Theory of Lehmer Transform and Its Applications in Identifying the Electroencephalographic Signature of Major Depressive Disorder

Masoud Ataei (✉ mataei@yorku.ca)
York University

Xiaogang Wang
York University

Research Article

Keywords: Lehmer Transform, Electroencephalographic Signature, Major Depressive Disorder

Posted Date: September 2nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-463014/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Theory of Lehmer Transform and Its Applications in Identifying the Electroencephalographic Signature of Major Depressive Disorder

Masoud Ataei¹,* and Xiaogang Wang¹

¹Department of Mathematics and Statistics, York University, Toronto, Ontario, Canada, ON M3J 1P3
*matei@yorku.ca

ABSTRACT

We propose a novel transform called Lehmer transform and establish theoretical results which are used to compress and characterize large volumes of highly volatile time series data. It will be shown that our proposed method could be used as a practical data-driven approach for analyzing extreme events in nonstationary and highly oscillatory stochastic processes such as biological signals. We demonstrate the advantage of the proposed transform in comparison with traditional methods such as Fourier and Wavelets transforms through an example of devising a classifier to discern the patients with major depressive disorder from the healthy subjects using their recorded EEG signals and provide the computational results. We show that the proposed transform can be used for building better and more robust classifiers with significant accuracy.

Introduction

In this paper, we introduce a new class of transforms referred to as the Lehmer transform, whose applications could potentially bridge the gap between parametric and nonparametric statistics. The proposed transform decomposes a function of the data sample into the so-called suddency moments, which in turn provides the means to construct parametric families of nonparametric statistical models.

The Lehmer transform has its mathematical root in the Lehmer mean function, named after the renowned number theorist Derrick H. Lehmer, which has primarily been studied in analytic number theory. To the best of our knowledge, however, the appealing properties of this function have not received much attention from statistics and machine learning communities in the past, leaving its wide range of potential applications largely unexploited in handling complex nonstationary time series such as biological signals.

Most existing works employing the Lehmer mean function for the purpose of data analysis, rely mainly on its special cases. For instance, Terziyan [21] formulated a distance metric using the harmonic and contra-harmonic means and evaluated its effectiveness in geographic information systems. In a similar manner, Somasundaram et al. [20] studied the disconnected graphs and their possible labeling schemes and Sluciak [19] developed state-dependent consensus algorithms by resorting to applications of some special cases of the Lehmer mean function. Besides, Gomes [11] successfully constructed a family of high-performance value at risk estimators through the use of this function and proved asymptotic normality of the constructed estimators.

On the other hand, being a generalization of the power mean function and having connections to other important classes of various means, the Lehmer mean function has received a lot of attentions in recent years. The function’s elementary properties like homogeneity, monotonicity and differentiability have been discussed in [1, 3], whereas its more advanced properties like Schur-convexity, Schur harmonic convexity and Schur power convexity have been the focus of more recent studies, e.g., see [8, 9, 10, 23, 24] and references therein. Also, the inflection points of the function have been studied in [18], and the results concerning its possible connections to Gini and Toader means have been provided in [5, 6, 7, 12, 22, 25].

In this article, we first define the Lehmer transform and then derive a function called action potential function which is the basis to establish a feature extraction scheme. The method is applied to an electroencephalography (EEG) dataset on major depressive disorder (MDD) which is a widely-used benchmark dataset for evaluation of different methods. Detailed comparison and discussions are also presented in this article.
Table 1. Some special cases of suddency moments.

| Suddency domain | Sample domain |
|-----------------|---------------|
| $-\infty$       | Minimum ($n \geq 1$) |
| 0               | Harmonic mean ($n \geq 1$) |
| $1/2$           | Geometric mean ($n = 2$) |
| 1               | Arithmetic mean ($n \geq 1$) |
| 2               | Contra-harmonic mean ($n \geq 1$) |
| $\infty$        | Maximum ($n \geq 1$) |

**Lehmer Transform**

Let $\mathbb{R}_{>0}$ and $\mathbb{R}$ denote the sets of strictly positive and extended real numbers, respectively. Given a signal $x = \{x_i \in \mathbb{R}_{>0}, i = 1, \ldots, n\}$, the Lehmer transform $\mathcal{L} : \mathbb{R} \to \mathbb{R}_{>0}$ is a mapping defined by

$$\mathcal{L}(s) = \begin{cases} \max_{i=1,\ldots,n} \{x_i\}, & \text{if } s = \infty, \\ \frac{1}{n} \sum_{i=1}^{n} x_i^s, & \text{if } s \in (-\infty, \infty), \\ \min_{i=1,\ldots,n} \{x_i\}, & \text{if } s = -\infty. \end{cases}$$

According to this definition, the Lehmer transform $\mathcal{L}(s)$ maps every point $s \in \mathbb{R}$ into some statistic contained within the range $[\min\{x\}, \max\{x\}]$, where we refer to the point $s$ as a suddency moment of the given signal $x$. Table 1 reports some of the widely-encountered suddency moments in data analysis. Furthermore, for any statistic $t \in [\min\{x\}, \max\{x\}]$, the inverse image $\mathcal{L}^{-1} : \mathbb{R}_{>0} \to \mathbb{R}$ provides a set that contains the associated suddency moment(s) of the signal such that

$$\mathcal{L}^{-1}(t) = \{s \in \mathbb{R} : \mathcal{L}(s) = t\}.$$ 

Provided that the signal $x$ contains at least two unique elements, the Lehmer transform $\mathcal{L}(s)$ would be a monotonously increasing function. Furthermore, the transform is differentiable in $P$-probability and its $n$th order derivative could be proven using the mathematical induction to be as follows:

$$\frac{\partial^n}{\partial s^n} \mathcal{L}(s) = \mathcal{L}(s) \left[ \sum_{k=0}^{n} \frac{1}{k!} \left( \sum_{j=0}^{k} (-1)^j \binom{k}{j} \Lambda^j \frac{\partial^n}{\partial s^n} \Lambda^{k-j} \right) \right],$$

where $\Lambda := \log \mathcal{L}(s)$.

In particular, the first derivative of the Lehmer transform is derived as

$$\frac{\partial}{\partial s} \mathcal{L}(s) = \mathcal{L}(s) \left[ \frac{\sum_{i=1}^{n} x_i \log x_i - \sum_{i=1}^{n} x_i^{s-1} \log x_i}{\sum_{i=1}^{n} x_i^{s-1}} \right] = \sum_{i=1}^{n} \sum_{k=i+1}^{n} (x_i - x_k) \left( \log \frac{x_i}{x_k} \right) (x_i x_k)^{s-1}.$$

**Theorem 1.** Given the signal $x$ satisfying the condition $x_i \neq x_j$ for some $i, j \in \{1, \ldots, n\}$ such that $i \neq j$, the inverse Lehmer transform is defined as

$$\mathcal{L}^{-1}(t) = \begin{cases} \infty, & \text{if } t = \max_{i=1,\ldots,n} \{x_i\}, \\ g(t), & \text{if } t \in \left( \min_{i=1,\ldots,n} \{x_i\}, \max_{i=1,\ldots,n} \{x_i\} \right), \\ -\infty, & \text{if } t = \min_{i=1,\ldots,n} \{x_i\}, \end{cases}$$

where $g(t)$ is some statistical function related to the Lehmer transform.
where

\[
g(t) = s_0 + \sum_{k=1}^{\infty} \frac{(t - L(s_0))^k}{k!} \left\{ \lim_{s \to s_0} \frac{\partial^{k-1}}{\partial s^{k-1}} \left( \frac{s - s_0}{L(s) - L(s_0)} \right)^k \right\},
\]

for some \( s_0 \in \mathbb{R} \).

**Proof.** Under the compatibility condition imposed on the given signal, the monotonicity property of \( L(s) \) implies that the function is injective, which in turn guarantees the existence of some inverse function denoted by \( L^{-1}(t) \). By further noting that \( L(s) \) is analytic for every point \( s \in \mathbb{R} \), the Lagrange-Bürmann formula \([4, 15]\) can be utilized to express the interior points of the inverse function as a Taylor expansion \( g(t) \) about some given point \( s_0 \in \mathbb{R} \), hence yielding the presented result. \( \blacksquare \)

**Action Potential Distribution Function**

Let us recall the principal branch of the Lambert-\( W \) function denoted by \( W_0(z) \), which is defined for every real \( z \in (-e^{-1}, \infty) \) as the inverse function of \( g(z) = ze^z \), whose schematics are depicted in Figure 1a. Here, \( \Omega \) denotes the Omega constant defined by \( \Omega = W_0(1) \approx 0.5671 \). The membrane voltage function (see Figure 1b) may then be expressed as some transformation of the following functional (for a constant \( c \)):

\[
V(t; c) = \begin{cases} 
W_0(t - \frac{1}{e}) - \frac{1}{ce^t}, & \text{if } t \in \left(0, \frac{1}{e}\right), \\
t^2e - \frac{1}{e}, & \text{if } t \in \left(\frac{1}{e}, \frac{e+1}{e}\right), \\
t^2e^{-t} - \frac{1}{e}, & \text{if } t \in \left(\frac{e+1}{e}, \infty\right). 
\end{cases}
\]

Inspired by the appealing properties of the membrane voltage function defined above, we propose a distribution function to account for the probabilistic phenomena that underlie the complex dynamics of the action potentials.

**Theorem 2.** Let us assume that a signal is given which has been transformed such that it would satisfy the following conditions:

\[
\mathcal{L}(\infty) = \exp\{\frac{W_0(\alpha \beta)}{\alpha \beta}\} \quad \text{and} \quad \mathcal{L}(-\infty) = \exp\{W_0(0)\}.
\]

Then, a real-valued random variable \( S \) is said to follow an action potential distribution if its probability density function is given by

\[
f_S(s) = C (1 + \alpha \beta \mathcal{L}(s)) e^{\beta \mathcal{L}(s)} (\mathcal{L}(s))^{-\frac{1}{\beta}} \mathcal{L}'(s),
\]

where \( C \) is a normalizing constant.

---

3/8
where \( \alpha \in (0, 1] \), \( \beta \in \mathbb{R}_{>0} \) and the normalization constant \( C \) is obtained using the following relation

\[
C = \left( \exp \left\{ \frac{W_0(\alpha \beta)}{\alpha^2 \beta} + \beta \exp\left\{ \frac{W_0(\alpha \beta)}{\alpha \beta} \right\} \right\} - \exp\{\beta\} \right)^{-1}.
\]

Proof. Let us consider the following function

\[
F(s) = A + B(s) \frac{1}{|e^\beta s|},
\]

where

\[
A = \frac{\exp\{\beta\}}{\exp\left\{ \frac{W_0(\alpha \beta)}{\alpha^2 \beta} + \beta \exp\left\{ \frac{W_0(\alpha \beta)}{\alpha \beta} \right\} \right\} - \exp\{\beta\}}
\]

and

\[
B = \frac{1}{\exp\left\{ \frac{W_0(\alpha \beta)}{\alpha^2 \beta} + \beta \exp\left\{ \frac{W_0(\alpha \beta)}{\alpha \beta} \right\} \right\} - \exp\{\beta\}}.
\]

Then, it is straightforward to show that \( F(s) \) is right-continuous and nondecreasing in the domain of suddenness moments, and further it satisfies the following properties

\[
\lim_{s \to -\infty} F(s) = 0 \quad \text{and} \quad \lim_{s \to \infty} F(s) = 1.
\]

Thus, \( F(s) \) is a distribution function. As a result, we may resort to the Caratheodory extension theorem which asserts that there exists a unique Lebesgue-Stieltjes measure associated with the distribution function \( F(s) \). In turn, this enables us to define a real-valued random variable \( S \) on a support \( s \in (-\infty, \infty) \), whose cumulative probability distribution function, i.e., \( F_S(s) \), coincides with \( F(s) \). The probability distribution function presented in the theorem is then derived by differentiating the \( F_S(s) \).

The action potential distribution function is a general framework to analyze complex nonstationary time series such as brain waves, e.g., see Figure 2 which depicts plots of action potential distribution function for an EEG signal with varying the parameters \( \alpha \) and \( \beta \). An important class of action potentials is referred to as the dendritic spikes, which plays a significant role in neuronal communication, memory and learning. Theoretically speaking, the dendritic spike dynamics would involve a number of different phase transitions, typically starting out from an initialization phase upon applying a stimulus. The dynamical system would then enter a rising phase if the membrane voltage rises above some threshold value, which is then followed by a phase transition into a falling phase. The final transition would occur at a reestablishment phase through which the membrane potential would eventually reside at its resting state.

In reality, however, a large body of evidence in neurophysiology suggests that the variations in electro-physiological measurements may distort the actual action potentials, yielding a substantially different phase transition scheme as compared to the theoretical one mentioned above. That is, different types of undershoot and overshoot may be encountered in dendritic spikes dynamics. The key element which we incorporate in our analysis of dendritic spikes, pertains to the point that membrane voltages could effectively be modelled using the Lambert-W function and its variants.

**Feature Extraction**

Let us first denote the 4-way tensor required for embedding the experiment data by the following convention \( x_{jm}^{(n)} \) for \( n = 1, \ldots, N, m = 1, \ldots, M, j = 1, \ldots, J \), and \( i = 1, \ldots, I \), where \( N \) and \( M \) denote the total number of the subjects and experiments, respectively. Also, let \( J \) and \( I \) represent the total number of active channels within each experiment and the sample points corresponding to each experiment, respectively. Note that, the values of \( I \) might be different throughout different experiments. However, we presume that such a fact would be clear from the context, hence eliminating the need to introduce any further subscript on \( I \). Accordingly, the measurements recorded on the \( j \)th channel of the \( m \)th experiment can be represented by \( x_{jm}^{(n)} = [x_{1jm}^{(n)}, x_{2jm}^{(n)}, \ldots, x_{Ijm}^{(n)}] \).

Thereafter, for each signal contained within the dataset, parameters of the action potential distribution are estimated using the maximum likelihood method, e.g., see Figure 3 which depicts plots of action potential probability density functions for EEG signals of a healthy as well as an MDD subject. It is worth mentioning that based on an extensive set of experiments conducted on EEG signals available from various domains of psychiatry and neuroscience, we have observed that there typically exist
several common features across all different datasets which can potentially be employed for the purposes of constructing fruitful classifiers. To this end, we introduce a set of five features which can possibly be used in construction of any EEG classifier. That is, the following quantities are computed for signals using their associated action potential distribution functions:

\[
F_S(M_1; \mathbf{x}_{jm}^{(n)}),
\]

\[
1 - F_S(M_2; \mathbf{x}_{jm}^{(n)}),
\]

\[
M_1 \int_{-\infty}^{-\infty} f_S(s; \mathbf{x}_{jm}^{(n)}) \log \left[ f_S(s; \mathbf{x}_{jm}^{(n)}) \right] ds,
\]

\[
M_2 \int_{M_1}^{\infty} f_S(s; \mathbf{x}_{jm}^{(n)}) \log \left[ f_S(s; \mathbf{x}_{jm}^{(n)}) \right] ds
\]

and

\[
\int_{M_2}^{\infty} f_S(s; \mathbf{x}_{jm}^{(n)}) \log \left[ f_S(s; \mathbf{x}_{jm}^{(n)}) \right] ds,
\]

where

\[
M_1 = \arg \max_{s \leq 1} f_S(s; \mathbf{x}_{jm}^{(n)})
\]

and

\[
M_2 = \arg \max_{s \geq 1} f_S(s; \mathbf{x}_{jm}^{(n)}).
\]

These quantities are then utilized to train a classifier for biological signals such as EEG measurements. More precisely, the inputs for the classification algorithm would form a matrix with \(N\) rows and \(M \times J \times 5\) columns.

**Application to Major Depressive Disorder**

The dataset we study in this work is comprised of EEG recordings of 34 patients suffering from MDD. Our goal would then be to investigate whether the patterns existing in the EEG recordings of the subjects could potentially be used to discern the MDD patients from healthy subjects. To this end, EEG recordings of 30 healthy subjects are further taken into consideration as the control group of this study. Each subject undergoes 3 different types of experiments while his/her EEG signals are recorded. More precisely, the experiments are conducted in 3 different setups: eyes closed (EC), eyes open (EO), and while the subject performs a specific task (T ASK). The duration for each of the experiments (EC) and (EO) is 5 minutes which comprises an average of 77000 signals per each channel (there exist 19 channel per each experiment). The experiment (T ASK), however, lasts for longer duration 10 minutes and its average number of signals sampled with frequency 256Hz is roughly 155000 on each channel; see [17] for more details on the conducted experiments.

In order to classify the subjects, we compare performance of three different algorithms; namely, Gaussian naive Bayes (GNB), convolutionary neural networks (CNN) and bootstrap-aggregated decision trees (BADT), whose performance have been reported in Table 2. Also, Table 3 reports the computational results for classification of the considered data using 10-fold cross validation averaged over 100 replications, which are further compared to the results reported by several authors in the literature including the most recent and prominent work of Mumtaz et al. [17]. It is noteworthy to mention that the classification algorithm proposed by Mumtaz et al. is mainly based on a logistic regression algorithm wherein the authors have used a combination of wavelet transforms (WT), short-time Fourier transforms (STFT) and empirical mode decomposition (EMD) as the feature extraction method and a rank-based algorithm for the purpose of feature selection. On the basis of Table 3, it is seen that our proposed method achieves higher accuracy as compared to the method based on the wavelet transforms as well as other previously-proposed approaches for classifying the MDD and healthy subjects. A distinguishing aspect of our proposed Lehmer-based classifier is that it yields relatively lower standard errors when classifying the instances. In turn, this makes the results obtained by using the proposed classifier reliable to a great extent, and such reliability can potentially lead to an effective applications of the Lehmer-based classifiers by practitioners who monitor the possible impacts of prescribed antidepressants’ treatment on their patients.
Table 2. Comparisons with different incorporated classifiers.

| Classifier                  | Accuracy | Sensitivity | Specificity |
|-----------------------------|----------|-------------|-------------|
| Lehmer Transform + GNB      | 93.8% (±1.3) | 93.9% (±1.3) | 93.8% (±1.3) |
| Lehmer Transform + CNN      | 87.9% (±1.5) | 84.2% (±1.9) | 97.2% (±1.3) |
| Lehmer Transform + BADT     | 90.5% (±1.3) | 94.1% (±1.2) | 87.0% (±1.5) |

Table 3. Comparisons with other algorithms from literature.

| Derived EEG Measure                          | Accuracy | Sensitivity | Specificity |
|----------------------------------------------|----------|-------------|-------------|
| ATR Index [13]                               | 61.7% (±9.7) | 70.0% (±15.3) | 54.0% (±17.2) |
| EEG Theta Coherence [16]                     | 70.7% (±6.5) | 75.7% (±9.1) | 65.7% (±8.3) |
| Coherence, PSD, PSD ratio [14]               | 72.1% (±7.6) | 80.0% (±13.5) | 65.0% (±12.3) |
| P300 (amplitude and latencies) [2]           | 74.2% (±13.1) | 70.0% (±15.6) | 75.0% (±7.7) |
| PSD, PSD ratios [14]                         | 54.5% (±8.0) | 55.0% (±14.7) | 50.0% (±15.6) |
| Wavelet Transform [17]                       | 87.5% (±7.1) | 95.0% (±4.3) | 80.0% (±8.8) |
| Lehmer Transform                             | 93.8% (±1.3) | 93.9% (±1.3) | 93.8% (±1.3) |

Discussion

We propose a new transform that is shown be suitable to handle highly oscillatory nonstationary time series. It can effectively summarize and extract features from multi-channel signals and form the basis for building a highly accurate classifier. Although our work is still exploratory in nature, we believe that the potential of this new transform can be utilized fully in order to discern subtle patterns that cannot be recognized by other traditional methods such as Fourier and Wavelets transforms.

Author Contributions Statement

M.A. developed the methodology and conceived, conducted and analyzed the experiments, X.W. developed the methodology and analyzed the experiments. All authors reviewed the manuscript.
Figure 2. Plots of action potential distributions for some EEG signal with varying the parameters $\alpha$ and $\beta$. 
Figure 3. Schematics of electrode placement and action potential distributions for control and MDD subjects.

(a) Electrode placement and mesh matrix representing the 10-20 international system.

(b) Plots of action potential probability density functions for a healthy subject when performing TASK.

(c) Plots of action potential probability density functions for a subject with MDD when performing TASK.