Supplementary Materials

I. Additional Notes

Remark 1: The valid range for the real proportion of outliers in measurements is $[0,\ 0.5]$, because if more than half of measurements are unreliable, it is meaningless to do data-based statistical inference. Hence, 0.5 is the upper bound. This fact is fundamental in mathematical and applied statistics. □

Remark 2: To guarantee figure cleanness, we distinguish different results only by different colors. Readers who have problems in identifying colors can change the shared source codes to generate different line types to display results. □

II. Additional Experiments

A. Breakdown Test

In the main body of the article (see Subsection VII-D "Sensitivity Analysis"), we fix the parameter ϵ of the algorithm and let the true proportion ϵ_{real} change from 0 to 0.5. In this supplementary experiment, we randomly fix the true proportion $\epsilon_{real} := 0.3$, and let the parameter ϵ of the algorithm vary in $[0,\ 0.5]$. We have the result in Fig. 1. Since $\epsilon_{real} := 0.3$ is fixed, the TMKF and the TK only generate one result, respectively. The result of the TMKF gives the lower reference line (in black), while that of the KF gives the upper one (in red). Fig. 1 shows that the parameter ϵ used in the algorithm is insensitive to the true proportion $\epsilon_{real} := 0.3$. This conclusion is consistent with that in Subsection VII-D "Sensitivity Analysis". Therefore, we finished validating the claims made in Remark 2 of the main body of the article.

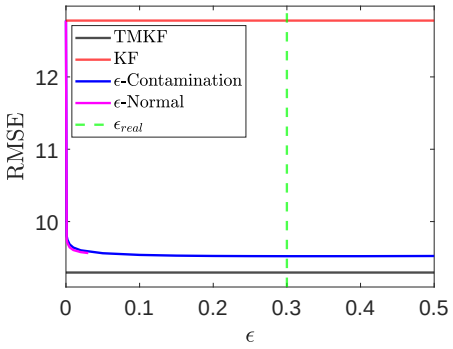


Fig. 1. Parameter sensitivity against the true proportion of outliers.

Remark 3: Note that in the ϵ-normal case, we only allow ϵ to vary from 0 to 0.0303. The reason can be found in Remark 2 of the main body of the article. □

B. Computational Complexity Analysis

As we can see in Algorithm 1, the most computationally intensive step is to solve (34) to obtain the worst-case scenario (i.e., $\boldsymbol{\Sigma}^*_{\boldsymbol{x},k}$ and $\boldsymbol{\Sigma}^*_{\boldsymbol{v},k}$) under the Wasserstein ambiguity sets. Problem (34) is a SDP which is numerically challenging to solve. Instead, if we use the moment-based ambiguity sets, we need to solve (36) to obtain the worst-case scenario. However, (36) can be analytically solved by Theorem 5. As a result, all the steps in Algorithm 1 have closed-form solutions, implying that the computational complexity is no longer an issue. This is the reason why we adopted the moment-based ambiguity sets throughout experiments.

Let $r := \max\{n, p, m\}$ where n is the dimension of the state vector $\boldsymbol{x}_k$, p of the process noise vector $\boldsymbol{w}_k$, and m of the measurement vector $\boldsymbol{y}_k$. Since for a usual state estimation problem $n \geq p$ and $n \geq m$, it is well-known that the (asymptotic) computational complexity of the canonical Kalman filter at each time step is $O(r^3) = O(n^3)$. This is because all computational operations at each time step of the Kalman filtering are just matrix addition/subtraction, matrix multiplicity, and matrix inverse; matrix multiplicity and matrix inverse operations admit cubic order of computational complexity in terms of the dimensions of the involved matrices. Therefore, likewise, the computational complexity of the proposed method is also $O(n^3)$ at each time step, given that the moment-based ambiguity sets are used.

Remark 4: The computational complexity of matrix multiplicity and matrix inverse operations can be reduced using advanced algorithms such as the Strassen algorithm. The mentioned "cubic order" are just results given by the conventional matrix multiplicity (e.g., definition) and matrix inverse (e.g., Gauss-Seidel) methods. □

For a demonstration, under the moment-based ambiguity sets, we augment the system matrices to improve the dimensions of the state vectors and the measurement vectors. Specifically, we let $\bar{\boldsymbol{F}}_k^{real}$, $\bar{\boldsymbol{G}}_k$, $\bar{\boldsymbol{H}}_k$, $\bar{\boldsymbol{Q}}_k$, and $\bar{\boldsymbol{R}}_k$ form a new uncertain linear system, where

$$\begin{aligned}
\bar{\boldsymbol{F}}_k^{real} &:= \mathrm{blkdiag}\{\boldsymbol{F}_k^{real}, \boldsymbol{F}_k^{real}, \boldsymbol{F}_k^{real}, \boldsymbol{F}_k^{real}, \boldsymbol{F}_k^{real}\},\\
\bar{\boldsymbol{G}}_k &:= \mathrm{blkdiag}\{\boldsymbol{G}_k, \boldsymbol{G}_k, \boldsymbol{G}_k, \boldsymbol{G}_k, \boldsymbol{G}_k\},\\
\bar{\boldsymbol{H}}_k &:= \mathrm{blkdiag}\{\boldsymbol{H}_k, \boldsymbol{H}_k, \boldsymbol{H}_k, \boldsymbol{H}_k, \boldsymbol{H}_k\},\\
\bar{\boldsymbol{Q}}_k &:= \mathrm{blkdiag}\{\boldsymbol{Q}_k, \boldsymbol{Q}_k, \boldsymbol{Q}_k, \boldsymbol{Q}_k, \boldsymbol{Q}_k\},
\end{aligned}$$

and $\bar{\boldsymbol{R}}_k := \mathrm{blkdiag}\{\boldsymbol{R}_k, \boldsymbol{R}_k, \boldsymbol{R}_k, \boldsymbol{R}_k, \boldsymbol{R}_k\}$. To be specific, for example,

$$\bar{\boldsymbol{F}}_k^{real} := \begin{bmatrix} \boldsymbol{F}_k^{real} & & & & \\ & \boldsymbol{F}_k^{real} & & & \\ & & \boldsymbol{F}_k^{real} & & \\ & & & \boldsymbol{F}_k^{real} & \\ & & & & \boldsymbol{F}_k^{real} \end{bmatrix}.$$

In this case, the dimension of the state vector is promoted from 2 to 10, and that of the measurement vector is promoted from 1 to 5.

We let $\alpha = 1$ and the true proportion of outliers equals to 0.05. Parameters of the candidate filters keep the same as used in the experiments of the main body of the article. The results are shown in Table I. We can see that the conclusions drawn from Table I are the same to those drawn from Table III in the main body of the article: the MKF is attractive for its performance and computational efficiency. (Besides, the calculation burden does not significantly increase as the scale of the problem rises.)

TABLE I
RESULTS WHEN $\alpha = 1$ AND ALSO OUTLIERS ($n = 10$)

Filter	RMSE	Avg Time	Filter	RMSE	Avg Time
TMKF	**7.23**	2.83e-5	τ-KF [27]	Comp. Slow	
KF	35.78	2.53e-5	WKF [28]	29.76	191.33e-5
HKF [33]	33.78	4.18e-5	MKF[Ours]	**26.17**	2.95e-5

Comp. Slow stands for Computationally Slow.

However, the difference between the performances of the TMKF and the MKF is still significant. Therefore, we are reminded again that robust filters are just remedial solutions against uncertain conditions: whether they are satisfactory or not depends on the real demands of the estimation accuracy for specific problems. If the prescribed accuracy cannot be obtained by robust solutions, what we can do is only to endeavor to improve the exactness of the nominal model (i.e., improve the modelling precision). To validate this point, we alternatively let $\alpha = 0.5$, i.e., the accuracy of the nominal model is improved. We have the results in Table II. It shows that the difference between the performances of the TMKF and the MKF becomes smaller.

TABLE II
RESULTS WHEN $\alpha = 0.5$ AND ALSO OUTLIERS ($n = 10$)

Filter	RMSE	Avg Time	Filter	RMSE	Avg Time
TMKF	**10.07**	3.14e-5	τ-KF [27]	Comp. Slow	
KF	28.33	3.29e-5	WKF [28]	25.30	193.51e-5
HKF [33]	23.18	3.31e-5	MKF[Ours]	**21.84**	3.20e-5

Comp. Slow stands for Computationally Slow.

C. Student's t-Distributed Measurement Noise

In this subsection, we investigate the performances of the candidate filters for Student's t-distributed measurement noises. The degree of freedom of the used Student's t-distribution is set to be 3. But the covariance of measurement noise $\boldsymbol{v}_k$ at each time step is kept unchanged as $\boldsymbol{R}_k$. Note that although the variance σ^2 of a t-distribution is determined by its degree of freedom ν through $\sigma^2 = \frac{\nu}{\nu-2}$ for $\nu \geq 3$, it can be scaled by constant coefficients. For example, supposing a random variable $\boldsymbol{T}$ follows a t-distribution with degree of freedom ν, the variance of $\frac{1}{\sqrt{\frac{\nu}{\nu-2}}}\boldsymbol{T}$ is unit.

Parameters of the candidate filters are tuned to perform best, respectively, for this new instance. (Details can be found in the shared source codes.) The results when only outliers exist are shown in Table III, while those when both model uncertainties and outliers exist are shown in Table IV. Note that in this case, the TMKF designed for Gaussian-noise models is no longer optimal for the t-noise true model (i.e., it reduces to the KF when $\alpha = 0$).

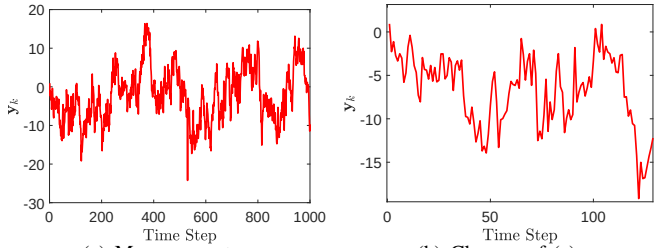
TABLE III
RESULTS WHEN $\alpha = 0$ AND ONLY OUTLIERS (T-DISTRIBUTED AND $n = 2$)

Filter	RMSE	Avg Time	Filter	RMSE	Avg Time
TMKF	**6.38**	1.58e-5	τ-KF [27]	7.07	26.00e-5
KF	**6.38**	1.18e-5	WKF [28]	**6.80**	121.51e-5
HKF [33]	**6.38**	1.66e-5	MKF[Ours]	**6.72**	1.31e-5

TABLE IV
RESULTS WHEN $\alpha = 1$ AND ALSO OUTLIERS (T-DISTRIBUTED AND $n = 2$)

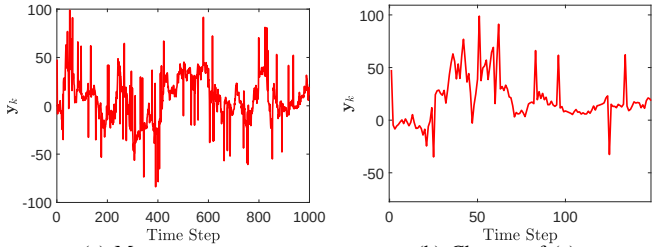
Filter	RMSE	Avg Time	Filter	RMSE	Avg Time
TMKF	**3.53**	1.16e-5	τ-KF [27]	8.73	22.62e-5
KF	13.40	0.85e-5	WKF [28]	**8.34**	115.05-5
HKF [33]	13.75	1.25e-5	MKF[Ours]	**8.40**	1.28e-5

As we can see, when there are no model uncertainties (see Table III), the TMKF, KF, and HKF have the same performance, and the τ-KF, WKF, and MKF perform worse than them. In other words, the Huber-based outlier-insensitive filter (HKF) no longer has advantage over the KF. This is because the measurements subject to t-distributed measurement noises do not have significantly outstanding outliers; see Fig. 2. In contrast, in the experiments in the main body of the article, we added significantly outstanding outliers; see Fig. 3. The two cases are all common in signal processing practices. Therefore, the outlier-robust methods (i.e., HKF and MKF) are more suitable for the cases where outliers significantly exist. Again, we see the price of robustness under uncertain conditions is sacrificing the optimality under perfect conditions because the MKF has larger RMSE than the HKF when there are no model uncertainties.



(a) Measurements (b) Closeup of (a)

Fig. 2. Measurements contaminated by t-distributed noises.



(a) Measurements (b) Closeup of (a)

Fig. 3. Measurements contaminated by significant outliers.

When there exist model uncertainties (see Table IV), as expected, the τ-KF, WKF, and MKF perform better because they are relatively robust against uncertainties. In this case, the MKF has smaller RMSE than the HKF, which verifies the claim that the sacrifice of optimality under perfect conditions might offer the robustness under uncertain conditions.

III. CONCLUSIONS

1) The proposed outlier-robust filtering frameworks that use influence functions in Theorem 6 are more suitable for

the cases that measurements contain significantly outstanding outliers and for the case that the noise models are unknown. If measurement noises are t-distributed, it means that the system model is exactly known so that we can derive optimal filters for t-distributed noises (theoretically, this is still possible no matter whether the mathematical derivation is easy or not; Cf. [1]). However, a filter that is optimal (or suitable) for t-distributed noise is likely to lose robustness for other types of noises.

2) The price of the robustness under uncertain conditions is sacrificing the optimality under perfect conditions.
3) Robust filters are just remedial solutions. If the modelling accuracy can be improved without much additional costs, the robust solutions are not the first-hand choices. Only when the model cannot be refined with acceptable costs or the performances of the robust filters are satisfactory, the robust filters can be chosen.

The highlights listed above verify the basic philosophy: nothing is free. We cannot expect a method to be satisfactory for all problems.

IV. Further Reading

Generally speaking, statistical signal processing deals with two kinds of problems: signal estimation and signal detection. The former concept corresponds to Parameter Estimation in mathematical and applied statistics, while the latter corresponds to Hypothesis Test. In other words, the two concepts, signal estimation and signal detection, are coined for signal processing practice when we try to use statistical definitions and methods, parameter estimation and hypothesis test, respectively. Likewise, the term Statistical Signal Processing is a conceptual counterpart of Statistics. (Recall that parameter estimation and hypothesis test are two main tasks of statistics.)

When a statistical model is not exact, we are motivated to consider Robust Statistics [2]. Accordingly, we study robust signal processing techniques: robust estimation (e.g., this article) and robust detection [3]. It is proved that robust statistics first identifies the least-favorable distribution from a distributional ambiguity set, and then applies classical statistical techniques over the least-favorable distribution [2], [3]. For example, recall the Huber's M-estimation: it is the maximum likelihood estimation for the least-favorable distribution [4] (Cf. classical estimation: the maximum likelihood estimation for the nominal distribution).

This basic philosophy for robust statistics gives the similarity between robust estimation (e.g., this article) and robust detection [3]. Specifically, robust estimation first explores the least-favorable distribution and then finds optimal estimate for it, whereas robust detection first identifies the least-favorable distributions and then applies optimal detector (i.e., optimal hypothesis tester) over them. Thus, typical distributional ambiguity sets in statistics such as ϵ-contamination sets, f-divergence (a.k.a. ϕ-divergence) sets, Wasserstein sets, are popular (and also natural) in both robust estimation and robust detection communities; Cf. this article and [3]. However, note that for robust estimation and robust detection, implications of and metrics for "least-favorable" are different. Given the same distributional ambiguity set(s), least-favorable distributions for robust estimation are not guaranteed to be the same as those for robust detection, e.g., comparing Lemma 1 of this article with (53) and (54) of [3]. Usually, only one least-favorable distribution is involved in a robust estimation problem. But for a robust detection problem, at least two least-favorable distributions exist (because at least two hypothesis classes are active for testing). Below lists some consensus in robust estimation and robust detection practices. They can be concluded from either (resp. both) this article or (resp. and) [3].

1) Robust solutions perform satisfactorily (but not optimal) for nominal models, and also satisfactorily for models near the nominal models. In contrast, optimal solutions for nominal models deteriorate significantly when true models deviate from nominal ones. The price of the robustness under uncertain conditions is sacrificing the optimality under perfect conditions.
2) Robust solutions are last-hand choices. Robust solutions do not reduce uncertainties but tolerate them. When accuracy of the nominal model can be improved (i.e., uncertainties can be reduced), efforts should be put on more exact modelling. Possible methods include, but are not limited to, adaptively estimating parameters of the model. For example, in state estimation contexts, it helps a lot if we can jointly estimate some parameters of the model when estimating the state; recall unknown-input filters. However, parameters of the nominal model cannot be estimated in all cases; see Subsection VI-A "Frameworks Addressing Parameter Uncertainties" and [3, Sec. III].
3) The distributional ambiguity sets can be neither too large nor too small. An extremely large set makes the robust solution too conservative, while an extremely small set cannot offer sufficient robustness. See Fig. 1 (b) in the main body of this article and [3, Sec. III]. However, the problem of optimally tuning the sizes of distributional ambiguity sets are open because for a real robust statistical signal processing problem, the true signal of interest is unknown (i.e., the training data set is unavailable). Therefore, practitioners can only try appropriate values for their specific robust statistical signal processing problems.

References

[1] L. Sun, W. K. Ho, K. V. Ling, T. Chen, and J. Maciejowski, "Recursive maximum likelihood estimation with t-distribution noise model," *Automatica*, vol. 132, p. 109789, 2021.

[2] P. J. Huber, *Robust Statistics (2nd edition)*. John Wiley & Sons, 2009.

[3] M. Fauß, A. M. Zoubir, and H. V. Poor, "Minimax robust detection: Classic results and recent advances," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2252–2283, 2021.

[4] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.