IJCNN 2007 Presentation

Symmetric Kernel Detector for Multiple-Antenna Aided Beamforming Systems

S. Chen, A. Wolfgang, C.J. Harris and L. Hanzo

School of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ, UK

Outline

- Motivations/overview for incorporating *a priori* knowledge, specifically **symmetry**, in kernel modelling

- Practical example of symmetry: multiple-antenna aided **beamforming** in wireless communication

- Proposed **symmetric kernel classifier** for beamforming detection and **orthogonal forward selection** algorithm based on Fisher ratio of class separability measure

- Simulation results and performance comparison

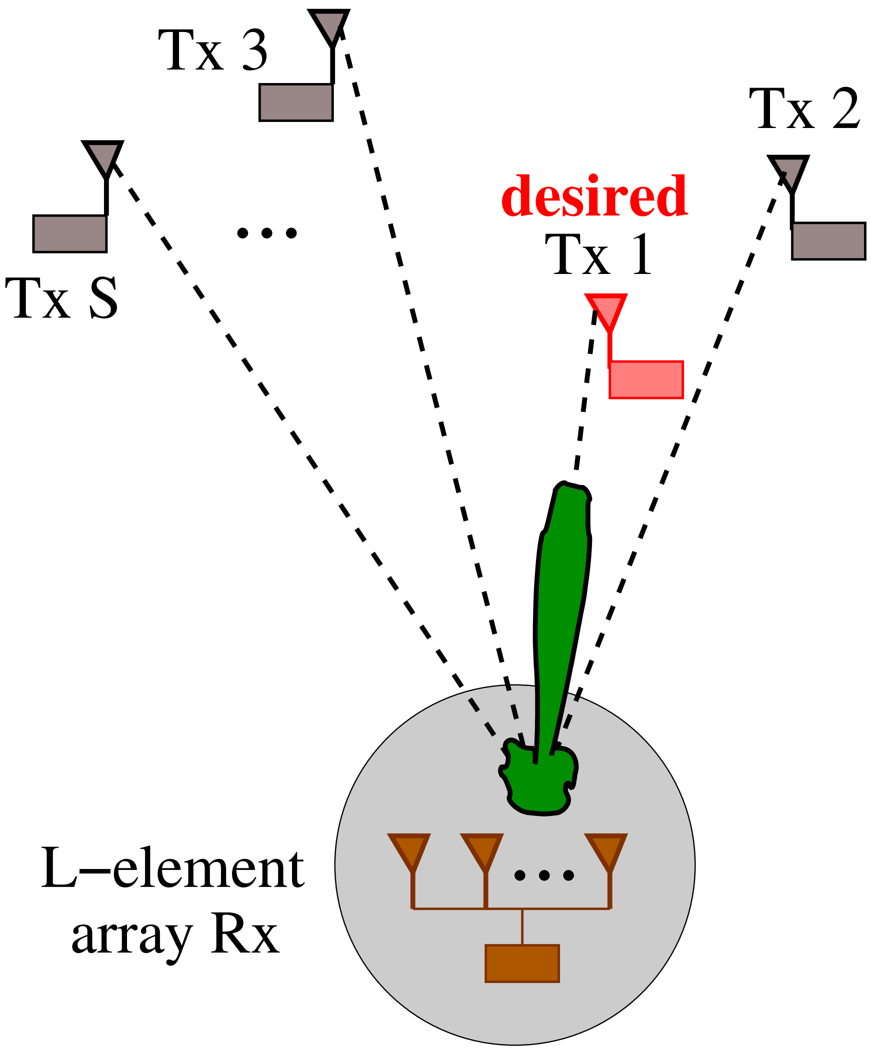
Motivations

- Standard kernel modelling constitutes **black-box** approach
  - Black-box modelling is appropriate, if no *a priori* information exists regarding underlying **data generating mechanism**
- Fundamental principle in data modelling however is to incorporate *a priori* information in **modelling process**
  - Many real-life phenomena exhibit inherent **symmetry**, which are hard to infer accurately from noisy data with black-box models
  - For **regression**, symmetric properties of underlying system have been exploited by imposing symmetry in RBF or **kernel** models
  - This leads to substantial improvements in achievable regression modelling **performance**

Motivations (continue)

- For **classification**, there appears lack of exploiting known properties of underlying system, such as symmetry
- Standard **support vector machine** and other kernel models have been adopted for detection in communication systems
  - Block-box kernel detector requires **more kernels** than number of necessary channel states
  - with notable **performance degradation** compared with **optimal Bayesian detection** solution
- We believe this gap can be bridged if inherent **odd symmetry** of underlying Bayesian solution is "copied" to kernel classifier
  - This motivates our novel **symmetric kernel classifier**

Multiple-Antenna Aided Beamforming



- System supports S users of same carrier with single transmit antenna, and receiver is equipped with a L-element linear **antenna array**
- Traditionally, **beamforming** is defined as **linear** processing, and optimal design for linear beamforming is linear **minimum bit error rate** solution
- If we are willing to extend beamforming process to **nonlinear**, significant performance improvement and larger user capacity can be achieved
- At cost of increased **complexity**

Signal Model

- Received **signal vector** $\mathbf{x}(k) = [x_1(k)\ x_2(k) \cdots x_L(k)]^T$ is expressed as

$$\mathbf{x}(k) = \mathbf{P}\,\mathbf{b}(k) + \mathbf{n}(k) = \bar{\mathbf{x}}(k) + \mathbf{n}(k)$$

- $\mathbf{n}(k) = [n_1(k)\ n_2(k) \cdots n_L(k)]^T$ is complex-valued channel white noise vector with $E[\mathbf{n}(k)\mathbf{n}^H(k)] = 2\sigma_n^2 \mathbf{I}_L$, and **system channel matrix**

$$\mathbf{P} = [A_1 \mathbf{s}_1\ A_2 \mathbf{s}_2 \cdots A_S \mathbf{s}_S]$$

  $\mathbf{s}_i$ is complex-valued **steering vector** of user i, and A_i is i-th complex-valued non-dispersive channel tap

- BPSK users $b_i(k) \in \{-1, +1\}$, $1 \leq i \leq S$, and transmitted symbol vector

$$\mathbf{b}(k) = [b_1(k)\ b_2(k) \cdots b_S(k)]^T$$

  User 1 is **desired** user

Optimal Bayesian Beamforming

- Denote $N_b = 2^S$ **legitimate sequences** of $\mathbf{b}(k)$ as $\mathbf{b}_q$, $1 \le q \le N_b$, and first element of $\mathbf{b}_q$, related to **desired user**, as $b_{q,1}$
- Noiseless **channel state** $\bar{\mathbf{x}}(k)$ takes values from set

$$\bar{\mathbf{x}}(k) \in \mathcal{X} \; = \; \{\bar{\mathbf{x}}_q = \mathbf{P}\,\mathbf{b}_q, 1 \le q \le N_b\}$$

- **Optimal decision** is $\hat{b}_1(k) = \mathrm{sgn}(y_{\mathrm{Bay}}(k))$, with **Bayesian detector**

$$y_{\mathrm{Bay}}(k) \; = \; f_{\mathrm{Bay}}(\mathbf{x}(k)) \; = \; \sum_{q=1}^{N_b} \mathrm{sgn}(b_{q,1})\beta_q e^{-\frac{\|\mathbf{x}(k)-\bar{\mathbf{x}}_q\|^2}{2\sigma_n^2}}$$

- State set can be divided into two **subsets** conditioned on value of $b_1(k)$

$$\mathcal{X}^{(\pm)} \; = \; \{\bar{\mathbf{x}}_i^{(\pm)} \in \mathcal{X}, 1 \le i \le N_{sb} : b_1(k) = \pm 1\}$$

where $N_{sb} = N_b/2 = 2^{S-1}$, and noise power is $2\sigma_n^2$

Symmetry of Bayesian Solution

- Optimal Bayesian beamforming solution has structure of **radial basis function** or **kernel** model with Gaussian kernel function
- Two state subsets are **symmetric**, as

$$\mathcal{X}^{(+)} = -\mathcal{X}^{(-)}$$

- Thus Bayesian detector has **odd symmetry**, as $f_{\text{Bay}}(-\mathbf{x}(k)) = -f_{\text{Bay}}(\mathbf{x}(k))$, and it takes form

$$y_{\text{Bay}}(k) = \sum_{q=1}^{N_{sb}} \beta_q \left(e^{-\frac{\|\mathbf{x}(k)-\bar{\mathbf{x}}_q^{(+)}\|^2}{2\sigma_n^2}} - e^{-\frac{\|\mathbf{x}(k)+\bar{\mathbf{x}}_q^{(+)}\|^2}{2\sigma_n^2}} \right)$$

  since all states are equiprobable, all coefficients β_q are equal
- **Standard** kernel model **does not guarantee** to have this symmetry, particularly when kernel model is trained using noisy data

Symmetric Kernel Classifier

- Consider generic **kernel model**

$$y_{\text{Sker}}(k) = f_{\text{Sker}}(\mathbf{x}(k)) = \sum_{i=1}^{M} \theta_i \phi_i(\mathbf{x}(k))$$

- where M is number of kernels, with novel **symmetric** kernel defined by

$$\phi_i(\mathbf{x}) = \varphi(\mathbf{x}; \mathbf{c}_i, \rho^2) - \varphi(\mathbf{x}; -\mathbf{c}_i, \rho^2)$$

- θ_i are real-valued kernel **weights**, $\mathbf{c}_i$ are complex-valued **centre vectors**, ρ^2 is positive kernel **variance**, and
- $\varphi(\bullet)$ is usual kernel function $\Rightarrow$ in **standard** kernel model, a kernel would simply be $\phi_i(\mathbf{x}) = \varphi(\mathbf{x}; \mathbf{c}_i, \rho^2)$
- Like Bayesian detector, symmetric kernel model has **odd symmetry**

$$f_{\text{Sker}}(-\mathbf{x}(k)) = -f_{\text{Sker}}(\mathbf{x}(k))$$

Training Model

- Given **training data** set $D_K = \{\mathbf{x}(k), d(k) = b_1(k)\}_{k=1}^{K}$, consider every $\mathbf{x}(i)$ as **candidate** kernel centre, i.e. $M = K$, $\mathbf{c}_i = \mathbf{x}(i)$ for $1 \le i \le K$
- By defining **modelling residual** $\varepsilon(i) = d(i) - y_{\text{Sker}}(i)$, kernel model over D_K can be written as

$$\mathbf{d} = \mathbf{\Phi}\,\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

- where $\mathbf{d} = [d(1)\ \ d(2) \cdots d(K)]^T$, $\boldsymbol{\varepsilon} = [\varepsilon(1)\ \ \varepsilon(2) \cdots \varepsilon(K)]^T$, $\boldsymbol{\theta} = [\theta_1\ \theta_2 \cdots \theta_M]^T$, and

$$\mathbf{\Phi} = [\boldsymbol{\phi}_1\ \boldsymbol{\phi}_2 \cdots \boldsymbol{\phi}_M] \in \mathcal{R}^{K \times M}$$

  is **regression matrix** with $\boldsymbol{\phi}_i = [\phi_i(\mathbf{x}(1))\ \phi_i(\mathbf{x}(2)) \cdots \phi_i(\mathbf{x}(K))]^T$
- The task becomes selecting small **subset** of M_{spa} **significant kernels**, where $M_{\text{spa}} \ll M$

Orthogonal Decomposition

- Let an **orthogonal decomposition** of $\mathbf{\Phi}$ be $\mathbf{\Phi} = \mathbf{\Omega}\,\mathbf{A}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & \alpha_{1,2} & \cdots & \alpha_{1,M} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \alpha_{M-1,M} \\ 0 & \cdots & 0 & 1 \end{bmatrix}, \quad \mathbf{\Omega} = \begin{bmatrix} \omega_{1,1} & \omega_{1,2} & \cdots & \omega_{1,M} \\ \omega_{2,1} & \omega_{2,2} & \cdots & \omega_{2,M} \\ \vdots & \vdots & \vdots & \vdots \\ \omega_{K,1} & \omega_{K,2} & \cdots & \omega_{K,M} \end{bmatrix}$$

- **Orthogonal matrix** $\mathbf{\Omega} = [\boldsymbol{\omega}_1\ \boldsymbol{\omega}_2 \cdots \boldsymbol{\omega}_M]$ has orthogonal columns satisfying $\boldsymbol{\omega}_i^T \boldsymbol{\omega}_l = 0$, if $i \neq l$
- Kernel model can alternatively be expressed as

$$\mathbf{d} = \mathbf{\Omega}\,\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$

$\boldsymbol{\gamma} = [\gamma_1\ \gamma_2 \cdots \gamma_M]^T = \mathbf{A}\,\boldsymbol{\theta}$ is weight vector in **orthogonal space** $\mathbf{\Omega}$

Fisher Ratio Class Separability

- Define two class sets $\mathbf{X}_{\pm} = \{\mathbf{x}(k) : d(k) = \pm 1\}$, having points $K_{\pm}$
- **Means** and **variances** of training samples $\mathbf{x}(k) \in \mathbf{X}_{\pm}$ in **direction** of **basis** $\boldsymbol{\omega}_l$ are

$$m_{+,l} = \frac{1}{K_+} \sum_{k=1}^{K} \delta(d(k) - 1)\omega_{k,l}, \ \ \sigma^2_{+,l} = \frac{1}{K_+} \sum_{k=1}^{K} \delta(d(k) - 1)\left(\omega_{k,l} - m_{+,l}\right)^2$$

$$m_{-,l} = \frac{1}{K_-} \sum_{k=1}^{K} \delta(d(k) + 1)\omega_{k,l}, \ \ \sigma^2_{-,l} = \frac{1}{K_-} \sum_{k=1}^{K} \delta(d(k) + 1)\left(\omega_{k,l} - m_{-,l}\right)^2$$

where $\delta(x) = 1$ if $x = 0$ and $\delta(x) = 0$ if $x \neq 0$

- **Fisher ratio** is defined as **ratio** of **interclass difference** and **intraclass spread** encountered in direction of $\boldsymbol{\omega}_l$

$$F_l = \frac{\left(m_{+,l} - m_{-,l}\right)^2}{\sigma^2_{+,l} + \sigma^2_{-,l}}$$

Construction Algorithm

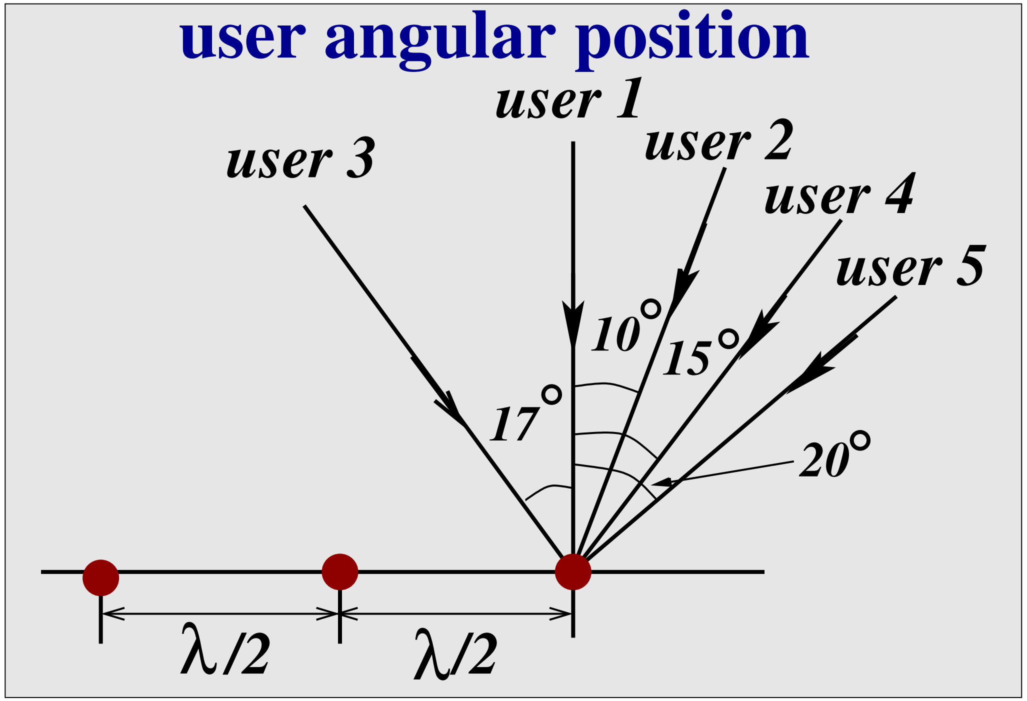
- **Orthogonal forward selection** with **Fisher ratio class separability**
  - At l-th stage, a **candidate** kernel is chosen as l-th kernel in selected model, if it produces largest F_l among remaining candidates
  - Procedure is terminated with a sparse M_{spa}-term model, when

$$\frac{F_{M_{\text{spa}}}}{\sum_{l=1}^{M_{\text{spa}}} F_l} < \xi$$

    where **threshold** ξ determines **sparsity** level of model selected
  - Appropriate value for ξ depends on application concerned, and can be determined empirically
- LS solution for sparse model **weight** vector $\boldsymbol{\theta}_{M_{\text{spa}}} = [\theta_1 \ \theta_2 \cdots \theta_{M_{\text{spa}}}]^T$ is available via $\boldsymbol{\gamma}_{M_{\text{spa}}} = \mathbf{A}_{M_{\text{spa}}} \boldsymbol{\theta}_{M_{\text{spa}}}$, given $\gamma_l = \boldsymbol{\omega}_l^T \mathbf{d} / \boldsymbol{\omega}_l^T \boldsymbol{\omega}_l$, $1 \le l \le M_{\text{spa}}$

Simulation Set Up

- Three-element antenna array with half wavelength spacing supports five BPSK equal-power users
- Simulated channel conditions are $A_i = 1 + j0$, $1 \le i \le 5$

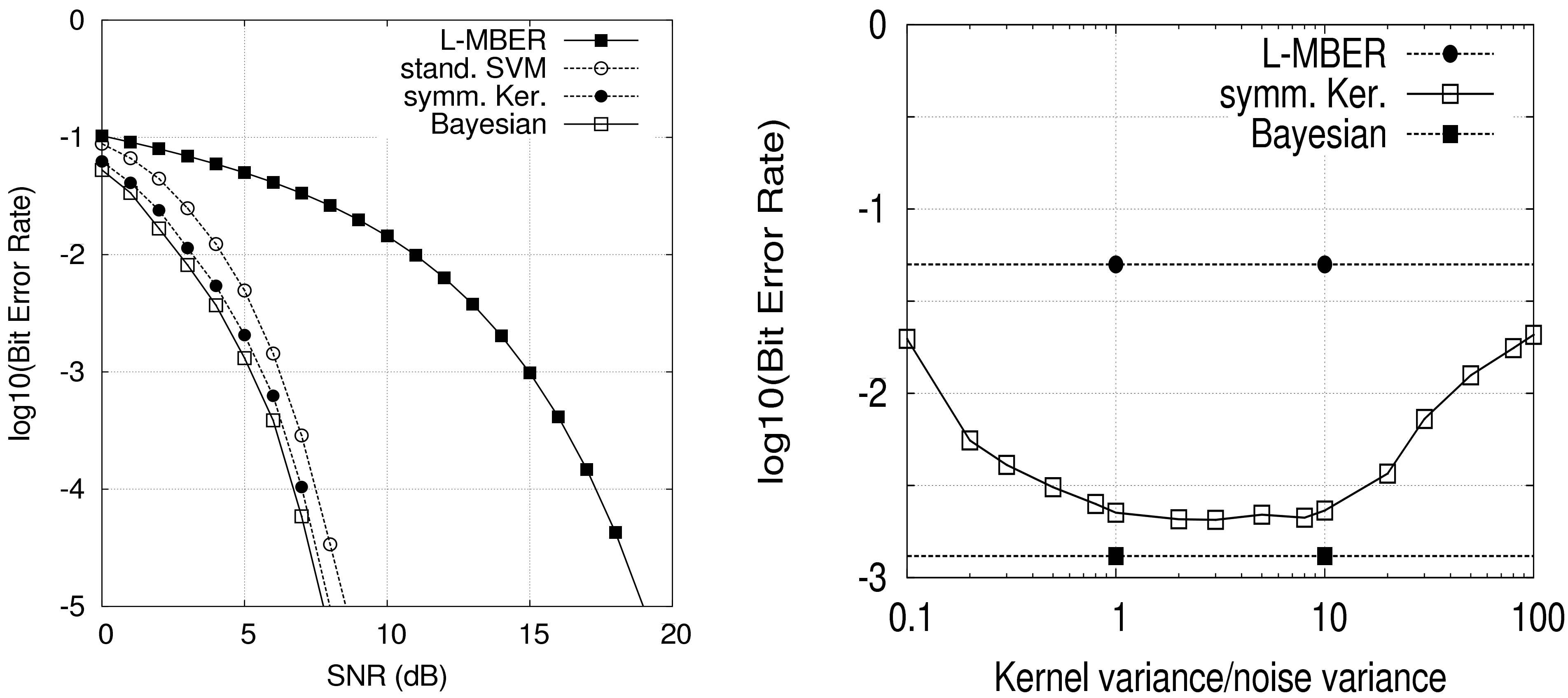


- $K = 600$ training samples are used to construct symmetric kernel classifier
- FRCSM-based OFS is used and kernel variance $\rho^2 = 3\sigma_n^2$
- As $\mathcal{X}^{(+)}$ has 16 states, we terminate kernel classifier construction at $M_{\mathrm{spa}} = 16$

Performance Comparsion

(a) Bit error rate performance comparison, where standard SVM classifier has 40 to 60 support vectors, and (b) Influence of kernel variance, where SNR= 5 dB

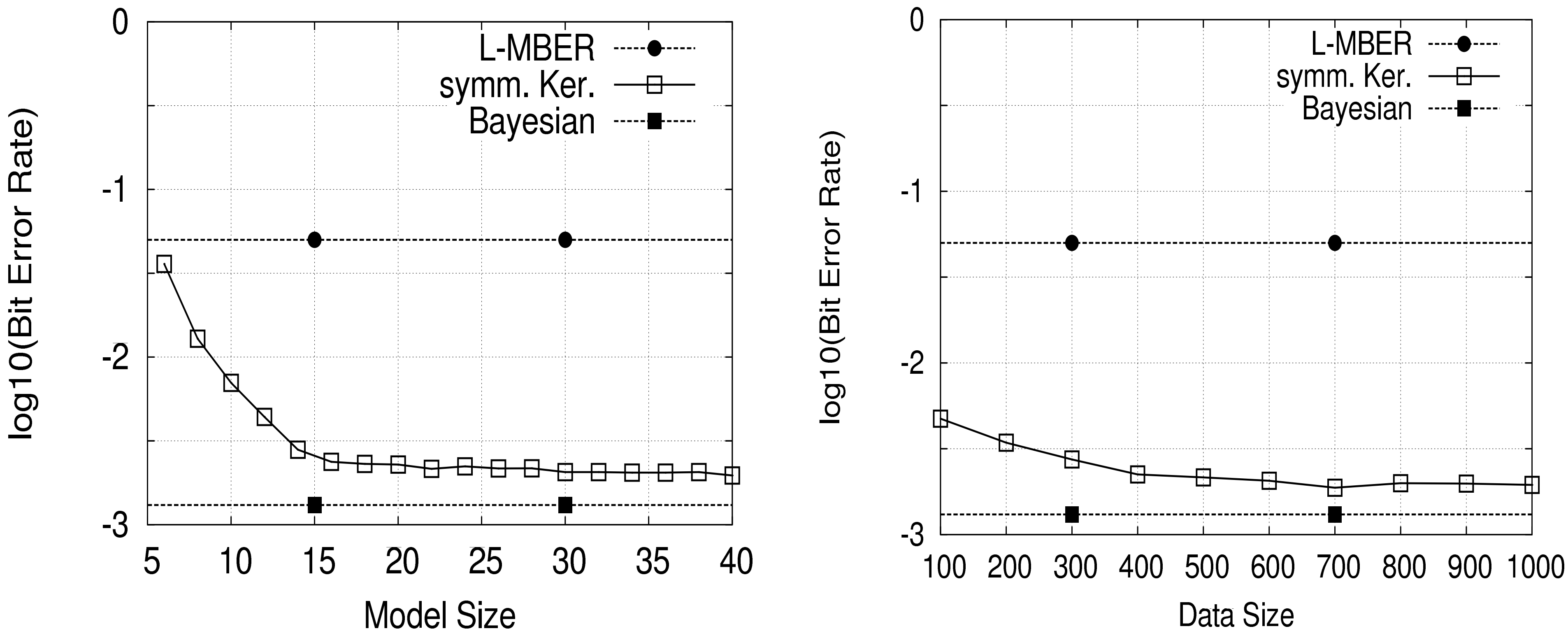
(a) (b)



Algorithm Investigation

(a) Influence of classifier's size, where $K = 600$, ρ^2 is variable depending on M_{spa} and SNR= 5 dB, and (b) Influence of training data length, where $M_{\text{spa}} = 16$, $\rho^2 = 3\sigma_n^2$ and SNR= 5 dB

(a) (b)



Conclusions

- A novel **symmetric kernel classifier** has been proposed for nonlinear beamforming
  - Explicitly exploit underlying **symmetry property** of optimal Bayesian solution
  - **Orthogonal forward selection** based on **Fisher ratio of class seaparability** to determine sparse kernel classifier
  - Substantially **outperform** previous solutions
- Proposed sparse symmetric kernel classifier is **generically applicable** to any problem exhibiting similar symmetry

THANK YOU.

The support of the United Kingdom Royal Academy of Engineering is gratefully acknowledged