WCCI 2008 Presentation

Fully Complex-Valued Radial Basis Function Networks for Orthogonal Least Squares Regression

Sheng Chen[†], Xia Hong[‡] and Chris J. Harris[†]

[†] **School of Electronics and Computer Science**
University of Southampton
Southampton SO17 1BJ, UK

[‡] **School of Systems Engineering**
University of Reading
Reading RG6 6AY, UK

Motivations

- Many practical applications involve multi-dimensional **complex-valued** signals, which lead to development of complex-valued neural networks

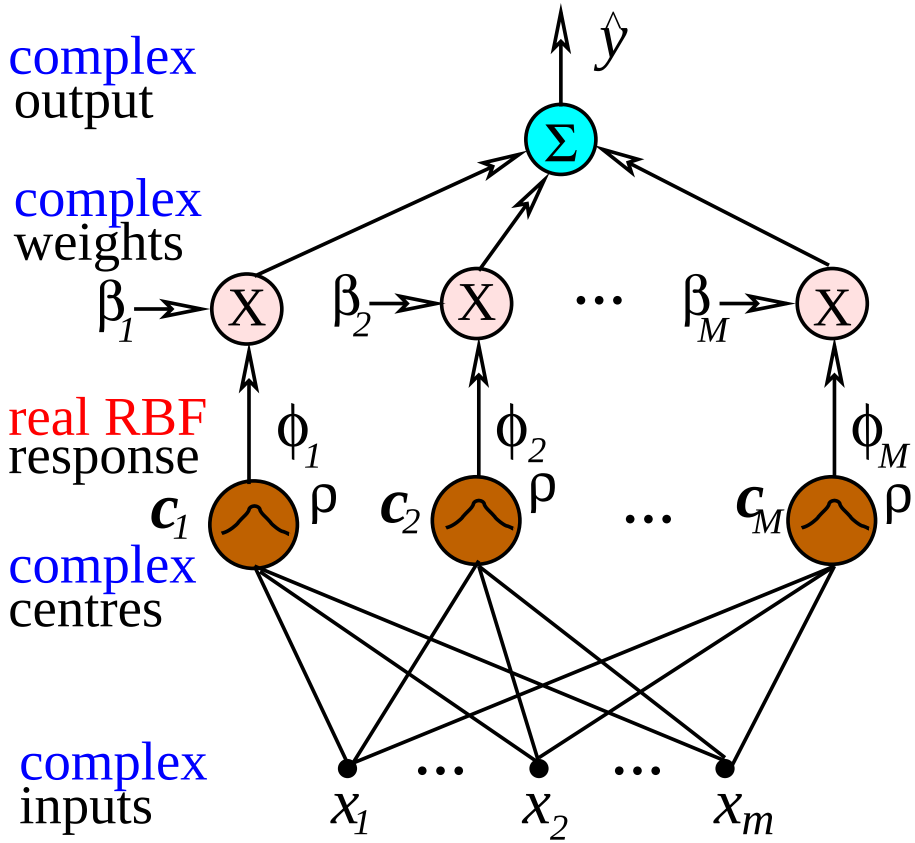
- Chen *et al.* (1994) developed a complex-valued **RBF** network

- **Real-valued** RBF response

$$\phi(\|\mathbf{x} - \mathbf{c}\|/\rho)$$

  can be interpreted as conditional **probability density function**

- Complex-valued RBF network with complex-valued RBF response is of theoretical and practical interests



Complex-Valued RBF Network

- Need to develop theoretic oriented complex-valued RBF node **response function**, but for practical purpose, we will use

$$\phi_i(\mathbf{x}) = \varphi(\|\Re[\mathbf{x}] - \Re[\mathbf{c}_i]\|/\rho) + j\varphi(\|\Im[\mathbf{x}] - \Im[\mathbf{c}_i]\|/\rho)$$

- $\Re[\bullet]$ and $\Im[\bullet]$ denote **real** and **imaginary** parts, $j = \sqrt{-1}$, $\mathbf{c}_i \in \mathcal{C}^m$ ith complex-valued RBF **centre**, and $\rho^2 > 0$ RBF variance

- Two choices for real-valued **basis function** $\varphi(\bullet)$

$$\varphi(\chi/1) = \chi^2 \log(\chi) \quad \text{and} \quad \varphi(\chi/\rho) = e^{-\chi^2/\rho^2}$$

- Almost all learning methods for real-valued RBF networks can be extended to complex-valued case
  - This presentation is for **regression** application
  - Another presentation in this session will consider **classification**

Complex-Valued RBF Regression

- Given **training** set $D_N = \{\mathbf{x}(k) \in \mathcal{C}^m, y(k) \in \mathcal{C}\}_{k=1}^N$, construct **complex-valued** RBF network

$$\hat{y}(k) = \sum_{i=1}^{M} \theta_i \phi_i(\mathbf{x}(k))$$

with modelling error $e(k) = y(k) - \hat{y}(k)$

- Given RBF variance ρ^2, use every $\mathbf{x}(k)$ as candidate RBF **centre**, i.e. $M = N \Rightarrow$ regression model over D_N

$$\mathbf{y} = \mathbf{\Phi}\,\boldsymbol{\theta} + \mathbf{e}$$

where $\mathbf{y} = [y(1) \cdots y(N)]^T$, $\mathbf{e} = [e(1) \cdots e(N)]^T$, RBF **weight** vector $\boldsymbol{\theta} = [\theta_1 \cdots \theta_M]^T$, complex-valued **regression** matrix $\mathbf{\Phi} = [\boldsymbol{\phi}_1\ \boldsymbol{\phi}_2 \cdots \boldsymbol{\phi}_M]$ with columns $\boldsymbol{\phi}_i = [\phi_i(\mathbf{x}(1))\ \phi_i(\mathbf{x}(2)) \cdots \phi_i(\mathbf{x}(N))]^T$

Orthogonal Decomposition

- **Orthogonal decomposition** $\mathbf{\Phi} = \mathbf{W}\mathbf{A}$, with orthogonal matrix

$$\mathbf{W} = [\mathbf{w}_1 \ \mathbf{w}_2 \cdots \mathbf{w}_M]$$

and **upper triangular** **complex-valued** matrix

$$\mathbf{A} = \begin{bmatrix} 1 & a_{1,2} & \cdots & a_{1,M} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{M-1,M} \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

- **Regression model** can alternatively be expressed as

$$\mathbf{y} = \mathbf{W}\mathbf{g} + \mathbf{e}$$

where new weight vector $\mathbf{g} = [g_1 \ g_2 \cdots g_M]^T$ satisfies $\mathbf{A}\boldsymbol{\theta} = \mathbf{g}$

Locally Regularised OLS algorithm

- **Regularised least square** criterion

$$J_R(\mathbf{g},\boldsymbol{\lambda}) = \mathbf{e}^H\mathbf{e} + \mathbf{g}^H\boldsymbol{\Lambda}\mathbf{g} = \mathbf{y}^H\mathbf{y} - \sum_{i=1}^{M}\left(\mathbf{w}_i^H\mathbf{w}_i + \lambda_i\right)|g_i|^2$$

  where $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_M\}$, and λ_i regularisation parameters

- OLS forward selection based on regularised **error reduction** ratio

$$[\mathrm{rerr}]_i = \left(\mathbf{w}_i^H\mathbf{w}_i + \lambda_i\right)|g_i|^2/\mathbf{y}^H\mathbf{y}$$

- **Evidence procedure** for updating regularisation parameters

$$\lambda_i^{\mathrm{new}} = \frac{\gamma_i^{\mathrm{old}}}{N - \gamma^{\mathrm{old}}}\frac{\mathbf{e}^H\mathbf{e}}{|g_i|^2},\ \ 1 \le i \le M$$

$$\gamma_i = \frac{\mathbf{w}_i^H\mathbf{w}_i}{\lambda_i + \mathbf{w}_i^H\mathbf{w}_i} \quad \text{and} \quad \gamma = \sum_{i=1}^{M}\gamma_i$$

D-optimality Experimental Design

- **Covariance** of LS estimate proportional to inverse of design matrix

$$Cov[\hat{\boldsymbol{\theta}}] \propto (\boldsymbol{\Phi}^H \boldsymbol{\Phi})^{-1}$$

- ***D*-optimality** selects subset model $\boldsymbol{\Phi}_{n_s}$ that maximises $\det(\boldsymbol{\Phi}_{n_s}^H \boldsymbol{\Phi}_{n_s})$

  Prevent selection of oversized **ill-posed** model and problem of high **estimate variances**

- Maximising $\det(\boldsymbol{\Phi}_{n_s}^H \boldsymbol{\Phi}_{n_s})$ identical to maximising $\det(\mathbf{W}_{n_s}^H \mathbf{W}_{n_s})$

$$\det(\boldsymbol{\Phi}^H \boldsymbol{\Phi}) = \det(\mathbf{W}^H \mathbf{W}) = \prod_{i=1}^{M} \mathbf{w}_i^H \mathbf{w}_i$$

  or equivalently to minimising $-\log \det(\mathbf{W}_{n_s}^H \mathbf{W}_{n_s})$

$$-\log\left(\det(\mathbf{W}^H \mathbf{W})\right) = \sum_{i=1}^{M} -\log(\mathbf{w}_i^H \mathbf{w}_i)$$

Combined LROLS and D-Optimality

- Combined LROLS and D-optimality algorithm adopts combined criterion

$$J_{RD}(\mathbf{g}, \boldsymbol{\lambda}, \beta) = J_R(\mathbf{g}, \boldsymbol{\lambda}) + \beta \sum_{i=1}^{M} - \log(\mathbf{w}_i^H \mathbf{w}_i)$$

- Selection based on combined regularised error reduction ratio

$$[\text{crerr}]_i = \left((\mathbf{w}_i^H \mathbf{w}_i + \lambda_i)|g_i|^2 + \beta \log(\mathbf{w}_i^H \mathbf{w}_i) \right) / \mathbf{y}^H \mathbf{y}$$

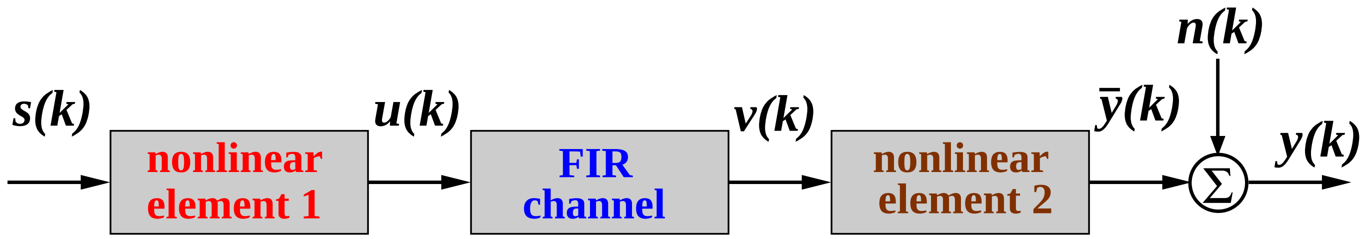
- There always exists an **optimal subset model** size n_s, such that

$$[\text{crerr}]_l \leq 0 \quad \text{for} \quad n_s + 1 \leq l \leq M$$

- Selection procedure **automatically** terminates with an n_s-term model

  very **sparse** model with excellent **generalisation** capability

A Modelling Example



- Transmitted data symbols $s(k) = s_R(k) + js_I(k)$ are 16-QAM
- 1st nonlinear element: transmitter nonlinear high power amplifier

$$u(k) = f_{\text{amp}}(s(k)) = \frac{2s(k)}{1 + |s(k)|^2} e^{j\frac{\pi}{3}\frac{|s(k)|^2}{1+|s(k)|^2}}$$

- FIR linear channel with transfer function

$$V(z)/U(z) = (0.3725{+}j0.2172)\left(1 - (0.35 + j0.7)z^{-1}\right)\left(1 - (0.5 + j)z^{-1}\right)$$

- 2nd nonlinear element: third-order complex-valued Volterra nonlinearity

$$\bar{y}(k) = f_{\text{Vol}}(v(k)) = v(k) + 0.2v^2(k) - 0.1v^3(k)$$

Underlying System

- Let $f(\bullet)$ be complex-valued mapping specified this nonlinear channel

$$y(k) = \bar{y}(k) + n(k) = f(\mathbf{x}(k)) + n(k)$$

$\mathbf{x}(k) = [s(k)\ s(k-1)\ s(k-2)]^T$ has $N_{\text{st}} = 16^3 = 4096$ states

$$\mathcal{X} = \{\bar{\mathbf{x}}_l,\ 1 \le l \le N_{\text{st}}\}$$

Noise-free channel output $\bar{y}(k)$ also has N_{st} values

$$\bar{\mathcal{Y}} = \{\bar{y}_l = f(\bar{\mathbf{x}}_l),\ 1 \le l \le N_{\text{st}}\}$$

- Identified model $\hat{y}(k) = \hat{f}(\mathbf{x}(k))$ over $\mathcal{X}$ also have N_{st} values

$$\hat{\mathcal{Y}} = \{\hat{y}_l = \hat{f}(\bar{\mathbf{x}}_l),\ 1 \le l \le N_{\text{st}}\}$$

Mean state error is defined as

$$\text{Mean State Errror} = \frac{1}{2N_{\text{st}}} \sum_{l=1}^{N_{\text{st}}} |\bar{y}_l - \hat{y}_l|^2$$

Identification Results

- Symbol power is scaled to 1.0, noise power 0.1, 600 training samples and 600 test samples
- Mean square error is defined by

$$\text{MSE} = \frac{1}{2N} \sum_{k=1}^{N} |y(k) - \hat{y}(k)|^2$$

- For thin-plate-spline basis function, appropriate weighting is $\beta = 10.0$
- For Gaussian basis function, appropriate weighting is $\beta = 10^{-6}$
- Results obtained using combined LROLS and D-optimality algorithm

basis function	ρ^2	n_s	training MSE	testing MSE	mean state error
Gaussian	3.0	50	0.128931	0.142484	0.035443
thin-plate-spline	NA	57	0.117874	0.146306	0.038081

Conclusions

- A **fully** complex-valued radial basis function network has been proposed
- Both RBF weights and RBF nodes' response are **complex-valued**
- Almost all **learning algorithms** for real-valued RBF network can be extended to this complex-valued RBF network
- **Regression** application is demonstrated using combined locally **regularised OLS** and ***D*-optimality** algorithm
- Effectiveness of proposed algorithm is tested by complex-valued nonlinear channel identification