A Brief Primer on the Legendre Transformation

Steven J. Kilner[1], David L. Farnsworth[2]

[1]Department of Mathematics, Monroe Community College, Rochester, USA
[2]School of Mathematics and Statistics, Rochester Institute of Technology, Rochester, USA

Email: dlfsma@rit.edu

Abstract

Guidance is offered for understanding and using the Legendre transformation and its associated duality among functions and curves. The genesis of this paper was encounters with colleagues and students asking about the transformation. A main feature is simplicity of exposition, while keeping in mind the purpose or application for using the transformation.

Keywords

Legendre Transformation, Dual Curve, Reflectivity, Line of Support, Tangent Line

1. Introduction

The Legendre transformation is a topic that we have frequently been asked about by colleagues and students from departments such as physics, chemistry, and economics. They have found that most explanations and applications of the transformation are confusing because of strange notation and, often, dependence on limiting procedures or integrals, which they thought were unintuitive or awkward. This paper was developed from our notes that we have used successfully in those situations. It contains the main ideas and sufficient examples for them to go forward with their work. The nature of this paper is expository. It is illustrated with eight prototypical examples, including one from physics and one from economics. It is as succinct as possible and at the post-Calculus I level. Its purpose is to clarify the issues that our colleagues and students were facing.

The main goal of the Legendre transformation is to change from coordinates and functions in one setting to different coordinates and functions in another setting. The portrayals are equivalent, but different insights might be gained from them. The Legendre transformation is a powerful mathematical tool that has the potential to convert one set of variables to another, potentially leading to

simpler equations, deeper insights, and aids in the development of various concepts in mathematical and applied fields. In applications, both sets of coordinates and functions and their settings are meaningful in their field of study, but in differing ways. Generically, the coordinates in one setting or space are x and y, the independent variable is x, and the function of interest is $f(x)$. In the other space, the generic coordinates are s and t, the independent variable is s, and the function is $g(s)$.

In Example 3.2, which is about a topic in economics, x represents the amounts of something that is produced and $f(x)$ represents the total cost of each of those amounts. The related set of quantities or variables are s, which is the selling price of each item, and $g(s)$, which is the maximum possible profit at that price. Determining the best quantity to produce and selecting the selling price to be set are related through the transformation.

Contrary to some appearances, the transformation is not difficult to understand or implement. Consider the differentiable curve: $y = f(x)$, which is either strictly concave up or down on an open and connected interval I. (These stringent conditions of differentiability and concavity are relaxed in Section 4.) The curve C is determined on I as the envelope of its tangent lines, *i.e.*, knowing the tangent lines is equivalent to knowing the identity of the points on the curve. See **Figure 1**. The tangent line to C at $a \in I$ is $y = f'(a)x - (af'(a) - f(a))$. Because of the strict concavity, as a varies over I, the set of tangent lines is in one-to-one correspondence with the points on C.

A non-vertical tangent line can be determined by its slope and the negation of its y-intercept, which are the coordinates of a new curve K. One reason to consider strictly concave functions is that there are no vertical tangents for these curves defined on open intervals. Additionally, many applications, such as
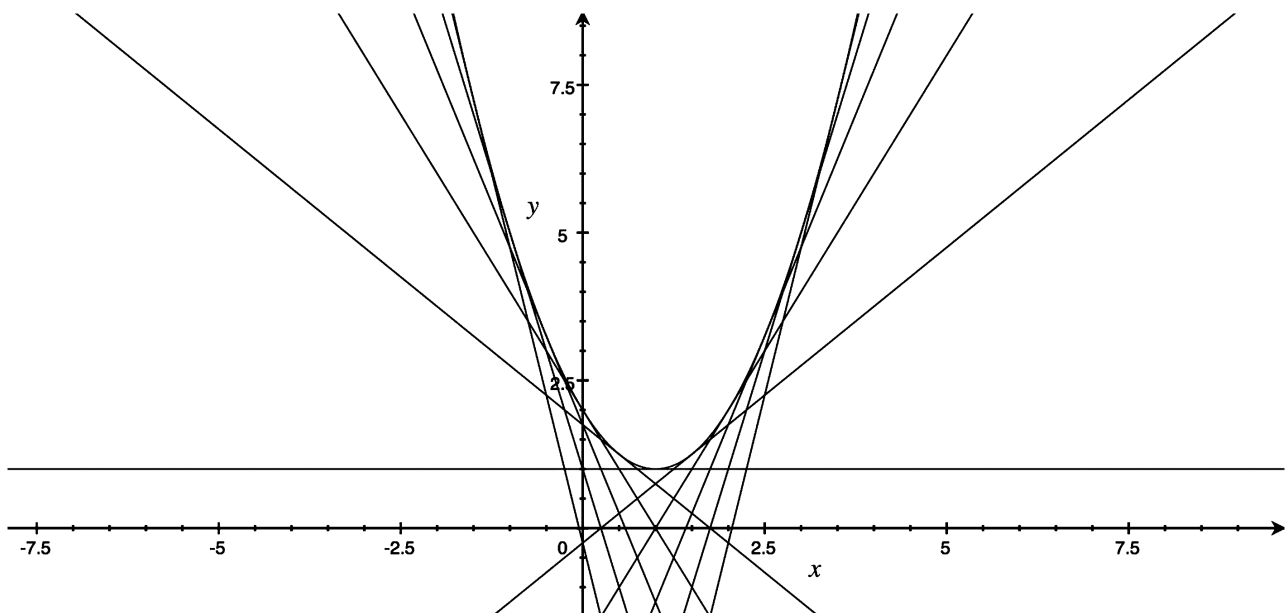

Figure 1. Each tangent line corresponds to exactly one point of the strictly concave differentiable curve $C : y = x^2 - 2x + 2$.

Examples 3.1 and 3.2, contain functions that are strictly concave up. The reasons for the negation are discussed in Section 5.

Definition 1.1. The Legendre transformation K of the strictly concave, differentiable curve $C : y = f(x)$ on a connected, open interval is

$$s = f'(a) \quad \text{and} \quad t = h(a) = af'(a) - f(a) \tag{1}$$

with parameter a. It is in a new space with coordinates (s, t).

The transformation is a way to catalog or record the tangent lines, and, in turn, the points of C. The points of C and the points of K correspond.

If the inverse function f'^{-1} can be found, then from (1), $a = f'^{-1}(s)$ and an explicit equation for K is

$$
\begin{aligned}
t = g(s) &\equiv h\left(f'^{-1}(s)\right) \\
&= f'^{-1}(s) f'\left(f'^{-1}(s)\right) - f\left(f'^{-1}(s)\right) \\
&= sf'^{-1}(s) - f\left(f'^{-1}(s)\right).
\end{aligned}
\tag{2}
$$

The function $g(s)$ is said to be the Legendre transformation of $f(x)$ or the function that is dual to $f(x)$, and is written

$$g(s) = \mathcal{L}(f(x))(s).$$

The term dual curve is used as well. Here is a mathematical example, where the initial curve is a semicircle.

Example 1.1. The curves $C : y = f(x) = \sqrt{1 - x^2}$ for $-1 < x < 1$ and $K : t = g(s) = -\sqrt{1 + s^2}$ for $s \in \mathbb{R}$ are dual curves. See **Figure 2** and **Figure 3**. The tangent line at point $\left(a, \sqrt{1 - a^2}\right)$ of C is

$$L : y = -\frac{a}{\sqrt{1 - a^2}} x - \frac{-1}{\sqrt{1 - a^2}}.$$
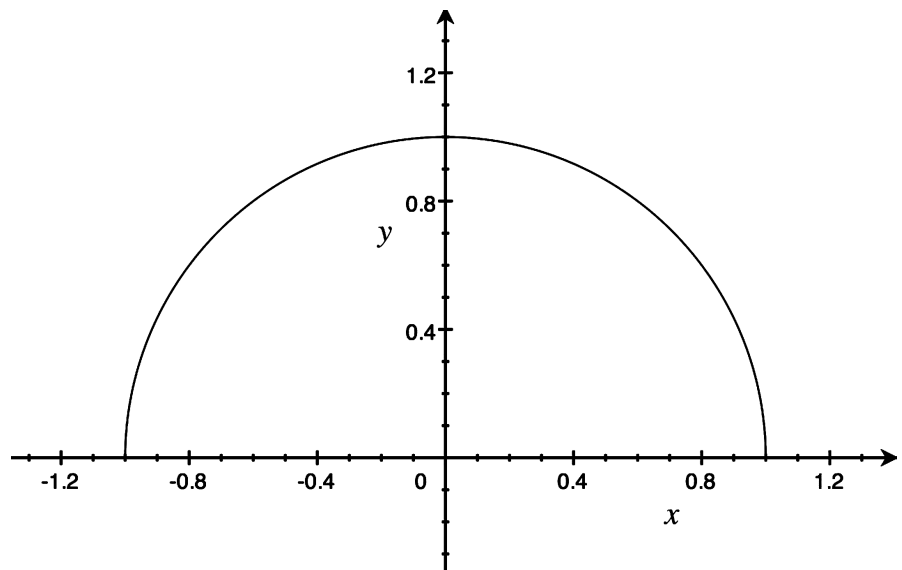
Thus,

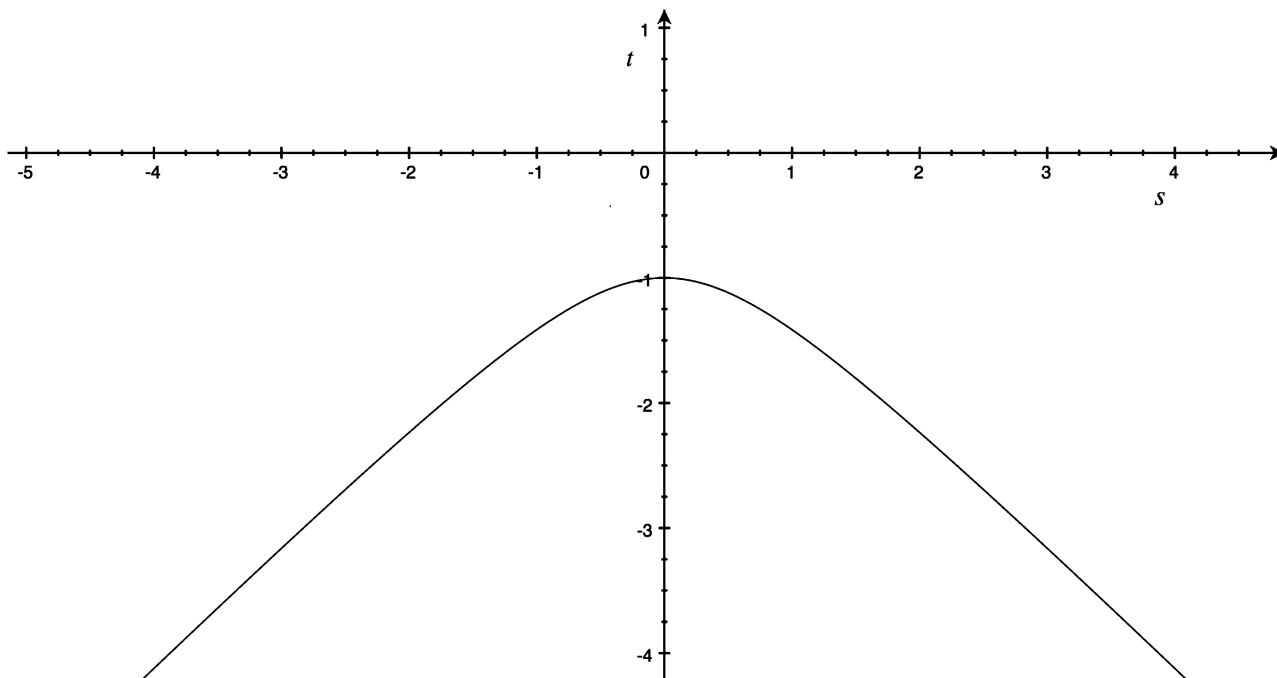**Figure 2.** The original curve for Example 1.1.

Figure 3. The dual curve for Example 1.1.

$$s = s(a) = -\frac{a}{\sqrt{1-a^2}} \quad \text{and} \quad t = h(a) = -\frac{1}{\sqrt{1-a^2}}$$

is a parametric depiction of K. Solving $s = -a/\sqrt{1-a^2}$ for $a^2 = s^2/(1+s^2)$ and substituting a^2 into $h(a) = -1/\sqrt{1-a^2}$ yields the dual curve $K : t = g(s) = -\sqrt{1+s^2}$ explicitly.

Section 2 discusses three properties of the Legendre transformation that are used subsequently and presents two more mathematical examples to display typical derivations. Section 3 contains a discussion of many applications, which exhibit both the practical and theoretical utility of the transformation. Examples 3.1 and 3.2 are key to physics and economics. Extensions to curves that are non-differentiable and not strictly concave up or down are shown through three examples in Section 4. Concluding remarks are in Section 5.

2. Properties and Examples

The following properties are among the many well-known ones of the Legendre transformation ([1], pp. 61-65, [2] [3]).

Property 2.1. The Legendre transformation is reflexive or involutive, *i.e.*, $\mathcal{L}(\mathcal{L}(f(x))) = f(x)$.

Proof. Replacing $a = f'^{-1}(s)$ with x in (1), shows that (2) can be expressed as $g(s) = sx - f(x)$ or

$$g(s) + f(x) = sx. \tag{3}$$

The symmetry of this equation in the variables s and x implies that f and g are dual functions of one another.

Property 2.2. The curves C and K have the same concavity.

Proof. Assume that f is twice differentiable. From (3), $g'(s) = x$, so that

$$1 = \frac{dx}{dx} = \frac{d}{dx}\frac{dg(s)}{ds} = \frac{d^2g(s)}{ds^2}\frac{ds}{dx} = \frac{d^2g(s)}{ds^2}\frac{d}{dx}\frac{df(x)}{dx} = \frac{d^2g(s)}{ds^2}\frac{d^2f(x)}{dx^2}.$$

Thus, the second derivatives of the functions for C and K share the same sign.

The next property illustrates the impact that translation has on a function's Legendre transformation.

Property 2.3. $\mathcal{L}(f(x+c_1)+c_2) = \mathcal{L}(f(x))(s) - c_1 s - c_2$, where c_1 and c_2 are independent of x.

Proof. Take the derivative of f to be invertible. The tangent line at $x = a$ is

$$y = f'(a+c_1)x - (af'(a+c_1) - f(a+c_1) - c_2),$$

So that $s = f'(a+c_1)$ and $a = f'^{-1}(s) - c_1$. Thus,

$$\begin{aligned} t &= af'(a+c_1) - f(a+c_1) - c_2 \\ &= (f'^{-1}(s) - c_1)s - f(f'^{-1}(s)) - c_2 \\ &= \mathcal{L}(f(x))(s) - c_1 s - c_2. \end{aligned}$$

To further illustrate the implementation of the definition of the dual curve, the present example shows that $y = e^x$ for $x \in \mathbb{R}$ and $t = s(\ln s - 1)$ for $s > 0$ are Legendre-transformation pairs of curves or functions. The function in the second example is the only function whose transformation has the same functional form point-by-point, *i.e.*, $\mathcal{L}(x^2/2)(s) = s^2/2$ for x and $s \in \mathbb{R}$. Examples 1.1, 2.1, and 2.2 illustrate Property 2.2 that concavity is preserved.

An extensive table of properties and dual pairs of curves appears in [2].

Example 2.1. The curve $C : y = f(x) = e^x$ for $x \in \mathbb{R}$ and the curve $K : t = g(s) = s(\ln s - 1)$ for $s > 0$ are dual curves. See **Figure 4** and **Figure 5**. The tangent line at point (a, e^a) of C is

$$y = e^a x - e^a(a-1).$$

From the slope and negation of the y-intercept, the corresponding point of the dual curve is $(e^a, e^a(a-1))$, which supplies a parametric representation of K with parameter a. Setting $s = e^a$ yields $a = \ln s$ for $s > 0$ and from (2) the explicit form

$$t = g(s) = s(\ln s - 1).$$

Slightly more efficiently, but perhaps more opaquely, the tangent line need not be displayed. From the slope of C, $s = y' = e^x$, obtain $x = \ln s$, and from (3), obtain $g(s) = sx - f(x)$, so

$$t = g(s) = s\ln s - e^{\ln s} = s(\ln s - 1).$$

Example 2.2. The curves $C : y = f(x) = x^2/2$ for $x \in \mathbb{R}$ and $K : t = g(s) = s^2/2$ for $s \in \mathbb{R}$ are dual to each other. Since $s = y'(x) = x$, from (3), $t = g(s) = sx - f(x) = s^2 - s^2/2 = s^2/2$.
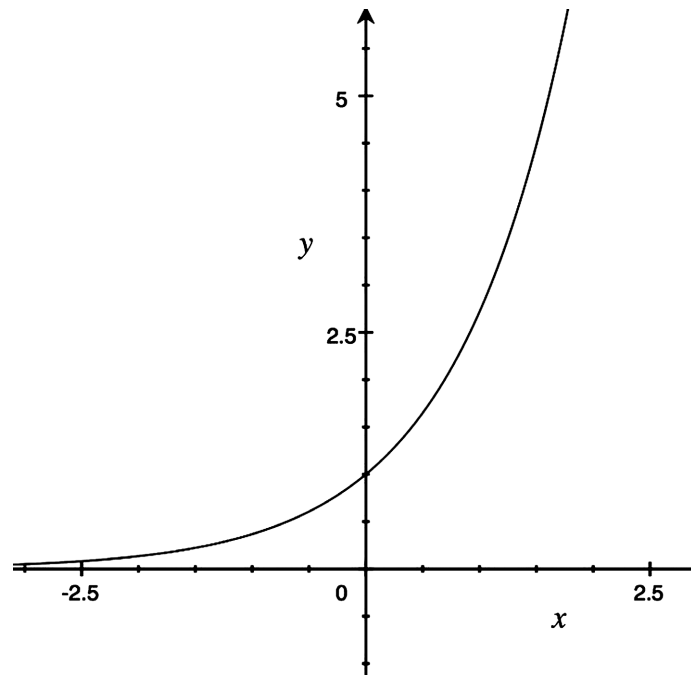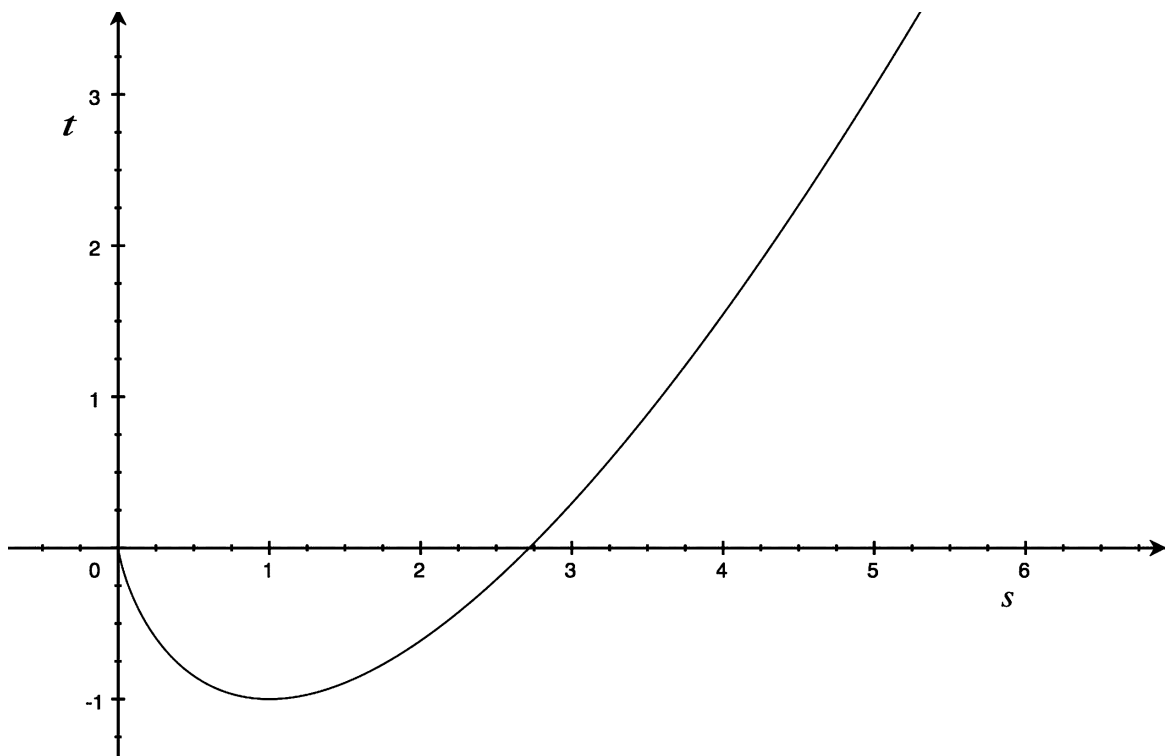
Figure 4. The original curve for Example 2.1.

Figure 5. The dual curve for Example 2.1.

3. Applications

Example 3.1 from classical mechanics and Example 3.2 from economics demonstrate important historical applications of the Legendre transformation. They reveal the fundamental nature and importance of the transformation. In the

economics example, two different ways of addressing a decision in terms of selling price and quantity are shown to be equivalent. In the physics example, the transformation is seen to be the bridge between two different foundational representations of mechanics.

One huge benefit of the transformation is practical. It can change variables from those that might be hidden from direct measurement or be theoretical to those that are measurable or amenable to control. In the example from economics, price may be set by the economic system or a governmental agency, but the quantity variable in the other space may be at the discretion of the company. An actuary might face the opposite situation of needing to find the correct selling price for insurance policies.

The tactic of utilizing the Legendre transformation to change variables to those that are measurable, consequential, or controllable appears in many areas. It is used as a tool in [4], where biochemical thermodynamical variables are transformed from those that are in the units of energy to variables in the more accessible units of pH. In a study of rainfall in [5], the transformation is used to show that high values of a parameter associated with one variable are linked to extreme outcomes of another variable. To determine the probabilities of detecting small birds by using various equipment, [6] used a Legendre transformation to obtain variables that are independent.

In Examples 3.1 and 3.2, the symbols are renamed from (x, y) and (s, t) in order to closely match the applications.

Example 3.1. The Lagrangian function is

$$L(x, \dot{x}) = \frac{1}{2} m \dot{x}^2 - V(x)$$

for a particle of mass m constrained to the x-axis, where its speed or velocity is $\dot{x} = dx/dt$, its kinetic energy is $\frac{1}{2} m \dot{x}^2$, and its potential energy is $V(x)$. The location x, which is independent of the speed $\dot{x}$, is treated as a constant in this Legendre transformation, so that it does not participate in the transformation. Symbols $\dot{x}$ and L replace x and f, respectively, in the generic set up. For brevity for finding the dual function, use the more efficient method, which was employed in Example 2.1. The slope of L is $s = m\dot{x}$, which is designated p and is the independent variable replacing the generic s. It is the particle's momentum. From (3), obtain $g(s) = s\dot{x} - \left(\frac{1}{2} m \dot{x}^2 - V(x)\right)$. By replacing the generic $g(s)$ with $H(x, p)$ and substituting $\dot{x} = p/m = p/m$ and $s = p$, obtain the dual function

$$H(x, p) = \frac{p}{2m} + V(x).$$

This is the Hamiltonian function, which is the conserved total energy. See [[1], pp. 65-67] and [3].

Example 3.2. The total cost as a function of quantity and the maximum total

profit as a function of selling price per item are dual curves. For a manufacturer of a particular item that can be shipped in arbitrarily sized lots, consider two formulations. In one, the variables are the quantity of items q in a lot and the cost c to the manufacturer of making the q items, thus $c = c(q)$. These replace the notation x and $y = f(x)$. In the other one, the variables are the selling price per item s that the market allows for that lot and the maximum total profit p for the lot. Thus, $p = p(s)$, where this replaces the notation for the generic $t = g(s)$.

Assume that $c = c(q)$ is strictly concave up over the open interval of q-values of interest, which is a reasonable assumption. Also assume that $c(q)$ is twice differentiable so $c'' > 0$ and that c' is invertible.

A lot's profit is the selling price per item times the number of items in the lot minus the cost of the lot, *i.e.*,

$$\text{Profit} = sq - c(q).$$

For each possible selling price s, $\mathrm{d}/\mathrm{d}q(\text{Profit}) = s - c'(q)$ and $\mathrm{d}^2/\mathrm{d}q^2(\text{Profit}) = -c''(q) < 0$. The maximum profit p is obtained at $q = q_s$, where $q_s = c'^{-1}(s)$ is the solution to $\mathrm{d}/\mathrm{d}q(\text{Profit}) = 0$, and hence $s = c'(q_s)$. This is a one-to-one correspondence between s and q. Thus,

$$p(s) = sq_s - c(q_s) = sc'^{-1}(s) - c(c'^{-1}(s)).$$

Comparing with (2), this expression $p(s)$ is the Legendre transformation of c.

4. Extending the Definition and the Applicability of $\mathcal{L}(f(x))$

In most applications, such as to cost $c(q)$ in Example 3.2, the functions are either strictly concave up or concave down. Occasionally, more complicated situations arise. The curve may be partly concave up and partly concave down, bi-valued, and even have self-intersections and cusps. There could be linear portions and corners. The Legendre transformation can be used for analyzing many of these functions and curves as the following three examples reveal.

Example 4.1 contains a function that is concave up in one part and concave down in the other part. Each part can be transformed separately, then combined for the transformed curve. The cusp in the transformation is an artifact of the vanishing second derivative of *f* at the origin. Because of reflectivity in Property 2.1, the inverse Legendre transformation of its transformation illustrates the ability of the Legendre transformation to handle the occurrence of a doubled-back curve and a cusp.

Example 4.1. Consider

$$y = f(x) = \frac{x^3}{3}$$

for $x \in \mathbb{R}$. See **Figure 6**, which shows that *f* is concave down for $x < 0$ and
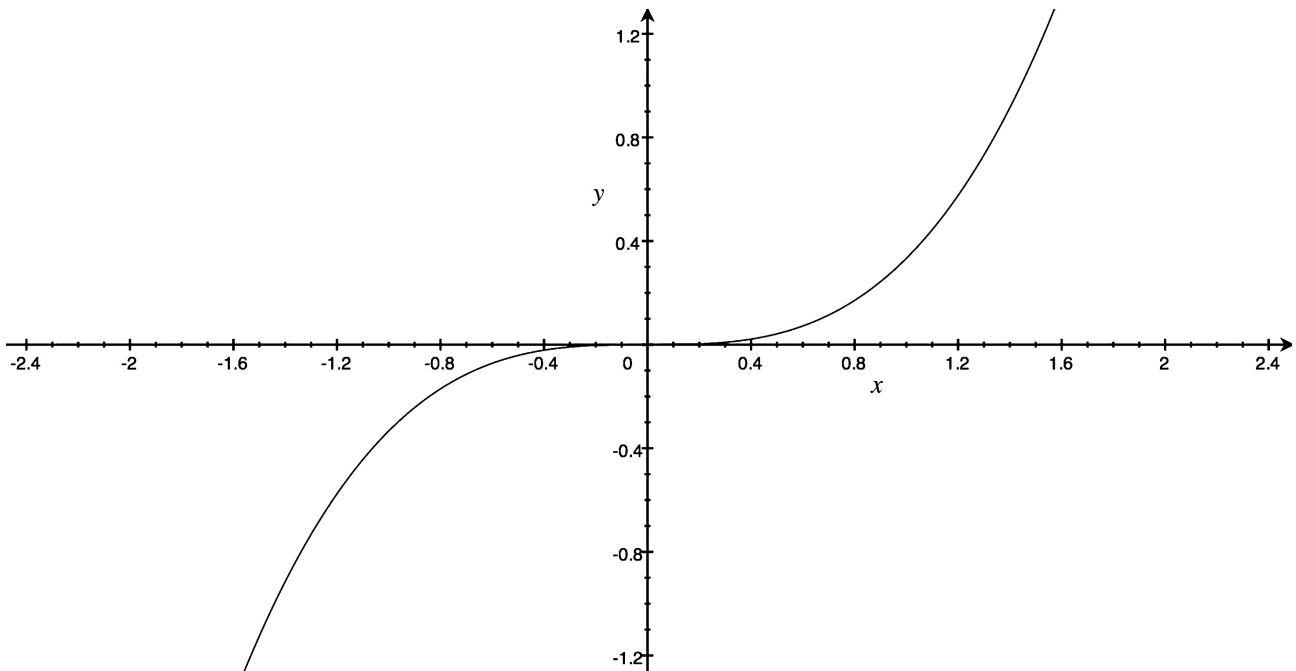
Figure 6. A curve with concave up and concave down parts in Example 4.1.

concave up for $x > 0$. The Legendre transformation is double valued for positive slopes s, because each value for slope of f applies to one tangent line for $x < 0$ and one for $x > 0$. At $x = a > 0$, the tangent line is $y = a^2 x - \frac{2}{3} a^3$. Then, $s = a^2$, $a = \sqrt{s}$, and $t = \frac{2}{3} s^{3/2}$. Similarly, for $x < 0$, $t = -\frac{2}{3} s^{3/2}$. The point $(x, y) = (0, 0)$ of f gives the single point $(s, t) = (0, 0)$, which is a cusp. See **Figure 7**.

The periodic function in Example 4.2 has concave down parts that are similar to each other and might come about in an application with a rectified wave.

Example 4.2. Consider

$$y = f(x) = \sqrt{1 - \left(x - (2n - 1)\right)^2}$$

for $2n - 2 < x < 2n$ and $n = 1, 2, 3, \cdots$. See **Figure 8**. Applying Example 1.1 and Property 2.3 with $c_1 = -(2n - 1)$ and $c_2 = 0$ to each period gives function

$$t = g(s) = \mathcal{L}(f(x))(s) = -\sqrt{1 + s^2} + (2n - 1)s. \tag{4}$$

For each value of n, the domain of the branch (4) is $\mathbb{R}$, because the slope takes all real values in each period of f. The end points where $x = 0, 2, 4, \cdots$ have vertical tangents, thus go to the points at infinity of each branch in the transformation. Each branch contains the point $(s, t) = (0, -1)$, because the maximum in each period of f has slope $s = 0$ and a tangent line with y-intercept 1. As n becomes very large, t in (4) approaches the t axis. See **Figure 9**. The self-intersection of $t = g(s)$ in a point (which is $(0, -1)$ in this example) and the appearance of the branches as curves rotating about that point is a
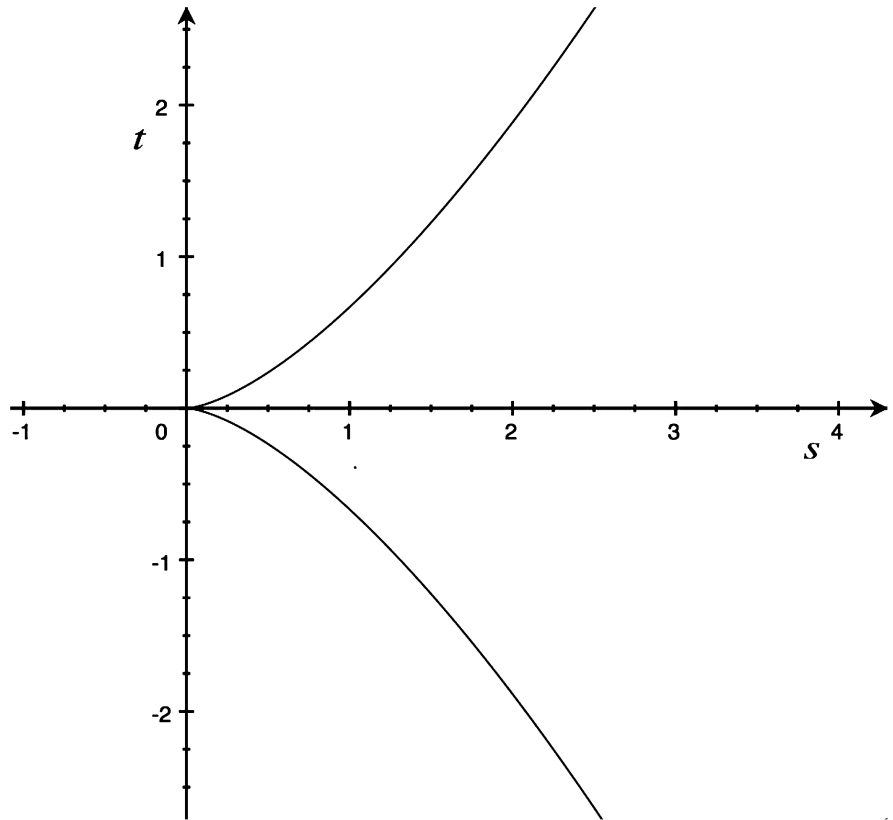
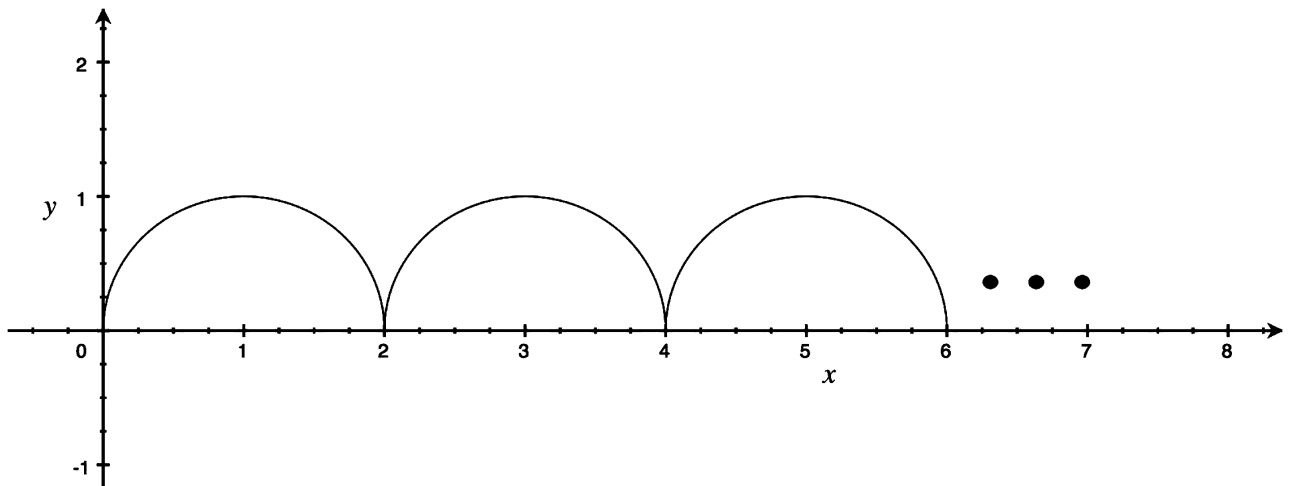**Figure 7.** A curve that doubles back and contains a cusp in Example 4.1.

Figure 8. The periods for $n = 1$, 2, and 3 are displayed for the rectified wave in Example 4.2.

signature of such a wave f.

Supporting lines are generalizations of tangent lines. A line of support to a curve intersects the curve locally in exactly one point or else in a line segment. Tangent lines are examples. Curves are determined by their lines of support in the same way that differentiable curves are by their tangent lines ([7], pp. 41-43, 205-212; [8], p. 34). Using supporting lines allows an extension of the Legendre transformation to piecewise linear curves, as shown in Example 4.3, which exhibits
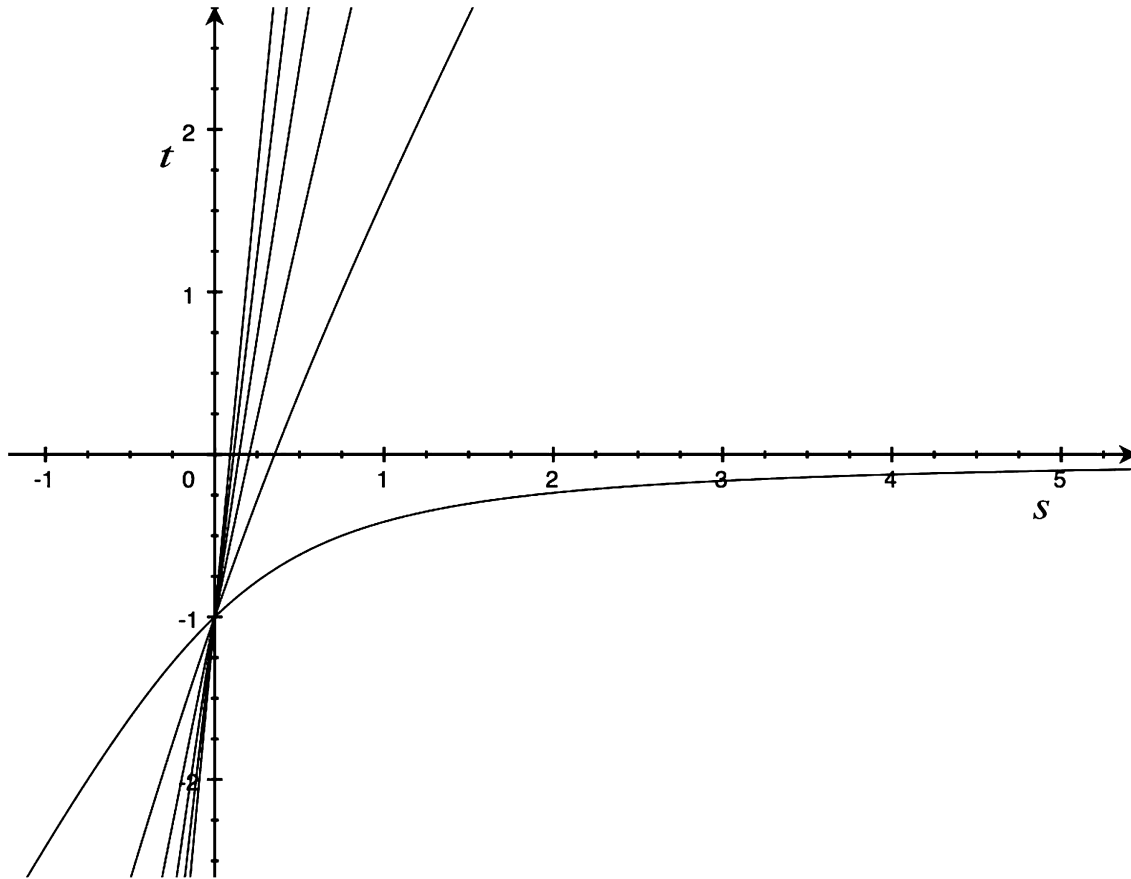
Figure 9. Moving counterclockwise about (0, −1), the branches for $n = 1$ though 6, respectively, for the dual curve in Example 4.2.

how linear portions (corner points) of a curve correspond to corner points (linear portions) of its Legendre transformation.

Example 4.3. Consider the following piecewise linear function consisting of three line-segments

$$y = f(x) = \begin{cases} -4x - 6 & x < -1 \\ x - 1 & -1 \le x < 1 \\ 3x - 3 & 1 \le x \end{cases}.$$

See **Figure 10**. The three linear portions give the three points $(s,t) = (-4,6),(1,1)$ and $(3,3)$.

For example, for $y = -4x - 6$, $s = -4$ and $t = 6$. Each of the two corners at $(x, y) = (-1, -2)$ and $(1, 0)$ produce a linear segment in the transformation. The corner at $(-1, -2)$ transforms to a line segment from the pencil of its supporting lines $y = s(x+1) - 2 = sx - (-s + 2)$ for $-4 \le s \le 1$, so that $t = -s + 2$. Similarly, the supporting lines at (1, 0) are $y = s(x-1) + 0$, so that $t = s$ for $1 \le s \le 3$. Thus,

$$t = g(s) = \mathcal{L}(f(x))(s) = \begin{cases} -s + 2 & -4 \le s < 1 \\ s & 1 \le s \le 3 \end{cases}.$$
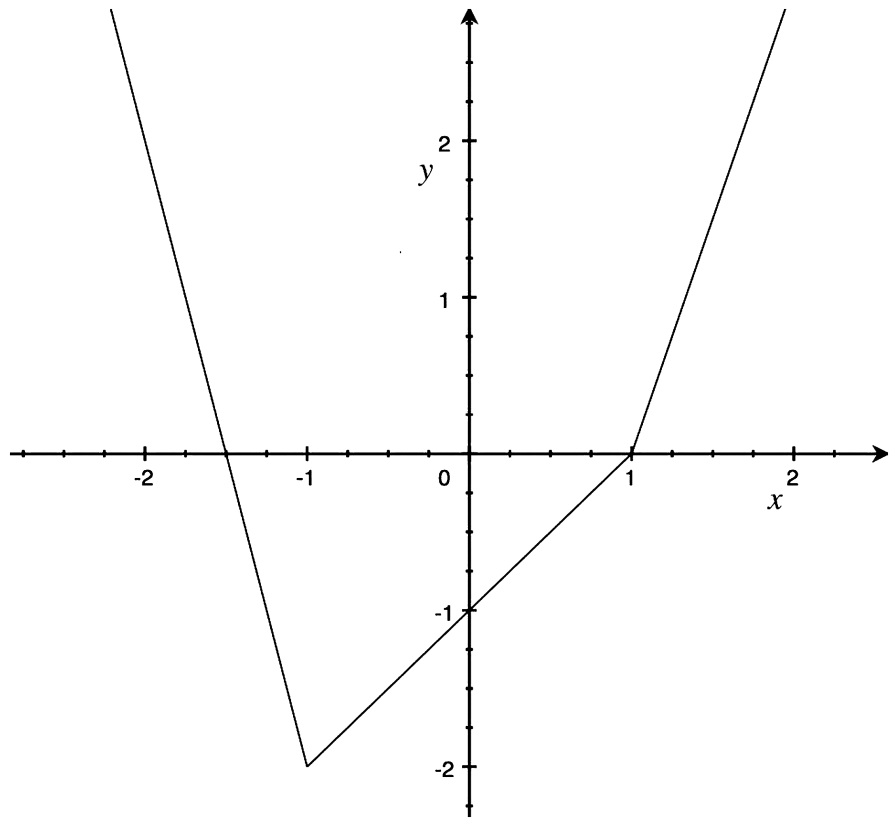
See **Figure 11**.

Figure 10. Piecewise linear curve in Example 4.3.
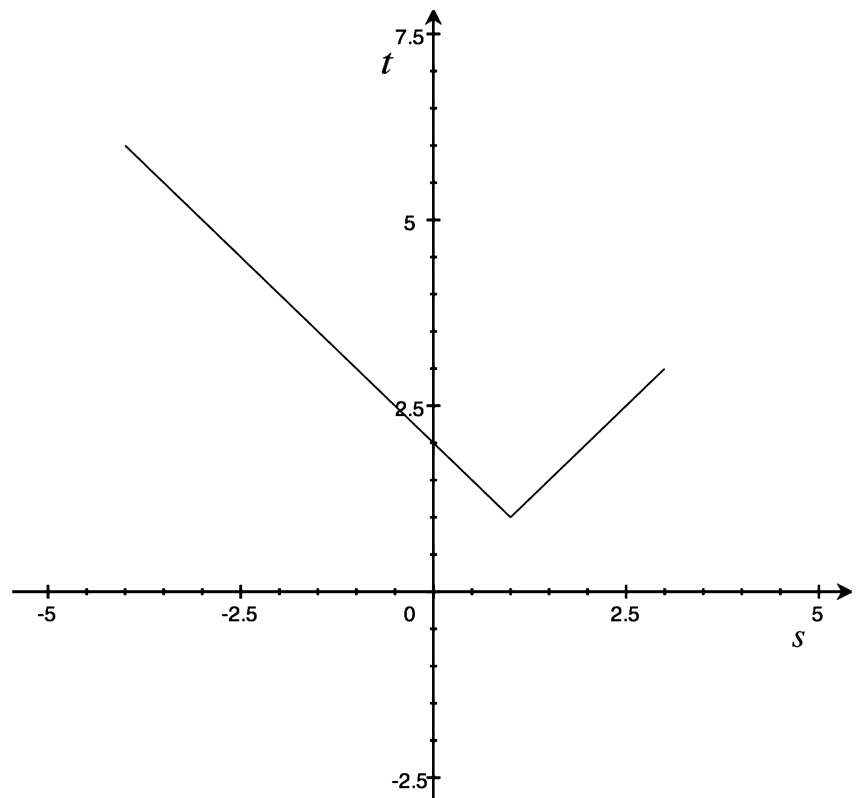

Figure 11. Dual curve consisting of line segments in Example 4.3.

5. Conclusions

The Legendre transformation is basically a simple idea: A second curve is created that has the same information as the original curve. Going back and forth between the curves is easy, since the parametrizations are functions of each other. The second curve uses coefficients from tangent and supporting lines of the other curve. In applications, the advantage of having the two curves is that different insights may be obtained from either one. In Example 3.2, selling price and quantity are functionally related but thinking in terms of one or the other might be preferable.

Alternative duality transformations can be based upon other pairs of coefficients that are in standard forms for lines. For the slope-intercept form

$$y = mx + b,$$

the new variables are m and $b = b(m)$. This transformation introduces a minus sign, compared to the Legendre transformation. In Examples 3.1 and 3.2, physical measurements would have been made negative in this transformation. This duality does not possess Property 2.2 of maintaining concavity because the minus sign reverses concavity. Going through any one of the proofs or examples with b, instead of g, illustrates how the minus sign moves through the process.

The standard form of a line that is based on the dot product,

$$ux + vy = (u,v) \bullet (x,y) = 1,$$

is widely used, especially in Minkowski geometries, that is, real normed vector or Banach spaces, where the variables are u and $v = v(u)$ [8] [9].

The concepts in the Legendre transformation are similar to those in other duality transformations, so studying it is helpful for understanding the others.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Arnold, V.I. (1978) Mathematical Methods of Classical Mechanics. 2nd Edition, Springer, New York. https://doi.org/10.1007/978-1-4757-1693-1

[2] Kolt, Q.T., Kilner, S.J. and Farnsworth, D.L. (2022) A Table of Legendre-Transformation Pairs with Methodologies for Construction, Authentication, and Approximation of Pairs. arxiv.org/abs/2208.05043. https://doi.org/10.48550/arXiv.2208.05043

[3] Zia, R.K.P., Redish, E.F. and McKay, S.R. (2009) Making Sense of the Legendre Transform. *American Journal of Physics*, **77**, 614-622. https://doi.org/10.1119/1.3119512

[4] Jinich, A., Sanchez-Lengeling, B., Ren, H., Goldford, J.E., Noor, E., Sanders, J.N., Segrè, D. and Aspuru-Guzik, A. (2020) A Thermodynamic Atlas of Carbon Redox Chemical Space. *Proceedings of the National Academy of Sciences of the United States of America*, **117**, 32910-32918. https://doi.org/10.1073/pnas.2005642117

[5] Lee, J., Paz, I., Schertzer, D., Lee, D.I. and Tchiguirinskaia, I. (2020) Multifractal Analysis of Rainfall-Rate Datasets Obtained by Radar and Numerical Model. *Journal of Applied Meteorology and Climatology*, **59**, 819-840. https://doi.org/10.1175/JAMC-D-18-0209.1

[6] Crewe, T.L., Deakin, J.E., Beauchamp, A.T. and Morbey, Y.E. (2019) Detection Range of Songbirds Using a Stopover Site by Automated Radio-Telemetry. *Journal of Field Ornithology*, **90**, 176-189. https://doi.org/10.1111/jofo.12291

[7] Lay, S.R. (2007) Convex Sets and Their Applications. Dover Publications, Mineola.

[8] Thompson, A.C. (1996) Minkowski Geometry. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781107325845

[9] Schneider, R. (2014) Convex Bodies: The Brunn-Minkowski Theory. 2nd Edition, Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781139003858