

Extraction of backgrounds in fluctuating systems

Ning-Ning Pang

Department of Physics, National Taiwan University, Taipei, Taiwan, Republic of China

Wen-Jer Tzeng

Department of Physics, Tamkang University, Tamsui, Taipei, Taiwan, Republic of China

(Received 3 July 2003; revised manuscript received 25 November 2003; published 24 March 2004)

We undertake an extensive analytical study on the “generalized detrended fluctuation analysis” method, designed to detect the scaling behaviors of fluctuating systems but exclude out the influences of the backgrounds (or the trends). Through our extensive studies, we systematically extract out the exact backgrounds (or the trends) of the fluctuating systems to any order, expressed in terms of the Legendre polynomial. Our results are exact and can be applied to any $(1+1)$ -dimensional continuous fluctuating systems.

DOI: 10.1103/PhysRevE.69.031108

PACS number(s): 05.40.-a, 05.45.Tp, 68.35.Ct, 89.75.Fb

Fluctuating systems have drawn considerable interest for the generic behaviors widespread in nature and society, such as molecular-beam-epitaxy interface growth [1,2], DNA sequences [3,4], physiology signals [5–7], atmospheric variability [8], and currency exchange rates [9]. Plenty of fluctuating systems contain noisy “signals,” which are statistically heterogeneous and exhibit various types of non-stationarity. For example, in meteorology the variations of the daily maximum temperatures from their average values have trends in time due to global warming [8], in financial markets the currency exchange rates obviously have trends in time due to the relative economic conditions in different countries [9], and in computational molecular biology the nucleotide concentrations in DNA sequences are heterogeneous and form the mosaic structures [3,4]. Peng *et al.* [3] then proposed a method “detrended fluctuation analysis” (DFA), designed to detect and quantify the long-range correlation in DNA sequences but exclude the artifact long-range correlation caused by the DNA’s mosaic structure (patchiness). The algorithm of the DFA method is given as follows. Let $y(i)$ be a sequence of length N . Then, a sliding observation window of length n is chosen and slid across the whole data set. Inside the observation window, the linear trend $\tilde{y}(i) = ai + b$ of the original profile $y(i)$ is eliminated, with the parameters a and b obtained from the least squares fit of the difference between the linear trend $\tilde{y}(i)$ and the original profile $y(i)$. Finally, one calculates the variance with respect to the linear trend from the original profile, takes the average over all the windows of length n , and then denotes it by $F_d^2(n)$, the square of the detrended fluctuation function. This method first demonstrated its usefulness in the investigation of long-range correlation in the noncoding regions of DNA sequences; since then, it has been widely adopted as a standard method in analyzing time series problems in various fields. In addition, its generalization by including higher-order trends in the analysis, the “generalized detrended fluctuation analysis” (GDFA) method, has recently been numerically studied in detail in Refs. [10,11].

On the other hand, recently much attention has been focused on the super-rough growth processes for their peculiar interfacial morphology [2,12–15]. It has been observed

both experimentally and numerically that some $(1+1)$ -dimensional growth processes are “super-rough,” since their saturated global interfacial widths relative to the average interfacial heights diverge faster than the lateral system size L . The most distinct feature of the super-rough interface is that its local spatiotemporal scaling is different from its global spatiotemporal scaling; namely, the local interfacial widths relative to the average interfacial heights, measured within a local window of size l ($\ll L$), have very different dynamic scaling behavior from that of the global interfacial widths. The main reason is that all the super-rough interfacial growth processes are associated with local interfacial orientational instability but, at the same time, with fixed or periodic boundary conditions restricting the development of the global interfacial width. We have proposed [14,15] a definition of the local interfacial width $\tilde{w}(l,t)$, the residual local interfacial width, as follows:

$$\tilde{w}^2(l,t) \equiv \overline{\langle \langle [h(x,t) - \tilde{h}_l(x,t)]^2 \rangle \rangle_l} \quad (1)$$

with $h(x,t)$ denoting the interface height from a flat substrate at position x and time t , $\tilde{h}_l(x,t)$ denoting the height at position x of a straight line segment obtained by least squares fit to the interfacial configuration in the local window of size l at a given time t , $\langle \cdot \cdot \cdot \rangle_l$ denoting the lateral spatial average over a local window of size l , $\langle \cdot \cdot \cdot \rangle_L$ denoting the sliding of the local window over the whole system of size L , and finally the overbar denoting the average over randomness. Clearly, $\tilde{w}(l,t)$ describes the width of the interfacial height fluctuation relative to the local interfacial orientation within a local window of size l . We then have shown [14,15] that, for the super-rough growth processes which form the mountains or the valleys in the interface configuration, the residual local interfacial widths retrieve the same dynamic scaling as the global interfacial widths. Although the DFA method is mainly used in the analysis of fluctuating time series and the residual local interfacial width is used in the study of the snapshot of $(1+1)$ -dimensional super-rough interface morphology at a certain time, the spirits of their mathematical formulation are exactly the same; i.e., they both are designed to eliminate the influences of the trends (or the backgrounds).

Although the G DFA method is so widely used, there remain some unsettled issues.

(1) For the correct estimate of the scaling exponents, high-order detrending are frequently required [5,6]. Then, it will involve the numerical least squares fit of many parameters at the same time and, thus, consume large CPU time.

(2) Many fluctuating time series problems and kinetic interfacial roughening processes are modeled by continuous stochastic differential equations. Thus, the explicit and exact expression of the trend (or the background) to any order can greatly help the analytical study of such equations.

Therefore, we are strongly motivated to take an extensive analytical study on the G DFA method and plan to extract out the *exact* expressions of the backgrounds (or the trends) in fluctuating systems to any order.

First, let $y(x)$ denote some continuous stochastic sequence of length L . For example, to describe time series problems, x and y simply represent time and the physical quantity to be measured, respectively; to describe the snapshot of the (1+1)-dimensional super-rough interface morphology at a certain time, x and y then denote the spatial coordinate along the substrate direction and the interfacial height, respectively. The variance of $y(x)$ in an observation window, centered at \hat{x} , of length l is then given as

$$w^2(l;\hat{x}) \equiv \langle [y(x) - \langle y(x) \rangle_{l;\hat{x}}]^2 \rangle_{l;\hat{x}}, \quad (2)$$

with $\langle \dots \rangle_{l;\hat{x}}$ denoting the average over an observation window, centered at \hat{x} , of length l . Using a p th degree polynomial to extract the contribution from the backgrounds (or the trends), the p th degree detrended variance of $y(x)$ in that observation window of length l is defined as

$$w_p^2(l;\hat{x}) \equiv \langle [y(x) - \hat{y}_p(x;\hat{x})]^2 \rangle_{l;\hat{x}}, \quad (3)$$

where the p th degree polynomial $\hat{y}_p(x;\hat{x})$ is obtained from the least squares fit to $y(x)$ within that observation window. Naively, people will naturally choose

$$\hat{y}_p(x;\hat{x}) = \sum_{q=0}^p M_q (x-\hat{x})^q \quad (4)$$

with the coefficient M_q obtained from the relation

$$0 = \frac{\partial w_p^2(l;\hat{x})}{\partial M_q} = -2 \left\langle \left[y(x) - \sum_{q'=0}^p M_{q'} (x-\hat{x})^{q'} \right] \times (x-\hat{x})^q \right\rangle_{l;\hat{x}}. \quad (5)$$

Then,

$$\begin{aligned} \langle y(x)(x-\hat{x})^q \rangle_{l;\hat{x}} &= \sum_{q'=0}^p M_{q'} \langle (x-\hat{x})^{q+q'} \rangle_{l;\hat{x}} \\ &= \sum_{q'=0}^p M_{q'} \left\{ \frac{\left(\frac{l}{2}\right)^{q+q'} [1+(-1)^{q+q'}]}{2(q+q'+1)} \right\}. \end{aligned} \quad (6)$$

Consequently, we have

$$M_q = \sum_{q'=0}^p [\mathbf{A}^{-1}]_{qq'} \langle y(x)(x-\hat{x})^{q'} \rangle_{l;\hat{x}}, \quad (7)$$

where the elements of the $p \times p$ matrix \mathbf{A} are

$$A_{qq'} = \frac{\left(\frac{l}{2}\right)^{q+q'} [1+(-1)^{q+q'}]}{2(q+q'+1)}. \quad (8)$$

Thus, to obtain the explicit form of M_q involves an inverse transform of the $p \times p$ matrix \mathbf{A} , which usually relies on computers for large values of p . However, the derivation of the explicit form of the coefficients M_q is necessary for any further analytical studies such as obtaining the exact relation between $w^2(l;\hat{x})$ [the original variance of $y(x)$ in an observation window, centered at \hat{x} , of length l] and $w_p^2(l;\hat{x})$ (its p th degree detrended variance in that observation window) or exploring the exact influences of the backgrounds (or the trends) on the fluctuating systems, etc.

Fortunately, with a closer look at Eq. (5), we note that this complication of the inverse matrix transform can actually be avoided if the basis of $\hat{y}_p(x;\hat{x})$ is deliberately chosen to be the polynomials of degree 0 to p , orthogonal to one another. Legendre polynomial $P_q(x)$ [16] is just a perfect candidate since it is a polynomial of degree q , given by

$$P_q(x) = \sum_{j=0}^{[q/2]} (-1)^j \frac{(2q-2j-1)!!}{(2j)!!(q-2j)!} x^{q-2j}, \quad (9)$$

and satisfies the orthogonal relations

$$\int_{-1}^1 dx P_q(x) P_{q'}(x) = \frac{2}{2q+1} \delta_{q,q'}. \quad (10)$$

Thus, we choose Legendre polynomial $P_q[2(x-\hat{x})/l]$, instead of $(x-\hat{x})^q$, to be the basis of $\hat{y}_p(x;\hat{x})$; namely,

$$\hat{y}_p(x;\hat{x}) = \sum_{q=0}^p C_q P_q \left(\frac{2(x-\hat{x})}{l} \right) \quad (11)$$

with the coefficient C_q obtained from the relation

$$0 = \frac{\partial w_p^2(l; \hat{x})}{\partial C_q} = -2 \left\langle \left[y(x) - \sum_{q'=0}^p C_{q'} P_{q'} \left(\frac{2(x-\hat{x})}{l} \right) \right] \times P_q \left(\frac{2(x-\hat{x})}{l} \right) \right\rangle_{l; \hat{x}}. \quad (12)$$

By employing Eq. (10) [the orthogonal relation of $P_q(x)$], the coefficient C_q is then explicitly derived as

$$C_q = (2q+1) \left\langle y(x) P_q \left(\frac{2(x-\hat{x})}{l} \right) \right\rangle_{l; \hat{x}} \\ = \sum_{j=0}^{\lfloor q/2 \rfloor} (-1)^j \frac{2^{q-3j} (2q-2j-1)!!}{j!(q-2j)!} \\ \times \left\langle y(x) \left(\frac{x-\hat{x}}{l} \right)^{q-2j} \right\rangle_{l; \hat{x}}. \quad (13)$$

Recall that the p th degree background (or the trend) $\hat{y}_p(x; \hat{x})$ is just $\sum_{q=0}^p C_q P_q[2(x-\hat{x})/l]$; thus, we have succeeded in *extracting the power series expansion of the background (or the trend) to any order*. Our result is applicable for any continuous and analytic background (or the trend) regardless of the shape. For illustration, the values of C_0 – C_3 are explicitly listed as follows,

$$C_0 = \langle y(x) \rangle_{l; \hat{x}}, \quad (14)$$

$$C_1 = \frac{6}{l} \langle y(x)(x-\hat{x}) \rangle_{l; \hat{x}}, \quad (15)$$

$$C_2 = \frac{30}{l^2} \langle y(x)(x-\hat{x})^2 \rangle_{l; \hat{x}} - \frac{5}{2} \langle y(x) \rangle_{l; \hat{x}}, \quad (16)$$

$$C_3 = \frac{140}{l^3} \langle y(x)(x-\hat{x})^3 \rangle_{l; \hat{x}} - \frac{21}{l} \langle y(x)(x-\hat{x}) \rangle_{l; \hat{x}}, \quad (17)$$

and the corresponding $\hat{y}_0(x; \hat{x})$ to $\hat{y}_3(x; \hat{x})$ are also displayed as follows

$$\hat{y}_0(x; \hat{x}) = \langle y(x) \rangle_{l; \hat{x}}, \quad (18)$$

$$\hat{y}_1(x; \hat{x}) = \langle y(x) \rangle_{l; \hat{x}} + 12 \langle y(x) \tilde{x} \rangle_{l; \hat{x}} \tilde{x}, \quad (19)$$

$$\hat{y}_2(x; \hat{x}) = \langle y(x) \rangle_{l; \hat{x}} \cdot \left(\frac{9}{4} - 15 \tilde{x}^2 \right) + 12 \langle y(x) \tilde{x} \rangle_{l; \hat{x}} \tilde{x} \\ - \langle y(x) \tilde{x}^2 \rangle_{l; \hat{x}} (15 - 180 \tilde{x}^2), \quad (20)$$

$$\hat{y}_3(x; \hat{x}) = \langle y(x) \rangle_{l; \hat{x}} \left(\frac{9}{4} - 15 \tilde{x}^2 \right) + \langle y(x) \tilde{x} \rangle_{l; \hat{x}} (75 \tilde{x} - 420 \tilde{x}^3) \\ - \langle y(x) \tilde{x}^2 \rangle_{l; \hat{x}} (15 - 180 \tilde{x}^2) - \langle y(x) \tilde{x}^3 \rangle_{l; \hat{x}} \\ \times (420 \tilde{x} - 2800 \tilde{x}^3) \quad (21)$$

with the dimensionless quantity $\tilde{x} \equiv (x-\hat{x})/l$.

In conclusion, we have undertaken an extensive analytical study on the GDFA method. Through our extensive studies, the exact backgrounds (or the trends) of the fluctuating systems are explicitly extracted out to any order, expressed in terms of the Legendre polynomial. Thus, no more numerical fitting is needed and large computation time can be saved. In addition, the obtained explicit expression of the background (or the trends) can greatly help the analytical investigation of fluctuating systems modeled by continuous stochastic differential equations. For example, it can be used to obtain the exact relation between the correlation function and the detrended width of fluctuations or to investigate the development of local orientational instability in the interfacial super-roughening phenomena. Moreover, our results can be used to show that the detrended width of fluctuation w_p only excludes out the influence of the background (or the trend) on the scaling behavior of the system and does retain the true scaling behavior originated from the stochastic nature of the system [17]. Note that all the obtained results are exact and thus can be applied to any (1+1)-dimensional continuous fluctuating systems.

The work of N.-N.P. was supported in part by the National Science Council of the Republic of China under Grant No. NSC 92-2112-M-002-017. The work of W.-J.T. was supported in part by the National Science Council of the Republic of China under Grant No. NSC 92-2112-M-032-011.

-
- [1] J. Krug, *Adv. Phys.* **46**, 139 (1997); T. Halpin-Healy and Y.-C. Zhang, *Phys. Rep.* **254**, 215 (1995).
 [2] S. Das Sarma, C.J. Lanczycki, R. Kotlyar, and S.V. Ghaisas, *Phys. Rev. E* **53**, 359 (1996).
 [3] C.-K. Peng *et al.*, *Phys. Rev. E* **49**, 1685 (1994).
 [4] S.V. Buldyrev *et al.*, *Phys. Rev. E* **51**, 5084 (1995).
 [5] S. Bahar *et al.*, *Europhys. Lett.* **56**, 454 (2001).
 [6] A. Bunde *et al.*, *Phys. Rev. Lett.* **85**, 3736 (2000).
 [7] M. Cernelc *et al.*, *J. Appl. Physiol.* **92**, 1817 (2002).
 [8] E. Koscielny-Bunde *et al.*, *Phys. Rev. Lett.* **81**, 729 (1998).
 [9] M. Ausloos and K. Ivanova, *Eur. Phys. J. B* **20**, 537 (2001).
 [10] J.W. Kantelhardt *et al.*, *Physica A* **295**, 441 (2001).
 [11] K. Hu *et al.*, *Phys. Rev. E* **64**, 011114 (2001).
 [12] H.-N. Yang, Y.-P. Zhao, G.-C. Wang, and T.-M. Lu, *Phys. Rev. Lett.* **76**, 3774 (1996).
 [13] J.H. Jeffries, J.K. Zuo, and M.M. Craig, *Phys. Rev. Lett.* **76**, 4931 (1996).
 [14] N.-N. Pang and W.-J. Tzeng, *Phys. Rev. E* **61**, 3212 (2000).
 [15] N.-N. Pang and W.-J. Tzeng, *Phys. Rev. E* **61**, 3559 (2000).
 [16] See, for example, *An Atlas of Functions*, edited by J. Spanier and K.B. Oldham (Springer-Verlag, Berlin, 1987); *Integrals and Series*, edited by A.P. Prudnikov, Y.A. Brychkov, and O.I. Marichev (Gordon and Breach, New York, 1986).
 [17] The detailed mathematical proof will be published elsewhere.