Extreme Mechanics Letters 6 (2016) 60–67

Contents lists available at ScienceDirect

Extreme Mechanics Letters

journal homepage: www.elsevier.com/locate/eml

Filtering noise from time series by means of elastic wave propagation and flux-corrected transport in a discrete medium

Juhua Zhang

Department of Biomedical Engineering, Beijing Institute of Technology, Beijing, 100081, China

ARTICLE INFO

Article history:
Received 12 June 2015
Received in revised form 26 September 2015
Accepted 6 October 2015
Available online 2 December 2015

Keywords:
Noise reduction
Time series
Filtering noise
Finite element
Flux-corrected transport
Elastic wave

ABSTRACT

We propose an efficient nonparametric method, named FEFCT, to eliminate noise from time-series data. We analyze noised time series data as a discrete wave packet that travels in one-dimensional finite-element grids discretized from a one-dimensional elastic medium. Then the techniques of flux-corrected transport are applied to filter the signal. The FEFCT is fast, accurate and performs well without Gaussian distribution assumption of noise. More importantly, when FEFCT is applied to infer regulatory network from human cell cycle data, it dramatically improves the association between genes, indicating FEFCT has captured the real signal in the complex biological system. Including but not limited to gene expression data, FEFCT can be easily extended to any other types of time-series data, which will be extremely important in the era of big data.

1. Introduction

Time series analysis is an important matter in many fields of science and engineering for pattern recognition, system identification, parameter estimation, and so on. When data is noisy, these activities may be impeded by noises. Therefore, noise reduction in experimental time series is an important issue. In the era of big data, this issue has become increasingly significant. Noise reduction is to remove noises effectively from data while preserving the real information as much as possible. The methods of filtering noise from time series may be classified as model based and model free. In the former case, the typical method, the Kalman filter [1] addresses the general problem by trying to estimate the state of a discrete-time system governed by a linear stochastic difference equation. In the latter case, wavelet transform [2–4] is the most widely adopted method. By adaptively thresholding the wavelet coefficients denoising operations can be performed. Most methods often assume that noises satisfy Gaussian distribution with zero mean. However, many recent studies [5,6] indicated that real data noise usually did not follow a Gaussian distribution. In this paper, we propose a conceptually new noise filter algorithm, which is based on the characteristic of elastic wave and computational fluid dynamics methods. The experiments show that the algorithm has the outstanding capability to reduce the noise and capture the real signal in the complex biological time series.

2. Theory of FEFCT

The proposed method is based on the elastic wave packet propagating in the one-dimensional elastic solid. There is no dispersion in this case; the phase and group velocity are always the same. Generally, the motion of the elastic solid Ω with the mass density $\rho = 1$ and elastic modulus $\mathcal{K} = 1$, satisfies

$$\frac{dv}{dt} = \frac{d\sigma}{dx} \tag{1}$$

E-mail address: jhzhang@bit.edu.cn.

http://dx.doi.org/10.1016/j.eml.2015.10.002
2352-4316/

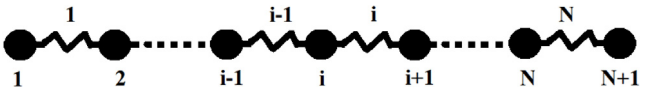

Fig. 1. The schematic diagram. A continuous body is discretized into the finite element system.

where v is the velocity, t is time, x is spatial position, and σ is the stress, in the context of this paper, substituted by the signal, i.e. the gene expression level.

2.1. Finite element method

The Lagrangian finite-element method is used to obtain the approximate solution of $\sigma(x, t)$ in Eq. (1). A one-dimensional elastic body Ω is uniformly divided into n segments, i.e. elements with length $L_i = 1,\ i = (1, \ldots, n)$. The nodes locate at two ends of each element, and the mass of the body is concentrated at the nodes. In this way, the continuous body is subdivided into the spring-partial system as shown in Fig. 1. In this system, the springs have zero mass, and the particles have zero volume. All nodes are labeled from 1 to n+1, and all elements are labeled from 1 to n. The element i is located between node i and node $i + 1$, where $i \in \{1, 2, \ldots, n\}$. We set the velocity of all nodes $v_i = 0$, where $i \in \{1, 2, \ldots, n + 1\}$ at time $t = 0$. Assume that in each experiment the expression level of each gene has d time sampling points $\{e_1, e_2, \ldots, e_d\}$, the elements labeled from $\frac{n}{2} + 1$ to $\frac{n}{2} + d$ are assigned the initial stress value $e_1, e_2, \ldots, e_d$ one by one. The initial stress in other elements is set to 0. Therefore, the time series of gene expression in the cell cycle is now transformed to the initial stress in the discrete linear elastic system.

Assume the stress $\sigma_i(t)$ in the ith element is known at time t, the stress $\sigma_i(t + \Delta t)$ at time $t + \Delta t$ can be obtained by time integration. The nodal acceleration a_i for a node i is

$$a_i = \frac{F_i}{m_i} = \frac{\sigma_{i-1} - \sigma_i}{m_i}, \quad (i = 2, \ldots, n),$$
$$a_1 = -\frac{\sigma_1}{m_1}, \qquad a_{n+1} = \frac{\sigma_n}{m_{n+1}}, \tag{2}$$

where $m_{i(i=2,\ldots,n)} = \rho L_i = 1$, $m_1 = m_{n+1} = 0.5$, is the nodal mass, F_i is the nodal force, and σ_i is the stress in the ith element. The central difference scheme is employed to obtain the trial nodal velocity v_i as follows,

$$v_i^*\left(t + \frac{\Delta t}{2}\right) = v_i\left(t - \frac{\Delta t}{2}\right) + a_i \Delta t. \tag{3}$$

To obtain the nodal velocities $v_i(t + \frac{\Delta t}{2})$ at time $t + \frac{\Delta t}{2}$, the trial nodal velocities are modified by the flux-corrected transport elucidated later. The nodal coordinate x_i at time $t + \Delta t$ is calculated by the nodal velocities,

$$x_i(t + \Delta t) = x_i(t) + v_i\left(t + \frac{\Delta t}{2}\right)\Delta t, \tag{4}$$

where the subscript $i = 1, \ldots, n + 1$. The integration time step Δt must satisfy the Courant condition

$$\Delta t = \eta \frac{\Delta x}{c}, \tag{5}$$

where Δx is the element length, and $c = \sqrt{\frac{\mathcal{K}}{\rho}} = 1$ is the wave speed of the material. The Courant coefficient η must be less than 1.0, and it is set to 0.75 in this work. The stress in the ith element at time $t + \Delta t$ satisfies Hooke's law

$$\sigma_i(t + \Delta t) = \sigma_i(t) + \mathcal{K}\Delta\epsilon_i = \sigma_i(t) + \Delta\epsilon_i, \tag{6}$$

where the strain rate $\dot{\epsilon}_i$ and strain increment $\Delta\epsilon_i$ in the ith element are calculated by

$$\Delta\epsilon_i = \dot{\epsilon}_i \Delta t = \frac{v_{i+1} - v_i}{x_{i+1} - x_i}\Delta t. \tag{7}$$

2.2. Flux-corrected transport

The nodal velocities calculated by Eq. (3) are modified by diffusion and anti-diffusion. The diffusive fluxes are calculated in accordance with the nodal velocities at time t,

$$f_i^0 = m_{i+1}v_{i+1}^0 - m_i v_i^0. \tag{8}$$

The diffusion is applied to the trial nodal velocity to get a new trail velocity as follows,

$$\widetilde{v}_i = v_i^* + \xi\frac{f_i^0 - f_{i-1}^0}{m_i}, \tag{9}$$

where $\xi \sim 0.5\eta(1 - \eta) = 0.093$. The influence of the diffusion must be automatically eliminated by anti-diffusive step. The anti-diffusive flux is obtained by the new trial nodal velocity,

$$f_i^1 = m_{i+1}\widetilde{v}_{i+1} - m_i\widetilde{v}_i. \tag{10}$$

Before one applies the anti-diffusive step to the new trial velocity, the anti-diffusive flux must be selected to maintain suitable residual diffusion or remove the diffusion,

$$f_i^1 = s \cdot max(0, min[s \cdot f_{i-1}^1, s \cdot f_i^1, s \cdot f_{i+1}^1]), \tag{11}$$

where $s = sign(f_i^1)$. The anti-diffusion is applied to the second trail velocity to obtain the updated nodal velocity at time $t + \frac{\Delta t}{2}$:

$$v_i = \widetilde{v}_i - \xi\frac{f_i^1 - f_{i-1}^1}{m_i}. \tag{12}$$

One-time cycle is finished from Eq. (2) to Eq. (12). Repeat the time cycle till arriving at the end time. The integration time (i.e. wave propagating time) step is controlled by the Courant condition Eq. (5), so the least integration time steps are $w/(2 \times 0.75)$. The pseudo code, which rigorously followed the method, is summarized in Algorithm 1 in Table 1.

3. Principles of the method

Shock waves are strong discontinuity. When shock wave propagation is simulated in linear elastic media with