

Dynamically sampled multivariate empirical mode decomposition

N. Rehman[✉], K. Naveed, M.W. Safdar, S. Ehsan and K.D. McDonald-Maier

A method for accurate multivariate local mean estimation in the multivariate empirical mode decomposition algorithm by using a statistical data-driven approach based on the Menger curvature measure and normal-to-anything variate-generation method is proposed. This is achieved by aligning the projection vectors in the direction of the maximum ‘activity’ of the input signal by considering the local curvature of the signal in multidimensional spaces, resulting in accurate mean estimation even for a very small number of projection vectors.

Introduction: Empirical mode decomposition (EMD) [1] decomposes an input signal $x(t)$ into a finite number M of multiple scales $c_m(t)$, also known as intrinsic mode functions, based on its intrinsic oscillations. In its original formulation, EMD can only process signals containing a single channel (univariate data). Its important extensions for multichannel data include: (i) bivariate EMD (BEMD) [2] and (ii) multivariate EMD (MEMD) [3]. A critical task in these extensions is the estimation of the multivariate local mean of the input signal based on its extrema. This is not a trivial task, however, since the concept of extrema for multidimensional spaces, such as complex space \mathbb{C} and quaternion space \mathbb{H} , is not defined. Therefore, these extensions operate by projecting the multivariate signal along different static but uniform directions in multidimensional space to obtain real-valued projections; interpolating the extrema locations of projected real-valued signals; and, finally, calculating their mean $m_p(t)$ as follows:

$$m_p(t) \approx \frac{1}{V_1 V_2 \dots V_{p-1}} \sum_{v_1=1}^{V_1} \sum_{v_2=1}^{V_2} \dots \sum_{v_{p-1}=1}^{V_{p-1}} e_{\{\phi_{v_1}, \phi_{v_2}, \dots, \phi_{v_{p-1}}\}} \quad (1)$$

where the vector $\{V_1, V_2, \dots, V_{p-1}\}$ and $e_{\{\phi_{v_1}, \phi_{v_2}, \dots, \phi_{v_{p-1}}\}}$, respectively, denote the number of direction vectors and the envelopes taken along the directions $\phi = \{\phi_{v_1}, \phi_{v_2}, \dots, \phi_{v_{p-1}}\}$, for a multivariate signal $s(t)$ with $p > 2$ number of channels [3]. For a very large number of projection vectors, V , the approximation in (1) holds; that is, convergence to the ‘true’ local mean is achieved. However, in practical cases, a small number of direction vectors should be used due to computational constraints, requiring an effective choice of those vectors as their distribution is critical to the accuracy of the local mean.

To this end, we present a method to generate projection vectors in multidimensional space in the directions where the input signal exhibits maximum dynamics (variations). These critical directions are obtained by first calculating the empirical distribution function (EDF) of the direction vectors that exhibit higher dynamics based on the local Menger curvature values of the signal. Subsequently, samples (of direction vectors) are drawn from the above calculated EDF using the normal-to-anything (NORTA) multivariate random vector generation method, resulting in a set of vectors where the signal dynamics is large.

Previous attempts for dynamically selecting the projection vectors in EMD include: (i) a global sampling strategy [4] which is only suited to stationary data, and (ii) the local dynamically sampled BEMD (DS-BEMD) scheme to generate direction vectors for bivariate data only [5]. Our proposed method is the first general extension of DS-BEMD for multivariate signals (containing any number of channels).

Measuring signal curvature in multidimensional space: The dynamics of a multivariate signal (containing p channels) can easily be visualised in a (p -dimensional) phase space, where the axes correspond to the input channels of the signal. Each multivariate signal, therefore, traces a unique curve in such a space.

In the proposed method, the first step involves estimating the local dynamics of an input multivariate signal in its multidimensional space. In mathematics, the *curvature* of a function can be used to quantify its deviation from a straight line and can therefore provide a useful measure to describe the local dynamics of such signals. A convenient means to measure the local curvature of a signal is through the Menger curvature κ , computed for a set of three points x, y and z on the curve and is equal to the reciprocal of the radius R of the circle

passing through those points

$$\kappa = \frac{1}{R} \quad (2)$$

Another way to compute κ is through the area A of the triangle xyz passing through those three points, with side lengths $|xy|$, $|yz|$ and $|zx|$

$$\kappa = 1/R = \frac{4A}{|xy||yz||zx|} \quad (3)$$

The definition of the Menger curvature in terms of the reciprocal of the radius of a circle passing through three points on a curve, as in (2), provides a useful interpretation of local signal dynamics: Fig. 1 illustrates this further where two osculating circles are drawn at points P and Q on the curve. At point P , the signal dynamics is slow as compared with the point Q and, therefore, the corresponding radius R_1 is greater than R_2 yielding a higher Menger curvature measure at Q , i.e. $\kappa_Q > \kappa_P$.

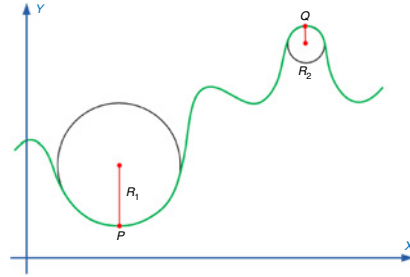


Fig. 1 Menger curvature estimation at points P and Q on curve via their respective radii (R_1 and R_2)

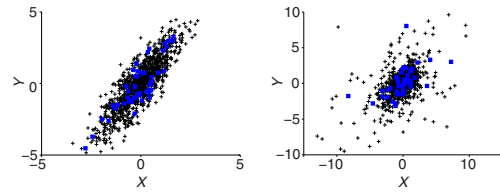


Fig. 2 NORTA-generated multivariate sample (blue squares) from input bivariate t -distribution (left) and normal distribution (right) data (shown in black cross)

In the proposed scheme, κ is computed at each point of a curve corresponding to the input signal; subsequently, only L number of points exhibiting the highest κ values are chosen and their angles $\{\phi^i\}_{i=1}^L$ are calculated. For a p -dimensional phase space, $\phi^i = \{\phi_1^i, \phi_2^i, \dots, \phi_{p-1}^i\}$ is a set of $p-1$ angles, which are required to locate a point on a curve (signal) in a p -dimensional space. The empirical cumulative distribution function (CDF) of the resulting angles $\{\phi_j^i\}_{i=1}^L$ is estimated, denoted by $F(\phi)$, which is to be used as a reference for selecting the finite number of angles (direction vectors) via the NORTA approach.

Generating vectors with arbitrarily specified marginal CDFs and correlations: Given the reference CDF $F(\phi)$ of angles (direction vectors) exhibiting high dynamics, the next stage is to generate the desired finite number of angles from the reference CDF $F(\phi)$. This problem is relatively simple for bivariate data, where a single angle ϕ_1 is sufficient to specify a point on a curve (signal) in a two-dimensional (2D) space: the inverse transform sampling (ITS) method was used to obtain desired angles from a given distribution function $F(\phi_1)$ in [5]. For a general multivariate case, it is not only the joint reference CDF that is important, the correlation among individual variates must also be considered, while choosing a finite number of direction vectors for the MEMD operation.

Hence, we employ the NORTA method to generate direction vectors (projection angles) for MEMD from the reference CDF $F(\phi)$ and the correlation structure between multiple variate ϕ_j . Fig. 2 (left to right) shows samples generated employing NORTA (blue squares) for input bivariate t -distribution and bivariate white Gaussian noise (WGN) (black dots), respectively. Note that the generated samples follow their respective desired distributions.

Let $F_{\phi_1}, F_{\phi_2}, \dots, F_{\phi_{p-1}}$ denote the desired marginal CDF of angles in a p -dimensional space, and ρ_{km} represent the correlation among ϕ_k

and ϕ_m . Then, the NORTA method operates as follows: (i) generate multivariate normal vector $\mathbf{Z} = \{Z_1, Z_2, \dots, Z_{p-1}\}$ with $Z_j \in \mathcal{N}(0, 1)$ and correlations $\rho_{km}(\mathbf{Z}) = \text{Cov}(Z_k, Z_m)$; (ii) compute $X_j = F^{-1}\{\Psi(Z_j)\}$, for $j = 1, 2, \dots, p-1$; where Ψ denotes the standard normal CDF function. Note that $\Psi(Z_j) \in \mathcal{U}(0, 1)$ and, therefore, $X_j = F^{-1}\{\Psi(Z_j)\}$ is similar to the ITS method for the generation of variates from F_{ϕ_j} . The resulting X_j follow the desired marginal CDFs F_{ϕ_j} , and also retain the original correlation structure ρ_{km} . It must be highlighted, however, that in the case of very small number of projection vectors $V \leq 6$, similar to the dynamically sampled BEMD method for bivariate data [5], it is preferable to use the first V number of angles corresponding to the maximum values in the histogram of ϕ , since the NORTA method may not be very accurate for such cases.

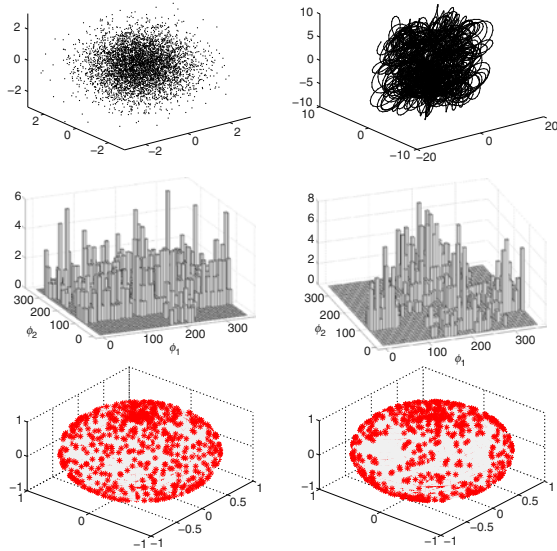


Fig. 3 Projection vectors (shown in bottom row) within DS-MEMD method for input circular WGN and synthetic trivariate data (shown in top row from left to right). Corresponding histograms are shown, respectively, in (middle left) and (middle right).

The key tasks involved in the NORTA method are evaluation of Ψ and F^{-1} , both of which are typically performed via numerical methods. The steps involved in the proposed method, referred to as dynamically sampled MEMD, are listed in Algorithm 1. A MATLAB implementation of NORTA along with its different variations is available in [6] and has been used in the implementation of our method.

Algorithm 1 Dynamically sampled MEMD (DS-MEMD)

- 1: Estimate the curvature κ of input multivariate signal of dimension p , using (3), for all points on the curve in the p -dimensional space.
 - 2: Select the first L points exhibiting top curvature values.
 - 3: Find L set of angles $\{\phi^i\}_{i=1}^L$ corresponding to the selected data points exhibiting high dynamics.
 - 4: **if** $V \leq 6$ **then**
 - 5: Generate $(p-1)$ -dimensional histogram h_ϕ with H number of bins of the angles $\{\phi^i\}_{i=1}^L$.
 - 6: Select the first $V \leq H$ angle set, $\{\phi^v\}_{v=1}^V$ corresponding to the peak values of h_ϕ .
 - 7: **else**
 - 8: Find the marginal CDFs F_{ϕ_j} for $j = 1, 2, \dots, p-1$ from $F(\phi)$.
 - 9: Use NORTA method (as detailed above) to generate desired angle set $\{X_j\}_{j=1}^V$.
 - 10: **end if**
-

Simulation results: Fig. 3 illustrates the operation of DS-MEMD for both the trivariate WGN and the non-circular synthetic combination of sinusoids; their scatter plots are shown, respectively, in Fig. 3 (top left) and (top right). The 2D histograms of the angles exhibiting fast dynamics in both cases (WGN and synthetic data) are shown in Fig. 3 (middle left) and (middle right), respectively. Observe that the WGN histogram is mostly random, whereas the one corresponding to the

synthetic data is concentrated at some specific angles. As a result, in both cases the proposed method selects angle set X , which follows the respective reference distributions shown by their 2D histograms; the resulting angles plotted as points on the surface of a unit sphere are shown in Fig. 3 (lower left) and (lower right).

To demonstrate the ability of the proposed method to improve the accuracy of local mean estimation in standard MEMD, we performed simulations on a trivariate data set as shown in Fig. 4 (visible in black solid line). Fig. 4 shows local mean estimates obtained from standard MEMD (dotted green lines) and the proposed DS-MEMD (dashed red lines) by using $V=4$ projection vectors, using a single sifting iteration. The plots of the estimates of true local mean or ground truth (dashed-dotted blue lines) are also shown, which were obtained by taking a very large number of direction vectors ($V=1024$) in standard MEMD. It can be noticed that the DS-MEMD provided more accurate estimates of the local mean compared with standard MEMD, as evident from the close proximity of its local mean to the ground truth, especially at points A and B which exhibit high dynamics.

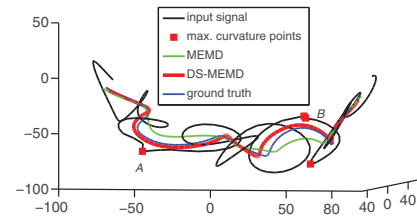


Fig. 4 Comparison of local mean estimation between MEMD (dashed green line) and proposed DS-MEMD (dotted blue line) for natural input trivariate signal (solid black line) for $V=4$.

Conclusion: We have presented a general data-driven scheme for choosing suitable direction vectors (angles) in multidimensional spaces for the computation of local mean within MEMD, by employing the Menger curvature measure and the NORTA variate-generation method. The proposed DS-MEMD method has been shown to yield more accurate local mean estimates than those obtained using MEMD, for a very small number of projection vectors. Given that the MEMD is computationally very expensive, which restricts its real-world application, the proposed method with its lower computational needs is expected to alleviate this problem and find new real-world and real-time applications.

© The Institution of Engineering and Technology 2015

Submitted: 24 April 2015 E-first: 3 November 2015

doi: 10.1049/el.2015.1176

One or more of the Figures in this Letter are available in colour online.

N. Rehman, K. Naveed and M.W. Safdar (*Department of Electrical Engineering, COMSATS Institute of Information Technology, Islamabad, Pakistan*)

✉ E-mail: naved.rehman@comsats.edu.pk

S. Ehsan and K.D. McDonald-Maier (*School of Computer Science and Electronic Engineering, University of Essex, Colchester, United Kingdom*)

References

- 1 Huang, N.E., Shen, Z., Long, S., *et al.*: ‘The empirical mode decomposition and Hilbert spectrum for non-linear and non-stationary time series analysis’, *Proc. R. Soc. Eng. A*, 1998, **454**, pp. 903–995
- 2 Rilling, G., Flandrin, P., Goncalves, P., *et al.*: ‘Bivariate empirical mode decomposition’, *IEEE Signal Process. Lett.*, 2007, **14**, pp. 936–939
- 3 Rehman, N., and Mandic, D.P.: ‘Multivariate empirical mode decomposition’, *Proc. R. Soc. Eng. A*, 2009, **466**, pp. 1291–1301
- 4 Ahrabian, A., Rehman, N., and Mandic, D.P.: ‘Bivariate empirical mode decomposition for unbalanced real world signals’, *IEEE Signal Process. Lett.*, 2013, **20**, pp. 245–248
- 5 Rehman, N., Safdar, M.W., Rehman, U., *et al.*: ‘Dynamically-sampled bivariate empirical mode decomposition’, *IEEE Signal Process. Lett.*, 2014, **21**, pp. 857–861
- 6 Henderson, S.G.: ‘Software’, accessed May 2014, <http://www.people.orie.cornell.edu/shane/NORTASoftware.html>