

Grid Evolution for Doubly Fractional Channel Estimation in OTFS Systems

Xiangjun Li, Pingzhi Fan, Qianli Wang, Zilong Liu

Abstract—In orthogonal time-frequency space communications, the performances of existing on-grid and off-grid channel estimation (CE) schemes are determined by the delay-Doppler (DD) grid density. In practice, multiple real-life DD channel responses might be co-located within a same DD grid interval, leading to performance degradation. A finer grid interval is needed to distinguish these responses, but this could result in a significantly higher CE complexity when traditional methods are used. To address this issue, a grid evolution method for doubly fractional CE is proposed by evolving the initially uniform coarse DD grid into a non-uniform dense grid. Simulation results show that our proposed method leads to improved computational efficiency, and achieves a good trade-off between CE performance and complexity.

Index Terms—OTFS, grid evolution, channel estimation, sparse Bayesian learning, fractional delay-Doppler.

I. INTRODUCTION

THE next generation communication systems must support reliable information exchanges in highly dynamic environments, e.g., vehicle-to-everything systems, high-speed railways, etc. In these environments, one needs to deal with the fast time-varying channels incurred by high mobility. The orthogonal frequency division multiplexing may be infeasible as its orthogonality can be easily destroyed by large Doppler. Against this background, orthogonal time-frequency space (OTFS) modulation has emerged in recent years due to its excellent performance in high mobility channels [1]–[5]. The basic principle of OTFS is to send the data symbols over delay-Doppler (DD) domain to achieve full-diversity.

Channel estimation (CE) plays an important role in OTFS system design as it has direct impact to the detection performance. A number of OTFS CE schemes have been proposed. An embedded frame structure and a threshold CE method are proposed in [2]. In [6], [7], the authors leveraged the sparsity in DD domain which permits the use of orthogonal matching pursuit (OMP) and sparse Bayesian learning (SBL) for CE. It is noted that these works assume that the real-life channel response is on the grid points. However, this assumption may not be valid in practice due to the fractional delay and Doppler values in real-world transmission. To address this problem, off-grid sparse Bayesian inference (OGSBI) [8] was adopted

in [9], leading to one-dimensional (1D) and two-dimensional (2D) off-grid CE scheme. The 1D scheme demonstrates a good performance, but with high complexity. In contrast, the 2D scheme strikes a balance between complexity and performance. Another 2D off-grid scheme was developed in [10] to eliminate some additional errors. Yet it still suffers from a high complexity. These CE methods generally set a fixed initial grid. A denser initial grid usually achieves better CE performance, but it also brings in a significant computational burden. Furthermore, multiple real-life channel responses may be co-located in the same DD grid interval even a denser grid is used and this could lead to deteriorated CE performance.

In this letter, inspired by the grid evolution (GE) method in [11], a GE method for doubly fractional CE is proposed to tackle the above problems. Different from the method in [11], our proposed GE is a 2D method, whereby the fission and adjustment of grid points are performed separately to avoid unnecessary calculations. Compared with the GE scheme in [12], the number of grid points in the proposed GE scheme is adjustable and hence it is generic and more flexible. Specifically, the proposed GE adaptively evolves from an initial uniform coarse grid to a non-uniform dense grid without a fixed interval. The GE framework contains three processes, i.e., the learning, the fission and the adjustment. The learning process estimates the channel response at the grid points. The fission process adds new grid points, and separates multiple real-life channel responses into different DD grid intervals. The adjustment process combines the off-grid parameters into the current grid for decreasing the modeling error. These processes iterate alternately to achieve adaptive grid refinement and therefore obtain more accurate CE performance. Compared with the previous schemes using a fixed uniform grid [7], [9], [10], the proposed GE scheme gives rise to significantly lower complexity due to smaller number of grid points. And if similar number of grid points are taken, the proposed GE scheme shows better CE performance.

II. SYSTEM MODELS

A. Signal Model

The DD channel response can be expressed as [1]–[3]

$$h(\nu, \tau) = \sum_{n=1}^P h_n \delta(\nu - \nu_n) \delta(\tau - \tau_n), \quad (1)$$

where P is the number of paths, $h_n, \tau_n \in (0, \tau_{\max})$ and $\nu_n \in (-\nu_{\max}, \nu_{\max})$ are the channel coefficient, delay, and

Xiangjun Li, Pingzhi Fan and Qianli Wang are with the School of Info Sci & Tech, Southwest Jiaotong University, Chengdu, China, and also with the Communications & Sensor Networks for Modern Transportation (CSNMT) Int. Cooperation Research Centre of China, Chengdu, China. Zilong Liu is with the School of Computer Science and Electronics Engineering, University of Essex, U. K. Corresponding author: Qianli Wang. Emails: lxj@my.swjtu.edu.cn; pzf@swjtu.edu.cn; qianli_wang@qq.com; zilong.liu@essex.ac.uk. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

Doppler of the n -th path, respectively. The symbol duration and bandwidth of OTFS system are NT and $M\Delta f$, respectively. N , M , T and Δf are number of time slots, number of subcarriers, slots duration and subcarrier spacing, respectively.

In this letter, a single pilot x_p with Doppler index k'_p and delay index l'_p in the DD grid is considered. To reduce interference between data and pilot, both the guard interval and the CE region in [9] is used. The sizes of the CE region are $N_T = (2k_{\max} + 1)$ and $M_T = (l_{\max} + 1)$ respectively, where $k_{\max} = \nu_{\max}NT$ and $l_{\max} = \tau_{\max}M\Delta f$. According to [2], [3], [9], the OTFS input-output relationship is

$$y_{\text{DD}}[k, l] = x_p \sum_{n=1}^P \tilde{h}_n w_\nu(k, k'_p, k_{\nu_n}) w_\tau^H(l, l'_p, l_{\tau_n}) + z[k, l], \quad (2)$$

where $y_{\text{DD}}[k, l]$ is the received signals in the DD domain, $k \in \{k'_p - k_{\max}, \dots, k'_p + k_{\max}\}$, $l \in \{l'_p, \dots, l'_p + l_{\max}\}$, $z[k, l] \sim \mathcal{CN}(0, \lambda^{-1})$ is the noise and λ^{-1} is the variance, $\tilde{h}_n = h_n e^{-j2\pi\nu_n\tau_n}$, $k_{\nu_n} = \nu_n NT$, $l_{\tau_n} = \tau_n M\Delta f$, $w_\nu(\cdot) = w_\tau(\cdot) = \mathcal{F}(\cdot)$. In [9], [10], $\mathcal{F}(\cdot)$ is denoted by

$$\mathcal{F}(\eta, \xi, \gamma) = \frac{1}{Q} \left[e^{-j(Q-1)\pi\frac{\eta-\xi-\gamma}{Q}} \frac{\sin(\pi(\eta-\xi-\gamma))}{\sin(\frac{\pi(\eta-\xi-\gamma)}{Q})} \right], \quad (3)$$

where $Q = N$ for $w_\nu(\cdot)$ and $Q = M$ for $w_\tau(\cdot)$.

B. Off-Grid Model and GE Model

A finer grid is usually used for sparse representation as that of [9]. In this case, the grid points of the initial uniform DD grid are $\{\{\tilde{\mathbf{k}}_\nu\} \times \{\tilde{\mathbf{l}}_\tau\}\} \in \mathbb{R}^{L_\nu \times L_\tau}$, $\tilde{\mathbf{k}}_\nu = [-k_{\max}, \dots, k_{\max}]^T \in \mathbb{R}^{L_\nu \times 1}$, $\tilde{\mathbf{l}}_\tau = [-l_{\max}, \dots, l_{\max}]^T \in \mathbb{R}^{L_\tau \times 1}$. Therefore, the initial delay and Doppler resolution are $r_\nu = \frac{2k_{\max}}{L_\nu - 1}$ and $r_\tau = \frac{l_{\max}}{L_\tau - 1}$, respectively. Let the number of grid points in the sampled DD grid be $L = L_\nu L_\tau$. After vectorizing the DD grid, all DD grid points can be expressed as $\tilde{\mathbf{S}} = \{\tilde{\mathbf{k}}_\nu, \tilde{\mathbf{l}}_\tau\}$, $\tilde{\mathbf{l}}_\tau = [l_0, l_1, \dots, l_i, \dots, l_{L-1}]^T$, $\tilde{\mathbf{k}}_\nu = [k_0, k_1, \dots, k_i, \dots, k_{L-1}]^T$. Let $i \in \{1, \dots, L\}$, $\mathbf{w}_\nu(k_i) = [w_\nu(k'_p - k_{\max}, k'_p, k_i), \dots, w_\nu(k'_p + k_{\max}, k'_p, k_i)]^T \in \mathbb{C}^{N_T \times 1}$, $\mathbf{w}_\tau(l_i) = [w_\tau(l'_p, l'_p, l_i), \dots, w_\tau(l'_p + l_{\max}, l'_p, l_i)]^T \in \mathbb{C}^{M_T \times 1}$. Then, the on-grid model based on Eq. (2) is

$$\mathbf{y} = x_p \Phi_{\text{I}}(\tilde{\mathbf{S}}) \tilde{\mathbf{h}} + \mathbf{z}, \quad (4)$$

where $\mathbf{y}, \mathbf{z} \in \mathbb{C}^{N_T M_T \times 1}$, $\mathbf{y} = \text{vec}(\mathbf{Y})$, $\mathbf{Y} \in \mathbb{C}^{N_T \times M_T}$ is the matrix form of $y_{\text{DD}}[k, l]$, $\tilde{\mathbf{h}} \in \mathbb{C}^{L \times 1}$, $\Phi_{\text{I}}(\tilde{\mathbf{S}}) = \psi \odot [\phi_{\text{I}}(k_0, l_0), \dots, \phi_{\text{I}}(k_{L-1}, l_{L-1})] \in \mathbb{C}^{N_T M_T \times L}$ is the on-grid part of the measurement matrix corresponding to the current DD grid, $\phi_{\text{I}}(k_i, l_i) = \text{vec}(\mathbf{w}_\nu(k_i) \mathbf{w}_\tau^H(l_i)) \in \mathbb{C}^{N_T M_T \times 1}$, $\psi = [\varphi_0, \dots, \varphi_{L-1}] \in \mathbb{C}^{N_T M_T \times L}$, $\varphi_i = e^{-\frac{j2\pi k_i l_i}{NM}} \mathbf{1}$, $\mathbf{1} \in \mathbb{C}^{N_T M_T \times 1}$ is the all-ones column vector, $\text{vec}(\cdot)$ is the vectorization operation, \odot is the dot product.

Since in practice the channel response generally does not fall exactly on the sampled DD grid points, an off-grid CE scheme was proposed in [9], [10]. Denote by $(k_i \in \tilde{\mathbf{k}}_\nu, l_i \in \tilde{\mathbf{l}}_\tau)$ the i -th grid point which is closest to

(k_{ν_n}, l_{τ_n}) in the discrete DD plane. Then a linear approximation in [9] can be obtained by first-order Taylor expansion, i.e.,

$$\Phi(k_{\nu_n}, l_{\tau_n}) = \phi_{\text{I}}(k_i, l_i) + \phi_\nu(k_i, l_i) \kappa_i + \phi_\tau(k_i, l_i) \iota_i + o(\kappa_i) + o(\iota_i), \quad (5)$$

where $\kappa_i = k_{\nu_n} - k_i$, $\iota_i = l_{\tau_n} - l_i$, $o(\kappa_i)$ and $o(\iota_i)$ are the modeling errors, $\phi_\nu(k_i, l_i) = \text{vec}(\mathbf{w}'_\nu(k_i) \mathbf{w}_\tau^H(l_i))$, $\phi_\tau(k_i, l_i) = \text{vec}(\mathbf{w}_\nu(k_i) (\mathbf{w}'_\tau(l_i))^H)$, $\mathbf{w}'_\nu(k_i)$ and $\mathbf{w}'_\tau(l_i)$ are the partial derivatives with respect to k_i and l_i respectively. Here, κ_i and ι_i are assumed to follow uniform distributions

$$\kappa_i \sim \mathcal{U}\left[-\frac{1}{2}\tilde{r}_{\nu-}, \frac{1}{2}\tilde{r}_{\nu+}\right], \iota_i \sim \mathcal{U}\left[-\frac{1}{2}\tilde{r}_{\tau-}, \frac{1}{2}\tilde{r}_{\tau+}\right], \quad (6)$$

where \tilde{r}_ν and \tilde{r}_τ respectively represent the Doppler and delay resolutions between a grid point and its adjacent grid points, $+$ and $-$ denote positive and negative directions respectively.

Let $\tilde{\boldsymbol{\kappa}} = [\kappa_0, \kappa_1, \dots, \kappa_{L-1}]^T \in \mathbb{C}^{L \times 1}$, $\tilde{\boldsymbol{\iota}} = [\iota_0, \iota_1, \dots, \iota_{L-1}]^T \in \mathbb{C}^{L \times 1}$. Similar to $\Phi_{\text{I}}(\tilde{\mathbf{S}})$, $\phi_\nu(k_i, l_i)$ and $\phi_\tau(k_i, l_i)$ can be arranged in order to obtain $\Phi_\nu(\tilde{\mathbf{S}})$ and $\Phi_\tau(\tilde{\mathbf{S}})$, respectively. Then the measurement matrix $\Phi(\tilde{\mathbf{S}}, \tilde{\boldsymbol{\kappa}}, \tilde{\boldsymbol{\iota}}) \in \mathbb{C}^{N_T M_T \times L}$ can be expressed as

$$\Phi(\tilde{\mathbf{S}}, \tilde{\boldsymbol{\kappa}}, \tilde{\boldsymbol{\iota}}) = \Phi_{\text{I}}(\tilde{\mathbf{S}}) + \Phi_\nu(\tilde{\mathbf{S}}) \text{diag}\{\tilde{\boldsymbol{\kappa}}\} + \Phi_\tau(\tilde{\mathbf{S}}) \text{diag}\{\tilde{\boldsymbol{\iota}}\}, \quad (7)$$

where $\text{diag}\{\cdot\}$ is the diagonal matrix operator. After absorbing the approximation error into the noise, the observation model can be written as

$$\mathbf{y} = x_p \Phi(\tilde{\mathbf{S}}, \tilde{\boldsymbol{\kappa}}, \tilde{\boldsymbol{\iota}}) \tilde{\mathbf{h}} + \mathbf{z}. \quad (8)$$

Note that $\tilde{\mathbf{S}}$ is constant in both the on-grid and off-grid models, meaning that r_ν and r_τ are constants. As mentioned before, the performance of the on-grid and off-grid models will deteriorate when multiple real-life channel responses are located in the same DD interval. Therefore, the proposed GE scheme uses a varying \mathbf{S} to decrease the modeling error, where $\mathbf{S} = \{\tilde{\mathbf{S}}_{\text{ini}}, \tilde{\mathbf{S}}_{\text{GE}}\}$, the $\tilde{\mathbf{S}}_{\text{ini}}$ is the \mathbf{S} with $r_\nu = r_\tau = 1$, $\tilde{\mathbf{S}}_{\text{GE}}$ represents the added grid points in the GE scheme. Therefore, the model in GE is,

$$\mathbf{y} = x_p \Phi(\mathbf{S}, \boldsymbol{\kappa}, \boldsymbol{\iota}) \mathbf{h} + \mathbf{z}, \quad (9)$$

where the sizes of $\boldsymbol{\kappa}$, $\boldsymbol{\iota}$ and \mathbf{h} are corresponding to \mathbf{S} .

III. GRID EVOLUTION CHANNEL ESTIMATION

To get the grid points \mathbf{S} and estimate \mathbf{h} in Eq. (9), GE method is adopted to increase the local resolution for a more accurate representation of the sparse channel responses in the DD domain. As shown in Fig. 1, for the off-grid scheme using a coarse DD grid, the modeling error will be large. There are two sources of modeling error. One is that larger off-grid gaps lead to large $o(\kappa_i)$ and $o(\iota_i)$. The second is because multiple channel responses are located in the same DD interval and they may not be distinguished. By using the GE method, more grid points will be generated around the true responses by decreasing the off-grid gaps to distinguish the close paths.



Fig. 1: The fission process for the GE scheme.

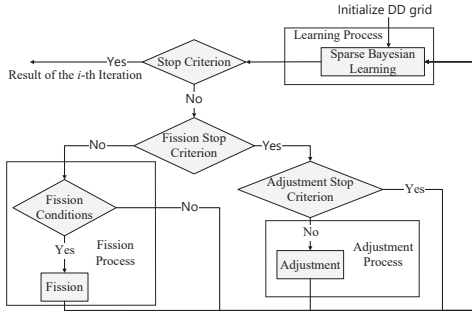


Fig. 2: The procedure of GE channel estimation.

Note that this procedure is different from direct grid refinement like that of chirp z-transform. The GE method is based on the compressive sensing thus the coefficients corresponding to these grid points will have to meet the sparsity constraint, resulting a sparse representation of the channel responses.

The GE method consists of the learning, the fission and the adjustment processes, as shown in Fig. 2. The fission process can distinguish multiple channel responses within the same initial DD interval. The adjustment process combines the off-grid parameters into the current grid to reduce the off-grid gap. The learning process estimates the channel response at the grid points by SBL. Alternate iterations of the learning-fission and alternate iterations of the learning-adjustment are performed sequentially to adaptively refine the DD grid for improving the local DD resolution and reducing modeling error. To strike a trade-off between modeling error and computational workload, stop criteria are proposed for fission and adjustment.

A. Learning Process

OGSBI in [8] is used to estimate the channel parameters in the learning process. According to [8], [9], the posterior conditional distribution of \mathbf{h} is assumed to be

$$p(\mathbf{h} | \mathbf{y}; \boldsymbol{\alpha}, \boldsymbol{\kappa}, \boldsymbol{\nu}, \lambda) = \mathcal{CN}(\mathbf{h} | \boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (10)$$

where $\boldsymbol{\alpha}$ is the hyperparameter that models the sparsity of \mathbf{h} and follows a Gamma distribution controlled by the parameter ρ , λ follows a Gamma distribution determined by a and b , The mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ for the (k) -th iteration are

$$\boldsymbol{\mu}^{(k)} = \lambda^{(k)} \boldsymbol{\Sigma}^{(k)} \boldsymbol{\Phi}^H(\mathbf{S}^{(k)}, \boldsymbol{\kappa}^{(k)}, \boldsymbol{\nu}^{(k)}) \mathbf{y}, \quad (11)$$

$$\boldsymbol{\Sigma}^{(k)} = \text{diag}(\boldsymbol{\alpha}^{(k)}) - \left(\boldsymbol{\alpha}^{(k)} \odot \left(\boldsymbol{\alpha}^{(k)} \odot \mathbf{C} \right)^T \right)^T, \quad (12)$$

where $\mathbf{C} = \boldsymbol{\Phi}^H(\mathbf{S}^{(k)}, \boldsymbol{\kappa}^{(k)}, \boldsymbol{\nu}^{(k)}) \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\Phi}(\mathbf{S}^{(k)}, \boldsymbol{\kappa}^{(k)}, \boldsymbol{\nu}^{(k)})$, $\boldsymbol{\Sigma}_y = \boldsymbol{\Phi}(\mathbf{S}^{(k)}, \boldsymbol{\kappa}^{(k)}, \boldsymbol{\nu}^{(k)}) \left(\boldsymbol{\alpha}^{(k)} \odot \boldsymbol{\Phi}^H(\mathbf{S}^{(k)}, \boldsymbol{\kappa}^{(k)}, \boldsymbol{\nu}^{(k)}) \right) + \lambda^{(k)} \mathbf{I}$, $\boldsymbol{\alpha} \odot \mathbf{C}$ is the dot product of the elements of the column vector $\boldsymbol{\alpha}$ with the corresponding rows of \mathbf{C} . Different from the traditional update formulation of $\boldsymbol{\Sigma}$, the diagonal matrix property is considered here to reduce the complexity.

The expectation-maximization algorithm is used to update each hyper-parameter for the $(k+1)$ -th iteration, i.e.,

$$\alpha(i)^{(k+1)} = \frac{\sqrt{1 + 4\rho \left(\boldsymbol{\Sigma}(i, i)^{(k)} + \left| \boldsymbol{\mu}(i)^{(k)} \right|^2 \right)} - 1}{2\rho}, \quad (13)$$

$$\lambda^{(k+1)} = \frac{2a - 2 + M_T N_T}{2b + \Delta y}, \quad (14)$$

where $i \in \{1, \dots, L\}$, $\Delta y = (\lambda^{(k)})^{-1} \sum_{i=1}^L 1 - \frac{\boldsymbol{\Sigma}(i, i)^{(k)}}{\alpha(i)^{(k)}} + \left\| \mathbf{y} - \boldsymbol{\Phi}(\mathbf{S}^{(k)}, \boldsymbol{\kappa}^{(k)}, \boldsymbol{\nu}^{(k)}) \boldsymbol{\mu}^{(k)} \right\|^2$, $\lambda > 0$, $a > 0$, $b > 0$. For updates of off-grid parameters, please refer to [8], [9]. The iteration will stop if $\frac{\|\boldsymbol{\alpha}^{(k+1)} - \boldsymbol{\alpha}^{(k)}\|}{\|\boldsymbol{\alpha}^{(k)}\|} < \delta$ or the maximum number of iterations K is reached, where δ is a tolerance.

B. Fission Process

Different from the 1D angular grid in [11], we consider 2D grid in this paper. If the fission is carried as that of [11], the grid \mathbf{S} will become $\{\{\bar{\mathbf{k}}_\nu, \mathbf{k}_{\text{new}}\} \times \{\bar{\mathbf{l}}_\tau, \mathbf{l}_{\text{new}}\}\}$, which is equivalent to inserting several rows and several columns of grid points into the initial DD grid. Obviously, except for a few points which are close to the channel responses, many grid points may be irrelevant to the channel responses, thus leading to extra calculation cost.

As shown in Fig. 1, we add grid points around the true channel responses. Three fission conditions are used like that of [11] to decide whether a fission of a grid should happen. For all the grid points $i \in \{1, \dots, L\}$, grid point with $|\boldsymbol{\mu}(i)| > \varepsilon \sqrt{\lambda^{-1}}$ needs fission for a higher representation accuracy, where ε is a weight related to miss detection or false-alarm probabilities. Only the grid point corresponding to the local maximum value can be selected for fission to prevent too many fission grid points corresponding to a large channel response. Finally, $\tilde{r}_\nu > r_{\min}$ or $\tilde{r}_\tau > r_{\min}$ should be satisfied.

When these three conditions are met, the fission needs to be carried out simultaneously in both the delay and the Doppler dimensions based on estimated off-grid parameters. Assume that the chosen grid point χ is (k_i, l_i) , and the estimation result of corresponding off-grid parameter is (\hat{k}_i, \hat{l}_i) . The off-grid parameters can indicate that the channel response is in a certain direction of grid point χ . Thus two new grid points $\chi_k = (k_{\text{fission}}, l_i)$ and $\chi_l = (k_i, l_{\text{fission}})$ will be generated after fission, where k_{fission} and l_{fission} are

$$k_{\text{fission}} = \begin{cases} k_i + \frac{1}{2} \tilde{r}_{\nu+}, \hat{k}_i > 0 \\ k_i - \frac{1}{2} \tilde{r}_{\nu-}, \hat{k}_i < 0 \end{cases}, \quad (15)$$

$$l_{\text{fission}} = \begin{cases} l_i + \frac{1}{2} \tilde{r}_{\tau+}, \hat{l}_i > 0 \\ l_i - \frac{1}{2} \tilde{r}_{\tau-}, \hat{l}_i < 0 \end{cases}. \quad (16)$$

If a grid point fissions, it will be necessary to add 1 or 2 elements corresponding to the grid point for κ , ι and h in Eq. (9). Owing to the OGSBI method we used for estimation, the fission of prior of h , i.e., α , is needed. Assume that the α value corresponding to grid points χ , χ_f , χ_k and χ_l are α_χ , α_{χ_f} , α_{χ_k} and α_{χ_l} , respectively, where χ_f is the grid point χ version after fission. To keep the prior $p(\alpha)$ unchanged after fission, $p(\alpha)$ should be a constant [11]. Therefore,

$$\alpha_{\chi_f} = \alpha_{\chi_k} = \alpha_{\chi_l} = \frac{1}{3}\alpha_\chi, \quad (17)$$

where α_χ is estimated by Eq. (13).

Note that too many fissions may reduce computational efficiency or even deteriorate the CE performance because of the high correlation between the grid points. Therefore, beyond the fission conditions, the fission stop criterion is introduced that no more fission will happen if the maximum number of iterations K_f for fission process is reached. K_f may be larger than $\log_2(1/r_{\min})$, which refers to the average iteration number for reaching a satisfying resolution.

C. Adjustment Process

If a dense grid is used in the uniform grid model, the Taylor approximation errors would be small. For the GE scheme, after the fission process, a non-uniform and locally fine DD grid with better CE performance will be obtained. However, the off-grid gap of the current DD grid cannot be ignored. This is because the CE performance may be affected by the fractional channel. Therefore, Adjustment process is proposed in this letter to further reduce the approximation errors.

It is not necessary that all grid points are adjusted. For all the grid points $i \in \{1, \dots, L\}$, if $|\mu(i)| > \varepsilon\sqrt{\lambda^{-1}}$, then we adjust the grid. The adjustment of a DD grid is given by

$$\begin{cases} k_i \leftarrow k_i + \kappa_i \\ l_i \leftarrow l_i + \iota_i \end{cases} \quad (18)$$

When the adjustment process is completed, the off-grid parameters k_i and l_i will be reduced, leading to decreased Taylor approximation errors $o(\kappa_i)$ and $o(\iota_i)$. Since κ_i and ι_i are integrated into the delay and Doppler values of a grid point, $\Phi(\mathcal{S}, \kappa, \iota)$ should be recalculated but this incurs a higher complexity. In this case, an adjustment stop criterion is required. The iteration of the adjustment process will stop if $\|[\kappa^{(k)}, \iota^{(k)}]\| < \delta_a$ or the maximum number of iterations K_a for the adjustment process is reached, where δ_a is a tolerance.

In addition to the above framework, another framework is that fission and adjustment are performed in the same iteration or execute iteratively. This letter does not use these two frameworks because the initial grid in GE scheme is coarse and it is quite possible that multiple paths are in a same grid interval. Therefore, a grid point in early iterations may represent a response of multiple paths and undergo fission in future iterations. After fission, the grid distribution will be reset thus the computational workload of adjustment of these grid point is wasted. In our framework, the paths are assumed to be distinguishable and represented by different grid points after the stop of fission process. Then the adjustment process could decrease the Taylor approximation errors for each path.

TABLE I: Simulation parameters

Parameter	Value
DD grid size	$N = M = 32$
Carrier frequency	$f_c = 4$ GHz
Subcarrier spacing	$\Delta f = 15$ kHz
Number of channel paths	$P = 5$
The channel coefficient	$h_i \sim \mathcal{CN}\left(0, \frac{\exp(-0.1l\tau_i)}{\sum_i \exp(-0.1l\tau_i)}\right)$
Maximum delay	$\tau_{\max} = 8.3 \times 10^{-6}$ s
Maximum relative velocity	500 km/h

IV. COMPLEXITY ANALYSIS AND SIMULATION RESULTS

The proposed GE scheme is compared with the on-grid, 1D off-grid [9], 2D off-grid-U [9] and 2D off-grid [10]. Table I presents the simulation parameters. The pilot pattern and guard interval in [9] are used. The power of single pilot is 30 dB higher than the average power of data. In GE, $a = b = 10^{-4}$, $\rho = 10^{-2}$, $\delta = 10^{-3}$, $\delta_a = 10^{-1}$, $K_f = 5$, $K_a = 50$, $K = 200$. If not stated otherwise, the resolution of the uniform grid scheme is $r = r_\nu = r_\tau = \frac{1}{4}$. The initial resolution in the GE scheme is 1, and $r_{\min} = \frac{1}{4}$.

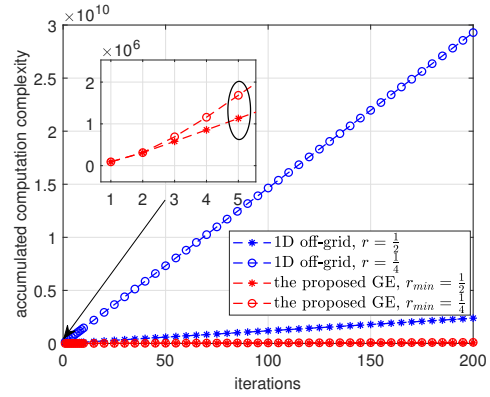


Fig. 3: Accumulated computational complexity comparison.

The accumulated complexity comparison at different numbers of iterations is shown in Fig. 3. The main complexity of the GE scheme is the same as that of the scheme in [9]. For the GE scheme, the initial uniform DD grid is coarse and L is small. The fission process stops at the iteration number in the black circle and the final L is still small. This leads to significantly reduced complexity of the proposed GE scheme.

In Fig. 4, the convergence of the different schemes is compared. Compared with the other two uniform grid schemes, the GE scheme can converge faster. Combining Fig. 3 and 4, it can be seen that with the same number of iterations, GE scheme has lower computational complexity and faster convergence while ensuring the NMSE performance.

The NMSE comparison between different methods is given in Fig. 5. The expected resolution is $r = \frac{1}{4}$, thus the number of grid points for all the uniform grid schemes is 561. The initial number of grid points for the GE scheme is 45 and finally reaches around 110. For a fair comparison, uniform grid schemes marked with '+' are also presented and this indicates that the number of grid points, in these uniform grid schemes, is approximately equal to 110 as that in the GE scheme. It can

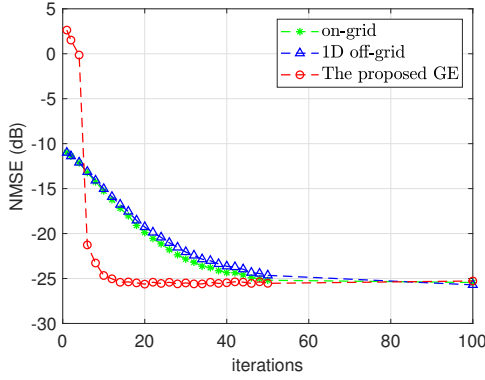


Fig. 4: Convergence comparison between the GE scheme and the uniform grid scheme.

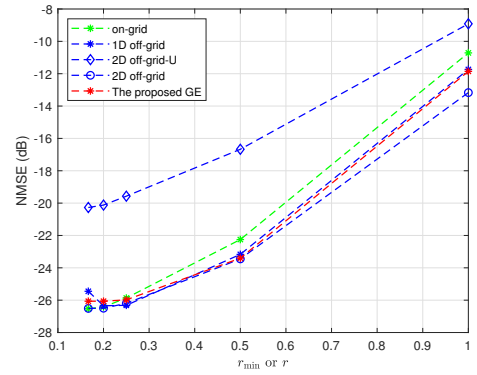


Fig. 6: NMSE under different r_{\min} or r .

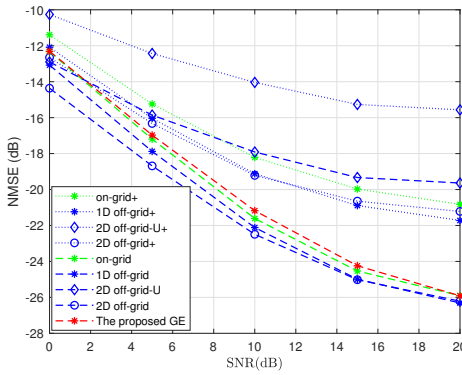


Fig. 5: NMSE comparison

be seen that the performance of the proposed GE with much fewer grid points can be very close to that of the uniform grid scheme with $r = \frac{1}{4}$, and significantly outperforms the uniform grid schemes when the number of grid points is comparable.

Note that the performance of GE method relies on the choice of r_{\min} . Therefore, to demonstrate the influence of r_{\min} , Fig. 6 shows the comparison of NMSE for different r_{\min} and r at SNR = 20 dB. It is found that the NMSE of GE is smaller if we take a smaller r_{\min} . And except for the 2D off-grid scheme that considers the tandem off-grid distortion to make the model error smaller, the GE scheme has the best performance when r is large. Considering all simulation results together, the GE scheme achieves a good trade-off between CE performance and complexity.

V. CONCLUSIONS

An improved GE method for doubly fractional CE has been proposed in this letter. SBL and grid refinement are combined to adaptively evolve uniform DD grid into non-uniform grid. The key idea of the proposed GE method consists of three processes, i.e., learning, fission and adjustment process. The learning process estimates the channel response at the current grid points, the fission process adds new grid points, and the adjustment process employs off-grid parameters to adjust the DD grid. The learning process iterates alternately with the other two processes, respectively. Simulation results show

that the proposed GE scheme achieves an excellent tradeoff between the CE performance and complexity.

REFERENCES

- [1] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," *IEEE Wireless Communications and Networking Conference (WCNC)*, San Francisco, CA, USA, 19-22, March, 2017, pp. 1-6.
- [2] P. Raviteja, K. T. Phan and Y. Hong, "Embedded pilot-aided channel estimation for OTFS in delay-Doppler channels," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4906-4917, May 2019.
- [3] P. Raviteja, K. T. Phan, Y. Hong and E. Viterbo, "Interference cancellation and iterative detection for orthogonal time frequency space modulation," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6501-6515, Oct. 2018.
- [4] G. Zhang, X. -Q. Jiang, H. Hai, L. Xu and S. Mumtaz, "Intelligent reflecting surfaces based offset index modulation for MIMO systems," *IEEE Transactions on Vehicular Technology* (Early Access), 2024.
- [5] Y. Yang, H. Hai, X. -Q. Jiang, Y. Wu and S. Mumtaz, "Low-Complexity Detectors for Space-Time Block Coded Differential Spatial Modulation," *IEEE Transactions on Vehicular Technology* (Early Access), 2024.
- [6] W. Shen, L. Dai, J. An, P. Fan, and R. W. Heath, "Channel estimation for orthogonal time frequency space (OTFS) massive MIMO," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4204-4217, Jul. 2019.
- [7] L. Zhao, W. Gao, and W. Guo, "Sparse bayesian learning of delay-doppler channel for OTFS system," *IEEE Communications Letters*, vol. 24, no. 12, pp. 2766-2769, 2020.
- [8] Z. Yang, L. Xie and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 38-43, Jan.1, 2013.
- [9] Z. Wei, W. Yuan, S. Li, J. Yuan and D. W. K. Ng, "Off-grid channel estimation with sparse Bayesian learning for OTFS systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 74077426, Sept. 2022.
- [10] Q. Wang, Y. Liang, Z. Zhang, P. Fan, "2D off-grid decomposition and SBL combination for OTFS channel estimation," *IEEE Trans on Wireless Communications*, vol.22, No.5, pp. 3084-3098, May 2023.
- [11] Q. Wang, Z. Zhao, Z. Chen and Z. Nie, "Grid evolution method for DOA estimation," *IEEE Transactions on Signal Processing*, vol. 66, no. 9, pp. 2374-2383, 1 May1, 2018.
- [12] Y. Shan, F. Wang, Y. Hao, J. Yuan, J. Hua and Y. Xin, "Off-grid channel estimation using grid evolution for OTFS systems," *IEEE Transactions on Wireless Communications* (Early Access), 2024.