Functions of continuous random variables[1]

1 $E(g(X))$

The purpose of this section is to prove that when $X\colon \Omega \to \mathbb{R}^m$ has a density f and $g\colon \mathbb{R}^m \to \mathbb{R}$ we have $Eg(X) = \int g(x)f(x)dx$.[2] We will assume that the sets $\{x \in \mathbb{R}^m\colon g(x) > t\}$ are Borel.

For discrete variables we proved that for discrete nonnegative X with integer values we have $E[X] = \sum_{n=0}^{\infty} P(X > n)$. We also showed that for any function $g\colon \mathbb{R} \to \mathbb{R}$ we have $Eg(X) = \sum_{x \in X(\Omega)} g(x)p(x)$ with p density of X.

In the continuous case we are going to use again the "reduction formula" reported in Baldi *Richiamo 3.11.* In the special case of a (possibly unbounded) rectangle the reduction says that you can integrate with respect to one variable at a time:

$$\int_{[a,b]\times[c,d]} f(x,y)dxdy = \int_a^b \left[\int_c^d f(x,y)dy\right] dx = \int_c^d \left[\int_a^b f(x,y)dx\right] dy$$

We may more appropriately call this the *Fubini Theorem.* We start by establishing the following result, useful in itself:

Lemma 1. *For $X\colon \Omega \to \mathbb{R}$ with density f*

$$EX = \int_0^\infty P(X > x)dx - \int_0^\infty P(X < -x)dx.$$

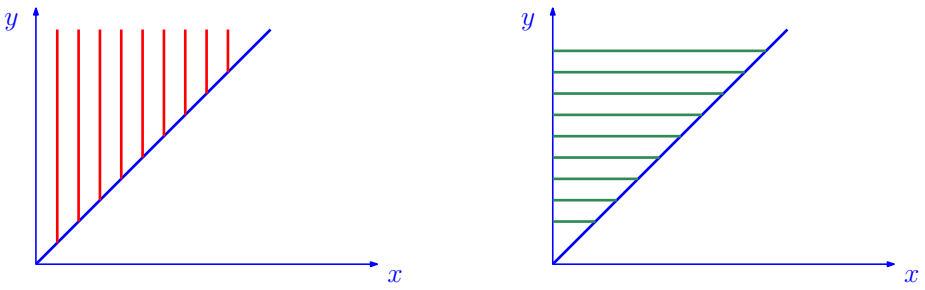
Proof. Assume first $X \geq 0$, in which case the result says

$$EX = \int_0^\infty P(X > x)dx.$$

The set $\{(x,y)\colon x \geq 0, y \geq x\}$ (left panel below) may also be written as $\{(x,y)\colon y \geq 0, 0 \leq x \leq y\}$ (right panel)

[1] S Modica 2025. Ref. Bertsekas-Tsitsiklis, *Introduction to Probability 2nd Edition* and Stroock, *A Concise Introduction to Analysis*.

[2] Recall that X having density f means that $f \geq 0$ and $P(X \in B) = \int_B f(x)dx$ (Riemann integral) for Borel sets B in $\mathbb{R}^n$.



therefore using Fubini we get

$$\begin{aligned}\int_0^\infty P(X>x)dx &= \int_0^\infty \left(\int_x^\infty f(y)dy\right)dx \\ &= \int_0^\infty \left(\int_0^y f(y)dx\right)dy \\ &= \int_0^\infty yf(y)dy = EX.\end{aligned}$$

Next consider general real-valued variables. That $\int_0^\infty P(X>x)dx = \int_0^\infty yf(y)dy$ is still true, and one shows similarly that $\int_0^\infty P(X<-x)dx = -\int_{-\infty}^0 yf(y)dy$;[3] putting the two together gives the result. □

Example (A merchant's optimization problem [4]). A merchant sells a product with a markup μ over buying cost c, that is at unit price $(1+\mu)c$ (a markup e.g. of 20% is $\mu = 0.2$). Monthly demand D is assumed exponential with parameter λ, and he restocks every first day of the month. With leftover stock $\overline{s}$ he has to choose new inventory level $s \geq \overline{s}$ for the month's trading, to maximize expected profit $E\pi(s)$. The quantity sold is $Q(s) = \min\{D, s\}$ so the profit is $(1+\mu)cQ(s) - c(s-\overline{s})$. Here c is irrelevant for maximization so we define $\pi(s) = (1+\mu)Q(s) - (s-\overline{s})$ (it is profit per unit cost).

To compute expected profit we need EQ; since $Q \geq 0$ we want $P(Q>x)$ to apply the above lemma. And $P(Q>x) = P(D>x) = e^{-\lambda x}$ if $x < s$ and zero otherwise. Therefore $EQ(s) = \int_0^s e^{-\lambda x}dx = \lambda^{-1}(1-e^{-\lambda s})$; note that $dEQ(s)/ds = e^{-\lambda s}$. From this we get

$$\begin{aligned}E\pi(s) &= (1+\mu)EQ(s) - (s-\overline{s}) \\ \frac{d}{ds}E\pi(s) &= (1+\mu)e^{-\lambda s} - 1\end{aligned}$$

[3] Indeed

$$\int_0^\infty P(X<-x)dx = \int_0^\infty \left[\int_{-\infty}^{-x} f(y)dy\right]dx = \int_{-\infty}^0 \left[\int_0^{-y} f(y)dx\right]dy = -\int_{-\infty}^0 yf(y)dy.$$

[4] Adapted from Letta, *Probabilità elementare.*

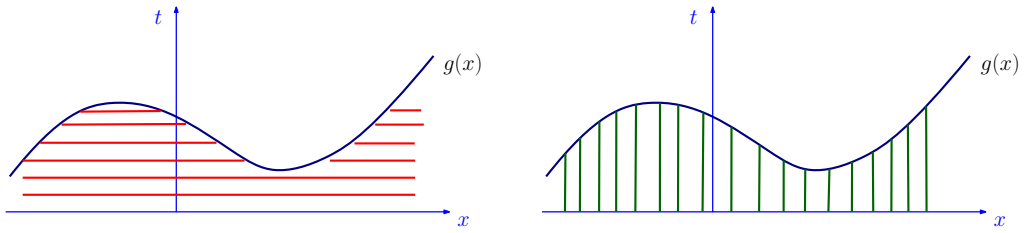
so $E\pi(s)$ (obviously concave) has a maximum at $s^* = \lambda^{-1}\ln(1+\mu) \approx \mu ED$. Therefore the solution is: if $s^* > \overline{s}$ then choose s^*; otherwise stay with $\overline{s}$. Observe that when $s^* > \overline{s}$ (so the merchant raises his inventory) one has $P(D > s^*) = e^{-\lambda s^*} = 1/(1+\mu)$, which is pretty close to 1: he chooses to run out of stock with high probability. The reason why s^* is so low is that with D exponential the marginal gain from stocking $(1+\mu)e^{-\lambda s}$ goes down fast, against a marginal cost of 1 which remains relatively high - so you choose to stay short on inventories.

Now given $X\colon \Omega \to \mathbb{R}^m$ and $g\colon \mathbb{R}^m \to \mathbb{R}$ consider a real-valued variable $g(X)$ such that the (Riemann) integral $\int_A f(x)dx$ over $A = \{x \in \mathbb{R}^m\colon g(x) > t\}$ exists for all t.

Proposition. *For $g(X)\colon \mathbb{R}^m \to \mathbb{R}$*

$$Eg(X) = \int_{\mathbb{R}^m} g(x)f(x)dx.$$

Proof. As before assume to start with that $g(X) \geq 0$. In this case we exploit the fact that the region below the graph of g, that is $\{(x,t)\colon x \in \mathbb{R}^m, 0 \leq t < g(x)\}$, is equal to the set $\{(x,t)\colon 0 \leq t < \infty, g(x) > t\}$, see the next picture for the case where X is real valued:



This, via Fubini Theorem, gives

$$\begin{aligned} Eg(X) &= \int_0^\infty P(g(X) > t)\,dt \quad \textit{from Lemma 1} \\ &= \int_0^\infty \left(\int_{\{x\colon\, g(x)>t\}} f(x)dx \right) dt \\ &= \int_{\mathbb{R}^m} \left(\int_{\{t\colon\, 0\leq t\leq g(x)\}} f(x)dt \right) dx \quad \textit{interchanging order as in the picture} \\ &= \int_{\mathbb{R}^m} g(x)f(x)dx. \end{aligned}$$

For the general case let $g^+ = \max\{g, 0\}$ and $g^- = \max\{-g, 0\}$ so that $g = g^+ - g^-$. Lemma 1 gives

$$Eg(X) = \int_0^\infty P(g(X) > t)dt - \int_0^\infty P(g(X) < -t)dt.$$

Now using the fact that for $t \geq 0$ we have $g(x) > t \iff g^+(x) > t$ we get $\int_0^\infty P(g(X) > t)dt = \int_0^\infty P(g^+(X) > t)dt = \int_{\mathbb{R}^m} g^+(x)f(x)dx$. And by a symmetrical argument - and using

$g < -t \iff g^- > t$ - we obtain $\int_0^\infty P(g(X) < -t)dt = \int_{\mathbb{R}^m} g^-(x)f(x)dx$. Therefore

$$\begin{aligned} Eg(X) &= \int_{\mathbb{R}^m} g^+(x)f(x)dx - \int_{\mathbb{R}^m} g^-(x)f(x)dx = \int_{\mathbb{R}^m} \left[g^+(x) - g^-(x)\right] f(x)dx \\ &= \int_{\mathbb{R}^m} g(x)f(x)dx. \end{aligned}$$ □

Remark. Remember that this is the result which delivers linearity of the expectation operator (as in **Baldi Prop. 3.35**). For example, setting $g(x, y) = x + y$ we derive

$$\begin{aligned} E(X+Y) &= \int_{\mathbb{R}^2} (x+y)f(x,y)dxdy = \int_{\mathbb{R}^2} xf(x,y)dxdy + \int_{\mathbb{R}^2} yf(x,y)dxdy \\ &= \int_{\mathbb{R}} xdx \int_{\mathbb{R}} f(x,y)dy + \int_{\mathbb{R}} ydy \int_{\mathbb{R}} f(x,y)dx \quad (Fubini) \\ &= \int_{\mathbb{R}} xf_X(x)dx + \int_{\mathbb{R}} yf_Y(y)dy = EX + EY. \end{aligned}$$

Of course it is still true that if $X \geq 0$ then $EX \geq 0$. The following also remains true as in the discrete case (Baldi Prop. 3.36) but the proof is more subtle:

Proposition. *Suppose $X \geq 0$. If $EX = 0$ then $P(X = 0) = 1$.*

Proof. Let $A_n = \{X \geq 1/n\}$. Then $0 = EX \geq \int_{A_n} xf(x)dx \geq P(A_n)/n$, which implies $P(A_n) = 0$. Thus $P(X < 1/n) = 1$ for all n; since $\{X = 0\} = \cap_n \{X < 1/n\}$ (a decreasing sequence) then continuity of probability gives $P(X = 0) = \lim_{n\to\infty} P(X < 1/n) = 1$. □

Now we can talk of Variance (with example from uniform), higher order moments, and observe that

Proposition. *If X and Y are independent then $EXY = EX \cdot EY$.*

Proof. Recall that independence for continuous variables is the property $f(x, y) = f_X(x)f_Y(y)$, so integrating we get (using reduction)

$$EXY = \int_{\mathbb{R}^2} xyf(x,y)dxdy = \int_{\mathbb{R}} xf_X(x)dx \cdot \int_{\mathbb{R}} yf_Y(y)dy.$$ □

It is still true that $\text{cov}(X, Y) = EXY - EXEY$, and therefore that independent variables have zero covariance also in the continuous model.

2 Density of $\varphi(X)$

Given a random variable X with values in $\mathbb{R}^m$ and density f and a function $\varphi\colon \mathbb{R}^m \to \mathbb{R}^m$, the purpose of this section is to find the density of $Y = \varphi(X)$. To do this we need to know

how to change variables in multiple integrals. As a bonus we obtain the standard normal density (univariate and bivariate).

2.1 Change of variables in multiple integrals

Recall how this is done for functions of one variable: start with an integral $\int h(x)dx$; set $x = \psi(u)$, monotone with $\psi' \neq 0$, and get

$$\int_{\psi(a)}^{\psi(b)} h(x)dx = \int_a^b h(\psi(u))\psi'(u)du.$$

The range of integration is clear: if u goes from a to b then $x = \psi(u)$ goes from $\psi(a)$ to $\psi(b)$.[5]

Now take an open set $U \subseteq \mathbb{R}^m$ and let $\psi\colon U \to \mathbb{R}^m$ be a diffeomorphism, that is a one to one, smooth function (some examples of diffeomorphisms are contained in Appendix).[6] Denote by $D\psi(u)$ the Jacobian of ψ, that is the $m \times m$ matrix of partial derivatives of ψ, with $D\psi(u)_{ij} = \partial\psi_i(u)/\partial u_j$.

Example 1. If $A = (a_{ij})$ is an $m \times m$ matrix and $u \in \mathbb{R}^m$ we have $DAu = A$. Indeed in this case the row $(Au)_i = \sum_j a_{ij}u_j$, which implies the result directly by differentiating with respect to $u_j,\ j = 1, \ldots, m$. This generalizes the one variable case $d(ax)/dx = a$ for $x \in \mathbb{R}$.

The following is a (hard) theorem in Analysis that we are not in a position to prove. For a diffeomorphism ψ the matrix $D\psi(u)$ is invertible, so $\det(D\psi(u)) \neq 0$, and the change of variable formula in $\mathbb{R}^m$ is now, with $x = \psi(u)$,

$$\int_{\psi(U)} h(x)dx = \int_U h(\psi(u))\,|\det(D\psi(u))|\,du. \tag{1}$$

2.1.1 The Standard Normal Density

The next example yields the most important density of all probability theory, that is *the Standard Normal density.*

Example 2 (Polar coordinates and the standard normal density). Consider the integral $\int_{\mathbb{R}^2} e^{-(x^2+y^2)/2}dxdy$. We use the letters (r, t) for $u \in \mathbb{R}^2$. Define the change of variables by

$$(x, y) = \psi(r, t) = (r\cos t, r\sin t)$$

which is a diffeomorphism from $U = (0, \infty) \times (0, 2\pi)$ to $\psi(U) = \mathbb{R}^2 \setminus \{(x, 0)\colon x \geq 0\}$ (the horizontal positive half line is removed because $t = 0$ is excluded, but this does not matter

[5]The formula follows immediately by observing that, with H primitive of h, the primitive of $h(\psi(u))\psi'(u)$ is $H(\psi(u))$. Thus RHS $= H(\psi(b)) - H(\psi(a)) =$ LHS.

[6]Let $B(x_0, r) = \{x \in \mathbb{R}^m\colon\ |x - x_0| < r\}$ (a ball around x_0); a set $U \subseteq \mathbb{R}^m$ is *open* if for any $x \in U$ there is r such that $B(x, r) \subseteq U$.

for integration because it is a negligible set). Now

$$\det\left(D\psi(r,t)\right) = \det\begin{pmatrix}\cos t & -r\sin t\\ \sin t & r\cos t\end{pmatrix} = r$$

so from equation (1) we get

$$\begin{aligned}\int_{\mathbb{R}^2} e^{-(x^2+y^2)/2}dxdy &= \int_{r=0}^{\infty}\int_{\theta=0}^{2\pi} re^{-r^2/2}drdt\\ &= \int_{\theta=0}^{2\pi} dt\int_{r=0}^{\infty} re^{-r^2/2}dr \quad (\textit{Fubini})\\ &= 2\pi\int_{r=0}^{\infty} re^{-r^2/2}dr = 2\pi.\end{aligned}$$

From Fubini again we have

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)/2}dxdy = \int_{\mathbb{R}} e^{-x^2/2}dx\ \cdot \int_{\mathbb{R}} e^{-y^2/2}dy = \left(\int_{\mathbb{R}} e^{-x^2/2}dx\right)^2$$

which implies $\int_{\mathbb{R}} e^{-x^2/2}dx = \sqrt{2\pi}$. The conclusion is that

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

is a density function (it is nonnegative and integrates to one). It is the *standard normal density,* the famous Gaussian curve that everybody "knows". Its graph has the shape of a bell. The Standard Normal random variable will generally be denoted by Z; the notation is $Z \sim N(0,1)$.

2.2 Density of $Y = \varphi(X)$

On some probability space $(\Omega, \mathcal{F}, P)$ there is a random variable $X\colon \Omega \to \mathbb{R}^m$ with density f. Recall that this means $P(X \in A) = \int_A f(x)dx$ for Borel sets A; recall also that by definition $\int_A f(x)dx = \int_{\mathbb{R}^m} 1_A(x)f(x)dx$ where 1_A is the indicator of A. We will omit the subscript in $\int_{\mathbb{R}^m}$.

With $\varphi\colon \mathbb{R}^m \to \mathbb{R}^m$ a diffeomorphism on $\mathbb{R}^m$, our aim is to determine the density of $Y = \varphi(X)$, say $g(y)$. Of course Y is a random variable from Ω to $\mathbb{R}^m$.

By definition g must satisfy $P(Y \in A) = \int_A g(y)dy = \int 1_A(y)g(y)dy$. But

$$P(Y \in A) = P(\varphi(X) \in A) = P(X \in \varphi^{-1}(A)) = \int_{\varphi^{-1}(A)} f(x)dx = \int 1_A(\varphi(x))f(x)dx,$$

the last equality since $1_{\varphi^{-1}(A)}(x) = 1_A(\varphi(x))$ (which is 1 if $\varphi(x) \in A$). So the property which

g must satisfy is

$$\int 1_A(y)g(y)dy = \int 1_A(\varphi(x))f(x)dx. \tag{2}$$

Now we apply (1) to the second member, setting $x = \varphi^{-1}(y)$; we get

$$\int 1_A(\varphi(x))f(x)dx = \int 1_A(y)f(\varphi^{-1}(y))\left|\det(D\varphi^{-1}(y))\right| dy,$$

and this gives what we were after:

$$g(y) = f(\varphi^{-1}(y))\left|\det(D\varphi^{-1}(y))\right|. \tag{3}$$

Example 3 (Example 3.28 p.121 from Baldi). Linear case $Y = AX + b$, that is $\varphi(x) = Ax + b$ for invertible A. Observe that $\varphi^{-1}(y) = A^{-1}(y - b)$ so $D\varphi^{-1}(y) = A^{-1}$; from $\det A^{-1} = 1/\det A$ we obtain $\left|\det(D\varphi^{-1}(y))\right| = |\det A|^{-1}$, and therefore

$$g(y) = |\det A|^{-1} f(A^{-1}(y - b))$$

which generalizes example 3.7 p. 105 (where we had $g(y) = |a|^{-1} f(a^{-1}(y - b))$.

In $\mathbb{R}^2$ we may avoid subscripts. In the next example we use (X, Y) instead of $X = (X_1, X_2)$ and (U, V) in place of Y, so we shall have $(U, V) = \varphi(X, Y)$.

Example 4. Example 3.29 from Baldi. With $A = \left(\begin{smallmatrix} 1 & 0 \\ -1 & 1 \end{smallmatrix}\right)$ we have $\varphi(x, y) = (x, y - x) = A\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$. Since $A^{-1} = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ it is $\varphi^{-1}(u, v) = A^{-1}\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right) = \left(\begin{smallmatrix} u \\ u+v \end{smallmatrix}\right)$; also $\det A^{-1} = \det\left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right) = 1$ so that the density $g(u, v)$ is given by

$$\begin{aligned} g(u, v) &= f\left(\varphi^{-1}(u, v)\right) = f(u, u + v) = \lambda^2 e^{-\lambda(u+v)} \text{ for } u, v > 0 \\ &= \lambda e^{-\lambda u} \cdot \lambda e^{-\lambda v} \end{aligned}$$

which implies that U and V that is X and $Y - X$ are independent.

Example 5 (Form Baldi exercise 3.22). Let (X, Y) have joint density

$$f(x, y) = \frac{4}{\theta^2}\, x\, y\, e^{-x^2/\theta} \quad 0 < y < x$$

(zero otherwise). We want the joint density of $(U, V) = \varphi(X, Y) = (X, X/Y)$; here $U = X$ and $V = X/Y$. We are after the density $g(u, v)$ of (U, V). Let us *keep in mind that* $0 < y < x$ means $u = x > 0$ and $v = x/y > 1$. To apply (3) we need $\varphi^{-1}(u, v)$. To find φ^{-1} we must

solve *in* (x, y) the system

$$\begin{cases} u = x \\ v = x/y \end{cases}$$

The first gives $x = u$; substituting into the second gives $y = u/v$; therefore

$$(x, y) = \varphi^{-1}(u, v) = (u, u/v)$$

(note that $\varphi^{-1} = \varphi$ here). Next

$$\det\left(D\varphi^{-1}(u, v)\right) = \det\begin{pmatrix} 1 & 0 \\ 1/v & -u/v^2 \end{pmatrix} = -u/v^2.$$

Since $f(\varphi^{-1}(u, v)) = f(u, u/v)$, (3) gives

$$g(u, v) = \frac{4}{\theta^2}\, u\, \frac{u}{v}\, e^{-u^2/\theta} \cdot \frac{u}{v^2} = \frac{4}{\theta^2}\, u^3\, e^{-u^2/\theta} \cdot \frac{1}{v^3} \quad u > 0, v > 1.$$

This also implies that $(U, V) = (X, X/Y)$ are independent, because $g(u, v) = g_1(u)g_2(v)$.

Example 6 (Example 3.26 from Baldi). This is here for two reasons: one is to get used to different names for variables; the other is that we can get g directly from (2). We start with an independent pair (W, T) with W exponential with $\lambda = 1/2$ and T uniform on $[0, 2\pi]$. We want the density of $(X, Y) = (\sqrt{W}\cos T, \sqrt{W}\sin T)$. First let $R = \sqrt{W}$; here $(x, y) = \varphi(r, t)$ with $\varphi(r, t) = (r\cos t, r\sin t)$ and we want $g(x, y)$.[7] Observe that since $R > r \iff W > r^2$ we have $P(R > r) = e^{-r^2/2}$, so that $P(R \le r) = 1 - e^{-r^2/2}$ (with density $re^{-r^2/2}$); therefore the joint density of (R, T) is

$$f(r, t) = \frac{1}{2\pi}\, r\, e^{-r^2/2} \quad r > 0, 0 \le t \le 2\pi.$$

Recall that the right member in equation (2) is $\int_{\varphi^{-1}(A)} f(x)dx$, and in our case $\varphi^{-1}(\mathbb{R}^2) = [0, \infty) \times [0, 2\pi)$. Therefore equation (2) becomes

$$\begin{aligned} \int_{\mathbb{R}^2} 1_A(x, y)g(x, y)dxdy &= \int_{r=0}^{\infty}\int_{t=0}^{2\pi} 1_A(r\cos t, r\sin t)f(r, t)drdt \\ &= \frac{1}{2\pi}\int_{t=0}^{2\pi}\int_{r=0}^{\infty} 1_A(r\cos t, r\sin t)re^{-r^2/2}drdt. \end{aligned}$$

But we know how to get this last integral from example 2: with the coordinate change

[7] To be sure: here the original X is (R, T) and Y is now (X, Y).

$(x, y) = \varphi(r, t) = (r\cos t, r\sin t)$ applied to the function $1_A(x, y)e^{-(x^2+y^2)/2}$ we obtain

$$\int_{\mathbb{R}^2} 1_A(x, y)e^{-(x^2+y^2)/2}dxdy = \int_{t=0}^{2\pi}\int_{r=0}^{\infty} 1_A(r\cos t, r\sin t)re^{-r^2/2}drdt.$$

Therefore the equation for g is $\int 1_A(x, y)g(x, y)dxdy = \frac{1}{2\pi}\int 1_A(x, y)e^{-(x^2+y^2)/2}dxdy$; hence we may take

$$g(x, y) = \frac{1}{2\pi}e^{-(x^2+y^2)/2}.$$

This is actually the density of the standard bivariate normal distribution.

Example 7. Example 3.30 from Baldi you do it on your own.

Example 8. Exercise 3.33(a) from Baldi.

Appendix: Elementary examples of diffeomorphisms and inversions

Here we see in some simple examples how to invert maps $f\colon \mathbb{R}^n \to \mathbb{R}^n$, mainly for $n = 2$. They are diffeomorphisms, that is one to one mappings with derivatives of any order.

For arbitrary n we use the notation $u = f(x),\ x, u \in \mathbb{R}^n$. The inverse of f is found by solving the equation $u = f(x)$ in x. It is a system of n equations in n unknowns. The solution gives x in terms of u, and the result gives the inverse function $x = f^{-1}(u)$.

Example 9. We recall the univariate case $n = 1$. If $u = f(x) = x^3$ we solve $u = x^3$ in x and find $x = u^{1/3}$; so $f^{-1}(u) = u^{1/3}$.

Example 10. Next the linear case: $f(x) = Ax$ with A invertible $n \times n$ matrix. In this case $f^{-1} = A^{-1}$, because solving $u = Ax$ in x gives $x = A^{-1}u$.

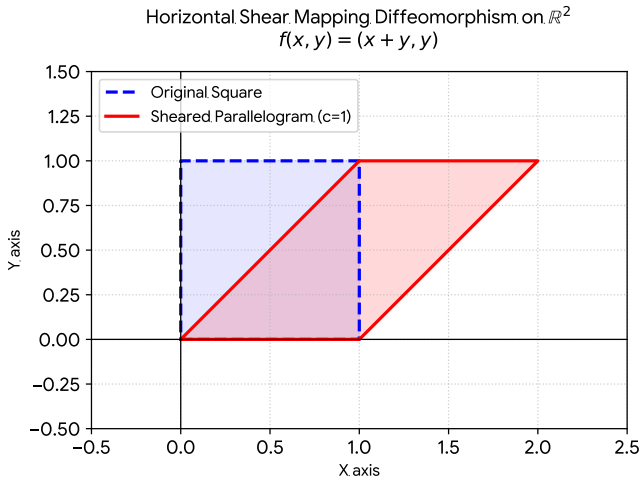
In the following examples we take $n = 2$, and use the notation

$$(u, v) = f(x, y).$$

Example 11 (The Shear Mapping). This is the linear map

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

Here $f(x, y) = A\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ with $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Thus $u = x + y$ and $v = y$. The y coordinate stays put and the x coordinate moves to the right by y. The action is nicely visible geometrically:



$A^{-1} = \left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right)$ so $A^{-1}\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right) = \left(\begin{smallmatrix} u-v \\ v \end{smallmatrix}\right)$ that is $x = u - v$ and $y = v$: the y coordinate had not moved and is still given by the identity map, and the x coordinate is moved back to the left.

Example 12. Next take

$$f(x, y) = \left(3y, 3x^7 - 6y\right)$$

The system to solve in (x, y) is now

$$\begin{cases} u = 3y \\ v = 3x^7 - 6y \end{cases}$$

The first gives $y = u/3$, and plugging this into the second gives $v = 3x^7 - 2u$ whence $x = ((2u + v)/3)^{1/7}$. Thus in this case

$$(x, y) = f^{-1}(u, v) = \left(\left(\frac{2u + v}{3} \right)^{1/7}, \frac{u}{3} \right).$$

Example 13. Consider lastly

$$f(x, y) = \left(\left(x^3 + y^3\right)^{1/3}, \left(x^3 - y^3\right)^{1/3} \right).$$

To invert this we have to solve in (x, y) the system

$$\begin{cases} u = \left(x^3 + y^3\right)^{1/3} \\ v = \left(x^3 - y^3\right)^{1/3} \end{cases}$$

Raise both members to the third power; then summing the two gives $u^3 + v^3 = 2x^3$, and subtracting the second from the first gives $u^3 - v^3 = 2y^3$. So we get

$$\begin{cases} x = \left(\frac{u^3 + v^3}{2} \right)^{1/3} \\ y = \left(\frac{u^3 - v^3}{2} \right)^{1/3} \end{cases}$$

which means

$$(x, y) = f^{-1}(u, v) = \left(\left(\frac{u^3 + v^3}{2} \right)^{1/3}, \left(\frac{u^3 - v^3}{2} \right)^{1/3} \right).$$