I. Determining the distributions of reflected rays for a general class of curved mirror surfaces

E.R. Vrscay

First version of this manuscript: May 10, 2024
Latest version: October 28, 2025

We consider the following problem in the xy-plane. Suppose that the curve $y = f(x)$ defines a reflective surface and that rays of light travelling downward and parallel to the y-axis hit the surface and are reflected according to the usual "law" of reflection, i.e., angle of incidence = angle of reflection, both angles measured with respect to the normal to the curve. Assuming that the function $f(x)$ is "suitable", i.e., convex at least in a region that would permit such reflection, we wish to determine the position $(0, d)$ at which a ray reflected from the point on the curve $(a, f(a)$ intersects the y-axis. Without loss of generality, assume that $f(0) = 0$. Without getting bogged down with the need for other assumptions, we'll simply state here that an acceptable reflection is one where the incident light ray hits the mirror surface $y = f(x)$ only once before intersecting the y-axis. In most cases, an unacceptable reflection is detected when the reflected ray is computed to intersect the y-axis at negative values – in such cases, the ray will almost certainly strike the mirror again, only to be deflected again. We do not worry about such cases. (That being said, it may well be interesting to pursue them in the future.)

A general situation is sketched in Figure 1 below. A vertical ray of light at $x = a$ hits the curve at $(a, f(a))$ (Point B) and is reflected toward the y-axis, intersecting it at $(0, d)$ (Point D). The normal $\mathbf{N}$ to the curve at $x = a$ (Point B) is shown. The slope of $\mathbf{N}$ is

$$m_N = -\frac{1}{f'(a)}\,. \tag{1}$$

(Of course, we assume that $f'(a)$ exists, i.e., is finite.) The equation of the normal $\mathbf{N}$ is easily found:

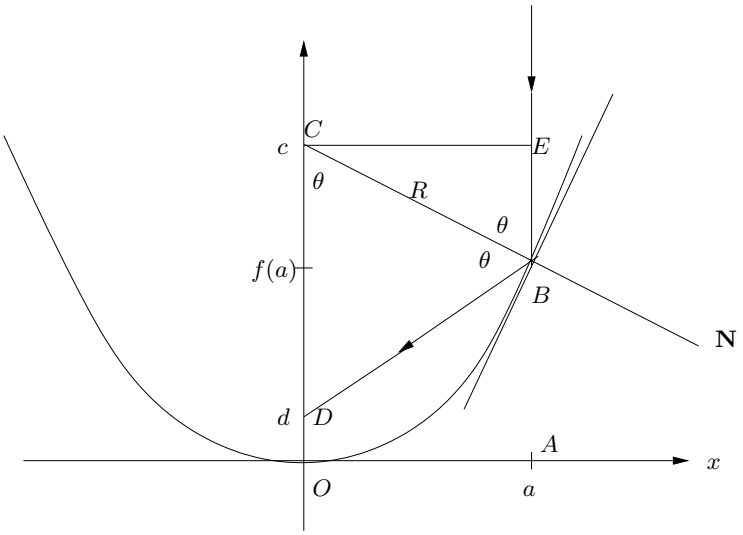

Figure 1

$$\frac{y - f(a)}{x - a} = -\frac{1}{f(a)} \quad \Longrightarrow \quad y = -\frac{1}{f'(a)}(x - a) + f(a)\,. \tag{2}$$

Now set $x = 0$ in the above equation to find the point $(0, c)$ where the normal $\mathbf{N}$ intersects the y-axis,

$$c = \frac{a}{f'(a)} + f(a)\,. \tag{3}$$

The (equal) angles θ of incidence and reflection measured with respect to the normal $\mathbf{N}$ are shown in the diagram. Note also that since the incoming light ray is parallel to the y-axis,

$$\angle BCD = \theta\,, \tag{4}$$

as shown in the diagram.

Triangle BDC is isoceles. We can find the length R of its base using Eq. (3):

$$R = \sqrt{a^2 + (c - f(a))^2} = \sqrt{a^2 + \left[\frac{a}{f'(a)}\right]^2} = a\sqrt{1 + \left(\frac{1}{f'(a)}\right)^2}\,. \tag{5}$$

We can also find angle θ by considering the triangle BEC formed by the normal $\mathbf{N}$, the incoming ray at $x = a$ and the horizontal line $y = c$ which intersects the incoming ray at E. Note that

$$||\overline{BE}|| = c - f(a) = \frac{a}{f'(a)} \quad \text{(from Eq. (3))}\,. \tag{6}$$

The triangle is shown below. Note that

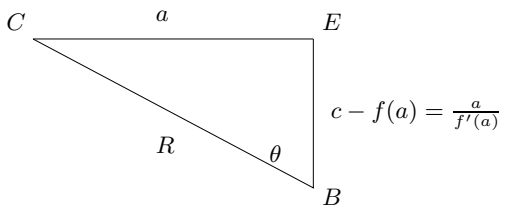

Figure 2

$$\theta = \mathrm{Tan}^{-1}\left(\frac{a}{a/f'(a)}\right) = \mathrm{Tan}^{-1}(f'(a))\,. \tag{7}$$

We now consider triangle BCD as an isoceles triangle with base length R and sides of equal length S where

$$S := c - d = \frac{R}{2}\sec\theta\,. \tag{8}$$

This triangle is shown below. From the earlier triangle, it follows that

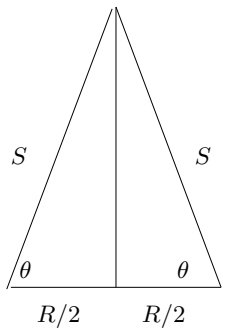

Figure 3

$$\sec\theta = \frac{R}{a/f'(a)} = \frac{Rf'(a)}{a}\,. \tag{9}$$

Substitution into Eq. (8) yields

$$\begin{aligned} S = c - d = \frac{R^2 f'(a)}{2a} &= \frac{a f'(a)}{2}\left[1 + \left(\frac{1}{f'(a)}\right)^2\right] \quad \text{from Eq. (5)} \\ &= \frac{a}{2}\left[f'(a) + \frac{1}{f'(a)}\right]. \end{aligned} \tag{10}$$

Rearranging and using Eq. (3), we have

$$d = c - S = \frac{a}{f'(a)} + f(a) - \frac{a}{2}\left[f'(a) + \frac{1}{f'(a)}\right], \tag{11}$$

which simplifies to the final result for the position of the reflected ray on the y-axis,

$$\boxed{d(a) = f(a) - \frac{a}{2} f'(a) + \frac{a}{2f'(a)}\,.} \tag{12}$$

Before considering some examples, we mention that in a later section the above result is derived using direction vectors. The direction vector method is very useful since it can be used to derive another result will be relevant to our solar cooker studies.

Example 1: It is well known that a parabola has a single focal point. Let $f(x) = rx^2$ in Eq. (12), where $r > 0$, to give

$$d_r(a) = ra^2 - a^2 r + \frac{a}{4ar} = \frac{1}{4r}\,, \tag{13}$$

which is independent of a. Note that we have denoted the r-dependence of d in the r-subscript.

Example 2: This is perhaps trivial, but let's consider the linear function $f(x) = rx$, i.e., a slanted and straight mirror with slope r starting at the origin. Then

$$d_r(a) = ra - \frac{ar}{2} + \frac{a}{2r} = \left(r - \frac{1}{2} + \frac{1}{2r}\right) a\,, \qquad a \geq 0\,. \tag{14}$$

Note that d is a linear function of a as expected on geometric grounds. When $r = 1$, i.e., the angle of inclination with respect to the x-axis is $\phi = \frac{\pi}{4}$, then $d = a$, as expected.

Example 3: Consider a circularly-shaped mirror, i.e.,

$$f(x) = -\sqrt{R^2 - x^2} + R\,, \quad -R \leq x \leq R\,. \tag{15}$$

Then

$$f'(x) = \frac{x}{\sqrt{R^2 - x^2}}\,. \tag{16}$$

From Eq. (12), we have, at least formally,

$$\begin{aligned} d_R(a) &= -\sqrt{R^2 - a^2} + R - \frac{a^2}{2}\frac{1}{\sqrt{R^2 - a^2}} + \frac{1}{2}\sqrt{R^2 - a^2} \\ &= R - \frac{1}{2}\sqrt{R^2 - a^2} - \frac{a^2}{2}\frac{1}{\sqrt{R^2 - a^2}} \\ &= R - \frac{1}{2}\frac{(R^2 - a^2) + a^2}{\sqrt{R^2 - a^2}} \\ &= R - \frac{R^2}{2}\frac{1}{\sqrt{R^2 - a^2}}\,, \qquad\qquad -R < a < R\,. \end{aligned} \tag{17}$$

Note that this equation will actually not be valid for a-values near R where d is predicted to be negative. Here, the rays would actually hit the mirror once again, and perhaps again and again. As a decreases from R toward 0, the first value at which the above formula is valid is when $d = 0$. Let's find this a-value:

$$R = \frac{R^2}{2\sqrt{R^2 - a^2}} \qquad \Longrightarrow \qquad \sqrt{R^2 - a^2} = \frac{R}{2} \qquad \Longrightarrow \qquad a = \frac{\sqrt{3}}{2}R\,. \tag{18}$$

For $a > \dfrac{\sqrt{3}}{2}R$, d increases with a. But as $a \to 0$, the above formula indicates that $d \to \dfrac{R}{2}$. As such, the result in Eq. (17) should be rewritten as follows,

$$d_R(a) = R - \frac{R^2}{2}\frac{1}{\sqrt{R^2 - a^2}}\,, \qquad -\frac{\sqrt{3}}{2}R \le a \le \frac{\sqrt{3}}{2}R\,. \tag{19}$$

In order to understand this result, let us consider the behaviour of $f(x) = R - \sqrt{R^2 - x^2}$ near $x = 0$. Since $f'(0) = 0$, we'll probably have to consider the quadratic approximation of $f(x)$ near $x = 0$. First of all, $f(0) = 0$. Then

$$f'(x) = x(R^2 - x^2)^{-1/2} \qquad \Longrightarrow \qquad f'(0) = 0\,. \tag{20}$$

Then

$$f''(x) = (R^2 - x^2)^{-1/2} + x\left(-\frac{1}{2}\right)(R^2 - x^2)^{-3/2}(-2x) \qquad \Longrightarrow \qquad f''(0) = \frac{1}{R}\,. \tag{21}$$

In other words,

$$f(x) \approx \frac{1}{2}f''(0)x^2 = \frac{1}{2R}x^2 \qquad \text{for } x \text{ near } 0\,, \tag{22}$$

i.e., $f(x)$ behaves locally like the parabola rx^2 with $r = \dfrac{1}{2R}$. From the result in Example 1, it follows that the focal point of this parabola is at

$$d_R = \frac{1}{4r} = \frac{R}{2}\,. \tag{23}$$

As such, we expect that incoming rays with near-zero a-values will be reflected close to the value given in Eq. (23).

In Figure 4 below is plotted the graph of $d(a)$ Eq. (19) for admissible positive values of a in the case $R = 1$. As expected, the graph of $d(a)$ intersects the a-axis at around 0.87 corresponding to the value $\sqrt{3}/2$ determined earlier. Note that for near-zero a-values, rays are reflected near the value $\dfrac{1}{2}$, as predicted by Eq. (23). As a increases, the d(a)-values of reflected rays move farther and farther away from the parabolic focal point value $\dfrac{1}{2}$ toward zero.

What is perhaps even more interesting is the distribution of these $d(a)$ values. For example, do they demonstrate some clustering about the parabolic focal point $\dfrac{1}{2}$? Or is the distribution of these values more evenly spread out, i.e., uniform, over the interval $[0, 0.5]$?

In an effort to answer this question, the distribution of $d(a)$ values is presented in Figure 5 below as a histogram plot. The roughly 1000 $d(a)$-values computed to produce the graph in Figure 4 were "binned" by dividing the a-interval $[0, 1]$ into $N = 100$ subintervals $I_k = [a_{k-1}, a_k)$ of equal length, and then determining the subinterval I_k in which each $d(a)$ belonged. The histogram represents (up to a constant) the number of $d(a)$ values in each subinterval I_k, $1 \le k \le N$. We see that the histogram shows a quite strong peaking at $d = 0.5$.

One may well ask if the nature of this distribution could have been conjectured (guessed?) from the nature of the plot of $d(a)$ in Figure 4. The answer is "Yes". To see that there is, in fact, a concentration of $d(a)$ values around the value 0.5, we should look at the graph of $d(a)$ in Figure 4 "sideways", i.e., at points on the vertical $d(a)$ axis looking rightward. To assist us with this "sideways" look, we can consider lines of constant d-values, i.e., $d = C$, where C is a constant, superimposed on the graph in Figure 4. Note that for C at or very near 0.5, many points of the graph of $d(a)$ lie close to this horizontal line. In other words, there are many values of $d(a)$

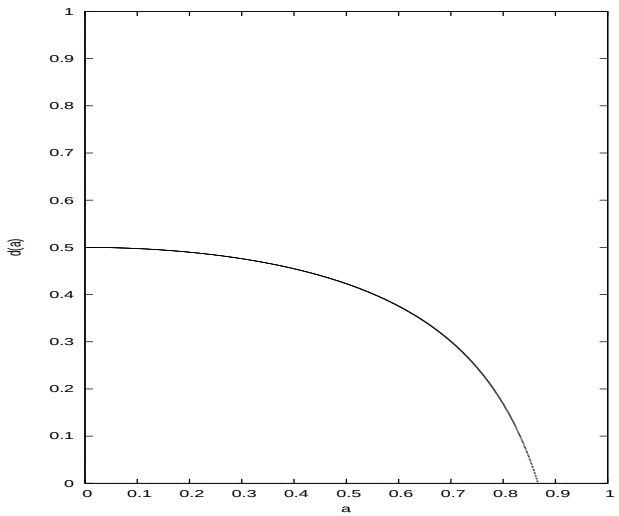

Figure 4: Plot of $d(a)$ vs. a for a circle of radius $R = 1$. Only admissible values of $a > 0$, i.e., those values for which $d(a) \geq 0$, are plotted.

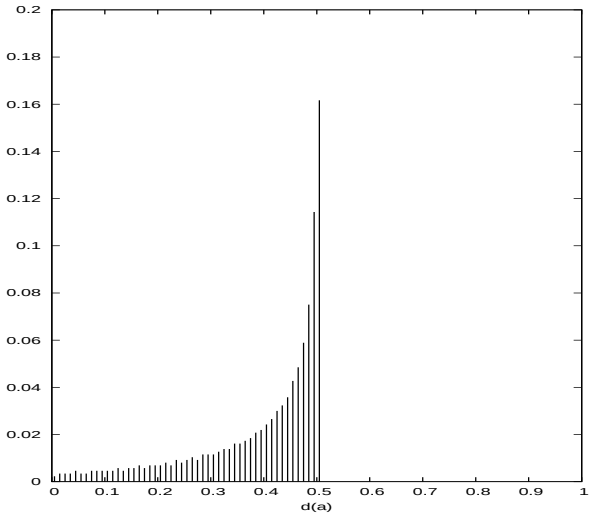

Figure 5: Histogram plot of values of $d(a)$ shown in Figure 4.

that lie close to C. In contrast, for C closer to 0, i.e., $C < 0.2$, the horizontal line $d = C$ cuts the graph of $d(a)$ very sharply and there are very few points of the graph that lie close to C.

(Here, we must qualify that the above discussion, which is admittedly informal, refers to the plot of a **finite** number of $d(a)$ values. As such, we can use the phrase "many points". When we move to the continuous case, i.e., graph of $d(a)$ for all real values between 0 and $\sqrt{3}$, we shall have to be more careful.)

The fact that for a-values near $a = 0$, the values of $d(a)$ do not vary significantly, i.e., remain close to a given horizontal line $d = C$, can clearly be related to the behaviour of the derivative function $d'(a)$ (assuming that it exists). For a-values at which $d'(a)$ is near zero, the function $d(a)$ will not change significantly. As such, we expect a clustering of $d(a)$ values. For a-values at which $d'(a)$ is sufficiently far from zero, the function $d(a)$ will change significantly and we expect a lack of clustering. One might go a step further and conjecture that the degree of clustering is proportional to the inverse of $|d'(a)|$. (This has the "aroma" of inverse functions, as we shall see in the next section.) We shall explore the distribution of reflected rays more theoretically in the next section of this paper.

Example 4: As a slight perturbation of the parabola in Example 1, consider the following even function,

$$f_p(x) = |x|^{2+p} \,. \tag{24}$$

For simplicity, we are considering the special case $r = 1$. And for further simplicity, we consider the case $x > 0$ so that we don't have to worry about the absolute value sign. Finally, for the moment, we first consider p-values of low magnitude, i.e., $\|p\| \ll 1$ but will relax this restriction later. Then

$$f_p(x) = x^{2+p} \,, \quad f'(x) = (2+p)x^{1+p} \,, \qquad x > 0 \,. \tag{25}$$

Substitution into Eq. (12) yields, after a little simplification,

$$d_p(a) = -\frac{p}{2}a^{2+p} + \frac{1}{2(2+p)a^p} \,, \qquad a > 0 \,. \tag{26}$$

When $p = 0$, i.e., the case of the parabola, the first term disappears and the second term becomes a constant, i.e., $d_0(a) = \dfrac{1}{4}$, in agreement with Example 1. When $p \neq 0$, $d_p(a)$ depends on a in some strange way. In fact, Eq. (26) is quite fascinating and worthy of detailed analysis. Unfortunately, such an endeavour is beyond the scope of this study – we shall be content with the examination of the $d_p(a)$ for some representative values of p. That being said, let us make a few quick observations as a kind of "teaser."

1. For $p > 0$, the second term in Eq. (26) decreases as $a \to \infty$ which implies that $d_p(a)$ is negative (and therefore inadmissible) for all a-values greater than a threshold value a_p^*. In fact, it is easy to compute this threshold value,

   $$a_p^* = [p(2+p)]^{1/(2+2p)} \,. \tag{27}$$

   But the second term will dominate for sufficiently small $a > 0$ so that, in fact, $d_p(a) \to \infty$ as $a \to 0^+$. That being said, we observe numerically that for very small p-values, $d_p(a)$ becomes close to the value 0.25 at quite small values of $a > 0$ – see Figure 6 below. (If we consider p to be a perturbation parameter, then perhaps some kind of "boundary layer" behaviour is going on here.)

2. For p negative, but not too negative, i.e., $-2 < a < 0$, the first term is positive which implies that $d_p(a)$ is an increasing function of a (since it is a sum of two increasing functions). Moreover, $d_p(a) \to \infty$ as $a \to \infty$.

   Note also that $d_p(a) \to 0^+$ as $a \to 0^+$. And that being said, we once again observe numerically that for very small $|p|$-values, $d_p(a)$ becomes close to the value 0.25 at quite small values of $a > 0$ – see Figure 8 below.

It would be nice to see if for very small p-values, both positive and negative, the reflected rays are somewhat concentrated near the focal point of the "unperturbed" parabola, $d = \dfrac{1}{4}$. This, once again, leads to the idea of the *distribution* of reflected rays. In the case of the parabola, i.e., $p = 0$, the distribution of reflected rays is a *Dirac delta distribution* at the point $d = \dfrac{1}{4}$.

In Figure 6 below, the positions $d_p(a)$ of the reflected rays on the y-axis are plotted as a function of a in the case $p = 0.01$. From Eq. (27), $d_p(a) < 0$ for $a \geq 6.918$. Recall that negative $d_p(a)$ values are considered invalid: In such cases, the reflected rays will actually hit the perturbed parabola at least a second time, violating the assumptions of our model. We notice that in the region $0 < a < 2$, the values of $d_p(a)$ are rather concentrated around $\dfrac{1}{4}$, the focal point of the "unperturbed" parabola – the graph of $d_p(a)$ decreases quite slowly. For $a > 2$, the graph of $d_p(a)$ decreases more quickly. (This suggests that the *derivative function* $d_p'(a)$ may play a role in characterizing the concentration of values. In fact, it might suggest that the *reciprocal* of the derivative, i.e., $1/d'(a)$ may play a more direct role. We shall indeed return to this idea.

The distribution of the values of $d_p(a)$ plotted in Figure 6 is shown as a rough histogram in Figure 7. We see that the histogram peaks at around d=0.25, the single focal point value for the unperturbed parabola,

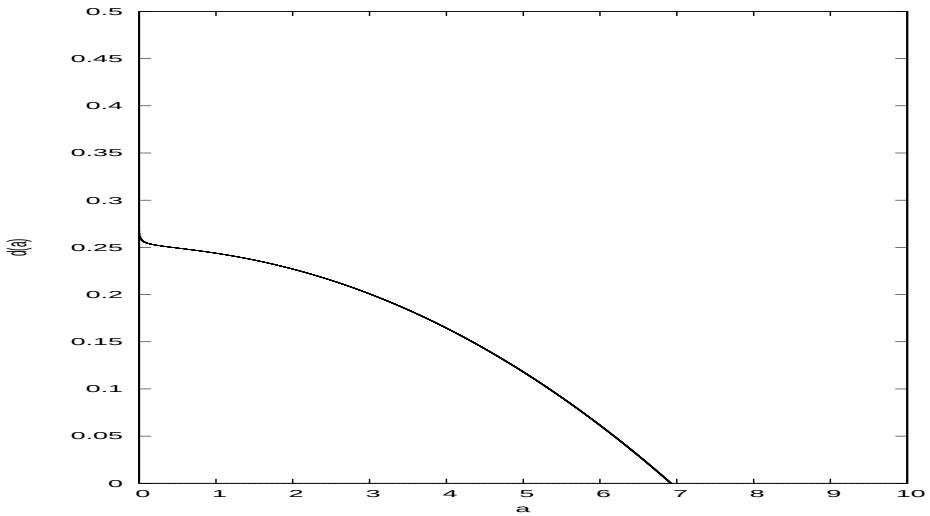

Figure 6: Plot of $d_p(a)$ vs. a for the "perturbed" parabola $f(x) = x^{2+p}$, with $p = 0.01$.

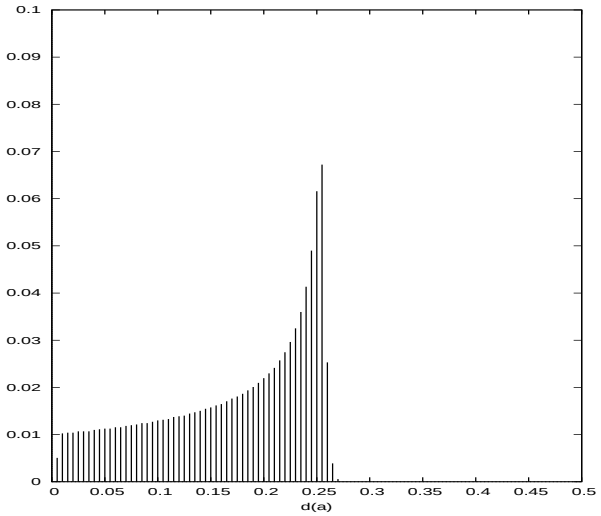

Figure 7: Histogram of values of $d_p(a)$ plotted in Figure 6.

and that the perturbation produces a significant broadening of this peak. We expect that as p increases, the histogram peaks of the distributions of $d_p(a)$ will broaden even further toward the left.

And what about negative p-values ? Let's keep in mind that if $p = -1$, the mirror surface is $f(x) = x$ which was studied in Example 2. In this special case, the distribution of the rays is over the entire non-negative y-axis. (In fact, the distribution is uniform over $(0, \infty)$ since d is a linear function of a.) From Eq. (17), we have that $d = a$.

In Figure 8 below, the positions $d_p(a)$ of the reflected rays on the y-axis are plotted as a function of a in the case of a slight negative perturbation, i.e., $p = -0.01$, and for $0 < a \leq 10$. We notice that in the region $0 < a < 2$, the values of $d_p(a)$ are rather concentrated around the value $\dfrac{1}{4}$, the focal point of the "unperturbed" parabola. In this case, however, and in contrast to the case for $p = 0.01$, the graph of $d_p(a)$ *increases* quite slowly. For $a > 2$, the graph of $d_p(a)$ increases more quickly. The distribution of the values of $d_p(a)$ shown in the plot in Figure 8 is presented in the histogram in Figure 9. We see that the histogram peaks at d=0.25, the single focal point value for the unperturbed parabola, and that the perturbation produces a broadening of this peak toward the right. The histogram in Figure 9 was truncated artificially because we computed $d(a)$

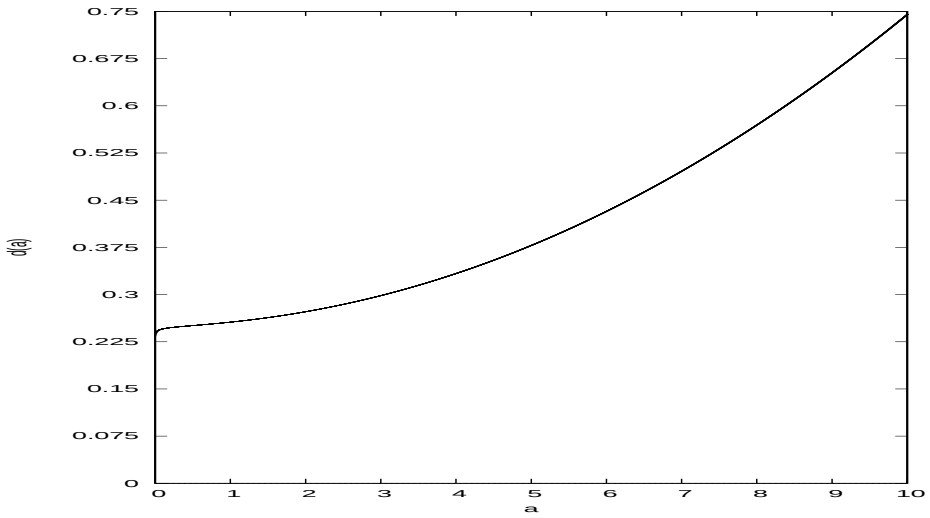

Figure 8: Plot of $d_p(a)$ vs. a for the "perturbed" parabola $f(x) = x^{2+p}$, with $p = -0.01$.

values only up to 0.75. Given that the graph of $d_p(a)$ is increasing for all $a > 0$, the histogram will, at least theoretically, continue indefinitely on the positive $d(a)$-axis.

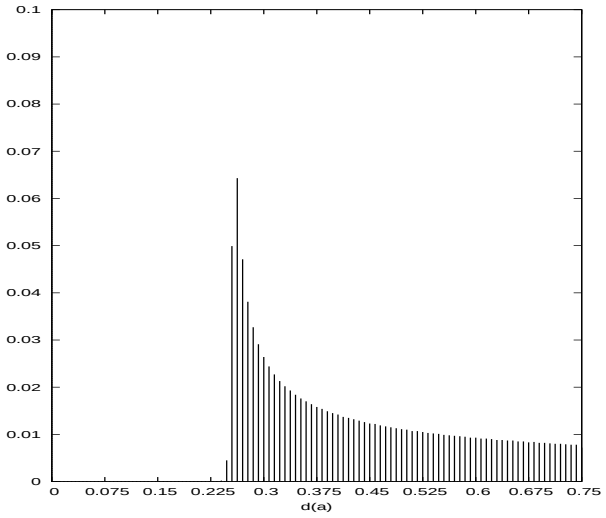

Figure 9: Histogram plot of values of $d(a)$ shown in Figure 8.

At this point, we comment on the qualitative difference between the plots of $d_p(a)$ in Figure 6, where $p = 0.01$, and Figure 7, where $p = 0.01$. As $|p| \to 0$, we expect the graphs of $d_{|p|}(a)$ and $d_{-|p|}(a)$ to approach the limiting horizontal curve $d_0(a) = \dfrac{1}{4}$, the single focal value of the "unperturbed" parabolic mirror function, $f(x) = x^2$.

Continuing our discussion of negative p-values, we expect that as $|p|$ increases, i.e., becomes more negative, the histograms of the distributions of $d(a)$ will broaden even further to the right. As p approaches -1, we expect the distributions to approach the uniform distribution over the interval $[0, \infty)$ which is associated with the linear mirror function $f(x) = x$.

The question of what happens for p-values of even larger magnitude, positive or negative, is a very interesting

one, and we may return to it in the future. As a kind of "teaser," we consider the value $p = 0.5$, i.e., the mirror surface is given by $f(x) = x^{5/2}$. The positions $d_p(a)$ of the reflected rays on the y-axis are plotted as a function of a in Figure 10 below. From Eq. (27), $d_p(a) < 0$ for $a > 0.9283$. (Recall that negative values of $d(a)$ are considered invalid.) In this case, there is a more noticeable "blowup" of $d_p(a)$ as $a \to 0^+$. (Numerically, we find that $d_p(0.0002) \cong 14.14$.)

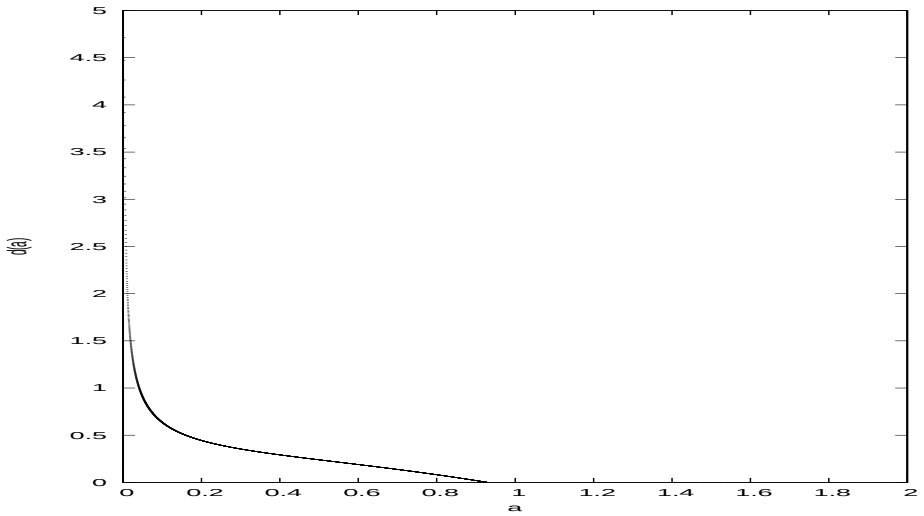

Figure 10: Plot of $d(a)$ vs. a for the "perturbed" parabola $f(x) = x^{2+p}$, with $p = 0.5$.

The distribution of the values of $d(a)$ shown in the plot in Figure 10 is shown as a histogram in Figure 11 below. We note that the distribution is now much more diffuse – in comparison with the case $p = 0.01$ from Figure 7, there is a significant increase of values attained near zero as well as a spreading of the distribution beyond $d = 1.0$.

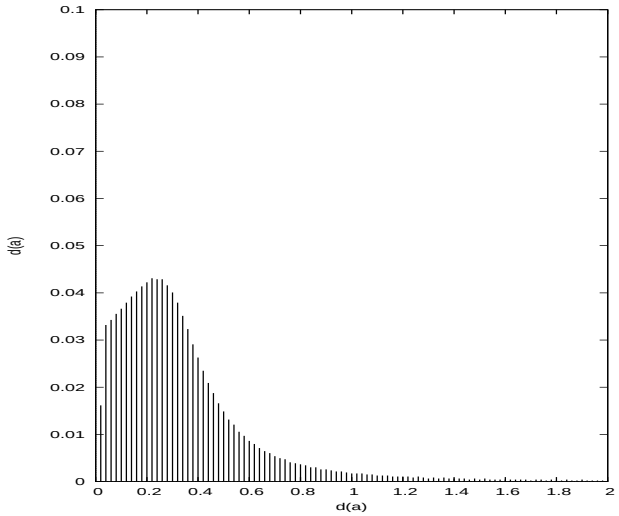

Figure 11: Histogram plot of values of $d(a)$ shown in Figure 10.

As mentioned earlier, it would be very interesting to continue this investigation. For example, recall that for r very large and positive, the function $f(x) = x^r$ approaches the shape of an infinite square well, with small values over the interval $(-1, 1)$ and large values outside it.

We now comment very briefly on the case of negative p-values with larger magnitudes. Recall that the case $p = -1$ corresponds to the linear mirror in Example 2. When $p < -1$, the mirror function $f(x)$ is no longer convex but concave. For example, when $p = -1.5$, $f(x) = \sqrt{x}$. A plot of reflection values $d(a)$ for this case is shown in Figure 11 below. The plot appears to be a slight "perturbation" of the linear case associated with the linear mirror surface $f(x) = x$. A histogram plot of the distribution of $d(a)$ values in Figure 12 is presented

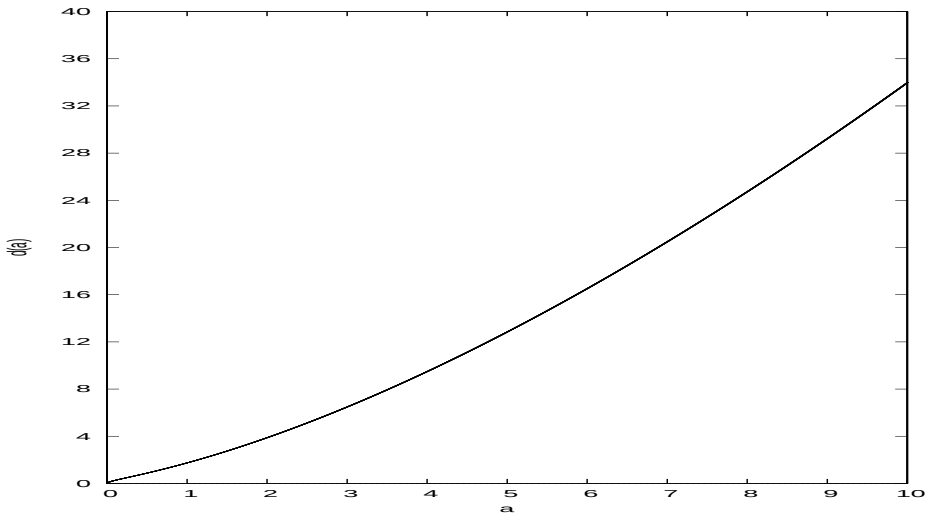

Figure 12: Plot of $d(a)$ vs. a for the concave mirror function $f(x) = \sqrt{x}$ corresponding to $p = -1.5$.

in Figure 13. One could view this histogram as a "perturbation" of the uniform distribution, with the most significant perturbation occurring over the interval $a \in [0, 2]$. Histogram values show that there is a peak of the $d(a)$ distribution at around 0.67. As $a \to 0^+$, the distribution goes to zero. This is presumably due to the extremely high values of the slope of the mirror surface function $f(x) = \sqrt{x}$ as $x \to 0^+$. Only at sufficiently large values of a does the slope of the function $f(x) = \sqrt{x}$ decrease to the point that a reasonable concentration of reflected rays can be produced on the y-axis.

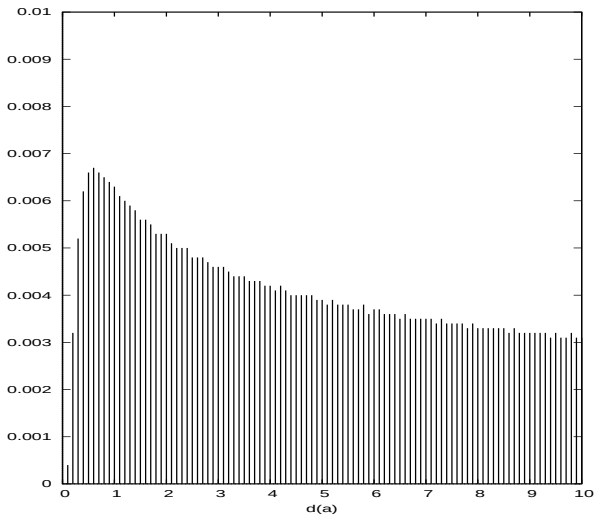

Figure 13: Histogram plot of values of $d(a)$ shown in Figure 12 for the concave mirror function $f(x) = \sqrt{x}$.

A closer examination of the distributions of reflected ray values $d(a)$

In the following discussion, we shall be concerned with the distribution of values of a real-valued function $g(x)$ where x is restricted to some appropriate domain of definition, typically an interval $[a, b] \subset \mathbb{R}$ to be denoted as $\mathcal{D}_g$. We'll define

$$\mathcal{R}_g = \{ y = g(x) \ \ \forall x \in \mathcal{D}_g \} . \tag{28}$$

Note that in this section we use "$g(x)$" to denote our function of concern – as opposed to "$d(a)$", which has been used to denote the y-intercepts of reflected rays.

The simplest case to consider is the constant function $g(x) = d$, where $d \in \mathbb{R}$, for all $x \in \mathcal{D}_g$. This situation is sketched in the plot at the left in Figure 14 below.

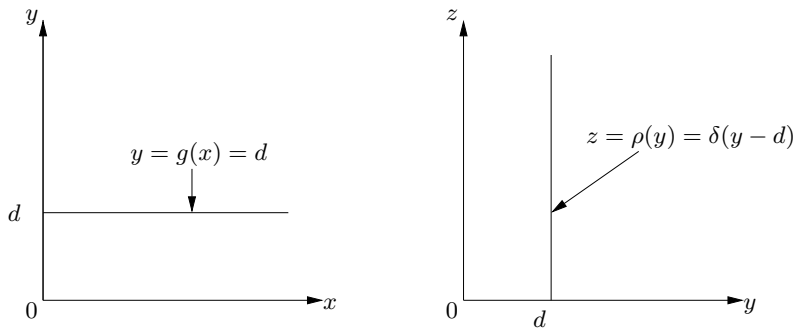

Figure 14. A constant function $g(x) = d$ and its associated value-distribution function, $\rho(y)$, which is the Dirac delta function $\delta(y - d)$.

Clearly, $g(x)$ assumes only one value, namely d, i.e., $\mathcal{R}_g = \{d\}$. The distribution of values assumed by g could not be simpler – it consists of the single point d. If we look at the graph of $g(x)$ from the side, considering the y-axis as our independent variable axis, we see a single peak at $y = d$. This has the "aroma" of a Dirac delta distribution. Indeed, the (normalized) distribution associated with g is the "Dirac delta function" supported on d, usually written as as $\delta(y - d)$.

The reader will see that this situation applies to Example 1 studied above, i.e., the parabolic mirror function $f(x) = rx^2$ for which the reflected ray function $d_r(a) = \dfrac{1}{4r}$, the single focal point of the parabola. The distribution of reflected rays is the Dirac delta function, $\delta\left(y - \dfrac{1}{4r}\right)$.

Let us now suppose that $g(x)$ assumes a finite number of distinct values d_i, $i \leq 1 \leq N$, for all $x \in \mathcal{D}_g$. Then $\mathcal{R}_g = \{d_1, \cdots, d_N\}$ is a finite set. The (normalized) distribution of values assumed by g will have the form

$$D(y) = \sum_{k=1}^{N} p_i \delta(y - d_i) , \tag{29}$$

where $0 \leq p_k \leq 1$ for all $1 \leq k \leq N$, and

$$\sum_{k=1}^{N} p_k = 1 . \tag{30}$$

Each p_k, $1 \leq k \leq N$, represents the fraction of x-values in $\mathcal{D}_g(x)$ at which $g(x) = d_k$. (If, for some k, $g(x) = d_k$ at only a finite set of $x \in \mathcal{D}_g$, then $d_k = 0$.) The above discussion also extends to the case that the set $\mathcal{R}_g$ is countably infinite, in which N may be replaced by "∞".

Before proceeding, let us state here, perhaps at the risk of over-repetition, that we are not simply concerned with what values $g(x)$ assumes for $x \in \mathcal{D}_g$. We are also concerned with the relative **distribution** of these values,

i.e., are there some values which are achieved more often than others?

In the more general case, indeed, the situation encountered with our reflected ray problems, the function $g(x)$ assumes a *continuum* of values as opposed to a finite number of values – typically, $\mathcal{R}_g$ is an interval $[c,d] \subset \mathbb{R}$ and therefore an uncountable set. Once again, we have characterized the set of values achieved by $g(x)$ for $x \in \mathcal{D}_g$. But what about the relative distribution of these values? Are there some which are achieved more often than others? To answer this question in such cases, it may be possible to characterize the distribution of g-values over $\mathcal{R}_g$ in terms of a non-negative **normalized density function** $\rho(y)$, $y \in \mathcal{R}_g$ satisfying the property,

$$\int_{\mathcal{R}_g} \rho(y)\, dy = \int_c^d \rho(y)\, dy = 1\,. \tag{31}$$

Let us provide a rather "loose" definition of $\rho(y)$. For simplicity, we'll assume that $g(x)$ is continuous or at least piecewise continuous. We consider the (uncountably infinite) set of all values achieved by $g(x)$ for $x \in \mathcal{D}_g$, namely, the interval $\mathcal{R}_g = [c,d]$. Because we are working with uncountable sets of values, we cannot consider particular values achieved by $g(y)$. Instead, we must consider (uncountable) sets of values, typically intervals of values achieved by $g(y)$. To illustrate, consider a subset $[r,s] \subset [c,d]$. We are interested in the following fraction,

$$f_{[r,s]} = \frac{\text{"size" of the set of all } x \in \mathcal{D}_g \text{ for which } f(x) \in [r,s]}{\text{"size" of the set } \mathcal{D}_g = [a,b]}\,. \tag{32}$$

For "size," we can use the usual Lebesgue measure so that the denominator is simply $b-a$. Since g is assumed to be at least piecewise continuous, the numerator should be expressible as a (usually but not necessarily finite) sum of lengths of subintervals, $[r_k, s_k]$, which are mapped to the interval $[r,s]$ by g We then claim that the normalized density function $\rho(y)$ satisfies the equation,

$$f_{[r,s]} = \int_r^s \rho(y)\, dy \quad \forall [r,s] \subseteq [c,d]\,. \tag{33}$$

Note that

$$f_{[c,d]} = \int_c^d \rho(y)\, dy = 1\,. \tag{34}$$

The density function $\rho(y)$ can be viewed as a weighting function – it will have higher values over intervals $[r,s]$ with values that are assumed by $g(x)$ more often than other intervals (of the same size).

Returning to our reflected ray problems studied in the above examples, we claim that the histograms presented in Figures 5,7,9,11 and 13 are (up to a constant) piecewise constant approximations to the density functions $\rho(y)$ associates with each particular problem. The histogram values were obtained by computing the fractions $f_{[y_{k-1},y_k]}$ where the subintervals $[y_{k-1}, y_k]$ were the "bins" used to count $d(a)$ values.

Let us now show that in the case that the function $g(x)$ is not only continuous but continuously differentiable, the function $\rho(y)$ can be related to the derivative function $g'(x)$. Our discussion is not intended to be "totally rigorous" and we shall have to make some additional assumptions as we proceed. In Figure 15 is sketched the graph of a function $g(x)$ over an interval $[a,b]$ which we shall consider to be $\mathcal{D}_g$. One is tempted to use the adjective "typical" to modify the word "function" in the previous sentence but this is actually not the case. The reader will note that $g(x)$ appears to be monotonically increasing on $[a,b]$, implying that it is a 1-1 function from $\mathcal{D}_g$ to the interval $[c,d]$ which we shall consider to define $\mathcal{R}_g$. We'll consider this rather special case first and then extend our results to the more general nonmonotonic case.

With reference to Figure 15, consider an infinitesimal element of width dx situated at (centered at?) a point $p \in (a,b)$. For the moment, we shall also assume that $g'(x) \neq 0$ for all $x \in \mathcal{D}_g$ which, of course, implies that $g'(p) \neq 0$. The element dx is mapped to an infinitesimal element of width $dy = |g'(p)|\, dx$ situated at the

point $y = g(p)$. This implies that

$$dx = \frac{1}{|g'(p)|}\, dy\,. \tag{35}$$

Since dy is situated at the point $q = g(p)$, we may rewrite the above equation as follows,

$$dx = \rho(q)\, dy \qquad \text{where} \quad \rho(q) = \rho(g(p)) = \frac{1}{|g'(p)|}\,. \tag{36}$$

Conversely, and perhaps more important, consider an infinitesimal element of g-values of width dy centered at the point $y = q = g(p)$. Then the infinitesimal element centered at the point $x = p$ and which is mapped to dy is given by dx in Eq. (36). Note that a **smaller value** of $|g'(p)|$ which implies, by the assumed continuity of $g'(x)$, smaller values of $|g'(x)|$ in a neighbourhood of p, implies a **larger value** of $\rho(g(p))$ which, in tern, implies a larger value for dx, i.e., a larger neighbourhood of values which are mapped by g to a neighbourhood of $q = g(p)$. Conversely a **larger value** of $|g'(p)|$ will be associated with a **smaller value** for dx, i.e., a smaller neighbourhood of values which are mapped by g to a neighbourhood of $q = g(p)$.

Recall that this behaviour was conjectured in our observations and subsequent discussions of Figure 4 – the $d(a)$ curve – and Figure 5 – the associated $\rho(y)$ curve of Example 3. But we are getting ahead of ourselves.

Some additional concerns

There remain a few technicalities that require attention. As before, the following discussion is not intended to be "totally rigorous."

1. "What happens if $g'(x) = 0$ at a point p, several points $\{p_k\}_{k=1}^{K}$ or possibly an infinity of points, either countable, i.e., $\{p_k\}_{k=1}^{\infty}$, or uncountable, i.e., $p \in [r, s] \subset \mathcal{D}_g$?"

   From Eq. (36), such points produce singularities in the density function $\rho(y)$. If $g'(x) = 0$ over an interval $I = [r, s] \subset \mathcal{D}_g$, then $g(x)$ is constant over I. From an earlier discussion, this will imply that our density function has a Dirac delta function component. Since the examples that we shall be studying do not involve this case, nor the case of an infinity of points at which $g'(x) = 0$, we shall not consider these cases any further.

   In the case of a single point p at which $g'(p) = 0$, it is quite likely that $\rho(y) \to \infty$ as $y \to g(p)$. (Recall that $g'(x)$ is assumed to be continuous.) However, this does not imply that $\rho(y)$ will not be integrable. We shall encounter this case in a couple of our Examples revisited below.

2. "What happens if a point $q \in \mathcal{R}_g$ has several preimages, i.e., g is not one-to-one on $\mathcal{R}_g$?"

   In the case that a given value $s \in \mathcal{R}_g$ has several preimages, i.e., $r_1, \cdots, r_K$ such that $g(r_1) = \cdots g(r_K) = s$, the value of $\rho(s)$ will be the sum of the individual contributions at the r_k, i.e.,

$$\rho(s) = \sum_{k=1}^{K} \frac{1}{|g'(r_k)|}\,. \tag{37}$$

   Those familiar with measure-preserving transformations will see an analogy here.

We now apply the above results to our reflected ray problem. Here, the function $g(x)$ becomes the reflected ray (y-intercept) function $d(a)$.

Example 1 revisited: The mirror surface is given by the parabolic function $f(x) = rx^2$, where $r > 0$ is a constant, with single focal point at $y = \dfrac{1}{4r}$ so that $d_r(a) = \dfrac{1}{4r}$. Here, we could consider any interval $[a_1, a_2]$ as

the domain $\mathcal{D}_g$ of g, or even the entire real line $\mathbb{R}$. The range of g consists of a single point, $\mathcal{R}_g = \left\{\frac{1}{4r}\right\}$. As such, our density function $\rho(y)$ is defined only at $y = \frac{1}{4r}$. That being said, we could define $\rho(y)$ to be zero at any $y \neq 0$. At $y = 0$, however, $\rho(0)$ has contributions of the form $\frac{1}{|d'(a)|}\infty$ for all $a \in \mathbb{R}$. This means that $\rho(0)$ is "infinitely infinite"! As discussed earlier, $\rho(y)$ the "Dirac delta function" at $y = \frac{1}{4r}$ which is not a function but a "distribution". Distributions are understood in terms of integration.

Example 2 revisited: The mirror surface given by the linear function $f(x) = rx$, where $r > 0$ is a constant. As expected from a geometrical argument, the reflected ray function $d_r(a)$, given in Eq. (14), is a linear function of a. This implies that

$$d_r(a) = r - \frac{1}{2} + \frac{1}{2r}\,, \tag{38}$$

a constant. This, in turn, implies that the density function $\rho(y)$ is constant, i.e.,

$$\rho(y) = \left[r - \frac{1}{2} + \frac{1}{2r}\right]^{-1}, \quad y \geq 0\,. \tag{39}$$

(Any positive constant will do. Typically, one uses $\rho(y) = 1$ unless there is a need to normalize it over a given interval.) This implies that the reflected rays are distributed uniformly over the non-negative y-axis.

Example 3 revisited: The circularly-shaped mirror function $f(x)$ with radius R, as given in Eq. (15) with associated reflected ray function $d_R(a)$ in Eq. (19). From Eq. (36), the density function associated with $d_R(a)$ is defined as follows,

$$\begin{aligned} \rho(d_R(a)) &= \frac{1}{|d_R'(a)|} \\ &= \frac{2}{R^2}\frac{(R^2 - a^2)^{3/2}}{a}\,, \qquad -\frac{\sqrt{3}}{2}R \leq a \leq \frac{\sqrt{3}}{2}R\,. \end{aligned} \tag{40}$$

In Figure 14 is presented a plot of ρ-values which were computed using Eqs. (19) and (40). One hundred values of a between 0.0 and $R = 1$ were used to compute $d_R(a)$ and $\rho(d_R(a))$. The plot is presented as an impulse plot so that it can be compared to Figure 5.

Aside: From a look at Eq. (40), the reader may be asking, "Why doesn't the plot of ρ 'blow up' at $a = 0$ since the RHS of the equation has an a in the denominator?" The answer is that ρ is not evaluated at a but rather at $d_R(a)$. The "blowup" on the RHS of (40) when $a = 0$ produces a "blowup" of ρ at the point $d_R(0) = \frac{R}{2} = \frac{1}{2}$, cf. Eq. (19).

Example 4 revisited: The "perturbed parabola" given by $f_p(x)$ defined in Eq. (24), where we consider only $x > 0$. The reflected ray function $d_p(a)$ given in Eq. (26) from which we can compute its associated density function as follows,

$$\rho(d_p(a)) = \frac{1}{|d_p'(a)|}\,, \tag{41}$$

where

$$d_p'(a) = -\frac{p(2+p)}{2}a^{1+p} - \frac{p}{2(2+p)a^{1+p}} \qquad a > 0\,. \tag{42}$$

In Figure 16 is presented a plot of $\rho(y)$ values for the case $p = 0.01$. One hundred a-values between 0.0 and 10.0 were used to compute this plot using Eqs. (26) and (41). It can be compared with the histogram approximation to $\rho(y)$ shown in Figure 7 which was obtained by "binning". Qualitatively, the two plots compare very well. (One should not worry about the difference in values since $\rho(y)$ is unique up to a constant multiple. The values in Figure 7 were not adjusted to take into consideration the sizes of the "bins.")

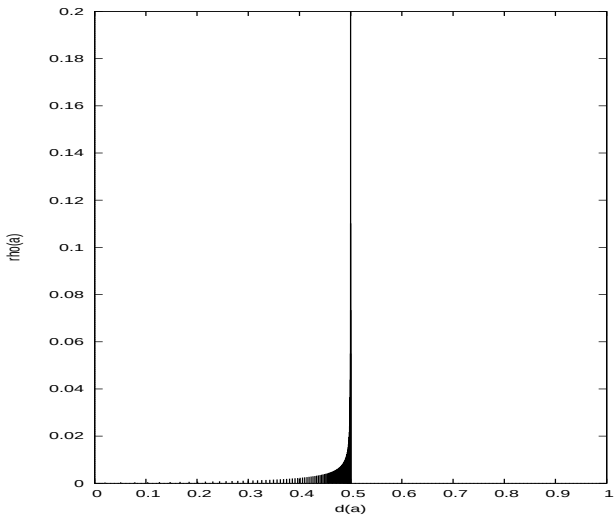

Figure 15: Plot of density function $\rho(y)$ for circular mirror function using Eqs. (19) and (40).

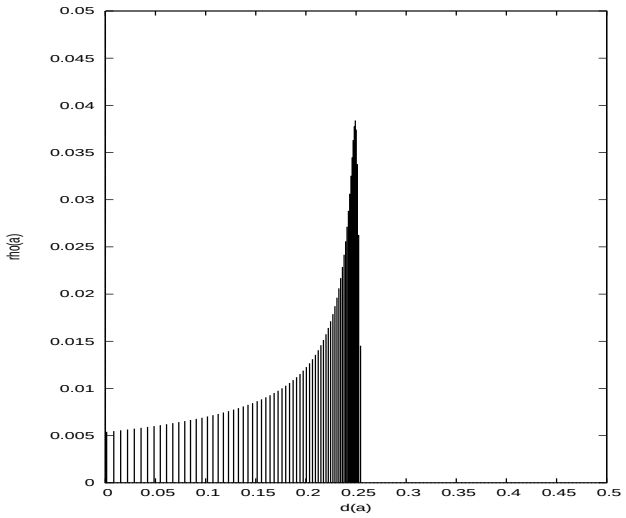

Figure 16: Plot of density function $\rho(y)$ for "perturbed parabola" with $p = 0.01$.

In Figure 17 is presented a plot of $\rho(y)$ values for the case $p = -0.01$. One hundred a-values between 0.0 and 10.0 were used to compute this plot using Eqs. (26) and (41). It can be compared with the histogram approximation to $\rho(y)$ shown in Figure 9 which was obtained by "binning". Once again, the two plots compare very well qualitatively.

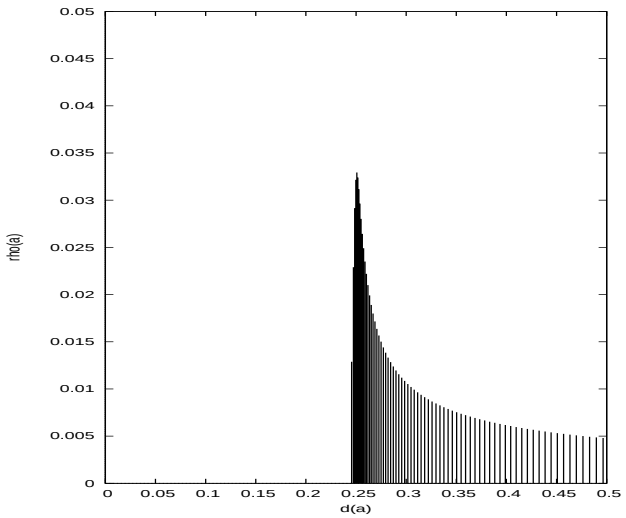

Figure 17: Plot of density function $\rho(y)$ for "perturbed parabola" with $p = -0.01$.

Plots of histograms for the final two cases considered earlier are presented below.

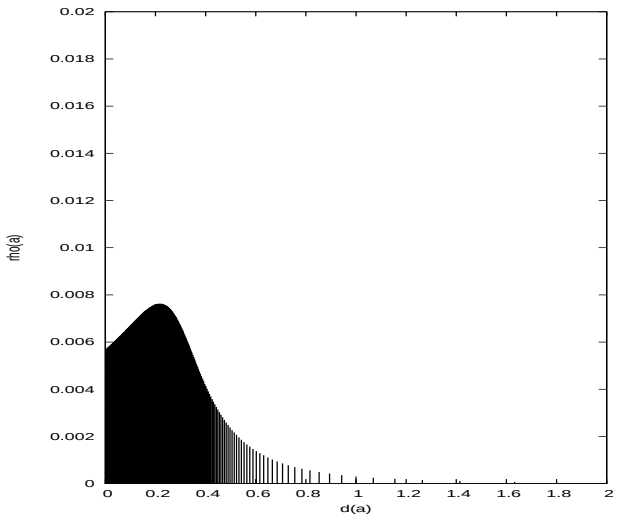

Figure 18: Plot of density function $\rho(y)$ for "perturbed parabola" with $p = 0.5$. Generated with 200 a-values between 0.0 and 1.0. To be compared with Figure 11.

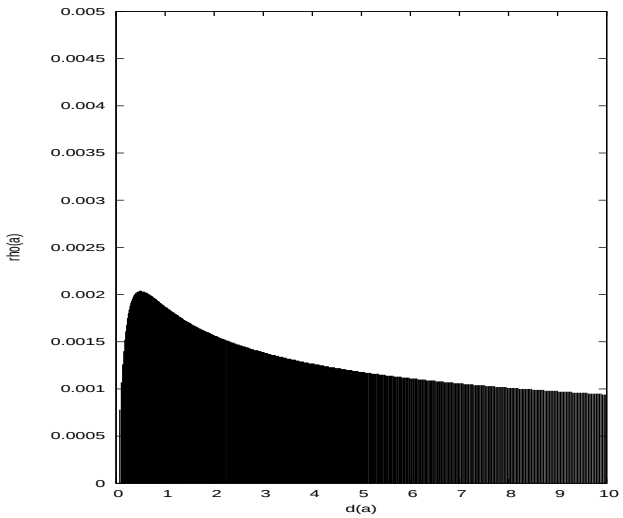

Figure 19: Plot of density function $\rho(y)$ for "perturbed parabola" with $p = -1.5$. Generated with 1000 a-values between 0.0 and 10.0. To be compared with Figure 13.

A direction vector method to derive Eq. (12) along with another important result

Here, we revisit the reflection problem originally sketched in Figure 1 to show how Eq. (12) for $d(a)$, the y-intercept D of the ray reflected at point $(a, f(a))$, can be derived using a method of direction vectors. We shall modify the situation slightly by removing the left component of the mirror, as shown in Figure 20, to allow for the possibility that the reflected ray will intersect the x-axis at point E. This result will also be important our study of solar ovens.

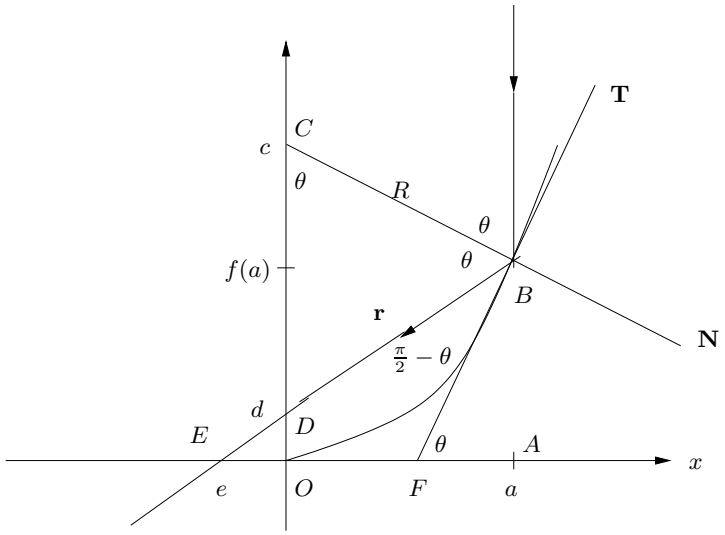

Figure 20

As before, the incident, downward-travelling ray at $x = a$ meets the mirror surface at an angle θ to the normal $\mathbf{N}$ at $(a, f(a))$. The reflected ray $\mathbf{r}$ leaves the point $(a, f(a))$ at an angle θ to $\mathbf{N}$.

Note that the angle between the incident ray and the tangent vector $\mathbf{T}$ at $(a, f(a))$ is $\frac{\pi}{2} - \theta$. This implies that

1. $\angle FBA = \dfrac{\pi}{2} - \theta$,

2. $\angle DBF = \dfrac{\pi}{2} - \theta$.

Recall that for any vector $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$, its unit direction direction vector $\hat{\mathbf{v}}$ can be written in the form

$$\hat{\mathbf{v}} = (\cos\gamma, \sin\gamma)\,, \tag{43}$$

where γ, its angle of orientation, is the angle which $\mathbf{v}$ makes with respect to positive x-axis measured counter-clockwise (when the tail of $\mathbf{v}$ is placed at the origin). This implies that the angle of orientation of the incoming ray at $x = a$ – with unit vector $\hat{\mathbf{v}} = (0, 1)$ – is $\gamma = \dfrac{3\pi}{2}$.

The orientation angle ϕ of the reflected ray $\mathbf{r}$, keeping in mind that it is travelling leftward in Figure 20, is given by

$$\phi = \frac{3\pi}{2} - 2\left(\frac{\pi}{2} - \theta\right) = 2\theta + \frac{\pi}{2}\,. \tag{44}$$

This implies that the unit direction vector $\hat{\mathbf{r}}$ corresponding to $\mathbf{r}$ is given by

$$\begin{aligned} \hat{\mathbf{r}} &= \left(\cos\left(2\theta + \frac{\pi}{2}\right), \sin\left(2\theta + \frac{\pi}{2}\right)\right) \\ &= (-\sin 2\theta,\, \cos 2\theta) \\ &= (-2\sin\theta\cos\theta, \cos^2\theta - \sin^2\theta)\,. \end{aligned} \tag{45}$$

We now make the connection between θ and $f'(a)$ The slope of the tangent vector $\mathbf{T}$ is

$$m_T = \tan\theta = \frac{\sin\theta}{\cos\theta} = f'(a)\,. \tag{46}$$

By drawing an appropriate right triangle with sides of length 1 and $f'(a)$ and hypotenuse of length $\sqrt{1 + f'(a)^2}$, we find that

$$\sin\theta = \frac{f'(a)}{\sqrt{1 + f'(a)^2}}\,, \qquad \cos\theta = \frac{1}{\sqrt{1 + f'(a)^2}}\,, \tag{47}$$

Substitution of these results into Eq. (45) yields

$$\hat{\mathbf{r}} = \left(-\frac{2f'(a)}{1 + f'(a)^2}, \frac{1 - f'(a)^2}{1 + f'(a)^2}\right)\,. \tag{48}$$

This implies that the reflected ray may be written in parametric form as follows,

$$\mathbf{r} = (x(t), y(t)) = (a, f(a) + t\left(-\frac{2f'(a)}{1 + f'(a)^2}, \frac{1 - f'(a)^2}{1 + f'(a)^2}\right)\,, \quad t \geq 0\,. \tag{49}$$

The components of $\mathbf{r}$ can be expressed separately as follows,

$$\begin{aligned} x(t) &= a - \frac{2f'(a)t}{1 + f'(a)^2} \\ y(t) &= f(a) + \frac{t(1 - f'(a)^2)}{1 + f'(a)^2}\,. \end{aligned} \tag{50}$$

We can now easily compute the coordinates of the x- and y-intercepts of the reflected ray $\mathbf{r}$ (assuming that they exist).

Computation of y-intercept $d(a)$: We find t^* such that $x(t^*) = 0$ and then compute $d(a) = y(t^*)$.

$$x(t^*) = 0 \quad \Longrightarrow \quad \frac{2f'(a)t^*}{1 + f'(a)^2} = a \quad \Longrightarrow \quad t^* = \frac{a(1 + f'(a)^2}{2f'(a)}\,. \tag{51}$$

Substitution of $t = t^*$ into Eq. (50) yields

$$\begin{aligned} d(a) = x(t^*) &= f(a) + \frac{a(1+f'(a)^2}{2f'(a)} \cdot \frac{1-f'(a)^2}{1+f'(a)^2} \\ &= f(a) + \frac{a}{2f'(a)} - \frac{a}{2} f'(a)\,. \end{aligned} \tag{52}$$

This is in agreement with the result in Eq. (12) earlier in this report.

Computation of x-intercept $e(a)$: We find t^* such that $y(t^*) = 0$ and then compute $e(a) = x(t^*)$.

$$y(t^*) = 0 \quad \Longrightarrow \quad f(a) = -t^* \frac{1-f'(a)^2}{1+f'(a)^2} \quad \Longrightarrow \quad t^* = \frac{f(a)(f'(a)^2+1)}{f'(a)^2-1}\,. \tag{53}$$

A couple of comments are in order here. First, for leftward reflection to occur, we must have $f'(a) > 0$. Note that the above equation for t^* "blows up" when $f'(a) = 1$. This is the case of an "infinitesimal" mirror inclined at angle $\dfrac{\pi}{4}$ to the x-axis. In this case, the vertical, downward light ray will be reflected horizontally, never intersecting the x-axis. For the reflected ray to intersect the x-axis, we must have $f'(a) > 1$. In this case t^* is positive and finite.

Substitution of $t = t^*$ into Eq. (50) yields, after a little "tidying," the final result,

$$\boxed{e(a) = y(t^*) = a - \frac{2f(a)f'(a)}{f'(a)^2-1}\,, \qquad f'(a) > 1\,.} \tag{54}$$

This result will be used in another report. That being said, let's provide a kind of teaser/preview for why it is useful.

In Figure 21 is sketched a schematic and scaled side view of a "solar oven" that has been constructed by the author. (A photo of the actual oven is shown in Figure 22.) The oven is represented by the square $BEFC$ which lies below the interval $-\dfrac{1}{2} \le x \le \dfrac{1}{2}$ on the x-axis. The two sides, $\overline{BE}$ and $\overline{FC}$, and bottom, $\overline{EF}$, of the oven, all of unit length, are insulated. The top of the oven is dotted since it represents a glass sheet which allows light to enter the oven in order to heat its interior. (What happens to light rays entering the oven from the top is another matter which beyond the scope of the present discussion.) The left and right line segments lying at angles $\dfrac{2\pi}{3}$ and $\dfrac{\pi}{3}$, respectively, to the (positive) x-axis, represent mirrors which reflect downward travelling light rays into the oven. Each of these two mirrors is also one unit in length.

The (common) length of the mirrors, i.e., unity, along with their angles of orientation with respect to the horizontal axis were chosen on the basis of two competing factors:

1. Try to collect as many downward travelling (sun)light rays as possible, i.e., make the horizontal distance between endpoints A and D as large as possible. For a given (equal) set of mirror lengths, $\overline{AD}$ is increased by rotating each of the sides toward the x-axis. For a given set of orientation angles, $\overline{AD}$ is increased by increasing the mirror lengths.

2. Try to use as little mirror material as possible, i.e., make the (equal) lengths of the two sides, $\overline{AB}$ and $\overline{CD}$, as small as possible but still guaranteeing that all light hitting the sides is reflected into the oven through the (glass) interval BC after only one contact with the mirrors.

We need to check only how the downward rays hitting points A and D are reflected according to Eq. (54). Point D is situated at $\left(1, \dfrac{\sqrt{3}}{2}\right)$. Hence $a = 1$ and $f(a) = \dfrac{\sqrt{3}}{2}$. The slope of the (flat) mirror represented by the line segment $\overline{CD}$ is $f'(a) = \sqrt{3}$. Substituting these values into Eq. (54) yields

$$e(a) = 1 - \frac{(2)(\frac{\sqrt{3}}{2})(\sqrt{3})}{(\sqrt{3})^2 - 1} = 1 - \frac{3}{2} = -\frac{1}{2}\,. \tag{55}$$

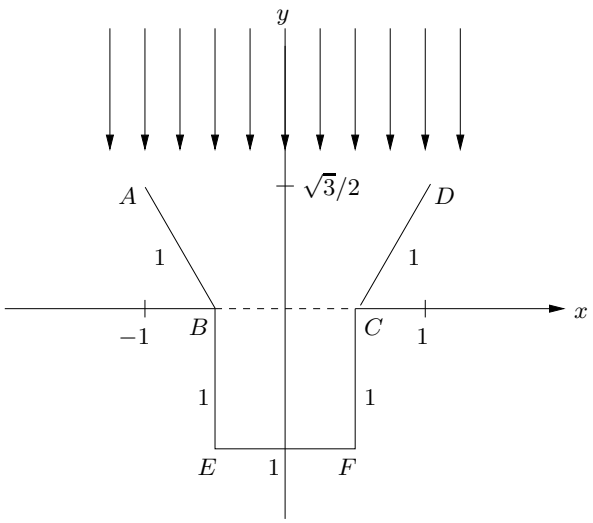

Figure 21. Schematic side view of a solar oven constructed by the author.

Therefore, the ray which hits point D is reflected to point B. This may be considered as a kind of "boundary point", as is point C. We conclude that all rays hitting the mirror between D and C are reflected to the glass surface between B and C, respectively. Mirror CD is sufficiently long to reflect all incoming rays to the glass surface BC. If we make CD longer by moving point D appropriately outward and upward (preserving the slope $\sqrt{3}$, downward travelling rays for which $a > 1$ will, after reflection, hit mirror AB and then be reflected outward/upward.

By symmetry, we can conclude that all rays hitting the mirror between A and B are reflected to the glass surface between B and C. (The reader is welcome to check this by explicit calculation.) Therefore, "Mission accomplished!"

Figure 22: Actual solar oven constructed by author. In order to compare it with Figure 21, visualize turning it and looking at it from the back.

Another note: Well after the above manuscript was completed, the author noticed that the result in Eq. (54) for the x-intercept of the reflected ray can, in fact, be also obtained using straightforward geometry. In other words, the direction vector method was not really necessary. But it's always good to be able to derive a result with more than one method.

Returning to our diagram in Figure 20 and examining $\triangle BEF$, we note that

$$\angle BEF = \pi - (\pi - \theta) - \left(\frac{\pi}{2} - \theta\right) = 2\theta - \frac{\pi}{2}\,. \tag{56}$$

Let us now consider $\triangle ABE$:

$$\frac{|\overline{AB}|}{|\overline{AE}|} = \frac{f(a)}{a-e} = \tan\left(2\theta - \frac{\pi}{2}\right) = \cot 2\theta\,. \tag{57}$$

It follows that

$$\begin{aligned} a - e &= f(a)\tan 2\theta \\ &= f(a)\,\frac{2\sin\theta\cos\theta}{\cos^2\theta - \sin^2\theta} \\ &= f(a)\,\frac{2f'(a)}{1-f'(a)^2} \qquad \text{from Eq. (47)}\,. \end{aligned} \tag{58}$$

A simple rearrangement yields the desired result,

$$e = a - \frac{2f(a)f'(a)}{f'(a)^2 - 1}\,. \tag{59}$$