Available online at www.sciencedirect.com
ScienceDirect
Computers and Mathematics with Applications 53 (2007) 276–286
An International Journal
computers & mathematics
with applications
www.elsevier.com/locate/camwa

Computational aspects of Worm-Like-Chain interpolation formulas

Ray W. Ogden[a,*], Giuseppe Saccomandi[b], Ivonne Sgura[c]

[a] *Department of Mathematics, University of Glasgow, Glasgow G12 8QW, Scotland, UK*
[b] *Dipartimento di Ingegneria dell'Innovazione, Sezione di Ingegneria Industriale, Università degli Studi di Lecce, 73100 Lecce, Italy*
[c] *Dipartimento di Matematica, Università degli Studi di Lecce, 73100 Lecce, Italy*

Received 13 September 2005; accepted 27 February 2006

Abstract

We show that the polynomial Worm-Like-Chain (WLC) interpolation formula introduced in [C. Bouchiat, M.D. Wang, J.-F. Allemand, T. Strick, S.M. Block, V. Croquette, Estimating the persistence length of a Worm-Like Chain molecule from force–extension measurements, Biophys. J. 76 (1) (1999) 409–413] in order to improve data fitting with respect to the WLC interpolation formula in [C. Bustamante, J.F. Marko, E.D. Siggia, S. Smith, Entropic elasticity of lambda-phage DNA, Science 265 (1994) 1599. Technical comment] is not unique. Ill-conditioning of the over-determined linear system associated with interpolation of synthetic data from [C. Bouchiat, M.D. Wang, J.-F. Allemand, T. Strick, S.M. Block, V. Croquette, Estimating the persistence length of a Worm-Like Chain molecule from force–extension measurements, Biophys. J. 76 (1) (1999) 409–413] is highlighted. Moreover, if the coefficients in the associated polynomial correction are considered as free parameters in the least squares fitting procedure for actual experimental data, then more than one solution with a low residual can be identified. For these reasons we propose modification of the interpolation formula in [C. Bustamante, J.F. Marko, E.D. Siggia, S. Smith, Entropic elasticity of lambda-phage DNA, Science 265 (1994) 1599. Technical comment] so that, for moderate extensions, quadratic contributions are included and no additional parameters are required. We show that relative errors for the fit of synthetically generated data are less than 2% and for actual data they are comparable with and sometimes better than those obtained by using polynomial WLC interpolation formulas.

Keywords: DNA data fitting; Worm-like chain; DNA elasticity; Least squares fitting; Ill-conditioning

1. Introduction

In [1] a procedure for studying the exact WLC model for a stretched DNA molecule was proposed. By following standard quantum mechanics arguments, the authors considered the Hamiltonian associated with the system defined in terms of the dimensionless parameter $y = FA/k_BT$, where F is the force applied, A is the macromolecule *persistence length*, $k_B = 1.38066 \times 10^{-23}$ J/K (1.38066×10^{-5} μm pN/K) is the Boltzmann constant, and T the absolute temperature.

Then, by first-order perturbation theory, they find that the relative elongation $z = x/L$ of the molecule with respect to the total contour length L can be found by studying the first (ground state) eigenvalue and eigenfunction

* Corresponding author. Tel.: +44 1413304550; fax: +44 1413304111.
E-mail addresses: rwo@maths.gla.ac.uk (R.W. Ogden), giuseppe.saccomandi@unile.it (G. Saccomandi), ivonne.sgura@unile.it (I. Sgura).

0898-1221/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2006.02.024

of the Hamiltonian. The boundary-value problem for the ordinary differential equation associated with the eigenvalue equation was solved numerically using a shooting-like method, and by standard iteration for the resulting nonlinear problem *synthetic* normalized force–extension data of the exact WLC were generated. Moreover, in [1] the authors proposed an analytical force–extension formula for the WLC model that, by interpolation on these data, gives a better fit than the original interpolation formula, namely

$$F_0(z) := \frac{k_B\, T}{A}\left(\frac{1}{4(1-z)^2} + z - \frac{1}{4}\right), \quad 0 \le z < 1, \tag{1}$$

proposed in [2]. In this pioneering paper the mathematical expression of (1) was proposed in order to describe the asymptotic data behaviour for $z \to 1$ as a second-order singularity, and the behaviour for small extensions $z \to 0$ as Hookean (linear).

As emphasized in [1], with respect to *synthetic data* this formula yields relative errors of about 10% at moderate extensions (in a neighbourhood of $z = 0.5$) and it was asserted that the fitting procedure for this model yields an overestimate in the value of the persistence length A with a typical error of about 5%. For this reason, if $E(z)$ represents the *exact* WLC model, it was proposed in [1] to modify the interpolation formula (1) by adding a polynomial correction, leading to

$$F_7(z) := \frac{k_B\, T}{A}\left(\frac{1}{4(1-z)^2} + z - \frac{1}{4} + \sum_{i=2}^{7} \alpha_i\, z^i\right), \quad 0 \le z < 1, \tag{2}$$

so that the residual $R(z) = E(z) - F_0(z)$ can be expressed as a seventh-order polynomial.

We are interested in the interpolation formula (2) because in the last decade it has been used in many biological papers as a model for force–extension experimental data obtained from DNA micro-manipulations in order to estimate the persistence length A and the contour length L of the macromolecule using nonlinear least squares fitting techniques. (For molecules other than DNA see, for example, [3,4], which are concerned with collagen and elastin, respectively. For further discussion of the phenomena in DNA see [5,6].)

The aim of this paper is to study some computational aspects associated with the WLC interpolation formula (1) and its modifications. We emphasize that the model (2) may be ill-conditioned and ill-posed. This implies, in particular, that if the coefficients in the interpolation formula $F_7(z)$ in (2) are considered as free parameters, more than one solution with a low residual can be identified in the fitting procedure. This problem is often overlooked in applications because the parameters introduced in the additional polynomial part of (2) are fixed at the specific values considered in [1]. However, there is no *a priori* justification for fixing these parameters. We propose an alternative modification of the basic WLC formula (1) that avoids the introduction of additional parameters. This new model is an improvement on (1) and, moreover, it is not ill-conditioned. In this way it is possible to achieve a good fit to experimental data and to confirm the persistence and contour lengths found using (2). Clearly this is a heuristic confirmation, but it is hoped that by emphasizing possible mathematical problems and introducing a new feasible interpolation formula for the WLC model a better understanding of the status of the various interpolation formulas may be obtained.

In the following we refer to (1) as formula WLC_0 and to (2) as formula WLC_7. We remark that our criticisms are not directed at the statistical mechanics of the WLC model (see [7,8]) but towards the interpolation formulas associated with this model.

2. Basic facts

To use the formula (2) as a general purpose model, the coefficients $\alpha_i, i = 2, \ldots, 7$ were calculated in [1] by imposing interpolation conditions on the synthetic data $(\mathbf{z}, \mathbf{y})$ generated, where $\mathbf{z} = [z_1, \ldots, z_m]$ and $\mathbf{y} = [y_1, \ldots, y_m]$ are vectors containing the data, m being the size of the data set. These conditions yield a linear system in the unknowns $\alpha_i, i = 2, \ldots, 7$. If the numerical values of $(\mathbf{z}, \mathbf{y})$ are as reported in [1], this system is given by

$$\mathsf{M}_b \mathbf{p} = \mathbf{b}, \tag{3}$$

where M_b is an $m \times 6$ Vandermonde-like matrix (see, e.g., [9]), $\mathbf{b} = [b_1, \ldots, b_m]$, with

$$b_i = y_i - [1/(1-z_i)^2 + 4z_i - 1]/4, \quad i = 1, \ldots, m,$$

and $\mathbf{p} = [\alpha_2, \alpha_3, \ldots, \alpha_7]$ is the vector of the parameters to be determined.