A Kalman Filter Design Based on the Performance/Robustness Tradeoff

Huan Xu, Shie Mannor
Department of Electrical and Computer Engineering, McGill University,
xuhuan@cim.mcgill.ca, shie.mannor@mcgill.ca

***Abstract*— We consider filter design of a linearly evolving system where the system parameters are subject to uncertainty. In contrast to robust design which focuses on a worst case analysis, we propose a design methodology which aims to achieve a good tradeoff between the nominal performance and robustness to the uncertainty. We prove that the proposed filter achieves similar steady-state stability properties as the robust filter. Simulation results show that unlike the robust filter and the Kalman filter, whose performances can be significantly affected by the problem setting, especially the relative magnitude of the admissible uncertainty, the performance of the proposed filter is suitable for a wider selection of parameters and hence can achieve a more flexible fiter design.**

I. Introduction

The Kalman filter addresses the optimal linear least mean-square-error estimation problem for linearly evolving systems, and is widely used in numerous fields including control, finance, communication and many others since its inception in the early 1960s [1]. Two characteristics of the Kalman filter contribute to its practical success. First, it is calculated in a recursive way with a low computational cost and memory requirement, and therefore is suitable for online application. Second, when the system is observable (a characteristic that can be easily verified), thie filter is guaranteed to converge to a stable steady state filter.

One central assumption of the Kalman filter is that the underlying state-space model is exactly known. In practice, this assumption is often violated, i.e., the parameters we use as the system dynamics (referred as *nominal parameters* hereafter) are only guesses of the unknown true parameters. It is reported [2]–[4] that, the performance of the filter designed based on the nominal parameters can deteriorate significantly, and the steady state estimator can be unstable even for observable systems.

To overcome this sensitivity to the modeling error of the Kalman filter, several so-called "robust" Kalman filters have been designed (e.g., [5]–[9]). Most of these designs require checking of certain existence conditions and hence are not suitable for online-operation. In [10], Sayed proposed a filtering framework based on a worst-case analysis (hereafter referred to as the *robust filter*). Sayed's algorithm is based on the observation that the standard Kalman filter is a minimization of the regularized residual norm at each iteration. Hence, the robust filter is designed to minimize the worst-possible regularized residual norm over the set of admissible uncertainty at each iteration. By doing this, the robust filter circumvents the existence condition checking, and takes a slightly modified recursive form comparing to the standard Kalman filter. Hence it can be implemented in real-time applications. In addition, under certain detectability and stabilizability conditions, the robust filter has desirable stability and bounded error-variance properties.

Empirical study shows that when the uncertainty is relatively large, the performance (measured by the steady-state error variance) of the robust filter is significantly better than the Kalman filter [10]. On the other hand, when the magnitude of the uncertainty is small, the Kalman filter can have a better performance than the robust filter, which we believe is due to the fact that the robust filter takes a worst-case view and hence can be overly conservative. Furthermore, even in cases where the steady-state error of the Kalman filter is larger, it usually has a faster transient response. This is well expected, since achieving robustness implies reducing sensitivity and hence the robust filter can have a slower response. Since both the Kalman filter and the robust filter are good in certain cases, a filter that exhibit a similar performance as the better one under all cases is desirable.

In this paper, we present a new filter design approach which is essentially an interpolation of the Kalman filter and the robust filter. To be more specific, similarly to the standard Kalman filter and the robust filter, the proposed filter is based on finding an one-step "optimal" smoothing filter. The "optimality" is obtained by minimizing the convex combination of the regularized residual norm under the nominal parameter and worst-case regularized residual norm over all admissible parameters. This approach leads to a filter that takes a similar recursive formula as the robust filter and hence is still online solvable. Furthermore, the proposed filter admits a same stability property as the robust filter. Intuitively speaking, this is due to the fact that if the worst case error goes high, the total cost increases, hence the proposed filter upper-bounds the worst-case error variance. Simulation results in Section V show that the proposed filter exhibits a similar performance to the better one between the Kalman filter and the robust filter. That is, when the uncertainty is small where the robust filter can be overly conservative, the performance of proposed filter is similar to the Kalman filter. On the other hand, when the uncertainty is large, which leads to a significant performance deterioration of the Kalman filter, the proposed filter has a similar performance to the robust filter. Therefore, the proposed filter is suitable for a wider range of problem setups compared to the Kalman filter and the robust filter.

The paper is organized as follows. In Section II, we modify the Robust Least Square Problem (RLSP) and show how to solve it. Based on this modified RLSP we give the recursive formula of the proposed filter in Section III. In Section IV we investigate the asymptotical behavior of the proposed filter, i.e., the stability and guaranteed error variance properties. Simulation results and comparisons with standard and robust Kalman filter are shown in Section V. In Section VI we give some concluding remarks.

Notations: Given a column vector z and a positive definite matrix W, $\|z\|_W^2$ stands for $z^\top W z$.

II. The modified Robust Least-Square Problem

In [11], the authors addressed the following bilevel optimization problem[1] on $\mathbf{x}$ which they named as *Robust Least-Square Problem*:

$$\min_{\mathbf{x}}: R(\mathbf{x}) \triangleq \mathbf{x}^\top Q\mathbf{x} + \max_{\|\mathbf{y}\|\le\phi(\mathbf{x})} \left[(A\mathbf{x}-\mathbf{b}+H\mathbf{y})^\top W(A\mathbf{x}-\mathbf{b}+H\mathbf{y})\right], \quad (1)$$

where $\|\cdot\|$ is the Euclidean norm; $Q>0$ and $W\ge 0$ are Hermitian weighting matrices and the function $\phi(\cdot)$ is a non-negative convex function of $\mathbf{x}$ which restricts the choice of $\mathbf{y}$. Although in general, bilevel optimization problems are NP-hard, this special case can be solved efficiently. Sayed showed that, Problem (1) can be converted into a minimization problem on a scalar variable λ. Furthermore, it is also proved that this scalar optimization problem is unimodal and hence a simple line search can solve Problem (1).

Many worst-case analysis problem can be formulated as Problem (1). However, worst-case analysis tends to be overly conservative. To avoid this, we consider the following modified Robust Least-Square Problem which minimizes the convex combination of the cost under the worst condition and the cost under the nominal condition.

$$\min_{\mathbf{x}}: \alpha\Big\{\mathbf{x}^\top Q\mathbf{x} + (A\mathbf{x}-\mathbf{b})^\top W(A\mathbf{x}-\mathbf{b})\Big\} + (1-\alpha)R(\mathbf{x}), \quad (2)$$

where $\alpha\in[0,1]$ is the weighting parameter. We show that, similar to RLSP, Problem (2) can be converted into a unimodal scalar optimization problem and efficiently solved.

We denote the objective function in Problem (2)by $C(\mathbf{x})$, and let $\underline{\lambda}\triangleq\|H^\top WH\|$. Further define the following matrix functions $W(\cdot)$, $\overline{W}(\cdot)$, vector function $\mathbf{x}^o(\cdot)$ and scalar function $G(\cdot)$ of $\lambda\in[\underline{\lambda},+\infty)$.

$$\begin{aligned}
W(\lambda) &\triangleq W + WH(\lambda I - H^\top WH)^\dagger H^\top W,\\
\overline{W}(\lambda) &\triangleq \alpha W + (1-\alpha)W(\lambda),\\
\mathbf{x}^o(\lambda) &\triangleq \arg\min_{\mathbf{x}}\Big\{\mathbf{x}^\top Q\mathbf{x} + (A\mathbf{x}-\mathbf{b})^\top\overline{W}(\lambda)(A\mathbf{x}-\mathbf{b})\\
&\quad + (1-\alpha)\lambda\phi^2(\mathbf{x})\Big\},\\
G(\lambda) &\triangleq \mathbf{x}^{o\top}(\lambda)Q\mathbf{x}^o(\lambda) + (1-\alpha)\lambda\phi^2\big(\mathbf{x}^o(\lambda)\big)\\
&\quad + \big(A\mathbf{x}^o(\lambda)-\mathbf{b}\big)^\top\overline{W}(\lambda)\big(A\mathbf{x}^o(\lambda)-\mathbf{b}\big).
\end{aligned}$$

Here, $(\cdot)^\dagger$ stands for the pseudo inverse of a matrix. The definition of $\mathbf{x}^o(\lambda)$ is valid, because the part in the curled bracket is strictly convex on $\mathbf{x}$. (Note that $Q>0$, $\phi(\cdot)$ is convex by assumption, and $\lambda\ge\underline{\lambda}$ implies $\overline{W}(\lambda)\ge 0$.)

[1] It can also be regarded as a Stackelberg Game

Theorem 1: 1) $G(\lambda)$ has a unique local minimum, and it is also globally minimal.
2) Let $\lambda^o\triangleq\arg\min_{\lambda\ge\|H^\top WH\|}G(\lambda)$, we have

$$\arg\min_{\mathbf{x}} C(\mathbf{x}) = \mathbf{x}^o(\lambda^o);\quad \min_{\mathbf{x}} C(\mathbf{x}) = G(\lambda^o).$$

Proof: Define $R(\mathbf{x},\mathbf{y})\triangleq(A\mathbf{x}-\mathbf{b}+H\mathbf{y})^\top W(A\mathbf{x}-\mathbf{b}+H\mathbf{y})$. Lemma 1 describes the property of $R(\mathbf{x},\mathbf{y})$; its proof can be found in [11].

Lemma 1: (a) The function $\max_{\|\mathbf{y}\|\le\phi(\mathbf{x})}R(\mathbf{x},\mathbf{y})$ is convex on $\mathbf{x}$.
(b) Given $\mathbf{x}$, the following equation holds

$$\max_{\|\mathbf{y}\|\le\phi(\mathbf{x})}R(\mathbf{x},\mathbf{y}) = \min_{\lambda\ge\underline{\lambda}}(A\mathbf{x}-\mathbf{b})^\top W(\lambda)(A\mathbf{x}-\mathbf{b}) + \lambda\phi^2(\mathbf{x}).$$

(c) The following scalar valued function

$$\lambda^o(\mathbf{x})\triangleq\arg\min_{\lambda\ge\underline{\lambda}}(A\mathbf{x}-\mathbf{b})^\top W(\lambda)(A\mathbf{x}-\mathbf{b}) + \lambda\phi^2(\mathbf{x})$$

is well defined and continuous.

Hence

$$\begin{aligned}
&\min_{\mathbf{x}} C(\mathbf{x})\\
&=\min_{\mathbf{x}}\Big\{\mathbf{x}^\top Q\mathbf{x} + \alpha(A\mathbf{x}-\mathbf{b})^\top W(A\mathbf{x}-\mathbf{b}) + (1-\alpha)\max_{\|\mathbf{y}\|\le\phi(\mathbf{x})}R(\mathbf{x},\mathbf{y})\Big\}\\
&=\min_{\mathbf{x},\lambda\ge\underline{\lambda}}\Big\{\mathbf{x}^\top Q\mathbf{x} + (A\mathbf{x}-\mathbf{b})^\top\overline{W}(\lambda)(A\mathbf{x}-\mathbf{b}) + (1-\alpha)\lambda\phi^2(\mathbf{x})\Big\}\\
&=\min_{\lambda\ge\underline{\lambda}}\min_{\mathbf{x}}\Big\{\mathbf{x}^\top Q\mathbf{x} + (A\mathbf{x}-\mathbf{b})^\top\overline{W}(\lambda)(A\mathbf{x}-\mathbf{b}) + (1-\alpha)\lambda\phi^2(\mathbf{x})\Big\}\\
&=\min_{\lambda\ge\|H^\top WH\|}G(\lambda).
\end{aligned}$$

We now show that $G(\cdot)$ is unimodal. Denote

$$H(\mathbf{x},\lambda)\triangleq\mathbf{x}^\top Q\mathbf{x} + (A\mathbf{x}-\mathbf{b})^\top\overline{W}(\lambda)(A\mathbf{x}-\mathbf{b}) + (1-\alpha)\lambda\phi^2(\mathbf{x}).$$

Note that $C(\mathbf{x})$ is strictly convex and goes to infinity whenever $\|\mathbf{x}\|\uparrow\infty$, which implies $C(\mathbf{x})$ is unimodal and has a unique global minimum. Also note, $H(\mathbf{x},\lambda)$ has the following property: fix one variable, then it is a unimodal function of the other variable and achieves unique minimum on its domain. This, combined with the continuity of $\lambda^o(\mathbf{x})$, establishes the unimodality of $G(\cdot)$ by applying Lemma C.2 in [11]. ■

Consider the case when $\phi(\mathbf{x})=\|E_a\mathbf{x}+E_b\|_2$ for some given E_a and E_b, which is of special interest since the smoothing step of the filter proposed in Section III can be converted into this form. Simple algebra yields

$$\mathbf{x}^o(\lambda)=\Big(Q+A^\top\overline{W}A+(1-\alpha)\lambda E_a^\top E_a\Big)^{-1}\Big(A^\top\overline{W}b+(1-\alpha)\lambda E_a^\top E_b\Big),$$

and $\mathbf{x}^o(\lambda^o)$ is the global optimal solution.

III. Filter design

We now consider the design of a state estimator for a linear system where the system parameters are subject to some uncertainties. Our goal is to achieve a good tradeoff between both the nominal performance and the worst case performance. Since the former can be a good indicator of the true performance of the filter and the latter guarantees the robustness of the filter even if the true parameters are far from the nominal ones, the proposed filter can lead to a better performance under the true parameters in many cases.

We consider the following system:

$$\begin{aligned}
\mathbf{x}_{i+1} &= (F_i+\delta F_i)\mathbf{x}_i + (G_i+\delta G_i)\mathbf{u}_i,\quad i\ge 0,\\
\mathbf{y}_i &= H_i\mathbf{x}_i + \mathbf{v}_i.
\end{aligned}$$

Here, the driving noises $\mathbf{u}_i$ and $\mathbf{v}_i$ are assumed to be uncorrelated, zero mean and white, with variance Q_i and R_i respectively. F_i and G_i are the nominal system parameters, and $(\delta F_i, \delta G_i) \in \left\{ (M_i \Delta_i E_{f,i}, M_i \Delta_i E_{g,i}) \middle| \|\Delta_i\| \leq 1 \right\}$, for some known matrix M_i, $E_{f,i}$ and $E_{g,i}$, i.e., we consider the set-inclusive uncertainty $\delta F_i, \delta G_i$ given by linear transformation of norm bounded matrix Δ_i. This formulation is standard in robust filter design [7] [11]. We use $\hat{\mathbf{x}}_{i|j}$ to denote the estimation of $\mathbf{x}_i$ given observation $\{\mathbf{y}_0, \cdots, \mathbf{y}_j\}$, and use $\hat{\mathbf{x}}_i$ to denote $\hat{\mathbf{x}}_{i|i-1}$.

The proposed filter is based on a correcting and propagating process. To be more specific, suppose at time i, let $\hat{\mathbf{x}}_{i|i}$ be the current estimate with error variance $P_{i|i}$, then $\hat{\mathbf{x}}_{i+1|i+1}$ is given by the following two steps:

Smoothing:

$$(\hat{\mathbf{x}}_{i|i+1}, \hat{\mathbf{u}}_{i|i+1}) := \arg\min_{\mathbf{x}_i, \mathbf{u}_i} \Big\{ \alpha \big[\|\mathbf{x}_i - \hat{\mathbf{x}}_{i|i}\|^2_{P^{-1}_{i|i}} + \|\mathbf{u}_i\|^2_{Q_i^{-1}} + \|\mathbf{y}_{i+1} - H_{i+1}\mathbf{x}_{i+1}\|^2_{R^{-1}_{i+1}} \big| \delta F_i = \delta G_i = 0 \big] + (1-\alpha) \max_{\delta F_i, \delta G_i} \big[\|\mathbf{x}_i - \hat{\mathbf{x}}_{i|i}\|^2_{P^{-1}_{i|i}} + \|\mathbf{u}_i\|^2_{Q_i^{-1}} + \|\mathbf{y}_{i+1} - H_{i+1}\mathbf{x}_{i+1}\|^2_{R^{-1}_{i+1}} \big] \Big\}.$$

Predicting:

$$\hat{\mathbf{x}}_{i+1|i+1} := F_i \hat{\mathbf{x}}_{i|i+1} + G_i \hat{\mathbf{u}}_{i|i+1}. \tag{3}$$

The smoothing step finds the optimal one-step-ahead smoothing state estimation and noise estimation, and the prediction step computes the current-step estimation.

Note that, the smoothing part can be converted into a modified RLSP. Using the result in Section II, with some algebra we get the recursion formula in measurement-update form. The readers may find that this is essentially a modified version of the robust filter.

Algorithm 1: **Measurement-Update form**

1) Initialize:

$$\begin{aligned} P_{0|0} &:= (\Pi_0^{-1} + H_0^\top R_0^{-1} H_0)^{-1} \\ \hat{\mathbf{x}}_{0|0} &:= P_{0|0} H_0^\top R_0^{-1} \mathbf{y}_0. \end{aligned}$$

2) Recursion:
Construct and minimize $G(\lambda)$ over $(\|M_i^\top H_{i+1}^\top R_{i+1}^{-1} H_{i+1} M_i\|, +\infty)$. Let the optimal value be λ_i^o. Compute the following values:

$$\begin{aligned} \hat{\lambda}_i &:= (1-\alpha)\lambda_i^o \qquad * \\ \overline{R}_{i+1} &:= R_{i+1} - \lambda^{o\,-1} H_{i+1} M_i M_i^\top H_{i+1}^\top \\ \hat{R}_{i+1}^{-1} &:= \alpha R_{i+1}^{-1} + (1-\alpha)\overline{R}_{i+1}^{-1} \qquad * \\ \hat{Q}_i^{-1} &:= Q_i^{-1} + \hat{\lambda}_i E_{g,i}^\top \left[I + \hat{\lambda}_i E_{f,i} P_{i|i} E_{f,i}^\top \right]^{-1} E_{g,i} \\ \hat{P}_{i|i} &:= (P_{i|i}^{-1} + \hat{\lambda}_i E_{f,i}^\top E_{f,i})^{-1} \\ &= P_{i|i} - P_{i|i} E_{f,i}^\top (\hat{\lambda}_i^{-1} I + E_{f,i} P_{i|i} E_{f,i}^\top)^{-1} E_{f,i} P_{i|i} \\ \hat{G}_i &:= G_i - \hat{\lambda}_i F_i \hat{P}_{i|i} E_{f,i}^\top E_{g,i} \\ \hat{F}_i &:= (F_i - \hat{\lambda}_i \hat{G}_i \hat{Q}_i E_{g,i}^\top E_{f,i})(I - \hat{\lambda}_i \hat{P}_{i|i} E_{f,i}^\top E_{f,i}) \\ P_{i+1} &:= F_i \hat{P}_{i|i} F_i^\top + \hat{G}_i \hat{Q}_i \hat{G}_i^\top \\ R_{e,i+1} &:= \hat{R}_{i+1} + H_{i+1} P_{i+1} H_{i+1}^\top \\ P_{i+1|i+1} &:= P_{i+1} - P_{i+1} H_{i+1}^\top R_{e,i+1}^{-1} H_{i+1} P_{i+1} \\ \hat{\mathbf{x}}_{i+1} &:= \hat{F}_i \hat{\mathbf{x}}_{i|i} \\ \mathbf{e}_{i+1} &:= \mathbf{y}_{i+1} - H_{i+1} \hat{\mathbf{x}}_{i+1} \\ \hat{\mathbf{x}}_{i+1|i+1} &:= \hat{\mathbf{x}}_{i+1} + P_{i+1|i+1} H_{i+1}^\top \hat{R}_{i+1}^{-1} \mathbf{e}_{i+1}. \end{aligned}$$

The formulas with $*$ are the modifications to the recursion formula of the robust filter. In addition, the cost function $G(\lambda)$ and hence its optimal solution λ^o are also different.

Before we rewrite the recursion formula in the prediction form (i.e, propagating $\{\hat{\mathbf{x}}_i, P_i\}$ directly) in Algorithm 2, we need the following two lemma.

Lemma 2:

$$P_{i+1} = F_i P_i F_i^\top - \overline{K}_i \overline{R}_{e,i}^{-1} \overline{K}_i^\top + \hat{G}_i \hat{Q}_i \hat{G}_i^\top, \tag{4}$$

where

$$\overline{K}_i \triangleq F_i P_i \overline{H}_i^\top, \qquad \overline{R}_{e,i} \triangleq I + \overline{H}_i P_i \overline{H}_i^\top,$$
$$\overline{H}_i \triangleq \begin{bmatrix} \hat{R}_i^{-1/2} H_i \\ \sqrt{\hat{\lambda}_i} E_{f,i} \end{bmatrix}.$$

Proof: First note

$$\begin{aligned} P_{i|i}^{-1} &= (P_i - P_i H_i^\top R_{e,i}^{-1} H_i P_i)^{-1} \\ &= \left(P_i - P_i H_i^\top (\hat{R}_i + H_i P_i H_i^\top)^{-1} H_i P_i \right)^{-1} \\ &= \left((P_i^{-1} + H_i^\top \hat{R}_i^{-1} H_i)^{-1} \right)^{-1} = P_i^{-1} + H_i^\top \hat{R}_i^{-1} H_i. \end{aligned} \tag{5}$$

Hence we have

$$\begin{aligned} P_{i+1} &= F_i \hat{P}_{i|i} F_i^\top + \hat{G}_i \hat{Q}_i \hat{G}_i^\top \\ &= F_i \left(P_{i|i}^{-1} + \hat{\lambda}_i E_{f,i}^\top E_{f,i} \right)^{-1} F_i^\top + \hat{G}_i \hat{Q}_i \hat{G}_i^\top \\ &= F_i \left(P_i^{-1} + H_i^\top \hat{R}_i^{-1} H_i + \hat{\lambda}_i E_{f,i}^\top E_{f,i} \right)^{-1} F_i^\top + \hat{G}_i \hat{Q}_i \hat{G}_i^\top \\ &= F_i \left(P_i^{-1} + \overline{H}_i^\top \overline{H}_i \right)^{-1} F_i^\top + \hat{G}_i \hat{Q}_i \hat{G}_i \\ &= F_i P_i F_i^\top - \overline{K}_i \overline{R}_{e,i}^{-1} \overline{K}_i^\top + \hat{G}_i \hat{Q}_i \hat{G}_i^\top. \end{aligned}$$

■

Lemma 3:

$$P_{i|i} H_i^\top \hat{R}_i^{-1} \equiv P_i H_i^\top R_{e,i}^{-1}.$$

Proof: From Equation (5) we have

$$\begin{aligned} P_{i|i}^{-1} (P_i H_i^\top R_{e,i}^{-1}) &= (P_i^{-1} + H_i^\top \hat{R}_i^{-1} H_i)(P_i H_i^\top R_{e,i}^{-1}) \\ &= H_i^\top (I + \hat{R}_i^{-1} H_i P_i H_i^\top) R_{e,i}^{-1} \\ &= H_i^\top \hat{R}_i^{-1} (\hat{R}_i + H_i P_i H_i^\top) R_{e,i}^{-1} = H_i^\top \hat{R}_i^{-1}. \end{aligned}$$

By left multiplying $P_{i|i}$ on both sides, the lemma follows. ■

Algorithm 2: **Prediction form**

1) Initialize:

$$\hat{\mathbf{x}}_0 := 0, \qquad P_0 := \Pi_0, \qquad \hat{R}_0 = R_0.$$

2) Calculating $P_{i|i}$ given $\hat{R}_i, H_i, P_i$:

$$\begin{aligned} P_{i|i} :=& (P_i^{-1} + H_i^\top \hat{R}_i^{-1} H_i)^{-1} \\ =& P_i - P_i H_i^\top (\hat{R}_i + H_i P_i H_i^\top)^{-1} H_i P_i \end{aligned}$$

3) Recursion: Construct and minimize $G(\lambda)$ over $(\|M_i^\top H_{i+1}^\top R_{i+1}^{-1} H_{i+1} M_i\|, +\infty)$. Let the optimal value be λ_i^o. Computing the following values:

$$\begin{aligned} \hat{\lambda}_i :=& (1-\alpha)\lambda_i^o \qquad * \\ \overline{R}_{i+1} :=& R_{i+1} - \lambda^{o-1} H_{i+1} M_i M_i^\top H_{i+1}^\top \\ \hat{R}_{i+1}^{-1} :=& \alpha R_{i+1}^{-1} + (1-\alpha)\overline{R}_{i+1}^{-1} \qquad * \\ \hat{Q}_i^{-1} :=& Q_i^{-1} + \hat{\lambda}_i E_{g,i}^\top \left[I + \hat{\lambda}_i E_{f,i} P_{i|i} E_{f,i}^\top \right]^{-1} E_{g,i} \\ \hat{P}_{i|i} :=& (P_{i|i}^{-1} + \hat{\lambda}_i E_{f,i}^\top E_{f,i})^{-1} \\ =& P_{i|i} - P_{i|i} E_{f,i}^\top (\hat{\lambda}_i^{-1} I + E_{f,i} P_{i|i} E_{f,i}^\top)^{-1} E_{f,i} P_{i|i} \\ \hat{G}_i :=& G_i - \hat{\lambda}_i F_i \hat{P}_{i|i} E_{f,i}^\top E_{g,i} \\ \hat{F}_i :=& (F_i - \hat{\lambda}_i \hat{G}_i \hat{Q}_i E_{g,i}^\top E_{f,i})(I - \hat{\lambda}_i \hat{P}_{i|i} E_{f,i}^\top E_{f,i}) \\ \overline{H}_i^\top :=& \left[H_i^\top \hat{R}_i^{-\top/2} \ \sqrt{\hat{\lambda}_i} \right] \\ \overline{R}_{e,i} :=& I + \overline{H}_i P_i \overline{H}_i^\top \\ \overline{K}_i :=& F_i P_i \overline{H}_i^\top \\ P_{i+1} :=& F_i P_i F_i^\top - \overline{K}_i \overline{R}_{e,i}^{-1} \overline{K}_i^\top + \hat{G}_i \hat{Q}_i \hat{G}_i^\top \\ \mathbf{e}_i :=& \mathbf{y}_i - H_i \hat{\mathbf{x}}_i \\ \hat{\mathbf{x}}_{i+1} :=& \hat{F}_i \hat{\mathbf{x}}_i + \hat{F}_i P_{i|i} H_i^\top \hat{R}_i^{-1} \mathbf{e}_i \\ =& \hat{F}_i \hat{\mathbf{x}}_i + \hat{F}_i P_i H_i^\top R_{e,i}^{-1} \mathbf{e}_i. \end{aligned}$$

IV. Steady-state Analysis

In this section we investigate steady-state features of the proposed filter to demonstrate its robustness. We show that, the proposed filter admits similar stability and bounded error-variance properties as the robust filter.

A. Stability

In this subsection we restrict our discussion to uncertainty models where all parameters are stationary, except Δ_i which is allowed to vary with time. Therefore we drop the subscript i for all stationary parameters. Further assume that $E_{g,i} \equiv 0$, i.e., the uncertainty only appears in F matrix. Hence, we have $\hat{Q} = Q$ and $\hat{G} = G$.[2] In addition, we fix the parameter $\lambda^o \triangleq (1+\beta)\|M^\top H^\top R^{-1} H M\|$ for some positive constant β, i.e., instead of choosing λ^o by optimizing $G(\lambda)$, we approximate it with a time-invariant constant. From the prediction formula, we have the following theorem.

Theorem 2: Assume that $\{F, \overline{H}\}$ is detectable and $\{F, GQ^{1/2}\}$ is stabilizable. Then, for any initial condition $\Pi_0 > 0$, the Riccati variable P_i converges to the unique solution of

$$P = FPF^\top - FP\overline{H}^\top (I + \overline{H} P \overline{H}^\top)^{-1} \overline{H} P F^\top + GQG^\top. \quad (6)$$

Furthermore, the solution P is semi-definite positive, and the steady state closed loop matrix $F_p \triangleq \hat{F}[I - PH^\top R_e^{-1} H]$ is stable.

[2]This setting is same as that of the robust filter.

Proof: Note that, for $\beta > 0$, $\overline{R}$ is strictly positive definite. Hence $\hat{R}$ is also strictly positive definite. The remaining follows a same line as the proof of Theorem 2 in [10]. We defer the detailed proof to the appendix. ■

B. Bounded Steady-State Error Variances

Further assume the system is quadratically stable, i.e, there exists a matrix $V > 0$ such that

$$V - [F + M\Delta E_f]^\top V [F + M\Delta E_f] > 0, \quad \|\Delta\| \leq 1.$$

Define

$$\mathcal{F} \triangleq \begin{bmatrix} F - F_p PH^\top \hat{R}^{-1} H & F - F_p - F_p PH^\top \hat{R}^{-1} H \\ F_p PH^\top \hat{R}^{-1} H & F_p + F_p PH^\top \hat{R}^{-1} H \end{bmatrix},$$

and

$$\mathcal{G} \triangleq \begin{bmatrix} G & -F_p PH^\top \hat{R}^{-1} H \\ 0 & F_p PH^\top \hat{R}^{-1} H \end{bmatrix}.$$

Denote

$$\hat{\mathcal{F}} \triangleq \mathcal{F} + \begin{bmatrix} M \\ 0 \end{bmatrix} \Delta [E_f \ \ E_f].$$

The following result is the same as Theorem 3 of [10]. Again we postpone the proof to the Appendix.

Theorem 3: Let $\tilde{x}_i$ be the estimation error, suppose a positive matrix $\mathcal{P}$ satisfies

$$\mathcal{P} - \hat{\mathcal{F}} \mathcal{P} \hat{\mathcal{F}}^\top - \mathcal{G} \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \mathcal{G}^\top \geq 0, \quad \forall \|\Delta\| \leq 1.$$

Then

$$\lim_{i \to \infty} \mathbb{E} \tilde{x}_i \tilde{x}_i^\top \leq \mathcal{P}_{11},$$

where $\mathcal{P}_{11}$ is the $(1,1)$ block entries of $\mathcal{P}$. Furthermore, such $\mathcal{P}$ is guaranteed to exist.

V. Simulation Study

In this section, we investigate the performance of the proposed filter and compare it to the robust filter and Kalman filter in three different parameter setups. These setups differ in the ratio between the nominal parameter and the uncertainty. We also investigate the performance of different α for different ratio setting.

We use the following numerical example, which is also used in [7] and [10], to compare the performance of the proposed filter with the robust filter and the standard Kalman filter:

$$\begin{aligned} \mathbf{x}_{i+1} &= (F + M\Delta_i E_f)\mathbf{x}_i + \mathbf{u}_i, \\ \mathbf{y}_i &= H\mathbf{x}_i + \mathbf{v}_i, \end{aligned}$$

where

$$F = \begin{bmatrix} 0.9802 & 0.0196 \\ 0 & 0.9802 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad H = [1 \ \ -1],$$

$$M = \begin{bmatrix} 0.0198 \\ 0 \end{bmatrix}, \quad E_f = [\ 0 \quad 5\ ], \quad E_g = [\ 0 \quad 0\ ],$$

$$R = 1, \quad Q = \begin{bmatrix} 1.9608 & 0.0195 \\ 0.0195 & 1.9608 \end{bmatrix}, \quad \mathbf{x}_0 \sim N(\mathbf{0}, I).$$

Observe that Δ_i is a scalar in this example, and we have

$$F + M\Delta_i E_f = \begin{bmatrix} 0.9802 & 0.0196 + 0.099\Delta_i \\ 0 & 0.9802 \end{bmatrix}.$$

We note that, in this case, the uncertainty only affects the $(1, 2)$ entry of matrix F. Furthermore, we notice that the nominal parameter and the uncertainty are of the same order of magnitude. The tradeoff parameter α for the proposed filter is set as 0.8 in the simulation.

The empirical average error variance is used to compare the performance of different filters. To be specific, 500 trajectories with 1000 steps each are generated. For the j-th trajectory, the observation series $\{\mathbf{y}_i^{(j)}\}$ is then filtered by a particular algorithm to get the state estimation series $\{\hat{\mathbf{x}}_i^{(j)}\}$. The expected error variance curve is approximated by

$$\mathbb{E}\|\mathbf{x}_i - \hat{\mathbf{x}}_i\|^2 \approx \frac{1}{500}\sum_{j=1}^{500}\|\mathbf{x}_i^{(j)} - \hat{\mathbf{x}}_i^{(j)}\|^2. \quad i = 1, \cdots, 1000.$$

In Figure 1(a), the uncertainty Δ is generated according to a uniform distribution in $[-1,\ 1]$, and is fixed for the whole trajectory. In Figure 1(b), the uncertainty is allowed to vary at each iteration, i.e., in each iteration, Δ is re-generated (according to a uniform distribution). In both cases, the proposed filter exhibits a similar steady-state performance to the robust filter, and a better transient performance. The transient performance of the Kalman filter is similar to the proposed filter, but its steady-state error variance for stationary uncertainty is about 2 dB worse. We also observe that, for the non-stationary case, both Kalman filter and the proposed filter achieve a better performance compared to the robust filter, probably due to the fact that time varying uncertainties cancel out.

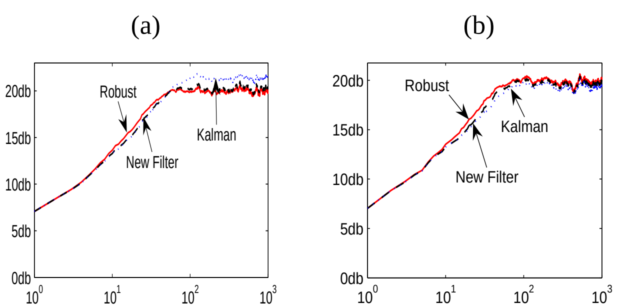

Fig. 1. Error variance curves: (a) fixed uncertainty; (b) time-varying uncertainty.

In Figure 2, we simulate the case where the uncertainty is enlarged, i.e., we set $M = col\{0.1980,\ 0\}$. Note that, in this case, the uncertainty is one order of magnitude larger than the nominal parameter. In [10], the authors observed that in such situation, the performance of the Kalman filter degraded significantly, which is also shown by our simulation result. In contrast to the standard Kalman filter, the steady-state error of the proposed filter is only slightly worse (about 1 dB) than that of the robust filter in the fixed uncertainty case, and is comparable to that of the Robust filter in the time-varying case. This shows that by adding the worst-case performance into the cost function to minimize, the proposed filter overcomes the sensitivity of the standard Kalman filter and achieves a comparable robustness as the robust filter.

In Figure 3, we investigate the case where the uncertainty

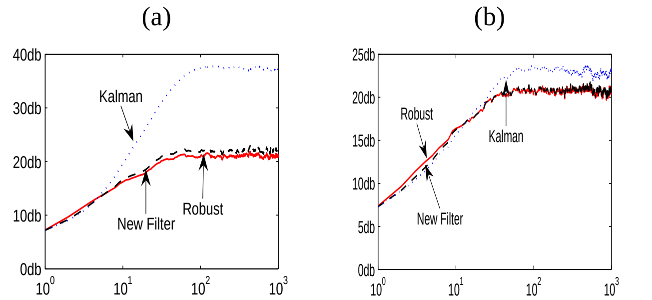

Fig. 2. Error variance curves for large uncertainty: (a) fixed uncertainty; (b) time-varying uncertainty.

is relatively small. To be specific, we set

$$F = \begin{bmatrix} 0.9802 & 0.3912 \\ 0 & 0.9802 \end{bmatrix}.$$

The $(1, 2)$ entry of the original F is enlarged 20 times, which means the uncertainty is one order smaller than the nominal parameter in this case. The robust filter achieves a steady-state error variance around 22 dB, while both the Kalman filter and the proposed filter achieves a steady-state error around 16 dB. This shows that the Robust filter could be overly conservative and achieve unsatisfactory performance when uncertainty is comparatively small, whereas the proposed new filter does not suffer from such conservativeness.

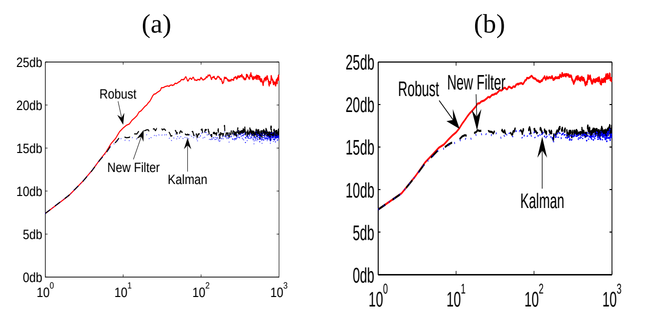

Fig. 3. Error variance curves for large nominal value: (a) fixed uncertainty; (b) time-varying uncertainty.

To investigate the relationship between the performance of the filter and the value of α, we simulate the steady-state error variance for $\alpha \in \{0,\ 0.2,\ 0.4,\ 0.6,\ 0.8,\ 0.99,\ 1\}$ [3] under different ratio setting $\gamma \in \{0.01, 0.05, 0.25, 1, 4, 20, 100\}$. Here, $\gamma = 1$ is the original numerical example. For $\gamma > 1$ we increase the magnitude of the uncertainty M by γ, whereas for $\gamma < 1$ we increase the magnitude of the nominal parameter $F_{1,2}$ by $1/\gamma$. From Figure 4, we see that, when γ is small, (i.e., uncertainty is relatively small), larger α achieves better performance. The smallest steady-state error is achieved by $\alpha = 1$, i.e., the standard Kalman filter. This shows that for relatively small uncertainty, focusing on robustness itself can degrade the performance of the filter. On the other hand, for large γ which stands for large uncertainty, the steady-state error for standard Kalman filter is huge. In contrast, we notice that even for $\alpha = 0.99$ which means the robust part plays a

[3] $\alpha = 0$ stands for a Robust filter, and $\alpha = 1$ stands for a Kalman filter.

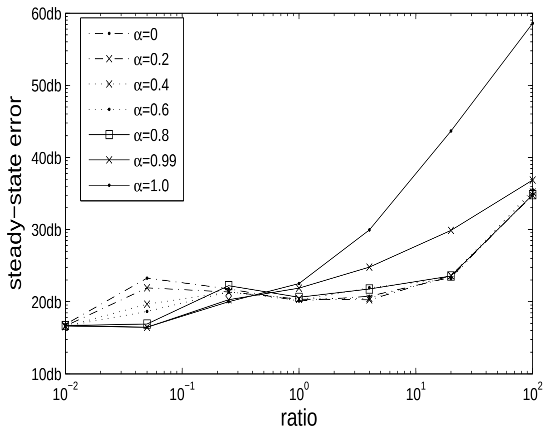

Fig. 4. Effect of α on steady-state error.

very small part in the objective function, the proposed filter achieves a much better performance. And for $\gamma \leq 0.8$ the proposed filter achieves a similar performance as the robust filter. The overall most-balanced filter is achieved by taking $\alpha = 0.8$, which has a nearly optimal performance for both small γ and large γ.

In short, the simulation study shows that the performances of both the Kalman filter and the robust filter is sensitive to the relative magnitude of the uncertainty. In contrast, in all three cases, the proposed filter exhibits a performance comparative to the better one, and therefore is suitable for a wider range of problems.

VI. Concluding Remarks

In this paper, we presented a new algorithm for state estimation of a linear system where the parameters are not exactly known. This filter is based on minimizing the weighted sum of the worst possible residual and the nominal residual. The resulting recursive form has a computational cost comparable to the robust filter, and can be easily implemented on-line. We show that, under certain technical conditions, the proposed filter converges to a stable steady-state estimator. Besides, the proposed filter achieves guaranteed bounded error variance if the underlying system is stable.

The performance of the proposed filter is compared with that of the robust filter and the Kalman filter. A simulation study shows that the proposed filter overcome both the sensitivity of the Kalman filter and the overly conservativeness of the robust filter, and hence achieves good performance under a wider range of parameters.

The main thrust of the proposed approach is to obtain more flexibility in filter design. As the simulation study showed, the performance of both the Kalman filter and the robust filter can varied significantly under different parameter settings. That is, they can behave rather poor in a non-suitable problem. Whether a problem setting is suitable for these two filters is not known beforehand, except a general guildline that small uncertainty favors Kalman filter and large uncertainty favors robust filter. Furthermore, the problem parameter itself can be varying. The proposed filter is of interest in the sense that its performance does not critically depend on the problem setting, and still behaves reasonably good even in a adversarial case.

References

[1] A. H. Sayed T. Kailath and B. Hassibi. *Linear Estimation*. Prentice-Hall, 2000.
[2] A. Gelb, editor. *Applied Optimal Estimation*. MIT Press, 1974.
[3] H. W. Sorenson, editor. *Kalman Filtering: Theory and Application*. IEEE Press, 1985.
[4] M. S. Grewal and A. P. Andrews. *Kalman Filtering: Theory and Practice*. Prentice-Hall, 1993.
[5] P. P. Khargonekar and K. M. Nagpal. Filtering and smoothing in an h_∞-setting. *IEEE Trans. on Automatic Control*, 36:151–166, 1991.
[6] T. Basar and P. Bernhard. *H_∞-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach.* Birkhauser, 1991.
[7] I. R .Petersen and A. V. Savkin. *Robust Kalman Filtering for Signals and Syustems with Large Uncertainties*. Birkauser, 1999.
[8] D. P. Bertsekas and I. B. Rhodes. Recursive state estimation for a set-membership description of uncertainties. *IEEE Trans. on Automatic Control*, 16:117–128, 1971.
[9] L. E. Ghaoui and G. Calafiore. Robust filtering for discrete-time systems with bounded noise and parametric uncertaitny. *IEEE Trans. on Automatic Control*, 46(7):1084–1089, 2001.
[10] A. H. Sayed. A framework for state-space estimation with uncertain models. *IEEE Trans. on Automatic Control*, 46(7):998–1013, July 2001.
[11] V. H. Nascimento A. H. Sayed and F. A. Cipparrone. A regularized robust design criterion for uncertain data. *SIAM Journal on Matrix Analysis and Its Applications*, 23(4):1120–1142, 2002.

Appendix

Appendix

Proof of Theorem 2

Proof: First, from the filtering recursion, we have the following closed loop formula for $\hat{\mathbf{x}}$

$$\begin{aligned}\hat{\mathbf{x}}_{i+1} =& \hat{F}_i\hat{\mathbf{x}}_i + \hat{F}_i P_i H^\top R_{e,i}^{-1}[\mathbf{y}_i - H\hat{\mathbf{x}}_i] \\ =& \hat{F}_i[I - P_i H^\top R_{e,i}^{-1} H]\hat{\mathbf{x}}_i + \hat{F}_i P_i H^\top R_{e,i}^{-1}\mathbf{y}_i.\end{aligned}$$

Now consider the closed loop gain

$$\begin{aligned}
&F_{p,i} \triangleq \hat{F}_i[I - P_i H^\top R_{e,i}^{-1} H] \\
=&F\left[I - \hat{\lambda}\big(P_i^{-1} + H^\top \hat{R}^{-1} H + \hat{\lambda} E_f^\top E_f\big)^{-1} E_f^\top E_f\right]\left[I - P_i H^\top R_{e,i}^{-1} H\right] \\
=&F\left[I - \hat{\lambda}(P_i^{-1} + \overline{H}^\top \overline{H})^{-1} E_f^\top E_f\right]\left[I - P_i H^\top R_{e,i}^{-1} H\right] \\
=&F(P_i^{-1} + \overline{H}^\top \overline{H})^{-1}\left[P_i^{-1} + \overline{H}^\top \overline{H} - \hat{\lambda} E_f^\top E_f\right]\left[I - P_i H^\top R_{e,i}^{-1} H\right] \\
=&F(P_i^{-1} + \overline{H}^\top \overline{H})^{-1}(P_i^{-1} + H^\top \hat{R}^{-1} H)\left[P_i - P_i H^\top R_{e,i}^{-1} H P_i\right] P_i^{-1} \\
=&F(P_i^{-1} + \overline{H}^\top \overline{H})^{-1}(P_i^{-1} + H^\top \hat{R}^{-1} H) \\
&\times \left[P_i - P_i H^\top (\hat{R}_i + H P_i H^\top)^{-1} H P_i\right] P_i^{-1} \\
=&F(P_i^{-1} + \overline{H}^\top \overline{H})^{-1} P_i^{-1},
\end{aligned}$$

and

$$\begin{aligned}
&F\left[I - P_i \overline{H}^\top \overline{R}_{e,i}^{-1} \overline{H}\right] \\
=&F\left[P_i - P_i \overline{H}^\top \overline{R}_{e,i}^{-1} \overline{H} P_i\right] P_i^{-1} \\
=&F\left[P_i - P_i \overline{H}^\top (I + \overline{H} P_i \overline{H}^\top)^{-1} \overline{H} P_i\right] P_i^{-1} \\
=&F(P_i^{-1} + \overline{H}^\top \overline{H})^{-1} P_i^{-1}.
\end{aligned}$$

Thus, we have

$$\hat{F}_i[I - P_i H^\top R_{e,i}^{-1} H] = F\left[I - P_i \overline{H}^\top \overline{R}_{e,i}^{-1} \overline{H}\right].$$

Note the positive definiteness of $\hat{R}$ guarantees that $\overline{H}$ is well defined. Hence, detectability of $\{F, \overline{H}\}$ and the stablizability of $\{F, GQ^{1/2}\}$ guarantee that P_i converges to the unique positive semi-definite solution P of Equation (6), which stabilizes the matrix

$$F[I - P\overline{H}^\top (I + \overline{H} P \overline{H}^\top)^{-1} \overline{H}].$$

Since this matrix equals to the steady state closed loop gain F_p, the stability is established. ■

Proof of Theorem 3

Proof: Define estimation error $\tilde{\mathbf{x}}_i \triangleq \mathbf{x}_i - \hat{\mathbf{x}}_i$, and

$$\delta\mathcal{F}_i \triangleq \begin{bmatrix} M\Delta_i E_f & M\Delta_i E_f \\ 0 & 0 \end{bmatrix}.$$

Hence the extended state equation holds:

$$\begin{bmatrix} \tilde{\mathbf{x}}_{i+1} \\ \hat{\mathbf{x}}_{i+1} \end{bmatrix} = (\mathcal{F} + \delta\mathcal{F}_i) \begin{bmatrix} \tilde{\mathbf{x}}_i \\ \hat{\mathbf{x}}_i \end{bmatrix} + \mathcal{G} \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix}. \tag{7}$$

Introduce a similarity transformation:

$$\mathcal{T} \triangleq \begin{bmatrix} I & I \\ 0 & I \end{bmatrix}, \qquad \mathcal{T}^{-1} = \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix}.$$

We have,

$$\mathcal{T}(\mathcal{F}+\delta\mathcal{F}_i)\mathcal{T}^{-1} = \begin{bmatrix} F & 0 \\ F_p P H^\top \hat{R}^{-1} H & F_p \end{bmatrix} + \begin{bmatrix} M\Delta_i E_f & 0 \\ 0 & 0 \end{bmatrix}.$$

Hence the first part (i.e., the nominal matrix, denote as $\tilde{\mathcal{F}}$) is stable since F and F_p are stable.

Furthermore, the following equality

$$E_f(zI - F)^{-1}M = [E_f \;\; 0](zI - \tilde{\mathcal{F}})^{-1} \begin{bmatrix} M \\ 0 \end{bmatrix},$$

shows that the extended system has a same $\mathcal{H}_\infty$-norm as the original system, and hence the extended system is quadratically stable. Hence, there exists a positive definite matrix $\mathcal{V}$ such that

$$\mathcal{V} - (\mathcal{F} + \delta\mathcal{F}_i)\mathcal{V}(\mathcal{F} + \delta\mathcal{F}_i)^\top > 0.$$

Thus, by scaling $\mathcal{V}$ large enough, there exists a positive $\mathcal{P}$ such that

$$\mathcal{P} \geq (\mathcal{F} + \delta\mathcal{F}_i)\mathcal{P}(\mathcal{F} + \delta\mathcal{F}_i)^\top + \mathcal{G} \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \mathcal{G}^\top. \tag{8}$$

Let

$$\mathcal{M}_i \triangleq \mathbb{E} \begin{bmatrix} \tilde{\mathbf{x}}_i \\ \hat{\mathbf{x}}_i \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_i \\ \hat{\mathbf{x}}_i \end{bmatrix}^\top,$$

then the following recursion formula holds

$$\mathcal{M}_{i+1} = (\mathcal{F} + \delta\mathcal{F}_i)\mathcal{M}_i(\mathcal{F} + \delta\mathcal{F}_i)^\top + \mathcal{G} \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \mathcal{G}^\top. \tag{9}$$

Subtracting Equation (9) from Equation (8) we get

$$\mathcal{P} - \mathcal{M}_{i+1} = (\mathcal{F} + \delta\mathcal{F}_i)(\mathcal{P} - \mathcal{M}_i)(\mathcal{F} + \delta\mathcal{F}_i)^\top + \mathcal{Q}_i,$$

for some $\mathcal{Q}_i \geq 0$. The quadratic stability of $\mathcal{F} + \delta\mathcal{F}_i$ implies that $\mathcal{P} - \mathcal{M}_\infty \geq 0$. ■