Capacity and Efficiency of Multi-Sensor Networks

Robert Distel
Doctor of Philosophy
April 2015

Supervisor: Dr Mark Moreland
Assessor:
# Contents

List of Common Symbols   v

1 Introduction 1
  1.1 Structure 1
  1.2 Introduction 2
  1.3 Problem Description 5
  1.4 Justification 10

2 Prior Work 11
  2.1 Overview 11
  2.2 Optimal Transmission under Bandwidth Constraints 13
  2.3 State Estimation and Data Fusion with multiple sensors 17
  2.4 System Impairments 19
  2.5 Control of Systems with Communications Impairments 23
  2.6 Delay and Message Loss 24
  2.7 Network Models 25
  2.8 Critical Results 27
  2.9 Encoding and Transmission 28
  2.10 Encoding of Innovations from a single sensor 30
  2.11 Encoding a Group of Sensors 36
  2.12 Joint Encoding Innovations 38
  2.13 Discussion 41

3 System Design Relationships 43
  3.1 Overview 43
  3.2 Introduction 43
  3.3 The Expected Covariance 47
  3.4 Sensor Scheduling 55
  3.5 System Relationships and Invariant Properties 60
  3.6 Conclusion 65

4 Applications of the System Design Relationships 67
  4.1 Overview 67
  4.2 Introduction 67
  4.3 Markov Queue Depth M/M/1/Q 68
  4.4 CSMA-CD Shared Channel Access 72
CONTENTS

4.5 AWGN and CDMA Channels ........................................ 76
4.6 Dense Cluster of Sensors using M/M/1 Queue .................... 78
4.7 Conclusion .......................................................... 84

5 Efficient transmission of measurements 87
   5.1 Overview ....................................................... 87
   5.2 Introduction ................................................... 87
   5.3 Transmission with lost Innovations ............................. 88
   5.4 Encoding and Decoding ......................................... 94
   5.5 Rescheduling Sensor Transmission .............................. 99
   5.6 Innovation-like Sequences .................................... 101
   5.7 Conclusion ..................................................... 105

6 Co-operative Transmission 107
   6.1 Overview ........................................................ 107
   6.2 Introduction .................................................... 107
   6.3 Joint Encoding .................................................. 109
   6.4 Co-operative Transmission of Observations from Identical Sensors ........................................ 113
   6.5 Example - An HF Transmission System ......................... 123
   6.6 Co-operative Transmission of Observations with Unknown Variance .................................... 128
   6.7 Example - A data fusion system using a single satellite uplink channel .................................. 131
   6.8 Co-operative Transmission of Non-Linear Observations .......... 137
   6.9 A Practical Implementation ..................................... 138
   6.10 Simulation ....................................................... 143
   6.11 Conclusion ..................................................... 145

7 Conclusion 147

8 Appendix 151
List of Figures

1.1 Multi-Sensor System ........................................ 4
1.2 Centralised Fusion Model .................................. 5
1.3 Distributed Fusion Model .................................. 6
1.4 University of Michigan, Michigan Micro Mote, coin for scale. .... 7
1.5 CICADA Mk III Naval Research Laboratory .................. 8
1.6 Remote Controlled Beetle - Sato et. al. 2009 ................. 9

2.1 Reference System Model .................................... 12

4.1 M/M/1/Q Model .............................................. 69
4.2 Block Diagram of CSMA-CD system with a Concealed Station ... 73
4.3 Block Diagram of CSMA-CD System ........................ 74
4.4 Graph of $4\pi \rho \log \left( E \left( \frac{1}{\pi \rho} \right) \right)$ for system with constant $\rho$ .... 82
4.5 Plot of $4\pi \rho_{\text{effective}} \log \left( \frac{1}{\pi \rho_{\text{effective}}} \right)$ for $\rho_{\text{effective}}$ ......... 84

5.1 Comparison of performance of log MMSE with innovations, quantisation and loss ........................................... 90
5.2 Block Diagram of Encoder/Decoder for Transmission of Innovations ...... 94
5.3 Performance of late joiner after missing first 10 samples - error scale is log10 MSE ............................................... 98

6.1 2kHz Bandwidth Sensor Performance ........................ 125
6.2 100kHz Bandwidth Sensor Performance ...................... 127
6.3 Satellite Coverage for Intelsat 19 ............................ 132
6.4 2KHz bandwidth Sensor Performance ........................ 135
6.5 100kHz bandwidth Sensor Performance ...................... 136
6.6 System Block Diagram from Conceptual System ............... 139
6.7 Diagram of Shared Transmission Channel Across Multiple Tracks .. 141
6.8 Sensor Performance for OFDM system with co-operative transmission 143
Abstract

This thesis examines the relationship between a distributed sensor information fusion system and the underlying information transport system for systems with large numbers of simple sensors. Practical methods for the prediction of system performance under changes to system parameters such as capacity, number of sensors, system state function, system and sensor noise are developed. A method for transmitting information more efficiently than measurements but convergent despite message loss is proposed. A method of co-operative transmission is developed such that a set of sensors can shape their RF transmission to allow the fusion process to take place within their co-operative interference allowing better performance than optimal joint encoding with less power and bandwidth on a shared channel. These proposals are demonstrated with the simulation of two reference designs, an HF transmission and EHF satellite, fusing multiple sensor observations.
# List of Common Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>lower case, a vector or scaler</td>
</tr>
<tr>
<td>$X$</td>
<td>upper case, a matrix</td>
</tr>
<tr>
<td>$A,F$</td>
<td>continuous and discrete state transition matrices for a linear system. $F = e^{A\tau}$ for a continuous system sampled at discrete intervals $\tau$</td>
</tr>
<tr>
<td>$x,z$</td>
<td>state and observation vectors for a linear system</td>
</tr>
<tr>
<td>$\hat{x}, \hat{y}$</td>
<td>mean of state estimator and innovation for a linear system</td>
</tr>
<tr>
<td>$H$</td>
<td>observation model for a linear system</td>
</tr>
<tr>
<td>$P$</td>
<td>covariance of the model state estimator $\hat{x}$</td>
</tr>
<tr>
<td>$R$</td>
<td>covariance of the model observations $z$</td>
</tr>
<tr>
<td>$\chi,Q$</td>
<td>continuous and discrete process noise matrices for a linear system. $\chi$ and $Q$ are related through the continuous time Lyapunov equation $AQ+QA+I = F\chi F'$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>an event, $\gamma \in {0,1}$</td>
</tr>
<tr>
<td>$\gamma_{1:k}$</td>
<td>a sequence of events, $\gamma = {\gamma_1, \gamma_2, \ldots, \gamma_k}$</td>
</tr>
<tr>
<td>$\Pr(\cdots)$</td>
<td>Probability of</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>probability of an event occurring, i.e. $\alpha = \Pr(\gamma = 1)$</td>
</tr>
<tr>
<td>$K$</td>
<td>Kalman Gain</td>
</tr>
<tr>
<td>$I$</td>
<td>Identity Matrix</td>
</tr>
<tr>
<td>$I$</td>
<td>Innovation Matrix</td>
</tr>
<tr>
<td>$E(\cdots)$</td>
<td>Expectation of</td>
</tr>
<tr>
<td>$\lambda(\cdots)$</td>
<td>Eigen value of</td>
</tr>
<tr>
<td>$\sigma(\cdots)$</td>
<td>Singular value of</td>
</tr>
<tr>
<td>$\rho(\cdots)$</td>
<td>Spectral radius of. Equivalent to max $</td>
</tr>
<tr>
<td>$\rho$</td>
<td>a density</td>
</tr>
<tr>
<td>$v$</td>
<td>discrete process noise. Normally distributed with $\text{var } v = Q$</td>
</tr>
<tr>
<td>$w$</td>
<td>discrete measurement noise. Normally distributed with $\text{var } w = R$</td>
</tr>
<tr>
<td>$U, V$</td>
<td>unitary matrices</td>
</tr>
<tr>
<td>$u, v$</td>
<td>eigen vectors</td>
</tr>
</tbody>
</table>
## List of Common Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$</td>
<td>a diagonal matrix</td>
</tr>
<tr>
<td>$T$</td>
<td>an upper triangular matrix</td>
</tr>
<tr>
<td>$L$</td>
<td>a lower triangular matrix</td>
</tr>
<tr>
<td>$S$</td>
<td>a symmetric matrix</td>
</tr>
<tr>
<td>$D(\cdots)$</td>
<td>a decoder</td>
</tr>
<tr>
<td>$E(\cdots)$</td>
<td>an encoder</td>
</tr>
<tr>
<td>$\Delta(\cdots)$</td>
<td>an quantiser</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>a small step or difference</td>
</tr>
<tr>
<td>$p(\cdots)$</td>
<td>a probability distribution</td>
</tr>
<tr>
<td>$e, E, \varepsilon, \epsilon$</td>
<td>an error</td>
</tr>
<tr>
<td>$1_m$</td>
<td>$\begin{bmatrix} 1 \ \vdots \ 1 \ 1 \cdots 1 \end{bmatrix}_{m \times m}$</td>
</tr>
<tr>
<td>$1_{m \times m}$</td>
<td>$\begin{bmatrix} 1 \ \vdots \ 1 \cdots 1 \end{bmatrix}_{m \times m}$</td>
</tr>
<tr>
<td>$e_{m,n}$</td>
<td>the $m$th eigenvector of the $n \times n$ identity matrix</td>
</tr>
<tr>
<td>$I \left( \tau_{k</td>
<td>k-1, \gamma_{1:k-1}} \right)$</td>
</tr>
<tr>
<td>$J$</td>
<td>generalised eigenvectors for a matrix in Jordan Canonical form $J T J^{-1}$ where $T$ is comprised of Jordan blocks</td>
</tr>
<tr>
<td>$E_s, n_o$</td>
<td>Energy per symbol and noise power per symbol</td>
</tr>
<tr>
<td>$K_{mn}$</td>
<td>the commutation matrix s.t. $K_{mn} \text{vec } A_{m \times n} = \text{vec } A'_{n \times m}$</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Structure

There are four main components of this thesis beyond the introductory and concluding material. Introductory material is covered in Chapters 1 and 2, Chapters 3 and 4 cover message loss and its impact on the system, Chapters 5 and 6 cover methods to more efficiently transit information from sensors for use, while Chapter 7 is the conclusion.

Chapter 1 is devoted to introducing the reader to the problem and describes the authors motivation and a broad overview without proceeding into detail.

Chapter 2 is devoted to reviewing prior work in areas related to the problem. It introduces some of the mathematical elements that will be used later in the thesis and useful summaries of critical results used later.

Chapter 3 is devoted to the consideration of the lost observation problem, in particular the relationship between the transmission system and the system performance when observations are lost. The greatest concern is the system behaviour with lost observations when the process causing observation loss is coupled to the number of sensors in the system. Relationships are developed between the system size and system performance under a number of operational constraints.

Chapter 4 takes the relationships developed in Chapter 3 and applies them to a number of common media access models in order to demonstrate how the communications infrastructure selection can impact the system performance, with all other properties identical. Of particular interest this demonstrates that the behaviour in the limit of a large number of interacting sensors can be quite different depending on the access control method used.

Chapter 5 addresses a common method of improving the bandwidth efficiency. It discusses a method to achieve bounded estimation errors without re-transmission or feedback while sending less information than pure measurements from sensors and
without the divergence seen when transmitting pure innovations. The chapter is fo-
cused on ensuring only the minimum amount of information is sent from a sensor such
that an estimator receiving the information will converge and that no retransmissions
or data need be sent by the central estimator.

Chapter 6 introduces the concept of co-operative transmission as a mechanism to
deal with the interference generated by a large number of sensors operating within a
(relatively) confined space and use the large number of transmitters to allow reception
of the fused state estimate at distances that are unachievable with the individual
sensors transmitters. The results of Chapter 6 build upon the previous results, using an
example system simulation to demonstrate the utility of the process. Within Chapter
6 an example system is progressively extended to incorporate material from prior
chapters.

1.2 Introduction

The problem of estimating the state of a system using multiple sensors has a long
and interesting history. Starting with R.E Kalman’s paper [1] in 1960 articles on
state estimation have proliferated. Now, over 5700 papers directly reference Kalman’s
paper. Another key author, Y. Bar-Shalom, published the earliest definitive book [2]
in 1995 on multi-sensor multi-target tracking, and is now directly cited by over 900
articles excluding references to earlier incarnations of the work as lecture notes (over
500 more references).

Wireless sensor systems for multi-sensor fusion have become an area of increasing
interest in the past 15 years. As will be seen in Chapter 2, multiple papers have been
written regarding routing of information, efficient modulation, power management,
distributed detection, quantisation, distributed parameter estimation, delay, message
loss and control of distributed systems. However, over the past years the literature
has divided up into two fundamentally different camps. One segment, mostly occu-
pyed by systems and control engineers as well as mathematicians has concentrated
on the development of algorithms and mathematical descriptions of idealized distrib-
uted systems. These systems are parameterised in various ways (stochastic models
for delay, loss, quantisation bandwidth limits) and analytic results generated for these
systems. The second grouping relates to computer science, and to a lesser extent
communications engineers, who have concerned themselves with transmission man-
agement, routing protocols, power management and media access. The literature is
more descriptive rather than analytical and is generally agnostic to the application of
the sensor network.

Between these two groups there is little literature that attempts to account for
the properties of the sensor network when developing and analysing algorithms for distributed estimation. This becomes obvious when algorithms are proposed that require distribution of small amounts of data to the majority of sensor nodes (a highly inefficient packet structure where $n$ sensors cause an $O(n^2)$ increase in traffic), rely on ordered or reliable delivery of information (which is not available within a bounded time delay), or involve “swarms” of sensors, and specific means to manage and model the network complexities are not provided.

Although analysis of various impairments are well represented in the literature to date there is little work that attempts to reconcile the types of impairment with the underlying network infrastructure that causes the impairment. Typical discussions by systems engineers and mathematicians assume that the impairments experienced in sensor networks are due to “radio frequency interference” that cause “packet errors”. Thus the analytic models used for these systems assume that the underlying statistical processes are independent from the system state.

The reality is that many assumptions made about the nature of the impairments are invalid in fielded systems. Many forms of system impairment relate to the methods used to access shared media (see [3]), implicit and explicit queues in the system (see [4] - noting that the TCP protocol implicitly forms logical queues) and the non-linear nature of capacity sharing on radio and fixed line channels. Further, and more importantly, these impairments depend on the state of the system. The number of sensors, the number of objects to be estimated and the rate at which the estimators are updated significantly impact the system performance through the mechanism of congestion. Neglecting congestion in the analysis and development of distributed sensor algorithms creates systems that are “brittle” in a systems engineering sense. That is, systems that are highly sensitive to small variations in their operational environment. Sensitive systems experience complete failure rather than graceful degradation under minor excursions from their designed operational envelope.

Figure 1.1 is an example of a multi-sensor system. At each point A, B and C a different form of impairment can occur. Node A is a wireless access point. Multiple sensors transmit information through node A using a wireless link access protocol. Node B represents a packet aggregation node (such as a bridge, router or some cryptographic equipment) where incoming traffic from multiple sources is stored and forwarded on a high capacity uplink. Node C represents a simple point to point link. This is a fixed capacity channel where traffic is enqueued at the origination point prior to transmission.

The wireless access point (Node A) experiences congestion impairments caused by contention between sensors in accessing the common wireless channel. Common chan-
nel access impairments are characterized by exponentially greater delays and packet loss (due to collisions) as utilization of the channel increases. Many fixed line and wireless protocols using statistical time division multiplexing exhibit this property (802.11a,g,n are modern examples). Even when ideal fair media access (such as slotted time division multiplexing in Link-16 and Link-22) is provided by the protocol a virtual queue is formed that serializes access to the shared media. Other wired and wireless multiplexing systems (code division and frequency division) exhibit co-channel interference which reduces the net channel capacity.

Packet aggregation nodes (Node B) represent points where network traffic from multiple sources is buffered. The stochastic behaviour of these buffers is well understood with analytic expressions for simple systems. The behaviour of the queues in terms of packet loss and delay is coupled with the amount of traffic offered to the queue, and the statistical nature of the traffic offered by different sources (distribution and independence). Different methods for processing traffic (within the same priority level) causes the system to exhibit different behaviour.

A point to point link (Node C) will exhibit similar behaviour to a packet aggregation node when a single sensor enqueues random traffic onto a single link. Even though the capacity and the availability of the link is fully deterministic the presence of random processes in the observed system (number of objects that are currently observable, probability of detection of an observable object, number of incorrectly attributed observations) affects the offered traffic on the link. This variation requires some form of message queuing discipline to adapt the randomised offered traffic rate.
1.3 Problem Description

The goal of this research is to quantify the coupling between the sensor network and the overall sensor system performance and then to develop mechanisms to fully exploit the sensor capabilities despite the restrictions of the network. In particular the relationship between the congestion of the sensor network and the sensor system is of interest. The impact of the network is to be considered in both the single remote sensor case and the multi-sensor fusion system. The goal is to understand the sensitivity that different systems, system architectures and algorithms have to variations in the characteristics of the transmission network.

Two distributed fusion architectures are considered: centralized fusion and distributed fusion. Figure 1.2 represents a centralised fusion system. $T_1$ through to $T_3$ represent targets that can be observed by one or more of four sensors $S_1$ through to $S_4$. Multiple sensors may observe the same target, multiple targets may be observed by the same sensor. In the centralised fusion model all information is presented at a single location where the information is “fused” to form the optimal estimate of the state of all the targets given the observations of the sensors. The communications network used by the centralised fusion system is modelled as a single tree rooted at the fusion centre. This represents the contention each sensor has with other sensors attempting to transmit information to the central node. It will be assumed that there is a single ingress bottleneck at some central point which will dominate the system performance.

![Figure 1.2: Centralised Fusion Model](image)

The other fusion architecture to be considered is the distributed fusion system (see Figure 1.3). $T_1$ through $T_3$ and $S_1$ through $S_4$ have the same meaning as above.
The additional observer node receives reports from each distributed tracking group and displays them to the end user. The distributed fusion model is a generalization of the centralized model. It is a set of trees, with each trees’ root located at one of the sensors. Each sensor is performing fusion with \( m \) other sensors, thus for \( n \) sensors there will be \( n \) queues processing traffic from \( m \) sensors. Congestion is assumed to occur only (or predominantly) on the final link to the sensor and consequently each queue can be treated independently, as if it were a centralized model. Thus the distributed case is reduced to \( n \) centralized cases with traffic \( m \). Although the distributed fusion model does not require all sensors to communicate with every other sensors it does require all sensors to communicate with the set of sensors that are tracking the same target. Furthermore it still requires all sensors to report (either directly or indirectly) to the central observer (the client of the sensor system). Thus there are \( j \) targets each observed by \( m \) sensors forming \( j \) sensor ensembles (a set of sensors jointly fusing the same target). This model ignores the additional traffic generated in order to discover peer sensors and perform target association. Sensors would need to transmit the list of targets to adjacent sensors that could potentially be observing the same targets causing a higher traffic level than this model would assume.

Information for fusion can be in the form of measurements or estimates. Optimal fusion algorithms exist for both, and include the handling of delayed and non-delayed measurements. Due to the behaviour of the encoding system there is a convergence in the type of information transmitted by the sensors. It can be shown that an optimal encoder in an error free network will transmit an “innovation-like” sequence of a sensor local estimator based on the current observation. In this case of a linear system this
information completely summarises the difference in estimator between the previous transmission and the current transmission, including transmission quantisation. Using the innovations and an agreed upon set of target dynamics either the original observation, or the current sensor state estimator can be recovered. Thus any discussion of system behaviour where sensor transmission occurs need only concern itself with the optimal fusion of the innovation sequences generated by the sensors. As the innovation sequences allow recovery of either local estimates or local observations any convenient fusion scheme can be used to analyse the optimal behaviour of the sensor network.

This investigation is particularly applicable to multi-sensor, multi-target tracking systems using centralized fusion where the number of sensors is very large and the communications system used by the sensors may impact on neighbouring sensors. The stochastic behaviour of the measurement process with missed reports and uncertain detection provide a fundamental basis for the queuing behaviour exhibited. This behaviour occurs even under conditions of totally deterministic network behaviour with no other traffic present. This behaviour is further amplified when optimal encoding is used, as the number of bits required to encode the information to a defined degree depends on not only how many sensors can see the same target, but additionally how long the target has been visible. Current developments in sensor solutions such as the University of Michigan’s M$^3$ project (Michigan Micro Mote see Fig. 1.4) encapsulates an optical sensor, processor, RF transceiver and battery in a 1 mm$^3$ package. The goal is to allow “Smart Dust” applications where the transmission, sensing and processing
resources of a large number of sensors distributed through a (relatively) small volume will provide a highly redundant, high performance sensor system. Concepts range from using microUAV, aerial deployment using small balloons (see Fig. 1.5) or even using remote controlled insects [5], [6](see Fig. 1.6)

A future system may well approach the limits specified by the size of antenna’s used for transmission and sensing with optical systems operating at IR or visible wavelengths allowing sizes that are sub millimetre in dimension. Given low cost and small size the opportunity arises to allow sensor clouds of thousands or more of interacting sensors to form a single virtual sensor of superior capability.

This thesis will introduce the concept of a “sparse” sensor network and a “dense” sensor network. The sparse network is well represented in literature. It comprises a sensor system where the density of sensors is such that the transmission systems used by the sensors introduces negligible interference with adjacent sensors. When analysing the performance of a sparse network the interaction between sensors can be neglected and the system may be analysed simply as a network of interconnected graph of dynamically changing point to point links. A dense network is the opposite. With a dense network adjacent sensors transmission systems will interact to the extent that modelling as point to point links is an unacceptable simplification.

A system may be designed with a dense network because:

- the selected transmission system has a long range with low losses. Compare the propagation of an HF based system versus an optical system. An HF system can provide worldwide coverage while IR systems can provide only direct line of site reception and is blocked by any intervening object.
the selected transmission media has low path loss. A system operating in free space has a lower propagation loss than a terrestrial system. Free space path loss follows the familiar inverse square law, however propagation over a ground plane with clutter the path loss can range is considerably more complex.

the selected application requires multiple sensors to observe the same system. Multiple sensors can be fused to improve the quality of the estimation of the system state. The multiple sensors must be close enough to observe the same state, and by being close may induce interference with each other. Sensors with reduced capability and lower cost may need to be closer than more sophisticated sensors, increasing the required density of the system.

The purpose of this research is to investigate the impact of the underlying network infrastructures congestion and co-channel interference on system performance and develop methods for minimizing the impact that network congestion and interference has upon the overall system. Of particular importance is the observation that congestion is coupled to the system through the number and state of the targets observed by the system; that is the act of observing the targets changes the congestion state of the system and reduces the capability of the system to observe additional targets, or observe the current target set as accurately.

\[ r^{-\beta_1} \left( 1 + \frac{r}{r_c} \right)^{-\beta_2} \]

where \( \beta_1, \beta_2, r_c \) depend on the environment and have been determined empirically to range \( 1.5 < \beta_1 < 0.5, -3 < \beta_2 < 0.5 \). A simpler model \( r^{-\beta} \) with \( 2.6 < \beta < 3.4 \) is also used for cellular networks [7].
1. Introduction

1.4 Justification

To date no work has been undertaken that deals with system level sensitivity of a dense distributed estimation system. In particular no attempt has been undertaken to link the system sensitivity to the underlying network characteristics. Prior works on sensitivity (for example [53], [57]) have focused on model sensitivity (how sensitive is the estimation to errors in the model). Furthermore there is little work that accounts for the network performance in evaluating and proposing estimation techniques. Techniques and analysis has focused on simplified models suitable for some aspects of wireless systems which do not adequately model actual wireless or fixed line system performance when the system is congested.

It can be argued that the increased capacity of fixed and wired networks makes the investigation of congestion behaviour in sensor systems irrelevant. The increased capacities allow systems to be over provisioned such that congestion never occurs. This argument ignores the fact that larger and more complex systems are proposed, estimating a greater number of parameters with a greater number of sensors than ever before. The provisioning of additional capacity has had the opposite effect. Rather than decreasing the congestion in networks the additional capacity has enabled an ever increasing number of applications which contend with the estimation application. Expectations for network capabilities has grown at a rate commensurate with capability, leaving the overall network capacity constrained in a manner similar to systems of past eras.

Thus congestion management is an evergreen topic, as relevant now as ever in the past. The design of robust systems is of critical importance for certain sets of applications, and the knowledge that a system fails in a “soft” fashion is invaluable.
Chapter 2

Prior Work

2.1 Overview

This chapter is intended to cover works related to the investigation, particularly in regard to the system impairments, optimal transmission, and network modelling. The intention is to provide an outline of the structure of existing papers rather than a detailed analysis of them. A general observation is that while many papers have some relevance (or are of key historical importance) very few papers have material specifically focused on the area of investigation. A good overview of the domain is given in [8] in which many of the problems that are associated with sensor network congestion are touched upon.

In order to assist in the following discussion, and to provide a consistent framework, a standard centralized fusion reference model for the system under investigation is provided.

The model in figure 2.1 lists elements and impairments that will be considered. Green sections are associated with data fusion and centralized station, red sections with channel encoding, purple with network impairments and blue with a control system. Not all portions of the system are present, or relevant for all papers, but this system model is useful for providing a consistent framework. The investigation includes all parts with the exception of the controller feedback path - for the purposes of the investigation there is assumed to be no feedback channel or control system.

The first 6 sections in this chapter provide a brief overview of the relevant literature and its place within the system. The sections are

2.2 Optimal transmission under bandwidth constraints - comprising of Generic quantiser design, Single channel coder/estimator and Optimal encoding over multiple channels.
2. Prior Work

2.3 State Estimation and Fusion with multiple sensors - comprising of Estimation (CEO problem) and the Fusion of multiple sensors over bandwidth limited channels.

2.4 System impairments - comprising of Message loss and Message delay.

2.5 Control of systems with communications impairments - comprising of references to Quantisation and Delay.

2.6 Delay and Message loss - comprising of references to the properties of estimation in the presence of Lost and Delayed measurements.

2.7 Network models - comprising of Congestion in wireless networks, Channel models and Jump linear systems.

Following these sections three critical papers ([18], [38], [108]) are considered in detail. These papers are selected as they provide useful results for encoding a single sensor and a group of sensors using joint encoding and innovations. In 2.10 the optimality of the transmission of innovations is discussed as well as the limitations. 2.11 discuss the rate distortion bounds for a group of sensors and 2.12 incorporates the use of innovations with a group of sensors. The most significant results from these papers are examined in detail and summarised for use in later chapters.
2. Prior Work

2.2 Optimal Transmission under Bandwidth Constraints

This section investigates literature related to the optimal transmission of information from one or more sensors in a bandwidth constrained environment. This is of importance in the investigation of congestion due to the tight coupling between message size and network congestion. Referring to figure 2.1 this section relates to the red elements. These elements provide the means to quantise a set of continuous valued information elements, transmit and recover the information element to within some specified error.

An enormous body of literature is devoted to the design of quantiser that minimize the reconstruction error of the quantised original. Early papers focus on scalar quantities with a known source distribution. Historical development of algorithms revolved around the development of efficient algorithms for the design of these quantisers. The original Joel Max paper [9] and the associated unpublished Lloyd paper [10] (written in 1957 but not published until 1981) develop an optimal quantisation of a scalar parameter drawn from known distribution. Proof that the uniform quantiser is asymptotically the lowest distortion quantiser as the number of bits increases was provided in [11]. Methods to design encoders and decoders based on channel characteristics (noise and distortion) [12] and with greater ease [13] were developed.

Vector quantiser design followed from the scalar case. Distortion costs are used to provide a measure such that a scalar quantity is minimized. Various measures are used to define the distortion cost functions. Linde’s [14] vector quantiser design provided a generalized approach to provide an optimised quantiser using any specified distortion measure. Methods for efficiently solving Linde’s method are proposed for designing optimal vector quantisers [15]. Typically, subsequent papers use the Euclidean Norm or a Weighted Euclidean Norm for determination of the distortion cost function.

Though the various forms of quantiser are quite general in their application, their use is sub-optimal in the case where there is correlation between quantised samples. However it is well known that a quantiser followed by an entropy encoder can approach Shannon’s bound [11]. The disadvantage of using entropy encoders is the requirement to collect samples prior to encoding in order to maximize the encoding efficiency, thus increasing latency.

The single channel coder estimator is a special case of the vector quantiser problem. The main difference is that the statistics of the source change over time and so non-recursive techniques (such as fixed quantisation) will not be able to cover all states. This is particularly true of systems where the state is free to cover $\mathbb{R}^n$ but the quantised estimate is bounded.

The concept of a sequence of outputs from a coder/estimator pair is introduced in [16]. Conditions for bounded and convergent behaviour are developed for a class
2. Prior Work

of coders. Basic results are obtained for some vector and scalar coders. [17] extends [16] by considering the noiseless plant observed in the presence of Gaussian noise. Further bounds for convergent behaviour are applied, although the state estimated is considered to be scalar.

The case of the noisy plant with noisy observations is considered in [18] and [19]. Conditions for bounded estimation dependent on data rate are provided for a class of sub-optimal estimator. The optimal case for an error free channel is considered but is not analytic. The structure is determined to be a Kalman filter followed by encoding of the state estimate based on the previous state estimates. The information transmitted is the “difference” between the posterior distribution at the transmitter given the new observation, and the prior distribution at the receiver. It contains all new information and any corrections required to prevent divergence between the receiver’s state estimate and the transmitter. In the sense that the transmission contains only new information then the transmission can be thought of as an innovation, but it contains additional information beyond the difference between the prior and posterior distributions of the estimator at the transmitter. The paper derives results for the scalar plant state case but claims that the results are readily extensible to the vector case and the results provided hold.

Further extension to the vector case is provided in [20] where sufficient conditions for convergence are defined. An upper bound for the error at convergence is found. A variation where the source observed is a two state HMM is investigated in [21]. The goal is to recover the state after transmission with the least error.

A different approach to rate allocation for a target tracking application is taken in [4]. This paper attempts to optimise the data rate and avoid congestion by designing a congestion management algorithm that is similar in intent to TCP. The impact of congestion is modelled (loosely) and used to derive a rate control algorithm that allows each source to “choose” the amount of information to send each period based on its local estimation of the networks congestion. This system is demonstrated and compared to a system using TCP. This paper is of interest as it attempts to address the issue of a network impacting on the performance of a tracking system, which is the core issue covered by this report.

An interesting variation on the problem is [22]. Rather than couching the problem in terms of the estimation of a stochastic process, the problem is one of estimation of an “uncertain system”. The theory of uncertain systems is a parallel to stochastic systems and originated in [23]. The primary concept is that the unknown system influences are not modelled as random numbers with PDF’s, but are merely constrained to belong to a bounded set. The outcome is similar to that of estimation on a bounded uniform prior. This is further developed through [24] and reproduces many of the same results
as classical Kalman Filtering, but from a different set of initial assumptions. Because of the closed loop - constrained - structure of the estimation problem in [22] an analytic bound with uniform quantisation is derived on the system data transmission rate to support the required constraints. Some extension is provided in [25] to the use of optimal quantisers using Voronoi spaces as the optimal encoder.

The channel effect in the presence of bit rate constraints and channel noise is investigated by [26]. It shows that the ideal estimators discussed in earlier papers will diverge (almost surely) in the presence of channel noise. Furthermore it presents a minimum bit rate constraint based on the dynamics of the system and the channel error rate below which it is impossible to keep the estimation error bounded. A related paper [27] discusses the problem when the communications channel includes feedback. The derivation is for noiseless plants and observations.

Of particular importance in the multi-sensor case is the optimal transmission of data where the data relates to observing the same process. The Slepian-Wolf encoder provides optimal encoding for two correlated sources on ideal channels. [28].

An alternative view is that two observers see the same object and encode it, though only a single observation needs be recovered. Wyner and Ziv show how to make a lossy encoder/decoder that achieves this in [29].

Other work provides bounds in the event of encoding and decoding in the presence of side information [30],[31],[32] for 2 or more observations. These results provide an extension to the single channel quantisers as they allow the accuracy to be improved by quantisation of multiple sources, but increase the efficiency by removing the redundant information between the channels.

The results above are only usable if the sensor system supports feedback of the fused data to allow correlation of sensor reports to centralized fused results. The innovation sequence transmitted by a sensor when using the previous fused estimate as the prior distribution contains only the adjustments caused by information specific to the local sensor at the time. Alternatively each sensor transforms its measurement space using orthogonal basis functions (i.e. Fourier transform, Walsh transform, wavelet transform) and transmits only a single transform coefficient. Other sensors are assumed to only transmit the other transform coefficients. cf. [33, 34]. An interesting point to note with [33, 34] is that for each orthogonal component that is estimated, a different number of sensors is used. Because the orthogonal components are based on a Fourier series, a greater number of sensors must estimate the lower orders than the higher orders. If a Walsh series is used as the basis (or a bi - orthogonal wavelet) the same number of sensors can be used for each component, up to the desired precision. An intuitive understanding of this is achieved by considering how much information is gained about an observation if the same basis is used multiple times versus using a
2. Prior Work

different (orthogonal) basis each time. Obviously an orthogonal basis generates more information each time as it is not correlated to previous observations, while observations using the same basis must be correlated and have more mutual information.

The area of investigation is related to these papers through the following question. Which is better for a congested system, halving the data rate through discarding half the measurements, or halving the data rate through discarding half the bits of all the measurements but preserving all measurements?

As a uniform quantiser followed by an entropy encoder (assuming uncorrelated samples) achieves Shannon’s bound then the rate distortion function lower bound can be established and a best case estimate of the Rate Distortion function made. This provides part of the information required to answer the question above. The difficulty of this assumption is that an optimal entropy encoder takes an unbounded number of samples and has an unbounded coding latency. Slepian-Wolf encoding allows redundancy in the information between multiple sensors to be exploited to reduce the total bandwidth and provides a bound for the case of two sensors (and implicitly any system of $2^N$ sensors).

An important theoretical observation that relates to the problem is that an optimal encoder for a sensor fusion system in a reliable channel is to determine the difference between the prior distribution of the estimator at the receiver and posterior distribution of the estimator at the transmitter, and then encode and transmit this difference. By only encoding and transmitting the difference the system is only transmitting “new” information and consequently has removed redundancy. The removal of redundancy optimises the system provided the transmitter can adequately estimate the receiver’s state. This impacts on two aspects of the channel. Firstly, the channel must be reliable to enable the transmitter to estimate the receiver’s state. It can be shown that the loss of a single optimal transmission can cause the divergence of the estimator. Secondly the amount of information transmitted varies over time. Earlier updates of the state estimates have larger covariances than later updates, and require more bits to transmit with the same error. Thus the required data rate is greatest when a target first appears, and reduces over time for an optimal system. This impacts the timing of congestion in the system. A system is most likely to congest when many new targets are introduced when an optimal compressor is used.
2. Prior Work

2.3 State Estimation and Data Fusion with multiple sensors.

In this section literature relating to state estimation using multiple sensors is considered. Of particular importance is literature that considers bandwidth constraints. In the context of figure 2.1 the discussion concerns the green section (although in many cases the red encoder section is inseparable).

The distributed estimation problem (often referred to as the CEO problem) is a closely related problem. These works are referenced as they cover a distributed estimation problem in a multi-dimensional, band limited context. They cover aspects of quantisation and multi-channel optimal encoding. This can be considered as the non-dynamic case of single target, multi-sensor tracking with sensor fusion.

The distributed estimation problem with arbitrary quantisation is covered in [35]. A quantisation rule for optimally transmitting updates to a locally known estimate is established. A simplification for vector estimation is covered in [36] using multiple linear quantisers and entropy encoders. The optimal rate allocation for vector Gaussian CEO problem is covered in [37]. A reformulation allows the problem to be cast as a convex optimization problem. Strong results for the quadratic Gaussian vector CEO problem are provided in [38] for the coupling of rate allocation and distortion. Estimation with different wireless access models in a rate and power constrained system is covered in [39].

A related area is the use of single bit quantisers and broadcast messages to distribute the estimate to all sensors. A basic estimator using single bit sensors is covered in [40], [41] while a simplified linearised system is covered in [42]. A distributed estimation where a single message is transmitted and made available to all other sensors is covered in a number of papers [43], [44], [45] and related to [46]. While single bit transmissions by themselves are of limited utility in a packet switched network (considering that a single message will comprise of a header and a single bit) a multiplex of multiple single bit “messages” sent over a synchronous stream could be efficiently implemented using variable length codes. Furthermore the single bit message offers a useful, and analytically simple, way of developing distributed quantisers.

Related to distributed one bit quantisation is the use of one bit quantisers with non-contiguous quantisation zones [33], [34]. This appears equivalent to each sensor estimating on separate orthogonal basis using Walsh functions as the basis set.

These results are of similar utility to the results associated with multi-channel encoding. Of particular interest are results such as [38] where upper bounds for the rate distortion function are generated. In the case of identical sensors with a quadratic cost an exact sum rate distortion is determined. It also shows that with correlated
observations that a decrease in bandwidth at one sensor can be exactly compensated for by increasing the bandwidth at other sensors to preserve the same net bandwidth overall.

The related problem of sensor fusion over limited capacity channels covers the case of estimation of a dynamic system.

In [47] scalar measurements are quantised using a non-linear quantiser. In practice this paper accurately encodes expected observations, while coarsely coding unexpected observations by using a A-law / µ-Law logarithmic encoder. This is conceptually similar to only coding the innovations using a Gaussian non-linear encoder which results in a system that was predicted by [19].

The fusion of two sensors with finite bandwidth constraints and bounds for accuracy of state estimation are covered in [48]. Using Slepian/Wolf encoding the scenario provides a firm lower limit for two sensor fusion using the transmission of innovations in a reliable network. An open question is what happens when sensor reports are lost (i.e. innovations lost). In the case of a single sensor this will cause an unbounded error but in the case of multiple sensors is this still true? A similar extension, but only for scalar measurements, is covered in [49]. 1-D measurements are transmitted and fused in the presence of non-unity detection, clutter and non-ideal data association.

The CEO case can be considered a specialization of the general transmission problem. This scenario is interested in the formation and transmission of an estimate formed by multiple observers. The total bandwidth required to achieve this is reduced compared to Slepian-Wolf encoding as it is not desired to recover all original measurements, rather only a single estimate (compared to the recovery of a single measurement given side information).

Track fusion of multiple sensors in bandwidth limited environment is covered in [50]. The solution is couched in terms of lossy Wyner-Ziv coding. A rate distortion function for the specific case of equivalent sensors with Gaussian noise is provided. Furthermore it claims that “on demand” track fusion performs worse than centralized fusion as the number of sensors increase for a fixed sum data rate due to the increasing degree of correlation between sensor reports as quantisation increases.

For the static system it would seem that these two processes should yield an identical result. However, the case of a dynamic system is not as clear-cut. A dynamic system can be considered to be a system that causes the entropy of the estimate to increase with time in the absence of observations. Transmitting the observation decreases the entropy of the estimate, but only to the degree that the transmission is accurately quantised. Consequently there are two opposing processes at work, one that requires frequent transmissions to deal with the dynamic nature of the system,
while the other is accuracy of the quantisation process. A final observation that will not be pursued further in the rest of this thesis is the considerable problem of allocation. The allocation of codes into the code book of joint encoders requires care to ensure that all encoders can encode observations without knowledge of the state of other encoders (i.e. the number of sensors observing is unknown to each sensors and randomly changes at the fuser). If each sensor “knows” how many sensors can currently “see” each target the optimal allocation of codes to achieve the optimal rate distortion bounds can be performed. However this requires that observations be allocated to tracks and this information be back propagated from the fusion centre to the sensors, an unrealistic requirement. A sub-optimal allocation of rate where each sensor independently determines its code-book and rate will lead to robust behaviour that is independent of the behaviours of other sensors. A further difficulty is that maximum likelihood decoding of at the receiver is a NP-complete problem in the same class as the maximum likelihood decoding of all block-codes, hence increasing the number of sensors becomes computationally infeasible.

An alternative perspective on the state estimation problem is the concept of analog fusion as discussed in [51] - Chapter 4 and the earlier conference paper [52]. These papers demonstrate that under specific conditions it takes less energy for a sensor to transmit the observations for use in the fusion centre in an analog form rather than with a digital encoding. Specifically a method of optimization of energy utilisation is demonstrated that uses the linear properties of the transmission media to perform a distributed fusion operation. This concept will be examined more closely in Chapter 6 of this thesis and extended.

2.4 System Impairments

The introduction of congestion into a network produces two main effects, message loss and delay. This section analyses papers on message loss and delay. There are two main threads to these papers. First, given that a particular impairment is present: what is the impact on a sensor or system of sensors? Secondly, what is the optimal way of mitigating the impairment? Considering figure 2.1 the impairment models occupy the purple elements.

Although not directly related to network message losses, a similar problem was addressed in early literature on data association. Early issues included the determination of stability of Kalman filtering under the assumption of non-unity detection probability (and also false detections). The problem of non-unity detection probability is identical to the problem of Bernoulli losses of messages on a communications link. An example of this (including derivations of stability criteria) is in [53] in the
2. Prior Work

context of incorrect data association false alarms and false detections.

The problem became active in closed loop control of jump linear systems. An example is [54] which defines the control problem when the jump and system state variables are known. This paper is later used in the context of packet loss estimation.

Using earlier results from control theory we see papers using single channel models and the Algebraic Riccati Equation to model state estimation in the presence of message deletions.

In [55] a dual sensor system is modelled with a Bernoulli channel causing message loss. It provides useful bounding relationships for stability of the estimate wp. 1. In [56] a simpler case with a single sensor and Bernoulli channel loss is considered. Bounds for stability and convergence of the estimated error covariance are determined. A further variation is the investigation in [57] of the impact of smoothing on the critical arrival rate and upon the probabilistic constraint (i.e. the covariance is bounded to within some limit with some probability). The above bounds are improved upon in [58]. In this paper an interesting approach to decomposing the covariance bound is used to derive tighter bounds for the case of the observation matrix $H$ being invertible on the observable subspace of $(F, H)$ then the critical probability is exactly $1 - \frac{1}{\rho(F)^2}$ rather than bounded by that quantity. A further improvement is provided in [59] in which the critical value is shown to be exactly $1 - \frac{1}{\rho(F)^2}$ under an even weaker set of assumptions than prior works. In particular it is shown that for an observable system the only requirement is that $F$ be diagonalisable (i.e. no repeated eigenvalues with the same magnitude). This elimination of the requirement for invertibility of $F$ significantly widens the applicability of [56]. Another interesting observation is made by [60] which investigates the problem of intermittent transmission of state estimates (as opposed to measurements) under the same assumptions of Bernoulli channel losses. Interestingly, the same critical value was found to exist for systems transmitting state as well as observations in terms of boundedness of the expected covariance.

In [61] under both a Bernoulli channel model and a bursty error model, the loss of packets is mitigated by splitting up the information and transmitting it separately. This allows for a more gradual performance reduction in the presence of packet loss (effectively a form of forward error correction). Another form of mitigation using the results of [56] is investigated in [62]. By more finely sampling the process $\rho(F)$ is reduced and the critical arrival rate altered. It is claimed that the finer sampling allows the system to remain bounded however the access model used is different. The Bernoulli model used in [56] allows arbitrarily long sequences of missing observations to occur while the access model described in [62] is deterministic (using a “round robin” access model). The upper bound derived in eq. (11) is none other than an upper bound of a Riccati equation (see [63]) derived by assuming that the covariance
of the prior is infinite. The improvement in performance is due to the deterministic access model more than the change in eigenvalues of the process function.

The results of [55] are extended by [64] by considering a HMM channel model. This allows previous results (for Bernoulli channels) to be supported within the same framework. Although the paper only defines this in the context of a single sensor, notes at the end suggest how this can be extended to multiple sensors. [65] extends this work further and provides stability analysis for Kalman filtering for the two state Markov Chain (i.e. Gilbert-Elliot Channel Model) and provides bounds for the peak covariance.

The case of optimal fusion in the case of two sensors with an unknown packet dropping statistics is considered in [66]. An innovation like message is used to update local estimates. This seems to be the first occurrence where the message transmitted is not actually a measurement but rather some statistic derived from the observers state. As such it is more applicable to distributed state estimation than prior results. The message sent summarizes the difference between the previous global state and the current local state.

The state augmentation process was first introduced in the original papers by Kalman for the discrete time sampled system. The state augmentation approach replaced the original $n \times n$ state transition matrix with a larger $mn \times mn$ state transition matrix where the maximum delay was $m$ intervals. The disadvantage of this approach was the evaluation of the filter now required inversion of much larger state transition matrices. The case of the delay process being stochastic is covered in [67].

In [68] an optimal filter where the delays are known for both observation and process model is developed. The algorithm is an extension of the Kalman-Bucy filter for a continuous time process with aperiodic observations. However, the solution of the PDE produced is only tractable for the case of scalar systems. As the system is couched as a multiple sensor system with independent delays for each sensor it is applicable for centralized fusion models.

A multi-sensor system with delayed measurements was tackled, in a simplified and discretised sense in [69] and [70]. Using the approach outlined in these papers a way of re-formulating the $mn \times mn$ state augmented matrix into $m$ lots of $n \times n$ matrices is outlined, with considerable savings in inverting the state transition matrix. The formulation is applied to discrete time Kalman filters with Gaussian noise processes, however the standard non-linear extensions are applicable. As the results are based on either plant delays or sensor measurement delays, the results are applicable to centralized fusion.

A sub-optimal fusion method with arbitrarily delayed measurements is considered in [71]. It is sub-optimal in the sense that it requires conditional independence of all
2. Prior Work

observations. It is still a centralized approach, although amenable to restructuring to track fusion. A similar paper [72] deals only with discretised delays. It is suboptimal in the same sense as [71] because of the assumption of independence of observations. It is described as the fusion of separate estimators that may include delayed observations. This model is similar in nature to tracklet fusion of a distributed system.

Delays are explicitly noted in the development of [73]. A global estimate is built using the delayed data and updated using locally available data. The algorithm uses discrete time delays and communicates only the innovations (rather than state estimates or measurements).

An optimal centralized fusion is provided by [74] which is significantly more efficient. It relies on a re-ordering approach where instead of augmenting the state transition matrix, multiple reduced order estimators are fused, reducing the computational cost. The re-ordering approach used is based on the early paper [75] which introduced the re-ordering concept. Another similar approach to optimal fusion with delayed measurements is given in [76]. Other processes for distributed fusion of a networked system are in [77] and [78].

A common factor for all these previous papers is that the delay is treated as a known quantity. The convergence of the estimate is assured assuming the delay is bounded.

The stability for a single sensor system with random delays is investigated in [79]. Although it only assumes a single sensor / single delay model it shows asymptotic stability and in the case of a linear system, superposition can be used to show stability for a centralized fusion system with random delays. The case is limited to a dual state HMM with 0 or 1 unit delay. Some extension is provided by later work [80] which includes some additional stability criteria. A generalisation of this work is performed in [81] where a similar approach to [80] is used and extended to a bounded 0 - N unit delay where the delay is a random process. An interesting point (and one that is elaborated on in section 5 of this paper) is that the process of fusing randomly delayed measurements is very similar to the process of fusing measurements from multiple sensors - note that most difficulties arise from the covariance terms that link the variance of the estimates, and how to estimate the coupling.

In [82] a HMM is used to model both a discrete delayed network and approximate a continuously delayed network. The problem is formulated as an estimation problem where the network is transporting vector measurements and a processor performs estimation based on the measurements. In the case analysed it is assumed that the measurements are not specifically time stamped. An optimal method for weighting the measurements is developed so that the measurements may then be applied to a state estimator. The theory is extended to the continuous delay distributions by
using an interpolation of the discrete HMM states. A similar problem where the delay is unknown is presented in [83]. In this case the delay is drawn from some known distribution where at each processing point precisely one (possibly duplicated) observation is received.

A useful extension of the two state HMM model for network delay is provided by [84]. This allows the integration of indeterminate measurements, random delay and packet dropout to be integrated into a single framework. Generation of the optimal filter structure uses standard state augmentation processes after the inclusion of the HMM model in the sensor framework. It would appear that the combination of [84] and [75] would provide an efficient implementation and “nice” mathematical properties.

An important point to note is that both the augmented and re-ordered [74] approaches to fusing delayed data provide equivalent and optimal performance at the limit as time $t \to \infty$ as they provide an identical covariance matrix as the undelayed case. That is, for optimal fusion systems, designed to account for delayed data, the introduction of delay has no impact on the final converged covariance. Thus the only impact of delay is that for a causal system the introduction of delay will require that the system predict further forward to provide useful state estimates and consequently the estimate will have a greater covariance.

A final point to note is that the literature addresses impairments in isolation and a priori. The coupled nature of the impairment with the system state typically is ignored, with the justification that the impairments are caused by wireless link behaviour and as such are independent of the network state. The motivation is that impairments that are independent of network state are analytically quite tractable with useful results for simple wireless systems experiencing only Gaussian noise. However these models do not deal with any system that experiences congestion, or where channels can co-interfere (CDMA and OFDM-CDMA systems and all protocols that use a CSMA-CD style of media access e.g. 802.11). In both systems the level of impairment increases with network traffic and must be dealt with in a stochastic sense.

2.5 Control of Systems with Communications Impairments

Although not directly related to this thesis, the related area of feedback control systems with delays and multiple sensors were also examined. This corresponds to the blue feedback path in figure 2.1. While mainly considering questions of controllability there is sufficient mathematical overlap between the estimation and control problems to warrant investigation.

A related active area is the control literature where the control of unstable systems in the presence of bandwidth limited channels is of interest. Although the results are
2. Prior Work

not always directly applicable to the estimation problem, the techniques are sufficiently interchangeable to allow the use of the material. Early work focused on scalar systems with ideal channels \[85\]. Later work focuses on controllability of MIMO systems in the presence of bandwidth constraints such as \[86\], \[87\].

State estimation and controllability bounds for a uniform quantiser are discussed in \[88\]. This is extended to an optimal controller and encoder in \[89\]. A summary of prior results in control domain is provided by \[90\]. A wide ranging review of a number of plant and encoder models is provided with a useful summary of results for control of stochastic plants. See also \[91\] for a broad discussion on many of the techniques available for dealing with closed loop systems with delays. Of particular interest are the LTI system models. Stability of a closed loop control system subject to random network delays is discussed in \[92\]. A useful unification between uncertain observations, Bernoulli channels and delayed observations is provided by \[84\]. A process for optimal filter design based on minimization of the $H_2$-norm is provided based on the channel model. LQG of system with time stamped messages with Markov delay distribution is covered by \[93\]. Optimization of bandwidth allocation with respect to control system parameters is undertaken in \[94\].

2.6 Delay and Message Loss

The impact of independent message losses was described in \[56\]. The derivation is straightforward, although complicated by not using the matrix inversion lemma to simplify the algebraic Riccati equation prior to determining its bounds (there are a number of simple analytic methods to derive loose upper and lower bounds when in this form). The standard approach of determining an upper bound of the expected covariance matrix that satisfies the Lyapunov equation is then used to show the existence of an expected covariance upper bound (which is actually the same upper bound discussed in \[116\]). The lower bound is formed using a Linear Matrix Inequality which requires numerical methods to evaluate. Although \[56\] deals only with the case of a constant loss hazard, the concept is readily extended to a hazard model using a Markov chain.

Interestingly the model used in \[56\] is claimed as being different from the Jump Linear models of previous authors (particularly referring to \[107\] but additionally, implicitly, to the earlier work by Tugnait \[103\]). The actual approach given known states of a Jump Linear System has more in common with Chang and Athans in \[104\]. In the case of \[104\] an optimal estimator is built assuming that the jump states are known (although in \[104\] this is not explicitly stated).

Stability under delay is covered in \[67\] through a modification of the state - aug-
2. Prior Work

Although it is well known that an augmented state allows the construction of an optimal estimator given delayed observations, Mateev in [67] shows that under bounded delay distribution the estimator will have finite covariance under all conditions that do not involve the loss of messages provided that the undelayed messages would have yielded a bounded estimator covariance. This confirms the intuitive insight that suggests that the optimal recursive estimator of a linear system is no worse than the optimal batch estimator. A batch estimate using all measurements except the last $N$ (where $N$ is finite) is no different from the optimal recursive estimator using all (out of order) observations save those from the last $N$. The problem of estimators for randomly delayed observations degenerates to the problem of estimators with missing observations. As shown in [56] the covariance of the estimator will be bounded provided some minimum number of observations is received (depending on observability and system dynamics) so that the behaviour of a system with bounded delays (as covered in [67]) will converge on that of a system with losses, where the hazard model for losses is constant, except for the last $N$ observations.

With this insight in hand we can infer that the lower bound of the covariance of the state estimate for a system with bounded delays will be the same as the lower bound of missing observations where the probability of missing an observation is the same as the probability of an observation being delayed beyond the $N$ step delay window. Furthermore, this estimate is easily tightened by calculating the covariance of the sequence of $N$ observations, given some initial starting covariance and the probability of the observation having occurred. The probability of an observation occurring is simply the probability that the observation is delayed less than $N$ steps. This lower bound of this sequence describes the reduction in accuracy caused by delay, as opposed to merely the fixed hazard rate. The point to note is that even total loss of $N$ messages will not cause the covariance to grow to infinity (to become unbounded) so there is no way that delay alone can cause the estimator covariance to become unbounded, albeit that it certainly increases.

2.7 Network Models

While the system impairments outlined earlier describe the specific impact on an estimation system, the network models define methods of modelling types of real network impairments as statistical processes. This section investigates papers that define the types of network impairments present in fixed and wireless networks. Models that relate real, physical impairments into models that are suitable for inclusion in the system model are included. Techniques for analysis of these models are covered, in particular jump linear systems.
2. Prior Work

Efforts at mitigating congestion in wireless sensor networks are regularly reported. These mitigation efforts are generally targeted at improving the real-time behaviour of inter-node routing and feeding back congestion information through the network. In most cases the implementation is naive with respect to the application (it has no knowledge of the application) and attempts to mitigate wireless network “hot spots”.

Tactics for mitigating congestion in wireless networks are investigated in [95]. An interesting observation is that wireless networks suffer congestion similar to, and in many cases worse than, a wired network. This follows from congestion occurring both at the logical “queue” level as well as at the physical layers (caused by co-channel interference) and access layer. One approach to congestion mitigation is to reduce the sampling rate. This is investigated in [96]. This approach drops congested packets, i.e. shortens queues (although it is not couched in these terms) in order to relieve congestion. Congestion control is attempted in [97] using feedback without any knowledge of the application. Congestion is mitigated by routing around congested network sections. Network Calculus is used in [98] to determine a network congestion control algorithm for a particular application.

The Bernoulli Loss or the Binary Symmetric Channel is a popular form of channel model. It represents a constant hazard model where the deletion of packets is represented by a constant independent process. This model was used by [56] and [99] in the establishment of asymptotic behaviour KF under packet loss. This is extended to two sets of sensor in [55]. It has the advantage of extreme simplicity and represents message hazards that are constant and independent to the quantity or size of messages. This makes it suitable for modelling fixed size messages in a wireless system in the presence of Gaussian noise.

A more sophisticated model is the “Bursty” Hazard or Gilbert-Elliot Channel Model. This is a two state Markov model where one state represents correct transmission and the other represents errored transmission. The model is proposed in [100] and [101] and is used to represent the error burst found in fading wireless channels and is used as the channel model in [65].

An alternative source of congestion (beyond simple link capacity) is the occurrence of congestion at the media access control layer. The MAC layer is responsible for arbitrating the use of the shared transmission resource between multiple transmitters. In this scenario enqueuing does not cause the loss of messages due to buffer overrun. Collisions, where multiple sensors transmit simultaneously are the cause of message loss. This is modelled in [3] where multiple media access methods are investigated and their impact on a simple dynamic estimation system are studied. In this case the delays and packet hazards do not follow the properties developed in other papers, and are different from the distributions investigated in this report. However it does show
a strong dependence on link utilization, albeit different from that of a Markov queued system.

A general Markov model can be used to represent channel behaviour when coupled with a jump linear system. The general Markov model coupled provides a convenient framework for both message delay (through message queues), channel loss (Gilbert-Elliot or Bernoulli) and errored measurements in a single framework. It is used as part of the analysis in [64]. An earlier form is used in [102] where it is used to model false alarms with incorrect observations used.

Jump linear systems provide an important theoretical underpinning for the analysis of many of the models described. The optimal estimator for a jump linear system when the jump state is unknown is developed in [103]. This also provides a practical way of approximating the optimal estimator. [104] provides a method of estimating the state provided the jump state is known (which is not made clear in the discussion in the paper). An alternative derivation is provided in [105], [106] and [107] for the optimal linear minimum mean square estimator.

These and other papers investigate the general problem of congestion in wireless networks and attempt to resolve it at a protocol level. While this will generally mitigate congestion in the system the consequences of the various strategies developed will have on a tracking system are barely touched or ignored. The unanswered question is whether the timeliness of the data is more or less important than reliable delivery. Additionally these network models address the system behaviour at the protocol layer. They answer the question - given a (semi) reliable link between nodes, what is the “best” method to route traffic between nodes? The difficulty with this starting premise is that the establishment of the point to point links in a mobile radio environment is not a given, and the radio links are not in any way point to point. Better approaches are complex and more typically analysed for the case of cellular mobile networks.

2.8 Critical Results

The current literature includes most of the parts required to address the research goal. However the material is isolated and will require further work and investigation to pull together. The combination of a Markov channel model and a jump linear system with known jump state allows the description of a realistic multi-sensor system with delay and drop. By generating the channel model parameters using the system state (number of targets, probability of target detection, number of sensors) and the network capabilities (process rate) the network state can be determined (the distribution of the network utilization) and hence the Markov model of the network. Using jump-linear techniques the impact of network parameters on a tracking system incorporating an
optimal linear mean square estimator, optimal delayed measurement processing and centralised data fusion can be determined. Given recent results in rate distortion bounds for multi-sensor jointly coded systems it is possible to relate the link message rate to the link distortion for some classes of systems and resolve if lowering sensor resolution or dropping sensor messages is the ideal method of dealing with congestion under varying conditions.

The following papers are selected for their particular relevance to either encoding, transmission or network impairment modelling. Although the papers in the previous sections touch upon one or more of these subjects, the following papers have results that will be of particular utility in the research to be undertaken. As the focus of the research is to characterise the impact of the transmission network, and in particular the congested network, on the system performance, it is critical to have an effective mathematical model. The model needs to be capable of handling both loss, delay, information fusion, be optimal for the class of linear systems, and be sufficiently analytically tractable.

The encoding of a message is the process whereby a real quantity of order $\mathbb{R}^N$ is formatted for transmission down a finite capacity transmission media. The process of encoding introduces an error at the point of reception of the message, that is independent of any errors introduced by the transmission media itself. In general the encoding process is a compromise between the size of the reception error, the time taken to encode the message and the size of the message. The reception error from one or more sources will cause a system error in the calculated state estimates.

The transmission of a message is not a perfect process. The message may be lost or delayed by a random period of time. These losses and delays contribute to errors in the systems state estimation. The encoding system used introduces a relationship between the reliability of transmission and the system error, potentially causing errors to grow in an unbounded fashion. The transmission modelling section is required to generate a model that represents the impact that common network impairments (congestion in particular) have upon the system error. In particular the relationship between the system loading and system error, as coupled through the network congestion is to be investigated and sensitivity determined.

### 2.9 Encoding and Transmission

In order to accurately estimate the impact of bandwidth allocation on system performance, a rate-distortion function is required. As the model is of a multi-sensor system the use of joint encoding is assumed. The challenge is then to identify the performance bounds of a joint encoding system and characterize them.
2. Prior Work

There are two important sets of papers that will characterize the encoding and transmission system. Firstly, papers by Nair [18],[19] provide the useful insight that an optimal encoding system for a linear estimation system will encode the innovations, not the measurements or state estimates. This confirms the heuristic argument that only the changes in the system state are transmitted, rather than the state or measurements. The innovation contains only the “new” information and consequently is all that needs to be sent. The results of [18] indicate that an “innovation-like” sequence of transmission provides the optimal encoding of the both state estimate and observations. The sequence is not of the innovation, but rather a quantised function of the posterior distribution based on all observations (actually in [18] the prior of the next estimator is used - which has the same information as the posterior of the current estimator provided the system is linear - selecting the symbol based on minimizing the error of the next prior in the case of a non-linear system would improve accuracy) and the prior distribution based on all previous transmitted symbols. The important distinction is that an innovation is based on the prior distribution based on all previous observations, not previous transmitted symbols. The sub optimal method discussed in [18] encodes this difference using a linear quantiser, although this is clearly sub-optimal (the distribution of the transmitted quantity is more closely related to a normal distribution than a uniform distribution) but serves its purpose well as a didactic example.

The second important paper is [38]. This paper covers the case of the design of a joint encoder to quantise Gaussian observations from multiple sensors, and as a result provides a rate distortion function for this process. This can be combined with the observation made in [108] that although both observations and state estimates from multiple sensors have significant cross correlation, the innovation does not. It is stated that the innovation depends on the mutual information (in effect the control input into the process) that is common to all sensors, while the variance of the innovation is dependent only on the properties of the local sensor. These two observations not only suggest a simple yet surprisingly effective fusion technique using fusion of the innovations, but also a method for optimally encoding this information by transmitting the innovation estimation.

Although these techniques may not be practicable for a real world system, for the purposes of providing a performance estimation of an optimally designed system they will suffice to give a reasonable bound on the system performance.
2. Prior Work

2.10 Encoding of Innovations from a single sensor.

An important consideration is how does the transmission of an “innovation-like” sequence behave on a non-ideal link where distortion and loss occur?

Consider the linear Gaussian system

\[ X_{k+1} = FX_k + v_k \]  
\[ Y_k =HX_k + w_k \]  
\[ E(v_k) = 0 \] 
\[ E(w_k) = 0 \] 
\[ \text{var}(v_k) = Q \] 
\[ \text{var}(w_k) = R \]

The MMSE is given by the Kalman Filter

\[ \hat{X}_{k+1|k+1} = F \hat{X}_k + K_k \left( Y_{k+1} - HF \hat{X}_k \right) \] 
\[ P_{k+1} = FP_kF^t + Q - K_kH \left( FP_kF^t + Q \right) \]

where

\[ K_k = \left( FP_kF^t + Q \right) H' \left( H \left( FP_kF^t + Q \right) H' + R \right)^{-1} \]

The innovation

\[ I_{k+1} = K_k \left( Y_{k+1} - HF \hat{X}_k \right) \]

has the well known properties

\[ E \left( I_{k+1} \right) = 0 \]
\[ \text{var} \left( I_{k+1} \right) = K_k \left( H \left( FP_kF^t + Q \right) H' + R \right)^{-1} K_k' \]
\[ E \left( I_k I_{k+j} \right) = \delta \left( j \right) \text{var} \left\{ I_k \right\} \]

That is, a sequence of innovations is a memoryless, Gaussian source. From an information theoretic perspective it is obvious that all new information is encapsulated by the innovation in the case of the linear, Gaussian system and the optimal encoding of the innovation sequence should satisfy the same requirements as outlined in [18]. The optimal MSE encoder for a one dimension memoryless Gaussian source is well known and can be generated for an arbitrary distortion using the Lloyd-Max quantiser algorithm (see [9]). A sub-optimal, but analytically tractable quantiser for the \( m \) dimensional case is readily formed by considering each dimension individually, but generally the multi-dimensional case needs to be treated using Linde’s method (see [14]) to generate an optimal vector quantiser.
2. Prior Work

The rate distortion function for encoding a memoryless vector Gaussian source is well known. Let $D_S$ be a diagonal matrix with distortions $d_1, \ldots, d_m$ arranged as

$$D_S = \text{diag} \{d_1, \ldots, d_m\}$$

and $\lambda_i \{P\}$ is the $i$th eigenvalue of $P$, the $m \times m$ variance matrix of the estimator of $X$. The rate distortion function is

$$R_S(D_S) = \frac{1}{2} \sum_{i=1}^{m} \log_2 \left\{ \frac{\lambda_i \{P\}}{d_i} \right\}$$

$$= \frac{1}{2} \sum_{i=1}^{m} \log_2 \{\lambda_i \{P\}\} - \frac{1}{2} \log_2 \{d_i\}$$

$$= \frac{1}{2} \log_2 \left\{ \prod_{i=1}^{m} |\lambda_i \{P\}| \right\} - \frac{1}{2} \log_2 \left\{ \prod_{i=1}^{m} d_i \right\}$$

$$= \frac{1}{2} \log_2 \left\{ \frac{|P|}{|D_S|} \right\}$$

Borrowing from [18] the encoder will use the previous state as well as the current state. Rather than projecting the encoder state forward we will track the errors back so that focus is maintained on the innovation like sequence properties. Note that the next transmission must include compensation for the error’s introduced by the previous transmissions quantisation. This is achieved by projecting the previous error term forward and subtracting it from the next innovation to be sent. This forward error correction process allows the focus to remain on the innovation process rather than the complexities of the quantiser.

Let

$$I_{s_{1:k}} = \delta_k (s_k, s_{1:k-1})$$

(2.19)

be the receiver innovation based on the received symbols $s_{1:k}$.

The symbols $s_{1:k}$ are selected using the encoder function $E$ and

$$s_k = E_k (I_{k|k}, s_{1:k-1})$$

(2.20)

where $s_{1:k-1}$ is the set of symbols already received at the decoder, i.e. the shared state information between the encoder and decoder, and $I_{k|k}$ is the $k$th innovation. The difference between the innovation at the transmitter and the innovation recovered at the receiver is

$$e_k = I_{k|k} - I_{s_{1:k}}$$

(2.21)

The received innovation is different from the “true” innovation at the transmitter, due to the finite accuracy with which they can be represented. The encoding process, the selection of the symbol $s_k$, of necessity loses some information. The process of recovering the state estimate at the receiver can be thought of as an integration of

The rate distortion function for encoding a memoryless vector Gaussian source is well known. Let $D_S$ be a diagonal matrix with distortions $d_1, \ldots, d_m$ arranged as

$$D_S = \text{diag} \{d_1, \ldots, d_m\}$$

and $\lambda_i \{P\}$ is the $i$th eigenvalue of $P$, the $m \times m$ variance matrix of the estimator of $X$. The rate distortion function is

$$R_S(D_S) = \frac{1}{2} \sum_{i=1}^{m} \log_2 \left\{ \frac{\lambda_i \{P\}}{d_i} \right\}$$

$$= \frac{1}{2} \sum_{i=1}^{m} \log_2 \{\lambda_i \{P\}\} - \frac{1}{2} \log_2 \{d_i\}$$

$$= \frac{1}{2} \log_2 \left\{ \prod_{i=1}^{m} |\lambda_i \{P\}| \right\} - \frac{1}{2} \log_2 \left\{ \prod_{i=1}^{m} d_i \right\}$$

$$= \frac{1}{2} \log_2 \left\{ \frac{|P|}{|D_S|} \right\}$$

Borrowing from [18] the encoder will use the previous state as well as the current state. Rather than projecting the encoder state forward we will track the errors back so that focus is maintained on the innovation like sequence properties. Note that the next transmission must include compensation for the error’s introduced by the previous transmissions quantisation. This is achieved by projecting the previous error term forward and subtracting it from the next innovation to be sent. This forward error correction process allows the focus to remain on the innovation process rather than the complexities of the quantiser.

Let

$$I_{s_{1:k}} = \delta_k (s_k, s_{1:k-1})$$

(2.19)

be the receiver innovation based on the received symbols $s_{1:k}$.

The symbols $s_{1:k}$ are selected using the encoder function $E$ and

$$s_k = E_k (I_{k|k}, s_{1:k-1})$$

(2.20)

where $s_{1:k-1}$ is the set of symbols already received at the decoder, i.e. the shared state information between the encoder and decoder, and $I_{k|k}$ is the $k$th innovation. The difference between the innovation at the transmitter and the innovation recovered at the receiver is

$$e_k = I_{k|k} - I_{s_{1:k}}$$

(2.21)

The received innovation is different from the “true” innovation at the transmitter, due to the finite accuracy with which they can be represented. The encoding process, the selection of the symbol $s_k$, of necessity loses some information. The process of recovering the state estimate at the receiver can be thought of as an integration of
2. Prior Work

the information transmitted by all prior received symbols. At the receiver the state estimate based on all prior received symbols is

\[ \hat{X}_{s:k} = F \hat{X}_{s:k-1} + I_{s:k} \]  (2.22)

The difference between the state estimate at the receiver and the “true” state estimate at the transmitter is the cumulative error.

The cumulative error at the receiver is

\[ E_k = \hat{X}_{k|k} - \hat{X}_{s:k} \]  (2.23)

\[ = \sum_{k=1}^{N} F^{N-k} I_{k|k} - \sum_{k=1}^{N} F^{N-k} I_{s:k} \]  (2.24)

\[ = F^{N} \sum_{k=1}^{N} F^{-k} \epsilon_k \]  (2.25)

We can write the encoder and decoder in terms of a quantiser function and a "pre-processing" function.

Define the optimal quantiser as

\[ s_k = \Delta(x_k : p_k, S) \]  (2.26)

that generates a symbol from an alphabet \( S = \{s^1, \ldots, s^n\} \) for a random variable \( x_k \) that is drawn from a distribution \( p_k \).

The inverse quantiser is

\[ \Delta^{-1}(s_k : p_k, S) \]  (2.27)

and the error of the quantiser is

\[ \epsilon_k = \Delta^{-1}(\Delta(x_k : p_k, S) : p_k, S) - x_k \]  (2.28)

The design of the quantiser/inverse quantiser pair is such that \( \text{var}(\epsilon_k) = D_{\Sigma_k} \) is minimized. Define \( f : \mathbb{R}^n \times S^{k-1} \rightarrow \mathbb{R}^n \) as a mapping from the tuple of the current innovation and all prior symbols.

The cumulative error for arbitrary \( f \) at the \( k \)th transmission becomes

\[ E_k = \sum_{k=1}^{N} F^{N-k} \left( \hat{I}_{k|k} - \Delta^{-1}(\Delta(f(\hat{I}_{k|k}, s_{1:k-1}) : p_k, S) : p_k, S) \right) \]  (2.29)

\[ = \sum_{k=1}^{N} F^{N-k} \left( \hat{I}_{k|k} - f(\hat{I}_{k|k}, s_{1:k-1}) - \epsilon_k \right) \]  (2.30)

\[ = \sum_{k=1}^{N} F^{N-k} (\hat{I}_{k|k} - \epsilon_k) - \sum_{k=1}^{N} F^{N-k} \left( f(\hat{I}_{k|k}, s_{1:k-1}) \right) \]  (2.31)
We can minimize $E_k$ and ensure that it is always bounded by setting

$$f \left( \hat{\mathbf{I}}_{k|k}, s_{1:k-1} \right) = \mathbf{i}_{k|k} - F \epsilon_{k-1}$$

(2.32)

and recursively generating $\epsilon_k$

$$\epsilon_{k-1} = \Delta^{-1} \left( \Delta \left( f \left( \hat{\mathbf{I}}_{k-1|k-1}, s_{1:k-2} \right) : p_{k-1}, S \right) : p_{k-1}, S \right) - f \left( \hat{\mathbf{I}}_{k-1|k-1}, s_{1:k-2} \right)$$

(2.33)

$$= \Delta^{-1} \left( \Delta \left( \hat{\mathbf{I}}_{k-1|k-1} - F \epsilon_{k-2} : p_{k-1}, S \right) : p_{k-1}, S \right) - \hat{\mathbf{i}}_{k-1|k-1} + F \epsilon_{k-2}$$

(2.34)

(2.35)

(2.36)

Re-writing we can form a telescoping series

$$E_k = \sum_{k=1}^{N} F^{N-k} \left( \hat{\mathbf{I}}_{k|k} - \epsilon_k \right) - \sum_{k=1}^{N} F^{N-k} \left( f \left( \hat{\mathbf{I}}_{k|k}, s_{1:k-1} \right) \right)$$

(2.37)

$$= \sum_{k=1}^{N} F^{N-k} \left( \hat{\mathbf{I}}_{k|k} - \epsilon_k \right) - \sum_{k=1}^{N} F^{N-k} \left( \hat{\mathbf{i}}_{k|k} - F \epsilon_{k-1} \right)$$

(2.38)

$$= - \sum_{k=1}^{N} F^{N-k} \left( \epsilon_k \right) + \sum_{k=1}^{N} F^{N-k} \left( F \epsilon_{k-1} \right)$$

(2.39)

$$= - \sum_{k=1}^{N} F^{N-k} \left( \epsilon_k \right) + \sum_{j=0}^{N-1} F^{N-j} \left( \epsilon_j \right)$$

(2.40)

$$= -\epsilon_N + F^{N} \epsilon_0$$

(2.41)

The first error term ($\epsilon_0$) can be set to 0 by ensuring that both encoder and decoder are initialized with the same initial conditions. Note that the distribution of $\epsilon_k$ is unimportant, and $\epsilon_0$ is eliminated by fiat so that the distribution of the cumulative error is governed by $\epsilon_N$ only. Unfortunately little can be said about the distribution of $\epsilon_N$ other than it approaches a uniform distribution as $|S|$ increases and it has mean 0 and variance $D_{\Sigma_N}$ as given in the definition of the quantiser above (although see [109] for conditions under which more can be determined).

The optimal quantiser is found when the quantiser’s distribution is identical to the quantity being quantised - i.e. if

$$p_k = p \left( \mathbf{I}_k - F \epsilon_{k-1} | y_{1:k}, s_{1:k-1} \right)$$

(2.42)

but the inverse quantiser at the receiver can only be built using information available at the receiver - i.e.

$$p_k = p \left( \mathbf{I}_k - F \epsilon_{k-1} | s_{1:k-1} \right)$$

(2.43)

so the source statistics must be approximated using the information available at the receiver.
2. Prior Work

Although $p(I_k|s_{1:k-1})$ is Gaussian, $p(F_{\epsilon_{k-1}}|s_{1:k-1})$ has an unknown distribution that approaches a linear distribution, preventing the easy construction of the optimal quantiser. If the distribution is approximated using the first two moments then

$$p_k = \mathcal{N}(\hat{I}_{k|s_{k-1}}, P_{k|s_{k-1}} + F \text{ var } \{\epsilon_{k-1}\} F')$$

(2.44)

$$= \mathcal{N}(\hat{I}_{k|s_{k-1}}, P_{k|s_{k-1}} + FD_{\Sigma_{k-1}} F')$$

(2.45)

Thus the optimal encoder in the class of encoders using quantised Gaussians has the rate distortion function

$$R_k (D_{\Sigma_k}) = \frac{1}{2} \log_2 \left\{ \frac{|P_{k|s_{k-1}} + FD_{\Sigma_{k-1}} F'|}{|D_{\Sigma_k}|} \right\}$$

(2.46)

where $R_k$ is the number of bits required to transmit the $k$th innovation such that the state estimate will have a distortion with an RMS error of $D_{\Sigma_k}$.

The cases covered by ([18]) do not cover loss of messages. Consider the cases where messages are lost according to a constant hazard model. In this case we have

$$\Pr \{\gamma_k = 1\} = \alpha$$

(2.47)

$$s_k = \begin{cases} \mathcal{E}_k (\hat{I}_{k|k}, s_{1:k-1}) : \gamma_k = 1 \\ s_0 : \gamma_k = 0 \end{cases}$$

(2.48)

where the symbol $s_0$ represents the (null) transmission signifying $\hat{I}_{k|s_{1:k-1}}$ and $\gamma_{1:k}$ is the sequence of drop/no drop events.

At the receiver the cumulative error becomes

$$E_k = \sum_{k=1}^{N} F^{N-k} \left( \left( \hat{I}_{k|k} - \Delta^{-1} \left( \Delta \left( f \left( \hat{I}_{k|k}, s_{1:k-1} \right) : p_k, S \right) : p_k, S \right) \right) \gamma_k \right)$$

(2.49)

$$= \sum_{k=1}^{N} F^{N-k} \left( \left( \hat{I}_{k|k} - f \left( \hat{I}_{k|k}, s_{1:k-1} \right) - \epsilon_k \right) \gamma_k \right)$$

(2.50)

$$= \sum_{k=1}^{N} F^{N-k} \left( \hat{I}_{k|k} - \left( f \left( \hat{I}_{k|k}, s_{1:k-1} \right) \gamma_k + \epsilon_k \gamma_k + (1 - \gamma_k) \hat{I}_{k|s_{1:k-1}} \right) \right)$$

(2.51)
2. Prior Work

The expected cumulative error becomes

\[
E (E_k) = E \left( \sum_{k=1}^{N} F^{N-k} \left( \hat{i}_{k|k} - \left( f \left( \hat{i}_{k|k}, s_{1:k-1} \right) \gamma_k + \epsilon_k \gamma_k + (1 - \gamma_k) \hat{i}_{k|s_{1:k-1}} \right) \right) \right) \tag{2.52}
\]

\[
= \sum_{k=1}^{N} F^{N-k} \left( \hat{i}_{k|k} - \left( f \left( \hat{i}_{k|k}, s_{1:k-1} \right) \gamma_k + \epsilon_k \gamma_k + (1 - \gamma_k) \hat{i}_{k|s_{1:k-1}} \right) \right) \tag{2.53}
\]

\[
= \sum_{k=1}^{N} F^{N-k} \left( \hat{i}_{k|k} - \left( f \left( \hat{i}_{k|k}, s_{1:k-1} \right) \alpha + \epsilon_k \alpha + (1 - \alpha) \hat{i}_{k|s_{1:k-1}} \right) \right) \tag{2.54}
\]

\[
= \sum_{k=1}^{N} F^{N-k} \left( \hat{i}_{k|k} - \left( \alpha \epsilon_k + (1 - \pi) \hat{i}_{k|s_{1:k-1}} \right) \right) \tag{2.55}
\]

\[
- \sum_{k=1}^{N} F^{N-k} \left( f \left( \hat{i}_{k|k}, s_{1:k-1} \right) \right) \tag{2.56}
\]

If we set

\[
f \left( \hat{i}_{k|k}, s_{1:k-1} \right) = \hat{i}_{k|k} - F \epsilon_{k-1} \tag{2.58}
\]

as before then the expected error becomes

\[
E (E_k) = \sum_{k=1}^{N} F^{N-k} \left( \hat{i}_{k|k} - \left( \alpha \epsilon_k + (1 - \alpha) \hat{i}_{k|s_{1:k-1}} \right) \right) \tag{2.59}
\]

\[
- \sum_{k=1}^{N} F^{N-k} \left( \hat{i}_{k|k} - F \epsilon_{k-1} \right) \tag{2.60}
\]

\[
= - \sum_{k=1}^{N} F^{N-k} \left( \epsilon_k + (1 - \alpha) \left( \hat{i}_{k|s_{1:k-1}} - \epsilon_k \right) \right) \tag{2.61}
\]

\[
+ \sum_{k=1}^{N} F^{N-k} F \left( \epsilon_{k-1} \right) \tag{2.62}
\]

\[
= - \sum_{k=1}^{N} F^{N-k} \epsilon_k + \sum_{k=1}^{N} F^{N-k} F \left( \epsilon_{k-1} \right) \tag{2.63}
\]

\[
- \sum_{k=1}^{N} F^{N-k} \left( (1 - \alpha) \left( \hat{i}_{k|s_{1:k-1}} - \epsilon_k \right) \right) \tag{2.64}
\]

\[
= - \epsilon_N + F^N \left( \epsilon_0 \right) \tag{2.65}
\]

\[
- (1 - \alpha) \sum_{k=1}^{N} F^{N-k} \left( \hat{i}_{k|s_{1:k-1}} - \epsilon_k \right) \tag{2.66}
\]
Thus the optimal process in the absence of transmission errors causes unbounded errors in the event of a transmission error. Both the expected error and expected error covariance are unbounded as \( k \to \infty \) but converge if either \( \lambda_{\text{max}} \{(F)\} < 1 \) or \( \alpha = 1 \). I.e. if a manoeuvre takes place and the message is lost then the expected error grows without bound.

The hint on how to rectify this is that divergence will not occur if \( \lambda_{\text{max}} \{(F)\} < 1 \) irrespective of manoeuvring or drop rate. A well known property of the Kalman Filter is that it will still track even if the process model does not match the actual process, provided the process noise is “large enough” (see [110]). In this the transmission process can be modelled by using a plant model that has a reduced maximum eigenvalue and a greater process noise. This causes the estimates to have a greater covariance than if a more accurate plant model was used, requiring more bits to encode for the same distortion and a correspondingly higher bit rate. The heuristic explanation is that by transmitting more bits we are introducing some redundancy in the information which will allow the system to recover in the event of a lost message, at the expense of a higher bit rate to compensate for possible errors.

The important points then are that the encoding of innovations provide optimal compression in an error free link, but this comes at the cost of unbounded errors when any information is lost on the link through the loss of individual measurements.

### 2.11 Encoding a Group of Sensors

Joint encoding is an important optimization of a communications system. The encoding of a group of sensors is couched as a co-operative transmission problem in [111] and a number of practical problems relating to the implementation of such a system are discussed. One observation is that the multiplexing of the channels cannot be performed as an RF synchronous system, differing path links will prevent a modulation system that requires phase coherency. However the paper discusses a method using
binary phase shift keying whereby three transmitters can be combined to transmit a form of fused information.

An important result is provided in [38] and extended in [112], [113] and [114]. In [38] an exact relationship for the special case of optimal joint encoding of Gaussian sources under a squared error metric is provided. While the paper develops a full framework for arbitrary distortion errors and distributions, for the purposes of this problem the most significant section is in section IV: The quadratic Gaussian CEO problem.

Note that the paper considers only the scalar case. Any quantiser developed will only quantise each orthogonal term - which will make the quantiser optimal only if there is no off diagonal terms in the covariance matrix. As this is not so, the quantiser is optimal only in the class of scalar quantisers - there will be vector quantisers that are superior. This can be avoided if it is possible to diagonalise the covariance through some unitary transform that can be known a priori by the receiver. In the case of a linear Gaussian system this is possible, as the covariance matrices can be calculated off-line in advance. Even in the case of intermittent observations, all that is required is the sequence of detection/non-detection events, not the actual covariances.

Since the covariance matrix for a linear system is known (or at least can be determined) then an independent set of observations can be formed and the translated observations can be transmitted. A final important observation is that in the presence of Gaussian noise, the jointly encoded transmitted value is the MMSE of the sensors, i.e. the mean of the sensors.

For the optimal case of a sensor system in which all sensors observe the target at the same time, and have a unitary probability of detection, then the rate distortion relationship can be significantly simplified. Using the rate distortion relationship developed in [38] and simplifying we note that each sensor will have the same covariance, and consequently the rate from each sensor will be identical. Furthermore, as all sensors are equal, all sensors will participate in the observation process.

\[
    r_{\Sigma} (d) = \frac{1}{2} \log_2 \left( \frac{\sigma_x^2}{d} \left( \frac{n}{\sigma^2 \left( \frac{1}{a_L} - \frac{1}{d} \right)} \right)^n \right) \\
    = \frac{1}{2} \left( \log_2 \left( \frac{\sigma_x}{d} \right) + n \log_2 \left( \frac{n}{\sigma^2 \left( \frac{1}{a_L} - \frac{1}{d} \right)} \right) \right)
\]

where \( \sigma_x \) is the variance of the process while \( \sigma \) is the variance of the sensors, \( n \) is the number of sensors and \( d \) is the desired distortion.
The lower limit on achievable distortion is
\[ d_L = \frac{1}{\left(\frac{1}{\sigma^2_x} + \frac{n}{\sigma^2}\right)} \] (2.73)
which is the MMSE of \( x \) given all observations.

Writing the desired distortion in units of \( d_L \)
\[ r_x (kd_L) = \frac{1}{2} \left( \log_2 \left( \frac{\sigma^2_x}{kd_L} \right) + n \log_2 \left( \frac{n}{\sigma^2 \left( \frac{1}{d_L} - \frac{1}{kd_L} \right)} \right) \right) \] (2.74)
\[ = \frac{1}{2} \left( \log_2 \left( \frac{\sigma^2_x}{kd_L} \right) + n \left( \log_2 \left( \frac{nk}{k-1} \right) - \log_2 \left( \frac{\sigma^2}{d_L} \right) \right) \right) \] (2.75)

Note that the lower bound rate distortion required to encode the Gaussian source \( x \) with variance \( \sigma^2_x \) and distortion \( kd_L \) is
\[ r_x (kd_L) = \frac{1}{2} \log_2 \left( \frac{\sigma^2_x}{kd_L} \right) \] (2.76)
then the “overhead” of transmitting using multiple sensors as opposed to a single sensor that could make all the observations itself is
\[ r_x (kd_L) - r_x (kd_L) = n \left( \log_2 \left( \frac{nk}{k-1} \right) - \log_2 \left( \frac{\sigma^2}{d_L} \right) \right) \] (2.77)
which for \( \sigma^2_x \) or \( n \) large approaches
\[ r_x (kd_L) - r_x (kd_L) = n \left( \log_2 \left( \frac{k}{k-1} \right) \right) \] (2.78)
This can be interpreted as an overhead per sensor of \( \log_2 \left( \frac{k}{k-1} \right) \) bits per sensor to achieve within a factor of \( k \) of the MMSE of \( x \) when compared to a single encoder with access to all observations achieving the same distortion.

2.12 Joint Encoding Innovations

The multi-sensor rate distortion results can be used to generate a rate distortion relationship for a multi-sensor system. It is immediately obvious that a system that transmits multiple individual measurements can be enhanced to jointly encode and transmit the mean observation. But the transmission of the observation is sub-optimal in an error free network. Is it possible to transmit the expected innovations? Can the expected innovation be used to form a fused state estimate?

The answer can be derived by considering [108]. The paper is about generalized fusion architectures and as an aside it mentions an “alternative distributed fusion
2. Prior Work

architecture” in section 3.3. This architecture uses “linearly processed data” that provides some nice properties for the cross covariance between sensors. With the aid of rewriting expression 8 in [108] in slightly more friendly notation it becomes apparent that the “linearly processed data” is none other than the innovation of the sensor using only the local data. Using Theorem 1 of [108] and the ideas expounded in section 3 of [108] it is possible to write a BLUE fuser that uses the innovations from the sensors.

In the simplified case of all m sensors having equal measurement noise covariance R and observation process H let the local estimates of x be mean \( \hat{x}_j^j \) and variance \( P_j^j \) and be written as

\[
\hat{x}_j^j = \hat{x}_j^{k|k-1} + P_j^j H' R^{-1} \left( \hat{x}_j^j - H \hat{x}_j^{k|k-1} \right)
\]

(2.79)

\[
\left( P_j^j \right)^{-1} = \left( P_k^j \right)^{-1} + H' R^{-1} H
\]

(2.80)

Let the local innovation be written as

\[
I_j^j = \hat{x}_j^j - \hat{x}_j^{k|k-1}
\]

(2.81)

From [108] the central estimator then is

\[
\hat{x}_k^k = \hat{x}_k^{k|k-1} + K \begin{pmatrix} I_1^k \\ \vdots \\ I_m^k \\ \hat{x}_k^{k|k-1} \end{pmatrix}
\]

(2.82)

where

\[
H = \begin{bmatrix} P_1^{k|k} H' R^{-1} H \\ \vdots \\ P_m^{k|k} H' R^{-1} H \end{bmatrix}
\]

(2.83)

\[
K = P_k^m H' S^{-1}
\]

(2.84)

\[
S = H P_k^m H' + \text{diag} \left\{ P_1^{k|k} H' R^{-1} H P_1^{k|k}, \ldots, P_m^{k|k} H' R^{-1} H P_m^{k|k} \right\}
\]

(2.85)

\[
P_k^m = P_k^m - K S K'
\]

(2.86)

Note that \( P_k^m \) is the fused prior. For the simplified case where all sensors “see” the target at the same time then \( P_j^j = P_k^m \) for all \( j \in [1, \ldots, m] \) sensors and so Eq. (2.82)
can be simplified to

\[ S = (P_k|k H'R^{-1}H \otimes 1_m) \cdot (P_k|k-1 \otimes 1) \]  
\[ \cdot (P_k|k H'R^{-1}H \otimes 1_m)' \]  
\[ + \text{diag} \{ P_k|k H'R^{-1}HP_{k|k} \} \]

\[ = P_k|k H'R^{-1}H P_{k|k-1} H'R^{-1} H P_{k|k} \otimes 1_{m \times m} \]  
\[ + \text{diag} \{ P_k|k h'R^{-1}hP_{k|k} \} \]

\[ = \begin{bmatrix} S_a + S_b & S_b & \cdots & S_b \\ S_b & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & S_b \\ S_b & \cdots & S_b & S_a + S_b \end{bmatrix} \]  

(2.87)

(2.88)

(2.89)

(2.90)

(2.91)

(2.92)

where

\[ S_a = P_k|k H'R^{-1}H P_{k|k-1} H'R^{-1} H P_{k|k} \]  

(2.93)

\[ S_b = P_k|k H'R^{-1}H P_{k|k} \]  

(2.94)

are symmetric matrices and

\[ K = (P_{k|k-1} \otimes 1) \cdot (P_{k|k-1} (P_{k|k} h'R^{-1}h)' \otimes 1_m)' \]

\[ \cdot \begin{bmatrix} S_a + S_b & S_b & \cdots & S_b \\ S_b & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & S_b \\ S_b & \cdots & S_b & S_a + S_b \end{bmatrix}^{-1} \]

\[ = (P_{k|k-1} (P_{k|k} h'R^{-1}h)' \otimes 1_m)' \otimes 1_m \]

\[ \cdot \begin{bmatrix} S_1 & S_2 & \cdots & S_2 \\ S_2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & S_2 \\ S_2 & \cdots & S_2 & S_1 \end{bmatrix} \]

(2.95)

(2.96)

(2.97)

and

\[ S_1 = S_a^{-1} - S_a^{-1} (S_b^{-1} + mS_a^{-1})^{-1} S_a^{-1} \]  

(2.98)

\[ S_2 = -S_a^{-1} (S_b^{-1} + mS_a^{-1})^{-1} S_a^{-1} \]  

(2.99)
and (2.97) follows from the matrix inversion lemma. Substitution into (2.82) yields

\[ \hat{x}_{k\mid k} = \hat{x}_{k\mid k-1} + \left( P_{k\mid k-1} \left( P_{k\mid k} H' R^{-1} H \right)' \otimes I_m \right) \begin{bmatrix} S_1 & S_2 & \cdots & S_2 \\ S_2 & \ddots & \vdots & \vdots \\ \vdots & \ddots & S_2 \\ S_2 & \cdots & S_2 & S_1 \end{bmatrix} \]  

(2.100)

\[
\begin{pmatrix} I_k' \\ \vdots \\ I_k' \end{pmatrix} - H \hat{x}_{k\mid k-1} = \left( P_{k\mid k-1} \left( P_{k\mid k} H' R^{-1} H \right)' \left( S_a^{-1} - m S_a^{-1} (S_b^{-1} + m S_a^{-1})^{-1} S_a^{-1} \right) \right) \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_2 \\ S_1 \end{bmatrix} + \hat{x}_{k\mid k-1} \]

(2.101)

\[
\begin{pmatrix} \sum_{j=1}^{m} I_k' \\ \vdots \\ \sum_{j=1}^{m} I_k' \end{pmatrix} - K H \hat{x}_{k\mid k-1} = \hat{x}_{k\mid k-1} + \left( P_{k\mid k-1} \left( P_{k\mid k} H' R^{-1} H \right)' \left( S_a/m + S_b \right)^{-1} \right) \bar{I}_k - K H \hat{x}_{k\mid k-1} \]

Note that (2.104) depends on the known covariances (which are calculated ahead of time) and the sample mean of the innovations. It is noted that this is a highly optimised scenario, the requirement that all sensors have equal independent measurement noises and can observe the target at the same time is optimistic. But for the purpose of establishing the upper bound for the system performance this is still suitable. The measurement process involves the sample mean of the innovations making it possible to use the Rate Distortion bound in [38] to directly express the error in the innovation and hence the error in the state estimate as a function of the message size and message error rate. In this case the noise induced due to the quantisation of \( I_k \) can be directly accounted for as an additive noise term added to the variance of \( I_k \) so that \( R \) becomes \( R + \Delta \) where \( \Delta \) is the noise caused by quantisation.

### 2.13 Discussion

Between these three sets of papers it is apparent that for the errorless bit rate constrained transmission link, it is possible to design the optimal linear MMSE multisensor fusion system and specify its rate distortion bound. It will transmit an “innovation like” sequence from each sensor that is jointly encoded and fuse the jointly encoded transmissions together. A similar system is proposed in [115] in which fusion of innovations is proposed. The deficiency in strictly using innovations is identified in this paper (the “random walk”) and a local feedback structure (the “Reference Tracker”) is used to compensate for this. The problem of unbounded divergence in
the presence of transmission errors is also identified, although no solution beyond re-
transmission is proposed. Although it is not known yet what the optimal encoding is
in the presence of transmission errors, an approach of using a modified system function
that is not divergent would appear promising to allow a rate distortion function for
an arbitrary errored link.

The prior work identified above has focused on parts of the overall system. Either
delay, rate/distortion or message loss have been examined in isolation and optimal
estimators generated that will cope with these impairments. However little has been
done to correlate the sources of these impairments, the actual link and traffic statistics,
with the actual observed impairments. Furthermore, there is little to guide the
configuration of network systems when a trade-off between various network character-
istics is required. Finally there has been no work which considers the sensitivity of a
multi-sensor networked system to variations in network parameters.
Chapter 3

System Design Relationships

3.1 Overview

This chapter develops relationships between the probability of a sensor making observations and the performance of the sensor system, in particular the performance of the MMSE for linear systems. The boundedness of the estimator is considered, then the criteria for ensuring that the performance of the system as measured by its MMSE remain invariant for changing observation probabilities and number of sensors. Results are developed that allow the engineer to quickly establish approximate behaviour of the system as sensor numbers and sensor reliability change. Sensor scheduling is considered. Given identical numbers of sensors and sensor reliability when is the best time for a set of sensors to take observations? How sensitive to sensor scheduling is the performance of the system? Self interference is considered in a cluster of sensors where the sensor performance varies with the distance of the sensor from the target.

3.2 Introduction

This chapter will explore the relationships between system characteristics, with particular focus on the transmission network characteristics of capacity, loss and size and how these properties interact with the performance of the state estimator. It is desirable to be able to provide analytic results that allow heuristic insight into the relationships between parameters without requiring complete numerical solutions. While numerical solutions are necessary for the actual determination of system performance based on particular conditions, they do not provide the engineer with any insight into how the system properties interact - only that under a particular set of parameters and assumptions the expected performance is thus.

We will start this exploration by considering how to characterise the system and
its performance. Through this section we are primarily concerned with developing the relationships between the system performance, the number of sensors and the transmission network performance - in terms of the reliability of messages being transmitted. Of necessity we are concerned primarily with linear systems and homogenous sensors. However we can provide analytic results for the special case of homogenous sensors using linearised range/bearing sensors randomly scattered through 2D sensor environments.

It is important to carefully balance the transmission network to the sensors and system under observation. This allows the minimisation of power consumption. Sensors are frequently heavily constrained by the amount of energy available.

Power constraints influence:

- the life span of the sensors. Drawing more power of necessity reduces the duration for which the sensor operates

- the type of power storage. Different peak power requirements can be satisfied by different energy storage technologies. Sensors that have very low peak power requirements can be powered by small solar arrays or radio-isotope batteries that do not develop high power outputs, but have lifespans of decades.

- their mass and volume. Less power to be stored allows smaller devices taking up less volume.

- they ease with which they can be transported and deployed. Smaller devices are more resilient, withstanding greater G-force shocks and vibration

- they impact on thermal management. Smaller low power devices have reduced self heating which can make thermal management simpler. This is particularly true of systems that must operate in vacuum, where thermal management is extremely challenging.

- they influence robustness and reliability. Smaller, simpler low powered systems comprising of fewer parts can generally be made more reliable (or derated further) - provided they are not made too small.

While some over-subscription of transmission resources may improve the performance of a sensor system, too much utilisation may reduce the utility of the sensor system to such an extent that the increased sensor count is not justified. By providing a set of relationships that connect the sensor network with the sensor systems performance it is possible to more closely balance these requirements such that the overall systems performance is improved.
3. System Design Relationships

While it is certainly possible to design transmission networks that are effectively unconstrained in bandwidth with bit rates approaching $10^9$ bps, it is much harder to provide these capabilities when transmission power is heavily constrained due to requirements for:

- **Longevity.** Sensors that are for implantation, distributed in the environment without ability for maintenance, very small or mobile sensors all require that the system operate as long as possible with as little stored energy as possible in as light or small volume as possible. In this case every element in the sensor system is designed to minimise power consumption. As the fundamental limit of transmission of information is governed by Shannon’s limit, there is a direct relationship between capacity and the energy allocated per transmitted bit.

- **“Stealthiness”**. Networks that use more than the minimum amount of energy to transport information are of necessity less more easily detected and monitored than more efficient systems. In a range of applications the ability to operate unnoticed for extended periods of time is of paramount importance and every attempt is made to ensure that nodes in the system emit as little signal as possible.

- **Conformance.** In all cases systems must be designed to operate within the regulatory environment of the end user. In many cases the amount of energy that is allowed to be emitted across the spectrum is strictly curtailed.\(^1\) Systems operate at much lower bit rates and transmit power levels than is actually achievable in order to conform with these regulations. The ability to maximise the performance of a ubiquitous sensor network while still adhering to required power limits is essential for the technology to be useful.

These restrictions are most prominent for sensors designed to operate either with satellite systems or with long range HF systems. In both cases sensors distributed over large areas can interfere and consequently must share a common transmission media. HF data networks can easily have ranges that extend 1000’s of km, while sensors sharing the same satellite footprint can cover a significant portion of the global hemisphere. However, due to the nature of these systems, the spectrum is shared across all sensors. Satellites generally operate in “bent-pipe” mode, where all mobile transmitters are amplified in the satellite and re-transmitted to a single base station. All spectrum allocated to the satellite is simply retransmitted after frequency shifting.

\(^1\) see CISPR-22 / EN 55022 (applicable to European markets) or FCC part 15 (Code of Federal Regulations - 47CFR15.109 - applicable to US markets for limitations to the amount of electromagnetic energy devices may radiate into the spectrum without specific exemptions or licensing
by wide-band amplifiers for processing at the ground station. In addition the available bandwidth can be very small as frequency division multiplexing is all that is available on the satellite uplinks for each independent user.\(^2\) HF data links at carrier frequencies between 3-30MHz do not require any external equipment to achieve almost global reach, but due to the small amount of spectrum suited for this mode of operation at any time\(^3\) the channel bandwidth available to any individual system is less than 3000Hz\(^4\), even further reducing capacity. The current “state of the art” beyond line of site (BLOS) HF data system\(^5\) can achieve bit rates of approximately 4000bps over links of over 1000 nautical miles. Satellite based systems based on EHF frequencies such as JTIDS/Link-16 support higher bit rates, but still less then 128Kbps. Designing networks that support the required data rates over such large geographic areas is extremely challenging, and as the amount of information produced increases the complexity of these networks correspondingly increases.

The goals of this section are to provide relationships for the design of ubiquitous sensor networks where the sensors are reporting to a central estimator through some transmission system. It is intended that multiple sensors will be monitoring the same state variables and that the purpose of using multiple sensors is to provide improved performance over that of a single sensor. In order to achieve this large numbers of sensors will be used and the information fused to provide better accuracy than a single sensor would achieve. It is assumed that the number of sensors is large compared to the available bandwidth and we then examine the relationships between the network and the sensor system performance.

It should be noted as seen in the previous chapter that there are many possible methods that allow a collection of sensors to transmit data. Transmission can be agnostic to the application of the data, relying on network management, network flooding, tree pruning, cycles or other routing strategies to ensure that generic packets reach the required endpoints in a varying network. Alternatively the data can be incrementally fused forming a distributed estimator as discussed in the previous chapter, eliminating redundancy and distributing the fusion workload. The management of the network, monitoring the change in the structure and ensuring the algorithms

\(^2\)Although other methods are used to manage access at the base station - both TDMA and CDMA are used in JTIDS tactical data links for example - but the satellite does not control this, the base station performs this function.

\(^3\)Less than 10% of the bandwidth available is usable at any time due to changing ionosphere conditions caused by night/day cycles and solar activity. Furthermore the amount of electromagnetic interference caused by man-made activities restricts the usability of lower spectral bands.

\(^4\)Aeronautical and military standard channel spacing defined by ITU-R BS.597-1 recommends 5KHz spacing - with guards allows about 3kHz of available channel while ICAO mandates 3KHz spacing.

\(^5\)Link-22 - currently in ratification process as STANAG 5522 see Link-22 Guide Book 4th Edition - Chapter 1
perform as effectively as the centralised estimator is complex and it will be shown in later chapters that it is possible to achieve identical or better performance using a co-operative transmission process without requiring communication between sensor nodes.

3.3 The Expected Covariance

The upper bounds of the covariance of the state estimator of a linear system under intermittent observation are well known, and can be used to demonstrate that the estimator is bounded and exists. However there is a lack of useful lower bounds that usefully predict system performance. The Bayesian Information Matrix is often used to predict the lower bound of a system with intermittent observation, however this bound does not predict important properties of the system with any accuracy, or even correctly distinguish if the system is bounded. In this section a set of bounds are developed that are both tight to the actual covariance, but furthermore are of similar form, indicating that they are both necessary and sufficient to predict stability of the system under intermittent observation.

A discrete time linear system is defined through the recurrence relationship

\[ x_{k+1} = Fx_k + \sqrt{Q}u_k \]  

(3.1)

The equivalent continuous time system is defined as

\[ \dot{x} = Ax + \sqrt{\chi}u \]  

(3.2)

where \( u_k, x_k \in \mathbb{R}^m \) with \( u_k \sim N(0, I_m) \), \( A, F \in \mathbb{R}^{m \times m}, Q, \chi \in \mathbb{S}_{++}^{m \times m} \). \( A \) and \( F \) are related through

\[ F = e^{A\tau} \]  

(3.3)

and

\[ Q = \int_0^{\tau} e^{At} \chi e^{A't} dt \]  

(3.4)

\[ AQ + QA' = e^{A\tau} \chi e^{A'\tau} - \chi \]  

(3.5)

\[ \text{vec } Q = \left( e^{A\tau} \otimes e^{A\tau} - I \otimes I \right) (A \otimes I + I \otimes A)^{-1} \text{vec } \chi \]  

(3.6)

Defining \( Q \in \mathbb{S}_{++}^{m \times m} \) s.t.

\[ (A \otimes I + I \otimes A)^{-1} \text{vec } \chi = \text{vec } Q^{-1} \]  

(3.7)

we can write

\[ \text{vec } Q = \left( e^{A\tau} \otimes e^{A\tau} - I \otimes I \right) \text{vec } Q^{-1} \]
3. System Design Relationships

Observations are related to the system state through

\[ z_k = H x_k + \sqrt{R} v_k \]  

(3.8)

where \( v_k \in \mathbb{R}^l \) with \( v_k \sim N(0, I_l) \) and \( R \in \mathbb{R}^{l \times l} \). The discrete observation matrix \( O(F, H) \) is defined as

\[
O(F, H) = \begin{bmatrix}
H \\
HF \\
\vdots \\
HF^{r_{\text{observable}}}
\end{bmatrix}
\]  

(3.9)

with \( \min r_{\text{observable}} \) s.t. \( O(F, H) \) is full rank. The discrete control matrix \( O(F, Q) \) is defined as

\[
O(F, Q) = \begin{bmatrix}
Q & FQ & \ldots & F^{r_{\text{controllable}}} Q
\end{bmatrix}
\]  

(3.10)

with \( \min r_{\text{controllable}} \) s.t. \( O(F, Q) \) is full rank. Provided \( r_{\text{observable}} \) and \( r_{\text{controllable}} \) exist then (3.1) is observable and controllable and consequently an estimator of \( x_k \) can be constructed.

**Remark 1**  Note that if a continuous system is observed by different observation processes such that the two observation processes have identical observation matrices then the observation processes are similar and estimators constructed will be linearly related.

The Kalman Filter implements the MMSE of (3.1) and calculates the prior variance of the estimator of \( x_{k+1|k} \), \( P_{k+1|k} \) as the recursion of (this is easily obtained from the posterior form through expansion and application of the matrix inversion lemma)

\[
P_{k+1|k} = F \left( P_{k|k-1}^{-1} + H' R^{-1} H \right)^{-1} F' + Q
\]  

(3.11)

We can collect all terms that relate to the observation period together and write

\[
P_{k+1|k} = e^{A \tau} \left( P_{k|k-1}^{-1} + H' R^{-1} H \right)^{-1} + Q^{-1} e^{A \tau} - Q^{-1}
\]  

(3.12)

where \( Q \) is defined (3.7). We have separated out the observation period - it is clearer when written in vec form

\[
\text{vec} \ P_{k+1|k} = (e^{A} \otimes e^{A})^T \text{vec} \left( P_{k|k-1}^{-1} + H' R^{-1} H \right)^{-1} + Q^{-1} - \text{vec} \ Q^{-1}
\]  

(3.13)

Note that in general \( H' R^{-1} H \) is not invertible for a particular observation interval, but we know that there exists \( r_{\text{controllable}} \) s.t. the observation matrix is full rank. The consequence is that if we were to increase \( \tau \) by a factor \( r_{\text{controllable}} \) and make a new observation process that incorporates the \( r_{\text{controllable}} \) observations into a single
observation vector we can form a new equivalent system that has $H'R^{-1}H$ invertible. A dual conjunctive decomposition can be performed to solve

$$YY' = Q, YDY' = H'R^{-1}H$$

which combined with the substitution

$$P_{k+1|k} = Y'P_{k+1|k}Y, \quad F^\tau = Y'e^{A\tau}Y'^{-1} = e^{A\tau}$$

yields a simplified form

$$P_{k+1|k} = F^\tau \left( \left( P_{k|k-1}^{-1} + D \right)^{-1} + I \right) F^\tau' - I$$

where $D$ is a diagonal matrix containing the generalised eigenvalues of $(H'R^{-1}H)$ and $Q$ (i.e. the solutions of the linear matrix pencil $|H'R^{-1}H - \lambda Q^{-1}| = 0$). Also note that the eigenvalues of $F^\tau$ are the same as $e^{A\tau}$. Finally note that $P_{k+1|k}$ is within a linear transform of $P_{k+1|k}$ and that +ve (semi) definite $P_{k+1|k}$ implies +ve semi definite $P_{k+1|k}$. This relationship separates the features of the system into two distinct parts, the system, fully captured by $F$ and the observation (including the impact of the stochastic part of the system) is fully characterised by $D$. We term $P$ as the normalised prior covariance and from this point drop the bold and refer to $F$, $P$ and $A$ rather than $F$, $P$ and $A$. From this point on we will only consider systems in this normalised form, as all other forms may be linearly transformed into this form.

**Remark 2** Note that some eigenvectors of $q$ may be in its nullspace. i.e. if $A$ is not invertible then $q$ will not be invertible. However, as the system must be controllable and observable the transform will still exists although $D$ will not be invertible it will remain diagonal.

We can now include the effect of intermittent observations. Intermittent observations are introduced through the indicator variable $\gamma_k \in \{0, 1\}, E(\gamma) = \alpha$ with $\gamma = 1$ indicating an observation. Consequently we can define the Prior Covariance conditioned on $\gamma_{1:k-1}$ as

$$P_{k+1|k, \gamma_{1:k}} = F^\tau \left( \left( P_{k|k-1, \gamma_{1:k-1}}^{-1} + \gamma_k D \right)^{-1} + I \right) F^\tau' - I$$
and after taking expectations over $\gamma_k$ yields the recursion
\[
E \left( P_{k+1|k,\gamma_{1:k}} \right) = F^T \left( E \left( \left( P_{k|k-1,\gamma_{1:k-1}}^{-1} + \gamma_k D \right)^{-1} \right) + I \right) F^{T'} - I
\]  
(3.18)

\[
= F^T \left( E_{\gamma_{1:k-1}|\gamma_k=1} \left( \left( P_{k|k-1,\gamma_{1:k-1}}^{-1} + \gamma_k D \right)^{-1} \right) + I \right) F^{T'} - I
\]  
(3.19)

Note that the eigenvalues of $(X^{-1} + D)^{-1}$ are concave w.r.t. $X$ and hence by Jensen’s Inequality $E \left( (X^{-1} + D)^{-1} \right) < (E(X)^{-1} + D)^{-1}$ where $X \in S_m^{m \times m}$ and the inequalities are interpreted in the Loewner sense. Consequently we get the well known recursive upper bound of the expectation. We now write $E \left( P_{k+1|k,\gamma_{1:k}} \right)$ as $E \left( P_{k+1|k} \right)$, the expected prior covariance given all (possibly missed) observations.

\[
P_{k+1|k} = F^T \left( \alpha E \left( \left( P_{k|k-1}^{-1} + D \right)^{-1} \right) + (1 - \alpha) E \left( P_{k|k-1} + I \right) \right) F^{T'} - I
\]  
(3.21)

Note the slightly different presentation to the more common ways of writing this upper bound. All dependence on the continuous process noise is now incorporated explicitly and scaled by the observation method and incorporated in a single diagonal matrix $D$. We can use this to quickly derive convergence criteria. We note that
\[
\alpha \left( P_{k|k-1}^{-1} + D \right)^{-1} > 0
\]  
(3.22)

and consequently
\[
P_{k+1|k} > F^T \left( (1 - \alpha) P_{k|k-1} + I \right) F^{T'} - I
\]  
(3.23)

If we write a recursion
\[
P_{\text{LB},k} = F^T \left( (1 - \alpha) P_{\text{LB},k-1} + I \right) F^{T'} - I = (1 - \alpha) F^T P_{\text{LB},k-1} F^{T'} + (F^T F^{T'} - I)
\]  
(3.24)

then for all $P_{k|k-1} = P_{\text{LB},k-1} \Rightarrow P_{k+1|k} > P_{\text{LB},k}$. We know that (3.24) has a positive (semi)definite solution for $\lim_{k \to \infty} P_{\text{LB},k} = P_{\text{LB},k+1}$ if $\lambda_{\text{max}}(F^T) < \frac{1}{1-\alpha}$. Noting that this is a lower bound implies that the covariance is unbounded if
\[
\alpha < 1 - \frac{1}{\lambda_{\text{max}}(F^T)^2}
\]  
(3.25)

but not the converse and duplicates Sinopoli’s result in ([56]) with an alternative method.
Remark 3 The CRLB is given by the recurrence

\[ P_{k+1|k} > P_{CRLB} = F^\tau \left( \left( P_{CRLB}^{-1} + \alpha D \right)^{-1} + I \right)^{-1} F^{\tau^*} - I \]  \hfill (3.26)

and is not useful for determining the existence of a stable covariance under intermittent observations. Note that the recursion converges for all \( \alpha > 0 \), while (3.24) has stricter requirements for convergence.

A useful re-arrangement of (3.21) that will be frequently used to solve for \( \lim_{k \to \infty} P_k = P \). We substitute

\[ X_k = P_{k+1|k} + I \]  \hfill (3.27)

and write

\[ X - F^{-\tau} X F^{-\tau^*} = \alpha \left( (X - I) - \left( (X - I)^{-1} + D \right)^{-1} \right) \]  \hfill (3.28)

\[ G = \left( (X - I) - \left( (X - I)^{-1} + D \right)^{-1} \right) \]  \hfill (3.29)

Note that we have separated \( \alpha \) from \( \tau \) into a Lyapunov like relationship. The goal of this formulation is that it allows us to define a goal error \( G \). Note that \( G \) is positive (semi)definite for all \( X \) and \( D \) positive (semi) definite. The arrangement above is similar to that used for numerically solving (3.21) for a Newton solver. In this approach the recursion is

\[ G_k = \left( \left( (X_k - I)^{-1} + D \right)^{-1} - (X_k - I) \right) \]  \hfill (3.30)

\[ X_{k+1} = \text{dylap} \left( F^{-\tau}, \alpha G_k \right) \]  \hfill (3.31)

where \text{dylap}() solves the discrete Lyapunov function. The convergence is rapid, more so than simply repeated evaluations of the Riccati recursion. Furthermore it highlights that \( X_{k+1} \) is linear with respect to \( G_k \) and \( \alpha \).

There are two results missing from the review above. Firstly we will investigate the sufficiency (3.25), it is established that it is necessary, but it is to be established that it is also sufficient. Secondly we will investigate the performance of lower bounds of (3.18) and develop a lower bound tighter than the BIM.

Theorem 1 The MMSE of a controllable/observable linear system under independent intermittent observations is bounded iff

\[ \alpha > \alpha_N = 1 - \frac{1}{\lambda_{\text{max}}(F)^2} \]  \hfill (3.32)
Proof. This is a strengthening of Sinopoli’s result ([56]), where he observes that the MMSE is only bounded if (3.25) is satisfied. However he does not prove the sufficiency of this. Suppose we have an initial covariance estimate $P_0$, $\lambda_{\text{max}}(P_0) < \infty$. We then miss $m$ observations where $\Pr(m = M) = (1 - \alpha)^m$, $\alpha$ is the probability of seeing a single observation. Note then that $P_{m|0} < \sum_{j=0}^{m} F_j (P_0 + I) F_j^T$. We can write the expected covariance at the next successful observation, noting that if the expected covariance at the next successful observation is unbounded the system is unbounded. We note that

$$E(P_{k|0}) = E \left( \sum_{j=0}^{m} F_j (P_0) F_j^T \right) - I$$

(3.33)

$$\text{vec} E(P_{k|0}) = \text{vec} E \left( \sum_{j=0}^{m} F_j (P_0) F_j^T \right) - I$$

(3.34)

$$= \sum_{m=0}^{\infty} \Pr(m = M) \sum_{j=0}^{m} (F \otimes F)^j \text{vec} P_0 - \text{vec} I$$

(3.35)

$$= \sum_{m=0}^{\infty} \Pr(m = M) (I - (F \otimes F))^{-1} (I - (F \otimes F)^m) \text{vec} P_0$$

(3.36)

$$- \text{vec} I$$

(3.37)

$$= (I - (F \otimes F))^{-1} \sum_{m=0}^{\infty} (1 - \alpha)^m (I - (F \otimes F)^m) \text{vec} P_0$$

(3.38)

$$- \text{vec} I$$

(3.39)

$$= (I - (F \otimes F))^{-1} \sum_{m=0}^{\infty} (I(1 - \alpha)^m - ((1 - \alpha)(F \otimes F))) \text{vec} P_0$$

(3.40)

$$- \text{vec} I$$

(3.41)

$$= (I - (F \otimes F))^{-1} \left( \frac{I}{\alpha} - (I - ((1 - \alpha)(F \otimes F)))^{-1} \right) \text{vec} P_0$$

(3.42)

$$- \text{vec} I$$

(3.43)

$$= \left( I - (F \otimes F)^{-1} \right)^{-1} \frac{(1 - \alpha)}{\alpha} (I - ((1 - \alpha)(F \otimes F)))^{-1} \text{vec} P_0$$

(3.44)

$$- \text{vec} I$$

(3.45)

Note that a positive definite solution will only exist if $(I - ((1 - \alpha)(F \otimes F)))$ has only positive eigenvalues which is the same requirement as the lower bound. Finally we note that this is the same as (3.25), i.e. we have shown $\alpha > 1 - \frac{1}{\lambda_{\text{max}}(F)^2}$ is both necessary and sufficient for convergence as both lower and upper bound converge to positive (semi) definite values. □
3. System Design Relationships

The second point of interest is a lower bound of (3.18) that is tighter than the Baysian Information Matrix (BIM). The motivation is that the BIM does not correctly predict the existence of the bound under intermittent observations and provides an overly optimistic prediction of the covariance of the MMSE.

**Theorem 2** Given the BIM sequence

\[ C_{k+1} = F \left( \left( C_k^{-1} + \alpha \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \right)^{-1} + I \right) F' - I \]  \tag{3.46}

\[ \mathcal{I} \left( \pi_{k|k-1, \gamma_{1:k-1}} \right) = C_k^{-1} \]  \tag{3.47}

A tight lower bound for (3.18) is given by

\[ C_{k+1} < P_{k+1|k} \]

\[ = F \left( \alpha \left( C_k^{-1} + \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \right)^{-1} \right. \]

\[ + (1 - \alpha) P_{k|k-1} + I \left. \right) F' - I \]

\[ < P_{k+1|k} \]  \tag{3.51}

For \( C_0 = E \left( P_{0|0} \right) \)

**Proof.** Firstly we demonstrate that (3.48) is a lower bound of 3.18. From the BIM we know that

\[ P_{k+1|k} \geq F \left( \left( P_{k|k-1}^{-1} + \alpha \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \right)^{-1} + I \right) F' - I \]  \tag{3.52}

\[ \geq F \left( \left( P_{k|k-1}^{-1} + \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \right)^{-1} + I \right) F' - I \]  \tag{3.53}

As \( \mathcal{I} \left( \pi_{k|k-1, \gamma_{1:k-1}} \right)^{-1} < P_{k|k-1} \) then

\[ E_{\gamma_{1:k-1}} \left( \left( P_{k|k-1, \gamma_{1:k-1}}^{-1} + \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \right)^{-1} \right) < \left( P_{k|k-1}^{-1} + \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \right)^{-1} \]  \tag{3.54}

\[ \left( P_{k|k-1}^{-1} + \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \right)^{-1} > \left( C_k^{-1} + \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \right)^{-1} \]  \tag{3.55}
3. System Design Relationships

consequently

\[ P_{k+1|k} > F \left( E_k \left( \prod_{j=0}^{k} \left( D_{j} + \frac{1}{1 - \gamma_{k}} \frac{P_{j|k}}{P_{j+1|k}} \right) \right) \right) + I \]  \( F' - I \)  \( (3.56) \)

\[ = F \left( \prod_{j=0}^{k} \left( \frac{D_{j}}{1 - \gamma_{k}} \frac{P_{j|k}}{P_{j+1|k}} \right)^{-1} \right) + I \]  \( F' - I \)  \( (3.57) \)

which leads to the recursive lower bound

\[ P_{k+1|k} = F \left( \prod_{j=0}^{k} \left( \frac{D_{j}}{1 - \gamma_{k}} \frac{P_{j|k}}{P_{j+1|k}} \right)^{-1} \right) + (1 - \alpha) P_{k|k-1} + I \]  \( F' - I \)  \( (3.58) \)

Next we demonstrate that \( C_{k+1} < P_{k+1|k} \). Assume \( C_{k} = P_{k|k-1} \), then we require

\[ \left( C_{k}^{-1} + \alpha \left[ \frac{D}{0} \right] \right)^{-1} \leq \left( \alpha \left( C_{k}^{-1} + \left[ \frac{D}{0} \right] \right)^{-1} + (1 - \alpha) C_{k} \right) \]  \( (3.59) \)

which after some re-arrangement yields

\[ \left[ \sqrt{D} \ 0 \right] C_{k} \left[ \sqrt{D} \ 0 \right] \leq \frac{1}{\alpha} \left[ \sqrt{D} \ 0 \right] C_{k} \left[ \sqrt{D} \ 0 \right] \]  \( (3.61) \)

which is true for \( \alpha < 1 \). Consequently, for all \( C_{0} = P_{0|0} \Rightarrow C_{k} < P_{k|k-1} \) and the bound is tighter than the BIM.  \( \blacksquare \)

There are two points to observe. Firstly, when looking for a simple design tool to estimate the acceptable loss of sensor reports for a system, the system performance is most closely associated with Lyapunov equations of the form

\[ P_{k} = (1 - \alpha) F \left( P_{k-1} + \frac{\alpha}{(1 - \gamma_{k})} \frac{P_{k-1}}{P_{k+1}} \right) F' + F F' - I \]  \( (3.62) \)

Furthermore the performance at the limit as \( \alpha \rightarrow 0 \) is most closely associated with the eigenvalues of the system characteristics \( \lambda(\tilde{F}) \), and almost independent of the sensor parameters \( (H, R) \) or motion \( (Q) \). The second observation is that both lower and upper bounds have identical form, to within a small correction term \( \left( P_{k+1}^{-1} + D \right)^{-1} \) which is a concave function. This implies that approximation as a Lyapunov function will be extremely close as \( \frac{\alpha}{1 - \alpha} \implies 0 \). Alternatively we can note that if we wish to
design for a particular covariance, the correction term \( K_k (P_{k-1}) \) is independent of \( \alpha \) and \( F \) and depends solely on the relationship between the quality of the sensor and the randomness of the motion. Note the similarity between (3.28) and (3.62), some re-arrangement yields

\[
(P_{k-1} + I) - F^{-1} (P_k + I) F^{-1} = \alpha \left( P_{k-1} - (P_{k-1}^{-1} + D)^{-1} \right)
\]

(3.63)

\[
= \alpha K_k (P_{k-1})
\]

(3.64)

The most important outcomes of this section are that we can summarise the observation process through the use of a single diagonal matrix, the sensor utility \( D \), which encapsulates the performance of the observer relative to the amount of random process noise. The final expected covariance as calculated through the recursion (3.30) is within a linear transform of the actual covariance. This property allows us to more easily focus on the behaviour of the (3.62) for small perturbations of \( \alpha \) for \( \lim_{k \to \infty} P_k \).

We will take advantages of these relationships in the following sections.

### 3.4 Sensor Scheduling

We can use the previous results to investigate the behaviour of multi-sensor systems under differing scheduling regimes. We will consider three types of multi-sensor systems

- For system \( S_1 \) all observations are scheduled such that they take place simultaneously.
- For system \( S_2 \) all observations are scheduled such that they are evenly spaced out in time.
- For system \( S_3 \) there is no scheduling, observations are independent random events and the spacing between observations is exponentially distributed.

\( S_1 \) represents a centrally controlled system and is typical of many simpler sensor networks where all sensors are slaved together. All sensors take observations at the same time with the same time base and the fusion of the results after allocation of measurements can be performed directly. Note that the measurements are maximally correlated as other than the errors introduced by the observation process they are all observing precisely the same state, consequently there is much redundant information present in each observation.

\( S_2 \) represents a centrally scheduled system, similar to \( S_1 \) but in contrast the measurements are minimally correlated. Each observation contains some information
about the change in state, consequently knowledge of the previous observation does not imply as much information about the next observation.

$S_3$ represents a loose federation of sensors where observation and transportation of observations proceeds in a laissez-faire fashion without centralised co-ordination. Fusion of measurements requires the use of non-fixed timebases and consequently is somewhat more involved.

Define $n$ as the number of identical sensors. In $S_1$ the sensors take all $n$ measurements at the same time, in system $S_2$ the sensors take identically spaced independent measurements, while in $S_3$ they take them at random intervals. We can write the system equations (after suitable normalisation) as

$$
S_1(n) \begin{cases}
 x_{k+1} = Fx_k + u_k \\
 z_{i,k} = Hx_k + \frac{1}{\gamma_{i,k}}\sqrt{D^{-1}}v_k \\
 \mathbb{E}_{\gamma_{1,n,k}} (P(S_1)_{k+1|k}) = \\
 \mathbb{E}_{\gamma_{1,n,k}} \left( F \left( \left( P(S_1)_{k|k-1,\gamma_{i,k},k-1} + \sum_{i=1}^{n} \gamma_{i,k}D \right)^{-1} + I \right) F' - I \right)
\end{cases}
$$

(3.65)

$$
S_2(n) \begin{cases}
 x_{k+1} = F^{1/n}x_k + u_k \\
 z_{i,k} = Hx_k + \frac{1}{\gamma_{i,k}}\sqrt{D^{-1}}v_k \\
 \mathbb{E}_{\gamma_{1,n,k}} (P(S_2)_{k+1|k}) = \\
 \mathbb{E}_{\gamma_{1,n,k}} \left( F^{1/n} \left( \left( P(S_2)_{k|k-1,\gamma_{i,k},k-1} + \gamma_{i,k}D \right)^{-1} + I \right) F^{1/n} - I \right)
\end{cases}
$$

(3.66)

$$
S_3(n) \begin{cases}
 x_{k+1} = F^{\tau_k}x_k + u_k \\
 z_{i,k} = Hx_k + \sqrt{D^{-1}}v_k \\
 \mathbb{E}_{\gamma_{1,n,k}} (P(S_3)_{k+1|k}) = \\
 \mathbb{E}_{\gamma_{1,n,k}} \left( F^{\tau_k} \left( \left( P(S_3)_{k|k-1,\gamma_{i,k},k-1} + D \right)^{-1} + I \right) F^{\tau_k} - I \right)
\end{cases}
$$

(3.67)

Pr ($\tau_k = t$) = exp ($-\lambda t$)

The first point to observe is that all three systems have very closely related criteria for convergence. From Theorem 4 after a little re-arranging we have $1 - \lambda_{\text{max}}(F)^{-2/n} < \alpha \rightarrow \lim_{k \rightarrow \infty} \mathbb{E}_{\gamma_{1,n,k}} (P(S_2)_{k+1|k}) < \infty$. Additionally we see that for $S_1$ we are only concerned with the case where there are no observations, i.e. $\prod_{i=1}^{n} \gamma_{i,k} = 0$. 
Pr \left( \prod_{i=1}^{n} \gamma_{i,k} = 0 \right) = (1 - \alpha)^n. \text{ Again, after a little bit of arranging we have}

\begin{equation}
1 - \lambda_{\text{max}} (F)^{-2/n} < \alpha \rightarrow \lim_{k \to \infty} E_{\gamma_{1:n,1:k}} \left( P(S_1)_{k+1|k} \right) < \infty \tag{3.68}
\end{equation}

the same as \( S_2 \). The above discussion relates to systems where observations are scheduled in a structured fashion, either all at once, or equispaced. We can extend this to where sensors randomly make measurements such that the time between measurements is exponentially distributed. The proof for \( S_3 \) is somewhat more involved and is attached below (Theorem 5). In summary, we can write the system in vec form and write a bound involving \( E((F \otimes F)^{\gamma_k}) \). Note that for \( F \) normal (i.e. it has no geometrically repeated eigenvalues) that the eigenvalues will be distributed according to a Pareto distribution and that the mean can be directly written. This yields the convergence constraint of \( \frac{1}{2} > \lambda_{\text{max}} (\ln F) \) or \( e^\lambda > \lambda_{\text{max}} (F)^2 \). Noting that the exponential distribution is the limiting case of the geometric distribution and writing \( \alpha = 1 - \exp^{-\lambda} \) it is no surprise that as we let \( n \to \infty \) that the two criteria correlate as it is the equivalent of having a large number of unreliable sensors. Effectively we have seen that \( S_1 \) and \( S_2 \) are the limiting cases of a more general system \( S_3 \).

Given that all three systems have equivalent convergence criteria is there any single sensor scheduling configuration that is superior? We will just consider the two limiting case systems, \( S_1 \) and \( S_2 \). The first difficulty is to define what is meant by best. Clearly the prior covariance of \( S_2 \) will always be smaller than that of \( S_1 \) for the same posterior due to the shorter interval. Similarly for the same prior covariance the posterior covariance of \( S_1 \) will be smaller due to the larger number of observations taking place. Clearly, evaluating only prior or posterior covariance will favour one system over the other. The arguably best method of evaluation is to consider when the estimates are required. If state estimates are only required at fixed times it seems sensible to take as many observations as possible at that time to ensure the best possible estimate and minimise the posterior covariance, even if that means that potentially the prior covariance is quite large. Alternatively, if the requirements for estimates are that they are required at unknown times it makes sense to spread the observations out, such that though the posterior covariance is never as small, the prior covariance never becomes too large. Another intuition is that if a system has large random inputs and small observation error then making multiple observations at the same time has reduced utility, there is little additional information gained, and at the limit where the observation is perfect all the observations will be identical and redundant.

We continue by considering a generalised arrangement that covers \( S_1 \) and \( S_2 \). In this arrangement we consider a hybrid of systems \( S_1 \) and \( S_2 \) such that from a set of \( n^d \) sensors we allow a subset of \( n \) sensors to report at the same time. The limits are
covered by \( n = n^\dagger \) and \( n = 1 \). From (3.16) we know that the relationship is governed by

\[
F = e^{A/n^\dagger}
\]

\[
P_{k+1|k} = F^n \left( P_{k|k} + I \right) F^{n^u} - I
\]

\[
P_{k+1|k+1} = \left( P_{k+1|k} + \sum_{j=1}^{n} \gamma_j D \right)^{-1}
\]

\[(3.69)\]

\[(3.70)\]

\[(3.71)\]

\[n^\dagger = \text{total number of sensors}\]

\[n = \text{number reporting at once}\]

\[A = \ln F^{1/n^\dagger}\]

which we can re-write the posterior and prior as

\[
P_{k+1|k+1} = \left( \sum_{j=1}^{n} \gamma_j D \right) + \left( F^n \left( P_{k|k} + I \right) F^{n^u} - I \right)^{-1}
\]

\[
P_{k+1|k} = F^n \left( P_{k|k-1} - 1 + \sum_{j=1}^{n} \gamma_j D \right)^{-1} + I \right) F^n - I
\]

\[(3.75)\]

\[(3.76)\]

Consider the prior covariance, this is the largest uncertainty that can occur and consequently the worse case scenario for the system.

Define

\[
P_{k+1|k,n} = F^n \left( P_{k|k-1}^{-1} + nD \right)^{-1} + I \right) F^n - I
\]

\[(3.77)\]

being the covariance after observing with \( n \) sensors at once. We wish to know if

\[
P_{k+1|k,n} > P_{k+1|k,1}
\]

\[(3.78)\]

for all \( P_{k|k-1} \). Note that for all \( X \) +ve definite that

\[
(X^{-1} + nD)^{-1} > \frac{1}{n} (X^{-1} + D)^{-1}
\]

\[(3.79)\]

and so

\[
P_{k+1|k,n} = F^n \left( P_{k|k-1}^{-1} + nD \right)^{-1} + I \right) F^n - I
\]

\[
> F^n \left( \frac{1}{n} P_{k|k-1}^{-1} + D \right)^{-1} + I \right) F^n - I
\]

\[(3.80)\]

\[(3.81)\]
A sufficient condition for \( P_{k+1|k,n^1} > P_{k+1|k,1} \) to hold is
\[
F^{n^1} \left( \frac{1}{n^1} \left( P_{k|k-1}^{-1} + D \right)^{-1} + I \right) F^{n^1} - I > F \left( \left( P_{k|k-1}^{-1} + D \right)^{-1} + I \right) F' - I \quad (3.82)
\]
\[
F^{n^1-1} \left( \frac{1}{n^1} \left( P_{k|k-1}^{-1} + D \right)^{-1} + I \right) F^{n^1-1} > \left( P_{k|k-1}^{-1} + D \right)^{-1} + I \quad (3.83)
\]
\[
F^{n^1-1} \left( \frac{1}{n^1} \left( P_{k|k-1}^{-1} + D \right)^{-1} + I \right) F^{n^1-1} > \left( P_{k|k-1}^{-1} + D \right)^{-1} + I \quad (3.84)
\]
\[
F^{n^1-1} \left( \frac{1}{n^1} \left( P_{k|k-1}^{-1} + D \right)^{-1} + I \right) F^{n^1-1} > \left( P_{k|k-1}^{-1} + D \right)^{-1} + I \quad (3.85)
\]
for all \( \left( P_{k|k-1}^{-1} + D \right)^{-1} \) which is true for
\[
F^{n-1} \left( \frac{1}{n} \left( P_{k|k-1}^{-1} + D \right)^{-1} + I \right) F^{n-1} > \left( P_{k|k-1}^{-1} + D \right)^{-1} + I \quad (3.86)
\]
\[
\frac{F^{n-1}}{\sqrt{n}} \left( \frac{1}{\sqrt{n}} \left( P_{k|k-1}^{-1} + D \right)^{-1} + I \right) F^{n-1} \frac{1}{\sqrt{n}} \left( \left( P_{k|k-1}^{-1} + D \right)^{-1} + I \right) > \frac{1}{\sqrt{n}} F^{n-1} \quad (3.87)
\]
Note that
\[
\frac{F^{n^1-1}}{\sqrt{n^1}} Z \frac{F^{n^1-1}}{\sqrt{n^1}} - Z = I - F^{n^1-1} F^{n^1-1} \quad (3.88)
\]
is of the form of a discrete time Lyapunov equation, and has a unique solution provided \( \lambda_{\text{max}}(F^{n-1}) < 1 \), the solution is +ve definite if \( \lambda_{\text{min}}(F^{n-1}) > 1 \). \( Z \) positive definite implies that there exists some \( \left( P_{k|k-1}^{-1} + D \right)^{-1} < Z \) s.t. (3.87) is not satisfied. Conversely if \( \lambda_{\text{max}}(F^{n^1-1}) > 1 \) and there is a unique solution, then that solution must be -ve (semi)definite which means there can be no positive semidefinite \( X \) s.t. (3.87) is satisfied. This implies that for \( \lambda_{\text{max}}(F^{n^1-1}) > 1 \) the inequality is always satisfied and that reading sensors consecutively is going to have a smaller covariance than that of reading them concurrently where there are no other considerations. Consequently for \( \frac{\lambda_{\text{max}}(F^{n^1-1})}{\sqrt{n^1}} > 1 \) then (3.87) is satisfied and by applying (3.82) demonstrates that \( P_{k+1|k,n^1} > P_{k+1|k,1} \).

It would appear that a generally “safe” mechanism would evenly space the observations, as this will give the least variation in state estimate variance as well as evenly distributing the utilisation of the communications network, there being no advantage in buffering up old observations for a transmission slot. Where the process has eigenvalues that are close to or identically one it appears that taking all measurements at once is superior when there is no other system impact. However note that when \( \lambda_{\text{max}}(F) e^{1/2} \) then there exists no solution for \( \frac{\lambda_{\text{max}}(F^{n-1})}{\sqrt{n}} < 1 \) where \( n > 1 \) and \( n \) is an integer. Thus for all cases with multiple sensors and processes with eigenvalues \( e^{1/2} \) it is best to take measurements from multiple sensors sequentially rather than all at the same time.
3. System Design Relationships

3.5 System Relationships and Invariant Properties

In the design process it is useful to be able to have an analytic relationship between system parameters that helps provide an intuitive understanding of the relationship between the physical characteristics of the system. This is particularly true when attempting to design a system to achieve performance goals. We can use the relationships between system components in order to understand the relationship between performance, reliability of observation and number of sensors. Assume we have a system with known performance such that

\[
(P_{n,\alpha} + I) - F^{-1/n} (P_{n,\alpha} + I) F^{-1/n} = \alpha (P_{n,\alpha} - (P_{n,\alpha}^{-1} + D)^{-1})
\]  

(3.89)

\[
= \alpha K (P_{n,\alpha})
\]  

(3.90)

for some combination of sensors \(n\) and observation reliability \(\alpha\). We would like to know how variations in the number of sensors and reliability of the sensors impacts the system performance, in particular we would like to know how to vary those parameters in order to achieve a particular performance. Consider the scenario where the figure of merit of the system is maximum mean squared error of an estimated state variable. We would like the system’s performance to remain bounded for differing performance requirements, sensor count and sensor reliability. In many cases we want to know if the system has some “reserve capacity”. Will its performance degrade significantly if either the number of sensors or the reliability of sensors changes? What number of sensors would provide equivalent performance if the sensor availability changed? Notice that the composition of the equation pulls all parameters affected by the number of sensors to the LHS while the sensor availability \(\alpha\) is only on the RHS.

Our interest in this relationship is for scenarios where we wish to know the degradation in performance given small changes in \(\alpha\) and \(n\)

\[
(P_{n,\alpha} + I) - F^{-1/n} (P_{n,\alpha} + I) F^{-1/n} = \alpha K (P_{n,\alpha})
\]  

(3.92)

Note that for \(P_{n,\alpha}\) to be a solution of (3.92) that for a given \(n\)

\[
\left( (P_{n,\alpha} + I) - F^{-1/n} (P_{n,\alpha} + I) F^{-1/n} \right) K (P_{n,\alpha})^{-1} = \Omega_n
\]  

(3.93)

the value of \(\alpha\) is constrained \(\lambda_{\min} (\Omega_n) < \alpha < \lambda_{\max} (\Omega_n)\).

Taking derivatives w.r.t. \(\alpha\)

\[
\frac{dP_{n,\alpha}}{d\alpha} - F^{-1/n} \left( \frac{dP_{n,\alpha}}{d\alpha} \right) F^{-1/n} = K (P_{n,\alpha}) + \alpha \frac{dK (P_{n,\alpha})}{d\alpha}
\]  

(3.94)
3. System Design Relationships

and expanding $\mathbf{K}(P_{n,\alpha})$

$$\frac{d\mathbf{K}(P_{n,\alpha})}{d\alpha} = \left( \frac{dP_{n,\alpha}}{d\alpha} - (I + P_{n,\alpha}D)^{-1} \frac{dP_{n,\alpha}}{d\alpha} (I + P_{n,\alpha}D)^{-1} \right)$$

(3.95)

yields

$$\left( \frac{dP_{n,\alpha}}{d\alpha} \right) - F^{-1/n} \left( \frac{dP_{n,\alpha}}{d\alpha} \right) F^{-1/n}$$

$$= \mathbf{K}(P_{n,\alpha}) + \alpha \left( \frac{dP_{n,\alpha}}{d\alpha} - (I + P_{n,\alpha}D)^{-1} \frac{dP_{n,\alpha}}{d\alpha} (I + DP_{n,\alpha})^{-1} \right)$$

(3.96)

(3.97)

taking vecs and re-arranging

$$\text{vec} \left( \frac{dP_{n,\alpha}}{d\alpha} \right) = \left( (1 - \alpha) I - \left( e^{-\frac{\alpha}{n}} \otimes e^{-\frac{\alpha}{n}} \right) \right)^{-1} \text{vec} \mathbf{K}(P_{n,\alpha})$$

(3.98)

which indicates that to a first order, as $P_{n,\alpha}$ becomes “large” we have

$$\text{vec} \left( \frac{dP_{n,\alpha}}{d\alpha} \right) \approx \left( (1 - \alpha) I - \left( e^{-\frac{\alpha}{n}} \otimes e^{-\frac{\alpha}{n}} \right) \right)^{-1} \text{vec} \mathbf{K}(P_{n,\alpha})$$

(3.99)

We define the sensitivity of $P_{n,\alpha}$ to variations in $\alpha$ as

$$S_{\alpha}(P_{n,\alpha}) = \text{tr} \left( \frac{\alpha}{P_{n,\alpha}} \frac{dP_{n,\alpha}}{d\alpha} \right)$$

(3.100)

$$= \alpha \text{tr} \left( P_{n,\alpha}^{-1} \frac{dP_{n,\alpha}}{d\alpha} \right)$$

(3.101)

$$= \alpha \text{vec'} \text{vec} \mathbf{K}(P_{n,\alpha})$$

(3.102)

$$= \alpha \left( \text{vec} I \right)' \sqrt{(P_{n,\alpha} \otimes P_{n,\alpha})^{-1}} \text{vec} \left( \frac{dP_{n,\alpha}}{d\alpha} \right)$$

(3.103)

Hence we can write

$$S_{\alpha}(P_{n,\alpha}) = \alpha \left( \text{vec} I \right)' \sqrt{(P_{n,\alpha} \otimes P_{n,\alpha})^{-1}} \text{vec} \left( \frac{dP_{n,\alpha}}{d\alpha} \right)$$

(3.104)

$$\approx -\alpha \text{vec'} I \left( \sqrt{(P_{n,\alpha} \otimes P_{n,\alpha})} \left( e^{-\frac{\alpha}{n}} \otimes e^{-\frac{\alpha}{n}} \right) \sqrt{(P_{n,\alpha} \otimes P_{n,\alpha})^{-1}} - (1 - \alpha) I \otimes I \right)^{-1} \right.$$

(3.105)

$$\text{vec} \left( \left( \sqrt{P_{n,\alpha}D} \sqrt{P_{n,\alpha}} \right)^{-1} + I \right)^{-1}$$

(3.106)

Note that the sensitivity depends largely on the eigenvalues of

$$\alpha \left( \sqrt{(P_{n,\alpha} \otimes P_{n,\alpha})} (F \otimes F)^{-1/n} \sqrt{(P_{n,\alpha} \otimes P_{n,\alpha})^{-1}} - (1 - \alpha) I \otimes I \right)^{-1}$$

(3.107)
3. System Design Relationships

We can write these directly by noting that \( \lambda(XYX^{-1}) = \lambda(Y) \) and writing

\[
\lambda_{ij} = \alpha \left( \lambda_j(F)^{-1/n} \lambda_i(F)^{-1/n} - (1 - \alpha) \right)^{-1}
\] (3.108)

Note that this is the same as the stability requirement, indicating that the system becomes increasingly sensitive as the system approaches the stability bounds. We see that the sensitivity increases the closer the system is to divergent.

We can compute the sensitivity to sensor count similarly. Noting that \( n \) is an integer we define the sensitivity as

\[
S_n(P_{n,\alpha}) = \text{tr} \left( n(P_{n+1,\alpha}^{-1} - P_{n,\alpha}) \right)
\] (3.109)

and expanding using vecs yields

\[
\frac{\text{vec}(P_{n+1,\alpha} - P_{n,\alpha})}{(n + 1) - n} = \left( (1 - \alpha) I \otimes I - \left( e^A \otimes e^{A'} \right)^{-1/n} + \alpha \left( P_{n,\alpha}^{-1} P_{n+1,\alpha}^{-1} + P_{n,\alpha} P_{n+1,\alpha}^{-1} \right) \right)^{-1} \cdot \left( e^A \otimes e^{A'} \right)^{-1/n+1} \vec(P_{n+1,\alpha} + I)
\] (3.110)

We can expand

\[
\left( e^A \otimes e^{A'} \right)^{-1/n} - \left( e^A \otimes e^{A'} \right)^{-1/n+1}
\] (3.111)

as a power series in \( \frac{1}{n} \) and taking the first term we get

\[
\left( e^A \otimes e^{A'} \right)^{-1/n} - \left( e^A \otimes e^{A'} \right)^{-1/n+1} \approx - \frac{1}{n^2} (A \otimes I + I \otimes A)
\] (3.112)

for \( n >> 1 \).

Using (3.115), for \( \alpha \to 0 \)

\[
\frac{\text{vec}(P_{n+1,\alpha} - P_{n,\alpha})}{(n + 1) - n} \approx - \left( (1 - \alpha) I \otimes I - \left( e^A \otimes e^{A'} \right)^{-1/n} \right)^{-1} \cdot \frac{1}{n^2} (A \otimes I + I \otimes A) \vec(P_{n+1,\alpha} + I)
\] (3.113)

giving an approximation of the sensitivity as

\[
S_n(P_{n,\alpha}) \approx - \text{vec}' \left( P_{n,\alpha}^{-1} \right) \left( (1 - \alpha) I \otimes I - \left( e^A \otimes e^{A'} \right)^{-1/n} \right)^{-1} \cdot \frac{1}{n} (A \otimes I + I \otimes A) \vec(P_{n+1,\alpha} + I)
\] (3.114)
which demonstrates that the sensitivity to the number of sensors is dominated by a $\frac{1}{n}$ term. Note that the sensitivity grows unbounded under the same conditions as (3.104).

The other relationship of interest is between the number of sensors at the loss ratio $\alpha$ for a fixed covariance. We assume that we have a target covariance $P_{n,\alpha}$ for a particular pair $\alpha, n$ and desire to find a new $\alpha^*, n^*$ for a particular $\alpha^*$ or $n^*$ such that the new covariance $P_{n^*,\alpha^*} + \Delta = P_{n,\alpha}$. We desire $\Delta$ positive semi definite such that some metric of $\Delta$ is minimised (i.e. $\text{tr} \Delta$ or $\lambda_{\text{max}} (\Delta)$ etc.)

\[
(P_{n,\alpha} + I) - F^{-1/n} (P_{n,\alpha} + I) F^{-1/n} = (P_{n,\alpha} - (P_{n,\alpha}^{-1} + D)^{-1})
\]

\[
(P_{n^*,\alpha^*} + I) - F^{-1/n^*} (P_{n^*,\alpha^*} + I) F^{-1/n^*} = (P_{n^*,\alpha^*} - (P_{n^*,\alpha^*}^{-1} + D)^{-1})
\]

Given $\alpha, n$ fixed how are $\alpha^*$ and $n^*$ related

\[
\frac{(P_{n,\alpha} + I) - F^{-1/n} (P_{n,\alpha} + I) F^{-1/n}}{\alpha^*} = K(P_{n,\alpha})
\]

\[
(I + \Delta) (P_{n,\alpha} + I) (I + \Delta)' - F^{-1/n^*} (I + \Delta) (P_{n,\alpha} + I) (I + \Delta)' F^{-1/n^*}
\]

and for $\Delta \to 0$ the relationship for the induced $p$ norm approaches

\[
\alpha^* > \| K(P_{n,\alpha})^{-1} ((P_{n,\alpha} + I) - F^{-1/n^*} (P_{n,\alpha} + I) F'^{-1/n^*}) \|_p
\]

or for trace

\[
\alpha^* > \frac{\text{tr} (P_{n,\alpha} + I) - \text{tr} (F^{-1/n^*} (P_{n,\alpha} + I) F'^{-1/n^*})}{\text{tr} (K(P_{n,\alpha}))}
\]

or for $\lambda_{\text{max}}$

\[
\alpha^* \lambda_{\text{max}} \left( \sqrt{(P_{n,\alpha} + I)}^{-1} K(P_{n,\alpha}) \sqrt{(P_{n,\alpha} + I)^{-1}} \right)
\]

\[
= \lambda_{\text{max}} \left( I - \sqrt{(P_{n,\alpha} + I)^{-1} F^{-1/n^*} (P_{n,\alpha} + I) F^{-1/n^*} (P_{n,\alpha} + I)^{-1}} \right)
\]

\[
\alpha^* > \frac{1 - \sigma_{\text{max}} \sqrt{(P_{n,\alpha} + I)^{-1} F^{-1/n^*} (P_{n,\alpha} + I)^{-1}}}{\lambda_{\text{max}} ((P_{n,\alpha} + I)^{-1} K(P_{n,\alpha}))}
\]

\[
> \frac{1 - \lambda_{\text{max}} (e)^{-2\Delta/n^*}}{\lambda_{\text{max}} ((P_{n,\alpha} + I)^{-1} K(P_{n,\alpha}))}
\]
Using the first two terms of the power series of $e^{-A/n^*}$ we get

$$\alpha^* > \frac{1}{n^*} \left( \frac{\text{tr} \left( 2A(P_{n,\alpha} + I) - \frac{1}{n} A \left( P_{n,\alpha} + I \right) A \right)}{\text{tr} \left( K(P_{n,\alpha}) \right)} \right) \quad (3.134)$$

and

$$\alpha^* > \frac{n^*}{\lambda_{\text{max}}(A)} \frac{\lambda_{\text{max}}(A)}{\left( P_{n,\alpha} + I \right)^{-1} K(P_{n,\alpha})} \quad (3.135)$$

These results provide three interesting tools to use in the design and understanding of distributed sensor networks with unreliable sensors.

Firstly we note that the sensitivity of the system to variations in the number of sensors and the rate of missing observations increases as the system approaches the (3.32) with $F = e^{\frac{A}{n}}$. Given a target performance $P_{n,\alpha}$ and a known number of sensors we can evaluate a bound on $\alpha$ using (3.93). We can use (3.129) to define a bound on $\alpha(n)$ given an existing system with known $P_{n,\alpha}$, $n$, $\alpha$ to predict the behaviour as the number of sensors changes. We can use (3.104) and (3.119) to evaluate the sensitivity of a system to perturbation in the probability of missed observations or changes in the number of sensors. In particular we notice that the sensitivity of systems to changes in the number of sensors decreases with the square of the number of sensors.

We can determine how many sensors $n$ are required to achieve a particular target prior covariance given a particular sensor utility $D$ by solving (3.89) for $n$ for a posterior target $P$.

Maximise integer $n$ s.t. $X < D$ where

$$\left( P - \frac{1}{\alpha} \left( (P + I) - F^{-1/n} (P + I) F^{-1/n} \right) \right)^{-1} - P^{-1} = X \quad (3.136)$$

After re-arranging we get

$$F^{-1/n} (P + I) F^{-1/n} > (P + I) - \alpha \left( P - (D + P^{-1})^{-1} \right) \quad (3.137)$$

and noting that $e^{-A/n} = F^{-1/n}$ and that $\ln \left( e^A S e^{A'} \right) = A + A' + \ln S$ where $S$ is symmetric we have

$$n > \lambda_{\text{max}} \left( \ln \left( I - \alpha \sqrt{(P + I)^{-1}} \left( P - (D + P^{-1})^{-1} \right) \sqrt{(P + I)^{-1}} \right)^{-1} \right) \quad (3.138)$$

which demonstrates the increase required in $n$ for a given $P$ as $\alpha \to 0$. Note that

$$\sqrt{(P + I)^{-1}} \left( P - (D + P^{-1})^{-1} \right) \sqrt{(P + I)^{-1}} < I \quad (3.139)$$
and consequently and using \( \lim_{X \to 0} \ln (I - X) = -X \) we can approximate the inequality to a first order as

\[
n > \frac{1}{\alpha} \lambda_{\text{max}} \left( (A + A') \sqrt{(P + I)} \left( P - (D + P^{-1})^{-1} \right)^{-1} \sqrt{(P + I)} \right)
\]

which highlights that to a first order \( \alpha n \) is a constant for a fixed system requirement \( \{A, P, D\} \).

### 3.6 Conclusion

In this chapter three main relationships between sensors, scheduling and transmission systems have been investigated. The purpose of the investigation is to allow the system designer to gain a more intuitive feel for how to design sensor systems using large numbers of low cost sensor platforms compared with a small number of more capable platforms.

In the first section we established a useful simplification of the recurrence relationship used to calculate the system covariance in the form of (3.28) and (3.30) for i.i.d missing observations. These become the foundations used to investigate the impact on performance of different methods of scheduling sensor observations. A useful simplification is that the system can be analysed to within a linear transform knowing only the “sensor utility”, the relative ratio of the “randomness” of the system compared to the accuracy of the sensors. This is intuitively satisfying as it suggests that a sensor need only be good enough to “just see” the variations of the observed system, and increasing its capability has little improvement beyond some critical point.

In the second section we have determined that optimal sensor scheduling strategy is dependent on the stability of the system being observed. A system that is stable (with eigenvalues less than or equal to one) benefits most from taking all measurements at once. A system that is unstable (eigenvalues > 1) may benefit from having equispaced observations, however if the eigenvalues are greater than \( \sqrt{\alpha} \) the system will always benefit from equispaced observations. The heuristic reason is that for stable systems it is beneficial to know the state as accurately as possible by taking all observations at once, secure in the knowledge that because the system is stable the amount of uncertainty that accumulates prior to the next set of observations is comparatively small. For an unstable system small errors in the estimate rapidly accumulate, thus suggesting that the better strategy is to take many less accurate, less correlated measurements as this restricts the worst case error to a greater extent. A final observation from this section is that a randomised observation process operates no worse than either method.
In the final section the sensitivity of the system to variations in the number of sensors and the sensor reliability is investigated. We demonstrate that the system performance remains constant provided the expected number of sensor reports received ($\alpha n$) remains constant. However, this rule of thumb starts to break down where the normalised covariance, as defined in (3.16), is significantly greater than $I$. 
Chapter 4

Applications of the System Design Relationships

4.1 Overview

In this chapter we use the relationships developed in the previous chapter to analyse the behaviour of large sensor networks. We consider the impact of commonly used media access methods (M/M/1/Q queue, CSMA-CD Shared Channel Access and AWGN and CDMA Channels) and consider how these methods impact on the MSE error of a system for a homogenous sensor system. We then extend the M/M/1/Q for a dense sensor cloud and consider the case of heterogeneous sensors where the sensor performance is dependent on the location of the sensor to provide an example of additional sensors degrading system performance.

4.2 Introduction

We can apply the results of Chapter 3 to answer questions about how existing and proposed sensor networks will impact on the performance of the estimator. We consider different network structures and examine how they impact the system performance. In particular we consider methods used to transmit data between system elements and the impact these methods have on system performance. Of particular interest is the relationship between the likelihood of an observation being received (\( \alpha \)) and the number of sensor in the system (\( N \)). In many network systems these two parameters have a fixed relationship. Using the results from Chapter 3 we see that there are also relationships between these parameters if a constant system performance is required.

We will consider 4 types of network systems and the impacts these systems have on the MSE error of the system.
4. Applications of the System Design Relationships

- Markov M/M/1/Q Queue
- CSMA-CD Shared Channel Access
- AWGN and CDMA Channels
- Dense Cluster of Sensors using M/M/1 Queue

These models are significant as they cover a number of important channel access strategies used by wired and wireless communications systems. Markov queues are observed in both wired and wireless systems when a centralized channel access strategy can be implemented. This may be through DiffServ style egress rate limiting \(^1\) or as a consequence of fixed channel allocation strategies \(^2\). The AWGN channel is of fundamental importance, and illustrates how dividing the capacity of a section of spectrum up amongst multiple channels reduces the total capacity. The CSMA-CD system is another decentralized access strategy that often occurs where a distributed network cannot monitor all network accesses \(^3\).

4.3 Markov Queue Depth M/M/1/Q

The Markov Queue is representative of many general purpose network architectures where access to network resources is independent between multiple access points. This may be because the network has a large amount of random broadly distributed traffic, so that access to the network can be assumed to be independent and quantified solely by the utilisation. A queue is used to synchronise access to the network, with traffic enqueued until network resources become available. The queue may be actually physically implemented as a shared transmission buffer or it may be a logical construct formed by a higher level protocol such as by using TCP transmission queues or round robin polling. In terms of the model 4.1 it is simply assumed that the queue has a finite depth, with messages removed from the queue as network resources become available and that a sensor can reliably enqueue a message provided space exists in the queue. The other significant property is that messages are assumed to be delivered if successfully enqueued, there are no collisions, interference or lost messages provided there is space to initially enqueue the message. The service times for each message and the arrival times are independent Poisson processes characterised by a message generation rate and message transmission rate.

\(^1\)IPv4 and IPv6 networks, see RFC 5865
\(^2\)Such as a TDMA multiplex as in JTIDS communications
\(^3\)Due to hidden/exposed terminal problem. See the Distributed Co-ordination Function (DCF) of the 802.11 family of protocols for an explanation of the problem and the Point Co-ordination Function (PCF) and Hybrid Co-ordination Function for examples of protocols that mitigate it.
Consider a communications system with a shared transmission medium and a queuing discipline for accessing the media. Let $q$ be the number of queue slots available and define $u$ the utilization as

$$u = \frac{\text{rate of messages generated}}{\text{rate of messages transmitted}} \quad (4.1)$$

The probability of the queue being $q$ deep is

$$p_0 = \left(1 + \sum_{j=1}^{q} u^j\right)^{-1} \quad (4.2)$$

$$= \frac{u - 1}{u^{q+1} - 1} \quad (4.3)$$

where

$$p_q = p_0 u^q \quad (4.4)$$

$$= \frac{u - 1}{u^{q+1} - 1} u^q \quad (4.5)$$
Consequently the probability of losing a message is the probability that the queue is $q$ deep when an attempt is made to enqueue a message (i.e. it is full)

$$Pr \text{ (loss)} = \frac{u - 1}{u^{q+1} - 1} u^q$$ \hspace{1cm} (4.6)

$$= 1 - \frac{1 - u^q}{1 - u^{q+1}} \hspace{1cm} (4.7)$$

If we have $n^*$ sources then the total utilisation is $n^* u$ and consequently the probability of loss becomes

$$Pr \text{ (loss|}n^* \text{ sensors)} = 1 - \frac{1 - (n^* u)^q}{1 - (n^* u)^{q+1}} \hspace{1cm} (4.8)$$

From (3.129) we know that to achieve $\text{tr} \left( P_{n,\alpha} \right) = \text{tr} \left( P_{n^*,\alpha^*} \right)$ then $n^*, \alpha^*$ must satisfy

$$\kappa = \frac{P_{n,\alpha} + I}{\text{tr} \left( K \left( P_{n,\alpha} \right) \right)} \hspace{1cm} (4.9)$$

$$\alpha^* > \text{tr} (\kappa) - \text{tr} \left( e^{-A/n^* \kappa} e^{-A'/n^*} \right) \hspace{1cm} (4.10)$$

The probability of successful reception is $1 - Pr \text{ (loss)}$ so consequently we set $\alpha^* n^* = \frac{1 - (n^* u)^q}{1 - (N^* u)^{q+1}}$ and evaluate

$$\frac{1 - (n^* u)^q}{1 - (n^* u)^{q+1}} > \left( \text{tr} (\kappa) - \text{tr} \left( e^{-A/n^* \kappa} e^{-A'/n^*} \right) \right) \hspace{1cm} (4.11)$$

Provided this inequality holds the system performance is not reduced ($\text{tr} \left( P_{n,\alpha} \right) \geq \text{tr} \left( P_{n^*,\alpha^*} \right)$) as the sensor count is increased.

We can gain some more insight by considering the case where $(n^* u)^q \gg 1$ ,

$$\frac{1 - (n^* u)^q}{1 - (n^* u)^{q+1}} \rightarrow \frac{1}{n^* u} \hspace{1cm} (4.12)$$

i.e. the scenario where the system is congested. At this point we can simplify (4.11) to

$$\text{tr} \left( e^{-A/n^* \kappa} e^{-A'/n^*} \right) > \text{tr} \left( \kappa - \frac{I}{n^* u} \right) \hspace{1cm} (4.13)$$

re-arranging and taking logs

$$\text{tr} \left( \ln \left( e^{-A/n^* \kappa} e^{-A'/n^*} \right) \right) > \text{tr} \left( \ln \left( \kappa - \frac{I}{n^* u} \right) \right) \hspace{1cm} (4.14)$$

and taking advantage of $e^{-A/n^* \kappa} e^{-A'/n^*}$ and $\kappa$ symmetric we can write

$$\text{tr} \left( -2A/n^* + \ln \kappa \right) > \text{tr} \left( \ln \left( \kappa - \frac{I}{n^* u} \right) \right) \hspace{1cm} (4.15)$$

$$\text{tr} \left( -A/n^* - A'/n^* \right) > \text{tr} \left( \ln \left( \kappa - \frac{I}{n^* u} \right) - \ln \kappa \right) \hspace{1cm} (4.16)$$

$$-\frac{\text{tr} \left( 2A \right)}{n^*} > \text{tr} \left( \ln \left( I - \kappa^{-1} \right) \right) \hspace{1cm} (4.17)$$
As $\kappa$ is symmetric we can use the inequality
\begin{equation}
\text{tr}\left(\ln\left(I - \frac{\kappa^{-1}}{n'u}\right)\right) < -\text{tr}\left(\frac{\kappa^{-1}}{n'u}\right)
\end{equation}
and we have
\begin{equation}
\frac{\text{tr}(2A)}{\text{tr}(\kappa^{-1})} < \frac{1}{u}
\end{equation}
as a sufficient condition for system improvement for all $n^*$. Note that $\kappa$ is set by the configuration of the system and is invariant while
\begin{equation}
\frac{\text{tr}(P_{n,\alpha} + I)}{\text{tr}\left(P_{n,\alpha} - (P_{n,\alpha}^{-1} + D)^{-1}\right)} > \frac{\text{tr}(P_{n,\alpha})}{\text{tr}\left(P_{n,\alpha} - (P_{n,\alpha}^{-1} + D)^{-1}\right)}
\end{equation}
implies that $\text{tr}(2A) \ll \frac{1}{u}$. This indicates that we will see improvements with large sensor numbers only if the eigen values of $A$ approach 0 or the initial utilisation approaches 0. Provided the bound is satisfied increasing the sensor count will always increase the system performance, although this may be a minor improvement.

We can also observe the rate at which $P_{n,\alpha}$ changes with $n$. Using
\begin{equation}
\frac{d}{dn} (P_{n,\alpha}) \approx -\frac{\alpha}{n^2} \left(\left(I - \frac{e^{-A/n}}{n}\right) \otimes \left(I - \frac{e^{-A/n}}{n}\right)\right)^{-1} \cdot (A \otimes I + I \otimes A) \text{vec} K\left(P_{n,\alpha}\right)
\end{equation}
and substituting $\alpha(n)$ back in we have
\begin{equation}
\frac{d}{dn} (P_{n,\alpha}) \approx -\frac{1-(n^*u)^2}{n^2}
\end{equation}
\begin{equation}
\left(\left(I - \frac{e^{-A/n}}{n}\right) \otimes \left(I - \frac{e^{-A/n}}{n}\right)\right)^{-1} \cdot (A \otimes I + I \otimes A) \text{vec} K\left(P_{n,\alpha}\right)
\end{equation}
which for $n \gg 1$ approaches
\begin{equation}
\frac{d}{dn} (P_{n,\alpha}) \approx -\frac{1}{(n^*)^3 u} (A \otimes I + I \otimes A) \text{vec} K\left(P_{n,\alpha}\right)
\end{equation}
indicating that although the system improves, the improvement is decreasing cubically. Additional sensors provide little improvement. This indicates that after adding
a relatively small number of sensors additional sensors at best provide only minor improvement, while if \( \text{tr} (2A) < \text{tr} \left( \frac{1}{w} \right) \) we cannot even be sure that additional sensors will improve the system.

In conclusion there is little advantage in adding additional sensors beyond the increase in either coverage or availability. If coverage is increased then the availability of the system is not improved (as the failure of a single sensor would reduce the coverage - the sensors being no longer redundant) while under either condition if the sensors continue to provide reports, while using the same communications network information will be lost, reducing the accuracy of the state estimation.

4.4 CSMA-CD Shared Channel Access

CSMA-CD (Carrier Sense Multiple Access with Collision Detection) is at the heart of many wireless transmission systems and in particular the widely deployed 802.11 family \(^4\) as part of the DCF and HCF implementation.\(^5\) In essence a single shared transmission resource is used when it is detected as being idle. The difficulty is that even with monitoring and handshake protocols it is impossible for all potential transmitters to determine that the channel is idle due to differing transmission paths and latencies. Consequently there always exists a possibility of two transmitters simultaneously transmitting and consequently interfering with each other. Consider 3 sensors A,B,C in figure 4.2. Due to obstacles transmitter B can monitor A and C for activity, A and C cannot monitor each other. Hence A cannot determine if C is transmitting and may transmit at the same time, accidentally jamming B from receiving from C. In addition both B and C may determine that the channel is idle, however there is a short period of time (the transit time) during which B may start transmitting before a message from C can reach B. With terrestrial systems where the maximum length of a link is of the order of 10000km the delay will be less than 10ms, while for a satellite service the delay can be of order 600ms for a geostationary service. For slow, low bit rate services this may correspond to only a small number of bits or a packet header being corrupted, and may be recoverable with error correction, but at higher bit rates this may correspond to the entire message, or even many messages.\(^6\) In this respect it differs from the simple Markov model as messages are not only queued

\(^4\) a.k.a. WiFi

\(^5\) The Distributed and Hybrid Co-ordination Functions are part of the 802.11 physical layer definition. These protocols manage access by the wireless stations to the radio spectrum, preventing multiple stations from transmitting simultaneously and interfering.

\(^6\) For example, each 802.11af Physical Protocol Data Unit is allocated no more than 5.484ms of transmit time and consists of not more than 4,692,480 bytes depending on modulation. It may contain one or more ethernet MAC frames so the disruption of a single PPDU may cause the loss of multiple packets being sent from a base-station to multiple users.
(when the channel is known to be in use) but in addition messages may be lost after being transmitted, with the probability of loss being related to how many transmitters are attempting to share the channel. Most transmission control protocols attempt to "fairly" allocate the channel resources, usually through attempting to ensure that data is despatched to the transmission channel in order of arrival.\(^7\)

In figure 4.3 each transmitter will locally enqueue messages for transmission while waiting for the transmission media to become idle. The Media Access control protocol used by the system then permits each channel to transmit one message onto the media one message at a time and attempts to fairly allocate the channel resources between each transmitter. As the number of transmitters increases the channel capacity available to each transmitter is reduced proportionally. For a single transmitter transmitting messages at rate \(r\) into a shared channel with capacity \(c\) and queue depth \(q\) shared equally between \(n\) transmitters the probability of successfully enqueuing a message is

\[
\left( \frac{1 - \left( \frac{r}{c/n} \right)^q}{1 - \left( \frac{r}{c/n} \right)^q} \right) = \left( \frac{1 - (nu)^q}{1 - (nu)^{q+1}} \right)
\]

which is of the same form as the expression in the previous section with the shared queue. The transmission process of managing the access to the transmission media is

\(^7\)With various degrees of success, for instance analysis of the 802.11 family indicates that channel sharing can be anything but fair (see [117])
Figure 4.3: Block Diagram of CSMA-CD System
not 100% reliable and may fail if more than one node transmits at a time. We may model this by using the results above and adding a probability of TX error occurring that is related to the number of transmitters. We assume that each node is capable of correctly determining that it has access to the media with probability \( p \). If a node incorrectly believes it has access to the channel it will erroneously start transmitting - corrupting the transmission. The probability of a transmission being successfully transmitted becomes \( p^n \) (i.e. all nodes have to make a correct determination of who has access to the channel - one must transmit - all others must be silent). The final probability of the message being successfully transmitted becomes the probability of sufficient queue depth being available and the probability of all transmitters correctly negotiating access to the transmission media

\[
\alpha (n) = \left( \frac{1 - (nu)^q}{1 - (nu)^{q+1}} \right) (p^n)
\]

(4.29)

From (3.129) and using (4.9) for the definition of \( \kappa \) we have the requirement that

\[
\left( \frac{1 - (nu)^q}{1 - (nu)^{q+1}} \right) (p^n) > \left( \text{tr} (\kappa) - \text{tr} (e^{-A/n} \kappa e^{-A/n}) \right)
\]

(4.30)

Using the same method as the previous section we can re-arrange and take logs to yield

\[
\text{tr} \left( \ln (e^{-A/n} \kappa e^{-A/n}) \right) > \text{tr} \left( \ln \left( \left( I - \left( \frac{1 - (nu)^q}{1 - (nu)^{q+1}} \right) (p^n) \left( \kappa^{-1} \right) \right) \kappa \right) \right)
\]

(4.31)

\[
\frac{\text{tr} (-A - A')}{n} > \text{tr} \left( \ln \left( I - \left( \frac{1 - (nu)^q}{1 - (nu)^{q+1}} \right) (p^n) \left( \kappa^{-1} \right) \right) \right)
\]

(4.32)

For \( \left( \frac{1 - (nu)^q}{1 - (nu)^{q+1}} \right) (p^n) \kappa^{-1} \ll 1 \) we can approximate

\[
\ln \left( I - \left( \frac{1 - (nu)^q}{1 - (nu)^{q+1}} \right) (p^n) \kappa^{-1} \right) \approx - \left( \frac{1 - (nu)^Q}{1 - (nu)^{Q+1}} \right) (p^n) \kappa^{-1}
\]

(4.33)

and write

\[
\frac{\text{tr} (A + A')}{\text{tr} (\kappa^{-1})} < \left( \frac{1 - (nu)^q}{1 - (nu)^{q+1}} \right) n (p^n)
\]

(4.34)

which as \( n \to \infty \) approaches

\[
\frac{\text{tr} (A + A')}{\text{tr} (\kappa^{-1})} < \frac{(p^n)}{u}
\]

(4.35)

As \( 0 < p < 1 \) there is always a point at which increasing the number of sensors reduces the performance for systems that are unstable (i.e. have eigenvalues of \( A \) that are \( > 0 \)).
Note that the relationship for ensuring system improvement is close to that of a single queued system, degraded by the probability of the system to access the transmission media correctly. Although CSMA-CD is not directly used, the principle that a media access method is used to arbitrate access to a shared resource is common. Other common methods are the use of RTS/CTS (ready to send / clear to send) handshakes to resolve two transmitters which do not have the ability to receive each others carrier, but these mechanisms can incur unacceptably high channel access latencies. When used with satellite systems the time taken to transmit from one mobile station to the satellite and received by the second mobile station will seriously impact the total capacity of the channel as the channel is unavailable while the handshake is being processed. The long latencies cause the utilisation of the channel to be degraded as the channel cannot be utilised while a channel access is pending. In order to mitigate this many satellite systems use time division access to share access to satellite resources\(^8\) use fixed bandwidth allocation which can only be varied slowly, if at all.

### 4.5 AWGN and CDMA Channels

The CDMA channel access process allows multiple transmitters to share the same spectrum without requiring them to centrally co-ordinate. Each transmitter appears as noise to each other receiver and consequently the capacity available for each transmitter can be estimated using Shannon’s law. Consider a noisy channel with channel capacity (from Shannon 1948) noise limited according to

\[
r_c = b \log_2 \left(1 + \frac{s}{n}\right)
\]

(4.36)

where \(s/n\) is the signal to noise ratio and \(r_c\) is the channel capacity and \(b\) is the channel bandwidth. As additional channels are added to the same spectrum the amount of noise in the spectrum increases - the transmissions of other channels interfere with the channel of interest. Consequently the channel capacity decreases as the channel count increases

\[
r_c = b \log_2 \left(1 + \frac{1}{1/ (s/n) + (n - 1)}\right)
\]

(4.37)

assuming that the signals from other channels correspond to non-correlated noise in the channel of interest. This is equivalent to using Pseudo Orthogonal codes to separate channels in a CDMA system where channels cause co-channel interference. Thus additional channels within the same piece of spectrum will reduce the capacity of each channel, and correspondingly increase its utilisation above its interference free baseline

\(^8\) such as JTIDS or Link 11
4. Applications of the System Design Relationships

(given that the traffic demand on the channel does not change and the channel capacity reduces). The effective channel utilisation $u_e$ in the presence of interference is given by

$$u_e = \frac{\log_2 (1 + s/n)}{\log_2 (1 + 1/(s/n) + (n - 1))} u$$

(4.38)

where $u$ is the channel utilisation when there are no other transmitters. From the M/M/1/Q case the probability of a message being successfully received in a system with a queue is

$$\alpha(N) = \frac{1 - (u_e)^Q}{1 - (u_e)^{Q+1}}$$

(4.39)

As the number of sensors is increased, each sensor has a reduced channel capacity available, increasing the utilisation of the sensor’s dedicated channel. We can consider the case of congestion limited behaviour where $s/n \gg \frac{1}{(n-1)}$, that is additional transmitters contribute more interference than is generated by the receiver noise or other noise sources.

$$u_e \approx \frac{\log_2 (1 + (s/n))}{\log_2 (1 + 1/(n - 1))} u$$

(4.40)

From (3.129) and using (4.9) for the definition of $\kappa$ we have the requirement that

$$\frac{1 - (u_e)^Q}{1 - (u_e)^{Q+1}} > \left( \text{tr} (\kappa) - \text{tr} \left( e^{-A/n} \kappa e^{-A/n} \right) \right)$$

(4.41)

$$\frac{\text{tr} (-A - A')}{n \text{tr} (\kappa^{-1})} \geq \frac{1 - (u_e)^Q}{1 - (u_e)^{Q+1}}$$

(4.42)

$$\frac{2 \text{tr} (A)}{N \text{tr} (\kappa^{-1})} \leq \frac{1 - \left( \frac{\log_2 (1 + (s/n))}{\log_2 (1 + 1/(n - 1))} u \right)^Q}{1 - \left( \frac{\log_2 (1 + (s/n))}{\log_2 (1 + 1/(n - 1))} u \right)^{Q+1}}$$

(4.43)

$$\log_2 (1 + s/n) \frac{2 \text{tr} (A)}{\text{tr} (\kappa^{-1})} \leq \frac{n}{(n - 1)}$$

(4.44)

The consequence is that if the original $s/n$ of the system was very poor (small) but the bandwidth of the channel large enough that the utilisation of the channel was still low ($u$ small) then adding additional channels will always improve performance. However there is a critical point, as the $s/n$ improves, holding the utilisation of the link constant and allowing the bandwidth to correspondingly reduce, that adding additional sensors causes enough degradation in channel performance that the additional sensor reports do not compensate for the reduction in reliability of sensor report delivery. This point is governed by the system properties encapsulated by $\frac{2 \text{tr} (A)}{\text{tr} (\kappa^{-1})}$.
4. Applications of the System Design Relationships

4.6 Dense Cluster of Sensors using M/M/1 Queue

The previous discussions considered sensor performance to be homogenous and independent of target state. In actuality this almost never happens. There is a very common form of sensor target state coupling that is of practical importance in a distributed sensor system. Most sensors have performance that degrades as the distance between the sensor and target increases, most typically the MSE of the sensor increases with the square of the distance, this is most notable when using linearised directional sensors (range/bearing linearised into Cartesian co-ordinates) - in this case the MSE of the decoupled X/Y estimate is dependent on the range.

Consider sensors uniformly scattered (with density \( \rho \)) through some region \( A \). Let \( d_{k,j} = \|x_k - x_{j,k}\| \) be the distance of the \( k \)th observation between the \( j \)th sensor and the target. Define the performance of the \( j \)th sensor for the \( k \)th observation as

\[
\begin{align*}
    z_{k,j} &= Hx_k + \|x_k - x_{j,k}\| \sqrt{Rv_{k,j}} \\
    z_{k,j} &= Hx_k + d_{k,j} \sqrt{Rv_{k,j}}
\end{align*}
\]

where \( x_k \) is the target location and \( x_{j,k} \) is the location of the \( j \)th sensor (of \( n \)) at time \( k \). We will assume that \( x_{j,k} \) is independently uniformly distributed over a area \( A \) s.t. for \( x_{j,k} \in A, \|x_k - x_{j,k}\| < d_{\text{max}} \). i.e. the sensors are uniformly distributed through a circular region and the performance of the sensors (the uncertainty) increase with the distance from the sensor. Note that in general \( d_{k,j} \) will depend on \( x_j \) if the set of reporting sensors is constant and the sensor location is fixed. However if the sensors are also manoeuvring and/or the set of reporting sensors changes from observation to observation we may argue that at the set of distances at interval \( d_{k,j} \) at the \( k \)th observation map approaches independence to the actual target state when there is a large number of sensors, but only a small changing subset can actually report on the target state.

Under the assumption of independence we have prior covariance (for \( n \) sensors) as

\[
P_{k+1|k} = F \left( \left( P_{\text{inv}_{k-1}}^{-1} + \sum_{j=1}^{n} d_{k,j}^{-2}D \right)^{-1} + I \right) F - I
\]

It is desirable to know the expected performance of the system at the limit of \( k \to \infty \) over all \( x_k \) and \( x_{j,k} \). Note that although \( d_{k,j} \) is a function of \( x_k \) that \( x_k \) and \( x_{j,k} \) are independent (the set of reporting sensors varies from observation to observation and the position of the sensors changes) and implies that \( d_{k,j} \) are independent. Consequently
we can write that we require

\[ E_{d,k;j}(P_{k+1|k}) = F \left( E_{d,k;j} \left( \left( P_{k|k-1}^{-1} + \sum_{j=1}^{n} d_{k,j}^{-2} D \right)^{-1} \right) + I \right) F - I \]  
(4.50)

\[ E(P_{k+1|d,k}) = F \left( E_{d,k;j} \left( \left( P_{k|\gamma,j,k-1,d_{j,k-1}}^{-1} + \sum_{j=1}^{n} d_{k,j}^{-2} D \right)^{-1} \right) + I \right) F - I \]  
(4.51)

where \( E_{d_k}(y) = \int y p(y|x_k) dx_k \).

Define \( \kappa : \kappa \in (0,1) \) as the information reduction factor s.t.

\[ E(P_{k+1|\gamma,j,k,d_{j,k}}) = F \left( E_{d,k;j,\gamma,j,k} \left( \left( P_{k|\gamma,j,k-1,d_{j,k-1}}^{-1} + \sum_{j=1}^{n} d_{k,j}^{-2} D \right)^{-1} \right) + I \right) F - I \]  
(4.52)

\[ = F \left( E_{\gamma,j,k} \left( \kappa \left( P_{k|\gamma,j,k-1,d_{j,k-1}}^{-1} + \kappa D \right)^{-1} \right) + I \right) F - I \]  
(4.53)

Define Z as the solution of

\[ Z = F \left( (Z^{-1} + D)^{-1} + I \right) F' - I \]  
(4.55)

and X as the solution of

\[ X = F \left( (X^{-1} + \kappa D)^{-1} + I \right) F' - I \]  
(4.56)

We can write

\[ \frac{Z}{\kappa} = F \left( \left( \frac{Z}{\kappa} \right)^{-1} + \kappa D \right)^{-1} + I \]  
(4.57)

or after a change of variable with \( Y = \frac{Z}{\kappa} \)

\[ Y = F \left( (Y^{-1} + \kappa D)^{-1} + \frac{I}{\kappa} \right) F' - \frac{1}{\kappa} \]  
(4.58)

We can use this to write a first order approximation for the solution X

\[ Y = X + \varepsilon = F \left( (X + \varepsilon)^{-1} + \frac{I}{\kappa} \right) F' - \frac{1}{\kappa} \]  
(4.59)

\[ = F \left( (X^{-1} + \kappa D)^{-1} + I \right) F' - I + \varepsilon \]  
(4.60)

which implies

\[ \varepsilon \approx \left( \frac{1}{\kappa} - 1 \right) (FF' - I) \]  
(4.61)
and consequently
\[ X = Y - \varepsilon \] (4.62)
\[ = \frac{1}{\varepsilon} Z - \varepsilon \] (4.63)
\[ \approx \frac{1}{\varepsilon} \left( Z + (I - FF') \right) - (I - FF') \] (4.64)

Note that \( \varepsilon \) is random variable independent of \( P_{k|\gamma_{j,k-1},d_{j,k-1}} \). Consequently we can write
\[ \varepsilon_k = \sum_{j=1}^{n} d_{k,j}^{-2} \] (4.65)

\[
\lim_{k \to \infty} \mathbf{E} \left( P_{k+1|d_{k,j}} \right) = \lim_{k \to \infty} \mathbf{F} \left( \mathbf{E} \left( \left( P_{k|d_{j,k-1}}^{-1} + \varepsilon_k D \right)^{-1} \right) + I \right) F - I
\]
\[
< \lim_{k \to \infty} \mathbf{F} \left( \mathbf{E} \left( \left( \mathbf{E} \left( P_{k|d_{j,k-1}}^{-1} + \varepsilon_k D \right)^{-1} \right) + I \right) F - I
\]
\[ = \mathbf{F} \left( \mathbf{E} \left( \left( X^{-1} + \varepsilon_k D \right)^{-1} \right) + I \right) F - I \] (4.68)
\[ \approx \mathbf{E} \left( \frac{1}{\varepsilon} \right) \left( Z \left( I - FF' \right) \right) - (I - FF') \] (4.69)

where \( X \) and \( Z \) are as defined above.

Consider a system where the accuracy of the measurements decreases with the distance. Many real sensors have errors that are linearly proportional to the distance from the sensed object (i.e. anything that senses angle and distance and then transforms these measurements using trigonometry will have errors dependent on the range). This translates into the variance of the measurement being scaled by \( \frac{1}{\varepsilon} = r^2 \). If we select the \( n \) closest sensor reports how will the system performance vary.

We consider the probability of \( n \) reporting sensors being contained within a region \( \mathcal{A} \), and then assuming a uniform distribution of sensors calculate the expected value of \( \varepsilon^{-1} = \left( \sum r^{-2} \right)^{-1} \). We can treat \( \mathcal{A} \) as the 2d circular plane centred on the target, with sensors uniformly distributed. The probability distribution of sensor distances, given that there is a single sensor within a distance \( r_{\text{max}} \) is
\[ \Pr \left( \text{sensor at distance } r \right) = \frac{2 - r/r_{\text{max}}}{r_{\text{max}}^2} \] (4.70)
and the density of sensors \( \rho \) is
\[ \rho = \frac{N}{2\pi r_{\text{max}}^2} \] (4.71)

The probability of \( n \) sensors being located in a volume \( \mathcal{A} \) is given by the Poisson distribution where the sensor density is set by \( \rho \). The probability given that there are
n sensors consequently is in a region $v$

$$\Pr (n = N) = \frac{\left(\frac{\rho v}{n!}\right)^n e^{-\left(\frac{\rho v}{n!}\right)}}{\int_0^\infty \left(\frac{\rho v}{n!}\right)^n e^{-\left(\frac{\rho v}{n!}\right)} dv}$$

$$= \frac{\rho^n (\rho a)^n}{n!} e^{-\left(\frac{\rho a}{n!}\right)} \tag{4.72}$$

with

$$E (v) = \int_0^\infty v \frac{(\rho v)^n}{n!} e^{-\left(\frac{\rho v}{n!}\right)} dv$$

$$= \frac{(n + 1)}{\rho} \tag{4.74}$$

and

$$E (v^2) = \int_0^\infty v^2 \frac{(\rho v)^n}{n!} e^{-\left(\frac{\rho v}{n!}\right)} dv$$

$$= \frac{(n + 2)(n + 1)}{\rho^2} \tag{4.76}$$

$$\text{var} (v) = \frac{(n + 1)}{\rho^2} \tag{4.78}$$

We can use the Unscented Transform (see for examples [118]) to calculate $E \left(\frac{1}{\zeta_n}\right)$

for the 2D planar case for varying values of $n$

$$v = 2\pi r^2 \tag{4.79}$$

$$E (r^2) = \frac{1}{2\pi} E (v) \tag{4.80}$$

$$\text{var} (r^2) = \left(\frac{1}{2\pi}\right)^2 \text{var} (v) \tag{4.81}$$

and evaluating

$$E \left(\frac{1}{\zeta_n}\right) = \frac{1}{2n} \sum_{k=1}^n \left( \frac{1}{\sum_{j=1}^n s_{k,j+}} + \frac{1}{\sum_{j=1}^n s_{k,j-}} \right) \tag{4.82}$$

and

$$s_{k,j+} = \frac{1}{2\pi} E (v_k) : k \neq j \tag{4.83}$$

$$= \frac{1}{2\pi} \frac{k + 1}{\rho} \tag{4.84}$$

$$s_{k,j+} = \frac{1}{2\pi} E (v_k) + \sqrt{\text{var} (v_k)} \tag{4.85}$$

$$= \frac{k + 1}{2\pi \rho} \left( 1 + \frac{1}{\sqrt{k + 1}} \right) \tag{4.86}$$

$$s_{k,j-} = \frac{k + 1}{2\pi \rho} \left( 1 - \frac{1}{\sqrt{k + 1}} \right) \tag{4.87}$$
Define \( u_n = \sum_{j=1}^{n} \frac{1}{s_{k,j}^+} \) so that

\[
\frac{1}{\left( \sum_{j=1}^{n} \frac{1}{s_{k,j}^+} \right)} = \frac{1}{2\pi \rho} \left( \frac{u_n + \frac{1}{(k+1) + \sqrt{k+1} - \frac{1}{k+1}}}{} \right) \quad (4.88)
\]

\[
= \frac{1}{2\pi \rho} \left( \frac{(k+1) (1 + \sqrt{k+1} u_n + 1)}{} \right) \quad (4.89)
\]

\[
\frac{1}{\left( \sum_{j=1}^{n} \frac{1}{s_{k,j}^-} \right)} = \frac{1}{2\pi \rho} \left( \frac{u_n + \frac{1}{(k+1) - \sqrt{k+1} - \frac{1}{k+1}}}{} \right) \quad (4.90)
\]

\[
= \frac{1}{2\pi \rho} \left( \frac{(k+1) (1 - \sqrt{k+1} u_n - 1)}{} \right) \quad (4.91)
\]

and consequently

\[
E \left( \frac{1}{x_n} \right) = \frac{1}{4\pi \rho} \frac{1}{n} \sum_{k=1}^{n} \left( \frac{(k+1) (1 + \sqrt{k+1} u_n + 1)}{(k+1) (1 + \sqrt{k+1} u_n - 1)} \right) \quad (4.92)
\]

where the term inside the \( \sum \) depends only on the number of sensors. From eq. (4.66) we see that the covariance is primarily dependent on \( E \left( \frac{1}{x_n} \right) \). We can graph the term solely dependent on the number of sensors \( n \) in Figure 4.4 on a log-log axis.

Figure 4.4: Graph of \( 4\pi \rho \log \left( E \left( \frac{1}{x_n} \right) \right) \) for system with constant \( \rho \)

Note that the improvement in performance is significantly less than that achieved by simply using \( n \) sensors at identical distances. Additional sensors do provide improvement, but not to the extent displayed by a set of \( n \) sensors that have performance independent of sensor position, or equivalently sensors that are equidistant from the object being measured. A further important point to notice is that \( \rho \) is the effective density of reporting sensors, that is the density of sensors that are successfully reporting observations to the information fuser. Consider the case of the likelihood of
a sensor successfully reporting an observation is given by \( \alpha = \mathbf{E}(\gamma_k) \) as considered in the previous sections. In this case we can define the effective density of sensors as

\[
\rho_{\text{effective}} = \alpha \rho
\]  

(4.93)

It is often the case, as is described above that the probability of receiving an observation is dependent on the number of interfering sensors. For example consider the case where the transmission of all the sensor reports must be enqueued for transmission such that they do not interfere. Under a \( M/M=1 \) queuing discipline (see eq. (4.2)) this implies that the probability of successfully transmitting a message is

\[
\alpha = \frac{1}{1 + nu}
\]  

(4.94)

\[
\rho_{\text{effective}} = \frac{\rho}{1 + nu}
\]  

(4.95)

where \( u \) is the channel utilisation of a single sensor and is a measure of the co-channel interference that multiple transmitting sensors would cause at the receiver. Re-arranging we get a new expression for \( \mathbf{E}\left(\frac{1}{x_n}\right) \) that includes the effect of \( M/M=1 \) queuing

\[
\mathbf{E}\left(\frac{1}{x_n}\right) = \frac{(1 + nu)}{4\pi \rho n} \sum_{k=1}^{n} (\cdots)
\]  

(4.96)

where the section in brackets is the same as eq. (4.92). As seen in Figure 4.5, due to the additional \( (1 + nu) \) term the benefits of additional sensors beyond some small \( n \) are completely eliminated. Once we incorporate the degradation in sensor performance caused by all the sensors not being able to get a perfect observation that increasing the number of sensors (taking the \( n \) best sensors) has a significantly reduced utility.

Note that the effectiveness increases slightly with increasing sensor count up to 32 sensors (which is equivalent to a total channel utilisation of \( 1/4 \)). Increasing to 64 sensors gives a minor improvement, and beyond that \( \frac{4\pi \rho n}{(1 + nu)} \mathbf{E}\left(\frac{1}{x_n}\right) \) starts to increase rapidly, indicating increased covariance of the estimator. The important point is that increased sensor count beyond a threshold will cause increased errors. Not all sensor reports are equivalent and as the probability of missing sensor reports increases then it becomes more likely that “good” reports are lost and “bad” reports are substituted.

How then does the number of sensors impact on the system performance? We see three mechanisms at work. Increasing the number of sensors improves the performance of the ideal system. Either all sensors taking measurements simultaneously, or consecutively will provide an improvement. The second affect is that not all sensors are of equal utility. If we use the \( n \) best sensors, where the sensors are evenly distributed through the observation space, the performance is degraded compared to \( n \) sensors at
4. Applications of the System Design Relationships

Figure 4.5: Plot of $4\pi \rho_{\text{effective}} E\left(\frac{1}{\kappa_n}\right)$ for $\rho_{\text{effective}}$

the optimal location. Finally note that transmitting the information from $n$ sensors requires more bandwidth/power than a single sensor. Because of this $\alpha$ and $n$ are related, with the probability of a message being lost strongly correlated to the number of sensors in use.

Considering (4.66) note that the improvement is not directly related to $E\left(\frac{1}{\kappa_n}\right)$ but also depends on the relationship between $Z$ and $(I - FF')$. A poor quality sensor (one with a small $D$) will have a large value of $\lambda_{\text{max}} (Z / (FF' - I))$, while a good quality sensor with a large $D$, capable of gaining much information for each observation will cause $Z$ to approach $(FF' - I)$. Thus a sensor with a small $D$, will benefit to a greater extent from an increase in the number of sensors as the first term in (4.66) dominates, while in the later case the fixed term is dominant.

4.7 Conclusion

This chapter demonstrates the utility of the relationships built in previous chapters. Firstly we demonstrate a method to separate the behaviour of the system based on its ideal characteristics (loss free with a single sensor) from its behaviour when placed in a multisensor environment with losses that depend on the number of sensors. The ideal characteristics are summarised in the term $\kappa$ (see eq. (4.9)) for a single sensor without loss. From this we can establish a bound for the required reliability $\alpha$ to maintain the expected covariance. Using the reliability for three different media access
control protocols we then used this bound to determine if increased sensor count will always improve system performance, using the M/M/1/q queuing discipline, the CSMA-CD method and the CDMA method. In all three cases we find there is an upper bound where increases in sensor count provide minimal improvement or may even cause system degradation.

For the M/M/1/q system eq. (4.11) gives the requirement for not reducing system performance as the number of sensors is altered, while eq.(4.19) gives the condition that if satisfies implies that adding sensors will always improve performance. Eq. (4.27) shows that the sensitivity of the covariance to the number of sensors is proportional to $\frac{1}{n^3}$, demonstrating that adding sensors is a poor way to improving performance.

The analysis of the CSMA-CD based system proceeded similarly, Eq. (4.30) gives the requirement for not reducing system performance and eq. (4.35) gives the requirement for improvement for $n$ large. It should be noted though that there will always be an $n$ which if exceeded may reduce the systems efficacy as the r.h.s of eq. (4.35) inequality will approach 0. The important point is that using CSMA-CD with a large number of sensors may cause the system to operate worse than a small number of sensors, and is a poor choice of media access control for dense sensor systems.

The analysis of the CDMA system eq. (4.42) gives the requirement for not reducing system performance and eq. (4.46) gives a requirement for improving performance for all $n$. The structure indicates that increasing sensor counts benefits a large bandwidth noisy channel more than a smaller bandwidth noise free channel, mainly as the noise channel is already operating in a noise limited mode, while the capacity of the noise free channel rapidly decreases as additional sensors are used.

The final system considered how the use of heterogeneous sensors changes the previous analysis. Rather than assuming all sensors have the same sensor quality ($D$) we assumed that the sensor quality is a function of the distance of the sensor from the object being sensed. The analysis was carried out over a 2 dimensional plane, assuming the sensor observation noise is dependent on the square of the distance, in a manner similar to that of a range and bearing tracker. Two cases were considered, firstly assuming a perfect access model where sensor density has no impact on message reception and secondly where sensor density impacts on message reception. For the case of an ideal system we see that the heterogeneous case does not perform as well as all sensors located equidistant when the sensors are uniformly distributed over the plane eq. (4.92). The reduction is explained by noting that although some sensors have a significantly better observation (they are closer) most sensors are added further away as the total number of sensors is increased (while maintaining the same sensor density). When congestion is taken into account by adding an additional loss factor dependent
on the number of sensors, reducing the effective number of sensors reporting. We saw that there is a sharp point of inflection where adding sensors makes the system considerably worse (see eq. (4.96)). The heuristic explanation is that not all sensor reports contain the same amount of information, but the likelihood of losing a report is the same for all reports. Loosing a report from a well positioned sensor is of greater impact than a poorly positioned sensor, and as more sensors are added it is likely that a lost sensor report due to congestion will be substituted with a lower quality report.

This final observation is of great import for the design of real sensor networks as it demonstrates that keeping the density of sensors low enough that they do not interfere is critical to achieving good performance when the sensors quality is dependent on being close to the object under observation. Simply increasing the sensor density will make the system less effective.
Chapter 5

Efficient transmission of measurements

5.1 Overview

In this chapter we will extend the results outlined in 2.10. We will demonstrate a method that allows the use of innovations without having the system diverge when messages are lost and establish bounds for the correct operation of this system. In 5.3 we describe the method for the transmission of innovations with message loss and establish the properties of the estimator formed. In 5.4 we continue the analysis assuming quantisation of the innovations and demonstrate that the quantised innovations can be transmitted without causing unbounded errors. 5.5 examines the consequence of sending many coarsely encoded innovations or few finely encoded innovations and 5.6 pulls the previous sections together into a single set of equations for use in Chapter 6.

5.2 Introduction

As discussed in Chapter 4, geographically dispersed sensor systems may be seriously limited in the amount of available bandwidth. Furthermore the links are unreliable, potentially degrading due to congestion and interference. The message size available for the transmission of observations can be quite small. Consider Link-11, used for the transmission of track data for both aircraft and naval vessels. Individual message frames are under 256 bits, while locations are represented as floating point numbers referenced to the WGS-84 geodesic.\(^1\) There is minimum information available on the

\(^1\)This is the same reference as is used in GPS navigation
5. Efficient transmission of measurements

accuracy of the estimation of the track data transmitted in the message.\(^2\) Up to 512K individual tracks can be represented in these systems, however the available bandwidth is not sufficient to realistically transmit this number of tracks. As a consequence efficiently transmitting information is of great importance. Any transmission system must deal with the reduced reliability of these systems, error rates of 10% are not atypical.\(^3\)

This motivates the consideration of more efficient methods of transmission of data, and in particular methods of transmitting the minimum set of information required to recover the state of the observation. Within this section the actual method of transmission is ignored, it is simply assumed that data can be transmitted with some degree of fidelity across point to point links. The management, co-ordination and configuration of these links is assumed to have zero cost. This assumption will be re-visited in the next chapter.

5.3 Transmission with lost Innovations

It is well known that just transmitting innovations requires the least amount of information to be transmitted between the sensor and the information consumer (see [115]). Each innovation contains only the information that a sensor has that differs from the expected observation the sensor would have taken. Consequently the innovation contains information about the control inputs to the system between observations. Each innovation provides a “summary” of all control inputs that occurred between the previous observation and the current observation. In addition an innovation can summarise all observations between two points in time - allowing the sensor system and the consumer of the information to know little about each other’s capabilities, provided all information contained in the innovation can be transmitted. This is termed the transmission of equivalent innovations, that is innovations that are equivalent to a sequence of observations based on the receiver’s state model without necessitating the transmission of all intervening observations.

If only innovations are sent a lost transmission will cause the receiver state to diverge. This requires that innovations only be transmitted over reliable networks and that all consumers of the sensor information receive all innovations generated over the observation period. Consequently there can be no “late joiners” - information consumers that join the network after observations started, or we require a mechanism that allows the late joiner to request a summary of all previous observations so that

\(^2\)See STANAG 5516 for more information on the message format of Link-11. Link-22 uses similarly formatted messages for compatibility.

\(^3\)For example Link-16 defines 3 reliability settings, Standard (80%), High (90%) and Auto-Retransmission, giving an indication of the expected error rates on a long range data link.
it can initialise its state. What is required is that the transmitter send sufficient additional information such that the loss of any individual transmission will not cause divergence.

Consider the following example. A Discrete time linear system as defined by (3.1) is formed with $F$ and $Q$ and a linear measurement noise $R$ as defined below

$$
F = \begin{bmatrix}
1 & 1 & 0 & 0 \\
-0.08 & 1.05 & 0 & 0 \\
0 & 0 & 1 & 1 \\
0 & 0 & -0.08 & 1.05
\end{bmatrix} \quad (5.1)
$$

$$
Q = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 5 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 5
\end{bmatrix} \quad (5.2)
$$

$$
R = \begin{bmatrix}
25 & 0 \\
0 & 25
\end{bmatrix} \quad (5.3)
$$

$$
H = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix} \quad (5.4)
$$

Note that this yields a discrete time process $F$ with complex eigenvalues with magnitude greater than 1. This is equivalent to an object oscillating around a central point, with the amplitude of oscillation increasing with time.

Figure 5.1 shows the performance of the MMSE on a logarithmic scale for no losses ($\alpha = 1$) and a 20% loss rate ($\alpha = .8$). Next note that if only innovations are sent when $\alpha = .8$ that the state diverges from the actual state. As the magnitude of the eigenvalues is greater than 1, missed innovations rapidly cause the system to diverge.

For a linear system transmitting measurements the MMSE is given by the Kalman Filter, with the steady state covariance given by the solution of

$$
P_{k+1|k} = F \left( P_{k|k-1}^{-1} + H' R^{-1} H \right) F' + Q \quad (5.5)
$$

$$
K_k = (FP_k F + Q) H' \left( H (FP_k F + Q) H' + H' R^{-1} H \right)^{-1} \quad (5.6)
$$

$$
K = \lim_{k \to \infty} K_k \quad (5.7)
$$

$$
P = \lim_{k \to \infty} P_k \quad (5.8)
$$

and innovation

$$
\tilde{y}_k = z_k - HF x_{k-1|k-1} \quad (5.9)
$$
Figure 5.1: Comparison of performance of log MMSE with innovations, quantisation and loss.
and innovation covariance

\[ S_k = H (FP_k F + Q) H' + R^{-1} \]  \hspace{1cm} (5.10)

At the receiver \((r)\) the sequence of innovations \(\tilde{y}_k\) is used to reconstruct the observations

\[ \tilde{z}_k = H F x^r_{k-1|k-1} + \tilde{y}_k \]  \hspace{1cm} (5.11)

and then may be used to form a receiver state estimate

\[ x^r_{k|k} = (I - K_k H) F x^r_{k-1|k-1} + K_k \tilde{z}_k \]  \hspace{1cm} (5.12)

\[ = F x^r_{k-1|k-1} + K_k \tilde{y}_k \]  \hspace{1cm} (5.13)

Note that if a single innovation is missed an error \(e_k = K_k \tilde{y}_k\) is introduced into the estimate of \(x_{k|k}\) and that at time \(k + n\) the error will become

\[ e_{k+n}(z_k) = F^n K_k \tilde{y}_k \]  \hspace{1cm} (5.14)

Eq. (5.14) immediately shows why innovations cannot be used in the presence of lost messages. A single lost message at time \(k\) will cause a non-decaying error if \(F\) has any eigenvalues greater than or equal to 1. This may still be unimportant if the expected error over many lost messages approaches 0. We may think of this as although messages are lost, they are as likely to cause errors in either direction and may cancel out.

Define the indicator \(\gamma_k \in \{0, 1\}\) to indicate the presence of a transmission with \(\gamma_k\) independent and with \(\mathbb{E}(\gamma_k) = \alpha\) the expected total error given observations \(z_{1:k}\)

\[ \mathbb{E}(e_k|z_{1:k}) = \sum_{n=0}^{k} e_k \Pr(\gamma_k = 0) \]  \hspace{1cm} (5.15)

is

\[ \mathbb{E}(e_k|z_{1:k}) = \mathbb{E}\left( \sum_{n=0}^{k} (1 - \alpha) F^n K_{k-n} \tilde{y}_{k-n} \right) \]  \hspace{1cm} (5.16)

\[ = \sum_{n=0}^{k} (1 - \alpha) F^n K_{k-n} \mathbb{E}(\tilde{y}_{k-n}) = 0 \]  \hspace{1cm} (5.17)

as the innovations are always 0 mean. Hence the expected error approaches 0 if the
innovations are 0 mean. Next we have the variance

\[
\text{vec var} (e_k|z_{1:k}) = \sum_{n=0}^{k} \text{vec} \mathbb{E} \left( (1 - \gamma_k)^2 F^n K_{k-n} (\tilde{y}_{k-n})' K_{k-n} F^n \right) \quad (5.18)
\]

\[
- \sum_{n=0}^{k} \text{vec} \mathbb{E} (e_k|z_{1:k}) \mathbb{E} (e_k|z_{1:k})'
\]

\[
= \sum_{n=0}^{k} \text{vec} \mathbb{E} \left( (1 - \gamma_k)^2 F^n K_{k-n} (\tilde{y}_{k-n})' K_{k-n} F^n \right) \quad (5.19)
\]

\[
> \sum_{n=0}^{k} (1 - \alpha) (F \otimes F)^n (K \otimes K) \text{vec} \mathbb{E} \left( (\tilde{y}_{k-n}) (\tilde{y}_{k-n})' \right) \quad (5.20)
\]

\[
= (1 - \alpha) \left( (I \otimes I) - (F \otimes F)^{k+1} \right) \quad (5.21)
\]

\[
((I \otimes I) - (F \otimes F))^{-1} (K \otimes K) \quad (5.22)
\]

\[
\text{vec} (H (FPF + Q) H' + R^{-1}) \quad (5.23)
\]

which is bounded only if \((I \otimes I) - (F \otimes F)) > 0\) i.e. \(F\) has eigenvalues \(< 1\). Consequently sending only innovations cannot work for most systems of interest if innovations may be lost as the variance of the expected error is unbounded. This is illustrated in Figure 5.1 as the innovation only sequence rapidly diverges away from the actual state.

Clearly we need to transmit additional information to prevent this divergence. Approaches that preserve bit-identical forward error correction can be used such as Fountain Codes combined with continuous retransmission\(^4\) or Forward Error Correction\(^5\) where the transmission is padded with additional redundant information in order to allow a late joiner to receive missing samples or recover from lost samples. These approaches do not take advantage of the approximate nature of the innovations. The transmitted innovation is only the symbol that provides the best estimate of the MMSE innovation that matches the sensor’s observation. It is not necessary to provide an exact re-transmission, it is only necessary that over time the sensor’s state estimate and the receiver’s state estimate converge “quickly” enough. This is particularly important as the state of the system evolves. It is not strictly necessary to know all innovations over a period if it is possible to recover the current state sufficiently accurately such that the application of further received innovations converge to the actual state.

Consider a system where we transmit a modified innovation sequence where \(\tilde{y}_k^+\) is

---

\(^4\)Such as Raptor and LT codes used in DVB-IPTV applications.

\(^5\)such as Block Codes and Punched Codes
5. Efficient transmission of measurements

the modified innovation and $\zeta \in (0, 1]$

$$\tilde{y}_k^\dagger = z_k - \zeta HF x_{k-1|k-1}$$  \hspace{1cm} (5.25)

$$\tilde{y}_k = \tilde{y}_k^\dagger - (1 - \zeta) HF x_{k-1|k-1}$$  \hspace{1cm} (5.26)

At a receiver we can reconstruct

$$x^{R}_{k|k} = Fx^{R}_{k-1|k-1} + K_k \left( \tilde{y}_k^\dagger - (1 - \zeta) HF x^{R}_{k-1|k-1} \right)$$  \hspace{1cm} (5.27)

$$= (I - K_k (1 - \zeta) H) Fx^{R}_{k-1|k-1} + K_k \tilde{y}_k^\dagger$$  \hspace{1cm} (5.28)

Effectively we transmit not only the innovation but a small amount of information about $x_{k-1}$ as well. If we choose $\zeta$ s.t.

$$\lambda_{\max} \left( (I - (1 - \zeta) K_k H) F \right) < 1$$  \hspace{1cm} (5.29)

then any error introduced through a missed innovation will decay.

$$e^\dagger_{k+n} = ((I - K_k (1 - \zeta) H) F)^n K_k \tilde{y}_k^\dagger$$  \hspace{1cm} (5.30)

Note that

$$E \left( \tilde{y}_k^\dagger \right) = (1 - \zeta) HF x_{k-1|k-1}$$  \hspace{1cm} (5.31)

and

$$\text{var} \left( \tilde{y}_k^\dagger \right) = (\zeta^2 H (FP_k F + Q) H' + R^{-1})$$  \hspace{1cm} (5.32)

Consequently

$$E \left( e^\dagger_1 | z_{1:k} \right) = \sum_{n=0}^{k} (1 - \alpha) F^n K_{k-n} E \left( \tilde{y}_{k-n}^\dagger \right)$$  \hspace{1cm} (5.33)

$$> (1 - \alpha) (1 - \zeta) K \left( I - ((I - K (1 - \zeta) H) F)^{k+1} \right) \bullet$$  \hspace{1cm} (5.34)

$$$$(I - (I - K (1 - \zeta) H) F)^{-1} HF E(x)$$  \hspace{1cm} (5.35)$$

and

$$\text{vec var} \left( e^\dagger_1 | y_{1:k} \right) > \begin{pmatrix} (I \otimes I) \\ - ((I - K_k (1 - \zeta) H) F) \otimes ((I - K_k (1 - \zeta) H) F)^{k+1} \end{pmatrix} \bullet$$  \hspace{1cm} (5.36)

$$\begin{pmatrix} (I \otimes I) \\ - ((I - K_k (1 - \zeta) H) F) \otimes ((I - K_k (1 - \zeta) H) F) \end{pmatrix}^{-1} \bullet$$  \hspace{1cm} (5.37)

$$(1 - \alpha) (K \otimes K) \text{vec} (\zeta^2 H (FPF + Q) H' + R^{-1})$$  \hspace{1cm} (5.38)
Consequently there is a systematic error which depends on the mean state estimate, and the lower bound of the variance of the systematic error is now bound. By selection of $\zeta$ we can modulate the transmission system between the pure transmission of observations and the pure transmission of innovations. At $\zeta = 1$ the transmission system is transmitting innovations and losses of transmissions will potentially cause divergence. As $\zeta$ is reduced some prior state information is “mixed” in, until if (5.29) is satisfied the system no longer diverges. Finally notice that the expected error caused by missed observations is given by (5.33) diminishes as $\zeta \to 0$.

5.4 Encoding and Decoding

The approach outlined in eq. (5.29) approach may be extended to include an encoding and decoding process. Observe that a small error in encoding will continue to grow without bound as a consequence of eq. (5.14). We define a local state estimate that is maintained at the sensor using only local observations as

$$x_{k|k}^l = (I - K_kH)Fx_{k-1|k-1}^l + K_kz_k$$ (5.39)

where $K_k$ is the Kalman gain. Consider the system in Figure 5.2. Define $\tilde{y}_k^l = D(c_k)$ and $c_k = E(\tilde{y}_k^l)$ as the decode and encode operators where $D(E(\tilde{y}_k^l)) - \tilde{y}_k^l$ is the coding error. Define $\mathcal{R} \subset \mathbb{R}$ as the subset of the state space which contains acceptable, transmittable states and $D(c_k) = \{d_1, ..., d_n\}$ are the decoded values of codewords $c_1, ..., c_n$ The encoding operator selects a symbol from the code book such that the encoding error is minimized. We define the receiver’s state estimate as our
local copy of the receiver’s state based on the encoded innovations transmitted to it

\[ x_{k|k}^r = Fx_{k-1|k-1}^r + KD(c_k) \]  (5.40)

Note that the receiver uses the Kalman gain at the limit - the receiver does not want/need to know how many previous transmissions have been made (consider the difficulties for a late joiner) and consequently we do not transmit the Kalman gain or estimator covariance. We want \( x^r \), the receiver state estimate to remain as close as possible to \( x^l \), our local state estimate, i.e. we want to select codewords \( c_k \) s.t.

\[
\min_{c_k} \| \epsilon_k \| = \left\| x_{k|k}^r - x_{k|k}^l \right\| 
\]

\[
= \min_{c_k} \left\| Fx_{k-1|k-1}^r + KD(c_k) - (I - K_k H) Fx_{k-1|k-1}^l - K_k z_k \right\| 
\]

\[
= \min_{c_k} \left\| F(\epsilon_{k-1} + x_{k-1|k-1}^r) + KD(c_k) - (I - K_k H) Fx_{k-1|k-1}^l - K_k z_k \right\| 
\]

\[
= \min_{c_k} \left\| \epsilon_{k-1} + K_k \left( HFX_{k-1|k-1}^l - z_k \right) + KD(c_k) \right\| 
\]

\[
= \min_{c_k} \left\| F\epsilon_{k-1} - K_k \bar{y}_k + KD(c_k) \right\| 
\]

The selected codeword consequently includes information about the new observation, plus cancels out the error introduced by the finite precision of transmitting the previous codeword.

We can use the result from (5.29) and (5.33) to introduce resilience to message loss. We proceed as above but we introduce \( \zeta \) and define

\[ x_{k|k}^r = (I - K (1 - \zeta) H) Fx_{k-1|k-1}^r + KD(c_k) \]  (5.46)

As before, the receiver uses the Kalman gain at the limit. Once again we want \( x^r \), the receiver state estimate to remain as close as possible to \( x^l \), our local state estimate, i.e. we want to select codewords \( c_k \) s.t.

\[
\min_{c_k} \| \epsilon_k \| = \left\| x_{k|k}^r - x_{k|k}^l \right\| 
\]

\[
= \min_{c_k} \left\| (I - K (1 - \zeta) H) Fx_{k-1|k-1}^r + KD(c_k) - (I - K_k H) Fx_{k-1|k-1}^l - K_k z_k \right\| 
\]

\[
= \min_{c_k} \left\| (I - K (1 - \zeta) H) F(\epsilon_{k-1} + x_{k-1|k-1}^r) + KD(c_k) - (I - K_k H) Fx_{k-1|k-1}^l - K_k z_k \right\| 
\]

\[
= \min_{c_k} \left\| (I - K (1 - \zeta) H) F\epsilon_{k-1} + (K_k - K (1 - \zeta)) HFX_{k-1|k-1}^l + KD(c_k) - K_k z_k \right\| 
\]

\[
= \min_{c_k} \left\| (I - K (1 - \zeta) H) F\epsilon_{k-1} - \left( K_k \bar{y}_k + (1 - \zeta) HFX_{k-1|k-1}^l \right) + KD(c_k) \right\| 
\]

\[
(5.49)
\]

\[
(5.50)
\]

\[
(5.51)
\]
In figure 5.1 we see that coarse quantisation using only an 8 bit quantiser can still give good results if no innovations are lost. Errors introduced in prior transmissions are cancelled in subsequent transmissions by selection of an appropriate $c_k$.

For a stable (in the mean) system we require for all $x_k^r \in \mathcal{R}$ and $\|\epsilon\| = \|\epsilon_{\text{max}}\|$ to satisfy

$$E_{\hat{y}_k} \left( \min_{c_k} \| (I - K (1 - \zeta) H) F \epsilon - (K_k \hat{y}_k + K (1 - \zeta) HF x_{k-1|k-1}^l) + K D (c_k) \| \right) < \| \epsilon \|$$  \hspace{1cm} (5.52)

i.e. we expect to reduce the error over all possible observations for all $x$ inside the acceptable state space. Note that $E (\hat{y}_k) = 0$. If we write $D (c_k)$ as a function of the maximum state variables and introduce a modified decoder $D^\# (c_k)$ s.t.

$$D (c_k) = (1 - \zeta) HF D^\# (c_k) : \zeta \neq 1$$  \hspace{1cm} (5.53)

then we can write the requirement as

$$\min_{c_k} \| F \epsilon - (1 - \zeta) KHF \left( \epsilon + x_{k-1|k-1}^l - D^\# (c_k) \right) \| < \| \epsilon \|$$  \hspace{1cm} (5.54)

Note that provided that there exists $c_k$ s.t. for all $x^l$ in $\mathcal{R}$ that $\| x^l \| < \| D^\# (c_k) \|$ then (5.54) is true provided

$$\| F \epsilon - (1 - \zeta) KHF \epsilon \| < \| \epsilon \|$$  \hspace{1cm} (5.55)

is true. Note that $\zeta$ gives the rate at which error decays (i.e. if $\zeta$ is close to one we are sending mostly the innovation, while as $\zeta$ becomes close to 0 we are sending mostly measurements). $\zeta$ is selected so that the maximum eigenvector of $(F - (1 - \zeta) KHF)$ is less than 1.

At the receiver the state estimate is

$$x^r_{k|k} = x^l_{k|k} + \epsilon_k.$$  \hspace{1cm} (5.56)

$$\text{var } x^r_{k|k} = P^r_{k|k} = \text{var } (x^l_{k|k} + \epsilon_k) = P^l_{k|k} + \text{var } \epsilon_k.$$  \hspace{1cm} (5.57)

From (5.47) we have

$$\epsilon_k = (I - K (1 - \zeta) H) F \epsilon_{k-1} - \left( K_k \hat{y}_k + K (1 - \zeta) HF x_{k-1|k-1}^l \right) + K D (c_k)$$  \hspace{1cm} (5.58)

Note that we pick $c_k$ s.t. $\epsilon_k$ is minimized. Consequently if we assume that the encoder is designed s.t. for all $x \in \mathcal{R}$ uniformly distributed in the operational state space

$$E (D (\mathcal{E} (x)) - x) = E (\cdot) = 0$$

$$\text{var } (D (\mathcal{E} (x)) - x) = \varepsilon \varepsilon^t$$
5. Efficient transmission of measurements

then we select \( c_k = \mathcal{E}(w) \), \( w \) s.t.

\[
0 = (I - K (1 - \zeta) H) F\epsilon_{k-1} - \left( K_k y_k + K (1 - \zeta) H F x_{k-1|k-1} \right) + K w \tag{5.59}
\]

\[
0 = (I - K (1 - \zeta) H) F\epsilon_{k-1} - \left( K_k y_k + K (1 - \zeta) H F x_{k-1|k-1} \right) + K (D (\mathcal{E}(w)) + \Delta) \tag{5.60}
\]

\[
0 = \epsilon_k + K \Delta \tag{5.61}
\]

\[
\mathbb{E}(\epsilon_k) = -K \mathbb{E}(\Delta) = 0 \tag{5.62}
\]

\[
\text{var}(\epsilon_k) = \text{var}(K \Delta) \tag{5.63}
\]

\[
= K \varepsilon \varepsilon' K' \tag{5.64}
\]

Note that \( \text{var}(\epsilon_k) \) depends on \( K \) and the encoder/decoder properties. We have consequently demonstrated that it is possible to build an encoder decoder pair that will minimise the error of the recovered state estimate for a particular set of system parameters and \( \zeta \) and not cause the estimate to diverge.

There are several advantages to this method when compared to other error correction processes.

- **Concealment for the information consumer** - The consumer of the information may not wish to disclose that they are using the information or broadcast their existence through a shared or observable transmission media.

- **Improved information firewalling** - In highly secure systems providing proof that the recipient is only transmitting re-transmission requests and cannot be suborned into disclosing other information is difficult to obtain. Systems that avoid this requirement (by providing only unidirectional data links) are simpler to certify and deploy.

- **Elimination of a class of denial of service attacks** - Systems that include acknowledgment / re-transmission requests have to be carefully designed to ensure that those mechanisms cannot be perverted such that forcing re-transmission requests or responding to re-transmission requests does not degrade the systems performance. Even when not maliciously generated re-transmissions can cause system degradation during periods of congestion such that the re-transmission causes even further congestion.

- **Allows multiple recipients without overhead** - As all information is made available over time to all recipients without the transmitter being aware of the number or even the state of the recipients, the system can easily scale.
Support of Late Joiners - A further use for this technique is to support “late joiners”, that is receivers that join the network at some point in time after the observations begin. With a strict innovation based system the sensors would need to be notified of the addition of another receiver and “replay” all the innovations in order to synchronize the transmitter and receiver together. If $\zeta < 1$ is used the late joiner will simply have a large initial error which will converge over time to the same error as all other receivers.

There are significant advantages to supporting late joiners. It allows the sensors and receivers to be completely autonomous of each other, with no requirement for a back channel for specific error control. The receiver does not require a protocol to monitor late joiners and provide them with additional data to bring them into the same state as existing receivers. The receiver will eventually come fully up to date.

![Figure 5.3: Performance of late joiner after missing first 10 samples - error scale is log10 MSE](image)

Consider again the system defined in (5.1). We construct the encoder and decoder according to 5.25 and ??, setting $\zeta$ s.t. 5.29 is satisfied. As we are transmitting
more information than the minimum required the estimation will converge even in the presence of missing observations and a late joining observer. Figure 5.3 demonstrates the behaviour of a late joining receiver for $\zeta = .4, .7$ and $.9$ for $\alpha = .5$. Note that after missing the first 10 samples and while missing half of the samples that the three systems converge to comparable values with 30 observations. Systems operating in this manner do not require the receiver of the state information to transmit to the sensor at all. Furthermore, although this discussion has been couched as a single sensor transmitting information to a single receiver, this can be used as a single link in a multi-sensor system provided the sensors are co-ordinated such that the same observation is not used multiple times.

5.5 Rescheduling Sensor Transmission.

It is convenient for the transmission process and the sensing process to be decoupled in the sensor. Sensor scheduling may be ad hoc, with sensor reports occurring at intervals that are unrelated to the underlying communications transmission system and the transmission of sensor reports may impact on the ability of the sensor communications system to transmit sensor reports from other elements. Furthermore sensors may not be able to generate a sensor report because of missed detections or difficulties with correlating sensor observations with system states. Consequently the quality of the state estimate at a sensor can vary dramatically over time, sometimes increasing, sometimes decreasing. A question to ask is: should a sensor transmit a single (potentially long) sensor report? Is it better to transmit a larger number of shorter less precise reports, each report potentially correcting some part of the encoding error of a previous report and contributing a small amount of additional information? We consider transmitting sensor reports assuming that there is no encoding/decoding error.

Using (3.6) and (3.11) we can write the limit of the covariance of a continuous system (3.2) sampled discretely as the solution of the recursion

$$
\text{vec } P = e^A \otimes e^A \text{vec } \left( P^{-1} + H' R^{-1} H \right)^{-1} + (A \otimes I + I \otimes A)^{-1} \left( (e \otimes e)^A - (I \otimes I) \right) \text{vec } \chi
$$

(5.65)

$$
+ (A \otimes I + I \otimes A)^{-1} \left( (e \otimes e)^A - (I \otimes I) \right) \text{vec } \chi
$$

(5.66)

If we consider taking $j$ times the number of observations, and require that after taking $j$ less precise observations that the state estimate have the same covariance we can write a new recursion with $\Xi$ s.t.

$$
\text{vec } P = e^{A/j} \otimes e^{A/j} \text{vec } \left( P^{-1} + \Xi^* \right)^{-1} + (A \otimes I + I \otimes A)^{-1} \left( (e \otimes e)^{A/j} - (I \otimes I) \right) \text{vec } \chi
$$

(5.67)

$$
+ (A \otimes I + I \otimes A)^{-1} \left( (e \otimes e)^{A/j} - (I \otimes I) \right) \text{vec } \chi
$$

(5.68)
5. Efficient transmission of measurements

where $P$ is the solution of (5.65). Rearranging

\[ e^{A(1-1/j)} \otimes e^{A(1-1/j)} \left( \text{vec} \left( P^{-1} + H' R^{-1} H \right)^{-1} + (A \otimes I + I \otimes A)^{-1} \text{vec} \chi \right) \]  
\[ = \text{vec} \left( P^{-1} + \Xi^+ \right)^{-1} + (A \otimes I + I \otimes A)^{-1} \text{vec} \chi \] (5.69) (5.70)

where

\[ (A \otimes I + I \otimes A) \text{vec} Q^+ = \text{vec} \chi \] (5.71)

we have

\[ \Xi^+ = \left( e^{A(1-1/j)} \left( (P^{-1} + H' R^{-1} H)^{-1} + Q^+ \right) e^{A(1-1/j)'} - Q^+ \right)^{-1} - P^{-1} \] (5.72)

i.e. $\Xi$ is the equivalent covariance of observations that take place more rapidly (i.e. we taken $k$ less accurate observations rather than a single observation). Note that $\Xi$ may not be full rank, however if the system $(F,H)$ is observable then the system $(F,\sqrt{\Xi})$ must also be. Consequently we could transmit a larger number of innovations based on observations with covariance $\Xi$ and have them converge to the same covariance and state estimate as transmitting a single innovation based on an observation with covariance $R$. Finally note that as $k \to \infty$ that $\Xi^+ \to 0$. i.e. each individual observation need be less and less accurate.

We can write an equivalent remote system

\[ P^r = e^{A/j} \left( (P^r)^{-1} + \Xi \right)^{-1} e^{A/j} + Q (j) \] (5.73)
\[ \Xi = e^{A/j} (P^r - Q (j))^{-1} e^{A/j} - (P^r)^{-1} \] (5.74)
\[ \text{vec} Q (j) = (A \otimes I + I \otimes A)^{-1} \left( (e \otimes e)^{A/j} - (I \otimes I) \right) \text{vec} q \] (5.75)

From (5.72) we know that if we set

\[ \Xi = \left( e^{A(1-1/j)} \left( \left( (P^l)^{-1} + H' R^{-1} H \right)^{-1} + Q^l \right) e^{A(1-1/j)'} - Q^l \right)^{-1} \] (5.76)
\[ - \left( P^l \right)^{-1} \] (5.77)

then the system is equivalent to the original system with samples being transmitted at a $j$ higher rate.

The implication is that any single observation made with a precision described by $H' R^{-1} H$ can be converted into a sequence of observations sent with lesser precision $\Xi$. If we are sending quantised innovations as in 5.4 we need devote fewer bits to encoding the innovation.
5.6 Innovation-like Sequences

Using innovations to transmit new information allows minimal information to be sent, but it requires that both transmitter and receiver have access to all prior observations, either directly or as a summary. This requires the transmitter to know exactly which messages were received and the final state estimate as calculated at the receiver. This is not always possible and an alternative would be to provide a method that does not require all prior information but rather a finite set of prior information. Consequently it is desirable to transmit observations that are very similar in characteristics to the innovations but satisfy two specific properties:

1. They do not require complete knowledge of all prior observations, rather only a finite set of prior state.
2. They should admit expression as weighted observations suitable for use in an information fuser.
3. The transmission should be calculated using the state estimate generated at the sensor rather than from the raw observations.

In this section we will demonstrate that the transmission of a sequence of innovations can be considered to the same as transmitting a sequence of observations for an augmented system. We are motivated to do this as if we can demonstrate the equivalence then all results pertaining to the fusion, transmission and processing of linear systems become applicable to systems transmitting modified innovations. Consider a sensor which generates a set of state estimators

\[
\begin{align*}
\{\hat{x}_{k|k}, P_{k|k}\}, \{\hat{x}_{k-1|k-1}, P_{k-1|k-1}\}, \cdots, \{\hat{x}_0, 0\}
\end{align*}
\]  

(5.78)

of a state augmented system with

\[
x_{k+1} = \begin{bmatrix} x_{k+1} \\ x_k \end{bmatrix}
\]

The actual measurements and system process are not germane, the sensor could be using non-linear measurements to generate the state estimates. The state estimates can be transformed into the equivalent sequence of observations \(\{z_k, R_k\}\) that would have generated the state estimates given a set of well known process parameters agreed upon between the sensor and a receiver. Furthermore it is convenient as will be seen in Chapter 6 to transmit \(\{R_k^{-1}z_k, R_k^{-1}\}\) for fusion of the measurements.

Consider an agreed upon state augmented system where

\[
F \in R^{n \times n} : F^{-1} \text{ exists}
\]  

(5.79)
5. Efficient transmission of measurements

\[ F = \begin{bmatrix} F & 0 \\ I & 0 \end{bmatrix} \]

\[ x_{k+1} = \begin{bmatrix} x_{k+1} \\ x_k \end{bmatrix} = F \begin{bmatrix} x_k \\ x_{k-1} \end{bmatrix} + \begin{bmatrix} I \\ 0 \end{bmatrix} u_k \]  \hspace{1cm} (5.80)

\[ u_k \sim N (0, Q) \]  \hspace{1cm} (5.82)

\( m \in \mathbb{Z}^+ \) is selected s.t. it satisfies both the stabilisability (5.83) and detectability (5.84) requirements of a non-augmented system \( \{ F, Q \} \) and \( \{ F, H \} \)

\[ \text{rank} \left( \begin{bmatrix} I & F & \cdots & F^{m-1} \end{bmatrix} \text{diag} (Q) \right) = n \]  \hspace{1cm} (5.83)

\[ \text{rank} \left( \begin{bmatrix} I \\ F \\ \vdots \\ F^{m-1} \end{bmatrix} \text{diag} (H') \right) = n \]  \hspace{1cm} (5.84)

The transmission system operates equivalently to 5.2, with the local sensor keeping track of the transmission errors while the receiver uses the received sequence to reconstruct the state estimate.

We define pseudo observations of the system as

\[ z_k = Hx_k + v_k \]  \hspace{1cm} (5.85)

and define the transmission sequence the sequence

\[ \{ R_k^{-1} z_k, R_k^{-1} \}, \{ R_{k-1}^{-1} z_{k-1}, R_{k-1}^{-1} \}, \ldots, \{ R_0^{-1} z_0, R_0^{-1} \} \]  \hspace{1cm} (5.86)

as the set of scaled equivalent measurements that would produce the sequence of state estimators at the receiver that matched

\[ \left\{ \tilde{x}_{k|k}, P_{k|k}^{-1} \right\}, \left\{ \tilde{x}_{k-1|k-1}, P_{k-1|k-1}^{-1} \right\}, \ldots, \left\{ \tilde{x}_0|0, P_0^{-1} \right\} \]  \hspace{1cm} (5.87)

Note that we are free to choose \( H, F \) and \( Q \), provided both transmitter and receiver are agreed. As was previously noted the transmission of innovations helps reduce the redundancy in the sequence, but pure innovations would not function in the presence of losses. Consequently we generate an innovation like sequence by defining the observation function as

\[ H = H \begin{bmatrix} I & -\zeta F \end{bmatrix} \]  \hspace{1cm} (5.88)

which makes \( z_k \) look like an innovation - its a function of the difference between \( x_k \) and \( x_{k-1} \).
Note that the system $\{H, F\}$ is detectable provided $\zeta \neq 1$ (see Theorem 7), consequently we can write for known $\hat{x}_{k|k}, P_{k|k}$

$$z_k = HS_{k|k}$$

$$\mathcal{I}_k = P_{k|k}^{-1} - P_{k|k-1}^{-1}$$

where $\mathcal{I}_k$ is the incremental information provided by the $k$th observation. At the receiver $P_{k|k-1}^{-1}$ is unknown, only the value derived from the previous recovered

$$\left(P_{k-1|k-1}^R\right)^{-1}$$

is known. The transmitter must use the value $\left(P_{k-1|k-1}^R\right)^{-1}$ in its calculations to correct for any error in the encoding of previous transmitted values. At the transmitter we calculate

$$P_{k|k-1}^R = FP_{k|k-1}^R F^t + Q$$

$$\mathcal{I}_k = P_{k|k}^{-1} - \left(P_{k|k-1}^R\right)^{-1}$$

Furthermore $H$ is not invertible. Note that $HH^t = I$, not $H^tH$ (which is a projection onto the range of $H$) so

$$\left(R_k^R\right)^{-1} = (H^t)^t \mathcal{I} H^t$$

$$= (H^t)^t \left(P_{k|k}^{-1} - \left(P_{k|k-1}^R\right)^{-1}\right) H^t$$

and

$$\left(R_k^R\right)^{-1} z_k^R = \left(R_k^R\right)^{-1} H\hat{x}_{k|k}$$

$$= (H^t)^t \left(P_{k|k}^{-1} - \left(P_{k|k-1}^R\right)^{-1}\right) H^t H\hat{x}_{k|k}$$

The receiver will receive $R_{k-1}^{-1}z_k$ and $R_k^{-1}$ from which it derives $\left(P_{k|k}^R\right)^{-1}$ and $\left(P^{-1}\hat{x}\right)_k^R$ using

$$\left(P_{k|k}^R\right)^{-1} = \left(P_{k|k-1}^R\right)^{-1} + H^t \left(R_k^R\right)^{-1} H$$

$$\left(P^{-1}\hat{x}\right)_k^R = \left(P_{k|k}^R\right)^{-1} \left(P^{-1}\hat{x}\right)_{k|k-1}^R + H^t \left(R_k^R\right)^{-1} z_k^R$$

The prior is formed by using

$$\left(P_{k|k-1}^R\right)^{-1} = \left(FP_{k-1|k-1}^R F + Q\right)^{-1}$$
5. Efficient transmission of measurements

\[(P^{-1}\hat{x})^R_{k|k-1} = (P^R_{k|k-1})^{-1} F P^R_{k-1|k-1} (P^{-1}\hat{x})^R_{k-1|k-1} \] (5.101)

Note that for

\[
P = \begin{bmatrix}
P_1 & P_{12} \\
P_{12}' & P_{22}
\end{bmatrix}
\] (5.102)

\[
P^{-1} = \begin{bmatrix}
A & B \\
B' & D
\end{bmatrix}
\] (5.103)

\[
(P^{-1}x) = \begin{bmatrix}
(P^{-1}x)_1 \\
(P^{-1}x)_2
\end{bmatrix}
\] (5.104)

and \( F \) and \( Q \) as described in (5.80) and (5.82) that

\[
(FP F + Q)^{-1} = \begin{bmatrix}
Q^{-1} & -Q^{-1}F \\
-F'Q^{-1} & P_1^{-1} + F'Q^{-1}F
\end{bmatrix}
\] (5.105)

\[
P_1^{-1} = A - BD^{-1}B'
\] (5.106)

and

\[
(FP F + Q)^{-1} FP (P^{-1}x) = \begin{bmatrix}
0 & 0 \\
I & P_1^{-1}P_{12}
\end{bmatrix} \begin{bmatrix}
(P^{-1}x)_1 \\
(P^{-1}x)_2
\end{bmatrix}
\]

\[
= \begin{bmatrix}
(P^{-1}x)_1 - BD^{-1}(P^{-1}x)_2 \\
0
\end{bmatrix}
\] (5.108)

allowing the prior to be calculated recursively in terms of the blocks of \( P^{-1} \) and \( (P^{-1}x) \).

Note that the differences between the transmitter and receiver decay. Let \( \varepsilon_k \) the difference between the state estimates, while \( \epsilon_k \) is the difference in the information matrices

\[
\varepsilon_k = (\hat{x}^R_{k|k} - \hat{x}_{k|k})
\] (5.109)

\[
= (P^R_{k|k}) (P^R_{k|k-1})^{-1} \hat{x}^R_{k|k-1} = (P^R_{k|k}) H' (R^R_{k})^{-1} z_k - \hat{x}_{k|k}
\] (5.110)

\[
= (P^R_{k|k}) (P^R_{k|k-1})^{-1} \hat{x}^R_{k|k-1} + (P^R_{k|k}) H' (R^R_{k})^{-1} H \hat{x}_{k|k} - \hat{x}_{k|k}
\] (5.111)

\[
= (P^R_{k|k}) (P^R_{k|k-1})^{-1} \hat{x}^R_{k|k-1} + (P^R_{k|k}) H' (R^R_{k})^{-1} H - (P^R_{k|k})^{-1} \hat{x}_{k|k}
\] (5.112)

\[
= (P^R_{k|k}) (P^R_{k|k-1})^{-1} (\hat{x}^R_{k|k-1} - \hat{x}_{k|k})
\] (5.113)

\[
= (P^R_{k|k}) (P^R_{k|k-1})^{-1} (F \hat{x}^R_{k-1|k-1} - F \hat{x}_{k-1|k-1} + u_k)
\] (5.114)

\[
= (P^R_{k|k}) (P^R_{k|k-1})^{-1} (F \varepsilon_{k-1} + u_k)
\] (5.115)
The position error includes a decaying term from the previous error plus a new term introduced by the actual change in system state. The actual error in the covariance is purely caused by the mismatch between the actual state estimator and the transmission system only transmitting a part of the state at each transmission. It does not grow over time as it is simply the projection of $\text{vec}I_k$ onto the nullspace of $H' \otimes H'$.

$$
\epsilon_k = P_{k|k}^{-1} - \left(P_{k|k}^R\right)^{-1}
= \left((P_{k|k-1}^R)^{-1} + I_k\right) - \left((P_{k|k-1}^R)^{-1} + H'(R_k^R)^{-1}H\right)
= I_k - H'(R_k^R)^{-1}H
= \left(I \otimes I - (H^+H)' \otimes (H^+H)^'\right) \text{vec}I_k
$$

We have demonstrated that the augmented system, transmitting innovations that are scaled by $\zeta$ is equivalent to a non-augmented system transmitting measurements provided $\zeta \neq 1$. Consequently we can apply all techniques that would be available to fuse measurements to form the linear unbiased estimator for the non-augmented system. These relationships will be further developed in the next chapter.

### 5.7 Conclusion

This chapter has developed the concept of an innovation-like sequence. We were initially motivated by the observations of ([18]) that although a sequence of innovations requires the least amount of information to be transmitted, lost innovations would cause unbounded errors. Using simulation we demonstrate that missed innovations cause unbounded error for an unstable system (one that has eigenvalues $> 1$).

In 5.3 we outline the initial concept and demonstrate that we can maintain convergence provided “enough” measurement information is mixed into the innovation sequence and show what the convergence criteria are. This mixing is performed according to 5.25 and 5.27. By taking a small linear term we add in a fraction of position information into an innovation sequence. Using this mixture of position and innovation information we demonstrate that this mitigates the loss of messages from a sequence of innovations and calculate the expected error.

In 5.4 we extend this further by demonstrating that there exists encoding/decoding strategies that maintain bounded error. In particular we note that 5.54 gives us a requirement on how to build the encoder. It states that the system will be stable provided the encoder can encode the full space of all possible values, i.e. that it must not saturate. We also demonstrate that the system allows “late joiners”, observers who have missed preceding messages to all converge to the same state in time, and
that the rate at which they converge is a function of the amount of measurement data added to the innovations. We show that even for an unstable system any introduced quantisation errors will be bounded and that a suitable quantiser can be constructed.

In 5.5 we demonstrate that we can always break down a system that has a single sensor into a system that uses many more less accurate sensors taking evenly spaced observations and that the systems are equivalent.

In 5.6 it is demonstrated that we can re-arrange our innovation-like sequence into an equivalent series of measurements of a state augmented system, consequently demonstrating the equivalence of the transmission of measurement sequences and modified innovation sequences. This is important as given equivalence all the previous results regarding intermittently observable systems and the fusing of results from multiple sensors are applicable to the new class of systems using pseudo-innovations. We will take advantage of this in the next chapter.

In combination this section demonstrates that for all linear observable systems we can always re-arrange the system to send pseudo innovations. Transmitting pseudo-innovations will always lead to estimators that are bounded when messages are lost and the requirements for bounding are the same as systems that send measurements. The rate at which the state estimator converges is dependent on the amount of measurement information transmitted.
Chapter 6

Co-operative Transmission

6.1 Overview

In this section we develop the concept of co-operative transmission and demonstrate that it can achieve power efficiency greater than optimal joint encoding. 6.2 introduces the problem and demonstrates the difficulties that confront a swarm of RF transmitters. 6.3 investigates the properties of joint encoding and demonstrates that joint encoding must implicitly or explicitly retain sufficient information to recover the measurement of any individual sensor. 6.4 develops the concept of co-operative transmission as a means to remove superfluous information from joint encoding and to take advantage of the co-interference identified in 6.2. To illustrate the concepts of Co-operative Transmission in this section we will alternate the development of theory with a practical examples. As the theory is developed in 6.4 and 6.6 practical examples and simulations are presented in 6.5 and 6.7. 6.6 extends the concept, addressing limitations. In 6.9 all the previous bought in and to provide a practical implementation that can provide extremely efficient spectral usage suitable for a large sensor swarm. Simulation results are presented in 6.10.

6.2 Introduction

In the previous chapter it was noted that the establishment of actual data links between nodes has some cost associated with it. The analysis of distributed systems usually starts with the assumption that given a set of point to point links some form of application protocol will need to be added to correctly fuse the sensor data. Solutions typically include some form of routing protocol incorporating neighbour discovery, formation of spanning trees rooted at the information consumer and incremental fusion of observations to reduce the amount of traffic forwarded to the information consumer.
6. Co-operative Transmission

Unfortunately this approach neglects the not inconsiderable problem of how to establish and maintain the point to point connectivity in the first place. The most important factor in the design of widely distributed radio network is self-interference.

Each transmitter in a radio network can potentially interfere with every other receiver. The design of cellular networks is a careful exercise in optimization of transmitter power and frequency to minimise this interference for a given topology and utilisation. This can be done for cellular networks largely due to the static nature of the basestations and the asymmetric design - base stations communicate with mobile units and mobiles with base stations, mobiles never talk to other mobiles. This asymmetry is exploited for protocols such as the 802.11 family (Wi-Fi) and the mobile telephony 3GPP consortium standards to achieve high data rates between subscribers and base stations. Unfortunately landmobile radio systems without stationary base stations cannot achieve the same capacities. Older digital radio technologies such as P-25 and TETRA can achieve rates in the 10’s of kbps, while mobile tactical UHF radio systems using mesh routing (such as the Harris AN/PRC117G family) can achieve data rates up to the megabit range only over a multi-node meshed network. The problems these systems experience are due to the difficulty of co-ordinating frequency and transmit power between a large number of mobile radios. As the radios move the RF environment changes, the set of neighbouring radios changes and consequently the optimum power and frequency plan is continuously changing.

An illustration of this can be seen by considering the following simplified problem. Given a randomly distributed set of transmitters distributed over 3 dimensions in a volume of free-space, what is the power detected by a receiver located within the transmitter cloud given a sensor density $\rho$ and transmitter power $P_{tx}$ for all transmitters located closer than $r_{max}$ and further than $r_{min}$ from the receiver.

$$P_{total} = \int_{r_{min}}^{r_{max}} \rho \frac{4\pi r^2}{r^2} P_{tx} dr$$

$$= \rho P_{tx} \frac{4\pi}{3} (r_{max} - r_{min})$$

Eq. (6.2) indicates that even for the simple case of free space propagation a receiver located within a volume of transmitters will experience an increasing level of interference as the sensor cloud increases in volume for a constant value of $P_{tx}$. This reveals the problem, even if the transmitted power is reduced to compensate for an increased density of sensors, as the sensor “cloud” increases in size the amount of interference generated still increases without bound.

For the case of a sensor system located on a surface in an urban environment the propagation model is closer to $1/\mu^{2.7}$ (the actual exponent is dependant on the
geometry and is only an approximation - but it is not the same as the free space exponent). In this case assuming transmitters randomly arranged in a circular region we have

\[ P_{\text{total}} = \int_{r_{\text{min}}}^{r_{\text{max}}} \int \rho \frac{2\pi r}{\rho r_{\text{min}}^{0.7}} P_{\text{tx}} dr \]

\[ = 1.4\pi \rho P_{\text{tx}} \left( \frac{1}{r_{\text{min}}^{0.7}} - \frac{1}{r_{\text{max}}^{0.7}} \right) \]

Eq. (6.5) is slightly better as the received noise power is bounded. The system still has to be configured such that the “nearest” transmitters are non-interfering, usually be allocating a different channel or operating frequency that the receiver is insensitive to - similar to dynamically allocating the cells of a cellular radio network. This requires solving the “four colour” problem repeatedly over a time varying distributed system and distributing this information throughout the network, a not inconsiderable problem in its own right - given that the distribution mechanism is precisely what is being changed dynamically.

We note that a digital radio mesh of landmobile transmitters requires sophisticated frequency and power management when used in a 2 dimensional environment to prevent transmitters causing to much interference for the receivers to operate and form node to node links. Without these node to node links no other application routing protocol can operate. In the 3 dimension environment using free-space propagation we see that the amount of interference is unbounded - there is no way to grow the network without increasing bandwidth (providing additional non-interfering channels).

### 6.3 Joint Encoding

In order to solve the problem of unbounded interference we will first consider the nature of what needs to be transmitted, and how efficiently it can be transmitted. Consider a system where the sensors are observing a single state variable distorted by observation noise. The consumer of the data is uninterested in the individual observations but rather only requires an estimate of the state variable. This problem is often examined in the context of the design of a digital system to transmit a “hidden” or unobservable variable which is corrupted by Gaussian noise and is referred to as the CEO problem. (see ??) For the case of a Gaussian r.v. corrupted by Gaussian noise an exact analytic result for the rate distortion is known, but general analytic solutions are lacking. Furthermore the reception and decoding of jointly encoded systems is
6. Co-operative Transmission

computationally complex making jointly encoded systems with sensor counts in the 10’s intractable as the maximum likelihood decoder of a set of block codes is an NP-hard problem.

We will consider the simplified case where multiple sensors are observing a variable $x \sim N(0, \sigma_x^2)$ corrupted by observation noise $y \sim N(0, \sigma_y^2)$ such that variables $z_i = x + y_i$ are observed by each sensor. At the central consumer the individual transmissions are not required, only the estimator of $x$, $\hat{x}$ is required. The distortion $d$, defined as $E(\hat{x} - x)^2$ is the mean squared error of the estimator. We will compare the rate distortion bounds of a single sensor and a group of sensors making independent observations of $x$. Firstly consider a single sensor that is able to take $n$ independent measurements and then transmit the estimate to the receiver. Define $d_L$ as the variance of the MMSE estimator

$$d_L = \frac{1}{\sigma_x^2 + \frac{n}{\sigma_y^2}}$$

(6.6)

The sensor can form the MMSE $\hat{x}$ and consequently we would like to transmit $\hat{x}$ with a distortion $d$ that approaches the MMSE error $d_L$ with a bit rate

$$r_{\text{SINGLE SENSOR}}(d) = \frac{1}{2} \log_2 \left( \frac{\sigma_x^2}{d - d_L} \right)$$

(6.7)

$$= \frac{1}{2} \log_2 \left( \frac{\sigma_x^2 \left( \frac{1}{\sigma_x^2} + \frac{n}{\sigma_y^2} \right)}{d \left( \frac{1}{\sigma_x^2} + \frac{n}{\sigma_y^2} \right) - 1} \right)$$

(6.8)

Now consider the second case of a group of $n$ sensors taking a single observation and transmitting the observation back using a joint encoder. In [38] the rate distortion bound for $n$ identical sensors (see eq. 10) using joint encoding and operating at the Berger-Tung limit can be written as

$$r_{\Sigma}(d) = \frac{1}{2} \left( \log_2 \left( \frac{\sigma_x^2}{d - d_L} \right) + n \log_2 \left( \frac{n}{\sigma_y^2 \left( \frac{1}{d_L} - \frac{1}{d} \right)} \right) \right)$$

(6.9)

which after some rearrangement can be written as

$$r_{\Sigma}(d) = \frac{1}{2} \left( \log_2 \left( \frac{\sigma_x^2}{d - d_L} \right) + n \log_2 \left( \frac{d_L nd}{\sigma_y^2 (d - d_L)} \right) \right)$$

We will contrast this to the previous example of a single sensor taking $n$ independent observations and then optimally quantising the result. Re-arranging

$$r_{\Sigma}(d) = r_{\text{SINGLE SENSOR}}(d) + \left( \frac{n - 1}{2} \right) \log_2 n$$

(6.10)

$$- \frac{1}{2} \left( \log_2 \left( \frac{\sigma_y^2}{n \sigma_x^2} + 1 \right) + (n - 1) \log_2 \left( \frac{\sigma_y^2}{\sigma_x^2} + n - \frac{\sigma_y^2}{d} \right) \right)$$

(6.11)
We can interpret this as a component that comprises of the single sensor rate

$$r_{\text{SINGLE SENSOR}} (d)$$  \hspace{1cm} (6.12)$$

plus additional terms. The first additional term allows pairs of sensors to be identified

$$\left( \frac{n-1}{2} \right) \log_2 (n)$$  \hspace{1cm} (6.13)$$

plus an additional term

$$(n-1) \log_2 \left( \frac{\sigma^2_y}{\sigma^2_x} + n - \frac{\sigma^2_y}{d} \right)$$  \hspace{1cm} (6.14)$$

that causes the bit rate to increase as the distortion is lowered plus a term that is independent of the distortion. We can compare the performance between the single sensor case and the distributed sensor and pick the distortion such that

$$\left( \frac{\sigma^2_y}{\sigma^2_x} + n - \frac{\sigma^2_y}{d} \right) = 1$$  \hspace{1cm} (6.15)$$

$$\frac{1}{\left( \frac{1}{\sigma^2_x} + \frac{n-1}{\sigma^2_y} \right)} = d$$  \hspace{1cm} (6.16)$$

forcing the final term to zero. This is equivalent to picking the distortion to be the same as that of the MMSE estimator for one less sensor. i.e. we have selected a distortion such that the lower bound variance of the joint encoder with $n$ sensors is the same as the lower bound variance of a single sensors taking $n-1$ identical observations. What remains for $d$ fixed is

$$r_{\text{SINGLE SENSOR}} + \frac{(n-1)}{2} \log_2 n - \frac{1}{2} \left( \log_2 \left( \frac{\sigma^2_y}{n\sigma^2_x} + 1 \right) \right)$$  \hspace{1cm} (6.17)$$

which as $n$ becomes large approaches

$$r_{\text{SINGLE SENSOR}} + \frac{(n-1)}{2} \log_2 (n) - \frac{1}{2} \frac{\sigma^2_y}{n\sigma^2_x}$$  \hspace{1cm} (6.18)$$

i.e. on the average the optimal CEO transmitter can achieve the rate of the optimal single sensor making $N$ independent observations plus every other sensor simply transmitting enough bits to identify itself\(^1\). Clearly there is sufficient information to not only recover the number of transmitting sensors, but to uniquely identify each sensor.

This gives us an insight as to how to further improve spectral efficiency. We need to avoid transmitting information about the identity or source of the observation. If we can recover the original set of observations we are sending too much information - we do not need this information, only the estimate.

\(^1\) $\frac{n-1}{2}$ sensors sending $\log_2 n$ bits allows us to determine the identity of each transmitting sensor
A common approach to this is to incrementally fuse the information. Information is routed through a sensor network along the vertices of a spanning tree, rooted at the information consumer. Each node is responsible for fusing its own observations and the fused observations of all contributing nodes together, and passing the fused estimate back towards the root of the tree. As the information is fused the identity of the source of the information is implicitly discarded, keeping the required datarate low. Although this approach makes the information fusion portion of the system relatively simple it pushes complexity into other parts of the system. In order to avoid using the same sensor report multiple times, the application must ensure that the spanning tree is maintained at all times, even when the nodes experience changes in geometry or availability. If the spanning tree must be regenerated, this needs to be synchronized to ensure that routing loops do not temporarily appear. Finally, the algorithm depends on the radio sub-system successfully maintaining the logical point to point links required. As demonstrated above, this can also be difficult as the system becomes large.

We can consider an alternative approach. As the number of sensors increases there will come a point where there will be more transmitters than number of different states to transmit. Suppose there are $\mathcal{N}$ states in the encoder and $n$ sensors where $n > \mathcal{N}$. Consider a system where each encoder state is allocated a distinct channel. Each sensor transmits using the same set of channels, such that sensors that transmit on the same channel will constructively interfere with other sensors. Provided received power levels are normalized the receiver will be able to estimate how many of each sensor are transmitting on a particular channel or equivalently, with a particular codeword. Consequently, rather than expecting each sensor to transmit identical messages using different channels, each sensor transmits each codeword s.t. the codewords for the same value constructively/additively interfere. In this way the amount of energy in each channel is proportional to the number of sensors in the state associated with the channel. A transmitter chooses the best symbol and transmits at a power level such that the receiver will receive the signal at some unit power level. Other transmitters choose the innovation that best matches their observation. At the receiver the same symbols interfere constructively generating a sequence which is a function of the sum of the total number of transmitters sending the symbol. The receiver now processes the sequence, using the power levels of each symbol to infer the number of sensors transmitting the particular symbol. Consequently as the number of sensors increase the power level of each symbol increases, but the required bandwidth does not, and the accuracy of the estimation improves.

We note that the bandwidth required for a single symbol is the bandwidth required to transmit the mean number of transmitters in this state. The sampling variance of this quantity will reduce as the number of sensors increase - hence the required
bandwidth will reduce. It is important that we are now receiving information about the total number of sensors reporting a state, but we are no longer able to determine which individual sensor is in that state. This is important as it helps explain why we appear to get additional information without using additional bandwidth. In a purely digital system we would need to allocate a separate channel for each sensor, consequently we would always be able to determine which sensor transmitted which piece of information. Given a sufficiently large number of sensors, the number of bits required to uniquely determine the identity of a sensor would be quite large. If we simply sum the power of each sensor together we lose this information, only the sum can be recovered. Channels can be separated by time, frequency or orthogonal codewords. The only requirement is that when multiple sensors transmit on the same channel we should be able to estimate the total amount of power in the channel - and consequently the number of transmitters on the channel.

6.4 Co-operative Transmission of Observations from Identical Sensors

In Chapter 4 of [51] a system for Energy Efficient Estimation is described. In the section headed Non-Orthogonal Access system with Vector Data a method is promoted that uses weighted observations to form an estimation of the observed state at the fusion centre. This section of the thesis addresses an equivalent problem from the perspective of a constant power system where the output power is controlled to achieve constant receive power, rather than modulated to achieve a constant estimation error. Using this approach we observe that we can exceed the performance of the equivalent optimal digital CEO system using the same transmitter power.

Even small improvements in transmitter efficiency can have considerable improvements in mobile system performance by allowing smaller payloads with smaller batteries and transmitters for equivalent system performance. For example a micro light UAV systems even small reductions in system mass have a major impact in system range. From the Breguet Range equation for a liquid fuelled aircraft we have

$$\text{Range} \propto \ln \left( \frac{\text{Fuel} + \text{Payload} + \text{Airframe}}{\text{Payload} + \text{Airframe}} \right)$$

(6.19)

which indicates that reduction in payload provides direct improvement in the UAV range, but additionally reduction in the required payload weight can also reduce the weight requirement of the corresponding airframe to carry the payload, providing even greater improvements. This becomes most significant in the smallest UAV systems where payload weight may be measured in grams and mass savings consequently have...
the greatest impact. In the case of an electrically powered hovering UAV we can use the Induced Power equation\(^2\) where \(F_N\) is the thrust, \(V_0\) is the exit velocity of air accelerated by the propeller, \(\rho\) is the density of air, \(A_p\) is the rotational area of the propeller, \(m\) is the mass of the helicopter and \(G\) is the acceleration due to gravity.

\[
\begin{align*}
P_{\text{min}} &= \frac{F_N V_0}{2} \left( 1 + \sqrt{1 + \frac{4 F_N}{V_0^2 \rho A_p}} \right) \quad (6.20) \\
&= mG \frac{V_0}{2} \left( 1 + \sqrt{1 + \frac{2 mG}{V_0^2 \rho A_p}} \right) \quad (6.21)
\end{align*}
\]
to get the relationship between the mass of a helicopter or quad-copter style drone and the amount of power required from the electric motors. Note that reducing mass has a non-linear effect on reducing the required power and consequently increases the duration and range proportional to \(m^{3/2}\). Once again reduction in payload mass allows the reduction in airframe mass providing benefits beyond those apparent in the equations.

Consider a system that transmits observations of the random variable \(z\) uniformly distributed in \(z \in R^m\). Let the vectors \(\varpi_1...\varpi_N : \mathcal{N} > m\) be the vertices of a convex polytope \(\mathcal{Z}\) in \(R^m\). Consequently for any \(z \in \mathcal{Z}\) there exists \(a_i\) s.t. \(z = \sum_{i=1}^{N} a_i \varpi_i\) for \(a_i \in [0,1] : \sum_{i=1}^{N} a_i = 1\). \(n\) sensors transmit the observations \(z_i\) from the \(i\)th sensor. Define \(z \sim N(0,\sigma_z) : z \in \mathcal{Z}\) restricted to being within the polytope. At the receiver we would like to form the estimator of \(z\), \(\hat{z}\). Define a normally distributed observation noise \(y_i \sim N(0,\sigma_y)\) and quantisation noise \(q_i \sim N(0,\sigma_q)\) such that the observation \(z_i\) is

\[
z_i = z + y_i + q_i \quad (6.22)
\]

The MMSE of \(z\) is

\[
\hat{z} = \frac{\sigma_y}{\sigma_z + \frac{n}{\sigma_y + \sigma_q}} \sum_{i=1}^{n} z_i 
\]

(6.23)

and the variance is

\[
\text{var} \hat{z} = \frac{1}{\sigma_z + \frac{n}{\sigma_y + \sigma_q}} \quad (6.24)
\]

which is termed the distortion \(d\) of the estimator.

The corresponding equivalent analogue system can be defined. As above, observations \(z_i\) are generated by each sensor, with the observations corrupted by observation noise \(y_i\) distributed as above. Rather than quantising the samples the sensor transmits the message in an analog form across multiple channels. The fixed total amount of power is used by the sensor on each transmission.

\(^2\)see [119] eq 3.52, p 219
We define the system as a transmitter that can place energy into individual transmission channels, a linear propagation network that causes energy placed in one channel to interfere with other channels and a receiver that can measure the amount of energy present in a channel. Define the vector \( s_i \in \mathbb{R}^N \) as the energy distributed across all \( N \) channels by the \( i \) transmitter. Note that the total amount of energy that can be transmitted for each observation symbol is \( E_s \) and consequently \( \|s_i\|_1 = 1 \). The propagation network is characterised by \( C \), the co-channel interference matrix which defines the amount of energy that a sensor will measure in a channel based on the energy injected into all channels. Due to the non-ideal nature of the sensors the transmitters received energy may be mis-allocated and double counted (for instance the sensor may respond to energy in other channels as well as its own - integration over all channels will yield a value having more energy than actually transmitted in channel). The elements of the matrix \( C \) define the amount of energy measured by the receiver in channel \( j \) given a transmitted energy in channel \( k \). Channels may interfere due to non-ideal band filters causing coupling between between close channels (in FDMA systems), inaccuracy in slot timing (in TDMA systems), or non-ideal orthogonality in the types of codes selected (CDMA systems), reflections and multi-path interference (from the geometry of the receiver and transmitter). We will make some assumptions about the co-channel interference:

- It is symmetric \( (c_{ij} = c_{ji}) \) and positive semi-definite. The mechanism causing interference couples energy from channel \( i \) into \( j \) equally to coupling \( j \) into \( i \). This is true for channels separated by TDMA, CDMA and FDMA mechanisms. This also implies that the energy measured in any channel cannot be reduced below 0.

- It is short term stationary and measurable. We can characterize a particular system in terms of \( C \) and design encoders and decoders to operate with it. It is measured by monitoring a pilot channel generated by the fusion centre of known characteristics. This can be used to generate the propagation matrix for the current channel conditions.

Note that \( \|C\| > 1 \) is acceptable as both the transmitters and receivers may not be strictly limited to single channels, so that the measured energy at the receiver is greater than the actual energy transmitted. We define an encoder function \( E_z : \mathbb{R}^n \rightarrow \mathbb{R}^N \) and a decoder function \( D_Z : \mathbb{R}^N \rightarrow \mathbb{R}^n \). We define channel noise \( \nu \in \mathbb{R}^N \) as identical independent Gaussian sources which are additively introduced into each channel of the sensor. Consequently \( \nu_i^2 \sim \chi^2(1) \) and \( \nu^2 \in \mathbb{R}^N \) is a vector of IID chi-squared r.v. We define \( n_o \) as the noise power of the channel noise per symbol and \( s_i \in \mathbb{R}^N \) as the energy received from the \( i \) th sensor.
We can write the system as
\[ z_i = z + y_i \]  
(6.25)
\[ y_i = C \mathcal{E}_z (z_i) E_s + \nu^2 n_o \]  
(6.26)
\[ z_i^R = \frac{D_Z (s_i)}{E_s} \]  
(6.27)
\[ z_i^R = D_Z \left( C \mathcal{E}_z (z_i) \right) + D_Z \left( \nu^2 n_o \right) E_s \]  
(6.28)
where \( z_i^R \) is the received and decoded value including noise. As the channel is linear we can sum the received energy from many sources.
\[ z^R = \sum_i z_i^R \]  
(6.29)
\[ = \sum_i \frac{D_Z (s_i)}{E_s} \]  
(6.30)
\[ = n E (z) \]  
(6.31)
It would be convenient if \( D_Z \) and \( \mathcal{E}_z \) were also linear such that we could write
\[ \sum_i \frac{D_Z (s_i)}{E_s} = \frac{D_Z}{E_s} \sum_i s_i \]  
(6.32)
With a linear encoder and decoder we can write
\[ z_i^R = D_Z C \mathcal{E}_z (z_i) + D_Z \nu^2 n_o E_s \]  
(6.33)
If we substitute
\[ D_Z = \begin{bmatrix} \omega_1 & \cdots & \omega_m & c_0 & \cdots & c_{(N-(m+1))} \end{bmatrix} C^{-1} \]  
(6.34)
then the decoder cancels out the cross channel interference, leaving the expression
\[ z_i^R = \begin{bmatrix} \omega_1 & \cdots & \omega_m & c_0 & \cdots & c_{(N-(m+1))} \end{bmatrix} C^{-1} C \mathcal{E}_z z_i \]  
(6.35)
\[ = \begin{bmatrix} \omega_1 & \cdots & \omega_n & c_0 & \cdots & c_{(N-(m+1))} \end{bmatrix} \mathcal{E}_z z_i \]  
(6.36)
where \( c_0 \) to \( c_{(N-2n-1)} \) are free constants. At the encoder we must select \( a_{i,j} \) s.t
\[ z_i = \sum_j a_{i,j} \omega_i \]  
(6.37)
\[ 1 = \sum_j a_{i,j} \]  
(6.38)
\[ 0 < a_i = \begin{bmatrix} a_{i,1} \\ \vdots \\ a_{i,m} \end{bmatrix} < 1 \]  
(6.39)
then we can write
\[ z_i = \begin{bmatrix} \varpi_1 & \cdots & \varpi_m \end{bmatrix} \begin{bmatrix} a_{i,1} \\ \vdots \\ a_{i,m} \end{bmatrix} \] (6.40)
and write
\[ \begin{bmatrix} a_{i,1} \\ \vdots \\ a_{i,m} \\ 0 \\ \vdots \end{bmatrix} = \mathcal{E} z_i (6.41) \]
and consequently
\[ z_i^R = \begin{bmatrix} \varpi_1 & \cdots & \varpi_m & c_0 & \cdots & c_{(N-(m+1))} \end{bmatrix} \mathcal{E} z_i + \text{(noise)} \] \hspace{1cm} (6.42)
\[ = \begin{bmatrix} \varpi_1 & \cdots & \varpi_m & c_0 & \cdots & c_{(N-(m+1))} \end{bmatrix} \begin{bmatrix} a_{i,1} \\ \vdots \\ a_{i,m} \\ 0 \\ \vdots \end{bmatrix} + \text{(noise)} \] \hspace{1cm} (6.43)
\[ = z_i + \text{(noise)} \] \hspace{1cm} (6.44)

Given this form we can write an estimator for \( z, \hat{z} \) and calculate the MSE. Define \( z^R = \sum_i^n z_i^R = nE(z) \) as the quantity decoded at the receiver
\[ z^R = \sum_i^n D_Z C \mathcal{E} z_i + D_Z \nu^2 n_o E_s \] (6.45)
\[ z^R = D_Z C \mathcal{E} z + D_Z \nu^2 n_o E_s \] (6.46)
\[ \frac{z^R}{n} = D_Z C \mathcal{E} z + D_Z C \mathcal{E} z \sum_i^n \frac{y_i}{n} + \frac{D_Z \nu^2 n_o}{n} E_s \] (6.47)
\[ z = \frac{z^R}{n} - \left( \sum_i^n \frac{y_i}{n} + \frac{D_Z \nu^2 n_o}{n} E_s \right) \] (6.48)
\[ = E(z) - \left( \sum_i^n \frac{y_i}{n} + \frac{D_Z \nu^2 n_o}{n} E_s \right) \] (6.49)

Note that each element of \( \nu^2 \) is \( \chi^2(1) \) and assumed independent and distributed with
unitary variance and consequently \( E(\nu^2) = 1 \).

\[
\hat{z} = \frac{z}{n} - \left( E(y_i) + \frac{DZ}{n} E(\nu^2) \frac{n_o}{E_s} \right) \tag{6.50}
\]

\[
\hat{z} = E(z) - \frac{1}{n} DZ 1_{m} \frac{n_o}{E_s} \tag{6.51}
\]

and

\[
\text{var}(z) = E\left( (z - \hat{z})(z - \hat{z})' \right) \tag{6.52}
\]

\[
= E\left( \sum_{i=1}^{n} \frac{y_i}{n} + \frac{DZ}{n} \left( (\nu^2 - E(\nu^2)) \frac{n_o}{E_s} \right) \right)^2 \tag{6.53}
\]

\[
= \frac{\sigma_y}{n} + \left( \frac{n_o}{n E_s} \right)^2 DZ E \left( (\nu^2 - E(\nu^2)) (\nu^2 - E(\nu^2))' \right) D_Z' \tag{6.54}
\]

\[
= \frac{\sigma_y}{n} + \left( \frac{n_o D_Z}{n E_s} \right) \text{var}(\nu^2) \left( \frac{n_o D_Z}{n E_s} \right)' \tag{6.55}
\]

Each element of \( \nu^2 \) is independent with \( \chi(1) \) distribution. Hence \( \text{var}(\nu^2) \) is of the form

\[
\text{var}(\nu^2) = \begin{bmatrix}
2 & 1 & \cdots & 1 \\
1 & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & 1 \\
1 & \cdots & 1 & 2
\end{bmatrix} \tag{6.56}
\]

The receiver does not generally have access to the noise power or the number of transmitters. the receiver must estimate \( n \) and \( n_o \). We can use some additional information to form these estimations. Firstly we note that each sensor transmits the same total power. Define

\[
D_n = \begin{bmatrix}
1 & \cdots & 1 & 0 & \cdots & 0
\end{bmatrix} C^{-1} \tag{6.57}
\]

and define

\[
n^R = D_n C E z_i + D_n \nu^2 \frac{n_o}{E_s} \tag{6.58}
\]

We can interpret the received signal \( C E z_i \) using the decoder \( D_N \) in order to extract
the total number of transmitting sensors using the decoder $D_N$ rather than $D_Z$. 

\[
\begin{align*}
n^R &= \left[ 1 \cdots 1 \ 0 \cdots 0 \right] C^{-1} C \sum_i^n \left[ \begin{array}{c} a_{i,1} \\ \vdots \\ a_{i,m} \\ 0 \\ \vdots \\ \end{array} \right] (6.59) \\
&= n + D_n \left( \nu^2 n_\sigma \over E_s \right) (6.61)
\end{align*}
\]

However we still have a dependency on $n_\sigma$. To eliminate this we define a final decoder $D_{n_\sigma}$ as

\[
D_{n_\sigma} = \left[ 0 \cdots 0 \ 1 \cdots 1 \right] C^{-1} (6.62)
\]

We can now estimate the received noise power as

\[
\begin{align*}
n^R &= \left[ 0 \cdots 0 \ 1 \cdots 1 \right] C^{-1} C \sum_i^n \left[ \begin{array}{c} a_{i,1} \\ \vdots \\ a_{i,m} \\ 0 \\ \vdots \\ \end{array} \right] (6.63) \\
&= n + D_{n_\sigma} \left( \nu^2 n_\sigma \over E_s \right) (6.64)
\end{align*}
\]

Note that the first term evaluates to zero - leaving only the noise term to contribute, allowing a direct estimation of the noise

\[
\begin{align*}
n^R &\approx D_{n_\sigma} \left( \nu^2 n_\sigma \over E_s \right) (6.65) \\
n_\sigma &= n^R E_s \over D_{n_\sigma} \nu^2 (6.66)
\end{align*}
\]

Note that the received $N^R$ is independent of the number of transmitters. We can incorporate these into the original estimate and write

\[
\begin{align*}
z &= \frac{z^R - n^R \over D_{n_\sigma} \nu^2 - \sum_i^n y_i}{n^R - n^R \over D_{n_\sigma} \nu^2} - \sum_i^n \frac{y_i}{n} (6.67) \\
\hat{z} &= \mathbb{E} \left( \frac{ (z^R D_{n_\sigma} - n^R D_{n_\sigma}) \nu^2 }{ (n^R D_{n_\sigma} - n^R D_{n}) \nu^2 } \right) (6.68)
\end{align*}
\]
6. Co-operative Transmission

We can use the Unscented Transform to evaluate the mean by noting that each $v_i$ is Gaussian with zero mean and unit variance. Choosing the sigma points and weights simply requires $2N$ points located at $\pm 1$ on each of the $N$ axis. The weight of each sigma point is $w_1 = \frac{1}{2N}$.

For the $i$ th sigma point we can write

$$v_i^2 = \nu_{i\times N}$$ (6.69)

Noting that the Unscented Transform propagates the mean and variance exactly we have

$$\hat{z} = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{z^R D_{n_o} - n_o^R D_Z}{(n^RD_{n_o}-n_o^RD_n)\nu_{i\times N}} \right)$$ (6.70)

and

$$\text{var} (z) = E \left( (z - \hat{z}) (z - \hat{z})' \right)$$ (6.71)

$$= \frac{1}{N} \sum_{i=1}^{N} \left( \frac{z^R D_{n_o} - n_o^R D_Z}{(n^RD_{n_o}-n_o^RD_n)\nu_{i\times N}} 1_i - \hat{z} \right) \left( \frac{z^R D_{n_o} - n_o^R D_Z}{(n^RD_{n_o}-n_o^RD_n)\nu_{i\times N}} 1_i - \hat{z} \right)'$$ (6.72)

Clearly performance is affected by the method of selection of $a_{i,j}$ as this is under-constrained if $m > n + 1$. At this point any optimization technique may be applied to form the encoder provided the linear relationship between the encoder/decoder is preserved, i.e. that

$$\sum DCE (z) = DCE \sum \mathcal{E} (z) = z$$ (6.73)

This remains true if the encoder function can be written as

$$E_j \in \mathcal{E} \text{ s.t. } DCE = I$$ (6.74)

$$\max_j z' E_j E_j z$$ (6.75)

$$a = E_j z$$ (6.76)

i.e. the process of encoding, where there are multiple possible linear encoding due to the underconstrained system is to select from the set of possible linear mappings one that maximises some parameter (in this case the $\|z\|_2$ norm). The resultant encoding still allows linear summation without constraining the choice of encoding method.

How well does a multi user digital system compare with a comparable multi-user analog system? We can compare these two systems for an equivalent set of channel resources. We note that the set of sensors is transmitting some joint information that has been corrupted by observation noise. The CEO problem [38] is a comparable problem as we can recover the sum of the observations if we know the central estimate and the number of sensors. From (6.9) we know that the lower bound rate for a central estimate to satisfy a distortion $d$ and $n$ sensors. As noted above this rate approaches
6. Co-operative Transmission

the rate that a single sensor would require given all \( n \) observations plus an overhead term that is comparable to the amount of information required to distinguish the individual sources.

Consider a system with bandwidth \( b \) and \( n \) channels. An ideal system will be able to provide a capacity \( c \) given AWGN channel interference.

\[
c = b \log_2 \left( 1 + \frac{\text{Power}}{\text{Noise}} \right)
\]

\[
= b \log_2 \left( 1 + \frac{(E_b \times c)}{(n_o \times b)} \right)
\]

\[
= b \log_2 \left( 1 + \frac{E_s}{n_o} \right)
\]

\[
= b \log_2 (1 + \text{SNR})
\]

where \( c \) is the channel capacity, \( b \) is the channel bandwidth, \( E_b \times c \) is the transmitter power and \( E_s \) is the energy per symbol. For a given rate \( r \) an encoder can encode a sample \( z_i \) from a source with mean \( z \) variance \( \sigma_z^2 \) with a distortion \( \sigma_y^2 + \sigma_q^2 = d \). If we wish to retain the same \( r \) and bandwidth \( b \) while increasing the number of sensors/channels \( n \) we must alter the channel \( E_b/n_o \) either by increasing the transmitter power or by reducing the receiver noise. Consider the achievable distortion that a CEO based digital system can achieve for a given bandwidth and SNR.

\[
b \log_2 (1 + \text{SNR}) = \frac{1}{2} \left( \log_2 \left( \frac{\sigma_z^2}{d} \right) + n \log_2 \left( \frac{n}{\sigma_y^2 \left( \frac{1}{d_L} - \frac{1}{d} \right)} \right) \right)
\]

after re-arranging for \( n \) large and \( d \) close to \( d_L \)

\[
b \log_2 (1 + \text{SNR}) \approx \frac{1}{2} \left( n \log_2 \left( \frac{n}{\sigma_y^2 \left( \frac{1}{d_L} - \frac{1}{d} \right)} \right) \right)
\]

\[
\frac{1}{d_{\text{CEO}}} \approx \frac{1}{d_L} - \left( \frac{n}{\sigma_y^2 (1 + \text{SNR})} \right) \frac{2b}{\pi}
\]

\[
= \frac{1}{\sigma_z^2} + \frac{n}{\sigma_y^2} \left( \frac{(1 + \text{SNR})^2}{(1 + \text{SNR})^2} - 1 \right)
\]

Note that as the number of sensors increases the distortion deviates further from the optimal behaviour of decreasing linearly with \( n \), but limits based on the SNR of the transmission system. Now consider the co-operative transmission system. Only a single transmission is required at the point where all the sensor’s transmissions are summed. Consequently though the channel transmission power remains the same, the noise bandwidth can be reduced to 1. Alternatively the bandwidth can remain
Co-operative Transmission

6. Co-operative Transmission

the same but after taking \( b \) observations the observations can be averaged together, effectively acting as a low pass filter to reduce the noise power. Note \( D_Z \) will implicitly depend on \( \sigma_z^2 \). Rather than a small amount of distortion introduced by quantisation, the encoding process must clip for large values that would exceed the power budget of the transmitter. As \( z \) is normally distributed, \( \sigma_z^2 \) will be less than the range of \( D_Z \).

For the purpose of this exercise we will define the decoder as \( \sigma_z D_Z \) to emphasise this. From (6.52) and substituting in the expression for SNR we have

\[
D_{\text{co-operative}} = \frac{\sigma_y^2}{n} + \left( \frac{\sigma_z}{nb \text{SNR}} \right) D_Z \text{var} (\nu^2) D'_Z \left( \frac{\sigma_z}{nb \text{SNR}} \right)
\]  

(6.85)

Where there are \( N \) degrees of freedom to be transmitted the available bandwidth \( b \) is reduced and the distortion becomes approximately

\[
D_{\text{co-operative}} = \frac{\sigma_y^2}{n} + \left( (N + 2) \frac{\sigma_z}{nb \text{SNR}} \right) D_Z \text{var} (\nu^2) D'_Z \left( (N + 2) \frac{\sigma_z}{nb \text{SNR}} \right)
\]  

(6.86)

\[
\frac{1}{D_{\text{co-operative}}} \approx \frac{1}{\sigma_z^2} + \frac{n}{\sigma_y^2} \left( \frac{(1 + \text{SNR}) \frac{2b}{2N} - 1}{(1 + \text{SNR}) \frac{2N}{2N}} \right)
\]  

(6.87)

Compared to the CEO system the co-operative transmission system is potentially superior if

\[
\frac{1}{D_{\text{CEO}}} < \frac{1}{D_{\text{co-operative}}}
\]  

(6.88)

\[
\frac{1}{\sigma_z^2} + \frac{n}{\sigma_y^2} \left( \frac{(1 + \text{SNR}) \frac{2b}{2N} - 1}{(1 + \text{SNR}) \frac{2N}{2N}} \right) < \frac{n}{\sigma_y^2 + \frac{n}{\sigma_y^2}} \left( (N + 2) \frac{1}{nb \text{SNR}} \right) D_Z \text{var} (\nu^2) D'_Z \left( (N + 2) \frac{1}{nb \text{SNR}} \right)
\]  

(6.89)

(6.90)

Note that as \( b \) increases that the inequality approaches

\[
\frac{1}{\sigma_z^2} < \frac{n}{\sigma_y^2}
\]  

(6.91)

and favours the co-operative system, while conversely if \( \sigma_z^2 \) is large while \( n \) is smaller the CEO estimator is superior for the same bandwidth. The choice depends on the number of sensors and the complexities of managing channel access for a large number of sensors and the encoder and decoder complexity for the CEO estimator.

To summarise, the amount of power required to achieve the required channel capacity for a CEO system operating at the Berger-Tung lower bound for a large number of sensor will increase without bound. Alternatively, for a co-operative transmission
system the transmitter power can remain constant and as the number of sensors increases the received transmission power will increase, allowing the formation of the estimate with reduced noise impact, the additional transmitters improve the signal to noise ratio rather than reduce it. This approach can only work where the messages to be sent are independent of the sensor state and linearizable, for instance this would not work with measurements of bearings as the use of bearing information requires the sensors position.

6.5 Example - An HF Transmission System

To illustrate this we can consider the example of a system designed to fuse measurements over an HF radio link. HF radio is capable of worldwide reception but has very little available bandwidth. HF channels are approximately 3kHz wide and spaced at 4KHz intervals. The HF band occupies 3-30MHz but due to the nature of the propagation of HF signals only a small fraction of this band is usable at a given time. Combined with the extremely long (and potentially global) reach of HF transmission it is apparent that HF bandwidth is highly contested and is often highly noisy. Data provided in [120] and [121] provided an example of the noise and path losses to be expected in this environment. Given this hostile and tightly constrained environment is it possible to transmit multiple sensor reports and fuse them at the receiver?

Consider a system allocated a fixed amount of spectrum. Each sensor is to transmit its independent measurements of a common parameter and the receiver is expected to recover the maximum likelihood estimate of the common observed parameter. Transmitters can only generate a fixed transmission power and we assume that for stationary transmitters and receivers that the propagation characteristics remain identical and stable over the short term (the period of a single observation). It is desired that the system’s performance should improve as more sensors are added to the system, however the maximum transmission power of a sensor, the available channel bandwidth and the amount of noise in the channel is fixed and beyond the control of the designer.

For a digital system the required transmission rate is set by the encoder and the source being transmitted. In the case of the source being a Gaussian source then the lower bound of the rate/distortion relationship is known analytically

\[ r = \frac{1}{2} \log_2 \left( \frac{\sigma^2}{d} \right) : \frac{\sigma^2}{d} > 1 \]  

(6.92)

and hence the variance of the error of the received samples caused by the encoding process will be not less than \( d \) for a transmission rate \( r \). The system is to take 4 non-correlated measurements drawn from a zero mean distribution of known variance \( \sigma^2 \)
and encode them. We assume that the system is using joint encoding to minimise the absolute amount of data and that the only traffic that is considered is that portion of the traffic that is not correlated between sensors (the random noise section) and that the correlated section can be assumed to be sent only once. We assume the data is organised such that 8 bits are allocated to the correlated portion of the data and consequently need only be accounted for once while the non-correlated portion is encoded using a further 8 bits that must be transmitted by each sensor for a total of 16 bits of precision. For \( n \) sensors to transmit measurements with distortion \( d \) the required rate is \( r_n \) (where the sensors are independent) and consequently we can calculate the lower bound of the SNR (Signal To Ratio, or SINAD - signal to noise plus distortion ratio) that the transmission system can support

\[
r = b \log_2 (1 + \text{SNR}) = b \log_2 (1 + E_s/n_o)
\]

(6.93)

(6.94)

where \( E_s \) is the energy per symbol while \( n_o \) is noise power per unit Hz. Noting that \( b, r \) and \( E_s \) are fixed all that remains is the maximum acceptable Noise Power spectral density \( n_o \) that the system can tolerate.

Alternatively the measurements can be transmitted in a co-operative form as described above. As each sensor transmits in the same channel the signals are mixed additively at the receiver.

Figure (6.1) shows the performance of a system using a 2kHz channel. The system assumes a -150dBW/Hz rural model for noise and a path loss of 130dB which corresponds to approximately continental coverage (see the example path losses in [121] ). For the digital system it is assumed that the correlated portion of the observations can be represented by 8 bits, while the non-correlated portion is encoded by a further 8 bits per sensor. The total number of bits to be transmitted over all sensors is not less than

\[
8 + 8N
\]

(6.95)

with the variance of the observed signal \( \sigma \) normalised to 1. 6.2 assumes the same scenario except that the transmission system is allocated a 100kHz wide channel. The transmission system has a cross channel co-channel interference matrix \( C \) of the form

\[
C = \frac{1}{N} \begin{bmatrix}
N & 1 & \cdots & 1 \\
1 & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & 1 \\
1 & \cdots & 1 & N
\end{bmatrix}
\]

which is comparable to that of a CDMA system. The
Figure 6.1: 2kHz Bandwidth Sensor Performance
decoder $D_Z$ was constructed with $N = 8$ as

$$2^7 \cdot \begin{bmatrix}
1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & -1
\end{bmatrix} \cdot C^{-1} \quad (6.96)$$

in order to represent signed 8 bit numbers (the top 8 correlating bits).

The simulated system generates a randomised sample set using the unit normal distribution. Firstly the required SNR is calculated using Shannon’s limit and assuming an ideal joint encoder with an ideal media access method. From the bounds of an ideal joint encoder we can determine the minimum aggregate bit rate. From Shannon’s limit we know the minimum SNR that can support the required bit rate, and from the assumption of ideal media access we can assume that all sensors can share that aggregated channel without loss in efficiency. Approximately 500,000 samples are taken over all sensors for each datapoint (the total number of samples is constant, as the number of sensors is increased the number of samples per sensor is proportionately decreased).

The solid blue line plots the RMS error of the estimator observed by the digital system. Although the aggregate bit rate is quite high the number of bits required for transmission from any individual sensor is small. The dashed red line represents the amount of transmitter power required by an individual transmitter to achieve the required aggregate bit rate, and represents the transmitters momentary peak power output. A transmitter must transmit the bits it is required to send at the specified power level in order for the receiver to be able to receive the signal at Shannon’s limit.

The dashed blue line is the average output power. Each transmitter only has to transmit a small number of bits in the bit stream. Consequently the average power consumed over a long period of time is considerably less than the peak power. The average power requirement is indicative of the power drained from batteries, while the peak output power is indicative of the peak load experienced by the transmitter. In the narrowband case note that the average power required per transmitter increases sharply as the number of sensors increases as the systems data rate is power limited. To achieve the required data rate on the narrow channel the SNR ratio must be improved, consequently the average power is increased. The wideband case does not exhibit this at the plotted sensor network sizes, the SNR ratio is not limiting the transmission rate and so the average power remains mainly constant.

The solid red line is the performance of the co-operative system under identical power constraints to the digital system for the same number of sensors, i.e. if the analog receiver received the same power and noise levels as the digital receiver for the
Figure 6.2: 100kHz Bandwidth Sensor Performance
same number of sensors its performance curve would match the red line. It should be noted that both systems require peak output power ratings that range up to extremely, infeasibly large levels (10KW and up) once the number of sensors becomes large.

The final green plot is the line that the co-operative analog system would achieve if each sensors transmit power level was locked at that of the two digital sensor system - i.e. the transmitter power used by each sensor is the same as the left most entry on the broken red plot. The green plot is significant as it shows that while keeping the peak output power at the minimum level (around 20W) the system can be constructed to operate with any number of sensors using precisely the same amount of bandwidth. Furthermore, provided the number of operational sensors exceeds 10 the performance of the system will exceed the performance achievable by a digital system for the same transmitter power level. As the number of sensors increase we see under both scenarios that the performance of the fixed transmitter power co-operative system approaches that of the digital system. We note that the co-operative system only starts providing comparable performance when the number of sensors exceeds the number of distinct transmission states ($N=8$), to ensure that each transmission state is occupied by multiple transmitters.

6.6 Co-operative Transmission of Observations with Unknown Variance

The previous section covers the case of $\sigma_y$ being a fixed known value, but what of the case of the observer with varying $\sigma_y$? The previous method of Section 6.4 has some serious limitations:

- The observation method has to be identically linear for all sensors.
- The observation method cannot be time varying as we require knowledge of the sensor covariance in order to decode the measurement.
- We require an explicit estimate of the number of contributing sensors $N$.

In this section we extend the method to remove these restrictions. Consider the case of a dynamical system where multiple sensor reports are to be fused to generate
a state estimate. Consider the linear stochastic system

\[ x_{k+1} = Fx_k + u_k \]  \hspace{1cm} (6.97)

\[ z_k = Hx_k + v_k \]  \hspace{1cm} (6.98)

\[ x_{k|k} = x_{k|k-1} + K_k(z_k - Hx_{k|k-1}) \]  \hspace{1cm} (6.99)

\[ K_k = P_{k|k-1} \left( H' S_k^{-1} \right) \]  \hspace{1cm} (6.100)

\[ S_k = HP_{k|k-1}H' + R \]  \hspace{1cm} (6.101)

\[ P^{-1}_{k|k} = P^{-1}_{k|k-1} + H'R^{-1}H \]  \hspace{1cm} (6.102)

\[ \text{var}(v_k) = R \]  \hspace{1cm} (6.103)

\[ \text{var}(x_{k|k-1}) = P_{k|k-1} \]  \hspace{1cm} (6.104)

where \( u, v \) are random vectors. \( z_k \) is a vector comprising the observations of many sensors, \( z_{k,1}, \ldots, z_{k,n} \) which have different linear observation operators \( H_i \) and noise covariance \( R_i \). We can re-write the observation process of the linear estimator as

\[
\begin{bmatrix}
  z_{k,1} \\
  \vdots \\
  z_{k,n}
\end{bmatrix}
= \begin{bmatrix}
  H & \vdots \\
  \vdots & \vdots \\
  H & \vdots
\end{bmatrix}
\begin{bmatrix}
  x_1 \\
  \vdots \\
  x_n
\end{bmatrix}
+ \begin{bmatrix}
  v_1 \\
  \vdots \\
  v_n
\end{bmatrix}
\]  \hspace{1cm} (6.105)

\[ R = \begin{bmatrix}
  R_1 & 0 & \cdots & 0 \\
  0 & R_2 & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \cdots & R_n
\end{bmatrix} \]  \hspace{1cm} (6.106)

for observation noise that is independent for each sensor. This is the form for a centralised fuser given multiple simultaneous observations. From the above definitions we can re-arrange the expressions for the covariance of the estimator

\[ P^{-1}_{k|k} = P^{-1}_{k|k-1} + H'R^{-1}H \]  \hspace{1cm} (6.107)

\[ = P^{-1}_{k|k-1} + H' \left( \sum_i R_i^{-1} \right) H \]  \hspace{1cm} (6.108)

as well as the update of the state estimator using the Kalman update.

Calculate \( S_k \) the residual covariance

\[ S_k^{-1} = \left( HP_{k|k-1}H' + R \right)^{-1} \]  \hspace{1cm} (6.109)

\[ = R^{-1} - R^{-1}H \left( P^{-1}_{k|k-1} + H'R^{-1}H \right)^{-1} H'R^{-1} \]  \hspace{1cm} (6.110)

\[ = R^{-1} - R^{-1}H \left( P^{-1}_{k|k-1} + H' \left( \sum_i R_i^{-1} \right)^{1/2} \right)^{-1} H'R^{-1} \]  \hspace{1cm} (6.111)
and $K_k$ the Kalman gain.

$$K_k = P_{k|k-1}^{-1} \left( - \left( H' \left( \sum_i R_i^{-1} \right) H \right) \left( P_{k|k-1}^{-1} + H' \left( \sum_i R_i^{-1} \right) H \right)^{-1} H'R^{-1} \right)$$  \hfill (6.112)

$$= P_{k|k-1} \left( I - \left( P_{k|k-1}^{-1} \left( H' \left( \sum_i R_i^{-1} \right) H \right)^{-1} + I \right)^{-1} \right) \cdot$$  \hfill (6.113)

$$= \begin{bmatrix} H'R_1^{-1} & \cdots & H'R_n^{-1} \end{bmatrix}$$  \hfill (6.114)

Note that we can re-arrange

$$P_{k|k-1}^{-1} \left( I - \left( P_{k|k-1}^{-1} \left( H' \left( \sum_i R_i^{-1} \right) H \right)^{-1} + I \right)^{-1} \right)$$  \hfill (6.115)

and write the estimator update in terms of $\sum_i R_i^{-1}$ and $\sum_i R_i^{-1} z_i$

$$P_{k|k}^{-1} = P_{k|k-1}^{-1} + H' \left( \sum_i R_i^{-1} \right) H$$  \hfill (6.117)

$$x_{k|k} = x_{k|k-1} + K_k \left( z_k - H x_{k|k-1} \right)$$  \hfill (6.118)

$$= \left( I - K_k H \right) x_{k|k-1} + K_k z_k$$  \hfill (6.119)

$$= \left( I - P_{k|k} H' \left( \sum_i R_i^{-1} \right) H \right) x_{k|k-1}$$  \hfill (6.120)

$$+ P_{k|k} \left[ \begin{array}{c} H'R_1^{-1} \cdot \cdots \cdot H'R_n^{-1} \end{array} \right] \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$$  \hfill (6.121)

$$= P_{k|k} P_{k|k-1}^{-1} x_{k|k-1} + P_{k|k} H' \left( \sum_i R_i^{-1} z_i \right)$$  \hfill (6.122)

Note that this allows us to express the fused update in terms of sums of terms transmitted by the sensors. If the sensor transmits $R_i^{-1}$ and $R_i^{-1} z_i$ in such a fashion as they additively interfere, then the receiver need only receive the summed value and process it.
6.7 Example - A data fusion system using a single satellite uplink channel.

For the next example we can consider a system built around satellite transmission to gain coverage of an extended area. Most current satellite systems are capable of retransmitting signals transmitted on an uplink channel onto a configured downlink channel. Each channel has a corresponding transponder frequency, the nominal centre frequency of the channel and a footprint, which is a geographical region that the satellites antenna system is “focused” on. An uplink associated with a particular transponder and footprint can be linked to a downlink to create a leased “circuit” provisioned by the satellite service vendor. Current commercial satellites can support hundreds of these circuits with bandwidths up to the 10’s of MHz. It is important to note that these circuits are RF analog circuits, in most cases the satellite does not attempt to demodulate and re-generate the transmitted signal, it simply acts as a broadband RF amplifier, amplifying and frequency shifting the signal from an input transponder to an output transponder. There are exceptions (the defunct Iridium system performed demodulation and switching in the satellite) but the current engineering practice is that of leaving the demodulation in the ground stations where they can be easily updated, while the satellite is provisioned with wideband solid state linear amplifiers and down converters in order to boost the reliability, simplicity and lifespan of the unmaintainable satellite. Spectral efficiency of satellite services is achieved through using beam steering and shaping, allowing the same spectrum to be re-used by multiple bases stations at different geographic locations.

A concrete example of this is the Intelsat services. This example uses information available at the Intelsat General service offerings as of 2014. The Intelsat 19 satellite offers both C and Ku Band coverage over the Australia/New Zealand region. It was launched in 2012 and supports 24 C band (5.925-6.425 GHz uplink, 3.7-4.2GHz downlink) and 34 Ku Band (14 - 14.5GHz uplink, 12.25-12.75 GHz downlink) transponders, each transponder capable of 36MHz bandwidth. The antenna system can support connection of the C band transponders to a single footprint covering the Pacific region while the Ku Band transponders can connect to one of 4 different footprints. The one depicted in the graphic above is the footprint covering Australia/New Zealand. Although a customer can purchase access for a complete transponder it is more typical that customers purchase a fraction of the available transponder capacity.

Noise in a satellite system is dominated by the thermal noise of the satellite and base station amplifiers and the extremely high path losses experienced by the signals (see [122]). This results in the performance dominated by the sensors transmission power and the thermal noise of the amplifiers in the satellite and base station. The
tables plotted below list the required EIRP power for the sensor base station (as opposed to the actual power). The EIRP is the effective isotropic radiated power, i.e. the power that would have been required if the sensor radiated power in all directions, which excludes the antenna gain. As the antenna gain for a modest parabolic reflector (10cm parabolic reflector at 14GHz can achieve antenna gains of 20dB) can be quite large the actual transmitter power need not be increased, only the antenna size.

We can construct such a system in a fashion similar to Section 6.3. In addition to the transmission of the mean, we also need the covariance. We form a transmission vector

\[
\begin{bmatrix}
a_{i,1} \\
\vdots \\
a_{i,m}
\end{bmatrix}
\]

such that the following relationships are satisfied

\[
\begin{bmatrix}
\varphi_1 & \cdots & \varphi_m
\end{bmatrix}
\begin{bmatrix}
a_{i,1} \\
\vdots \\
a_{i,m}
\end{bmatrix} = H_i^T R_i^{-1} z_i \tag{6.123}
\]

\[
\begin{bmatrix}
\psi_1 & \cdots & \psi_m
\end{bmatrix}
\begin{bmatrix}
a_{i,1} \\
\vdots \\
a_{i,m}
\end{bmatrix} = \text{vec } H_i^T R_i^{-1} H_i \tag{6.124}
\]
We define the decoder $D_{Z^2}$ as

$$D_{Z^2} = \begin{bmatrix} \psi_1 & \cdots & \psi_m \end{bmatrix} C^{-1} \quad (6.125)$$

Combined together we have the linear programming problem for $a_{i,j}$. Given

$$\varpi_1, \ldots, \varpi_m, \xi_1, \ldots, \xi_m \quad (6.126)$$

how to solve for $a_{i,j} \in [0, 1]$ and maximise $\|a\|_2$ while satisfying

$$\begin{bmatrix} \varpi_1 & \cdots & \varpi_m \\ \psi_1 & \cdots & \psi_m \\ 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_{i,1} \\ \vdots \\ a_{i,m} \end{bmatrix} = \begin{bmatrix} H_i' R_i^{-1} z_i \\ \text{vec } H_i' R_i^{-1} H_i \\ 1 \end{bmatrix} \quad (6.127)$$

As above we require $n_o$ and can use the same method to determine it. We do not explicitly require $n$ so we only require

$$n_o = \frac{n_o^R E_s}{D_{n_o} \nu^2} \quad (6.128)$$

We define the the received observations as

$$H' R^{-1} z = H' \sum_i^n R_i^{-1} z_i \quad (6.129)$$

$$(H' R^{-1} z)_R = H' R^{-1} z + D_Z \nu^2 n_o \quad (6.130)$$

and the received variance as

$$\text{vec } \left( H' \sum_i R_i^{-1} \right) = \text{vec } \left( H' \left( \sum_i R_i^{-1} \right) H \right) + \text{vec } \left( H \sum_i (R_i^{-1})' H' \right) \quad (6.131)$$

$$= D_{Z^2} C E_{Z^2} \sum_i \begin{bmatrix} a_{i,1} \\ \vdots \\ a_{i,m} \end{bmatrix} + \frac{(D_{Z^2} + K_{mn} D_{Z^2})}{2} \nu^2 n_o \quad (6.132)$$

$$\text{vec } H' R^{-1} H = \text{vec } H' R^{-1} H + \frac{(D_{Z^2} + K_{mn} D_{Z^2})}{2} \nu^2 n_o \quad (6.133)$$

where $K_{mn}$ is the commutation matrix. The formation of the decoder enforces that the received $R_R^{-1}$ is symmetric.

$$\widehat{H' R^{-1} z} = (H' R^{-1} z)_R - n_o^R D_Z E \left( \frac{\nu^2}{D_{n_o} \nu^2} \right) \quad (6.134)$$

$$\text{var } (H' R^{-1} z) = n_o^R D_Z \text{var } \left( \frac{\nu^2}{D_{n_o} \nu^2} \right) D_Z n_o^R \quad (6.135)$$
6. Co-operative Transmission

\[ \text{vec } H^T R^{-1} H = \text{vec } H^T R^{-1} H - \frac{n^R_0}{2} \mathbb{E} \left( \frac{(D_{Z*Z} + \kappa_{mn} D_{Z*Z}) \nu^2}{D_{n_o} \nu^2} \right) \]  
\[ \text{var} \left( \text{vec } H^T R^{-1} H \right) = \left( \frac{n^R_0}{4} \right) \bullet \]  
\[ \left( D_{Z*Z} + \kappa_{mn} D_{Z*Z} \right) \text{var} \left( \frac{\nu^2}{D_{n_o} \nu^2} \right) (D_{Z*Z} + \kappa_{mn} D_{Z*Z})' \]  

Note that \( \mathbb{E} \left( \frac{\nu^2}{D_{n_o} \nu^2} \right) \) and \( \text{var} \left( \frac{\nu^2}{D_{n_o} \nu^2} \right) \) can be easily numerically evaluated using the Unscented Transform (see [118]) and our knowledge of the distribution of \( \nu \).

Consequently we can transmit reports from multiple sensors and by correctly formulating the transmission format perform the data fusion at the receiver by exploiting the linear properties of the transmission channel. This allows many sensors to transmit over a set of common use channels provided the channel characteristics are sufficiently similar for all sensors.

The simulated system generates a randomised sample set using the unit normal distribution identical to the previous section. Firstly the required SNR is calculated using Shannon’s limit and assuming an ideal joint encoder with an ideal media access method. From the bounds of an ideal joint encoder we can determine the minimum aggregate bit rate, from Shannon’s limit we know the minimum SNR that can support the required bit rate, and from the assumption of ideal media access we can assume that all sensors can share that aggregated channel without loss in efficiency. Approximately 500,000 samples are taken over all sensors for each datapoint (the total number of samples is constant, as the number of sensors is increased the number of samples per sensor is proportionately decreased). The solid blue line plots the RMS error of the estimator observed by the digital system. The dashed red line represents the amount of power required to transmit at the aggregate bit rate, and represents the transmitters momentary peak power output. A transmitter must transmit the bits it is required to send at the specified power level in order for the receiver to be able to receive the signal at Shannon’s limit. The dashed blue line is the average output power. Each transmitter only has to transmit a small number of bits in the bit stream. Consequently the average power consumed over a long period of time is less than the peak power. The average power requirement is indicative of the power drained from batteries, while the peak output power is indicative of the peak load experienced by the transmitter. Note that the average and peak powers are the EIRP power, and do not consider the transmitter antenna gain. This will reduce the actual power reduced by the transmitter and improve battery life at the expense of a more complex transmitter antenna configuration. In addition to the problems associated with the design of transmitters with a large ratio between the average and peak powers there is also limits on the maximum received power that a satellites sensitive receivers can tolerate. The
Figure 6.4: 2KHz bandwidth Sensor Performance
Figure 6.5: 100kHz bandwidth Sensor Performance
receivers will typically have between 30db and 40db acceptable input range of power before non-linear effects are introduced into the amplifiers. This is seen in the satellites acceptable SFD which represents the acceptable power density incident on the satellite. For example Intelsat 19 can handle an SFD of between -103dBW and -66dBW before nonlinearities can result. The solid red line is the performance of the analog system under identical power constraints to the digital system for the same number of sensors, i.e. if the analog receiver achieved the same level as the digital receiver for the same number of sensors its performance curve would match the red line. The final green plot is the line that the analog system would achieve if each sensors transmit power level was locked at that of the two digital sensor system. Furthermore, provided the number of operational sensors exceeds 64 the performance of the system will exceed the performance achievable by a digital system for the same transmitter power level. As the number of sensors increase we see under both scenarios that the performance of the fixed transmitter power analog system approaches that of the digital system. We note that the analog system only starts providing comparable performance when the number of sensors exceeds the number of distinct analog transmission states \( N = 40 \), to ensure that each transmission state is occupied by multiple transmitters. Note that the decoder averages the received \( H'R^{-1}H \) by summing it with the transpose - this ensures the received value is always symmetric irrespective of noise.

This and the previous example demonstrate that large numbers of small co-operative sensors can outperform a smaller number of high quality sensors, but in order to achieve this methods of efficiently transmitting the sensor reports are required. By using a co-operative transmission system such that the individual sensor reports can be combined additively at the receiver allows individual sensors to use less transmission power and bandwidth than the equivalent digital system. Reduction in transmission power requirements allows reduction in power storage, and consequential reductions in the weight, size and cost of individual sensors.

6.8 Co-operative Transmission of Non-Linear Observations

There are still two additional improvements that can be applied to this system. In Chapter 5 we developed two techniques to reduce the amount of information sent by a sensor while still allowing the receiver to support both lost transmission and late joiners. By changing \( \zeta \) the amount of information sent by the sensor can be smoothly varied between sending only observations to only innovations, or any convenient amount between the two extremes. In section 5.5 we demonstrated that we can
generate a pseudo innovation that provides information between that of an observation and an actual information. From (5.88) we know that we can re-arrange a single observation such that it is a pseudo-innovation, an innovation that summarises a subset of the previous observations. Furthermore we can select a value $\zeta$ to reduce the impact of prior observations on the transmitted value, allowing the system to gracefully deal with late joiners or changes in the number of sensors. More importantly we can form the innovation in a linear form, even if the true observation was non-linear as the underlying process is still linear. As noted, provided the transmitter and receiver are agreed on the parameters of the linear process used for the transmission model there is no requirement that this linear process actually match the real process. Provided the sensor can form local state estimates and posterior covariances, those estimates can be used for the transmission and fusion at the expense of sub-optimal encoding due to the innovations no longer being zero-mean. In exchange sensors can be heterogeneous and non-linear.

By incorporating both these methods we can:

- Permit non-linear sensors and sensors using non-linear process models, provided the sensors all have an agreed upon linear transmission model.

- Reduce the impact of noise at the receiver. The encoder must be able to encode the entire possible range of values in a linear fashion. Consequently received noise will be amplified in the same way. Reducing the possible input range by only transmitting innovations reduces the gain applied to received noise.

- Allow late joiners. Not all sensors need transmit all the time, a fusion centre can gain the benefits of innovation like sequences but allow individual sensors to join late and drop out.

Noting that incorporating these methods is as simple as creating a new $R_i$ and $z_i$ and changing the fusion centre to use the state augmented dynamical model we can directly incorporate the techniques of Chapter 5 by transmitting the updated forms of $H' R_i^{-1} H$ and $H' R_i^{-1} z_i$ and perform information fusion on the state augmented model. Transmission occurs as previously but note that the range of values of $z_i$ has now been reduced, reducing the impact of noise on the system performance.

We will demonstrate this method with an example.

6.9 A Practical Implementation

In this final section a practical system that transmits weighted innovations is examined and some of the engineering issues considered. We consider the requirements of the
Each sensor has limited power available. It is assumed that its sensor is passive and processor efficient such that the majority of power needs to be allocated to transmission. The sensors transmission system (the combination of the transmitter and antenna) is characterised by its EIRP (effective isotropic radiated power).

The problem of track allocation of observations is neglected. It is assumed that there exists a mechanism that allocates observations to tracks at the sensor with sufficient reliability. Furthermore it is assumed that there exists a method for allocation of tracks to channels.

The channel properties are time variable, but observable. Typical techniques for determining the channel properties are the use of pilot channels used to measure channel fading and multipath reflections. There are many more sensors than there are tracks to be observed or independent state variables to be transmitted. This is required so that the co-channel interference $C$ can be estimated at the sensor as well as the channel loss.

In the forward path of the system a reference signal is transmitted at a known power level and with known timing information. Note that the reference signal need not be
6. Co-operative Transmission

transmitted by the actual data fuser and does not need any information about the tracks or fusion centre. In satellite systems the reference signal may be generated by the satellite itself and is used by the ground stations to estimate the channel properties. The timing information is used to synchronise the transmitters CDMA “chips”. As an alternative timing information can be obtained from GPS references or through an alternative absolute time source.

The channel is a time varying linear media that causes attenuation and multipath reflections. It can be modelled as a FIR filter with the tap values varying slowly with time. At each sensor a channel estimator estimates the channel properties based on the received reference signal. This is used to construct an appropriate transmitter filter that will precompensate for (some) of the distortion and loss that would be experienced by the transmitted signal. This process is commonly used in WCDMA and CDMA-2000 systems to precompensate for loss in the uplink of mobile telephony systems. As the transmission chip rate can be very low (only a single “symbol” per observation is required) the impact of fast fading can be mitigated by spreading each symbol over time and receiver antenna diversity, while feed forward power control helps reduce slow fade. Timing synchronisation is required such that the transmitter chips are correctly aligned between transmitters, but this is a less stringent requirement than RF synchronisation as the base band chip rate is orders of magnitude slower than the RF modulation frequency.

The sensor generates the innovations and power is allocated to the parameters as described previously. The signal is modulated to precompensate for the measured channel characteristics and the uplinked parameters are mixed in the channel and received at the fusion centre as a sequence of summed pseudo-innovations and measurement covariances. At the fusion centre receiver the aggregate power level is monitored, and if higher than its safe limit (the point at which the receiver may be damaged, or become nonlinear) the reference signal from the fusion centre is increased, causing the channel estimators to “back off” the transmitter power at the sensors.

Each sensor may transmit innovations covering one or more tracks, and divide its transmit power across those tracks. The division may be equal or weighted by the confidence that the sensor has in the track. By directing more transmitter power at tracks that the sensor has a greater confidence in, the track will be given greater weight when fused with other sensor estimates. This allows a “fair” distribution of sensor transmission resources over all possible transmittable reports. Each track is spread across the channel spectrum by a (pseudo) orthogonal code such as Walsh or Gold code and will cause interference with other tracks such that as the number of tracks monitored by the system increases the performance of the individual tracks will degrade. The use of CDMA implies that interference in the channel will be seen as
non-correlated noise for all noise sources that do not share the same “chip” sequence as will other tracks that are unrelated.

Each track is assigned $m+1$ orthogonal codewords for each $m$ degrees of freedom in the innovation vector from a total pool of $N$ codewords. The codewords may be separate pseudo orthogonal codes (for a CDMA system). The codewords are characterised by their co-channel interference and their selection should attempt to minimise this co-channel interference. Each sensor then transmits its sequence of (pseudo) innovations, encoding the innovation by distributing its total transmit power across all $m+1$ codewords such that the weighted sum of the codewords is equal to the innovation. The total transmitted signal is compensated for loss and multipath effects as required using the parameters derived by the open loop channel estimator to ensure that transmitters on different paths generate signals that are heard at the receiver at approximately the same power level.

Each sensor maintains an augmented local state estimate $\{x_k, P_k\}$ based only on its local observations. The local state estimates are used to generate innovation like sequences using the observation process (5.88). We can form a pseudo innovation sequence that contains the part of the innovation of the linear state estimate that is not described by an approximate linear system with $\zeta \neq 1$ selected such that a late joiner will converge on the state estimate sufficiently quickly. Provided both the transmitter and receiver agree on the parameters of the augmented linear approximation $F$ and $Q$ the pseudo innovations can be used to recover and fuse the state estimates. The pseudo observations are generated using (5.95) and (5.97). These represent equivalent observations with innovation like characteristics that will provide at least the same amount of information when used in the reference system and allow recovery of the local estimate $\{x_k^R, P_k^R\}$ using (5.99) and (5.98). As the pseudo measurements are
transmitted in the form of \( \{ R_k^{-1} z_k, R_k^{-1} \} \) they may be fused directly using (6.122) and (6.117). Consequently the fused estimate is the best linear unbiased estimate of the sensors local state estimates (irrespective of the sensors actual measurement process).

Consider a sensor that has an estimate of the likelihood that the transmitted (pseudo)measurement is correctly allocated to the system observed. This may be due to the sensor performing multi-hypothesis estimation, or because the sensor is transmitting pseudo measurements based of a non-linear state estimation process. Consider the \( i \)th sensor that has formed a non-linear estimator for the state variable \( x_i \sim p(x_i|z_{1:k}) \) and formed the local linear estimate based on the pseudo measurements transmitted, \( x_i^R \sim N(\hat{x}_i^R, P_i^R) \) where \( \{ \hat{x}_i^R, P_i^R \} \) are formed according to (5.99) and (5.98). We can describe a weight for the pseudo observation as the probability that the observation is a true observation as

\[
\alpha_i = \int \Pr(x_i|x_i^R) \, dx_i \quad (6.139)
\]

\[
= \int N(x : \hat{x}_i^R, P_i^R) \, p(x|z_{1:k}) \, dx \quad (6.140)
\]

For each sensor we note that the expected fused posterior distribution at the receiver is given by

\[
\mathbb{E}\left(P_{k|j}^{-1}\right) = \mathbb{E}\left(P_{k|j-1}^{-1}\right) + \sum_{i=1}^{N} \alpha_i R_i^{-1} \quad (6.141)
\]

\[
\mathbb{E}\left((P^{-x}_{k|j})_{k|j}\right) = \mathbb{E}\left((P^{-x}_{k|j})_{k|j-1}\right) + H' \sum_{i=1}^{N} \alpha_i R_i^{-1} z_{i,k} \quad (6.142)
\]

Normally the sensors transmitted signal satisfies (6.127), so that each sensors transmission is summed with equal weight. However, as (6.127) is a linear function, simply reducing the total power allocated by a factor \( \alpha_i \) will have the effect of weighting the observation as per (6.139). The benefit is that sensors with less certain measurements can reduce the amount of energy spent transmitting these measurements and allocate more energy to the transmission of measurements that have a greater level of significance. This also indicates that if a particular sensor experiences fading that cannot be compensated by feed-forward power estimation the weight allocated to the sensor in the fusion process will be reduced. However if the receiver experiences a fade which effects all channels, the impact will be less pronounced (beyond that caused by the reduction in SNR). Finally we note that as the sensor must achieve a particular power level at the receiver, the amount of power required at the sensor depends solely on the propagation loss of the transmission system.

Note that while the expectations of the posterior is exact the prior recursions are an upper bound. Using the definitions of (5.102) and taking expectations on (5.105)
and (5.107) we obtain

\[ E\left( P_{k|k-1}^{-1} \right) = \begin{bmatrix} Q^{-1} & -Q^{-1}F \\ -F'Q^{-1} & E(A) - E(BD^{-1}B') + F'Q^{-1}F \end{bmatrix} \] \tag{6.143}

and

\[ E\left( (P^{-1}\hat{x})_{k|k-1} \right) = \begin{bmatrix} 0 & E(P^{-1}x)_{k-1} - E(BD^{-1}(P^{-1}x)_{k-2}) \\ E(P^{-1}x)_{k-1} - E(B)E(D^{-1})E(P^{-1}x)_{k-2} \end{bmatrix} \] \tag{6.145}

where \( B, D \) are the blocks of \( P^{-1} \) as defined in (5.103).

### 6.10 Simulation

The final simulation incorporates elements from all previous discussions. By using pseudo-innovations as described in (5.6) the amount of information to be transmitted
is minimised. The use of the pseudo innovations reduces the range of values that needs to be transmitted, reducing the impact of noise on the recovered estimator. By reducing the amount of information and restricting the range of values to be transmitted both the co-operative shared channel and discrete joint encoded separated channels can select efficient transmission encodings that transmit less redundant information. By selecting an appropriate $\zeta$ the estimator maintains stability even in the presence of missing observations without any other form of error management such as retransmission requests.

For comparison purposes a joint encoder is assumed to operate at the rate distortion bound of a central estimator as described in [38]. As the usage of a digital system requires each sensor to be allocated a discrete channel it is assumed that the number of sensors can be easily determined and consequently the summation required to perform the data fusion can be achieved through the use of the central estimator multiplied by the number of sensors. The co-operative encoding uses the same form of information fusion but allows the information to be summed within the channel rather than requiring separate channels for each sensor. The performance of both systems under identical conditions is compared, the impact of channel noise on the different transmission systems being demonstrated.

The sensors utilise a common 1kHz bandwidth channel. All transmitters have a fixed maximum transmit power such that the received SNR is constant. The channel access method is assumed ideal and imposes no overhead, allowing each sensor using digital modulation to access an even share of the channel bandwidth at the Hartley-Shannon limit. The quantisation noise generated by the equivalent digital system is assumed to satisfy the limit specified. The system is generating the MMSE for an intermittently observable target ($\alpha = .7$) with $\zeta$ set to .95. The sensor system is assumed to use a CEO based transmission system that achieves the lower bound transmission rate.

Figure 6.8 shows that under these constraints there exists a point where the use of co-operative transmission exceeds the performance achievable at the rate distortion bound of a joint encoded system. There exists a point where the received error is reduced compared to the use of a joint encoded central estimator. The simulation is similar to the previous examples being generated using 500,000 randomised observations for each scenario. Under the conditions set for the scenario, 100 sensors would be better deployed using co-operative encoding than using an optimal CEO estimator. The decoding of CEO transmissions is NP-complete and appears to be a NP-hard problem which will not scale to extremely large numbers of sensors.
6.11 Conclusion

This Chapter has introduced the concept of Co-operative transmission, a method that allows a large number of sensors to pool their transmitter power in order to generate a state estimate based on their collective observations. In 6.2 we see that by eq.(6.2) and (6.5) do not allow us to increase the number of transmitting sensors without bound. Depending on the RF environment they will rapidly jam each other with their mutual interference.

The primary contribution of this thesis follows in the subsequent sections with the development of a Co-operative transmission system that uses less power and provides greater accuracy than a jointly encoded sensor system. This is developed in 6.9 and the performance is presented in 6.10. Because the transmitters are allowed to constructively interfere the amount of power required by each individual sensor can be significantly less than that required for a purely digital channel access method, vindicating the choice of a merged channel access, transmission and fusion strategy.

Although Walsh codes are unsuited for use in digital CDMA uplinks due to multi-path interference and timing accuracy causing the codes to not be orthogonal they are highly suited in this application. As each transmitter is transmitting using multiple orthogonal codes the timing relationship between different Walsh codes is preserved, and consequently multi-path interference can be controlled with conventional techniques (Rake receivers). Furthermore we want constructive interference between transmitters so that the total energy allocated to each code is summed. The receiver operates by summing up the total received power, and the received power in each orthogonal code. As the receive and transmit paths are similar a sensor can estimate the amount of power required for the power received at the receiver to be a constant value. This allows the transmitter to reduce the transmit power and consequently improve battery life, as well as not overly weighting the contributions of closer sensors.

The most significant outcome is that there is no limit to the number of sensors that can be accommodated by a fixed bandwidth channel provided the data fusion process can be expressed as weighted sum of values. In the system described the summation is of Gaussian distributed pseudo innovations, but this could be extended to more complex distributions such SoG or where each codeword describes the weight associated with an individual particle in a particle filter or for a single hypothesis in a multi-hypothesis estimator.

A further point to note is that the amount of power required by a single sensor is less than the power required for an equivalent digital transmission system. For the digital system, a sensor must transmit with enough power to satisfy the SNR requirements for the chosen modulation scheme. Too little power will cause the data to be irrecoverably
lost, and the power used for transmission wasted. In the co-operative transmission system, even if the amount of power used by a single sensor is considerably less than a digital system, provided the aggregate of all sensors transmissions is sufficient no sensors individual contribution is completely wasted. Thus individual sensors can be configured to use much less power to transmit their observations.

The single requirement is that the fusion process be expressible in a manner such that it becomes a summation in the transmission media. In this way the performance is limited by the noise floor and linearity of the transmitters rather than the characteristics of the channel as the number of sensors becomes large. As discussed in previous sections a system using co-operative transmission can out-perform a system using digital channels and either joint encoding or using CEO encoder as these encoders, by necessity must include additional information to allow the receiver to separate the individual transmissions. The co-operative encoder makes it impossible for the receiver to determine what any individual sensor transmitted, or even how many sensors exist. Nor does it require a channel access control to allocate sensors to channels allowing it to scale to much larger numbers of sensors than a digital system.
Chapter 7

Conclusion

The concern throughout this thesis has been how to take advantage of an ubiquitous sensor system. A “dense” system where the number of sensors able to observe an event is much greater than the capacity of the communications infrastructure to transmit the information naively. While current technology is not yet pushing these limits it is certainly foreseeable that the computing capability and sensor acuity will continue to improve toward fundamental physical limits (such as Landauer’s limit, an application of the Second Law of Thermodynamics as applied to computing) while the ability to transmit information in a digital form over a noisy channel is already close to the theoretical limits (with Turbo codes able to approach within a dB of the Hartley-Shannon limit of a noisy channel). The balance of power consumption, size and volume will shift inexorably towards the transmission system as the computing system still has opportunity to shrink before reaching physical limits.

From this perspective it would seem that the ability to improve the utilisation of channel resources by reducing power requirements will be one of the enabling factors required for utilisation of ubiquitous networks. As demonstrated in the first part of this thesis there is little room for improvement in system capability when systems are built with increasing numbers of sensors without concurrent improvements in communications capacity. Optimization of observation and transmission timings can yield modest improvements, and certain forms of media access control have negative impacts, but overall the additional sensor capability cannot be effectively utilised to provide increased accuracy. The dream of multiple low cost, low accuracy sensor replacing the capability of a single high cost, high accuracy sensor would founder on the extremely high communications bandwidth required for fusing the information into a single estimation.

The better approach to the utilisation of additional sensors is to not only provide additional information but to also improve the ability to transmit that information to
the observer by cooperatively forming the channel modulation. This allows sensors to
assist in transmitting the fused estimate rather than compete for bandwidth resources.
By forming a co-operative transmission system a collection of sensors can achieve lower
power requirements.

In Chapter 3 of this thesis we examined the implications of data-transmission side-
effecting on state estimation. In order to efficiently discuss this a revised expression for
the recursion of the prior covariance of a linear dynamical system was developed for a
system that has missing observations. With this framework we were able to develop
analytic expressions suitable for use as first order approximations in the design of sensor
networks. In particular we were able to demonstrate the coupling between additional
sensors and improved estimation quality and the type of message loss process. This
process confirmed that for simpler sensor cluster architectures, the addition of sensors
rarely provided any performance benefit, or the benefits where marginal at best. In
some common cases the addition of sensors reduced the performance of the system.

Of particular note is that we can bundle the intrinsic performance of the sensor
system as a single metric which relates the ability of a sensor compared to uncertainty
of the system being observed. Note that (3.16) shows that we can bundle the process
noise and sensor noise into a single matrix $D$, independent of the number of sensors $n$.
This is applicable to establishing the required number of sensors to achieve a particular
covariance given a given sensor utility expressed as $D$. For a fixed set of system
requirements $\{\alpha, P, D, F\}$ we can determine the required sensor count, furthermore its
noted that to a first approximation that $\alpha n$ held constant will maintain a constant
system performance.

Chapter 4 extended on the results of Chapter 3 and demonstrated that the media
access method used by the sensors to co-ordinate access to the transmission network
has an impact on the performance of the estimator system. Of particular interest it
was demonstrated that CSMA-CD based systems have a rapid reduction in system
sensitivity to additional sensors as the sensor count is increased, the relation being
inverse cubic. The other important observation is that for a common class of problems,
where the accuracy of the measurement is related through an inverse square law to the
distance of the sensor from the object being measured. In this case increased sensor
density reduces the performance of the system. Low accuracy measurements may
block the reception of better quality measurements, causing the system performance
to degrade rapidly.

These two chapters demonstrate that the sensor network and the way the net-
work interacts with the sensors can significantly alter the performance of a sensor
system, either reducing the benefits of additional sensors or causing the network to
significantly degrade. This is the motivation for the second section of the thesis. To
demonstrate a method that will always allow the performance of the system to improve with additional sensors without limit.

In Chapter 5 we started to explore techniques to mitigate problems in message loss and to improve the efficiency of the digital link between sensors and an information consumer. By carefully restructuring the measurement process the system can be tuned to provide various levels of message loss resilience and bandwidth efficiency without requiring retransmission or bit exact error correction. The major contribution in this section is the Innovation-Like sequence, which allows the generation of transmission symbols that can arbitrarily approach an innovation in terms of information efficiency but allow recovery from lost transmission without retransmission in a bounded number of messages.

The ability to reduce the amount of data transmitted allows denser swarms of sensors through the reduction of the amount of transmitted data. Furthermore, as the techniques discussed do not require retransmission it is possible for a fusion centre to be a passive member of the sensor network, joining and leaving the network while the sensors and other outside observers remain unaware.

In Chapter 6 we introduce the co-operative transmitter swarm and demonstrate that CEO and Joint Encoding methods as currently understood cannot achieve the same spectral efficiency as these systems intrinsically include sufficient information to identify the source of all observations down to the individual sensor. While it is possible to determine some information about an individual sensors observations it is apparent that more information than is necessary is being transmitted. This is heuristically understandable by considering that a CEO estimator each sensor must be granted a transmission channel to transmit its observation. Thus every transmission by the sensor encodes two pieces of information. Firstly information to be used by the CEO estimator for generating the estimate. Secondly it is providing the identity of the estimator sending that information. Thus an observer can determine which observer produced each observation as this can be inferred by the allocation of sensors to channels. Co-operative transmission does not allow an observer to allocate sensor measurements to sensors, and consequently requires less power to generate the same accuracy of estimate.

The combination of co-operative transmission with Innovation-Like sequences allows the efficient fusion of observations even when the observations are taking place with non-linear sensors such as distance bearing trackers and non-linear processes, provided the actual process models used by the transmitter and fuser are known and linear, as the innovation like sequence allows the efficient transmission of the state estimate and state estimate covariance.

Co-operative transmission is only of assistance for certain classes of sensor network,
those which have a great number of redundant low quality sensors from which a single high quality state estimate is to be formed. It is particularly helpful where bandwidth or transmitter power is in short supply for an individual sensor and it scales easily to large sensor collections. Some potential applications would be for multistatic radar and lidar based systems where each sensor is a low resolution transmitter or receiver but given a large number of sensors distributed over a wide baseline would allow good performance of a sensor “cloud”. In addition it has the nice properties of redundancy and resilience without requiring the receiver to co-ordinate with the sensors to correct transmission errors. It is not a substitute for CEO or joint encoded digital systems for smaller numbers of sensors and all systems have a transition point where one scheme or the other is optimal for both spectral efficiency and sensor power consumption.

An interesting consequence of co-operative transmission is that it allows us to sidestep the issue of channel capacity. Rather than each sensor competing for capacity to transmit an even smaller piece of incremental information it allows the use of a single channel that only requires as much bandwidth as that required by the ideal sensor to transmit the completely fused state estimate. The system described in this thesis simply uses a asynchronous receiver measuring power levels in orthogonal channels to perform the linear summation operations required for data fusion. It is possible that synchronous encoding schemes could be developed to further improve the encoding efficiency as could polarisation. A photonic system could be formed by observing the sensor network with a single detector, somewhat like observing the aggregate light from a swarm of fireflies.

In conclusion the techniques outlined allow the engineer to design and implement dense sensor networks that scale optimally and simply without destructive self interference and complex NP-hard decoding methods while achieving performance that is equivalent to or better than current techniques.
Chapter 8

Appendix

**Theorem 3** \( E \left( (X^{-1} + Y)^{-1} \right) \leq \left( E(X)^{-1} + Y \right)^{-1} \) for all \( X, Y \) positive (semi) definite.

**Proof.** Consider \( (X^{-1} + Y)^{-1} \). We can always write \( Y = W \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} W' \) where \( W \) is invertible for all \( X, Y \) positive (semi) definite. Consequently we need to determine only if

\[
W'^{-1} \left( Z^{-1} + \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \right)^{-1} W^{-1}
\]

(8.1)

\[
Z = WXW'
\]

(8.2)

Consider that we can decompose \( Z \in S^{m \times m} \) using an \( LDL' \) decomposition as

\[
Z = \begin{bmatrix} U_1 & 0 \\ U_{12} & U_{22} \end{bmatrix} \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix} \begin{bmatrix} U'_1 & U'_{12} \\ 0 & U'_{22} \end{bmatrix}
\]

(8.3)

where \( U_1 \) is unitary and \( D_1, D_2 \) are eigenvalues of \( Z \). Using 8.3 we obtain

\[
\left( Z^{-1} + \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \right)^{-1}
\]

(8.4)

\[
= \left( (LDL')^{-1} + \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \right)^{-1}
\]

(8.5)

\[
= \begin{bmatrix} U_1 & 0 \\ U_{12} & U_{22} \end{bmatrix} \begin{bmatrix} D_1 - D_1 \left( D_1 + (U'_1U_1)^{-1} \right)^{-1} D_1 & 0 \\ 0 & D_2 \end{bmatrix} \begin{bmatrix} U'_1 & U'_{12} \\ 0 & U'_{22} \end{bmatrix}
\]

(8.6)

\[
= \begin{bmatrix} U_1 & 0 \\ U_{12} & U_{22} \end{bmatrix} \begin{bmatrix} (D_1^{-1} + I)^{-1} & 0 \\ 0 & D_2 \end{bmatrix} \begin{bmatrix} U'_1 & U'_{12} \\ 0 & U'_{22} \end{bmatrix}
\]

(8.7)
Note that \((D_1^{-1} + I)^{-1}\) is a concave function. Consequently the eigenvalues of

\[
\left( Z^{-1} + \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \right)^{-1}
\]

must be concave with respect to \(Z\), as must they be with respect to \(X\) as \(X\) is linearly related to \(Z\), and so by Jensen’s inequality we have (where all subsequent matrix inequalities are treated in the Loewner sense) we must have

\[
\mathbf{E} \left( (X^{-1} + D)^{-1} \right) \leq \left( \mathbf{E} (X)^{-1} + D \right)^{-1}
\]

\(8.9\)

Theorem 4 Any linear Gaussian system with \(n\) observers converges irrespective of the sensor scheduling provided the expected scheduling rate \(\lambda(n)\) for \(n\) sensors satisfies

\[
\frac{2\lambda_{\text{max}}(A)}{n} < \lambda(n)
\]

\(8.10\)

or equivalently the probability of any particular sensor report being received \(\alpha(n)\) is

\[
1 - \frac{1}{\lambda_{\text{max}}(e^{2A/n})} < \alpha(n)
\]

\(8.11\)

provided the loss process is independent where \(S_1, S_2\) system is are as defined in \((3.65)\) and \((3.66)\).

Proof. Taking expectations for \(S_1\) we have

\[
\mathbf{E} \left( P(S_1)_{k+1|k} \right) = F \left( \sum_{i=1}^{n} \alpha^i (1 - \alpha)^{n-i} \mathbf{E} \left( P(S_1)_{k|k-1, y_{i,1:k-1}}^{-1} + iH'R^{-1}H \right)^{-1} \right) F' + Q
\]

\(8.12\)

\[
> F \left( \sum_{i=1}^{n} (1 - \alpha)^{n-i} \mathbf{E} \left( P(S_1)_{k|k-1}^{-1} + nH'R^{-1}H \right)^{-1} \right) F' + Q
\]

\(8.13\)

\[
< F \left( \sum_{i=1}^{n} (1 - \alpha)^{n} P(S_1)_{k|k-1}^{-1} + H'R^{-1}H \right)^{-1} F' + Q
\]

\(8.14\)
8. Appendix

Note that the upper and lower bounds differ only in terms of the the observation $nH'R^{-1}H$ and by Theorem 2 $P(S_1)_{k+1|k}$ converges iff

$$1 - \frac{1}{\lambda_{\max}(F)^2} < 1 - (1 - \alpha)^n$$

(8.16)

$$1 - \frac{1}{\lambda_{\max}(F)^2/n} < \alpha$$

(8.17)

Taking expectations for $S_2$ we have

$$E(P(S_2)_{k+1|k}) = F^{1/n} \left( \alpha E\left( \left( P(S_2)_{k|k-1}^{-1} + \gamma_{i,k}H'R^{-1}H\right)^{-1} \right) + (1 - \alpha) E(P(S_2)_{k|k-1}) \right) F^{1/(8.18)}$$

(8.18)

$$+ Q(n)$$

(8.19)

which by Theorem 1 converges iff

$$1 - \frac{1}{\lambda_{\max}(F^{1/n})^2} < \alpha$$

(8.20)

which is equivalent to $S_1$. i.e. both systems utilizing different ways of scheduling $n$ sensors have identical convergence criteria. ■

**Theorem 5** Given independent, exponentially distributed observations at rate $\lambda$ the mean covariance of the estimator will be finite provided $\frac{\lambda}{2} > \lambda_{\max}(A)$ where the system is defined as (3.65) , (3.66) and (3.67),

**Proof.** Note that we can define a single sensor system as

$$F_k = \exp(At_k) = F^{t_k}$$

(8.21)

$$\text{vec} \ Q_k = (A \otimes I + I \otimes A)^{-1} \text{vec}(e^{At_k}q e^{At_k'} - q)$$

(8.22)

$$\Pr(t_k = t) = \exp(-\lambda t)$$

(8.23)

$$S_3 \left\{ \begin{array}{l}
    x_{k+1} = F_k x_k + \sqrt{Q_k} u_k \\
    y_k = H x_k + \frac{1}{\tau_k} \sqrt{R} v_k \\
    \mathbb{E} (P_{k+1|k}) = \mathbb{E} \left( F_k \left( P_{k|k-1}^{-1} + H'R^{-1}H\right)^{-1} F_k' + Q_k \right)
\end{array} \right\}$$

(8.24)

where $t_k$, the time between successful observations is exponentially distributed. Taking
vecs we can write

\[
\mathbf{E}\left(\text{vec } P(S^i_{k+1|k})\right) = \mathbf{E}\left(e^{At_k} \otimes e^{At_k} \text{vec}\left(\left(P_{k|k-1}^{-1} + I_M\right)\right)
+ (A \otimes I + I \otimes A)^{-1} \text{vec}\left(e^{At_k}q e^{At_k'} - q\right)\right)
= \mathbf{E}\left((e^A \otimes e^A)^t_k\right)\left(\text{vec}\left(\left(P_{k|k-1}^{-1} + I_M\right)\right)
+ (A \otimes I + I \otimes A)^{-1} \text{vec } q\right)
- (A \otimes I + I \otimes A)^{-1} \text{vec } q
\]  

(8.25)

(8.26)

(8.27)

To proceed further we need to be able to evaluate \(\mathbf{E}\left((e^A \otimes e^A)^t_k\right)\). Note that we can write in Jordan form as

\[
F^t_k = e^{At_k} = Je^{(D+N)t_k} \pi J^{-1}
\]

(8.28)

\[
e^{Ak} \otimes e^{Ak} = Je^{(D+N)k}J^{-1} \otimes Je^{(D+N)k}J^{-1}
= (J \otimes J)\left(e^{(D+N)k} \otimes e^{(D+N)k}\right)(J \otimes J)^{-1}
\]

(8.29)

(8.30)

for eigenvectors \(J\) (including generalized eigenvectors), \(D\) a diagonal matrix as \(N\) a nil-potent matrix where \(D + N\) forms the Jordan Blocks of \(A\).

\[
= (J \otimes J)\left(e^{Dk} \otimes e^{Dk}\right)\left(e^{Nk} \otimes e^{Nk}\right)(J \otimes J)^{-1}
\]

(8.31)

Note \(e^{Nk}\) is unit upper triangular with all eigenvalues having a magnitude of 1- the only effect of \(k\) is to alter the eigenvectors. Consider the diagonal entries \(D_j\) of \(D\)

\[
\Pr(t_{k,j} = D_j t) = \frac{\lambda}{D_j} \exp\left(-\frac{\lambda t}{D_j}\right)
\]

(8.32)

Note that \(e^{tk}\) is Pareto distributed so we can directly write

\[
\mathbf{E}\left(e^{tk}\right) = \frac{\lambda}{\lambda - D_j} : D_j < \lambda
\]

(8.33)

\[
\mathbf{E}\left(e^{tk}\right) = \left(I - \frac{1}{\lambda} D\right)^{-1}
\]

(8.34)

where the mean exists only for \(D_j < \lambda\). Consequently we can decompose \(\mathbf{E}\left(e^{Ak} \otimes e^{Ak}\right)\) as

\[
\mathbf{E}\left(e^{Ak} \otimes e^{Ak}\right) = (J \otimes J)\mathbf{E}\left(\left(e^{Dk} \otimes e^{Dk}\right)\left(e^{Nk} \otimes e^{Nk}\right)\right)(J \otimes J)^{-1}
= (J \otimes J)\left(I - \frac{1}{\lambda} (D \otimes I + I \otimes D)\right)^{-1} \mathbf{E}\left(e^{Nk} \otimes e^{Nk}\right)(J \otimes J)^{-1}
\]

(8.35)

(8.36)

(8.37)

(8.38)
If $F$ is normal (i.e. - has no generalized eigenvectors) then $N = \begin{bmatrix} 0 & \cdots & \cdots \\ \vdots & \ddots & \ddots \end{bmatrix}$ and

$$E \left( e^{Nk} \otimes e^{Nk} \right) = (I \otimes I)$$

and we can write

$$E \left( e^{Ak} \otimes e^{Ak} \right) = \left( I \otimes I \right) - \frac{1}{\lambda} \left( A \otimes I + I \otimes A \right)$$

(8.39)

while if $F$ is not normal we can write

$$\lambda_{\text{max}} \left( E \left( e^{Ak} \otimes e^{Ak} \right) \right) < \lambda_{\text{max}} \left( \left( I \otimes I \right) - \frac{1}{\lambda} \left( A \otimes I + I \otimes A \right) \right)$$

(8.40)

Consequently as $E \left( \left( P_{k|k-1}^{-1} + I_M \right)^{-1} \right)$ is independent of $t_k$ we can write (for all $F$)

$$E \left( \text{vec} \, P_{k+1|k} \right) \leq \left( I \otimes I - \frac{1}{\lambda} \left( A \otimes I + I \otimes A \right) \right)^{-1} \times$$

$$\left( \text{vec} \left( E \left( \left( P_{k|k-1}^{-1} + I_M \right)^{-1} \right) \right) + (A \otimes I + I \otimes A)^{-1} \text{vec} \, q \right)$$

(8.41)

$$- (A \otimes I + I \otimes A)^{-1} \text{vec} \, q$$

(8.42)

Define $Y_k = E \left( P_{k|k-1}^{-1} \right)$

$$\text{vec} \, Y_{k+1} < \left( I \otimes I - \frac{1}{\lambda} \left( A \otimes I + I \otimes A \right) \right)^{-1} \times$$

$$\left( \text{vec} \left( Y_k^{-1} + I_M \right)^{-1} \right) + (A \otimes I + I \otimes A)^{-1} \text{vec} \, q$$

(8.43)

which will converge for all

$$\left( I - \frac{1}{\lambda} \left( A \otimes I + I \otimes A \right) \right) > 0$$

(8.47)

i.e. $\frac{1}{\lambda} > \lambda_{\text{max}} \left( A \right)$. Noting that the exponential distribution is the limiting case of the geometric distribution, we can write the characteristic $\alpha$ for the equivalent discrete geometric distribution and set $\alpha = 1 - \exp^{-2\lambda_{\text{max}} \left( A \right)}$. Using our previously established convergence requirement

$$1 - \lambda_{\text{max}} \left( F \right)^{-2/n} < \alpha : n = 1$$

(8.48)

and substituting $F = e^A$ and taking the expression for a geometric distribution at its limit $\alpha = 1 - \exp^{-\lambda}$ we have equivalent requirements

$$1 - \lambda_{\text{max}} \left( e^A \right)^{-2} < 1 - \exp^{-\lambda}$$

(8.49)

$$\lambda_{\text{max}} \left( e^A \right)^{-2} > \exp^{-\lambda}$$

(8.50)

$$\lambda_{\text{max}} \left( A \right) < \frac{\lambda}{2}$$

(8.51)
which indicates that the result converges at the limit with the previous results. How will increasing the number of sensors affect this result? For \( n \) sensors we can write the requirement as (and setting the rate \( \lambda \) as dependent on the number of sensors)

\[
\lambda_{\text{max}} (A) < \frac{n\lambda(n)}{2}
\]  

(8.52)

Noting that the exponential distribution is the limiting case of the discrete geometric distribution we can write the \( \alpha \) equivalent by setting

\[
1 - \exp (-\lambda(n)) = \alpha (n)
\]  

(8.53)

and writing

\[
\lambda_{\text{max}} \left( \exp \left( \frac{A}{\sqrt{n}} \right) \right) < \exp (\lambda(n))
\]  

(8.54)

\[
\lambda_{\text{max}} (F^{-\frac{2}{\pi}}) > \exp (-\lambda(n))
\]  

(8.55)

\[
\lambda_{\text{max}} (F^{-\frac{2}{\pi}}) > 1 - \alpha (n)
\]  

(8.56)

\[
\alpha (n) > 1 - \lambda_{\text{max}} \left( F^{-\frac{2}{\pi}} \right)
\]  

(8.57)

which is consistent with the previous results. ■

**Theorem 6** The steady state RMS error of the state estimator will be reduced beyond that of the ideal single sensor system with increased number of sensors provided that \( \alpha (n) \) satisfies

\[
\alpha (n) > \frac{1 - \lambda_{\text{max}} \left( e^{-A/n} (I + \tilde{\theta}^{\dagger}) e^{-A/n} \right)}{1 - \lambda_{\text{max}} \left( e^{-A} (I + \tilde{\theta}^{\dagger}) e^{-A} \right)}
\]  

(8.58)

where

\[
\tilde{\theta}^{\dagger} = \sqrt{P^{-1}} \tilde{\theta} \sqrt{P^{-1}}
\]  

(8.59)

\( P \) is the prior covariance for a single sensor and no losses and \( \tilde{\theta} \) is

\[
\text{vec} \tilde{\theta} = \left( (e \otimes e)^{A/n} - (I \otimes I) \right)^{-1} \text{vec} Q(n)
\]  

(8.60)

and

\[
P = e^{A} \left( P^{-1} + D \right)^{-1} e^{A^T} + Q
\]  

(8.61)

**Proof.** Consider \( N \) sensors transmitting at equispaced intervals. From previous results we can write out the equations

\[
P = F(n) \left( P^{-1} + I_{M} \right)^{-1} F(n)^{T} + Q(n)
\]  

(8.62)

\[
F(n) = e^{A/n}
\]  

(8.63)

\[
Q(n) = \int_{0}^{1/n} e^{A\tau} \chi e^{A\tau} d\tau
\]  

(8.64)
Taking vecs we get
\[
\text{vec } Q(n) = \int_0^{1/n} (e^A \otimes e^A)^T \text{vec } \chi \, d\tau \\
= \left((e \otimes e)^{A/n} - (I \otimes I)\right) (A \otimes I + I \otimes A)^{-1} \text{vec } \chi \\
= \left((e \otimes e)^{A/n} - (I \otimes I)\right) \text{vec } \vartheta
\] (8.65)
(8.66)
(8.67)

As before we assume observation \(\gamma_k = 1\) indicates a successful observation and \(\mathbb{E}(\gamma_k) = \alpha\). We can define an upper bound given \(n\) sensors.

\[
\mathbb{E}(P_k(n, \alpha)) = F(n) \left( \mathbb{E} \left( \gamma_k \left( P_{k-1}(n, \alpha)^{-1} + D \right)^{-1} \right) + (1 - \gamma_k) P_{k-1}(n, \alpha) \right) F(n)' \\
+ Q(n)
\] (8.68)
(8.69)
(8.70)
\[
\mathbb{E}(\gamma_k) = \alpha
\] (8.71)
\[
\mathbb{E}(P_k(n, \alpha)) < P_k(n, \alpha) = F(n) \left( \alpha \left( P_{k-1}(n, \alpha)^{-1} + D \right)^{-1} \right) + Q(n)
\] (8.72)
\[
\overline{P_{n,\alpha}} = \lim_{k \to \infty} P_k(n, \alpha)
\] (8.73)

Taking vecs and expanding we can write this as
\[
\text{vec } \overline{P_{n,\alpha}} = \alpha (e^A \otimes e^A)^{1/n} \text{vec } \left( \overline{P_{n,\alpha}}^{-1} + D \right)^{-1} \\
+ (1 - \alpha) (e^A \otimes e^A)^{1/n} \text{vec } \overline{P_{n,\alpha}} \\
+ \left((e^A \otimes e^A)^{1/n} - (I \otimes I)\right) \text{vec } \vartheta
\] (8.74)
(8.75)
(8.76)

Note that
\[
\text{vec } P_{1,1} = (e^A \otimes e^A) \text{vec } \left( P_{1,1}^{-1} + D \right)^{-1} \\
+ \left((e^A \otimes e^A) - (I \otimes I)\right) \text{vec } \vartheta
\] (8.77)
(8.78)

and define
\[
\text{vec } \Theta = \text{vec } \left( P_{1,1}^{-1} + D \right)^{-1}
\] (8.79)

then for all \(\text{vec } \overline{P_{n,\alpha}} < \text{vec } P_{1,1}\) we can write the inequality
\[
\text{vec } \overline{P_{n,\alpha}} < \alpha (e^A \otimes e^A)^{1/n} \text{vec } \Theta \\
+ (1 - \alpha) (e^A \otimes e^A)^{1/n} \text{vec } \overline{P_{n,\alpha}} \\
+ \left((e^A \otimes e^A)^{1/n} - (I \otimes I)\right) \text{vec } \vartheta
\] (8.80)
(8.81)
(8.82)
8. Appendix

Solving (8.80) we have

\[ \text{vec} \mu_{n,\alpha} < (I \otimes I) - (1 - \alpha) (e^A \otimes e^A)^{1/n} \] \[ \cdot (\alpha (e^A \otimes e^A)^{1/n} \text{vec} \Theta + ((e^A \otimes e^A)^{1/n} - (I \otimes I) \text{vec} \vartheta) ) \] \[ (8.83) \]

We desire a relationship \( \alpha (n) \) s.t. \( \text{vec} \mu_{n,\alpha} < \text{vec} P_{1,1} \). Using (8.77) and (8.79) and substituting into (8.80) we require

\[ \text{vec} P_{1,1} > (I \otimes I) - (1 - \alpha) (e^A \otimes e^A)^{1/n} \] \[ \cdot (\alpha (e^A \otimes e^A)^{1/n} \text{vec} \Theta + ((e^A \otimes e^A)^{1/n} - (I \otimes I) \text{vec} \vartheta) ) \] \[ (8.85) \]

which after some re-arrangement becomes

\[ \text{vec} P_{1,1} - (e^A \otimes e^A)^{-1/n} \text{vec} (P_{1,1} + \vartheta) \]

\[ < \alpha \left( \text{vec} P_{1,1} - (e^A \otimes e^A)^{-1/n} \text{vec} (P_{1,1} + \vartheta) \right) \] \[ (8.90) \]

noting that

\[ e^{-A/n} = \sqrt{P_{1,1}^{-1}} e^{-A/n} \sqrt{P_{1,1}} \] \[ (8.92) \]

has the same eigenvalues as \( e^{-A/n} \) we can write an equivalent relationship

\[ \text{vec} I - (e^{A} \otimes e^{A})^{-1/n} \text{vec} (I + \vartheta) \]

\[ < \alpha \left( \text{vec} I - (e^{A} \otimes e^{A})^{-1/n} \text{vec} (I + \vartheta) \right) \] \[ (8.93) \]

If \( A \) is normal (i.e. has no repeated eigenvectors) then we can directly write an eigenvalue relationship for \( \alpha \) and the maximum eigenvalue (as the eigenvectors of \( (e^{A} \otimes e^{A})^{-1/n} \) and \( (e^{A} \otimes e^{A})^{-1} \) will be the same) then we can write

\[ \lambda_{\text{max}} \left( (I - (e^{-A/n} (I + \vartheta) e^{-A/n})) \left( I - \left( e^{-A} \left( I + \vartheta \right) e^{-A} \right) \right)^{-1} \right) \] \[ (8.95) \]

\[ = \lambda_{\text{max}} \left( I - (e^{-A/n} (I + \vartheta) e^{-A/n}) \right) \] \[ (8.96) \]

\[ = \frac{1 - \lambda_{\text{max}} \left( e^{-A/n} (I + \vartheta) e^{-A/n} \right)}{1 - \lambda_{\text{max}} (e^{-A} (I + \vartheta) e^{-A})} \] \[ (8.97) \]
In the event that $A$ is not normal then

$$\lambda_{\text{max}}\left(\left(I - \left(e^{-A^T/n} \left(I + \vartheta^T\right) e^{-A^T/n}\right)\right) \left(I - \left(e^{-A^T} \left(I + \vartheta^T\right) e^{-A^T}\right)^{-1}\right)\right)^{-1}$$

which implies the condition is not as tight.

**Lemma 1** For $A$ square and $S$ symmetric

$$\text{tr} \left( \ln \left( e^A Se^{A^T} \right) \right) = \text{tr} \left( 2A + \ln \left( S \right) \right)$$

**Proof.** Note that as $S$ is symmetric then $e^A Se^{A^T}$ is symmetric. For symmetric $S$ we can write

$$\text{tr} \left( \ln \left( S \right) \right) = \ln |S|$$

using the properties of tr, ln and $|\cdot|$. Hence we can write

$$\text{tr} \ln \left( e^A Se^{A^T} \right) = \ln |e^A Se^{A^T}|$$

and by $|AB| = |A||B|$ for all $A, B$ square matrices we have

$$= \ln |e^A||S||e^{A^T}|$$

$$= \ln |e^A| + \ln |e^{A^T}| + \ln |S|$$

$$= \ln |e^Ae^{A^T}| + \ln |S|$$

by $e^A e^B = e^{A+B}$ only if $A$ and $B$ are commutative we have $e^A e^{A^T} = e^{A+A^T}$

$$= \ln |e^{A+A^T}| + \ln |S|$$

and returning to traces

$$= \text{tr} \left( \ln \left( e^{A+A^T} \right) \right) + \text{tr} \left( \ln \left( S \right) \right)$$

and taking logs and combining traces

$$= \text{tr} \left( A + A' + \ln \left( S \right) \right)$$

$$= \text{tr} \left( 2A + \ln \left( S \right) \right)$$
Theorem 7 If the system \( \{F, H\} \) is detectable then a sufficient condition for detectability of the augmented system
\[
\left\{ \begin{bmatrix} F & 0 \\ I & 0 \end{bmatrix}, \begin{bmatrix} H & -\zeta F \end{bmatrix} \right\}
\]
(8.110)
is is that \( \zeta \neq 1 \)

**Proof.** Consider \( A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{m \times n} \). The definition of detectability states that system \( \{A, B\} \) is detectable provided for all \( Bx = 0 \) and \( Ax = kx \) that either \( k < 1 \) or \( x = 0 \). Noting that for \( Ax = kx \) to hold that \( x \) must be a right eigenvector we can write the requirements as no eigenvector of \( A \) is in the nullspace of \( B \) unless the eigenvalue is \( < 1 \) or the eigenvector is null. Expanding we evaluate
\[
\begin{bmatrix} F & 0 \\ I & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} Fv_1 \\ v_1 \end{bmatrix}
\]
(8.111)
Note that the augmented system is block triangular, all diagonal blocks are 0 except for the top left block. Hence all eigenvectors associated with the 0 blocks will be in the nullspace and we can write the kernel as
\[
\ker \left( \begin{bmatrix} F & 0 \\ I & 0 \end{bmatrix} \right) = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}
\]
(8.112)
We can ignore the augmented kernel space. The augmented states are not controllable but they are stabilisable given their \( \leq 1 \) eigenvalues. Consider the remaining eigenvectors. If we define \( v_i \) as a right eigenvector of \( F \) such that \( Fv_i = \lambda_i v_i \) we can write
\[
\begin{bmatrix} Fv_i \\ v_i \end{bmatrix} = \begin{bmatrix} \lambda_i v_i \\ v_i \end{bmatrix}
\]
(8.113)
and consequently be equating terms we have the eigenvalues and eigenvectors of the block form as
\[
\begin{bmatrix} F & 0 \\ I & 0 \end{bmatrix} \begin{bmatrix} v_i \\ bv_i \end{bmatrix} = \begin{bmatrix} v_i \\ bv_i \end{bmatrix} = \begin{bmatrix} \lambda_i v_i \\ v_i \end{bmatrix}
\]
(8.114)
\[
\begin{bmatrix} F & 0 \\ I & 0 \end{bmatrix} \begin{bmatrix} v_i \\ \frac{v_i}{\lambda_i} \end{bmatrix} = \lambda_i \begin{bmatrix} v_i \\ \frac{v_i}{\lambda_i} \end{bmatrix}
\]
(8.115)
Writing
\[
H \begin{bmatrix} I & -\zeta F \end{bmatrix} v = H \begin{bmatrix} I & -\zeta F \end{bmatrix} \begin{bmatrix} v_i \\ \frac{v_i}{\lambda_i} \end{bmatrix}
\]
(8.116)
\[
= H \left( 1 - \zeta \right) v_i
\]
(8.117)
From the requirement that \( \{ F, H \} \) be observable we know that \( Hv_i \neq 0 \) and consequently for the augmented system to be observable \( \zeta \neq 1 \). All other eigenvectors of the augmented system are in the nullspace, and having eigenvalues less than one can be ignored. ■
Bibliography


Author/s:
Distel, Robert

Title:
Capacity and efficiency of multi-sensor networks

Date:
2015

Persistent Link:
http://hdl.handle.net/11343/91088

File Description:
Capacity and efficiency of multi-sensor networks