
Gareth W. Peters$^{1,2}$, Ido Nevat$^3$, Jinhong Yuan$^4$ and Ian Collings$^3$

$^1$ School of Mathematics and Statistics, University of New South Wales, Sydney, 2052, Australia; email: peterga@maths.unsw.edu.au
$^2$ CSIRO Sydney, Locked Bag 17, North Ryde, New South Wales, 1670, Australia
$^3$ CSIRO, Wireless & Networking Tech. Lab, Sydney, Australia.
$^4$ School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia.

Abstract

This paper presents a flexible stochastic model developed for a class of cooperative wireless relay networks, in which the relay processing functionality is not known at the destination. The challenge is then to perform online system identification in this wireless relay network. To address this challenging problem we develop a novel class of statistical models and a computationally efficient algorithm that can be performed in real time processing, to undertake system identification for each relay channel in the presence of partial Channel State Information (CSI). We also develop a lower bound result and several sub-optimal though computationally efficient solutions to the identification problem, for comparison. We provide several examples for different non-linear relay functionalities.

Keywords: Relay networks, System Identification, Gaussian processes, Kernel methods, Iterated Conditioning on the Modes, co-operative wireless relay network.
I. BACKGROUND ON WIRELESS COMMUNICATIONS AND RELAY SYSTEM IDENTIFICATION

Cooperative communications systems have been proposed to exploit the spatial diversity gains inherent in multiuser wireless systems without the need of multiple antennas at each node, see [1] and [2]. This is achieved by having the users relay each others messages and, thus, forming multiple transmission paths to the destination. Simply put, such a system 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 [3], [2]. The relay function can be optimized for different design objectives [4], [5], [6]. For example, in the estimate and forward (EF) scheme, in the case of BPSK signaling, the optimal relay function is the hyperbolic tangent [7]. Other criteria for which the optimal relay function is non-linear include: capacity maximisation [7], minimum error probability at the receiver [8], SNR maximisation [6], rate maximisation [9] and minimisation of the average error probability [10].

In ad-hoc 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. Therefore, in order for the destination to perform detection of the data symbols, it also needs to perform estimation of the wireless channels and the relay functionality. This is a challenging problem due to uncertain functional form of each relays processing on the received signal. To address this problem we introduce to the wireless communications setting, semi-parametric modelling procedures based on Gaussian processes (GP), which will allow us to efficiently solve the joint detection and system identification problem.

II. BACKGROUND ON GAUSSIAN PROCESSES

Gaussian processes (GP) define a family of stochastic processes that allow one to undertake flexible semi-parametric modelling of causal relationships without a priori specification of the structure of the causal relationship. That is we may utilise a GP regression model as a flexible family of regression models, in which the relationship between the predictors (independent variables) and the responses (dependent variables) is not specified in advance. Instead, this relationship along with the parameters of this semi-parametric model are learnt jointly from the observed data. This is ideal in the context of system identification in relay networks, where there is potential for several possible highly non-linear relay function transformations to be applied to a transmitted signal.

In the wireless communications setting, we develop a stochastic model in which we define a distribution over a function space. This distribution over the class of continuous smooth functions $f(\cdot) \in C$ when constructed in the estimation framework we develop acts as a priori knowledge of the functional form of the unknown relay system transforms. The typical relay system transformations that one can encounter in practice include linear amplification such as in an amplify and forward [11], hyperbolic tan which has the property that it minimises the MMSE estimate at the relay [12].
These relay functions can be either static in time or in more advanced relay systems they may adapt over time, reacting and updating the transformations applied to the received signals before retransmission to account for time varying channel characteristics. The methodology we develop in this paper to perform this system identification is on-line and efficient, flexible enough to handle both the static and adaptive relaying functionalities.

In particular the distribution we specify on the function space of smooth continuous functions is constructed via a GP, which can be considered simplistically as the analog of an infinite dimensional Gaussian random vector, in which each component of the vector corresponds to a point of evaluation of the unknown random function. Therefore, analogously to the multi-variate Gaussian random vector, the infinite dimensional GP prior on the function space used to characterize the distribution over possible smooth continuous unknown relay functions is characterized sufficiently by a mean function and a covariance function. The covariance function, must be specified carefully to ensure that finite realizations of the process correspond appropriately to multi-variate Gaussian random vectors. As such, each kernel evaluation for the covariance function, that goes into the construction of the covariance matrix for such a realization must a a minimum result in a positive definite covariance matrix. One way to ensure this is to work with kernel functions which are in the family of Mercer kernels [13], these make up a GP which will be defined in a reproducing Kernel Hilbert space [14]. As discussed in tutorials by [14], [15] the Gaussian process literature is well established in spatial and temporal modelling, since the works of [16]. Examples which include successful development of GP models in engineering include forecasting of non-linear dynamic systems [17], system identification of dynamical systems [18], non-linear identification, equalization and separation of signals [19] and many more.

In summary, we will use a GP as a prior probability distribution over the unknown functional forms of the relay functions. This will enable us to perform inference on the relay functionality and perform data detection at the receiver. Hence, summarizing this concept for the wireless relay setting, the idea of utilizing a GP is to work with the unknown relay functions at the \( l \)-th relay, without parameterizing this function for the received signal, \( f^{(l)}(r) \). Instead, we place a GP prior, \( P(f^{(l)}(r)) \), directly on the space of functions that we wish to consider in the system identification. The probabilistic nature of the GP model allows to directly define the space of admissible functions relating inputs to outputs, by simply specifying the mean and covariance functions of the process. In the flexible models we develop in this paper, the mean function of each relay depends on the received signal vector for each transmitted symbol in \( r \). The covariance function of each relay expresses the expected covariance between the value of the relay function \( f^{(l)}(\cdot) \) for any two received symbols \( r_i \) and \( r_j \), thereby capturing the temporal dependence. The covariance function can also be used to capture dependence spatially between relay nodes to reflect dependence between local relay nodes over space and time. In this framework, the observations correspond to an (incomplete) realisation of the process. Although a parametric form is
actually pre-supposed (albeit not on the functional itself but on the family from which it can come from), the model is very powerful and flexible, with the Gaussian assumption keeping most derivations analytically tractable and simple. For good introductions to GP see [14], [20], [21].

A. Statistical Properties of Semi-Parametric Modelling via a Gaussian Process

In this section we provide a brief review of Gaussian Process regression and the statistical concepts that are utilised from this research area in the system identification solution we develop. We begin with the formal definition of a Gaussian process:

**Definition 1 [14]:** 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 [14]. The selection of a covariance function which is a Mercer kernel ensures the resulting Gramm-matrix, denoted by $K$ and defined in Equation (5), which is constructed from the evaluation of the covariance function pointwise and acts as the covariance matrix, is strictly positive definite.

In the context of system identification that we address in this paper, we formulate the problem in the context of a semi-parametric regression model for each relay. 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

$$f(\cdot) \sim \mathcal{GP}(m(\cdot; \theta), C(\cdot, \cdot; \Psi)).$$

(1)

As discussed recently in [22] 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 \( \{r_t\}_{t=1:T} \), the corresponding random vector for the function at these points given by \( f_{1:T} = [f(r_1), \ldots, f(r_T)] \) is distributed according to the following multivariate Gaussian distribution,

$$f_{1:T} \sim p(f_{1:T}|R_{1:T} = r_{1:T}) = \mathcal{N}(f_{1:T}; [m(r_1; \theta), \ldots, m(r_T; \theta)], K_{1:T})$$

(2)

with each component \( [K_{1:T}]_{i,j,t} = C(R_i(t), R_j(t)) = \text{Cov}[f(R_i(t))f(R_j(t))]. \) Therefore, we see that the covariance matrix, constructed from the kernel covariance function measures the similarity between pairs of function values. In this paper we consider stationary covariance functions in which for all combinations of $R_i(t)$ and $R_j(t)$ the covariance function is always just a function of the difference $C(R_i(t), R_j(t)) = C(R_i(t) - R_j(t))$. Therefore, as discussed in [22] the elegance of a GP framework is that the properties of the unknown function to be estimated are expressed directly in terms of the...
covariance function, rather than implicitly via basis functions such as in a basis expansion model. Hence, summarising this concept we observe that if we consider two scalar inputs \( R_i(t) \) and \( R_j(t) \) separated by a distance of \(||R_i(t) - R_j(t)||_2\). Firstly, as we draw different realizations from the GP for \( f(R_i(t)) \) and \( f(R_j(t)) \) we get different fluctuations depending on the function drawn. The degree to which this fluctuation in the values drawn occurs, is directly affected by the the choice of kernel. Under a stationary kernel for which the values of the kernel is high when this distance is high results in corresponding sampled function values \( f(R_i(t)) \) and \( f(R_j(t)) \) which are highly correlated and therefore similar, restricting the roughness of the functions obtained. Conversely, under a stationary kernel for which the values of the kernel are low when this distance is high, results in corresponding sampled function values \( f(R_i(t)) \) and \( f(R_j(t)) \) which are approximately uncorrelated and therefore not similar, allowing rough continuous functions to be obtained.

B. Contributions, Outline and Notation

The main contributions of this work are as follows:

- We propose a novel stochastic model for the problem of relay identification based on a flexible semi-parametric Gaussian process prior on the functional form of each relays transform.
- We develop a formulation of this stochastic model under a novel sequential Bayesian estimation framework. In doing so we present three estimation scenarios and a comparative lower bound on their performance in order of complexity:
  1) the first involves the optimal identification of the unknown relay function after integration of the nuisance parameters, for each frame of transmitted symbols, associated with each relay’s unknown AWGN realizations. We explain how this solution is computationally expensive for on-line real-time solutions;
  2) the second involves an efficient, though sub-optimal on-line and real-time solution to statistical estimation challenge of system identification for each of the the unknown relay functions. This is based on utilizing practical system knowledge of pilot symbols in each transmission frame, zero forcing of the relay noise and partial channel state information;
  3) the third involves an efficient, though sub-optimal on-line and real-time solution to statistical estimation challenge of system identification for each of the the unknown relay functions. This is based on utilizing practical system knowledge of pilot symbols in each transmission frame, zero forcing of the relay noise and perfect channel state information;
- Finally, we develop a novel and highly efficient estimation procedure for these systems based on Maximum A Posteriori (MAP) estimation jointly for the unknown relay function and unknown model parameters achieved via an adaption of the Iterated Conditioning on the Modes algorithm of [23] to the GP stochastic models developed.
1) **Outline**: The paper is organised as follows: in Section III we introduce a stochastic system model for the wireless relay system and the associated Bayesian model. In Section IV we present the optimal detection algorithm. In Section V we present a low complexity sub-optimal algorithms. Section VI presents results and analysis and conclusions are provided in Section VII.

2) **Notation**: The following notation is used throughout: random variables are denoted by upper case letters and their realizations by lower case letters. In addition, bold will be used to denote a vector or matrix quantity, upper subscripts will refer to the relay node index and lower subscripts refer to the element of a vector or matrix.

### III. Bayesian System Model and Relay Identification

In this section we introduce the system model and a Bayesian model for inference on the system model parameters. In our system model, the channels in the relay network are modelled stochastically, where we do not know *a priori* the realized channel coefficient values. Instead, we consider partial channel state information (CSI), in which we assume known statistics of the distribution of the channel coefficients.

#### A. System model and assumptions

Here we present the system model and associated assumptions. The system model is depicted in Fig. 1.

1. We will generically denote the frame index for the $t$-th frame using $t \in \{1, \ldots, T\}$.

2. Assume a wireless relay network with a single source node, transmitting sequences of $K$ pilot symbols per frame. We will denote this set of pilot symbols for frame $t$ as $s = s_{1:K}$. These symbols are transmitted from a source to a single destination via $L$ relay nodes.

3. The sequence of $K$ symbols, $s \in \Omega$, are taken from a digital constellation with cardinality $M$.

4. 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. We assume that all channels are independent with a coherence interval larger than the codeword length $K$.

5. Assume a general model for the 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 an assumption 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 \( \{ G^{(l)}(\cdot) \sim F(\cdot, \sigma^2) \}_{l=1}^L \) where \( F(\cdot) \) is the distribution of the channel coefficients, where the estimated channel, \( \hat{g}^{(l)} \), is assumed known at the receiver.

5) The received signal at the \( l \)-th relay is a random vector given by

\[
R^{(l)} = SH^{(l)} + W^{(l)}, \quad l \in \{1, \ldots, L\},
\]

where \( H^{(l)} \) is the channel coefficient (scalar random variable) between the transmitter and the \( l \)-th relay, \( S \) is the transmitted code-word (vector random variable) and \( W^{(l)} \) is the noise realization (vector random variable) associated with the relay.

6) The transformation (relay function) of the received signal \( R^{(l)} \), performed by the \( l \)-th relay is assumed unknown. Therefore to perform detection at the receiver one must also perform system identification for each relay node jointly with the detection. 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,

\[
f^{(l)}(\cdot) \sim GP\left(\mu_{\theta^{(l)}}(\cdot), C_{D^{(l)}}(\cdot, \cdot)\right).
\]

Where \( f^{(l)}(\cdot) \) is defined to be the random vector function (relay function). A realization of this random vector function will be denoted by \( f^{(l)}(R^{(l)}) = \left[ f^{(l)}(R_1^{(l)}), \ldots, f^{(l)}(R_K^{(l)}) \right] \), 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 \( \mu_{\theta^{(l)}}(\cdot) \) parameterised by \( \theta^{(l)} \) and covariance function \( C_{D^{(l)}}(\cdot, \cdot) \) constructed from a kernel parametrised by \( D^{(l)} \).

We denote time series observations of the function evaluation at the \( l \)-th relay

\[
f^{(l)}_{1:T} = \left( \begin{array}{c}
    f^{(l)}(R_1^{(l)}(1)), \ldots, f^{(l)}(R_K^{(l)}(1)), \\
    f^{(l)}(R_1^{(l)}(2)), \ldots, f^{(l)}(R_K^{(l)}(2)), \\
    \vdots, \\
    f^{(l)}(R_1^{(l)}(T)), \ldots, f^{(l)}(R_K^{(l)}(T)),
  \end{array} \right)_{t=1}^{T}.
\]

7) To ensure a parsimonious and estimatable statistical model, particularly when \( L \) is large, we assume the all relay functions will have the same class of mean and covariance functions. Furthermore, we assume the mean function to be parameterized identically across all models and relays.

8) We consider the following model structure for the relay functionality:

- The choice of mean function considered will be restricted to linear constant and trend models of the form \( \mu_{\theta^{(l)}}(R_k^{(l)}) = \theta_1^{(l)} + \theta_2^{(l)} R_k^{(l)} \). This assumption is consistent with the forms of relay function considered in the literature such as the AF of [24] or the EF of [12].

- The choice of kernel function, that constructs the Gram matrix \( K^{(l)} \) of the \( l \)-th relay, is jointly estimated along with the kernel parameters from the following squared exponential model:
\( C(R_i^{(l)}, R_j^{(l)}) = \exp \left(-\frac{||R_i^{(l)}-R_j^{(l)}||^2}{2\sigma^2}\right), \forall i, j \in \{1, \ldots, K\}. \)

This widely used kernel produces smooth functions with the properties that the covariance function is stationary and non-degenerate [14]. Using this kernel, the corresponding Gram matrix for the \( l \)-th relay can be expressed as

\[
K_{1:T}^{(l)} = \begin{bmatrix}
C(R_1^{(l)}(1), R_1^{(l)}(1)) & \cdots & C(R_1^{(l)}(1), R_K^{(l)}(T)) \\
\vdots & \ddots & \vdots \\
C(R_K^{(l)}(T), R_1^{(l)}(1)) & \cdots & C(R_K^{(l)}(T), R_K^{(l)}(T))
\end{bmatrix},
\]

(5)

\[
\mu_{l,T}^{(l)} = \begin{pmatrix}
\mu_{l,1^{(l)}}(R_1^{(l)}(1)), \ldots, \mu_{l,1^{(l)}}(R_K^{(l)}(1)) \\
\mu_{l,2^{(l)}}(R_1^{(l)}(2)), \ldots, \mu_{l,2^{(l)}}(R_K^{(l)}(2)) \\
\vdots & \ddots & \vdots \\
\mu_{l,L^{(l)}}(R_1^{(l)}(T)), \ldots, \mu_{l,L^{(l)}}(R_K^{(l)}(T))
\end{pmatrix}^T
\]

(6)

- Conditional on the mean functions and covariance functions, \( \mu_{l,1^{(l)}}(\cdot), C_{D_i^{(l)}}(\cdot, \cdot) \) and \( \mu_{l,m^{(m)}}(\cdot), C_{D^{(m)}}(\cdot, \cdot) \) we consider realizations of each Gaussian Process function to be statistically independent temporally and spatially. Therefore we are assuming the following model structure:

\[
\mathbb{E} \left[ f^{(l)}(X) f^{(m)}(Y)^T | \mu_{l,1^{(l)}}(\cdot), C_{D_i^{(l)}}(\cdot, \cdot), \mu_{l,m^{(m)}}(\cdot), C_{D^{(m)}}(\cdot, \cdot) \right] = \Sigma
\]

for all \( X, Y \) inputs and all relays \( l, m \), where \( \Sigma \) is a diagonal covariance matrix. This gives us spatial independence between the functionality of each relay.

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

\[
Y^{(l)} = f^{(l)}(R^{(l)}) G^{(l)} + V^{(l)}, \ l \in \{1, \ldots, L\},
\]

(7)

where the scalar random variable \( G^{(l)} \) is the channel coefficient between the \( l \)-th relay and the receiver, \( f^{(l)}(r^{(l)}) = \begin{bmatrix} f^{(l)}(r_1^{(l)}), \ldots, f^{(l)}(r_K^{(l)}) \end{bmatrix}^T \) is the memoryless relay processing function (with possibly different functions at each of the relays) and the random vector \( V^{(l)} \) is the noise realization associated with the receiver.

We define

\[
Y_{1:T} = \begin{pmatrix} Y_{1:T}^{(1)}, \ldots, Y_{1:T}^{(L)} \end{pmatrix},
\]

where

\[
Y_{1:T}^{(l)} = \begin{pmatrix}
Y_1^{(l)}(1), \ldots, Y_K^{(l)}(1) \\
Y_1^{(l)}(2), \ldots, Y_K^{(l)}(2) \\
\vdots \\
Y_1^{(l)}(T), \ldots, Y_K^{(l)}(T)
\end{pmatrix}^T
\]

(8)

10) All received signals are corrupted by 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 \( W_i^{(l)} \sim \mathcal{CN}(0, \sigma_w^2). \)
At the receiver this is denoted by random variable $V_i^{(l)} \sim CN(0, \sigma_v^2)$. Additionally, we assume the following properties:

$$
E[W_i^{(l)}\overline{V}_j^{(m)}] = E[V_i^{(l)}\overline{W}_j^{(m)}] = E[V_i^{(l)}\overline{W}_j^{(m)}] = 0,
$$

$\forall i, j \in \{1, \ldots, K\}, \forall l, m \in \{1, \ldots, L\}, i \neq j, l \neq m$, where $\overline{W}_j$ denotes the complex conjugate of $W_j$.

We define

$$
W_{1:T} = \left( W_{1:1}, \ldots, W_{1:T} \right), \quad W_{1:l}^{(t)} = \left( W_{1}^{(t)}(1), \ldots, W_{K}^{(t)}(1), W_{1}^{(t)}(2), \ldots, W_{K}^{(t)}(2), \ldots, W_{1}^{(t)}(T), \ldots, W_{K}^{(t)}(T) \right)^T.
$$

(9)

Given this system model specification, we can now develop a Bayesian hierarchal GP model for the relay system which will allow us to perform jointly detection and system identification.

**B. Bayesian Real-Time Relay Identification**

Here we present the relevant aspects of the Bayesian model and associated assumptions. First we present the posterior parameters required to be estimated, followed by the prior distributional choices, finishing with the posterior distribution for the directed acyclic graph in Fig. 2. Given the posterior distribution we formulate the relay system identification problem in Section IV. In the remainder of the paper we consider scenarios in which the transmitted symbols are known pilot symbols within each frame. Furthermore, the received signal at the relay is obtained by conditioning on knowledge of either partial CSI or perfect CSI. Finally, we also condition on knowledge of the channel between the relay and the receiver either according to partial CSI or perfect CSI. These assumptions are explicitly presented in Section IV where we will present the system identification solution. With this in mind the Bayesian model is developed to perform system identification at each relay.

The posterior parameters and functions of interest in relay identification are given by the parameter vector after observing $T$ frames,

$$
(f(\cdot), \theta, D) = \left( f^{(1)}(\cdot), \ldots, f^{(L)}(\cdot), \theta^{(1:L)}, D^{(1:L)} \right).
$$

(10)

To make clear the structure of the posterior distribution, we now develop the prior model. We utilize the prior choices for the sequence of symbols and the unknown channel coefficients as specified in [25], detailed below. 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.

Next we present the full prior specification.

**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 $\theta^{(l)} = [\theta_1^{(l)}, \theta_2^{(l)}]$, where $\theta_1^{(l)} \sim N(0, 1)$, $\theta_2^{(l)} \sim N(0, 100)$ for all $l$. 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 l-th relay function is then given by $f^{(l)}(\cdot) \sim \mathcal{G}\mathcal{P} \left( \mu^{(l)}_{\theta^{(l)}}(\cdot), C^{(l)}_{D^{(l)}}(\cdot, \cdot) \right)$, with $\mu^{(l)}_{\theta^{(l)}}$ parameterised by $\theta^{(l)}$ and covariance function $C^{(l)}_{D^{(l)}}(\cdot, \cdot)$ constructed from a kernel parametrised by $D^{(l)}$.

**Posterior Model Structure**

The combination of the likelihood model, priors for model parameters and symbols and the hierarchical priors for the GP prior, when combined under Bayes’ Theorem, result in a full posterior distribution given by Equation (11), from which all statistical inference can be conducted. In particular the relay function identification

$$p(\theta, D, w^{1:L}_{1:T}, f^{1:L}(\cdot) | y^{1:L}_{1:T})$$

$$= \prod_{l=1}^{L} p(\theta^{(l)}, D, w^{(l)}_{1:T}, f^{(l)}(\cdot) | y_{1:T}^{(l)})$$

$$= \prod_{l=1}^{L} \mathcal{G}\mathcal{P} \left( f^{(l)}(\cdot); \mu^{(l)}_{\theta^{(l)}}(\cdot), C^{(l)}_{D^{(l)}}(\cdot, \cdot) \right) p(\theta^{(l)}) p \left( D^{(l)} \right)$$

$$= \prod_{l=1}^{L} \left( \prod_{t=1}^{T} p \left( y^{(l)}(t) | f^{(l)}(r^{(l)}(t)), \theta^{(l)}, D^{(l)}, w^{(l)}(t) \right) \mathcal{G}\mathcal{P} \left( f^{(l)}(\cdot); \mu^{(l)}_{\theta^{(l)}}(\cdot), C^{(l)}_{D^{(l)}}(\cdot, \cdot) \right) \right)$$

$$\times p \left( w^{(l)}(t) \right) p \left( \theta^{(l)} \right) p \left( D^{(l)} \right)$$

$$= \prod_{l=1}^{L} \left( \prod_{t=1}^{T} \prod_{k=1}^{K} p \left( y^{(l)}_{k}(t) | f^{(l)}(r^{(l)}_{k}(t)), \theta^{(l)}, D^{(l)}, w^{(l)}_{k}(t) \right) \mathcal{G}\mathcal{P} \left( f^{(l)}(\cdot); \mu^{(l)}_{\theta^{(l)}}(\cdot), C^{(l)}_{D^{(l)}}(\cdot, \cdot) \right) \right)$$

$$\times p \left( w^{(l)}_{k}(t) \right) p \left( \theta^{(l)} \right) p \left( D^{(l)} \right).$$

Note, we have included auxiliary parameters $w^{1:L}_{1:k}$ 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 [25].

We note that the posterior model we develop is non-standard as we define this distribution jointly over a high-dimensional parameter space in $\mathcal{R}^d$, corresponding to $d = \text{dim} \{ \theta, D, W \}$ and posterior on the space of all possible relay functions $f$ constructed via the choice of kernel that we make. That is the marginal posterior for $p(f(\cdot) | y)$ is defined to take support on a function space, obtained via the GP prior structure we developed.
IV. PROBLEM FORMULATION FOR SYSTEM IDENTIFICATION AND INFERENCE

Now we can specify the marginal posterior distributions of particular interest to the identification problem using the Bayesian model presented in Equation (11). We note that typically in the standard GP regression framework, the values of the predictor, in our wireless model given by \( r_k^{(l)}(t) \) are known or observed input predictor values, one corresponding to each of the observations of the symbols at each of the relays. However, in our model, these will be stochastic due to the additive noise at the relay. Hence, the scenario we consider here for the relay function identification setting is non-standard in this regard as these locations at which we wish to estimate the relay functions are random and unobserved variables. This is due to the fact that the input predictor to our Bayesian relay function identification problem is a random process based on the realization of the random vector \( \mathbf{R}^{(l)} \). In the following section we specify three different solutions to the estimation problem, given this non-standard framework.

This section has three parts: The first presents the most general problem formulation of the identification problem in the case of which partial CSI is considered, and unknown noise realisation. We present a general off-line solution for this problem which could be solved for example via techniques such as MCMC. However, such a solution would be off-line and computationally expensive. The second problem reduces the first problem to a real time on-line solution by considering partial CSI and ZF relay noise (conditional on noise mean) and an efficient algorithm under the Iterated Conditioning on the Modes (ICM) framework. The third presents a lower bound on the estimation accuracy of system identification under the scenario in which we have perfect CSI and consider ZF relay noise (conditional on noise mean).

A. Problem Formulation I

The most general framework for system identification in the wireless relay model presented is based on the posterior estimation of the GP mean and covariance functions, parameterized uniquely by the parameters \( \theta, D \). Then we refine this procedure and the posterior distribution to simplify the estimation using constraints based on knowledge of the Partial Channel State information and a zero-forcing approach.

System Identification Definition 1: We define the system identification problem to be the estimation of the relay processing function, for each \( l \in \{1, \ldots, L\} \), to be estimation of \( f^{(l)}(\cdot) \) at any vector of locations \( \mathbf{R}^{(l)} \). This identification problem can be stated as the solution to the following

\[
\left\{ \hat{f}^{(1:L)}(\cdot), \hat{\theta}^{(1:L)}, \hat{D}^{(1:L)} \right\} = \arg \max_{f^{(1:L)}(\cdot), \theta^{(1:L)}, D^{(1:L)}} \int \prod_{l=1}^{L} \left( \prod_{t=1}^{T} \prod_{k=1}^{K} dw_k^{(l)}(t) p\left( y_k^{(l)}(t) | f^{(l)}(r_k^{(l)}(t)), \theta^{(l)}, D^{(l)}, w_k^{(l)}(t) \right) \right) \times \mathcal{GP} \left( f^{(l)}(\cdot) ; \mu_{\theta^{(l)}}, \sigma_{\theta^{(l)}} \right) p\left( \theta^{(l)} \right) p\left( D^{(l)} \right). 
\]

(12)
Remark 1: This solution is optimal from the perspective that it minimises the variance in the relay identification. Statistical estimation procedures we will compare in both low SNR and high SNR scenarios. The loss of statistical model estimation problem for relay function system identification. Therefore, under these assumptions we may define the following bound on the estimation performance. For advanced statistical solutions to such problems see [25].

In this paper we focus on an alternative class of solutions that has a much lower computational complexity, which admit simple real-time on-line recursive estimation algorithms. Instead of performing numerical integration followed by numerical optimization to perform relay identification, we instead solve the following problem defined in Problem II via an iterative optimisation procedure known as Iterated Conditioning on the modes [23].

B. Problem Formulation II

In formulation of this sub-optimal, though on-line real-time computationally and statistically efficient estimation procedure, we shall make the following assumptions.

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

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

As a result of these two standard simplifying assumptions on the partial CSI and the ZF approximation, the received signal at the relay is given by

$$
\begin{align*}
    r_{ZF}^{(l)}(s, H^{(l)} = \hat{h}^{(l)}, W^{(l)} = 0) &= s\hat{h}^{(l)}, \quad l \in \{1, \cdots, L\}, \\
    Y^{(l)}(s, H^{(l)} = \hat{h}^{(l)}, G^{(l)} = \hat{g}^{(l)}, W^{(l)} = 0) &= f^{(l)}(s\hat{h}^{(l)}) \hat{g}^{(l)} + V^{(l)}, \quad l \in \{1, \cdots, L\},
\end{align*}
$$

where $f^{(l)}(\cdot) \sim GP\left(\mu_{\theta}^{(l)}(\cdot), C_{\theta}^{(l)}(\cdot, \cdot)\right)$. In the results section where we study the performance of our estimation procedures we will compare in both low SNR and high SNR scenarios the loss of statistical efficiency in estimation obtained by making these assumptions, by comparing to the optimal lower bound on the estimation performance. Therefore, under these assumptions we may define the following statistical model estimation problem for relay function system identification.

**System Identification Definition 2:** Under assumptions I and II on the partial CSI and zero forcing, the resulting identification problem can be stated as the following

$$
\begin{align*}
\left\{ \hat{f}^{(1:L)}(\cdot), \hat{g}^{(1:L)}, \hat{D}^{(1:L)} \right\} = \\
\arg\max_{f^{(1:L)}(\cdot), g^{(1:L)}, D^{(1:L)}} \prod_{l=1}^{L} \left( \prod_{t=1}^{T} \prod_{k=1}^{K} \mathbb{P}\left(y^{(l)}_{k}(t) | f^{(l)}(r^{(l)}_{k}(t)) \right) \right) \mathbb{GP}\left(f^{(l)}(\cdot) ; \mu_{\theta}^{(l)}(\cdot), C_{\theta}^{(l)}(\cdot, \cdot)\right) \mathbb{P}\left(D^{(l)}\right)
\end{align*}
$$

12
Next we present the third possible estimation procedure which can act as a lower bound on this estimation, in which we make the more restrictive assumption of perfect CSI.

C. Problem Formulation III

As a lower bound on the identification accuracy we present the following lower bound estimation, based on the following two system assumptions, which would not be considered practical, though meaningful here due to the lower bound on performance they provide.

Assumption III: We consider perfect Channel State Information (PCSI) which means that we still consider the channels $G^{(l)}$ and $H^{(l)}$ to be random vectors, for which in each frame of transmission, for each relay up-link and down-link channel, we know the exact channel realizations.

Through Assumption II and Assumption III, we obtain a realistic lower bound on statistical accuracy and performance that one can achieve in this system identification framework when applying a solution to Problem II. This is because they result in knowledge of the following form,

\[
\mathbf{r}^{(l)}_{ZF} \left( \mathbf{s}, \mathbf{H}^{(l)} = h^{(l)}, \mathbf{W}^{(l)} = 0 \right) = sh^{(l)}, \quad l \in \{1, \ldots, L\},
\]

\[
\mathbf{Y}^{(l)} \left( \mathbf{s}, \mathbf{H}^{(l)} = h^{(l)}, \mathbf{G}^{(l)} = g^{(l)}, \mathbf{W}^{(l)} = 0 \right) = f^{(l)} \left( sh^{(l)} \right) g^{(l)} + \mathbf{V}^{(l)}, \quad l \in \{1, \ldots, L\}.
\]

This allows us to make the following third system identification definition for the lower bound on performance.

System Identification Definition 3: Under assumptions II and III, the resulting lower bound on the identification problem can be stated as the following

\[
\left\{ \hat{\mathbf{f}}_{LB}^{(1:L)} \left( \cdot \right), \hat{\mathbf{g}}_{LB}^{(1:L)} \left( \cdot \right), \hat{\mathbf{D}}_{LB}^{(1:L)} \left( \cdot \right) \right\} = \arg \max_{\mathbf{f}^{(1:L)} \left( \cdot \right), \mathbf{g}^{(1:L)} \left( \cdot \right), \mathbf{D}^{(1:L)} \left( \cdot \right)} \prod_{l=1}^{L} \left( \prod_{T=1}^{T} \prod_{K=1}^{K} p \left( y_{k}^{(l)} \left( t \right) | f^{(l)} \left( \hat{y}_{k}^{(l)} \left( t \right) \right) \right) \right) \mathcal{GP} \left( \mathbf{f}^{(l)} \left( \cdot \right) ; \mathbf{\mu}_{\mathbf{f}}^{(l)} \left( \cdot \right), \mathbf{C}_{\mathbf{f}}^{(l)} \left( \cdot, \cdot \right) \right) p \left( \mathbf{\theta}^{(l)} \right) p \left( \mathbf{D}^{(l)} \right).
\]

We note that the system identification aspect of the estimation problem involve estimation of the Maximum a Posteriori (MAP) relay function values locally at each of these predictor locations given by the map estimator $\hat{\mathbf{f}}^{(l)} \left( \cdot \right)$. One may then utilise the results of this estimation for prediction of the unknown mean structure of the function $f^{(l)} \left( \cdot \right)$, conditional on estimated hyper parameters for the mean and covariance function, at new input values of the received signal, $r_{s}$, via the identity

\[
\bar{f}_{s}^{(l)} \triangleq \mathbb{E} \left[ f^{(l)} \left( r_{s} \right) | y_{T}^{(l)} \left( \cdot \right), \hat{\mathbf{g}}^{(l)} \left( \cdot \right), \hat{\mathbf{D}}^{(l)} \left( \cdot \right), \hat{\mathbf{R}}_{T}^{(l)} \left( \cdot \right), r_{s} \right] = \mathbf{K}^{(l)} \left( r_{s}, \hat{\mathbf{R}}_{T}^{(l)} \right) \left( \mathbf{K}^{(l)}_{T} \right)^{-1} y_{T}^{(l)}.
\]
V. SYSTEM IDENTIFICATION ESTIMATION ALGORITHM VIA ITERATED CONDITIONAL MODES

Having formulated the relay identification and prediction problem, as well as the ICM based estimation procedure, we now address algorithmic procedures to solve efficiently Equation (12). Hence, we provide solutions to problems II and III in three different on-line real-time recursive computationally efficient estimation procedures. Each of these estimation procedures is based on a methodology developed in [23] known as Iterated Conditioning on the Modes (ICM) which we develop and adapt for a computationally efficient solution to the GP system identification problem.

A. Understanding Iterated Conditioning on the Modes Estimation for System Identification

Originally developed in [26] and [23] for efficient MAP estimation in very high dimensional Bayesian models developed for Markov random fields in the analysis of images with speckle noise, this estimation procedure has developed in many other areas of estimation, see [27] and [28] for example.

As discussed in [29] ICM is fundamentally a deterministic optimization method that finds the joint posterior modal estimators corresponding to the MAP estimates. We illustrate this simple algorithm on a generic two parameter example developed in [29] before extending the concept to the solution to the relay identification problem in wireless communications. In this example a generic bivariate posterior distribution \( p(\theta_1, \theta_2 | y_{1:T}) \) is considered and the aim is to find the MAP estimate corresponding to the mode, which satisfies

\[
\frac{\partial}{\partial \theta_1} p(\theta_1, \theta_2 | y_{1:T})|_{\theta_2=\hat{\theta}_2} = \frac{\partial}{\partial \theta_2} p(\theta_1, \theta_2 | y_{1:T})|_{\theta_1=\hat{\theta}_1} = 0. \tag{18}
\]

This is equivalent to

\[
p(\theta_2 | y_{1:T}) \frac{\partial}{\partial \theta_1} p(\theta_1 | \theta_2, y_{1:T})|_{\theta_2=\hat{\theta}_2} = p(\theta_1 | y_{1:T}) \frac{\partial}{\partial \theta_2} p(\theta_2 | \theta_1, y_{1:T})|_{\theta_1=\hat{\theta}_1} = 0, \tag{19}
\]

assuming that \( p(\theta_1 | y_{1:T}) \neq 0 \) and \( p(\theta_2 | y_{1:T}) \neq 0 \). Hence, given the full conditional posterior distributions \( p(\theta_1 | \theta_2, y_{1:T}) \) and \( p(\theta_2 | \theta_1, y_{1:T}) \) and solutions for their modes \( \hat{\theta}_1 = \hat{\theta}_1(\theta_2, y_{1:T}) \) and \( \hat{\theta}_2 = \hat{\theta}_2(\theta_1, y_{1:T}) \). The algorithm then proceeds by initializing the mode estimates and successively iterating over updates of conditional mode estimates, where at iteration \( j \), the update of \( \hat{\theta}_1^{(j)} \) conditions on the mode estimate of \( \hat{\theta}_2^{(j-1)} \) and then the update of \( \hat{\theta}_2^{(j)} \) conditions on the mode estimate \( \hat{\theta}_1^{(j)} \). This is repeated either for a fixed number of iterations or until a convergence criterion is satisfied. In cases in which the posterior full conditionals are unimodal, this procedure is guaranteed to converge to the global maximum, in other multi-modal settings, the ICM algorithm is guaranteed to converge to a maximum, though it may be a local maximum.

Hence, all variants of the ICM estimation procedure have the following in common, they first involve specification of a multivariate posterior distribution, deconstructed as a set of full conditional posterior distributions. In this paper, these full conditional posterior distributions are given, for the \( l \)-th relay,
based on the full posterior in Equation (11), after applying Assumptions I and II or Assumptions II and III,

\[
p \left( f^{(l)}(\cdot) \mid \theta, D, y_{1:T}^{(l)}} \right) \propto \prod_{t=1}^{T} \prod_{k=1}^{K} p \left( y_k(t) \mid f^{(l)}(t), \theta^{(l)}, D^{(l)}, w_k(t) \right) GP \left( f^{(l)}(t) ; \mu_{\theta^{(l)}}(t), \gamma_{D^{(l)}}(\cdot, \cdot) \right); \quad (20a)
\]

\[
p \left( \theta \mid D, f^{(l)}(\cdot), y_{1:T}^{(l)}} \right) \propto \prod_{t=1}^{T} \prod_{k=1}^{K} GP \left( f^{(l)}(t) ; \mu_{\theta^{(l)}}(t), \gamma_{D^{(l)}}(\cdot, \cdot) \right) p \left( \theta^{(l)} \right); \quad (20b)
\]

\[
p \left( D \mid \theta, f^{(l)}(\cdot), y_{1:T}^{(l)}} \right) \propto \prod_{t=1}^{T} \prod_{k=1}^{K} GP \left( f^{(l)}(t) ; \mu_{\theta^{(l)}}(t), \gamma_{D^{(l)}}(\cdot, \cdot) \right) p \left( D^{(l)} \right). \quad (20c)
\]

Next, given these full conditional posterior distributions, the ICM algorithm initializes the estimate of the MAP solutions, denoted by \( \left( \hat{f}^{(l)}_{1:T}, \hat{\theta}^{(0)}, \hat{D}^{(0)} \right) \), after having observed \( T \) frames of \( K \) symbols for the \( l \)-th relay, defined as the relay identification estimates in problem formulations I, II and III. Then the \( j \)-th iteration of the ICM algorithm successively updates each estimate of \( \left( \hat{f}^{(l)}_{1:T}, \hat{\theta}^{(j)}, \hat{D}^{(j)} \right) \) based on the solutions at iteration \( (j - 1) \), and the solutions to the following sequence of MAP estimates for the full conditional posteriors,

\[
Vec[\hat{f}^{(l)}_{1:T}] = \arg \max_{f^{(l)}_{1:T}} p \left( f^{(l)}(\cdot) \mid \hat{\theta}^{(j-1)}, \hat{D}^{(j-1)}, y_{1:T}^{(l)} \right); \quad (21a)
\]

\[
\hat{\theta}^{(j)} = \arg \max_{\theta} p \left( \theta \mid \hat{D}^{(j-1)}, \hat{f}^{(l)}_{1:T}, \hat{y}_{1:T}^{(l)} \right); \quad (21b)
\]

\[
\hat{D}^{(j)} = \arg \max_{D} p \left( D \mid \hat{\theta}^{(j)}, \hat{f}^{(l)}_{1:T}, \hat{y}_{1:T}^{(l)} \right). \quad (21c)
\]

Iteration of the procedure defined above is analogous to a Gibbs sampler, except instead of sampling recursively from each full conditional posterior distribution in order to construct a Markov chain to sample from the posterior in Equation (11), we maximize successively the full conditional, based on the previous estimated maximums. Repeating this procedure can guarantee convergence to a maximum, see [23], where it is also discussed that ICM can be made equivalent to instantaneous freezing in a stochastic optimization procedure known as simulated annealing [30].

Iterating this procedure successively produces a sequence of MAP estimates which converge to an optimum solution, typically for a small number of iterations \( J \). Additionally, the number of ICM iterations \( J \) required to obtain the optimal MAP solutions for each set of \( T \) frames of observations of \( K \) symbols in the problem formulation developed in this paper is designed to be independent of \( T \) and \( K \). The reason for this is due to the conjugacy we exploit for the model MAP estimation of the vector component which grows linear in dimension with \( KT \), corresponding to the vector \( Vec[f^{(l)}_{1:T}] \). In the next section we present Theorem 1 which derives closed form analytic results for the full conditional posterior distribution MAP estimates for the full conditionals for \( Vec[f^{(l)}_{1:T}] \), \( \hat{\theta} \) and an analytic expression for the gradient of the kernel bandwidth for \( \hat{D} \).
B. Generic ICM Estimation for Wireless Relay System Identification via GP’s

To develop an ICM algorithm, we need to construct a block Gibbs framework for posterior inference, and obtain expressions for the mode of the full conditional posterior distributions. This is equivalent to finding the conditional MAP estimate. In this problem we will exploit conjugacy properties of the posterior model developed in Equation (12). We begin by specifying the posterior full conditional distributions and then analytic expressions for their modes in Theorem 1.

**Theorem 1:** Conditional on partial CSI and relay noise ZF according to Equation (15), the full conditional distributions of the relay identification system model are given by:

1) The full conditional posterior distribution for the $l$-th relay function in (20a) is given by

$$p \left( \text{Vec} \left[ f_{1:T}^{(l)} \right] | \theta^{(l)}, D^{(l)}, y_{1:T}^{(l)} \right) = N \left( \mathbf{M}, \Sigma \right),$$

where

$$\mathbf{M} = \left( \left(K_{1:T}^{(l)} \right)^{-1} + \frac{1}{\sigma^2} \mathbf{I}_{K \times T} \right)^{-1} \left( \left(K_{1:T}^{(l)} \right)^{-1} \mathbf{M}_0 + \frac{1}{\sigma^2} \mathbf{I}_{K \times T} \text{Vec} \left[ y_{1:T}^{(l)} \right] \right),$$

with

$$\mathbf{M}_0 = \left[ \begin{array}{c} \mu_{\theta^{(l)}} \left( r_{1}^{(l)} \left( 1 \right) \right), \ldots, \mu_{\theta^{(l)}} \left( r_{K}^{(l)} \left( 1 \right) \right), \ldots, \mu_{\theta^{(l)}} \left( r_{1}^{(l)} \left( T \right) \right), \ldots, \mu_{\theta^{(l)}} \left( r_{K}^{(l)} \left( T \right) \right) \end{array} \right]^{T},$$

and

$$\Sigma = \left( \left(K_{1:T}^{(l)} \right)^{-1} + \frac{1}{\sigma^2} \mathbf{I}_{K \times T} \right)^{-1}.$$ (25)

The conditional MAP estimate of the relay function in (21a), denoted by Vec $\left[ \hat{f}_{1:T}^{(l)} \right]$, is given by

$$\text{Vec} \left[ \hat{f}_{1:T}^{(l)} \right] = \arg \max_{f_{1:T}^{(l)}} p \left( \text{Vec} \left[ f_{1:T}^{(l)} \right] | \theta^{(l)}, D^{(l)}, y_{1:T}^{(l)} \right) = \mathbf{M}.$$ (26)

2) The full conditional for $\theta^{(l)}$ in (20b) can be expressed as

$$p \left( \theta^{(l)} | D^{(l)}, f_{1:T}^{(l)} \left( \cdot \right), y_{1:T}^{(l)} \right) \propto N \left( \text{Vec} \left[ f_{1:T}^{(l)} \right] ; \mu_{1:T}^{(l)}, K_{1:T}^{(l)} \right) p \left( \theta^{(l)} \right).$$ (27)

The conditional MAP estimate of $\theta^{(l)} = \left[ \theta_{1}^{(l)}, \theta_{2}^{(l)} \right]$ in (21b), denoted by $\hat{\theta}^{(l)}$, is given by

$$\theta_1 = \frac{-\theta_2 \left[ \left( \Sigma_{\theta} \right)^{-1} \right]_{2,1} + \sum_{i=1}^{K} \sum_{t=1}^{T} r_i(t) \left[ \left( K_{1:T} \right)^{-1} \right]_{i,t} - \frac{1}{2} \sum_{i=1}^{K} \sum_{t=1}^{T} f(r_i(t)) \left[ \left( K_{1:T} \right)^{-1} \right]_{i,t}} {1 + \sum_{i=1}^{K} \sum_{t=1}^{T} r_i(t) \left[ \left( K_{1:T} \right)^{-1} \right]_{i,t}}$$

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

(28)
where
\[
\Phi = -\left(\left[(\Sigma \Theta)^{-1}\right]_{2,1} + \sum_{i=1}^{K} \sum_{t=1}^{T} r_i(t) \left[(K_{1:T})^{-1}\right]_{i,t}\right) \left(\left[(\Sigma \Theta)^{-1}\right]_{1,2} + \sum_{i=1}^{K} \sum_{t=1}^{T} r_i^2(t) \left[(K_{1:T})^{-1}\right]_{i,t}\right) \\
+ \left[(\Sigma \Theta)^{-1}\right]_{2,2} + \sum_{i=1}^{K} \sum_{t=1}^{T} r_i^2(t) \left[(K_{1:T})^{-1}\right]_{i,t}.
\]
(29)

3) The full conditional for \(D^{(l)}\) in in (20c) can be expressed as
\[
p \left(D^{(l)}|f_{1:T}^{(l)}, \theta^{(l)}, y_{1:T}^{(l)}\right) \propto N \left(\text{Vec} \left[f_{1:T}^{(l)}; \mu_{1:T}^{(l)}, K_{1:T}^{(l)} \right] p \left(D^{(l)}\right) \right).
\]
(30)
The conditional MAP estimate of \(D^{(l)}\) in (21c), denoted by \(\hat{D}^{(l)}\) is given as the numerical solution to
\[
\left[-\frac{1}{2} \text{Tr} \left(K_{1:T}^{-1} dK_{1:T} dD^{(l)}\right) + \frac{1}{2} \left(\text{Vec} [f_{1:T}^T - \mu_{1:T}]^T (K_{1:T})^{-1} dK_{1:T} dD^{(l)} (K_{1:T})^{-1} \left(\text{Vec} [f_{1:T}^T - \mu_{1:T}]\right)\right) \right] \left[D^{(l)}\right] = 0,
\]
(31)
where we have
\[
\frac{dK_{1:T}}{dD^{(l)}} = \frac{1}{D^{(l)}} \Phi_{1:T} \odot K_{1:T},
\]
(32)
and \(\odot\) is the Kronecker product, and the \(i, j\)-th element of \(\Phi_{1:T}\) is given by \([\Phi_{1:T}]_{i,j} = ||R_i^{(l)} - R_j^{(l)}||^2\).

The proof of Theorem 1 is provided in Appendix 1.

C. Computationally Efficient Recursive On-Line ICM-GP System Identification

To solve these problems in real-time, we require a low complexity on-line algorithm. We consider developing a version of Iterative Conditioning on the Mode (ICM) \([31]\) which exploits different structural properties of the GP model. This approach is generically defined under several possible computationally efficient solutions, each based on the full conditional posterior MAP estimation results developed in Theorem 1.

We begin with the most computationally inefficient approach, though it represents the optimal solution as it utilises all information obtained to update the estimated function. This is detailed in Online Estimation I.

Online Estimation I - Full Information

This approach is optimal since, for each relay \(l\), having observed \(t\)-frames of \(K\)-symbols of observed data \(Y_{1:t}^{(l)} = (Y_{1}^{(l)}(1), \ldots, Y_{K}^{(l)}(1), \ldots, Y_{1}^{(l)}(t), \ldots, Y_{K}^{(l)}(t))\), we evaluate the Gramm-Matrix based on
\[
K_{1:t}^{(l)} = \begin{bmatrix}
C \left(R_{1}^{(l)}(1), R_{1}^{(l)}(1)\right) & \cdots & C \left(R_{1}^{(l)}(1), R_{K}^{(l)}(t)\right) \\
\vdots & \ddots & \vdots \\
C \left(R_{K}^{(l)}(t), R_{1}^{(l)}(1)\right) & \cdots & C \left(R_{K}^{(l)}(t), R_{K}^{(l)}(t)\right)
\end{bmatrix}.
\]
(33)
This approach is optimal in terms of utilising all the observed information in the estimation of the ICM algorithm. However for \(J\) iterations of the ICM algorithm it will be of complexity \(O(JK^2 t^2)\) in memory.
usage and $O(JK^3t^3)$ in computational complexity. This is primarily due to the cost of inverting the Gram matrix in each update stage of the ICM algorithm. However, in this approach each estimation that is performed based on a new frame of length $K$ of observed symbols, the estimation of the matrix of function values, $f_{1:t}^{(l)}$, is completely updated based on all past history. To see this consider for example the evaluation of Equation (23) for updating the ICM MAP estimate of the function values given by,

$$M = \left(\left(K_{1:T}^{(l)}\right)^{-1} + \frac{1}{\sigma^2_V}I_{K \times T}\right)^{-1} \left(\left(K_{1:T}^{(l)}\right)^{-1}M_0 + \frac{1}{\sigma^2_V}I_{K \times T}\vec{y}_{1:T}^{(l)}\right). \quad (34)$$

This expression involves the evaluation of the Gram-matrix $K_{1:t}^{(l)}$ and its inverse, which become rapidly computationally expensive though analytically exact for the update of the MAP estimate of the function values $f_{1:t}^{(l)}$ at each iteration of the ICM algorithm. We circumvent this growing computational complexity with two alternative approaches presented next and compare their performance to this optimal solution. 

**Online Estimation II - Frame-by-Frame**

This approach is sub-optimal though significantly more computationally efficient and viable for a real-time online algorithm. In this proposed approach, for each relay $l$, having observed the $t$-th frame of $K$-symbols of observed data $Y_t^{(l)} = (y_1^{(l)}(t), \ldots, y_K^{(l)}(t))$, we evaluate the Gram-Matrix based on just the current frame of data,

$$K_t^{(l)} = \begin{bmatrix} C \left(R_1^{(l)}(t), R_1^{(l)}(t)\right) & \cdots & C \left(R_1^{(l)}(t), R_K^{(l)}(t)\right) \\ \vdots & \ddots & \vdots \\ C \left(R_K^{(l)}(t), R_1^{(l)}(t)\right) & \cdots & C \left(R_K^{(l)}(t), R_K^{(l)}(t)\right) \end{bmatrix}. \quad (35)$$

This approach is sub-optimal in terms of utilising only the current frame of observed information in the estimation of the ICM algorithm update at frame $t$ for $f_t^{(l)}$. Effectively it is equivalent to partitioning the Gram matrix as follows,

$$K_{1:t}^{(l)} = K_1^{(l)} \oplus K_2^{(l)} \oplus \cdots \oplus K_t^{(l)} = \begin{bmatrix} K_1^{(l)} & 0 & \cdots & 0 \\ 0 & K_2^{(l)} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & K_t^{(l)} \end{bmatrix}.$$ 

However, for $J$ iterations of the ICM algorithm at frame $t$ it will be of complexity $O(JK^2)$ in memory usage and $O(JK^3)$ in computational complexity. In addition, since the relay functions being estimated are not time varying, the estimates obtained on previous frames ($f_{1:t-1}^{(l)}$) and their uncertainty can be combined in several different approaches for example via an update on-line according to the following recursions,

$$m_t = m_{t-1} + \frac{1}{t} \left(\tilde{f}_t^{(l)} - m_{t-1}\right)$$

$$\Phi_t = \Phi_{t-1} + \frac{1}{t} \left(\tilde{f}_t^{(l)} - m_t\right)\left(\tilde{f}_t^{(l)} - m_{t}\right)' - \Phi_{t-1}. \quad (36)$$
Here, the matrix of function estimates, denoted by \( \tilde{f}_l(t) \), corresponds to function values obtained for frame \( t \) from the ICM estimates \( f_l(t) \), quantised to a grid of predictor values \( R_1, \ldots, R_S \) which partition the convex hull of the relay symbols based on the constellation being transmitted. These quantised values are then included in an online average of the function values for each frame, where the mean function vector at these quantised grid points is denoted by \( m_t \) and its uncertainty measured by a covariance in the estimate at these quantised values is denoted by \( \Phi_t \).

**Online Estimation III - Sliding Window**

This approach provides a trade-off in computational complexity and optimality between the proposed online estimation approaches I and II. It is sub-optimal though significantly more computationally efficient and viable for a real-time online algorithm than approach I, though it recursively uses previous frames observed symbols as opposed to the block wise analysis in approach II.

Under this proposed approach, for each relay \( l \), having observed \( \tau \) symbols of observed data \( Y^{(l)}_{\tau} = \left( Y^{(l)}_{1}(1), \ldots, Y^{(l)}_{K}(1), \ldots, Y^{(l)}_{k}(\tau) \right) \), we evaluate the Gramm-Matrix according to a sliding window based on the past block of \( S \) observed symbols. In the example provided we consider blocks of length \( K \) (ie. length of the frame) and an update of the function estimates for each newly observed symbols observation. This corresponds to utilizing a Gramm-matrix for each update with structure given by,

\[
K^{(l)}_{\tau} = \begin{bmatrix}
\hat{K}^{(l)}_{\tau-1} & \tilde{k}^{(l)}_{\tau-1}(R^{(l)}(\tau)) \\
\tilde{k}^{(l)}_{\tau-1}(R^{(l)}(\tau))^T & C(R^{(l)}(\tau), R^{(l)}(\tau))
\end{bmatrix},
\]

(37)

where we have dropped the frame index on the received signals which is redundant in this specification. We define the last row and column of the new Gramm matrix as the vector \( \tilde{k}^{(l)}_{\tau-1}(R^{(l)}(\tau)) = [C(R^{(l)}(\tau - S + 1), R^{(l)}(\tau)), \ldots, C(R^{(l)}(\tau - 1), R^{(l)}(\tau))] \). In addition the modified Gramm matrix \( \tilde{K}^{(l)}_{\tau-1} \) corresponds to the “down sized” version of the regularised matrix \( K^{(l)}_{\tau-1} \) obtained by removing the first row and column, ensuring the sliding window structure and the fixed dimensionality of \( K^{(l)}_{\tau} \) for all times \( \tau \).

This particular approach is efficient as it allows one to utilise a very efficient computational evaluation of inverse Gramm matrix \( K^{(l)}_{\tau-1} \) in Equation (23) which is used for updating the ICM MAP estimate of the function values. This efficiency is obtained utilising recursive knowledge of the previous updates Gramm matrix and its inverse \( \left[ K^{(l)}_{\tau-1} \right]^{-1} \) under an approach described in [19], which utilises the Sherman - Morrison - Woodbury matrix inversion lemma [32] in two applications as detailed below for the context of this paper.

1) Evaluate the inverse of the down sized Gramm matrix \( \tilde{K}^{(l)}_{\tau-1} \).
2) Evaluate the inverse of the up sized Gramm matrix \( K^{(l)}_{\tau} \) based on the inverse of \( \tilde{K}^{(l)}_{\tau-1} \).

Dropping the relay index \( l \) for convenience, the inverse of the down sized matrix is obtained using
the previous evaluation of the Gramm matrix, decomposed according to
\[
\mathcal{K}_{\tau-1} = \begin{bmatrix}
\mathcal{C}(R(\tau - S +1), R(\tau - S +1)) & \tilde{k}_{\tau-1}(R(\tau - 1)) \\
\tilde{k}_{\tau-1}(R(\tau - 1)) & \mathcal{K}_{\tau-1}
\end{bmatrix},
\]
and its inverse decomposed as
\[
\mathcal{K}_{\tau-1}^{-1} = \begin{bmatrix}
[\mathcal{K}_{\tau-1}^{-1}]_{11} & [\mathcal{K}_{\tau-1}^{-1}]_{12:1S} \\
[\mathcal{K}_{\tau-1}^{-1}]_{12:1S}^T & [\mathcal{K}_{\tau-1}^{-1}]_{22:SS}
\end{bmatrix},
\]
where \([\mathcal{K}_{\tau-1}^{-1}]_{22:SS}\) denotes the lower \((2 \times S)(2 \times S)\) sub-block of the inverse matrix \(\mathcal{K}_{\tau-1}^{-1}\). Under these decompositions, the inverse for the down sized Gramm matrix is obtained using the following matrix inversion lemma identity [33]
\[
\tilde{k}_{\tau-1}^{-1} = [\mathcal{K}_{\tau-1}^{-1}]_{22:SS} - \frac{1}{[\mathcal{K}_{\tau-1}^{-1}]_{11}} [\mathcal{K}_{\tau-1}^{-1}]_{12:1S} [\mathcal{K}_{\tau-1}^{-1}]_{12:1S}^T.
\]

Having obtained the inverse of the down sized Gramm matrix in terms of values known at iteration \(\tau - 1\), the inverse of the up sized Gramm matrix proceeds according to the following decompositions again via application of the matrix inversion lemma [32]. Again, dropping the relay index \(l\) for convenience, the inverse of the up sized matrix is obtained using the previous evaluation of the Gramm matrix, decomposed according to
\[
\mathcal{K}_{\tau} = \begin{bmatrix}
\tilde{k}_{\tau-1} & \tilde{k}_{\tau}(R(\tau)) \\
\tilde{k}_{\tau}(R(\tau))^T & \mathcal{C}(R(\tau), R(\tau))
\end{bmatrix},
\]
and its inverse decomposed as
\[
\mathcal{K}_{\tau}^{-1} = \begin{bmatrix}
[\mathcal{K}_{\tau}^{-1}]_{11:(S-1)(S-1)} & [\mathcal{K}_{\tau}^{-1}]_{12:1S} \\
[\mathcal{K}_{\tau}^{-1}]_{12:1S}^T & [\mathcal{K}_{\tau}^{-1}]_{SS}
\end{bmatrix},
\]
where \([\mathcal{K}_{\tau}^{-1}]_{11:(S-1)(S-1)}\) denotes the upper \((2 \times S)(2 \times S)\) sub-block of the inverse matrix \(\mathcal{K}_{\tau}^{-1}\). Applying the matrix inversion lemma, one can obtain further decomposition according to the previously evaluated down sized Gramm matrix inverse as follows,
\[
\mathcal{K}_{\tau}^{-1} = \begin{bmatrix}
\tilde{k}_{\tau-1}^{-1} \left( I + \tilde{k}_{\tau}(R(\tau)) \tilde{k}_{\tau}(R(\tau))^T \tilde{k}_{\tau-1}^{-1H} [\mathcal{K}_{\tau}^{-1}]_{SS} \right) & -\tilde{k}_{\tau-1}^{-1} \tilde{k}_{\tau}(R(\tau)) [\mathcal{K}_{\tau}^{-1}]_{SS} \\
-\left( \tilde{k}_{\tau-1}^{-1} \tilde{k}_{\tau}(R(\tau)) \right)^T [\mathcal{K}_{\tau}^{-1}]_{SS} & [\mathcal{K}_{\tau}^{-1}]_{SS}
\end{bmatrix},
\]
with \([\mathcal{K}_{\tau}^{-1}]_{SS} = \left( \mathcal{C}(R(\tau), R(\tau)) - \tilde{k}_{\tau}(R(\tau))^T \tilde{k}_{\tau-1}^{-1} \tilde{k}_{\tau}(R(\tau)) \right)^{-1}\).

In addition, since the relay functions being estimated are not time varying, as detailed in approach II, the estimates obtained on previous frames \(\left(f_{1:1:1}^{(l)}\right)\) and their uncertainty can be combined in several different approaches for example via the on-line update mechanism described in Equation (36).

VI. SIMULATION RESULTS

In this section, we present the performance of the proposed algorithms via Monte Carlo simulations.
A. Simulation Set-up

The simulation settings for all the simulations are as follows:

- 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 $\sigma^2_G = \sigma^2_F = 0.2$;
- We define the receive SNR as the ratio of the average received signal power to the average noise power,

$$\text{SNR} = 10 \log \frac{\text{Tr} \left[ \mathbb{E} \left( (G(l)F(l)s(l)) (G(l)F(l)s(l))^H \right) \right]}{\text{Tr} \left[ \mathbb{E} \left( (G(l)V(l) + W(l)) (G(l)V(l) + W(l))^H \right) \right]} = 10 \log \frac{1}{\sigma^2_V + \sigma^2_W};$$

- 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 over the solutions to Equations (21a-21c);
- The relay functions tested corresponded to absolute, linear, sinusoidal, hyperbolic-tan and demodulated;
- Simulations were performed for Full kernel matrix online estimation approach I under both perfect CSI and partial CSI; partial kernel matrix online estimation approach II with frame by frame estimation under both perfect CSI and partial CSI; and a sliding window with 99% overlap online estimation approach III under both perfect CSI and partial CSI.

B. Analysis of the Online Relay Identification Algorithms

We begin this section with analysing the results for the case in which the GP system identification we develop is applied to an increasing number of frames of transmitted symbols, according to the Full Information Approach I in which the kernel matrix is not truncated in any manner. These are the results with the highest computational complexity but with out any reduction of the kernel matrix, hence these results were produced form the full set of observed information over time, as presented diagrammatically in Subplot (a) in Figure (3).

Analysis of the performance of the ICM algorithm can be studied in several ways, we first presented in Figure (4) 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 $j = 1, \ldots, 50$ for the online estimation after receiving 20 frames, each with 200 symbols. 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.
Next we summarize the findings of the analysis for the perfect CSI and partial CSI studies of each of the different relay function we tested: absolute Figure (5), linear Figure (6), sinusoidal Figure (7), tanh Figure (8) and demodulated Figure (9). These results are presented each in four sub-plots, the first two are obtained from the online Full Information Approach I and the last two are for the least information, most efficient online Frame-by-Frame Approach II. In comparing these results we get a representative comparison of the relative performance of undertaking the computationally more expensive estimation approach I versus the computationally more efficient, though truncated information approach II. The results we present are the posterior MAP estimated relay function estimated under our GP-ICM frameworks evaluated at the PAM constellation points via Equation (17) (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 for the online Full Information Approach I.
- 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 observed that in the Full Information case 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.
- We also note that when looking at the Frame-by-Frame results in which we perform the running average over estimates on each frame, according to Equation (36), at fixed grid points set by the constellation transmission symbols space, we observe a smoothing of the estimates which can help resolve the local resolution when compared to the estimation of the relay function at each observed constellation point as undertaken when producing results in Approach I. This smoothing approach could also be applied to Approach 1.

Next, we present summarised results in Table (I), which are each based on the absolute error in the MAP estimated relay functions under Approaches I, II and III over 100 frames. We observe that as expected, there is a clear trade-off between computationally efficiency and accuracy in the estimation.
The online Full-Information Approach I is most accurate followed by the Overlapping-Sliding-Window which in this simulation had a 99% overlap, with the least accurate, though most computationally efficient, approach corresponding to the online Fram-by-Fram estimation.

VII. CONCLUSIONS

In this paper we considered the problem of semi-blind relay identification in the complicated 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 this wireless relay network.

In particular we demonstrated that our GP identification via ICM can resolve this problem of system identification in a computationally efficient real-time online algorithm for many different relay functional forms which have desirable characteristics with respect to transmission functionality and quality of service.

REFERENCES


In this section we provide a proof for the expressions in Theorem 1.

Proof: The posterior distribution decomposes according to Equation (11), 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 (20a) is given by

\[
p \left( f^{(l)}_{1:T}, \theta^{(l)}, D^{(l)}, y^{(l)}_{1:T} \right) = \frac{p \left( y^{(l)}_{1:T} | f^{(l)}_{1:T} \right) p \left( f^{(l)}_{1:T}, \mu^{(l)}_{\theta^{(l)}}, r^{(l)}_t(t), K^{(l)}_{1:T} \right)}{\int p \left( y^{(l)}_{1:T} | f^{(l)}_{1:T} \right) p \left( f^{(l)}_{1:T}, \mu^{(l)}_{\theta^{(l)}}, r^{(l)}_t(t), K^{(l)}_{1:T} \right) d f^{(l)}_{1:T}}.
\]

(43)

With a matrix variate normal likelihood model for \( y^{(l)}_{1:T} | f^{(l)}_{1:T} \) 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 multivariate Gaussian distributions which admit standard conjugacy properties, see [34]. This results in

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

(44)

where \( M \) and \( \Sigma \) are defined in (23-25).

2) The full conditional for \( \theta^{(l)} \) in (20b) can be expressed as

\[
p \left( \theta^{(l)} | D^{(l)}, f^{(l)}_{1:T} (\cdot), y^{(l)}_{1:T} \right)
\]

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

(45)

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:

\[
\nabla_{\theta^{(l)}} \left[ - \log \left( 2 \pi \det \left[ \Sigma^{(l)}_{1:T} \right] \right) \frac{1}{2} \left( \text{Vec} \left[ f^{(l)}_{1:T} \right] - \mu^{(l)}_{\theta^{(l)}} \right)^T \left( \Sigma^{(l)}_{1:T} \right)^{-1} \left( \text{Vec} \left[ f^{(l)}_{1:T} \right] - \mu^{(l)}_{\theta^{(l)}} \right) + \log p \left( \theta^{(l)} \right) \right] = 0
\]

(46)

which gives

\[
\nabla_{\theta^{(l)}} \left[ \left( \text{Vec} \left[ f^{(l)}_{1:T} \right] - \mu^{(l)}_{\theta^{(l)}} \right)^T \left( \Sigma^{(l)}_{1:T} \right)^{-1} \left( \text{Vec} \left[ f^{(l)}_{1:T} \right] - \mu^{(l)}_{\theta^{(l)}} \right) + \log p \left( \theta^{(l)} \right) \right] =
\]

\[
\left( \begin{array}{cccccc}
1 & 1 & 1 & 1 & 1 & 1 \\
r_1^{(1)} & \ldots & r_K^{(1)} & 1 & \ldots & r_K^{(T)}
\end{array} \right)
\]

(47)

which produces the following linear system of equations with a unique solution:

\[
\left( \theta_1 + \theta_2 \right) \sum_{i=1}^{K} \sum_{t=1}^{T} r_i(t) \left( \left( \Sigma_{\theta} \right)^{-1} \right)_{i,t} = - \frac{1}{2} \sum_{i=1}^{K} \sum_{t=1}^{T} f(r_i(t)) \left( \left( \Sigma_{\theta} \right)^{-1} \right)_{i,t},
\]

(48)
3) The full conditional for $D$ derivative with respect to $\log$
Deriving the MAP estimate is achieved by applying the choice given in Section III-A item 8, ensures that this parameter is a scalar random variable.
where we clarify that the Gramm matrix $\mathcal{K}$
complete this we derive the
where this expression is obtained by using standard matrix derivative identities, see [32]. To
which gives

$$
\theta_1 = -\frac{1}{2} \left[ \left( \Sigma_\Theta \right)^{-1} \right]_{2,1} + \sum_{i=1}^{K} \sum_{t=1}^{T} r_i(t) \left[ \left( \mathcal{K}_{1:T} \right)^{-1} \right]_{i,t} - \frac{1}{2} \sum_{i=1}^{K} \sum_{t=1}^{T} f(r_i(t)) \left[ \left( \mathcal{K}_{1:T} \right)^{-1} \right]_{i,t},
$$

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

where

$$
\Phi = - \left[ \left( \Sigma_\Theta \right)^{-1} \right]_{2,1} + \sum_{i=1}^{K} \sum_{t=1}^{T} r_i(t) \left[ \left( \mathcal{K}_{1:T} \right)^{-1} \right]_{i,t} \left[ \left( \Sigma_\Theta \right)^{-1} \right]_{1,2} + \sum_{i=1}^{K} \sum_{t=1}^{T} r_i^2(t) \left[ \left( \mathcal{K}_{1:T} \right)^{-1} \right]_{i,t} + \left[ \left( \Sigma_\Theta \right)^{-1} \right]_{2,2} + \sum_{i=1}^{K} \sum_{t=1}^{T} r_i^2(t) \left[ \left( \mathcal{K}_{1:T} \right)^{-1} \right]_{i,t}.
$$

(50)

3) The full conditional for $D^{(l)}$ in in (20c) can be expressed as

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

(51)

where we clarify that the Gramm matrix $\mathcal{K}^{(l)}_{1:T}$ is implicitly a function of $D^{(l)} = D^{(l)}$, and the kernel choice given in Section III-A item 8, ensures that this parameter is a scalar random variable. 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:

$$
\frac{d}{dD} \left[ - \frac{1}{2} \log |\mathcal{K}_{1:T}| - \frac{1}{2} \left( \text{Vec} \left[ f^{(l)}_{1:T} \right] - \mu_{1:T} \right)^T (\mathcal{K}_{1:T})^{-1} \left( \text{Vec} \left[ f^{(l)}_{1:T} \right] - \mu_{1:T} \right) + \log p \left( D^{(l)} \right) \right] \| D^{(l)} \|_{(a,b)}
= \left[- \frac{1}{2} \text{Tr} \left( \mathcal{K}^{-1}_{1:T} \frac{d\mathcal{K}_{1:T}}{dD^{(l)}} \right) \right] + \frac{1}{2} \left( \text{Vec} \left[ f^{(l)}_{1:T} \right] - \mu_{1:T} \right)^T (\mathcal{K}_{1:T})^{-1} \frac{d\mathcal{K}_{1:T}}{dD^{(l)}} (\mathcal{K}_{1:T})^{-1} \left( \text{Vec} \left[ f^{(l)}_{1:T} \right] - \mu_{1:T} \right) \| D^{(l)} \|_{(a,b)}
$$

(52)

where this expression is obtained by using standard matrix derivative identities, see [32]. To complete this we derive the

$$
\frac{d\mathcal{K}_{1:T}}{dD^{(l)}} = \frac{1}{D^{(l)}} \Phi_{1:T} \odot \mathcal{K}_{1:T},
$$

(53)

where the $i, j$-th element of $\Phi_{1:T}$ is given by $[\Phi_{1:T}]_{i,j} \triangleq \|R_i^{(l)} - R_j^{(l)}\|^2$. 

\]
Fig. 1: Cooperative wireless relay network with $L$ relays.

Fig. 2: Directed Acyclic Graphical Model structure for the hierarchical Bayesian model developed.

Fig. 3: Construction of Kernel matrix.
Fig. 4: Convergence of hyperparameters

![Graphs showing convergence of hyperparameters for different approaches with perfect and imperfect CSI.]

Fig. 5: ABS relay function

(a) Approach I: perfect CSI  (b) Approach I: imperfect CSI  (c) Approach II: perfect CSI  (d) Approach II: imperfect CSI

![Graphs showing ABS relay function for different approaches with perfect and imperfect CSI.]

<table>
<thead>
<tr>
<th>Relay function</th>
<th>Full Information</th>
<th>Frame-by-Frame</th>
<th>Sliding Window</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>perfect CSI</td>
<td>imperfect CSI</td>
<td>perfect CSI</td>
</tr>
<tr>
<td>SNR</td>
<td>SNR</td>
<td>SNR</td>
<td>SNR</td>
</tr>
<tr>
<td>ABS</td>
<td>0.91</td>
<td>3.97</td>
<td>0.97</td>
</tr>
<tr>
<td>LINEAR</td>
<td>0.57</td>
<td>0.76</td>
<td>0.76</td>
</tr>
<tr>
<td>SIN</td>
<td>0.14</td>
<td>0.41</td>
<td>0.15</td>
</tr>
<tr>
<td>TANH</td>
<td>0.18</td>
<td>0.21</td>
<td>0.19</td>
</tr>
<tr>
<td>DEM</td>
<td>0.79</td>
<td>0.85</td>
<td>0.83</td>
</tr>
</tbody>
</table>

TABLE I: Comparison of absolute error in MAP estimation for relay functions under each estimation approach, for perfect and imperfect CSI.
Fig. 6: LINEAR relay function

Fig. 7: SIN relay function
Fig. 8: TANH relay function

Fig. 9: Demodulate relay function