Learning a Potential Function From a Trajectory

David R. Brillinger

***Abstract*—This letter concerns the use of stochastic gradient systems in the modeling of the paths of moving particles and the consequent estimation of a potential function. The work proceeds by setting down a parametric or nonparametric model for the potential function. The method is simple, direct, and flexible, being based on a linear model and the least squares. Explanatories, attractors, and repellors may be included directly. The large sample distribution of the estimated potential function is provided, under specific assumptions. There are direct extensions to updating, sliding window, adaptive, robust, and real-time variants. An example analyzing the path of an elk is presented.**

***Index Terms*—Mobility model, monitoring, potential function, stochastic differential equation, stochastic gradient system.**

I. Introduction

LOCATION signals of moving objects, obtained for example by GPS or LORAN, have become common in practice. Typically, one has scattered positions along trajectories of the objects. The questions of how to summarize, how to predict, and how to simulate make such movements arise. This happens particularly when a number of paths are involved or the path of an object is a tangle. See Fig. 1, which shows 1571 locations over a period of a month along the track of an elk in Starkey Project in Oregon. (Reference [1] provides the project's website address.)

This letter provides a unified approach for dealing with movement modeling and associated data. The fields in which movement data have arisen include animal tracking [2], [3] and soccer [4]. There are papers developing a statistical potential approach to tracks. These include [2], [3], and references therein. This letter provides some formal background missing in those papers, discussion, and an example.

Let $\boldsymbol{r}$ denote a point in R^p. (In the mathematical expressions below, all the vectors appearing are column vectors and set in boldface.) A potential function, $V(\boldsymbol{r})$, is a real-valued function of location. Its use can lead to simpler representations of motion than those based on modeling velocities directly. One can note that in the overdamped case, the equation of motion of a particle in the potential field, $V(\boldsymbol{r})$, is

$$d\mathbf{r}(t) = -\nabla V(\mathbf{r}(t))dt \quad (1)$$

having assumed $V(\boldsymbol{r})$ differentiable and with ∇ denoting the gradient. [The negative sign in (1) is traditional.] The entity $d\boldsymbol{r}(t)/dt$ is called a vector field. When $p = 2$, the level surfaces of the potential function are conveniently displayed in contour form and its gradient as arrows on a grid (see Figs. 2 and 3).

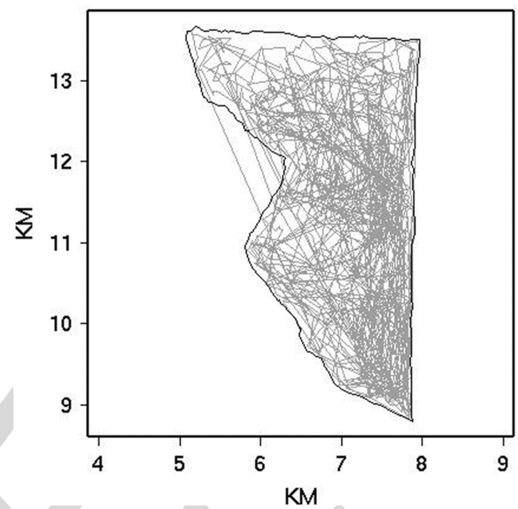

Fig. 1. Path of an elk around the NE pasture of the Starkey Experimental Forest in Oregon. Locations were estimated every two hours and are joined by consecutive straight lines.

The estimation method to be presented can be motivated by stochastic gradient systems, that is, systems that can be written in the time invariant case as

$$d\boldsymbol{r}(t) = -\nabla V(\boldsymbol{r}(t))dt + \boldsymbol{\sigma}(\boldsymbol{r}(t))d\boldsymbol{B}(t) \quad (2)$$

for some differentiable V with $\boldsymbol{B}(t)$ a p-dimensional Brownian motion and $\boldsymbol{\sigma}$ a p by p matrix. Expression (2) is a particular case of the stochastic differential equation (SDE)

$$d\boldsymbol{r}(t) = \boldsymbol{\mu}(\boldsymbol{r}(t))dt + \boldsymbol{\sigma}(\boldsymbol{r}(t))d\boldsymbol{B}(t). \quad (3)$$

What distinguishes the traditional SDE work from the present study is that the drift term μ here has the special form $-\nabla V$ for some real-valued function V. It will be seen that the modeling situation is simplified when such a V is assumed to exist.

II. Problem and Approach

The basic problem assumes the model (2) and seeks to learn $V(\boldsymbol{r})$ given data $(\boldsymbol{r}(t_i), i = 1, \cdots, n)$. These data will be viewed as locations at successive times, $\{t_i\}$, of an object moving along a trajectory of the process (2). One seeks both vector field and potential function estimates.

Manuscript received December 12, 2006; revised April 2, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Brian M. Sadler.

The author is with the Statistics Department, University of California Berkeley, Berkeley, CA 94720 USA (e-mail: brill@stat.berkeley.edu).

Digital Object Identifier 10.1109/LSP.2007.900032

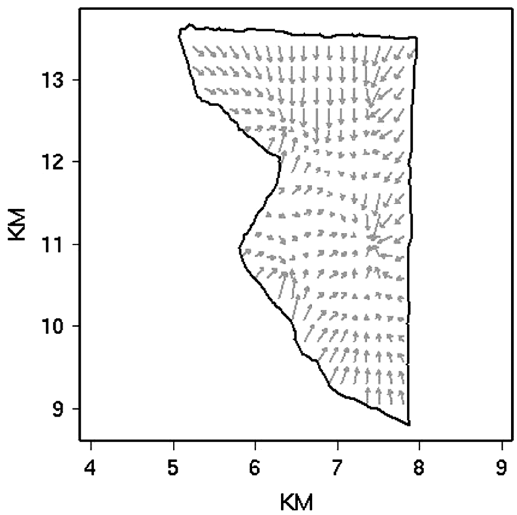

Fig. 2. Estimated vector field for the path of Fig. 1.

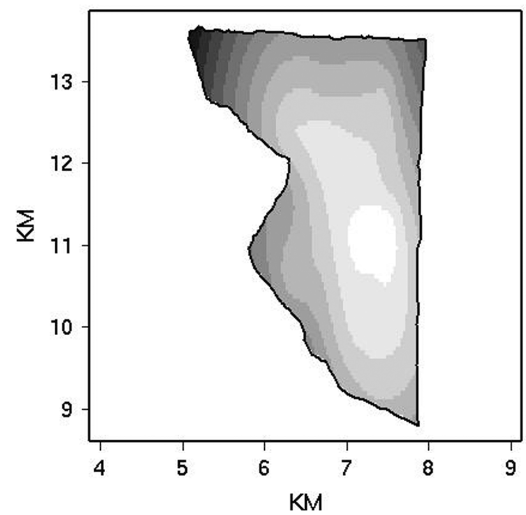

Fig. 3. Estimated potential for the path of Fig. 1. The lighter shading corresponds to smaller values.

Supposing that $\nabla V(\boldsymbol{r})$ is a smooth function of $\boldsymbol{r}$, and that the observation times are close together, one can set down the following approximation to (2):

$$\boldsymbol{r}(t_{i+1}) - \boldsymbol{r}(t_i) = -\nabla V(\boldsymbol{r}(t_i))(t_{i+1} - t_i) + (t_{i+1} - t_i)^{1/2}\boldsymbol{\sigma}\boldsymbol{Z}_{i+1} \quad (4)$$

for $i = 1, 2, 3, \cdots, n$, with $\boldsymbol{\sigma}$ a p by p covariance matrix and with the $\boldsymbol{Z}_i$ independent p-dimensional variates having mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{I}$. The reason for the multiplier $(t_{i+1}-t_i)^{1/2}$ is that for real-valued Brownian, $\mathrm{Var}(dB(t)) = dt$.

The approximation

$$(\boldsymbol{r}(t_{i+1}) - \boldsymbol{r}(t_i))/(t_{i+1} - t_i) = \mu(\boldsymbol{r}(t_i)) + \boldsymbol{\sigma}\boldsymbol{Z}_{i+1}/(t_{i+1} - t_i)^{1/2} \quad (5)$$

for the SDE (3) was employed in [2] and [3] for elk and movement and is employed in the example of this letter. In [2], an early attempt was made at estimating a potential function by numerical integration and simulation. The question was asked whether the vector field, $\boldsymbol{\mu}$, had the form $-\nabla V(\boldsymbol{r})$. This may be studied by comparing an unrestricted estimate of $\boldsymbol{\mu}$ with one assuming the existence of a potential function. The approach of papers [2] and [3] was informal.

III. Potential Functions

A basic issue is how to describe mathematically a potential function, $V(\boldsymbol{r}), \boldsymbol{r}$ in R^p. Suppose one exists. For introductory purposes in the development, suppose V is linear in a vector-valued parameter $\boldsymbol{\beta}$. Write $V(\boldsymbol{r}) = \boldsymbol{\phi}(\boldsymbol{r})^T\boldsymbol{\beta}$ with $\boldsymbol{\phi}$ an L by 1 vector of functions of known form and $\boldsymbol{\beta}$ an L by 1 unknown parameter. Examples of such a V follow. The gradient of V is the p by 1 vector $\nabla\boldsymbol{\phi}(\boldsymbol{r})^T\boldsymbol{\beta}$.

Example 1: Polynomial expansion.

Consider $V(\boldsymbol{r}) = \sum \boldsymbol{\beta}_m \boldsymbol{r}^m$, where $\boldsymbol{m} = (m_1, \cdots, m_p)$ and $\boldsymbol{r}^m = x_1^\wedge m_1 \ldots x_p^\wedge m_p$, with $\sum$ over $m_1, \cdots, m_p \geq 0$ and $1 \leq m_1 + cldots + m_p \leq M$.

One could employ a trigonometric polynomial, a spline function, or a wavelet expansion here. Many functions, V, can be well approximated by taking M large. In practice, one might employ M_n with M_n increasing with n.

Example 2: Node based.

Consider nodal points $\boldsymbol{u}_l, l = 1, \cdots, L$ in R^p and set $V(\boldsymbol{r}) = \sum \boldsymbol{\beta}_l K(\boldsymbol{r} - \boldsymbol{u}_l)$ for some real-valued differentiable kernel K. As a specific example of K, one has the radial basis thin plate splines [5, pp. 30–34]

$$K(\boldsymbol{r}) = |\boldsymbol{r}|^{2q-p}\log|\boldsymbol{r}| \text{for } p \text{ even}, = |\boldsymbol{r}|^{2q-p} \text{for } p \text{ odd}. \quad (6)$$

Here q denotes the order of differentiability of $K, 2q - p > 0$, and $|\boldsymbol{r}| = (\boldsymbol{r}^T r)^{1/2}$. An expression like (6) leads to a smooth representation for V.

Example 3: Attraction and repulsion.

Consider a region A and a point $\boldsymbol{r}$ outside A. Potential functions can be set down leading to attraction or repulsion from A. Specifically, if one lets $d_A(\boldsymbol{r})$ denote the minimum distance from the point $\boldsymbol{r}$ outside A to A and sets $V(\boldsymbol{r}) = \beta d_A(\boldsymbol{r})^\alpha$, then for $\alpha > 0$, one has attraction to A and repulsion if $\alpha < 0$. One can reverse attraction and repulsion by changing the sign of d_A. It can be convenient to use $V(\boldsymbol{r}) = \beta_1 \log d_A(\boldsymbol{r}) + \beta_2 d_A(\boldsymbol{r})$ for similar purposes.

The functional forms of Examples 1–3 may be added together to provide other forms.

Reference [6] considers the observed trajectory of a monk seal near the island of Molokai employing the mixed function

$$V(\boldsymbol{r}) = \gamma_1 x + \gamma_2 y + \gamma_{11} x^2 + \gamma_{12} xy + \gamma_{22} y^2 + C/d(x, y) \quad (7)$$

where $\boldsymbol{r} = (x, y)^T$ is in R^2 and represents the location of the animal on the ocean surface. The value $d(x, y)$ is the distance from the location $\boldsymbol{r}$ to the nearest point on the island. The γ's and C are unknown parameters to be estimated. The final term in (7) keeps the seal off of the island. A different monk seal is studied in [7], and a different, now time dependent, potential function is employed

$$V(\boldsymbol{r},t) = \alpha \log d(\boldsymbol{r},t) + \beta d(\boldsymbol{r},t)$$

with $d(\boldsymbol{r},t)$ being the animal's distance from an attractor at time t. The attractor switched depending on whether the animal was on an outbound or an inbound journey.

Reference [4] studies the motion of a soccer ball during a very exciting World Cup moment. The potential function used is

$$\alpha \log d(\boldsymbol{r}) + \beta d(\boldsymbol{r}) + \gamma_1 x + \gamma_2 y + \gamma_{11} x^2 + \gamma_{12} xy + \gamma_{22} y^2$$

with $d(\boldsymbol{r})$ the shortest distance to the goalmouth from $\boldsymbol{r} = (x,y)^T$. The first two terms lead to attraction to the goalmouth and the remaining to general motion on the field.

Potential function and vector field estimates are provided in each of the papers just referenced. The function $V(\boldsymbol{r})$ is linear in the parameter, and least squares is employed as the estimation procedure in each case. The function could be nonlinear, and then, nonlinear least squares could be employed. Reference [8] develops asymptotic results pertinent to the nonlinear case. Alternately, the $\{Z_i\}$ in (5) could be non-Gaussian and maximum likelihood estimation employed.

IV. Estimation

The representation (4) with $\boldsymbol{r}$ in R^p and $\nabla V(\boldsymbol{r}) = \nabla \boldsymbol{\phi}(\boldsymbol{r}_i)^T \boldsymbol{\beta}$ will be employed. The values $\boldsymbol{r}(t_i)$ will be written $\boldsymbol{r}_i$. Consider the p by 1 vector $(\boldsymbol{r}_{i+1} - \boldsymbol{r}_i)/(t_{i+1} - t_i)^{1/2}$. Following expression (5), the model has the form $(\boldsymbol{r}_{i+1} - \boldsymbol{r}_i)/(t_{i+1} - t_i)^{1/2} = -\nabla\phi(\boldsymbol{r}_i)\ ^T \beta(t_{i+1} - t_i)^{1/2} + \boldsymbol{\sigma Z}_{i+1}$

$i = 1, \cdots, n-1$ involving the L by 1 vector $\boldsymbol{\beta}$, the L by p matrix $\nabla \boldsymbol{\varphi}(\boldsymbol{r}_i)$, the p by p matrix $\boldsymbol{\sigma}$, and the p by 1 vector $\boldsymbol{Z}_{i+1}$. Suppose $\boldsymbol{\sigma} = \boldsymbol{\sigma I}$ with σ positive and $\boldsymbol{I}$ the p by p identity matrix. Stack the $n-1$ values $(\boldsymbol{r}_{i+1} - \boldsymbol{r}_i)/(t_{i+1} - t_i)^{1/2}, i = 1, \cdots, n-1$ vertically to form the $(n-1)p$ by 1 array $\boldsymbol{Y}_n$. Stack the $n-1$ matrices $-\nabla \boldsymbol{\varphi}(\boldsymbol{r}_i)^T (t_{i+1} - t_i)^{1/2}$ to form the $(n-1)p$ by L matrix $\boldsymbol{X}_n$. Stack the $n-1$ values $\sigma \boldsymbol{Z}_{i+1}$ to form $\boldsymbol{\varepsilon}_n$. Then one has the regression model

$$\boldsymbol{Y}_n = \mathbf{X}_n \boldsymbol{\beta} + \varepsilon_n \tag{8}$$

with the difference from ordinary regression that $\boldsymbol{Y}_n$ and $\boldsymbol{X}_n$ are statistically dependent. Using a generalized inverse, if necessary, one can compute an ordinary least-squares estimate $\boldsymbol{b} = (\boldsymbol{X}_n^T \boldsymbol{X}_n)^{-1} \boldsymbol{X}_n^T \boldsymbol{Y}_n$ of $\boldsymbol{\beta}$, and then, if $\boldsymbol{\varphi}(\boldsymbol{r})^T \boldsymbol{\beta}$ is estimable, $\boldsymbol{\varphi}(\boldsymbol{r})^T \boldsymbol{b}$ is a reasonable estimate of $V(\boldsymbol{r})$.

Supposing the individual entries of ε_n to be independent, zero mean, variance σ^2 variates, asymptotic properties of $\boldsymbol{\varphi}(\boldsymbol{r})^T \boldsymbol{b}$ may be obtained from [9, Theorem 3]. The theorem is given in the Appendix.

Let y_j denote the jth row of $\boldsymbol{Y}_n$. Let $\boldsymbol{x}_j^T$ denote the jth row of $\boldsymbol{X}_n^T$. One can compute $s_n^2 = ((n-1)p)^{-1} \sum (\boldsymbol{y}_j - \boldsymbol{x}_j^T \boldsymbol{b})^T (\boldsymbol{y}_j - \boldsymbol{x}_j^T \boldsymbol{b})$ as an estimate of σ^2 and, for example, set down a confidence interval for $\boldsymbol{\varphi}(\boldsymbol{r})^T \boldsymbol{\beta}$ using the results of [9]. Specifically, provided lim log $\lambda_{\max}(\boldsymbol{X}_n^T \boldsymbol{X}_n)/n \to 0$ almost surely, one has $s_n \to \sigma$ and by a Slutsky Theorem

$$(\boldsymbol{\varphi}(\boldsymbol{r})^T (\boldsymbol{X}_n^T \boldsymbol{X}_n)^{-1} \boldsymbol{\varphi}(\boldsymbol{r}))^{-1/2} \boldsymbol{\varphi}(\boldsymbol{r})^T (b - \boldsymbol{\beta})/s_n \to N(0,1)$$

with N(0,1) the standard normal. This leads to the approximate $100(1-\alpha)\%$ confidence interval

$$\boldsymbol{\varphi}(\boldsymbol{r})^T \boldsymbol{\beta} = \boldsymbol{\varphi}(\boldsymbol{r})^T \boldsymbol{b} \pm z_{\alpha/2} (\boldsymbol{\varphi}(\boldsymbol{r}) (\boldsymbol{X}_n^T \boldsymbol{X}_n)^{-1} \boldsymbol{\varphi}(\boldsymbol{r})^T)^{1/2} s_n$$

where $z_{\alpha/2}$ denotes the $100\alpha/2$ percent point of the standard normal. As mentioned in [9], one could use the F distribution to construct an approximate confidence region for a collection of values $\{\boldsymbol{\phi}(\boldsymbol{r}_k)^T \boldsymbol{\beta}\}$.

V. Example

The Starkey Project is a large area in Oregon set aside to study the interactions of elk, deer, cows, and man sharing an environment [1]. Fig. 1 shows a sampled trajectory of one of the elk in the NE Pasture. There were 1571 GPS locations and times of location obtained with a time interval of approximately two hours between successive locations. It is recognized that the theory connecting the sampled times case to the continuous time case expects the times to be close together. It is still anticipated that the discrete model studied is of interest in its own right and will provide results of practical use.

A potential function $V(\boldsymbol{r})$ was approximated by a thin plate radial basis spline employing the kernel function of (6) with p and $q = 2$, and $L = 36$. The x and y components of the $\boldsymbol{u}_l$ were taken to be the $100m/7m = 1, \cdots, 6$ percentiles of the standardized x and y values. These reason values were chosen was for illustrative purposes.

The coefficients $\boldsymbol{\beta}_l$ were estimated by ordinary least squares employing the model (5) with $\boldsymbol{\sigma} = \boldsymbol{\sigma I}$. The results are provided in Figs. 2 and 3. One sees the confusion of Fig. 1 much reduced. A point of attraction appears near the point (7.5,11.0). When one looks at a topographic plot of elevations, the point of attraction appears to be a valley/canyon of sorts. Fig. 3 provides an image plot of the potential function. Now one sees the point of attraction immediately.

The confusion of Fig. 1 has been referred to. An empirical gradient plot is similarly confused.

VI. Extensions and Conclusion

Various generalizations of the letter's results may be mentioned. One could set down an expansion for V employing wavelet functions. One could consider updating methods for real-time work, e.g., those based on a Kalman filter. One could envisage a potential function as a spatial state variable and the paths of objects determined by the measurement equation. If the potential function is changing slowly, one could consider a sliding window estimate [10]. Estimates that are robust to non-normailty and resistant estimation can be considered. In video analysis, one might consider the model $\boldsymbol{I}(\boldsymbol{r},t) = \boldsymbol{I}_0(\boldsymbol{r}) + \delta(\boldsymbol{r}(t) - \boldsymbol{r})$ with t indexing the video frames and δ the Dirac delta. The term I_0 represents a stationary background and $\boldsymbol{r}(t)$ the location of an object moving around in the scene [11].

This letter presents an estimation method for handling moving objects. The computations may be implemented by the least-squares algorithm. The model may be viewed as parametric or nonparametric.

APPENDIX

Because of the statistical dependence of the location of the object at time t_i on past locations, one needs special arguments to get the asymptotic distribution. For the simple cases of the letter, results based on martingale arguments are available in [9]. A result is [9, Theorem 3]. Consider the regression model $y_j = \boldsymbol{x}_j^T\boldsymbol{\beta} + \varepsilon_j, j = 1, 2, \ldots$ with the $\{\varepsilon_j\}$ martingale differences with respect to an increasing sequence of σ-fields $\{F_N\}$. Suppose that sup $_nE(\|\boldsymbol{\varepsilon}_N\|^\alpha|F_{N-1}) < \infty$ almost surely for some $\alpha < 2$. Suppose further that lim $_{n\to\infty}$ var$(\boldsymbol{\varepsilon}_N|F_{N-1}) = \sigma^2$ almost surely for some nonstochastic σ. Define $\boldsymbol{X}_N = [\boldsymbol{x}_1 \ldots \boldsymbol{x}_N]^T$. Assume that $\boldsymbol{x}_j$ is a F_{j-1} -measurable random variable and that there exists a nonrandom positive definite symmetric L by L matrix $\boldsymbol{B}_N$ for which $\boldsymbol{B}_N^{-1}(\boldsymbol{X}_N^T\boldsymbol{X}_N)^{1/2} \to \boldsymbol{I}$, sup $_{1\le i\le N}\|\boldsymbol{B}_N^{-1}\boldsymbol{x}_N\| \to 0$ in probability. Then as $N \to \infty$

$$(\boldsymbol{X}_N^T\boldsymbol{X}_N)^{1/2}(\boldsymbol{b}-\boldsymbol{\beta}) \to N(0, \sigma^2\boldsymbol{I})$$

in distribution.

Note that zero-mean independent observations like the successive entries of ε_n of (9) form a martingale difference sequence with respect to the σ-field generated by the preceding locations.

Reference [15] shows that under the further assumption, lim log $\lambda_{\max}$ $(\boldsymbol{X}_n^T\boldsymbol{X}_n)/n \to 0$ almost surely, one has $s_n \to \sigma$ almost surely.

ACKNOWLEDGMENT

The author would like to thanks A. A. Ager and H. K. Preisler for involving him with Starkey data sets and ideas that proved useful.

REFERENCES

[1] Starkey Project. [Online]. Available: http://www.fs.fed.us/pnw/starkey.

[2] D. R. Brillinger, H. K. Preisler, A. A. Ager, and J. G. Kie, , A. K. M. E. Saleh, Ed., "The use of potential functions in modelling animal movement," in *Data Analysis from Statistical Foundations*. Huntington, NY: Nova, pp. 369–386. [Online]. Available: http://www.stat.berkeley.edu/~brill/Papers/fraser8.pdf, (2001a).

[3] D. R. Brillinger, H. K. Preisler, A. A. Ager, J. Kie, and B. S. Stewart, Modelling Movements of Free-ranging Animals, Univ. California Berkeley Statistics, Tech. Rep. 610. [Online]. Available: http://www.stat.berkeley.edu/~brill/Preprints/610.pdf, (2001b).

[4] D. R. Brillinger, "A potential function approach to the flow of play in soccer," *J. Quant. Anal. Sports* vol. 3, no. 1, 2007. [Online]. Available: http://www.stat.berkeley.edu/~brill/Papers/goal21.pdf.

[5] G. Wahba, *Spline Models for Observational Data*. Philadelphia, PA: SIAM, 1980.

[6] D. R. Brillinger, B. Stewart, and C. Littnan, E. P. Liski, J. Isotalo, S. Puntanen, and G. P. H. Styan, Eds., "A meandering hylje," in Festschrift for Tarmo Pukkila on His 60th Birthday, Dept. Mathematics, Statistics and Philosophy, Univ. Tampere. [Online]. Available: http://www.stat.berkeley.edu/~brill/Papers/hylje.pdf, 2006a.

[7] D. R. Brillinger, B. Stewart, and C. Littnan, "Three months journeying of a Hawaiian monk seal," *Lecture Notes Statist.* [Online]. Available: http://www.stat.berkeley.edu/~brill/Papers/monk32.pdf 2006b.

[8] T. L. Lai, "Asymptotic properties of nonlinear least squares estimates in stochastic regression models," *Ann. Statist.*, vol. 10, pp. 1917–1930, 1994.

[9] T. L. Lai and C. Z. Wei, "Least squares estimation in stochastic regression models with applications to identification and control of dynamic systems," *Ann. Statist.*, vol. 10, pp. 154–166, 1982.

[10] M. Belge and E. L. Miller, "A sliding window RLS-like adaptive algorithm for filtering alpha-stable noise," *IEEE Signal Process. Lett.*, vol. 7, no. 4, pp. 86–89, Apr. 2000.

[11] V. Zlokolica, A. Pizurica, and W. Philips, "Noise estimation for video processing based on spatio-temporal gradients," *IEEE Signal Process. Lett.*, vol. 13, no. 6, pp. 337–340, Jun. 2006.