Soft Information Aided Diagonal Kalman Filter for Joint Channel Estimation and Detection in Massive MIMO Systems

Xuanxiang Hu, Zheng Wang, Yongming Huang
School of Information Science and Engineering
Southeast University, Nanjing, China
Email: huxx@seu.edu.cn, wznuaa@gmail.com

Zhen Gao
School of Information and Electronics
Beijing Institute of Technology, Beijing, China
Email: gaozhen16@bit.edu.cn

***Abstract*—This paper proposes a soft-information-aided diagonal Kalman filter (SIA-DKF) for joint channel estimation and symbol detection (JED) in massive multiple-input multiple-output (MIMO) systems. Specifically, building upon the Kalman filter (KF) framework for JED, the proposed SIA-DKF incorporates the symbol means and variances obtained from variational inference (VI) into the KF, improving its ability to track channels and mitigating mismatches in the update step. Furthermore, SIA-DKF not only exploits favorable propagation in massive MIMO to simplify the VI for symbol detection, but also employs a diagonal approximation of the covariance matrix at each channel update step, thereby significantly reducing the computational complexity. Simulations confirm that SIA-DKF improves the bit error rate (BER) with lower complexity.**

***Index Terms*—Soft-information-aided diagonal Kalman filter, joint channel estimation and symbol detection, massive MIMO, favorable propagation.**

I. Introduction

As the demand for spectrum and computational efficiency increases in B5G and 6G systems [1], joint channel estimation and symbol detection (JED) is drawing attention for its iterative coupling of estimation and detection. Specifically, the decisions from the detector can refine the channel estimate, which is then fed back to the detector for detection performance enhancement. Consequently, extensive JED studies have appeared, such as iterative receivers based on the turbo principle [2] [3] and Bayesian inference methods employing expectation propagation (EP) [4] or variational inference (VI) [5].

However, most JED studies treat time slots independently without considering the temporal priors across slots, which incur repeated channel re-estimation and avoidable computational costs. To address this, some studies incorporate temporal priors into the channel model, thereby preserving estimation accuracy while avoiding redundant computations. For example, [6] proposed a method that incorporates temporal correlation priors into EP channel modeling, enabling channel tracking in single-input single-output (SISO) systems through forward prediction and backward smoothing. Meanwhile, a JED method based on QR detection and Kalman filter (KF) was proposed [7], where the channel is modeled as an autoregressive (AR) polynomial. Compared with traditional JED methods, KF not only provides the channel priors for detection but also utilizes the detection to recursively update the polynomial coefficients.

In this paper, we propose a low-complexity JED method based on the KF framework for massive MIMO systems, referred to as the soft-information-aided diagonal Kalman filter (SIA-DKF). On the one hand, we integrate symbol means and variances obtained by VI into the update step, thereby enhancing the reliability of the Kalman gain $\mathbf{K}_t$ compared with traditional KF-based JED. On the other hand, we not only exploit favorable propagation in massive MIMO to simplify the VI by decoupling it across coordinates, but also apply a diagonal approximation to the covariance matrix at each update, thereby achieving an advanced JED algorithm.

II. System Model

Consider an uplink time-division duplexing (TDD) multiuser MIMO system, in which the base station (BS) is equipped with N_r receive antennas and serves N_t single-antenna users transmit simultaneously. Each user transmits over $T = T_p + T_d$ time slots: T_p pilot slots followed by T_d data slots. The received signals at slot t can be expressed as

$$\mathbf{y}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{w}_t = (\mathbf{I}_{N_r} \otimes \mathbf{x}_t^T)\mathbf{h}_t + \mathbf{w}_t, \qquad (1)$$

where $\mathbf{H}_t \in \mathbb{C}^{N_r \times N_t}$ is the CSI matrix, $\mathbf{x}_t \in \mathbb{C}^{N_t}$ and $\mathbf{w}_t \sim \mathcal{CN}(0, \sigma_w^2 \mathbf{I}_{N_r})$. Here, the CSI matrix is vectorized as $\mathbf{h}_t \triangleq vec(\mathbf{H}_t)$ and the $\otimes$ denotes the Kronecker

Corresponding author: Zheng Wang (email: wznuaa@gmail.com)

product. Without loss of generality, a first-order autoregressive (AR(1)) model is adopted to characterize the channel's temporal evolution, as follows [8]:

$$\mathbf{h}_t = \alpha \mathbf{h}_{t-1} + \mathbf{v}_t, \mathbf{v}_t \sim \mathcal{CN}(0, \mathbf{Q}). \tag{2}$$

Here $\alpha = \mathcal{J}_0(2\pi f_d)$ is the temporal correlation coefficient, where $\mathcal{J}_0$ denotes the zeroth-order Bessel function and f_d denotes the normalized Doppler. The process noise $\mathbf{v}_t$ has covariance $\mathbf{Q} = (1-\alpha^2)\mathbf{R}_h$, and $\mathbf{R}_h = \mathbb{E}\{\mathbf{h}_t \mathbf{h}_t^H\}$ denotes the channel correlation matrix. Building upon this time-varying channel model, KF-based JED schemes adopt a "prediction–symbol detection–channel update" framework [7], as follows.

Prediction: Given the channel state $\hat{\mathbf{h}}_{t-1|t-1}$ and the posterior covariance matrix $\mathbf{P}_{t-1|t-1}$,

$$\hat{\mathbf{h}}_{t|t-1} = \mathbf{F}\hat{\mathbf{h}}_{t-1|t-1} = \alpha \hat{\mathbf{h}}_{t-1|t-1}, \tag{3}$$

$$\mathbf{P}_{t|t-1} = \mathbf{F}\mathbf{P}_{t-1|t-1}\mathbf{F}^H + \mathbf{Q} = \alpha^2 \mathbf{P}_{t-1|t-1} + \mathbf{Q}, \tag{4}$$

where $\mathbf{F} \triangleq \alpha \mathbf{I}_{N_r N_t}$ is the state transition matrix, and $\mathbf{h}_{t|t-1}$, $\mathbf{P}_{t|t-1}$ denote the channel state prediction and the prior covariance matrix, respectively.

Symbol detection: Given the channel state prediction $\hat{\mathbf{H}}_{t|t-1} = \text{reshape}(\hat{\mathbf{h}}_{t|t-1}, N_r, N_t)$, linear detection methods such as zero forcing (ZF) or minimum mean square error (MMSE) is typically employed,

$$\tilde{\mathbf{x}}_t^{\text{ZF}} = (\hat{\mathbf{H}}_{t|t-1}^H \hat{\mathbf{H}}_{t|t-1})^{-1} \hat{\mathbf{H}}_{t|t-1} \mathbf{y}_t, \tag{5}$$

$$\tilde{\mathbf{x}}_t^{\text{MMSE}} = (\hat{\mathbf{H}}_{t|t-1}^H \hat{\mathbf{H}}_{t|t-1} + \frac{\sigma_w^2}{\sigma_x^2})^{-1} \hat{\mathbf{H}}_{t|t-1}^H \mathbf{y}_t. \tag{6}$$

Next, the hard decision $\hat{\mathbf{x}}_t$ is obtained by direct quantization $\hat{\mathbf{x}}_t^{\text{ZF}} = \lceil \tilde{\mathbf{x}}_t^{\text{ZF}} \rfloor_{\mathcal{Q}} \in \mathcal{O}^{N_t}$ or $\hat{\mathbf{x}}_t^{\text{MMSE}} = \lceil \tilde{\mathbf{x}}_t^{\text{MMSE}} \rfloor_{\mathcal{Q}} \in \mathcal{O}^{N_t}$.

Channel update: Given the detection output $\hat{\mathbf{x}}_t$ and the measurement matrix $\mathbf{A}_t = \mathbf{I}_{N_r} \otimes \hat{\mathbf{x}}_t^T$, the channel state $\hat{\mathbf{h}}_{t|t}$ and the posterior covariance $\mathbf{P}_{t|t}$ are updated as follows:

$$\hat{\mathbf{h}}_{t|t} = \hat{\mathbf{h}}_{t|t-1} + \mathbf{K}_t \mathbf{u}_t, \quad \mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t \mathbf{A}_t)\mathbf{P}_{t|t-1}, \tag{7}$$

where $\mathbf{u}_t = \mathbf{y}_t - \mathbf{A}_t \hat{\mathbf{h}}_{t|t-1}$ and the Kalman gain $\mathbf{K}_t$ can be computed as

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{A}_t^H \mathbf{S}_t^{-1}. \tag{8}$$

Here $\mathbf{K}_t$ represents the optimal linear weight between the prior of AR(1) and the observation equation (1) in Kalman filter, which is determined by the covariance

$$\mathbf{S}_t = \mathbf{A}_t \mathbf{P}_{t|t-1} \mathbf{A}_t^H + \sigma_w^2 \mathbf{I}_{N_r}. \tag{9}$$

To be more specific, the covariance $\mathbf{S}_t$ reflects the uncertainty of the update process. However, the traditional KF-based JED scheme fails to account for the detection errors inherent to hard-decision feedback when calculating $\mathbf{S}_t$, thereby leading to an unreliable Kalman gain $\mathbf{K}_t$ and impairing overall tracking performance.

Algorithm 1: Conventional KF-based JED

Input: $\boldsymbol{y}(t)$, $\mathbf{x}_p$, T_p, T_d

Initialize: $\hat{\mathrm{h}}(0) = \mathbf{0}$, $\mathbf{P}(0) = \mathbf{I}$;

for $t = 1$ *to* T **do**

- **Prediction:** given $\hat{\mathbf{h}}_{t-1|t-1}$
- update $\hat{\mathbf{h}}_{t|t-1}$ and $\mathbf{P}_{t|t-1}$ via (3) and (4)
- **Symbol detection:**
- if $t < T_p$: let $\mathbf{x}_t = \mathbf{x}_p$
- else: compute $\hat{\mathbf{x}}_t$ via ZF (5) or MMSE (6)
- **Channel update:** $\mathbf{A}_t = \mathbf{I}_{N_r} \otimes \hat{\mathbf{x}}_t$
- calculate $\mathbf{K}_t$ and $\mathbf{S}_t$ via (8) and (9)
- update $\hat{\mathbf{h}}_{t|t}$ and $\mathbf{P}_{t|t}$ via (7)

Output: $\hat{\mathbf{h}}_{t|t}, \hat{\mathbf{x}}_t$

III. Joint Channel Estimation and Detection Based on Soft Information

To improve the reliability of the Kalman gain $\mathbf{K}_t$, we now propose a soft-information-aided KF-based JED. In particular, unlike the conventional KF-based JED, we adopt the maximum a posteriori (MAP) criterion for symbol detection to obtain $\hat{\mathbf{x}}_t$, i.e.,

$$\hat{\mathbf{x}}_t = \arg \max_{\mathbf{x}_t \in \mathcal{O}^{N_t}} p(\mathbf{x}_t | \mathbf{y}_t). \tag{10}$$

Subsequently, we leverage the posterior distribution $p(\mathbf{x}_t|\mathbf{y}_t)$ to compute the soft information, which includes the symbol mean $\boldsymbol{\mu}_t$ and variance $\boldsymbol{\sigma}_t^2$. This soft information is then fed back into the channel update, thereby improving the reliability of the Kalman gain $\mathbf{K}_t$ and the overall channel estimation performance.

Specifically, our soft-information-aided KF-based JED approach follows a "prediction-detection-soft information computation-update" framework, as follows.

Prediction: Given the channel state $\hat{\mathbf{h}}_{t-1|t-1}$ and covariance matrix $\mathbf{P}_{t-1|t-1}$ at $t-1$ slot, we can predict the channel state $\hat{\mathbf{h}}_{t|t-1}$ and the prior covariance matrix $\mathbf{P}_{t|t-1}$ using equation (3)-(4) under AR(1) priors.

Symbol detection: Using the MAP criterion, we detect $\hat{\mathbf{x}}_t$ from the posterior distribution $p(\mathbf{x}_t|\mathbf{y}_t)$. However, the exact posterior distribution is intractable to compute in massive MIMO systems. To overcome the issue, we adopt VI to approximate the true posterior $p(\mathbf{x}_t|\mathbf{y}_t)$ with $q(\mathbf{x}_t)$ by maximizing the evidence lower bound (ELBO), i.e.

$$\mathcal{L}(q) = \int q(\mathbf{x}_t) \ln \left\{ \frac{p(\mathbf{x}_t, \mathbf{y}_t)}{q(\mathbf{x}_t)} \right\} d\mathbf{x}_t. \tag{11}$$

Moreover, to avoid optimizing over the entire $q(\mathbf{x}_t)$ distribution, we adopt the *mean-field theory* factorization:

$$q(\mathbf{x}_t) = \prod_{i=1}^{N_t} q_i(x_{t,i}). \tag{12}$$

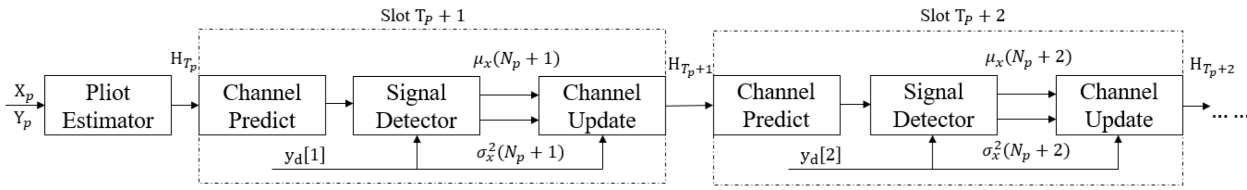

Fig. 1. System diagram of the proposed joint channel estimation and signal detection method for massive MIMO.

Then, substituting the factorization into (11), the optimal solution $q_i^\star(x_{t,i})$ for each transmit antenna at slot t is iteratively updated using coordinate ascent variational inference (CAVI)

$$\ln q_i^\star(x_{t,i}) = \mathbb{E}_{q(\mathbf{x}_t \backslash x_{t,i})}[\ln p(\mathbf{x}_t, \mathbf{y}_t)] + \text{const}, \quad (13)$$

where the joint distribution can be expressed as

$$\begin{aligned} \ln p(\mathbf{x}_t, \mathbf{y}_t) &= \ln p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t) \\ &= -\tfrac{\|\mathbf{y}_t - \mathbf{H}_{t|t-1}\mathbf{x}_t\|^2}{2\sigma_n^2} + \sum \ln\left\{\tfrac{\mathbb{I}_{x_i\in\mathcal{O}}}{\sqrt{M}}\right\}. \end{aligned} \quad (14)$$

Here $\mathbb{I}_{x_i\in\mathcal{O}}$ is the indicator function and M represents the index of the quadrature amplitude modulation (QAM). At this point, the i-th detection output can be obtained by $\hat{x}_{t,i} = \arg\max_{x_{t,i}\in\mathcal{O}} q_i^\star(x_{t,i})$.

Soft-information computation: Given the normalized posterior

$$q_i(x_{t,i}) = \frac{q_i^\star(x_{t,i})}{\sum_{z_{t,i}\in\mathcal{O}} q_i(z_{t,i})}, \quad x_{t,i}\in\mathcal{O}, \quad (15)$$

the soft information for stream i is computed by

$$\mu_{t,i} \triangleq \mathbb{E}_q[x_{t,i}] = \sum_{x_{t,i}\in\mathcal{O}} x_{t,i} q_i(x_{t,i}), \quad (16)$$

$$\sigma_{t,i}^2 \triangleq Var_q[x_{t,i}] = \sum_{x_{t,i}\in\mathcal{O}} |x_{t,i}|^2 q_i(x_{t,i}) - \mu_{t,i}^2. \quad (17)$$

Collecting (16)-(17) across streams yields the symbol means $\boldsymbol{\mu}_t = [\mu_{t,1}, ..., \mu_{t,N_t}]^T$ and variances $\boldsymbol{\sigma}_t^2 = [\sigma_{t,1}^2, ..., \sigma_{t,N_t}^2]$.

Channel update: Given the symbol means $\boldsymbol{\mu}_t$ and variances $\boldsymbol{\sigma}_t^2$, we can model $\mathbf{x}_t - \boldsymbol{\mu}_t$ as having zero-mean with covariance $\mathbf{C}_x(t) \triangleq \text{diag}(\boldsymbol{\sigma}_t^2)$ to represent the symbol-wise detection uncertainly, and incorporate it into the channel update by defining the equivalent noise covariance

$$\mathbf{R}_t \triangleq \hat{\mathbf{H}}_{t|t-1}\mathbf{C}_x(t)\hat{\mathbf{H}}_{t|t-1}^H + \sigma_w^2\mathbf{I}_{N_r}. \quad (18)$$

Then we can compute covariance $\mathbf{S}_t$ with $\mathbf{A}_t = \mathbb{I}_{N_r}\otimes\boldsymbol{\mu}_t$ and $\mathbf{R}_t$ by

$$\mathbf{S}_t = \mathbf{A}_t\mathbf{P}_{t|t-1}\mathbf{A}_t^H + \mathbf{R}_t. \quad (19)$$

Using $\mathbf{S}_t$ in (19), the Kalman gain $\mathbf{K}_t$ can be computed by $\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{A}_t^H\mathbf{S}_t^{-1}$. Finally, the channel state $\hat{\mathbf{h}}_{t|t}$ and the posterior covariance $\mathbf{P}_{t|t}$ are updated following the same formulation in (7). For a better understanding, our proposed JED framework is illustrated in Fig.1.

IV. Complexity Reduction of Soft-Information-Aided KF-based JED

With respect to the proposed soft-information-aided KF-based JED framework, we further reduce complexity in both symbol detection and channel update.

A. Efficient Detection under Favorable Propagation

In the symbol detection step of the framework, the calculation of likelihood in (14), i.e.,

$$\ln p(\mathbf{y}_t|\mathbf{x}_t) = -\tfrac{\|\mathbf{y}_t - \mathbf{H}_{t|t-1}\mathbf{x}_t\|^2}{2\sigma_n^2} + \text{const}, \quad (20)$$

which is computationally expensive due to repeated matrix–vector products with $\mathbf{H}_{t|t-1}$. To alleviate this burden, we rewrite the system model in (1) as [9]:

$$\mathbf{r}_t = \mathbf{J}_t\mathbf{x}_t + \tilde{\mathbf{w}}_t, \quad (21)$$

where $\mathbf{r}_t = \frac{\mathbf{H}_t^H\mathbf{y}_t}{N_r}$, $\mathbf{J}_t = \frac{\mathbf{H}_t^H\mathbf{H}_t}{N_r}$, and $\tilde{\mathbf{w}}_t = \frac{\mathbf{H}_t^H\mathbf{w}_t}{N_r}$. For each coordinate i, it follows that

$$r_{t,i} = J_{t,ii}x_{t,i} + g_{t,i}, \quad (22)$$

with $g_{t,i} \triangleq \sum_{k\neq i} J_{t,ik}x_{t,k} + \tilde{w}_{t,i}$. On the other hand, under the favorable propagation in massive MIMO, the columns of $\mathbf{H}_t$ are nearly orthogonal, such that [10]

$$\lim_{N_r\to\infty} \frac{\mathbf{H}_t^H\mathbf{H}_t}{N_r} \approx \mathbf{I}_{N_r}, \quad (23)$$

which implies that the $r_{t,i}$ is conditionally independent of $\mathbf{x}_{t,\backslash i}$ given $x_{t,i}$. Next, following the mean-field factorization in (12), the residual $g_{t,i}$ can be modeled as a Gaussian distribution with $(\mu_{g_{t,i}}, \sigma_{g_{t,i}}^2)$. Then, the likelihood $p(r_{t,i}|\mathbf{x}_t)$ can be simplified to $p(r_{t,i}|x_{t,i})$, i.e.

$$\ln p(r_{t,i}|x_{t,i}) \approx -\frac{\left(r_{t,i}|J_{t,ii}x_{t,i} + \mu_{g_{t,i}}\right)^2}{2\sigma_{g_{t,i}}^2} + \text{const}, \quad (24)$$

which reduces the computation of likelihood in (20) from $\mathcal{O}(N_rN_t^2)$ to $\mathcal{O}(N_t^2)$ for each i. By doing this, the joint distribution in (14) can be rewritten as

$$\ln p(\mathbf{x}_t, \mathbf{r}_t) = \sum_{i=1}^{N_t} \left(\ln p(r_{t,i}|x_{t,i}) + \ln p(x_{t,i})\right). \quad (25)$$

Next, taking the expectation in (13), we can easily get the optimal posterior by

$$\ln q_i^\star(x_{t,i}) = \ln \mathbb{I}_{x_{t,i}\in\mathcal{O}} - \frac{\left(r_{t,i} - J_{t,ii}x_{t,i} - \mu_{g_{t,i}}\right)^2}{2\,\sigma_{g_{t,i}}^2} + \text{c}, \quad (26)$$

and subsequently obtain the detection by $\hat{x}_{t,i} = \arg\max_{x_{t,i}\in\mathcal{O}} q_i^\star(x_{t,i})$.

B. Low-Complexity Channel Update via Diagonal Approximation

As for the channel update step, the computations of $\mathbf{S}_t$ in (19) and $\mathbf{K}_t$ in (8) still may incur substantial computational cost. To further reduce the computational complexity, we adopt a diagonal approximation to the covariance $\mathbf{S}_t$. Specifically, we introduce the information form of covariance matrix $\mathbf{P}_{t|t}$ update in [11], i.e.,

$$\mathbf{P}_{t|t}^{-1} = \frac{1}{\alpha}\mathbf{P}_{t-1|t-1}^{-1} + \mathbf{A}_t^H \mathbf{R}_t^{-1}\mathbf{A}_t. \quad (27)$$

For a better understanding of the diagonal approximation, we first provide the diagonally dominance of the equivalent noise covariance matrix $\mathbf{R}_t$ in Lemma 1, i.e,

Lemma 1: According to the favorable propagation in massive MIMO, the equivalent noise covariance matrix $\mathbf{R}_t$ in (18) is diagonally dominant by

$$|[\mathbf{R}_t]_{ii}| \gg \sum_{j\neq i} |[\mathbf{R}_t]_{ij}|. \quad (28)$$

Proof: Consider the $\mathbf{R}_t \triangleq \hat{\mathbf{H}}_{t|t-1}\mathbf{C}_x(t)\hat{\mathbf{H}}_{t|t-1}^H + \sigma_w^2 \mathbf{I}_{N_r}$, which can be decomposed into

$$[\mathbf{R}_t]_{ij} = \begin{cases} \sum\limits_{k=1}^{N_t} |[\hat{\mathbf{H}}_{t|t-1}]_{ik}|^2 \sigma_{t,k}^2 + \sigma_w^2, & i = j, \\ \sum\limits_{k=1}^{N_t} [\hat{\mathbf{H}}_{t|t-1}]_{ik} [\hat{\mathbf{H}}_{t|t-1}]_{jk}^* \sigma_{t,k}^2, & i \neq j. \end{cases} \quad (29)$$

Under the favorable propagation condition of massive MIMO in (23), i.e.

$$|[\hat{\mathbf{H}}_{t|t-1}]_{ik}|^2 \approx 1, \quad [\hat{\mathbf{H}}_{t|t-1}]_{ik}[\hat{\mathbf{H}}_{t|t-1}]_{jk}^* \approx 0, \quad (30)$$

the diagonal elements of $\mathbf{R}_t$ are approximately

$$[\mathbf{R}_t]_{ii} \approx \sum_{k=1}^{N_t} \sigma_{t,k}^2 + \sigma_w^2, \quad (31)$$

while the off-diagonal elements vanish asymptotically as $N_r \to \infty$, completing the proof. ■

Then, under the i.i.d. symbol assumption [12], i.e., $\mathbb{E}[x_{t,i}] = 0$ and $\mathbb{E}[|x_{t,i}|^2] = \rho$ for all i, the symbol covariance is $\mathbf{R}_x = \mathbb{E}[\mathbf{x}_t\mathbf{x}_t^H] = \rho \mathbf{I}_{N_t}$, where ρ is the average per-stream symbol power. Since $\mathbf{A}_t = \mathbf{I}_{N_r} \otimes \mathbf{x}_t$, the measurement matrix inherits this diagonal structure, i.e.,

$$\mathbb{E}[\mathbf{A}_t^H \mathbf{A}_t] = \mathbf{R}_x \otimes \mathbf{I}_{N_r}. \quad (32)$$

From the diagonal dominance of $\mathbf{R}_t$ and $\mathbf{A}_t$ as shown in Lemma 1 and the equation (32), we can derive that $\mathbf{A}_t^H \mathbf{R}_t^{-1}\mathbf{A}_t$ in (27) is diagonally dominant. Motivated by the derivation, we can conclude that $\mathbf{P}_{t|t}^{-1}$ is diagonally dominant with the recurrence relation of (27), and approximate $\overline{\mathbf{P}}_{t|t}$ by its diagonal part

$$\overline{\mathbf{P}}_{t|t} = \mathrm{diag}(\mathrm{diag}(\mathbf{P}_{t|t})). \quad (33)$$

Algorithm 2: Soft-information-aided diagonalized Kalman Filter (SIA-DKF)

Input: $\boldsymbol{y}(t)$, $\boldsymbol{x}_p(t)$, T_p, T_d

Initialize: $\widehat{\mathrm{h}}(0) = \mathbf{0}$, $\mathbf{P}(0) = \mathbf{I}$;

for $t = 1$ *to* T **do**

  Prediction:

  update $\hat{\mathbf{h}}_{t|t-1}$ and $\overline{\mathbf{P}}_{t|t-1}$ via (3)-(4)

  Symbol detection:

  if $t < T_p$: using $\mathbf{x}_p$ to update

  else: rewrite system as (22)

  iteratively update $q(x_{t,i})$ via (26)

  Soft information computation:

  compute $\boldsymbol{\mu}_x(t)$ and $\mathbf{C}_x(t)$ via (16)-(17)

  Channel update:

  calculate $\mathbf{R}(t)$ via (18)

  calculate $\mathbf{S}(t)$ and $\mathbf{K}_t$ via (34) and (8)

  update $\hat{\mathbf{h}}_{t|t}$, $\overline{\mathbf{P}}_{t|t}$ via (7) and (33)

Output: $\hat{\mathbf{h}}_{t|t}, \hat{\mathbf{x}}_t$

Next, substituting (33) into the computation of $\mathbf{S}_t$, we can obtain the diagonal approximation of the $\mathbf{S}_t$ by

$$\begin{aligned} \mathbf{S}_t &\approx \mathbf{A}_t \overline{\mathbf{P}}_{t|t-1}\mathbf{A}_t^H + \mathrm{diag}\big(\mathrm{diag}(\mathbf{R}_t)\big) \\ &= \mathrm{diag}\left(s_{1,t}, \ldots, s_{N_r,t}\right). \end{aligned} \quad (34)$$

Here $s_{m,t} = \sum_{i=1}^{N_t} p_{m,i,t|t-1} |\mu_{t,i}|^2 + [\mathbf{R}_t]_{m,m}$, and $p_{m,i,t|t-1}$ denotes the (m,i)-th element of the block matrix $\overline{\mathbf{P}}_{t|t-1} = \alpha\overline{\mathbf{P}}_{t-1|t-1}$. Using the approximation in (34) to compute the Kalman gain $\mathbf{K}_t$, the matrix multiplications and the inversion in channel update step are replaced by element-wise operations, which significantly reduces the complexity of traditional KF from $\mathcal{O}(N_r^2 N_t + N_r^3 + N_r N_t)$ to $\mathcal{O}(N_r N_t)$. In a nutshell, the process of the SIA-DKF algorithm is shown in Algorithm 2. Meanwhile, a detailed comparison of the computational complexity between the proposed SIA-DKF and the traditional method is also provided in Table I, where T_{iter} denotes the number of VI iterations. In our experiments, the VI process converges when T_{iter} is set between 5 and 8 during the detection stage [10].

TABLE I
THE COMPUTATIONAL COMPLEXITIES OF SIA-DKF AND TRADITIONAL ALGORITHM PER UPDATE

JED methods	Channel estimation	Detection
LS-MMSE	$\mathcal{O}_{\mathrm{LS}}(N_r N_t^2)$	$\mathcal{O}_{\mathrm{MMSE}}(N_r N_t^2 + N_t^3)$
KF-MMSE	$\mathcal{O}_{\mathrm{KF}}(N_r N_t^2 + N_r^3)$	$\mathcal{O}_{\mathrm{MMSE}}(N_r N_t^2 + N_t^3)$
SIA-DKF	$\mathcal{O}_{\mathrm{DKF}}(N_r N_t)$	$\mathcal{O}_{\mathrm{VI}}(N_r N_t^2 + T_{iter} N_t^2)$

V. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed SIA-DKF method. We adopt the transpose of randomly selected N_t columns from an $T_p \times T_p$ DFT matrix as the orthogonal pilot matrix. Unless otherwise specified, the default simulation parameters are set as $N_r = 128, N_t = 16, T_p = N_t, T_d = 40 * T_p, f_d = 0.01$, and 64-QAM modulation is employed.

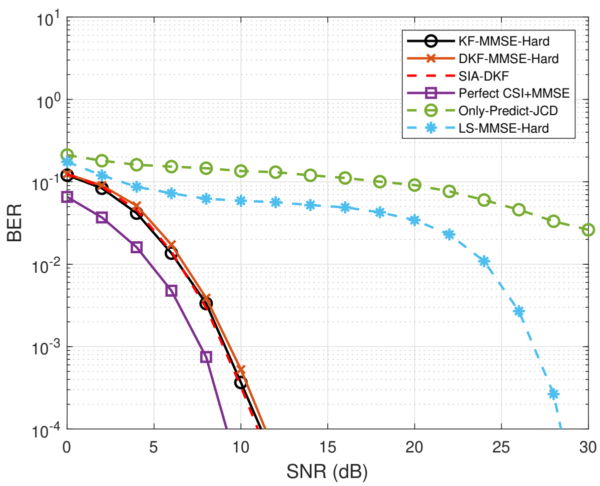

Fig. 2. BER versus SNR under 64-QAM scheme with $N_r = 128, N_t = 16, T_p = N_t, T_d = 40T_p, f_d = 0.01$.

As observed in Fig. 2, in time-varying scenarios, BER curves obtained with pilot-only channel estimation or conventional iterative JCD methods exhibit a slow decay as the SNR increases. In contrast, the proposed SIA-DKF achieves superior BER performance with significantly lower computational complexity.

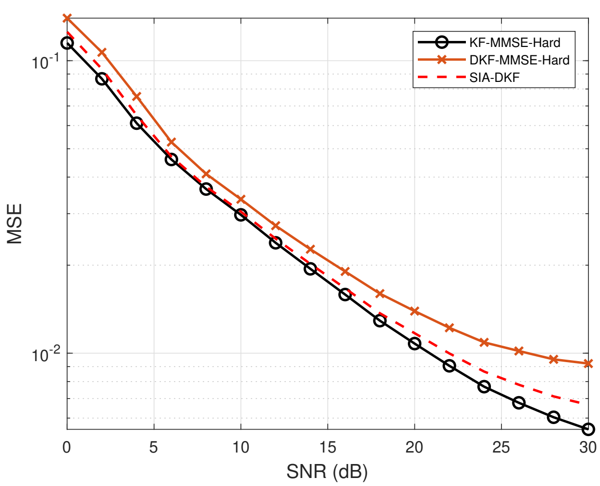

Fig. 3. MSE versus SNR under 64-QAM scheme with $N_r = 128, N_t = 16, T_p = N_t, T_d = 40T_p, f_d = 0.01$.

Fig. 3 illustrates the channel estimation mean-square error for three joint algorithms. As observed, SIA-DKF maintains estimation accuracy close to the full-state baseline.

VI. CONCLUSION

This paper proposes a soft-information-aided diagonalized Kalman filtering (SIA-DKF) algorithm for joint channel estimation and signal detection. By leveraging variational inference and diagonalizing the innovation covariance matrix, the method mitigates model mismatch while reducing computational complexity. Simulation results show that SIA-DKF achieves better BER performance with lower complexity than conventional KF-based JED methods.

VII. ACKNOWLEDGMENT

This work was supported in part by the National Key R&D Program of China under Grants No.2023YFC2205501, in part by National Science and Technology Major Projects of China under Grant 2025ZD1301800, and in part by National Natural Science Foundation of China under Grants No.62371124.

REFERENCES

[1] Y. Huang, ”Challenges and Opportunities of Sub-6 GHz Integrated Sensing and Communications for 5G-Advanced and Beyond,” in Chinese Journal of Electronics, vol. 33, no. 2, pp. 323-325, March 2024.

[2] L. Luo, J. Zhang and Z. Shi, “Iterative (turbo) joint channel estimation and signal detection for Quadrature OFDMA systems,” *IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, Cannes, France, 2008, pp. 1-5.

[3] J. Ylioinas and M. Juntti, “Iterative Joint Detection, Decoding, and Channel Estimation in Turbo-Coded MIMO-OFDM,” *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1784-1796, May 2009.

[4] A. Karataev, C. Forsch and L. Cottatellucci, “Bilinear Expectation Propagation for Distributed Semi-Blind Joint Channel Estimation and Data Detection in Cell-Free Massive MIMO,” *IEEE Open Journal of Signal Processing*, vol. 5, pp. 284-293, 2024.

[5] S. M. S. Sadough and Z. Chamideh, “Efficient Variational Bayesian Method for Joint Channel Estimation and Signal Detection in OFDM-Based AF Relay Networks,” *IEEE Communications Letters*, vol. 19, no. 10, pp. 1786-1789, Oct. 2015.

[6] E. Conti, A. Piemontese, G. Colavolpe and A. Vannucci, “Expectation Propagation for Flat-Fading Channels,” *IEEE Wireless Commun. Lett.*, vol. 12, no. 11, pp. 1846-1850, Nov. 2023.

[7] H. Hijazi and L. Ros, “Joint data QR-detection and Kalman estimation for OFDM time-varying Rayleigh channel complex gains,” *IEEE Trans. Commun.*, vol. 58, no. 1, pp. 170-178, January 2010.

[8] Xiaobo Zhou, Yunxin Zhao and Xinhua Zhuang, “Kalman filter channel tracking algorithm based on AR model in WCDMA systems,” *International Conference on Signal Processing Proceedings*. Beijing, China, 2000, pp. 1759-1762 vol.3.

[9] L. He, Z. Wang, Z. Gao, L. Liu and Y. Huang, “Favorable-Propagation-Exploited Variational Inference for Massive MIMO Detection,” *IEEE Trans. Veh. Technol.*, vol. 73, no. 9, pp. 14074-14079, Sept. 2024.

[10] F. Rusek et al., “Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays,” *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40-60, Jan. 2013

[11] C. D’souza and R. Zanetti, “Information Formulation of the UDU Kalman Filter,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 1, pp. 493-498, Feb. 2019.

[12] O. Edfors, M. Sandell, J. . -J. van de Beek, S. K. Wilson and P. O. Borjesson, “OFDM channel estimation by singular value decomposition,” *IEEE Transactions on Communications*, vol. 46, no. 7, pp. 931-939, July 1998.