Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ORIGINAL ARTICLE

Cross-Correlation and Averaging: An Equivalence Based on the Classical Probability Density

Attila Genda*[1] | Alexander Fidlin[1] | Oleg Gendelman[2]

[1]Institute of Engineering Mechanics, Karlsruhe Institute of Technology, Karlsruhe, Germany
[2]Mechanical Engineering, Technion – Israel Institute of Technology, Haifa, Israel

Correspondence
*Attila Genda. Email: attila.genda@kit.edu

Present Address
Building 10.23, 1st floor, Room 101, Kaiserstr. 10, 76131, Karlsruhe, Germany

Summary

The averaging method is a widely used technique in the field of nonlinear differential equations for effectively reducing systems with "fast" oscillations overlaying "slow" drift. The method involves calculating an integral, which can be straightforward in some cases, but can also require simplifications such as series expansions. We propose an alternative approach that relies on the classical probability density (CPD) of the "fast" variable. Further, we demonstrate the equivalence between the averaging integral and the cross-correlation product of the CPD and the target function. This equivalence simplifies handling many problems, particularly those involving piecewise-defined target functions. We propose an effective numerical method to calculate the averaged function, exploiting the well-known mathematical properties of cross-correlation products.

KEYWORDS:
averaging, cross-correlation, multiple time scales, system reduction, slow-fast dynamical systems, classical probability density

1 | INTRODUCTION

The averaging method is widely used in nonlinear differential equations to reduce complex systems with "fast" oscillations and "slow" drifts. It is an essential tool for analyzing and synthesizing such systems, as it enables the effective representation of the underlying dynamics. The technique has been extensively studied and applied in various fields, including physics[1–4], engineering[5–8], and biology[9].

The basic idea behind the averaging method is to transform a rapidly oscillating system into a slowly varying one maintaining the significant properties of the original one. Then, the averaged system can give more analytical insight into the system's "slow" dynamics or might reduce computational costs. The simplification is achieved by calculating the time average of the rapidly oscillating variables over a period of the "fast" motion.

While the idea of averaging is much older and has been used many years before, in 1934 *Krylov and Bogulyubov* developed first a general averaging approach and showed that the solution of the averaged system approximates the exact dynamics[10,11].

Over the past few decades, the averaging method has been the subject of numerous studies, and various approaches have been proposed to address its limitations and improve its accuracy. For example, in[12], the author proposed modifying the classical averaging method that accounts for higher-order terms in the expansion. The averaging technique has also been applied in stochastic differential equations[13]: a stochastic averaging method for studying the effects of time-delayed feedback control on quasi-integrable Hamiltonian systems subjected to Gaussian white noise was proposed in[14]. In[15], the authors proposed a stochastic averaging technique to analyze randomly excited single-degree-of-freedom (SDOF) strongly nonlinear systems with delayed feedback fractional-order proportional-derivative (PD) control.

The growth of the knowledge on the averaging method resulted in the publication of numerous articles[16–18], book chapters[19–21] and comprehensive monographs such as *Averaging Methods in Nonlinear Dynamical Systems*[22] and *Nonlinear oscillations in mechanical engineering*[23]. These works offer an in-depth look at the theory and practical applications of the averaging method in various areas of study.

In addition, the averaging method has also been applied to various real-world problems, including the analysis of power electronic systems[24,25] and the study of climate dynamics[26].

Parallel to its analytic applications, the averaging method finds its numerical applications as well[27]. In[28] *Leimkuhler and Reich* described a reversible staggered time-stepping method for simulating long-term dynamics formulated on two or more time scales.

The above examples demonstrate the versatility and importance of the averaging method in studying nonlinear differential equations. Despite its numerous applications and improvements, the method still presents many challenges and opportunities for further research.

Although calculating the averaging integral is typically straightforward, obtaining an analytic expression can pose algebraic difficulties in many cases. Additionally, there may be a need to perform averaging using numerical or experimental data. Traditional quadrature methods for averaging require time series data, which can be particularly challenging to obtain for fast motions. However, measuring the probability of finding a particle at a specific position can often be a more accessible approach. For example, a camera with a sufficiently long exposure time can generate an image with pixel brightness proportional to the CPD.

In the present paper, an alternative formula is proposed to evaluate the averaging integral for the case of one dependent variable that can be decomposed into the sum of a "fast" motion and a "slow" drift. We show that the averaging integral is equivalent to the cross-correlation of the "fast" motion's CPD and the original function subjected to averaging. In general, the average of the function is not necessarily equal to the function of the average.

The CPD is a well-known concept in classical and quantum mechanics[29] describing the probability that a particle, following a certain motion, can be found at a given position. For simple, periodic motions, the CPD can be derived analytically[30–32]. However, the evaluation is often possible numerically only for more complex motions.

In the following, we will interchangeably use the terms probability density function (PDF) and CPD since their mathematical properties are identical.

The paper is structured as follows: in Sec. 2, the equivalence of the averaging integral and the CPD-based cross-correlation of the target function is proven, followed by some important implications on the efficient calculation and properties of the averaged function. In Sec. 3, the CPDs of several types of oscillations are derived. In Sec. 4, different methods are described for the efficient numerical calculation of CPDs and the averaged function itself. Sec. 5 summarizes the paper's results and gives scope for applications and further research of the averaging method based on cross-correlation. The appendix contains additional information on the moments and partial moments of the arcsine distribution.

2 | THEOREM ON CROSS-CORRELATION BASED AVERAGING

Main result. We consider the averaging of a scalar-valued function of the form $f(x_S + x_F)$ where the "slow" variable is denoted by x_S and the "fast" variable by x_F. We assume, furthermore, that the "fast" variable x_F can be given as a periodic function of the time $g(t)$ with the time period T (for almost periodic functions[33], let $T \to \infty$). Hence, $x_F = g(t)$, thus, one has $f(x_S + g(t))$. We prove that the time average of this function is equivalent to the cross-correlation of $f(x)$ and the CPD $\rho(x)$ of the "fast" variable $x_F(t)$, i.e.

$$\frac{1}{T}\int_{0}^{T} f(x_S + g(t))\mathrm{d}t = \int_{-\infty}^{\infty} f(x)\rho(x - x_S)\mathrm{d}x. \tag{1}$$

Definition 1. Let $g_i : (a, b) \to \mathbb{R}$ be either a strictly monotonically increasing C^1 function with the parameter $d_i = 0$ when the sign of its derivative is positive or a strictly monotonically decreasing C^1 function with the parameter $d_i = 1$ when the sign of its derivative is negative. Then, its CPD is defined[34] by

$$\rho_i : \begin{cases} \mathbb{R} & \to \mathbb{R}^{0+} \\ x & \mapsto \frac{(-1)^{d_i}}{g_i'[g_i^{-1}(x)]} \frac{1}{b-a} \mathbf{1}_{(g_i(a),g_i(b))}(x), \end{cases} \tag{2}$$

where slightly abusing the notation (during the whole article) to set the value of ρ_i to 0 outside $(g_i(a), g_i(b))$, we use the indicator function defined as

$$\mathbf{1}_X(x) = \begin{cases} 1 & x \in X, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Note that the strict monotony guarantees the existence of the inverse. If the function $g_i(x) = C_i$ is constant on $x \in (a, b)$, its CPD is given by

$$\rho_i(x) = \delta(x - C_i), \tag{4}$$

where $\delta(\cdot)$ denotes the Dirac distribution.

Definition 2. Let g be a piecewise, continuously differentiable, periodic function with the time period T defined by

$$g : \begin{cases} (t_{i-1}, t_i) & \to \mathbb{R} \qquad\qquad \text{for } i = 1 \ldots n, \\ x & \mapsto g_i(x), \end{cases} \tag{5}$$

with $t_0 = 0$ and $t_n = T$ such that all g_i are either strictly monotonously increasing, decreasing, or constant on its domain of definition. We further define $\Delta T_i := t_i - t_{i-1}$. Then the CPD of g is defined by the weighted average

$$\rho(x) := \frac{1}{T} \sum_{i=1}^{n} \Delta T_i \rho_i(x). \tag{6}$$

Theorem 1. For a bounded function f and an at least peace-wisely continuously differentiable periodic function g with period T, the averaging operator

$$\tilde{f}(x_S) = \langle f(x_S + g(t)) \rangle = \frac{1}{T} \int_0^T f(x_S + g(t)) \mathrm{d}t \tag{7}$$

is equivalent to the cross-correlation integral

$$(\rho \star f)(x_S) = \int_{-\infty}^{\infty} f(x) \rho(x - x_S) \mathrm{d}x, \tag{8}$$

where $\rho(x)$ denotes the CPD of the "fast" variable $g(t)$.

Proof. Starting at time 0 and ending at T, the time period can be divided into n intervals, such that on n_I pieces of intervals, the function is strictly monotonously increasing, on n_D pieces of intervals strictly monotonously decreasing, and on $n - n_I - n_D$ pieces of intervals it is constant. Let us denote the division points by $t_0 = 0, t_1, \ldots t_n = T$. We denote the intervals by

$$T_i = (t_{i-1}, t_i) \subset \mathbb{R} \qquad \text{for } i = 1 \ldots n, \tag{9}$$

and their length by

$$\Delta T_i = t_i - t_{i-1} \in \mathbb{R}^+ \qquad \text{for } i = 1 \ldots n. \tag{10}$$

We define the index sets

$$S_I := \{i \in S_I | g_i \text{ is strictly monotonically increasing on } T_i\}, \tag{11}$$

$$S_D := \{i \in S_D | g_i \text{ is strictly monotonically decreasing on } T_i\}, \tag{12}$$

$$S_C := \{i \in S_C | g_i \text{ is constant on } T_i\}. \tag{13}$$

By applying the three categories, Eq. (6) can be written as

$$\rho(x) := \frac{1}{T}\left(\sum_{i\in S_I}\frac{1}{g_i'[g_i^{-1}(x)]}\mathbf{1}_{g_i(T_i)}(x) + \sum_{i\in S_D}\frac{-1}{g_i'[g_i^{-1}(x)]}\mathbf{1}_{g_i(T_i)}(x) + \sum_{i\in S_C}\Delta T_i\delta(x-C_i)\right). \tag{14}$$

Inserting it into Eq. (8) we have

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\rho(x-x_S)\mathrm{d}x = \int_{-\infty}^{\infty} f(x)\frac{1}{T}\left(\sum_{i\in S_I}\frac{1}{g_i'[g_i^{-1}(x-x_S)]}\mathbf{1}_{g_i(T_i)}(x-x_S)\right.\\ \left.+\sum_{i\in S_D}\frac{-1}{g_i'[g_i^{-1}(x-x_S)]}\mathbf{1}_{g_i(T_i)}(x-x_S) + \sum_{i\in S_C}\Delta T_i\delta(x-x_S-C_i)\right)\mathrm{d}x. \end{aligned}\tag{15}$$

Since $f(x)$ is bounded and $\rho_i(x) \leq 1$ by definition, the dominated convergence theorem assures that integration and summations signs can be interchanged (even if the number of intervals goes to infinity), hence

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\rho(x-x_S)\mathrm{d}x &= \frac{1}{T}\left(\sum_{i\in S_I}\int_{-\infty}^{\infty}\frac{f(x)}{g_i'[g_i^{-1}(x-x_S)]}\mathbf{1}_{g_i(T_i)}(x-x_S)\mathrm{d}x - \sum_{i\in S_D}\int_{-\infty}^{\infty}\frac{f(x)}{g_i'[g_i^{-1}(x-x_S)]}\mathbf{1}_{g_i(T_i)}(x-x_S)\mathrm{d}x\right.\\ &\quad\left.+\sum_{i\in S_C}\Delta T_i\int_{-\infty}^{\infty} f(x)\delta(x-x_S-C_i)\right)\mathrm{d}x \qquad (16)\\ &= \frac{1}{T}\left(\sum_{i\in S_I}\int_{g_i(t_{i-1})+x_S}^{g_i(t_i)+x_S}\frac{f(x)}{g_i'[g_i^{-1}(x-x_S)]}\mathrm{d}x - \sum_{i\in S_D}\int_{g_i(t_i)+x_S}^{g_i(t_{i-1})+x_S}\frac{f(x)}{g_i'[g_i^{-1}(x-x_S)]}\mathrm{d}x + \sum_{i\in S_C}\Delta T_i f(x_S+C_i)\right), \qquad (17)\end{aligned}$$

where the last summation term is obtained by using the sifting property of the Dirac distribution. Note that in case of decreasing intervals, the lower boundary of $g(T_i)$ is at $g(t_i)$, and the upper boundary is at $g(t_{i-1})$. Thus a change in the integration boundaries will cancel out the minus sign as follows

$$\int_{-\infty}^{\infty} f(x)\rho(x-x_S)\mathrm{d}x = \frac{1}{T}\left(\sum_{i\in S_I}\int_{g_i(t_{i-1})+x_S}^{g_i(t_i)+x_S}\frac{f(x)}{g_i'[g_i^{-1}(x-x_S)]}\mathrm{d}x + \sum_{i\in S_D}\int_{g_i(t_{i-1})+x_S}^{g_i(t_i)+x_S}\frac{f(x)}{g_i'[g_i^{-1}(x-x_S)]}\mathrm{d}x + \sum_{i\in S_C}\Delta T_i f(x_S+C_i)\right). \tag{18}$$

Now, in every non-constant interval of f, we introduce the following variable transformation, respectively

$$x = x_S + g_i(t), \qquad x_S = x - g_i(t), \qquad \mathrm{d}x = g_i'(t)\mathrm{d}t, \qquad t = g_i^{-1}(x-x_S) \tag{19}$$

and we substitute it into Eq. (18)

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\rho(x-x_S)\mathrm{d}x &= \frac{1}{T}\left(\sum_{i\in S_I}\int_{t_{i-1}}^{t_i}\frac{f(x_S+g_i(t))}{g_i'[g_i^{-1}(g_i(t))]}g_i'(t)\mathrm{d}t + \sum_{i\in S_D}\int_{t_{i-1}}^{t_i}\frac{f(x_S+g_i(t))}{g_i'[g_i^{-1}(g_i(t))]}g_i'(t)\mathrm{d}t + \sum_{i\in S_C}\Delta T_i f(x_S+C_i)\right) \qquad (20)\\ &= \frac{1}{T}\left(\sum_{i\in S_I}\int_{t_{i-1}}^{t_i} f(x_S+g_i(t))\mathrm{d}t + \sum_{i\in S_D}\int_{t_{i-1}}^{t_i} f(x_S+g_i(t))\mathrm{d}t + \sum_{i\in S_C}\int_{t_{i-1}}^{t_i} f(x_S+\underbrace{g_i(t)}_{=C_i})\mathrm{d}t\right) \qquad (21)\\ &= \frac{1}{T}\int_0^T f(x_S+g(t))\mathrm{d}t. \qquad (22)\end{aligned}$$

□

Remark 1. In technically relevant applications, almost periodic "fast" motions, such as, for example, the sum of two sines with incommensurable frequencies, often arise. Choosing the interval boundaries will influence the result if one wants to average such functions on a finite time interval. $g(t)$ being almost periodic, no time period exists in this case. However, one can take $T \to \infty$ to obtain a uniquely-defined integral (cf. Sec. 3.2.2).

This result is important for several reasons; it facilitates the calculation of averaged values in the case of piecewise-defined functions. Further, the numerical calculation of averaged values also becomes simpler since, numerically, the CPD is very easy to obtain; it is enough to evaluate the "fast" movement on a time period in N pieces of equidistantly positioned time instants and to make the histogram of the obtained data. It is well known that for $N \to \infty$, the histogram approaches the PDF/CPD[35].

Corollary 1. Since Eq. (8) is the cross-correlation of the functions $f(x)$ and $\rho(x)$, the averaging problem in Eq. (1) can be transformed into the Fourier domain if the product $f(x)\rho(x - x_S)$ is $L^1(\mathbb{R})$. (By the boundedness of $f(x)$, this criterion can always be ensured by multiplying $f(x)$ with a window function $w(x)$ to restrict it to the technically relevant region. The area under the curve of $\rho(x)$ is 1 by definition, thus $w(x)f(x)\rho(x - x_S) \in L^1(\mathbb{R})$). The cross-correlation integral in the Fourier domain becomes a product, and through inverse Fourier transformation, the averaged function can be obtained rapidly, i.e.

$$\mathcal{F}\{\rho \star f\} = \overline{\mathcal{F}\{\rho\}} \cdot \mathcal{F}\{f\}. \tag{23}$$

The numerical calculation of the averaged value of $f(x)$ can also be performed using effective methods relying on the numerical equivalents of the Fourier transformation, for example, the fast Fourier transform (FFT) algorithm.

Corollary 2. Eq. (8) remains valid if the "fast" variable x_F depends on x_S, i.e. $x_F = g(t, x_S)$.

Proof. It is easy to see that x_S plays the role of a constant through the calculations; thus, the proof remains valid if we allow the dependency of the "fast" variable on the "slow" one. □

Corollary 3. Assume that $m_1 = \int_{-\infty}^{\infty} x\rho(x)\mathrm{d}x = 0$. By this and the fact that $m_0 = \int_{-\infty}^{\infty} \rho(x)\mathrm{d}x = 1$, affine functions, i.e., of the form $f(x) = ax + b$, remain unchanged under the application of the cross-correlation integral. Hence, under the above assumptions, affine functions are eigenfunctions of the averaging operator with eigenvalue $\lambda = 1$.

Proof.

$$\tilde{f}(y) = \int_{-\infty}^{\infty} (ax + b)\rho(x - y)\mathrm{d}x = a \underbrace{\int_{-\infty}^{\infty} x\rho(x - y)\mathrm{d}x}_{=y} + b \underbrace{\int_{-\infty}^{\infty} \rho(x - y)\mathrm{d}x}_{=1} = ay + b. \tag{24}$$

□

Definition 3. The k^{th} *moment* of a random variable X described by its PDF $\rho(x)$ is defined by

$$m_k = \mathrm{E}(X^k) = \int_{-\infty}^{\infty} x^k \rho(x)\mathrm{d}x. \tag{25}$$

Definition 4. The k^{th} *partial moment* of a random variable X described by its PDF $\rho(x)$ is defined by

$$m_k(x) = \mathrm{E}_x(X^k) = \int_{-\infty}^{x} y^k \rho(y)\mathrm{d}y. \tag{26}$$

Lemma 1. All partial moments of $\rho(x)$ are bounded and exist if the range of $g(t)$ is bounded.

Proof. Let $F(x) = \int_{-\infty}^{x} \rho(\tilde{x})\mathrm{d}\tilde{x}$ denote the cumulative density function of X. Since the range of $g(t)$ is bounded, $\rho(x)$ has compact support with $x_l := \inf g(t)$ and $x_u := \sup g(t)$. In the range of interest, we have $x_l < x < x_u$. Since $\rho(x)$ is a PDF, $\int_{-\infty}^{\infty} \rho(x)\mathrm{d}x = 1$. Let us define

$$L_k(x) = \min_{y \in (x_l, x)} y^k \quad \text{and} \quad U_k(x) = \max_{y \in (x_l, x)} y^k. \tag{27}$$

Then, we have

$$L_k(x)F(x) = L_k(x) \int_{x_l}^{x} \rho(y)\mathrm{d}y = \int_{-\infty}^{x} L_k(x)\rho(y)\mathrm{d}y \leq \int_{-\infty}^{x} y^k \rho(y)\mathrm{d}y \leq \int_{-\infty}^{x} U_k(x)\rho(y)\mathrm{d}y = U_k(x) \int_{x_l}^{x} \rho(y)\mathrm{d}y = U_k(x)F(x). \tag{28}$$

Thus all partial moments are bounded and therefore exist. By inserting $x = x_u$, it is also shown that all moments exist. □

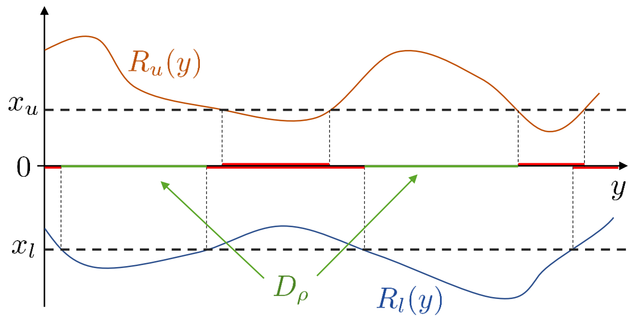


Figure 1 Visual interpretation of the definition of D_ρ

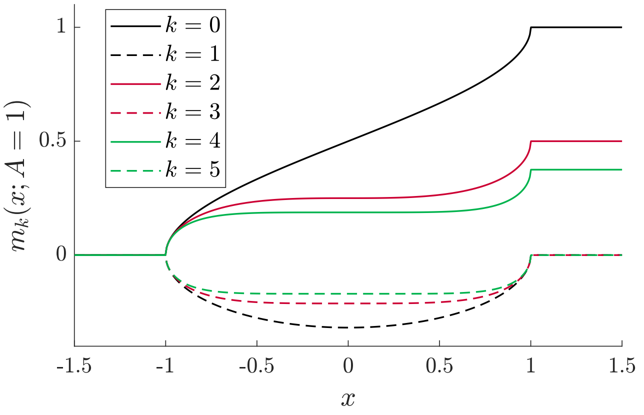


Figure 2 Partial moments of the arcsine distribution with $A = 1$

Theorem 2. Assume that $f(x)$ is a real analytic function and has the domain of convergence $D(y) = (y + R_l(y), y + R_u(y))$ when expanded into Taylor series around y with the non-positive valued function $R_l(y)$ and non-negative valued one $R_u(y)$. Further, assume that the range of the "fast" variable $g(t)$ is $[x_l, x_u]$, i.e., $\sup g(t) = x_u$ and $\inf g(t) = x_l$. Without loss of generality, we assume $m_1 = 0$, thus $x_l \leq 0 \leq x_u$. We define the set

$$D_\rho = \{(y \in \mathbb{R} | (R_l(y) < x_l) \wedge (x_u < R_u(y))\}, \tag{29}$$

i.e., the set of points around which the convergence radius of $f(x)$ is large enough that the support of $\rho(x)$ fits into it (cf. Fig. 1). Then, the following holds

$$\tilde{f}(y) = \int_{-\infty}^{\infty} f(x)\rho(x-y)\mathrm{d}x = \sum_{k=0}^{\infty} m_k \frac{f^{(k)}(y)}{k!} \quad \text{for } y \in D_\rho, \tag{30}$$

where $(\cdot)^{(k)}(x)$ denotes the k^{th} derivative.

Proof. Taylor expansion of $f(x)$ around y and interchanging the summation and integral signs (allowed due to dominated convergence) yields

$$\tilde{f}(y) = \int_{-\infty}^{\infty} \left(\sum_{k=0}^{\infty} \frac{f^{(k)}(y)}{k!}(x-y)^k \right) \rho(x-y)\mathrm{d}x = \sum_{k=0}^{\infty} \left(\frac{f^{(k)}(y)}{k!} \int_{-\infty}^{\infty} x^k \rho(x)\mathrm{d}x \right) = \sum_{k=0}^{\infty} m_k \frac{f^{(k)}(y)}{k!} \quad \text{for } y \in D_\rho. \tag{31}$$

□

Eq. (30) demonstrates that the feasibility of obtaining an analytic expression for the averaged value of $f(x_S + g(t))$ depends solely on whether the corresponding moments of $\rho(x)$ are known. Often the moments can be obtained with the moment-generating function or the characteristic function of the corresponding CPD/PDF. Furthermore, the moments can be estimated easily if equidistant experimental/simulation time series of the "fast" motion is available:

$$\hat{m}_k = \frac{\sum_{i=1}^{n} x_i^k}{n}, \tag{32}$$

where x_i is the i^{th} time instance in the n-element data series. The variance of the moments estimator is obtained by

$$\mathrm{Var}(\hat{m}^k) = \mathrm{Var}\left(\frac{\sum_{i=1}^{n} X_i^k}{n} \right) = \frac{\mathrm{Var}(X_1^k)}{n} = \frac{m_{2k} - m_k^2}{n}. \tag{33}$$

Eq. (33) shows that the estimator gets more accurate with a larger sample size. However, higher moments are more sensitive regarding the tails of the distribution. Therefore, noisy measurement data requires a larger sample size to obtain accurate estimates for higher moments.

Theorem 2 is especially important in two cases: a) the target function is a polynomial; thus, only a finite number of moments are needed to obtain the average, or b) the support of ρ is small, and for $k \to \infty$ we have $m^k \to 0$. We prove this second statement in the following.

Corollary 4. We keep the assumptions of Theorem 2 and further assume that ρ has short support, i.e., $x_u - x_l = \varepsilon$. We also assume that all derivatives of the target function $f^{(k)}(y)$ are of $\mathcal{O}(1)$. Then, the difference between the original and the averaged target function $\tilde{f}(y) - f(y)$ is uniformly of $\mathcal{O}(\varepsilon^2)$ for $y \in D_\rho$.

Proof. By the fact that ρ is a PDF and by the assumption $m_1 = 0$, Theorem 2 yields

$$\tilde{f}(y) = f(y) + \sum_{k=2}^{\infty} m_k \frac{f^{(k)}(y)}{k!} \quad \text{for } y \in D_\rho. \tag{34}$$

We also have $-\varepsilon \le x_l \le 0 \le x_u \le \varepsilon$ and make use of Lemma 1 by setting $L_k(x) = -|x_l|^k$ and $U_k(x) = x_u^k$, thus the moments are bounded by

$$-\varepsilon^k \le m_k \le \varepsilon^k, \tag{35}$$

hence

$$\tilde{f}(y) = f(y) + \mathcal{O}(\varepsilon^2), \tag{36}$$

and

$$|\tilde{f}(y) - f(y)| = \mathcal{O}(\varepsilon^2) \quad \text{for } y \in D_\rho. \tag{37}$$

□

It is an important implication since it shows that under the above assumptions, sufficiently smooth functions are not altered much by averaging if the "fast" variable has a small range. Furthermore, since $\varepsilon < 1$ we also have $\lim_{k\to\infty} \varepsilon^k = 0$.

The averaging method might also be applied when defining the target function piecewise. Similar results can be formulated using partial moments of the "fast" motion's CPD.

Theorem 3. Assume that the range of the "fast" variable $g(t)$ is $[x_l, x_u]$, i.e. $x_u = \sup g(t)$ and $x_l = \inf g(t)$. Without loss of generality, we assume $m_1 = 0$, thus $x_l \le 0 \le x_u$. Let the target function $f(x)$ be composed of m pieces of analytic functions $f_i(x)$ with $i \in \mathcal{I} := \{1, \dots, m\}$, i.e.,

$$f(x) = \sum_{i=1}^{m} f_i(x) \mathbf{1}_{(x_{i-1}, x_i)}(x), \tag{38}$$

with $f_i(x)$ convergent on the domains $D_i(y) = (y + R_{l,i}(y), y + R_{u,i}(y))$ when expanded into Taylor series around y, where $R_{l,i}(y)$ and $R_{u,i}(y)$ are non-positive and non-negative functions, respectively. Let the domain boundaries be given by $x_0, \dots, x_m$ with $x_0 = -\infty$ and $x_m = \infty$. We denote the domains by $d_i := [x_{i-1}, x_i]$. Further, we define the set-valued function

$$\mathcal{I}_A(y) := \{i \in \mathcal{I} | d_i \cap [y + x_l, y + x_u] \neq \emptyset\} \tag{39}$$

denoting the indices of the set of active functions, i.e., those where $\rho(x - y) > 0$ for any $x \in [x_{i-1}, x_i]$. We define

$$D(y) := \bigcap_{i \in \mathcal{I}_A(y)} D_i(y) \quad \Rightarrow \quad D(y) = (y + R_l(y), y + R_u(y)) \tag{40}$$

with

$$R_l(y) := \max_{i \in \mathcal{I}_A(y)} R_{l,i}(y) \quad \text{and} \quad R_u(y) := \min_{i \in \mathcal{I}_A(y)} R_{u,i}(y). \tag{41}$$

Furthermore, we define D_ρ as in Theorem 2. Then,

$$\tilde{f}(y) = \sum_{i \in \mathcal{I}_A(y)} \sum_{k=0}^{\infty} \frac{f_i^{(k)}(y)}{k!} \left(m_k(x_i - y) - m_k(x_{i-1} - y) \right) \quad \text{for } y \in D_\rho, \tag{42}$$

where $m_k(x)$ denotes the k^{th} partial moment as defined in Definition 4.

Proof. As long as $x \in D(y)$, we can evaluate $f(x)$ using its Taylor expansion around any y by

$$f(x) = \sum_{i \in \mathcal{I}_A(y)} f_i(x) \mathbf{1}_{(x_{i-1}, x_i)}(x) = \sum_{i \in \mathcal{I}_A(y)} \left(\sum_{k=0}^{\infty} \frac{f_i^{(k)}(y)}{k!} (x-y)^k \right) \mathbf{1}_{(x_{i-1}, x_i)}(x), \quad \text{for } x \in D(y). \tag{43}$$

Inserting Eq. (43) into the averaging integral Eq. (1) we have

$$\tilde{f}(y) = \int_{-\infty}^{\infty} f(x) \rho(x-y) \mathrm{d}x = \int_{-\infty}^{\infty} \left(\sum_{i \in \mathcal{I}_A(y)} \left(\sum_{k=0}^{\infty} \frac{f_i^{(k)}(y)}{k!} (x-y)^k \right) \mathbf{1}_{(x_{i-1}, x_i)}(x) \right) \rho(x-y) \mathrm{d}x \tag{44}$$

$$= \int_{-\infty}^{\infty} \sum_{i \in \mathcal{I}_A(y)} \left(\sum_{k=0}^{\infty} \frac{f_i^{(k)}(y)}{k!} x^k \rho(x) \right) \mathbf{1}_{(x_{i-1}-y, x_i-y)}(x) \mathrm{d}x, \tag{45}$$

and by dominated convergence, summation, and integral signs can be interchanged, leading to

$$\tilde{f}(y) = \sum_{i \in \mathcal{I}_A(y)} \sum_{k=0}^{\infty} \frac{f_i^{(k)}(y)}{k!} \left(\int_{-\infty}^{\infty} x^k \rho(x) \mathbf{1}_{(x_{i-1}-y, x_i-y)}(x) \mathrm{d}x \right) \tag{46}$$

$$= \sum_{i \in \mathcal{I}_A(y)} \sum_{k=0}^{\infty} \frac{f_i^{(k)}(y)}{k!} \left(\int_{x_{i-1}-y}^{x_i-y} x^k \rho(x) \mathrm{d}x \right) \tag{47}$$

$$= \sum_{i \in \mathcal{I}_A(y)} \sum_{k=0}^{\infty} \frac{f_i^{(k)}(y)}{k!} \left(m_k(x_i - y) - m_k(x_{i-1} - y) \right) \quad \text{for } y \in D_\rho. \tag{48}$$

□

Eq. (42) shows that the knowledge of the partial moments of the "fast" variable's CPD is sufficient to calculate the average of the target function.

The following two examples demonstrate the usefulness of Theorem 2 and Theorem 3.

Example 1. Calculate the average of $f(x+g(t))$ with $f(x) = x^2$ and $g(t) = A\sin(\omega t)$. The average can be obtained classically by calculating

$$\tilde{f}(x) = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} (x + A\sin\omega t)^2 \mathrm{d}t = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} \left(x^2 + 2xA\sin\omega t + A^2 \sin^2\omega t \right) \mathrm{d}t = x^2 + \frac{A^2}{2}. \tag{49}$$

Alternatively, the averaged function can be calculated using Eq. (30). It is well known that the CPD of a harmonic motion (cf. Eq. (120)) with amplitude A is given by the arcsine distribution[35] with half-width A. The derivation of its moments is given in the appendix. For now, the moments are relevant up to the second order: $m_0 = 1$, $m_1 = 0$, and $m_2 = A^2/2$. Furthermore, $f'(x) = 2x$ and $f''(x) = 2$ and $f^{(k)}(x) = 0$ for $k > 2$. Thus by Eq. (30)

$$\tilde{f}(x) = m_0 f(x) + m_1 f'(x) + m_2 \frac{f''(x)}{2} = x^2 + \frac{A^2}{2}. \tag{50}$$

Example 2. Calculate the average of $f(x+g(t))$ with

$$f(x) = \begin{cases} x & \text{for } |x| < 1 \\ 0 & \text{otherwise.} \end{cases} \qquad \text{and} \qquad g(t) = A\sin(\omega t) \quad \text{with } A < 1. \tag{51}$$

Using Theorem 3 we have $m = 3$,

$$x_0 = -\infty, \qquad x_1 = -1, \qquad x_2 = 1, \qquad x_3 = \infty, \tag{52}$$

and

$$f_1(y) = 0, \qquad f_2(y) = y, \qquad f_2'(y) = 1, \qquad f_3(y) = 0 \qquad \text{with} \quad D(y) = \mathbb{R} \quad \text{for } y \in \mathbb{R}, \tag{53}$$

hence, $D_\rho = \mathbb{R}$. The active set is

$$\mathcal{I}_A(y) = \begin{cases} \{1\} & \text{for } y < -1 - A, \\ \{1,2\} & \text{for } -1 - A < y < -1 + A, \\ \{2\} & \text{for } -1 + A < y < 1 - A, \\ \{2,3\} & \text{for } 1 - A < y < 1 + A, \\ \{3\} & \text{for } 1 + A < y, \end{cases} \tag{54}$$

and by Eq. (42),

$$\tilde{f}(y) = y\left(m_0(1-y) - m_0(-1-y)\right) + m_1(1-y) - m_1(-1-y), \tag{55}$$

and using Eq. (A.13) we find

$$\tilde{f}(y) = \begin{cases} 0 & \text{for } y \leq -1 - A, \\ y/2 + \pi^{-1}\left[\sqrt{A^2 - (1+y)^2} + y \arcsin\left((1+y)/A\right)\right] & \text{for } -1 - A < y \leq -1 + A, \\ y & \text{for } -1 + A < y \leq 1 - A, \\ y/2 - \pi^{-1}\left[\sqrt{A^2 - (1-y)^2} - y \arcsin\left((1-y)/A\right)\right] & \text{for } 1 - A < y \leq 1 + A, \\ 0 & \text{for } 1 + A < y. \end{cases} \tag{56}$$

For an example with $A = 0.5$, see Fig. 15.

3 | OBTAINING THE CLASSICAL PROBABILITY DENSITY

An analytic expression of the CPD of the motion $x_F(t) \equiv g(t)$ is rarely available. However, in some simple cases, it can be derived[30,32]. In this section, we recapitulate some of the most important CPDs that rely on a particle's undamped motion in a potential and provide a novel example for calculating a more complex motion consisting of the sum of two incommensurable harmonics. Following this, we also provide some ideas on determining the CPD numerically.

3.1 | CPD of undamped, free oscillations

The PDF/CPD of a function in the form as given in Eq. (2) originates from probability theory and is not related to particle motion[34]. However, an equivalent, physical definition can also be given[30]. Consider that the particle performs a unidirectional motion $x(t)$ from $x(a) = x_a$ to $x(b) = x_b$ with $a < b$. The particle spends $\mathrm{d}t$ amount of time in an infinitesimally small region of this interval $\mathrm{d}x$, where

$$\mathrm{d}t = \frac{\mathrm{d}x}{\mathrm{d}x/\mathrm{d}t} = \frac{\mathrm{d}x}{v(x)}, \tag{57}$$

being inversely proportional to the particle's velocity. The probability $\rho(x)$ of finding the particle in this infinitesimal region is the ratio of the time spent here to the total amount of time $b - a$ needed from x_a to x_b, that is

$$\rho(x)\mathrm{d}x \equiv \text{Probability}[(x, x + \mathrm{d}x)] = \frac{\mathrm{d}t}{b-a} = \frac{1}{b-a}\frac{\mathrm{d}x}{v(x)}, \tag{58}$$

hence,

$$\rho(x) = \frac{1}{b-a}\frac{1}{v(x)}. \tag{59}$$

Often the time and not the displacement dependency of the velocity is known, i.e., $v(t) = \mathrm{d}x/\mathrm{d}t \equiv g'(t)$, so we have to express t in terms of x and rewrite Eq. (59) as

$$\rho(x) = \frac{1}{b-a}\frac{1}{g'(g^{-1}(x))} \quad \text{for } x \in (x_a, x_b). \tag{60}$$

Eq. (60) is useful when the velocity is known as a function of the time, whereas Eq. (59) is preferred when the velocity is known as a function of the displacement. This latter one is when an undamped particle of unit mass oscillates freely in a potential well between turning points x_a and x_b with the time period τ. Based on the conservation of energy, we have

$$E = T + V = \frac{1}{2}v^2 + V(x), \tag{61}$$

thus insertion of

$$v(x) = \sqrt{2(E - V(x))} \tag{62}$$

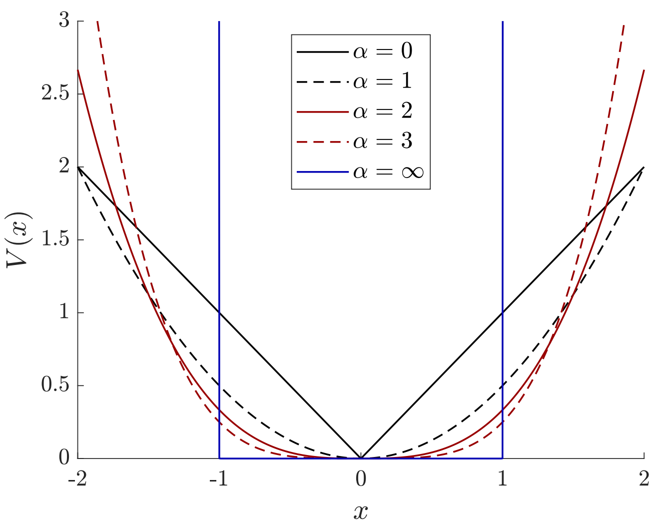
into Eq. (59) yields

$$\rho(x) = \frac{2}{\tau}\sqrt{\frac{1}{2(E - V(x))}}. \tag{63}$$

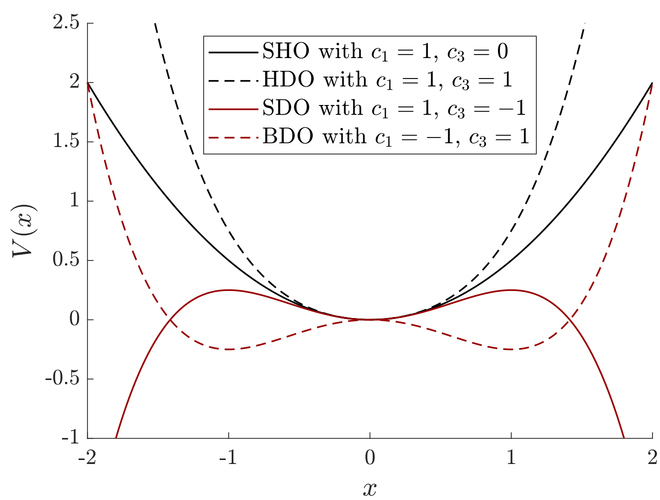
The factor $2/\tau$, including the time period of the oscillation, normalizes the area of $\rho(x)$ to one. τ is given by the integral[1]

$$\tau = 2\int_{x_a}^{x_b} \frac{1}{\sqrt{2(E - V(x))}}\mathrm{d}x. \tag{64}$$

In the following, we give some examples of potentials usually found in applications. Fig. 3 shows examples of the so-called purely nonlinear oscillators (PNOs) and for the Duffing type oscillators (DTOs).



(a) Potentials of PNOs $V(x) = |x|^{\alpha+1}/(\alpha+1)$ for different values of the parameter α



(b) Potentials of DTOs $V(x) = c_1\frac{x^2}{2} + c_3\frac{x^4}{4}$ for different values of the parameters c_1 and c_3

Figure 3 Common potentials. The corresponding probability density functions are calculated in the next sections.

3.1.1 | Purely Nonlinear Oscillators

PNOs have a potential that, after non-dimensionalization, can be written as

$$V(x) = \frac{1}{\alpha+1}|x|^{\alpha+1}, \tag{65}$$

where α is a positive real number. The particle has its turning points x_a and x_b at $\pm A$ that depends on the particle's initial energy E_0:

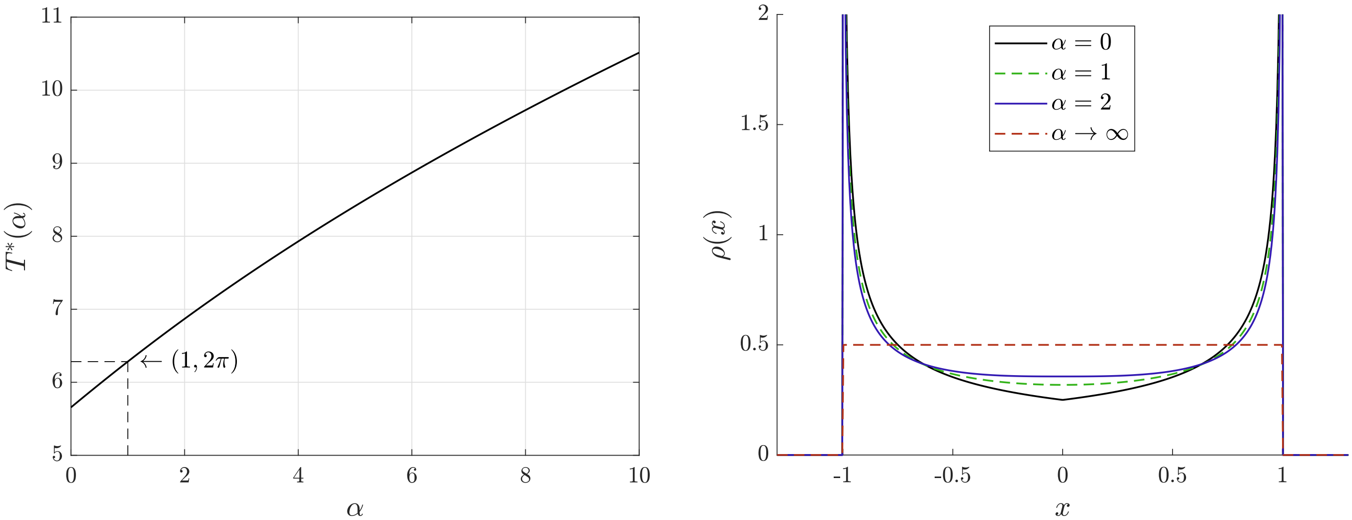
$$A = \left(E_0(\alpha+1)\right)^{1/(\alpha+1)}. \tag{66}$$

To determine the CPD, the only challenging task is to calculate the time period's value; the rest is readily given by Eq. (63). Fortunately, the problem was solved by many authors in the past[36–39], and the time period of PNOs is known to be

$$\tau = \underbrace{\sqrt{\frac{8\pi}{\alpha+1}}\frac{\Gamma\left(\frac{1}{\alpha+1}\right)}{\Gamma\left(\frac{1}{2}+\frac{1}{\alpha+1}\right)}}_{T^*(\alpha):=} A^{(1-\alpha)/2} = T^*(\alpha)A^{(1-\alpha)/2}, \tag{67}$$

where Γ denotes the gamma function. $T^*(\alpha)$, a factor only depending on α, is depicted in Fig. (4a). Thus, the CPD is given by

$$\rho_{\text{PNO}}(x) = \frac{2}{T^*(\alpha)A^{(1-\alpha)/2}} \frac{1}{\sqrt{\frac{2}{\alpha+1}\left(A^{\alpha+1}-x^{\alpha+1}\right)}}\mathbf{1}_{|x|<A}(x) = \frac{\alpha+1}{2\sqrt{\pi}} \frac{\Gamma\left(\frac{1}{2}+\frac{1}{\alpha+1}\right)}{\Gamma\left(\frac{1}{\alpha+1}\right)} \frac{A^{(\alpha-1)/2}}{\sqrt{A^{\alpha+1}-|x|^{\alpha+1}}}\mathbf{1}_{|x|<A}(x). \tag{68}$$



(a) Amplitude independent factor of a PNO's time period $T^*(\alpha)$[39]

(b) CPD of PNOs with different values of α with initial displacement $A = 1$

Figure 4 Purely nonlinear oscillators

Important values of α include 1 and the limiting cases $\alpha \to 0^+$ and $\alpha \to \infty$, which correspond to the cases of a simple (linear) harmonic oscillator (SHO), a constant restoring force oscillator and the so-called "infinite well"[30] oscillator, respectively. Further interesting cases are $\alpha = 2, 3, \dots$, which might arise if, during a system's linearization around an equilibrium position, the linear restoring term vanishes, but the remaining force can be approximated well by only using one further purely nonlinear term.

After lengthy calculations with the help of the software ®Mathematica, the moments of ρ_{PNO} can be obtained. It turns out that the integral can be represented in terms of the hypergeometric function, i.e.,

$$m_{\text{PNO},k}(x) = \int_{-\infty}^{x} y^k \rho_{\text{PNO}}(y)\mathrm{d}y = C(\alpha, A)\int_{-\infty}^{x} \frac{y^k}{\sqrt{A^{\alpha+1}-|y|^{\alpha+1}}}\mathrm{d}y = C(\alpha, A)\frac{y^{k+1}{}_2F_1\left(\frac{1}{2}, \frac{k+1}{\alpha+1}; \frac{k+1}{\alpha+1}+1; \frac{|y|^{\alpha+1}}{A^{\alpha+1}}\right)}{(k+1)A^{\frac{\alpha+1}{2}}}\Bigg|_{-A}^{x}, \tag{69}$$

where ${}_2F_1$ is defined as

$${}_2F_1(a, b; c; z) := \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 \frac{u^{b-1}(1-u)^{c-b-1}}{(1-uz)^a}\mathrm{d}u, \tag{70}$$

and based on Eq. (68) the constant $C(\alpha, A)$ is given by

$$C(\alpha, A) := \frac{\alpha+1}{2\sqrt{\pi}} \frac{\Gamma\left(\frac{1}{2}+\frac{1}{\alpha+1}\right)}{\Gamma\left(\frac{1}{\alpha+1}\right)} A^{\frac{\alpha-1}{2}}. \tag{71}$$

After the simplification of Eq. (69), the k^{th} partial moments of the purely nonlinear oscillator's motion are given by

$$m_{\text{PNO},k}(x) = \frac{\alpha+1}{2\sqrt{\pi}A(k+1)} \frac{\Gamma\left(\frac{1}{2}+\frac{1}{\alpha+1}\right)}{\Gamma\left(\frac{1}{\alpha+1}\right)} \left(x^{k+1}{}_2F_1\left(\frac{1}{2}, \frac{k+1}{\alpha+1}; \frac{k+1}{\alpha+1}+1; \left(\frac{|x|}{A}\right)^{\alpha+1}\right) + (-1)^k A^{k+1}\sqrt{\pi}\frac{\Gamma\left(\frac{k+1}{\alpha+1}+1\right)}{\Gamma\left(\frac{k+1}{\alpha+1}+\frac{1}{2}\right)}\right). \tag{72}$$

Evaluation of Eq. (72) at $x = A$ yields the k^{th} moment of the PNO's motion:

$$m_{\text{PNO},k} = \begin{cases} 0 & \text{for } k \text{ odd}, \\ \frac{\alpha+1}{k+1} \frac{\Gamma\left(\frac{1}{2}+\frac{1}{\alpha+1}\right)\Gamma\left(\frac{k+1}{\alpha+1}+1\right)}{\Gamma\left(\frac{1}{\alpha+1}\right)\Gamma\left(\frac{k+1}{\alpha+1}+\frac{1}{2}\right)} A^k & \text{for } k \text{ even}. \end{cases} \tag{73}$$

3.1.2 | Duffing Type oscillators

Undamped Duffing type oscillators [39] are given in the form

$$\ddot{x} + c_1 x + c_3 x^3 = 0, \tag{74}$$

with real coefficients c_1 and c_3. Based on the signs of these coefficients, three interesting cases are possible

- hardening Duffing oscillator (HDO) for $c_1 > 0$ and $c_3 > 0$,
- softening Duffing oscillator (SDO) for $c_1 > 0$ and $c_3 < 0$,
- bistable Duffing oscillator (BDO) for $c_1 < 0$ and $c_3 > 0$.

Based on [39], the time period of the HDO is given by

$$\tau_{\text{HDO}} = \frac{4K\left(\frac{c_3A^2}{2(c_1+c_3A^2)}\right)}{\sqrt{c_1+c_3A^2}}, \tag{75}$$

where K denotes the complete elliptic integral of the first kind with the elliptic parameter m. Using Eq. (63) and (75) the CPD of the HDO is obtained as

$$\rho_{\text{HDO}}(x) = \frac{\sqrt{c_1+c_3A^2}}{2K\left(\frac{c_3A^2}{2(c_1+c_3A^2)}\right)} \frac{1}{\sqrt{c_1A^2+\frac{c_3}{2}A^4-c_1x^2-\frac{c_3}{2}x^4}}, \tag{76}$$

where $\pm A$ are the turning points of the oscillating particle.

Similarly, for $|A| < \sqrt{c_1/|c_3|}$ the time period of the SDO can be given by

$$\tau_{\text{SDO}} = \frac{4K\left(\frac{c_3A^2}{2(c_1-|c_3|A^2)}\right)}{\sqrt{c_1-|c_3|A^2}}, \tag{77}$$

hence, the CPD is

$$\rho_{\text{SDO}}(x) = \frac{\sqrt{c_1-|c_3|A^2}}{2K\left(\frac{c_3A^2}{2(c_1-|c_3|A^2)}\right)} \frac{1}{\sqrt{c_1A^2-\frac{|c_3|}{2}A^4-c_1x^2+\frac{|c_3|}{2}x^4}}. \tag{78}$$

The BDO has two subcases, depending on the particle's energy and thus on its turning points x_a and x_b. For

$$x_a < -\sqrt{\frac{2|c_1|}{c_3}} < \sqrt{\frac{2|c_1|}{c_3}} < x_b, \tag{79}$$

the particle passes through both potential wells, called "full-swing" or "out-of-well mode." We denote this case with BDO_1. However, for

$$-\sqrt{\frac{2|c_1|}{c_3}} < x_a < x_b < 0 \qquad \text{or} \qquad 0 < x_a < x_b < \sqrt{\frac{2|c_1|}{c_3}}, \tag{80}$$

the particle oscillates only in one of the potential wells, which is also called the "half-swing" or "in-well mode," and we denote it by BDO_2 [39]. When started with zero velocity and initial displacement $x_0 = A$, the particle has the time periods

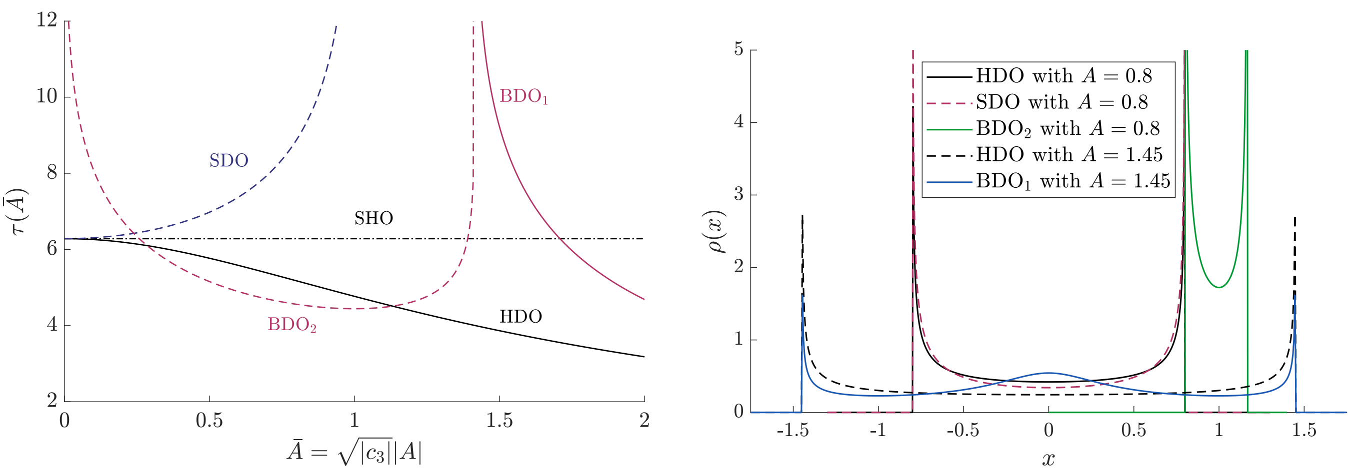
$$\tau_{\mathrm{BDO}_1} = \frac{4K\left(\frac{c_3A^2}{2(c_3A^2-|c_1|)}\right)}{\sqrt{c_3A^2-|c_1|}}, \qquad \tau_{\mathrm{BDO}_2} = \frac{2K\left(\frac{2(c_3A^2-|c_1|)}{c_3A^2}\right)}{\sqrt{\frac{c_3}{2}}A}, \tag{81}$$

respectively. Then, the CPDs become

$$\rho_{\mathrm{BDO}_1}(x) = \frac{\sqrt{c_3A^2-|c_1|}}{2K\left(\frac{c_3A^2}{2(c_3A^2-|c_1|)}\right)} \frac{1}{\sqrt{c_1A^2+\frac{c_3}{2}A^4-c_1x^2-\frac{c_3}{2}x^4}}, \tag{82}$$

$$\rho_{\mathrm{BDO}_2}(x) = \frac{\sqrt{2c_3}A}{K\left(\frac{2(c_3A^2-|c_1|)}{c_3A^2}\right)} \frac{1}{\sqrt{c_1A^2+\frac{c_3}{2}A^4-c_1x^2-\frac{c_3}{2}x^4}}, \tag{83}$$

respectively. In all four cases, the time periods are functions of c_1 and c_3A^2. Through non-dimensionalization, $|c_1| = 1$ can be achieved; thus, the time period becomes a univariate function. Its values are depicted in Fig. 5a.



(a) Time periods of different types of non-dimensionalized ($|c_1| = 1$) undamped DTOs depicted against the parameter $\bar{A} = \sqrt{|c_3|}|A|$ [39]

(b) CPDs of the three different types of DTOs with different initial displacement $A = 0.8$ and $A = 1.45$ for $|c_1| = |c_3| = 1$

Figure 5 Duffing type oscillators

The partial moments of the CPDs defined in Eqs. (76), (78), (82), (83) can be obtained by calculating

$$m_{\mathrm{DUFF},k}(x) = \int_{-\infty}^{x} y^k \rho_{\mathrm{DUFF}}(y)\mathrm{d}y, \tag{84}$$

where ρ_{DUFF} denotes any of the cases HDO, SDO, BDO_1 and BDO_2. After lengthy calculations with the help of the software ®Mathematica, it turns out that the integral is solvable for $V(x) < E_0 := c_1 A^2/2 + c_3 A^4/4$ in terms of the Appell hypergeometric function, i.e.

$$m_{\mathrm{DUFF},k}(x) = C(c_1, c_3, A) \int\limits_{x_a(c_1,c_3,E_0)}^{x} \frac{y^k}{\sqrt{2E_0 - c_1 y^2 - c_3 \frac{y^4}{2}}} \mathrm{d}y \tag{85}$$

$$= C(c_1, c_3, A) \left. \frac{y^{k+1} F_1\left(\frac{k+1}{2}; \frac{1}{2}, \frac{1}{2}; \frac{k+3}{2}; \frac{c_3 y^2}{\sqrt{c_1^2+4c_3E_0}-c_1}, -\frac{c_3 y^2}{\sqrt{c_1^2+4c_3E_0}+c_1}\right)}{\sqrt{2E_0}(k+1)} \right|_{x_a(c_1,c_3,E_0)}^{x} \quad \text{for } x_a(c_1,c_3,E_0) < x < x_b(c_1,c_3,E_0), \tag{86}$$

with the constant $C(c_1, c_3, A)$ given by Eqs. (76,78,82) and (83), respectively, and with the lower turning point $x_a(c_1, c_3, E_0)$ and the upper turning point $x_b(c_1, c_3, E_0)$. The Appell hypergeometric function is defined as

$$F_1(\alpha; \beta, \beta'; \gamma; x, y) = \frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\gamma-\alpha)} \int\limits_0^1 u^{\alpha-1}(1-u)^{\gamma-\alpha-1}(1-ux)^{-\beta}(1-uy)^{-\beta'} \mathrm{d}u. \tag{87}$$

The turning points of the HDO and BDO_1 are

$$x_{a,\mathrm{HDO/BDO}_1} = -A = -\sqrt{\sqrt{\frac{c_1^2}{c_3^2} + \frac{4E_0}{c_3}} - \frac{c_1}{c_3}}, \quad \text{and} \quad x_{b,\mathrm{HDO/BDO}_1} = A = \sqrt{\sqrt{\frac{c_1^2}{c_3^2} + \frac{4E_0}{c_3}} - \frac{c_1}{c_3}}. \tag{88}$$

The turning point of SDO is

$$x_{a,\mathrm{SDO}} = -A = -\sqrt{-\frac{c_1}{c_3} - \sqrt{\frac{c_1^2}{c_3^2} + \frac{4E_0}{c_3}}}, \quad \text{and} \quad x_{b,\mathrm{SDO}} = A = \sqrt{-\frac{c_1}{c_3} - \sqrt{\frac{c_1^2}{c_3^2} + \frac{4E_0}{c_3}}}, \tag{89}$$

while the turning points of BDO_2 are

$$x_{a,\mathrm{BDO}_2} = \sqrt{-\frac{c_1}{c_3} - \sqrt{\frac{c_1^2}{c_3^2} + \frac{4E_0}{c_3}}}, \quad \text{and} \quad x_{b,\mathrm{BDO}_2} = \sqrt{-\frac{c_1}{c_3} + \sqrt{\frac{c_1^2}{c_3^2} + \frac{4E_0}{c_3}}}. \tag{90}$$

Since $\rho_{\mathrm{HDO}}(x)$ is even, its odd moments are all 0. Its even moments are

$$m_{\mathrm{HDO},k} = \frac{\sqrt{c_1 + c_3 A^2}}{K\left(\frac{c_3 A^2}{2(c_1+c_3A^2)}\right)} \frac{A^{k+1} F_1\left(\frac{k+1}{2}; \frac{1}{2}, \frac{1}{2}; \frac{k+3}{2}; 1, \frac{c_1^2}{2E_0c_3}\left(\sqrt{1+\frac{4E_0c_3}{c_1^2}} - 1\right) - 1\right)}{\sqrt{2E_0}(k+1)} \tag{91}$$

$$= \frac{\sqrt{c_1 + c_3 A^2}}{K\left(\frac{c_3 A^2}{2(c_1+c_3A^2)}\right)} \frac{A^{k+1}\sqrt{\pi}\Gamma\left(\frac{k+3}{2}\right) {}_2F_1\left(\frac{1}{2}, \frac{k+1}{2}; \frac{k}{2}+1; \frac{c_1^2}{2E_0c_3}\left(\sqrt{1+\frac{4E_0c_3}{c_1^2}} - 1\right) - 1\right)}{\Gamma\left(\frac{k}{2}+1\right)\sqrt{2E_0}(k+1)}. \tag{92}$$

By the same argument, the odd moments of $\rho_{\mathrm{SDO}}(x)$ and $\rho_{\mathrm{BDO}_1}(x)$ are also all 0, and its even moments are given by

$$m_{\mathrm{SDO},k} = \frac{\sqrt{c_1 - |c_3|A^2}}{K\left(\frac{c_3 A^2}{2(c_1-|c_3|A^2)}\right)} \frac{A^{k+1}\sqrt{\pi}\Gamma\left(\frac{k+3}{2}\right) {}_2F_1\left(\frac{1}{2}, \frac{k+1}{2}; \frac{k}{2}+1; \frac{c_1^2}{2E_0|c_3|}\left(1 - \sqrt{1-\frac{4E_0|c_3|}{c_1^2}}\right) - 1\right)}{\Gamma\left(\frac{k}{2}+1\right)\sqrt{2E_0}(k+1)}, \tag{93}$$

and

$$m_{\mathrm{BDO}_1,k} = \frac{\sqrt{c_3 A^2 - |c_1|}}{K\left(\frac{c_3 A^2}{2(c_3A^2-|c_1|)}\right)} \frac{A^{k+1}\sqrt{\pi}\Gamma\left(\frac{k+3}{2}\right) {}_2F_1\left(\frac{1}{2}, \frac{k+1}{2}; \frac{k}{2}+1; -\frac{c_1^2}{2E_0c_3}\left(\sqrt{1+\frac{4E_0c_3}{c_1^2}} + 1\right) - 1\right)}{\Gamma\left(\frac{k}{2}+1\right)\sqrt{2E_0}(k+1)}. \tag{94}$$

$\rho_{\mathrm{BDO}_2}(x)$ possesses exceptional characteristics as it exhibits asymmetry. The complexity of the distribution prevents us from presenting its moments in a closed form. Instead, we leave this task to the reader as an exercise.

Example 3. We are interested in the escape of two particles coupled by a strong linear spring of stiffness $k \gg 1$ in a quadratic-quartic potential well $V(x) = x^2/2 - x^4/4$. The equations of motion are given by

$$\ddot{x}_1 + V'(x_1) + k(x_1 - x_2) = 0, \tag{95}$$

$$\ddot{x}_2 + V'(x_2) + k(x_2 - x_1) = 0, \tag{96}$$

$$x_1(0) = x_2(0) = 0, \tag{97}$$

$$\dot{x}_1(0) = -v_0, \tag{98}$$

$$\dot{x}_2(0) = v_0. \tag{99}$$

Introducing the new variables center of mass and relative displacement

$$y_1 = \frac{x_1 + x_2}{2}, \qquad \text{and} \qquad y_2 = x_2 - x_1, \tag{100}$$

we obtain

$$\ddot{y}_1 + \frac{V'\left(y_1 - \frac{y_2}{2}\right) + V'\left(y_1 + \frac{y_2}{2}\right)}{2} = 0, \tag{101}$$

$$\ddot{y}_2 + \left(2k + 1 - 3y_1^2\right) y_2 - \frac{y_2^3}{4} = 0, \tag{102}$$

$$y_1(0) = \dot{y}_1(0) = 0, \tag{103}$$

$$y_2(0) = 0, \tag{104}$$

$$\dot{y}_2(0) = 2v_0. \tag{105}$$

Since $V'(x)$ is even, $y_1(t) = 0$ is a solution; however, its stability needs to be clarified and will depend on the values of v_0 and k. By inserting $y_1(t)$ in Eq. (102), the problem is reduced to an SDO with parameters $c_1 = 2k + 1$ and $c_3 = -1/4$. The total energy is given by

$$E_0 = \frac{1}{2}\dot{y}_2^2(0) = 2v_0^2, \tag{106}$$

which determines the amplitude of the vibrations

$$A = 2\sqrt{2k+1}\sqrt{1 - \sqrt{1 - \frac{2v_0^2}{(2k+1)^2}}}. \tag{107}$$

Since $k \gg 1$, the vibrations in y_2 are fast, and we can average Eq. (101). By Theorem 1 it is not necessary to exactly determine $y_2(t)$; its CPD suffices. Theorem 2 shows that only the first three moments will play a role in the averaging since $V'(x)$ is a polynomial of degree three. Since $y_2(t)$ is symmetric, the odd moments are zero, and the only moment left (besides the trivial zeroth one) is the second one, given by Eq. (93). Since $c_1 \gg |c_3|$, the motion does not differ much from a harmonic motion having the second moment given by Eq. (A.8)

$$m_{y_2,2} = \frac{A^2}{2}. \tag{108}$$

After rescaling the moment due to the factor $1/2$, insertion into Eq. (108) yields

$$\left\langle V'\left(y_1 - \frac{y_2(t)}{2}\right)\right\rangle = \left\langle V'\left(y_1 + \frac{y_2(t)}{2}\right)\right\rangle = V'(y_1) + \frac{1}{8}V'''(y_1)m_{y_2,2} = \left(1 - \frac{3}{8}A^2\right) y_1 - y_1^3, \tag{109}$$

resulting in the averaged differential equation

$$\ddot{y}_1 + \left(1 - \frac{3}{8}A^2\right) y_1 - y_1^3 = 0. \tag{110}$$

Linear stability analysis yields the stability condition

$$\frac{8}{3} \overset{!}{>} A^2, \tag{111}$$

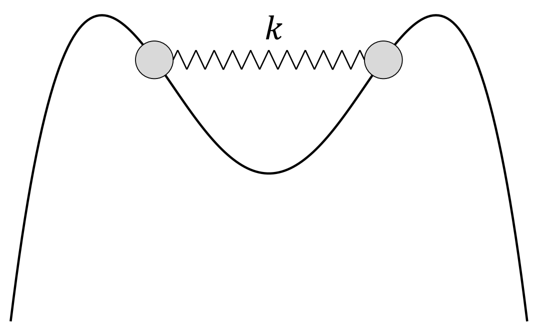


Figure 6 Problem setting – a pair of coupled particles in a quadratic-quartic potential well

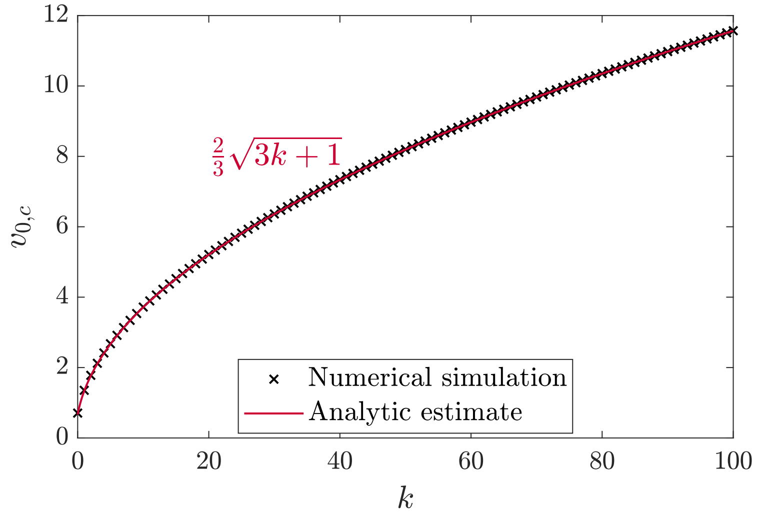


Figure 7 The critical value of the initial velocity $v_{0,\mathrm{c}}$ depicted against the stiffness of the linear spring k

which is equivalent to

$$v_0 \overset{!}{<} v_{0,\mathrm{c}} := \frac{2}{3}\sqrt{3k+1}. \tag{112}$$

A comparison of the analytic estimate with direct numerical simulations is shown in Fig. 7. The numerical simulations were obtained by disturbing the initial conditions by setting $y_1(0) = 0.005$ and integrating the system up to 1000 time units. If the particle pair escapes, the solution $y_1(t) = 0$ is categorized as unstable.

3.2 | CPD of forced oscillations

When excitation is present, the energy conservation principle can no longer be applied in the simple form as before, and the use of Eq. (59) becomes less viable. Instead, using Eq. (60) is more appropriate. However, this approach has a drawback: it requires knowledge of the analytic solution of the forced oscillation, which is typically only available for a few, but important cases.

A specific case of nonlinear systems is called *partially strongly damped systems*[23]. These systems consist of "slow" master variables and strongly damped "slaves." In the standard form, such systems can be represented as

$$\dot{x} = \varepsilon X(x, y, t) \tag{113}$$

$$\dot{y} = K(x)y + \varepsilon Y(x, y, t) \tag{114}$$

$$x(0) = x_0, \quad y(0) = y_0, \tag{115}$$

with the assumption

$$\max\left\{ eigenvalue\left[\frac{1}{2}(K + K^\top)\right]\right\} = -1. \tag{116}$$

One can introduce the "slow" variables ξ and η and perform averaging that yields

$$\dot{\xi}_0 = \varepsilon\langle X(\xi_0, \eta_0, t)\rangle_t \tag{117}$$

$$\eta_0 = \exp(K(\xi_0)t)y_0, \quad \xi_0(0) = x_0, \tag{118}$$

$$||x - \xi_0|| = O(\varepsilon), 0 \le t \le O(\varepsilon^{-1}). \tag{119}$$

One implication is that their forced response determines the time evolution of the strongly damped variables. Often, the equations are such that the variable of interest can be described by the sum of a "slow" variable $x_S \equiv \xi_0$ and a "fast" one $x_F \equiv \eta_0$. In the following, we focus on this case assuming only one variable of the form $x = x_S + g(t)$; thus, Theorem 1 is applicable. We present two typical "fast" motions often encountered in practical applications.

3.2.1 | CPD of an SHO

Arguably, the most important analytically solvable case is the one of a harmonic response of amplitude A. Fortunately, both previous investigations on the purely nonlinear oscillators ($\alpha = 1$) and the Duffing type oscillators ($c_3 = 0$) include the above case. Thus, we have

$$\rho_{\mathrm{SHO}}(x) = \frac{1}{\pi}\frac{1}{\sqrt{A^2 - x^2}}\mathbf{1}_{(-A,A)}(x). \tag{120}$$

See the appendix for more information on the moments of ρ_{SHO}.

3.2.2 | CPD of the sum of two harmonic functions with incommensurable frequencies

Let us consider now the CPD of a particle performing the following motion

$$g(t) = \underbrace{A_1 \sin(\omega_1 t + \beta_1)}_{=:g_1(t)} + \underbrace{A_2 \sin(\omega_2 t + \beta_2)}_{=:g_2(x)}, \tag{121}$$

where $A_1, A_2 > 0$ and $\omega_1/\omega_2 \notin \mathbb{R}\backslash\mathbb{Q}$, i.e., ω_1 and ω_2 are incommensurable. This motion is obtained as the solution of a harmonically excited undamped harmonic oscillator, where ω_1 is the eigenfrequency of the oscillator and ω_2 is the excitation frequency. The same expression emerges as the particular solution of a damped harmonic oscillator under bi-harmonic excitation (BHO).

To obtain the CPD of $g(t)$, the analytic formula (60) cannot be used anymore since the inverse of the function cannot be given by a closed formula. In addition, the motion is not periodic, implying the assembly of infinitely many terms in Eq. (6). If one wants to sample $g(t)$ starting at $t = a$ and ending at $t = b$, the resulting CPD would depend on the values of a and b. However, if the length of the function sample, $b - a$ goes to ∞, so Eq. (6) converges to a particular limiting function $\rho_{\mathrm{BHO}}(x)$ which can be obtained by convolution of the CPDs/PDFs $\rho_1(x)$ and $\rho_2(x)$. This statement is true because $g(t)$ has the same PDF as the random variable $X := X_1 + X_2$, where X_1 and X_2 are independent random variables following the arcsine distribution centered at 0 with half-width A_1 and A_2. Based on[40], the PDF of a random variable consisting of the sum of two independent random variables can be calculated by the convolution of the PDFs of the summands. The following elliptic integral has to be solved to obtain an analytic value for the PDF of $g(t)$:

$$\rho_{\mathrm{BHO}}(x) = (\rho_1 * \rho_2)(x) = \int_{-\infty}^{\infty} \rho_1(\tau)\rho_2(x-\tau)\mathrm{d}\tau = \int_{-\infty}^{\infty} \frac{1}{\pi\sqrt{A_1^2 - \tau^2}}\mathbf{1}_{(-A_1,A_1)}(\tau)\frac{1}{\pi\sqrt{A_2^2 - (\tau - x)^2}}\mathbf{1}_{(-A_2+x,A_2+x)}(\tau)\mathrm{d}\tau, \tag{122}$$

Without loss of generality, we assume that $A_1 \geq A_2$, and we introduce the polynomial

$$G(\tau) := \left(A_1^2 - \tau^2\right)\left(A_2^2 - (\tau - x)^2\right) \tag{123}$$

with roots

$$\tau_1 = -A_1, \qquad \tau_2 = A_1, \qquad \tau_3 = x - A_2, \qquad \tau_4 = x + A_2. \tag{124}$$

Thus, depending on the value of x integral (122) can be reduced to

$$\rho_{\mathrm{BHO}}(x) = \begin{cases} 0 & x < -A_1 - A_2 \\ \int_{-A_1}^{x+A_2} G^{-1/2}(\tau)\mathrm{d}\tau & -A_1 - A_2 < x < -A_1 + A_2 \\ \int_{x-A_2}^{x+A_2} G^{-1/2}(\tau)\mathrm{d}\tau & -A_1 + A_2 < x < A_1 - A_2 \\ \int_{x-A_2}^{A_1} G^{-1/2}(\tau)\mathrm{d}\tau & A_1 - A_2 < x < A_1 + A_2 \\ 0 & A_1 + A_2 < x \end{cases} \tag{125}$$

The elliptic integral can be calculated by transforming Eq. (122) to Legendre's normal form[41] given as

$$\rho_{\mathrm{BHO}}(x; A_1, A_2) = C(x, A_1, A_2)\int_0^{\frac{\pi}{2}} \frac{\mathrm{d}\phi}{\sqrt{1 - m(x, A_1, A_2)\sin^2\phi}}, \tag{126}$$

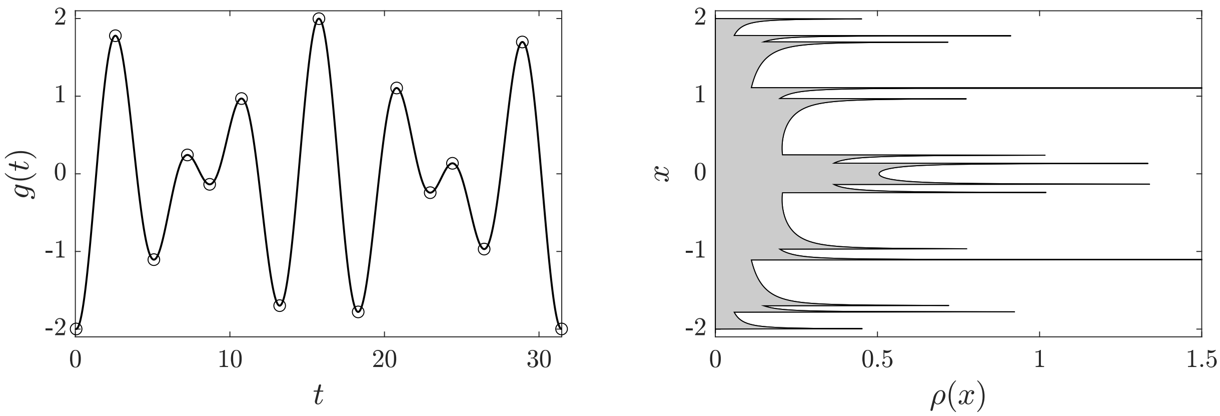


Figure 8 Periodic, bi-harmonic motion given by $g(t) = -\cos(t) - \cos(1.4t - 0.1)$ and its numerically obtained CPD

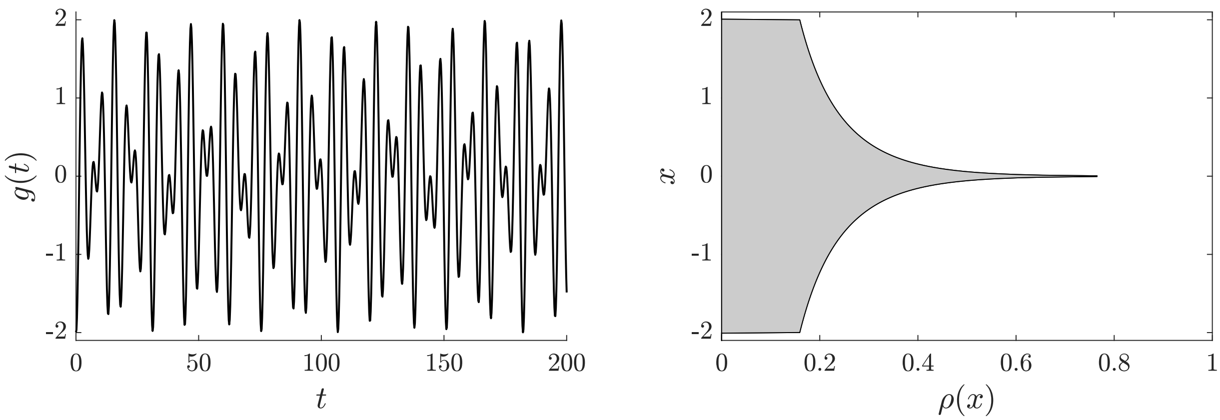


Figure 9 Aperiodic, bi-harmonic motion given by $g(t) = -\cos(t) - \cos(\sqrt{2}t - 0.1)$ and its analytically obtained CPD

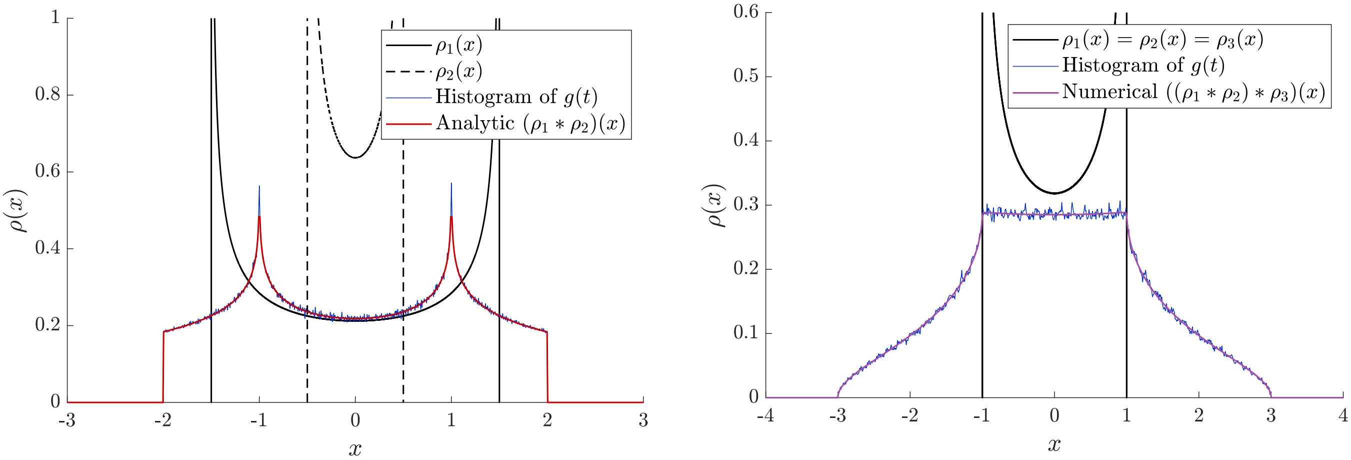
where C and the parameter m are functions of the independent variable x and the parameters A_1 and A_2. To obtain the normal form in Eq. (126), the following rational transformation has to be applied

$$\tau = \frac{a_3(a_2 - a_4) - a_4(a_2 - a_3)\sin^2\phi}{(a_2 - a_4) - (a_2 - a_3)\sin^2\phi}, \tag{127}$$

where $a_1 > a_2 > a_3 > a_4$ are the roots of $G(\tau)$, i.e., the same values as $\tau_1, \dots, \tau_4$, but in descending order. Substituting Eq. (127) in Eq. (125) we obtain

$$\rho_{\mathrm{BHO}}(x) = \begin{cases} 0 & \text{for } x < -A_1 - A_2, \\ \frac{1}{\pi^2\sqrt{A_1A_2}}\boldsymbol{K}\left(\frac{(A_1+A_2)^2-x^2}{4A_1A_2}\right) & \text{for } -A_1 - A_2 < x < -A_1 + A_2, \\ \frac{2}{\pi^2\sqrt{(A_1+A_2)^2-x^2}}\boldsymbol{K}\left(\frac{4A_1A_2}{(A_1+A_2)^2-x^2}\right) & \text{for } -A_1 + A_2 < x < A_1 - A_2, \\ \frac{1}{\pi^2\sqrt{A_1A_2}}\boldsymbol{K}\left(\frac{(A_1+A_2)^2-x^2}{4A_1A_2}\right) & \text{for } A_1 - A_2 < x < A_1 + A_2, \\ 0 & \text{for } A_1 + A_2 < x, \end{cases} \tag{128}$$

where $K(m)$ is the complete elliptic integral of the first kind with modulus m. Figs. 8-9 give examples for the CPD of a periodic bi-harmonic motion with $\omega_2/\omega_1 = 1.4$ and an aperiodic one with $\omega_2/\omega_1 = \sqrt{2}$.



(a) CPD $\rho(x)$ of the quasi-periodic motion $g(t) = 1.5\sin(t) + 0.5\sin(\sqrt{2}t)$. The histogram-based numerical approximation (blue) is compared to the exact analytic solution (red). $\rho_1(x)$ and $\rho_2(x)$ represent the CPDs of the individual harmonics in $g(t)$

(b) CPD $\rho(x)$ of $g(t) = \sin(t)+\sin(\sqrt{2}t)+\sin(\sqrt{3}t)$. The black line represents the CPD of the individual harmonic terms, the arcsine distribution with half-width 1. The convolution-based numerical estimate is given in purple, while the histogram-based estimate is given in black

Figure 10 CPD of poly-harmonic functions. Two different numerical approaches are shown

If $g(t)$ is periodic, the frequency ratio can be written as $\omega_2/\omega_1 = a/b$ with $a, b \in \mathbb{N}$ being relative primes to each other. Thus, the time period is given by $\tau = 2\pi a/\omega_1$. The PDF of $g(t)$ depends strongly on the values of A_1, A_2, a, and b and the initial phase difference of the two harmonics $\beta_2 - \beta_1$. Since the sample necessary to describe the PDF has a finite length, at every turning point of $g(t)$ the PDF ρ has a singularity (cf. Fig. 8). Contrary to this, for an aperiodic function, the sample has to be infinitely long, resulting in the limiting case with only one ($A_1 = A_2$) or two peaks ($A_1 \neq A_2$).

4 | NUMERICAL METHODS

If $g(t)$ is more complicated than in the above examples, the usage of numerical methods might become necessary since Eq. (6) cannot be solved analytically anymore. The numerical approximation of the PDF is also necessary if $g(t)$ is not known explicitly but is only given as a numerical solution of an ODE. In the following, two general and one specific method are described highlighting their advantages and disadvantages.

4.1 | Histogram-based approximation of the CPD

Arguably the simplest numerical method to obtain $\rho(x)$ from $g(t)$ that instinctively comes into mind is based on sampling, i.e., on the construction of a histogram. $g(t)$ is evaluated at a large number of equidistantly placed values, and the obtained data is plotted in a histogram (cf. Fig. 10)[30]. Subsequently, linear/cubic splines can be fitted on the data in order to be able to evaluate the result at arbitrary values.

The histogram converges to the PDF[40] as the number of evaluation points goes to infinity. The advantage of the method is that it can be applied flexibly on long or short intervals, and the possibility of evaluating $g(t)$ (without explicitly having a formula for it) is sufficient. The convergence to the CPD depends mainly on the number of function evaluations.

The main disadvantage of the method is its relatively slow convergence.

4.2 | Spline-based approximation of the CPD

Another method that instinctively comes to mind once Eqs. (2) and (6) are known, is the usage of splines (cf. Fig. 8). Since we can only handle monotonically increasing or decreasing intervals, $g(t)$ has to be truncated into pieces $g_i(t)$ such that all are either monotonically increasing or decreasing. To achieve this, one can split $g(t)$ at its local extrema. Then, on each piece, $g_i(t)$ is evaluated at a sufficiently large number of points $x_{N_i} \in \mathbb{R}^{N_i}$, which not necessarily have to be distributed equidistantly. After that, natural cubic splines are fitted on the data $(x_{N_i}, g_i(x_{N_i}))$.

Further on, the derivatives and the inverses of these splines are needed. A straightforward way to get the derivative is to calculate it piecewise from the polynomial pieces of the splines. Fitting splines on the data $(g_i(x_{N_i}), x_{N_i})$ is a direct method for obtaining the inverse. The spline fitting is performed easily using ®MATLAB's curve fitting tool. The fits are combined as prescribed by Eq. (2), and their weighted sum is calculated as given by Eq. (6), thus the result is a fit that can be evaluated at arbitrary values.

The advantage of the method is the more exact estimation of the CPD, even with relatively few spline nodes.

The disadvantage is that for non-periodic, highly-oscillating functions, the estimation of the CPD on long intervals can be very slow due to the large number of pieces created between local extrema of $g(t)$.

4.3 | Convolution-based approximation of the CPD for special cases

In special cases, when $g(t)$ is given by the sum of n independent functions $g_i(t)$ with $i = 1, \ldots, n$ with individual CPDs $\rho_i(x)$, the resulting CPD $\rho(x)$ might be obtained by consecutive $n-1$ times numerical convolution of all $\rho_i(x)$ (cf. Fig. (10b)[1]. In order to do so, the CPDs are evaluated at many equidistant points, and $n-1$ numerical convolutions are performed such that at the end, all $\rho_i(x)$ are contained in the result. Since convolution is associative, it does not matter in which order the CPDs are convoluted with each other. Subsequently, cubic splines are fitted on the data, and by numerical quadrature, the fit's area is normalized to one to compensate for numerical inaccuracies.

The spline-based method is advantageous for periodic $g(t)$ with the known time period, whereas the histogram-based method is better suited for non-periodic $g(t)$. In exceptional cases, the convolution-based method can obtain significantly more accurate results than the histogram-based method.

In any of the three presented methods, the final result is a fit that MATLAB can use just as any in-built function. However, keeping the number of data points within limits is essential since the evaluation time might increase unnecessarily.

4.4 | Numerical averaging

Once the CPD $\rho(x)$ is obtained in either analytic or numerical form, the cross-correlation described in Eq. (1) has to be calculated to obtain the average of $f(x_S + g(t))$. If analytically, the solution is not accessible by direct integration or Fourier transformation; a numerical one must be found.

Three basic approaches are given in the following:

1. Direct numerical quadrature of

$$\frac{1}{T}\int_{0}^{T} f(x_S + g(t))\mathrm{d}t \tag{129}$$

with N_x fixed values of x_S followed by subsequent spline fitting on the obtained data. The equidistant evaluation of $f(x_S + g(t))$ for many values of t ($t_i = 0, T/N_t, 2T/N_t, \ldots, (N_t - 1)T/N_t$) and calculation their mean is the midpoint rule Riemann sum with $\Delta x =$ const. It is well known that for sufficiently smooth (C^2) functions, this quadrature rule has a quadratic convergence rate, thus the mean average error (MAE), defined in Eq. (137) is of $O(N_t^{-2})$ and is independent of the spatial resolution N_x (cf. Fig. 12b). The computational cost is $O(N_t N_x)$. In Fig. (12a), sublinear dependency on N_t and N_x can be observed based on a numerical experiment due to MATLAB's algorithm, which evaluates vector data structures sublinearly proportional to their size.

The direct numerical quadrature approach is problematic if $T \to \infty$, and it might be generally slow since new function evaluations are needed for every new value of x_S.

However, this is the only possible way if the function to average is given only as $f(x_S, x_F, t)$.

[1]By independent, we mean that $g(t)$ has the same CPD as the sum of the independent random variables X_i described by their PDFs $\rho_i(x)$.

2. Numerical computation of $(\rho \star f)(x_S)$. Much faster than the previous method since $f(x)$ does not have to be evaluated repeatedly. Instead, $\rho(x)$ has to be obtained. Once $f(x)$ and $\rho(x)$ are discretized at N_x equidistant grid points, their numerical cross-correlation can be performed. The numerical convolution of two vectors of length n_x is an operation of cost $O(N_x^2)$ when done by its definition.

3. Numerical cross-correlation, taking advantage of the properties of the Fourier transform, can be evaluated using the FFT with computational cost $O(N_x \log N_x)$, which is the most significant advantage of this method (cf. Fig. 13a). Indeed, MATLAB's inbuilt cross-correlation function `xcorr` itself uses FFT. For sufficiently smooth $f(x)$ and bounded $\rho(x)$, the MAE of cross-correlation-based numerical averaging is inversely proportional to the discretization step length since the numerical quadrature, in this case, corresponds to a Riemann sum with the left rule, i.e., MAE$\sim N_x^{-1}$ (cf. Fig. 13). If f or ρ are less regular, the convergence rate might be worse than linear (cf. Fig. 14).

4.5 | Benchmark example

The following will demonstrate these properties using two benchmark examples with known analytic solutions. In the first case, let the "fast" motion be given by

$$g(t) = \frac{2A}{\pi} \arcsin \sin \left(\frac{\pi}{2}\omega t\right), \tag{130}$$

which is a triangle wave taking values between $-A$ and A and time period of $\tau = 4/\omega$, we assume $A < \pi$. It might be interpreted as the motion of a massless particle with energy $E_0 = A^2\omega^2$ in an "infinite well," i.e., a PNO with $\alpha \to \infty$ of width $2A$.

The target function $f(x)$ to be averaged will be a truncated sine force field given by

$$f(x) = \begin{cases} \sin x & \text{for } |x| < \pi, \\ 0 & \text{otherwise.} \end{cases} \tag{131}$$

The problem to be solved by averaging is

$$\tilde{f}(x) = \frac{\omega}{4} \int\limits_0^{\frac{4}{\omega}} \sin\left(x + \frac{2A}{\pi}\arcsin\sin\left(\frac{\pi}{2}\omega t\right)\right) \cdot \mathbf{1}_{(-\pi,\pi)}\left(x + \frac{2A}{\pi}\arcsin\sin\left(\frac{\pi}{2}\omega t\right)\right) \mathrm{d}t. \tag{132}$$

The calculation of Eq. (132) in this form is not trivial. However, the average can be easily obtained using Theorem 1. The CPD is given by

$$\rho(x) = \begin{cases} \frac{1}{2A} & \text{for } |x| < A, \\ 0 & \text{otherwise,} \end{cases} \tag{133}$$

and so the average is determined by

$$\tilde{f}(x_S) = \int\limits_{-\infty}^{\infty} g(x)\rho(x - x_S)\mathrm{d}x = \frac{1}{2A} \int\limits_{-A+x_S}^{A+x_S} \sin x \cdot \mathbf{1}_{(-\pi,\pi)}(x)\mathrm{d}x. \tag{134}$$

The result will be piecewise defined; thus, we introduce the boundary points

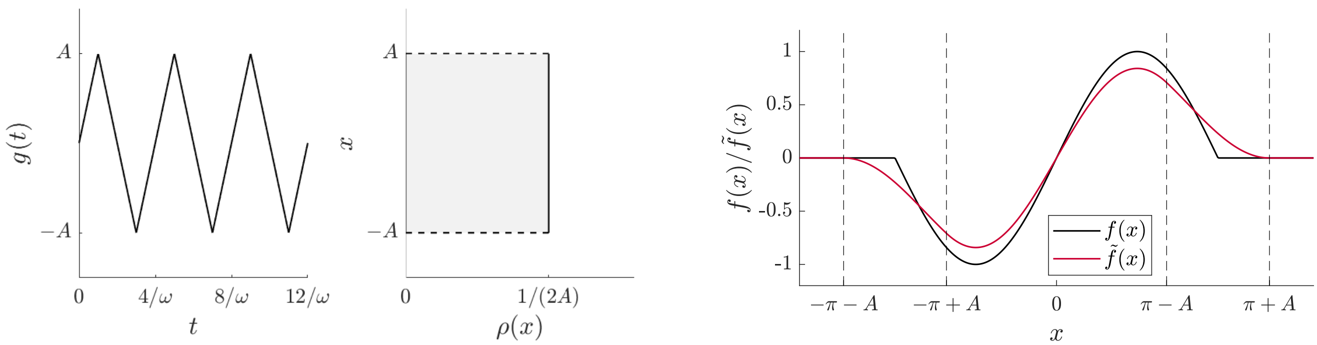
$$d_1 = -\pi - A, \qquad d_2 = -\pi + A, \qquad d_3 = \pi - A, \qquad d_4 = \pi + A. \tag{135}$$

We denote the five intervals defined by $d_1, \dots, d_4$ as

$$D_1 = \{x \in \mathbb{R} | x \le d_1\}, \qquad D_i = \{x \in \mathbb{R} | d_{i-1} \le x < d_i\} \quad \text{for } i = 2 \dots 4, \qquad D_5 = \{x \in \mathbb{R} | d_4 \le x\}.$$

Evaluating the integrals in the different domains, we finally find

$$\tilde{f}(x_S) = \begin{cases} 0 & x_S \in D_1, \\ -\frac{1}{2A}\left(1 + \cos(x_S + A)\right) & x_S \in D_2, \\ \mathrm{si}(A)\sin x_S & x_S \in D_3, \\ \frac{1}{2A}\left(1 + \cos(x_S - A)\right) & x_S \in D_4, \\ 0 & x_S \in D_5, \end{cases} \tag{136}$$



(a) Triangle wave function with amplitude A and its CPD (b) $f(x)$ and its averaged values $\tilde{f}(x)$ with $A = 1$.

Figure 11 Benchmark problem

with $\mathrm{si}(A) := \sin(x)/x$. This solution will be compared to numerical ones taking $A = 1$.

In Fig. 12a, the computational cost of the direct numerical integration of Eq. (129) by Riemann sums is depicted against the resolution of the spatial (N_x) and temporal (N_t) discretization in a log-log plot. The computational cost shows a sublinear dependency in N_x and N_t. In Fig. 12b, a log-log plot depicts the numerical solution's mean absolute error (MAE) against N_x and N_t. The MAE is defined as

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N} |\hat{X}_i - X_i|}{N}, \tag{137}$$

where $\hat{X}$ denotes the estimates, while X stands for the exact values. In the direct numerical quadrature of Eq. (129), N_x has practically no effect on the accuracy, while the absolute error decreases quadratically with increasing temporal resolution.

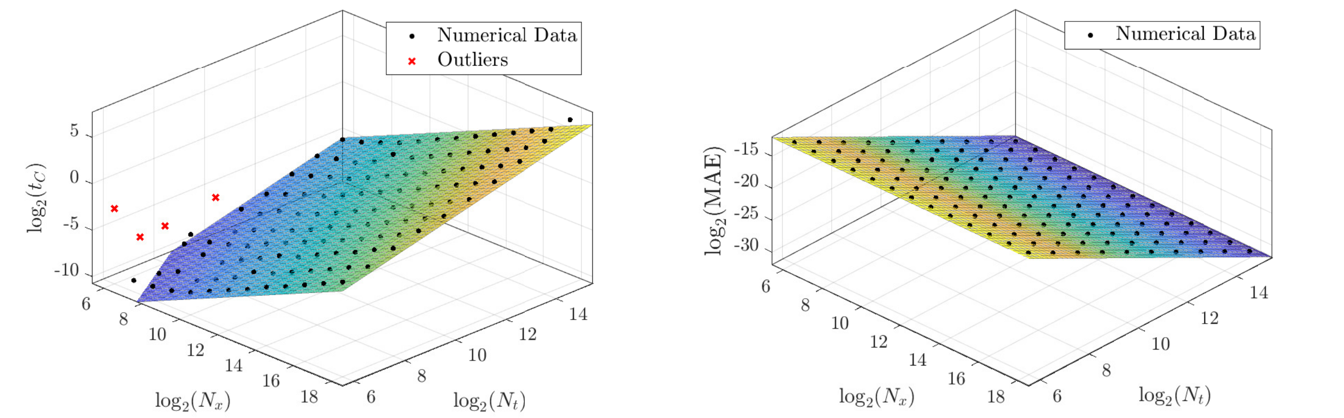
In Fig. 13, the computational cost and the MAE are depicted against the resolution of the spatial discretization on a log-log scale. Due to MATLAB's FFT-based cross-correlation algorithm, the computational cost grows only almost linearly with the problem size, while the MAE is inversely proportional to the spatial resolution.

The second benchmark example is given in *Example 2.*, $A = 0.5$ is used for numerical calculations. This problem is more challenging numerically than the previous one since $f(x)$ is not continuous, and the CPD of the arcsine distribution has singularities at its boundaries. However, the MAE converges to zero when the resolution of the spatial discretization tends to infinity (cf. Fig. 14). The MAE is not evenly distributed: it becomes the largest at the domain boundaries due to the discontinuities of $f(x)$ (cf. Fig. 15).

5 | CONCLUSIONS

An efficient alternative to standard averaging based on cross-correlation has been proposed in this article. The method is applicable if the dependent variable can be written as the sum of a "slow" and a "fast" variable, which is often the case since various methods of nonlinear dynamics, like multiple scales, averaging, or Blekhman's direct separation of motions are explicitly based on this assumption. In order to perform averaging on a function with the above "slow-fast" variable in its argument, it is sufficient if the CPD of the "fast" variable is known; its explicit time dependency itself is not required, although often available and may be used to obtain the CPD itself. In some cases, this might be a significant advantage over the classical way of evaluating the integral in Eq. (129) since $g(t)$ does not have to be known explicitly. This fact also provides more insight into what averaging is. Information on the exact time history of the "fast" variable is discarded, and only the probability of its location is used.

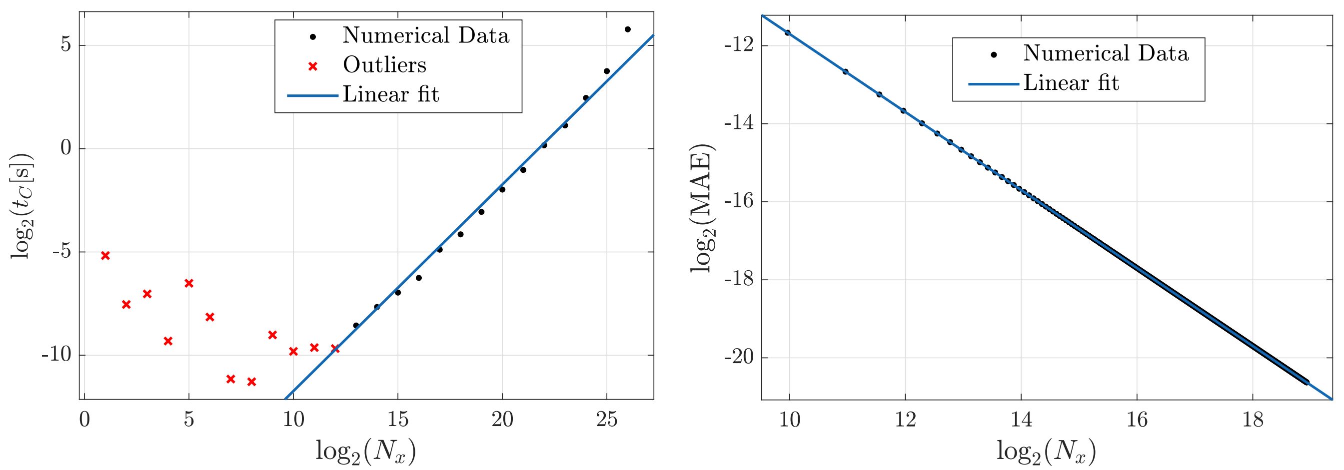
A further alternative for the representation of averages of analytic functions has been provided by a specific type of expansion where the moments of the "fast" variable's CPD and the target function's derivatives are used. By utilizing the moment generating and characteristic functions of random variables, it is possible to perform averaging with "fast" variables with the same CDP as the sum of independent random variables. It is well known that the PDF of such random variables can be calculated by the convolution of the individual PDFs, which simplifies the product of their moment-generating/characteristic functions.



(a) $\log_2 t_C \approx -22 + 0.9411 \log_2 N_x + 0.7031 \log_2 N_t$ with $R^2 = 0.9888$

(b) $\log_2 \mathrm{MAE} \approx -0.9962 + 0.003 \log_2 N_x - 2 \log_2 N_t$ with $R^2 = 1$

Figure 12 Plane fits on the computational time t_C and mean absolute error (MAE) depicted against the spatial and temporal discretization resolution in a log-log plot.



(a) Linear fit $\log_2 t_C \approx 0.9963 \log_2 N_x - 21.32$ with $R^2 = 0.9714$. $\mathcal{O}(N_x \log N_x)$ computational cost is expected

(b) Linear fit $\log_2 \mathrm{MAE} \approx -0.9999 \log_2 N_x - 1.701$ with $R^2 = 1$

Figure 13 Computational time t_C and mean absolute error (MAE) depicted against the spatial resolution of the discretization in a log-log plot.

Furthermore, several explicit formulas for CPDs have been derived in this study, and it has been shown how they might be utilized to perform averaging on piecewise defined polynomials. For other analytically not accessible cases, an efficient, FFT-based numerical method has been proposed.

The cross-correlation-based averaging method can be extended for the case of more than one dependent variable. However, it might be disadvantageous for periodic motions since in more than one spatial dimension, the CPD is not a function anymore, but it degenerates to a distribution on a set with zero measure, yet with unit hyper-volume. Still, for aperiodic motions, such as, for example, the motion of an undamped particle tossed in 45° angle in a two-dimensional rectangular "infinite-well" with incommensurable side length (ideally elastic ball in billiard table), the CPD becomes a function, and the multidimensional cross-correlation can be calculated. In this case, the particle is found at any point of the potential well with the same constant velocity inversely proportional to the area of the well.

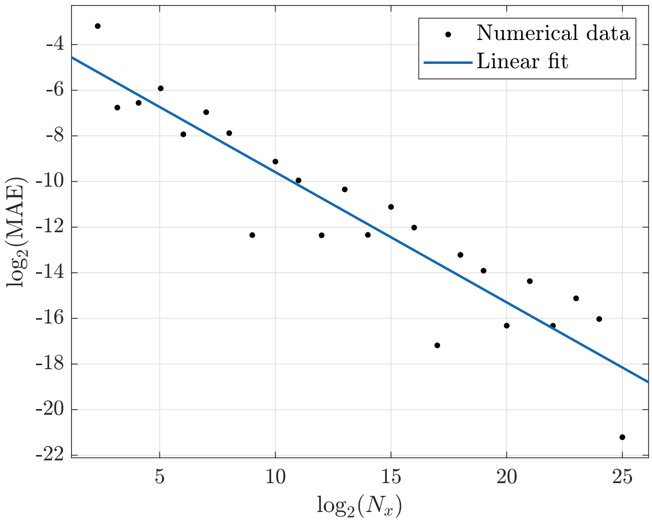


Figure 14 Linear fit $\log_2 \mathrm{MAE} \approx -0.5709 \log_2 N_x - 3.88$

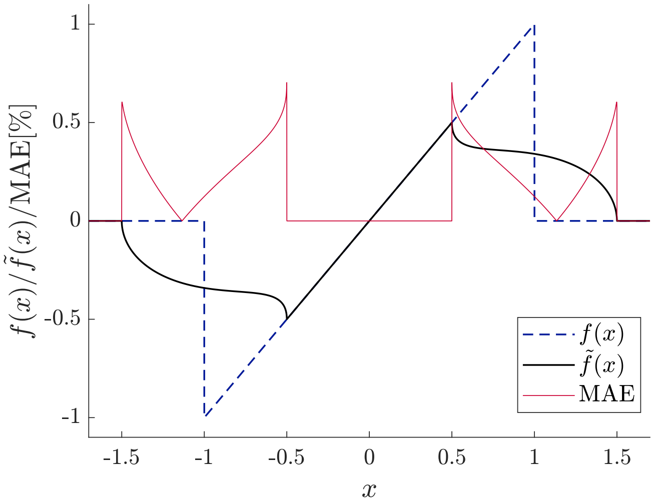


Figure 15 Absolute error [%] of the numerical cross correlation with $N_x = 2^{13} + 1$ grid points on $(-1.7, 1.7)$. $f(x) = x \cdot \mathbf{1}_{(-1,1)}(x)$ and $g(t) = 0.5 \sin t$. The analytic expression of the averaged function $\tilde{f}(x)$ is given in Eq. (56)

We aim to attract readers' attention to the model reduction technique proposed in this paper. The method has several potential applications in areas of science where multiple time scales are present. Indeed, without proving the validity of the equivalence, the method has been used in previous research of the authors [42,43] to reduce the complexity of the underlying escape problems, respectively.

One of the most interesting potential applications is when the CPD can be measured experimentally. In many cases, obtaining the high-frequency compound of the motion may be very difficult because it needs fast, high-resolution measuring equipment. On the other hand, the CPD can be easily obtained as a "cloud" picture, approximating the linger time by the brightness. With the appropriate image processing, the CPD can be estimated and used to obtain the semi-empirical equations governing the "slow" system's evolution.

It may be an exciting hint to future work to build a potential bridge between the classical dynamics of high-frequency excited systems and quantum mechanics. Further applications are in the development of control strategies [44].

Future research might also consider finding other cases of motion where explicit CPDs can be derived. The investigation of the goodness of approximations of CPDs, for example, through polynomials, might also be a possible direction of further research.

A possible generalization of this paper's results includes the extension of the cross-correlation integral to stochastic "fast" motions: Would the result on the equivalency of the time integral with a spatial cross-correlation hold in the stochastic case as well? If so, Theorem 1 may also be interpreted in the framework of the ergodic theory: a function's time average is exchanged by its spatial average.

ACKNOWLEDGMENT

We want to thank PD Gudrun Thäter for proofreading this paper. Her assistance helped to refine and improve the quality of this work.

FUNDING ACKNOWLEDGMENT

This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Project No. 508244284.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

References

[1] L.D. Landau, E.M. Lifshitz, *Mechanics 3rd ed.*, Butterworth, Oxford, **1976**.

[2] Wiktor Eckhaus, *Journal of Mathematical Analysis and Applications* **1975**, *49* (3), 575–611.

[3] O. V. Gendelman, *Nonlinear Dynamics* **2018**, *93* (1), 79–88.

[4] Vladimir I. Arnold, Valery V. Kozlov, Anatoly I. Neishtadt, *Mathematical Aspects of Classical and Celestial Mechanics 3rd ed.*, of *Encyclopaedia of Mathematical Sciences*, Springer Berlin, Heidelberg, **2006**, Original Russian edition published by URSS, Moscow, 2002.

[5] Alexander Fidlin, Olga Drozdetskaya, *Procedia IUTAM* **2016**, *19*, 43–52.

[6] Jimmy Aramendiz, Alexander Fidlin, Kan Lei, *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik* **2021**, *101* (7), e201800293.

[7] Simon Schröders, Alexander Fidlin, *Nonlinear dynamics* **2021**, *103* (3), 2315–2327.

[8] J. Guckenheimer, P. Holmes, *Nonlinear Oscillations Dynamical Systems and Bifurcations of Vector Fields*, Springer-Verlag, Berlin, **1990**.

[9] Guillaume Ballif, Frédérique Clément, Romain Yvinec, *SIAM Journal on Applied Mathematics* **2022**, *82* (1), 359–380.

[10] N.M. Krylov, N.N. Bogoliubov, *New Methods of Nonlinear Mechanics in their Application to the Investigation of the Operation of Electronic Generators, I (in Russian)*, United Scientific and Technical Press, Moscow, **1934**.

[11] NN Bogolyubov, Yu A Mitropolskii, Asymptotic methods in the theory of nonlinear oscillations, , Foreign Technology Div Wright-Patterson AFB, Ohio, **1955**.

[12] Vadim Strygin, *Journal of Applied Mathematics and Mechanics* **1984**, *48*, 767–769.

[13] R Khasminski, *Kybernetika* **1968**, *4* (3), 260–279.

[14] Z.H. Liu, W.Q. Zhu, *Journal of Sound and Vibration* **2007**, *299* (1), 178–195.

[15] L. Chen, X. Liang, W. Zhu, et al., *Sci. China Technol. Sci.* **2019**, *62*, 287–297.

[16] Yuri Kifer, *Modern dynamical systems and applications* **2004**, 385–403.

[17] Yuri Kifer, *Inventiones mathematicae* **1992**, *110* (1), 337–370.

[18] Yuri Kifer, *Ergodic Theory and Dynamical Systems* **2004**, *24* (3), 847–871.

[19] V M Volosov, *Russian Mathematical Surveys* **1962**, *17* (6), 1.

[20] John Guckenheimer, Philip Holmes, *Averaging and Perturbation from a Geometric Viewpoint*, Springer New York, New York, NY, **1983**, pp. 166–226.

[21] Yu. A. Mitropolskii, Nguyen Van Dao, *Averaging Method*, Springer Netherlands, Dordrecht, **1997**, pp. 282–326.

[22] J.A. Sanders, F. Verhulst, J. Murdock, *Averaging Methods in Nonlinear Dynamical Systems*, of *Applied Mathematical Sciences*, Springer New York, **2007**.

[23] Alexander Fidlin, *Nonlinear oscillations in mechanical engineering*, Springer Science & Business Media, **2005**.

[24] P.T. Krein, J. Bentsman, R.M. Bass, B.L. Lesieutre, *IEEE Transactions on Power Electronics* **1990**, *5* (2), 182–190.

[25] Jian Sun, H. Grotstollen, *IEEE Transactions on Power Electronics* **1997**, *12* (3), 537–546.

[26] Yuri Kifer, in *Stochastic Climate Models*, (Eds: Peter Imkeller, Jin-Song von Storch), Birkhäuser Basel, Basel, **2001**, pp. 171–188.

[27] Ernst Hairer, Christian Lubich, Gerhard Wanner, *Geometric numerical integration 2nd ed.*, Vol. 31 of *Springer Series in Computational Mathematics*, Springer-Verlag, Berlin, **2006**, Structure-preserving algorithms for ordinary differential equations.

[28] Ben Leimkuhler, Sebastian Reich, *Journal of Computational Physics* **2001**, *171* (1), 95–114.

[29] D.J. Griffiths, D.F. Schroeter, *Introduction to Quantum Mechanics*, Cambridge University Press, **2018**.

[30] R. W. Robinett, *American Journal of Physics* **1995**, *63* (9), 823–833.

[31] C. C. Real, J. G. Muga, S. Brouard, *Comment on "Quantum and classical probability distributions for position and momentum," by R. W. Robinett [Am. J. Phys. 63 (9), 823–832 (1995)]*, **1997**. https://doi.org/10.1119/1.18791.

[32] G. Yoder, *American Journal of Physics* **2006**, *74* (5), 404.

[33] Harald Bohr, *Acta Mathematica* **1925**, *46* (1), 101–214.

[34] Jay L. Devore, Kenneth N. Berk, *Modern Mathematical Statistics with Applications 2nd ed.*, of *Springer Texts in Statistics*, Springer New York, NY, **2012**.

[35] Maurice G, Alan Stuart, J. Keith Ord, *Kendall's Advanced Theory of Statistics, Volume 1, Distribution Theory 6th ed.*, John Wiley & Sons, **1994**.

[36] A. M. Lyapunov, *Stability of Motion*, GITTL, Moscow, **1950**.

[37] A. Gelb, W. E. Vander Velde, *Multiple-Input Describing Functions and Nonlinear System Design*, McGraw-Hill, New York, **1968**.

[38] Ali H Nayfeh, Delores T Mook, *Nonlinear oscillations*, Wiley, New York, **1979**.

[39] Ivana Kovacic, *Nonlinear Oscillations: Exact Solutions and their Approximations 1st ed.*, Springer Cham, Switzerland, **2020**, Hardcover ISBN: 978-3-030-53171-3; Softcover ISBN: 978-3-030-53174-4; eBook ISBN: 978-3-030-53172-0.

[40] Robert V. Hogg, Joseph W. McKean, Allen T. Craig, *Introduction to mathematical statistics 6th ed.*, Prentice Hall, Upper Saddle River, New Jersey, **2004**.

[41] Harry Bateman, *Higher Transcendental Functions [Volumes I-III], Vol. I-III*, McGraw-Hill Book Company, New York, **1953**.

[42] Attila Genda, Alexander Fidlin, Oleg Gendelman, *Nonlinear Dynamics* **2021**, *104* (1), 91–102.

[43] Attila Genda, Alexander Fidlin, Oleg Gendelman, *On the Escape of a Resonantly Excited Couple of Colliding Particles from a Potential Well under Bi-harmonic Excitation*, **2022**. https://enoc2020.sciencesconf.org/375919/document.

[44] Alexander Fidlin, Jon Juel Thomsen, *International Journal of Non-Linear Mechanics* **2008**, *43* (7), 569–578.

[45] I. Bronstein, K. Semendjajew, G. Musiol, H. Mühlig, *Taschenbuch der Mathematik*, Verlag Harri Deutsch, Frankfurt am Main, **2001**.

[46] Robert L. Winkler, et al., *Management Science* **1972**, *19* (3), 290–296.

How to cite this article: A. Genda, A. Fidlin, and O. Gendelman (2023), Cross-Correlation and Averaging: An Equivalence Based on the Classical Probability Density *(to be filled out by the editor of the journal)*

APPENDIX

The arcsine distribution's moments and partial moments are obtained in the following. First, the moment generating function $M_X(t)$ of the standard arcsine distribution (a special case of the Beta distribution with parameters $\alpha = \beta = 1/2$[35]), given on $(0, 1)$ is derived. Let $X \sim \mathrm{Beta}(1/2, 1/2)$ be a random variable following the standard arcsine distribution, of which the moment-generating function is defined as

$$M_X(t) = \mathrm{E}\left[e^{tX}\right] = \int_0^1 e^{tx} \frac{1}{\pi\sqrt{x(1-x)}} \mathrm{d}x = {}_1F_1\left(\frac{1}{2}; 1; t\right), \tag{A.1}$$

where ${}_1F_1(a; b; t)$ denotes the confluent hypergeometric function defined by

$${}_1F_1(a; b; t) = \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)} \int_0^1 e^{tu} u^{a-1} (1-u)^{b-a-1} \mathrm{d}u. \tag{A.2}$$

By a linear transformation, one can change the distribution's location and scaling, i.e., by $Y = \alpha X + \beta$, the random variable is shifted to the negative direction by β and stretched by α. The moment-generating function then becomes

$$M_{\alpha X+\beta}(t) = \mathrm{E}\left[e^{(\alpha X+\beta)t}\right] = e^{\beta}\mathrm{E}\left[e^{\alpha X t}\right] = e^{\beta t} M_X(\alpha t) = e^{\beta t}{}_1F_1\left(\frac{1}{2}, 1, \alpha t\right) = \exp\left(\frac{\alpha t}{2} + \beta t\right) I_0\left(\frac{\alpha z}{2}\right), \tag{A.3}$$

with $I_0(x)$ denoting the modified Bessel function of the first kind of the zeroth order. The moments are given by

$$m_n = M^{(n)}_{\alpha X+\beta}(t)\Big|_{t=0}. \tag{A.4}$$

For the confluent hypergeometric function, the following identity holds

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}{}_1F_1(a, b, t) = \frac{(a)_k}{(b)_k}{}_1F_1(a+k, b+k, t), \tag{A.5}$$

where $(a)_k$ denotes the rising factorial, i.e. $(a)_k = a(a+1)(a+2)\dots(x+k-1)$ and $(a)_0 = 1$. By making use of the identity, one obtains the k^{th} derivative of Eq. (A.3)

$$M^{(k)}_{\alpha X+\beta}(t) = \sum_{j=0}^{k} \binom{k}{j} \alpha^j \beta^{k-j} \frac{\left(\frac{1}{2}\right)_j}{(1)_j} e^{\beta t}{}_1F_1\left(\frac{1}{2} + j, 1 + j, \alpha t\right). \tag{A.6}$$

Inserting $t = 0$, one obtains the moments

$$m_k = \sum_{j=0}^{k} \binom{k}{j} \alpha^j \beta^{k-j} \frac{\left(\frac{1}{2}\right)_j}{(1)_j} = \beta^k + \sum_{j=1}^{k} \binom{k}{j} \alpha^j \beta^{k-j} \frac{(2j-1)!!}{2^j j!}, \quad \text{for } k \geq 1. \tag{A.7}$$

where $(2j-1)!! = (2j-1)(2j-3)\dots 3 \cdot 1$ is the double factorial. In case of a centered arcsine distribution with $\alpha = 2A$ and $\beta = -A$ one has

$$m_k = A^k \left((-1)^k + \sum_{j=1}^{k} \binom{k}{j} (-1)^{k-j} \frac{(2j-1)!!}{j!} \right) = \begin{cases} A^k \frac{1}{2^k} \binom{k}{k/2} & \text{for } k \text{ even,} \\ 0 & \text{for } k \text{ odd,} \end{cases} \quad \text{for } k \geq 1. \tag{A.8}$$

The partial moments of the centered arcsine distribution can be determined by solving the integral

$$m_k(x; A) = \int_{-A}^{x} \frac{y^k}{\pi\sqrt{A^2 - y^2}} \mathrm{d}y = A^k \int_{-\frac{\pi}{2}}^{\arcsin\frac{x}{A}} \sin^k t \mathrm{d}t, \tag{A.9}$$

where the change of variables $y = A \sin t$ has been applied. The indefinite integral of $\sin^k t$ can be obtained by using recursively the identity[45]

$$\int \sin^k t \mathrm{d}t = -\frac{\sin^{k-1} t \cos t}{k} + \frac{k-1}{k} \int \sin^{k-2} t \mathrm{d}t. \tag{A.10}$$

Thus,

$$\int \sin^k t \mathrm{d}t = \begin{cases} -\cos t \left(\sum_{j=0}^{\frac{k-1}{2}} p(k,j) \sin^{k-2j-1} t \right) & \text{for } k \text{ even,} \\ -\cos t \left(\sum_{j=0}^{\frac{k}{2}-1} p(k,j) \sin^{k-2j-1} t \right) + p(k, \frac{k}{2}-1)t & \text{for } k \text{ odd.} \end{cases} \tag{A.11}$$

with

$$p(k,j) = \begin{cases} \dfrac{\binom{k}{k/2}}{2^{2j+2}(k-2j-1)\binom{k-1-2j}{(k-1)/2-j}} & \text{for } k \text{ even,} \\ \dfrac{2^{2j}\binom{k-2j-1}{(k-1)/2}}{\binom{k-1}{(k-1)/2}k} & \text{for } k \text{ odd.} \end{cases} \tag{A.12}$$

Hence, Eq. (A.9) becomes

$$m_k(x;A) = \begin{cases} 0 & \text{for } x < -A \\ -\pi^{-1}\sqrt{A^2-x^2}A^{k-1} \sum_{j=0}^{(k-1)/2} p(k,j)x^{k-2l-1} & \text{for } |x| < A \text{ and } k \text{ even,} \\ -\pi^{-1}\sqrt{A^2-x^2}A^{k-1} \sum_{j=0}^{k/2-1} p(k,j)x^{k-2l-1} + p(k,k/2-1)\left(\pi^{-1}\arcsin(x/A)+1/2\right) & \text{for } |x| \leq A \text{ and } k \text{ odd,} \\ m_k & \text{for } A < x, \end{cases} \tag{A.13}$$

where m_k is defined in Eq. (A.8). The arcsine distribution's first six partial moments ($k = 0, \ldots, 5$) are depicted in Fig. 2.