Analysis a large data by using Weyl function

تحليل البيانات الكبيرة باستخدام دالة ويل

د. سعاد إبراهيم المقصبي- أستاذ الرياضيات المشارك - كلية العلوم- جامعة بنغازي
ا. أسماء مفتاح محمد - محاضر مساعد- كلية الآداب والعلوم- قمينس - جامعة بنغازي
ا. أماني ميلاد الشطشاط - محاضر - كلية التقنية الكهربائية والإلكترونية - بنغازي
ا. عائشة امبيه الزهاوي - محاضر- كلية العلوم- جامعة بنغازي

Dr: Suad I. Al-Mugassabi - Associate Professor of Mathematics - Faculty of Science - University of Benghazi.
Emahl: *Souad.mugassabi@uob.edu.ly*
MA: Asma M. Muhammad - Assistant Lecturer - College of Arts and Sciences - Qminis - University of Benghazi.
Email: *aisha.mhamed@uob.edu.ly*
MA: Amani M. Al Shattat - Lecturer - Faculty of Electrical and Electronic Technology – Benghazi.
Email: *Amani.milad@ceet.edu.ly*
MA: Aisha A. Al-Zahawi - Lecturer - Faculty of Science - University of Benghazi.
Email: aisha.mhamed@uob.edu.ly

تاريخ تسليم البحث	تاريخ قبول البحث
2022 / 8 / 25	2022 / 9 /13

الملخص: تحليل كميات كبيرة من البيانات باستخدام دالة ويل. وقد تبين أن "متغيرات التردد" توفر معلومات مهمة تضيع في تقنيات أخرى. تنطبق هذه الطريقة على دراسات الحالة المزاجية للمعلومات في وسائل الإعلام وعلى الإنترنت. تم تقديم تحليل كمي للمعلومات الموجودة على الإنترنت فيما يتعلق بتسريب النفط لشركة برينتش بتروليوم في الولايات المتحدة. يعد التحليل السريع للمعلومات الواردة من الشبكة الاجتماعية (تويتر) حول الشؤون الجارية أمرًا مهمًا للمعلقين الحكوميين والسياسيين. وآخرون.

الكلمات المفتاحية: دالة ويل ، التردد الزمني ، تحليل البيانات ، الشبكات

Abstract: Analyze large amounts of data using Weyl functions. "Frequency variables" have been shown to provide important information that is lost in other techniques. This method applies to mood studies of information in the media and on the Internet. Presented is a quantitative analysis of information on the Internet regarding his BP oil spill in the United States. Rapid analysis of information from the social network (Twitter) on current affairs is important for government and political commentators. etc.

Key words: Weyl function, Time-frequency, Analysis data, Networks.

1. Introduction

Weyl functions and more generally time-frequency methods have long been studied in the context of mathematics, physics and engineering (especially signal processing) (e.g. [2,9,1]). There is a lot of research (e.g. [3,4,6,7,8]) on how to compress information in large amounts of data. In time-frequency analysis, each variable has a corresponding "frequency variable", providing very important information that is lost in other techniques. In this article, we apply these techniques to rapidly analyze and compress information from large datasets. This example is the same as using the [5] Wigner function, but in this article we will use the Weyl function.

2. Main results

Our example is a sentiment study on media and information on the internet. Here, we present a quantitative analysis of information on the Internet about BP oil spills in the United States. The first step is to describe the data in terms of the real function $F(x)$ of the variable x. In this example, x represents the information location (any integer value of x represents an internet address) and $F(x)$ is the number of comments at address x on BP's oil spill. Here we need a criterion for the order of the addresses (such as geographic location). We can include times t in the function (see Figure 1), but for simplicity we have not explored how the information varies as a function of time. We also consider a function of one variable, but more variables can explain the information more precisely (e.g. another variable y can be used to quantify the intensity of BP's feelings about the oil spill). The real function $F(x)$ is defined above for discrete values of x in the finite interval of the real line. By interpolation, assuming that $F(x)$ vanishes very quickly outside this finite interval (or assuming that $F(x)$ is a periodic function of x), then for all real values of x can be defined as $F(x)$ is also a real function, but it is convenient to transform it into a complex function $f(x)$ using the well-known Hilbert transform (called the "analytic signal" in signal analysis). Although $f(x)$ and $F(x)$ contain exactly the same information, it is easier to apply this method to the complex function $f(x)$. The Hilbert transform is available in any computer library (such as MATLAB). [5]

We study the Weyl function

$$W(x, \nu_x) = \int_{-\infty}^{\infty} f(y - \tfrac{1}{2}x) f^*(y - \tfrac{1}{2}x)\, e^{\nu_x y i}\, dy \qquad (1)$$

Corresponding to the function $f(x)$. Regions in the $x - \nu_x$ plane with large Weyl functions show strong activity compared to the information we are investigating at appropriate addresses and at appropriate frequencies. We searched for 'BP oil spill in USA' in Google news in [5]. We have used 19741 articles on this topic in 37 different websites (alphabetically sorted by $x = 1,2,\cdots,37$). These articles were published in two months. For simplicity we do not consider how the data change as a function of time. From these data we created a real function F(x) for $x = 1, 2, \cdots, 37$. By interpolation, we constructed the function F(x) for all real values in the interval $(1, 37)$. As already mentioned, for the remainder of x, we assume the function equals zero. Then we used the Hilbert transform to get the complex function $f(x)$. From this we calculated the Weyl function using Eq.(1). In Figure 1, we plot the Weyl function again at time t. The result for real function F(x) is shown in Figure 2 and the complex function $f(x)$ is plotted in Figure 3.

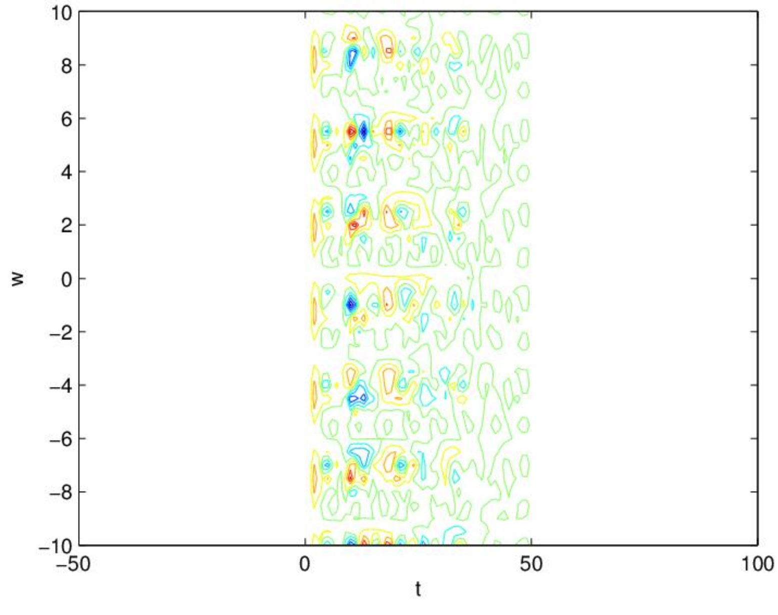

Figure 1: The Weyl function against to time (*t*).

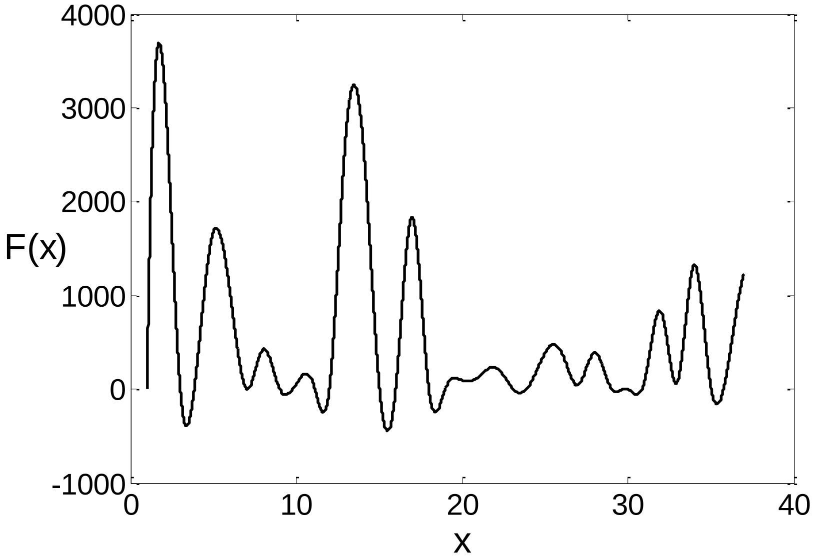

Figure 2: The real function *F*(*x*).

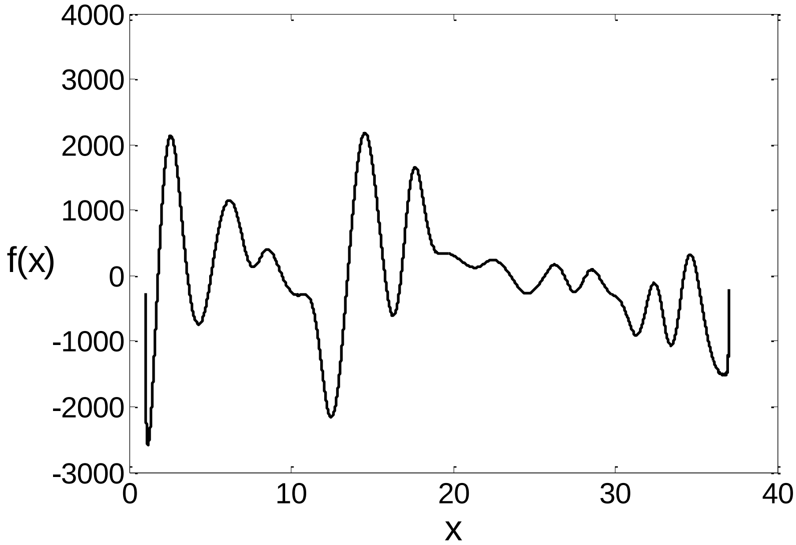

Figure 3: The complex function *f*(*x*).

The Weyl function is shown in Figure 4 with contour diagram. Figure 5 shows a contour diagram of the Weyl function is shown. The results indicate that there were many articles about the BP oil spill in the addresses labelled with $1 \le x \le 5$ and also in the addresses labelled with $11 \le x \le 18$. The Weyl function takes high values at high frequencies $15 \le \nu_x \le 40$. The merit of our approach is the fact that we can detect uniform or non-uniform behaviour in a neighbourhood of x (these are his website here).

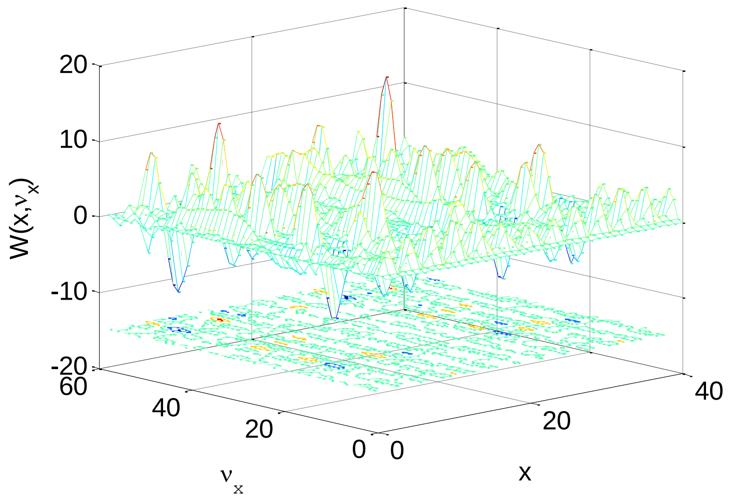

Figure 4: The Weyl function with contour diagram.

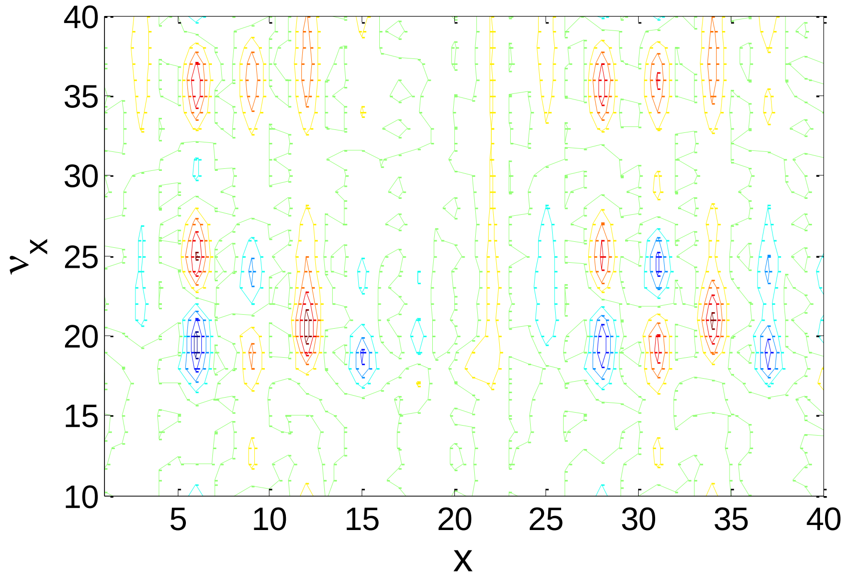

Figure 5: The contour diagram of Weyl function.

The regions with negative values are a well-known feature of the Weyl function related to interference. This interference is associated with many peaks in the function $F(x)$, which in the present context indicate a non-uniform (oscillating) distribution of the articles at different addresses.

3. Conclusions

Given the rapidly growing size and complexity of data sets generated by scientific experiments and observations of natural, social, and financial phenomena, the amount of data to be analyzed poses a major challenge in the future. The Weyl function is widely used in mathematics, physics, and engineering. In the current context, Weyl functions can be used to extract meaningful information from large amounts of data in near-real time. In particular, the frequency variable represents the uniformity of the variable's behavior in the environment. We considered the data as a function of the variable x (and corresponding frequency ν_x). More variables are available to describe the information more precisely. For example, in a BP oil spill incident, the value $1, \ldots, M$ can quantify the degree of opposition to the event. In this case the Weyl function is $W(x, y, \nu_x, \nu_y)$. Additionally, you can include a time variable that shows how your emotions change as a function of time. Our approach has a wealth of potential applications. In addition to the examples discussed so far, another important application is commercial (or national) security. For example, we can count keywords in emails from company employees to look for anomalies outside of normal patterns. The frequency variable does precisely that, by detecting non-uniform behavior among similar employees. Therefore the Wigner function approach can be very useful in this context. There are other distributions related to the Weyl function that can be used to extract information from large datasets. For example, the ambiguities or properties or Wigner functions (eg, refs [2,5]) are related to the Weyl functions by the two-dimensional Fourier transform. These two functions are related to the distribution instant of the data. The first two moments (which give the mean and standard deviation) and the other higher moments are widely used in data analysis, especially financial data. The aim of this paper is to show that, in the context of data analysis, the Weyl function reveals information through frequency variables that are not easily discernible over time. In summary, the Weyl function approach is a powerful technique for extracting information from large amounts of data.

References

- Chountasis, S., & Vourdas, A. (1998). Weyl functions and their use in the study of quantum interference. *Physical Review A*, *58*(2), 848.
- Cohen, L. (1995). Time-frequency analysis. *New Jersey: Prentice hall.*
- Cook, K., Earnshaw, R., & Stasko, J. (2007). Guest editors' introduction: Discovering the unexpected. *IEEE computer graphics and applications*, *27*(5), 15-19.
- Earnshaw, R. A., de Silva, M., & Excell, P. S. (2015). Ten unsolved problems with the internet of things. *International Conference on Cyberworlds (CW)* (pp. 1-7). IEEE.
- Earnshaw, R. A., Lei, C., Li, J., Mugassabi, S., & Vourdas, A. (2012). Large-scale data analysis using the Wigner function. *Physica A: Statistical Mechanics and Its Applications*, *391*(7), 2401-2407.
- Earnshaw, R., Guedj, R., Van Dam, A., & Vince, J. (Eds.). (2001). Frontiers of human-centered computing, online communities and virtual environments. *Springer Science & Business Media.*
- Magkonis, G., & Jackson, K. (2019). Identifying networks in social media: The case of# Grexit. *Networks and Spatial Economics*, *19*(1), 319-330.
- Qiu, Q., Thompson, A., Calderbank, R., & Sapiro, G. (2015). Data representation using the Weyl transform. *IEEE Transactions on Signal Processing*, *64*(7), 1844-1853.
- Tolimieri, R., & An, M. (1998). Time-frequency representations. *Springer Science & Business Media.*