Improved Maximum Entropy Method applied to Real-time Time-Dependent Density Functional Theory

M. Toogoshi, S. S. Kano and Y. Zempo
Computer and Information Sciences, Hosei University, 3-7-2 Kajino, Koganei, Koganei 184-8584 Japan
E-mail: 16t9001@cis.k.hosei.ac.jp

Abstract. The maximum entropy method (MEM) is one of the key techniques for spectral analysis. The main feature is to describe spectra in low frequency with short time-series data. We adopted MEM to analyze the spectrum from the dipole moment obtained by the time-dependent density functional theory (TDDFT) calculation in real time, which is intensively studied and applied to computing optical properties. In the MEM analysis, we proposed that we use the concatenated data set made from several-times repeated raw data together with the phase. We have applied this technique to spectral analysis of the TDDFT dipole moment of oligo-fluorene with \( n = 8 \). As a result, the higher resolution can be obtained without any peak shift due to the phase jump. The peak position is in good agreement to that of FT with just raw data. This paper presents the efficiency and characteristic features of this technique.

1. Introduction
Time-dependent density functional theory (TDDFT) is a powerful tool for analyzing optical properties of medium-to-large sized molecule. We employ a real-time and real-space technique to solve the time-dependent Kohn-Sham equations. Within the framework of this approach, the wave functions are calculated by the finite difference method on real spatial grids [1] without using explicit bases such as plane waves or Gaussian. In our procedure to calculate optical properties, we use the time-series data, namely the dynamic dipole moment, from whose Fourier transform (FT) optical properties are calculated in a usual technique. The spectral resolution depends on the length of the dipole moment. To obtain the good resolution, the computational cost is quite expensive.

To solve this difficulty, we focus on Maximum Entropy Method (MEM)[2], which is one of the key techniques of spectrum analysis. The principles of MEM are methods of the information theory for estimating unknown probability distributions based on the information about their expected values. MEM is widely used to the solution of a variety of problems related to spectral estimation such as earth and planetary science[3], spectroscopy[4]. The main feature is to obtain a fairly high resolution and accuracy with a relatively small number of time-series data. MEM is based on the Fourier pair of the autocorrelation and power spectrum. We applied this technique to spectral analysis of time-dependent dipole moments of molecules, which is calculated by real-time TDDFT. In the comparison of MEM and FT spectrum for several molecules, we realize that MEM needs less time step than FT to obtain the optical spectrum with same resolution[6]. As a new improved MEM, we proposed to use the concatenated data set made from several-times
repeated raw data[7] together with the phase to avoid the side effect of the artificial periodicity. With this improvements, we have successively obtained the much better spectral resolution of the target peak. Thus, this improvements will make a further advantage of MEM.

We applied this technique to spectral analysis of the TDDFT dipole moment of typical some molecules such as benzene, oligo-fluorene and other materials. In the analysis of optical properties, we are interested in the lower energy peak, corresponding to the band gap area. The results show the higher resolution and the emphasized peak near the band gap without being affected by the artificial periodicity. In this paper, we present characteristic features of this technique, comparing with the FT and the conventional MEM.

The paper is organized as follows. In section 2, we briefly describe a procedure of our TDDFT calculation. Then, we recall a basis of MEM and our new techniques. In section 3, results of absorption spectra of molecules are presented and discussed. Finally, we summarize the efficiency and characteristic features of our technique.

2. Method

2.1. Time-dependent density functional theory

In this section, we briefly describe the procedure of our TDDFT calculation[8]. The basis is the density functional theory (DFT) [9] with the local density approximation (LDA). The total energy of the ground state can be derived from the Kohn-Sham equation (KS) [10]. DFT is much less successful in describing optical responses and absorption spectra where electronic excited states are involved. However, this difficulty is, in principle, solved by the extension of DFT to its time-dependent version, TDDFT, which was established by Runge and Gross [11].

In analogy to the time-independent case, the TDDFT equation of motion coupled with pseudopotentials is given by

$$\left\{ -\frac{1}{2}\nabla^2 + V_{\text{ion}}^{ps}(\mathbf{r}) + V_H(\mathbf{r}, t) + V_{XC}[\rho(\mathbf{r}, t)] + V_{\text{ext}}(\mathbf{r}, t) \right\} \psi_j(\mathbf{r}, t) = i\frac{\partial}{\partial t} \psi_j(\mathbf{r}, t),$$

(1)

where $V_{\text{ion}}^{ps}$ is an ionic pseudopotential, $V_H$ is the Hartree potential, and $V_{XC}$ is the exchange-correlation potential. Since the exact time-dependent xc kernel is not known, the originally non-local time-dependent xc kernel is replaced with a time-independent local one. This is reasonable when the density varies slowly with time. This approximation allows the use of a standard local ground-state xc functional in the TDDFT framework. The Hartree and exchange-correlation potentials can be determined from the electronic charge density,

$$\rho(\mathbf{r}, t) = \sum_j |\psi_j(\mathbf{r}, t)|^2.$$  

(2)

The summation is over all occupied states $j$. The Hartree potential is determined by $\nabla^2 V_H = -4\pi \rho$, and as the xc potential $V_{XC}$, the usual local density approximation (LDA) is used in our study. For the ionic potential, we employed the pseudopotential $V_{\text{ion}}^{ps}$ in the separable form so that the only valence electrons were considered[12]. Prior to the calculation of optical responses, we first solved the usual, time-independent formulation of the pseudopotential-DFT method[13] to obtain the optimized electronic structure[14]. Then, we applied an external field $V_{\text{ext}}$ to the system as a perturbation and followed linear responses of the system in real time. Although the xc functional we adopted is originally for electronic ground states, we also employed it in our practical calculations. This method of TDDFT has been effectively used, for the cases where the potential was the time-dependent, to study behaviors of electrons in oscillating electric and magnetic fields and hence excited-state reactions[15]. In our calculations, the real-time and real-space technique was adopted in solving Eq.(1) by means of the finite difference approach[1]. The uniform grid was used in our study for the simplicity.
The time dependent wave function is given by \( \psi(t) = \exp[-iHt] \psi(0) \), with the initial wave function \( \psi(0) \), where \( H \) is the Hamiltonian of the system, and \( k \) is a small wave number corresponding to the external perturbation in the \( z \) direction. In the linear response, the time-dependent polarizability is proportional to the dipole matrix element \( \mu_\xi(t) = \langle \psi(t) | \xi | \psi(t) \rangle \), where \( \xi \) represents \( x, y \) and \( z \). The polarizability \( \alpha_\xi(\omega) \) is numerically calculated using Fourier transformation (FT) of \( \mu_\xi(t) \) and averaged as \( \alpha = (\alpha_x + \alpha_y + \alpha_z)/3 \). The oscillator strength distribution \( S(\omega) \) is related to the imaginary part of the polarizability,

\[
S(\omega) = \frac{2\omega}{\pi} \text{Im} \alpha(\omega).
\]

2.2. Maximum entropy method

In this section, we briefly describe the procedure and derivation of MEM. The entropy in information theory has been recognized as a measure of uncertainty[16, 17]. Any inferences made from incomplete information should be with the probability distribution which maximizes the entropy under the constraints of available information[18].

A framework of the present study is based on Burg’s method[2]. The time-series data \( \mu \) to be analyzed is given by real-time TDDFT, \( \mu(n) = \mu(n\Delta t) \). If samples obey Gaussian distribution, the entropy \( h \) of a Gaussian process is given by

\[
h / \Delta t = \int_{-\pi/\Delta t}^{\pi/\Delta t} \log P(\omega) d\omega,
\]

where \( \Delta t \) is a sampling interval, and \( P(\omega) \) is the power spectrum. We find \( \hat{P}(\omega) \) which Eq.(5) is maximized under the constraints of the Wiener-Khintchin’s theorem,

\[
P(\omega) = \Delta t \sum_{m=-\infty}^{\infty} C_m \exp(-i\omega m \Delta t)
\]

\[
C_m = \frac{1}{2\pi} \int_{-\pi/\Delta t}^{\pi/\Delta t} P(\omega) \exp(i\omega m \Delta t) d\omega.
\]

The solution is

\[
\hat{P}(\omega)/\Delta t = \left| \frac{\beta}{1 + \sum_{m=1}^{M} a_m \exp(-i\omega m \Delta t)} \right|^2,
\]

where \( M \) is the maximum length of autocorrelation \( C_M \). The parameters \( a_k \) and \( \beta \) are the Lagrange multipliers and they are the solution of Yule-Walker equation[19],

\[
\begin{pmatrix}
C_0 & C_1 & \cdots & C_M \\
C_1 & C_0 & \cdots & C_{M-1} \\
\vdots & \vdots & \ddots & \vdots \\
C_M & C_{M-1} & \cdots & C_0
\end{pmatrix}
\begin{pmatrix}
1 \\
a_1 \\
\vdots \\
a_M
\end{pmatrix}
= 
\begin{pmatrix}
|\beta|^2 \\
0 \\
\vdots \\
0
\end{pmatrix}.
\]

We adopt the Levinson-Durbin algorithm to solve Eq.(9) efficiently. To solve Eq.(9), the autocorrelation \( C_m \) at the lag \( m \) is

\[
C_m = \frac{1}{N} \sum_{n=0}^{N-1-m} \mu^*(n) \mu(n + m),
\]
where $N$ is the length of the time-series data, and a symbol $*$ is a complex conjugate. Generally, the autocorrelation shows the periodicity of the signal based on a lag. Given the large lag, it makes the estimation of lower frequency components more reliable. However, the values of high-order autocorrelations are almost close to zero with the too large lag $M$. The effective value of $M$ is restricted by the number of the data $N$. Thus, we need a large enough $N$ for a reasonable spectral resolution of the low frequency part.

2.3. Improvements of MEM

As a new improved MEM, we propose to use the concatenated data set made from several-times repeated raw data together with the phase to avoid the side effect of the artificial periodicity.

The information of low energy parts is important for the optical spectrum analysis. To emphasize this part, we need a large $M$ which is the maximum lag of autocorrelation. However, the unphysical results such as peak splits and false peaks are appeared by selecting too large value of $M$ from our experience. Also, we cannot take the larger $M$ where $M$ is limited by the total number of time step $N$. To solve this difficulty, we duplicate the raw data and repeat it many times as shown in Fig. 1 (b). The lower energy resolution is attributed to the maximum lag $M$. With the repeated signal, the value of $M$ may be selected large enough without causing additional computations, which save our calculation cost.

At the connection point, there may be a phase jump for each frequency component. This phase jump is considered as a noise. The side effect of phase jump may be added, for example the peak shifts. To reduce the side effect fo the phase jump, we propose that an appropriate phase for target frequency is introduced together with the repeated data

$$\tilde{\mu}(n) = \mu'(n) \exp(ik\phi),$$

where $\mu'(n)$ is the several repeated raw data, and $\tilde{\mu}(n)$ is the concatenated data $\mu'(n)$ with an appropriate phase $\phi$ for the target frequency ($-\pi \leq \phi \leq \pi$). Even if we introduce this phase, the total power of the spectrum doesn’t change. Then an absolute value of exponential in Eq. (11) is equal to 1.

Figure 1(c) shows the repeated data $N = 5000 \times 4$ with the phase $\phi = 0.25\pi$ in the case of the oligomer of fluorene with $n = 8$. In this case, $k$ in Eq.(11) is equal to zero, then the value of a x-axis is in the area of $[0, 5)$; one in the area of $[5, 10)$; two in the area of $[10, 15)$; three in the area of $[15, 20)$. Since the phase $\phi$ is a constant value, it is necessary to choose the appropriate phase for the target frequency to reduce the side effect of the phase jump. We decide this phase to match the bandgap peak in our study.

3. Results and Discussion

In this section, we show the calculated absorption spectrum of the total oscillator strength of improved MEM. Then, we discuss characteristics of our method compared to the result of conventional MEM, and FT.

3.1. Benzene, Naphthalene, Anthracene and Tetracen

As a simple example to confirm the efficiency of our proposed method, we applied improved MEM to benzene, naphthalene, anthracene, and tetracene. The molecular structure is based on the ground state. The time evolution is carried out according to the real-time TDDFT [14, 8]. The spectrum $S(E)$ is calculated from the time-series data in our MEM technique. Results are shown in Fig.3.1, where the solid line is the result of benzene, the dashed line is that of naphthalene, the dash-dot line is that of anthracene, and the dotted line is that of tetracene. These spectra are normalized by the intensity of first peak. Figure 3.1(a) shows results of improved MEM. The parameters($M$ and $\phi$) of our MEM are $M = 1800$ and $\phi = -0.25\pi$ at
Figure 1. Comparison of the time-series data for oligo-fluorene with \( n = 8 \), which is made by real-time TDDFT. (a) Real time evolution up to \( N = 20000 \) for FT. (b) Four times repeated data of the dipole moment with \( N = 5000 \times 4 \). (c) The data (b) together with the phase \( \phi = 0.25\pi \). The solid line is the real part of the data, and the dashed line is the imaginary part of the data.

benzene; \( M = 2500 \) and \( \phi = -0.20\pi \) at naphthalene; \( M = 2000 \) and \( \phi = -0.23\pi \) at anthracene; \( M = 3000 \) and \( \phi = -0.21\pi \) at tetracene. For both MEM calculations, the same number of the time step \( N = 10000 \times 100 \) is adopted. Figure 3.1(b) shows results of FT with \( N = 10000 \). Our method provides the good resolution compared with that of FT. Also, we can see the clear spectrum in comparison of FT. This is a one of the feature of our MEM.

3.2. Fluorene
Poly(9,9-dialkyl-fluorene) and their substituted derivatives are used for one of the Organic LED materials, and expected to be basic materials for a blue emission. Their electronic structures have been extensively studied [20, 21, 22, 23, 24]. We employed the oligomer of fluorene (oligo-FL) with \( n = 8 \) and carried out the MEM calculation. The molecule size is fairly big, then the calculation for each step is quite expensive. To obtain the low energy part of the spectrum, of course, it takes a long time to perform the calculation. If we can save the calculation cost to figure out the lower energy part, it is quite effective. To this signal analysis, we applied our improved MEM. Figure 3 shows the comparison of the absorption spectrum with four methods. The solid line shows the result of improved MEM with \( N = 5000 \times 100, M = 2500 \) and \( \phi = 0.25\pi \). The dashed line is that of FT for \( N = 20000 \), the dash-dot line is our conventional
Figure 2. Comparison of improved MEM and FT spectrum using benzene(solid line), naphthalene(dashed line), anthracene(dash-dot line) and tetracene(dotted line). (a) is results of improved MEM using $N = 10000 \times 100$ together with $M = 1800$ and $\phi = -0.25\pi$ at benzene; $M = 2500$ and $\phi = -0.20\pi$ at naphthalene; $M = 2000$ and $\phi = -0.23\pi$ at anthracene; $M = 3000$ and $\phi = -0.21\pi$ at tetracene. (b) is results of FT with $N = 10000$. All the spectra are normalized at the first low energy peak.

MEM spectrum with just the real time steps $N = 5000$ and $M = 2500$. The dotted line is that of MEM with $N = 5000 \times 100$ and $M = 2500$. These spectra are normalized at the second low energy peak.

FT(dashed line) and MEM(dash-dot line) provide the broad peak at the bandgap area. With only the repeated signal, the strength of the first peak emphasizes, and the peak position shifts for the side effect of the phase jump[7]. On the other hand, we can obtain the clear first peak of our new method. In this case, we decide the matched phase corresponding to the first peak. Thus, the position of the second peak differs from that of other methods. If we choose the mismatched phase for the target peak, the position of that peak shifts, and the strength decreases. The bandgap peak at $2.53\ eV$ for improved MEM may correspond to the experimentally observed one at $3.56\ eV$ for poly-FL. This discrepancy is due to the inherent problem to DFT[8].

3.3. C60 fullerene
Our new MEM is not always works well for the target frequency. In this section, we show the case of our MEM does not work well.

Figure 4 shows that the result of C60 fullerene as the bad example of our MEM calculation. The solid line is the result of our MEM with hundred repeated signals $N = 10000 \times 100$ together with the phase $\phi = -0.125\pi$ and $M = 6000$. The dashed line shows that of FT with $N = 10000$. The result of FT has better resolution for first peak in comparison of that of MEM.

If the length of data $N$ is not enough, we obtain the quite broad peak like Lorentz function in our MEM calculation. In particular, we can see that the first peak is not separated due to the lack of the information of that peak. We cannot obtain the low energy component which is not included in the data $N$. Thus, we have to calculate a certain time evolution until the information of the low energy part is obtained.

4. Concluding remarks
We employed MEM to obtain optical spectra of molecules, and the spectral resolution was quite improved, which is better than that of FT with the same number or total time steps. As a new aspect, we introduced the use of repeated the same signal together with the phase to make apparently long data, in which larger time lag was involved. In particular, it is efficient to find the low energy peak. We obtained quite better resolutions for target peak without adding practical machine time. This means that we can save our calculation cost efficiently. However,
Figure 3. Comparison of the absorption spectrum for oligo-fluorene with $n = 8$. The dashed line is the result of FT with $N = 20000$, the dash-dot line is that of MEM with $N = 5000$, the dotted line is the that of MEM with repeated repeated data $N = 5000 \times 100$, and the solid line is that of improved MEM with $N = 5000 \times 100$, $M = 2500$ and $\phi = 0.25\pi$. Spectra are normalized at the second low energy peak.

Figure 4. Absorption spectrum for C60 fullerene. The dashed line is the result of FT with $N = 10000$, and the solid line is that of improved MEM with $N = 10000 \times 100$, $M = 6000$ and $\phi = -0.125\pi$. Spectra are normalized at the highest peak in the area of 0-10 eV. The experimental absorption peaks at 3.8, 4.8 and 5.8 eV [14].
our new MEM is not always works well for target peak. With this method for the spectrum analysis, we cannot obtain the low energy component which is not included in the data. We have to calculate a certain level of the time evolution until the information of the low energy part is obtained.

Acknowledgments
Authors would like to express their acknowledgements to Professor H. Hanaizumi for useful discussions and encouragements. This work was partially supported by JSPS Grants-in-Aid for Scientific Research (C) Grant number 25390158, Sumitomo Chemical, Co. Ltd. and Simulatio Corporation.

References
[1] Chelikowsky J, Troullier N, Wu K and Saad Y 1994 Phys. Rev. B50, 11355
[2] Burg J P 1968 Advanced Study Institute on Signal Processing, 15-0, NATO, Erschede, Netherlands
[3] Sakurai T 2002 Earth Planets Space 54, 153
[4] Vartiainen E M, Peiponen K E, Kishida H and Koda T 1996 J. Opt. Soc. Am. B13, 2106.
[5] Smith C R and Erickson G J 1987 "Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems", ed Smith C R and Erickson G J, D. Reidel Publishing Company(Holland)
[6] Toogoshi M, Kato M, Kano S S, Zempo Y 2014 J. Phys. Conf. Ser. 510, 012027
[7] Toogoshi M, Kano S S, Zempo Y 2015 J. Phys. Conf. Ser. 640, 012069
[8] Zempo Y, Akino, Ishida M, Ishitobi M and Kurita Y 2008 J. Phys. Condens. Matter 20, 064231
[9] Hohenberg P and Kohn W 1964 Phys. Rev. 136, B864
[10] Kohn W and Sham L. J 1965 Phys. Rev. 140, A1133
[11] Runge E and Gross E K U 1984 Phys. Rev. Lett. 52, 997
[12] Troullier N and Martins J L 1991 Phys. Rev. B43, 1993
[13] e.g. Pickett W E 1989 Comput. Phys. Reports 9, 115
[14] Yabana K and Bertisch G F 1996 Phys. Rev. B54, 4484
[15] Tateyama Y, Oyama N, Ohno T and Miyamoto Y 2006 J. Chem. Phys. 124, 124507
[16] Shannon C 1948 Bell Syst. Tech. J. 27, 379 and 623
[17] Jaynes E T 1957 Phys. Rev. 106, 620
[18] Jaynes E T 1982 Proc. IEEE 70, 939
[19] Haykin S and Kesler S 1979 "Non linear Methods of Spectral Analysis", ed Haykin S, Springer(Berlin)
[20] e.g. Hummer K, Puschning P, Sagmeister S and Ambrosch-Draxl C 2006 Mod. Phys. Lett. B20, 261 ; and the references therein.
[21] Yu W L, Cao Y, Pei J, Huang W and Heeger A J 1990 Appl. Phys. Lett. 75 3270
[22] Wang J F, Feng J K, Ren A M, Liu X D, Ma Y G, Lu P and Zhang H X 2004 Macromolecules 37, 3451
[23] Brière J F and Côté M 2004 J. Chem. Phys. B108, 3123
[24] Grice A W, Bradley D D C, Bernius, M T, Inbasekaran M, Woo EP and Wu W W 1998 Appl. Phys. Lett. 73, 629