# Testing for the heterogeneity of complex fluid samples in nanobiophysics

Kui Zhang (Advisor: Gustavo Didier)
*Department of Mathematics, Tulane University, New Orleans - LA, USA*

## I. SCALING LAWS IN NANOBIOPHYSICS

Improvements in light microscopy, fluorescence techniques, nanoparticle synthesis and high-speed video have ushered in a flurry of experimental activity [1]. Single particle tracking has become a common tool in many scientific areas, such as colloid physics [2], the study of nanobiophysical systems, both *in vivo* and *in vitro* [7], and the microrheology of complex fluids [3–6]. In particular, the latter is the application that motivates the research in this dissertation.

Of primary concern in the analysis of particle path data is the mean squared displacement (MSD) $\langle X^2(t)\rangle$, where $X$ is the tracer particle's position. A basic dynamic characterization of the latter is given by the relation

$$\langle X^2(t)\rangle \propto \theta t^{\alpha}, \quad \theta, \alpha > 0, \quad \boldsymbol{\xi} := (\log_2 \theta, \alpha), \qquad (1)$$

where $\theta$ and $\alpha$ are called, respectively, the diffusivity constant and the diffusion exponent. The parameter value $\alpha = 1$ corresponds to classical diffusion. If $\alpha \neq 1$, the process $X$ is called an anomalous diffusion, as statistically observed in most microrheological experiments.

The dominant statistical technique in the biophysical literature for estimating the parameters $\theta$ and $\alpha$ is based on the so-named sample pathwise mean squared displacement ($\widehat{\text{MSD}}$). For a tracer bead sample path $X(j)$, $j = 1, \ldots, n$, the statistic

$$\overline{\mu}_2(h) := \frac{1}{n-h}\sum_{j=1}^{n-h}\{X(j+h) - X(j)\}^2 \qquad (2)$$

is the $\widehat{\text{MSD}}$ at $h$, i.e., the statistical counterpart of the MSD $\mu_2(h) = \langle X^2(h)\rangle$. An estimator $(\widehat{\log_2 \theta}, \widehat{\alpha})$ is obtained by means of the linear regression

$$\log_2 \overline{\mu}_2(h_k) = \log_2 \theta + \alpha \log_2(h_k) + \varepsilon_k, \quad k = 1, \ldots, m, \qquad (3)$$

possibly over several independent particle paths, where $\{\varepsilon_k\}_{k=1,\ldots,m}$ is a random vector with an unspecified distribution. Plots of $\widehat{\text{MSD}}$ curves as a function of the lag $h$, often on a log-log scale, are widely reported as part of diffusion analysis (e.g., [8, 9]).

One key interest in the microrheological study of complex fluids is to test for heterogeneity, both within and between fluid samples, by means of the observed diffusion of embedded tracer particles. However, the biophysical literature lacks reliable asymptotic results for the distribution of the $\widehat{\text{MSD}}$. The purpose of this dissertation is to provide a broad framework for the testing of fluid heterogeneity based on accurate mathematical results.

## II. CONTRIBUTIONS

### II.1. Asymptotic distribution of the pathwise sample MSD

Consider the random vector

$$\Big(\overline{\mu}_2(h_1), \ldots, \overline{\mu}_2(h_m)\Big). \qquad (4)$$

In [10], we show that for a broad class of Gaussian, stationary increment processes the convergence in distribution of the $\widehat{\text{MSD}}$ occurs at different rates, and the limiting distribution may be Gaussian and non-Gaussian (Rosenblatt-type), all depending on the value of the diffusivity exponent $\alpha$. More precisely,

$$\left(\frac{n-h_k}{\eta(n-h_k)\zeta(h_k)}\big(\overline{\mu}_2(h_k) - \big\langle X^2(h_k)\big\rangle\big)\right)_{k=1,\ldots,m} \stackrel{d}{\to} \mathbf{Z}, \qquad (5)$$

where $n \gg h_k \to +\infty$, $k = 1, ..., m$. The rates of convergence in (5) are given by

$$\begin{cases} 0 < \alpha < 3/2: & \eta(n) = \sqrt{n},\ \zeta(h) = h^{\alpha+1/2}; \\ \alpha = 3/2: & \eta(n) = \sqrt{n\log(n)},\ \zeta(h) = h^2; \\ 3/2 < \alpha < 2: & \eta(n) = n^{\alpha-1},\ \zeta(h) = h^2. \end{cases} \qquad (6)$$

### II.2. WLS estimator of $\alpha$ with bias correction

To test heterogeneity more accurately, we would like to increase the performance of $\widehat{\alpha}$, namely, reduce it variance and bias. Research shows that the variances of $\log_2 \overline{\mu}_2(h_k)$ increases exponentially with respect to $h_k$. For this reason, we consider a weighted, rather than ordinary, least squares approach (WLS vs OLS) to the linear model in (3). From the fact that [11]

$$\langle \log_2 \overline{\mu}_2(h)\rangle \neq \log_2 \langle \overline{\mu}_2(h)\rangle = \alpha \log_2(h) + \log_2(\theta),$$

we correct the bias of each term $\log_2 \overline{\mu}_2(\cdot)$ by subtracting the expectation of the second term of the Taylor expansion of $\log_2 \overline{\mu}_2(h)$. This idea is presented below in the form of pseudo-code.

**Input**: observed particle paths
**Step 1**: compute the estimator $\widehat{\alpha}_{\text{OLS}}$
**Step 2**: correct the bias of $\log_2 \overline{\mu}_2(\cdot)$ and choose weights $\mathbf{w}$ based on $\widehat{\alpha}_{\text{OLS}}$
**Step 3**: compute the estimator $\widehat{\alpha}_{\text{WLS}}$ and $\widehat{\theta}_{\text{WLS}}$

The comparative performance of the WLS and OLS estimators is illustrated in FIG 1 and 2 in terms of mean squared error, variance, and bias. The combination of finite-sample correction and asymptotic results

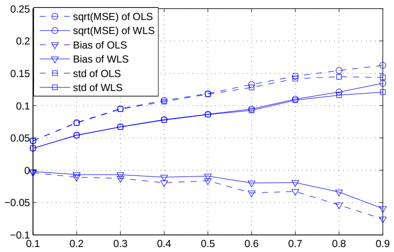


FIG. 1. $\widehat{\alpha}_{\text{OLS}}$ v.s. $\widehat{\alpha}_{\text{WLS}}$

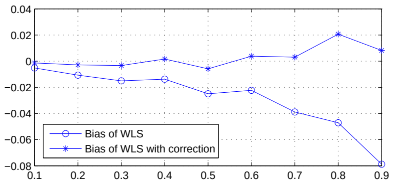


FIG. 2. $\widehat{\alpha}_{\text{WLS}}$ with bias correction

also leads to accurate confidence intervals and hypothesis test statistics by means of an approximation of $\sigma^2_{\widehat{\alpha}_{\text{WLS}}}$ (see FIG 3).

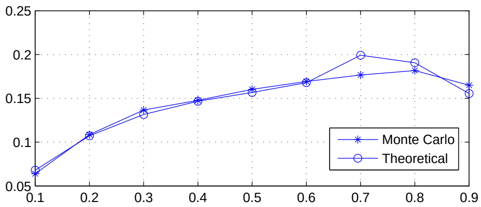


FIG. 3. standard deviation of $\widehat{\alpha}_{\text{WLS}}$: Monte Carlo simulation vs theoretical value

## III. EXPECTED CONTRIBUTIONS

### III.1. MSD-based heterogeneity testing

Assume $\nu_1$ and $\nu_2$ paths are obtained from fluids I and II, respectively. We will build upon the asymptotic and finite-sample frameworks described in Section II.1 and II.2 to construct an encompassing methodology for the testing of intra- and inter-sample heterogeneity. This will consist of:
$(i)$ testing the null hypotheses (equalities between different paths' parameters from a fluid)

$$H_{0,i} : \boldsymbol{\xi}_{1,i} = \boldsymbol{\xi}_{2,i} = \cdots = \boldsymbol{\xi}_{\nu_i,i}, \quad i = 1, 2, \tag{7}$$

where a chi-square test can be applied.
$(ii)$ testing the null hypothesis

$$H_0 : \boldsymbol{\xi}_{\text{I}} = \boldsymbol{\xi}_{\text{II}}, \tag{8}$$

for which a $z$-test can be applied.

### III.2. Wavelet-based heterogeneity testing

We will also invertigate the testing of fluid heterogeneity by means of wavelet techniques. A great advantage of wavelet-based estimation is that, unlike the result in Section II, its limiting distribution is Gaussian and rate of convergence is $\sqrt{n}$ if the number of vanishing moment of wavelet is no less than 2. However, a potential drawback (shown by numerical simulation) of wavelet-base estimator is its mean squared error increases if number of vanishing moment increases. We will compare $\widehat{\alpha}_{\text{wave}}$ and $\widehat{\alpha}_{\widehat{\text{MSD}}}$ in terms of their finite-sample and asymptotic performance in the testing of heterogeneity.


[1] I. M. Sokolov, Physics **1**, 8 (2008).
[2] S. T. Hess, T. Girirajan, and M. D. Mason, Biophysical Journal **91**, 4258 (2006).
[3] T. Mason and D. Weitz, Physical Review Letters **74**, 1250 (1995).
[4] J. Suh, M. Dawson, and J. Hanes, Advanced Drug Delivery Reviews **57**, 63 (2005).
[5] H. Matsui, V. Wagner, D. Hill, U. Schwab, T. Rogers, B. Button, R. Taylor, R. Superfine, M. Rubinstein, B. Iglewski, and R. Boucher, Proceedings of the National Academy of Sciences **103**, 18131 (2006).
[6] S. Lai, Y. Wang, R. Cone, D. Wirtz, and J. Hanes, PLoS One **4**, e4294 (2009).
[7] E. Barkai, Y. Garini, and R. Metzler, Physics Today **65**, 29 (2012).
[8] M. Valentine, P. Kaplan, D. Thota, J. Crocker, T. Gisler, R. Prudhomme, M. Beck, and D. A. Weitz, Physical Review E **64**, 061506 (2001).
[9] O. Lieleg, I. Vladescu, and K. Ribbeck, Biophysical Journal **98**, 1782 (2010).
[10] G. Didier and K. Zhang, To appear in Journal of Time Series Analysis.
[11] D. Veitch and P. Abry, IEEE Transactions on Information Theory **45**, 878 (1999).