Understanding the damped SHM without ODEs

Chiu-king Ng
E-mail: ngchiuking@phy.hk

Abstract

Instead of solving ordinary differential equations (ODEs), the damped simple harmonic motion (SHM) is surveyed qualitatively from basic mechanics and quantitatively by the instrumentality of a graph of velocity against displacement. In this way, the condition $b \geq \sqrt{4mk}$ for the occurrence of the non-oscillating critical damping and heavy-damping is derived. Besides, we prove in the under-damping, the oscillation is isochronous and the diminishing amplitude satisfies a rule of "constant ratio". All are done on a non-ODE basis.

1. Introduction

When displaced and released, a mass will oscillate indefinitely in simple harmonic motion (SHM) if it is acted on by a restoring force obeying Hook's law and free from any dissipative forces. More realistically, a damping force is always present and hence the oscillations will stop eventually. Damped simple harmonic motion is a standard subject in classical mechanics, and a complete analytical solution comes from solving Newton's second law of motion in the form of an ordinary differential equation (ODE) [1]. Harder mathematics seems to be involved, so pre-university physics curricula usually introduce the damping simply by listing the main results with a brief explanation and perhaps an experimental verification, even if the undamped SHM is taught in depth [2, 3]. The aim of this paper is to attempt to fill the gap, we put forward an alternative approach which is mainly geometric, algebraic and argument-based to allow students, after learning the undamped SHM, to have a more instructive introduction to the damped SHM. We reason, from basic mechanics, how the velocity of a damped oscillator varies with position.

Thus, the trajectory of the oscillator in a phase space diagram (velocity against displacement) is *deduced*, and subsequently, some essential features of the damped SHM are derived. Phase space portrait has already been used widely in advanced texts to describe oscillatory motions; nevertheless, we find it a good *tool* to serve our purpose. Instead of using equations to plot, we take the process of deducing the graph as a useful means for pursuing a deeper understanding of the damped SHM.

2. The oscillator

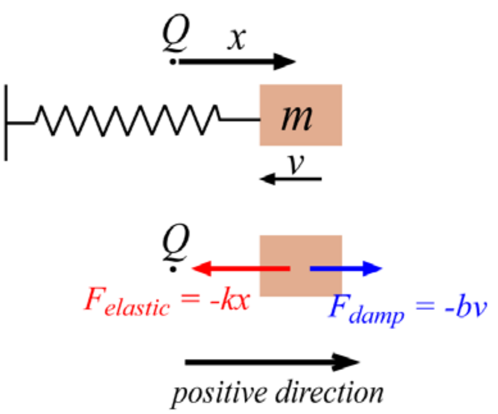

Figure 1. A mass m is acted on by two forces, $F_{elastic} = -kx$ and $F_{damp} = -bv$, where x is the displacement from the equilibrium position Q and v is the velocity. The direction to the right is taken to be positive.

In figure 1, a mass m oscillating horizontally about an equilibrium point Q is subjected to two forces: an elastic spring force ($F_{elastic}$) which always directs towards Q and a damping force (F_{damp}) which always opposes the motion. The elastic force obeys Hooke's law and is expressed as $-kx$, where k is a positive constant and x is the displacement of m from Q. Assumed to be proportional to velocity, the damping force has the form $-bv$, where b is a positive constant giving the degree of damping and v is the velocity of m. Accordingly, the net force acting on m is $F_{net} = -bv - kx$. Here, we only discuss the motion which starts with the initial condition: m is released from rest at a position on the right hand side of the equilibrium position. We take the direction to the right to be positive.

3. The damped motion and its *v*-*x* graph

The total mechanical energy of an undamped SHM is conserved, i.e., $mv^2/2 + kx^2/2 = kA^2/2$, or $(m/k)v^2 + x^2 = A^2$, where A is the amplitude of oscillation. Graphically, this equation draws an ellipse in a graph of v against x. In figure 2, we draw such an ellipse as well as the straight line

$$L: \quad -bv - kx = 0 \qquad (1)$$

on the same *v*-*x* graph. One would recognize the LHS of L is F_{net}. Yes, it is, but surely L is irrelevant to the ellipse because it contains the damping constant b. Our strategy is, by applying the physics principles, to deduce the *v*-*x* curve of a small damping from a modification of the ellipse with the aid of the line L.

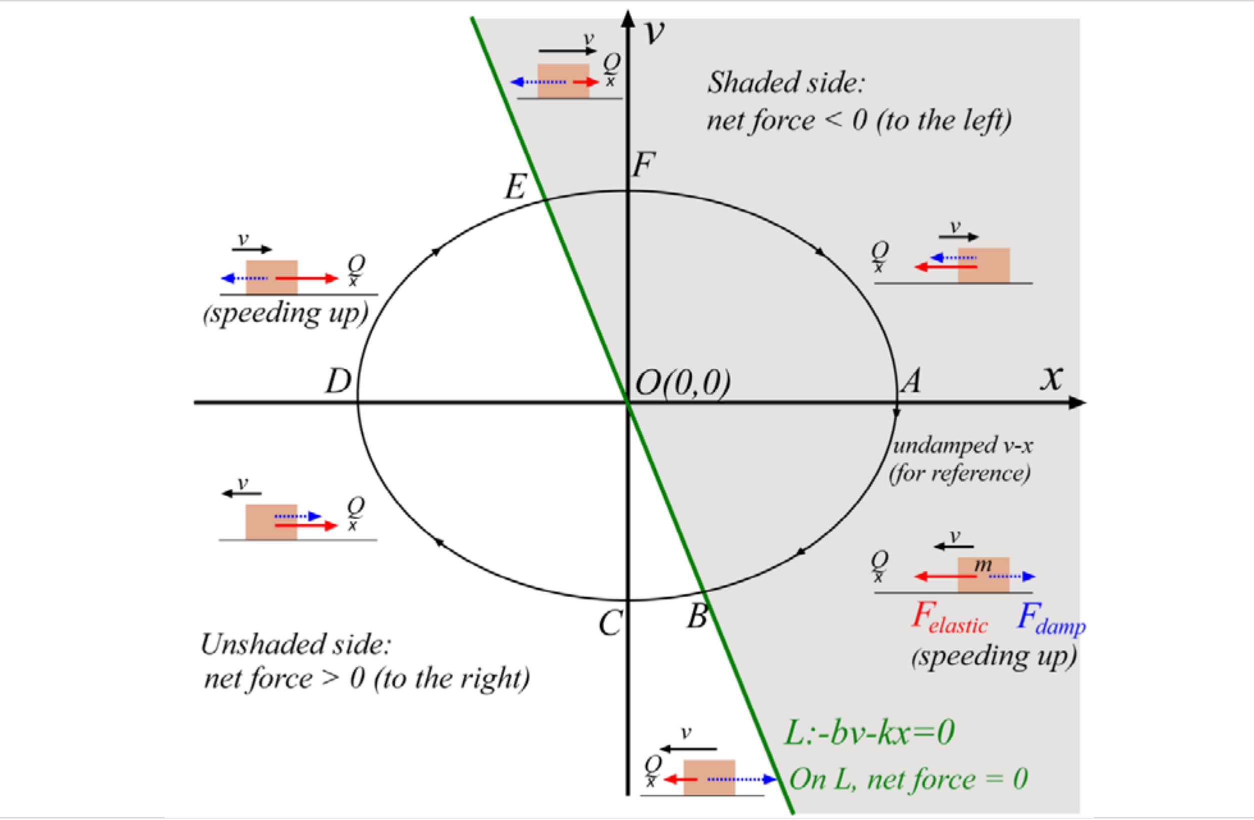

Figure 2. The ellipse represents the *v*-*x* plot of an undamped SHM. The *v*-*x* plot of a damped SHM (not shown) is deduced by taking the ellipse as a reference and using the line L:$-bv - kx = 0$. On the shaded (unshaded) side of L, the net force on m is pointing to the left (right). On L, the net force is zero. In the force diagrams, the red and dashed blue arrows represent the elastic and damping force, respectively. The regions AB and DE correspond to a speeding up of the oscillator, while the other four correspond to a slowing down.

In figure 2, the two axes, together with *L*, divide the whole plane into six regions. The three on the right side of *L* (the shaded half plane) satisfy $-bv - kx < 0$, implying the net force on *m* is pointing to the left when its *v-x* trajectory orbits in this half plane. In contrast, the three regions on the left side of *L* satisfy $-bv - kx > 0$, so the net force is pointing to the right. To help us understand how the net force is produced at different positions, the velocity *v* (black arrow), the two forces $F_{elastic}$ (red arrow) and F_{damp} (dashed blue arrow) are drawn in accordance with their directions and magnitudes in each of the six regions. For instance, in the region *BC, m* is moving to the left but at a position very close to *Q*, the leftward elastic force then is in magnitude smaller than the opposing damping force, yielding a rightward net force.

While using figure 2 to interpret motion, two additional points are worth reminding. First, an object speeds up when it is moving in the same direction as its net force, otherwise it slows down. In view of this, only regions *AB* and *DE* among the six in figure 2 correspond to a speeding up of the oscillator. In the other four regions, the oscillator slows down. Secondly, one should be aware of the difference between the equilibrium position *Q* and the origin *O(0, 0)*. Whenever *m* passes through *Q*, its *v-x* trajectory crosses the *v*-axis at a nonzero intercept. The whole *v*-axis is $x = 0$ (the point *Q*) while the origin *O(0, 0)* means specifically a stop at *Q*.

We reason the damped *v-x* curve as follows. Suppose *m* is released at rest at the right extremity $x = A$, so the curve starts from that point and orbits in a clockwise direction because *v* is negative and *x* is smaller thereafter. The part of the curve at $x = A$ is no different from the ellipse since the damping force is zero ($v = 0$) when the motion starts. With the same notations as the ellipse, the curve will cut *L*, the *v*- axis, the *x*-axis …. at *B*, *C*, *D* …., respectively (see figure 2). Leaving from *A*, the curve orbits downwards, but less rapidly than the ellipse. Its shape becomes flatter and flatter because when *m* gains speed, it becomes more and more difficult to accelerate further under the opposition of the stronger and stronger damping force. The stage of the elastic force prevailing over the damping force ends when the trajectory reaches its intersection with *L* (point *B*), at where $-bv - kx = 0$, meaning the damping force just balances the elastic force. At *B*, the slope of the curve is zero because *m* attains a maximum speed there. After *B*, *m* begins to slow down (*v* becomes less negative), the curve diverts upwards and the speed is reduced to zero at *D*, a smaller *x*-intercept than *A.* In other words, deceleration has already begun before *m* reaching the equilibrium position *Q*. In the case of no damping, acceleration and deceleration occur symmetrically about *Q*. A damping spoils this symmetry: *after release m accelerates slowly over a short distance, deceleration begins before the equilibrium position and the speed is reduced to zero rapidly*, thus causing a smaller *x*-intercept *D*.

From *A* to *D*, the oscillator completes the first half cycle. By applying the arguments similarly, the curve continues for more cycles. For instance, the process of deducing the curve of the next half cycle is basically a repetition of the first: after *D*, the curve orbits upwards with a flatter and flatter shape, it reaches a maximum at *E* and then orbits downwards until cutting the positive *x*-axis vertically again at an even smaller intercept.

With the clues obtained so far, it is ready to make a reasonable *sketch* of the damped *v*-*x* graph, like that in figure 3. The curve spirals around the origin *O(0,0)* with a decreasing amplitude. In section 5, we will show the diminishing amplitude satisfies a rule of "*constant ratio*" and the oscillation is *isochronous*. In the next section, the case of not overshooting beyond the equilibrium position is discussed.

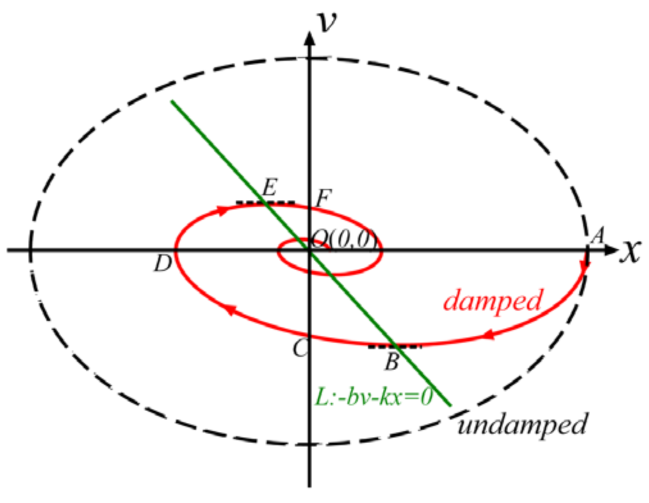

Figure 3. The spiral curve is a *v*-*x* plot of a damped harmonic oscillator. It is sketched under the arguments based on the laws of mechanics. Whenever the curve cuts the line *L*, its slope is zero because the velocity at that moment is constant over a short interval of *x*.

4. Non-overshooting

The previous section outlines the motion of a damped oscillator, whose amplitude diminishes with each oscillation. Based on the scenario described before, one may speculate on what the situation will be if the damping b is very large. Clearly, when b is larger, the line L in figure 3 will be more horizontal (slope of L is $-k/b$), making point B closer to the x- axis; the v-intercept C, being always higher than B, will be much closer to the origin O (see figure 3). As this trend continues further, it is certain that C will get closer and closer to O. What value of b will make C coincide with O? Logically, at this point we can only say: either the damping b is infinitely large or b is greater than a finite value. In the next paragraph, we will find exactly what this value of b should be. Whatever the case, if C is at O, then the v-x trajectory will look like that portrayed in figure 4. It is noteworthy that a v-x trajectory of the damped SHM can reach but cannot pass through the origin $O(0,0)$ since when reaching there both the velocity and acceleration are zero. Such a trajectory implies mass m will come to a final stop when it comes to the equilibrium position Q for the first time, in other words, m will not overshoot to the other side and there will be no oscillations at all.

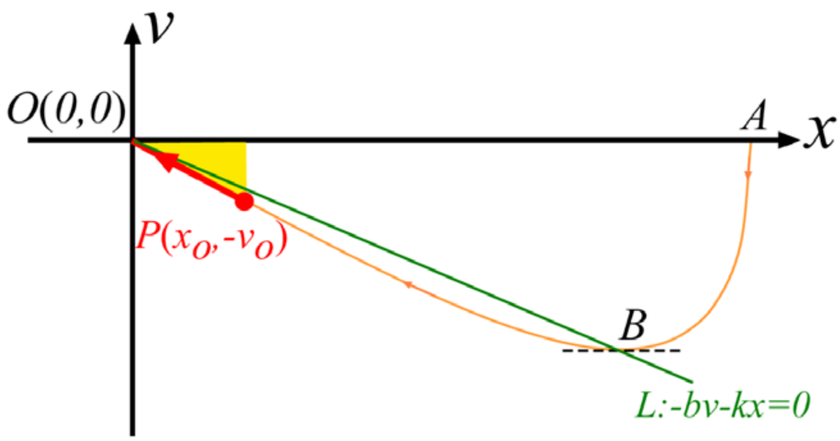

Figure 4. If the degree of damping is too large, it is reasonably believed that the oscillator will move directly towards and stop at the equilibrium position Q. If so, its v-x curve should have the end part OP, where P is a point arbitrarily close to the origin O.

What value of b will cause non-overshooting? We try to figure out the answer by an argument based on the energy-work relationship. Let $P(x_o, -v_o)$ be a point on the towards-origin trajectory in figure 4, where both x_o and v_o are positive. The total mechanical energy of m at that point, $mv_o^2/2 + kx_o^2/2$, must be entirely dissipated via the work done against the damping force over the distance x_o if m stops the first time it comes to the equilibrium position. The area under a F-x graph is work done and the v-x graph is transformed to the F_{damp} –x graph when the constant $-b$ is multiplied to the v-axis, so the work done against the damping force equals the product $b \times$ area bounded by the curve and the x-axis from O to P, where the area is a positive value. In general, this area is not easy to evaluate, except when P is very close to O, then the curve between them can legitimately be regarded as a straight line and the area (the yellow triangle in figure 4) is simply equal to $v_o x_o/2$. In this case, the energy-work condition becomes $mv_o^2/2 + kx_o^2/2 = bv_o x_o/2$, or

$$m(v_o/x_o)^2 - b(v_o/x_o) + k = 0. \qquad (2)$$

Our line of reasoning is that: oscillator does not overshoot provided that its v-x trajectory moves to the origin; equivalently, we say there must be points $P(x_o,-v_o)$ on the trajectory which are arbitrarily close to the origin, so the value v_o/x_o in equation (2) must be real, thus requiring a nonnegative discriminant of equation (2), that is, $(-b)^2 - 4mk \geq 0$, or

$$b \geq \sqrt{4mk}\ . \qquad (3)$$

In brief, non-overshooting occurs only when b is greater than or equal to the finite value $\sqrt{4mk}$. The minimum value of b for oscillator undergoing non-overshooting is $b = \sqrt{4mk}$, the oscillator then is called *critically damped*. When $b > \sqrt{4mk}$, the oscillator is called *heavily damped or over-damped*, which is similar to the critically damped but the oscillator takes a longer time to return to the equilibrium position. It is difficult for the oscillator to gain any speed if it is heavily damped, so it can only return at a very low average speed. When $b < \sqrt{4mk}$, under any circumstances the oscillator overshoots. The oscillator must oscillate a few times before dying out, it is described as *lightly damped* or *under-damped* and the damped motion discussed in section 3 actually belongs to this case.

5. Isochronism and constant ratio

In the previous sections, we have seen the damped *v-x* trajectory always intersects the line *L*: $-bv - kx = 0$ with a zero slope. Now, we prove a more general theorem: the trajectory intersects an origin-passing straight line of equation *L'*: $-\beta v - kx = 0$, where β is a constant of any value, at points with the same slope. The inclination of *L'* can be altered freely by fitting a value of β. Whenever the curve meets *L'*, the net force on *m*, $F_{net} = -bv - kx$ can be combined with the equation of *L'*, giving $F_{net} = (\beta - b)v$. Hence at that moment the acceleration of *m*, $a = F_{net}/m = [(\beta - b)/m]v$. After a very short time interval Δt, the velocity change $\Delta v = a\Delta t = [(\beta - b)/m]v\Delta t$. Because of $v\Delta t = \Delta x$, we get $\Delta v = [(\beta - b)/m]\Delta x$, or $\Delta v/\Delta x = (\beta - b)/m$. The ratio $\Delta v/\Delta x$, which is equal to dv/dx when the changes are infinitely small, is the slope of the *v-x* curve at its intersection with *L'*. Being not dependent on any particular coordinates, the slope is therefore the same at all the intersection points of the curve and *L'*. In particular, when $\beta = b$, slope = 0 and when $\beta = 0$ (the *v*-axis), slope = $-b/m$.

L' is an example of the lines called *isoclines*, which are often used to solve an ODE by a graphical method [4].

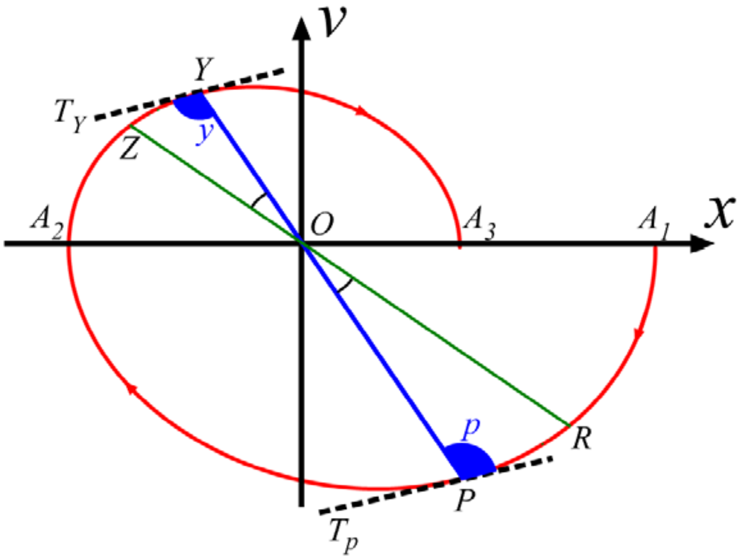

Figure 5. The red curve represents a complete cycle of a lightly damped SHM. The arc from A_1 to A_2 is geometrically similar to that from A_2 to A_3.

Figure 5 shows a *v*-*x* curve of one complete lightly damped oscillation. As the above theorem is applied to the straight line *YOP*, T_Y and T_P, the two tangents to the curve at its two intersection points, are found to be parallel. Hence, $\angle y$ and $\angle p$, the two alternate angles of *YOP* making with T_Y and T_P, are equal. Importantly, line *YOP* is arbitrary, so when it is made to rotate about *O* to any inclination with its two ends, *Y* and *P*, keeping always on the curve, the two alternate angles vary but remain equal. Such a result reflects the fact that the lower and upper arc curve exactly in the same way, meaning they are similar. In terms of the portions of the curve, we find that in figure 5 arcs *YZ* and *PR* are similar; that in figure 6 the solid red arcs (dashed blue arcs) are themselves similar to each other. In addition, the two sector-like figures *OYP* and *OPR* in figure 5 are similar since they have the same vertically opposite angle at *O* and two similar arcs.

The shape of the spiral curve repeats exactly itself for every half cycle as the size of the curve is shrinking. Part of the reason is that the asymmetrical acceleration and deceleration have as a whole completed once for every half cycle.

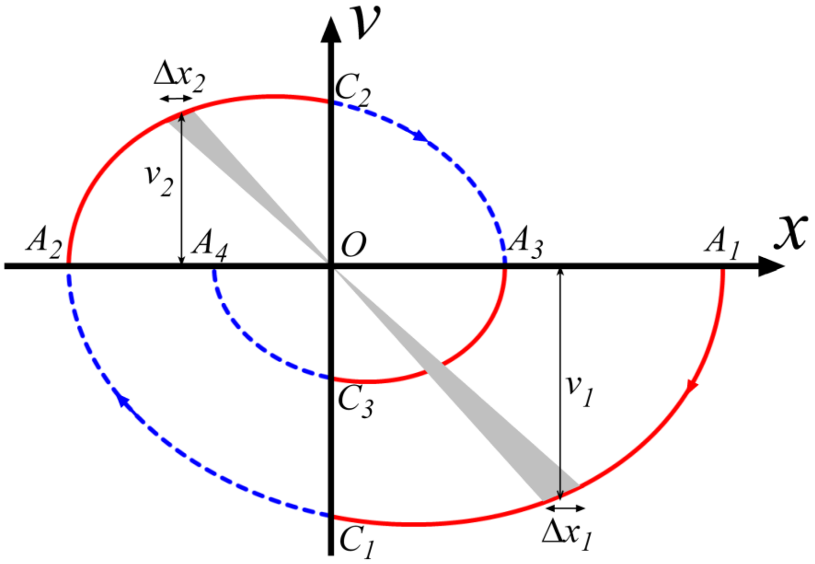

Figure 6. The two shaded thin triangles are similar. The sold red (dashed blue) quarter arcs of the spiral curve are themselves similar to each other.

In figure 6, symbols A_1, A_2, A_3,… are used to denote the amplitudes in the sequence of their appearances on the trajectory. Based on geometric similarity, we are going to prove two characteristic features of the lightly damped SHM.

(1) The two shaded thin triangles in figure 6 are similar, and exactly the same as the two sector-like figures in figure 5 with the intersection angle at *O* made to be infinitely small. In geometry, it is known that the ratios of corresponding sides in two similar figures are equal. Besides, these two infinitely thin triangles are formed by two intersecting lines and two parallel sides, giving their two corresponding sides are in the same ratio as their projections on any line since they make the same angle with that line. Because of these reasons, we have $|\Delta x_1/\Delta x_2| = |v_1/v_2|$ or $|\Delta x_1/v_1| = |\Delta x_2/v_2|$, where Δx_1 and Δx_2 are the projections of their outermost sides (those on curve) on the *x*-axis and v_1 and v_2 are the projections of these two infinitely thin triangles on the *v*-axis. Physically, v_1 (v_2) is the velocity of *m* as it moves across the short distance Δx_1 (Δx_2), so the equality $|\Delta x_1/v_1| = |\Delta x_2/v_2|$ means the times spent in traversing these short distances are the same. The distance from A_1 to A_2 can be segmented into infinitely many small parts Δx_1s, to each of them there is a corresponding Δx_2 at the "opposite side" on the returning path from A_2 to A_3,

and the times spent in traversing each pair of these short distances are the same. Summing up over the segments, we find the time for *m* travelling from A_1 to A_2 is the same as it returns from A_2 to A_3 and, by similar arguments, the same from A_3 to A_4 and so on. This is to suggest that the semi-period (or the period) is a constant although the amplitude diminishes with time.

Unlike the semi-periods, the quarter periods are not all equal. When time is counted from an extremity, only the first quarter period is equal to the third and the second is equal to the fourth. In figure 6, the arcs of quarter cycles are shown in either solid red or dashed blue lines. Only those shown with the same type of lines are geometrically similar to each other, so the quarter periods they correspond to are equal. Mass *m* travels the same distance from *Q* to an extremity as it returns from that extremity to *Q*, but the average speed on the returning path must be comparatively slower owing to the energy loss caused by the damping. This leads to a slightly longer third (first) quarter period than the second (fourth). In spite of this, all the quarter periods are amplitude-independent.

Undamped SHM exhibits isochronism (period is independent of amplitude), but we cannot say therefore the damped does too. They are governed by two different equations of motion, and the former is only a particular case of the latter. Nonetheless, the damped SHM is proved to be isochronous, but from first principles.

(2) In figure 6, the solid red arcs of the spiral curve and the two axes form three similar figures, they are $OA_1C_1 \sim OA_2C_2 \sim OA_3C_3$. Their corresponding sides are in the same ratio, hence

$$OC_1 : OC_2 : OC_3 = OA_1 : OA_2 : OA_3. \quad (4)$$

As well, the dashed blue arcs of the spiral curve and the two axes form another three similar figures, they are $OA_2C_1 \sim OA_3C_2 \sim OA_4C_3$. Therefore,

$$OC_1 : OC_2 : OC_3 = OA_2 : OA_3 : OA_4. \quad (5)$$

Putting equations (4) and (5) together results in $OA_1 : OA_2 : OA_3 = OA_2 : OA_3 : OA_4$, or

$$OA_1/OA_2 = OA_2/OA_3 = OA_3/OA_4. \quad (6)$$

With more oscillations involved, equation (6) is extended to

$$OA_1/OA_2 = OA_2/OA_3 = \ldots = OA_n/OA_{n+1} = \ldots, \qquad (7)$$

where n is a positive integer.

A larger damping constant b brings about a larger value of this ratio, which is equal to unity when $b = 0$. The ratio of the amplitude at one side to the next at the other side is unchanged throughout the process of dying out, suggesting the amplitude decays by a fixed percentage over equal time intervals. This evidences it is an *exponential decay* [5].

6. The *x*-*t* graphs

The *x*-*t* graphs are more familiar to students. Through analysing the *v*-*x* graphs and the qualitative discussions in the previous sections, we have gathered some pieces of valuable information to sketch or examine the *x*-*t* graphs.

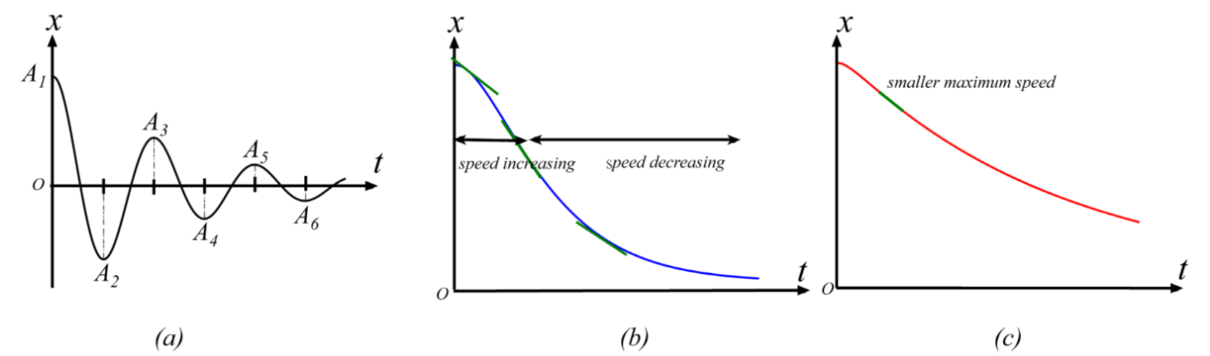

Figure 7. Typical *x*-*t* graphs of a *(a)* lightly damped, *(b)* critically damped and *(c)* heavily damped oscillator. The classification of these three cases and the main features shown in the graphs are the results obtained from some qualitative arguments and non-ODE mathematics.

Figure 7 illustrates the typical *x*-*t* graphs of a *(a)* lightly damped, *(b)* critically damped and *(c)* heavily damped oscillator. A lightly damped oscillator ($b < \sqrt{4mk}$) oscillates with a decaying amplitude under a rule of "constant ratio" (in this example, $OA_1/OA_2 = OA_2/OA_3 = \ldots = 1.5$) and a constant semi-period. Although almost indiscernible, in theory the shape of each "crest" and "trough" of the decaying sinusoidal curve is not exactly left-right symmetrical (left part is steeper), corresponding to the unequal quarter periods.

Figures 7*(b)* and *(c)* show a non-oscillating critically damped ($b = \sqrt{4mk}$) and heavily damped ($b > \sqrt{4mk}$) oscillator, respectively. After release, they return to the equilibrium position Q with a zero speed. As the slope of an *x*-*t* graph is velocity, the *x*-*t* graphs of these two types should display explicitly the two stages of at first the magnitude of slope increasing from zero to a maximum over a short period of time and then decreasing gradually to zero. The accelerating period for a heavily damped oscillator is shorter, thus the maximum speed attained is smaller. An overall slower speed means a longer time to return.

7. Conclusion

Our non-ODE approach yields a few of the results of the damped SHM, but they are essential and quite sufficient for one to acquire a non-superficial understanding of the subject. Our approach could not guide us to derive all the analytical expressions, but it could lead us to think the problem more from a physics point of view. Teachers could select the most advantageous parts for their students, the qualitative outline of the damped motion in terms of forces in section 3 and perhaps the three cases in section 4 are adequate most of the time, while section 5 may inspire the students more.

The *v*-*x* curve of a lightly damped SHM is an example of the famous *logarithmic spirals*, which often occur in nature [6]. In some advanced topics, e.g., chaos physics, nonlinear dynamics, “phase space portraits” is an indispensible tool [7, 8]. In the author’s opinion, giving students a chance to glimpse into and experience a preliminary use of this tool is worthwhile in pre-university physics education.

Lastly, we believe this nonstandard approach of introducing the damped SHM is far from complete, refining the details and enriching the contents are necessary in future work.

ORCID iDs

Chiu-King Ng https://orcid.org/0000-0002-1290-1039

References

[1] Marion J B and Thornton S T 1995 *Classical Dynamics of Particles and Systems* 4th edn (Fort Worth: Harcourt Brace College Publisher) pp 116-7

[2] IOP Teaching Advanced Physics Episode 306: Damped simple harmonic motion http://tap.iop.org/vibration/shm/306/page_46606.html

[3] IBO 2007 Physics Guide http://www.ncbis.net/Editor/EditorImages/Physics.pdf

[4] Wikipedia 2015 Isocline https://en.wikipedia.org/wiki/Isocline

[5] Nuffield Foundation 2015 Some useful equations for half-lives http://www.practicalphysics.org/go/Guidance_87.html

[6] Wikipedia 2015 Logarithmic spiral https://en.wikipedia.org/wiki/Logarithmic_spiral

[7] Baker G L and Gollub J P 1996 *Chaotic Dynamics: An Introduction* 2nd edn (New York: Cambridge University Press) pp 7-14

[8] Strogatz S H 2001 *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering* 1st edn (Boulder: Westview Press) pp 145-50