“Minimum average risk” as a new peak-detection algorithm applied to myofibrillar dynamics

S.Yu. Sokolov, A.A. Grinko, A.V. Tourovskaya, F.B. Reitz, O. Yakovenko, G.H. Pollack,*, F.A. Blyakhman

A Department of Physics, Ural State University, 51 Lenin Street, Yekaterinburg 620083, Russia
b Department of Bioengineering, WD-12, PO Box 357962, University of Washington, Seattle, WA 98195, USA

Received 22 May 2001; received in revised form 26 July 2002; accepted 29 July 2002

Abstract

We present a new peak-detection algorithm based on the method of “minimum average risk” proposed by Kolmogorov and developed for signal processing in various fields. In this method, translations of features within a signal scan are quantified by minimizing the integrated pointwise product of each scan relative to the first derivative of the immediately previous scan. We have adapted this method for use in a new algorithm to monitor dynamic changes of sarcomere length in single myofibrillar sarcomeres of striated muscles, but the algorithm can also be used more generally for peak localization. We find that this method results in sub-nanometer precision and higher signal-to-noise ratio than current methods. At an equal noise level, the RMS deviation of the minimum average risk algorithm was 1.3 times lower than that of the center of mass method with modeled data and 3–4 times lower with actual data.

Keywords: Peak localization; Myofibril; Sarcomere length

1. Introduction

Increasingly, the field of biomechanics is making forays into the nanoscale, e.g. optical traps are used to stretch individual titin molecules [1] and atomic force microscopes are used to measure intermolecular binding forces [2]. In cases such as these, well-established methods such as electronic feedback or optical beam deflection suffice to transduce the measurement of interest. Data pertaining to changes between the monitored points, however, is less readily quantified. For example, what part of the stretched titin molecule is unfolding, and when?

Monitoring the contraction dynamics of individual sarcomeres presents just this problem, on an initially more approachable scale. Sarcomeres are the contiguous subunits that compose myofibrils, the contractile organelles of muscle cells (cf. Fig. 1, inset; Fig. 2a). Previous methods for quantifying small (nanometer-sized) sarcomeric length variations within individual isolated myofibril include simple curve-fitting [3] or tracking feature cen-
troids computed from digitized images [4]. We present an improved peak-detection algorithm and demonstrate the tracking-over time of individual sarcomere length dynamics with sub-nanometer precision.

The new algorithm is based on the method of minimum average risk (MAR) proposed by Kolmogorov [5,6] and further developed for signal processing in various other fields [7–9]. As implemented herein, a 2-sarcomere-wide subsection of each successive digitized line scan of a myofibril is selected to bracket a given A-band. This section is then multiplied pointwise by the first derivative of the corresponding section of the immediately previous scan. This procedure is repeated for varying lateral translations, and the translation yielding the minimum integrated product is taken as the shift in A-band position. The distance between the adjacent A-bands yields sarcomere length.

The inherent tendency of this technique to weight edge features and exclude low-frequency drifts results in a high signal-to-noise ratio, and we anticipate that it will be general in its applicability.

2. Methods

2.1. Experimental apparatus

The apparatus used for experiments on isolated myofibrils is schematized in Fig. 1 (cf. [10]). Individual myofibrils are isolated from bumblebee flight muscle by fine mincing followed by manual selection under phase contrast microscopy of the occasionally occurring single myofibrils. Micropositionable levers are used to pick up individual myofibrils and wrap their ends onto the levers, providing a suspended length of ~20 sarcomeres.

Stretch-release protocols may then be applied while sarcomere dynamics are monitored. During a 30–50 s cycle in which the myofibril is stretched and released, intensity line scans are acquired every 50 ms.

These intensity profiles are then provided as input to custom Labview® software for peak motion determination and calculation of resulting sarcomere length dynamics (National Instruments, Austin, TX). Labview® 4.0 for the Macintosh was used to develop this software. Labview® 4.0 or
later would be required to use the software, which is available upon request from the authors. Typical resulting scans are shown in Fig. 2.

2.2. Center of mass algorithm

As shown in Fig. 2, the A-bands appear as peaks in the intensity profiles (actually corresponding to a drop in transmitted intensity). In the center of mass algorithm (CM; [4]), a lower boundary is set at the signal mean minus 30–70% of the root mean square (RMS) signal value. The centroid of each peak area above this boundary is calculated, and position differences between adjacent centroids yield traces of individual sarcomere length as a function of time, the dynamics of which may then be analyzed.

2.3. MAR algorithm

The MAR algorithm as presented here operates on repeated scans of an intensity peak, precisely quantifying peak movement between scans. Peak position information is thus available. In our application, the positions of peaks corresponding to adjacent myofibrillar A-bands are determined, and their difference yields a measure of sarcomere length.

In practice, the MAR algorithm is implemented by finding the optimal position of a scan relative to the first derivative of the immediately previous scan. Optimal registration occurs when the integrated pointwise product of these functions is minimized.

The algorithm originates from a manipulation of the conventional risk function

\[ R(X|Y) = \int C(X_0 - X)p(X_0|Y)\,dX_0, \quad (1) \]

an integral over all guesses \( X_0 \) of the true peak position, where \( Y \) represents measured data, \( X \) represents estimations of the actual value underlying the data, \( C(X_0 - X) \) is a losses function (typically a constant minus the Dirac delta function [11]), and \( p(X_0|Y) \) is a conventional (a posteriori) density of the value probability of the \( X_0 \) parameter. The average risk function is a probability-weighted average of the conventional risk function over all possible data \( Y \):

\[ R(X) = \int R(X|Y)p(Y)\,dY. \quad (2) \]

The form of the conventional risk function essentially that of an inverted function \( p(X_0|Y) \). The point \( X \) of minimum \( R(X|Y) \) thus corresponds to the point \( X_0 \) of maximal \( p(X_0|Y) \).

As

\[ p(X_0|Y) = Kp(Y|X_0)p(X_0), \quad (3) \]
where $K$ is a constant and $p(X_0)$ is considered to be uniform in the absence of a prior information, the point of maximal $p(Y|X_0)$—the likelihood function—corresponds to the point of maximal $p(X_0|Y)$. The distribution $p(Y|X_0)$ is available from the obtained data $Y$ as described below.

As the sources of noise are numerous (mechanical and thermal noise, defects and soiling of the optic system, inhomogeneity of photodiodes on the array, errors in quantization and sampling of the analog-to-digital converter, temporal and temperature instability of the electronic equipment parameters, etc.), by the central limit theorem [12], the resulting noise samples are well modeled by white Gaussian noise. The multidimensional probability density of such noise is

$$p(Y|X_0) = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^M \exp\left\{-\sum_{i=1}^M (Y(t_i) - S(t_i, X_0))^2 \over 2\sigma^2\right\},$$

where $t_i$ is a discrete spatial coordinate, $S(t_i, X_0)$ is the signal in the absence of noise for a given peak position $X_0$, $M$ the number of points per scan, and $\sigma$ the RMS noise. The maximum of this distribution may be found by setting the derivative of its logarithm with respect to $X_0$ equal to zero, giving the equation

$$\sum_{i=1}^M (Y(t_i) - S(t_i, X_0)) \frac{\delta S(t_i, X_0)}{\delta X_0} = 0,$$  

(5)

or

$$\sum_{i=1}^M Y(t_i) \frac{\delta S(t_i, X_0)}{\delta X_0} = \sum_{i=1}^M S(t_i, X_0) \frac{\delta S(t_i, X_0)}{\delta X_0}.$$  

(6)

The point $X_0$ at which this equation is satisfied is determined more precisely via fitting the discrete data with a fifth-order polynomial and interpolating.

### 2.4. Signal modeling

To evaluate the algorithms in question, a model scan was generated using the positive half-cycle of the sinusoidal equation

$$Y_i(t, k) = A \sin\left(\frac{(t - X_{ik})\pi}{T_i + \Delta T_{ik}}\right) + \sigma n(t)$$  

(7)

for each “sarcomere”, where subscript $i$ denotes an individual sarcomere, $k$ denotes an individual scan, $Y_i(t, k)$ is the signal corresponding to each sarcomere scan, $A$ the amplitude of the signal, $X_{ik}$ the spatial coordinate of the beginning of each sarcomere, $T_i$ represents initial sarcomere length, $\Delta T_{ik}$ the length change of each sarcomere, $\sigma$ the noise RMS and $n(t)$ the white Gaussian noise of RMS $= 1$.

### 3. Results and discussion

RMS and peak-to-peak noise on the outputs of the MAR and CM algorithms are shown in Table 1. In the table, “static” model data reflect the contribution of the white noise added to the model function. “Dynamic” data were obtained by computationally including a 10% extension and release of the model “myofibril”. “Actual” data ($N = 26$, shown as ± standard error of the mean) were obtained by scanning a myofibril held stationary on the experimental apparatus. Model sarcomere length is taken to be 3300 nm for the purpose of expressing noise percentiles in practical terms. The model data represent 300-fold averaging, and the signal-to-noise ratio is $A1/\sqrt{2s}$.

Figs. 3 and 4 show graphically the noise reduction of as much as fourfold by the MAR relative to the CM method. In Fig. 3, a stepwise translation ramp (trace i) is applied to a myofibril as a test signal not dissimilar from those obtained from biological specimens [1,13,14]. Resulting sarcomere length changes are then analyzed using the CM (trace ii) and MAR (trace iii) algorithms. The signal-to-noise difference is readily qualitatively visible, as further illustrated by the figure inset, in which the analyzed traces are shown superimposed.

Fig. 4 shows the tracking of the length of a single cardiac myofibrillar sarcomere undergoing spontaneous oscillatory contractions. Cardiac myofibrils maintained at an intermediate activation level have been found to display periodic, wave-like contractile oscillations that propagate...
Table 1
Relative performance of MAR and CM algorithms

<table>
<thead>
<tr>
<th></th>
<th>Model data, static</th>
<th>Model data, dynamic</th>
<th>Actual data, static</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MAR</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMS noise (nm)</td>
<td>0.81</td>
<td>0.88</td>
<td>0.55±0.03</td>
</tr>
<tr>
<td>Peak-to-peak noise (nm)</td>
<td>3.8</td>
<td>5.04</td>
<td>2.40±0.02</td>
</tr>
<tr>
<td><strong>CM</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMS noise (nm)</td>
<td>1.04</td>
<td>1.15</td>
<td>2.15±0.08</td>
</tr>
<tr>
<td>Peak-to-peak noise (nm)</td>
<td>5.3</td>
<td>6.8</td>
<td>9.57±0.53</td>
</tr>
</tbody>
</table>

Computed static and dynamic functions with added white noise serve as model myofibril data (see Section 2.4). “Actual” data are obtained by scanning a stationary myofibril.

Fig. 3. Sarcomere length vs. time with stepwise ramp. Sarcomere length as tracked by (A) the minimum risk algorithm and (B) the center of mass method.

Fig. 4. Spontaneous oscillations of cardiac myofibril. A suspended cardiac myofibril maintained at an intermediate activation level is projected onto the linear photodiode array and monitored as it oscillates (cf. [15]). Tracking of single sarcomere length by (a) the center of mass and (b) the MAR algorithms is shown.

along the myofibril [15]. Data with cardiac myofibrils were obtained as above (Fig. 1). The periodicity of the contractions remains obscure in the CM-analyzed data, while the MAR-analyzed data appear clearly sinusoidal.

It is instructive to compare the MAR algorithm with the CM algorithm, as there are several fundamental aspects of the MAR that contribute to its superior performance. First, noise on slowly changing portions of a signal contribute more to
feature localization error than that on rapidly changing portions [16,17]. The MAR algorithm takes advantage of this differential noise sensitivity by weighting the signal by its first derivative. Note that the advantage of the MAR algorithm was less marked with the model data used in part because the model data had relatively smooth edges, and thus less weighting of feature edges by their derivative. The comparison between model and actual data also point up the fact that the MAR algorithm does use all available data points, as opposed, for example, to the CM algorithm which involves a thresholding step. The model data are fully used by both algorithms, while the CM algorithm discards part of the actual data upon thresholding, again leading to a more marked difference between algorithms using the actual data.

The exact nature of the mechanism of muscle contraction remains controversial. Large-scale preparations are too far removed from molecular detail, while molecular preparations have lost the spatial organization of intact tissue. With the aid of the improved algorithm outlined herein, we have been able to make measurements at the level of the intact organelle and subparts thereof, and even to make forays into the study of molecular scale events, including spontaneously generated steps similar to those of Fig. 3[13,14].

The precise quantification of the spontaneous oscillation of each individual cardiac sarcomere as demonstrated above opens the door to a host of questions. Do adjacent sarcomeres exhibit cooperativity? What calcium dynamics accompany such length variations? How do differing changes in sarcomere length under the same tension modify our understanding of the length–tension relationship?

We anticipate that this high-resolution method may be profitably employed in these and other applications in which intra-endpoint positional information is desired. Notably included in this category are force vs. length measurements in which the compliance of sample-apparatus connections is in question; these regions may be excluded by measuring displacements within the sample.

References