Problem 1: A best fit line

The collection of 5 data points $(-1,6)$, $(0,3)$, $(1,0)$, $(2,-3)$, $(3,-4)$ lies close to a line of negative slope; see Figure 1. We are going to compute that line.

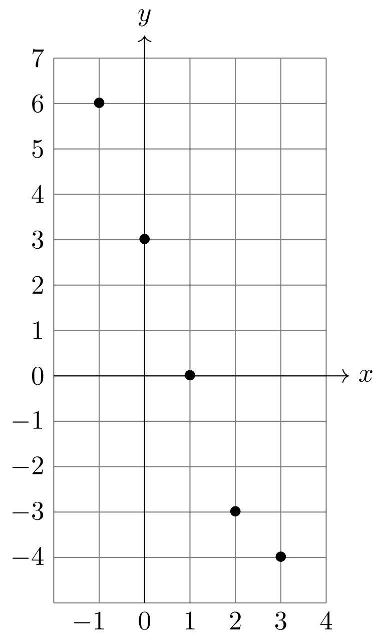

Figure 1: Five data points: $(-1,6)$, $(0,3)$, $(1,0)$, $(2,-3)$, $(3,-4)$.

Suppose the line of best fit (in the least squares sense) is written as $y = mx + b$.

(a) Write down explicit 5-vectors $\mathbf{X}$ and $\mathbf{Y}$ so that for the 5-vector $\mathbf{1}$ whose entries are all equal to 1, the projection of $\mathbf{Y}$ into the plane $V = \mathrm{span}(\mathbf{X}, \mathbf{1})$ in $\mathbf{R}^5$ is $m\mathbf{X} + b\mathbf{1}$.

(b) Compute an orthogonal basis of $V = \mathrm{span}(\mathbf{X}, \mathbf{1})$ having the form $\{\mathbf{1}, \mathbf{v}\}$ for a 5-vector $\mathbf{v}$, and find scalars t and s so that $\mathbf{Proj}_V(\mathbf{Y}) = t\mathbf{v} + s\mathbf{1}$.

(c) By expressing $\mathbf{v}$ from (b) as a linear combination of $\mathbf{X}$ and $\mathbf{1}$, use your answer to (b) to find m and b so that the equation $y = mx + b$ gives the line of best fit. (As a safety check on your computations, you may want to plot your line on the above figure to see that it is a good fit for the data.)

Solution:

(a) The vectors are $\mathbf{X} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \\ 3 \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} 6 \\ 3 \\ 0 \\ -3 \\ -4 \end{bmatrix}$. (Given such vectors, the "m" and "b" given in the equation of the line of best fit are precisely those values for which the magnitude of the "vector of errors," i.e., $\|\mathbf{Y} - (m\mathbf{X} + b\mathbf{1})\|$, is minimized. This is in turn just another way of saying that $m\mathbf{X} + b\mathbf{1}$ is the vector in $V = \mathrm{span}(\mathbf{X}, \mathbf{1})$ which is closest to $\mathbf{Y}$; by definition, this is the projection of $\mathbf{Y}$ into V.)

(Note: entries can be rearranged, provided that both $\mathbf{X}$ and $\mathbf{Y}$ are rearranged in the same way.)

(b) The vector $\mathbf{v}$ can be taken to be $\mathbf{X} - \mathbf{Proj_1}(\mathbf{X})$, by the properties of the orthogonal projection. Furthermore,

$$\mathbf{Proj_1}(\mathbf{X}) = \left(\frac{\mathbf{X}\cdot\mathbf{1}}{\mathbf{1}\cdot\mathbf{1}}\right)\mathbf{1} = \overline{x}\mathbf{1},$$

with $\overline{x}$ equal to the average of the entries x_i in $\mathbf{X}$. (This is true because the numerator $\mathbf{X}\cdot\mathbf{1}$ is the sum of the entries of $\mathbf{X}$, while the denominator $\mathbf{1}\cdot\mathbf{1}$ is the number of entries, or 5.) We have the average $\overline{x} = (1/5)(-1+0+1+2+3) = 1$, so $\mathbf{v} = \mathbf{X} - (1)\mathbf{1} = \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}$. (This $\mathbf{v}$ is what is called $\widehat{\mathbf{X}}$ in the course text.)

Now that we have an orthogonal basis $\{\mathbf{1}, \mathbf{v}\}$ for V, we may compute $\mathbf{Proj}_V(\mathbf{Y})$ (using the formula given in the Orthogonal Projection Theorem):

$$\mathbf{Proj}_V(\mathbf{Y}) = \mathbf{Proj_v}(\mathbf{Y}) + \mathbf{Proj_1}(\mathbf{Y}) = \left(\frac{\mathbf{Y}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v} + \left(\frac{\mathbf{Y}\cdot\mathbf{1}}{\mathbf{1}\cdot\mathbf{1}}\right)\mathbf{1} = \left(\frac{\mathbf{Y}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v} + \overline{y}\mathbf{1}.$$

We compute the dot products $\mathbf{Y}\cdot\mathbf{v} = -12 - 3 - 3 - 8 = -26$ and $\mathbf{v}\cdot\mathbf{v} = 4 + 1 + 0 + 1 + 4 = 10$; and also $\overline{y} = 2/5 = 0.4$. Thus, $\mathbf{Proj}_V(\mathbf{Y}) = -(2.6)\mathbf{v} + (0.4)\mathbf{1}$. (That is, $t = -2.6$ and $s = 0.4$.)

(c) Using the fact from (b) that $\mathbf{v} = \mathbf{X} - \mathbf{1}$, we may rewrite our computed linear combination for $\mathbf{Proj}_V(\mathbf{Y})$ as a combination of $\mathbf{X}$ and $\mathbf{1}$, with the final outcome being $m\mathbf{X} + b\mathbf{1}$ for the sought-after coefficients m and b:

$$\begin{aligned}
\mathbf{Proj}_V(\mathbf{Y}) &= -(2.6)\mathbf{v} + (0.4)\mathbf{1} \\
&= -(2.6)(\mathbf{X} - \mathbf{1}) + (0.4)\mathbf{1} \\
&= -(2.6)\mathbf{X} + (2.6)\mathbf{1} + (0.4)\mathbf{1} \\
&= -(2.6)\mathbf{X} + (2.6 + 0.4)\mathbf{1} = -(2.6)\mathbf{X} + (3)\mathbf{1}.
\end{aligned}$$

Hence, the line of best fit is $y = -(2.6)x + 3$. (That is, $m = -2.6$ and $b = 3$.) A plot of the data shows it is reasonably close to that line.

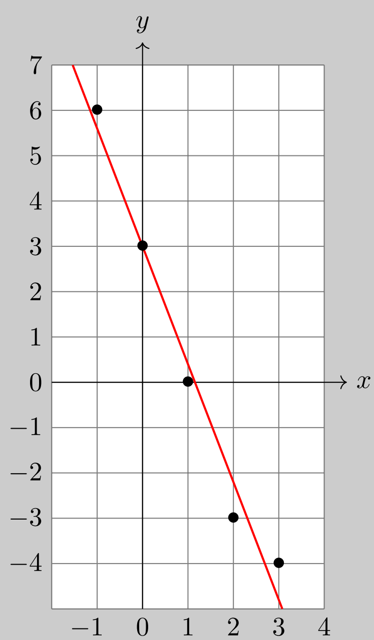

Figure 2: Line of best fit for given data.

Problem 2: A linear mathematical model via closest vector and dot products

A researcher measures the basal metabolic rate[1], height, and weight for 100 people and expresses the result as vectors:

$$\mathbf{B}, \mathbf{W}, \mathbf{H} \in \mathbf{R}^{100}$$

Here, the ith entry of $\mathbf{H}$ is the height of the ith person in inches, and similarly for $\mathbf{B}$ (basal metabolic rate in kilocalories per day) and $\mathbf{W}$ (weight in pounds).

The researcher would like to work out a linear formula to estimate the basal metabolic rate in terms of height and weight. In mathematical terms, she would like to find $a, b \in \mathbf{R}$ for which

$$a\mathbf{H} + b\mathbf{W} \text{ is as close to } \mathbf{B} \text{ as possible.}$$

(a) Suppose that the vectors were in $\mathbf{R}^3$ rather than $\mathbf{R}^{100}$. Draw a picture to explain why the a, b we are looking for must satisfy

$$\mathbf{B} - (a\mathbf{H} + b\mathbf{W}) \text{ is perpendicular to } \mathbf{H}, \mathbf{W}.$$

(We know this is true in $\mathbf{R}^{100}$ by the Orthogonal Projection Theorem; the point is to understand it intuitively with a picture in $\mathbf{R}^3$.)

(b) Use the orthogonality as discussed in (a) (which must hold for 100-vectors) and dot products to write down a system of linear equations for a, b (whose coefficients involve dot products among 100-vectors).

(c) The researcher computes that $\mathbf{H} \cdot \mathbf{H} = 1/2, \mathbf{W} \cdot \mathbf{W} = 3$ and $\mathbf{H} \cdot \mathbf{W} = 3/2$; also $\mathbf{B} \cdot \mathbf{W} = 3$ and $\mathbf{B} \cdot \mathbf{H} = 2$. Using the vanishing of dot products against $\mathbf{H}$ and $\mathbf{W}$ arising from (a), solve for a and b. (In the real world, such dot products would usually be "ugly" numbers; we made them clean, as we do on exams, so the answer comes out cleanly without using a calculator.)

Observe that the solution did not require knowledge of the 100-element vectors– just knowledge about their dot products! (Of course, to *compute* those dot products one has to know the 100-vectors, but the point is that the *only* way the knowledge of the 100-vectors is relevant is solely to compute those dot products.)

Solution:

(a) In order for a point in the plane (or line) through the origin spanned by $\mathbf{H}$ and $\mathbf{W}$ to be as close to the point $\mathbf{B}$ as possible, such a point should be the foot of the perpendicular drawn from $\mathbf{B}$ to this plane (or line) through the origin — since geometry tells us that the shortest distance from a point to a plane (or line) through the origin is the perpendicular distance. Thus, we want a and b so that the displacement vector $\mathbf{B} - (a\mathbf{H} + b\mathbf{W})$ is perpendicular to the plane (or line) through the origin spanned by $\mathbf{H}$ and $\mathbf{W}$; i.e., is perpendicular to both $\mathbf{H}$ and $\mathbf{W}$.

(b) $\mathbf{B} - (a\mathbf{H} + b\mathbf{W})$ is perpendicular to $\mathbf{H}, \mathbf{W}$ precisely when

$$(\mathbf{B} - (a\mathbf{H} + b\mathbf{W})) \cdot \mathbf{H} = 0, \;\; (\mathbf{B} - (a\mathbf{H} + b\mathbf{W})) \cdot \mathbf{W} = 0,$$

which we can rewrite (by properties of dot products) as

$$(\mathbf{H} \cdot \mathbf{H})a + (\mathbf{W} \cdot \mathbf{H})b = \mathbf{B} \cdot \mathbf{H}, \;\; (\mathbf{H} \cdot \mathbf{W})a + (\mathbf{W} \cdot \mathbf{W})b = \mathbf{B} \cdot \mathbf{W}.$$

(c) The system of equations is

$$(1/2)a + (3/2)b = 2, \;\; (3/2)a + 3b = 3.$$

This is solved by the method from high school algebra to give $a = -2$ and $b = 2$.

[1] rate at which the body uses energy, measured in kilocalories per day, if the person is at rest

Problem 3: Level sets of multivariable functions

(a) Describe and sketch the level sets of $\ln(y - x^2)$ on the region where $y > x^2$, relating each level set to the parabola $y = x^2$.

(b) Describe and sketch the level sets of $\cos(x^2 + y^2)$.

(c) Express the surface graph of $f(x, y) = x^2 + y^2$ in $\mathbf{R}^3$ as a level set of a function $h(x, y, z)$.

(d) (Extra) By using polar coordinates, describe the part of the graph of $f(x, y) = x^2 + y^2$ from (c) that lies over a line in the xy-plane through the origin, and use that to sketch the actual surface graph. (Don't "cheat" by looking on a computer; the point is to learn for yourself how to use restriction over well-chosen lower-dimensional subspaces, such as lines through the origin in $\mathbf{R}^2$, to build up a mental model of what happens over the entire domain.)

Solution:

(a) The region $y > x^2$ is the part of $\mathbf{R}^2$ "above" the parabola $y = x^2$ (away from the shaded region in Figure 3). The level sets correspond to the condition $\ln(y - x^2) = c$, or in other words $y - x^2 = e^c$, which is to say $y = x^2 + e^c$ with any $c \in \mathbf{R}$. The number e^c can be any positive value, so the level sets are the parabolas $y = x^2 + a$ for $a > 0$; this is the collection of "nested" parabolas obtained by translating $y = x^2$ upwards; some of these are shown (for various $c = \ln(a)$) in different colors in Figure 3.

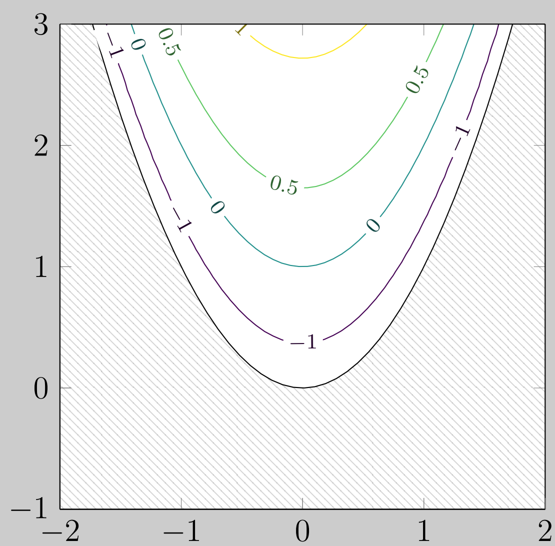

Figure 3: The level sets $\ln(y - x^2) = c$ for various c.

(b) The level sets are $\cos(x^2 + y^2) = c$ for $-1 \le c \le 1$. For each c, the corresponding level set is an infinite collection of circles centered at the origin (including the origin as a point when $c = 1$), with $x^2 + y^2$ varying through the values which cos carries to c (and the radii of those circles being the square roots of the values for $x^2 + y^2$ on the level set).

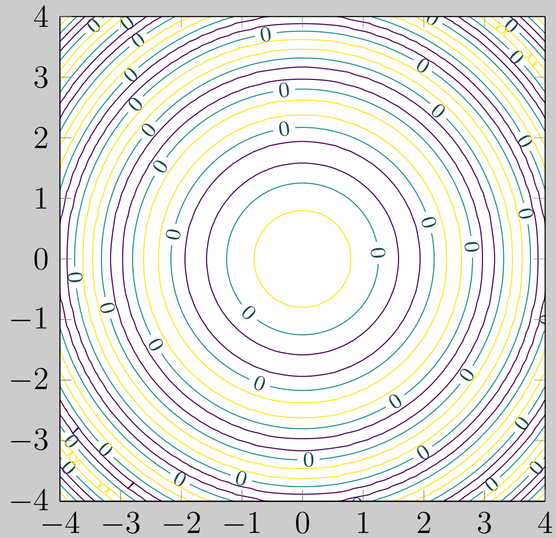

Figure 4: The level sets of $\cos(x^2+y^2)$ are nested circles (those of a fixed color).

(c) The graph is the region $z = x^2 + y^2$ in $\mathbf{R}^3$, so this is the level set $z - (x^2 + y^2) = 0$ for the function $h(x, y, z) = z - (x^2 + y^2)$.

(d) The part of the graph over a line in the xy-plane making an angle θ with the x-axis is given by substituting into $f(x, y)$ the expressions $x = \pm r\cos\theta$ and $y = \pm r\sin\theta$ (using a common sign: we need to allow the sign to account for both halves of the line, on either side of the origin). This yields $(\pm r\cos\theta)^2 + (\pm r\sin\theta)^2 = r^2$ over the two points at a distance r from the origin along that line in both directions, so it is the graph of the parabola $g(r) = r^2$ as we let $r \geq 0$ vary.

This parabola is the outcome regardless of the angle θ! Hence, it says that the surface graph is obtained by spinning that parabola in the xz-plane ($\theta = 0$) around the z-axis (to sweep out all possible angles θ), so it is a "cone-like" shape but with round bottom at the origin (rather than a sharp tip). Figure 5 shows the surface graph and the parabolic slices of the surface graph (in red, green, and blue) for several angles θ. If you zoom in, you'll see that these parabolas all meet at a common point, which is the bottom of the rounded tip. (It may look as if the red and green parabolas have their minimum points somewhere else, but that is an optical illusion caused by a rendering a 3-dimensional situation on a flat piece of paper or computer screen.)

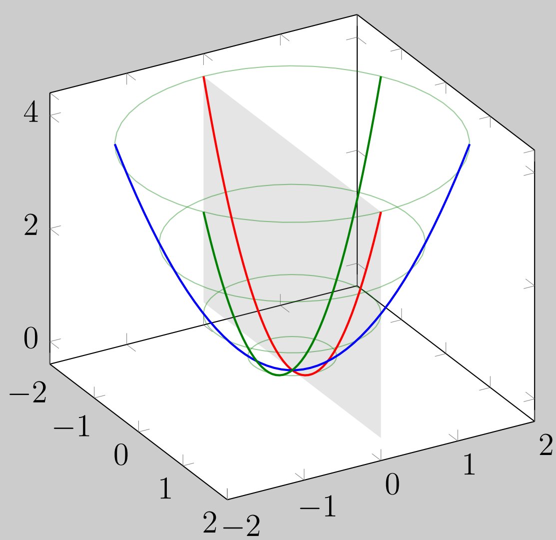

Figure 5: The part of $z = x^2 + y^2$ lying over a line (through the origin) in the xy plane is always a parabola.

Problem 4: Computations with vector-valued functions

For the functions $\mathbf{f} : \mathbf{R}^n \to \mathbf{R}^m$ and $\mathbf{g} : \mathbf{R}^m \to \mathbf{R}^p$ below, compute $\mathbf{g} \circ \mathbf{f} : \mathbf{R}^n \to \mathbf{R}^p$ by working out its component functions; in each part also state the values of n, m, and p.

(a) $\mathbf{f}(x, y) = (e^x \cos(y), e^x \sin(y))$, $\mathbf{g}(v, w) = (v^2 - w^2, 2vw)$

(b) $\mathbf{f}(x, y) = (x^2 - y^2, 2xy)$, $\mathbf{g}(v, w) = (e^v \cos(w), e^v \sin(w))$

(c) $\mathbf{f}(t) = (1 - t^2, 2t, 1 + t^2)$, $\mathbf{g}(x, y, z) = x^2 + y^2 - z^2$

Solution:

(a) Here $\mathbf{f} : \mathbf{R}^2 \to \mathbf{R}^2$, so $n = m = 2$; and $\mathbf{g} : \mathbf{R}^2 \to \mathbf{R}^2$, so $p = 2$. For $\mathbf{g} \circ \mathbf{f}$, we have

$$\begin{aligned}
\mathbf{g}(\mathbf{f}(x, y)) = \mathbf{g}(e^x \cos(y), e^x \sin(y)) &= ((e^x \cos(y))^2 - (e^x \sin(y))^2, 2e^x \cos(y) e^x \sin(y)) \\
&= (e^{2x}(\cos^2(y) - \sin^2(y)), e^{2x}(2\cos(y)\sin(y))) \\
&= (e^{2x}\cos(2y), e^{2x}\sin(2y)).
\end{aligned}$$

(It is unnecessary to make the final observation via double-angle formulas, but is geometrically informative.)

(b) Here the roles of $\mathbf{f}$, $\mathbf{g}$ are simply interchanged from those in (a); so again $n = m = p = 2$. However, for $\mathbf{g} \circ \mathbf{f}$ we now have $\mathbf{g}(\mathbf{f}(x, y)) = \mathbf{g}(x^2 - y^2, 2xy) = (e^{x^2 - y^2}\cos(2xy), e^{x^2 - y^2}\sin(2xy))$. This is very different from (a)!

(c) Here $\mathbf{f} : \mathbf{R} \to \mathbf{R}^3$, so $n = 1$ and $m = 3$; meanwhile $\mathbf{g} : \mathbf{R}^3 \to \mathbf{R}$, so $p = 1$. For $\mathbf{g} \circ \mathbf{f}$ (a scalar-valued function of one variable), we have $\mathbf{g}(\mathbf{f}(t)) = \mathbf{g}(1 - t^2, 2t, 1 + t^2) = (1 - t^2)^2 + (2t)^2 - (1 + t^2)^2 = 1 - 2t^2 + t^4 + 4t^2 - (1 + 2t^2 + t^4) = 0$.