SYSTEM IDENTIFICATION IN WIRELESS RELAY NETWORKS VIA GAUSSIAN PROCESS ITERATED CONDITIONING ON THE MODES ESTIMATION

Ido Nevat\textsuperscript{1}, Gareth W. Peters\textsuperscript{2}, Jinhong Yuan\textsuperscript{3} and Iain B. Collings\textsuperscript{1}

Wireless & Networking Technology Lab, CSIRO, Sydney, Australia\textsuperscript{1},
School of Mathematics and Statistics, University of NSW, Sydney, Australia\textsuperscript{2}
School of Electrical Engineering and Telecommunications, University of NSW, Sydney, Australia\textsuperscript{3}.

ABSTRACT

We present a flexible class of stochastic models that are developed for cooperative wireless relay networks systems, in which the relay processing functionality is not known at the destination. The challenge is then to perform system identification in this wireless relay network. We first construct a statistical model based on a representation of the system using Gaussian Processes. We then develop a computationally efficient algorithm which is based on the Iterated Conditioning on the Modes estimation to undertake system identification for each relay in the presence of partial Channel State Information (CSI). We evaluate the identification performance for different non-linear relay functionalities.

Index Terms— Relay networks, System identification, Gaussian processes, Kernel methods.

1. INTRODUCTION

Cooperative communications systems have been proposed to exploit the spatial diversity gains inherent in multiuser wireless systems without the need for multiple antennas at each node [1]. This is achieved by having the users relay each others messages, thus, forming multiple transmission paths to the destination. Simply put, such a system first broadcasts a signal from a transmitter at the source through a wireless channel, the signal is then received by each relay node and a relay “strategy” is applied before the signal is retransmitted to the destination. A number of relay strategies have been studied in the literature [1]. The relay function can be optimized for different design objectives [2]. For example, in the estimate and forward (EF) scheme, in the case of BPSK signaling, the optimal relay function is the hyperbolic tangent [3]. Other criteria for which the optimal relay function is non-linear include: capacity maximisation [3], minimum error probability at the receiver [4], SNR maximisation [2] and rate maximisation [5].

In general, the system identification problem may arise in several ways, see [6], [7] and references within. For example, in an ad-hoc or sensor networks, it is possible that the destination does not have a priori knowledge of the relay functionality utilised by each of the relays in the system [8]. Alternative scenarios that this problem may arise include networks which can adapt or cognitively learn suitable relay transmission functionality to optimise quality of service constraints such as capacity, throughput, bit-error-rate and transmission power [9]. This is a challenging problem due to the uncertain functional form of each relays processing on the received signal. To address this problem we introduce a semi-parametric modelling procedures based on Gaussian processes (GP). This approach coupled with Iterated Conditioning on the Modes (ICM), which is an iterative optimization method that finds the joint posterior modal estimators corresponding to the MAP estimates procedure [10], will allow us to efficiently solve the system identification problem.

2. GAUSSIAN PROCESS MODELLING

We provide a brief review of Gaussian Process regression and the statistical concepts. We begin with the formal definition of a Gaussian process:

\textbf{Definition 1} [11]: A Gaussian Process is a collection of random variables, any finite number of which have a joint Gaussian distribution.

Furthermore, a GP is completely specified by the equivalent of sufficient statistics for a process, in this instance a mean function, denoted \( m(x; \theta) \) and parameterised by \( \theta \), and a covariance function, denoted \( C(x, x'; \Psi) \). The covariance function is typically selected as a Mercer kernel which is parameterised by \( \Psi \), for discussion on the properties of kernels including stationarity and homogeneity in space or time see [11]. In particular we encode our a-priori belief in the functional form of the relay transform in terms of a prior distribution on a function space via a Gaussian process, denoted by the following prior \( f() \sim GP(m(: \theta), C (: , : \Psi)) \).

As discussed recently in [12] this allows us to encode assumptions about properties such as smoothness and continuity of the possible relay functions we consider.

Formally, this prior model ensures that for any finite set of predictor values or inputs to the unknown regression function \( \{r_i \}_{1:T} \), the corresponding random vector for the function evaluated at these points, given by \( f_{1:T} = [f(r_1), \ldots, f(r_T)] \), is distributed according to the following multivariate Gaussian distribution,

\[
p(f_{1:T} | R_{1:T} = r_{1:T}) = N(f_{1:T} | m(r_1; \theta), \ldots, m(r_T; \theta), \mathbf{K}_{1:T})
\]

where

\[
\mathbf{K}_{1:T} = \mathbf{C} (R_1(t), \ldots, R_T(t)) = \text{Cov} [f(R_1(t)) f(R_T(t))].
\]

3. SYSTEM MODEL

Here we present the system model, depicted in Fig. 1 and associated assumptions. We will generically denote the frame index for the \( t \)-th frame using \( t \in \{1, \ldots, T\} \).

1. Assume a wireless relay network with a single source node, transmitting sequences of \( K \) pilot symbols per frame. We denote the set of pilot symbols for frame \( t \) as \( s_{1,K} \). These symbols are transmitted from a source to a single destination via \( L \) relay nodes.
2. There are \( L \) relays which cannot receive and transmit on the same time slot and on the same frequency band.
We thus consider a half duplex system model in which the data transmission is divided into two steps. In the first step, the source node broadcasts a code word $s$ from the codebook to all the $L$ relay nodes. In the second step, the relay nodes then transmit their signals to the destination node on orthogonal non-interfering channels.

3. Assume a general model for the Channel State Information (CSI) in which the estimates formed from the unknown realised channel coefficients for each relay link are known at the receiver i.e. partial CSI. This involves assumptions regarding the channel coefficients as follows:

- From source to relay there are $L$ i.i.d. channels parameterized by $\{H(l) \sim F(\hat{h}(l), \sigma_H^2)\}_{l=1}^L$, where $F(\cdot)$ is the distribution of the channel coefficients, where the estimated channel, $\hat{h}(l)$, is assumed known at the receiver.
- From relay to destination there are $L$ i.i.d. channels parameterized by $\{G(l) \sim F(\hat{g}(l), \sigma_G^2)\}_{l=1}^L$, where $F(\cdot)$ is the distribution of the channel coefficients, where the estimated channel, $\hat{g}(l)$, is assumed known at the receiver.

4. The received signal at the $l$-th relay is a random vector given by $\mathbf{R}_l = s H(l) + \mathbf{W}(l)$, $l \in \{1, \ldots, L\}$, where $H(l)$ is the channel coefficient (scalar random variable) between the transmitter and the $l$-th relay, $s$ is the transmitted code-word and $\mathbf{W}(l)$ is the noise realization (vector random variable) associated with the relay.

5. The transformation (relay function) of the received signal $\mathbf{R}_l$, performed by the $l$-th relay is assumed unknown. The unknown system model at each relay node $l$ will be modelled by a distribution over a function space as specified by a Gaussian Process prior, $f(l) \sim \mathcal{GP}(\mu_{\theta_1}(\cdot), \mathcal{C}_{D_0} (\cdot, \cdot))$, where $f(l)$ is defined to be the random vector function (relay function). A realization of this random vector function will be denoted by $f(l)(\mathbf{R}_l) = f(l)(R_{1}(l)), \ldots, f(l)(R_{K}(l))$, which is evaluated for the received signal at the $l$-th relay. The distribution of possible functions to be considered is controlled by the GP mean function $\mu_{\theta_1}(\cdot)$ parameterised by $\theta_1$ and covariance function $\mathcal{C}_{D_0} (\cdot, \cdot)$ constructed from a kernel parametrised by $\mathbf{D}_0$.

We denote time series observations of the function evaluation at the $l$-th relay

$$f(l)_{1:T} = \underbrace{f(l)(R_{1}(l(1)), \ldots, f(l)(R_{K}(l(1))))}_{t=1} \ldots \underbrace{f(l)(R_{1}(l(T))), \ldots, f(l)(R_{K}(l(T)))}_{t=T}. \tag{1}$$

6. To ensure a parsimonious statistical model, particularly when $L$ is large, we assume the all relay functions will have the same class of mean and covariance functions.

7. We consider the following model structure for the relay functionality:

- Mean function: the choice of mean function considered will be restricted to linear constant and trend models of the form

$$\mu_{\theta_1}(l)(R_{K}(l)) = \theta_{1} + \theta_2 R_{K}(l).$$

This assumption is consistent with the forms of relay function considered in the literature such as the AF of [13] or the EF of [14].

- Kernel function: the choice of kernel function, that constructs the Gram matrix $\mathcal{K}_l(\cdot)$ of the $l$-th relay, is jointly estimated along with the kernel parameters from the following squared exponential model:

$$\mathcal{C}(R_{1}(l), R_{2}(l)) = \exp\left(-\frac{|R_{1}(l) - R_{2}(l)|^2}{2\sigma^2}\right).$$

This widely used kernel produces smooth functions with the properties that the covariance function is stationary and non-degenerate [11]. Using this kernel, the corresponding Gram matrix for the $l$-th relay can be expressed as

$$\mathcal{K}_l = \begin{bmatrix} \mathcal{C}(R_{1}^{(1)}, R_{1}^{(1)}) & \cdots & \mathcal{C}(R_{1}^{(1)}, R_{K}^{(1)}) \\ \vdots & \ddots & \vdots \\ \mathcal{C}(R_{K}^{(1)}, R_{1}^{(1)}) & \cdots & \mathcal{C}(R_{K}^{(1)}, R_{K}^{(1)}) \end{bmatrix}.$$  \tag{2}

8. Conditional on matrix $f = (f(1)(R_{1}(1)), \ldots, f(L)(R_{L}(1)))$, the received signal at the destination, from the $l$-th relay, is a random vector given by

$$\mathbf{Y}(l) = f(l)(\mathbf{R}(l)) \mathcal{G}(l) + \mathbf{V}(l), \quad l \in \{1, \ldots, L\}, \tag{3}$$

where the scalar random variable $\mathcal{G}(l)$ is the channel coefficient between the $l$-th relay and the receiver, $f(l)(\mathbf{R}(l)) = \begin{bmatrix} f(l)(R_{1}(1)) & \cdots & f(l)(R_{K}(1)) \end{bmatrix}^T$ is the memoryless relay processing function (with possibly different functions at each of the relays) and the random vector $\mathbf{V}(l)$ is the noise realization associated with the receiver. We define $\mathbf{Y}_{1:T} = \begin{bmatrix} \mathbf{Y}_{1}(1) & \cdots & \mathbf{Y}_{L}(T) \end{bmatrix}$, where

$$\mathbf{Y}_{1:T} = \begin{bmatrix} Y_{1}(1) & \cdots & Y_{K}(1) \\ \vdots & \ddots & \vdots \\ Y_{1}(T) & \cdots & Y_{K}(T) \end{bmatrix}^T.$$  \tag{4}

9. All received signals are corrupted by i.i.d. zero-mean additive white complex Gaussian noise. At the $l$-th relay the noise corresponding to the $l$-th transmitted symbol is denoted by random variable $W_{l}(1) \sim \mathcal{CN}(0, \sigma_{w}^2)$. At the receiver this is denoted by random variable $V_{l}(1) \sim \mathcal{CN}(0, \sigma_{v}^2)$.

Given this system model specification, we can now develop a Bayesian hierarchical GP model for the relay system which will allow us to perform system identification. The posterior parameters and functions of interest in relay identification are given by the parameter vector after observing $T$ frames, $(\mathbf{F}(\cdot), \theta, \mathbf{D}) = (f(1)(\cdot), \ldots, f(L)(\cdot), \theta(1:L), \mathbf{D}(1:L))$, and presented as a directed acyclic graph in Fig. 2. We now develop the prior model. The prior choices for the relay functionality for $f$ are given by a GP with hyper priors for the mean and covariance functions as specified below.
Prior Model Structure
1. The priors of the hyper-parameters associated with the linear mean function are given by \( \theta = \{ \theta(1), \ldots, \theta(L) \} \), with \( \theta(i) = [\theta_1(i), \theta_2(i)] \), where \( \theta_1(i) \sim N(0,1) \), \( \theta_2(i) \sim N(0,100) \) for all \( i \). Note, here we assume a vague prior for the gradient of the mean function.
2. The priors of the hyper-parameters, associated with the kernel function \( C \), considered in the construction of the covariance function is specified by \( D = \{ D(1), \ldots, D(L) \} \), where \( D(i) \sim U[0, 10] \).
3. The hierarchical prior for the \( i \)-th relay is then given by \( \mathbf{f}(i) \sim \mathcal{GP}(\mu_{\theta(i)}(\cdot), \mathbf{C}_{\theta(i)}(\cdot, \cdot)) \), with GP mean function \( \mu_{\theta(i)} \) parameterised by \( \theta(i) \) and covariance function \( \mathbf{C}_{\theta(i)}(\cdot, \cdot) \) constructed from a kernel parameterised by \( D(i) \).

Posterior Model Structure
The combination of the likelihood model, priors for model parameters and the hierarchical priors for the GP prior, when combined under Bayes’ Theorem, result in a full posterior distribution given by

\[
\begin{align*}
&= \prod_{l=1}^{L} \prod_{i=1}^{T} p\left( \mathbf{f}(i), D, \mathbf{w}_i, \mathbf{f}(i) \mid y_{1:T} \right) \\
&\propto \prod_{l=1}^{L} \prod_{i=1}^{T} p\left( \mathbf{y}_{1:T} \mid \mathbf{f}(i), \mathbf{w}_i, \mathbf{f}(i) \right) \times \mathcal{GP}\left( \mathbf{f}(i) \mid \mu_{\theta(i)}(\cdot), \mathbf{C}_{\theta(i)}(\cdot, \cdot) \right) \\
&\quad \times \mathcal{GP}\left( \mathbf{C}_{\theta(i)}(\cdot, \cdot) \mid \mathbf{D}(i) \right) \\
&= \prod_{l=1}^{L} \prod_{i=1}^{T} \prod_{k=1}^{K} p\left( y_{k}(i) \mid f(i) \right) f(i) \left( h_{k}(i) \right) \left( \theta(i), D(i), w_{k}(i) \right) \\
&\quad \times \mathcal{GP}\left( f(i) \mid \mu_{\theta(i)}(\cdot), \mathbf{C}_{\theta(i)}(\cdot, \cdot) \right) \times \mathcal{GP}\left( \mathbf{C}_{\theta(i)}(\cdot, \cdot) \mid \mathbf{D}(i) \right) \\
&\propto \mathcal{GP}\left( \mathbf{f}(i) \mid \beta_{\theta(i)}(\cdot), \mathbf{C}_{\theta(i)}(\cdot, \cdot) \right) \times \mathcal{GP}\left( \mathbf{C}_{\theta(i)}(\cdot, \cdot) \mid \mathbf{D}(i) \right).
\end{align*}
\]

Note, we have included auxiliary parameters \( \mathbf{w}_{1:L,k} \) to represent the unknown noise realizations at the \( L \) relays for each transmitted sequence of symbols. The augmentation of these auxiliary parameters in the posterior specification allows us to obtain closed form expressions for the likelihood model, in particular a Gaussian form which will be relevant when combined with the GP prior for the relay functionality. Without the introduction of these auxiliary nuisance parameters we would be unable to derive a closed form expression for the relay function likelihoods, see discussions in [15].

Assumption I: Consider partial Channel State Information (CSI) which means that although we still consider the channels \( G(i) \) and \( H(i) \) to be unknown random vectors, for the sake of low complexity relay identification, we condition inference in this section on a noisy estimate of the sufficient statistics of the channels, given by \( \mathbb{E}[G(i)] = \tilde{g}(i) \mathbb{E}[H(i)] = \tilde{h}(i) \).

Assumption II: We consider application of a zero forcing (ZF) condition for the relay thermal noise given by \( \mathbb{E}[\mathbf{w}(i)] = \mathbf{0} \).

As a result of these two simplifying assumptions, the received signal at the relay is given by

\[
\begin{align*}
\mathbf{r}(i) = \mathbb{E}[\mathbf{r}(i)] + \mathbb{E}[\mathbf{w}(i)] = \mathbf{s}(i) \tilde{h}(i), \\
\mathbb{E}[\mathbf{y}(i)] = \mathbb{E}[\mathbf{r}(i)] = \mathbb{E}[\mathbf{s}(i) \tilde{h}(i) + \mathbf{v}(i)],
\end{align*}
\]

where \( \mathbf{f}(i) \sim \mathcal{GP}(\mu_{\theta(i)}(\cdot), \mathbf{C}_{\theta(i)}(\cdot, \cdot)) \).

1: Under assumptions I and II, the resulting identification problem can be stated as the following

\[
\begin{align*}
\arg \max_{\mathbf{f}(i)} \mathbb{E}[\mathbf{y}(i) \mid \mathbf{r}(i)] = \mathbb{E}[\mathbf{s}(i) \tilde{h}(i) + \mathbf{v}(i)],
\end{align*}
\]

4. SYSTEM IDENTIFICATION VIA ICM
Having formulated the relay identification problem, we now develop an efficient solution via ICM procedure [10]. The ICM estimation procedure has the following steps: the first involves specification of a multivariate posterior distribution, deconstructed as a set of full conditional posterior distributions. These full conditional posterior distributions are given for the \( r \)-th relay, based on the full posterior in Equation (4), after applying Assumptions I and II

\[
\begin{align*}
\mathbf{p}\left( \mathbf{f}(i) \mid \theta, D, \mathbf{y}_{1:T}(i) \right) \propto \prod_{l=1}^{K} \prod_{k=1}^{K} \mathbb{E}\left[ \mathbf{y}_{k}(i) \mid \mathbf{f}(i) \right] f(i) \left( h_{k}(i) \right) \left( \theta, D, w_{k}(i) \right) \\
\propto \mathcal{GP}\left( \mathbf{f}(i) \mid \mu_{\theta(i)}(\cdot), \mathbf{C}_{\theta(i)}(\cdot, \cdot) \right) \times \mathcal{GP}\left( \mathbf{C}_{\theta(i)}(\cdot, \cdot) \mid \mathbf{D}(i) \right) \\
\end{align*}
\]

Next, given these full conditional posterior distributions, the \( j \)-th iteration of the ICM algorithm successively updates each estimate of \( \hat{\mathbf{f}}_{1:T}(i), \hat{\theta}(i), \hat{D}(i) \) based on the solutions at iteration \( (j-1) \), and the solutions to the following sequence of MAP estimates for the full conditional posteriors.

\[
\begin{align*}
\hat{\mathbf{f}}_{1:T}(i) = \arg \max_{\mathbf{f}(i)} \mathbb{E}[\mathbf{f}(i) \mid \theta(j-1), \hat{D}(j-1), \mathbf{y}_{1:T}(i)]; \quad \theta(j) = \arg \max_{\theta} \mathbb{E}[\mathbf{f}(i) \mid \theta(j), \hat{D}(j-1), \mathbf{y}_{1:T}(i)]; \\
\hat{D}(j) = \arg \max_{D} \mathbb{E}[\mathbf{f}(i) \mid \theta(j), \hat{D}(j-1), \mathbf{y}_{1:T}(i)].
\end{align*}
\]
Theorem 1. The full conditional distributions in (7) and conditional MAP estimates in (8a-8c) are given by:

1. The full conditional of \( p \left( f^{(l)} (\cdot) \left| \theta, D, y^{(1:L)} \right. \right) \) in (7) is

\[
p \left( \text{Vec} \left[ f^{(1:T)}_{1:1} \right] \left| \theta^{(l)}, D^{(l)}, y^{(1:T)}_{1:1} \right. \right) = N (M, \Sigma),
\]

where

\[
M = \left( \left( K^{(l)}_{1:T} \right)^{-1} + \frac{1}{\sigma^2} I_{K \times T} \right)^{-1} \times \left( \left( K^{(l)}_{1:T} \right)^{-1} M_0 + \frac{1}{\sigma^2} I_{K \times T} \text{Vec} \left[ y^{(1:T)}_{1:1} \right] \right),
\]

with

\[
M_0 = \left( \mu^{(l)}_1 \left( r^{(1)} (1) \right), \ldots, \mu^{(l)}_K \left( r^{(1)} (1) \right), \ldots, \right)_{t=1}^{T},
\]

\[
\mu^{(l)}_i \left( r^{(1)} (T) \right), \ldots, \mu^{(l)}_K \left( r^{(1)} (T) \right),
\]

and

\[
\Sigma = \left( \left( K^{(l)}_{1:T} \right)^{-1} + \frac{1}{\sigma^2} I_{K \times T} \right)^{-1}.
\]

The conditional MAP estimate in (8a), is given by

\[
\text{Vec} \left[ \hat{f}^{(1:T)}_{1:1} \right] = \arg \max \left( \text{Vec} \left[ f^{(1:T)}_{1:1} \right] | \theta^{(l)}, D^{(l)}, y^{(1:T)}_{1:1} \right) = M.
\]

2. The full conditional for \( p \left( \theta | D, f^{(l)} (\cdot), y^{(1:T)} \right) \) in (7) is

\[
p \left( \theta^{(l)} | D^{(l)}, f^{(l)}_{1:T} (\cdot), y^{(1:T)} \right) \propto N \left( \text{Vec} \left[ f^{(1:T)}_{1:1} \right]; \mu^{(l)}, K^{(l)}_{1:T} \right) p \left( \theta^{(l)} \right).
\]

The conditional MAP estimate of \( \theta^{(l)} \) is \( \left[ \theta^{(l)}_1, \theta^{(l)}_2 \right] \) in (8b):

\[
\theta_1 = -\theta_2 \left( \left( \Sigma_{\theta} \right)^{-1} \right)_{1,1} + \sum_{i=1}^{K} \sum_{t=1}^{T} r_i (t) \left( \left( K^{(l)}_{1:T} \right)^{-1} \right)_{i,t},
\]

\[
\theta_2 = -\frac{1}{2 \Phi} \left( -\sum_{i=1}^{K} r_i (t) f(r_i (t)) \left( \left( K^{(l)}_{1:T} \right)^{-1} \right)_{i,t} \right) + \sum_{i=1}^{K} \sum_{t=1}^{T} r_i (t) \left( \left( K^{(l)}_{1:T} \right)^{-1} \right)_{i,t} \times \left( \left( \Sigma_{\theta} \right)^{-1} \right)_{1,2} + \sum_{i=1}^{K} r_i^2 (t) \left( \left( K^{(l)}_{1:T} \right)^{-1} \right)_{i,t},
\]

where

\[
\Phi = \left( \left( \Sigma_{\theta} \right)^{-1} \right)_{1,1} + \sum_{i=1}^{K} \sum_{t=1}^{T} r_i (t) \left( \left( K^{(l)}_{1:T} \right)^{-1} \right)_{i,t} \times \left( \left( \Sigma_{\theta} \right)^{-1} \right)_{1,2} + \sum_{i=1}^{K} r_i^2 (t) \left( \left( K^{(l)}_{1:T} \right)^{-1} \right)_{i,t} + \left( \left( \Sigma_{\theta} \right)^{-1} \right)_{2,2} + \sum_{i=1}^{K} r_i^2 (t) \left( \left( K^{(l)}_{1:T} \right)^{-1} \right)_{i,t},
\]

and \( \otimes \) is the Kronecker product, and the \( i, j \)-th element of \( \Phi_{1:T} \) is given by \( \Phi_{i,j} \equiv \left\| R_{i,j}^{(1:T)} \right\|^2 \).

The proof can be found in the Appendix.

5. SIMULATION RESULTS

We present the performance of the proposed algorithms via Monte Carlo simulations. The prior distribution for all the channels is Rayleigh fading, and the channels are assumed to be both spatially and temporally independent; The channels uncertainty was set to \( \sigma^2 = 0.2 \). The SNR is set to 0 dB (low SNR) and 10 dB (high SNR); The results are obtained from simulations over \( T = 100 \) transmitted frames with \( K = 200 \) symbols per frame; In all simulations 16PAM constellations were considered; The ICM algorithm iterated \( J = 50 \) times; The relay functions tested corresponded to absolute \( f(x) = |x| \), linear \( f(x) = ax + b \), sinusoidal \( f(x) = \sin(wx + \phi) \), hyperbolic-tan \( f(x) = \tanh(wx + \phi) \) and demodulated; We first present in Figure (3) a random selection of representative results for the estimation of the MAP model hyper-parameters \( (\theta_1, \theta_2, D) \), jointly estimated with the relay function identification under ICM. We present these estimates versus the ICM iterations \( J = 1, \ldots, 50 \). Two important features are evident, the first that the results converge to a set of optimal values and secondly that this occurs relatively rapidly, with very few iterations of ICM required. This is characteristic of all the examples we tested.

The results we present are the posterior MAP estimated relay function estimated under our GP-ICM frameworks evaluated at the PAM constellation points (black). In addition we present the 95% posterior confidence intervals on
these MAP estimates (grey). The true relay function is depicted in red. We compare these estimates to the true relay functional form utilised in the simulations. For all the non-linear relay functions considered, the estimation under both perfect and partial CSI is highly accurate. The Demodulated relay function which had the linear trend with stairs overlayed, was most difficult to perform system identification as it contained a global feature of the linear trend as well as local fine scale features corresponding to the stairs function. We observe that with perfect CSI, the estimation was relatively accurate. However, for the partial CSI settings, the estimation of the global feature of the trend was evident, though the resolution of the local features of the stairs was diminished and therefore learning such intricate features will require many more frames of estimation. The results presented were for 100 frames, with an increase over time to 500 or more, the estimation will resolve these local features. As expected in any regression based analysis, the functional forms were most difficult to estimate at the extremities of the convex hull of the received PAM constellation points. Theoretically, this can be proven to result in the largest predictive uncertainty in the estimated function, leading in this case to most uncertainty in the estimated relay functional forms. This was observed in all settings and for all functions, and is most poignant in the Demodulated function example.

Finally, we present summarised results in Table (1), which are each based on the absolute error in the MAP estimated relay functions.

6. CONCLUSIONS

We solved the problem of relay identification in the setting of wireless relay systems. We have developed a novel and flexible stochastic model for a class of cooperative wireless relay networks, in which the relay processing functionality is not known at the destination. Working under this modelling framework, we developed and demonstrated the performance of our estimation procedure aimed at performing online system identification. In particular we demonstrated that our GP identification via ICM can resolve this problem via a computationally efficient algorithm for many different relay functional forms.

7. REFERENCES

[1] J. Laneman, D. Tse, and G. Wornell, “Cooperative diversity in wireless networks: Efficient protocols and outage behavior,” Information Theory, IEEE Transactions on, vol. 50, no. 12, pp. 3062–3080, 2004.

[2] K. Gomadam and S. Jafar, “Optimal Relay Functionality for SNR Maximization in Memoryless Relay Networks,” Arxiv preprint cs.IT/0510002, 2006.

[3] L. Ghabeli and M. Arel, “A new achievable rate and the capacity of some classes of multilevel relay network,” EURASIP Journal on Wireless Communications and Networking, vol. 2008, p. 38, 2008.

[4] D. Crouse, C. Berger, S. Zhou, and P. Willett, “Optimal memoryless relays with noncoherent modulation,” IEEE Trans. Signal Processing, vol. 56, no. 12, p. 1, 2008.

[5] S. Yao, M. Khormuji, and M. Skoglund, “Sawtooth relaying,” IEEE Communication Letters, vol. 12, no. 9, pp. 612–615, 2008.

[6] G. Pillonetto and G. De Nicolao, “A new kernel-based approach for linear system identification,” Automatica, vol. 46, no. 1, pp. 81–93, 2010.

[7] A. Chiuso and G. Pillonetto, “A bayesian approach to sparse dynamic network identification,” 2011.
[8] C. Volosencu, “Identification in sensor networks,” in Proceedings of the 9th WSEAS, 2008, pp. 175–183.
[9] A. Khan and V. Kuhn, “Power Optimization in Adaptive Relay Networks,” Proc. IEEE Global Communications Conference, vol. 4, 2010.
[10] J. Besag, “Statistical analysis of dirty pictures,” Journal of the Royal Statistical Society, vol. 39, part 2, pp. 159–195, 1974.
[11] C. Rasmussen and C. Williams, Gaussian Processes for Machine Learning (Adaptive Computations and Machine Learning). The MIT Press, 2005.
[12] M. Lazaro-Gredilla, J. Quinonero-Candela, C. Rasmussen, and A. Figueiras-Vidal, “Sparse spectrum gaussian process regression,” The Journal of Machine Learning Research, vol. 11, pp. 1865–1881, 2010.
[13] Y. Zhao, R. Adve, and T. Lim, “Improving amplify-and-forward relay networks: optimal power allocation versus selection,” pp. 1234–1238, 2006.
[14] K. Gomadam and S. Jafar, “Optimal relay functionality for SNR maximization in memoryless relay networks,” IEEE Journal on Selected Areas in Communications, vol. 25, no. 2, pp. 390–401, 2007.
[15] G. Peters, I. Nevat, S. Sisson, Y. Fan, and J. Yuan, “Bayesian symbol detection in wireless relay networks via likelihood-free inference,” Signal Processing, IEEE Transactions on, vol. 58, no. 10, pp. 5206–5218, 2010.

Appendix

Proof. The posterior distribution decomposes according to Equation (4), and we can therefore derive the following quantities for a given relay:

1. The full conditional posterior distribution for the $l$-th relay function in (7) is given by

   \[
   p(f^{(l)}_l, \theta^{(l)}, D^{(l)}, y^{(l)}_{1:T}) = \frac{p(y^{(l)}_{1:T} | f^{(l)}_l, \theta^{(l)}) p(f^{(l)}_l | \mu^{(l)}_1, \kappa^{(l)}_1)}{p(y^{(l)}_{1:T} | f^{(l)}_l, \theta^{(l)}) p(f^{(l)}_l | \mu^{(l)}_1, \kappa^{(l)}_1) df^{(l)}_l},
   \]

   With a matrix variate normal likelihood model for $y^{(l)}_{1:T} | f^{(l)}_l$ and a GP prior on the function which will result in a matrix variate prior for the function over the symbols in each frame, i.e. $K \times T$. After vectorizing the observation random matrix and the prior random matrix, we obtain multi variate Gaussian distributions which admit standard conjugacy properties. This results in Equation (9).

2. The full conditional for $\theta^{(l)}$ in (7) can be expressed as

   \[
   p(\theta^{(l)} | D^{(l)}, f^{(l)}_l, y^{(l)}_{1:T}) \propto p(\{y^{(l)}_{1:T} | f^{(l)}_l, \theta^{(l)}\}) p(\{f^{(l)}_l | \mu^{(l)}_1, \kappa^{(l)}_1\}) \]

   \[
   \times p(\theta^{(l)}) = N(\{f^{(l)}_l | \mu_1^{(l)}, \kappa_1^{(l)}\} p(\theta^{(l)})).
   \]

Deriving the MAP estimate is achieved by applying the log transform to the posterior, taking partial derivative with respect to each parameter, setting to 0 and solving as follows:

\[
\nabla \theta^{(l)} \left[ -\frac{1}{2} \left( \mathrm{Vec} \left( y^{(l)}_{1:T} \right) - \mu^{(l)}_1 \right)^T K^{-1} \left( \mathrm{Vec} \left( y^{(l)}_{1:T} \right) - \mu^{(l)}_1 \right) + \log p(\theta^{(l)}) - \log (2\pi \det K_1^{(l)}) \right] = 0,
\]

which produces the following linear system of equations with a unique solution in (13-14).

3. The full conditional for $D^{(l)}$ in (7) can be expressed as

\[
p(D^{(l)} | f^{(l)}_l, \theta^{(l)}, y^{(l)}_{1:T}) \propto N(\{f^{(l)}_l | \mu^{(l)}_1, \kappa^{(l)}_1\} p(D^{(l)})
\]

where we clarify that the Gramm matrix $K_1^{(l)}$ is implicitly a function of $D^{(l)}$. Deriving the MAP estimate is achieved by applying the log transform to the posterior, taking derivative with respect to $D^{(l)}$, setting to 0 and solving as follows:

\[
\frac{d}{dD} \left[ \log p(D^{(l)}) - \frac{1}{2} \log |K_1^{(l)}| \right.
\]

\[
- \frac{1}{2} \left( \mathrm{Vec} \left( f_{1:T} \right)^T - \mu_{1:T} \right)^T \left( K_1^{(l)} \right)^{-1} \left( \mathrm{Vec} \left( f_{1:T} \right)^T - \mu_{1:T} \right)
\]

\[
\times \left[ D^{(l)} \in (a, b) \right]
\]

\[
= -\frac{1}{2} \mathrm{Tr} \left( K_1^{(l)} \frac{dK_1^{(l)}}{dD^{(l)}} \right)
\]

\[
+ \frac{1}{2} \left( \mathrm{Vec} \left( f_{1:T} \right)^T - \mu_{1:T} \right)^T \left( K_1^{(l)} \right)^{-1} \frac{dK_1^{(l)}}{dD^{(l)}} \left( K_1^{(l)} \right)^{-1} \left( \mathrm{Vec} \left( f_{1:T} \right)^T - \mu_{1:T} \right) \bigg|_{D^{(l)} \in (a, b)},
\]

where this expression is obtained by using standard matrix derivative identities.

---

**Table 1**: Comparison of absolute error in MAP estimation for relay functions.

| Relay function | perfect CSI | imperfect CSI |
|---------------|-------------|---------------|
|               | high SNR    | low SNR       | high SNR   | low SNR   |
| ABSOLUTE      | 0.91        | 3.97          | 0.97      | 4.1       |
| LINEAR        | 0.57        | 0.76          | 0.76      | 0.83      |
| SIN           | 0.14        | 0.41          | 0.15      | 0.45      |
| TANH          | 0.18        | 0.21          | 0.19      | 0.24      |
| DEM           | 0.79        | 0.85          | 0.83      | 0.91      |

**Fig. 6**: Demodulate relay function.