Decomposition Into Granules Speeds Up Data Processing Under Uncertainty

Andrzej Pownuk and Vladik Kreinovich
Computational Science Program
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
ampownuk@utep.edu, vladik@utep.edu

Abstract

In many real-life situations, uncertainty can be naturally described as a combination of several components, components which are described by probabilistic, fuzzy, interval, etc. granules. In such situations, to process this uncertainty, it is often beneficial to take this granularity into account by processing these granules separately and then combining the results.

In this paper, we show that granular computing can help even in situations when there is no such natural decomposition into granules: namely, we can often speed up processing of uncertainty if we first (artificially) decompose the original uncertainty into appropriate granules.

1 Need to Speed Up Data Processing Under Uncertainty: Formulation of the Problem

Need for data processing. One of the main reasons for data processing is that we are interested in a quantity y which is difficult (or even impossible) to measure or estimate directly. For example, y can be a future value of a quantity of interest.

To estimate this value y, we:

- find easier-to-measure and/or or easier-to-estimate quantities $x_1, \ldots, x_n$ which are related to y by a known dependence $y = f(x_1, \ldots, x_n)$,
- measure or estimate x_i's, and
- use the known relation $y = f(x_1, \ldots, x_n)$ to predict y.

Need to take uncertainty into account. Due to measurement uncertainty, the measurement results $\widetilde{x}_i$ are, in general, different from the actual values x_i

of the corresponding quantities. This is even more true to the results of expert estimates.

Therefore, the value $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ that we obtain by processing the measurement/estimation results is, in general, different from the desired value $y = f(x_1, \ldots, x_n)$. It is therefore important to estimate the resulting uncertainty $\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y$; see, e.g., [1].

Measurement or estimation errors are usually relatively small. Measurement and estimation errors are usually assumed to be relatively small, so that terms quadratic in measurement errors can be safely ignored [1].

If we expand the expression

$$\Delta y = f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(\widetilde{x}_1 - \Delta x_1, \ldots, \widetilde{x}_n - \Delta x_n)$$

in Taylor series and ignore terms which are quadratic (or higher order) in terms of Δx_i, then we get

$$\Delta y = \sum_{i=1}^{n} c_i \cdot \Delta x_i,$$

where $c_i \stackrel{\text{def}}{=} \dfrac{\partial f}{\partial x_i}$.

This simplified expression enables us to estimate the uncertainty in the result y of data processing based on the known information about the uncertainties Δx_i.

How to describe uncertainty. For measurements, we usually have a large number of situations when we performed the measurement with our measuring instrument and we also measured the same quantity with some more accurate measuring instrument – so that we have a good record of past values of measurement errors. For example, we may record the temperature outside by a reasonably cheap not-very-accurate thermometer, and we can find the measurement errors by comparing these measurement results with accurate measurements performed at a nearby meteorological station.

Based on such a record, we can estimate the probability of different values of the measurement error. Thus, it is reasonable to assume that for each i, we know the distribution of the measurement error $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$.

Measurement errors corresponding to different variables are usually independent.

In this paper, we consider an ideal case when we know:

- the exact dependence $y = f(x_1, \ldots, x_n)$,
- the probability distribution of each of the variables Δx_i, and
- the values c_i.

Thus, we know the probability distribution of each of the terms $t_i = c_i \cdot \Delta x_i$. So, we arrive at the following problem:

- we know the probability distributions of each of n independent random variables $t_1, \ldots, t_n$,
- we are interested in the probability distribution of their sum $t = \sum_{i=1}^{n} t_i$.

How this problem is solved now. The usual way to represent a probability distribution is by its probability density function (pdf) $\rho(x)$. The pdf of the sum $t = t_1 + t_2$ of two independent random variables with pdfs $\rho_1(t_1)$ and $\rho_2(t_2)$ is equal to

$$\rho(t) = \int \rho_1(t_1) \cdot \rho_2(t - t_1)\, dt_1.$$

A straightforward way of computing each value $\rho(t)$ is by replacing the integral with a sum. If we use N different points, then we need N computations to compute the sum corresponding to each of the N points, thus we need the total of N^2 computation steps; see, e.g., [2].

A faster computation can be done if we use characteristics functions $\chi_i(\omega) \stackrel{\text{def}}{=} E[\exp(\mathrm{i} \cdot \omega \cdot t_i)]$, where E denotes the expected value. Then, from $t = t_1 + t_2$, we conclude that

$$\exp(\mathrm{i} \cdot \omega \cdot t) = \exp(\mathrm{i} \cdot \omega \cdot t_1) \cdot \exp(\mathrm{i} \cdot \omega \cdot t_2)$$

and thus, since t_1 and t_2 are independent, that

$$E[\exp(\mathrm{i} \cdot \omega \cdot t)] = E[\exp(\mathrm{i} \cdot \omega \cdot t_1)] \cdot E[\exp(\mathrm{i} \cdot \omega \cdot t_2)],$$

i.e., $\chi(\omega) = \chi_1(\omega) \cdot \chi_2(\omega)$. Here:

- computing each characteristic function $\chi_i(\omega)$ by Fast Fourier Transform requires $O(N \cdot \ln(N))$ computational steps,
- computing point-by-point multiplication requires N steps, and
- the inverse Fourier Transform to reconstruct $\rho(t)$ from its characteristic function also takes $O(N \cdot \ln(N))$ steps.

So overall, we need $O(N \cdot \ln(N))$ steps, which is smaller than N^2.

Can we do it faster? For large N, the time N needed for point-wise multiplication is still huge, so it is reasonable to look for the ways to make it faster.

2 Analysis of the Problem and Our Idea

Processing can be faster if both distributions are normal. If both t_i are normally distributed, then we do not need to perform these computations: we know that the sum of two normal distributions with mean μ_i and variances V_i is also normal, with mean $\mu = \mu_1 + \mu_2$ and variance $V = V_1 + V_2$.

In this case, we need two computational steps instead of $O(N)$.

Other cases when we can speed up data processing. Same holds for any *infinitely divisible* distribution, with characteristic function

$$\chi(\omega) = \exp(\mathrm{i} \cdot \mu \cdot \omega - A \cdot |\omega|^\alpha).$$

For example:

- For $\alpha = 2$, we get normal distribution.
- For $\alpha = 1$, we get Cauchy distribution, with the probability density function

$$\rho(x) = \frac{1}{\pi \cdot \Delta} \cdot \frac{1}{1 + \dfrac{(x-\mu)^2}{\Delta^2}}$$

  for an appropriate $\Delta > 0$.

Indeed, in this case, once we know the distributions for t_1 and t_2, then, based on the corresponding characteristic functions $\chi_1(\omega) = \exp(\mathrm{i} \cdot \mu_1 \cdot \omega - A_1 \cdot |\omega|^\alpha)$ and $\chi_2(\omega) = \exp(\mathrm{i} \cdot \mu_2 \cdot \omega - A_2 \cdot |\omega|^\alpha)$, we can conclude that the characteristic function $\chi(\omega)$ for the sum $t_1 + t_2$ has the form

$$\chi(\omega) = \chi_1(\omega) \cdot \chi_2(\omega) = \exp(\mathrm{i} \cdot \mu \cdot \omega - A \cdot |\omega|^\alpha),$$

where $\mu = \mu_1 + \mu_2$ and $A = A_1 + A_2$. So, in this case too, we need two computational steps instead of $O(N)$:

- one step to add the means μ_i, and
- another step to add the values A_i.

Our idea. Our idea is to select several values $\alpha_1, \ldots, \alpha_k$ – e.g., $\alpha_1 = 1$ and $\alpha_2 = 2$ – and approximate each random variable t_i by a sum

$$t_{a,i} = t_{i1} + \ldots + t_{ij} + \ldots + t_{ik}$$

of infinitely divisible random variables t_{ij} corresponding to the selected values of α_j.

The characteristic function $\chi_{ij}(\omega)$ for each variable t_{ij} has the form

$$\chi_{ij}(\omega) = \exp(\mathrm{i} \cdot \mu_{ij} \cdot \omega - A_{ij} \cdot |\omega|^{\alpha_j}).$$

Thus, the characteristic function $\chi_{a,i}(\omega)$ of the sum $t_{a,i} = \sum\limits_{j=1}^{k} t_{ik}$ is equal to the product

$$\chi_{a,i}(\omega) = \prod_{j=1}^{k} \chi_{ij}(\omega) = \exp\left(\mathrm{i} \cdot \mu_i \cdot \omega - \sum_{j=1}^{k} A_{ij} \cdot |\omega|^{\alpha_j}\right),$$

where $\mu_i \stackrel{\text{def}}{=} \sum_{j=1}^{k} \mu_{ij}$.

From $\chi_1(\omega) \approx \chi_{a,1}(\omega)$ and $\chi_2(\omega) \approx \chi_{a,2}(\omega)$, we conclude that the characteristic function $\chi(\omega) = \chi_1(\omega) \cdot \chi_2(\omega)$ for the sum $t = t_1 + t_2$ is approximately equal to the product of the approximating characteristic functions:

$$\chi(\omega) \approx \chi_a(\omega) \stackrel{\text{def}}{=} \chi_{a,2}(\omega) \cdot \chi_{a,2}(\omega) = \exp\left(\mathrm{i} \cdot \mu \cdot \omega - \sum_{j=1}^{k} A_j \cdot |\omega|^{\alpha_j}\right),$$

where $\mu = \mu_1 + \mu_2$ and $A_j = A_{1j} + A_{2j}$.

In this case, to find the approximating distribution for the sum t, we need to perform $k+1$ arithmetic operations instead of N:

- one addition to compute μ and
- k additions to compute k values $A_1, \ldots, A_k$.

Comment. A similar idea can be applied to the case of fuzzy uncertainty; see [3] for details.

3 How to Approximate

Natural idea: use Least Squares. We want to approximate the actual distribution $\rho_i(t)$ for each of the variables t_i by an approximate approximate distribution $\rho_{a,i}(t)$. A reasonable idea is to use the Least Squares approximation, i.e., to find a distribution $\rho_{a,i}(t)$ for which the value $\int (\rho_i(t) - \rho_{a,i}(t))^2\, dt$ is the smallest possible.

Let us reformulate this idea in terms of the characteristic functions. The problem with the above idea is that while for $\alpha = 1$ and $\alpha = 2$, we have explicit expressions for the corresponding probability density function $\rho_{a,i}(t)$, we do not have such an expression for any other α. Instead, we have an explicit expression for the characteristic function $\chi(\omega)$. It is therefore desirable to reformulate the above idea in terms of characteristic functions.

We want to approximate the characteristic function $\chi_i(\omega)$ by an expression $\chi_{a,i}(\omega)$ of the type $\exp\left(-\sum_j c_j \cdot f_j(\omega)\right)$ for some fixed functions $f_j(\omega)$; in our case, $f_0(\omega) = -\mathrm{i} \cdot \omega$ and $f_j(\omega) = |\omega|^{\alpha_j}$ for $j \geq 1$.

This can be done, since, due to Parceval theorem, the least squares (L^2) difference $\int (\rho_i(t) - \rho_{a,i}(t))^2\, dt$ between the corresponding pdfs $\rho_i(t)$ and $\rho_{a,i}(t)$ is proportional to the least squares difference between the characteristic functions:

$$\int (\rho_i(t) - \rho_{a,i}(t))^2\, dt = \frac{1}{2\pi} \cdot \int (\chi_i(\omega) - \chi_{a,i}(\omega))^2\, d\omega.$$

So, minimizing the value $\int(\rho_i(t)-\rho_{a,i}(t))^2\,dt$ is equivalent to minimizing

$$I \stackrel{\text{def}}{=} \int(\chi_i(\omega)-\chi_{a,i}(\omega))^2\,d\omega.$$

How to approximate: computational challenge and its solution. The problem with the above formulation is that the Least Squares method is very efficient is we are looking for the coefficients of a linear dependence. However, in our case, the dependence of the expression $\chi_{a,i}(\omega)$ on the parameters μ_i and A_{ij} is non-linear, which makes computations complicated.

How can we simplify computations? We can borrow the idea from the case of normal distributions: in this case, we start with the maximum likelihood methods, in which we maximize the probability, and take negative logarithms of the pdfs – which results in the known Least Squares method [2]. In our more general case too, if we take the negative logarithm of the characteristic function, we get a linear function of the unknowns:

$$-\ln(\chi_{a,i}(\omega)) = -\mathrm{i}\cdot\mu_i\cdot\omega+\sum_{j=1}^{k}A_{ij}\cdot|\omega|^{\alpha_j}.$$

To use this idea, let us reformulate the objective function

$$\int(\chi_i(\omega)-\chi_{a,i}(\omega))^2\,d\omega$$

in terms of the difference between the negative logarithms. We are interested in situations in which the approximation is good, i.e., in which the difference $\varepsilon_i(\omega)\stackrel{\text{def}}{=}\chi_{a,i}(\omega)-\chi_i(\omega)$ is small. Then, $\chi_{a,i}(\omega)=\chi_i(\omega)+\varepsilon_i(\omega)$, hence

$$-\ln(\chi_{a,i}(\omega)) = -\ln(\chi_i(\omega)+\varepsilon_i(\omega)) = -\ln\left(\chi_i(\omega)\cdot\left(1+\frac{\varepsilon_i(\omega)}{\chi_i(\omega)}\right)\right) =$$

$$-\ln(\chi_i(\omega))-\ln\left(1+\frac{\varepsilon_i(\omega)}{\chi_i(\omega)}\right).$$

Since $\varepsilon_i(\omega)$ is small, we can ignore terms which are quadratic and higher order in $\varepsilon_i(\omega)$ and get

$$\ln\left(1+\frac{\varepsilon_i(\omega)}{\chi_i(\omega)}\right)\approx\frac{\varepsilon_i(\omega)}{\chi_i(\omega)}.$$

Thus, in this approximation,

$$(-\ln(\chi_i(\omega)))-(-\ln(\chi_{a,i}(\omega)))=\frac{\varepsilon_i(\omega)}{\chi_i(\omega)},$$

hence

$$\varepsilon_i(\omega)=\chi_{a,i}(\omega)-\chi_i(\omega)=\chi_i(\omega)\cdot((-\ln(\chi_{a,i}(\omega)))-(-\ln(\chi_i(\omega)))),$$

so the minimized integral takes the form

$$I = \int (\chi_i(\omega) - \chi_{a,i}(\omega))^2 \, d\omega = \int \chi_i^2(\omega) \cdot ((-\ln(\chi_i(\omega))) - (-\ln(\chi_{a,i}(\omega)))^2 \, d\omega,$$

or, equivalently, the form

$$I = \int (f_i(\omega) - f_{a,i}(\omega))^2 \, d\omega,$$

where we denoted

$$f_i(\omega) \stackrel{\text{def}}{=} -\chi_i(\omega) \cdot \ln(\chi_i(\omega))$$

and

$$f_{a,i}(\omega) \stackrel{\text{def}}{=} -\chi_i(\omega) \cdot \ln(\chi_{a,i}(\omega)).$$

In our case

$$f_{a,i}(\omega) = -\mathrm{i} \cdot \mu_i \cdot \omega \cdot \chi_i(\omega) + \sum_{j=1}^{k} A_{ij} \cdot \chi_i(\omega) \cdot |\omega|^{\alpha_j}.$$

In other words, we need to find the coefficients c_k by applying the Least Squares method to the approximate equality

$$-\ln(\chi_i(\omega)) \cdot \chi_i(\omega) \approx -\mathrm{i} \cdot \mu_i \cdot \omega \cdot \chi_i(\omega) + \sum_{j=1}^{k} A_{ij} \cdot \chi_i(\omega) \cdot |\omega|^{\alpha_j}.$$

4 Resulting Algorithm

Problem: reminder. We know the probability distributions for $t_1 = c_1 \cdot \Delta x_1$ and $t_2 = c_2 \cdot \Delta x_2$, We want to find the probability distribution for

$$t = t_1 + t_2 = c_1 \cdot \Delta x_1 + c_2 \cdot \Delta x_2.$$

Motivations: reminder. By repeating this procedure several times, we get:

- the probability distribution for $c_1 \cdot \Delta x_1 + c_2 \cdot \Delta x_2 = \sum_{i=1}^{2} c_i \cdot \Delta x_i$,
- then the distribution for $(c_1 \cdot \Delta x_1 + c_2 \cdot \Delta x_2) + c_3 \cdot \Delta x_2 = \sum_{i=1}^{3} c_i \cdot \Delta x_i$,
- then the distribution for

$$\left(\sum_{i=1}^{3} c_i \cdot \Delta x_i \right) + c_4 \cdot \Delta x_4 = \sum_{i=1}^{4} c_i \cdot \Delta x_i,$$

  etc.,

- until we get the desired probability distribution for the measurement error

$$\Delta y = \sum_{i=1}^{n} c_i \cdot \Delta x_i.$$

Preliminary step. We select the values $\alpha_1 < \ldots < \alpha_k$. For example, we can have these values uniformly distributed on the interval $[1, 2]$, by taking $\alpha_j = 1 + \dfrac{j-1}{k-1}$. For example:

- for $k = 2$, we get $\alpha_1 = 1$ and $\alpha_2 = 2$,
- for $k = 3$, we get $\alpha_1 = 1$, $\alpha_2 = 1.5$, and $\alpha_3 = 2$.

Comment. A (slightly) better selection of the values α_j is described in the Appendix.

First step: computing characteristic functions. First, we apply Fourier transforms to the given probability distributions $\rho_i(t)$, and get the corresponding characteristic functions $\chi_i(\omega)$.

Second step: approximating characteristic functions. For each of the two characteristic functions, to find the parameters μ_i, $A_{i1}, \ldots, A_{ik}$, we use the Least Squares method to solve the following system of approximate equations:

$$-\ln(\chi_i(\omega)) \cdot \chi_i(\omega) \approx \chi_{a,i}(\omega) \stackrel{\text{def}}{=} -\mathrm{i} \cdot \mu_i \cdot \omega \cdot \chi_i(\omega) + \sum_{j=1}^{k} A_{ij} \cdot \chi_i(\omega) \cdot |\omega|^{\alpha_j}$$

for values $\omega = \omega_1, \omega_2, \ldots, \omega_N$.

Comment. If the resulting approximation error $\int (\chi_i(\omega) - \chi_{a,i}(\omega))^2 \, d\omega$ is too large, we can increase k – and thus, get a better approximation.

Final step: describing the desired probability distribution for $t = t_1 + t_2$. As a good approximation for the characteristic function $\chi(\omega)$ of the probability distribution for the sum $t = t_1 + t_2$, we can take the expression

$$\chi_a(\omega) = \exp\left(\mathrm{i} \cdot \mu \cdot \omega - \sum_{j=1}^{k} A_j \cdot |\omega|^{\alpha_j}\right),$$

where $\mu = \mu_1 + \mu_2$ and $A_j = A_{1j} + A_{2j}$ for $j = 1, \ldots, k$.

5 Numerical Example

We tested our method on several examples, let us provide one such example.

Let us assume that t_1 is normally distributed with 0 mean and standard deviation1. For this distribution, the characteristic function takes the form

$$\chi_1(\omega) = \exp\left(-\frac{1}{2}\cdot\omega^2\right).$$

As t_2, let us take the Laplace distribution, with probability density $\rho_2(t) = \frac{1}{2}\cdot\exp(-|t|)$ and the characteristic function $\chi_2(\omega) = \frac{1}{1+\omega^2}$.

To approximate both distributions, we used $k = 3$, with $\alpha_1 = 1$, $\alpha_2 = 1.5$, and $\alpha_3 = 2$. In this case, the first distribution is represented exactly, with

$$\mu_1 = 0, \quad A_{11} = A_{12} = 0, \quad \text{and} \quad A_{13} = \frac{1}{2}.$$

To find the optimal approximation for the characteristic function of the Laplace distribution, we used the values ω uniformly distributed on the interval $[-5, 5]$. As a result, we get the following approximation:

$$\mu_2 = 0, \quad A_{21} = -0.162, \quad A_{22} = 1.237, \quad \text{and} \quad A_{23} = -0.398.$$

Thus, for the characteristic function of the sum $t = t_1 + t_2$ we get

$$\mu = 0, \quad A_1 = -0.162, \quad A_2 = 1.237, \quad \text{and} \quad A_3 = 0.102.$$

By applying the inverse Fourier transform to this distribution, we get an approximate probability density function $\rho_a(t)$ for the sum. The comparison between the actual probability distribution $\rho(t)$ and the approximate pdf $\rho_a(t)$ is given on Fig. 1. The corresponding mean square error $\sqrt{\int(\rho(t)-\rho_a(t))^2\,dt}$ is equal to 0.01.

Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, and by an award "UTEP and Prudential Actuarial Science Academy and Pipeline Initiative" from Prudential Foundation.

The authors are thankful to Witold Pedrycz and Shyi-Ming Chen for their encouragement.

References

[1] S. G. Rabinovich, *Measurement Errors and Uncertainty: Theory and Practice*, Springer Verlag, Berlin, 2005.

[2] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.

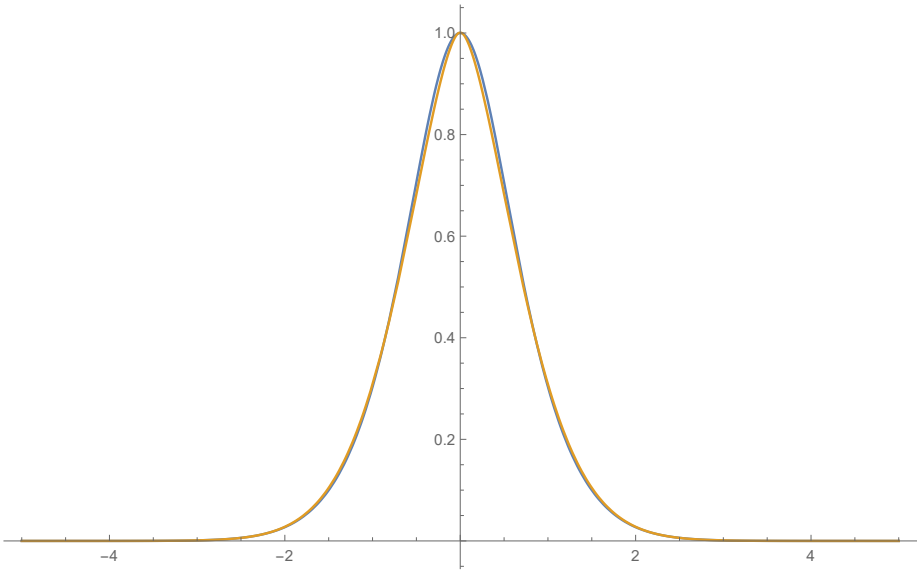

Figure 1: How good is the proposed approximation

[3] C. D. Stylios, A. Pownuk, and V. Kreinovich, "Sometimes, it is beneficial to process different types of uncertainty separately", *Proceedings of the Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2015 and 5th World Conference on Soft Computing*, Redmond, Washington, August 17–19, 2015.

A Non-Uniform Distribution of α_j is Better

Idea. If we select two values α_j too close to each other, there will be too much correlation between them, so adding the function corresponding to the second value does not add much information to what we know from a function corresponding to the first value.

We are approximating a general function (logarithm of a characteristic function) as a linear combination of functions $|t|^{\alpha_j}$. If two values α_j and α_{j+1} are close, then the function $|t|^{\alpha_{j+1}}$ can be well approximated by a term linear in $|t|^{\alpha_j}$, thus, the term proportional to $|t|^{\alpha_{j+1}}$ is not needed.

It therefore makes sense to select the values α_j in such as way that for each j, the part of $|t|^{\alpha_{j+1}}$ that cannot be approximated by terms proportional to $|t|^{\alpha_j}$ should be the largest possible.

Let us reformulate this idea in precise terms. For every two functions $f(t)$ and $g(t)$, the part of $g(t)$ which cannot be represented by terms $a \cdot f(t)$ (proportional to $f(t)$) can be described as follows. It is reasonable to describe

the difference between the two functions $f(t)$ and $g(t)$ by the least squares (L^2) metric $\int (f(t)-g(t))^2\,dt$. In these terms, the value of a function itself itself can be described as its distance from 0, i.e., as $\int (f(t))^2\,dt$.

When we approximate a function $g(t)$ by a term $a\cdot f(t)$, then the remainder $g(t)-a\cdot f(t)$ has the value $\int (g(t)-a\cdot f(t))^2\,dt$. The best approximation occurs when this value is the smallest, i.e., when it is equal to $\min\limits_a \int (g(t)-a\cdot f(t))^2\,dt$. Out of the original value $\int (g(t))^2\,dt$, we have unrepresented the part equal to $\min\limits_a \int (g(t)-a\cdot f(t))^2\,dt$. Thus, the relative size of what cannot be represented by terms $a\cdot f(t)$ can be defined as a ratio

$$R(f(t),g(t))=\frac{\min\limits_a \int (g(t)-a\cdot f(t))^2\,dt}{\int (g(t))^2\,dt}.$$

Let us simplify the resulting expression. This expression can be simplified if we find the explicit expression for a for which the value $\int (g(t)-a\cdot f(t))^2\,dt$ is the smallest possible. Differentiating the minimized expression with respect to a and equating the derivative to 0, we conclude that

$$-\int (g(t)-a\cdot f(t))\cdot f(t)\,dt=0,$$

i.e., that

$$a\cdot \int (f(t))^2\,dt=\int f(t)\cdot g(t)\,dt,$$

and

$$a=\frac{\int f(t)\cdot g(t)\,dt}{\int (f(t))^2\,dt}.$$

For this a, the value $\int (g(t)-a\cdot f(t))^2\,dt$ takes the form

$$\int (g(t)-a\cdot f(t))^2\,dt=\int (g(t))^2\,dt-2a\cdot\int f(t)\cdot g(t)\,dt+a^2\cdot\int (f(t))\,dt.$$

Substituting the above expression for a into this formula, we conclude that

$$\int (g(t)-a\cdot f(t))^2\,dt=\int (g(t))^2\,dt-\frac{2(\int f(t)\cdot g(t)\,dt)^2}{\int (f(t))^2\,dt}+\frac{(\int f(t)\cdot g(t)\,dt)^2}{\int (f(t))^2\,dt},$$

i.e., that

$$\int (g(t)-a\cdot f(t))^2\,dt=\int (g(t))^2\,dt-\frac{(\int f(t)\cdot g(t)\,dt)^2}{\int (f(t))^2\,dt}.$$

Thus, the desired ratio takes the form

$$R(f(t),g(t))\stackrel{\text{def}}{=}\frac{\min\limits_a \int (g(t)-a\cdot f(t))^2\,dt}{\int (g(t))^2\,dt}=1-\frac{(\int f(t)\cdot g(t)\,dt)^2}{(\int (f(t))^2\,dt)\cdot(\int (g(t))^2\,dt)}.$$

Thus, we arrive at the following optimization problem.

Resulting optimization problem. To make sure that the above remainders are as large as possible, it makes sense to find the values $\alpha_1^{\text{opt}} < \ldots < \alpha_k^{\text{opt}}$ that maximize the smallest of the remainders between the functions $f(t) = |t|^{\alpha_j}$ and $g(t) = |t|^{\alpha_{j+1}}$:

$$\min_j R\left(|t|^{\alpha_j^{\text{opt}}}, |t|^{\alpha_{j+1}^{\text{opt}}}\right) = \max_{\alpha_1 < \ldots < \alpha_k} \min_j R(|t|^{\alpha_j}, |t|^{\alpha_{j+1}}).$$

Solving the optimization problem. Let us consider an interval $[-T, T]$ for some T. Since the function is symmetric, it is sufficient to consider the values from $[0, T]$.

For $f(t) = t^\alpha$ and $g(t) = t^\beta$, the integral in the numerator of the ratio is equal to

$$\int_0^T f(t) \cdot g(t)\, dt = \int_0^T t^\alpha \cdot t^\beta\, dt = \int_0^T t^{\alpha+\beta}\, dt = \frac{T^{\alpha+\beta+1}}{\alpha+\beta+1}.$$

Similarly, the integrals in the denominator take the form

$$\int_0^T f^2(t)\, dt = \int_0^T t^{2\alpha}\, dt = \frac{T^{2\alpha+1}}{2\alpha+1}$$

and

$$\int_0^T g^2(t)\, dt = \int_0^T t^{2\beta}\, dt = \frac{T^{2\beta+1}}{2\beta+1},$$

so

$$R = 1 - \frac{\dfrac{T^{2(\alpha+\beta+1)}}{(\alpha+\beta+1)^2}}{\dfrac{T^{2\alpha+1}}{2\alpha+1} \cdot \dfrac{T^{2\beta+1}}{2\beta+1}}.$$

One can see that the powers of T cancel each other, and we get

$$R = 1 - \frac{(2\alpha+1)\cdot(2\beta+1)}{(\alpha+\beta+1)^2},$$

or, equivalently, if we denote $r \stackrel{\text{def}}{=} \dfrac{\beta+0.5}{\alpha+0.5}$, we get

$$R = R(r) \stackrel{\text{def}}{=} 1 - 4 \cdot \frac{r}{(1+r)^2}.$$

The derivative of the function $R(r)$ is equal to

$$\frac{dR}{dr} = -4 \cdot \frac{(1+r)^2 - 2\cdot(1+r)}{(1+r)^4} = -4 \cdot \frac{(1+r)\cdot(1+r-2)}{(1+r)^4} =$$

$$4 \cdot \frac{(1+r)\cdot(r-1)}{(1+r)^4} = 4 \cdot \frac{r-1}{(1+r)^3}.$$

So this derivative is positive for all $r > 1$. Thus, the function $R(r)$ is monotonically increasing, and looking for the values α_j^{opt} for which $\min\limits_j R(|t|^{\alpha_j}, |t|^{\alpha_{j+1}})$ is the largest is equivalent to looking for the values α_j^{opt} for which the smallest $\min\limits_j \dfrac{\alpha_{j+1}+0.5}{\alpha_j+0.5}$ of the ratios $r = \dfrac{\alpha_{j+1}+0.5}{\alpha_j+0.5}$ attains the largest possible value:

$$\min_j \frac{\alpha_{j+1}^{\text{opt}}+0.5}{\alpha_j^{\text{opt}}+0.5} = \max_{\alpha_1<\ldots<\alpha_k} \min_j \frac{\alpha_{j+1}+0.5}{\alpha_j+0.5}.$$

One can check that this happens when $\alpha_j + 0.5 = 1.5 \cdot \left(\dfrac{5}{3}\right)^{(j-1)/(k-1)}$. Indeed, in this case, $\min\limits_j \dfrac{\alpha_{j+1}+0.5}{\alpha_j+0.5} = \left(\dfrac{5}{3}\right)^{1/(k-1)}$. We cannot have it larger: if we had $\min\limits_j \dfrac{\alpha_{j+1}+0.5}{\alpha_j+0.5} > \left(\dfrac{5}{3}\right)^{k-1}$, then we would have $\dfrac{\alpha_{j+1}+0.5}{\alpha_j+0.5} > \left(\dfrac{5}{3}\right)^{k-1}$ for all j. Here,

$$\alpha_k + 0.5 = (\alpha_1+0.5)\cdot\frac{\alpha_2+0.5}{\alpha_1+0.5}\cdot\frac{\alpha_3+0.5}{\alpha_2+0.5}\cdot\ldots\cdot\frac{\alpha_k+0.5}{\alpha_{k-1}+0.5}.$$

The first factor $\alpha_1 + 0.5$ is ≥ 1.5, each of the other $k-1$ terms is greater than $\left(\dfrac{5}{3}\right)^{1/(k-1)}$, so for their product, we get

$$\alpha_k + 0.5 > 1.5 \cdot \left(\left(\frac{5}{3}\right)^{1/(k-1)}\right)^{k-1} = 1.5 \cdot \frac{5}{3} = 2.5,$$

while we assumed that all the values α_j are from the interval $[1, 2]$, and so, we should have $\alpha_k + 0.5 \leq 2.5$.

Resulting optimal values of α_j. Thus, the optimal way is to not to take the values uniformly distributed on the interval $[1, 2]$, but rather take the values

$$\alpha_j^{\text{opt}} = 1.5 \cdot \left(\frac{5}{3}\right)^{(j-1)/(k-1)} - 0.5$$

for which the logarithms $\ln(\alpha_j^{\text{opt}}+0.5) = \dfrac{j-1}{k-1}\cdot\ln\left(\dfrac{5}{3}\right) = \ln(1.5)$ are uniformly distributed.

Comment. It is worth mentioning that there is intriguing connection between these values α_j and music: for example, the twelves notes on a usual Western octave correspond to the frequencies:

- f_1,
- $f_2 = f_1 \cdot 2^{1/12}$,
- $f_3 = f_1 \cdot 2^{2/12}$, ...,
- $f_{12} = f_1 \cdot 2^{11/12}$, and
- $f_{13} = f_1 \cdot 2$,

for which the logarithms $\ln(f_j)$ are uniformly distributed. Similar formulas exist for five-note and other octaves typical for some Oriental musical traditions.