

# Plotting and Fitting Power-Laws

Ryan Shijié Dù

February 6, 2023

## 1 Introduction

In applied math, data often follow power-laws of the form:

$$y(x) = Cx^\gamma. \tag{1}$$

Here  $\gamma$  is the power-law exponent. For example, in turbulence theory, conserved quantities often distribute over scales (wave-numbers  $k$ ) following a power-law. One result of this kind is the famed “Kolmogorov 5/3 law” which states that the kinetic energy spectrum  $E(k)$  of 3D hydrodynamics turbulence follows the law (Frisch 1995):

$$E(k) = Ck^{-5/3}. \tag{2}$$

Another example is in the field of numerical analysis. Numerical integration schemes commonly have errors asymptotically proportional to powers of discretization step size. To give two concrete cases: 1. composite trapezoidal rule for integration has error asymptotically following

$$e(h) = Ch^2 \tag{3}$$

and 2. explicit Runge–Kutta 4<sup>th</sup> order method (Rk4) for solving differential equations has

$$e(h) = Ch^4 \tag{4}$$

where  $h$  is (spatial or temporal) discretization step size (Süli and Mayers 2003). It suffices to say there are more examples than the ones listed, and power-laws are constantly being plotted and analyzed.

Power-laws are commonly visualized using a log-log plot, where the  $x$  and  $y$ -axis are in log scale and power-laws appear as straight lines. To obtain the power-law exponent, it is common to use linear regression on the log-log scale data. In §2 this note we will examine the application of linear regression on data distributed following power-laws and conclude that while it is appropriate in many cases, the choice of cost function deserves more case-specific consideration. §3 gives our suggestions for plotting data following power-laws. This necessitates a discussion of when it is appropriate to diagnose a power-law exponent from data.

## 2 Diagnosing power-law exponent from data

We will work with the general power-law

$$y(x) = Cx^\gamma. \tag{5}$$

We assume that the data  $z$  follows (5) with small deviations:

$$z(x) = Cx^\gamma + \epsilon. \tag{6}$$

Assumptions on the form of the error and how we measure it will dictate the best regression to use. Only in some cases is the common linear regression on the log-log scale appropriate.

### 2.1 Scaled $L^2$ norm

One interpretation of regression with least square error is that it is the orthogonal projection onto the selected basis functions. The projection minimizes the functional  $L^2$  norm:

$$\int |\log z - (\log C + \gamma \log x)|^2 d(\log x). \tag{7}$$

We focus on

$$\log z - (\log C + \gamma \log x) = \log z - \log y = \log z - \log(z - \epsilon). \tag{8}$$

We assume that the deviation  $\epsilon$  is small so that Taylor expansion is appropriate. We have

$$\log z - \log(z - \epsilon) \approx \frac{\epsilon}{z} \tag{9}$$

$$\Rightarrow \log z - (\log C + \gamma \log x) \approx \frac{z - Cx^\gamma}{z}. \tag{10}$$

We replace the integrand in (7)

$$\int \left| \frac{z - Cx^\gamma}{z} \right|^2 d(\log x). \tag{11}$$

A change of variable to  $dx$  gives

$$\int \left| \frac{z - Cx^\gamma}{z} \right|^2 \frac{1}{x} dx. \tag{12}$$

Without the  $1/x$  factor, we see that linear regression on the log-log scale is in fact minimizing the  $L^2$  error relative to the values of the data  $z$ . This normalization is desirable in some cases: because the data follow power-laws, their magnitude can change significantly over a large range of  $x$ . If the error scale as the magnitude of the data, it is appropriate to minimize the relative error. However, if we know *a priori* the error magnitude is uniform, the normalized error is not appropriate.

## 2.2 Density of data

The  $1/x$  factor comes from the change of variable in the integration. We will show that it directly relates to the density of the discrete data. We have focused on the integral functional norm, but we only have discrete data. One way to relate them is to use the discrete data to numerically approximate the integration. A factor will appear if the data are distributed unevenly. Take the example of turbulence spectra, we have data for each integer wavenumber  $x$ . This means in the log-log scale, the data appear at each  $\log x$  and have gaps of approximate size  $1/x$ . The simple sum error used in linear regression on the log-log scale

$$\sum_i |\log z_i - (\log C + \gamma \log x_i)|^2 \tag{13}$$

is a discrete approximation of the scaled integral

$$\int x |\log z - (\log C + \gamma \log x)|^2 d(\log x). \tag{14}$$

Intuitively, the  $x$  factor is due to the denser data for larger  $x$ . Apply the calculations in §2.1 we have that it is approximately equivalent to

$$\int \left| \frac{z - Cx^\gamma}{z} \right|^2 dx \tag{15}$$

the integrated relative error squared. Because the data are distributed evenly in  $x$ , the  $1/x$  factor disappears in integration with regards to  $x$ . (13) and (15) connect the commonly used linear regression on the log-log scale to a more interpretable relative  $L^2$  error on the unscaled data. For diagnosing the power law exponent of turbulence spectra, this interpretation is the most appropriate. This is the interpretation assumed in DÙ and Bühler 2023 when they diagnose power laws for turbulence spectra.

In other cases, the data is collected with an even distance in  $\log x$ . For example, to check the convergence of a numerical scheme, one usually uses results with discretization step  $h, h/2, h/4, \dots$ . In these situations, the simple sum (13) is an approximation of (12). The  $1/x$  appears because the data distribute unevenly in  $x$ .

## 2.3 Maximum likelihood interpretation

An alternative interpretation of inference using least square error is that it maximizes the log-likelihood function assuming the deviations are independent and Gaussian. Further assuming that the errors have uniform standard deviations in the log-log scale, we recover the error (13). To retain the Gaussian nature of the errors, we assume again that the error is small so that the linear Taylor expansion is a sufficient approximation. Then for  $x$  and  $z$ , this means the errors are independent Gaussians with standard deviations that scale as  $z$ . We have the cost function

$$\sum_i \left| \frac{z_i - Cx_i^\gamma}{z_i} \right|^2. \tag{16}$$

This is the underlying interpretation for the regression in Figure 3 of DÙ and Bühler 2023.

Alternatively, assuming that the error is uniform in the regular scale, maximum likelihood minimizes

$$\sum_i |z_i - Cx_i^\gamma|^2 \iff \sum_i |z_i (\log z_i - (\log C + \gamma \log x_i))|^2. \quad (17)$$

We see that the maximum likelihood interpretation recovers the result of the functional  $L^2$  norm interpretation except the scaling due to data density. Indeed, with the assumption of independence, the maximum likelihood interpretation takes no account of the distribution of data in  $x$ .

## 2.4 $\ell^1$ norm and KL divergence

We pay special attention to the right-hand-side equation in (17). Instead of the squared sum, we sum the absolute values ( $\ell^1$  norm). We have

$$\sum_i |z_i (\log z_i - \log y_i)| = \sum_i \left| z_i \log \left( \frac{z_i}{y_i} \right) \right| = D_{\text{KL}}[z||y]$$

where  $D_{\text{KL}}$  denotes the Kullback–Leibler (KL) divergence. Note the KL divergence is not a norm. The loss of symmetry between  $y$  and  $z$  comes from the choice of  $z$  as the center for the Taylor expansion. We could understand this surprising connection by noting that many probability distributions of interests have power-law tails (e.g.: Cauchy distribution). This opens the tools of applying the tools of probability and information theory to our problem.

## 3 Plotting data following power-laws

We have addressed how to diagnose the power-law exponent from data. However, it is not always appropriate to do the diagnostic in the first place. In many cases, we know the power-law exponent the data should follow. Then we should plot the data and the theoretical power-law on the same plot and compare them, instead of fitting a power-law to the data. We will showcase this point using some examples. They will also serve as examples of plotting data that follows power-laws.

### 3.1 Turbulence spectra

Data measuring turbulence spectra are commonly plotted on the log-log scale. We often know the power-law it should follow from theory. Figure 1 is an example of such a plot, and it does several things right. Since we know the power-law, we should not measure the exponent from data. Instead, this figure plots the power-law alongside the data and we can compare the data and theory easily. Additionally, this plot labels its axes correctly. It is more stylistic to label the axes using raw values, instead of labeling them using  $\log x$  and  $\log y$  and the correspondent log values. And it is *incorrect* to label the numbers using the unscaled numbers ( $10^1, 10^2, \dots$ ), and label the axes  $\log x$  and  $\log y$ . That would mean  $\log x = 10^1, 10^2, \dots$  and is absurd. In MATLAB, the numerical labels are automatically correct, and one only have to make sure that the text labels are not in log. Finally, since we have discrete data, it is better to plot the data as discrete data points and not connect them.

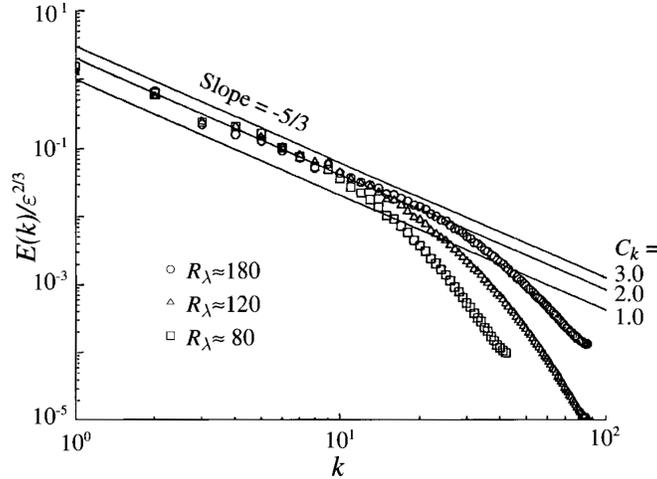


Fig. 5.9. Energy spectra in the space domain for three values of  $R_\lambda$  as indicated, obtained from a  $256^3$  computer simulation. Three different values for the Kolmogorov constant  $C_k$  are tried. Courtesy T. Sanada and K. Ishii.

Figure 1: Example plot from Frisch 1995, Fig. 5.9.

### 3.2 Numerical errors

The same principle applies to plotting numerical errors. Figure 2 shows an example of a clear plot. The axes are labeled correctly. Power-laws are not diagnosed, instead, the theoretical power-law is plotted alongside the data so that their agreements and differences are clear at one glance. The data are plotted using dots to signify their discreteness.

## 4 Concluding remarks

This note treats the topics of when and how to diagnose power-law exponents from data. We conclude that not all situations require the diagnostics of power-law exponents via regression. And in situations when their diagnostics are appropriate, there are subtleties in choosing the correct cost functions to minimize for regression. One must take care when plotting and analyzing data that follow power-laws.

## Acknowledgment

We were inspired to think more about fitting power-law while working with Oliver Bühler on Dù and Bühler 2023. We thank him for pointing out that relative error is sometimes the appropriate measure of error. We learned many lessons on plotting from Aleksandar Donev, some of them are included in this note. We thank him for teaching us to think carefully before we plot.

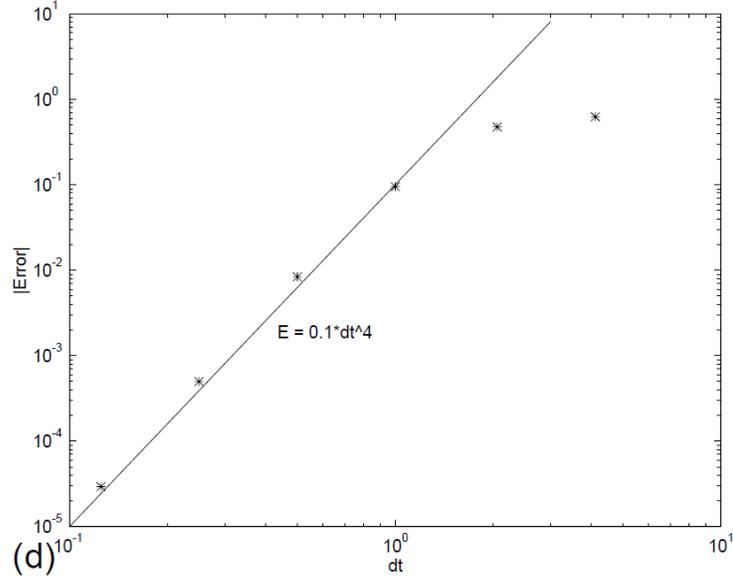


Figure 2: Example plot from Figure 1d of Milewski and Tabak 1999. Original label: Propagation of an oblique solitary wave, with parameters  $\epsilon = 0.1$  and  $k = \frac{1}{\sqrt{5}}(2, 1)$ , in a grid with periods  $L = 14\pi$  and  $H = 28\pi$  and  $256 \times 256$  points. (d) Log-log plot of the normalized error after one tour of the periodic box, as a function of  $\Delta t$ .

## References

- Dù, Ryan Shìjié, and Oliver Bühler. 2023. “The impact of frequency bandwidth on a one-dimensional model for dispersive wave turbulence.” *Submitted*.
- Frisch, Uriel. 1995. *Turbulence: The Legacy of A. N. Kolmogorov*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139170666>.
- Milewski, Paul A., and Esteban G. Tabak. 1999. “A PseudoSpectral Procedure for the Solution of Nonlinear Wave Equations with Examples from Free-Surface Flows.” *SIAM Journal on Scientific Computing* 21, no. 3 (January): 1102–1114. ISSN: 1064-8275. <https://doi.org/10.1137/s1064827597321532>.
- Süli, Endre, and David F. Mayers. 2003. *An Introduction to Numerical Analysis*. Cambridge: Cambridge University Press. ISBN: 978-0-521-00794-8. <https://doi.org/10.1017/CBO9780511801181>.