Simultaneous state and input estimation

with applications in vehicle problems

Grigorios Gakis

Department of Engineering
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Magdalene College June 2022

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Grigorios Gakis
June 2022

Acknowledgements

This thesis is the result of several years of work and it would not have been possible without the support of colleagues and friends. I would first and foremost like to thank my supervisor Prof. Malcolm Smith. His love for the subject is an inspiration and his devotion to asking fundamental questions a guiding example. I am very grateful to McLaren Automotive Limited ("McLaren Automotive") for the support of an industrial CASE studentship and providing invaluable simulation data. In particular, I would like to thank Mark Salmon and Simon Lacey from McLaren Automotive for helping to shape the directions this project has developed in and generously sharing insights and ideas. I would also like to thank all the members of the Control Group at the University of Cambridge for an extremely welcoming and stimulating environment.

It has been a great pleasure to share these exciting years with friends from the university. I wish to thank in particular my close friends Dimitris, Theo and Fotis for our countless and most exciting discussions in engineering and beyond. I also wish to acknowledge friends from Magdalene and other colleges including Vangelis, Ilias, Petros, Pedro, Nikita, Nana, Ashley, Sophie, Harri, Abbie, Giles, Millie and Silvia. I am most grateful to my sister Katerina, parents Sofia and Miltos, grandmother Vasoula and relatives Petros, Danai, Mina for their love, care, support and encouragement. Thank you!

Abstract

Engineering is about design of materials, structures, systems and interconnections to obtain a desired behaviour. In control engineering the focus is often on systems equipped with sensors whose output is used to provide feedback control and achieve the desired behaviour. A central paradigm in control is the separation principle, that is the optimal control action is achieved by applying the control law that is optimal under an assumption of full information to an "optimal estimate" of the system state. While in the context of linear-quadratic problems there is a well developed theory of optimal estimation and control, research on systems with inputs that are not measured directly is still ongoing. Motivated by automotive applications where it is not always feasible or practical to have sensors that measure all vehicle inputs, we aim to advance the theory on simultaneous state and input estimation and apply it to commercially important automotive examples.

In particular, we formulate a deterministic estimation problem to find the input and state of a linear continuous time dynamical system which minimises a weighted integral squared error between the resulting output and the measured output. A completion of squares approach is used to find the unique optimum in terms of the solution of a Riccati differential equation. The optimal estimate is obtained from a two-stage procedure that is reminiscent of the Kalman filter. The first stage is an end-of-interval estimator for the finite horizon which may be solved in real time as the horizon length increases. The second stage computes the unique optimum over a fixed horizon by a backwards integration over the horizon. A related tracking problem is solved in an analogous manner. Making use of the solution to both the estimation and tracking problems a constrained estimation problem is solved which shows that the Riccati equation solution has a least squares interpretation that is analogous to the meaning of the covariance matrix in stochastic filtering. We show that the estimation and tracking problems considered here include the Kalman filter and the linear quadratic regulator as special cases. The infinite horizon case is also considered for both the estimation and tracking problems. Stability and convergence conditions are provided and the optimal solutions are shown to take the form of left inverses of the original system.

Motivated by the intrinsically discrete nature of operation of modern computers and sensors, we then focus on systems in which the output is measured only at a discrete

sequence of times. We derive two forms for the zero informational input limit for the discrete time Kalman filter in the case that there is direct feedthrough (of full column rank) of the input to the measurements. The first form is complementary to a zero informational limit filter derived recently by Bitmead, Hovd, and Abooshahab for the case where the first Markov parameter has full column rank and there is no direct feedthrough of the process noise. This form of the limit filter is closely related to a filter proposed by Gillijns and De Moor who used a constrained optimisation problem to estimate an unknown input in a standard Kalman filter with feedthrough of the unknown disturbance; more precisely, the filters coincide if the process noise covariance in Gillijns and De Moor is set to zero. A second form of the limit filter is derived from the first which takes the form of a standard Kalman filter without unknown inputs. This form is used to derive necessary and sufficient conditions for convergence and stability of the filter. These consist of a controllability condition and a minimum phase condition.

We consider a deterministic estimation problem to find the input and state of a linear continuous time dynamical system with discrete time measurements which minimises a weighted sum squared error between the resulting output and the measured output. Similarly to the estimation problem with continuous time measurements we use a completion of squares approach to find the unique optimum. The optimal estimate is obtained in terms of the solution of a Riccati difference equation from a two-stage procedure that is reminiscent of the discrete time Kalman filter. The first stage is an end-of-interval estimator for the finite horizon which may be solved in real time as the horizon length increases. The second stage computes the unique optimum over a fixed horizon by a backwards recursion over the horizon. The infinite horizon case is also considered and stability conditions are provided. We apply the algorithm to two automotive examples, the first is on (offline) road elevation mapping and the second is on (online) slip estimation. In both examples we assume that the vehicle is equipped with basic sensors (e.g. accelerometers and gyroscopes (IMU), global positioning system (GPS)) and use very simple vehicle models.

Table of contents

List of figures

Chapter 1

Introduction

1.1 Motivation and scope

Over the past century powerful mathematical techniques have been developed for the analysis and control of dynamical systems. They are frequently used to design information systems (e.g. algorithms in computer code) due to the wide availability of data (e.g. through sensing devices) and computational power. They have been applied to a plethora of problems across diverse fields ranging from engineering to physics, biology, medicine and economics. We are primarily motivated by automotive design problems, in particular we are interested in understanding the dynamic behaviour of a vehicle, mapping its environment and designing feedback control laws for high performance, safety and autonomous features. Central to those goals is the formulation of estimation problems which fuse information from multiple vehicle on-board sensors to estimate vehicle states and inputs. More specifically, we are most interested in the estimation of longitudinal and lateral tyre slips, the road-tyre coefficient of friction, aerodynamic loads and road profile mapping. We consider basic vehicle on-board sensors including accelerometers and gyroscopes (IMU), a global positioning system (GPS), wheel rotation sensors and drivetrain torque sensors. Important vehicle exogenous inputs (e.g. tyre forces, slips and road profile inputs) are very difficult to measure directly but they nevertheless appear in direct feedthrough terms in typical sensor measurements (e.g. IMU). Therefore, we wish focus on problems of simultaneous input and state estimation in dynamical systems where a direct measurement of the driving inputs is not feasible. Furthermore, we wish to understand the limits of what is possible given the available vehicle sensors and generate estimates that are reliable at all times on a high performance vehicle, especially in highly dynamic situations at the limits of the available road grip. We will develop algorithms that can be applied both online (e.g. for vehicle control) and offline (e.g.

for data analysis) that exploit our knowledge of the dynamics of the system (e.g. Newton's second law and kinematics of the vehicle, empirical tyre model).

There is already an extensive literature devoted to model based estimation of states and inputs. Within this literature there is a wide variety of problem formulations and assumptions: deterministic or stochastic, discrete or continuous time, whether there is feedthrough of inputs to outputs, and whether real time estimates of the inputs are sought. In deterministic formulations an optimisation problem is posed based on the available data and its solution corresponds to the estimated states and inputs. Stochastic formulations treat the states and inputs as random variables with known prior statistics and the estimates are given by their posterior expectations. A close connection between the solutions of the two formulations for a quadratic cost function and Gaussian statistics was observed from early on. Estimation solutions are also closely connected to control solutions even though the respective problems appear upon first thought conceptually distant. More specifically, in least squares optimisation results in estimation can be transformed and applied to control and vice versa. The mathematical duality between estimation and control and the close connection between deterministic and stochastic estimation formulations are central in existing literature and will influence the development of the ideas within this thesis.

1.2 Thesis outline

We will now briefly summarise the contents of the thesis. Starting in Section 2.1 we introduce some basics on the standard theory of linear quadratic estimation, present a few applications of historical significance and review the latest literature on systems with unknown inputs. In Section 2.2 we introduce the stochastic problem formulations of discrete time Kalman filtering and smoothing, derive their solutions and present some standard theory on the Riccati equation. In Section 2.3 we present the extended Kalman filter (EKF) and unscented Kalman filter (UKF) which are extensions of the Kalman filter to nonlinear problems and discuss Monte Carlo methods in the context of filtering and smoothing.

In Chapter 3 we initially formulate a deterministic optimisation problem for the simultaneous estimation of states and inputs in a continuous time system (Section 3.2) and then formulate and solve the dual control problem (Section 3.3). In Section 3.4 we pose a new constrained optimisation problem which requires the state to pass through a prescribed point at a given time. This allows us to interpret solutions of Riccati equations deterministically as weights which determine how well state positions fit the data. In Section 3.5 we show that the standard Kalman filter and linear-quadratic regulator (LQR) solutions are special cases of the estimation and tracking problems of Sections 3.2 and 3.3. In Sections 3.6 and 3.7 we consider infinite horizon limits for both estimation (i.e.steady state filtering) and control (i.e. infinite horizon control) problems.

In Chapter 4 we focus on systems in which the output is measured only at a discrete sequence of times. In Section 4.1 we consider the zero informational input limit of the discrete time Kalman filter for systems with full column rank feedthrough to the output. The filter takes a recursive form which closely relates to the filter of Gillijns and De Moor [20] but is derived directly by computing a limit. In Section 4.2 we consider a deterministic problem formulation for continuous time systems in which we impose a zero hold assumption on the input and then derive the optimal solution. We propose a heuristic algorithm to extend the results to nonlinear systems. We derive necessary and sufficient conditions for the convergence and stability of both filters in Chapter 4.

In Chapter 5 we implement the algorithm of Section 4.2 to two vehicle examples. The first is on (offline) road profile mapping in Section 5.1 and the second is on (online) slip estimation in Section 5.2. In both examples we assume that the vehicle is equipped with basic on-board sensors (e.g. accelerometers and gyroscopes (IMU), global positioning system (GPS)) and use very simple vehicle models (i.e. quarter car model, vehicle bicycle model). In Chapter 6 we discuss the contributions of the thesis and interesting future research directions.

Chapter 2

Preliminaries

2.1 Introduction

2.1.1 Linear quadratic estimation

The close connection between the deterministic concept of estimation by least squares and the minimisation of a mean square error in a statistical sense has long been appreciated (see [49], [54], [57], [33]). Kalman followed earlier work of Wiener and Kolmogorov in taking a statistical view of signal estimation, though his approach specified a linear state-space model (see Fig. 2.1) for the system and estimation of the state rather than estimation of a signal from a noisy measurement with assumptions on the spectral characteristics. The resulting filter was initially introduced for discrete time systems [37] and then for continuous time systems first by analogy with optimal control [36] and later as a limit of the discrete time equations [39]. The discrete (continuous) time filter has the form of a recursive algorithm (differential equation) which allows for easy implementation in real time applications and a lot of its success can be attributed to this property. Very important is Kalman's work [38] in deriving conditions for the stability of time invariant systems. A (deterministic) least squares formulation and derivation of the continuous time Kalman filter was given recently by Willems [60]. As noted in [60] the possibility of such a derivation had been "system theory folklore" ever since the first appearance of Kalman's work and he provides a number of earlier references in this direction. The work by Willems inspired the analogous study in discrete time systems in [7] which focuses exclusively on minimal systems.

Over the last decades the Kalman filter has proven to be a powerful and useful tool in a range of problems. An early and historically significant application is in spacecraft tracking and guidance for the Apollo Project by NASA in the 1960s. It is through this project that the Extended Kalman Filter (EKF) for nonlinear systems was developed and

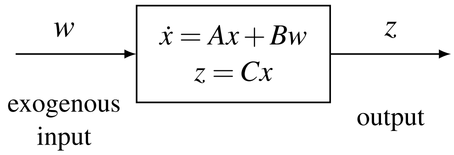

Fig. 2.1 A linear dynamical system with state x, exogenous input w and output z.

subsequently became standard practice in the aerospace industry and sensor system design [1]. Nowadays commercial aircraft are equipped with attitude and heading reference systems (AHRS) which integrate sensor data to estimate and display aircraft attitude, roll, pitch and yaw on the electronic flight instrument system [2]. The Global Positioning System (GPS) deployed by the U.S. Department of Defence also uses a Kalman filter with a large state vector that includes the states of satellites, clocks and parameters relating to signal propagation delays [1]. Interestingly, the signals from the GPS are used along with an inertial navigation system (INS) (i.e. accelerometers and gyroscopes) daily for navigation applications in vehicles, ships and smartphones. The combination of GPS and INS highlights the ability of the Kalman filter to optimally fuse information from different sources. More specifically, the INS produces accurate estimates that slowly drift over time while the GPS has the complimentary properties [3]. In the automotive industry the Kalman filter has been applied for slip control in anti-lock braking systems (ABS). For example, the PhD thesis [4] developed a UKF to estimate sideslips for the braking systems of heavy goods vehicles (HGVs) motivated by the significantly longer stopping distances of HGVs compared to passenger cars. The use of the filter improved both the stopping distance and directional stability in combined braking and cornering situations. Other examples of implementations can be found in the areas of weather forecasting, pollution estimation, seismology, time series analysis and prediction in economics and trading, noise filtering and enhancement of images and speech, model identification, simultaneous localisation and mapping, state estimation of power systems, orbit determination of planets and satellites, battery state of charge (SoC) estimation, autopilot systems, object tracking in computer vision, medical diagnosis and human neural decoding.

2.1.2 Systems with unknown inputs

An active branch of study that emerged following Kalman's initial contribution focused on estimation in systems with unknown inputs. Noise-free systems were considered first and observers were designed for systems with a full column rank first Markov parameter and zero feedthrough matrix [14], [21], [45], [59]. This work was extended to systems with a non-zero feedthrough matrix in [22] while the first Markov parameter rank condition was relaxed in [24]. Noisy discrete time systems were considered next under the same system matrix conditions. Various filter derivations and properties were given in [43], [12], [13], [41], [19]. This was extended in [23] to systems with a non-zero feedthrough matrix using the technique in developed in [22] for noise-free systems. In [20] an additional least-squares procedure is used for input estimation to derive a filter for systems with a full column rank feedthrough matrix. More recently [11], [61], [62] inspired by [20] and [12] describe a procedure to derive filters which relax the above matrix conditions. In [5], [18] filters have been derived for linear discrete time systems in the zero informational limit for the process noise as a method to treat the estimation problem for unknown inputs. In [5] there is no direct feedthrough of the process noise to the measurements and there is an assumption that the first Markov parameter has full column rank.

Early work on observability of systems with unknown inputs focused on continuous and deterministic systems without feedthrough of the inputs to the measurements [3]. Procedures for the construction of full [14], reduced [21], [45] and minimal [59] order observers are developed assuming that the first Markov parameter (CB in the notation of (2.21), (2.22) for continuous time systems) is full column rank, the latter being closely related to the use of the first derivative of the output for the estimation of the unknown inputs. The approach is extended to the case where there is a non-zero feedthrough matrix in [22] again assuming use of the first derivative of the measurement output. The rank condition assumed in [14], [21], [45] and [59] is relaxed in [24] at the expense of using higher order derivatives of the measured output for input reconstruction.

Starting with [43] discrete time filters are derived for stochastic systems with unknown inputs that do not feedthrough to the measurements, again assuming that the first Markov parameter has full column rank. The problem is posed as a constrained optimisation with a free gain matrix parameter. The work continued with [13] that reformulated the problem as a state estimation of a singular stochastic system which is solved by employing a generalised least squares approach. Later, [12] shows the optimality of the filter in [43] among the set of recursive filters and produces stability results, while [41] verifies optimality among the set of all linear filters. In [19] the scope of the filter in [43] is expanded by simultaneously estimating both the state and unknown input.

In [23] the case of a rank deficient feedthrough matrix is considered. The unknown input decoupling technique developed in [22] to transform the original system is used to construct an optimal filter. In [26] and [25] the input is treated as a stochastic process with a wide-sense representation. Paper [47] considers the system in [43] with some additional partial measurements of the inputs with no noise. In [56] the work of [47] is continued to derive results on existence, optimality and asymptotic stability. In [40] a system is considered that includes constant biases in addition to unknown inputs. A solution is proposed based on a system augmentation and transformation.

In [20] a discrete time stochastic system with a full column rank input feedthrough matrix is considered. The input is estimated based on a generalised least squares approach and state estimation is posed as a constrained minimisation problem using Lagrange multipliers. The approach is reminiscent of [43] and others but applied to a different problem. The paper does not present any asymptotic stability results. Further, the detectability of the equivalent system with known inputs is assumed without proof (a condition that will be shown to be necessary but insufficient for asymptotic stability in this thesis). The work of [11], [61] and [62] attempts to produce more general filters which do not require the input feedthrough matrix to be full column rank. A linear recursive filter is proposed that is inspired by both [20] and [12]. More general multi-step delay filters can be derived at the cost of increased complexity.

The recent work of [5] takes a new approach to the problem of estimating highly uncertain inputs. The unknown input is treated as white noise whose covariance becomes unbounded. More precisely the limit is taken in the filter equations as the inverse of the input covariance tends to zero, which is the zero informational limit. It is shown that the resulting filter equations take the same form as those of Gillijns and De Moor [19]. Both [5] and [19] consider the case where there is no direct feedthrough of unknown input to the measurement vector and with the assumption that the first Markov parameter is full rank.

2.2 Kalman filter and smoother

2.2.1 Stochastic formulation

In engineering literature real time estimation is often referred to as filtering, the name originates from notions of sorting out entities or entities passing through a barrier (e.g. filtering impurities in water) and modern usage often has the more abstract notion of extracting useful information from a large amount of data. Smoothing (prediction) on the other hand refers to the estimation at a past (future) time using information up to the present time and smoothed estimates usually appear smoother across time since data from the distant future are often not very informative of the present. In this section we will introduce a stochastic problem formulation for filtering and smoothing in discrete time linear systems which was introduced by Kalman [37] (see [1], [51], [27] and references within for a more in depth analysis). Consider the linear, finite-dimensional, stochastic, discrete time system with the state space description:

$$x_{k+1} = Ax_k + Bw_k, \tag{2.1}$$

$$z_k = Cx_k + v_k \tag{2.2}$$

where the subscript $k \in \mathbb{N}^+$ (non-negative natural numbers) is a discrete time index, $x_k \in \mathbb{R}^n$ is the system state, $z_k \in \mathbb{R}^p$ is the vector of measurements, $w_k \in \mathbb{R}^m$ is the process noise or input and $v_k \in \mathbb{R}^p$ is the measurement noise, the system matrices $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}, C \in \mathbb{R}^{p\times n}$ are assumed to be known. We assume that w_k and v_k are independent, zero mean, Gaussian white noise processes with covariances $Q > 0$ and $R > 0$, i.e. $\mathbb{E}(w_k w_l^T) = Q\delta_{kl}$, $\mathbb{E}(v_k v_l^T) = R\delta_{kl}$ and $\mathbb{E}(w_k v_l^T) = 0$ for all $k, l \in \mathbb{N}^+$ where δ_{kl} is the Kronecker delta and the initial state x_0 is Gaussian random variable with mean $\hat{x}_{0|-1}$ and covariance $P_{0|-1} > 0$. We denote the set of measurements $\{z_0, \ldots, z_j\}$ by Z_j and introduce the notation for the conditional state mean $\hat{x}_{k|j} = \mathbb{E}(x_k|Z_j)$ and covariance $P_{k|j} = \mathbb{E}((x_k - \hat{x}_{k|j})(x_k - \hat{x}_{k|j})^T|Z_j)$.

The Gaussian form of the distributions allows for tractable and exact estimation through computationally efficient algorithms. The standard Kalman filtering and smoothing algorithms are derived in the next two sections. The Gaussian assumption used under this formulation can be applicable to many problems as a result of the central limit theorem, namely by arising from a large sum of non-Gaussian disturbances.

2.2.2 Filtering algorithm

In this section we will derive a filtering algorithm for this problem. Using the assumptions and notation of Section 2.2.1 we have:

$$\hat{x}_{k+1|k} = \mathbb{E}(Ax_k + Bw_k|Z_k) = A\mathbb{E}(x_k|Z_k) + B\mathbb{E}(w_k|Z_k) = A\hat{x}_{k|k}, \tag{2.3}$$

$$P_{k+1|k} = \mathbb{V}(Ax_k + Bw_k|Z_k) = A\mathbb{V}(x_k|Z_k)A^T + B\mathbb{V}(w_k|Z_k)B^T = AP_{k|k}A^T + BQB^T. \tag{2.4}$$

We call these equations the propagation (or prediction) step of the Kalman filter. We now compute the conditional expectations of the state given a measurement and a prior. We first note that for a bivariate Gaussian random variable $\begin{bmatrix} X_1 & X_2 \end{bmatrix}^T$ with mean and covariance:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

the conditional probability density function of X_1 given knowledge of $X_2 = x_2$ is a Gaussian random variable with mean and covariance:

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \tag{2.5}$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T. \tag{2.6}$$

Applying this conditional probability lemma to the augmented "prior" vector of random variables $\begin{bmatrix} x_{k|k-1} & z_{k|k-1} \end{bmatrix}^T$ gives the conditional expectations:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(z_k - C\hat{x}_{k|k-1}), \tag{2.7}$$

$$P_{k|k} = (\mathrm{I} - K_k C)P_{k|k-1}, \tag{2.8}$$

$$K_k = P_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1} \tag{2.9}$$

where we have introduced the "Kalman gain" K_k for convenience. We call these equations the update (or measurement) step of the Kalman filter. We note that the update equation shrinks the uncertainty envelope of the estimate, i.e $P_{k|k} < P_{k|k-1}$ (see Section 2.2.4). The recursive application of the prediction and update steps gives the filtered estimates of the state at all times.

2.2.3 Smoothing algorithm

In this section we will derive a smoothing algorithm which computes expectations over a fixed length interval of horizon length N given measurements in the same interval. First assume that the filtered estimates $x_{k|k}$ and $P_{k|k}$ have already been computed as described in Section 2.2.2. Now let $x_{k+1} = x^*_{k+1}$ be known, applying the conditional probability lemma to the augmented vector $\begin{bmatrix} x_{k|k} & x_{k+1|k} \end{bmatrix}^T$ gives the following mean and covariance (i.e. expectations given measurements up to time k and state x_{k+1}):

$$\hat{x}^*_{k|k} = \hat{x}_{k|k} + \Phi_k(x^*_{k+1} - A\hat{x}_{k|k}) = \Phi_k x^*_{k+1} + (\mathrm{I} - \Phi_k A)\hat{x}_{k|k}, \tag{2.10}$$

$$P^*_{k|k} = (\mathrm{I} - \Phi_k A)P_{k|k}, \tag{2.11}$$

$$\Phi_k = P_{k|k}A^T P^{-1}_{k+1|k} \tag{2.12}$$

where we have introduced Φ_k for convenience. Assuming $\hat{x}_{k+1|N}$ and $P_{k+1|N}$ are known, then by averaging over x_{k+1} the mean and covariance of $x_{k|N}$ are given by:

$$\begin{aligned} \hat{x}_{k|N} &= \mathbb{E}(\Phi_k x_{k+1} + (\mathrm{I} - \Phi_k A)\hat{x}_{k|k}|Z_N) = \Phi_k \mathbb{E}(x_{k+1}|Z_N) + (\mathrm{I} - \Phi_k A)\hat{x}_{k|k} \\ &= \Phi_k \hat{x}_{k+1|N} + (\mathrm{I} - \Phi_k A)\hat{x}_{k|k} = \hat{x}_{k|k} + \Phi_k(\hat{x}_{k+1|N} - A\hat{x}_{k|k}), \end{aligned} \tag{2.13}$$

$$\begin{aligned} P_{k|N} &= \mathbb{V}(\Phi_k x_{k+1} + (\mathrm{I} - \Phi_k A)\hat{x}_{k|k}|Z_N) + P^*_{k|k} = \Phi_k \mathbb{V}(x_{k+1}|Z_N)\Phi_k^T + (\mathrm{I} - \Phi_k A)P_{k|k} \\ &= \Phi_k P_{k+1|N}\Phi_k^T + (\mathrm{I} - \Phi_k A)P_{k|k} = P_{k|k} + \Phi_k(P_{k+1|N} - P_{k+1|k})\Phi_k^T. \end{aligned} \tag{2.14}$$

We may interpret this step as a backwards in time Kalman filter where the model is given by (2.10)-(2.12) (note that this model is time varying even though the original system is not). The end-of-interval estimates $\hat{x}_{N|N}$ and $P_{N|N}$ are available from the filtering stage, it follows from the backwards in time recursive application of the smoothing equations that the smoothed estimates $\hat{x}_{k|N}$ and $P_{k|N}$ can be computed for all $k = \{0, \ldots, N-1\}$.

2.2.4 Riccati equation

Consider the non-linear matrix difference equation:

$$P_{k+1} = AP_kA^T + BQB^T - AP_kC^T(CP_kC^T + R)^{-1}CP_kA^T \tag{2.15}$$

in the unknown matrix function $P : \mathbb{N} \to \mathbb{R}^{n\times n}$ where $A, B, C, R > 0, Q > 0$ are fixed real known matrices. The recursion is known as a *Riccati difference equation* (RDE) and is central to discrete time estimation problems, including the Kalman filter. To see this substitute from (2.8) into (2.4) which gives the required recursion. Assuming $P_0 > 0$, then the RDE has a unique positive definite solution $P_k > 0$ for all $k \geq 0$ (see [6, p. 165]). In steady-state the RDE (2.15) is given by the *algebraic Riccati equation* (ARE):

$$P = APA^T + BQB^T - APC^T(CPC^T + R)^{-1}CPA^T \tag{2.16}$$

in the unknown matrix $P \in \mathbb{R}^{n\times n}$. A real symmetric nonnegative definite solution of the ARE is said to be a *strong solution* if all the eigenvalues of $A - AKC$ are on or inside the unit circle, where $K = PC^T(CPC^T + R)^{-1}$. If all the eigenvalues are strictly inside the unit circle, the solution is said to be a *stabilizing solution* [4], [10].

Lemma 1 *The strong solution of algebraic Riccati equation exists and is unique if and only if (C, A) is detectable. The strong solution is the only nonnegative definite solution of algebraic Riccati equation if and only if (C, A) is detectable and (A, B) has no uncontrollable mode outside the unit circle. Furthermore, the strong solution coincides with the stabilising* *solution if and only if (C,A) is detectable and (A,B) has no uncontrollable mode on the unit circle. The stabilising solution is positive definite if and only if (C,A) is detectable and (A,B) has no uncontrollable mode inside, or on the unit circle.*

Proof: See [15, Theorem 3.2] and [10, Theorem 3.1]. □

Lemma 2 *Let (C,A) be detectable, P_k be the solution of the Riccati difference equation (2.15) with initial condition P_0, and P be the unique strong solution of the algebraic Riccati equation. Then if (A,B) has no uncontrollable mode on the unit circle or $P_0 \geq P$, then P_k asymptotically converges to P as $k \to \infty$.*

Proof: See [15, Theorem 4.1, Theorem 4.2]. □

The equivalent continuous time equation is the non-linear matrix differential equation:

$$\dot{P} = AP + PA^T - PC^T R^{-1} CP + BQB^T \tag{2.17}$$

in the unknown matrix function $P : \mathbb{R} \to \mathbb{R}^{n \times n}$. It is known as a *Riccati differential equation* (RDE) and is fundamental to continuous time estimation problems. Assuming $P(0) > 0$, then the RDE has a unique positive definite solution $P(t) > 0$ for all $t \geq 0$ (see [6, p. 165]). In steady-state the RDE (2.17) is given by the *algebraic Riccati equation* (ARE):

$$AP + PA^T - PC^T R^{-1} CP + BQB^T = 0 \tag{2.18}$$

in the unknown matrix $P \in \mathbb{R}^{n \times n}$. Similarly the differential equation:

$$\dot{S} = FS + SF^T + SH^T R^{-1} HS - GQG^T \tag{2.19}$$

in the unknown matrix function $S : \mathbb{R} \to \mathbb{R}^{n \times n}$ is an RDE central to control problems, e.g. linear-quadratic-regulator (LQR). Assuming $S(0) > 0$, then the RDE (2.19) has a unique positive definite solution $S(t) > 0$ for all $t \leq 0$. This follows from the positive time case by considering the transformation $t \to -t$. In steady-state the RDE (2.19) is given by the *algebraic Riccati equation* (ARE):

$$FS + SF^T + SH^T R^{-1} HS - GQG^T = 0 \tag{2.20}$$

in the unknown matrix $S \in \mathbb{R}^{n \times n}$.

Lemma 3 *The ARE (2.18) has a unique solution P^∞ that is stabilizing, i.e. $A - P^\infty C^T R^{-1} C$ is Hurwitz, if and only if (C,A) is detectable and (A,B) has no uncontrollable modes on the* *imaginary axis. If these conditions hold $P^\infty \geq 0$. Furthermore, P^∞ is nonsingular if and only if (A,B) has no stable uncontrollable modes.*

Proof: See [63, Theorem 13.7], [9, p. 985], [48] or [44]. □

Lemma 4 *Let the conditions of Lemma 3 hold such that P^∞ is the unique positive semi-definite and stabilising solution of the ARE (2.18). Then the unique positive definite solution $P(t) > 0$ for all $t \geq 0$ of the RDE (2.17) with the initial condition $P(0) > 0$ has a limit as $t \to \infty$ which is given by P^∞.*

Proof: See [9], [35], [8] and [46, Theorem 3.7] noting that the null space of $P(0)$ is empty by assumption. □

Remark 1 *These results carry over easily to the reversed time case. Namely the ARE (2.20) has a unique solution S^∞ that is "anti-stabilizing", i.e. $-F - S^\infty H^T R^{-1} H$ is Hurwitz, if and only if $(H,-F)$ is detectable and (F,G) has no uncontrollable modes on the imaginary axis. If these conditions hold $S^\infty \geq 0$. Furthermore, S^∞ is nonsingular if and only if (F,G) has no unstable uncontrollable modes. The unique positive definite solution $S(t) > 0$ for $t \leq 0$ of the RDE (2.19) with the terminal condition $S(0) > 0$ has a limit as $t \to -\infty$ which is given by S^∞.*

2.2.5 Filtering in systems with feedthrough

We now consider the same problem formulation for a system which has an additional feedthrough of inputs to outputs:

$$x_{k+1} = Ax_k + Bw_k, \tag{2.21}$$

$$z_k = Cx_k + Dw_k + v_k \tag{2.22}$$

where $D \in \mathbb{R}^{p\times m}$ is assumed to be known and introduce the notation $\hat{w}_{k|j} = \mathbb{E}(w_k|Z_j)$, $P^{ww}_{k|j} = \mathbb{E}((w_k - \hat{w}_{k|j})(w_k - \hat{w}_{k|j})^T|Z_j)$ and $P^{xw}_{k|j} = (P^{wx}_{k|j})^T = \mathbb{E}((x_k - \hat{x}_{k|j})(w_k - \hat{w}_{k|j})^T|Z_j)$. The filtering recursions for this system can be similarly derived [16] and are given by:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_{x,k}(z_k - C\hat{x}_{k|k-1}), \tag{2.23}$$

$$\hat{w}_{k|k} = K_{w,k}(z_k - C\hat{x}_{k|k-1}), \tag{2.24}$$

$$P_{k|k} = P_{k|k-1} - K_{x,k}\Theta_k K_{x,k}^T, \tag{2.25}$$

$$P^{ww}_{k|k} = Q - K_{w,k}\Theta_k K_{w,k}^T, \tag{2.26}$$

$$P^{xw}_{k|k} = (P^{wx}_{k|k})^T = -K_{x,k}\Theta_k K_{w,k}^T, \tag{2.27}$$

$$K_{x,k} = P_{k|k-1} C^T \Theta_k^{-1}, \tag{2.28}$$

$$K_{w,k} = QD^T \Theta_k^{-1}, \tag{2.29}$$

$$\Theta_k = CP_{k|k-1}C^T + DQD^T + R, \tag{2.30}$$

$$\hat{x}_{k+1|k} = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} \hat{x}_{k|k} \\ \hat{w}_{k|k} \end{bmatrix}, \tag{2.31}$$

$$P_{k+1|k} = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} P_{k|k} & P^{xw}_{k|k} \\ P^{wx}_{k|k} & P^{ww}_{k|k} \end{bmatrix} \begin{bmatrix} A & B \end{bmatrix}^T \tag{2.32}$$

where (2.23)–(2.30) is the update step and (2.31)–(2.32) is the propagation step.

2.3 Nonlinear filters

The Kalman filter and smoother exactly compute state expectations in linear systems given Gaussian assumptions on the noise and initial state. In most problems of interest the systems have dynamics that are not linear and are not known precisely and the Gaussian assumptions do not apply. In all but a few special cases, these problems do not have exact closed form solutions and the statistics are computationally intractable. In this section we will review widely used algorithms which aim to approximate the true solutions and have been shown to be successful in real world problems. We will start with the extended and unscented Kalman filters which attempt to approximate only the mean and covariance of the probability distributions. Then we will discuss Monte Carlo algorithms which attempt to approximate the probability distributions directly through a large set of samples.

2.3.1 Extended Kalman filter

The “extended Kalman filter" (EKF) proposes that the nonlinear system is linearised at the filtered state estimate and the standard Kalman filter update and propagation steps are applied. This approach uses the first order of the (multi-dimensional) Taylor series expansion of the nonlinear dynamics but ignores all higher order terms. This simple extension of the Kalman filter comes at the cost of accuracy, especially when the system nonlinearities are significant relative to the level of state uncertainty. The accuracy of the estimation can be improved by including higher order terms of the expansion, for example the second term of the expansion can be included at the computational cost of finding the Hessian matrices. For systems derived from the discretisation of continuous time systems, reducing the time-step size can also improve accuracy at a computational cost. The computational cost also increases sharply with the dimensions of the state and outputs.

A common manifestation of the errors introduced in the approximation is filter inconsistency, namely the filter becomes unjustifiably confident about the state position (i.e. the state covariance is underestimated). Monte-Carlo simulations can be used here to evaluate how confident the filter can justifiably be (we will discuss the limitations of using Monte Carlo simulations for estimation in a later section). A common consequence of filter inconsistency is filter divergence, this occurs when a biased state estimate leads to an inaccurate linearisation and because the filter is too confident on the position of the state, the measurements are not trusted, they are effectively discarded and the state estimate diverges further from the true value. A heuristic approach often used in practice to resolve the inconsistency and divergence problem is to add process noise, which artificially increases the state covariance. The extended Kalman filter does not come with any performance guarantees and its assumptions pose fundamental constrains on the accuracy of the estimation.

2.3.2 Unscented Kalman filter

The “unscented Kalman filter” (UKF) developed in [30] proposes an alternative approach to extend the Kalman filter to nonlinear systems. It is based on a selection of a small number of non-random samples of the state and input, called sigma points, to parametrise the mean and covariance of the transformed system state and output . At a similar computational cost to the EKF and without requiring the direct computation of Jacobians, the UKF is able to capture the transformed state and output mean and covariance up to second order.

We will now briefly present the key ideas of the algorithm in [30]. Consider a random variable $X \in \mathbb{R}^L$ with mean x and covariance P_x and a second random variable Y related to X through the transformation $Y = f(X)$. A general closed form solution for the density of Y does not exist. Let the Taylor series expansion of $f(X)$ about x be given (up to first order) by:

$$Y = f(x) + \nabla f(X - x) + \dots \tag{2.33}$$

then taking expectations gives the mean and covariance:

$$y = f(x), \tag{2.34}$$

$$P_y = (\nabla f) P_x (\nabla f)^T \tag{2.35}$$

where we have truncated the series after the first term. This is equivalent to linearising as it is done in the EKF and it is accurate if the second and higher order terms are negligible. Now consider the small number of weighted samples (sigma points) given by:

$$x_i = x \qquad W_0 = \frac{\lambda}{L+\lambda} \qquad \text{for } i = 0 \tag{2.36}$$

$$x_i = x + P_i \qquad W_i = \frac{1}{2(L+\lambda)} \qquad \text{for } i = 1,\dots,L \tag{2.37}$$

$$x_i = x - P_i \qquad W_i = \frac{1}{2(L+\lambda)} \qquad \text{for } i = L+1,\dots,2L \tag{2.38}$$

where P_i denotes the ith column of $\sqrt{(L+\lambda)P_x}$ and λ is a constant. The weighted sample mean and covariance of x_i are x and P_x, i.e. the same as the mean and covariance of X. Transform the sigma points through the function, i.e. evaluate $y_i = f(x_i)$, and then compute their weighted sample mean and covariance:

$$y = \sum_{i=0}^{2L} W_i y_i, \tag{2.39}$$

$$P_y = \sum_{i=0}^{2L} W_i (y_i - y)(y_i - y)^T. \tag{2.40}$$

The mean and covariance of Y are given correctly up to second order by y and P_y respectively [29], [32], [28]. This is an improvement over the linearisation approach used in the EKF, which calculates the mean correctly only up to first order. In particular, in the limit as $L+\lambda \to 0$ (note that λ can be freely chosen) all the samples x_i tend to x (i.e. $x_i \to x$) and the computed mean and covariance tend to:

$$y = f(x) + \frac{1}{2}\nabla^2 f P_x, \tag{2.41}$$

$$P_y = (\nabla f) P_x (\nabla f)^T \tag{2.42}$$

which are the expectations truncating the Taylor series expansion after the second order term. This approximation is used in the second order Gauss filter and requires the additional computational effort relative to the EKF of computing the Hessian matrix. It should be noted that because any square root is an orthogonal transformation of another [55], all choices of matrix square roots capture the mean and covariance. The Cholesky decomposition is often used as a numerally efficient and stable method [29].

This method of approximating the mean and covariance of a transformed random variable is known as the unscented transform (UT) and is the central idea behind the UKF. More precisely, in the propagation step of the filter the UT method is applied directly to find the statistics at the next time step, in the update step it is applied to the augmented state and output vector and the conditional probability lemma then applies directly. The UKF algorithm can be found in detail at [30], [58] and [31]. It is worth noting that white noise added to a transformed random variable, e.g.process and measurement noise, only increases the variance of the output by the variance of the white noise. When this was realised the UKF algorithm was adapted since it is computationally advantageous over unnecessarily generating and transforming more sigma points and then computing expectations.

Interesting estimation applications of the UKF and EKF algorithms are presented in [58], including nonlinear system identification, duel estimation and training of neural networks where the UKF consistently outperformed the EKF at a comparable complexity. More recently [2] the EKF and UKF were applied in a vehicle estimation application and tested in simulations and experiments with a BMW 5 Series. It was observed that 1) the UKF was more robust and accurate, especially in cases of large sampling time-steps, and 2) the EKF diverged in some tests where the filter was initialised correctly. Lastly, we should note a practical advantage of the UKF over the EKF algorithm. Namely, it does not require finding analytic expressions for the Jacobian (and potentially Hessian) of nonlinear models. It is thus easier to implement quickly in multiple complex models.

2.3.3 Monte Carlo methods

The EKF and UKF algorithms capture only the mean and covariance of the transformed probability distributions (up to first or second order). This approach fundamentally limits the achievable estimation accuracy, especially when the probability distributions are not adequately described only through the first two moments. Further, the accuracy achieved by these algorithms cannot be increased with increased computational power and convergence guarantees do not exist. In this section we will discuss particle methods, which are Monte Carlo algorithms that give approximate solutions to inference problems [17]. They are normally nonlinear and seek out the true probability distributions (unlike the EKF and UKF algorithms). They attempt to address limitations of the EKF and UKF algorithms but they are more computationally expensive and hence potentially less suitable for real time applications. Lastly, we will discuss some of the current challenges with particle methods in both filtering and smoothing, especially in relation to problems with a large horizon.

Consider a hidden Markov model (HMM) and a known prior state distribution. Similarly to the Kalman filter and smoother, inference given output measurements is based on computing conditional posterior distributions. Applying Bayes theorem and the Markov assumption we can derive recursive expressions for filtering (which have a prediction and update step) and smoothing. However, we do not make any further assumptions, e.g. a linear Gaussian model as in the Kalman filter, and thus in most cases closed form expressions for the posterior distributions do not exist. Particle methods are numerical methods where a large number of random samples called particles is used to approximate the posterior distributions for both filtering and smoothing.

Sampling directly from complex and high dimensional target distributions for the state trajectory is not practical. Instead, we can sample from a proposal distribution which is easy to sample, e.g. a multivariate Gaussian, and associate a weight to each sample based on the target distribution, this method is known as importance sampling (IS). Furthermore, by selecting a proposal distribution which allows us to sample (and update the weights) sequentially at each time step the computational cost increases only linearly with the number of time steps, this is known as sequential importance sampling (SIS). It is worth noting that if the hidden Markov model is a nonlinear state space model with Gaussian disturbances and noises, then the proposal distribution can be chosen to be a Gaussian such that all the weights are equal, which has some advantages we will now discuss.

The samples of an approximation using importance sampling are weighted and they are not approximately distributed according to the target distribution. When there is a large disparity in the sample weights, some of the samples have a negligible weight and hence contribution to the approximation, i.e the effective sample size (ESS) is lower, but still have an equal impact on the computational effort. To address this, techniques which sample from the important sampling approximation are used (e.g. systematic resampling, residual resampling, multinomial resampling [42]) and are sometimes applied only when an ESS threshold criterion is satisfied. Resampling in essence rejects particles of low weight with high probability and hence computational effort is not expended on areas of low probability density. The downside is that every resampling step reduces the number of unique sample trajectories especially in the distant past (degeneracy), but it is fundamentally a consequence of attempting to represent a distribution on a space of increasing dimension (i.e. expanding time horizon) using a finite number of samples. Hence, resampling aims to reduce the weight variance to improve the sample representation in the future at the expense of sample representation in the present and past, this poses a major challenge in smoothing applications with long horizons.

Chapter 3

Estimation and control for continuous time systems: A deterministic view

3.1 Introduction

Our goal in this chapter is to pose and solve a filtering/estimation problem for the simultaneous estimation of inputs and states in a continuous time linear finite dimensional dynamical system. We assume that a model of the system is available. The output of the dynamical system is the vector of *all* variables that are measured (e.g. by means of sensors). The filter should make use only of these measured outputs for estimation and produce the best estimate of the system variables treating the exogenous inputs and states on an equal footing. The meaning of "best" is to minimise a weighted integral squared error between output and measured output. The problem set-up is illustrated in Fig. 3.1. The filtered signals $w_1(t), x_1(t)$ provide best estimates of w and x at a given time instant t based on measurements up to that time, while the estimates $\hat{w}(t), \hat{x}(t)$ provide the best estimates over a time interval $[0, T]$.

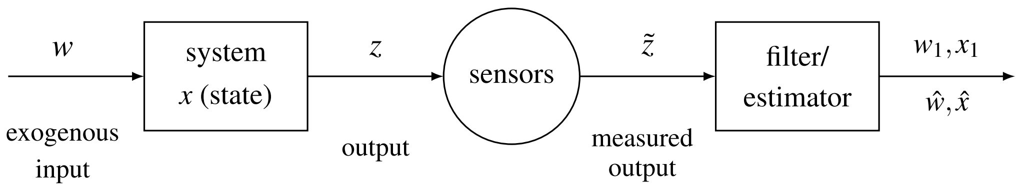

Fig. 3.1 Block diagram of a dynamical system with state x, exogenous input w and output z which is measured with sensors interconnected with an estimator.

The problem formulation differs from the Kalman filter [39] in two respects: (1) there is no notion of "process noise"; (2) the problem is purely deterministic. In regard to (1), if there is a noisy measurement of an exogenous input, our formulation advocates that this should be included in the measurement vector along with the other measured variables. In regard to (2), our approach is inspired by Willems who showed in [60] that the continuous time Kalman filter admits a deterministic least squares formulation.

The estimation problem of Fig. 3.1 has a strong motivation in engineering applications. For example, the motion of an automobile is determined by the external forces acting on it (principally tyre and aerodynamic forces). These are very difficult to measure directly but may be estimated together with the system state from an appropriate set of sensors. Similar considerations apply to other types of vehicles, e.g. aerial or nautical vehicles or vessels. In the field of structural dynamics an example would be the estimation of forces applied to a structure together with the resulting vibrational displacements in the structure.

The use of the standard Kalman filter in such examples imposes a Gaussian assumption on the exogenous input which may not be justified. A similar issue arises in the deterministic formulation of the Kalman filter [60] since the optimisation problem requires a best fit of the observed signal using an input disturbance that has minimum integral squared norm. Such an assumption may not be physically motivated. The formulation in Fig. 3.1 dispenses with this requirement.

Our solution to the problem of Fig. 3.1 is otherwise directly inspired by the approach of Willems. In particular, we rewrite the cost functional in a convenient form to determine the unique optimal solution by means of completion of squares. This involves the construction of a dynamical system which turns out to be an "end-of-interval estimator" to generate the signals $w_1(t), x_1(t)$. The form of the solution involves a feedback from a projected error signal with a gain based on the solution of a Riccati differential equation that is reminiscent of the Kalman filter and can be solved in real time. A further construction is required to determine the optimal solution $\hat{w}(t), \hat{x}(t)$ on a fixed interval $[0, T]$. The latter corresponds to the "smoothed" estimate in regular Kalman filtering and is not computable in real time since it involves the further solution of a system backwards in time.

In the stochastic approach to the Kalman filter the Riccati equation solution is a state covariance. In this work we wish to go beyond [60] to obtain a least squares interpretation of the Riccati differential equation solution $P_1(t)$ arising in our filter. To do this we pose and solve a constrained optimisation problem which requires the state to pass through a prescribed point at a given time. The optimal cost for the new problem is increased by a term which is the norm squared error between the state and the optimal state at the given time weighted by the inverse of a matrix Lyapunov differential equation solution $P_2(t)$. The latter coincides with $P_1(t)$ if the prescribed time for the state constraint is at the end of the interval, i.e. $t = T$. This allows the interpretation that if $P_1(t)$ is small, the measurements suggest strongly that $x_1(t)$ is an accurate estimate of the state at time t given measurements up to that time, while conversely if its inverse $P_1(t)^{-1}$ is near singular then there exist trajectories with $x(t)$ far from $x_1(t)$ that fit the measurements $\tilde{z}$ up to time t almost as well. There is a similar interpretation for $P_2(t)$ valid for any time t within the fixed horizon length. These interpretations are closely analogous to the meaning of the filtered and smoothed state covariances in the stochastic formulation.

In order to solve the constrained optimisation problem, a tracking problem is introduced that has a close relation to the estimation problem and has independent interest. In one respect the estimation and tracking problems are identical, namely it is desired to minimise the weighted integral squared error between the output of the dynamical system and another signal that is given, i.e. a set of sensor measurements or a desired trajectory. In the estimation problem we seek the input and state that is the best explanation of the observed trajectory given the measurements made, while in the tracking problem we seek the input (and state) that gives the closest output trajectory of the system to the desired one. In another respect the two problems differ in that there is a quadratic penalty in the cost criterion on the initial state (estimation problem) or on the final state (tracking problem). This results in the two problems having solutions which are dual to each other: in the tracking problem the first stage of the solution solves a Riccati differential equation backwards in time after which the optimal control and state trajectory are found by integrating forwards in time, which is the opposite way round to the estimation problem. This duality is reminiscent of, and closely related to, the well-known duality between the Kalman filter and the linear quadratic regulator.

Throughout the chapter we make an assumption that the direct feedthrough matrix of the system has full column rank. It is possible that this could be relaxed at the expense of making the filter equations more involved. However, we demonstrate in Section 3.5 that the formulation is still sufficiently general to include the regular Kalman filter, the Kalman filter with direct feedthrough of process noise to measurements and the linear quadratic regulator as special cases. We mention that the full column rank assumption is closely related to the need to avoid differentiating the measured output in the estimation problem.

The chapter goes on to consider the infinite horizon case for the estimation problem in Section 3.6. It is shown under mild conditions that the limiting form of the end-of-interval estimator can be written as a linear system solved forwards in time with the system matrices determined via the solution of an algebraic Riccati equation. We show that the end-of-interval estimator of the input is a stable left inverse of the original system. We also show that the unique solution of the estimation problem has a limiting form which includes a second stage of processing via an anti-stable system, equivalently a system that is stable in the backwards time direction. We show that the series connection of the end-of-interval estimator (with judiciously chosen output) and this anti-causal “smoother" is also a left inverse of the original system.

Section 3.7 considers the infinite horizon tracking problem. On a finite horizon the tracking problem solution has a natural two-stage form where the first stage involves a backwards-in-time integration and the second stage has an integration forwards in time. This form is maintained in the infinite horizon limit with, under mild conditions, the first stage being an anti-stable system (equivalently a stable anti-causal system) and the second stage being a stable system. The analysis takes care to show that the optimal control converges for any fixed time t to the solution described above. Moreover, we show that the first stage system is an anti-stable left inverse of the original system, and that the series connection of the first stage's computation of a modified output and the second stage system is also a left inverse of the original system.

3.2 Estimation

Consider the linear, finite-dimensional, continuous time system with the state space description:

$$\dot{x} = Ax + Bw, \tag{3.1}$$

$$z = Cx + Dw \tag{3.2}$$

where $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}, C \in \mathbb{R}^{p\times n}$ and $D \in \mathbb{R}^{p\times m}$ (full column rank) are fixed known matrices[1] and $w \in \mathscr{L}_{2,e}^m$, $x \in \mathscr{L}_{2,e}^n$ and $z \in \mathscr{L}_{2,e}^p$ are input, state and output related through this linear system. We consider the problem to estimate w and x from the measurement of the signal z, which is the same as estimating w and $x(0)$ since x is generated by (3.1). We assume that the state x and driving input w are not measured directly, other than (indirectly) through the measurement of z (i.e. all measurements of the system are made through the output vector z). Each element of z may correspond to an individual sensor or multiple entries of z may be generated by a single device. To pose our problem precisely we will denote by $\tilde{z}$ the actual measured output signal in an experiment (see Fig. 3.1). We introduce the performance index:

$$C_T(w, x(0)) = \int_0^T \|\tilde{z}(t) - z(t)\|_{R^{-1}}^2 dt + \|x(0) - \gamma\|_{\Gamma^{-1}}^2 \tag{3.3}$$

[1]The assumption that the system matrices are constant is for notational convenience. We note that all finite horizon results in Sections 3.2-3.5 are valid if the system matrices A, B, C, D are time-varying, with no change required in the proofs.

where $0 < R \in \mathbb{R}^{p \times p}$, $0 < \Gamma \in \mathbb{R}^{n \times n}$, $\gamma \in \mathbb{R}^n$, $0 < T \in \mathbb{R}$ are specified. The problem we wish to solve is:

$$\inf_{w,x(0)} C_T(w,x(0)) \tag{3.4}$$

subject to (3.1) and (3.2). In particular we wish to compute the optimal w and $x(0)$ which we will denote by $\hat{w}$ and $\hat{x}(0)$. A key step in our solution of (3.4) is a "completion of squares" construction for the performance index (3.3) which is given in the following lemma.

Lemma 5 *Consider the system:*

$$\dot{x}_1 = (A_1 - K_1 C_1)x_1 + (B_1 + K_1)\tilde{z}, \tag{3.5}$$
$$\dot{P}_1 = A_1 P_1 + P_1 A_1^T - K_1 R K_1^T + B_1 R B_1^T, \tag{3.6}$$
$$K_1 = P_1 C_1^T R^{-1}, \tag{3.7}$$
$$w_1 = D^\dagger(\tilde{z} - z_1), \tag{3.8}$$
$$z_1 = C x_1 \tag{3.9}$$

with the initial conditions $P_1(0) = \Gamma$ and $x_1(0) = \gamma$, where we have defined:

$$A_1 = A - B_1 C, \tag{3.10}$$
$$B_1 = BD^\dagger, \tag{3.11}$$
$$C_1 = \Pi_c C, \tag{3.12}$$
$$\Pi_c = \mathrm{I} - \Pi, \tag{3.13}$$
$$\Pi = DD^\dagger, \tag{3.14}$$
$$D^\dagger = (D^T R^{-1} D)^{-1} D^T R^{-1}. \tag{3.15}$$

Then the RDE (3.6) has a unique positive definite solution $P_1(t) > 0$ for all $t \in [0,T]$. Furthermore, the performance index defined in (3.3) is given by:

$$C_T(w,x(0)) = \int_0^T \|\Pi_c(\tilde{z}(t) - z_1(t))\|^2_{R^{-1}} dt$$
$$+ \int_0^T \|\Pi(RB_1^T P_1(t)^{-1}(x(t) - x_1(t)) + \tilde{z}(t) - z(t))\|^2_{R^{-1}} dt + \|x(T) - x_1(T)\|^2_{P_1(T)^{-1}} \tag{3.16}$$

(w_1 is defined here for convenience and will be first used in Theorem 2).

Proof: The parallel projection Π satisfies $\Pi^2 = \Pi$ and:

$$\Pi_c^T R^{-1} D = 0. \tag{3.17}$$

Hence the following identities hold:

$$K_1 D = 0, \tag{3.18}$$

$$K_1 C_1 = K_1 C. \tag{3.19}$$

From (3.2) $w = D^\dagger (z - Cx)$. Substituting into (3.1) gives:

$$\dot{x} = A_1 x + B_1 z = (A_1 - K_1 C_1)x + (B_1 + K_1)z \tag{3.20}$$

using (3.18) and (3.19). From (3.20) and (3.5) we obtain:

$$\dot{x} - \dot{x}_1 = (A_1 - K_1 C_1)(x - x_1) - (B_1 + K_1)(\tilde{z} - z). \tag{3.21}$$

We note that $\Pi_c(z - z_1) = C_1(x - x_1)$ from which it follows that:

$$C_1(x - x_1) + (\tilde{z} - z) = \Pi_c(\tilde{z} - z_1) + \Pi(\tilde{z} - z). \tag{3.22}$$

Hence from (3.17) and (3.22):

$$\|C_1(x - x_1) + (\tilde{z} - z)\|^2_{R^{-1}} = \|\Pi_c(\tilde{z} - z_1)\|^2_{R^{-1}} + \|\Pi(\tilde{z} - z)\|^2_{R^{-1}}. \tag{3.23}$$

Using (3.6), (3.21) and (3.23) we verify the calculation:

$$\begin{aligned}
&\frac{d}{dt}\Big(\|x - x_1\|^2_{P_1^{-1}}\Big) = \frac{d}{dt}\Big((x - x_1)^T P_1^{-1}(x - x_1)\Big) \\
&= 2(x - x_1)^T P_1^{-1}(\dot{x} - \dot{x}_1) - (x - x_1)^T P_1^{-1}\dot{P}_1 P_1^{-1}(x - x_1) \\
&= 2(x - x_1)^T P_1^{-1}((A_1 - K_1 C_1)(x - x_1) - (B_1 + K_1)(\tilde{z} - z)) \\
&\qquad - (x - x_1)^T (P_1^{-1}A_1 + A_1^T P_1^{-1} - C_1^T R^{-1} C_1 + P_1^{-1} B_1 R B_1^T P_1^{-1})(x - x_1) \\
&= -2(x - x_1)^T (P_1^{-1}B_1 + C_1^T R^{-1})(\tilde{z} - z) - (x - x_1)^T (P_1^{-1} B_1 R B_1^T P_1^{-1} + C_1^T R^{-1} C_1)(x - x_1) \\
&= -\|R B_1^T P_1^{-1}(x - x_1) + \Pi(\tilde{z} - z)\|^2_{R^{-1}} + \|\Pi(\tilde{z} - z)\|^2_{R^{-1}} \\
&\qquad - \|C_1(x - x_1) + (\tilde{z} - z)\|^2_{R^{-1}} + \|(\tilde{z} - z)\|^2_{R^{-1}} \\
&= -\|R B_1^T P_1^{-1}(x - x_1) + \Pi(\tilde{z} - z)\|^2_{R^{-1}} - \|\Pi_c(\tilde{z} - z_1)\|^2_{R^{-1}} + \|(\tilde{z} - z)\|^2_{R^{-1}}
\end{aligned} \tag{3.24}$$

where in the penultimate step we have noted that $B_1\Pi = B_1$. Integrating (3.24) in the interval $[0,T]$ gives the required expression on noting that $RB_1^T = \Pi RB_1^T$. □

Theorem 1 *The optimisation problem in (3.4) has a unique solution $\hat{w}$, $\hat{x}(0)$ which is obtained as follows: first integrate (3.5)–(3.7) forwards in time in the interval $0 \leq t \leq T$ with initial conditions $P_1(0) = \Gamma$ and $x_1(0) = \gamma$; then integrate:*

$$\dot{\hat{x}} = A_2\hat{x} + B_2\tilde{z}_2 \tag{3.25}$$

backwards in time with terminal condition:

$$\hat{x}(T) = x_1(T) \tag{3.26}$$

to compute $\hat{x}(0)$ (and indeed $\hat{x}$); and lastly set:

$$\hat{w} = D^{\dagger}(\tilde{z}_2 - C_2\hat{x}) \tag{3.27}$$

where:

$$A_2 = A - B_2C_2, \tag{3.28}$$
$$B_2 = B_1, \tag{3.29}$$
$$C_2 = C - RB_1^TP_1^{-1}, \tag{3.30}$$
$$\tilde{z}_2 = \tilde{z} - RB_1^TP_1^{-1}x_1. \tag{3.31}$$

Furthermore, the minimum of the performance index (3.3) is:

$$\inf_{w,x(0)} C_T(w,x(0)) = \int_0^T \|\Pi_c(\tilde{z}(t) - z_1(t))\|^2_{R^{-1}}dt \tag{3.32}$$
$$= \int_0^T \|\tilde{z}(t) - \hat{z}(t)\|^2_{R^{-1}}dt + \|\hat{x}(0) - \gamma\|^2_{\Gamma^{-1}} \tag{3.33}$$

where we have denoted the optimal output by:

$$\hat{z} = C\hat{x} + D\hat{w}. \tag{3.34}$$

Proof: We note by an application of [6, Theorem 1, p. 40] that (3.25) may be integrated on the interval $[0,T]$ to yield $\hat{x}$. Next we verify that (3.1) driven by $w = \hat{w}$ from the initial state $x(0) = \hat{x}(0)$ generates the state trajectory $x = \hat{x}$, i.e. $\dot{\hat{x}} = A\hat{x} + B\hat{w}$. This is easily seen by substituting for $\hat{w}$ to obtain (3.25). We now claim that the following lower bound:

$$C_T(w,x(0)) \geq \int_0^T \|\Pi_c(\tilde{z}(t) - z_1(t))\|^2_{R^{-1}} dt \tag{3.35}$$

holds for all w and $x(0)$. To see this note that all terms in (3.16) are non-negative and the first term is independent of w and $x(0)$. We proceed to claim that for $x(0) = \hat{x}(0)$ and $w = \hat{w}$ the last two terms in (3.16) are zero. To see that the second term is zero we substitute $w = \hat{w}$ from (3.27) and $x = \hat{x}$ into (3.16) with z defined in (3.2), i.e. $z = \hat{z} = C\hat{x} + D\hat{w}$. The third term is zero with $x = \hat{x}$ from (3.26). We therefore conclude that $w = \hat{w}$, $x(0) = \hat{x}(0)$ achieve a minimum of the performance index with the minimum given by (3.33). The second line in (3.33) is given by substitution of the optimal solution into (3.3). It remains to show that this solution is unique, which we will now establish by contradiction. Let:

$$x = \hat{x} + \delta x, \tag{3.36}$$

$$w = \hat{w} + \delta w \tag{3.37}$$

be another solution that satisfies (3.1) with δx, δw not identically zero in the interval $[0,T]$. Substituting (3.36) and (3.37) into (3.1) gives:

$$\dot{\delta x} = A\delta x + B\delta w \tag{3.38}$$

by noting that $\hat{x}$, $\hat{w}$ also satisfy (3.1) by construction. The difference in the output is given by:

$$\delta z = C\delta x + D\delta w. \tag{3.39}$$

We now substitute (3.36) and (3.37) into the performance index (3.16) which gives:

$$C_T(w,x(0)) = \int_0^T \|\Pi_c(\tilde{z}(t) - z_1(t))\|^2_{R^{-1}} dt + \int_0^T \|\Pi(RB_1^T P_1(t)^{-1}\delta x(t) - \delta z(t))\|^2_{R^{-1}} dt + \|\delta x(T)\|^2_{P_1(T)^{-1}} \tag{3.40}$$

using the fact that with $x = \hat{x}$ and $w = \hat{w}$ the integrand in the second term of (3.16) is identically zero in the interval $[0,T]$. Under the assumption that the trajectory in (3.36), (3.37) is a solution to the optimisation problem, the last two terms in (3.40) have to be zero, which gives:

$$D^{\dagger}(RB_1^T P_1^{-1}\delta x - \delta z) = 0 \tag{3.41}$$

on the interval $[0,T]$ since D is full column rank, and:

$$\delta x(T) = 0 \tag{3.42}$$

since $P_1(T) > 0$. Substituting δz from (3.39) into (3.41) gives:

$$\delta w = -D^{\dagger} C_2 \delta x \tag{3.43}$$

using (3.30). Substituting δw from (3.43) into (3.38) gives:

$$\dot{\delta x} = A_2 \delta x \tag{3.44}$$

using (3.28). Solving (3.44) backwards in time with the terminal condition (3.42) gives $\delta x = 0$ identically in the interval $[0,T]$, and from (3.43) we have $\delta w = 0$ identically in the same interval, which results in a contradiction. □

We now turn our attention to the filtered estimates (i.e. end-of-interval estimates) of the state and input, namely $\hat{x}(T)$ and $\hat{w}(T)$. In real time applications, the horizon T is itself a variable. It would appear at first glance that a new optimisation problem needs to be solved at every T to compute end-of-interval estimates. The following result shows that this is not the case.

Theorem 2 *Consider the system (3.5)–(3.8) where $P_1(0) = \Gamma$ and $x_1(0) = \gamma$. Then for any fixed T:*

$$\hat{x}(T) = x_1(T), \tag{3.45}$$

$$\hat{w}(T) = w_1(T). \tag{3.46}$$

Proof: The result follows from (3.8), (3.26), (3.27), (3.30) and (3.31). □

Theorem 2 shows that integrating (3.5)–(3.7) as the measurements $\tilde{z}$ become available is sufficient to recover the end-of-interval estimates without computing $x(0)$ or w. We further note that the filter is non-anticipating, meaning that the end-of-interval estimates $x_1(T)$ and $w_1(T)$ do not depend on future measurements, i.e. $\tilde{z}(t)$ for $t > T$. This property is required of any filter to be applied in a real time application.

3.3 Tracking

In this section we will consider a related tracking problem (see [63, Ch. 14] for standard results in optimal control). Assume the state q and input u satisfy:

$$\dot{q} = Fq + Gu, \tag{3.47}$$

$$y = Hq + Ju \tag{3.48}$$

where $F \in \mathbb{R}^{n\times n}, G \in \mathbb{R}^{n\times m}, H \in \mathbb{R}^{p\times n}$ and $J \in \mathbb{R}^{p\times m}$ (full column rank) are fixed known matrices and $u \in \mathscr{L}_{2,e}^{m}$, $q \in \mathscr{L}_{2,e}^{n}$ and $y \in \mathscr{L}_{2,e}^{p}$ are input, state and output related through this linear system. We wish to find an input u such that the output y best tracks a desired signal $\tilde{y} \in \mathscr{L}_{2,e}^{p}$ over a finite horizon T together with a penalty on the deviation of the terminal state from a desired state ξ for a given but arbitrary initial state η. More precisely, we introduce the performance index:

$$W_T(u) = \int_0^T \|\tilde{y}(t) - y(t)\|_{R^{-1}}^2 dt + \|q(T) - \xi\|_{\Xi^{-1}}^2 \tag{3.49}$$

where $\xi \in \mathbb{R}^n$, $0 < \Xi \in \mathbb{R}^{n\times n}$ and propose the problem:

$$\inf_u W_T(u) \tag{3.50}$$

subject to (3.47), (3.48) and $q(0) = \eta$. We denote the optimal solution to (3.50) by $\hat{u}$. This problem differs from the standard finite horizon LQ tracking problem [53, Ch. 8] (see also [52] and [63, Ch. 15]). In particular, note that this formulation has a full column rank input feedthrough matrix and no penalty on the input norm. We first give a completion of squares result similar to Lemma 5 which we then use to solve (3.50) in Theorem 3.

Lemma 6 *Consider the system:*

$$\dot{q}_1 = (F_1 + M_1 H_1)q_1 + (G_1 - M_1)\tilde{y}, \tag{3.51}$$

$$\dot{S}_1 = F_1 S_1 + S_1 F_1^T + M_1 R M_1^T - G_1 R G_1^T, \tag{3.52}$$

$$M_1 = S_1 H_1^T R^{-1}, \tag{3.53}$$

$$u_1 = J^{\dagger}(\tilde{y} - y_1), \tag{3.54}$$

$$y_1 = Hq_1 \tag{3.55}$$

with the terminal conditions $S_1(T) = \Xi$ and $q_1(T) = \xi$, where we have defined the matrices:

$$F_1 = F - G_1 H, \tag{3.56}$$
$$G_1 = GJ^\dagger, \tag{3.57}$$
$$H_1 = \Lambda_c H, \tag{3.58}$$
$$\Lambda_c = \mathrm{I} - \Lambda, \tag{3.59}$$
$$\Lambda = JJ^\dagger, \tag{3.60}$$
$$J^\dagger = (J^T R^{-1} J)^{-1} J^T R^{-1}. \tag{3.61}$$

Then the RDE (3.52) has a unique positive definite solution $S_1(t) > 0$ for all $t \in [0,T]$. Furthermore, the performance index in (3.49) is given by:

$$\begin{aligned} W_T(u) = &\int_0^T \|\Lambda_c(\tilde{y}(t) - y_1(t)))\|^2_{R^{-1}} dt \\ &+ \int_0^T \|\Lambda(RG_1^T S_1(t)^{-1}(q(t) - q_1(t)) - \tilde{y}(t) + y(t))\|^2_{R^{-1}} dt + \|\eta - q_1(0)\|^2_{S_1(0)^{-1}}. \end{aligned} \tag{3.62}$$

Proof: We sketch the outline of two proofs. A direct proof is a completion of squares argument analogous to Lemma 5. It differs from Lemma 5 only in the signs of some terms. An indirect proof is to recognise that Lemma 6 is the time reversed Lemma 5. The transformations $\frac{d}{dt} \to -\frac{d}{dt}$, $A \to -F$, $B \to -G$, $C \to H$, $D \to J$, $P_1 \to S_1$, $x_1 \to q_1$, $z_1 \to y_1$, $\tilde{z} \to \tilde{y}$ and consequential correspondences $A_1 \to -F_1$ etc. suffice to give the result. Furthermore, $S_1(t) > 0$ for all $t \in [0,T]$ is guaranteed by the reversed time Lemma 20 (see Remark 1). □

Theorem 3 *The optimisation problem in (3.50) has a unique solution $\hat{u}$ which is obtained as follows: first integrate (3.51)–(3.53) backwards in time with terminal conditions $S_1(T) = \Xi$ and $q_1(T) = \xi$; then integrate:*

$$\dot{\hat{q}} = F\hat{q} + G\hat{u} \tag{3.63}$$

forwards in time with $\hat{u}$ given by the feedback law:

$$\hat{u} = J^\dagger(\tilde{y}_2 - H_2\hat{q}) \tag{3.64}$$

and initial condition $\hat{q}(0) = \eta$*, where we have defined:*

$$H_2 = H + RG_1^T S_1^{-1}, \tag{3.65}$$

$$\tilde{y}_2 = \tilde{y} + RG_1^T S_1^{-1} q_1. \tag{3.66}$$

Furthermore, the minimum of the performance index (3.49) is:

$$\inf_u W_T(u) = \int_0^T \|\Lambda_c(\tilde{y}(t) - y_1(t))\|^2_{R^{-1}} dt + \|\eta - q_1(0)\|^2_{S_1(0)^{-1}}. \tag{3.67}$$

Proof: We note that (3.63) with $\hat{u}$ given by (3.64) and $\hat{q}(0) = \eta$ may be integrated [6, Theorem 1, p. 40] on the interval $[0,T]$ to yield $\hat{q}$, while $\hat{u}$ can be computed by substituting $\hat{q}$ into (3.64). We now claim that the following lower bound:

$$W_T(u) \geq \int_0^T \|\Lambda_c(\tilde{y}(t) - y_1(t))\|^2_{R^{-1}} dt + \|\eta - q_1(0)\|^2_{S_1(0)^{-1}} \tag{3.68}$$

holds for all u. To see this note that all terms in (3.62) are non-negative and the first and third terms are independent of u. We proceed to claim that for $u = \hat{u}$ the second term in (3.62) is zero. To see this we substitute $u = \hat{u}$ and $q = \hat{q}$ into (3.62) with y defined in (3.48), i.e. $y = \hat{y} = H\hat{q} + J\hat{u}$, and noting that $\Lambda^2 = \Lambda$. We therefore conclude that $u = \hat{u}$ is a solution to the optimisation problem and the minimum of the performance index (3.49) is given by (3.67). It remains to show that this solution is unique, which we will now establish by contradiction. Let:

$$q = \hat{q} + \delta q, \tag{3.69}$$

$$u = \hat{u} + \delta u \tag{3.70}$$

be another solution that satisfies (3.47) with δq, δu not identically zero in the interval $[0,T]$. Substituting (3.69) and (3.70) into (3.47) gives:

$$\dot{\delta q} = F\delta q + G\delta u \tag{3.71}$$

using (3.63). The difference in the output is given by:

$$\delta y = H\delta q + J\delta u. \tag{3.72}$$

We now substitute (3.71) and (3.72) into the performance index (3.62) which gives:

$$W_T(u) = \int_0^T \|\Lambda_c(\tilde{y}(t) - y_1(t))\|^2_{R^{-1}} dt + \int_0^T \|\Lambda(RG_1^T S_1(t)^{-1}\delta q(t) + \delta y(t))\|^2_{R^{-1}} dt + \|\eta - q_1(0)\|^2_{S_1(0)^{-1}} \quad (3.73)$$

using the fact that with $q = \hat{q}$ and $u = \hat{u}$ the integrand in the second term of (3.62) is identically zero in the interval $[0,T]$. Under the assumption that the trajectory in (3.69), (3.70) is a solution to the optimisation problem, the second term in (3.73) has to be zero, which gives:

$$J^\dagger(RG_1^T S_1^{-1}\delta q + \delta y) = 0 \quad (3.74)$$

on the interval $[0,T]$ since J is full column rank. Substituting δy from (3.72) into (3.74) gives:

$$\delta u = -J^\dagger H_2 \delta q \quad (3.75)$$

using (3.65). Substituting δu from (3.75) into (3.71) gives:

$$\dot{\delta q} = (F - G_1 H_2)\delta q. \quad (3.76)$$

Solving (3.76) forwards in time with the initial constraint $\delta q(0) = 0$, which follows since $q(0) = \hat{q}(0) = \eta$, gives $\delta q = 0$ identically in the interval $[0,T]$. Using (3.75) we have $\delta u = 0$ identically in the same interval, which results in a contradiction. □

3.4 Constrained estimation

We now turn our attention to the constrained optimisation problem:

$$\inf_{w,x(0)} C_T(w,x(0)) \text{ subject to } x(\tau) = \zeta \quad (3.77)$$

for $\zeta \in \mathbb{R}^n$ and $0 \leq \tau \leq T$ where (3.1) and (3.2) hold. Here $C_T(w,x(0))$ is defined as in (3.3) with the same meaning for $\tilde{z}$, γ and Γ. Again we wish to compute the optimal w and $x(0)$ which we will denote by $\hat{w}$ and $\hat{x}(0)$. A solution of this optimisation problem will show how the optimal cost increases compared to the unconstrained value when we demand that the state passes through a prescribed point at a given time. This will give an indication in a least squares sense of how “likely" it is that the state passes through the optimum point for the unconstrained problem. For example, if there is a sharp rise in the cost when the state is required to pass through a different point, then we may have more confidence in the value of the unconstrained optimum state at that time. We will first give Lemma 7 before deriving the solution to the optimisation problem (3.77) in Theorem 4.

Lemma 7 *Let x_1, P_1, z_1, A_2, B_2, C_2, $\tilde{z}_2$ be defined as in Lemma 5 and Theorem 1 and consider the system:*

$$\dot{x}_2 = A_2 x_2 + B_2 \tilde{z}_2, \tag{3.78}$$
$$\dot{P}_2 = A_2 P_2 + P_2 A_2^T - B_2 R B_2^T, \tag{3.79}$$
$$w_2 = D^\dagger(\tilde{z}_2 - C_2 x_2) \tag{3.80}$$

with the terminal conditions $P_2(T) = P_1(T)$, $x_2(T) = x_1(T)$. (Note that x_2, w_2 are the optimal trajectories of the optimisation problem given by (3.4), namely $x_2 = \hat{x}$, $w_2 = \hat{w}$ as defined in Theorem 1.) Then $C_T(w, x(0))$ is equivalently given by:

$$C_T(w,x(0)) = \int_0^T \|\Pi_c(\tilde{z}(t) - z_1(t))\|^2_{R^{-1}} dt + \int_0^\tau \|\Pi(\tilde{z}_2(t) - C_2 x(t) - Dw(t))\|^2_{R^{-1}} dt$$
$$+ \int_\tau^T \|\Pi(\tilde{z}_3(t) - C_3 x(t) - Dw(t))\|^2_{R^{-1}} dt + \|x(\tau) - x_2(\tau)\|^2_{P_2^{-1}(\tau)} \tag{3.81}$$

where C_3 and $\tilde{z}_3$ are given by:

$$C_3 = C_2 + R B_2^T P_2^{-1}, \tag{3.82}$$
$$\tilde{z}_3 = \tilde{z}_2 + R B_2^T P_2^{-1} x_2. \tag{3.83}$$

Proof: Using (3.2), (3.30) and (3.31) we obtain:

$$\Pi(R B_1^T P_1^{-1}(x - x_1) + \tilde{z} - z) = \Pi(\tilde{z}_2 - C_2 x - Dw). \tag{3.84}$$

Substituting (3.84) into (3.16) gives:

$$C_T(w,x(0)) = \int_0^T \|\Pi_c(\tilde{z}(t) - z_1(t))\|^2_{R^{-1}} dt + \int_0^\tau \|\Pi(\tilde{z}_2(t) - C_2 x(t) - Dw(t))\|^2_{R^{-1}} dt$$
$$+ \int_\tau^T \|\Pi(\tilde{z}_2(t) - C_2 x(t) - Dw(t))\|^2_{R^{-1}} dt + \|x(T) - x_1(T)\|^2_{P_1^{-1}(T)}. \tag{3.85}$$

We next note that Lemma 6 remains true for time varying matrices. The proof will apply Lemma 6 to the last two terms of (3.85) in the interval $[\tau, T]$ rather than $[0, T]$. First we make the following notational substitutions: $q \to x$, $u \to w$, $\tilde{y} \to \Pi\tilde{z}_2$, $F \to A$, $G \to B$, $H \to \Pi C_2$,

$J \to D$ and setting $\xi = x_1(T)$, $\Xi = P_1(T)$. Making these replacements in (3.56)–(3.61) and then (3.53) gives $F_1 = A_2$, $G_1 = B_2$, $H_1 = 0$, $\Lambda = \Pi$, $J^\dagger = D^\dagger$ and $M_1 = 0$. Substituting these into (3.51), (3.52) and (3.54) with the notational substitutions $q_1 \to x_2$, $S_1 \to P_2$ and $u_1 \to w_2$ gives (3.78)–(3.80). We next note that the expression in (3.49) with the notational substitutions and the lower limit replaced by τ is the same as the last two terms in (3.85). Therefore, using Lemma 6, we can replace these terms by the expression in (3.62), which gives (3.81) using (3.82) and (3.83) and the fact that the first integral on the right hand side of (3.62) is zero since $(I - \Pi)\Pi = 0$. □

Theorem 4 *The optimisation problem (3.77) has a unique solution $\hat{w}$, $\hat{x}(0)$ which is obtained as follows: first integrate (3.5)–(3.7) forwards in time in the interval $0 \le t \le T$ with initial conditions $P_1(0) = \Gamma$ and $x_1(0) = \gamma$ which gives x_1 and P_1 in the interval $0 \le t \le T$; then integrate (3.78)–(3.80) backwards in time in the interval $\tau \le t \le T$ with terminal conditions $P_2(T) = P_1(T)$ and $x_2(T) = x_1(T)$ which gives x_2 and P_2 in the interval $\tau \le t \le T$; then integrate:*

$$\dot{\hat{x}} = A\hat{x} + B\hat{w} \tag{3.86}$$

backwards in time in the interval $0 \le t \le \tau$ with the feedback law:

$$\hat{w} = D^\dagger(\tilde{z}_2 - C_2\hat{x}); \tag{3.87}$$

and the terminal condition $x(\tau) = \zeta$ to find $\hat{x}$, $\hat{w}$ in the interval $0 \le t \le \tau$; then integrate (3.86) forwards in time in the interval $\tau \le t \le T$ with the feedback law:

$$\hat{w} = D^\dagger(\tilde{z}_3 - C_3\hat{x}) \tag{3.88}$$

and the initial condition $x(\tau) = \zeta$ to find $\hat{x}$, $\hat{w}$ in the interval $\tau \le t \le T$. The minimum of the performance index is:

$$\int_0^T \|\Pi_c(\tilde{z}(t) - z_1(t))\|^2_{R^{-1}} dt + \|\zeta - x_2(\tau)\|^2_{P_2^{-1}(\tau)}. \tag{3.89}$$

Proof: We proceed similarly to the proof of Theorem 1. The first and fourth terms of the performance index in (3.81) are independent of $x(0)$ and w, subject to the constraint $x(\tau) = \zeta$ (i.e. they are constants with respect to the variables of the optimisation problem). The two integrands in (3.81) are identically zero in their respective intervals for $x = \hat{x}$ and $w = \hat{w}$, which can be verified by substitution. Uniqueness is proven similarly to Theorem 1 for the intervals $0 \le t \le \tau$ and $\tau \le t \le T$ separately. □

The solution to the constrained optimisation problem (3.77) given in Theorem 4 introduced the vector and matrix variables x_2, w_2 and P_2. We recall that x_2 and w_2 coincide with the state and input trajectories on the interval $[0,T]$ that minimise the performance criterion (3.3) as shown in Theorem 1. We may now provide an interpretation of the matrix variable P_2. The unique minimum of the constrained optimisation problem (3.77) is given in (3.89) and consists of two terms. The first term coincides with the minimum of the unconstrained problem given in (3.32). The second term is a quadratic which is zero when $\zeta = x_2(\tau)$, in which case we recover the results of Theorem 1. Consider now the eigenvector-eigenvalue decomposition of $P_2(\tau)$. Components of $\zeta - x_2(\tau)$ in those eigenvector directions of $P_2(\tau)$ which have small eigenvalues (i.e. large eigenvalues of $P_2^{-1}(\tau)$) give a large contribution to the second term in (3.89). Hence the measurements provide high confidence that the state $x(\tau)$ should be close to $x_2(\tau)$ in those directions. Fig. 3.2 illustrates the interpretation of $P_2(\tau)$ in the 2-D case. The figure shows an ellipse with centre $x_2(\tau)$ whose axes are aligned with the eigenvectors of $P_2(\tau)$ and lengths given by the corresponding eigenvalue square roots. All points on the ellipse increase the minimum performance index (3.89) by 1.

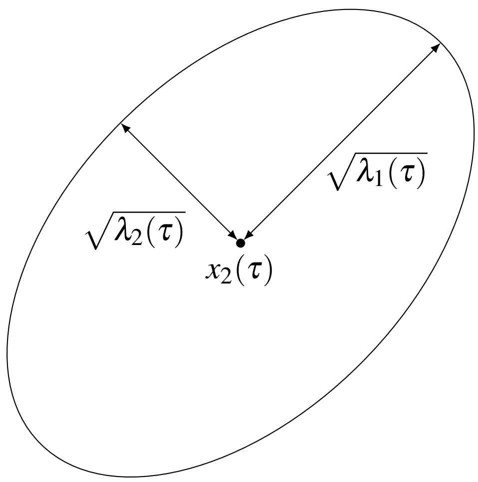

Fig. 3.2 An ellipse with semi-axes of length given by the eigenvalue square roots of $P_2(\tau)$, $\sqrt{\lambda_1(\tau)}$ and $\sqrt{\lambda_2(\tau)}$, and aligned with the corresponding eigenvectors.

3.5 Special cases

3.5.1 Standard Kalman filter

We now show how the continuous time Kalman filter can be derived as a special case of the filter in Theorems 1 and 2. We therefore consider a system described by:

$$\dot{x} = Ax + Bw \tag{3.90}$$

$$z = Cx \tag{3.91}$$

with noisy measurement $\tilde{z}$ of z. Note that we assume as standard that sensor measurements of the state are not directly affected by the process noise w. In the standard Kalman filter the process noise w can be interpreted as a small magnitude disturbance to the system. To translate into our framework we need to incorporate a weighted 2-norm constraint on w in the performance index (3.3). In particular, we consider the following optimisation problem:

$$\inf_{w,x(0)} \left(\int_0^T \|\tilde{z}(t) - z(t)\|^2_{R^{-1}} dt + \int_0^T \|w\|^2_{Q^{-1}} dt + \|x(0) - \gamma\|^2_{\Gamma^{-1}} \right). \tag{3.92}$$

To translate this into the framework of this work we introduce a virtual measurement of w which is equal to zero. More precisely, we consider an augmented system with (real and virtual) outputs given by:

$$z_a = \begin{bmatrix} C \\ 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ \mathrm{I} \end{bmatrix} w \tag{3.93}$$

and for which we have the measurement:

$$\tilde{z}_a = \begin{bmatrix} \tilde{z} \\ 0 \end{bmatrix}. \tag{3.94}$$

We define an augmented block diagonal weighting matrix R_a given by:

$$R_a = \begin{bmatrix} R & 0 \\ 0 & Q \end{bmatrix}. \tag{3.95}$$

The following result is obtained by applying Lemma 5 and Theorem 1 to this augmented system.

Theorem 5 *Consider the system:*

$$\dot{x}_1 = Ax_1 + K(\tilde{z} - Cx_1), \tag{3.96}$$

$$\dot{P}_1 = AP_1 + P_1A^T - KRK^T + BQB^T, \tag{3.97}$$

$$K = P_1C^TR^{-1} \tag{3.98}$$

for $P_1(0) = \Gamma$ and $x_1(0) = \gamma$. The optimisation problem (3.92) where z is defined by (3.90)–(3.91) has a unique solution $\hat{w}$, $\hat{x}(0)$ where $\hat{w}$ is defined by the feedback law:

$$\hat{w} = QB^TP_1^{-1}(\hat{x} - x_1) \tag{3.99}$$

and $\hat{x}(t)$ is obtained by solving $\dot{\hat{x}} = A\hat{x} + B\hat{w}$ backwards on the interval $[0,T]$ with terminal condition $\hat{x}(T) = x_1(T)$. Furthermore, the optimal cost (3.92) is given by:

$$\int_0^T \|\tilde{z} - Cx_1\|^2_{R^{-1}} dt. \tag{3.100}$$

Proof: Replacing (3.2) by (3.93), $\tilde{z}$ and R in (3.3) by (3.94) and (3.95), and applying Lemma 5 gives equations (3.96)–(3.98) after some simplification. Equations (3.99) and (3.100) are obtained by substituting from (3.93), (3.94) and (3.95) into (3.27) and (3.33) respectively. □

The filter (3.96)–(3.98) is an end-of-interval estimator (cf. Theorem 2) in the sense that $\hat{x}(T) = x_1(T)$, $\hat{w}(T) = w_1(T)$ and takes the form of the standard Kalman filter with gain K. The above result reduces to that given in [60] with $R = \mathrm{I}$ and $Q = \mathrm{I}$. It is interesting to note that by substituting for $\hat{w}$ from (3.99) we obtain an equation for the optimal state estimate:

$$\dot{\hat{x}} = A\hat{x} + BQB^TP_1^{-1}(\hat{x} - x_1) \tag{3.101}$$

where $\hat{x}(T) = x_1(T)$ that coincides with the standard form for the smoothed estimate in Kalman filtering (see [34, eqn. 34(a)]). Similarly by specialising (3.79) to the present case we have the equation:

$$\dot{P}_2 = (A + BQB^TP_1^{-1})P_2 + P_2(A + BQB^TP_1^{-1})^T - BQB^T \tag{3.102}$$

where $P_2(T) = P_1(T)$, which is the corresponding form for the smoothed covariance (see [34, eqn. 34(b)]).

3.5.2 Kalman filter with input feedthrough

We consider the extension of the standard Kalman filter to the case where there is direct feedthrough of the input to the measurements. In particular we consider a system described by:

$$\dot{x} = Ax + Bw, \tag{3.103}$$

$$z = Cx + Dw \tag{3.104}$$

with noisy measurement $\tilde{z}$ of z. As in section 3.5.1 we incorporate a weighted 2-norm constraint on w in the performance index (3.3) and consider the optimisation problem (3.92) with z given by (3.104) rather than (3.91). To solve this we proceed similarly and consider an augmented output given by:

$$z_a = \begin{bmatrix} C \\ 0 \end{bmatrix} x + \begin{bmatrix} D \\ \mathrm{I} \end{bmatrix} w \tag{3.105}$$

and for which we have the measurement $\tilde{z}_a$ as in (3.94) and we define an augmented weighting matrix as in (3.95).

Theorem 6 *Consider the system:*

$$\dot{x}_1 = Ax_1 + Bw_1 + K_x(\tilde{z} - Cx_1 - Dw_1), \tag{3.106}$$

$$\dot{P}_1 = AP_1 + P_1A^T - K_xRK_x{}^T + (B - K_xD)P_w(B - K_xD)^T, \tag{3.107}$$

$$w_1 = K_w(\tilde{z} - Cx_1), \tag{3.108}$$

$$K_w = QD^T(DQD^T + R)^{-1}, \tag{3.109}$$

$$K_x = P_1C^TR^{-1}, \tag{3.110}$$

$$P_w = (\mathrm{I} - K_wD)Q \tag{3.111}$$

with $P_1(0) = \Gamma$ and $x_1(0) = \gamma$. The optimisation problem (3.92) where z is given by (3.103)–(3.104) has a unique solution $\hat{w}$, $\hat{x}(0)$ where $\hat{w}$ is defined by the feedback law:

$$\hat{w} = K_w(\tilde{z} - C\hat{x}) + P_wB^TP_1^{-1}(\hat{x} - x_1) \tag{3.112}$$

and $\hat{x}(t)$ is obtained by solving $\dot{\hat{x}} = A\hat{x} + B\hat{w}$ backwards on the interval $[0,T]$ with terminal condition $\hat{x}(T) = x_1(T)$. Furthermore, the optimal cost (3.92) is:

$$\int_0^T \|\tilde{z} - Cx_1\|^2_{(DQD^T+R)^{-1}} dt. \tag{3.113}$$

Proof: We will apply Lemma 5 replacing (3.2) by (3.105), $\tilde{z}$ and R in (3.3) by (3.94) and (3.95). Substituting into (3.15) gives:

$$D^\dagger = \begin{bmatrix} K_w & \mathrm{I} - K_w D \end{bmatrix} \tag{3.114}$$

where we have used the definition (3.109) and the matrix inversion identities (3.1) and (3.2) of Section 6.3 in [1]. Substituting into (3.10), (3.11), (3.12) and (3.8) using (3.114) gives:

$$A_1 = A - BK_wC, \tag{3.115}$$

$$B_1 = B\begin{bmatrix} K_w & \mathrm{I} - K_w D \end{bmatrix}, \tag{3.116}$$

$$C_1 = \begin{bmatrix} \mathrm{I} - DK_w \\ -K_w \end{bmatrix} C \tag{3.117}$$

and (3.108) respectively. Noting the symmetry $DK_wR = (DK_wR)^T$ we find after substituting into (3.7) using (3.117) and the definition (3.110) that:

$$K_1 = K_x \begin{bmatrix} \mathrm{I} - DK_w & -RK_w^T Q^{-1} \end{bmatrix}. \tag{3.118}$$

We now verify:

$$\begin{aligned}
&\begin{bmatrix} \mathrm{I} - DK_w \\ -K_w \end{bmatrix}^T \begin{bmatrix} R & 0 \\ 0 & Q \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{I} - DK_w \\ -K_w \end{bmatrix} \\
&\quad = R^{-1}(\mathrm{I} - DK_w) + K_w^T D^T R^{-1}(-\mathrm{I} + DK_w + R(DQD^T + R)^{-1}) \\
&\quad = R^{-1}(\mathrm{I} - DK_w) && (3.119) \\
&\quad = (DQD^T + R)^{-1} && (3.120) \\
&\quad = R^{-1}(R - DP_wD^T)R^{-1}. && (3.121)
\end{aligned}$$

From (3.7) using (3.117), (3.119) and (3.110) we obtain:

$$K_1C_1 = K_x(\mathrm{I} - DK_w)C. \tag{3.122}$$

Similarly from (3.7) using (3.117), (3.121) and (3.110) we obtain:

$$K_1 R_a K_1^T = K_x R K_x^T - (K_x D) P_w (K_x D)^T. \quad (3.123)$$

Substituting into (3.5) using (3.115), (3.122), (3.116), (3.118) and (3.108) gives (3.106). Using (3.109), (3.110) and (3.111) we obtain:

$$K_w C P_1 = P_w (K_x D)^T. \quad (3.124)$$

By substituting from (3.114) we obtain:

$$D^{\dagger} R_a D^{\dagger T} = P_w. \quad (3.125)$$

Substituting into (3.6) using (3.115), (3.123), (3.124) and (3.125) gives (3.107). Equation (3.112) follows from (3.27) using (3.114) and (3.125). Equation (3.113) follows from the first expression in (3.33) using (3.114) and (3.120). □

We again note that by substituting for $\hat{w}$ from (3.112) we obtain the differential equation for the optimal state estimate:

$$\dot{\hat{x}} = A\hat{x} + BK_w(\tilde{z} - C\hat{x}) + BP_w B^T P_1^{-1}(\hat{x} - x_1) \quad (3.126)$$

where $\hat{x}(T) = x_1(T)$. Similarly (3.79) specialises to:

$$\dot{P}_2 = (A - BK_w C + BP_w B^T P_1^{-1})P_2 + P_2(A - BK_w C + BP_w B^T P_1^{-1})^T - BP_w B^T. \quad (3.127)$$

Finally, we point out that the augmented input feedthrough matrix in (3.105) is full column rank for all D. Furthermore, the regular Kalman filter can be recovered trivially by setting $D = 0$ in (3.106)–(3.111), (3.126) and (3.127).

3.5.3 Standard linear quadratic regulator

We now show how the solution of the standard linear quadratic regulator (LQR) on a finite time horizon can be derived as a special case of the tracking problem in Theorem 3. In the standard LQR we wish to find a low energy input u that brings the state q to the origin. More precisely, we consider the optimisation problem:

$$\inf_u \left(\int_0^T \|q(t)\|^2_{R_q^{-1}} + \|u(t)\|^2_{R_u^{-1}} dt + \|q(T)\|^2_{\Xi^{-1}} \right) \quad (3.128)$$

where the state q and input u satisfy:

$$\dot{q} = Fq + Gu \tag{3.129}$$

and the initial state $q(0) = \eta$ is known. To put this into the form required to apply Lemma 6 and Theorem 3 we choose:

$$y = \begin{bmatrix} \mathrm{I} \\ 0 \end{bmatrix} q + \begin{bmatrix} 0 \\ \mathrm{I} \end{bmatrix} u, \tag{3.130}$$

$\tilde{y} = 0$, $\xi = 0$ and:

$$R = \begin{bmatrix} R_q & 0 \\ 0 & R_u \end{bmatrix}. \tag{3.131}$$

Theorem 7 *Consider the RDE:*

$$\dot{S_1} = FS_1 + S_1F^T + S_1R_q^{-1}S_1^T - GR_uG^T \tag{3.132}$$

with the terminal condition $S_1(T) = \Xi$. The optimisation problem (3.128) has a unique solution $\hat{u}$ which is defined by the feedback law:

$$\hat{u} = -R_uG^TS_1^{-1}\hat{q} \tag{3.133}$$

and $\hat{q}(t)$ is obtained by solving $\dot{\hat{q}} = F\hat{q} + G\hat{u}$ forwards in the interval $[0,T]$ with the initial condition $\hat{q}(0) = \eta$. Furthermore, the optimal cost (3.128) is:

$$\|\eta\|^2_{S_1(0)^{-1}}. \tag{3.134}$$

Proof: We apply Lemma 6 with y, R given by (3.130), (3.131). Note that $J^\dagger = \begin{bmatrix} 0 & \mathrm{I} \end{bmatrix}$, $F_1 = F$, $G_1 = G\begin{bmatrix} 0 & \mathrm{I} \end{bmatrix}$ and $H_1 = \begin{bmatrix} \mathrm{I} & 0 \end{bmatrix}^T$. Since $\tilde{y} = 0$, $q_1(T) = \xi = 0$ we have $q_1(t) = 0$ for all t. Equation (3.132) follows from (3.52). Applying Theorem 3 we obtain (3.133) and (3.134) by substituting into (3.64) and (3.67) respectively. □

This is recognised as the classical LQR result on a finite time horizon.

3.6 Steady state filtering

3.6.1 Stability

We first consider the convergence properties of the filter of Lemma 5 (end of interval estimator) as $T \to \infty$. In order to express convergence conditions directly in terms of A, B, C, D we first need to establish the following two Lemmas.

Lemma 8 *Let D have full column rank.* $s_0 \in \mathbb{C}$ *is an uncontrollable mode of* (A_1, B_1) *if and only if it is an uncontrollable mode of* (A, B).

Proof: If s_0 is an uncontrollable mode of (A, B) then there exists $0 \neq x \in \mathbb{C}^n$ such that $x^* A = x^* s_0$ and $x^* B = 0$. Hence $x^*(A - BD^\dagger C) = x^* s_0$ and $x^* BD^\dagger = 0$. The converse follows since $D^\dagger$ has full row rank. □

Lemma 9 *Let D have full column rank.* $s_0 \in \mathbb{C}$ *is an unobservable mode of* (C_1, A_1) *if and only if it is an invariant zero of the system (3.1)–(3.2), i.e.* $\begin{bmatrix} A - s_0 \mathrm{I} & B \\ C & D \end{bmatrix}$ *does not have full column rank.*

Proof: The proof is a more general result to [63, Lemma 13.9] to all system modes. □

Theorem 8 *Suppose* (A, B) *has no uncontrollable mode* $s_0 \in \mathbb{C}$ *with* $Re(s_0) = 0$*, the system (3.1)–(3.2) has no invariant zero* $s_0 \in \mathbb{C}$ *with* $Re(s_0) \geq 0$ *and* $\tilde{z}(t) \in \mathscr{L}_\infty^p[0, \infty)$*. Then the ARE:*

$$A_1 P_1^\infty + P_1^\infty A_1^T - K_1^\infty R K_1^{\infty T} + B_1 R B_1^T = 0 \tag{3.135}$$

where $K_1^\infty = P_1^\infty C_1^T R^{-1}$ *has a unique solution* P_1^∞ *such that:*

$$A^\infty = A_1 - K_1^\infty C_1 \tag{3.136}$$

is Hurwitz. Furthermore, $P_1^\infty \geq 0$ *and* $P_1(t) \to P_1^\infty$ *as* $t \to \infty$ *where* $P_1(t)$ *is given by (3.6) with the initial condition* $P_1(0) = \Gamma$*. Consider the system:*

$$\dot{x}_1^\infty = A^\infty x_1^\infty + B^\infty \tilde{z}, \tag{3.137}$$

$$w_1^\infty = D^\dagger (\tilde{z} - C x_1^\infty) \tag{3.138}$$

with any initial condition $x_1^\infty(0) \in \mathbb{R}^n$ *where:*

$$B^\infty = B_1 + K_1^\infty. \tag{3.139}$$

Then:

1. $x_1(t) - x_1^\infty(t) \to 0$ *as* $t \to \infty$*;*

2. $w_1(t) - w_1^\infty(t) \to 0$ *as* $t \to \infty$

where $x_1(t)$ *and* $w_1(t)$ *are given by (3.5) and (3.8) with the initial condition* $x_1(0) = \gamma$*.*

Proof: The claims in regard to (3.135) follow directly by applying Lemmas 19 and 20 to the RDE (3.6) and expressing the convergence conditions in terms of A,B,C,D using Lemmas 8 and 9. The convergence results 1) and 2) follow from Lemma 26. □

We remark that the system (3.137)–(3.138) is the "limiting form" of the end-of-interval estimator of Lemma 5 in which $P_1(t)$ is replaced by P_1^∞. We do not assert any convergence property other than 1) and 2) in Theorem 8.

3.6.2 The steady state filter as a stable left inverse

We adopt the notation introduced in [63, Ch. 3] and denote the transfer function of (3.1)–(3.2) by:

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = C(s\mathrm{I} - \mathrm{A})^{-1}B + D. \tag{3.140}$$

We consider the transfer function of the steady state filter (3.137)–(3.138):

$$\left[\begin{array}{c|c} A^\infty & B^\infty \\ \hline -D^\dagger C & D^\dagger \end{array}\right]. \tag{3.141}$$

We will now show that under certain conditions (3.140) is stable left inverse of the system. Conditions for the existence and stability of inverses of linear time invariant systems are available in [50] and an example of an inverse when D is full column rank can be found in [63, Ch. 3].

Theorem 9 *Suppose* (A,B) *has no uncontrollable mode* $s_0 \in \mathbb{C}$ *with* $Re(s_0) = 0$ *and the system (3.1)–(3.2) has no invariant zero* $s_0 \in \mathbb{C}$ *with* $Re(s_0) \geq 0$*. Then (3.141) is a stable left inverse of (3.140).*

Proof: To see this we consider the cascade connection of (3.141) with (3.140) and verify the calculation:

$$\left[\begin{array}{c|c} A^\infty & B^\infty \\ \hline -D^\dagger C & D^\dagger \end{array}\right] \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = \left[\begin{array}{cc|c} A^\infty & B^\infty C & B^\infty D \\ 0 & A & B \\ \hline -D^\dagger C & D^\dagger C & D^\dagger D \end{array}\right] = \left[\begin{array}{cc|c} A^\infty & B^\infty C & B \\ 0 & A & B \\ \hline -D^\dagger C & D^\dagger C & \mathrm{I} \end{array}\right] \tag{3.142}$$

where we have used the transfer matrix product operation in [63, Sec. 3.6] and noting that $K_1^\infty D = 0$ (cf. (3.18)). The product is equivalently given by:

$$\left[\begin{array}{cc|c} A^\infty & 0 & 0 \\ 0 & A & B \\ \hline -D^\dagger C & 0 & \mathrm{I} \end{array}\right] = \mathrm{I} \tag{3.143}$$

using the similarity transformation:

$$\begin{bmatrix} \mathrm{I} & -\mathrm{I} \\ 0 & \mathrm{I} \end{bmatrix}$$

after some simplification and noting that $K_1^\infty C_1 = K_1^\infty C$ (cf. (3.19)). We recall from Theorem 8 that A^∞ is Hurwitz. □

3.6.3 Filtering error

We now return to the series connection of the original system, sensors and filter as shown in Fig. 3.1. Suppose w and x are *any* input and state which satisfy equations (3.1)–(3.2). This may or may not be the trajectory which generated the measured output $\tilde{z}$ in an experiment. We may define the state estimation errors $e = x - x_1$ and $e^\infty = x - x_1^\infty$. Let the conditions of Theorem 8 hold. Then the following equations follow from (3.1), (3.2), (3.5) and (3.137):

$$\dot{e} = (A_1 - K_1 C_1)e + (B_1 + K_1)(z - \tilde{z}), \tag{3.144}$$

$$\dot{e}^\infty = A^\infty e^\infty + B^\infty (z - \tilde{z}). \tag{3.145}$$

We note that e^∞ is the state of an asymptotically stable linear time invariant system driven by the error $z - \tilde{z}$ and that $e^\infty(t) - e(t) \to 0$ as $t \to \infty$ (by Lemma 26).

3.6.4 The infinite time smoother as a left inverse

Lemma 10 *Suppose (A,B) has no uncontrollable mode $s_0 \in \mathbb{C}$ with $Re(s_0) \leq 0$ and the system (3.1)–(3.2) has no invariant zero $s_0 \in \mathbb{C}$ with $Re(s_0) \geq 0$. Then $P_1^\infty > 0$. Furthermore, let:*

$$A_2^\infty = A - B_1 C_2^\infty, \tag{3.146}$$

$$C_2^\infty = C - R B_1^T (P_1^\infty)^{-1}. \tag{3.147}$$

Then $-A_2^\infty$ *is Hurwitz.*

Proof: $P_1^\infty > 0$ follows by applying Lemma 19 to the ARE (3.135) using Lemmas 8 and 9 to express the convergence conditions in terms of A,B,C,D. Now note that:

$$A_2^\infty = A_1 + B_1 R B_1^T (P_1^\infty)^{-1} \tag{3.148}$$

by substituting (3.147) into (3.146) and then using (3.10). We may then verify that:

$$A_2^\infty P_1^\infty + P_1^\infty (A^\infty)^T = 0 \tag{3.149}$$

by substituting for A^∞, A_2^∞ and using (3.135). Hence $(A^\infty)^T$ and $-A_2^\infty$ are similar which means that $-A_2^\infty$ is Hurwitz. □

We now assume that the conditions of Lemma 10 hold and consider the cascade connection of two systems. The first system has input $\tilde{z}$ and output $\tilde{z}_2^\infty$. It is given by:

$$\dot{x}_1^\infty = A^\infty x_1^\infty + B^\infty \tilde{z}, \tag{3.150}$$

$$\tilde{z}_2^\infty = \tilde{z} - R B_1^T (P_1^\infty)^{-1} x_1^\infty \tag{3.151}$$

and has the transfer function:

$$\left[\begin{array}{c|c} A^\infty & B^\infty \\ \hline -RB_1^T (P_1^\infty)^{-1} & \mathrm{I} \end{array}\right]. \tag{3.152}$$

(Note that (3.150) coincides with (3.137).) The second system is driven by the output of the first and has output $\hat{w}^\infty$. It is given by:

$$\dot{\hat{x}}^\infty = A_2^\infty \hat{x}^\infty + B_2 \tilde{z}_2^\infty, \tag{3.153}$$

$$\hat{w}^\infty = D^\dagger (\tilde{z}_2^\infty - C_2^\infty \hat{x}^\infty) \tag{3.154}$$

and has the transfer function:

$$\left[\begin{array}{c|c} A_2^\infty & B_2 \\ \hline -D^\dagger C_2^\infty & D^\dagger \end{array}\right]. \tag{3.155}$$

This cascade connection is the "limiting form" of the construction for the optimal estimator of Theorem 1 and is shown next to be a left inverse of the original system. We do not assert any formal convergence property for this cascade connection.

Theorem 10 *Suppose (A,B) has no uncontrollable mode $s_0 \in \mathbb{C}$ with $Re(s_0) \leq 0$ and the system (3.1)–(3.2) has no invariant zero $s_0 \in \mathbb{C}$ with $Re(s_0) \geq 0$. Then the cascade connection of (3.152) with (3.155):*

$$\left[\begin{array}{c|c} A_2^{\infty} & B_2 \\ \hline -D^{\dagger}C_2^{\infty} & D^{\dagger} \end{array}\right]\left[\begin{array}{c|c} A^{\infty} & B^{\infty} \\ \hline -RB_1^T(P_1^{\infty})^{-1} & \mathrm{I} \end{array}\right] \tag{3.156}$$

is a left inverse of the system (3.1)–(3.2).

Proof: To see this consider the cascade connection of (3.156) with (3.140) which is given by:

$$\left[\begin{array}{ccc|c} A_2^{\infty} & -B_1RB_1^T(P_1^{\infty})^{-1} & B_2C & B \\ 0 & A^{\infty} & B^{\infty}C & B \\ 0 & 0 & A & B \\ \hline -D^{\dagger}C_2^{\infty} & -D^{\dagger}RB_1^T(P_1^{\infty})^{-1} & D^{\dagger}C & \mathrm{I} \end{array}\right] \tag{3.157}$$

where we have used the transfer matrix product operation and noting that $B_2D = B$ and $B^{\infty}D = B$ (cf. proof of Theorem 9). The product is equivalently given by:

$$\left[\begin{array}{cc:c|c} A_2^{\infty} & -B_1RB_1^T(P_1^{\infty})^{-1} & 0 & 0 \\ 0 & A^{\infty} & 0 & 0 \\ \hdashline 0 & 0 & A & B \\ \hline -D^{\dagger}C_2^{\infty} & -D^{\dagger}RB_1^T(P_1^{\infty})^{-1} & 0 & \mathrm{I} \end{array}\right] = \mathrm{I} \tag{3.158}$$

using the similarity transformation:

$$\begin{bmatrix} \mathrm{I} & 0 & -\mathrm{I} \\ 0 & \mathrm{I} & -\mathrm{I} \\ 0 & 0 & \mathrm{I} \end{bmatrix}$$

after some simplification. □

Remark 2 *The conditions of Theorem 10 can be written is several alternative ways, e.g. (C_1,A_1) detectable and $(-A_1,B_1)$ stabilizable, or equivalently $(-A,B)$ stabilizable. We can interpret this as a forwards in time detectability condition for the first stage of the inversion and a backwards in time stabilisability condition for the second stage.*

Remark 3 *It is interesting to note that there is a natural state transformation given by $x_2 = (P_1^{\infty})^{-1}x_1$ which leads to the alternative state space representation of the transfer* *function in (3.152):*

$$\left[\begin{array}{c|c} (-A_2^\infty)^T & (P_1^\infty)^{-1}B^\infty \\ \hline -RB_1^T & \mathrm{I} \end{array}\right]. \tag{3.159}$$

We will make use of the analogous state transformation for the infinite horizon tracking problem in the next section.

3.7 Infinite horizon tracking

3.7.1 An anti-stable left inverse

We begin by considering the convergence of the construction of Lemma 6 as the horizon length increases. We show that $q_1(t)$ and $u_1(t)$ converge *for any fixed t* to the state and input of an anti-stable time invariant system solved backwards in time. For convenience we first restate Lemmas 8 and 9 with the relevant notational substitutions.

Lemma 11 *Let J have full column rank. $s_0 \in \mathbb{C}$ is an uncontrollable mode of (F_1, G_1) if and only if it is an uncontrollable mode of (F, G).* □

Lemma 12 *Let J have full column rank. $s_0 \in \mathbb{C}$ is an unobservable mode of (H_1, F_1) if and only if it is an invariant zero of the system (3.47)–(3.48).* □

Theorem 11 *Suppose (F, G) has no uncontrollable mode $s_0 \in \mathbb{C}$ with $Re(s_0) = 0$, the system (3.47)–(3.48) has no invariant zero $s_0 \in \mathbb{C}$ with $Re(s_0) \leq 0$ and $\tilde{y}(t) \in \mathscr{L}_\infty^p[0, \infty)$. Then the ARE:*

$$F_1 S_1^\infty + S_1^\infty F_1^T + M_1^\infty R M_1^{\infty T} - G_1 R G_1^T = 0 \tag{3.160}$$

where $M_1^\infty = S_1^\infty H_1^T R^{-1}$ has a unique solution S_1^∞ such that $-F^\infty$ is Hurwitz where:

$$F^\infty = F_1 + M_1^\infty H_1. \tag{3.161}$$

Furthermore, $S_1^\infty \geq 0$ and $S_1(t, T) \to S_1^\infty$ as $T \to \infty$ for any fixed $t \geq 0$ where $S_1(t, T)$ equals $S_1(t)$ in (3.52) with the terminal condition $S_1(T) = \Xi$. Consider the system:

$$\dot{q}_1^\infty = F^\infty q_1^\infty + G^\infty \tilde{y}, \tag{3.162}$$

$$u_1^\infty = J^\dagger(\tilde{y} - H q_1^\infty) \tag{3.163}$$

with any terminal condition $q_1^\infty(T) \in \mathbb{R}^n$ *where:*

$$G^\infty = G_1 - M_1^\infty. \tag{3.164}$$

Then:

1. $q_1(t,T) - q_1^\infty(t,T) \to 0$ *as* $T \to \infty$ *for any fixed* $t \geq 0$*;*
2. $u_1(t,T) - u_1^\infty(t,T) \to 0$ *as* $T \to \infty$ *for any fixed* $t \geq 0$

where $q_1(t,T)$ *equals* $q_1(t)$ *in (3.51) with the terminal condition* $q_1(T) = \xi$ *and* $u_1(t,T)$ *equals* $u_1(t)$ *in (3.54).*

Proof: The claims in regard to (3.160) follow directly by applying Remark 1 to the RDE (3.52) and expressing the convergence conditions in terms of F, G, H, J using Lemmas 11 and 12. The convergence results 1) and 2) follow from the time reversed Lemma 26. □

Remark 4 *It is interesting to note the contrasting form of of 1) and 2) in Theorems 8 and 11. At first sight this is unexpected since the estimation and tracking problems are dual to each other. The difference arises since the infinite horizon limit of the time window* $[0,T]$ *is taken to be* $[0,\infty)$ *in both cases, namely the left hand limit is fixed at the origin while the right hand limit tends to* ∞*, which is not symmetric since the two problems are dual by time reversal.*

We now consider the transfer function of the system (3.162)–(3.163):

$$\left[\begin{array}{c|c} F^\infty & G^\infty \\ \hline -J^\dagger H & J^\dagger \end{array}\right]. \tag{3.165}$$

Theorem 12 *Suppose* (F,G) *has no uncontrollable mode* $s_0 \in \mathbb{C}$ *with* $Re(s_0) = 0$ *and the system (3.47)–(3.48) has no invariant zero* $s_0 \in \mathbb{C}$ *with* $Re(s_0) \leq 0$*. Then (3.165) is an anti-stable left inverse of the system (3.47)–(3.48).*

Proof: To see this we consider the cascade connection of (3.165) with the transfer function of the system (3.47)–(3.48) and then perform the transfer matrix product and similarity transformation similarly to the proof of Theorem 9. □

3.7.2 The infinite horizon controller

We now consider the convergence properties of the unique solution $\hat{u}$ of the tracking problem (3.50), which is given in Theorem 3, in the infinite time horizon limit, i.e. as $T \to \infty$.

Theorem 13 *Suppose (F,G) has no uncontrollable mode $s_0 \in \mathbb{C}$ with $Re(s_0) \geq 0$, i.e. it is stabilizable, the system (3.47)–(3.48) has no invariant zero $s_0 \in \mathbb{C}$ with $Re(s_0) = 0$ and $\tilde{y}(t) \in \mathscr{L}_\infty[0,\infty)$. Then the ARE:*

$$S_2^\infty F_1 + F_1^T S_2^\infty - S_2^\infty G_1 R G_1^T S_2^\infty + H_1^T R^{-1} H_1 = 0 \tag{3.166}$$

has a unique solution S_2^∞ that is stabilising, i.e. F_2^∞ in (3.169) is Hurwitz, and $S_2^\infty \geq 0$. Consider the system:

$$\dot{\hat{q}}^\infty(t) = F\hat{q}^\infty(t) + G\hat{u}^\infty(t), \tag{3.167}$$

$$\hat{u}^\infty(t) = J^\dagger(\tilde{y}_2^\infty(t) - H_2^\infty \hat{q}^\infty(t)) \tag{3.168}$$

with the initial condition $\hat{q}^\infty(0) = \eta$ where:

$$F_2^\infty = F - G_1 H_2^\infty, \tag{3.169}$$

$$G_2^\infty = H_1^T R^{-1} - S_2^\infty G_1, \tag{3.170}$$

$$H_2^\infty = H + RG_1^T S_2^\infty, \tag{3.171}$$

$$\tilde{y}_2^\infty(t) = \tilde{y}(t) + RG_1^T q_2^\infty(t), \tag{3.172}$$

$$q_2^\infty(t) = \int_t^\infty e^{F_2^{\infty T}(\tau - t)} G_2^\infty \tilde{y}(\tau) d\tau. \tag{3.173}$$

Then the unique optimal control input $\hat{u}(t,T)$ of the tracking problem (3.50) with the initial condition $q(0) = \eta$ (i.e. $\hat{u}$ as defined in (3.64)) converges as $T \to \infty$, i.e. $\lim_{T\to\infty} \hat{u}(t,T)$ exists for any fixed $t \geq 0$, and the limit is given by $\hat{u}^\infty(t)$.

Proof: Let S_1, q_1 be defined by (3.51), (3.52) with the terminal conditions $S_1(T) = \Xi$, $q_1(T) = \xi$. We introduce the variables:

$$S_2 = S_1^{-1}, \tag{3.174}$$

$$q_2 = S_1^{-1} q_1. \tag{3.175}$$

Hence S_2, q_2 are generated by solving:

$$\dot{S}_2 = -S_2 F_1 - F_1^T S_2 + S_2 G_1 R G_1^T S_2 - H_1^T R^{-1} H_1, \tag{3.176}$$

$$\dot{q}_2 = \Big(-F_1^T + S_2 G_1 R G_1^T\Big) q_2 + \Big(S_2 G_1 - H_1^T R^{-1}\Big)\tilde{y} \tag{3.177}$$

backwards in time with the given terminal conditions:

$$S_2(T) = S_1^{-1}(T) = \Xi^{-1}, \tag{3.178}$$

$$q_2(T) = S_1^{-1}(T)q_1(T) = \Xi^{-1}\xi. \tag{3.179}$$

We now apply Remark 1 to the ARE (3.166) and the RDE (3.176). The conditions of the theorem are obtained in terms of F, G, H, J using Lemmas 11 and 12. Furthermore F_2^∞ is Hurwitz and:

$$\lim_{T\to\infty} S_1^{-1}(t,T) = \lim_{T\to\infty} S_2(t,T) = S_2^\infty \tag{3.180}$$

for all $\Xi > 0$ and for any fixed $t \geq 0$. We introduce the anti-stable and time-invariant equation:

$$\dot{q}_2^\infty = -F_2^{\infty T} q_2^\infty - G_2^\infty \tilde{y}. \tag{3.181}$$

We now consider the time reversed equations (3.177) and (3.181) for a given T (i.e. solved forwards in time). These equations take the form of (A.28) and (A.29) on the interval $[0,T]$, where we note that $u(t)$ depends on T, but with $\sup_{0\leq t\leq T} |u(t)|_\infty \leq \|\tilde{y}(t)\|_\infty$ for any T. Now choose any $\varepsilon > 0$ and find the T_0 guaranteed by Lemma 26. Then the time reversed solutions are within ε in norm for $T_0 \leq t \leq T$ providing $T_0 \leq T$. Hence, for any $T > T_0$, $|q_2(t) - q_2^\infty(t)|_\infty < \varepsilon$ for $0 \leq t \leq T - T_0$. It follows that:

$$\lim_{T\to\infty} q_2(t,T) = q_2^\infty(t) \tag{3.182}$$

for any fixed $t \geq 0$, where $q_2^\infty(t)$ is given by the convolution form (3.173). We now return to compute the limit of the unique optimal input $\hat{u}(t,T)$ given in Theorem 3, i.e. $\lim_{T\to\infty} \hat{u}(t,T)$. Taking the limit in (3.65), (3.66) and substituting from (3.180), (3.182) gives:

$$\lim_{T\to\infty} H_2(t,T) = H_2^\infty, \tag{3.183}$$

$$\lim_{T\to\infty} \tilde{y}_2(t,T) = \tilde{y}_2^\infty(t) \tag{3.184}$$

where $\tilde{y}_2^\infty(t)$ and H_2^∞ are given by (3.169)–(3.173). Rewriting (3.63) for the infinite horizon with a notational substitution and taking the limit in (3.64) gives the feedback law (3.168). □

Remark 5 *Evaluating q_2^∞ for all finite t using (3.173) is costly even if it is possible. It can be approximated for any finite t_0 as accurately as required by:*

$$q_2^\lambda(t_0) = \int_{t_0}^{t_0+\lambda} e^{F_2^{\infty T}(\tau - t_0)} G_2^\infty \tilde{y}(\tau) d\tau \tag{3.185}$$

for a sufficiently large $\lambda > 0$ since F_2^∞ is Hurwitz. Integrating (3.181) forwards in time for $t > t_0$ with the initial condition $q_2^\lambda(t_0)$ obtained from (3.185) gives $q_2^\infty(t)$ approximately for t near t_0 but errors amplify since (3.181) is anti-stable. A practical compromise is to evaluate (3.185) at regular intervals and to integrate (3.181) within those intervals.

3.7.3 The steady state controller as an unstable right inverse

Theorem 14 *Suppose (F,G) has no uncontrollable mode $s_0 \in \mathbb{C}$ with $Re(s_0) \geq 0$, i.e. it is stabilizable, and the system (3.47)–(3.48) has no invariant zero $s_0 \in \mathbb{C}$ with $Re(s_0) = 0$. Then the infinite horizon controller is given by the cascade connection of an anti-stable system with input $\tilde{y}$ and output $\tilde{y}_2^\infty$ and transfer function:*

$$\left[\begin{array}{c|c} (-F_2^\infty)^T & -G_2^\infty \\ \hline RG_1^T & I \end{array}\right] \tag{3.186}$$

followed by a stable system with input $\tilde{y}_2^\infty$ and output $\hat{u}$ and transfer function:

$$\left[\begin{array}{c|c} F_2^\infty & G_1 \\ \hline -J^\dagger H_2^\infty & J^\dagger \end{array}\right]. \tag{3.187}$$

Furthermore, their cascade connection, given by:

$$\left[\begin{array}{c|c} F_2^\infty & G_1 \\ \hline -J^\dagger H_2^\infty & J^\dagger \end{array}\right] \left[\begin{array}{c|c} (-F_2^\infty)^T & -G_2^\infty \\ \hline RG_1^T & I \end{array}\right] \tag{3.188}$$

is a left inverse of the system (3.47)–(3.48).

Proof: Apply the transfer matrix product operation and a system similarity transformation similarly to the proof of Theorem 10. Note that F_2^∞ is Hurwitz from Theorem 13. □

Chapter 4

Estimation in systems with discrete time measurements

4.1 A stochastic formulation: The zero informational input limit

4.1.1 Introduction

In this section we aim to simultaneously estimate the inputs and states of a *stochastic discrete time* linear finite dimensional dynamical system where no direct measurement of the driving inputs is available. In particular, we consider a system with the state space description:

$$x_{k+1} = Ax_k + Bw_k \tag{4.1}$$

$$y_k = Cx_k + Dw_k + v_k \tag{4.2}$$

where the system matrices are assumed known and the initial state x_0 and the measurement noise v_k for all $k \in \mathbb{N}^+$ are uncorrelated Gaussian random variables with known probability distributions. There is a wide variety of problem formulations and assumptions in the literature of this field, we have given a brief account in Section 2.1.2. We approach this problem similarly to [5], namely we let the input w_k be a Gaussian white noise process of known covariance and consider the limit as the inverse of the input covariance tends to zero (the zero informational limit). We consider the case that there is direct feedthrough of the unknown inputs to the measurement vector through a matrix that is full column rank. In comparison, [5] considers systems without feedthrough and a full column rank Markov parameter. The approach differs from that of Chapter 3, which formulates an optimisation problem for *deterministic continuous time* systems. The filter recursions take an interesting form which closely relates to the filter of Gillijns and De Moor [20]. However, the filter of [20] was derived using a hybrid approach, involving both the computation of expectations and the formulation of an optimisation problem. More specifically, the contributions of this section are as follows:

1. To derive directly a (first) form of the Kalman filter in the zero informational limit on the input when the input feedthrough matrix is full column rank; to note that this limit filter is closely related to the recursive filter of [20] and thereby to provide a simpler notion of optimality for that filter.

2. To show that the limit filter equations can be transformed to an alternative (second) form which is a standard Kalman filter for a new system.

3. To show that the second form of the limit filter allows necessary and sufficient conditions for the stability and convergence of the filter to be stated, which may be expressed as a controllability condition and a minimum phase condition in terms of the invariant zeros of the original system.

4.1.2 Estimation

We derive the zero informational limit of the Kalman filter with feedthrough (see Section 2.2.5), namely the limit as the information matrix $Q^{-1} \to 0$, under the assumption that D has full column rank. We take $P_{0|-1}$, $\hat{x}_{0|-1}$ and $z_0, z_1, z_2, \dots$ as given and consider the limits of $\hat{x}_{k|k}$, $\hat{w}_{k|k}$, $P_{k|k}$, $P^{ww}_{k|k}$, $P^{xw}_{k|k}$, $\hat{x}_{k+1|k}$ and $P_{k+1|k}$ as the information matrix $Q^{-1} \to 0$. We show that the limits exist for all k and introduce the notation:

$$\bar{x}_{k|k} = \lim_{Q^{-1} \to 0} \hat{x}_{k|k} \tag{4.3}$$

and similarly for $\bar{w}_{k|k}$, $\bar{P}_{k|k}$, etc. We name the resulting recursive equations the *limit filter*. We first derive a form of the limit filter written directly in terms of the system matrices A, B, C, D. For convenience we state the matrix inversion lemma (see [1, Sec. 6.3]) which is used in the proof.

Lemma 13 *Let $S = S^T > 0$, $T = T^T > 0$ and H be arbitrary of compatible dimension. Then:*

$$S - SH^T(HSH^T + T)^{-1}HS = (S^{-1} + H^T T^{-1} H)^{-1}, \tag{4.4}$$

$$SH^T(HSH^T + T)^{-1} = (S^{-1} + H^T T^{-1} H)^{-1} H^T T^{-1}. \tag{4.5}$$

Theorem 15 *Let D have full column rank. Then $\hat{x}_{k|k}(Q)$, $\hat{w}_{k|k}(Q)$, $P_{k|k}(Q)$, etc. for all k have well-defined limits $\bar{x}_{k|k}$, $\bar{w}_{k|k}$, $\bar{P}_{k|k}$, etc. as $Q^{-1} \to 0$ given recursively by:*

$$\bar{x}_{k|k} = \bar{x}_{k|k-1} + L_{x,k}(z_k - C\bar{x}_{k|k-1} - D\bar{w}_{k|k}), \tag{4.6}$$

$$\bar{w}_{k|k} = L_{w,k}(z_k - C\bar{x}_{k|k-1}), \tag{4.7}$$

$$\bar{P}_{k|k} = \bar{P}_{k|k-1} - L_{x,k}(\bar{\Sigma}_k - D\bar{P}^{ww}_{k|k}D^T)L^T_{x,k}, \tag{4.8}$$

$$\bar{P}^{ww}_{k|k} = L_{w,k}\bar{\Sigma}_k L^T_{w,k}, \tag{4.9}$$

$$\bar{P}^{xw}_{k|k} = -L_{x,k}\bar{\Sigma}_k L^T_{w,k}, \tag{4.10}$$

$$L_{x,k} = \bar{P}_{k|k-1}C^T\bar{\Sigma}_k^{-1}, \tag{4.11}$$

$$L_{w,k} = (D^T\bar{\Sigma}_k^{-1}D)^{-1}D^T\bar{\Sigma}_k^{-1}, \tag{4.12}$$

$$\bar{\Sigma}_k = C\bar{P}_{k|k-1}C^T + R, \tag{4.13}$$

$$\bar{x}_{k+1|k} = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} \bar{x}_{k|k} \\ \bar{w}_{k|k} \end{bmatrix}, \tag{4.14}$$

$$\bar{P}_{k+1|k} = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} \bar{P}_{k|k} & \bar{P}^{xw}_{k|k} \\ \bar{P}^{wx}_{k|k} & \bar{P}^{ww}_{k|k} \end{bmatrix} \begin{bmatrix} A & B \end{bmatrix}^T \tag{4.15}$$

with $\bar{P}_{0|-1} = P_{0|-1}$ and $\bar{x}_{0|-1} = \hat{x}_{0|-1}$.

Proof: We will proceed inductively, first considering the recursive expressions (2.23)–(2.32) in the limit as $Q^{-1} \to 0$. Suppose for a given k, $\hat{x}_{k|k-1}(Q)$ and $P_{k|k-1}(Q)$ tend to well-defined limits $\bar{x}_{k|k-1}$ and $\bar{P}_{k|k-1}(Q)$ as $Q^{-1} \to 0$. We will show that $\hat{x}_{k|k}(Q)$, $\hat{w}_{k|k}(Q)$, $P_{k|k}(Q)$, $P^{ww}_{k|k}(Q)$, $P^{xw}_{k|k}(Q)$, $\hat{x}_{k+1|k}(Q)$, $P_{k+1|k}(Q)$ have well-defined limits given by (4.6)–(4.15). Substituting for Θ_k from (2.30) into (2.29) and using (4.5) gives:

$$K_{w,k} = (Q^{-1} + D^T\Sigma_k^{-1}D)^{-1}D^T\Sigma_k^{-1} \tag{4.16}$$

where:

$$\Sigma_k = CP_{k|k-1}(Q)C^T + R. \tag{4.17}$$

Taking the limit as $Q^{-1} \to 0$ in (4.16), $K_{w,k}$ approaches $L_{w,k}$, where $\bar{\Sigma}_k$ is defined in (4.13). It follows that (2.24) has a well-defined limit given by (4.7). Now substitute for $K_{x,k}$ in (2.27) and let $L_{x,k}$ be *defined* as in (4.11). Then (2.27) has a well-defined limit given by (4.10). Applying (4.4) to (2.30) gives:

$$\Theta_k^{-1} = \Sigma_k^{-1} - \Sigma_k^{-1}D(D^T\Sigma_k^{-1}D + Q^{-1})^{-1}D^T\Sigma_k^{-1}. \tag{4.18}$$

Substituting for (2.28) and then (4.18) into (2.23) and taking the limit as $Q^{-1} \to 0$, $\hat{x}_{k|k}$ has a well-defined limit given by:

$$\bar{x}_{k|k} = \bar{x}_{k|k-1} + \bar{P}_{k|k-1} C^T \bar{\Sigma}_k^{-1} (I - D(D^T \bar{\Sigma}_k^{-1} D)^{-1} D^T \bar{\Sigma}_k^{-1})(z_k - C\bar{x}_{k|k-1}). \tag{4.19}$$

Making use of the expressions for $\bar{w}_{k|k}$ and $L_{x,k}$ in (4.19) gives (4.6). Substituting for $K_{w,k}$ from (2.29) and Θ_k from (2.30) into (2.26) and applying (4.4) gives:

$$P_{k|k}^{ww} = (Q^{-1} + D^T \Sigma_k^{-1} D)^{-1}. \tag{4.20}$$

Taking the limit as $Q^{-1} \to 0$, $P_{k|k}^{ww}$ has a well-defined limit given by:

$$\bar{P}_{k|k}^{ww} = (D^T \bar{\Sigma}_k^{-1} D)^{-1}. \tag{4.21}$$

This expression can be equivalently written in terms of $L_{w,k}$ in the form (4.9). Substituting for $K_{x,k}$ from (2.28) and Θ_k^{-1} from (4.18) into (2.25), taking the limit as $Q^{-1} \to 0$, and then using the expressions for $L_{x,k}$ and $\bar{P}_{k|k}^{ww}$ from (4.21), $P_{k|k}$ has a well-defined limit given by (4.8). Taking the limit as $Q^{-1} \to 0$ in (2.31) and (2.32), $\hat{x}_{k+1|k}$ and $P_{k+1|k}$ have well-defined limits given by (4.14) and (4.15) respectively. The result follows by induction since, by definition, $\bar{x}_{0|-1} = \hat{x}_{0|-1} = \lim_{Q^{-1} \to 0} \hat{x}_{0|-1}$ and $\bar{P}_{0|-1} = P_{0|-1} = \lim_{Q^{-1} \to 0} P_{0|-1}$. □

A block diagram illustrating the limit filter equations (4.6), (4.7) and (4.14) is shown in Fig. 4.1, where $L_{w,k}$ and $L_{x,k}$ are defined through (4.8)–(4.13) and (4.15). Fig. 4.2 shows the recursive form of the limit filter.

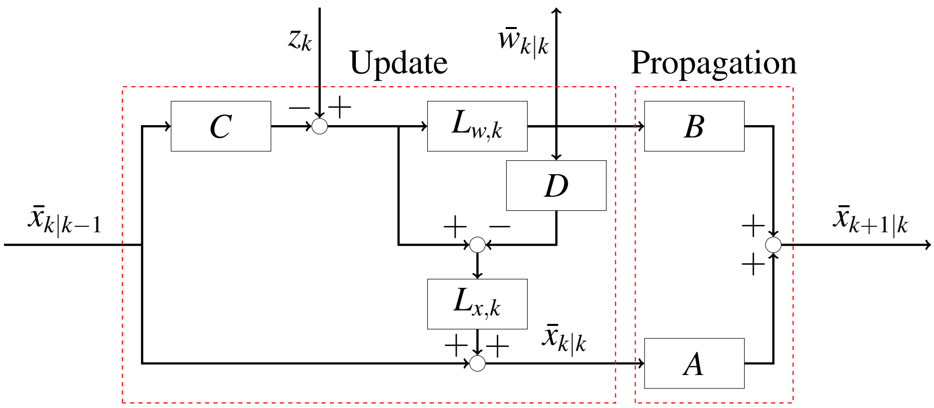

Fig. 4.1 Block diagram of the limit filter recursion with decomposition into update and propagation steps.

We now derive an alternative (second) form for the limit filter which takes the form of the standard Kalman filter for a new system with matrices A_1, B_1, C_1. Note that A_1, B_1, C_1,

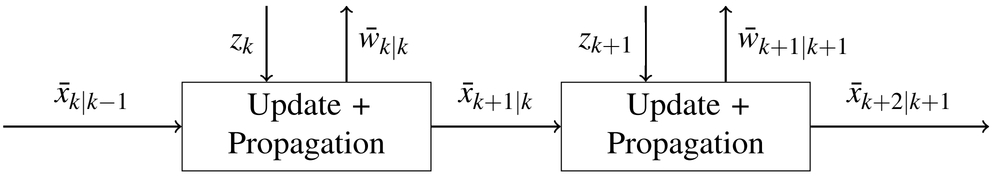

Fig. 4.2 Block diagram of an interconnection of limit filter recursion steps.

Π_c, Π, and $D^\dagger$ are defined in (3.10)–(3.15). This form will be very convenient to analyse convergence and stability. We first establish the following lemma.

Lemma 14 *Suppose D has full column rank, $R = R^T > 0$, $S = S^T \geq 0$ and Π_c is given by (3.13), then:*

$$(S+R)^{-1} - (S+R)^{-1}D(D^T(S+R)^{-1}D)^{-1}D^T(S+R)^{-1} = \Pi_c^T\left(\Pi_c S \Pi_c^T + R\right)^{-1}\Pi_c. \quad (4.22)$$

Proof: We first introduce the projection:

$$\Pi_S = D(D^T(S+R)^{-1}D)^{-1}D^T(S+R)^{-1}. \quad (4.23)$$

We observe that Π given in (3.14) and Π_S are parallel projections onto the same space, namely the column space of D, but have different null spaces. Further note that $\Pi^2 = \Pi_S\Pi = \Pi$, $\Pi_S^2 = \Pi\Pi_S = \Pi_S$ and:

$$(I-\Pi)\Pi_S = 0, \quad (4.24)$$

$$(I-\Pi_S)\Pi = 0, \quad (4.25)$$

$$\Pi_S^T R^{-1}(I-\Pi) = 0, \quad (4.26)$$

$$\Pi^T(S+R)^{-1}(I-\Pi_S) = 0. \quad (4.27)$$

We first claim that:

$$\left((I-\Pi)S(I-\Pi)^T + R\right)^{-1}(I-\Pi) = (S+R)^{-1}(I-\Pi_S). \quad (4.28)$$

To see this note that:

$$(I-\Pi)S(I-\Pi)^T + R = (I-\Pi)(S+R) + \left(R - (I-\Pi)S\right)\Pi^T \quad (4.29)$$

using $\Pi R = R\Pi^T$. Therefore using (4.24) and (4.27):

$$
\begin{aligned}
&\Big((I-\Pi)S(I-\Pi)^T + R\Big)(S+R)^{-1}(I-\Pi_S) \\
&= (I-\Pi)(I-\Pi_S) + \Big(R-(I-\Pi)S\Big)\Pi^T(S+R)^{-1}(I-\Pi_S) \\
&= (I-\Pi)
\end{aligned}
\tag{4.30}
$$

from which (4.28) follows. Further from (4.27) and (4.28):

$$
\Pi^T\Big((I-\Pi)S(I-\Pi)^T + R\Big)^{-1}(I-\Pi) = 0. \tag{4.31}
$$

Hence, substituting Π_S from (4.23) into (4.28) gives (4.22). □

Theorem 16 *Let D have full column rank. Then the limit filter recursions (4.6)–(4.15) are equivalent to:*

$$
\begin{aligned}
\bar{x}_{k|k} &= \bar{x}_{k|k-1} + \bar{K}_k(z_k - C_1\bar{x}_{k|k-1}), && (4.32)\\
\bar{P}_{k|k} &= (I - \bar{K}_k C_1)\bar{P}_{k|k-1}, && (4.33)\\
\bar{x}_{k+1|k} &= A_1\bar{x}_{k|k} + B_1 z_k, && (4.34)\\
\bar{P}_{k+1|k} &= A_1\bar{P}_{k|k}A_1^T + B_1 R B_1^T, && (4.35)\\
\bar{K}_k &= \bar{P}_{k|k-1}C_1^T(C_1\bar{P}_{k|k-1}C_1^T + R)^{-1} && (4.36)
\end{aligned}
$$

with the input estimate:

$$
\bar{w}_{k|k} = D^{\dagger}(z_k - C\bar{x}_{k|k}). \tag{4.37}
$$

Proof: We first establish several identities that are needed in the proof.
(1) We claim that:

$$
\bar{K}_k = L_{x,k}(I - DL_{w,k}). \tag{4.38}
$$

To see this note that on substitution from (4.11) and (4.12) the right hand side of (4.38) becomes:

$$
\bar{P}_{k|k-1}C^T(\bar{\Sigma}_k^{-1} - \bar{\Sigma}_k^{-1}D(D^T\bar{\Sigma}_k^{-1}D)^{-1}D^T\bar{\Sigma}_k^{-1}) \tag{4.39}
$$

which equals:

$$\bar{P}_{k|k-1}C^T\Big((I-\Pi)^T\big((I-\Pi)S(I-\Pi)^T+R\big)^{-1}(I-\Pi)\Big) \tag{4.40}$$

after applying (4.22) with $S = C\bar{P}_{k|k-1}C^T$ and noting that $\bar{\Sigma}_k = S+R$ from (4.13). Using the transpose of (4.31), the $I-\Pi$ on the right in (4.40) can be removed, which gives (4.36) on substituting from (3.12). This establishes (4.38).

(2) From (4.38) it is easy to see that:

$$\bar{K}_k D = 0 \tag{4.41}$$

since $L_{w,k}D = I$. Hence:

$$\bar{K}_k C = \bar{K}_k C_1. \tag{4.42}$$

(3) We claim that:

$$\bar{K}_k C_1 \bar{P}_{k|k-1} = L_{x,k}(\bar{\Sigma}_k - D\bar{P}^{ww}_{k|k}D^T)L^T_{x,k}. \tag{4.43}$$

To see this note that on substitution from (4.11) and (4.21) the right hand side of (4.43) becomes:

$$L_{x,k}(I-D(D^T\bar{\Sigma}_k^{-1}D)^{-1}D^T\bar{\Sigma}_k^{-1})C\bar{P}_{k|k-1} \tag{4.44}$$

which equals $\bar{K}_k C\bar{P}_{k|k-1}$ using (4.38). Applying (4.42) establishes the claim.

(4) We claim that:

$$L_{w,k} = D^\dagger(I - C\bar{K}_k). \tag{4.45}$$

To see this consider:

$$\begin{aligned} L_{w,k} - D^\dagger(I-C\bar{K}_k) &= L_{w,k} - D^\dagger + D^\dagger C L_{x,k}(I - DL_{w,k}) \\ &= L_{w,k} - D^\dagger + D^\dagger(\bar{\Sigma}_k - R)\bar{\Sigma}_k^{-1}(I-DL_{w,k}) \\ &= -D^\dagger R\bar{\Sigma}_k^{-1}(I-DL_{w,k}) \\ &= -(D^TR^{-1}D)^{-1}D^T\bar{\Sigma}_k^{-1}(I-D(D^T\bar{\Sigma}_k^{-1}D)^{-1}D^T\bar{\Sigma}_k^{-1}) \\ &= 0 \end{aligned} \tag{4.46}$$

where we have used (4.38), (4.11), (4.13), (3.15) and (4.12).

(5) We claim that:

$$\bar{P}^{xw}_{k|k} = -\bar{P}_{k|k}(D^{\dagger}C)^{T}. \tag{4.47}$$

To see this substitute from (4.11) and (4.45) into (4.10) and then use (4.33) and (4.42).

(6) Lastly we claim that:

$$\bar{P}^{ww}_{k|k} = D^{\dagger}(C\bar{P}_{k|k}C^{T} + R)(D^{\dagger})^{T}. \tag{4.48}$$

Substituting (4.13) and (4.45) into (4.9) gives:

$$\bar{P}^{ww}_{k|k} = D^{\dagger}\Big((I - C\bar{K}_k)C\bar{P}_{k|k-1}C^{T} + R - C\bar{K}_k R - (I - C\bar{K}_k)\bar{\Sigma}_k(C\bar{K}_k)^{T}\Big)(D^{\dagger})^{T}. \tag{4.49}$$

Using (4.42) note that:

$$\bar{\Sigma}_k\bar{K}_k^{T} = (C\bar{P}_{k|k-1}C_1^{T} + R)\bar{K}_k^{T}, \tag{4.50}$$

$$\bar{K}_k\bar{\Sigma}_k\bar{K}_k^{T} = \bar{K}_k(C_1\bar{P}_{k|k-1}C_1^{T} + R)\bar{K}_k^{T} = \bar{P}_{k|k-1}C_1^{T}\bar{K}_k^{T} \tag{4.51}$$

from which it follows that:

$$(I - C\bar{K}_k)\bar{\Sigma}_k\bar{K}_k^{T} = R\bar{K}_k^{T}. \tag{4.52}$$

Noting that $D^{\dagger}R\bar{K}_k^{T} = 0$ from (4.41) the last two terms in (4.49) are zero. The claim follows using (4.42) and (4.33) on the first term. We now use the identities to establish the theorem. Substituting (4.7) into (4.6) and using (4.38) gives (4.32). Substituting (4.43) into (4.8) gives (4.33). Substituting (4.45) into (4.7) and using (4.32) gives (4.37). Substituting (4.37) into (4.14) gives (4.34). Substituting (4.47) and (4.48) into (4.15) gives an expansion of (4.35). □

A block diagram illustrating the limit filter equations (4.32) and (4.34) is shown in Fig. 4.3.

4.1.3 Convergence

We will now study the behaviour of the limit filter as the horizon extends to infinity (i.e. $k \to \infty$). Substituting from (4.33) into (4.35) gives the Riccati difference equation:

$$\bar{P}_{k+1|k} = A_1\bar{P}_{k|k-1}A_1^{T} + B_1RB_1^{T} - A_1\bar{P}_{k|k-1}C_1^{T}(C_1\bar{P}_{k|k-1}C_1^{T} + R)^{-1}C_1\bar{P}_{k|k-1}A_1^{T}. \tag{4.53}$$

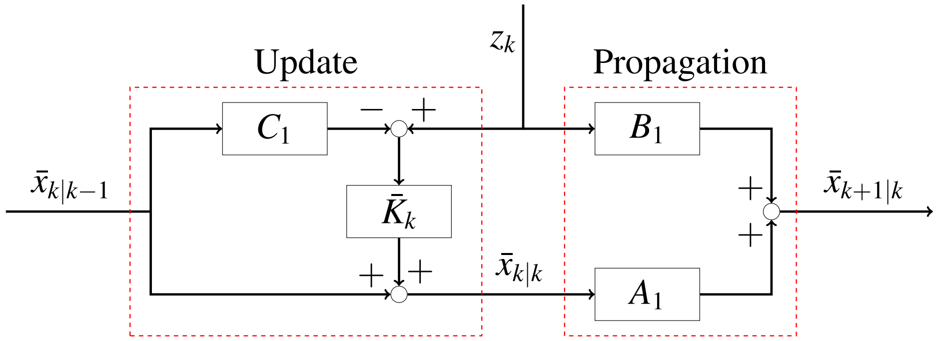

Fig. 4.3 An equivalent block diagram of the limit filter recursion with decomposition into update and propagation steps.

If $\bar{P}_{k+1|k}$ converges to $\bar{P}_\infty$ as $k \to \infty$, then $\bar{P}_\infty$ satisfies the algebraic Riccati equation:

$$\bar{P}_\infty = A_1 \bar{P}_\infty A_1^T + B_1 R B_1^T - A_1 \bar{P}_\infty C_1^T (C_1 \bar{P}_\infty C_1^T + R)^{-1} C_1 \bar{P}_\infty A_1^T. \tag{4.54}$$

A real symmetric nonnegative definite solution of (4.54) is said to be a *strong solution* if all the eigenvalues of $A_1 - A_1 \bar{K}_\infty C_1$ are on or inside the unit circle, where $\bar{K}_\infty$ is given by:

$$\bar{K}_\infty = \bar{P}_\infty C_1^T (C_1 \bar{P}_\infty C_1^T + R)^{-1}. \tag{4.55}$$

If all the eigenvalues are strictly inside the unit circle, the solution is said to be a *stabilizing solution* [4], [10]. The standard form taken by Theorem 16 and (4.54) allows well-known stability and convergence conditions in terms of A_1, B_1 and C_1 to be stated in the next two lemmas.

Lemma 15 *Let D have full column rank, then:*

1. *the strong solution of (4.54) exists and is unique if and only if (C_1, A_1) is detectable,*
2. *the strong solution is the only nonnegative definite solution of (4.54) if and only if (C_1, A_1) is detectable and (A_1, B_1) has no uncontrollable mode outside the unit circle,*
3. *the strong solution coincides with the stabilising solution if and only if (C_1, A_1) is detectable and (A_1, B_1) has no uncontrollable mode on the unit circle,*
4. *the stabilising solution is positive definite if and only if (C_1, A_1) is detectable and (A_1, B_1) has no uncontrollable mode inside, or on the unit circle.* □

Proof: See [15, Theorem 3.2] and [10, Theorem 3.1]. □

Lemma 16 *Let D have full column rank and:*

1. (A_1, B_1) *have no uncontrollable mode on the unit circle,*

2. (C_1, A_1) *be detectable,*

3. $P_{0|-1} > 0$

or

1. (C_1, A_1) *be detectable,*

2. $P_{0|-1} \geq \bar{P}_\infty$

then $\bar{P}_{k+1|k}$ *given by the Riccati difference equation (4.53) asymptotically converges to the unique strong solution* $\bar{P}_\infty$ *of the algebraic Riccati equation (4.54) as* $k \to \infty$.

Proof: See [15, Theorem 4.1, Theorem 4.2]. □

We will now express the convergence conditions in terms of the original system matrices A, B, C and D using Lemmas 17 and 18. We begin with the controllability of modes of (A_1, B_1).

Lemma 17 *Let D have full column rank.* $\lambda_0 \in \mathbb{C}$ *is an uncontrollable mode of* (A_1, B_1) *if and only if it is an uncontrollable mode of* (A, B).

Proof: If λ_0 is an uncontrollable mode of (A_1, B_1) then there exists $0 \neq x \in \mathbb{C}^n$ such that $x^*A = x^*\lambda_0$ and $x^*B = 0$. Hence $x^*(A - BD^\dagger C) = x^*\lambda_0$ and $x^*BD^\dagger = 0$. The converse follows since $D^\dagger$ has full row rank. □

We now turn to the detectability of (C_1, A_1).

Lemma 18 *Let D have full column rank. Then* (C_1, A_1) *is detectable if and only if the system matrix* $\begin{bmatrix} A - zI & B \\ C & D \end{bmatrix}$ *has full column rank for all* $z \in \mathbb{C}$ *with* $|z| \geq 1$.

Proof: The lemma and proof are the same as [63, Lemma 13.9] except that modes on or outside the unit circle are considered (rather than the imaginary axis). □

Theorem 17 *Let D have full column rank. Then* $\bar{P}_{k+1|k}$ *given by the Riccati difference equation (4.53) with* $P_{0|-1} > 0$ *asymptotically converges to the unique stabilising solution* $\bar{P}_\infty$ *of the algebraic Riccati equation (4.54) as* $k \to \infty$*, providing the system with realisation* (A, B, C, D) *has (1) no uncontrollable modes on the unit circle, and (2) no invariant zeros on or outside the unit circle.*

Proof: This follows from Lemmas 15, 16, 17 and 18. □

Remark 6 *Condition (2) in Theorem 17 that there are no invariant zeros on or outside the unit circle is a type of minimum phase condition.*

Remark 7 *From Lemma 18 we see that a necessary condition for (C_1, A_1) to be detectable is that (C, A) is detectable, under the assumption that D is full column rank, though it is clearly not sufficient, e.g. $A = -B = C = D = 1$. We note that the detectability of (C, A) is assumed in [20].*

Remark 8 *The assumption that D is full column rank requires that $p \geq m$, i.e. the number of measurements is no less than the number of unknown inputs. It is interesting to note that the conditions of Theorem 17 may still hold even if $p = m$.*

4.2 A deterministic formulation: The zero hold input assumption

4.2.1 Introduction

In this section we pose and solve a filtering/ estimation problem for the simultaneous estimation of the inputs and states in a continuous time linear finite dimensional dynamical system. The output of the dynamical system is the vector of all variables that are measured at a discrete sequence of times. The filter should make use only of these discrete time output measurements. This is in contrast to the formulation of Chapter 3, where the filter has access to measurements that are continuous time signals. Similarly to Chapter 3 we make the assumption that the direct feedthrough matrix of the system has full column rank. The formulation proposed here is motivated by practical considerations, namely that the operation of modern computers and even sensors (e.g. angular rotation sensors in vehicle wheels) is intrinsically discrete in nature. The filter should produce the best estimate of the system states and exogenous inputs. The meaning of "best" is to minimise a weighted sum squared error between the output at the discrete sequence of times and the corresponding output measurements. The problem set-up is illustrated in Fig. 4.4, where $\tilde{z}$ is the measurement of the output at a discrete sequence of times. We adopt the same notation as in Chapter 3, namely the filtered signals $w_1(t), x_1(t)$ provide best estimates of w and x at a given time instant t based on measurements up to that time, while the estimates $\hat{w}(t), \hat{x}(t)$ provide the best estimates over a fixed time interval.

To ensure the estimation problem is well posed we assume that the exogenous input is piecewise constant (i.e. zero-order hold) within the measurement intervals. This is the simplest sensible assumption that can be imposed on the form of the input. Alternative

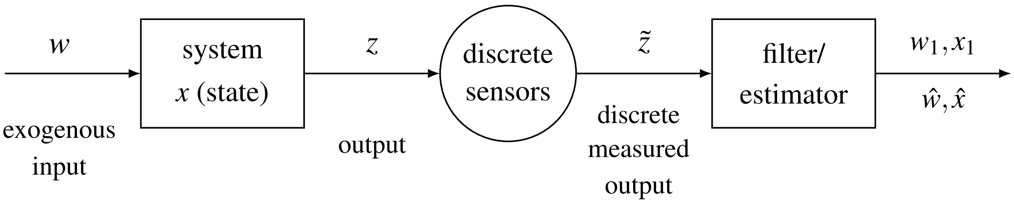

Fig. 4.4 Block diagram of a dynamical system with state x, exogenous input w and output z which is measured with sensors at a discrete sequence of times interconnected with an estimator.

assumptions for the form of the input are likely to give rise to more complicated algorithms. Some possible alternatives will later be discussed.

Our formulation here differs from Section 4.1 in two respects, the system dynamics are in continuous time and the estimation problem considered is deterministic in nature. However, they are not as different as they appear at first for the following reasons. The zero-order hold assumption imposed on the input allows us to integrate the continuous time system and in effect replace it by a discrete time system. Furthermore, we will show in this section that the close relation between the stochastic and deterministic formulations of the Kalman filter and smoother extends to problems with unknown inputs. To see this compare the filtering forms of Section 4.1 and Section 4.2 under an appropriate mapping of the matrices.

Our solution to the problem of Fig. 4.4 is inspired by both the work of Willems in [60] and more recently French in [7] for the continuous and discrete time Kalman filter respectively. In particular, we rewrite the cost functional in a convenient form to determine the unique optimal solution by means of completion of squares. This involves the construction of a forwards in time algebraic recursion which generates the filtered signals $w_1(t), x_1(t)$ evaluated at the discrete sequence of times the output measurements are available. The filtering recursion has a two-stage form of “update" (containing a projected error signal with a gain based on the solution of a Riccati algebraic equation) and “propagation” that is reminiscent of the discrete time Kalman filter and can be solved in real time. A further backwards in time recursion is required to determine the optimal solution $\hat{w}(t), \hat{x}(t)$ on a fixed interval.

4.2.2 Estimation

We start by considering the same system of Chapter 3, namely the linear continuous time system with the state space description:

$$\dot{x} = Ax + Bw, \tag{4.56}$$

$$z = Cx + Dw \tag{4.57}$$

where $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}, C \in \mathbb{R}^{p\times n}$ and $D \in \mathbb{R}^{p\times m}$ (full column rank) are fixed known matrices[1] and $w \in \mathscr{L}_{2,e}^{m}$, $x \in \mathscr{L}_{2,e}^{n}$ and $z \in \mathscr{L}_{2,e}^{p}$ are input, state and output related through this linear system. Let the system output be measured at the discrete finite sequence of times $t_0 < t_1 < \cdots < t_N \in \mathbb{R}$ where $N \in \mathbb{N}$ and denote by $\tilde{z}_i$ the actual output measurement at time t_i. We consider the problem to estimate w and x from the finite set of measurements $\tilde{z}_i$ (see Fig. 4.4). We introduce the performance index:

$$C(w,x(t_0)) = \|x(t_0) - \gamma\|^2_{\Gamma^{-1}} + \sum_{i=0}^{N} \|\tilde{z}_i - z(t_i)\|^2_{R^{-1}} \tag{4.58}$$

where $0 < R \in \mathbb{R}^{p\times p}$, $0 < \Gamma \in \mathbb{R}^{n\times n}$, $\gamma \in \mathbb{R}^n$ are specified. The problem we wish to solve is:

$$\inf_{w,x(t_0)} C(w,x(t_0)) \tag{4.59}$$

subject to

$$w(t) = w(t_i) \text{ for all } t \in [t_i, t_{i+1}) \tag{4.60}$$

where x, w and z satisfy (4.56)–(4.57). Our solution to (4.59) is based on two "completion of squares" arguments for the performance index (4.58) which are given in the following two lemmas.

Lemma 19 *Let $x_i^- \in \mathbb{R}^n$ and $P_i^- \in \mathbb{R}^{n\times n}$ where $P_i^- > 0$ be known, C_1, Π_c, Π and $D^\dagger$ are given in (3.12)–(3.15) and define x_i^+, P_i^+ by the update equations:*

$$x_i^+ = x_i^- + K_i(\tilde{z}_i - C_1 x_i^-), \tag{4.61}$$

$$P_i^+ = \left(\mathrm{I} - K_i C_1\right) P_i^-, \tag{4.62}$$

$$K_i = P_i^- C_1^T (C_1 P_i^- C_1^T + R)^{-1}. \tag{4.63}$$

Then $P_i^+ > 0$ and:

$$\|e_i^-\|^2_{(P_i^-)^{-1}} + \|\Pi_c v_i\|^2_{R^{-1}} = \|e_i^+\|^2_{(P_i^+)^{-1}} + \|\Pi_c \tilde{z}_i - C_1 x_i^-\|^2_{(C_1 P_i^- C_1^T + R)^{-1}} \tag{4.64}$$

where:

$$e_i^- = x(t_i) - x_i^-, \tag{4.65}$$

$$e_i^+ = x(t_i) - x_i^+, \tag{4.66}$$

$$v_i = z(t_i) - \tilde{z}_i \tag{4.67}$$

[1] The assumption that the system matrices are constant is for notational convenience.

for any $x(t_i)$, $z(t_i)$ satisfying:

$$\Pi_c z(t_i) = C_1 x(t_i). \tag{4.68}$$

(Note: (4.68) always holds if (4.57) is satisfied for some choice of w.)

Proof: First note that $K_i \Pi = 0$. Then from (4.61):

$$x_i^+ = \big(I - K_i C_1\big) x_i^- + K_i \Pi_c \tilde{z}_i \tag{4.69}$$

and from (4.68):

$$x(t_i) = \big(I - K_i C_1\big) x(t_i) + K_i \Pi_c z(t_i) \tag{4.70}$$

and hence the difference is given by:

$$e_i^+ = \big(I - K_i C_1\big) e_i^- + K_i \Pi_c v_i. \tag{4.71}$$

It can be verified directly that:

$$\big(I - K_i C_1\big)^{-1} = I + P_i^- C_1^T R^{-1} C_1 \tag{4.72}$$

and hence $P_i^+ > 0$ follows from:

$$(P_i^+)^{-1} = (P_i^-)^{-1} \big(I - K_i C_1\big)^{-1}. \tag{4.73}$$

Using (4.71) and (4.73):

$$(P_i^+)^{-1} e_i^+ = (P_i^-)^{-1} e_i^- + (P_i^+)^{-1} K_i \Pi_c v_i. \tag{4.74}$$

Using (4.63), (4.73) and (4.72):

$$\begin{aligned} K_i^T (P_i^+)^{-1} K_i &= K_i^T \Big((P_i^-)^{-1} + C_1^T R^{-1} C_1 \Big) K_i \\ &= (C_1 P_i^- C_1^T + R)^{-1} C_1 P_i^- C_1^T R^{-1} \\ &= R^{-1} - (C_1 P_i^- C_1^T + R)^{-1}. \end{aligned} \tag{4.75}$$

Therefore using (4.71), (4.74) and then (4.75) we verify:

$$\begin{aligned}\|e_i^+\|^2_{(P_i^+)^{-1}} &= (e_i^+)^T(P_i^+)^{-1}e_i^+ = (e_i^-)^T(P_i^-)^{-1}e_i^- - (C_1e_i^-)^T(C_1P_i^-C_1^T+R)^{-1}C_1e_i^- \\ &+2(\Pi_c v_i)^T(C_1P_i^-C_1^T+R)^{-1}C_1e_i^- + (\Pi_c v_i)^T R^{-1}\Pi_c v_i - (\Pi_c v_i)^T(C_1P_i^-C_1^T+R)^{-1}\Pi_c v_i \\ &= \|e_i^-\|^2_{(P_i^-)^{-1}} + \|\Pi_c v_i\|^2_{R^{-1}} - \|\Pi_c v_i - C_1 e_i^-\|^2_{(C_1P_i^-C_1^T+R)^{-1}} \end{aligned} \quad (4.76)$$

from which the required result follows using (4.68). □

Lemma 20 *Let $x_i^+ \in \mathbb{R}^n$, $P_i^+ \in \mathbb{R}^{n\times n}$ where $P_i^+ > 0$ be known and A_{1_d}, B_{1_d} given by:*

$$A_{1_d} = e^{A\Delta t_i} - B_{1_d}C, \quad (4.77)$$

$$B_{1_d} = \int_0^{\Delta t_i} e^{A(\Delta t_i-\tau)}d\tau B_1 \quad (4.78)$$

with A_{1_d} nonsingular where $\Delta t_i = t_{i+1} - t_i$ and B_1 is given by (3.11). Define x_{i+1}^-, P_{i+1}^- by the propagation equations:

$$x_{i+1}^- = A_{1_d}x_i^+ + B_{1_d}\tilde{z}_i, \quad (4.79)$$

$$P_{i+1}^- = A_{1_d}P_i^+A_{1_d}^T + B_{1_d}RB_{1_d}^T. \quad (4.80)$$

Then $P_{i+1}^- > 0$ and:

$$\|e_i^+\|^2_{(P_i^+)^{-1}} + \|\Pi v_i\|^2_{R^{-1}} = \|e_{i+1}^-\|^2_{(P_{i+1}^-)^{-1}} + \|\Pi v_i - RB_{1_d}^T(P_{i+1}^-)^{-1}e_{i+1}^-\|^2_{(R-RB_{1_d}^T(P_{i+1}^-)^{-1}B_{1_d}R)^{-1}} \quad (4.81)$$

for any x, w, z satisfying (4.56), (4.57) and (4.60) where e_i^-, e_i^+, v_i are given by (4.65)–(4.67). (Note: the condition that A_{1_d} is nonsingular is required for the inverses in (4.81) to exist.)

Proof: First note that $B_{1_d} = B_{1_d}\Pi$. Then from (4.56), (4.57) and (4.60):

$$x(t_{i+1}) = A_{1_d}x(t_i) + B_{1_d}\Pi z(t_i) \quad (4.82)$$

and hence using (4.79):

$$e_{i+1}^- = A_{1_d}e_i^+ + B_{1_d}\Pi v_i. \quad (4.83)$$

Substituting from (4.83) gives:

$$\Pi v_i - RB_{1_d}^T(P_{i+1}^-)^{-1}e_{i+1}^- = (R - RB_{1_d}^T(P_{i+1}^-)^{-1}B_{1_d}R)R^{-1}\Pi v_i - RB_{1_d}^T(P_{i+1}^-)^{-1}A_{1_d}e_i^+ \quad (4.84)$$

hence:

$$\|\Pi v_i - RB_{1_d}^T (P_{i+1}^-)^{-1} e_{i+1}^-\|^2_{(R-RB_{1_d}^T (P_{i+1}^-)^{-1} B_{1_d} R)^{-1}} = \|\Pi v_i\|^2_{R^{-1}} - \|B_{1_d} \Pi v_i\|^2_{(P_{i+1}^-)^{-1}} - 2(B_{1_d}\Pi v_i)^T (P_{i+1}^-)^{-1} A_{1_d} e_i^+ + \|RB_{1_d}^T (P_{i+1}^-)^{-1} A_{1_d} e_i^+\|^2_{(R-RB_{1_d}^T (P_{i+1}^-)^{-1} B_{1_d} R)^{-1}}. \quad (4.85)$$

Substituting from (4.83) gives:

$$\|e_{i+1}^-\|^2_{(P_{i+1}^-)^{-1}} = \|A_{1_d} e_i^+\|^2_{(P_{i+1}^-)^{-1}} + \|B_{1_d}\Pi v_i\|^2_{(P_{i+1}^-)^{-1}} + 2(B_{1_d}\Pi v_i)^T (P_{i+1}^-)^{-1} A_{1_d} e_i^+ \quad (4.86)$$

and adding (4.85) and (4.86) gives:

$$\|\Pi v_i - RB_{1_d}^T (P_{i+1}^-)^{-1} e_{i+1}^-\|^2_{(R-RB_{1_d}^T (P_{i+1}^-)^{-1} B_{1_d} R)^{-1}} + \|e_{i+1}^-\|^2_{(P_{i+1}^-)^{-1}} = \|\Pi v_i\|^2_{R^{-1}} + \|A_{1_d} e_i^+\|^2_{(P_{i+1}^-)^{-1}} + \|RB_{1_d}^T (P_{i+1}^-)^{-1} A_{1_d} e_i^+\|^2_{(R-RB_{1_d}^T (P_{i+1}^-)^{-1} B_{1_d} R)^{-1}}. \quad (4.87)$$

To establish the claim it remains to show that:

$$\|e_i^+\|^2_{(P_i^+)^{-1}} = \|A_{1_d} e_i^+\|^2_{(P_{i+1}^-)^{-1}} + \|RB_{1_d}^T (P_{i+1}^-)^{-1} A_{1_d} e_i^+\|^2_{(R-RB_{1_d}^T (P_{i+1}^-)^{-1} B_{1_d} R)^{-1}} \quad (4.88)$$

which follows directly from the matrix inversion lemma (see [1, pg. 139]) and noting that the various inverses exist. □

The “completion of square” lemmas can be combined to give an alternative equivalent form for the performance index $C(w,x(t_0))$ in (4.58). The new form is given in Lemma 21 and is the key step in solving the optimisation problem (4.59)–(4.60).

Lemma 21 *Let x, w, z satisfy (4.56), (4.57) and (4.60) and A_{1_d} be nonsingular. Compute x_i^-, P_i^-, x_i^+ and P_i^+ for $i = 0$ to N using the forward recursions (4.61)–(4.62) and (4.79)–(4.80) setting the initial conditions $x_0^- = \gamma$ and $P_0^- = \Gamma$. Then the performance index $C(w,x(t_0))$ in (4.58) is given by:*

$$C(w,x(t_0)) = \|e_N^+\|^2_{(P_N^+)^{-1}} + \|\Pi v_N\|^2_{R^{-1}} + \sum_{i=0}^{N-1} \|\Pi v_i - RB_{1_d}^T (P_{i+1}^-)^{-1} e_{i+1}^-\|^2_{(R-RB_{1_d}^T (P_{i+1}^-)^{-1} B_{1_d} R)^{-1}} + \sum_{i=0}^{N} \|\Pi_c \tilde{z}_i - C_1 x_i^-\|^2_{(C_1 P_i^- C_1^T + R)^{-1}} \quad (4.89)$$

where e_i^-, e_i^+, v_i have been defined in (4.65)–(4.67).

Proof: Taking the sum of (4.64) for $i = 0$ to N, adding this to the sum of (4.81) from $i = 0$ to $N-1$, adding $\|\Pi v_N\|^2_{R^{-1}}$ on both sides and noting the cancellations, gives the required expression. Further note that we have also used the Pythagorean expression:

$$\|v_i\|^2_{R^{-1}} = \|\Pi v_i\|^2_{R^{-1}} + \|\Pi_c v_i\|^2_{R^{-1}} \tag{4.90}$$

which follows from the orthogonality condition:

$$\Pi_c^T R^{-1} \Pi = 0. \tag{4.91}$$

□

Theorem 18 *Let x, w, z satisfy (4.56), (4.57) and (4.60) and A_{1_d} be nonsingular. Compute x_i^-, P_i^-, x_i^+ and P_i^+ for $i = 0$ to N using the forward recursions (4.61)–(4.62) and (4.79)–(4.80) setting the initial conditions $x_0^- = \gamma$ and $P_0^- = \Gamma$. Then the optimisation problem (4.59)–(4.60) has a unique solution $w(t) = \hat{w}$, $x(t_0) = \hat{x}(t_0)$ given by solving:*

$$\hat{x}(t_N) = x_N^+, \tag{4.92}$$

$$\hat{x}(t_i) = x_i^+ + K_2 e_{i+1}^- \tag{4.93}$$

recursively backwards in time from $i = N$ to 0 to find $\hat{x}(t_0)$ (and indeed $\hat{x}(t_i)$ for all i) and then computing the input $\hat{w}(t)$ in the intervals $t_i \le t < t_{i+1}$ for $i = 0$ to $N-1$ by:

$$\hat{w}(t) = D^\dagger(\tilde{z}_i - C\hat{x}(t_i) + K_3 e_{i+1}^-) \tag{4.94}$$

and at $t = t_N$ by:

$$\hat{w}(t_N) = D^\dagger(\tilde{z}_N - C\hat{x}(t_N)) \tag{4.95}$$

where:

$$K_2 = P_i^+ A_{1_d}^T (P_{i+1}^-)^{-1}, \tag{4.96}$$

$$K_3 = R B_{1_d}^T (P_{i+1}^-)^{-1} \tag{4.97}$$

and e_i^- is defined in (4.65).

Proof: We first verify that the initial state $x(t_0)$ and input $w(t)$ as computed in (4.92)–(4.94) generate the intermediate states $x(t_i)$ given by (4.92)–(4.93). To see this we first note using (3.1) and (4.60) that:

$$x(t_{i+1}) = A_{1_d} x(t_i) + B_{1_d} z(t_i). \tag{4.98}$$

Substituting from (4.93), (3.2) and (4.94) gives:

$$
\begin{aligned}
&x(t_{i+1}) - A_{1_d}x(t_i) - B_{1_d}z(t_i) \\
&= x(t_{i+1}) - A_{1_d}(x_i^+ + K_2 e_{i+1}^-) - B_{1_d}(Cx(t_i) + \Pi(\tilde{z}_i - Cx(t_i) + K_3 e_{i+1}^-)) \\
&= x(t_{i+1}) - (A_{1_d}x_i^+ + B_{1_d}\tilde{z}_i) - (A_{1_d}K_2 + B_{1_d}K_3)e_{i+1}^- \\
&= x(t_{i+1}) - x_{i+1}^- - e_{i+1}^- = 0 \qquad (4.99)
\end{aligned}
$$

for every $i = 0$ to $N-1$ using $B_{1_d} = B_{1_d}\Pi$, (4.79) and noting that $A_{1_d}K_2 + B_{1_d}K_3 = \mathrm{I}$ from (4.80). We then note that all terms in (4.89) are nonnegative, and the last term is independent of w and $x(0)$. It is easy to verify that: the first term is zero by substitution from (4.92); the second term is zero by substitution from (4.95); and the third term is zero by substitution from (4.94) on noting that $\Pi K_3 = K_3$. Hence the system trajectory generated by the $x(t_0)$, $w(t)$ constructed in the theorem give a minimum of the performance index. Uniqueness can be shown by considering a variation of the above optimal trajectory, noting that the weighting matrices in (4.89) are strictly positive definite. □

Remark 9 *The assumption that A_{1_d} is nonsingular has been imposed to ensure that the inverses in Lemma 20 exist. The condition is satisfied for a sufficiently small time step Δt_i. Similar conditions are imposed in the deterministic formulation of the discrete time Kalman filter to ensure that the state weighting matrices remain strictly positive definite (see [7]). It is thought that Theorem 18 can be extended to the case when A_{1_d} is singular at the cost of mathematical complexity. Here we will be satisfied with a proof when the condition that A_{1_d} is nonsingular is imposed.*

In real time applications the number of measurements N and horizon length of the optimisation problem (4.59)–(4.60) increase. It might appear at first sight that a fresh solution of the optimisation problem needs to be calculated at each step. However, the optimal end-of-interval input and state estimates w_N^+ and x_N^+ are given by $w_1(t_N)$ and $x_1(t_N)$ respectively, which are computed by forward recursions only. Similarly, we can see that the optimal estimates at time t_i based on knowledge of $\tilde{z}(t_0), \ldots, \tilde{z}_i$ only are given by x_i^+ and w_i^+. Thus, to apply the algorithm in real time as measurements $\tilde{z}_i$ come in one at a time, all that is required is to execute a single step of the update and propagation equations of Lemmas 19 and 20. In summary, from (4.92) and (4.95) it follows that at $t = t_i$ the filtered estimates are given by:

$$x_1(t_i) = x_i^+, \qquad (4.100)$$

$$w_1(t_i) = D^\dagger(\tilde{z}_i - Cx_i^+) \qquad (4.101)$$

where x_i^+ is computed by the forwards in time recursions (4.61)–(4.62) and (4.79)–(4.80):

$$x_i^+ = x_i^- + K_i(\tilde{z}_i - C_1 x_i^-), \tag{4.102}$$
$$P_i^+ = \big(I - K_i C_1\big)P_i^-, \tag{4.103}$$
$$K_i = P_i^- C_1^T (C_1 P_i^- C_1^T + R)^{-1}, \tag{4.104}$$
$$x_{i+1}^- = A_{1_d} x_i^+ + B_{1_d}\tilde{z}_i, \tag{4.105}$$
$$P_{i+1}^- = A_{1_d} P_i^+ A_{1_d}^T + B_{1_d} R B_{1_d}^T \tag{4.106}$$

with the initial conditions $x_0^- = \gamma$, $P_0^- = \Gamma$ and where we have defined:

$$A_{1_d} = e^{A\Delta t_i} - B_{1_d} C, \tag{4.107}$$
$$B_{1_d} = \int_0^{\Delta t_i} e^{A(\Delta t_i - \tau)} d\tau B_1, \tag{4.108}$$
$$B_1 = BD^\dagger, \tag{4.109}$$
$$C_1 = \Pi_c C, \tag{4.110}$$
$$\Pi_c = I - \Pi, \tag{4.111}$$
$$\Pi = DD^\dagger, \tag{4.112}$$
$$D^\dagger = (D^T R^{-1} D)^{-1} D^T R^{-1} \tag{4.113}$$

In offline applications the number of measurements N and horizon length are fixed. The optimal input and state estimates over the interval, denoted by $\hat{w}$ and $\hat{x}$, were derived in Theorem 18. In summary, the construction of the smoothed estimates $\hat{w}$, $\hat{x}$ has a two stage form. First the filtered estimates are computed by a forwards recursion (4.102)–(4.113). Then from (4.92)–(4.97) the backwards recursion:

$$\hat{x}(t_i) = x_i^+ + P_i^+ A_{1_d}^T (P_{i+1}^-)^{-1} e_{i+1}^-, \tag{4.114}$$
$$\hat{w}(t_i) = D^\dagger(\tilde{z}_i - C\hat{x}(t_i) + RB_{1_d}^T (P_{i+1}^-)^{-1} e_{i+1}^-) \tag{4.115}$$

with terminal condition $\hat{x}(t_N) = x_N^+$ uses the filtered estimates to compute the smoothed estimates. Computing the smoothed estimates requires additional computational effort but estimates may be substantially improved. It should be noted that the smoothed input and state signals are trajectories of the system, the same is not true for the filtered signals. This is easy to verify by substituting the input and state signals into the differential equation (4.56) and noting whether the dynamics are satisfied. The above can be easily observed in the simulation example of Chapter 5 on road profile mapping (see Fig. 5.2 and Fig. 5.3).

4.2.3 Convergence

We will now study the asymptotic behaviour of the filter by noting that the filtering equations (4.102)–(4.106) take the form of the standard discrete time Kalman filer. Standard theory can be applied by appropriate matrix substitutions. We substitute from (4.103) into (4.106), which gives the Riccati difference equation (RDE):

$$P_{i+1}^- = A_{1_d} P_i^- A_{1_d}^T - A_{1_d} P_i^- C_1^T (C_1 P_i^- C_1^T + R)^{-1} C_1 P_i^- A_{1_d}^T + B_{1_d} R B_{1_d}^T. \tag{4.116}$$

If P_i^- converges to P_∞^- as $i \to \infty$, then P_∞^- satisfies the algebraic Riccati equation (ARE):

$$P_\infty^- = A_{1_d} P_\infty^- A_{1_d}^T - A_{1_d} P_\infty^- C_1^T (C_1 P_\infty^- C_1^T + R)^{-1} C_1 P_\infty^- A_{1_d}^T + B_{1_d} R B_{1_d}^T. \tag{4.117}$$

Lemma 22 *Let D have full column rank, then:*

1. *the strong solution of (4.117) exists and is unique if and only if (C_1, A_{1_d}) is detectable,*
2. *the strong solution is the only nonnegative definite solution of (4.117) if and only if (C_1, A_{1_d}) is detectable and (A_{1_d}, B_{1_d}) has no uncontrollable mode outside the unit circle,*
3. *the strong solution coincides with the stabilising solution if and only if (C_1, A_1) is detectable and (A_{1_d}, B_{1_d}) has no uncontrollable mode on the unit circle,*
4. *the stabilising solution is positive definite if and only if (C_1, A_{1_d}) is detectable and (A_{1_d}, B_{1_d}) has no uncontrollable mode inside, or on the unit circle.* □

Proof: See Lemma 15. □

Lemma 23 *Let D have full column rank and:*

1. *(A_{1_d}, B_{1_d}) have no uncontrollable mode on the unit circle,*
2. *(C_1, A_{1_d}) be detectable,*
3. *$P_0^- > 0$*

or

1. *(C_1, A_{1_d}) be detectable,*
2. *$P_0^- \geq P_\infty^-$*

then P_{i+1}^- given by the Riccati difference equation (4.116) asymptotically converges to the unique strong solution P_∞^- of the algebraic Riccati equation (4.117) as $i \to \infty$.

Proof: See Lemma 16. □

4.2.4 Nonlinear extension

Formulation

We will now turn our attention to estimation in nonlinear systems by considering the same problem formulation (4.58)–(4.60) we introduced for linear systems. More precisely, let $x(t) \in \mathbb{R}^n$, $w(t) \in \mathbb{R}^m$ and $z(t) \in \mathbb{R}^p$ satisfy the nonlinear dynamics:

$$\dot{x} = f(x,w), \tag{4.118}$$

$$z = h(x,w) \tag{4.119}$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$. Consider the optimisation problem:

$$\inf_{w,x(t_0)} C(w,x(t_0)) \tag{4.120}$$

subject to

$$w(t) = w(t_i) \text{ for all } t \in [t_i,t_{i+1}) \tag{4.121}$$

with the performance index:

$$C(w,x(t_0)) = \|x(t_0) - \gamma\|^2_{\Gamma^{-1}} + \sum_{i=0}^{N} \|\tilde{z}_i - z(t_i)\|^2_{R^{-1}} \tag{4.122}$$

where $0 < R \in \mathbb{R}^{p\times p}$, $0 < \Gamma \in \mathbb{R}^{n\times n}$, $\gamma \in \mathbb{R}^n$ are specified. The optimisation problem here is identical to the previous section with the exception that the dynamics (4.118)–(4.119) are nonlinear. We will propose a heuristic algorithm for this problem without deriving any optimality or convergence guarantees. In essence, we propose applying the algorithm of Theorem 18, with the linear system matrices replaced by local linear approximations to the nonlinear system along the estimated trajectory.

Filtering

We begin by noting that the solution of Theorem 18 holds for piecewise linear systems (we previously assumed constant system matrices for notational convenience). Here we will approximate the nonlinear system by a piecewise linear system and then apply the solution of Theorem 18. We now introduce a local linear approximation to (4.118)–(4.119) near

$x = x(t_i)$, $w = w(t_i)$ which is given by:

$$\dot{x} = f_i + A_i(x - x(t_i)) + B_i(w - w(t_i)), \tag{4.123}$$

$$z = h_i + C_i(x - x(t_i)) + D_i(w - w(t_i)) \tag{4.124}$$

where f_i, h_i, A_i, B_i, C_i and D_i are constant parameters of appropriate dimensions. The true state $x(t_i)$ and input $w(t_i)$ are unknown and cannot be used to generate a local linear approximation, but they can be replaced by the best available estimates. Possible methods of linearisation (i.e. computation of the constant parameters f_i, h_i, A_i, B_i, C_i and D_i) are discussed in the next sections. Then the forward recursion of Theorem 18 is given by:

$$x_i^+ = x_i^- + K_i(\tilde{z}_i - h_i), \tag{4.125}$$

$$w_i^+ = w_{i-1}^+ + D_i^\dagger(\tilde{z}_i - h_i), \tag{4.126}$$

$$P_i^+ = \left(\mathrm{I} - K_i C_1\right) P_i^-, \tag{4.127}$$

$$x_{i+1}^- = x_i^+ + A_3 f_i, \tag{4.128}$$

$$P_{i+1}^- = A_{1_d} P_i^+ A_{1_d}^T + B_{1_d} R B_{1_d}^T, \tag{4.129}$$

$$K_i = P_i^- C_1^T (C_1 P_i^- C_1^T + R)^{-1} \tag{4.130}$$

where:

$$A_{1_d} = e^{A_i \Delta t_i} - B_{1_d} C, \tag{4.131}$$

$$A_3 = \int_0^{\Delta t} e^{A_i(\Delta t - \tau)} d\tau, \tag{4.132}$$

$$B_{1_d} = A_3 B_1, \tag{4.133}$$

$$B_1 = B_i D^\dagger, \tag{4.134}$$

$$C_1 = \Pi_c C_i, \tag{4.135}$$

$$\Pi_c = I - \Pi, \tag{4.136}$$

$$\Pi = D_i D^\dagger, \tag{4.137}$$

$$D^\dagger = (D_i^T R^{-1} D_i)^{-1} D_i^T R^{-1} \tag{4.138}$$

with the initial conditions $x_0^- = \gamma$ and $P_0^- = \Gamma$. The filtered input and state estimates are computed by forward recursion only and are given by $w_1(t_i) = w_i^+$ and $x_1(t_i) = x_i^+$.

Smoothing

We now consider the computation of the smoothed input and state estimates $\hat{w}$ and $\hat{x}$ over a fixed horizon. For a given piecewise linear approximation the backwards recursion of Theorem 18 is given by:

$$\hat{x}(t_i) = x_i^+ + K_2 e_{i+1}^-, \tag{4.139}$$

$$\hat{w}(t_i) = w_i^+ + D^\dagger(\tilde{z}_i - h_i + K_3 e_{i+1}^-) \tag{4.140}$$

where:

$$e_i^- = \hat{x}(t_i) - x_i^-, \tag{4.141}$$

$$K_2 = P_i^+ A_{1_d}^T (P_{i+1}^-)^{-1}, \tag{4.142}$$

$$K_3 = RB_{1_d}^T (P_{i+1}^-)^{-1} \tag{4.143}$$

with the terminal condition $\hat{x}(t_N) = x_N^+$. Note that the piecewise linear approximation used for filtering and smoothing may be different. It remains to discuss how the piecewise linear approximations are generated. Two approaches are discussed in the next two sections.

Explicit linearisation

The true input and state trajectories are of course unknown and cannot be used to generate local linear approximations. Here we propose to instead explicitly linearise the nonlinear system at the best available estimates of the input and state (i.e. linearising along the estimated system trajectory). For the forwards recursion the best available estimates are the filtered estimates $x_1(t_i)$ and $w_1(t_i)$, while for the backwards recursion it is the smoothed estimates $\hat{x}(t_i)$ and $\hat{w}(t_i)$. The explicit linearisation is given by evaluating the nonlinear system functions $f(x,w)$ and $h(x,w)$ and computing their Jacobians (with respect to the state and input)[2] at the best available estimates. More specifically, for filtering we have:

$$f_i = f(x_1(t_i), w_1(t_i)), \tag{4.144}$$

$$h_i = h(x_1(t_i), w_1(t_i)), \tag{4.145}$$

$$A_i = \frac{\partial f(x, w_1(t_i))}{\partial x}\Big|_{x=x_1(t_i)}, \tag{4.146}$$

$$B_i = \frac{\partial f(x_1(t_i), w)}{\partial w}\Big|_{w=w_1(t_i)}, \tag{4.147}$$

$$C_i = \frac{\partial h(x, w_1(t_i))}{\partial x}\Big|_{x=x_1(t_i)}, \tag{4.148}$$

$$D_i = \frac{\partial h(x_1(t_i), w)}{\partial w}\Big|_{w=w_1(t_i)} \tag{4.149}$$

while for smoothing we replace the filtered estimates $x_1(t_i)$, $w_1(t_i)$ by the smoothed estimates $\hat{x}(t_i)$, $\hat{w}(t_i)$. This linearisation approach is similar to the Extended Kalman filter (EKF).

Implicit linearisation

We will now propose an alternative approach to linearising the nonlinear system which is based on sampling the nonlinear functions $f(x,w)$ and $h(x,w)$. It is reminiscent of the unscented Kalman filter (UKF) but differs both in terms of interpretation and application. In particular, we sample the nonlinear functions at a deterministically chosen set of points and associate weights to each. We then generate linear function approximations based on the weighted samples. Let $\lambda \in \mathbb{R}$, $0 < V \in \mathbb{R}^{L \times L}$ where $L = n + m$ and introduce the samples:

$$\mathscr{Q}_j = \begin{bmatrix} \mathscr{X}_j \\ \mathscr{W}_j \end{bmatrix} = \begin{cases} q & \text{for } j = 0, \\ q + \lambda \sqrt{V}_{*,j} & \text{for } j = 1, \dots, L, \\ q - \lambda \sqrt{V}_{*,j-L} & \text{for } j = L+1, \dots, 2L \end{cases} \tag{4.150}$$

and associated weights:

$$W_0 = 1 - 2LW_j \qquad \text{for } j = 0, \tag{4.151}$$

$$W_j = \frac{1}{2\lambda^2} \qquad \text{for } j = 1, \dots, 2L \tag{4.152}$$

[2] We have assumed that the functions are differentiable, i.e. $f \in C^1(\mathbb{R}^n \times \mathbb{R}^m, \mathbb{R}^n)$, $h \in C^1(\mathbb{R}^n \times \mathbb{R}^m, \mathbb{R}^p)$.

where q is the augmented vector of the best state and input estimate and evaluate:

$$\mathscr{Y}_j = \begin{bmatrix} f(\mathscr{X}_j, \mathscr{W}_j) \\ h(\mathscr{X}_j, \mathscr{W}_j) \end{bmatrix} \tag{4.153}$$

for all the samples $\mathscr{Q}_j$. Then a local linear approximation is given by:

$$y = \begin{bmatrix} f_i \\ h_i \end{bmatrix} = \sum_{j=0}^{2L} W_j \mathscr{Y}_j, \tag{4.154}$$

$$Y = \sum_{j=0}^{2L} W_j (\mathscr{Y}_j - y)(\mathscr{Q}_j - q)^T, \tag{4.155}$$

$$\begin{bmatrix} A_i & B_i \\ C_i & D_i \end{bmatrix} = YV^{-1}. \tag{4.156}$$

Using a Taylor series expansion note that the linearisation is independent of the tuning parameter λ up to second order. Furthermore, the matrices A_i, B_i, C_i and D_i are independent of the tuning parameter V up to second order and the vectors f_i and h_i up to first order. In the limit as $V \to 0$ this approach coincides with explicitly linearising the system at the point q. In filtering, V can be chosen to coincide with the solution of the RDE, i.e. P_i^+ at time t_i, to account for second order effects in estimating f_i and h_i.

Chapter 5

Estimation theory applied to vehicle examples

In this chapter we will apply the algorithm of Section 4.2 to two vehicle examples. The first example is on (offline) road elevation mapping and the second is on (online) slip estimation.

5.1 Road profile mapping

5.1.1 Model

We consider the problem of constructing (offline) an accurate map of a road profile using a vehicle equipped with a global positioning system (GPS) sensor and basic suspension sensors but without any dedicated optical sensors (e.g. profilometer). We consider the quarter-car suspension model of Fig. 5.1 with dynamical equations:

$$m_s \ddot{x}_s = -k_s(x_s - x_u) - c_s(\dot{x}_s - \dot{x}_u),$$
$$m_u \ddot{x}_u = k_s(x_s - x_u) + c_s(\dot{x}_s - \dot{x}_u) - k_t(x_u - x_r)$$

where m_s and m_u are sprung and unsprung masses, k_s and c_s are suspension spring and damper constants, k_t is a tire stiffness constant and the coordinates x_s, x_u and x_r represent displacements of m_s, m_u and road respectively. For simplicity we have neglected inertial and aero forces applied to the suspension, those can be included by modelling the dynamics of the entire vehicle. The road profile x_r is treated as an unknown system input.

We begin by considering a sensor set consisting of accelerometers located on m_s and m_u and a strut deflection sensor which generates measurements every Δt and a GPS sensor located on m_s that measures x_s. Hence, the state, input and output vectors are given by:

$$x =: \begin{bmatrix} x_s & \dot{x}_s & x_u & \dot{x}_u \end{bmatrix}^T, \tag{5.1}$$

$$w =: \begin{bmatrix} x_r \end{bmatrix}, \tag{5.2}$$

$$z =: \begin{bmatrix} x_s & x_s - x_u & \ddot{x}_s & \ddot{x}_u \end{bmatrix}^T. \tag{5.3}$$

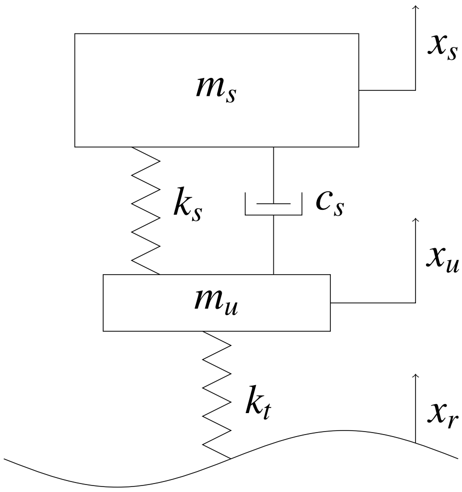

Fig. 5.1 Diagram of a quarter vehicle suspension model.

5.1.2 Stability analysis

We now consider the convergence properties of the filter assuming the model used is identical to the system which generates the measurements (i.e. vehicle and suspension). Note that among the outputs in (5.3) it is $\ddot{x}_u$ (i.e. the use of an unsprung mass accelerometer sensor) which ensures that D is full column rank. The detectability condition on (C_1, A_{1_d}) in Lemma 22 and 23 is up to first order in Δt equivalent to:

$$\begin{bmatrix} 1-z & \Delta t & 0 & 0 & 0 \\ -\frac{k_s \Delta t}{m_s} & 1-z-\frac{c_s \Delta t}{m_s} & \frac{k_s \Delta t}{m_s} & \frac{c_s \Delta t}{m_s} & 0 \\ 0 & 0 & 1-z & \Delta t & 0 \\ \frac{k_s \Delta t}{m_u} & \frac{c_s \Delta t}{m_u} & -\frac{(k_s+k_t)\Delta t}{m_u} & 1-z-\frac{c_s \Delta t}{m_u} & \frac{k_t \Delta t}{m_u} \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ -\frac{k_s}{m_s} & -\frac{c_s}{m_s} & \frac{k_s}{m_s} & \frac{c_s}{m_s} & 0 \\ \frac{k_s}{m_u} & \frac{c_s}{m_u} & -\frac{(k_s+k_t)}{m_u} & -\frac{c_s}{m_u} & \frac{k_t}{m_u} \end{bmatrix} \tag{5.4}$$

full column rank for all $z \in \mathbb{C}$ with $|z| \geq 1$. Denoting by ρ_n the n^{th} row of (5.4) and noting that elementary row operations preserve rank, we perform sequentially the following operations:

$$\begin{aligned}
&\rho_1 \to \Delta t^{-1}(\rho_1 - (1-z)\rho_5), \\
&\rho_6 \to -\rho_6 + \rho_5, \\
&\rho_3 \to \Delta t^{-1}(\rho_3 - (1-z)\rho_6), \\
&\rho_8 \to m_u \rho_8 - k_s \rho_5 + (k_s + k_t)\rho_6 - c_s(\rho_1 - \rho_3)
\end{aligned}$$

followed by reordering and rescaling to obtain a matrix whose first five rows are the identity matrix, which is obviously full column rank for all z. In a similar way, we can show that the matrix (A_{1_d}, B_{1_d}) has full row rank for all z. It therefore follows that the stability and convergence results of Lemma 22 and 23 hold. It can be verified that the conditions hold even without measurement of $x_s - x_u$ (strut deflection) and $\ddot{x}_s$ (sprung mass acceleration) but fails if measurement of x_s (GPS sensor) is removed.

A similar analysis to the above can be carried out for other choices of sensors. For example, a profilometer which measures $x_s - x_r$ could be considered and in this case another non-zero entry appears in the D matrix. Thus the profilometer could be considered as an alternative to measurement of $\ddot{x}_u$ to ensure that D is full rank. Similar conclusions hold in this case, namely the required detectability holds if x_s is measured but not if only measurements of $\ddot{x}_u$ and $x_s - x_r$ are available.

5.1.3 Simulations

We demonstrate the performance of the filter and smoother on simulated data. We set the vehicle model parameters to $m_s = 350\,\text{kg}$, $m_u = 50\,\text{kg}$, $k_s = 20\,\text{kN/m}$, $c_s = 1\,\text{kNs/m}$ and $k_t = 200\,\text{kN/m}$. We assume that the strut deflection and acceleration measurements are available every 1ms and the GPS measurements every 1s. Simulated white noise is added to the GPS, strut deflection and accelerometer signals of standard deviation given by 1m, 10^{-2}m and 10^{-1}m/s^2 respectively to generate the measurement signals. We choose the measurement weighting matrix accordingly, namely we take $R = \text{diag}\{1, 10^{-4}, 10^{-2}, 10^{-2}\}$. (In practise the true level of noise is unknown and R can be treated as a tuning parameter.) We let the initial states for the sprung and unsprung mass position have an error of 1 m and set the initial state weighting matrix to $\Gamma = \text{diag}\{1, 0.1, 1, 0.1\}$.

We consider two alternative road profiles, a square wave (period 1 s, amplitude 0.1 m) and a ramp (slope 0.1 m/s). The true, filtered and smoothed position signals are plotted in Fig. 5.2 (square wave) and Fig. 5.3 (ramp). Note that for both road profiles the filter reduces the large initial estimation error, which is expected given the system satisfies the conditions derived in Section 4.2.3. Furthermore, the smoother generates a plausible state and input system trajectory which in effect differs from the true signal only by an offset due to the inaccuracy of the few available GPS measurements.

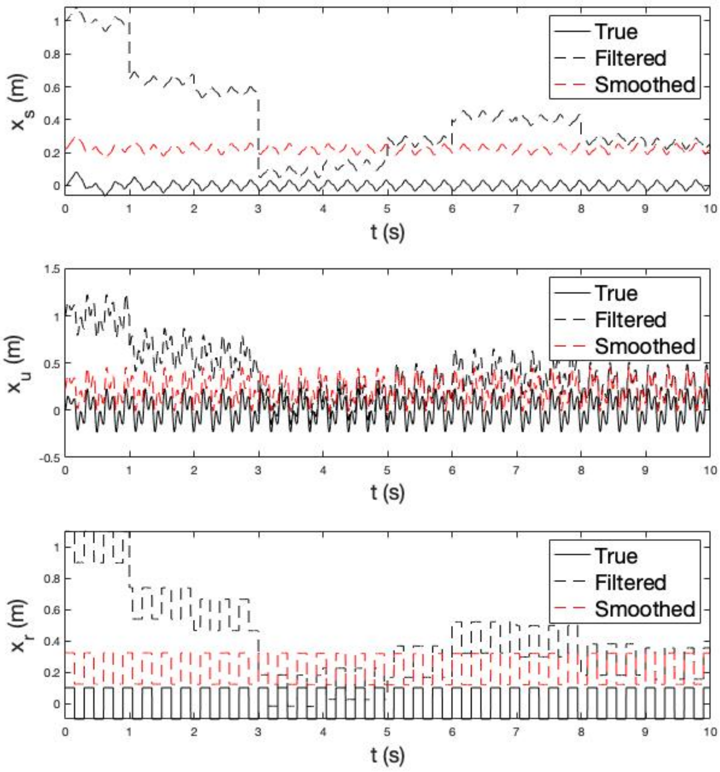

Fig. 5.2 True, filtered and smoothed position signals of sprung mass, unsprung mass and road for a square wave road profile.

5.2 Vehicle slip estimation

5.2.1 Model

We consider the problem of (online) vehicle slip estimation using a simple vehicle model on data generated from a high fidelity vehicle simulation. The simulation data are provided by McLaren Automotive using the "McLaren Integrated Data Analysis and Simulation" (MIDAS) simulation tool. We apply the nonlinear version of the algorithm of Section 4.2

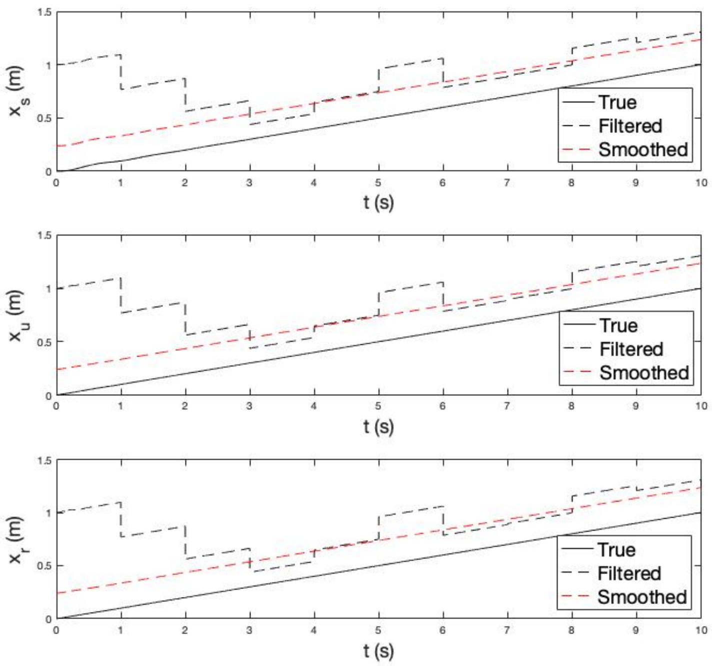

Fig. 5.3 True, filtered and smoothed position signals of sprung mass, unsprung mass and road for a ramp road profile.

using the implicit linearisation method. The high fidelity simulation signals supplied to the filter include the steering angle, longitudinal and lateral accelerations and yaw rate. We consider the 2D bicycle model of a vehicle given in Fig. 5.4. In the vehicle frame of reference the acceleration of the centre of mass is:

$$\ddot{\mathbf{r}} = a_x\mathbf{i}' + a_y\mathbf{j}' = (\dot{u}_x - \omega u_y)\mathbf{i}' + (\dot{u}_y + \omega u_x)\mathbf{j}' \tag{5.5}$$

where u_x, u_y are the longitudinal and lateral velocities of the centre of mass in the vehicle reference frame and ω is the yaw rate. Applying Newton's 2nd Law gives:

$$ma_x = F_{Fx}\cos(\delta) - F_{Fy}\sin(\delta) + F_{Rx} - C_d u_x^2, \tag{5.6}$$

$$ma_y = F_{Fx}\sin(\delta) + F_{Fy}\cos(\delta) + F_{Ry}, \tag{5.7}$$

$$I_{zz}\dot{\omega} = l_F F_{Fx}\sin(\delta) + l_F F_{Fy}\cos(\delta) - l_R F_{Ry} \tag{5.8}$$

where m is the vehicle mass, I_{zz} is the vehicle moment of inertia (around the vertical axis), δ is the steering angle, F_{Fx}, F_{Rx}, F_{Fy} and F_{Ry} are front and rear longitudinal and lateral tyre forces, l_F, l_R are vehicle length dimensions as shown in Fig. 5.4 and C_d is the aerodynamic drag coefficient. (For simplicity we have assumed that drag acts in the longitudinal direction and is quadratic in the longitudinal velocity.)

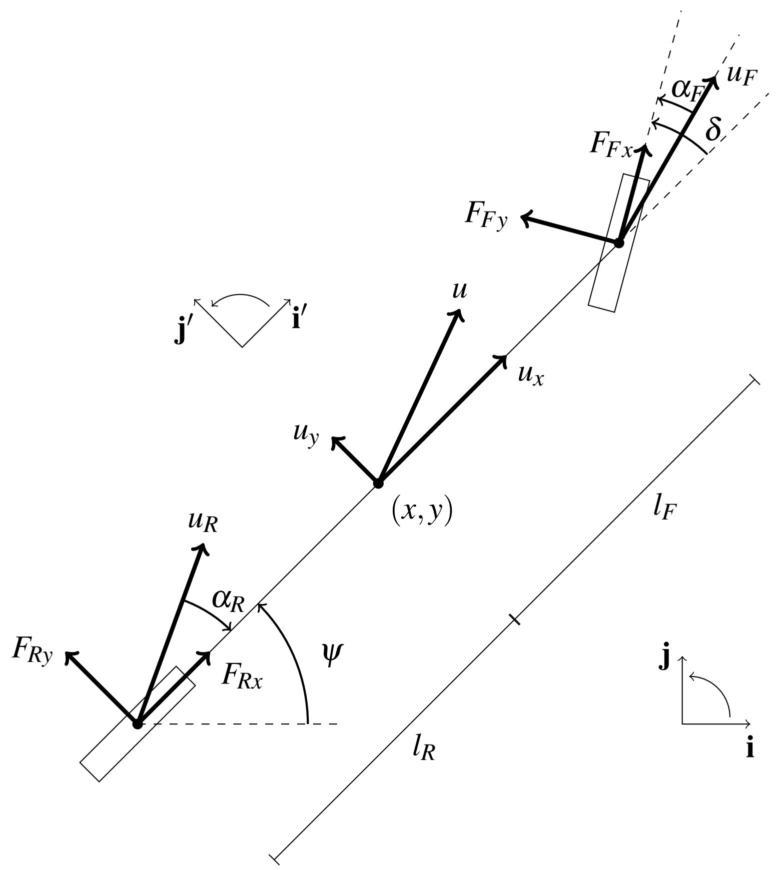

Fig. 5.4 Diagram of a bicycle model of a vehicle.

We treat the longitudinal and lateral velocities and yaw rate as model states and the steering angle and tyre forces as model inputs. In particular, we define the vector signals:

$$x =: \begin{bmatrix} u_x & u_y & \omega \end{bmatrix}^T, \tag{5.9}$$

$$w =: \begin{bmatrix} \delta & F_{Fy} & F_{Ry} & F_{Fx} & F_{Rx} \end{bmatrix}^T \tag{5.10}$$

which satisfy a nonlinear vector state equation $\dot{x} = f(x,w)$ derived from (5.6)–(5.8). We now define the front and rear tyre slip angles:

$$\alpha_F = \delta - \arctan\left(\frac{u_y + \omega l_F}{u_x}\right), \tag{5.11}$$

$$\alpha_R = -\arctan\left(\frac{u_y - \omega l_R}{u_x}\right). \tag{5.12}$$

We assume that the front and rear vertical tyre forces F_{Fz}, F_{Rz} are constant and given by:

$$F_{Fz} = \frac{l_R}{l_F + l_R} mg, \tag{5.13}$$

$$F_{Rz} = \frac{l_F}{l_F + l_R} mg. \tag{5.14}$$

We model the longitudinal tyre forces F_{Fx}, F_{Rx} using a heuristic switch logic. When the vehicle is in "driving mode", the front longitudinal tyre force is assumed to be approximately zero, i.e. $F_{Fx} \approx 0$. When the vehicle is in "braking mode", the front and rear longitudinal tyre forces are assumed to be approximately equal, i.e. $F_{Fx} \approx F_{Rx}$. We determine whether the vehicle is in driving or braking mode based on the sign of F_{Rx}. Based on these assumptions we derive the approximation:

$$F_{Fx} - H(-F_{Rx})F_{Rx} \approx 0 \tag{5.15}$$

where H is the Heaviside function:

$$H(x) = \begin{cases} 1 & \text{for } x \geq 0, \\ 0 & \text{for } x < 0. \end{cases}$$

Furthermore, we assume that the lateral tyre forces F_{Fy}, F_{Ry} approximately satisfy:

$$F_{Fy} - F_{Fz} D_F \sin(C_F \arctan(B_F \alpha_F)) \cos(G_F F_{Fx}) \approx 0, \tag{5.16}$$

$$F_{Ry} - F_{Rz} D_R \sin(C_R \arctan(B_R \alpha_R)) \cos(G_R F_{Rx}) \approx 0. \tag{5.17}$$

We let $z = h(x, w)$ denote a nonlinear vector output function which consists of δ, a_x, a_y and ω augmented by the functions on the left hand side of (5.15), (5.16) and (5.17) and where a_x and a_y are given by (5.6) and (5.7) respectively. We then let:

$$\tilde{z} =: \begin{bmatrix} \tilde{\delta} & \tilde{a}_x & \tilde{a}_y & \tilde{\omega} & 0 & 0 & 0 \end{bmatrix}^T \tag{5.18}$$

where $\tilde{\delta}$, $\tilde{a}_x$, $\tilde{a}_y$ and $\tilde{\omega}$ are the steering angle, longitudinal and lateral acceleration and yaw rate signals generated by the high fidelity vehicle simulation.

5.2.2 Simulations

We apply the same filtering algorithm of Section 4.2 using approximate values for the vehicle and tyre model parameters. The simulated measurement signals $\tilde{\delta}$, $\tilde{a}_x$, $\tilde{a}_y$ and $\tilde{\omega}$ are available at a frequency of 100Hz. We treat the weighting matrix R as a tuning parameter and manually choose to assign the diagonal matrix $R = \text{diag}\{10^{-2}, 10^{-1}, 10^{-1}, 10^{-2}, 10^{4}, 10^{4}, 10^{4}\}^2$. We offset the initial filter estimate of the longitudinal velocity such that it has an error of 10 m/s. The true (i.e. from a high fidelity vehicle simulation) and filtered signals for the longitudinal velocity, lateral velocity, front slip angle and rear slip angle are plotted in Fig. 5.5, Fig. 5.6, Fig. 5.7, and Fig. 5.8 respectively for a 90s simulation around a track.

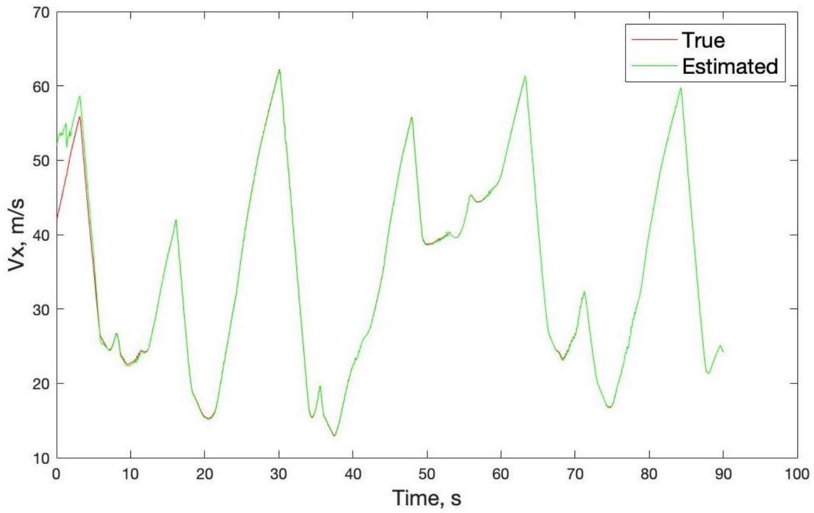

Fig. 5.5 True and filtered longitudinal vehicle velocity signals.

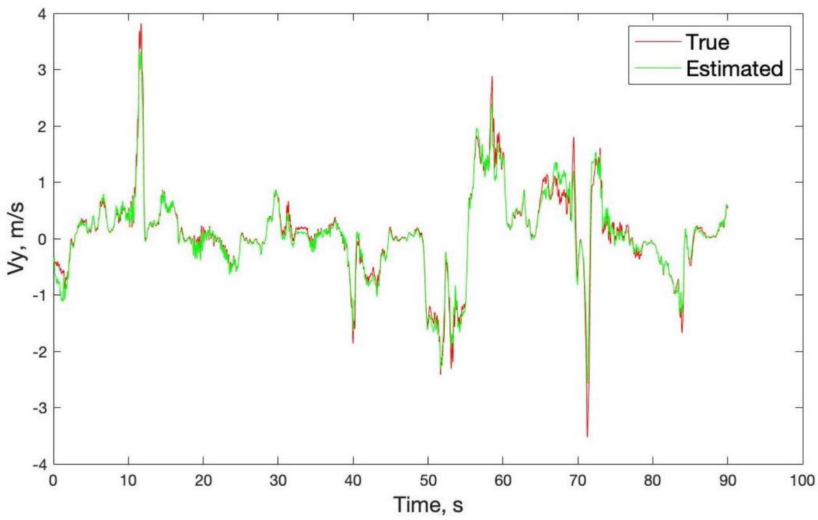

Fig. 5.6 True and filtered lateral vehicle velocity signals.

Note that the filter reduces the initial error in the longitudinal velocity and tracks closely all the required signals. In this simulation example it appears unintuitive at first sight that acceleration signals can be used successfully to estimate velocity signals. The apparent paradox is resolved after noting that using a vehicle and tyre model we can obtain velocity information from acceleration signals. We have performed many simulations where we vary the initial state estimate error, the model constant parameters and the vehicle manoeuvre. The filter performs similarly and surprisingly well even for large errors and highly dynamic vehicle manoeuvres and is robust against parameter variations.

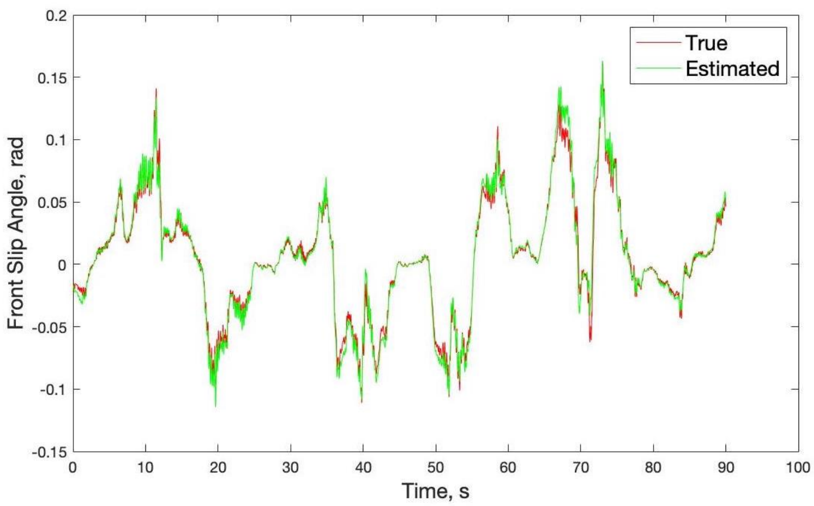

Fig. 5.7 True and filtered front tyre slip angle signals.

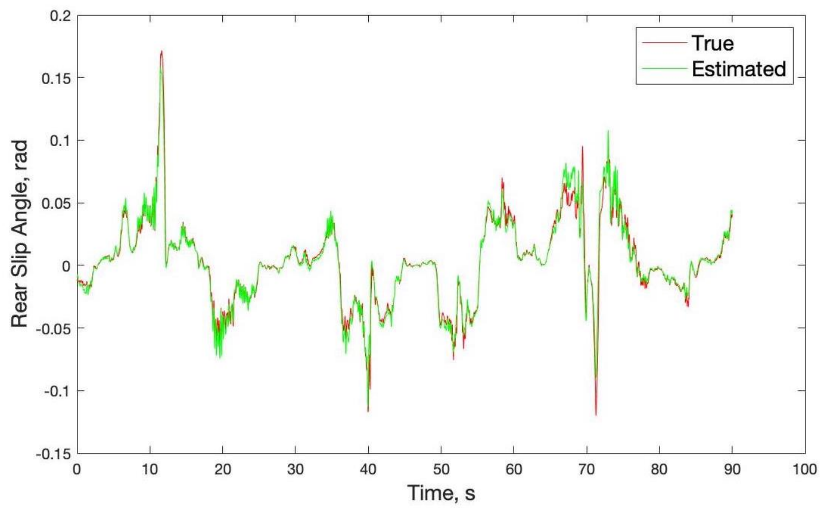

Fig. 5.8 True and filtered rear tyre slip angle signals.

Chapter 6

Discussion

6.1 Contributions

This thesis has explored directions of theoretical and practical interest in the field of estimation and control. In Chapter 3 we proposed a framework for estimation in which the output of the dynamical system comprises *all* variables that are measured, and the variables to be estimated comprise, equally, system states and exogenous inputs. This framework is quite general in that if an exogenous input is measured then we may include a direct feedthrough component in the output vector to reflect this. This estimation problem was solved for linear systems with a full column rank feedthrough matrix. The unique optimum solution on a finite horizon takes a two-stage form in which the first stage provides an end-of-interval estimator which can be solved in real time as the horizon length increases. It was shown that the full rank assumption is general enough to include the Kalman filter and, for the dual tracking problem, the linear quadratic regulator as special cases. Generalising this result to the case where this condition, or similar full rank assumptions on the Markov parameters, does not hold remains an open problem. Furthermore, the solution of a matrix Lyapunov differential equation $P_2(t)$ is shown to have an analogous (deterministic) interpretation to the smoothed covariance in the stochastic case. This has been achieved by considering the least-squares estimation problem with an additional constraint that the state passes through a prescribed point at a given time in the fixed horizon. To solve this problem a tracking problem was also considered which is dual to our estimation problem. We also considered the time invariant limiting forms of the estimation and tracking problems. Conditions were given for the convergence of the finite horizon solutions to these limits. Stability of the end-of-interval estimator on the infinite horizon requires a minimum phase condition (i.e. that there are no invariant zeros of the system in the closed right half plane) as well as the absence of uncontrollable modes on the imaginary axis. The time invariant systems were shown to be stable or anti-stable left inverses of the original system under appropriate conditions.

In Chapter 4 we derived two forms for the zero informational limit of the discrete time Kalman filter with feedthrough of the process noise to the measurements. We have shown that the recursions in the first form are closely related to those of [20] and hence we have provided a simpler notion of optimality for that recursive filter. The second form takes the form of the standard Kalman filter for a modified system (Theorem 16). This form is convenient to derive conditions for the asymptotic convergence of the limit filter to steady state form which are expressed in terms of the original system matrices as a minimum phase and a controllability condition (Theorem 17). Additionally, we proposed a deterministic formulation for simultaneous state and input estimation in continuous time systems with discrete time measurements. To pose the problem well we assumed that the exogenous input is piecewise constant (i.e. zero-order hold) within measurement intervals. We derived the solution to this problem for a fixed length horizon and the asymptotic form of the filter.

In Chapter 5 we apply the filtering and smoothing algorithms of Section 4.2 to two vehicle examples of interest. The first example is on (offline) road elevation mapping and the second is on (online) slip estimation. The first example is a linear example that serves to demonstrate the benefits of smoothing compared to filtering in offline applications. Furthermore, it shows how the use of an estimator and basic suspension sensors can be used in place of a dedicated sensor (i.e. profilometer) without compromising high estimation accuracy. Lastly, it shows the relevance of simultaneous state and input estimation algorithms to mapping problems of practical interest. The second example is a challenging nonlinear example of major practical interest. More specifically, accurate slip estimation can enhance vehicle safety and performance by increasing traction, reducing braking distances and improving stability control. It is particularly important in advancing self-driving capabilities. The example demonstrates how it is possible to achieve accurate and robust slip estimation with a surprisingly simple vehicle model and very basic sensors. This is in contrast to [2] which requires additional torque sensors and develops a complicated model for the suspension and wheel dynamics. The proposed approach which is based on the simultaneous state and input estimation algorithm reduces computational cost by dramatically decreasing the model dimensionality and complexity and significantly improving ease of tuning.

6.2 Further work

This work has sparked a lot of new and intriguing ideas that require further development. Many of the ideas developed in Chapter 3 for systems with continuous time measurements, e.g. the constrained estimation problem, have not been extended to the case of discrete time measurements. We speculate that this is possible if we assume that the exogenous input is piecewise constant (i.e. zero-order hold). It is also interesting to consider alternative assumptions on the input behaviour which may arguably be more realistic. They may lead to more accurate estimation algorithms but may also require more complicated algorithms. For example we may wish to assume continuity in the input (e.g. first-order hold) or its derivatives. It is also possible to consider penalising deviations from a zero-order (or first-order) hold input trajectory by augmenting the performance index without assuming a zero-order hold. The tracking formulation of Chapter 3 assumes full knowledge of the initial state and the absence of any input disturbances. It is of theoretical interest to consider a deterministic formulation of control in the face of uncertainty and to establish if the separation principle holds. Another major challenge of theoretical and practical importance is estimation and control when the feedthrough matrix (or another Markov parameter) is not full column rank. In the literature few results of limited applicability currently exist in this topic and virtually none relating to stability. Solving this challenge by developing estimation (or control) algorithms and stability conditions that can be easily implemented can have an immediate impact on applied problems.

Estimation and control problems that are outside the idealised linear-quadratic domain present further challenges. While the extended and unscented Kalman filters attempt to give approximate solutions, the recursions involved are correct only up to first or second order. A solution that is both very interesting and successful but limited only to systems defined on matrix Lie groups (e.g. robot navigation) is the invariant Kalman filter (IKF). The IKF applies the Kalman filtering algorithm to a deliberately chosen set of coordinates and the resulting distributions are no longer Gaussian in the original coordinates. It is particularly interesting to consider whether it is possible to extend the ideas behind the IKF to systems with different nonlinearities. It is thought that the theory of Koopman operators which uses (nonlinear) observable functions of the state may offer a path to achieve this for estimation and control problems alike. Monte Carlo methods are a successful alternative to address the challenges of nonlinear estimation. To ensure that the effective sample size (ESS) at the present time remains sufficiently high for accurate filtering, they use resampling techniques which have the downside of reducing the number of unique sample trajectories (degeneracy) (see discussion in Chapter 2). This limitation is particularly important for problems with large horizon lengths. It is interesting to explore whether it is possible to develop techniques which simultaneously exploit the benefits of Monte Carlo methods and Kalman filtering and smoothing. An approach that has been proposed and is inspired by both particle methods and the Kalman filter is the Gaussian sum filter which replaces the Gaussian distribution by a sum of Gaussians which resemble particles. This approach is not a Monte Carlo method since it does not rely on random sampling and it does not address the degeneracy problem. A promising direction is to consider particle filtering and smoothing techniques with a resampling stage that is based on a Gaussian mixture model (GMM) and exploits classical GMM fitting techniques.

Lastly, a range of vehicle applications can be explored based on the developments of this thesis. An example that stands out is the estimation of aero loads a vehicle experiences. This is an inherently difficult problem due to the complexity of performing accurate simulations which couple computational fluid dynamics (CFD) with vehicle dynamics. An alternative that can be explored is the approach considered within this thesis, namely the application of filtering and smoothing algorithms for simultaneous state and input estimation using a simple vehicle model and basic vehicle on-board sensors. This approach can be used to either verify CFD simulations (offline) or for real time stability control.

References

[1] Anderson, B. D. O. and Moore, J. B. (1979). *Optimal filtering*. Prentice-Hall.

[2] Antonov, S., Fehn, A., and Kugi, A. (2011). Unscented kalman filter for vehicle state estimation. *Vehicle System Dynamics*, 49(9):1497–1520.

[3] Basile, G. and Marro, G. (1969). On the observability of linear, time-invariant systems with unknown inputs. *Journal of Optimization theory and applications*, 3(6):410–415.

[4] Bitmead, R. R., Gevers, M. R., Petersen, I. R., and Kaye, R. J. (1985). Monotonicity and stabilizability-properties of solutions of the riccati difference equation: Propositions, lemmas, theorems, fallacious conjectures and counterexamples. *Systems & Control Letters*, 5(5):309–315.

[5] Bitmead, R. R., Hovd, M., and Abooshahab, M. A. (2019). A Kalman-filtering derivation of simultaneous input and state estimation. *Automatica*, 108, article 108478.

[6] Brockett, R. W. (1970). *Finite dimensional linear systems*. Wiley.

[7] Buchstaller, D., Liu, J., and French, M. (2020). The deterministic interpretation of the Kalman filter. *International Journal of Control*.

[8] Callier, F. and Willems, J. (1981). Criterion for the convergence of the solution of the Riccati differential equation. *IEEE Transactions on Automatic Control*, 26(6):1232–1242.

[9] Callier, F. M., Winkin, J., and Willems, J. L. (1994). Convergence of the time-invariant Riccati differential equation and LQ-problem: mechanisms of attraction. *International journal of control*, 59(4):983–1000.

[10] Chan, S., Goodwin, G., and Sin, K. (1984). Convergence properties of the Riccati difference equation in optimal filtering of nonstabilizable systems. *IEEE Transactions on Automatic Control*, 29(2):110–118.

[11] Cheng, Y., Ye, H., Wang, Y., and Zhou, D. (2009). Unbiased minimum-variance state estimation for linear systems with unknown input. *Automatica*, 45(2):485–491.

[12] Darouach, M. and Zasadzinski, M. (1997). Unbiased minimum variance estimation for systems with unknown exogenous inputs. *Automatica*, 33(4):717–719.

[13] Darouach, M., Zasadzinski, M., Onana, A. B., and Nowakowski, S. (1995). Kalman filtering with unknown inputs via optimal state estimation of singular systems. *International journal of systems science*, 26(10):2015–2028.

[14] Darouach, M., Zasadzinski, M., and Xu, S. J. (1994). Full-order observers for linear systems with unknown inputs. *IEEE transactions on automatic control*, 39(3):606–609.

[15] De Souza, C., Gevers, M., and Goodwin, G. (1986). Riccati equations in optimal filtering of nonstabilizable systems having singular state transition matrices. *IEEE Transactions on Automatic Control*, 31(9):831–838.

[16] Deshpande, A. S. (2017). Bridging a gap in applied Kalman filtering: Estimating outputs when measurements are correlated with the process noise [focus on education]. *IEEE Control Systems Magazine*, 37(3):87–93.

[17] Doucet, A. and Johansen, A. M. (2009). A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering*, 12(3):656–704.

[18] Gakis, G. and Smith, M. C. (2021). A limit Kalman filter for systems with highly uncertain inputs. *Submitted to Automatica*.

[19] Gillijns, S. and De Moor, B. (2007a). Unbiased minimum-variance input and state estimation for linear discrete-time systems. *Automatica*, 43(1):111–116.

[20] Gillijns, S. and De Moor, B. (2007b). Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough. *Automatica*, 43(5):934–937.

[21] Hou, M. and Muller, P. (1992). Design of observers for linear systems with unknown inputs. *IEEE Transactions on automatic control*, 37(6):871–875.

[22] Hou, M. and Muller, P. (1994). Disturbance decoupled observer design: A unified viewpoint. *IEEE Transactions on automatic control*, 39(6):1338–1341.

[23] Hou, M. and Patton, R. (1998a). Optimal filtering for systems with unknown inputs. *IEEE transactions on Automatic Control*, 43(3):445–449.

[24] Hou, M. and Patton, R. J. (1998b). Input observability and input reconstruction. *Automatica*, 34(6):789–794.

[25] Hsieh, C.-S. (2000). Robust two-stage Kalman filters for systems with unknown inputs. *IEEE Transactions on Automatic Control*, 45(12):2374–2378.

[26] Hsieh, C.-S. and Chen, F.-C. (1999). Optimal solution of the two-stage Kalman estimator. *IEEE Transactions on automatic control*, 44(1):194–199.

[27] Jazwinski, A. H. (2007). *Stochastic processes and filtering theory*. Courier Corporation.

[28] Julier, S., Uhlmann, J., and Durrant-Whyte, H. F. (2000). A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on automatic control*, 45(3):477–482.

[29] Julier, S. and Uhlmann, J. K. (1996). A general method for approximating nonlinear transformations of probability distributions. *Citeseer*.

[30] Julier, S. J. and Uhlmann, J. K. (1997). New extension of the kalman filter to nonlinear systems. *Signal processing, sensor fusion, and target recognition VI*, 3068:182–193.

[31] Julier, S. J. and Uhlmann, J. K. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422.

[32] Julier, S. J., Uhlmann, J. K., and Durrant-Whyte, H. F. (1995). A new approach for filtering nonlinear systems. *Proceedings of 1995 American Control Conference*, 3:1628–1632.

[33] Kailath, T. (1974). A view of three decades of linear filtering theory. *IEEE Transactions on information theory*, 20(2):146–181.

[34] Kailath, T. and Frost, P. (1968). An innovations approach to least-squares estimation–part II: Linear smoothing in additive white noise. *IEEE Transactions on Automatic Control*, 13(6):655–660.

[35] Kailath, T. and Ljung, L. (1976). The asymptotic behavior of constant-coefficient Riccati differential equations. *IEEE Transactions on Automatic Control*, 21(3):385–388.

[36] Kalman, R. E. (1960a). Contributions to the theory of optimal control. *Bol. soc. mat. mexicana*, 5(2):102–119.

[37] Kalman, R. E. (1960b). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions of the ASME*, 82(1):35–45.

[38] Kalman, R. E. (1963). New methods in Wiener filtering theory. In *Proceedings of the First Symposium on Engineering Applications of Random Function Theory and Probability, edited by JL Bogdanoff and F. Kozin, John Wiley & Sons, New York*.

[39] Kalman, R. E. and Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering, Transactions of the ASME*, 83(3):95–108.

[40] Keller, J.-Y. and Darouach, M. (1999). Two-stage Kalman estimator with unknown exogenous inputs. *Automatica*, 35(2):339–342.

[41] Kerwin, W. S. and Prince, J. L. (2000). On the optimality of recursive unbiased state estimation with unknown inputs. *Automatica*, 36(9):1381–1383.

[42] Kitagawa, G. (1996). Monte carlo filter and smoother for non-gaussian nonlinear state space models. *Journal of computational and graphical statistics*, 5(1):1–25.

[43] Kitanidis, P. K. (1987). Unbiased minimum-variance linear state estimation. *Automatica*, 23(6):775–778.

[44] Kucera, V. (1972). A contribution to matrix quadratic equations. *IEEE Transactions on Automatic Control*, 17(3):344–347.

[45] Kudva, P., Viswanadham, N., and Ramakrishna, A. (1980). Observers for linear systems with unknown inputs. *IEEE Transactions on Automatic Control*, 25(1):113–115.

[46] Kwakernaak, H. and Sivan, R. (1972). *Linear optimal control systems*. Wiley-interscience New York.

[47] Li, B. (2013). State estimation with partially observed inputs: A unified Kalman filtering approach. *Automatica*, 49(3):816–820.

[48] Molinari, B. (1977). The time-invariant linear-quadratic optimal control problem. *Automatica*, 13(4):347–357.

[49] Mortensen, R. (1968). Maximum-likelihood recursive nonlinear filtering. *Journal of Optimization Theory and Applications*, 2(6):386–394.

[50] Moylan, P. (1977). Stable inversion of linear systems. *Transactions on Automatic Control*, 22(1).

[51] Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

[52] Park, J. H., Han, S., and Kwon, W. H. (2008). LQ tracking controls with fixed terminal states and their application to receding horizon controls. *Systems & Control Letters*, 57(9).

[53] Sontag, E. D. (2013). *Mathematical control theory: Deterministic finite dimensional systems*, volume 6. Springer Science & Business Media.

[54] Sorenson, H. W. (1970). Least-squares estimation: from Gauss to Kalman. *IEEE spectrum*, 7(7):63–68.

[55] Strang, G. (2006). *Linear Algebra and its applications*. Thomson, Brooks Cole.

[56] Su, J., Li, B., and Chen, W.-H. (2015). On existence, optimality and asymptotic stability of the Kalman filter with partially observed inputs. *Automatica*, 53:149–154.

[57] Swerling, P. (1971). Modern state estimation methods from the viewpoint of the method of least squares. *IEEE Transactions on Automatic Control*, 16(6):707–719.

[58] Wan, E. A. and Van Der Merwe, R. (2000). The unscented kalman filter for nonlinear estimation. *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*, pages 153–158.

[59] Wang, S.-H., Wang, E., and Dorato, P. (1975). Observing the states of systems with unmeasurable disturbances. *IEEE transactions on Automatic Control*, 20(5):716–717.

[60] Willems, J. C. (2004). Deterministic least squares filtering. *Journal of econometrics*, 118(1-2):341–373.

[61] Yong, S. Z., Zhu, M., and Frazzoli, E. (2015). Simultaneous input and state estimation with a delay. *54th IEEE Conference on Decision and Control (CDC)*, pages 468–475.

[62] Yong, S. Z., Zhu, M., and Frazzoli, E. (2016). A unified filter for simultaneous input and state estimation of linear discrete-time stochastic systems. *Automatica*, 63:321–329.

[63] Zhou, K., Doyle, J. C., and Glover, K. (1996). *Robust and optimal control*. Prentice hall New Jersey.

Appendix A

Notation and lemmas

A.1 Spaces and norms

A real scalar, a real m dimensional vector and a real $m \times l$ dimensional matrix are denoted by $\mathbb{R}$, $\mathbb{R}^m$ and $\mathbb{R}^{m\times l}$ respectively. A square symmetric matrix $\Theta = \Theta^T \in \mathbb{R}^{m\times m}$ is positive (semi-positive or negative) definite and is denoted by $\Theta > 0$ ($\Theta \geq 0$ or $\Theta < 0$) if $\theta^T \Theta \theta > 0$ ($\theta^T \Theta \theta \geq 0$ or $\theta^T \Theta \theta < 0$) for all $0 \neq \theta \in \mathbb{R}^m$ and we denote by $\|\theta\|^2_{\Theta^{-1}}$ the norm on $\mathbb{R}^m$ defined by $\|\theta\|^2_{\Theta^{-1}} := \theta^T \Theta^{-1} \theta$ if $\Theta > 0$. Furthermore, we define the vector norm and induced matrix norm:

$$|x(t)|_\infty = \max_j |x_j(t)|, \tag{A.1}$$

$$|G(t)|_\infty = \max_i \sum_j |G_{ij}(t)| \tag{A.2}$$

where $x(t) \in \mathbb{R}^n$ and $G(t) \in \mathbb{R}^{n\times n}$. $\mathscr{L}^m_{2,e}$ denotes the space of vector signals of dimension m whose Lebesgue integrated squared 2-norm exists on any finite interval. Let $\mathscr{L}^m_\infty$ denote the space of Lebesque integral m dimensional vector functions of bounded ∞-norm. We define the signal norms:

$$\|x(t)\|_\infty = \sup_{t\geq 0} |x(t)|_\infty, \tag{A.3}$$

$$\|G(t)\|_\infty = \sup_{t\geq 0} |G(t)|_\infty \tag{A.4}$$

for signals which belong to the corresponding Lebesgue space $\mathscr{L}^n_\infty[0,\infty)$ or $\mathscr{L}^{n\times n}_\infty[0,\infty)$. (Strictly we should take the *essential supremum* in (A.3) and (A.4) though this will always coincide with the supremum for signals encountered here.) We further define the norm:

$$\|G(t)\|_1 = \max_i \sum_j \int_0^\infty |G_{ij}(\tau)| d\tau. \tag{A.5}$$

A.2 Asymptotically stable linear systems with bounded input

Lemma 24 *Suppose $x(t) \in \mathbb{R}^n$ satisfies:*

$$\dot{x}(t) = A(t)x(t) + f(t) \tag{A.6}$$

where $A(t)$ is continuously time varying and $\lim_{t\to\infty} A(t) = A$ with A Hurwitz, $f(t) \in \mathscr{L}_\infty^n[0,\infty)$ and $x(0) \in \mathbb{R}^n$. Then $x(t)$ is uniformly bounded, i.e. $\|x(t)\|_\infty < \infty$.

Proof: First we set:

$$M_1 = \|e^{At}\|_1 \tag{A.7}$$

noting that $M_1 < \infty$ since A is Hurwitz. Next we choose $\delta > 0$ such that $\delta M_1 < 1$. Since $A(t) \to A$ we can find t_0 such that $|A(t) - A|_\infty < \delta$ for all $t > t_0$. We next consider the free and forced solution of (A.6) on the interval $[0, t_0]$. We define:

$$M_2 = \left\{ \sup_{0 \le t \le t_0} |x(t)|_\infty : \dot{x}(t) = A(t)x(t), x(0) = x_0 \right\} \tag{A.8}$$

where $M_2 = M_2(x_0)$ and:

$$M_3 = \left\{ \sup_{0 \le t \le t_0} |x(t)|_\infty : \dot{x}(t) = A(t)x(t) + f(t), x(0) = 0, \|f(t)\|_\infty \le 1 \text{ for } t \in [0, t_0] \right\}. \tag{A.9}$$

We note that $M_2 < \infty$ and $M_3 < \infty$ follows from [6, Theorem 1, p. 40]. Hence:

$$|x(t)|_\infty < M_2 + M_3 \|f(t)\|_\infty \text{ for all } t \in [0, t_0]. \tag{A.10}$$

We next define:

$$M_4 = \|e^{At}\|_\infty. \tag{A.11}$$

We can see that $M_4 < \infty$ as follows. Let $A = TJT^{-1}$ be a Jordan decomposition and let $\bar{\lambda}$ be the largest real part among the eigenvalues of A. Then:

$$|e^{At}|_\infty \leq |T|_\infty |T^{-1}|_\infty e^{-\bar{\lambda}t}\left(1+t+\cdots+\frac{t^{n-1}}{(n-1)!}\right) \tag{A.12}$$

which is uniformly bounded since A is Hurwitz and thus $\bar{\lambda} < 0$. We now consider the solution of (A.6) for $t \geq t_0$. We can write:

$$x(t) = e^{A(t-t_0)}x(t_0) + \int_{t_0}^{t} e^{A(t-\tau)}u(\tau)d\tau \tag{A.13}$$

where we have defined:

$$u(t) = (A(t)-A)x(t)+f(t). \tag{A.14}$$

Then:

$$|x(t)|_\infty = \left|e^{A(t-t_0)}x(t_0) + \int_{t_0}^{t} e^{A(t-\tau)}u(\tau)d\tau\right|_\infty \tag{A.15}$$

$$\leq |e^{A(t-t_0)}x(t_0)|_\infty + \left|\int_{t_0}^{t} e^{A(t-\tau)}f(\tau)d\tau\right|_\infty + \left|\int_{t_0}^{t} e^{A(t-\tau)}(A(t)-A)x(t)d\tau\right|_\infty \tag{A.16}$$

$$\leq M_4(M_2+M_3\|f(t)\|_\infty)+M_1(\|f(t)\|_\infty+\delta \sup_{t_0\leq\tau\leq t}|x(\tau)|_\infty). \tag{A.17}$$

Combining (A.17) with (A.10) we obtain:

$$|x(t)|_\infty \leq \max\{1,M_4\}(M_2+M_3\|f(t)\|_\infty)+M_1(\|f(t)\|_\infty+\delta \sup_{0\leq t\leq t_1}|x(t)|_\infty) \tag{A.18}$$

for all $t \in [0,t_1]$ and any t_1. Since this is true for all t we can replace the LHS of (A.18) by $\sup_{0\leq t\leq t_1}|x(t)|_\infty$. Therefore:

$$\sup_{0\leq t\leq t_1}|x(t)|_\infty \leq \frac{1}{1-\delta M_1}\left(\max\{1,M_4\}(M_2+M_3\|f(t)\|_\infty)+M_1\|f(t)\|_\infty\right) \tag{A.19}$$

Since this is true for all t_1 the RHS is an upper bound for $\|x(t)\|_\infty$ which completes the proof. □

A.3 Stable linear systems with asymptotically vanishing input

Lemma 25 *Suppose $x(t) \in \mathbb{R}^n$ satisfies:*

$$\dot{x}(t) = Ax(t) + f(t) \tag{A.20}$$

where $A \in \mathbb{R}^{n\times n}$ is Hurwitz, $f(t) \in \mathscr{L}_\infty^n[0,\infty)$, $\lim_{t\to\infty} f(t) = 0$ and $x(0) \in \mathbb{R}^n$. Then $\lim_{t\to\infty} x(t) = 0$.

Proof: We consider the solution to (A.20):

$$x(t) = e^{At}x(0) + \int_0^t e^{A(t-\tau)} f(\tau)d\tau. \tag{A.21}$$

Since A is Hurwitz it follows that $\lim_{t\to\infty} e^{At} = 0$ and thus without loss of generality we set $x(0) = 0$ in (A.21). We set:

$$M = \int_0^\infty |e^{At}|_\infty dt \tag{A.22}$$

where $M < \infty$ using (A.12). Choose any $\varepsilon > 0$. We first set $\delta = \varepsilon/2M$. Since $\lim_{t\to\infty} f(t) = 0$ we can find t_0 such that $|f(t)|_\infty < \delta$ for all $t > t_0$. Then for $t > t_0$:

$$|x(t)|_\infty = \left| \int_0^t e^{A(t-\tau)} f(\tau)d\tau \right|_\infty \tag{A.23}$$

$$\le \int_0^{t_0} |e^{A(t-\tau)}|_\infty |f(\tau)|_\infty d\tau + \int_{t_0}^t |e^{A(t-\tau)}|_\infty |f(\tau)|_\infty d\tau \tag{A.24}$$

$$\le \|f(t)\|_\infty \int_{t-t_0}^t |e^{A\tau}|_\infty d\tau + \delta \int_0^{t-t_0} |e^{A\tau}|_\infty d\tau \tag{A.25}$$

$$< \|f(t)\|_\infty \left(\sup_{t-t_0\le\tau\le t} |e^{A\tau}|_\infty \right) t_0 + \delta M. \tag{A.26}$$

We note using (A.12) that $\lim_{t\to\infty} |e^{At}|_\infty = 0$ and thus there exists $t_1 > t_0$ such that:

$$\sup_{t-t_0\le\tau\le t} |e^{A\tau}|_\infty \le \frac{\varepsilon}{2\|f(t)\|_\infty t_0} \tag{A.27}$$

for all $t > t_1$. It follows that $|x(t)|_\infty < \varepsilon$ for all $t > t_1$. □

A.4 Convergence of asymptotically stable linear systems with bounded input

Lemma 26 *Suppose $x(t), x_1(t) \in \mathbb{R}^n$ satisfy:*

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \tag{A.28}$$

$$\dot{x}_1(t) = Ax(t) + Bu(t) \tag{A.29}$$

where $A(t)$, $B(t)$ are continuously time varying, $\lim_{t\to\infty} A(t) = A$ *with A Hurwitz,* $\lim_{t\to\infty} B(t) = B$*, $u(t) \in \mathscr{L}_\infty^m[0,\infty)$ and $x(0), x_1(0) \in \mathbb{R}^n$. Then* $\lim_{t\to\infty}(x(t) - x_1(t)) = 0$. *More precisely, given any $\varepsilon > 0$, $\exists T_0$ such that $|x(t) - x_1(t)|_\infty < \varepsilon$ for all $t > T_0$, where T_0 depends on $\|u(t)\|_\infty$ but not $u(t)$ itself.*

Proof: First note from Lemma 24 that $\|x(t)\|_\infty$ is finite. Moreover it can be seen from the proof of Lemma 24 that $\|x(t)\|_\infty$ has an upper bound which depends on $\|B(t)\|_\infty \|u(t)\|_\infty$ but otherwise does not depend on $u(t)$ (see (A.19)). Now write:

$$\dot{x}(t) - \dot{x}_1(t) = A(x(t) - x_1(t)) + (A(t) - A)x(t) + (B(t) - B)u(t). \tag{A.30}$$

The conclusion follows from Lemma 25 by noting that the choice of t_0 and t_1 can be made independent of the choice of $u(t)$ for a given bound on $\|u(t)\|_\infty$. □