A Physical Model for Intuiting Linear Regression

Phillip Cervantes, Patricia Purdue, and Jeff Steele

Physics Department, Colorado College, Colorado Springs, CO 80903

(Dated: July 10, 2019)

Abstract

The application of linear regression to data analysis is traditionally a tedious calculation delegated to a computer. This article presents a physical model to facilitate an intuitive way of thinking about the meaning of a fit, particularly the slope. In this simple model, data points are envisioned as beads applying torques on a massless rod. As an illustration of how this model can help colleagues directly address special questions about fits, the model is applied to a situation in which repeated measurements of the dependent variable (y-values) are made at each of the independent variable values (x-values), and vice versa. The question posed was, "Does the slope depend on whether a linear fit is done with *all* of the data points or on data where the repeated measurements have been averaged first?" The answer can be calculated mathematically or by a computer; however, this paper provides a method for intuiting an answer using introductory-level physics. In addition, the model arrives at the coefficient of correlation formula in a very straightforward way, and as a final perk, it makes the recollection of the least-squared slope formula and coefficient of correlation much easier.

I. INTRODUCTION

When teaching data analysis to students or discussing the subject with colleagues, it is common to encounter questions about hypothetical data collections and how different fit procedures would impact the fit of a line. To approach such questions mathematically is often difficult and time consuming, often starting with the lumbering motion of looking up the formula for a least-squares fit. This paper presents a physical way of thinking about a linear least-squares data fit and includes an example of how the model can be useful.

In particular, two methods of fitting a line to "clustered" data will be compared. To illustrate the question, consider an example in which several measurements of a wire's resistance are made at a particular temperature, and this procedure is replicated for several different temperatures. A typical way for a scientist to handle such data is to average the points in each cluster (the resistance measurements taken at each temperature, in this case) and then make a weighted fit of the averaged data values; call this the "standard" approach. But how would the slope calculation change if a scientist elects to simply fit a line to all of the data points without averaging the clusters first? Will the slope of this "full data fit" agree with the standard approach? The physical model presented in this paper can shed some light on this question, as will be shown. For simplicity, this analysis will be restricted to data with error along the y-axis only or along the x-axis only, but not both.

The physical model is presented in Sec. II, including a derivation of the standard least-squares-fit slope from the model. Section III will explore the clustered data example mentioned above. Additionally, Sec. III C gives a straightforward way of thinking about and remembering the coefficient of correlation, R^2. Finally, Sec. IV shows that the model can be applied to weighted fits as well.

II. A PHYSICAL MODEL FOR LINE FITTING

First, the physical model will be described, and then it will be used to derive the well-known least-squares-fit formula[1] for the slope B of a line,

$$B = \frac{N \sum_i x_i y_i - \sum_i x_i \sum_i y_i}{N \sum_i x_i^2 - \left(\sum_i x_i\right)^2} \; , \tag{1}$$

for a set of N data points of the form (x_i, y_i). Note that each of these sums is from $i = 1$ to N. This abbreviated notation will be used throughout for simplicity, except in cases where the full notation is needed for clarity.

In this physical model, each of these data points is treated as if it is a bead of mass m on a horizontal, massless rod resting on a fulcrum. Each point's x-value gives the bead's horizontal location on the rod. A point's mass times its y-value represents a vertical force the bead exerts on the rod. For an unweighted fit, the masses of all the beads are equal and can be assigned a value of 1, which means that, in effect, y-value alone can be treated as the force. This simplification is used throughout the rest of this section and the next. Weighted fits will be addressed in Sec. IV.

The fulcrum of the rod is located at the x-coordinate of the center of mass of the beads, which is equal to the average of the x-values in an unweighted fit. Positive y-values are chosen to be forces acting vertically upward on the rod, and negative y-values to be forces acting vertically downward. An example of assigning data values to the model is shown in Fig. 1, where two data points of a fictitious data set are displayed. It will be shown that the slope of the best fit line for this data is given by the angular acceleration α of the rod due to the beads.

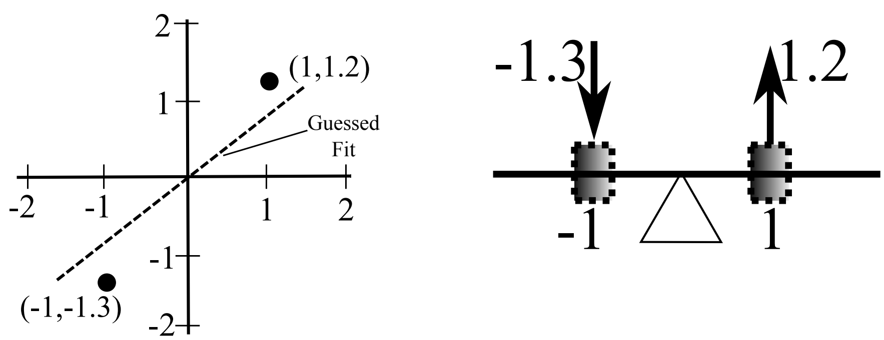


FIG. 1: A physical model for line fitting, in which data points are thought of as beads of mass m applying forces to a massless rod. The fulcrum of the rod is placed at the x-coordinate of the "center of mass" of the data points. The x-coordinates of the data points indicate the position of the beads, and the y-coordinates represent forces.

The first step to proving this is to calculate the net torque that the beads exert on the rod. As stated above, the fulcrum of the rod is located at the x-coordinate of the center of mass, denoted x_{cm}. In the case of unweighted data (equally weighted beads of mass $m = 1$), this is also the average of the x-values of the data, $x_{\text{cm}} = \bar{x} = \frac{1}{N}\sum_i x_i$. Recall the usual torque formula $\tau = rF_{\perp}$, where $F_{\perp}$ is the force perpendicular to the rod applied at a distance

r from the fulcrum. In this model, $F_\perp$ corresponds to a given data point's y-value, and the distance r becomes $(x_i - \bar{x})$. As a result, the net torque exerted on the rod by the beads is

$$\tau = \sum_i (x_i - \bar{x}) y_i \,. \tag{2}$$

Expanding this expression, substituting in the expression for $\bar{x}$, and simplifying gives

$$\tau = \sum_i x_i y_i - \frac{1}{N} \sum_i x_i \sum_i y_i \,. \tag{3}$$

Next, consider the moment of inertia of the beads on the massless rod. The moment of inertia for a point mass m is $I = mr^2$, where r is the distance from the rotation axis. With equally weighted beads of mass 1, the rotational inertia for this rod-and-bead system is

$$I = \sum_i (x_i - \bar{x})^2 \,. \tag{4}$$

Again, expanding and substituting in the expression for $\bar{x}$ gives

$$I = \sum_i x_i^2 - \frac{1}{N} \left(\sum_i x_i \right)^2 \,. \tag{5}$$

The angular acceleration can then be expressed as

$$\alpha = \frac{\tau}{I} = \frac{\sum_i x_i y_i - \frac{1}{N} \sum_i x_i \sum_i y_i}{\sum_i x_i^2 - \frac{1}{N} \left(\sum_i x_i \right)^2} \,. \tag{6}$$

Multiplying both the numerator and denominator by N reproduces Eq. (1), where the slope B is equal to the angular acceleration α of the rod due to the beads. Thus, the least-squares-fit slope formula can be derived from simple physics equations applied to this rod-and-bead model.

Equation (1) can be further simplified by shifting to a reference frame in which the vertical axis passes though the center of mass so that $x_{\text{cm}} = \bar{x} = 0$. (The center-of-mass frame is an obvious such choice, but the condition $y_{\text{cm}} = 0$ is not necessary to simplify the slope equation.) In this new reference frame, the slope equation reduces to

$$B = \frac{\tau}{I} = \frac{\sum_i x_i y_i}{\sum_i x_i^2} \,. \qquad [x_{\text{cm}} = \bar{x} = 0] \tag{7}$$

Equation (7) will be used throughout the paper to show the utility of this model.

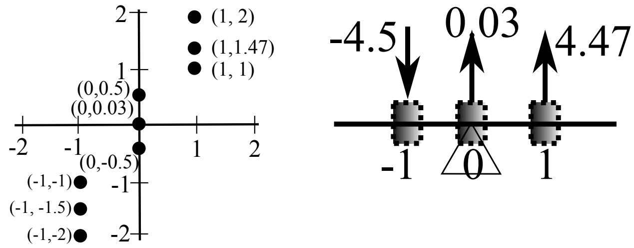
Note that one only needs to remember this rod-and-bead model to figure out the formula for the slope of a line. Determining the angular acceleration of a system is a skill physicists usually have readily at their disposal and can use intuitively. Up to now, only unweighted fits have been discussed, but Sec. IV will show that the rod-and-bead model still applies to weighted fits, with the mass of each bead corresponding to the the assigned weight of the associated data point.

As a first example of how this physical model can elucidate an aspect of line fitting, note that data with x-values near the center of mass do not significantly affect the slope. This rod-and-bead model makes this obvious since these points would correspond to masses resting near the fulcrum, which of course would not provide much torque or rotational inertia. This can also be quantitatively confirmed with Eq. (7) by noticing neither the numerator or denominator are significantly affected by data with x-values near zero. Instructors commonly advise students revisiting an experiment to try to gather extra data at the extremes rather than at the center of their graphs. The logic behind this instruction is easily explained by this physical model.

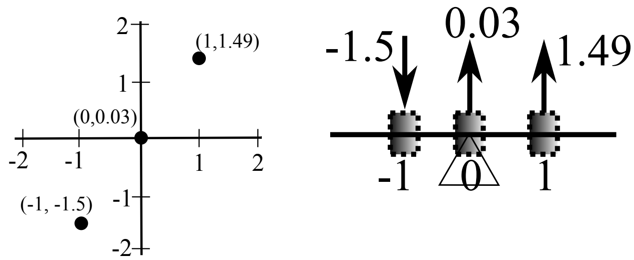
III. PHYSICAL MODEL APPLIED TO CLUSTERED DATA

A. Data with Vertical Clusters

Now the model can be applied to the question posed in the introduction: How does averaging or not averaging data clusters affect the fit of the line? Consider data where there are multiple y-values for each x-value, such as that shown on the graph on the left side of Fig. 2a. The data is conveniently plotted in a reference frame with $x_{\mathrm{cm}} = 0$. In a full data fit approach, let each x-value be represented by one bead. If there is a cluster of n data points at a given x-value, the corresponding bead on the rod will have a mass of n units. In the example of Fig. 2a, the beads each have three unit masses. The force each bead exerts on the massless rod is equal to the sum of the y-values of the points in the associated cluster. The slope of this line would be given by Eq. (7): $B = \frac{\tau}{I} = \frac{(-1)(-4.5)+(0)(.03)+(1)(4.47)}{3(-1)^2+3(0)^2+3(1)^2} = \frac{8.97}{6} = 1.495$.



(a)



(b)

FIG. 2: (a) Sample data with vertical clusters of data, graphed (left) and modeled (right). Note that, in the model, the y-value "forces" are summed for points with the same x-value. (b) Same data with clusters averaged, graphed and modeled.

Alternatively, the standard approach calls for averaging the y-values of each data cluster and fitting a line to those average values. Figure 2b shows the averaged data values and the corresponding physical model. Notice that the beads now have unit mass. Both the torque and the rotational inertia decrease proportionally so the result is exactly the same, $B = \frac{\tau}{I} = \frac{(-1)(-1.5)+(0)(.03)+(1)(1.49)}{(-1)^2+(0)^2+(1)^2} = \frac{2.99}{2} = 1.495$. This calculation was done explicitly to demonstrate an application of the model.

As to the question of whether it matters whether a full data fit or a standard fit is used, it does not for this case. However, that conclusion would be different if the number of data points in each cluster were not the same. The slope for the full data fit will be more influenced by the larger clusters, and agreement between the standard and full data fit is not guaranteed.

This can also be shown mathematically, without the usage of the model, by directly manipulating Eq. (7), which is implicitly a full data fit when applied directly. Suppose there are N clusters, which means that there are N values for the independent variable x, labeled $\{x_1, x_2, \ldots, x_N\}$. For each value of x, there are n measurements of the y value,

$\{y_{11}, y_{12}, \ldots, y_{1n}, y_{21}, \ldots, y_{2n}, \ldots, y_{Nn}\}$. Using this notation, Eq. (7) becomes

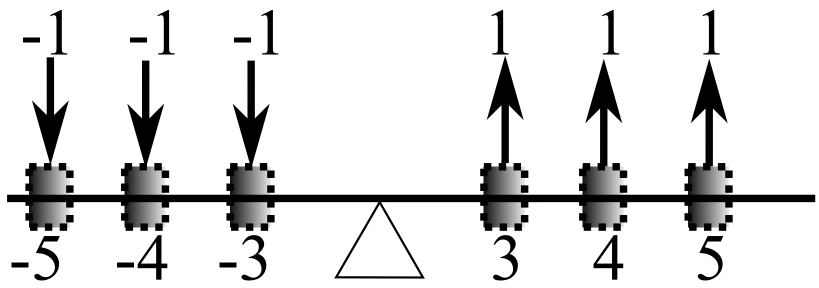
$$B = \frac{\sum_{i=1}^{N}\sum_{j=1}^{n} x_i y_{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{n} x_i^2} = \frac{\sum_{i=1}^{N} x_i \sum_{j=1}^{n} y_{ij}}{\sum_{i=1}^{N} n x_i^2} = \frac{\sum_{i=1}^{N} x_i n \bar{y}_i}{\sum_{i=1}^{N} n x_i^2} = \frac{\sum_{i=1}^{N} x_i \bar{y}_i}{\sum_{i=1}^{N} x_i^2} \ , \tag{8}$$

where $\bar{y}_i$ is the average of the y values in the i-th group. Notice that the original expression for the slope [Eq. (7)] is recovered except with the average y-values instead of actual y-values, which is how the standard-fit slope is calculated. So as long as each cluster of points has the same number of members, it doesn't matter whether a standard or full fit is done; the result is the same. Otherwise, the results of the standard and full data fits will not match exactly.

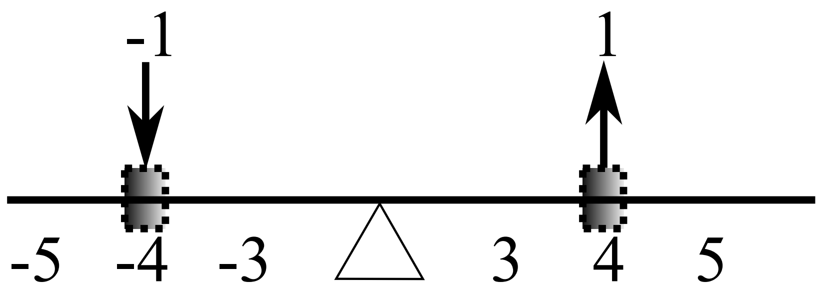
B. Data with Horizontal Clusters

1. Discussion

For data with horizontal clusters, in contrast to data with vertical clusters, full data and standard fits generally do not agree, even when the number of data points in each cluster is the same. In fact, for data with horizontal clusters, the standard fit will yield a steeper slope than a full data fit. This is easy to see using this model, but first, consider a simple example. Imagine data points $(-5, -1), (-4, -1), (-3, -1), (3, 1), (4, 1), (5, 1)$. Figure 3 dispenses with the graph of the hypothetical data and goes straight to the model to compute the respective slopes.



(a)



(b)

FIG. 3: (a) Modeled version of sample data with horizontal data clusters. (b) Same data with clusters averaged before modeling.

In Fig. 3a, the data points are represented as six beads with force (y-values) of magnitude 1 on each bead; the forces on the left are negative while those on the right are positive. A full data fit gives $B = \frac{(-5)(-1)+(-4)(-1)+(-3)(-1)+(3)(1)+(4)(1)+(5)(1)}{(-5)^2+(-4)^2+(-3)^2+(3)^2+(4)^2+(5)^2} = \frac{24}{100}$. In the standard approach, the fit uses the average of the data clusters, as shown in Fig. 3b. In this case, the average x-position on the right cluster will be $x = 4$ and the one on the left will be $x = -4$. The average force (y-value) is -1 and $+1$, on the left and right bead, respectively. For this fit, the slope is $B = \frac{(-1)(-4)+(1)(4)}{(-4)^2+(4)^2} = \frac{8}{32}$, which is slightly steeper than the full-data-fit slope.

This difference in slope can be explained with the physical model, where one can see that both the torque (numerator) and rotational inertia (denominator) for the full data fit are larger than that of the standard fit. In the example, the torque for the full data fit is larger by a multiplicative factor of three due to the contributions of the three values from each cluster. The moment of inertia of the full data fit, however, is larger by slightly more than a factor of three. To understand this, consider that, in the case of the full data fit, the denominator of the slope is the moment of inertia of all the data points (I_{Full}). When the data points in each cluster are averaged for the standard fit, some information is lost. In this case, the lost information contains the contribution due to the internal moment of inertia of each cluster ($I_{\text{int},i}$) about its center of mass. Each cluster in the standard fit contributes only one term for the moment of inertia, namely the cluster's center of mass about the fulcrum ($I_{\text{cm},i}$). In contrast, the full data fit includes two terms ($I_{\text{cm},i}+I_{\text{int},i}$) from each cluster. Thus, the standard fit's effective moment of inertia (I_{eff}) is not proportional to the full moment of inertia because of the excluded internal terms. As a result, the numerator and denominator of the standard fit are not changed by the same multiplicative factor, and the result is a steeper slope in the standard-fit case. This will be shown analytically below.

Because the internal moments of inertia of the clusters are the source of the discrepancy between the two methods of calculating the slope, the two methods should approximately agree when those internal terms are small. This will occur when the clusters are sufficiently compact, which makes sense intuitively. More detail on these findings appears in the following subsubsection to help substantiate this simple model, but hopefully, it will not distract the reader away from this intentionally intuitive physical model for thinking about line fits.

2. *Analytic Comparison*

To see the slope discrepancy mathematically, again let N be the number of clusters. In the case of horizontal clusters, there will be N values of y, denoted $\{y_1, y_2, \ldots, y_N\}$. For each value of y, let there be n measurements of the x value, giving $\{x_{11}, x_{12}, \ldots, x_{1n}, x_{21}, \ldots, x_{2n}, \ldots, x_{Nn}\}$. Using this notation, Eq. (7) for the full data fit becomes

$$B_{\text{Full}} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{n} x_{ij}y_i}{\sum_{i=1}^{N}\sum_{j=1}^{n} x_{ij}^2} = \frac{\tau_{\text{Full}}}{I_{\text{Full}}} \,. \qquad (9)$$

Note that the labeling of the numerator and denominator as τ_{Full} and I_{Full}, respectively, follows directly from Eqs. (2) and (4) being summed over all the data points (with mass 1) in a reference frame with $x_{\text{cm}} = \bar{x} = 0$.

The full-fit moment of inertia I_{Full} can be calculated via a different but more complicated method than using Eq. (4). Doing this will illustrate the different contributions to the moment of inertia and how that relates to the differences between the standard and full fit techniques. This alternate method consists of first finding the moment of inertia of each cluster about its own center of mass, then using the parallel-axis theorem to find the corresponding moment of inertia about the model's fulcrum (at $x_{\text{cm}} = \bar{x} = 0$), and finally summing over all the clusters, as follows.

First, the internal moment of inertia of the i-th cluster about its center of mass is

$$I_{\mathrm{int},i} = \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 \,, \tag{10}$$

where $\bar{x}_i$ is the average of the x-values in the i-th cluster and all points are assumed to have a mass of 1, similar to Eq. (4). This expression can be expanded and simplified to give

$$I_{\mathrm{int},i} = \sum_{j=1}^{n} x_{ij}^2 - \frac{1}{n}\left(\sum_{j=1}^{n} x_{ij}\right)^2 = \frac{1}{2n}\sum_{a=1}^{n}\sum_{b=1}^{n}(x_{ia} - x_{ib})^2 \,, \tag{11}$$

where the final equality is given by the useful (and general) relationship

$$\frac{1}{2}\sum_{a=1}^{n}\sum_{b=1}^{n}(x_a - x_b)^2 = n\sum_{j=1}^{n} x_j^2 - \left(\sum_{j=1}^{n} x_j\right)^2 \,. \tag{12}$$

Note that the $2n$ in the denominator of Eq. (11) is a little deceptive since it seems to indicate that $I_{\mathrm{int},i}$ gets smaller as n increases. However, that is not the case. The factor of 2 corrects for double counting, while the factor of n compensates for the fact the rotational inertia comes from the x-value distances from their cluster averages, rather than their distances from each other, which the argument of the sum contains.

The next step is to find this cluster's moment of inertia about the fulcrum. The parallel-axis theorem indicates that a term of the form md^2, where m is the mass of the cluster and d is the distance ($\bar{x}_i$) from the cluster's center of mass to the fulcrum, must be added to $I_{\mathrm{int},i}$. The mass of the cluster is n because each point has mass 1 and the cluster has n members. So this added term is $I_{\mathrm{cm},i} = n\bar{x}_i^2$.

Summing over all the clusters gives the full moment of inertia,

$$I_{\mathrm{Full}} = \sum_{i=1}^{N}\left(I_{\mathrm{cm},i} + I_{\mathrm{int},i}\right) = \sum_{i=1}^{N}\left(n\bar{x}_i^2 + \sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2\right) \,. \tag{13}$$

Expanding and simplifying this expression will reproduce the much simpler form $I_{\mathrm{Full}} = \sum_{i=1}^{N}\sum_{j=1}^{n} x_{ij}^2$ that appears in Eq. (9).

To see how this relates to the slope difference between the two line-fitting methods, consider the standard-fit slope, found by plugging the average x-values into Eq. (7) to get

$$B_{\mathrm{Std}} = \frac{\sum_{i=1}^{N}\bar{x}_i y_i}{\sum_{i=1}^{N}\bar{x}_i^2} \,. \tag{14}$$

It is straightforward to show that the numerator, or $\tau_{\rm Std}$, is proportional to $\tau_{\rm Full}$,

$$\tau_{\rm Std} = \sum_{i=1}^{N} \bar{x}_i y_i = \sum_{i=1}^{N} \left(\frac{1}{n} \sum_{j=1}^{n} x_{ij} \right) y_i = \frac{\tau_{\rm Full}}{n} . \tag{15}$$

However, the denominator of $B_{\rm Std}$, denoted $I_{\rm eff}$, is *not* simply $I_{\rm Full}/n$ because the contribution of the internal moments of inertia of the clusters is no longer included,

$$I_{\rm eff} = \sum_{i=1}^{N} \bar{x}_i^2 = \sum_{i=1}^{N} \frac{1}{n} I_{{\rm cm},i} = \frac{1}{n} \left(I_{\rm Full} - \sum_{i=1}^{N} I_{{\rm int},i} \right) . \tag{16}$$

Thus,

$$B_{\rm Std} = \frac{\tau_{\rm Std}}{I_{\rm eff}} = \frac{\tau_{\rm Full}}{I_{\rm Full} - \sum_i I_{{\rm int},i}} . \tag{17}$$

For completeness, the standard-fit slope can be expressed in terms of the original data points only,

$$B_{\rm Std} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{n} x_{ij} y_i}{\sum_{i=1}^{N} \left(\sum_{j=1}^{n} x_{ij}^2 - \frac{1}{2n} \sum_{a=1}^{n} \sum_{b=1}^{n} (x_{ia} - x_{ib})^2 \right)} , \tag{18}$$

by using the definitions of the various terms plus Eq. (12). By comparing Eq. (9) for $B_{\rm Full}$ and Eq. (17) or (18) for $B_{\rm Std}$, it is clear that the slope's denominator in the standard fit is always smaller than that for the full data fit (in the case of horizontal clusters of equal size) because of an extra subtractive term, and hence the standard fit will always produce a steeper slope than the full fit in these circumstances.

As mentioned at the end of Sec. III B 1, the standard and full fits will agree reasonably well when the clusters are sufficiently compact, meaning that the typical width of the clusters is significantly less than the width of the entire data set. In that case, the differences $|x_{ij} - \bar{x}_i|$ are generally small compared to $|x_{ij}|$ for most clusters. Some terms near the fulcrum may have $|x_{ij} - \bar{x}_i| \gtrsim |x_{ij}|$, but the magnitude of those terms will be quite small (because of their proximity to the fulcrum) compared to the typical size of the terms from the other clusters. So even if $|x_{ij} - \bar{x}_i| \ll |x_{ij}|$ is not true for all terms, the relationship $\sum_i I_{{\rm int},i} \ll \sum_i I_{{\rm cm},i}$ will hold as long as it's true for most. As a result, the $\sum_i I_{{\rm int},i}$ contribution should be negligible, giving $I_{\rm eff} \approx I_{\rm Full}$ and $B_{\rm Std} \approx B_{\rm Full}$.

C. R^2 Interpretation

A natural question to ask is, "What happens when the two axes are reversed? Is the new slope the inverse of the old one? That is, does $B_{xy} \times B_{yx} = 1$?" The answer is usually "no."

Applying Eq. (7) as an expression for B_{xy} and for B_{yx} gives

$$B_{xy} \times B_{yx} = \frac{\sum_i x_i y_i}{\sum_i x_i^2} \times \frac{\sum_i y_i x_i}{\sum_i y_i^2} = \frac{(\sum_i x_i y_i)^2}{\sum_i x_i^2 \sum_i y_i^2} \; . \tag{19}$$

This is actually the coefficient of correlation, R^2, in the center-of-mass frame. (Note: In this case, the center-of-mass frame *is* necessary because the x- and y-axes are being swapped in the slope formula. Therefore, a reference frame in which both $x_{\rm cm} = 0$ and $y_{\rm cm} = 0$ is needed.) For comparison, the standard expression for the coefficient of correlation[1] is

$$R^2 = \frac{\left[\sum_i (x_i - \bar{x})(y_i - \bar{y})\right]^2}{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2} \; , \tag{20}$$

where $\bar{x}$ and $\bar{y}$ represent the average values of $\{x_i\}$ and $\{y_i\}$, respectively. In the center-of-mass frame, $\bar{x}$ and $\bar{y}$ are both zero ($\bar{x} = x_{\rm cm} = 0$ and $\bar{y} = y_{\rm cm} = 0$ for unweighted data), reducing this expression to the final term in Eq. (19). Thus, the coefficient of correlation naturally falls out from this model. If the data are perfectly linearly correlated, $R^2 = 1$, and consequently $B_{xy} \times B_{yx} = 1$ in that special case.

IV. PHYSICAL MODEL APPLIED TO WEIGHTED FITS

The slope of a weighted fit[1] is given by

$$B = \frac{\sum_i w_i \sum_i w_i x_i y_i - \sum_i w_i x_i \sum_i w_i y_i}{\sum_i w_i \sum_i w_i x_i^2 - (\sum_i w_i x_i)^2} \; , \tag{21}$$

where w_i are the weights assigned to the data points. This expression can be derived from the rod-and-bead model using the same method as in Sec. II but with a few modifications. Before, it was assumed that all data points had the same weight and were assigned unit mass. Now, the weight w_i of each point becomes the mass of the corresponding bead in the model, and the force the bead exerts on the rod is now $w_i y_i$ instead of just y_i. Also, the average of the x-values is no longer equal to the center of mass, so the $\bar{x}$ terms appearing in the equations of Sec. II revert to $x_{\rm cm}$.

In this case, the torque exerted on the rod by the beads is

$$\tau = \sum_i (x_i - x_{\rm cm})(w_i y_i) = \sum_i w_i x_i y_i - \frac{1}{M} \sum_i w_i x_i \sum_i w_i y_i \; , \tag{22}$$

where $M = \sum_i w_i$ is the total mass of the beads and $x_{\rm cm} = \frac{1}{M} \sum_i w_i x_i$. The moment of inertia is similarly

$$I = \sum_i w_i (x_i - x_{\rm cm})^2 = \sum_i w_i x_i^2 - \frac{1}{M} \left(\sum_i w_i x_i \right)^2 \; . \tag{23}$$

Using these two expressions to calculate the angular acceleration $\alpha = \tau/I$ and multiplying both numerator and denominator by M reproduces Eq. (21). As in the unweighted case, the data can be moved to a frame with $x_{\text{cm}} = 0$ for simplicity, and the slope equation reduces to

$$B = \frac{\sum_i w_i x_i y_i}{\sum_i w_i x_i^2} \,. \qquad [x_{\text{cm}} = 0] \tag{24}$$

When weights do not vary too much, as in typical undergraduate experiments, the slope will not be too different from unweighted fits. In the scheme of this model, the torque and rotational inertia both increase or decrease almost proportionally and the angular acceleration (i.e., slope) isn't impacted much. In general, much of the previous discussion about unweighted fits should apply to weighted fits as well.

V. CONCLUSION

It has been shown that line fitting can be thought of in a physical framework to help facilitate discussion between scientists. This physical rod-and-bead model allowed questions regarding clustered data to be readily addressed. While this manuscript contained more math than preferred in light of the goal, it was included to convince the reader of the fidelity of the model and its application. Once comfortable with the implementation of the model, the reader will hopefully be able to intuit conclusions more efficiently than trying multiple fits on a computer or staring at the summations in Eqs. (1) and (21). The coefficient of correlation R^2 emerged naturally from the model, and as such, is a natural quantity to consider using. Finally, it was demonstrated that the model can also be applied to weighted data, where the beads in the model take on a more literal meaning of the term "weight."

---

[1] John R. Taylor, *An Introduction to Error Analysis*, 2nd edition (University Science Books, Sausalito, 1997).