Minimax Adaptive Estimation for Finite Sets of Linear Systems

Olle Kjellqvist and Anders Rantzer

Abstract— For linear time-invariant systems with uncertain parameters belonging to a finite set, we present a purely deterministic approach to multiple-model estimation and propose an algorithm based on the minimax criterion using constrained quadratic programming. The estimator tends to learn the dynamics of the system, and once the uncertain parameters have been sufficiently estimated, the estimator behaves like a standard Kalman filter.

I. INTRODUCTION

A. Problem Statement

In this article, we consider output prediction for linear systems of the form

\[
x_{t+1} = Fx_t + Gu_t + w_t
\]

\[
y_t = Hx_t + v_t, \quad 0 \leq t \leq N - 1,
\]

where \( x_t \in \mathbb{R}^n \), \( u_t \in \mathbb{R}^p \) and \( y_t \in \mathbb{R}^m \) are the states and the measured input and output at time-step \( t \), respectively. \( w_t \in \mathbb{R}^n \) and \( v_t \in \mathbb{R}^m \) are unmeasured process disturbance and measurement noise. The model, \((F,H,G)\) is fixed but unknown, belonging to some finite set

\[
\{(F_1,H_1,G_1), \ldots, (F_K,H_K,G_K)\}.
\]

of triplets of real-valued matrices. In particular, we are interested in strictly causal estimation of \( y_N \), such that the gain from disturbance trajectories \((w_t,v_t)_{t=0}^{N-1}\) to pointwise estimation error \( (y_N - Hx_N) \) in some weighted \( \ell_2 \)-norm is bounded by a constant \( \gamma_N > 0 \). This means that given positive definite matrices \( P_0 \in \mathbb{R}^{n \times n}, R \in \mathbb{R}^{m \times m} \) and \( Q \in \mathbb{R}^{n \times n} \) and a nominal value of the initial state, \( \hat{x}_0 \),

\[
|\hat{y}_N - Hx_N|^2 \leq \gamma_N^2,
\]

should hold for all disturbances and models compatible with the measurement history \((y_t,u_t)_{t=0}^{N-1}\). This approach is different from the Bayesian approach to filtering where one takes the conditional expectation as the estimate \( \hat{y}_N \). The interest in worst-case gain is motivated by robust feedback-control from estimates. In such settings instability or lack of performance due to model errors is a larger concern than robustness to outliers.

We are grateful to our colleague Dr. Carolina Bergeling for comments that greatly improved the manuscript. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme under grant agreement No 834142 (ScalableControl).

Both authors are with the Department of Automatic Control, Lund University, Lund, Sweden olle.kjellqvist@control.lth.se, anders.rantzer@control.lth.se

B. Background

Simultaneous estimation of states and parameters in linear systems is a bilinear estimation problem. The Maximum-likelihood approach leads to estimates which cannot be put in recursive form and must be obtained by iteration [1]. A recursive method can be obtained by parametrizing the dynamical equations and the observer and learning the parameters using the sequential prediction error approach. Alternatively, one can augment the state vector with the uncertain parameters and apply nonlinear filtering methods such as the Extended Kalman filter [2]. Unfortunately, optimality guarantees for such methods are difficult to obtain. One exception is when the system can be modeled as a finite set of linear systems and the noise is Gaussian, then the Maximum-likelihood estimates can be put on a recursive form [3].

Solutions based on the multiple-model approach have been tremendously successful in modeling and estimating complex engineering systems. In essence, it consists of two parts: 1) design simpler models for a finite set of possible operating regimes. 2) Run a filter for each model and cleverly combine the estimates. Multiple-model adaptive estimation has been around since the ’60s [4], [5] and has been an active research field since. The estimation approach easily extends to systems where the active model can switch (hybrid systems) by matching a Kalman filter with each possible trajectory. In that case, the number of filters will grow exponentially, which has sparked research into more efficient methods. Notable numerically tractable and suboptimal algorithms for estimation in hybrid systems are the Generalized Pseudo Bayesian [6], [7], and the Interacting Multiple Model [8]. The algorithms have been coupled with extended and unscented Kalman filters to deal with non-linear systems [9], and [10] studied robustness to identification error. In [11], the authors pointed out that methods based on Kalman filters are sensitive to noise distributions and proposed an Interactive Multiple Model algorithm based on particle filters to handle non-Gaussian noise at the expense of a 100 fold increase in computation. Recently, machine-learning approaches to classification have been combined with the Interacting Multiple Model estimator [12], [13] and showed improved accuracy in simulations.

The Bayesian approach to the Multiple-model estimation problem involves assigning probability distributions to disturbances \((w_t,v_t)\) and models \((F,G,H)\). The estimate is taken as the expected value of \( y_N \) conditioned on past measurements. If the disturbances are zero-mean and Gaussian, then the conditional expectation can be computed as the weighted average of Kalman filter estimates (one for each model),
weighted by the conditional probability that its model is active.

It is evident in practice that the estimator’s performance depends on the quality of the model set. The models must be distinguishable using measured signals, and the models should accurately describe the operating regimes. Since the estimates can be susceptible to non-Gaussian noise, it is surprising that deterministic approaches similar to those studied by the control community in the ’80s and ’90s have gathered little attention. Recent progress to minimax adaptive control of linear systems with uncertain parameters belonging to a finite set [14] under the assumption of perfect measurements has inspired this research into compatible estimation techniques.

C. Contribution

In this paper, we formulate the multiple-model estimation problem as a deterministic, two-player dynamic game. In particular, this formulation allows for online computation of the worst-case gain from disturbances to estimation error and tractable synthesis of suboptimal estimators that minimize the worst-case gain. Deterministic dynamic games have played a key role in solving and understanding $\mathcal{H}_\infty$ filtering [15], [16]; our goal in this work has been to take a first step towards extending the advantages of that framework to the multiple model setting.

D. Outline

The outline is as follows: First, we introduce notation in Section II, then we introduce minimax multiple-model filtering and the main results in Section III. In Section IV, we present a simplified form for time-invariant systems. In Section II, then we introduce minimax multiple-model filtering and the main results in Section III. In Section IV, we present a simplified form for time-invariant systems. We illustrate the theory through a numerical example in Section V. Section VI contains concluding remarks, and supporting lemmata are given in the Appendix.

II. NOTATION

The set of $n \times m$-dimensional matrices with real coefficients is denoted $\mathbb{R}^{n \times m}$. The transpose of a matrix $A$ is denoted $A^\top$. For a symmetric matrix $A \in \mathbb{R}^{n \times n}$, we write $A \succ (\succeq) 0$ to say that $A$ is positive (semi)definite. Given $x \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$, $|x|_A^2 := x^\top A x$. For a vector $x_t \in \mathbb{R}^n$ we denote the sequence of such vectors up to time $t$ by $x^t := (x_k)_{k=0}^t$.

III. MINIMAX MULTIPLE MODEL FILTERING

In contrast to the Bayesian approach, our approach is fully deterministic; similarly to [15], [16], we do not make explicit assumptions on the distribution of the noise trajectories $w_t$ and $v_t$. We will instead construct a two-player dynamic game between a minimizing player that chooses the estimate, and a maximizing player that chooses dynamics and disturbances. Recall that we are interested in characterizing an estimator $\hat{y}_N$ such that the gain from disturbances to the pointwise estimation error is bounded by $\gamma_N$. I.e., (2) holds for all disturbances consistent with (1) and the data $(y^{N-1}, u^{N-1})$.

Since the disturbances are unknown, we cannot evaluate (2) directly. However, define

$$J_N(y^{N-1}, u^{N-1}, \hat{y}_N)$$

weighted by the conditional probability that its model is active. It is evident in practice that the estimator's performance depends on the quality of the model set. The models must be distinguishable using measured signals, and the models should accurately describe the operating regimes. Since the estimates can be susceptible to non-Gaussian noise, it is surprising that deterministic approaches similar to those studied by the control community in the ’80s and ’90s have gathered little attention. Recent progress to minimax adaptive control of linear systems with uncertain parameters belonging to a finite set [14] under the assumption of perfect measurements has inspired this research into compatible estimation techniques.

C. Contribution

In this paper, we formulate the multiple-model estimation problem as a deterministic, two-player dynamic game. In particular, this formulation allows for online computation of the worst-case gain from disturbances to estimation error and tractable synthesis of suboptimal estimators that minimize the worst-case gain. Deterministic dynamic games have played a key role in solving and understanding $\mathcal{H}_\infty$ filtering [15], [16]; our goal in this work has been to take a first step towards extending the advantages of that framework to the multiple model setting.

D. Outline

The outline is as follows: First, we introduce notation in Section II, then we introduce minimax multiple-model filtering and the main results in Section III. In Section IV, we present a simplified form for time-invariant systems. In Section II, then we introduce minimax multiple-model filtering and the main results in Section III. In Section IV, we present a simplified form for time-invariant systems. We illustrate the theory through a numerical example in Section V. Section VI contains concluding remarks, and supporting lemmata are given in the Appendix.

II. NOTATION

The set of $n \times m$-dimensional matrices with real coefficients is denoted $\mathbb{R}^{n \times m}$. The transpose of a matrix $A$ is denoted $A^\top$. For a symmetric matrix $A \in \mathbb{R}^{n \times n}$, we write $A \succ (\succeq) 0$ to say that $A$ is positive (semi)definite. Given $x \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$, $|x|_A^2 := x^\top A x$. For a vector $x_t \in \mathbb{R}^n$ we denote the sequence of such vectors up to time $t$ by $x^t := (x_k)_{k=0}^t$.

III. MINIMAX MULTIPLE MODEL FILTERING

In contrast to the Bayesian approach, our approach is fully deterministic; similarly to [15], [16], we do not make explicit assumptions on the distribution of the noise trajectories $w_t$ and $v_t$. We will instead construct a two-player dynamic game between a minimizing player that chooses the estimate, and a maximizing player that chooses dynamics and disturbances. Recall that we are interested in characterizing an estimator $\hat{y}_N$ such that the gain from disturbances to the pointwise estimation error is bounded by $\gamma_N$. I.e., (2) holds for all disturbances consistent with (1) and the data $(y^{N-1}, u^{N-1})$.

Since the disturbances are unknown, we cannot evaluate (2) directly. However, define

$$J_N(y^{N-1}, u^{N-1}, \hat{y}_N)$$

where the maximization is performed subject to the constraints (1). Then (2) holds if and only if

$$J_N(y^{N-1}, u^{N-1}, \hat{y}_N) \leq 0.$$

In this setting, $w_t = x_{t+1} - F x_t - G u_t$ and $v_t = y_t - H x_t$ are uniquely determined by the states, the measurements and the active model. Inserting into (3), we get

$$J_N(y^{N-1}, u^{N-1}, \hat{y}_N)$$

holds, where $\hat{y}_N$ are functions of past data $y^{N-1}$ and $u^{N-1}$. This constitutes a two-player dynamic game and would be linear quadratic if not for the model being chosen by the maximizing player. The intuition behind (5) makes sense in the following way. The minimizing player is penalized for selecting the true (noisless) output, and the maximizing player is penalized for selecting a model which requires large disturbances $w$ and $v$ to be compatible with the data. As $N$ increases, the penalty for selecting a model different from the truth grows too large, resulting in a learning mechanism. It turns out that the cost associated with the disturbance trajectories required to explain each model should accurately describe the operating regimes. Since the disturbances are unknown, we cannot evaluate (2) directly. However, define

$$J_N(y^{N-1}, u^{N-1}, \hat{y}_N)$$

where the maximization is performed subject to the constraints (1). Then (2) holds if and only if

$$J_N(y^{N-1}, u^{N-1}, \hat{y}_N) \leq 0.$$
and assume that $H_i P_{N,i} H_i^\top \prec \gamma^2_N I$. Then the cost (4) is equivalent to

$$J_N(y_{N-1}, u_{N-1}, \hat{y}_N) = \max_i \left\{ |\hat{y}_N - H_i \tilde{x}_{N,i}|^2 (1 - \gamma^2_N H_i P_{N,i} H_i^\top)^{-1} - \gamma^2_N c_{N,i} \right\}.$$  

(6)

$\tilde{x}_{N,i}$ is the Kalman filter estimate of $x_N$ using the $i$th model, and $c_{N,i}$ are generated according to

$$\begin{align*}
\tilde{x}_{0,i} &= x_0 \\
\tilde{x}_{t+1,i} &= F_t \tilde{x}_{t,i} + K_t (y_t - H_t \tilde{x}_{t,i}) + G_t u_t \\
K_t &= F_t P_t H_t^\top (R + H_t P_t H_t^\top)^{-1} \\
c_{0,i} &= 0 \\
c_{t+1,i} &= |H_t \tilde{x}_{t,i} - y_t|^2 (R + H_t P_t H_t^\top)^{-1} + c_{t,i}.
\end{align*}$$

Proof. We will perform the maximization over state-trajectories in (4) in two steps. First over past trajectories $(x^{N-1})$ and then over the future state $x_N^1$. The right-hand side of (4) becomes

$$\sup_{x^{N-1},i} \left\{ |\hat{y}_N - H_i x_N|^2 - \gamma^2_N \inf_{x^{N-1}} \left\{ |x_0 - \hat{x}_0|^2_{P_{0,i}} \\
\quad + \sum_{t=0}^{N-1} \left( |x_{t+1} - F_t x_t - G_t u_t|^2_{Q_{t-1}} + |y_t - H_t x_t|^2_{R_{t-1}} \right) \right\} \right\},$$

where $i = 1, \ldots, K$ is an index for the active model $(F_t, H_t, G_t)$. Apply Lemma 4 to get

$$J_N(y_{N-1}, u_{N-1}, \hat{y}_N) = \sup_{x^{N-1},i} \left\{ |\hat{y}_N - H_i x_N|^2 - \gamma^2_N \left( |x_0 - \hat{x}_0|^2_{P_{0,i}} \\
\quad + \sum_{t=0}^{N-1} \left( |x_{t+1} - F_t x_t - G_t u_t|^2_{Q_{t-1}} + |y_t - H_t x_t|^2_{R_{t-1}} \right) \right) \right\}.$$  

For fix $\hat{y}_N$ and $i$, the assumption $H_i P_{N,i} H_i^\top \prec \gamma^2_N I$ guarantees that we maximize a concave function of $x_N$ and we apply Lemma 5 with $A = H_i$, $X = I$, $Y = P_{N,i}$ to conclude,

$$J_N(y_{N-1}, u_{N-1}, \hat{y}_N) = \max_i |\hat{y}_N - H_i \tilde{x}_{N,i}|^2 (1 - \gamma^2_N H_i P_{N,i} H_i^\top)^{-1} - \gamma^2_N c_{N,i}.$$  

\hfill $\Box$

Remark 1. Theorem 1 holds also for time-varying systems, if $F_t$ and $H_t$ are replaced by $F_{t,i}$ and $H_{t,i}$. Further, $P_t, Q$ and $R$ can be time-varying and differ between models.

Remark 2. Equation (6) is monotonically increasing in $\gamma_N$ and the smallest $\gamma_N^*$ such that $J_N(y_{N-1}, u_{N-1}, \hat{y}_N) \leq 0$ can be found efficiently through bisection.

The below Corollary follows from Theorem 1 and describes how to compute the minimax estimator as a convex quadratic program.

$$\begin{align*}
\max_{y_{N-1}^i} \{ \ldots \} &= \max_{y_{N-1}^i} \{ \max_{y_{N-1}^i} \{ \ldots \} \}. \\
\text{The maximizing argument is given by } x_N^1(\hat{y}_N, i) &= (H_i^\top H_i - \gamma^2_N P_{N,i})^{-1}(H_i^\top \hat{y}_N - P_{N,i} \tilde{x}_{N,i}).
\end{align*}$$

Corollary 2. With assumptions as in Theorem 1, consider the convex program

$$\begin{align*}
\text{minimize} & \quad t \\
\text{subject to:} & \quad |\hat{y}_N - H_i \tilde{x}_{N,i}|^2 (1 - \gamma^2_N H_i P_{N,i} H_i^\top)^{-1} - \gamma^2_N c_{N,i} \leq t \\
& \quad \forall i = 1, \ldots, K.
\end{align*}$$

The minimizing argument $\hat{y}_N^*$ satisfies (5).

Remark 3. If the model set is a singleton, then $\hat{y}_N^* = H x_N = H \tilde{x}_N$ is the estimate generated by the Kalman filter, which is a well known result [16].

A. On $c_{N,i}$ and the relation to conditional probability.

It is known (see for instance [3]) that if $u_t$ and $v_t$ are uncorrelated Gaussian white noise with covariances $Q$ and $R$, the conditional probability that the measured output $y_N$ has been generated by the model $(F_t, G_t, H_t)$ and the input $u_N$ can be expressed as

$$p(i|y_N, u_N) = \frac{\alpha_N e^{-|y_N - H_i \tilde{x}_{N,i}|^2_{P_{N,i}}}}{\det(2\pi R_{N,i})^{1/2}}.$$  

\alpha_N is some normalization constant independent of $i$, and

$$R_{N,i} = R + H_i P_{N,i} H_i^\top,$$

with $P_{N,i}$ as in Theorem 1. Taking $c_{N,i}$ as in Theorem 1 we see that the conditional probability is proportional to $e^{c_{N,i}}$,

$$p(i|y_N, u_N) \propto e^{-c_{N,i}} \prod_{t=1}^{N} \det(2\pi R_{t,i})^{-1/2}.$$  

IV. STATIONARY SOLUTION

For a set of time-invariant systems, we summarize a simple version of the filter in the below theorem.

Theorem 3. Consider matrices $F_1, \ldots, F_K \in \mathbb{R}^{n \times n}$, $H_1, \ldots, H_K \in \mathbb{R}^{m \times n}$, and positive definite $Q, P_0 \in \mathbb{R}^{n \times n}, R \in \mathbb{R}^{m \times m}$. Assume that the algebraic Riccati equations

$$P_i = Q + F_i (P_i - P_i H_i H_i^\top (R + H_i P_i H_i^\top)^{-1} H_i P_i) F_i^\top,$$

have solutions $H_i P_i H_i^\top \prec \gamma^2_N I$. Then a minimax strategy $\hat{y}_N^*$ for the game defined by

$$\min_{\hat{y}_N} \max_{x^{N-1},i} \left\{ |\hat{y}_N - H_i x_N|^2 - \gamma^2_N |x_0 - \hat{x}_0|^2_{P_{0,i}} \\
\quad - \gamma^2_N \sum_{t=0}^{N-1} \left( |x_{t+1} - F_t x_t - G_t u_t|^2_{Q_{t-1}} + |y_t - H_t x_t|^2_{R_{t-1}} \right) \right\},$$

and (1), is the minimizing argument

$$\min_{\hat{y}_N} \left\{ |\hat{y}_N - H_i \tilde{x}_{N,i}|^2 (1 - \gamma^2_N H_i P_{N,i} H_i^\top)^{-1} - \gamma^2_N c_{N,i} \right\}.$$
\( \hat{x}_{N,i} \) is the Kalman filter estimate of \( x_N \) using the \( i \)th model, and \( c_{N,i} \) are generated according to

\[
\hat{x}_{0,i} = x_0 \\
\hat{x}_{t+1,i} = F_i \hat{x}_{t,i} + K_i (y_t - H_i \hat{x}_{t,i}) + G_i u_t \\
K_i = F_i P_{i,N} H_i^T (R + H_i P_{i,N} H_i^T)^{-1} \\
c_{0,i} = 0 \\
c_{t+1,i} = |H_i \hat{x}_{t,i} - y_t|^2 (R + H_i P_{i,N} H_i^T)^{-1} + c_{t,i}.
\]

**Proof.** This is a special case of Theorem 1, by replacing \( P_0 \) with \( P_i \).

V. EXAMPLE

In this example, we compare a minimax estimator synthesized using Corollary 2, bisecting over \( \gamma_N \), to find the estimator \( \hat{y}_N \) such that (2) is satisfied for the smallest possible \( \gamma_N \). We compare this to a Bayesian multiple-model estimator [3] and calculate the corresponding bound \( \gamma_N \) using Theorem 1 and bisection. Consider the uncertain linear system

\[
x_{t+1} = F x_t + w_t, \quad y_t = x_t + v_t, \quad F \in \{-1, 1\}.
\]

The weights in (2) are chosen to be \( Q = R = P_0 = 1 \). We generate data \( y^{N-1} \) by simulating the system with \( F = 1 \) and \( w_t, v_t \) as independent Gaussian white noise with intensity 1. For \( N = 5 \) we find

\[
P_{5,1} = P_{5,-1} = 1.62, \\
\hat{x}_{5,1} = -2.34, \quad \hat{x}_{5,-1} = 1.50, \\
c_{5,1} = 3.56, \quad c_{5,-1} = 8.11.
\]

In Fig. 1, we illustrate (6) for \( N = 5 \) and the estimates. Note that \( \gamma = 1.51 \) can be guaranteed for the minimax estimator, but not the Bayesian. Fig. 2 contains a comparison between the smallest \( \gamma_N \) so that (2) can be guaranteed for the minimax estimator and the Bayesian estimator when \( N = 1 \ldots 20 \).

VI. CONCLUSIONS

We stated the minimax criterion for output prediction, where the dynamics belong to a finite set of linear systems and proposed a minimax estimation strategy. The strategy can be implemented as a convex program, and the resulting estimate is a weighted interpolation of Kalman filter estimates. We showed in a numerical example how to apply the theoretical results to compute the worst-case gain from disturbances to error for any multi-model estimation algorithm online and how to generate estimates that minimize the said gain.

By running a minimax estimator in parallel to another estimator, we can measure the worst-case performance level of the other estimator. A large difference in performance levels indicates that the nominal estimator may be highly sensitive to errors in the noise model.

Predetermining the smallest achievable gain from disturbances to estimation errors is still an open research problem, that is, finding necessary and sufficient conditions such that

\[
\sup_{y^{N-1}} J_N^* (y^{N-1}, u^{N-1}) \leq 0.
\]

In future work, we plan to develop a Multi-model adaptive estimator with a prescribed \( \ell_2 \)-gain bound from disturbance to error and methods for infinite sets of linear systems.

APPENDIX — SUPPORTING LEMMATA

Lemma 4. The cost function

\[
V_{N,i}(x_N, y^{N-1}) = \min_{x^{N-1}} \left\{ \frac{1}{2} x_0 \right\} + \sum_{k=1}^{N-1} \left( |x_{t+1} - F_i x_t - G_i u_t|^2 + |y_t - H_i x_t|^2 R_{t-1} \right)
\]

(7)
under the dynamics (1), is of the form

\[ V_{t,i}(x, y_{t-1}) = |x - \bar{x}_{t,i}|^2_{P_{t,i}} + c_{t,i}, \]

where \( P_{t,i} \) and \( c_{t,i} \) are generated as

\[
P_{0,i} = P_0 \\
P_{t+1,i} = Q + F_t P_{t,i} F_t^T - F_t P_{t,i} H_t^{-1} (R + H_t P_{t,i} H_t^T)^{-1} H_t P_{t,i} F_t^T \\
\bar{x}_{0,i} = x_0 \\
\bar{x}_{t+1,i} = \bar{F}_t \bar{x}_{t,i} + K_{t,i}(y_t - H_t \bar{x}_{t,i}) + G_t u_t \\
K_{t,i} = F_t P_{t,i} H_t^{-1} (R + H_t P_{t,i} H_t^T)^{-1} \\
c_{0,i} = 0 \\
c_{t+1,i} = |H_t \bar{x}_{t,i} - y_{t}|^2_{(R + H_t P_{t,i} H_t^T)^{-1}} + c_{t,i}.
\]

**Proof.** The proof builds on forward dynamic programming [17], and is similar to one given in [18] but differs in the assumption that \( F_t \) is not invertible. Further, the constant terms \( c_{t,i} \) are explicitly computed. The cost function \( V_N^3 \) can be computed recursively:

\[
V_I(x, y^I) = |x - x_0|_{P_0}^2 \\
V_{t+1}(x, y^I) = \min_\xi |x - F_t \xi - G_t u_t|_{Q_{t-1}}^2 + |y_t - H_t \xi|_{R_{t-1}}^2 + V_t(\xi, y^{t-1}).
\]  

(8)

With a slight abuse of notation, we assume a solution of the form \( V_I(x) = |x - \bar{x}_t|_{P_{t-1}}^2 + c_t \) and solve for the minimum

\[
V_{t+1}(x) = \min_\xi |x - G_t u_t|^2_{Q_{t-1}} + |\xi|^2_{F_t^T Q_{t-1} F_t + H_t^T R_{t-1} H_t + P_{t-1}} - 2(F_t^T Q_{t-1}(x - G_t u_t) + H_t^T R_{t-1} y_t + P_{t-1} \bar{x}_t)^T \xi \\
+ |y_{t}|_{R_{t-1}}^2 + |\bar{x}|_{P_{t-1}}^2.
\]

Assume at this stage \( S_t := F_t^T Q_{t-1} F_t + H_t^T R_{t-1} H_t + P_{t-1} \succ 0 \), then the minimizing \( \xi^* \) is a stationary point

\[
\xi^* = S_t^{-1}(F_t^T Q_{t-1}(x - G_t u_t) + H_t^T R_{t-1} y_t + P_{t-1} \bar{x}_t)
\]

and the resulting partial cost

\[
|x - \bar{x}_{t+1}|_{P_{t-1}}^2 + c_{t+1} = |x - G_t u_t|_{Q_{t-1}}^2 + |y_{t}|_{R_{t-1}}^2 + |\bar{x}_{t+1}|^2_{P_{t-1}} - |F_t^T Q_{t-1}(x - G_t u_t) + H_t^T R_{t-1} y_t + P_{t-1} \bar{x}_t|_{S_t}^2 + c_t.
\]

(10)

Since this should hold for arbitrary \( x \) and

\[
x - \bar{x}_{t+1} = (x - G_t u_t) - (\bar{x}_{t+1} - G_t u_t),
\]

we get

\[
P_{t+1} = Q - Q^{-1} F S_{t}^{-1} F^T Q^{-1} \\
\bar{x}_{t+1} = G_t u_t = P_{t+1} Q^{-1} F S_{t}^{-1} (H_t^T R_{t-1} y_t + P_{t-1} \bar{x}_t)
\]

The expression for calculating \( P_{t+1} \) can be further simplified using the Woodbury identity,

\[
P_{t+1} = (Q + F_t H_t^T R_{t-1} H_t + P_{t-1})^{-1} F_t^T Q^{-1} \\
P_{t+1} = Q + F_t P_{t} F_t^T - F_t P_{t} H_t^T (R + H_t P_{t} H_t^T)^{-1} H_t P_{t} F_t^T,
\]

where we used the Woodbury matrix identity twice. Inserting these expressions into (10), applying the Woodbury matrix identity to \( S_{t}^{-1} F_t^T (Q - F S_{t}^{-1} F_t^T) - S_{t}^{-1} = (S_t - F_t^T Q^{-1} F_t) - (H_t^T R_{t-1} H_t + P_{t-1})^{-1} \)

\[
c_{t+1} = \begin{cases} 
\left|H_t \bar{x}_t - y_t\right|^2_{(R + H_t P_{t} H_t^T)^{-1}} + c_t \\
\left|H_t \bar{x}_t - y_t\right|^2_{(R + H_t P_{t} H_t^T)^{-1}} + c_t 
\end{cases}
\]

Next we show that \( \bar{x} \) can be formulated as a state-observer

\[
\bar{x}_{t+1} - G u_t = P_{t+1} Q^{-1} F S_{t}^{-1} (H_t^T R_{t-1} y_t + P_{t-1} \bar{x}_t) \\
= P_{t+1} Q^{-1} F S_{t}^{-1} (H_t^T R_{t-1} y_t + P_{t-1} \bar{x}_t) \\
+ P_{t+1} Q^{-1} F S_{t}^{-1} (H_t^T R_{t-1} H_t + P_{t-1})^{-1} \bar{x}_t
\]

Use the matrix inversion lemma \( (A + BCD)^{-1} BC = A^{-1} B (C + DA^{-1} B)^{-1} \).

\[
P_{t+1} Q^{-1} F S_{t}^{-1} \bar{x}_t \\
= -(Q^{-1} + Q^{-1} F S_{t}^{-1} F^T (Q^{-1})^{-1} Q^{-1} F S_{t}^{-1}) \\
= -(Q^{-1} + (Q^{-1} F) (S_t - F_t^T Q^{-1} F_t)^{-1}) \\
= F (H_t^T R_{t-1} H_t + P_{t-1})^{-1}.
\]

Insert in to the previous expression and conclude

\[
\bar{x}_{t+1} = F \bar{x}_t + K_t (y_t - H \bar{x}_t) + G_t u_t,
\]

where

\[
K_t = F P_{t} H_t^T (R + H_t P_{t} H_t^T)^{-1}
\]

**Lemma 5.** For \( x \in \mathbb{R}^n, v, y \in \mathbb{R}^m \), a non-zero matrix \( A \in \mathbb{R}^{n \times m} \), positive-definite matrices \( X \in \mathbb{R}^{n \times n} \) and \( Y \in \mathbb{R}^{m \times m} \), and a positive real number \( \gamma_N > 0 \) such that

\[
A^T X^{-1} A - \gamma_N^2 Y^{-1} \prec 0,
\]

it holds that

\[
\max_v \{|x - Av|_{X^{-1}}^2 - \gamma_N^2 |y - v|_{Y^{-1}}^2\} = \max_v \{|x - Ay|_{(X - \gamma_N^{-2} AYA^T)^{-1}}^2\}. \tag{11}
\]

**Proof.** Expanding the left-hand side of (11) and equating the
gradient with $0$ we get
\[
\max_v \left\{ |x - Av|^2_{X^{-1}} - \gamma_N^2 |y - v|^2_{Y^{-1}} \right\}
= \max_v \left\{ |v|^2_{A^TX^{-1}A - \gamma_N^2 Y^{-1}} - 2v^T(A^TX^{-1}x - \gamma_N^2 Y^{-1})y \right\}
= |x|_{X^{-1}}^2 - \gamma_N^2 |y|_{Y^{-1}}^2
- |A^TX^{-1}x - \gamma_N^2 Y^{-1}|((A^TX^{-1}A - \gamma_N^2 Y^{-1})^{-1}A^TX^{-1}x
+ |y|^2_{\gamma_N^2 Y^{-1}} - \gamma_N^2 Y^{-1}((A^TX^{-1}A - \gamma_N^2 Y^{-1})^{-1}Y^{-1}\gamma_N^2 Y^{-1})
- 2x^T X^{-1} A (A^TX^{-1}A - \gamma_N^2 Y^{-1})^{-1} (-\gamma_N^2 Y^{-1})y
= |x|^2_{(X - \gamma_N^2 AY A^T)^{-1}} + |Ay|^2_{(X - \gamma_N^2 AY A^T)^{-1}} - 2x^T (X - \gamma_N^2 AY A^T)^{-1} Ay
= |x - Ay|^2_{(X - \gamma_N^2 AY A^T)^{-1}}.
\]

REFERENCES

[1] Y. Bar-Shalom, “Optimal simultaneous state estimation and parameter identification in linear discrete-time systems,” IEEE Transactions on Automatic Control, vol. 17, no. 3, pp. 308–319, 1972.

[2] G. C. Goodwin and K. S. Sin, Adaptive filtering prediction and control / Graham C. Goodwin and Kwai Sang Sin., ser. Prentice-Hall information and system sciences series. Englewood Cliffs, N.J.: Prentice-Hall, 1984.

[3] I. L. Crassidis and J. L. Junkins, Optimal Estimation of Dynamic Systems, Second Edition (Chapman & Hall/CRC Applied Mathematics & Nonlinear Science), 2nd ed. Chapman & Hall/CRC, 2011.

[4] D. Magill, “Optimal adaptive estimation of sampled stochastic processes,” IEEE Transactions on Automatic Control, vol. 10, no. 4, pp. 434–439, 1965.

[5] D. G. Lainiotis, “Partitioning: A unifying framework for adaptive systems, i: Estimation,” Proceedings of the IEEE, vol. 64, no. 8, pp. 1126–1143, 1976.

[6] G. Ackerson and K. Fu, “On state estimation in switching environments,” IEEE Transactions on Automatic Control, vol. 15, no. 1, pp. 10–17, 1970.

[7] C. B. Chang and M. Athans, “State estimation for discrete systems with switching parameters,” IEEE Transactions on Aerospace and Electronic Systems, vol. AES-14, no. 3, pp. 418–425, 1978.

[8] H. A. P. Blom and Y. Bar-Shalom, “The interacting multiple model algorithm for systems with markovian switching coefficients,” IEEE Transactions on Automatic Control, vol. 33, no. 8, pp. 780–783, 1988.

[9] A. Akca and M. Önder Efe, “Multiple model kalman and particle filters and applications: A survey,” IFAC-PapersOnLine, vol. 52, no. 3, pp. 73–78, 2019, 15th IFAC Symposium on Large Scale Complex Systems LSS 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405896319300977

[10] K. Xiong, C. Wei, and L. Liu, “Robust multiple model adaptive estimation for spacecraft autonomous navigation,” Aerospace Science and Technology, vol. 42, pp. 249–258, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1270963815000371

[11] G. Ronghua, Q. Zheng, L. Xiangnan, and C. Junliang, “Interacting multiple model particle-type filtering approaches to ground target tracking,” Journal of Computers, vol. 3, 07 2008.

[12] D. Li, P. Zhang, and R. Li, “Improved IMM algorithm based on XGBoost,” Journal of Physics: Conference Series, vol. 1748, p. 032017, Jan 2021. [Online]. Available: https://doi.org/10.1088/1742-6596/1748/3/032017

[13] L. Deng, D. Li, and R. Li, “Improved IMM algorithm based on RNNs,” Journal of Physics: Conference Series, vol. 1518, p. 012055, Apr 2020. [Online]. Available: https://doi.org/10.1088/1742-6596/1518/1/012055

[14] A. Rantzer, “Minimax adaptive control for a finite set of linear systems,” 2021.

[15] X. Shen and L. Deng, “Game theory approach to discrete $H_\infty$ filter design,” IEEE Transactions on Signal Processing, vol. 45, no. 4, pp. 1092–1095, 1997.

[16] T. Basar and P. Bernhard, $H_\infty$-Optimal Control and Related Minimax Design Problems — A dynamic Game Approach. Birkhauser, 1995.

[17] H. Cox, “On the estimation of state variables and parameters for noisy dynamic systems,” IEEE Transactions on Automatic Control, vol. 9, no. 1, pp. 5–12, 1964.

[18] G. C. Goodwin, J. A. De Dona, and M. M. Seron, Constrained Control and Estimation — An Optimization Approach. Springer-Verlag, 2005.