

## Iterative projection algorithm for retrieval of angle-resolved single-molecule dipoles from high-harmonic spectra

Siqi Sun,<sup>1</sup> Yanqing He,<sup>1</sup> Lixin He,<sup>1,2,\*</sup> Jianchang Hu,<sup>1</sup> Pengfei Lan,<sup>1,2,†</sup> and Peixiang Lu<sup>1,2,3</sup>

<sup>1</sup>Wuhan National Laboratory for Optoelectronics and School of Physics, Huazhong University of Science and Technology, Wuhan 430074, China

<sup>2</sup>Optical Valley Laboratory, Hubei 430074, China

<sup>3</sup>CAS Center for Excellence in Ultra-intense Laser Science, Shanghai 201800, China



(Received 5 September 2022; revised 15 November 2022; accepted 23 February 2023; published 9 March 2023)

High-order harmonic spectroscopy (HHS), as an emerging technique, provides a way to detect molecular structures with ångström spatial resolution and ultrafast electron dynamics in molecules on an attosecond timescale. However, the implementation of HHS still encounters great difficulties due to the coherent average of single-molecule emissions from different alignment angles. Here, we provide a machine learning algorithm based on the iterative projection method to fully retrieve the complex single-molecule dipole of a linear molecule from the harmonic spectra measured at different delays and polarization angles between the alignment and driving pulses. We demonstrated our algorithm both theoretically and experimentally using the N<sub>2</sub> molecule. The results show that our algorithm can accurately retrieve the complex single-molecule dipole from the harmonic spectra with good noise stability and robustness. From the retrieved single-molecule dipole, the contributions of multiple orbitals in high harmonic generation are identified even with a low degree of alignment.

DOI: [10.1103/PhysRevA.107.033105](https://doi.org/10.1103/PhysRevA.107.033105)

### I. INTRODUCTION

Probing the structure and internal dynamics of the molecules is essential for understanding the physics of complex chemical and biological reactions. To this end, a high spatial resolution of the order of an ångström as well as an attosecond temporal resolution [1–3] are required. High-order harmonic generation (HHG) through laser-molecule interaction provides one way to achieve this goal which is usually called high-order harmonic spectroscopy (HHS). The HHS relies on a built-in pump-probe process [4,5]: an electron is first tunnel ionized (pump) by the strong laser field, subsequently accelerated in the external electric field, and finally brought back to recombine with (probe) the parent ion by releasing its kinetic energy by emitting harmonic photons. The returning electron acts as an ultrafast “probe” during recombination, recording information about the molecular structure and dynamics in the harmonic spectrum. As a result, the high-harmonic spectrum contains abundant information on the structure and dynamics of the emitting molecule, which provides a promising way for detecting the molecular structure and electron dynamics with both attosecond temporal resolution and ångström spatial resolution in the same measurement. Up to now, HHS has been used in molecular orbital tomography [6,7], probing molecular vibration [8], and proton dynamics in molecules [9], decoding the underlying attosecond multielectron dynamics [10], probing the structure

of the excited molecules [11], imaging the attosecond wave packet [12], and monitoring the attosecond charge migration in molecules [13,14].

In HHS, the molecular frame information is encoded in the harmonic emission on the single-molecule level [15–19]. However, in the experiment, the molecules are generally randomly distributed in space, the single-molecule information will be lost after averaging over the random distribution. Even if the laser-induced molecular alignment technique was developed [20–28] to address this issue, the molecular ensemble can never be perfectly aligned in the experiment. The measurement is still an average response over the individual emissions from the molecule directed at different angles. The average response over the molecular alignment distribution can deviate the measurement from the actual single-molecule response, reducing the accuracy or destroying the fact of reconstruction [14]. Therefore, it is essential to unravel the internal angular single-molecule information from the measurement. Previously, the so-called rotational coherence spectroscopy (RCS) used the laser-induced molecular alignment technique to obtain the time-varying photonic or electronic signals modulated by a time-dependent molecular alignment distribution. It can retrieve the single-molecule information, for example, molecular-frame angular distributions of photoelectrons [29], fixed-in-space tunneling or multiphoton ionization rates [30–33], and the time dependence of molecular alignment distributions [16,34]. Under these circumstances, except for [29], the information to be retrieved is positive real quantities, which is different from the HHS. In HHS, it encounters significant challenges, primarily because of the coherent nature of HHG [35,36]. In most

\*helx\_hust@hust.edu.cn

†pengfeilan@hust.edu.cn

previous studies of HHG from aligned molecules, this internal coherence is addressed either under the perfect alignment approximation by ignoring angle dependence [6,7,12,13] (i.e., assigning measurements to the response of the molecules with maximal distribution probability), or by using an incoherent treatment to ignore the harmonic phases [19,37]. However, such processing effectively overlooks the most important coherent nature of HHG. Recently, the same idea of RCS has been expanded to fully retrieve the angle-resolved complex dipole in the molecular frame [38,39], but they all remained in the theoretical stage. Methods for achieving a true measurement of the single-molecule harmonic response have been long desired, but are still out of reach. Recently, a machine learning algorithm has been shown to be very useful in high-order harmonic spectroscopy for some inverse problems, for example, retrieving molecular axis distribution [16], detecting multiple chiral centers in chiral molecules [40], and reconstructing band structure and pulse waveform for high-order harmonic spectroscopy in solids [41].

Here, we develop a robust machine learning algorithm based on the iterative projection method to fully retrieve the alignment-angle-resolved complex dipole moments of HHG for linear molecules at the single-molecule level from the harmonic spectra measured at different delays and polarization angles between the alignment and driving pulses. Our algorithm is first tested on simulated spectra to verify the accuracy and noise stability. Then, we apply our algorithm to real spectra measured in the experiment of  $N_2$  and successfully retrieve the single-molecule dipole moments. The results show that the harmonics in the plateau region are mainly contributed to by the highest occupied molecular orbital (HOMO) of  $N_2$ . While for harmonics near the cutoff region, the contribution of HOMO-1 is clearly identified even if the alignment degree is low in our experiment.

## II. THEORY

In this section, we present essential elements for retrieving the single-molecule dipole moment from the harmonic spectra measured at different delays and polarization angles between the alignment and driving pulses. These include the physical model of laser-induced nonadiabatic field-free molecular alignment, the connection between the single-molecule response and angle-resolved time-dependent harmonic spectra, the reconstruction problem, which is formalized as an optimization problem, and our machine learning algorithm based on iterative projection.

### A. Laser-induced nonadiabatic field-free molecular alignment and convolution of the single-molecule response

Theories of laser-induced molecular alignment have been widely studied in the literatures [26–28,42]. Using the rigid

rotor approximation, the rotational motion of the linear molecule with initial state  $\Psi_{JM}$  excited by a short laser pulse can be described by the time-dependent Schrödinger equation

$$i \frac{\partial \Psi_{JM}(\theta, \phi, \tau)}{\partial \tau} = \left[ \eta J^2 - \frac{E(\tau)^2}{2} (\alpha_{\parallel} \cos^2 \theta + \alpha_{\perp} \sin^2 \theta) \right] \times \Psi_{JM}(\theta, \phi, \tau). \quad (1)$$

Here,  $E(\tau)$  is the electric field of the alignment pulse.  $\eta$  is the rotational constant of the molecule.  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  are the anisotropic polarizabilities in the parallel and perpendicular directions with respect to the molecular axis, respectively.

The time-dependent molecular axis distribution  $\rho(\theta, \phi, \tau)$  can be written as a weighted average of the modulus square of the wave function  $\Psi_{JM}(\theta, \psi, \tau)$ ,

$$\rho(\theta, \psi, \tau) = \sum_{JM} \Gamma_{JM} |\Psi_{JM}(\theta, \psi, \tau)|^2, \quad (2)$$

where  $\Gamma_{JM}$  is the statistical weight of the initial states. The initial states are given by the Boltzmann distribution corrected by the spin statistical weights as the molecules are initially in thermal equilibrium.

The harmonic signal emitted from aligned molecules can be expressed as the modules square of the convolution of the time-dependent molecular axis distribution  $\rho(\theta, \phi, \tau)$  with the dipole moment of the single-molecule response  $D_q(\Theta)$  [15,43,44]:

$$I_q(\alpha, \tau) = \left| \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} D_q(\Theta) \rho(\theta, \phi, \tau) \sin \theta d\theta d\phi \right|^2, \quad (3)$$

where  $q$  is the harmonic order.  $\theta$  and  $\phi$  are the polar and azimuthal angles of the molecular axis with respect to the polarization of the alignment laser, which is the  $z$  axis of the laboratory coordinate system and the  $x$  axis of the laboratory coordinate system.  $\Theta$  is the polar angle in the molecular coordinate system which means the angle between the molecular axis and the polarization of the driving laser.  $\alpha$  is the polarization angle between the driving laser and the alignment laser. The delay  $\tau$  and polarization angle  $\alpha$  are scanned to obtain the harmonic spectra (see Fig. 1). Angle  $\Theta$  can be expressed in the laboratory coordinate system as

$$\cos \Theta = \sin \theta \sin \alpha \cos \phi + \cos \theta \cos \alpha. \quad (4)$$

Since perfect alignment cannot be achieved in the experiment. The measured signal  $I_q(\alpha, \tau)$  will deviate from the single-molecule response  $D_q(\Theta)$ . Therefore, it is essential to disentangle the internal angular coherence from the measurements.

### B. Problem formalization

To retrieve the single-molecule dipole from the averaged harmonic signal, we have to deal with an inverse problem of Eq. (3). It is difficult to solve this inverse problem directly because of the nonlinearity and ill-posedness of Eq. (3). In our reconstruction, we first rewrite Eq. (3) to deal with the nonlinearity of the problem. Considering that the molecular axis

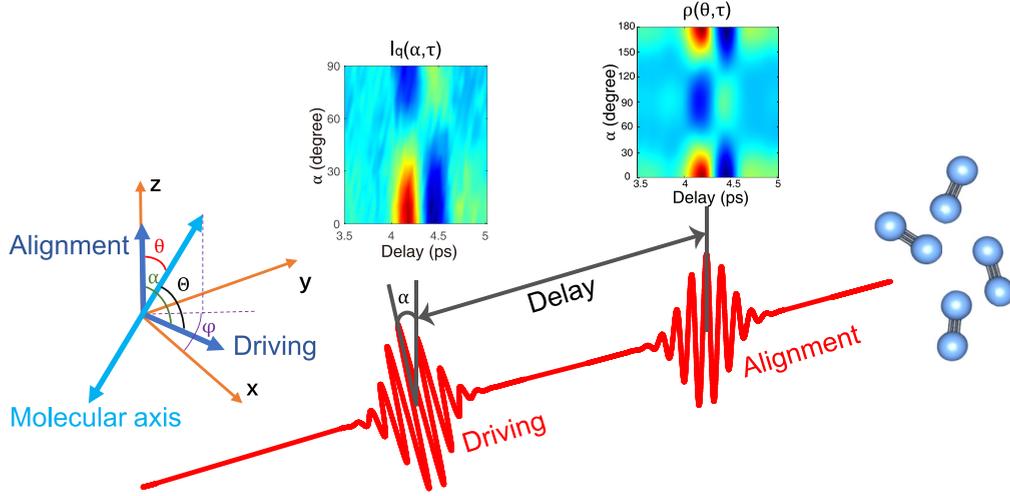


FIG. 1. Schematic diagram of our experiment. The alignment pulse polarized along the  $z$  axis is first applied at  $\tau = 0$  to align the molecules. The high-order harmonics are measured by scanning the delay and the polarization direction of the driving pulse in the  $x$ - $z$  plane.  $\rho(\theta, \tau)$  is the time-dependent molecular axis distribution after the alignment pulse interacting with the molecules.  $I_q(\alpha, \tau)$  is the harmonic signal generated by the driving pulse.

distribution is independent of  $\phi$  in the linearly polarized alignment laser, Eq. (3) can be expanded as

$$\begin{aligned} I_q(\alpha, \tau) &= \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} D_q(\theta, \phi, \alpha) d\phi \rho(\theta, \tau) \sin \theta d\theta \left( \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} D_q(\theta, \phi, \alpha) d\phi \rho(\theta, \tau) \sin \theta d\theta \right)^* \\ &= \int_{\theta_1=0}^{\pi} \int_{\theta_2=0}^{\pi} \int_{\phi_1=0}^{2\pi} \int_{\phi_2=0}^{2\pi} D_q(\theta_1, \phi_1, \alpha) D_q^*(\theta_2, \phi_2, \alpha) d\phi_1 d\phi_2 \rho(\theta_1, \tau) \rho(\theta_2, \tau) \sin \theta_1 \sin \theta_2 d\theta_1 d\theta_2. \end{aligned} \quad (5)$$

Here, we define

$$M_{q,\alpha}(\theta_1, \theta_2, \phi_1, \phi_2) \equiv \text{Re} [D_q(\theta_1, \phi_1, \alpha) D_q^*(\theta_2, \phi_2, \alpha)], \quad (6)$$

$$\rho(\theta_1, \theta_2, \tau) \equiv \rho(\theta_1, \tau) \rho(\theta_2, \tau) \sin \theta_1 \sin \theta_2. \quad (7)$$

Then Eq. (5) becomes

$$I_q(\alpha, \tau) = \int_{\theta_1=0}^{\pi} \int_{\theta_2=0}^{\pi} \int_{\phi_1=0}^{2\pi} \int_{\phi_2=0}^{2\pi} M_{q,\alpha}(\theta_1, \theta_2, \phi_1, \phi_2) d\phi_1 d\phi_2 \rho(\theta_1, \theta_2, \tau) d\theta_1 d\theta_2. \quad (8)$$

Note that the imaginary part of  $D_q(\theta_1, \phi_1, \alpha) D_q^*(\theta_2, \phi_2, \alpha)$  vanishes after convolution because of its asymmetry upon the exchange of  $\theta_1, \theta_2$  and  $\phi_1, \phi_2$ , thus is left out in Eq. (6).

By discretizing Eq. (8) and writing it in matrix form, we can obtain a linear relationship between  $M_{q,\alpha}(\theta_1, \theta_2, \phi_1, \phi_2)$  and  $I_q(\alpha, \tau)$ ,

$$\mathbf{I}_{q,\alpha} = \mathbf{P} \mathbf{M}_{q,\alpha}, \quad (9)$$

with

$$\mathbf{I}_{q,\alpha} \equiv \begin{pmatrix} I_{q,\alpha}(\tau_1) \\ \vdots \\ I_{q,\alpha}(\tau_K) \end{pmatrix}, \quad (10)$$

$$\mathbf{P} \equiv d\theta_1 d\theta_2 \begin{pmatrix} \rho(\theta_1^1, \theta_2^1, \tau_1) & \cdots & \rho(\theta_1^1, \theta_2^N, \tau_1) & \rho(\theta_1^2, \theta_2^1, \tau_1) & \cdots & \rho(\theta_1^2, \theta_2^N, \tau_1) & \cdots & \rho(\theta_1^N, \theta_2^N, \tau_1) \\ \rho(\theta_1^1, \theta_2^1, \tau_2) & \cdots & \rho(\theta_1^1, \theta_2^N, \tau_2) & \rho(\theta_1^2, \theta_2^1, \tau_2) & \cdots & \rho(\theta_1^2, \theta_2^N, \tau_2) & \cdots & \rho(\theta_1^N, \theta_2^N, \tau_2) \\ \vdots & & & & & & & \\ \rho(\theta_1^1, \theta_2^1, \tau_K) & \cdots & \rho(\theta_1^1, \theta_2^N, \tau_K) & \rho(\theta_1^2, \theta_2^1, \tau_K) & \cdots & \rho(\theta_1^2, \theta_2^N, \tau_K) & \cdots & \rho(\theta_1^N, \theta_2^N, \tau_K) \end{pmatrix}, \quad (11)$$

$$\mathbf{M}_{q,\alpha} \equiv d\phi_1 d\phi_2 \sum_{i=1}^J \sum_{j=1}^J \begin{pmatrix} M_{q,\alpha}(\theta_1^1, \theta_2^1, \phi_1^i, \phi_2^j) \\ \vdots \\ M_{q,\alpha}(\theta_1^1, \theta_2^N, \phi_1^i, \phi_2^j) \\ M_{q,\alpha}(\theta_1^2, \theta_2^1, \phi_1^i, \phi_2^j) \\ \vdots \\ M_{q,\alpha}(\theta_1^2, \theta_2^N, \phi_1^i, \phi_2^j) \\ \vdots \\ M_{q,\alpha}(\theta_1^N, \theta_2^N, \phi_1^i, \phi_2^j) \end{pmatrix}. \quad (12)$$

Here,  $K$  is the sampling number of time points.  $N$  is the sampling number of  $\theta_1$  and  $\theta_2$  and  $J$  is the sampling number of  $\phi_1$  and  $\phi_2$ . The fully discretized  $M_{q,\alpha}(\theta_1, \theta_2, \phi_1, \phi_2)$  has billions of elements. Therefore, it cannot be directly solved and we have to reduce the dimension of  $M_{q,\alpha}(\theta_1, \theta_2, \phi_1, \phi_2)$ . Note that, in the molecular coordinate system, the single-molecule dipole only depends on the angle  $\Theta$ . Thus, we can define

$$R_q(\Theta_1, \Theta_2) \equiv D_q^*(\Theta_1)D_q(\Theta_2), \quad (13)$$

and then discretize  $R_q(\Theta_1, \Theta_2)$  to write it into matrix form.

We define

$$\mathbf{R}_q \equiv \begin{pmatrix} R_q(\Theta_1^1, \Theta_2^1) & R_q(\Theta_1^1, \Theta_2^2) & \cdots & R_q(\Theta_1^1, \Theta_2^N) \\ R_q(\Theta_1^2, \Theta_2^1) & R_q(\Theta_1^2, \Theta_2^2) & \cdots & R_q(\Theta_1^2, \Theta_2^N) \\ \vdots & \vdots & \ddots & \vdots \\ R_q(\Theta_1^N, \Theta_2^1) & R_q(\Theta_1^N, \Theta_2^2) & \cdots & R_q(\Theta_1^N, \Theta_2^N) \end{pmatrix}, \quad \mathcal{G}(\mathbf{R}_q) \equiv \begin{pmatrix} R_q(\Theta_1^1, \Theta_2^1) \\ \vdots \\ R_q(\Theta_1^1, \Theta_2^N) \\ R_q(\Theta_1^2, \Theta_2^1) \\ \vdots \\ R_q(\Theta_1^2, \Theta_2^N) \\ \vdots \\ R_q(\Theta_1^N, \Theta_2^N) \end{pmatrix}, \quad (14)$$

$$\mathbf{R}'_q \equiv \text{Re}(\mathbf{R}_q). \quad (15)$$

Here,  $\mathcal{G}$  represents a transformation that transforms a matrix into a vector and  $\mathcal{G}^{-1}$  represents the inverse transformation. Note that in Eq. (12),  $M_{q,\alpha}(\theta_1, \theta_2, \phi_1, \phi_2)$  is discretized using a grid defined in the laboratory coordinate system. According to Eq. (4), we can obtain the corresponding grid in the molecular coordinate system. Then, we can use nearest-neighbor interpolation to replace the element of  $M_{q,\alpha}$  with the element of  $\mathbf{R}'_q$ . This can be expressed by an interpolation matrix  $\mathbf{C}_{q,\alpha}$ , which is computed by counting the nearest-neighbor points (TNNP) for each grid point in  $M_{q,\alpha}(\theta_1, \theta_2, \phi_1, \phi_2)$ . If we define

$$\mathcal{N}[(\theta_1^{n_1}, \theta_2^{n_2}, \phi_1^i, \phi_2^j), (\Theta_1^{n_3}, \Theta_2^{n_4})] = \begin{cases} 0, & \text{if } (\Theta_1^{n_3}, \Theta_2^{n_4}) \text{ is not TNNP of } (\theta_1^{n_1}, \theta_2^{n_2}, \phi_1^i, \phi_2^j). \\ 1, & \text{if } (\Theta_1^{n_3}, \Theta_2^{n_4}) \text{ is TNNP of } (\theta_1^{n_1}, \theta_2^{n_2}, \phi_1^i, \phi_2^j). \end{cases} \quad (16)$$

Here  $(\theta_1^{n_1}, \theta_2^{n_2}, \phi_1^i, \phi_2^j)$  is the grid point of  $M_{q,\alpha}(\theta_1, \theta_2, \phi_1, \phi_2)$  and  $(\Theta_1^{n_3}, \Theta_2^{n_4})$  is the grid point of  $\mathbf{R}'_q$ . Then the matrix element of  $\mathbf{C}_{q,\alpha}$  can be expressed as

$$C_{q,\alpha}^{n_1 n_2 n_3 n_4} \equiv C_{q,\alpha}(\theta_1^{n_1}, \theta_2^{n_2}, \Theta_1^{n_3}, \Theta_2^{n_4}) = \sum_{i,j} \mathcal{N}[(\theta_1^{n_1}, \theta_2^{n_2}, \phi_1^i, \phi_2^j), (\Theta_1^{n_3}, \Theta_2^{n_4})], \quad (17)$$

and the matrix  $\mathbf{C}_{q,\alpha}$  is defined as

$$\mathbf{C}_{q,\alpha} \equiv \begin{pmatrix} C_{q,\alpha}^{1111} & \cdots & C_{q,\alpha}^{111N} & C_{q,\alpha}^{1121} & \cdots & C_{q,\alpha}^{112N} & \cdots & C_{q,\alpha}^{11NN} \\ \vdots & & & & & & & \\ C_{q,\alpha}^{1N11} & \cdots & C_{q,\alpha}^{1N1N} & C_{q,\alpha}^{1N21} & \cdots & C_{q,\alpha}^{1N2N} & \cdots & C_{q,\alpha}^{1NNN} \\ C_{q,\alpha}^{2111} & \cdots & C_{q,\alpha}^{211N} & C_{q,\alpha}^{2121} & \cdots & C_{q,\alpha}^{212N} & \cdots & C_{q,\alpha}^{21NN} \\ \vdots & & & & & & & \\ C_{q,\alpha}^{2N11} & \cdots & C_{q,\alpha}^{2N1N} & C_{q,\alpha}^{2N21} & \cdots & C_{q,\alpha}^{2N2N} & \cdots & C_{q,\alpha}^{2NNN} \\ \vdots & & & & & & & \\ C_{q,\alpha}^{NN11} & \cdots & C_{q,\alpha}^{NN1N} & C_{q,\alpha}^{NN21} & \cdots & C_{q,\alpha}^{NN2N} & \cdots & C_{q,\alpha}^{NNNN} \end{pmatrix}. \quad (18)$$

Then, Eq. (9) becomes

$$\mathbf{I}_{q,\alpha} = \mathbf{P}\mathbf{C}_{q,\alpha}\mathcal{G}(\mathbf{R}'_q). \quad (19)$$

Now we can use  $R_q(\Theta_1, \Theta_2)$  to describe the HHG signal and keep the equation linear. This approximation is fairly accurate because the discrete grid is dense and the single-molecule dipole is fairly smooth. When the polarizations of the alignment and driving pulses are parallel, which means that  $\alpha = 0$ , the interpolation matrix  $\mathbf{C}_{q,\alpha}$  is a unit matrix. This specific case was used in the previous work to retrieve the single-molecule dipole moment [39]. However, the method in [39] cannot use signals with  $\alpha \neq 0$ , so the amount of data that can be used is small, which leads to poor accuracy and stability of reconstruction. With Eq. (19), we can solve these problems by adding HHG signals from other angles. To this end, we further concatenate  $\mathbf{I}_{q,\alpha}$  and  $\mathbf{P}\mathbf{C}_{q,\alpha}$  with different  $\alpha$  by defining

$$\mathbf{S}_q \equiv \begin{pmatrix} \mathbf{I}_{q,1} \\ \mathbf{I}_{q,2} \\ \vdots \\ \mathbf{I}_{q,O} \end{pmatrix}, \mathbf{K}_q \equiv \begin{pmatrix} \mathbf{P}\mathbf{C}_{q,1} \\ \mathbf{P}\mathbf{C}_{q,2} \\ \vdots \\ \mathbf{P}\mathbf{C}_{q,O} \end{pmatrix}. \quad (20)$$

Here,  $O$  is the sampling number of  $\alpha$ . Then we have

$$\mathbf{S}_q = \mathbf{K}_q\mathcal{G}(\mathbf{R}'_q). \quad (21)$$

We now formalize the retrieval problem as an optimization problem. Here, we assume that the molecular axis distribution  $\rho(\theta, \tau)$  is already known; thus, the matrix  $\mathbf{P}$  is known in advance. In the experiment,  $\rho(\theta, \tau)$  can be determined following the previous works [43,45]. Equation (21) indicates that the reconstruction problem can be formalized as linear regression problem. Meanwhile, Eqs. (13) and (15) have to be satisfied to decompose  $R_q(\Theta_1, \Theta_2)$  into  $D_q^*(\Theta)$ . Then we can write the reconstruction problem as

$$\min_{\mathbf{R}'_q} \|\mathbf{S}_q - \mathbf{K}_q\mathcal{G}(\mathbf{R}'_q)\|_2^2, \quad s.t. \quad \begin{cases} \mathbf{R}_q = \mathbf{R}_q^\dagger, \\ \text{rank}(\mathbf{R}_q) = 1, \\ \mathbf{R}'_q = \text{Re}(\mathbf{R}_q), \end{cases} \quad (22)$$

which minimizes the square error of the input signal relative to the reconstructed signal. The constraints of Eq. (22) mean that  $\mathbf{R}_q$  is a Hermitian matrix whose rank is 1 and  $\mathbf{R}'_q$  has to be the real part of  $\mathbf{R}_q$ . As a result, the  $\mathbf{R}_q$  is defined as a product of two column vectors that are conjugate to each other. This is an equivalent form of the definition of  $\mathbf{R}_q$  and  $\mathbf{R}'_q$  in Eqs. (14) and (15). Note that this nonconvex constraint was ignored in previous work [39].

### C. Retrieval algorithm

In this subsection, we introduce our retrieval algorithm based on the iterative projection method. This includes the segmentation of the solution space, definition of the projections, description of the iterative procedure and generalization of the heteroscedastic situation.

#### 1. Segmentation of the solution space

Equation (22) is a high-dimensional nonconvex optimization problem. To solve it, a proper algorithm must be used. Iterative algorithms based on projection [46] have been proved

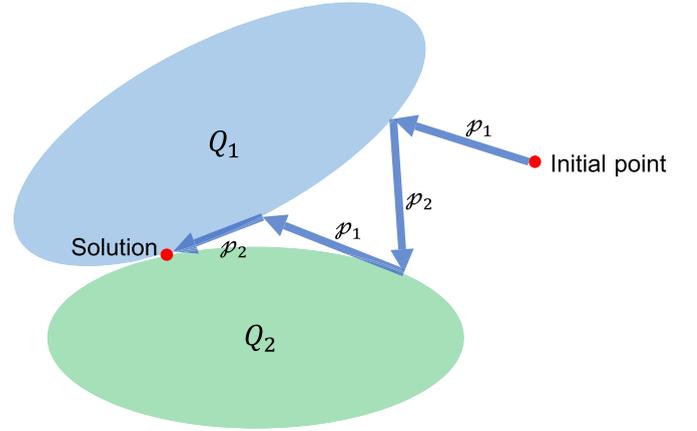


FIG. 2. Schematic diagram of the convergence process of the iterative projection method.  $Q_1$  and  $Q_2$  represent two different sets that segment the solution space into two parts.  $\mathcal{P}_1$  and  $\mathcal{P}_2$  represent the projections corresponding to  $Q_1$  and  $Q_2$  which satisfy Eq. (25).

to be very effective in solving nonconvex optimization problems [47]. Recently, it was widely used in phase retrieval [48], reconstructing state mixtures [49], diffractive imaging [50–52], frequency-resolved optical gating [53], and image superresolution [54]. Here, we provide an algorithm for solving Eq. (22) based on the iterative projection. The basic idea of the projection method is to divide the solution space into several sets and each set satisfies a subset of the constraints. Then the solution to the feasibility problem is at the intersection of all sets (see Fig. 2). For Eq. (22), according to its optimization objective, one of the subsets of the solution space can be expressed as

$$Q_1 \equiv \{a | \mathbf{K}_q \text{Re}(a) = \mathbf{S}_q\}. \quad (23)$$

This means the real part of the matrices in  $Q_1$  is the solutions of the linear equation Eq. (21), while the imaginary part is not restricted. Another set that corresponds to the constraints of Eq. (22) can be expressed as

$$Q_2 \equiv \{b | \exists x \in \{x | x = x^\dagger, \text{rank}(x) = 1\}, b = \text{Re}(x)\}. \quad (24)$$

The matrices in  $Q_2$  are the real part of a Hermitian matrix whose rank is 1.  $Q_1$  and  $Q_2$  segment the solution space into two parts. Under this partition, the solution of Eq. (22) thus belongs to the intersection of  $Q_1$  and  $Q_2$ .

#### 2. Definition of the projections

For searching the intersection of  $Q_1$  and  $Q_2$ , we should define the idempotent optimal projection for each set, which means that the Euclidean distance before and after projection should be minimal. The requirement for projections can be expressed as

$$\min_{\mathcal{P}_i} \|\mathbf{A} - \mathcal{P}_i(\mathbf{A})\|_F^2, \quad s.t. \quad \begin{cases} \mathcal{P}_i[\mathcal{P}_i(\mathbf{A})] = \mathcal{P}_i(\mathbf{A}), \\ \forall \mathbf{A} = \mathbf{A}^\dagger, \mathcal{P}_i(\mathbf{A}) \in Q_i, \end{cases} \quad (25)$$

where  $\mathbf{A}$  is an arbitrary Hermitian matrix.  $Q_i$  represents the sets that divide the solution space.  $\mathcal{P}_i$  corresponds to an operation that projects an arbitrary Hermitian matrix  $\mathbf{A}$  into the set  $Q_i$ . The constraints in Eq. (25) indicate that the projection

$\mathcal{P}_i$  is idempotent. In addition, the optimization objective of Eq. (25) ensures the Euclidean distance (here is the Frobenius norm of the matrix) between  $\mathbf{A}$  and  $\mathcal{P}_i(\mathbf{A})$  is minimal.

We have to solve Eq. (25) and make the projection  $\mathcal{P}_i$  concrete. For the projection  $\mathcal{P}_1$  corresponding to  $Q_1$ , we can solve the following optimization problem:

$$\min_{\delta} \|\delta\|_F^2, \quad s.t. \quad \mathbf{K}_q\{\mathcal{G}(\delta) + \text{Re}[\mathcal{G}(\mathbf{A})]\} = \mathbf{S}_q. \quad (26)$$

The solution  $\delta$  is the residual error of  $\mathbf{A}$ . Then, the solution  $\delta$  of Eq. (26) is used to define the projection  $\mathcal{P}_1$ ,

$$\mathcal{P}_1(\mathbf{A}) = \delta + \mathbf{A}. \quad (27)$$

Here,  $\delta$  can be understood as the change of  $\mathbf{A}$ . Minimizing  $\|\delta\|_F^2$  means the projection  $\mathcal{P}_1$  is optimal.

$$\mathbf{F} \equiv \begin{pmatrix} B_{00}(\theta_1^1, \theta_2^1) & B_{10}(\theta_1^1, \theta_2^1) & B_{11}(\theta_1^1, \theta_2^1) & \cdots & B_{MM}(\theta_1^1, \theta_2^1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ B_{00}(\theta_1^1, \theta_2^N) & B_{10}(\theta_1^1, \theta_2^N) & B_{11}(\theta_1^1, \theta_2^N) & \cdots & B_{MM}(\theta_1^1, \theta_2^N) \\ B_{00}(\theta_1^2, \theta_2^1) & B_{10}(\theta_1^2, \theta_2^1) & B_{11}(\theta_1^2, \theta_2^1) & \cdots & B_{MM}(\theta_1^2, \theta_2^1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ B_{00}(\theta_1^2, \theta_2^N) & B_{10}(\theta_1^2, \theta_2^N) & B_{11}(\theta_1^2, \theta_2^N) & \cdots & B_{MM}(\theta_1^2, \theta_2^N) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ B_{00}(\theta_1^N, \theta_2^N) & B_{10}(\theta_1^N, \theta_2^N) & B_{11}(\theta_1^N, \theta_2^N) & \cdots & B_{MM}(\theta_1^N, \theta_2^N) \end{pmatrix}. \quad (29)$$

Note that the column vectors of  $\mathbf{F}$  are nonorthogonal in Euclidean space. Thus, they are orthonormalized to construct a set of orthonormal basis vectors in Euclidean space and form a new transformation matrix  $\mathbf{H}$ . Such a procedure preserves the symmetry of the original Legendre polynomial basis set. Consequently, the residual error  $\delta$  is expanded by the orthonormal basis in  $\mathbf{H}$  and the Frobenius norm of the original matrix  $\delta$  is equal to the l2-norm of the expansion coefficient  $\epsilon$

$$\|\delta\|_F^2 = \|\mathbf{H}\epsilon\|_2^2 = \|\epsilon\|_2^2. \quad (30)$$

With Eq. (30), Eq. (26) is equivalent to the following equation in the column space of matrix  $\mathbf{H}$ :

$$\min_{\epsilon} \|\epsilon\|_2^2, \quad s.t. \quad \mathbf{K}_q\{\mathbf{H}\epsilon + \text{Re}[\mathcal{G}(\mathbf{A})]\} = \mathbf{S}_q. \quad (31)$$

Thus, the projection  $\mathcal{P}_1$  can be defined as

$$\mathcal{P}_1(\mathbf{A}) = \mathcal{G}^{-1}(\mathbf{H}\epsilon) + \mathbf{A}. \quad (32)$$

Equation (31) can be solved using the following equation:

$$\epsilon = (\mathbf{K}_q\mathbf{H})^+\{\mathbf{S}_q - \mathbf{K}_q \text{Re}[\mathcal{G}(\mathbf{A})]\}. \quad (33)$$

Here, the Moore-Penrose pseudoinverse [55] is used to solve the linear system in the constraint of Eq. (31) and minimizes the l2-norm of  $\epsilon$ . This will give us a unique minimum-norm solution. In practice, a low-rank approximation of  $\mathbf{K}_q\mathbf{H}$  is required to calculate the Moore-Penrose pseudoinverse for numerical stability and solve the ill condition of this equation. Finally, combining Eq. (32) with Eq. (33) the projection  $\mathcal{P}_1$  can be expressed as

$$\mathcal{P}_1(\mathbf{A}) = \mathcal{G}^{-1}(\mathbf{H}(\mathbf{K}_q\mathbf{H})^+\{\mathbf{S}_q - \mathbf{K}_q \text{Re}[\mathcal{G}(\mathbf{A})]\}) + \mathbf{A}. \quad (34)$$

Sometimes the solution requires additional symmetry; for example, the single-molecule dipole  $D_q(\Theta)$  should have a mirror symmetry of 90 degrees in our case. As a result, a set of orthonormal bases with corresponding symmetry is essential to expand the  $\delta$ . Specifically, we construct the symmetrized two-dimensional (2D) Legendre polynomial basis functions

$$B_{mn}(\theta_1, \theta_2) \equiv \frac{1}{\sqrt{2}}[L_m(\theta_1)L_n(\theta_2) + L_m(\theta_2)L_n(\theta_1)] \quad (m \geq n), \quad (28)$$

where  $L_{m(n)}(\theta)$  is the normalized Legendre polynomial of degree  $m(n)$ . Assuming that the highest order of the polynomial is  $M$ , we can obtain the transformation matrix  $\mathbf{F}$ , which is

By now the projection  $\mathcal{P}_1$  has been properly defined. Similarly, the projection  $\mathcal{P}_2$  corresponding to  $Q_2$  also needs to be solved by an optimization problem, as expressed as

$$\min_{\mathbf{B}} \|\mathbf{A} - \text{Re}(\mathbf{B})\|_F^2, \quad s.t. \quad \begin{cases} \mathbf{B} = \mathbf{B}^\dagger, \\ \text{rank}(\mathbf{B}) = 1, \end{cases} \quad (35)$$

where  $\mathbf{A}$  is an input arbitrary Hermitian matrix. The constraints of Eq. (35) mean the solution  $\mathbf{B}$  is an Hermitian matrix whose rank is one.

Such an optimization problem can be solved by the iterative projection method as well. Same as Eqs. (23) and (24), the solution space of Eq. (35) can be further segmented into two subsets. According to the constraints in Eq. (35), one of the subsets of the solution space  $Q_3$  can be expressed as

$$Q_3 \equiv \{c | c = c^\dagger, \text{rank}(c) = 1\}, \quad (36)$$

which means the matrices in  $Q_3$  are Hermitian matrices whose rank are one. The other subset that relates to the optimization objective in Eq. (35) is defined as

$$Q_4 \equiv \{d | \text{Re}(d) = \text{Re}(\mathbf{A})\}, \quad (37)$$

which means the real part of the matrices in  $Q_4$  are equal to that of the input matrix  $\mathbf{A}$ . Then, the projection  $\mathcal{P}_3$  that satisfies Eq. (25) can be expressed as

$$\mathcal{P}_3(\mathbf{\Delta}) = \mathbf{V}\mathbf{V}^\dagger \mathbf{E}, \quad (38)$$

where  $\mathcal{P}_3$  projects any Hermitian matrix to  $Q_3$ .  $\mathbf{\Delta}$  is an arbitrary Hermitian matrix.  $E$  is the maximum eigenvalue of the input matrix  $\mathbf{\Delta}$  and  $\mathbf{V}$  is the corresponding eigenvector.

The definition of  $\mathcal{P}_4$  is

$$\mathcal{P}_4(\mathbf{\Delta}) = \text{Re}(\mathbf{A}) + \text{Im}(\mathbf{\Delta})i, \quad (39)$$

where  $\mathcal{P}_4$  projects any Hermitian matrix to  $Q_4$ . Here, by replacing the real part of  $\mathbf{\Delta}$  with the real part of  $\mathbf{A}$ , Eq. (25) can be easily satisfied.

After solving Eq. (35), the projection  $\mathcal{P}_2$  can be defined as

$$\mathcal{P}_2(\mathbf{A}) = \text{Re}(\mathbf{B}). \quad (40)$$

Here, with  $\mathcal{P}_3$  and  $\mathcal{P}_4$ , an internal iterative projection is used to implement  $\mathcal{P}_2$ .

### 3. Procedure of the iterative projection method

Once the projections  $\mathcal{P}_1$  and  $\mathcal{P}_2$  are defined, the solution of Eq. (22) can be searched in the solution space. Here, we use the Projection Onto Convex Sets (POCS) algorithm [46]. When there are two sets, it can be denoted as an iterative formula

$$\mathbf{R}'_q{}^{(n+1)} = \mathcal{P}_2[\mathcal{P}_1(\mathbf{R}'_q{}^{(n)})], \quad (41)$$

where  $\mathbf{R}'_q{}^{(n)}$  is the solution from the previous iteration and  $\mathbf{R}'_q{}^{(n+1)}$  is the solution from this iteration. The measured signal fed into the algorithm is represented as a vector  $\mathbf{S}_q$  which is defined in Eq. (20). The molecular axis distribution is used to construct the matrix  $\mathbf{P}$ . These are used to implement the projection  $\mathcal{P}_1$ . The difference between  $\mathbf{R}'_q{}^{(n)}$  and  $\mathbf{R}'_q{}^{(n+1)}$  is defined as the Frobenius norm  $\|\mathbf{R}'_q{}^{(n)} - \mathbf{R}'_q{}^{(n+1)}\|_F^2$ . The iteration stops once the difference is smaller than the tolerance. Figure 2 vividly describes the convergence process of the algorithm. Starting from the initial point, it converges to the solution after repeated projection between the two sets. The same formula can be used to solve Eq. (35) by using projections  $\mathcal{P}_3$  and  $\mathcal{P}_4$  which are defined in Eqs. (38) and (39), respectively.

After solving Eq. (22), we can use the following equations to extract  $D_q(\Theta)$  from  $\mathbf{R}'_q$  [39]:

$$|D_q(\Theta)| = \sqrt{\text{Re}[R'(\Theta, \Theta)]}$$

$$\phi(\Theta_2) - \phi(\Theta_1) = \pm \arccos \left[ \frac{\text{Re}[R'(\Theta_1, \Theta_2)]}{|D_q(\Theta_1)| |D_q(\Theta_2)|} \right], \quad (42)$$

where  $\phi(\Theta)$  is the phase of  $D_q(\Theta)$ . Here, we need to emphasize that, unlike in [39], our method completely includes constraints on  $R'(\Theta_1, \Theta_2)$ , so  $D_q(\Theta)$  can be extracted using Eq. (42).

### 4. Generalization of the heteroscedastic situation

Equation (22) assumes that the input signal is homoscedastic, which means that each measured  $\mathbf{S}_q$  has the same variance. Here, we provide a way to extend the algorithm to the heteroscedastic situation. For the heteroscedastic situation, the problem can be explained as a weighted linear regression model with some constraints. The linear regression model is represented as

$$\mathbf{S}_q = \mathbf{K}_q \mathcal{G}(\mathbf{R}'_q) + \mathbf{v}_q. \quad (43)$$

Here,  $\mathbf{v}_q$  is the heteroscedastic noise. The inverse of its variance can be used as the weight  $\omega_{q,\alpha}^k = \frac{1}{\text{Var}(\mathbf{v}_{q,\alpha}^k)}$ . Then we

can define the weight matrix  $\mathbf{W}_q$  as

$$\mathbf{\Omega}_q \equiv \begin{pmatrix} \omega_{q,1}^1 \\ \vdots \\ \omega_{q,1}^K \\ \omega_{q,2}^1 \\ \vdots \\ \omega_{q,2}^K \\ \vdots \\ \omega_{q,O}^K \end{pmatrix}, \quad \mathbf{W}_q \equiv \text{diag}(\mathbf{\Omega}_q). \quad (44)$$

Here  $\text{diag}$  represents a transformation that transfers a vector to a diagonal matrix. The optimization problem can be represented as

$$\min_{\mathbf{R}'_q} \sum_{\alpha} [\mathbf{S}_q - \mathbf{K}_q \mathcal{G}(\mathbf{R}'_q)]^T \mathbf{W}_q [\mathbf{S}_q - \mathbf{K}_q \mathcal{G}(\mathbf{R}'_q)],$$

$$\text{s.t.} \quad \begin{cases} \mathbf{R}_q = \mathbf{R}_q^\dagger, \\ \text{rank}(\mathbf{R}_q) = 1, \\ \mathbf{R}'_q = \text{Re}(\mathbf{R}_q). \end{cases} \quad (45)$$

By multiplying each term in the Eq. (43) with  $\sqrt{\mathbf{W}_q}$ , we have

$$\sqrt{\mathbf{W}_q} \mathbf{S}_q = \sqrt{\mathbf{W}_q} \mathbf{K}_q \mathcal{G}(\mathbf{R}'_q) + \sqrt{\mathbf{W}_q} \mathbf{v}_q. \quad (46)$$

This is also a linear regression model and its noise term  $\sqrt{\mathbf{W}_q} \mathbf{v}_q$  is homoscedastic. Considering the constraints of  $\mathbf{R}'_q$ , the corresponding optimization form can be represented as

$$\min_{\mathbf{R}'_q} \sum_{\alpha} \|\sqrt{\mathbf{W}_q} \mathbf{S}_q - \sqrt{\mathbf{W}_q} \mathbf{K}_q \mathcal{G}(\mathbf{R}'_q)\|_2^2,$$

$$\text{s.t.} \quad \begin{cases} \mathbf{R}_q = \mathbf{R}_q^\dagger, \\ \text{rank}(\mathbf{R}_q) = 1, \\ \mathbf{R}'_q = \text{Re}(\mathbf{R}_q). \end{cases} \quad (47)$$

Equation (47) can be solved with the algorithm introduced above.

## III. RESULTS AND DISCUSSION

In this section, we first verify the validity and stability of our algorithm with numerical results, then introduce our experimental scheme, and finally implement our algorithm on experimental data.

### A. Numerical results

To test the validity and stability of our algorithm, we first perform our algorithm on theoretically calculated HHG spectra. We calculate the single-molecule dipole by the quantitative rescattering (QRS) theory [44,56–58]. The driving pulse which generates the high harmonics has a peak intensity of  $1.3 \times 10^{14}$  W/cm<sup>2</sup> and a duration of 40 fs (full width at half maximum, FWHM). The calculated amplitude and phase of the 25th harmonic (H25) are shown in Fig. 3(a) with solid lines. Then, we calculate the molecular axis distribution  $\rho(\theta, \phi, \tau)$  of N<sub>2</sub> molecules. In our calculation, the alignment pulse has a peak intensity of  $3 \times 10^{13}$  W/cm<sup>2</sup> and a duration of 50 fs (FWHM). The rotational temperature of the gas is 100 K. The calculated molecular axis distribution is shown

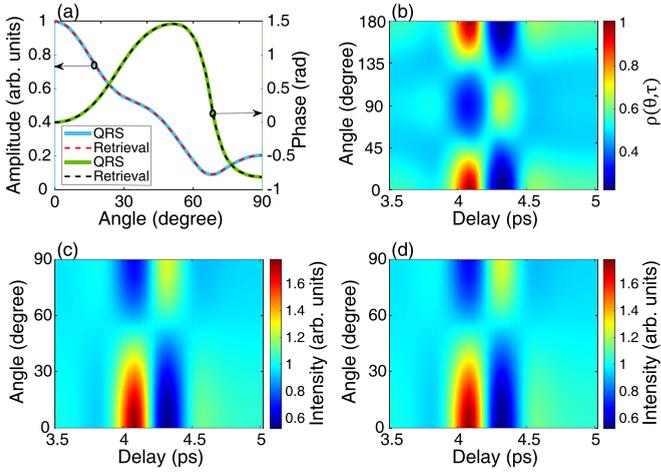


FIG. 3. (a) The calculated amplitude (blue solid line) and phase (green solid line) of the single-molecule dipole  $D(\Theta)$  of H25 for  $N_2$  using QRS theory. Red and black dashed lines are the retrieved results. (b) Calculated time-dependent molecular axis distribution. (c) Calculated time-dependent harmonic signal at different polarization angles between the alignment and driving pulses for H25 after convoluting the calculated single-molecule dipole  $D(\Theta)$  in (a) with the time-dependent molecular axis distribution in (b). (d) Same as (c), but is calculated with the retrieved single-molecule dipole.

in Fig. 3(b). The maximal degree of alignment ( $\cos^2 \theta$ ) is approximately 0.44. The HHG signal  $I_q(\alpha, \tau)$  then can be calculated using Eq. (3) to simulate experimental data. To facilitate the calculation, the corresponding random alignment signal was used to normalize the calculation result, which means that we do not care about the absolute amplitude but the angle distribution of the single-molecule dipole. We mention

that the absolute amplitude can be easily reproduced with our reconstruction result multiplied by a real factor. Figure 3(c) shows the simulated signal of H25 as a function of the time delay and polarizing angle between the alignment and driving pulses. With the simulated HHG signals, we first demonstrate our algorithm in a noiseless case. The single-molecule dipole retrieved for H25 of  $N_2$  is shown in Fig. 3(a) as the dashed lines. Without the noise, one can see that both the amplitudes and phases are accurately retrieved. The HHG signal reproduced with the retrieved dipole moment is shown in Fig. 3(d), which is also in excellent agreement with the input result in Fig. 3(c).

Considering the presence of the noise in the experiment, we next investigate the robustness of the algorithm against the noise. we added Gaussian random noise to our numerical results to simulate the real spectra measured in the experiment. In the simulations, the signal-noise ratio (SNR) is used to assess the level of the Gaussian random noise, which is defined as

$$\text{SNR} = 10 \log_{10} \left( \frac{\text{Var}(S)}{\text{Var}(\epsilon)} \right), \quad (48)$$

where  $\text{Var}(S)$  and  $\text{Var}(\epsilon)$  are the variance of the signal and the noise, which describe the average power of the signal and the noise respectively. The unit of the SNR is decibel (dB). Here, we still take H25 as an example. Figures 4(a) to 4(d) show the simulated spectra for  $\text{SNR} = 20, 10, 5, 0$ , respectively. The amplitude and phase of H25 retrieved from the noisy data are shown in Figs. 4(e) and 4(f). One can see that the single-molecule dipole can be accurately retrieved with the SNR changing from 0 to 20 dB. Additional tests for other harmonics are shown in Appendix A. All results show good noise stability. In our experiment, the SNR of the measured HHG spectra is estimated to be about 10 dB. Thus

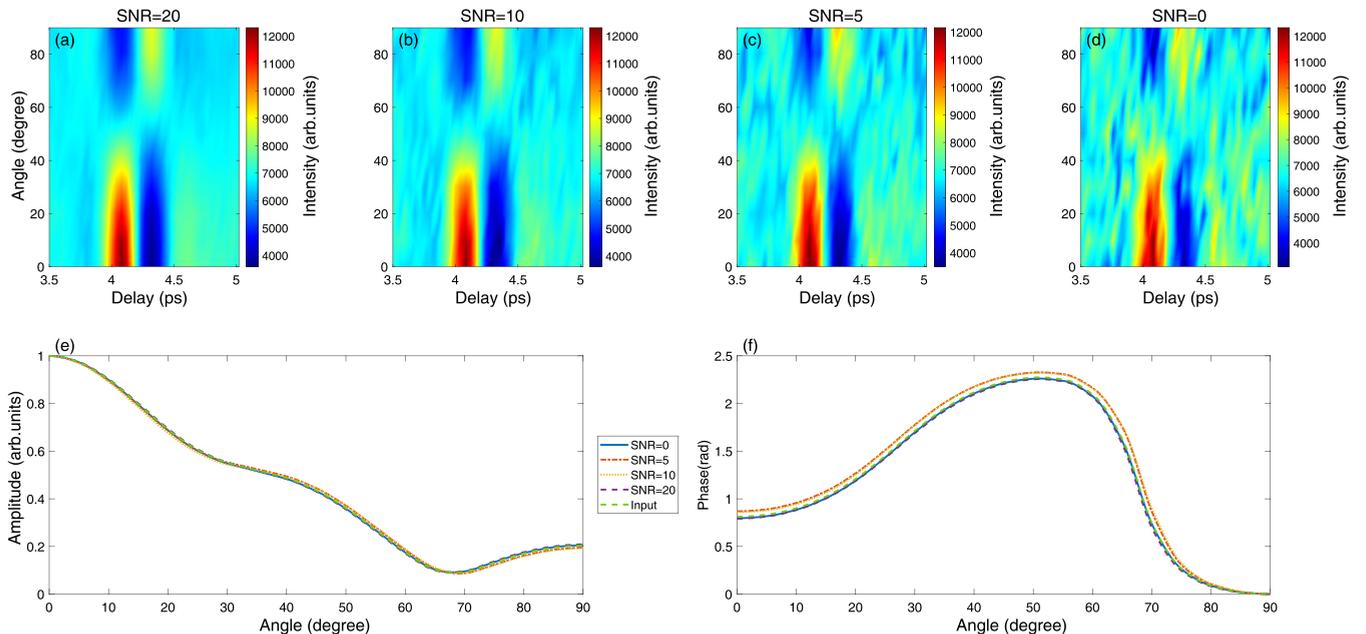


FIG. 4. (a)–(d) are calculated spectra of H25 with  $\text{SNR} = 20, 10, 5, 0$ , respectively. The dashed purple (dark gray) line, dotted yellow line, dash-dotted red line, and blue solid line in (e,f) are the amplitude and phase of the single-molecule dipoles retrieved with the spectra in (a)–(d). The dashed green (light gray) line is the input single-molecule dipole calculated by QRS theory.

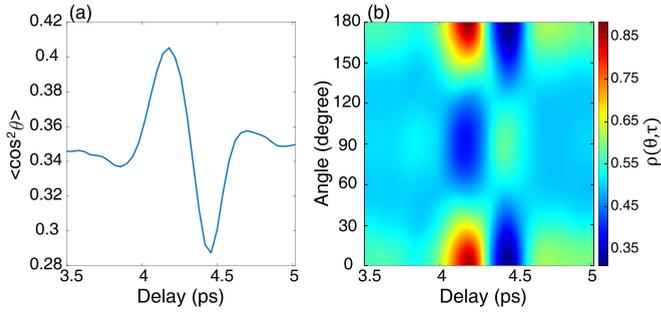


FIG. 5. (a) Time-dependent degrees of molecular alignment  $\langle \cos^2 \theta \rangle$  around the half-rotational revival of  $N_2$  are reconstructed using the experimental data. (b) Reconstructed time-dependent molecular axis distribution  $\rho(\theta, \tau)$ .

the proposed algorithm can be well used in our experimental data. We also test the noise stability for noise in the molecular axis distribution. Our results show that our algorithm can accurately retrieve the complex single-molecule dipole from the harmonic spectra with good noise stability (see Appendix B).

### B. Experiment results

In this section, we demonstrate the implementation of our algorithm on experimental data. Our experiment is carried out by using a commercial Ti: sapphire laser system (Astrella-USP-1K, Coherent, Inc.), which delivers 35-fs, 800-nm pulses at a repetition rate of 1 kHz. The output laser is split into two pulses to produce the alignment and driving pulses. An alignment pulse with moderate intensity and polarization parallel to the  $z$  axis is used to induce the nonadiabatic alignment of molecules along its polarization. Later, the intense driving

pulse with adjustable time delay and polarization is applied to interact with the molecular ensemble to generate high-order harmonics. A motorized delay line and a half-wave plate are installed in the arm of the driving pulse to adjust its time delay and polarization with respect to the alignment one. The delay and polarizations of the driving pulses are scanned to obtain the harmonic spectra like that in Fig. 4(a). The generated high harmonics are detected by a homemade flat-field soft x-ray spectrometer [59,60].

To retrieve the single-molecule dipole, we first need to determine the time-dependent molecular axis distribution in our experiment. We do this following the previous work [43]. The time-dependent degrees of molecular alignment  $\langle \cos^2 \theta \rangle$  reconstructed by experimental data are shown in Fig. 5(a). The maximum alignment degree is approximately 0.4. Figure 5(b) shows the time-dependent molecular axis distribution  $\rho(\theta, \tau)$  reproduced with experimental data. With  $\rho(\theta, \tau)$  determined, we can then use our algorithm to retrieve the single-molecule dipole with experimental data. Figures 6(a) and 6(c) show the experimental signals of H19 and H23 measured at different delays and polarization angles between the alignment and driving pulses. The solid lines in Figs. 6(e) to 6(h) are the retrieved dipole amplitudes and phases of H19 and H23, respectively. Figures 6(b) and 6(d) show the corresponding HHG signals reproduced with the retrieved dipoles. One can see that the reconstruction results agree well with the experimental results. We also perform theoretical calculations using the QRS theory to compare with our retrievals. In our QRS calculations, the peak intensity of the driving pulse is  $1 \times 10^{14}$  W/cm<sup>2</sup>. The amplitude and phase of calculated dipoles of H19 and H23 are shown as dashed lines in Figs. 6(e) to 6(h). In our calculation, only the HOMO orbital is used. One can see that the calculation falls within the bounds of the

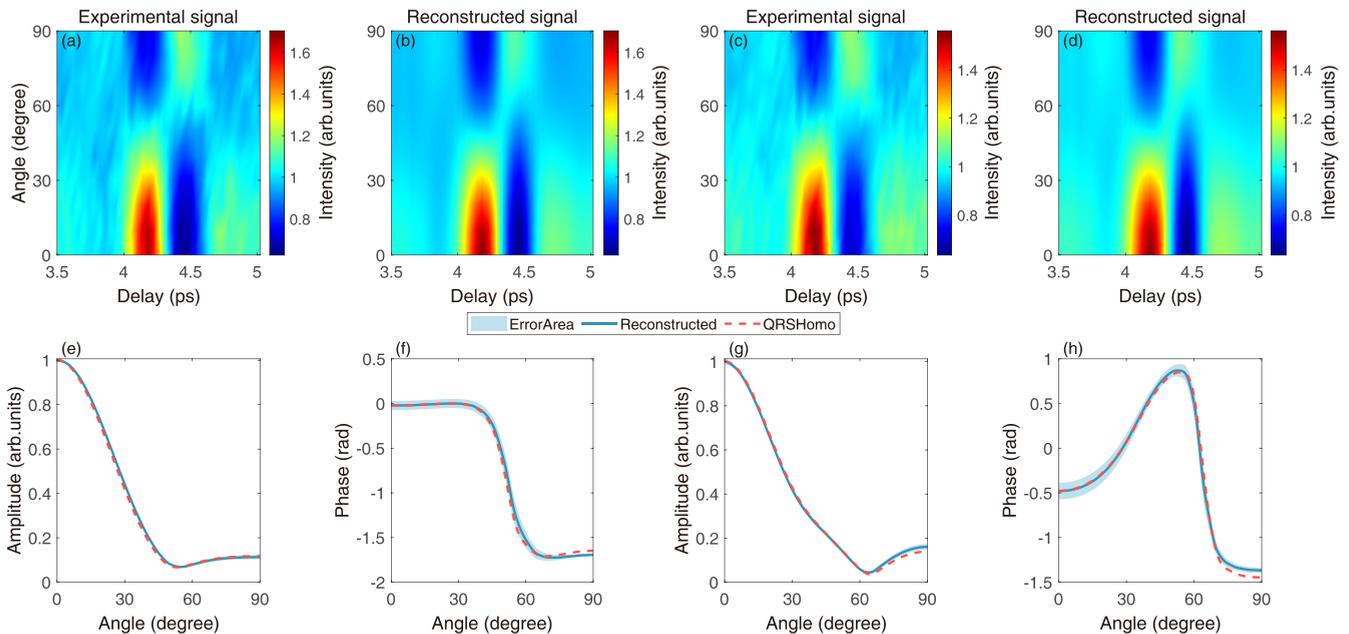


FIG. 6. (a,b) are the harmonic signals of H19 measured in our experiment and calculated with the retrieved single-molecule dipole. Blue solid lines in (e,f) are the retrieved amplitude and phase of the single-molecule dipole for the experimental signal in (a). Red dashed lines in (e,f) are the amplitude and phase of the single-molecule dipole of H19 for the HOMO orbital which is calculated by QRS theory. (c,d) and (g,h) Same as (a,b) and (e,f), but for H23.

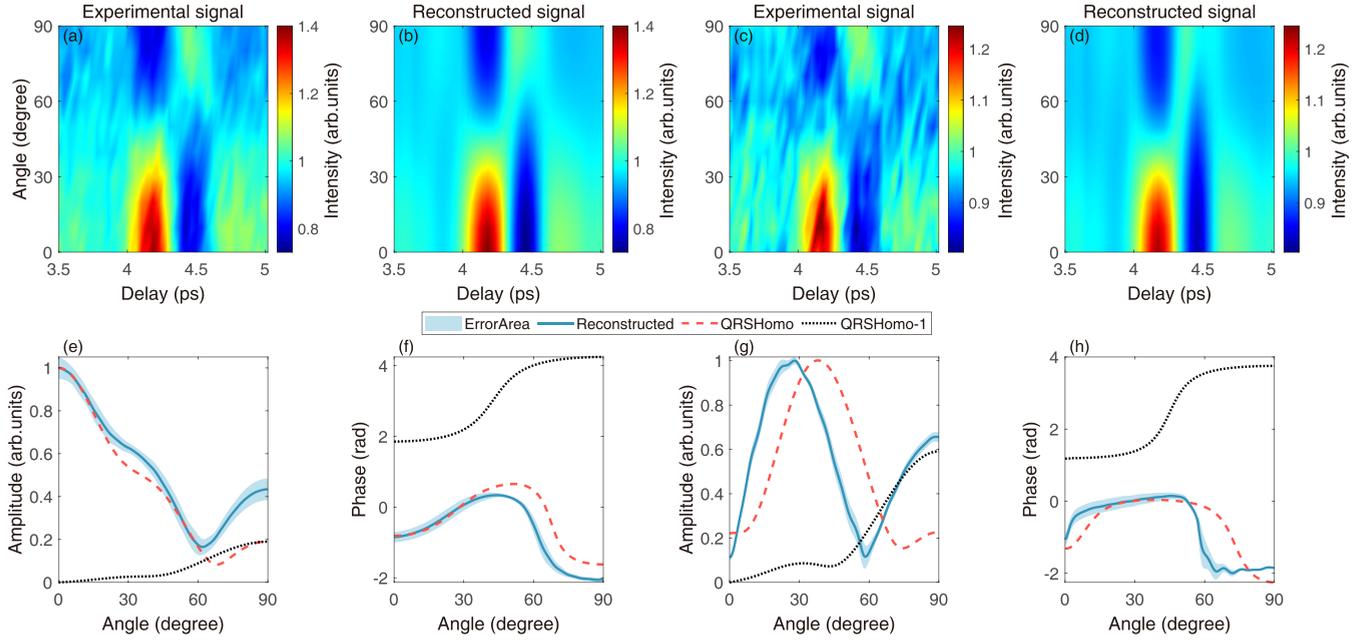


FIG. 7. (a,b) are the harmonic signals of H25 measured in our experiment and calculated with the retrieved single-molecule dipole. Blue solid lines in (e,f) are the retrieved amplitude and phase of the single-molecule dipole for the experimental signal in (a). Red dashed lines and black dotted lines in (e,f) are the amplitude and phase of the single-molecule dipole of H25 for HOMO and HOMO-1 orbitals which is calculated by QRS theory. (c,d) and (g,h) Same as (a,b) and (e,f), but for H31.

experimental uncertainty for almost every alignment angle, and agrees at a quantitative level with the experiment. These results indicate that the HOMO orbital dominates the HHG process.

We also performed our algorithm for harmonics in the cutoff region. Figures 7(a) and 7(c) show the experimental signals of H25 and H31. The solid lines in Figs. 7(e) to 7(h) are the retrieved dipole amplitudes and phases of H25 and H31, respectively. The HHG signals reproduced with the retrieved dipoles also agree well with the experimental results shown in Figs. 7(b) and 7(d). However, the retrieved harmonic dipoles show a significant difference compared to the theoretical calculations for HOMO [see Figs. 7(e) to 7(h)]. The retrieval results show a much larger enhancement after 60 degrees than the theoretical results for HOMO. To understand this difference, we further performed simulations with deeper orbitals taken into account [61]. It was reported in previous works that the HOMO-1 orbital plays an important role in the HHG from  $N_2$  when the molecule is perpendicularly aligned to the polarization of the driving laser [62]. The calculation results of HOMO-1 for H25 and H31 are shown in Figs. 7(e) to 7(h) with black dashed lines. From Figs. 7(e) to 7(h), one can see that the contribution of HOMO-1 is most prominent after 60 degrees due to its  $\pi_g$  symmetry, which is consistent with our retrievals. In addition, we also note that the contribution of HOMO-1 is small for the lower order. It will not significantly effect the harmonic spectra below H25 according to our calculation result. This also agrees with our experimental results. Note that experimental results of other harmonics are also shown in Appendix A.

Finally, we emphasize that the contribution from deeper molecular orbitals of  $N_2$  in HHG can be directly observed in the measured HHG signals in the previous experiment [62].

However, to observe this signature, it usually requires a high alignment degree which is not observed in our experiment due to the low alignment degree [see Figs. 7(a) and 7(c)]. But using our algorithm, the effects of multiple orbitals can still be clearly identified from our retrieved results.

#### IV. CONCLUSION

In summary, we developed a robust machine learning algorithm based on the iterative projection method to retrieve the single-molecule dipole moment from the harmonic spectra measured at different delays and polarization angles between the alignment and driving pulses. We introduce a linearization technique to efficiently use signals with nonzero polarization angles between the alignment and driving pulses. This could increase the accuracy and stability. Our numerical test shows that the accuracy and noise stability of our algorithm are good and robust enough to deal with the experimental data. We also demonstrate that our algorithm works well on the experimental data. The retrieved single-molecule dipoles agree well with the theoretical calculations with QRS theory. Moreover, multiple orbitals effects in the cutoff region is clearly identified from our retrievals even if the alignment degree is pretty low.

In addition, our algorithm can also be applied to other more complex linear molecules. We do not limit the complexity of the single-molecule dipole as in previous work [38] since the algorithm can directly solve the high-dimensional optimization problem. The current algorithm can be directly applied to nonsymmetric linear molecules by replacing the molecular alignment distribution with the molecular oriented distribution. However, for nonlinear molecules, the single-molecule dipole moment needs to be described with a second Euler

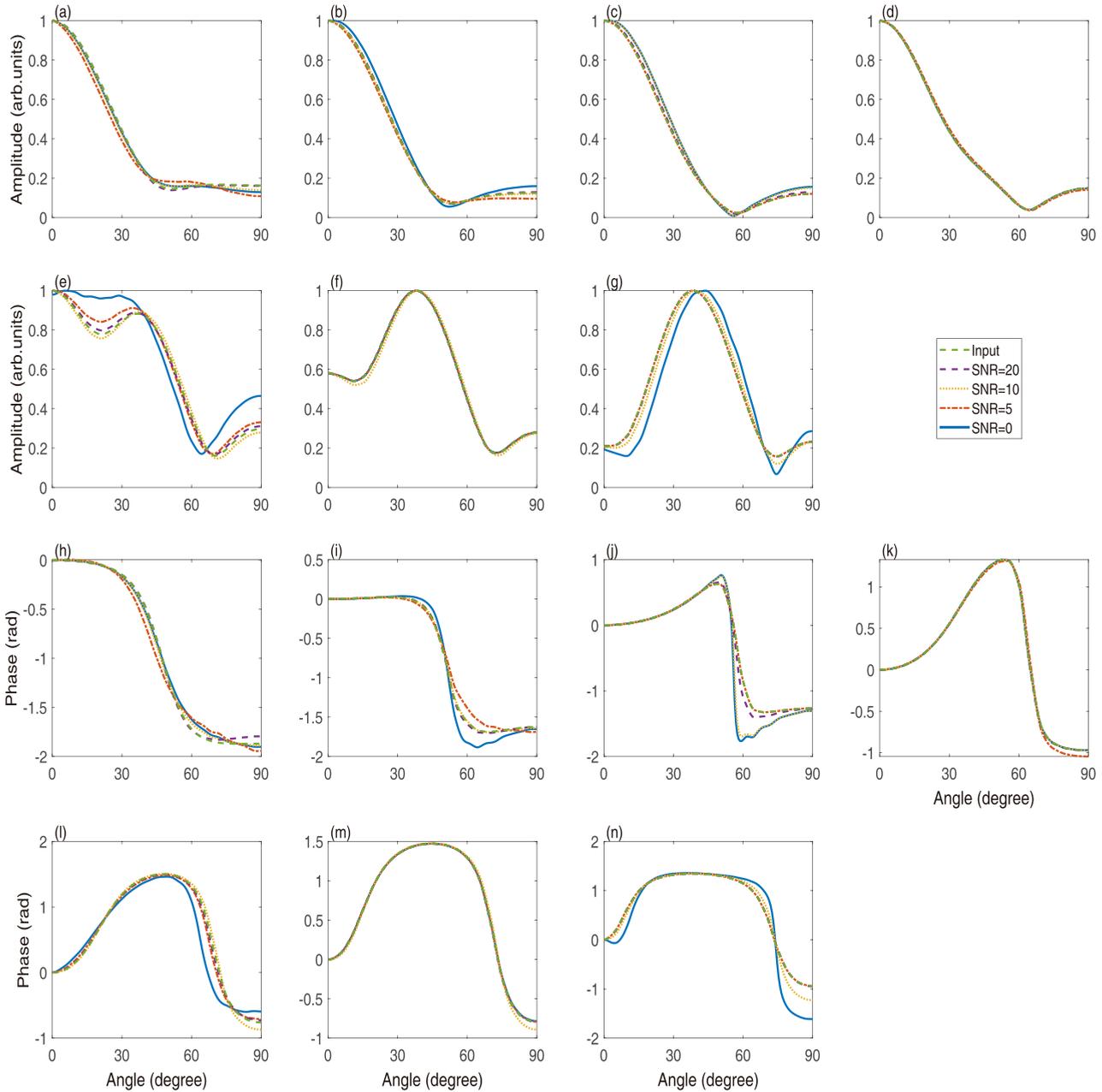


FIG. 8. The amplitude (a)–(g) and phase (h)–(n) of the single-molecule dipoles for H17, H19, H21, H23, H27, H29, H31. The dashed purple (dark gray) line, dotted yellow line, dash-dotted red line, and blue solid line are the retrieved results for SNR = 20, 10, 5, 0 of signals. The dashed green (light gray) line is the input single-molecule dipole calculated by QRS theory.

angle  $\chi$ . Since the one-dimensional (1D) alignment technique is used, the reconstructed result will be the integral dipole moment with respect to Euler angle  $\chi$ . Our work provides an algorithm for retrieving the single-molecule dipole of the linear molecule that can be used as a tool in detecting molecular structures and ultrafast dynamics.

#### ACKNOWLEDGMENTS

This work was supported by National Key Research and Development Program of China (Grant No. 2019YFA0308300), National Natural Science Foundation of China (Grants No. 91950202, No. 12225406, No. 12074136,

and No. 12021004), and the Natural Science Foundation of Hubei Province (Grant No. 2021CFB330).

#### APPENDIX A: THEORETICAL AND EXPERIMENTAL RESULTS OF OTHER HARMONICS

Here we present theoretical and experimental results of other harmonics to further illustrate the capability of our algorithm. Figure 8 shows the theoretical results of H17, H19, H21, H23, H27, H29, and H31. The theoretical results of all harmonics agree well with the ground truth and perform good noise stability and accuracy. Figure 9 shows the experimental results of H17 and H21 in the plateau region. In spite of the

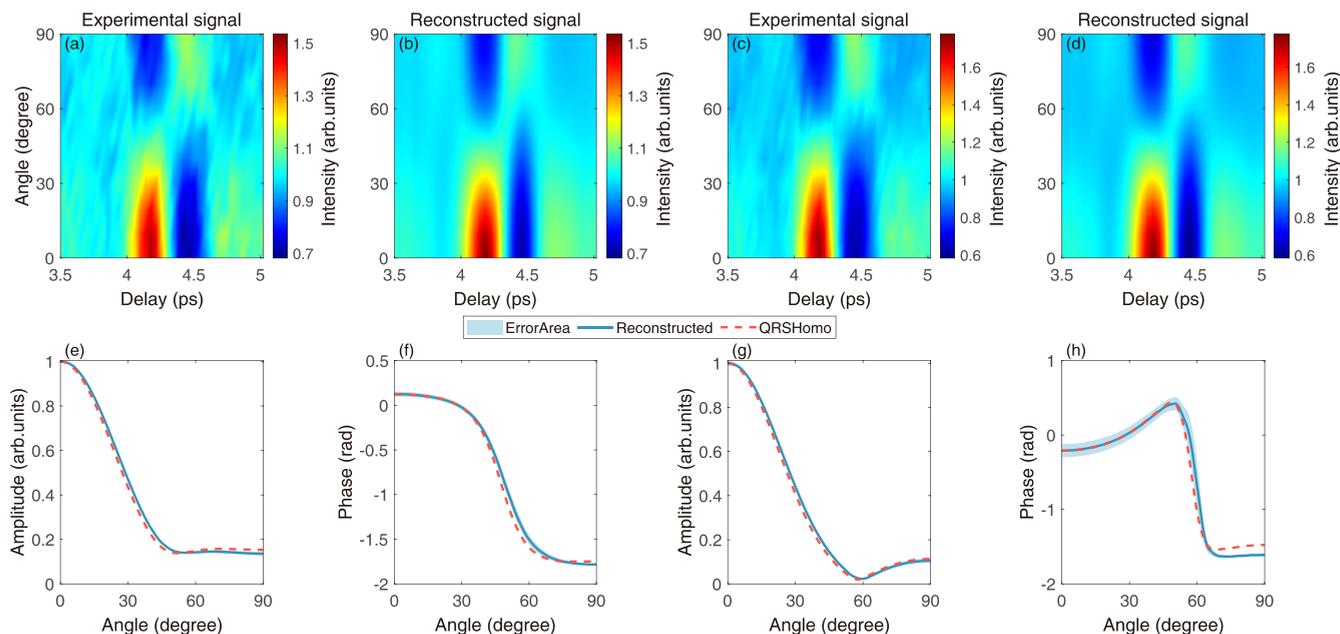


FIG. 9. (a,b) are the harmonic signals of H17 measured in our experiment and calculated with the retrieved single-molecule dipole. Blue solid lines in (e,f) are the retrieved amplitude and phase of the single-molecule dipole for the experimental signal in (a). Red dashed lines in (e,f) are the amplitude and phase of the single-molecule dipole of H17 for the HOMO orbital, which is calculated by QRS theory. (c,d) and (g,h) Same as (a,b) and (e,f), but for H21.

slight difference, the calculations well reproduce the general trend of the experimental retrievals, which also indicates a dominant contribution of the HOMO orbital in the generation of H17 and H21. Figure 10 shows the experimental results

of H27 and H29 in the cutoff region. Both theoretical calculations and experimental reconstructions indicate that the contribution of the HOMO-1 orbital significantly effect the harmonic spectra after H25.

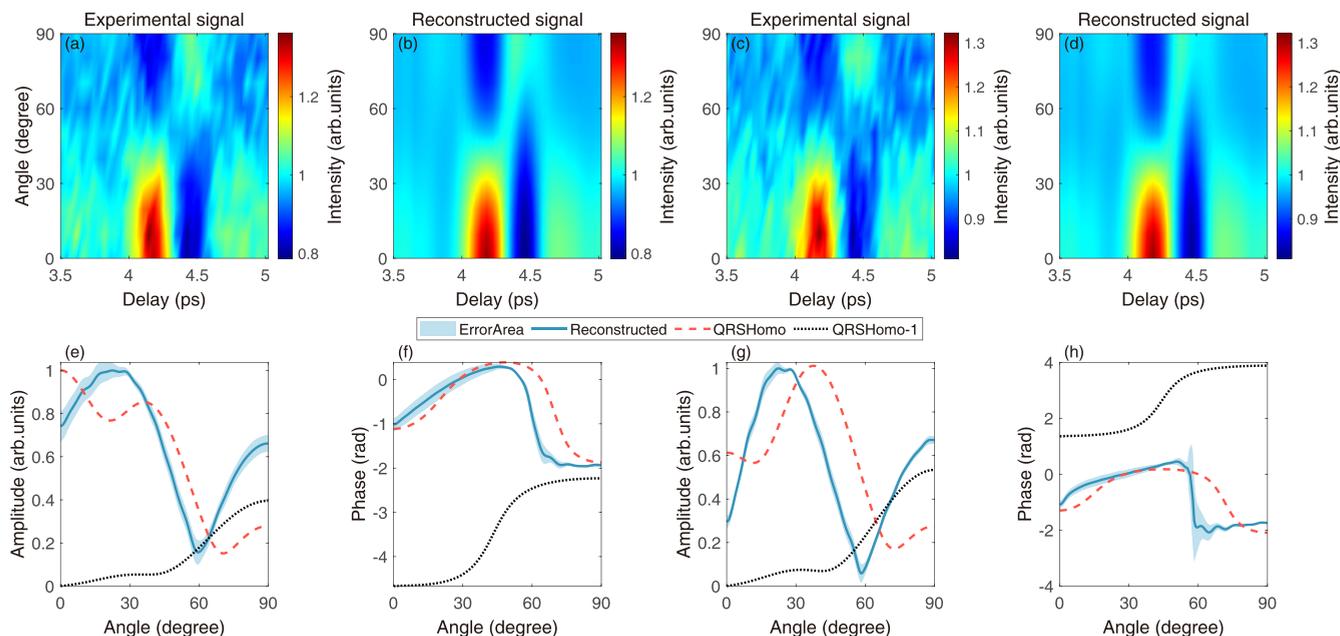


FIG. 10. (a,b) are the harmonic signals of H27 measured in our experiment and calculated with the retrieved single-molecule dipole. Blue solid lines in (e,f) are the retrieved amplitude and phase of the single-molecule dipole for the experimental signal in (a). Red dashed lines and black dotted lines in (e,f) are the amplitude and phase of the single-molecule dipole of H27 for HOMO and HOMO-1 orbitals which is calculated by QRS theory. (c,d) and (g,h) Same as (a,b) and (e,f), but for H29.

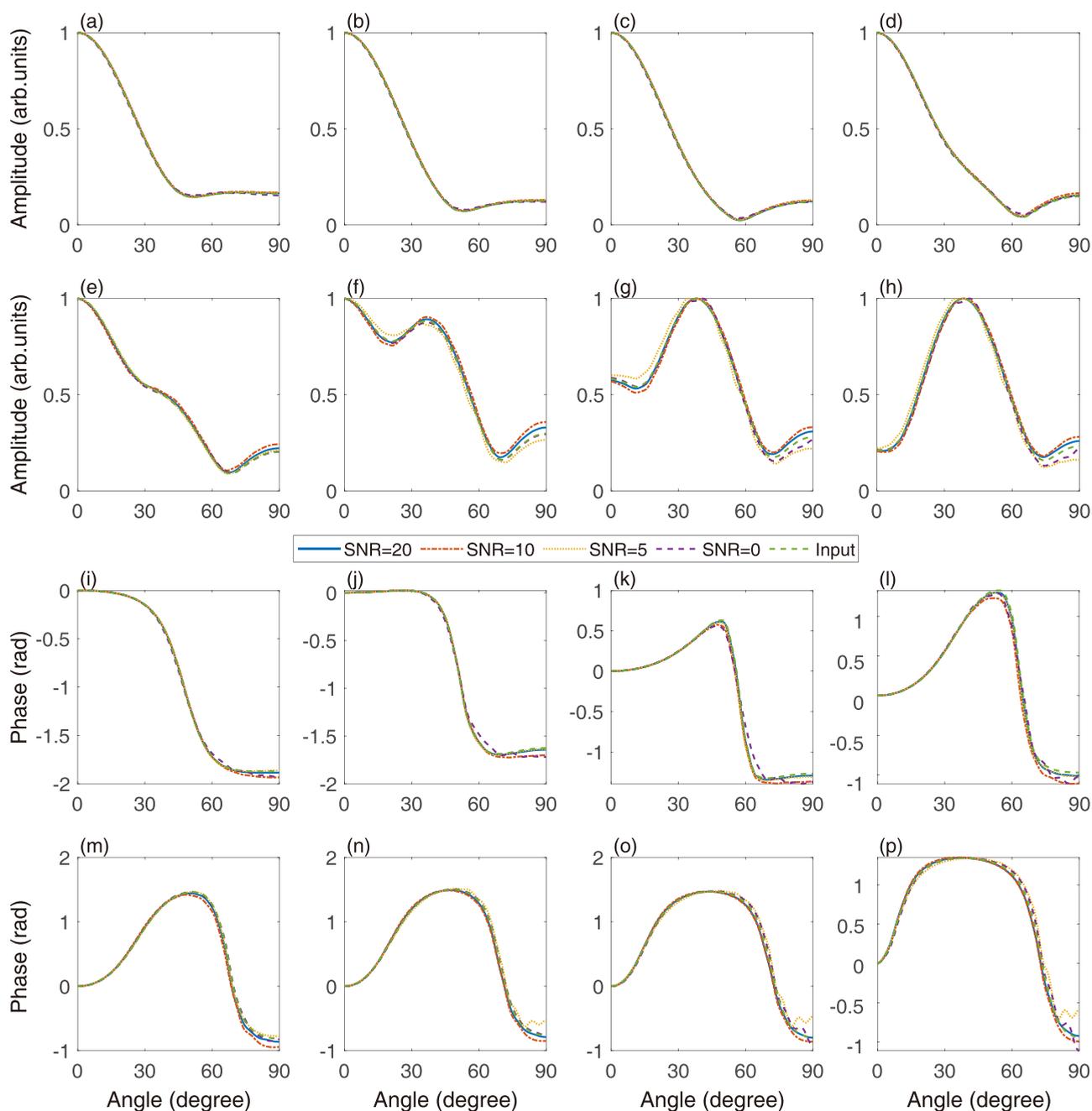


FIG. 11. The amplitude (a)–(h) and phase (i)–(p) of the single-molecule dipoles for H17, H19, H21, H23, H25, H27, H29, H31. The dashed purple (dark gray) line, dotted yellow line, dash-dotted red line, and blue solid line are the retrieved results for SNR = 0, 5, 10, 20 of alignment axis distribution. The dashed green (light gray) line is the input single-molecule dipole calculated by QRS theory.

#### APPENDIX B: NOISE STABILITY FOR NOISE IN THE MOLECULAR AXIS DISTRIBUTION

To demonstrate the noise stability for noise in the molecular axis distribution of our algorithm, we performed some numerical simulations. Gaussian random noise is added in the

molecular axis distribution with SNR = 20, 10, 5, 0, respectively. Figure 11 shows the reconstructed results compared with the ground truth for H17, H19, H21, H23, H25, H27, H29, and H31. All results show that our algorithm has good noise stability for noise in the molecular axis distribution.

[1] F. Remacle and R. D. Levine, An electronic time scale in chemistry, *Proc. Natl. Acad. Sci. USA* **103**, 6793 (2006).

[2] J. Breidbach and L. S. Cederbaum, Migration of holes: Numerical algorithms and implementation, *J. Chem. Phys.* **126**, 034101 (2007).

- [3] A. I. Kuleff, S. Lünemann, and L. S. Cederbaum, Electron-correlation-driven charge migration in oligopeptides, *Chem. Phys.* **414**, 100 (2013).
- [4] P. B. Corkum, Plasma Perspective on Strong Field Multiphoton Ionization, *Phys. Rev. Lett.* **71**, 1994 (1993).
- [5] M. Lewenstein, P. Balcou, M. Yu. Ivanov, A. L'Huillier, and P. B. Corkum, Theory of high-harmonic generation by low-frequency laser fields, *Phys. Rev. A* **49**, 2117 (1994).
- [6] J. Itatani, J. Levesque, D. Zeidler, H. Niikura, H. Pépin, J. C. Kieffer, P. B. Corkum, and D. M. Villeneuve, Tomographic imaging of molecular orbitals, *Nature (London)* **432**, 867 (2004).
- [7] C. Vozzi, M. Negro, F. Calegari, G. Sansone, M. Nisoli, S. De Silvestri, and S. Stagira, Generalized molecular orbital tomography, *Nature Phys* **7**, 822 (2011).
- [8] W. Li, X. Zhou, R. Lock, S. Patchkovskii, A. Stolow, H. C. Kapteyn, and M. M. Murnane, Time-resolved dynamics in  $N_2O_4$  probed using high harmonic generation, *Science* **322**, 1207 (2008).
- [9] S. Baker, J. S. Robinson, C. A. Haworth, H. Teng, R. A. Smith, C. C. Chirilă, M. Lein, J. W. G. Tisch, and J. P. Marangos, Probing proton dynamics in molecules on an attosecond time scale, *Science* **312**, 424 (2006).
- [10] O. Smirnova, Y. Mairesse, S. Patchkovskii, N. Dudovich, D. Villeneuve, P. Corkum, and M. Yu. Ivanov, High harmonic interferometry of multi-electron dynamics in molecules, *Nature (London)* **460**, 972 (2009).
- [11] A.-T. Le and C.-D. Lin, Imaging a chemical reaction, *Nature Photon* **4**, 671 (2010).
- [12] S. Haessler *et al.*, Attosecond imaging of molecular electronic wavepackets, *Nat. Phys* **6**, 200 (2010).
- [13] P. M. Kraus *et al.*, Measurement and laser control of attosecond charge migration in ionized iodoacetylene, *Science* **350**, 790 (2015).
- [14] L. He *et al.*, Filming movies of attosecond charge migration in single molecules with high harmonic spectroscopy, *Nat. Commun.* **13**, 4595 (2022).
- [15] L. He, P. Lan, A.-T. Le, B. Wang, B. Wang, X. Zhu, P. Lu, and C. D. Lin, Real-Time Observation of Molecular Spinning with Angular High-Harmonic Spectroscopy, *Phys. Rev. Lett.* **121**, 163201 (2018).
- [16] Y. He, L. He, P. Lan, B. Wang, L. Li, X. Zhu, W. Cao, and P. Lu, Direct imaging of molecular rotation with high-order-harmonic generation, *Phys. Rev. A* **99**, 053419 (2019).
- [17] D. Pavičić, K. F. Lee, D. M. Rayner, P. B. Corkum, and D. M. Villeneuve, Direct Measurement of the Angular Dependence of Ionization for  $N_2$ ,  $O_2$ , and  $CO_2$  in Intense Laser Fields, *Phys. Rev. Lett.* **98**, 243001 (2007).
- [18] K. Yoshii, G. Miyaji, and K. Miyazaki, Dynamic Properties of Angle-Dependent High-Order Harmonic Generation from Coherently Rotating Molecules, *Phys. Rev. Lett.* **101**, 183902 (2008).
- [19] K. Yoshii, G. Miyaji, and K. Miyazaki, Retrieving Angular Distributions of High-Order Harmonic Generation from a Single Molecule, *Phys. Rev. Lett.* **106**, 013904 (2011).
- [20] J. Ortigoso, M. Rodriguez, M. Gupta, and B. Friedrich, Time evolution of pendular states created by the interaction of molecular polarizability with a pulsed nonresonant laser field, *J. Chem. Phys.* **110**, 3870 (1999).
- [21] H. Stapelfeldt and T. Seideman, Colloquium: Aligning molecules with strong laser pulses, *Rev. Mod. Phys.* **75**, 543 (2003).
- [22] M. Artamonov and T. Seideman, Theory of three-dimensional alignment by intense laser pulses, *J. Chem. Phys.* **128**, 154313 (2008).
- [23] L. Holmegaard, J. H. Nielsen, I. Nevo, H. Stapelfeldt, F. Filsinger, J. Küpper, and G. Meijer, Laser-Induced Alignment and Orientation of Quantum-State-Selected Large Molecules, *Phys. Rev. Lett.* **102**, 023001 (2009).
- [24] R. Boll, D. Anielski, C. Bostedt, J. D. Bozek, L. Christensen, R. Coffee, S. De, P. Declava, S. W. Epp, B. Erk *et al.*, Femtosecond photoelectron diffraction on laser-aligned molecules: Towards time-resolved imaging of molecular structure, *Phys. Rev. A* **88**, 061402(R) (2013).
- [25] D. Rolles *et al.*, Femtosecond x-ray photoelectron diffraction on gas-phase dibromobenzene molecules, *J. Phys. B: At. Mol. Opt. Phys.* **47**, 124035 (2014).
- [26] J. G. Underwood, B. J. Sussman, and A. Stolow, Field-Free Three Dimensional Molecular Axis Alignment, *Phys. Rev. Lett.* **94**, 143002 (2005).
- [27] A. Rouzée, S. Guérin, V. Boudon, B. Lavorel, and O. Faucher, Field-free one-dimensional alignment of ethylene molecule, *Phys. Rev. A* **73**, 033418 (2006).
- [28] X. Ren, V. Makhija, and V. Kumarappan, Multipulse Three-Dimensional Alignment of Asymmetric Top Molecules, *Phys. Rev. Lett.* **112**, 173602 (2014).
- [29] C. Marceau, V. Makhija, D. Platzer, A. Yu. Naumov, P. B. Corkum, A. Stolow, D. M. Villeneuve, and P. Hockett, Molecular Frame Reconstruction Using Time-Domain Photoionization Interferometry, *Phys. Rev. Lett.* **119**, 083401 (2017).
- [30] I. Thomann, R. Lock, V. Sharma, E. Gagnon, S. T. Pratt, H. C. Kapteyn, M. M. Murnane, and W. Li, Direct measurement of the angular dependence of the single-photon ionization of aligned  $N_2$  and  $CO_2$ , *J. Phys. Chem. A* **112**, 9382 (2008).
- [31] J. Mikosch, A. E. Boguslavskiy, I. Wilkinson, M. Spanner, S. Patchkovskii, and A. Stolow, Channel- and Angle-Resolved Above Threshold Ionization in the Molecular Frame, *Phys. Rev. Lett.* **110**, 023004 (2013).
- [32] P. Sándor, A. Sissay, F. Mauger, P. M. Abanador, T. T. Gorman, T. D. Scarborough, M. B. Gaarde, K. Lopata, K. J. Schafer, and R. R. Jones, Angle dependence of strong-field single and double ionization of carbonyl sulfide, *Phys. Rev. A* **98**, 043425 (2018).
- [33] P. Sándor, A. Sissay, F. Mauger, M. W. Gordon, T. T. Gorman, T. D. Scarborough, M. B. Gaarde, K. Lopata, K. J. Schafer, and R. R. Jones, Angle-dependent strong-field ionization of halomethanes, *J. Chem. Phys.* **151**, 194308 (2019).
- [34] S. J. Weber, M. Oppermann, and J. P. Marangos, Role of Rotational Wave Packets in Strong Field Experiments, *Phys. Rev. Lett.* **111**, 263601 (2013).
- [35] O. Smirnova and M. Ivanov, Towards a one-femtosecond film, *Nature Phys* **6**, 159 (2010).
- [36] E. Lindroth, F. Calegari, L. Young, M. Harmand, N. Dudovich, N. Berrah, and O. Smirnova, Challenges and opportunities in attosecond and XFEL science, *Nat. Rev. Phys.* **1**, 107 (2019).
- [37] V. Makhija, X. Ren, D. Gockel, A.-T. Le, and V. Kumarappan, Orientation resolution through rotational coherence spectroscopy, [arXiv:1611.06476v2](https://arxiv.org/abs/1611.06476v2).

- [38] B. Wang, Y. He, X. Zhao, L. He, P. Lan, P. Lu, and C. D. Lin, Retrieval of full angular- and energy-dependent complex transition dipoles in the molecular frame from laser-induced high-order harmonic signals with aligned molecules, *Phys. Rev. A* **101**, 063417 (2020).
- [39] X. Wang, A.-T. Le, Z. Zhou, H. Wei, and C. D. Lin, Theory of retrieving orientation-resolved molecular information using time-domain rotational coherence spectroscopy, *Phys. Rev. A* **96**, 023424 (2017).
- [40] O. Neufeld, O. Wengrowicz, O. Peleg, A. Rubio, and O. Cohen, Detecting multiple chiral centers in chiral molecules with high harmonic generation, *Opt. Express* **30**, 3729 (2022).
- [41] N. D. Klimkin, Á. Jiménez-Galán, R. E. F. Silva, and M. Ivanov, Deep neural networks for high harmonic spectroscopy in solids, [arXiv:2106.08638v2](https://arxiv.org/abs/2106.08638v2).
- [42] S. Pabst, P. J. Ho, and R. Santra, Computational studies of x-ray scattering from three-dimensionally-aligned asymmetric-top molecules, *Phys. Rev. A* **81**, 043425 (2010).
- [43] Y. He, L. He, P. Wang, B. Wang, S. Sun, R. Liu, B. Wang, P. Lan, and P. Lu, Measuring the rotational temperature and pump intensity in molecular alignment experiments via high harmonic generation, *Opt. Express* **28**, 21182 (2020).
- [44] A.-T. Le, R. R. Lucchese, S. Tonzani, T. Morishita, and C. D. Lin, Quantitative rescattering theory for high-order harmonic generation from molecules, *Phys. Rev. A* **80**, 013401 (2009).
- [45] P. Wang *et al.*, Rotational echo spectroscopy for accurate measurement of molecular alignment, *Opt. Lett.* **47**, 1033 (2022).
- [46] Y. Censor, W. Chen, P. L. Combettes, R. Davidi, and G. T. Herman, On the effectiveness of projection methods for convex feasibility problems with linear inequality constraints, *Comput. Optim. Appl.* **51**, 1065 (2012).
- [47] S. Gravel and V. Elser, Divide and conquer: A general approach to constraint satisfaction, *Phys. Rev. E* **78**, 036706 (2008).
- [48] V. Elser, Phase retrieval by iterated projections, *J. Opt. Soc. Am. A* **20**, 40 (2003).
- [49] P. Thibault and A. Menzel, Reconstructing state mixtures from diffraction measurements, *Nature (London)* **494**, 68 (2013).
- [50] S. Marchesini, H. He, H. N. Chapman, S. P. Hau-Riege, A. Noy, M. R. Howells, U. Weierstall, and J. C. H. Spence, X-ray image reconstruction from a diffraction pattern alone, *Phys. Rev. B* **68**, 140101(R) (2003).
- [51] D. R. Luke, Relaxed averaged alternating reflections for diffraction imaging, *Inverse Probl.* **21**, 37 (2004).
- [52] P. Thibault, V. Elser, C. Jacobsen, D. Shapiro, and D. Sayre, Reconstruction of a yeast cell from x-ray diffraction data, *Acta Crystallogr A* **62**, 248 (2006).
- [53] D. J. Kane, Principal components generalized projections: A review [invited], *J. Opt. Soc. Am. B* **25**, A120 (2008).
- [54] Z. Tang, M. Deng, C. Xiao, and J. Yu, Projection onto Convex Sets Super-Resolution Image Reconstruction Based on Wavelet Bi-Cubic Interpolation, in *Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology*, Vol. 1 (IEEE, New York, 2011), pp. 351–354.
- [55] R. Penrose, A generalized inverse for matrices, *Mathematical Proceedings of the Cambridge Philosophical Society* **51**, 406 (1955).
- [56] A.-T. Le, R. D. Picca, P. D. Fainstein, D. A. Telnov, M. Lein, and C. D. Lin, Theory of high-order harmonic generation from molecules by intense laser pulses, *J. Phys. B: At. Mol. Opt. Phys.* **41**, 081002 (2008).
- [57] C. D. Lin, A.-T. Le, Z. Chen, T. Morishita, and R. Lucchese, Strong-field rescattering physics—self-imaging of a molecule by its own electrons, *J. Phys. B: At. Mol. Opt. Phys.* **43**, 122001 (2010).
- [58] T. Morishita, A.-T. Le, Z. Chen, and C. D. Lin, Accurate Retrieval of Structural Information from Laser-Induced Photoelectron and High-Order Harmonic Spectra by Few-Cycle Laser Pulses, *Phys. Rev. Lett.* **100**, 013903 (2008).
- [59] L. He, P. Lan, Q. Zhang, C. Zhai, F. Wang, W. Shi, and P. Lu, Spectrally resolved spatiotemporal features of quantum paths in high-order-harmonic generation, *Phys. Rev. A* **92**, 043403 (2015).
- [60] L. He *et al.*, Monitoring ultrafast vibrational dynamics of isotopic molecules with frequency modulation of high-order harmonics, *Nat. Commun.* **9**, 1108 (2018).
- [61] C. Jin, A.-T. Le, and C. D. Lin, Analysis of effects of macroscopic propagation and multiple molecular orbitals on the minimum in high-order harmonic generation of aligned CO<sub>2</sub>, *Phys. Rev. A* **83**, 053409 (2011).
- [62] B. K. McFarland, J. P. Farrell, P. H. Bucksbaum, and M. Gühr, High harmonic generation from multiple orbitals in N<sub>2</sub>, *Science* **322**, 1232 (2008).