

Impacts of the measurement equation modification of the adaptive Kalman filter on joint polarization and laser phase noise tracking

Shaohua Hu (胡少华)¹, Jing Zhang (张静)^{1*}, Qun Liu (刘群)¹, Linchangchun Bai (柏林长春)¹, Xingwen Yi (易兴文)², Bo Xu (许渤)¹, and Kun Qiu (邱昆)¹

¹Key Laboratory of Optical Fiber Sensing and Communications, University of Electronic Science and Technology of China, Chengdu 611731, China

²School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510275, China

*Corresponding author: zhangjing1983@uestc.edu.cn

Received October 22, 2021 | Accepted November 18, 2021 | Posted Online December 13, 2021

Kalman filtering (KF) has good potential in fast rotation of state of polarization (RSOP) tracking. Different measurement equations cause the diverse RSOP tracking performances. We compare the conventional KF (CKF) and the modified KF (MKF), which have different measurement equations. Semi-theoretical analysis indicates the lower conditional variances of measurement residuals and process noise of MKF. Compared with CKF, the MKF has > 3 dB optical signal-to-noise ratio (OSNR) improvement at the 10 MHz scrambling rate in simulation. For MKF, more significant tracking speed improvement exists for lower OSNR. MKF can be smoothly combined with an adaptive algorithm, which outperforms adaptive CKF throughout the simulations.

Keywords: optics communication; polarization de-multiplexing; coherent transmission; adaptive filtering.

DOI: [10.3788/COL202220.020603](https://doi.org/10.3788/COL202220.020603)

1. Introduction

Polarization division multiplexing plays an important role to double the spectral efficiency for high speed coherent optical fiber transmissions. The polarization de-multiplexing algorithms can be divided into blind equalization and data-aided multi-input multi-output (MIMO) algorithms. The blind algorithms are mainly based on the characteristic of constellation points, such as constant modulus algorithm (CMA), multi-modulus algorithm (MMA), and cascaded MMA (CMMA), together with their variants^[1-3]. These kinds of blind algorithms are based on the constant modulus or certain radii, which enables de-multiplexing in the presence of laser phase noise, frame asynchronism, and inter-symbol interference^[4]. However, these blind algorithms have low convergence speed, singularity problem, and step-length sensitivity^[5] due to the intrinsic shortcoming of such modulus-based cost functions. Besides, the independent component algorithm (ICA) is also effective for blind source separation^[6,7], which solves the singularity^[8] but has the inherent polarization alignment problem^[9]. The singularity problem can be solved by Stokes space estimation or as the initialization of the elements in the Jones matrix for the CMA-based blind equalizations. However, the fast rotation of state of polarization (RSOP), whose speed may exceed several hundred

mega radians per second in some extreme cases^[10], may disable the conventional MIMO fiber system and bring enormous challenges to existing digital signal processing.

The Kalman filtering (KF) has been widely used in system control, which has been introduced into optical communications in recent years. KF has both blind^[11,12] and data-aided two-stage equalization forms^[13] for both coherent and direct detection schemes^[14,15]. Blind KF avoids the decision-error-induced divergence during iteration, but is modulation format dependent. The two-stage data-aided KF uses the training sequence at the first stage and decision-directed tracking at the second stage, enabling general de-multiplexing under all kinds of modulation formats without singularity or channel logical confusion at the first stage. However, the performance deterioration of the second stage may occur in the presence of decision error, especially for low SNR cases. In our previous work, we have proposed modified KF (MKF) with higher speed of convergence and stability by modifying the measurement equation of the conventional KF (CKF)^[14]. However, the improvements from measurement equation modification have not been analyzed, which will be beneficial for different kinds of KF.

In this paper, we analyze the noise tolerance of CKF and MKF in theory by introducing conditional variance of measurement

residual and process noise. We find that the enhanced noise tolerance of MKF compared with CKF results from the decrease of the mean square error in each iteration by numerical simulations. Besides, the covariance reduction indicates the noise tolerance improvement of MKF, which results in better tracking ability and covariance initialization flexibility. Furthermore, we find the noise covariance matching method can be smoothly combined with MKF to adaptively update the covariance matrices of measurement noise and process noise. We call it adaptive MKF corresponding to adaptive CKF. The simulation comparison between adaptive CKF and adaptive MKF is conducted in virtual photonics instrument (VPI)-MATLAB co-simulation on a single carrier quadrature amplitude modulation (QAM)-16 coherent optical fiber transmission system. Compared with adaptive CKF, adaptive MKF has > 3 dB optical SNR (OSNR) improvement at the scrambling rate of 10 MHz, and around 73.5% scrambling rate improvement at the OSNR of 24 dB. Particularly, the adaptive MKF retains the successful joint tracking of the fast RSOP and laser phase noise in the lower OSNR cases. For higher OSNR, either the MKF or adaptive MKF outperforms CKF or adaptive CKF in initialization flexibility.

2. Theoretical Comparisons between CKF and MKF

Before the comparison, we denote the channel model and variables of the transmission system. The polarization diversity multiplexing (PDM) coherent optical transmission is assumed to be linear and the discrete-time form of the received signal $y(n)$ after dispersion compensation can be expressed as

$$y(n) = H_{\text{CD}}^{-1}\{H_{\text{CD}}[\mathbf{H}x(n) + N]\}, \quad (1)$$

where $x(n)$ is discrete-time form of the transmitted signal, N is the additive white Gaussian noise (AWGN), and $H_{\text{CD}}(\cdot)$ and $H_{\text{CD}}^{-1}(\cdot)$ denote the fiber dispersion and its digital-domain compensation. The channel response matrix \mathbf{H} representing laser phase noise and RSOP can be expressed as

$$\mathbf{H} = \begin{bmatrix} H_{xx} & H_{xy} \\ H_{yx} & H_{yy} \end{bmatrix} = \begin{bmatrix} \cos \kappa(n) e^{j\xi(n)} & -\sin \kappa(n) e^{j\eta(n)} \\ \sin \kappa(n) e^{-j\eta(n)} & \cos \kappa(n) e^{-j\xi(n)} \end{bmatrix} e^{j\phi(n)}, \quad (2)$$

where $\phi(n)$ is the Wiener laser phase noise, and the RSOP can be characterized with a three-parameter model denoted by the amplitude ratio angle κ and the phase angles η and ξ ^[16–18]. We can merge the five-parameter estimation of laser phase noise and Jones matrix in Eq. (2) into four-parameter estimation by multiplying the laser phase noise into the matrix.

The model of CKF and MKF is defined as follows. During the iteration process of KF, the convergence can be achieved by minimizing the difference between the predicted signal and the observed signal. It is assumed that \hat{z}_{CKF} and \hat{z}_{MKF} are signals predicted by CKF and MKF. In CKF, \hat{z}_{CKF} can be derived by

$$\hat{z}_{\text{CKF}} = \begin{bmatrix} \hat{x}_x(n) \\ \hat{x}_y(n) \end{bmatrix} = \hat{\mathbf{G}} \begin{bmatrix} y_x(n) \\ y_y(n) \end{bmatrix} = \hat{\mathbf{G}}\mathbf{y}, \quad (3)$$

whereas in MKF, \hat{z}_{MKF} can be derived by

$$\hat{z}_{\text{MKF}} = \begin{bmatrix} \hat{y}_x(n) \\ \hat{y}_y(n) \end{bmatrix} = \mathbf{H} \begin{bmatrix} x_x(n) \\ x_y(n) \end{bmatrix} = \mathbf{H}\mathbf{x}, \quad (4)$$

where the subscripts x and y in Eqs. (3) and (4) stand for the signal on two orthogonal polarization states, i.e., x -Pol. and y -Pol, and $\hat{\mathbf{H}}$ and $\hat{\mathbf{G}}$ are the estimated channel response matrix and the estimated inverse channel response matrix, respectively. Then, the measurement residual δ_{CKF} and δ_{MKF} for CKF and MKF can be expressed as

$$\delta_{\text{CKF}} = \mathbf{z}_{\text{CKF}} - \hat{z}_{\text{CKF}} = \begin{bmatrix} x_x(n) \\ x_y(n) \end{bmatrix} - \hat{\mathbf{G}} \begin{bmatrix} y_x(n) \\ y_y(n) \end{bmatrix}, \quad (5)$$

$$\delta_{\text{MKF}} = \mathbf{z}_{\text{MKF}} - \hat{z}_{\text{MKF}} = \begin{bmatrix} y_x(n) \\ y_y(n) \end{bmatrix} - \hat{\mathbf{H}} \begin{bmatrix} x_x(n) \\ x_y(n) \end{bmatrix}. \quad (6)$$

The difference in measurement residual between CKF and MKF is illustrated in Fig. 1. For CKF, the noisy terms $\hat{\mathbf{G}}$ and the predicted measurement $y(n)$ are multiplied together to influence the measurement residual. While MKF differs from CKF, only one noise-related factor contributes to the prediction \hat{z}_{MKF} , and the two noisy terms $y(n)$ and $\hat{\mathbf{H}}$ additively act on the measurement residual. We show the equalization model and state estimation process in Fig. 2.

Apart from the measurement residual, the KF algorithm also minimizes the process noise. \mathbf{G} and \mathbf{H} are defined as the state matrix of CKF and MKF, respectively. The process noises $\Delta_{\mathbf{G}}$

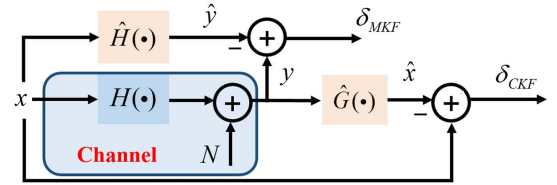


Fig. 1. Measurement residuals of CKF and MKF.

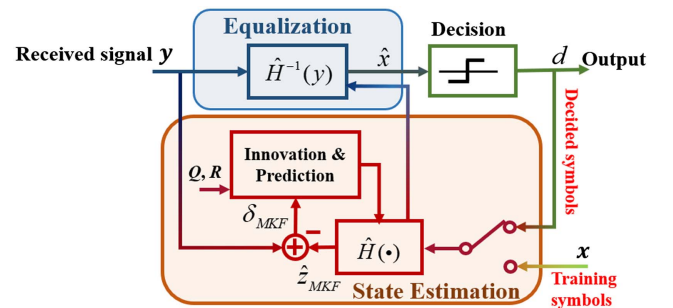


Fig. 2. Equalization model and state estimation process of MKF.

and Δ_H in each iteration for CKF and MKF can be expressed in Eqs. (7) and (8):

$$\hat{\mathbf{G}} = \mathbf{G} + \Delta_G = \mathbf{H}^{-1} + \Delta_G, \quad (7)$$

$$\hat{\mathbf{H}} = \mathbf{H} + \Delta_H. \quad (8)$$

To concentrate on the noise tolerance analysis of the two schemes, we degenerate the theoretical analysis into single polarization. Substituting $\hat{\mathbf{G}}$ in Eq. (5) with Eq. (7), we derive the measurement residual of CKF as

$$\begin{aligned} \delta_{\text{CKF}} &= x - \hat{x} = x - \hat{\mathbf{G}}(\mathbf{H} \cdot x + N) \\ &= x - (\mathbf{H}^{-1} + \Delta_G)(\mathbf{H} \cdot x + N) \\ &= -(\mathbf{G} + \Delta_G)N - \Delta_G Hx = -\hat{\mathbf{G}}N - \Delta_G Hx. \end{aligned} \quad (9)$$

Similarly, we have the measurement residual δ_{MKF} ,

$$\begin{aligned} \delta_{\text{MKF}} &= y - \hat{y} = y - \hat{\mathbf{H}}x = (\mathbf{H}x + N) - \hat{\mathbf{H}}x \\ &= (\mathbf{H}x + N) - (\mathbf{H} + \Delta_H)x = N - \Delta_H x. \end{aligned} \quad (10)$$

According to Eqs. (9) and (10), the process noise for CKF and MKF versus the measurement residual can be expressed as

$$\Delta_G = -(\delta_{\text{CKF}} + \mathbf{H}^{-1}N)/y = -(\delta_{\text{CKF}} + \mathbf{H}^{-1}N)/(\mathbf{H}x + N), \quad (11)$$

$$\Delta_H = (N - \delta_{\text{MKF}})/x. \quad (12)$$

We introduce the conditional variance of measurement residuals and process noises for CKF and MKF to evaluate the noise tolerance and to explore their interaction during each prediction and innovation iteration. We define the conditional variance of δ_{CKF} when sequence x is transmitted as

$$D(\delta_{\text{CKF}}|x) = E_{\Delta_G}[D(\hat{\mathbf{G}}N + \Delta_G Hx)]. \quad (13)$$

Similarly, we have

$$D(\delta_{\text{MKF}}|x) = E_{\Delta_H}[D(N - \Delta_H x)], \quad (14)$$

$$D(\Delta_G|x) = E_{\delta_{\text{CKF}}}\{D[(\delta_{\text{CKF}} + \mathbf{H}^{-1}N)/(\mathbf{H}x + N)]\}, \quad (15)$$

$$D(\Delta_H|x) = E_{\delta_{\text{MKF}}}\{D[(N - \delta_{\text{MKF}})/x]\}, \quad (16)$$

where $E(\cdot)$ and $D(\cdot)$ are the expectation and the second central moment. The conditional variances are derived from the property of

$$D(A|C) = E_B[D(A|B,C)] + D_B[E(A|B,C)],$$

where the second term on the right hand is small enough and can be assumed to be zero since the process noise and the measurement residual of a converged equalizer in a certain iteration can be regarded as a zero-mean random variable. Larger

conditional variance leads to worse noise tolerance particularly at the hard decision stage, and hence the filter has higher probability suffering from the noise-dominated divergence.

To avoid the complicated analytical derivation in variance calculation, numerical simulations are conducted to compare the noise tolerance between CKF and MKF based on Eqs. (13)–(16), where a 16-QAM sequence with 2^{16} complex symbols is transmitted through the AWGN channel. For simplicity, we compare the process noises by setting the channel-scaling factor, $H = 1$, which leads to $G = H^{-1} = 1$. Figures 3(a) and 3(b) show the conditional variance curves of measurement error and process noise varying with process noise and measurement error at different SNRs. To fully compare the conditional variance, we set the x axis in logarithmic scale, which covers the cases with both slighter and stronger noise. We take specific values of two variables to illustrate the rationality of the abscissa of Figs. 3(a) and 3(b). Since the I- or Q-component of the transmitted QAM signal ranges from -3 to 3 , $\delta_{\text{KF}} = 6$ in Fig. 3(a) indicates that the measurement residual is on the same order of magnitude as the transmitted signal, which is large enough for a convergent Kalman filter. In Fig. 3(b), $\Delta_{\text{KF}} = 10^{[10 \log_{10}(\Delta_{\text{KF}})]/10} = 10^{3/10} = 2$ also provides the conditional variance under the situation with larger process noise. As shown in Figs. 3(a) and 3(b), both the conditional variances of process noise and measurement error for MKF are always smaller than those of CKF. The gaps between MKFs and CKFs occur in practical SNRs and tend to be more obvious for lower SNR cases.

To compare the SNR robustness between MKF and CKF, we further depict the ratio of $D(\delta_{\text{CKF}}|\Delta_G, x)$ and $D(\delta_{\text{MKF}}|\Delta_H, x)$ versus SNRs at different process noise levels, as shown in Fig. 4. All of the conditional variance ratios are larger than one, which indicates that MKF always introduces a smaller measurement residual, especially for channels with higher noise power.

Therefore, for the general SNR region, we always have

$$D(\delta_{\text{MKF}}|\Delta_H, x) \leq D(\delta_{\text{CKF}}|\Delta_G, x), \quad (17)$$

$$D(\Delta_H|\delta_{\text{MKF}}, x) \leq D(\Delta_G|\delta_{\text{CKF}}, x), \quad (18)$$

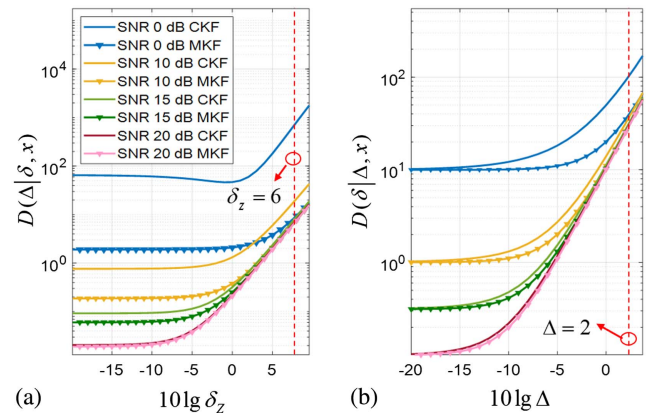


Fig. 3. (a) Variance of process noise versus the measurement error; (b) variance of measurement noise versus the process noise.

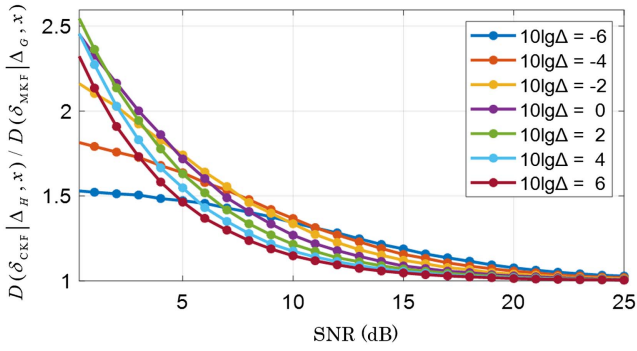


Fig. 4. Variance rate of measurement errors versus SNR.

which means that MKF has less equalization enhanced instability than CKF. By averaging each case over $\Delta_{H(G)}$ and $\delta_{(\text{MKF})\text{CKF}}$ in Fig. 2, for the given SNRs, we have

$$D(\Delta_H|x) < D(\Delta_G|x), \quad (19)$$

$$D(\delta_{\text{MKF}}|x) < D(\delta_{\text{CKF}}|x). \quad (20)$$

Inequalities (19) and (20) also hold in all practical channel noise powers, as shown in Fig. 3, which indicates MKF outperforms CKF in noise tolerance, especially for large noise. This also explains the better noise tolerance for MKF compared with CKF in our previous experiment^[19].

For stability and real implementation, we introduce the noise covariance matching method into MKF, since it has lower complexity and is more practical to implement than other adaptive algorithms^[20,21]. The adaptive estimations of covariance matrices of process and measurement noise \mathbf{Q}_k and \mathbf{R}_k are expressed in a compact form as Eqs. (21) and (22):

$$\mathbf{R}_k = \alpha_R \mathbf{R}_{k-1} + (1 - \alpha_R) \frac{\text{tr}(\check{\delta}_{\text{MKF},k} \check{\delta}_{\text{MKF},k}^H + \mathbf{M}_n \mathbf{P}_k \mathbf{M}_n^H)}{\text{tr}(\mathbf{R}_{k-1})} \mathbf{R}_{k-1}, \quad (21)$$

$$\mathbf{Q}_k = \alpha_Q \mathbf{Q}_{k-1} + (1 - \alpha_Q) \frac{\text{tr}(\mathbf{K}_k \check{\delta}_{\text{MKF},k} \check{\delta}_{\text{MKF},k}^H \mathbf{K}_k^H)}{\text{tr}(\mathbf{Q}_{k-1})} \mathbf{Q}_{k-1}, \quad (22)$$

where

$$\mathbf{M}_n = \begin{bmatrix} x_x(n) & x_y(n) & 0 & 0 \\ 0 & 0 & x_x(n) & x_y(n) \end{bmatrix}, \quad (23)$$

$$\mathbf{s}_k = [\hat{H}_{xx}(k), \hat{H}_{xy}(k), \hat{H}_{yx}(k), \hat{H}_{yy}(k)], \quad (24)$$

and $\check{\delta}_{\text{MKF},k} = [y_x, y_y]^T - \mathbf{M}_n \mathbf{s}_k^T$ is the measurement residual after the update of \mathbf{s}_k . “tr” represents the trace of the matrix. \mathbf{K} and \mathbf{P} are the Kalman gain matrix and the error autocorrelation matrix. α_R and α_Q are forgetting factors that are helpful to balance the tracking speed and estimation accuracy.

When the channel is memorial, the state vector can be extended to multi-tap. When obtaining the multi-tap state

vector \mathbf{s} , we can reshape the state vector to the channel response matrix \mathbf{A} and determine its pseudo-inverse matrix as $\mathbf{G} = \mathbf{A}^\dagger = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H$ if the columns of \mathbf{A} are linearly dependent. Besides, the inter-symbol interference can also be mitigated by conventional adaptive equalizers with the least mean squares (LMS) or recursive least squares (RLS) algorithms.

3. Simulation Comparisons of the Tracking Abilities

In this section, we conduct numerical simulations to testify the proposed modified adaptive Kalman filter for joint polarization and phase noise tracking. The simulation system is a conventional polarization-multiplexed coherent 16-QAM optical transmission system established in VPI 9.1. System parameters are shown in Table 1. The block of RSOP is set based on Eq. (2), where the angle $\kappa = 2\pi f t$ varies in a sinusoidal form with the scrambling frequency f . At the receiver, frequency-domain group velocity dispersion compensation, fourth power FFT-based frequency offset compensation, and match filter are applied before the Kalman filter. The adaptive KF is adopted for polarization de-multiplexing and phase recovery. Both CKF and MKF adopt the covariance matching adaptive algorithm with the initial covariance of 0.5.

3.1. Noise tolerance

We first compare the tracking performance between adaptive CKF and adaptive MKF at different OSNRs, as shown in Fig. 5, where we also compare the performance of the extended CKF (ECKF) and pilot-aided averaging window method (AWM) as a reference, where the ECKF still regards the measurement variable as the transmitted symbols but estimates the Jones matrix and laser phase noise $\phi(n)$ separately. The

Table 1. Simulation Parameter Setup.

Parameter	Value
Baud rate	28 GBaud
Pulse shaping	RRC
Training length	100
Signal linewidth	100 kHz
PMD coefficient	0.1 ps/km ^{1/2}
Dispersion	16 ps/(nm·km)
Sampling rate	112 GSa/s
Rolling factor	0.1
Freq. offset	100 MHz
Fiber length	400 km
Linewidth (LO)	100 kHz

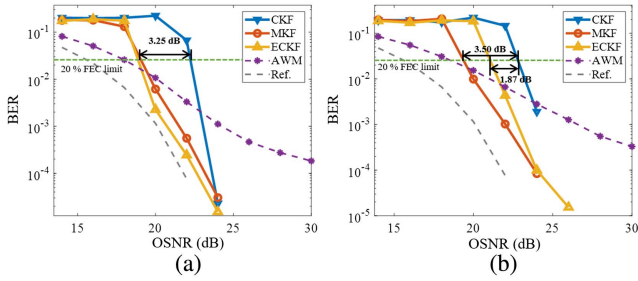


Fig. 5. BER of the recovered signal versus OSNR. The scrambling rate is (a) 10 MHz and (b) 20 MHz. The reference curves (gray dashed lines) are under the back-to-back (B2B) case without RSOP.

AWM inserts three pilot symbols between every 17 symbols and smooths the estimation noise by time averaging. The pilot overhead of AWM is 15%, while for KF the number of training symbols is 100. The overhead of KF is much less than AWM. The scrambling rate of RSOP is 10 MHz and 20 MHz, corresponding to Figs. 5(a) and 5(b), both of which have the linewidth of 100 kHz. The performances are evaluated at the 20% forward error correction (FEC) limit, i.e., the bit error rate (BER) of 2.5×10^{-2} .

When the OSNR is below 18 dB, the divergence of KF mainly results from the noise sensitivity of the second stage of Kalman filter. The BER performance of MKF and CKF mainly differs from each other within 18 to 24 dB since the noise tolerance of KF differs in this range. We find MKF outperforms CKF and ECKF for both cases and has > 3 dB OSNR improvement over CKF at the 20% FEC limit. When the scrambling rate increases to 20 MHz, 3.5 dB OSNR gain is observed compared with CKF, while 1.87 dB OSNR gain is observed compared with ECKF. AWM works better than KF at extremely low OSNRs, which benefits from its very large overhead. Once KF converges, the BERs rapidly drop below those of AWM.

3.2. Scrambling rate tolerance

We then compare the tracking ability of the adaptive CKF and MKF with lower and larger noise powers. Figures 6(a) and 6(b) show the BER performances for CKF and MKF versus scrambling rate at the OSNRs of 24 dB and 20 dB. The trends can be divided into three segments, i.e., the optimum stage

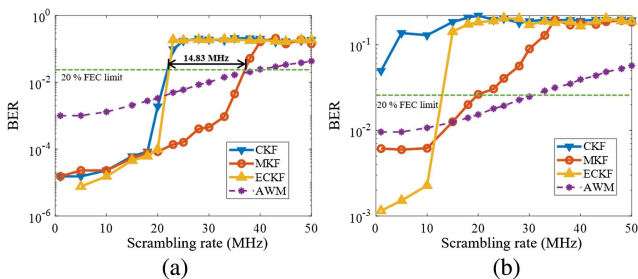


Fig. 6. (a) Signal qualities versus scrambling rates at the OSNR of 24 dB; (b) signal qualities versus scrambling rates at the OSNR of 20 dB.

(the scrambling rate is lower than 10 MHz), the transition stage (the scrambling rate from 20 MHz to 35 MHz at 24 dB OSNR, from 10 MHz to 20 MHz at 20 dB OSNR), and the divergence stage (the higher scrambling rate).

Focusing on the transition stage, i.e., the extreme tracking performance for CKF and MKF, we find that CKF and ECKF diverge sharply at the OSNR of 24 dB, whereas the performance of MKF deteriorates gradually. From Fig. 6(a), we find that MKF can converge under the scrambling rate up to 35 MHz, around 73.5% enhancement over CKF, i.e., 14.83 MHz. When the OSNR is 20 dB, CKF diverges, but MKF still has > 15 MHz scrambling rate tolerance, as shown in Fig. 6(b). This is because the enhanced channel noise has higher probability to induce cycle-slip or algorithm divergence.

At the optimum stage for the OSNR of 24 dB, the CKF and MKF have similar performances. For the OSNR of 20 dB, MKF can tolerate higher scrambling rate than ECKF at 20% FEC limit, while CKF fails to track the fast RSOP in the megahertz (MHz) level. MKF and ECKF outperform AWM in both cases, even though MKF is a kind of linear KF. The results are consistent with Fig. 5, which shows that MKF can work stably in relatively lower OSNR.

3.3. Parameter initialization flexibility

Finally, we investigate the parameter initialization tolerance with and without the Q/R adaptive algorithm. We show the BERs versus the initial Q and R with hot maps in Fig. 7, where the OSNR is 30 dB. In Fig. 7, the required R is proportional to but larger than Q when the Q/R adaptive algorithm is not enabled. Comparing the BER performance of MKF with CKF, we find MKF outperforms CKF, as shown in the blue regions of the

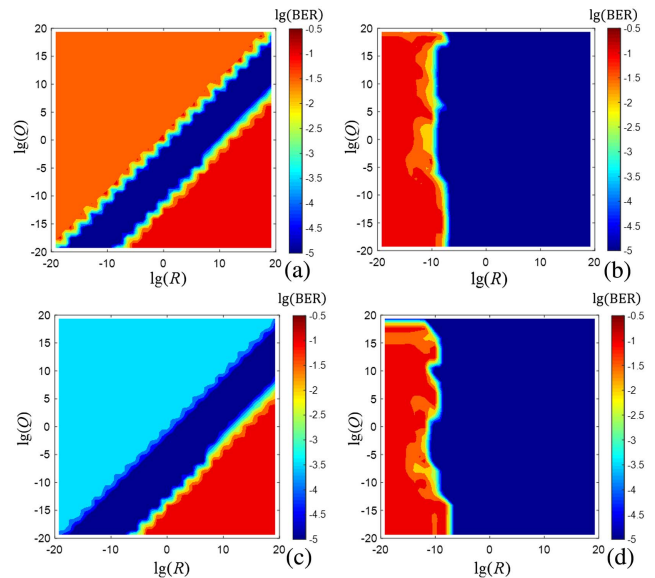


Fig. 7. Initialization tolerance of Q/R parameter for linear CKF and MKF at the OSNR of 30 dB and 10 MHz scrambling rate. (a) CKF without Q/R adaptive algorithm; (b) CKF with Q/R adaptive algorithm; (c) MKF without Q/R adaptive algorithm; (d) MKF with Q/R adaptive algorithm.

Table 2. Performance Comparison for Kalman Filters.

Filter Type	Min. OSNR at 10 MHz RSOP (dB)	Max. RSOP at 24 dB OSNR (MHz)	Max. RSOP at 20 dB OSNR (MHz)	Initialization Tolerance without Adaptive Algorithm
CKF	22	22		Worse
MKF	19	37	20	Better

hot maps in Figs. 7(a) and 7(c), which indicates a higher flexibility of MKF in Q/R initialization that is mainly brought from the improved noise tolerance of MKF.

When the adaptive algorithm is applied to the Kalman filters, the convergence regions are only dependent on the initial R . We find that the Q/R adaptive algorithm can also effectively improve the flexibility of MKF, since both the convergence regions of CKF and MKF are expanded significantly. This indicates that the adaptive algorithm can improve the robustness of MKF at higher OSNRs. While for lower OSNRs, adaptive MKF succeeds in tracking the fast RSOP but adaptive CKF fails, as discussed in Section 3.2.

The simulation results in Section 3 are concluded in Table 2 for a clear comparison. Based on the discussion above, our findings are as follows. (1) MKF has higher noise tolerance than CKF especially with lower OSNRs. (2) The modification improves the tracking speed and initialization flexibility, which benefits from the improved noise tolerance. (3) The noise covariance matching approach is effective for improving the initialization flexibility of MKF.

4. Conclusions

Both theoretical analysis and simulation results have shown that the MKF has higher robustness to noise-induced decision error and covariance matrices initialization that are related to the tracking ability in the presence of fast RSOP. Higher tracking speed (up to 50% improvement from CKF), better OSNR tolerance (> 3 dB improvement from CKF), and superior initialization flexibility are observed from MKF. The MKF combined with the covariance matching method outperforms the adaptive CKF, especially in the case of lower OSNR.

Acknowledgement

This work was supported by the National Key Research and Development Program of China (No. 2018YFB1801704), National Natural Science Foundation of China (NSFC) (Nos. 61871082 and 61871408), State Key Laboratory of Advanced Optical Communication Systems and Networks, Shanghai Jiao Tong University (No. 2020GZKF014), Fundamental Research Funds for the Central Universities (Nos. ZYGX2020ZB043 and ZYGX2019J008), and Open Fund of IPOC (BUPT) (No. IPOC2020A011).

References

1. J. Zhang, J. Yu, N. Chi, and H.-C. Chien, "Time-domain digital pre-equalization for band-limited signals based on receiver-side adaptive equalizers," *Opt. Express* **22**, 20515 (2014).
2. X. Zhou, K. Zhong, J. Huo, L. Gao, Y. Wang, L. Wang, Y. Yang, J. Yuan, K. Long, L. Zeng, A. P. T. Lau, and C. Lu, "112 Gb/s transmission over 80 km SSMF using PDM-PAM4 and coherent detection without optical amplifier," *Opt. Express* **24**, 17359 (2016).
3. C. Zhu, A. V. Tran, S. Chen, L. B. Du, T. Anderson, A. J. Lowery, and E. Skafidas, "Improved two-stage equalization for coherent Pol-Mux QPSK and 16-QAM systems," *Opt. Express* **20**, B141 (2012).
4. P. Johannisson, M. Sjödin, and M. Karlsson, "A modified CMA for PS-QPSK," in *Signal Processing in Photonic Communications* (Optical Society of America, 2011), paper SPTuB3.
5. L. Lu, Y. Lei, J. Lei, X. Zou, and W. Sheng, "Analysis of singularity problem from CMA induced by PDL in polarization demultiplexing," in *Asia Communications and Photonics Conference* (2013), paper AF2F.40.
6. P. Johannisson, H. Wymeersch, M. Sjödin, A. S. Tan, E. Agrell, P. A. Andrekson, and M. Karlsson, "Convergence comparison of the CMA and ICA for blind polarization demultiplexing," *J. Opt. Commun. Netw.* **3**, 493 (2011).
7. X. Li, W.-D. Zhong, A. Alphones, C. Yu, and Z. Xu, "Channel equalization in optical OFDM systems using independent component analysis," *J. Lightwave Technol.* **32**, 3206 (2014).
8. H. Zhang, Z. Tao, L. Liu, S. Oda, T. Hoshida, and J. C. Rasmussen, "Polarization demultiplexing based on independent component analysis in optical coherent receivers," in *34th European Conference on Optical Communication* (IEEE, 2008), p. 1.
9. N. Nabavi and T. J. Hall, "Demultiplexing by independent component analysis in coherent optical transmission: the polarization channel alignment problem," in *2015 Photonics North* (IEEE, 2015), p. 1.
10. P. M. Krümmrich, D. Ronnenberg, W. Schairer, D. Wienold, F. Jenau, and M. Herrmann, "Demanding response time requirements on coherent receivers due to fast polarization rotations caused by lightning events," *Opt. Express* **24**, 12442 (2016).
11. Q. Xiang, Y. Yang, J. Cao, Q. Zhang, and Y. Yao, "Polarization tracking for Stokes-vector modulation formats using Kalman filter," in *CLEO Pacific Rim Conference* (Optical Society of America, 2018), paper W3A.19.
12. Z. Zheng, N. Cui, H. Xu, X. Zhang, W. Zhang, L. Xi, Y. Fang, and L. Li, "Window-split structured frequency domain Kalman equalization scheme for large PMD and ultra-fast RSOP in an optical coherent PDM-QPSK system," *Opt. Express* **26**, 7211 (2018).
13. T. Marshall, B. Szafraniec, and B. Nebendahl, "Kalman filter carrier and polarization-state tracking," *Opt. Lett.* **35**, 2203 (2010).
14. N. Cui, Z. Zheng, X. Zhang, W. Yi, R. Guo, W. Zhang, X. Tang, H. Xu, and L. Xi, "Joint blind equalization of CD and RSOP using a time-frequency domain Kalman filter structure in Stokes vector direct detection system," *Opt. Express* **27**, 11557 (2019).
15. W. Yi, Z. Zheng, N. Cui, X. Zhang, L. Qiu, N. Zhang, L. Xi, W. Zhang, and X. Tang, "Joint equalization scheme of ultra-fast RSOP and large PMD compensation in presence of residual chromatic dispersion," *Opt. Express* **27**, 21896 (2019).
16. N. Cui, X. Zhang, Z. Zheng, H. Xu, W. Zhang, X. Tang, L. Xi, Y. Fang, and L. Li, "Two-parameter-SOP and three-parameter-RSOP fiber channels: problem and solution for polarization demultiplexing using Stokes space," *Opt. Express* **26**, 21170 (2018).

17. B. Wang and Z. Li, "Uniformly high-speed semi-open loop polarization control and PMD suppression," *Chin. Opt. Lett.* **18**, 050601 (2020).
18. S. Wang and Z. Li, "Polarization scrambling characteristic analysis based on density of polarization states statistics," *Chin. Opt. Lett.* **18**, 060604 (2020).
19. Y. Jiang, X. Yi, S. Hu, X. Huang, W. Tang, W. Zhou, X. Huang, J. Zhang, and K. Qiu, "Polarization de-multiplexing using a modified Kalman filter in CO-OFDM transmissions," *Chin. Opt. Lett.* **17**, 030603 (2019).
20. Q. Zhang, Y. Yang, Q. Xiang, Q. He, Z. Zhou, and Y. Yao, "Noise adaptive Kalman filter for joint polarization tracking and channel equalization using cascaded covariance matching," *IEEE Photonics J.* **10**, 7900911 (2018).
21. N. Cui and X. Zhang, "Fast tracking of polarization impairments using DSP algorithms in fiber for a coherent optical communication system," *Microwave Opt. Technol. Lett.* **63**, 2453 (2021).