace of $\mathbb{R}^n$ and $\mathrm{Im}(\mathbf{A})$ is a subspace of $\mathbb{R}^m$. The **rank-nullity theorem** is that $\dim(\ker(\mathbf{A})) + \dim(\mathrm{Im}(\mathbf{A})) = \dim(\mathbb{R}^n) = n$.

The **fundamental theorem of linear algebra** [3, Theorem 3.4] tells us that

$$\mathrm{Im}(\mathbf{A})^\perp = \mathrm{Ker}(\mathbf{A}^T), \qquad \mathrm{Ker}(\mathbf{A})^\perp = \mathrm{Im}(\mathbf{A}^T).$$

To ease notation in this chapter, on occasion we will use the ordered tuple $(x_1, x_2, \ldots, x_n)$ to denote the vector $\begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix}^T$.

Example: linear independence

Consider the vectors

$$\mathbf{v_1} = (-4, 5, 1), \quad \mathbf{v_2} = (-4, -2, 1), \quad \mathbf{v_3} = (5, 5, -4).$$

To decide if $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$ is linearly independent, we determine all solutions to

$$x\mathbf{v}_1 + y\mathbf{v}_2 + z\mathbf{v}_3 = \mathbf{0},$$

which corresponds to the system of linear equations

$$\begin{aligned} -4x - 4y + 5z &= 0, \\ 5x - 2y + 5z &= 0, \\ x + y - 4z &= 0. \end{aligned}$$

Using `rref([-4 -4 5 0; 5 -2 5 0; 1 1 -4 0])` in MATLAB, we obtain

$$\left(\begin{array}{ccc|c} -4 & -4 & 5 & 0 \\ 5 & -2 & 5 & 0 \\ 1 & 1 & -4 & 0 \end{array}\right) \equiv \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right),$$

implying $x = y = z = 0$. Hence, the set $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$ is linearly independent.

Consider instead the vectors

$$\mathbf{u_1} = (5, 3, 5), \quad \mathbf{u_2} = (2, -4, 4), \quad \mathbf{u_3} = (-2, -9, 1).$$

Using `rref([5 2 -2 0; 3 -4 -9 0; 5 4 1 0])` in MATLAB shows that

$$\left(\begin{array}{ccc|c} 5 & 2 & -2 & 0 \\ 3 & -4 & -9 & 0 \\ 5 & 4 & 1 & 0 \end{array}\right) \equiv \left(\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 3/2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right),$$

which has infinitely many non-zero solutions, so the set $\{\mathbf{u_1}, \mathbf{u_2}, \mathbf{u_3}\}$ is *not* linearly independent.

In particular, note that

$$\mathbf{u}_1 - \tfrac{3}{2}\mathbf{u}_2 + \mathbf{u}_3 = \mathbf{0},$$

implying $\mathbf{u}_3 = -\mathbf{u}_1 + \frac{3}{2}\mathbf{u}_2$. The vector $\mathbf{u}_3$ is a linear combination of the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$, making $\mathbf{u}_3$ "redundant" in the sense that

$$\mathrm{Sp}(\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}) = \mathrm{Sp}(\{\mathbf{u}_1, \mathbf{u}_2\}).$$

Example: bases and spans

In general, a basis for $\mathrm{Sp}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ can be found by applying Gaussian elimination to the matrix whose rows are $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and then reading the resulting rows as vectors.

For example, consider the vectors in $\mathbb{R}^3$ given by

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ -4 \\ -2 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 1 \\ 4 \\ 5 \end{pmatrix}.$$

The corresponding matrix and row reduced form is

$$\begin{pmatrix} 1 & -4 & -2 \\ 1 & 4 & 5 \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 & \frac{3}{2} \\ 0 & 1 & \frac{7}{8} \end{pmatrix},$$

so $\mathrm{Sp}((1, -4, -2), (1, 4, 5)) = \mathrm{Sp}((1, 0, \frac{3}{2}), (0, 1, \frac{7}{8}))$.

The effect of this is more observable when the vectors are linearly dependent. For example, let

$$\mathbf{u}_1 = (-4, 5, 5), \quad \mathbf{u}_2 = (0, 3, -4), \quad \mathbf{u}_3 = (-4, 2, 9).$$

To find a basis for $\mathrm{Sp}((-4,5,5),(0,3,-4),(-4,2,9))$, the following MATLAB code can be used:

```
format rational;
rref([-4 5 5; 0 3 -4; -4 2 9])
```

which gives

$$\begin{pmatrix}-4 & 5 & 5\\ 0 & 3 & -4\\ -4 & 2 & 9\end{pmatrix} \equiv \begin{pmatrix}1 & 0 & -35/12\\ 0 & 1 & -4/3\\ 0 & 0 & 0\end{pmatrix}.$$

So $\mathrm{Sp}((-4,5,5),(0,3,-4),(-4,2,9)) = \mathrm{Sp}((1,0,-35/12),(0,1,-4/3))$. Note that we started with 3 vectors and ended with 2 because one of the original vectors could be written as a linear combination of the others.

The span of a single vector corresponds to a line passing through the origin and the point corresponding to that vector. For example, in $\mathbb{R}^2$,

$$\mathrm{Sp}((1,2)) = \{x(1,2) : x \in \mathbb{R}\} = \{(x,2x) : x \in \mathbb{R}\},$$

which is the line passing through the origin and the point $(1,2)$.

The span of two linearly independent vectors is a plane passing through the origin and the two corresponding points. For example, in $\mathbb{R}^3$,

$$\mathrm{Sp}((1,2,1),(1,0,3)) = \{x(1,2,1)+y(1,0,3) : x,y \in \mathbb{R}\} = \{(x+y,2x,x+3y) : x,y \in \mathbb{R}\},$$

which is the plane passing through the origin and the points $(1,2,1)$ and $(1,0,3)$.

Example: kernel and image

Consider the matrix $\mathbf{A}$, where

$$\mathbf{A} = \begin{pmatrix}-4 & -4 & 5\\ 5 & -2 & 5\\ 1 & 1 & -4\end{pmatrix}.$$

We kernel of $\mathbf{A}$ can be found by applying Gaussian elimination to $\mathbf{A}$ and reading off the associated solution to $\mathbf{A}\mathbf{x} = \mathbf{0}$. We have

$$\begin{pmatrix}-4 & -4 & 5\\ 5 & -2 & 5\\ 1 & 1 & -4\end{pmatrix} \equiv \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix},$$

so $\mathrm{Ker}(\mathbf{A}) = \{(x_1,x_2,x_3) \in \mathbb{R}^3 : x_1 = 0, x_2 = 0, x_3 = 0\} = \{\mathbf{0}\}$. The image of $\mathbf{A}$ can be found by applying Gaussian elimination to $\mathbf{A}^T$, and reading the rows as vectors

$$\begin{pmatrix}-4 & 5 & 1\\ -4 & -2 & 1\\ 5 & 5 & -4\end{pmatrix} \equiv \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}.$$

So $\mathrm{Im}(\mathbf{A}) = \mathrm{Sp}((1,0,0),(0,1,0),(0,0,1)) = \mathbb{R}^3$.

Now consider the matrix $\mathbf{B}$, where

$$\mathbf{B} = \begin{pmatrix}4 & -8\\ 2 & -4\end{pmatrix}.$$

We use MATLAB to reduce $\mathbf{B}$ and $\mathbf{B}^T$:

```
format rational;
B = [4 -8; 2 -4];
rref(B)
rref(B')
```

This gives

$$\mathbf{B} \equiv \begin{pmatrix} 1 & -2 \\ 0 & 0 \end{pmatrix}, \qquad \mathbf{B}^T \equiv \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & 0 \end{pmatrix}.$$

From the first matrix, we have

$$\begin{aligned} \mathrm{Ker}(\mathbf{B}) &= \{(x_1, x_2) \in \mathbb{R}^2 : x_1 - 2x_2 = 0\} \\ &= \{(x_1, x_2) \in \mathbb{R}^2 : x_1 = 2x_2\} \\ &= \{(2x_2, x_2) \in \mathbb{R}^2 : x_2 \in \mathbb{R}\} \\ &= \{x_2(2,1) \in \mathbb{R}^2 : x_2 \in \mathbb{R}\} \\ &= \mathrm{Sp}((2,1)). \end{aligned}$$

From the second matrix, we have

$$\mathrm{Im}(\mathbf{B}) = \mathrm{Sp}((1, \tfrac{1}{2})) = \mathrm{Sp}(2,1).$$

In this case, the image and kernel are the same, but this need not be true in general. Geometrically, they can be described as the line passing through the origin and the point $(2,1)$ in $\mathbb{R}^2$.

Now consider the matrix $\mathbf{C}$, where

$$\mathbf{C} = \begin{pmatrix} 5 & -1 & 11 \\ 2 & 2 & 2 \\ 3 & -4 & 10 \\ 3 & 2 & 4 \end{pmatrix}.$$

As $\mathbf{C}$ has three columns, its kernel is a subspace of $\mathbb{R}^3$. We have

$$\begin{pmatrix} 5 & -1 & 11 \\ 2 & 2 & 2 \\ 3 & -4 & 10 \\ 3 & 2 & 4 \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

so

$$\begin{aligned} \mathrm{Ker}(\mathbf{C}) &= \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 + 2x_3 = 0, x_2 - x_3 = 0\} \\ &= \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 = -2x_3, x_2 = x_3\} \\ &= \{(-2x_3, x_3, x_3) \in \mathbb{R}^3 : x_3 \in \mathbb{R}\} \\ &= \{x_3(-2,1,1) \in \mathbb{R}^3 : x_3 \in \mathbb{R}\} \\ &= \mathrm{Sp}((-2,1,1)), \end{aligned}$$

which is the line in $\mathbb{R}^3$ passing through the origin and the point $(-2,1,1)$.

For $\mathrm{Im}(\mathbf{C})$, we have

$$\begin{pmatrix} 5 & 2 & 3 & 3 \\ -1 & 2 & -4 & 2 \\ 11 & 2 & 10 & 4 \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 & 7/6 & 1/6 \\ 0 & 1 & -17/12 & 13/12 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

so

$$\mathrm{Im}(\mathbf{C}) = \mathrm{Sp}((1,0,7/6,1/6),(0,1,-17/12,13/12)) = \mathrm{Sp}((6,0,7,1),(0,12,-17,13)),$$

which is the plane in $\mathbb{R}^4$ passing through the origin and the points $(6,0,7,1)$ and $(0,12,-17,13)$.

Consider a function $\mathbf{h}\colon \mathbb{R}^n \to \mathbb{R}^m$, let $\mathbf{c} \in \mathbb{R}^m$, and let $\mathcal{H} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{h}(\mathbf{x}) = \mathbf{c}\}$ be the $\mathbf{c}$-level set of $\mathbf{h}$. The **tangent space** of $\mathcal{H}$ at $\mathbf{p} \in \mathcal{H}$ is the set

$$T\mathcal{H}(\mathbf{p}) = \mathrm{Ker}(D\mathbf{h}(\mathbf{p})).$$

The **normal space** of $\mathcal{H}$ at $\mathbf{p} \in \mathcal{H}$ is the set

$$N\mathcal{H}(\mathbf{p}) = \mathrm{Im}(D\mathbf{h}^T(\mathbf{p})).$$

Note that the derivative (or Jacobian) of $\mathbf{h}$ can be written as

$$D\mathbf{h}(\mathbf{x}) = \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial h_m}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_n} \end{pmatrix} (\mathbf{x}) = \begin{pmatrix} \nabla h_1^T(\mathbf{x}) \\ \vdots \\ \nabla h_m^T(\mathbf{x}) \end{pmatrix}.$$

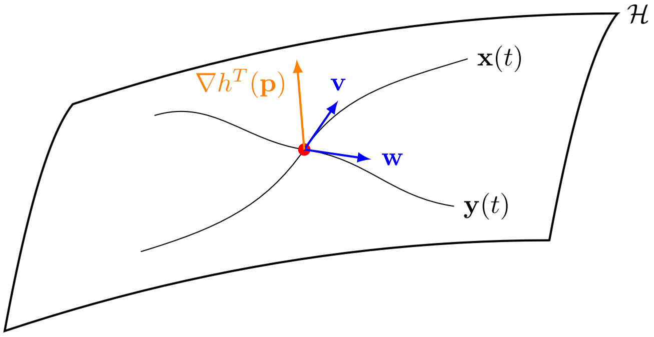
So, $D\mathbf{h}^T$ is the matrix whose k-th column is ∇h_k and hence

$$N\mathcal{H}(\mathbf{p}) = \mathrm{Sp}(\nabla h_1, \ldots, \nabla h_m).$$

To illustrate these ideas, consider a function $h\colon \mathbb{R}^3 \to \mathbb{R}$, and consider a c-level set of h,

$$\mathcal{H} = \{\mathbf{x} \in \mathbb{R}^3 : h(\mathbf{x}) = c\},$$

which corresponds to a surface in $\mathbb{R}^3$, depicted generically below.



Let $\mathbf{x}, \mathbf{y}\colon \mathbb{R} \to \mathbb{R}^3$ be parametrisations of two curves on $\mathcal{H}$, which intersect transversally at a point $\mathbf{p} = \mathbf{x}(r) = \mathbf{y}(s) \in \mathcal{H}$, which is marked by the red dot in the figure above. Let $\mathbf{v}$ denote the derivative of $\mathbf{x}$ at r and let $\mathbf{w}$ denote the derivative of $\mathbf{y}$ at s, so that $\mathbf{v} = D\mathbf{x}(r)$ and $\mathbf{w} = D\mathbf{y}(s)$. These vectors are represented in blue above and, when placed at $\mathbf{p}$, they are tangent to the surface. They span the tangent space of $\mathcal{H}$ at $\mathbf{p}$. Note that they are independent because the curves $\mathbf{x}, \mathbf{y}$ intersect transversally at $\mathbf{p}$.

Applying the chain rule to $h(\mathbf{x}(t)) = c$ at $t = r$ we get

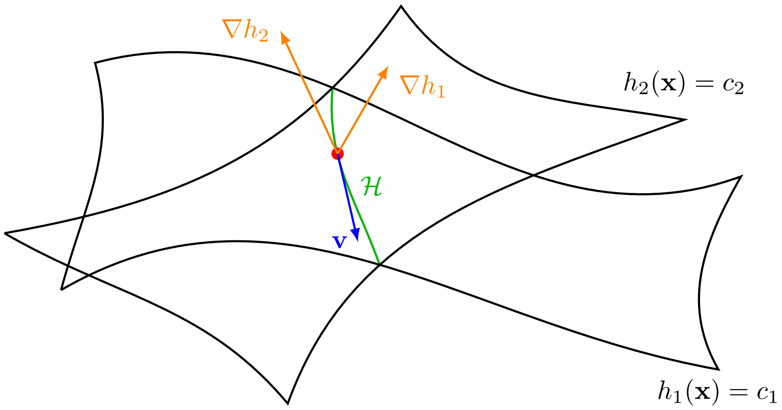
$$\begin{aligned} h(\mathbf{x}(t)) = c &\implies Dh(\mathbf{x}(r)) = 0 \\ &\implies Dh(\mathbf{x}(r))D\mathbf{x}(r) = 0 \\ &\implies \nabla^T h(\mathbf{p})\mathbf{v} = 0, \end{aligned}$$

and likewise, we have $\nabla^T h(\mathbf{p})\mathbf{w} = 0$, which shows that the vector ∇h at $\mathbf{p}$ (depicted in orange) is orthogonal to both tangent vectors $\mathbf{v}$ and $\mathbf{w}$ at $\mathbf{p}$.

Now consider instead a function $\mathbf{h}\colon \mathbb{R}^3 \to \mathbb{R}^2$, and consider a $\mathbf{c}$-level set of $\mathbf{h}$,

$$\mathcal{H} = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{h}(\mathbf{x}) = \mathbf{c}\},$$

which corresponds to the intersection of two surfaces in $\mathbb{R}^3$, which is generically a curve (depicted below in green).



Let $\mathbf{x}\colon \mathbb{R} \to \mathbb{R}^3$ be a parametrisation of this curve. Now let $\mathbf{p} = \mathbf{x}(r)$ be any point on $\mathcal{H}$ (marked by the red dot), and let $\mathbf{v}$ denote the derivative of $\mathbf{x}$ at r, so that $\mathbf{v} = D\mathbf{x}(r)$ (marked in blue). As $\mathbf{x}(t)$ is the the $\mathbf{c}$-level set of $\mathbf{h}$, we have $\mathbf{h}(\mathbf{x}(t)) = \mathbf{c}$, for all t. Applying the chain rule to this we obtain,

$$\begin{aligned} \mathbf{h}(\mathbf{x}(t)) = \mathbf{c} &\implies D\mathbf{h}(\mathbf{x}(r)) = \mathbf{0} \\ &\implies D\mathbf{h}(\mathbf{x}(r))D\mathbf{x}(r) = \mathbf{0} \\ &\implies \begin{pmatrix} \nabla h_1^T(\mathbf{p})\mathbf{v} \\ \nabla h_2^T(\mathbf{p})\mathbf{v} \end{pmatrix} = \mathbf{0}, \end{aligned}$$

which shows that both (orange) vectors ∇h_1, ∇h_2 at $\mathbf{p}$ are orthogonal to the tangent vector v at $\mathbf{p}$. The vectors ∇h_1, ∇h_2 span the normal space at $\mathbf{p}$.

Example: tangent and normal spaces

In Section 5.1, we considered the following optimisation problem:

$$\begin{aligned} \text{minimise} \quad & f(\mathbf{x}) = x_1^2 + x_2^2 \\ \text{subject to} \quad & h(\mathbf{x}) = (x_1 - 0.3)^2 - 2x_2 = 3. \end{aligned}$$

Note that

$$Dh(\mathbf{x}) = \begin{pmatrix} 2(x_1 - 0.3) & -2 \end{pmatrix}.$$

Let $\mathcal{H}$ denote the 3-level set of h, so that

$$\mathcal{H} = \{(x_1, x_2) : h(\mathbf{x}) = 3\}.$$

Let $\mathbf{p} = \begin{pmatrix} p_1 & p_2 \end{pmatrix}^T \in \mathcal{H}$. We have $Dh(\mathbf{p}) = \begin{pmatrix} 2(p_1 - 0.3) & -2 \end{pmatrix}$, and so

$$\begin{aligned} T\mathcal{H}(\mathbf{p}) = \operatorname{Ker}(Dh(\mathbf{p})) &= \{(x_1, x_2) \in \mathbb{R}^2 : Dh(\mathbf{p})\mathbf{x} = 0\} \\ &= \{(x_1, x_2) \in \mathbb{R}^2 : 2(p_1 - 0.3)x_1 - 2x_2 = 0\} \\ &= \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = (p_1 - 0.3)x_1\} \\ &= \{(x_1, (p_1 - 0.3)x_1) \in \mathbb{R}^2 : x_1 \in \mathbb{R}\} \\ &= \operatorname{Sp}((1, p_1 - 0.3)), \end{aligned}$$

which is the line in $\mathbb{R}^2$ passing through the origin with slope $p_1 - 0.3$.

The normal space of $\mathcal{H}$ at $\mathbf{p}$ is

$$N\mathcal{H}(\mathbf{p}) = \operatorname{Im}(Dh^T(\mathbf{p})) = \operatorname{Sp}((2(p_1 - 0.3), -2)) = \operatorname{Sp}((p_1 - 0.3, -1)),$$

which is the line in $\mathbb{R}^2$ passing through the origin and the point $(p_1 - 0.3, -1)$.

A point $\mathbf{p}$ is called a **regular point** of $\mathbf{h}$ if the set

$$\{\nabla h_1(\mathbf{p}), \nabla h_2(\mathbf{p}), \dots, \nabla h_m(\mathbf{p})\}$$

is linearly independent.

Example: regular and non-regular points

Consider $h(\mathbf{x}) = (x_1 - 0.3)^2 - 2x_2 - 3$. If $\mathbf{p}$ were *not* a regular point of h, we would require $\{\nabla h(\mathbf{p})\}$ to be linearly dependent, and as a one-element set, this amounts to having $\nabla h(\mathbf{p}) = \mathbf{0}$. We have

$$\nabla h(\mathbf{x}) = \begin{pmatrix} 2(p_1 - 0.3) \\ -2 \end{pmatrix}.$$

which is never equal to $\mathbf{0}$. So all points are regular points of $h(\mathbf{x}) = (x_1 - 0.3)^2 - 2x_2 - 3$.

For $g(\mathbf{x}) = x_1^2 + x_2^2$, we have

$$\nabla g(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}.$$

This equals zero for $x_1 = x_2 = 0$, so the point $\mathbf{0}$ is a non-regular point of g.

For the vector function $\mathbf{h}\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\mathbf{h}(\mathbf{x}) = \begin{pmatrix} x_1 x_2 \\ x_1^2 + x_2^2 \end{pmatrix},$$

we have

$$\nabla h_1(\mathbf{x}) = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}, \quad \nabla h_2(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}.$$

The set $\{\nabla h_1(\mathbf{x}), \nabla h_2(\mathbf{x})\}$ is linearly dependent precisely when $\nabla h_1(\mathbf{x})$ is a scalar multiple of $\nabla h_2(\mathbf{x})$, i.e., when there is $\lambda \in \mathbb{R}$ such that $\nabla h_1(\mathbf{x}) = \lambda \nabla h_2(\mathbf{x})$. The equations $x_2 = 2\lambda x_1$ and $x_1 = 2\lambda x_2$ imply $|x_1| = |x_2|$, so all points of the form $(x, \pm x)$ with $x \in \mathbb{R}$ are non-regular points.