Justification for Employing Support Vector Detector in Large-Scale MIMO Systems

Tianpei Chen

Department of Electrical and Computer Engineering

McGill University, Montreal, Quebec, Canada

November 27, 2015

Abstract

This report provides the justifications for employing support vector regression (SVR) in the detection problem of Large-Scale MIMO (LS-MIMO) systems, which have medium or full loading factor (number of transmit antennas is close or equal to the number of receive antennas). Under this condition, LS-MIMO detection in the view of statistical learning is a multidimensional linear regression problem, given a small training data set. We compare linear detectors (LD) (zero forcing (ZF) and minimum mean square error (MMSE)) with SVR in view of statistical learning theory and demonstrate the better performance that SVR possess over LD in the high spectrum efficiency LS-MIMO systems with medium and full loading factor.

Index Terms

Large-Scale MIMO, full loading, statistical learning theory, support vector regression, linear detectors

I. Large-Scale MIMO System Model and Equivalent Regression Estimation Problem

In this section, we present the system model of Large-Scale MIMO (LS-MIMO) systems and the equivalent regression estimation model in view of statistical learning.

A. System Model

Consider a uncoded complex LS-MIMO uplink spatial multiplexing (SM) system, uplink means the data streams are transmitted from users to base stations (BS). Let N_t denotes the number of transmit antennas and N_r denotes the number of receive antennas, $N_r \geq N_t$, define loading factor $\alpha = \frac{N_t}{N_r}$. Typically LS-MIMO systems have hundreds of receive antennas at BS.

Independent bit sequences, which are modulated to complex symbols are transmitted over flat and slow fading channel. The discrete time model of LS-MIMO systems in the complex domain is given by

$$\mathbf{y}_c = \mathbf{H}_c \mathbf{s}_c + \mathbf{n}_c, \tag{1}$$

where $\mathbf{y}_c \in \mathbb{C}^{N_r \times 1}$ denotes the receive symbol vector, $\mathbf{H}_c \in \mathbb{C}^{N_r \times N_t}$ denotes the Rayleigh fading propagation channel matrix. Each component of $\mathbf{H}_c$ is independent identically distributed (i.i.d), circularly symmetric complex Gaussian (CSCG) random variables with zero mean and unit variance. $\mathbf{s}_c \in \mathbb{C}^{N_t \times 1}$ denotes the transmit symbol vector. Each transmit symbol is equally probable and mutually independent, taken from a finite signal constellation alphabet $\mathbb{O}$ (4-QAM, 16-QAM and 64-QAM), $|\mathbb{O}| = M$. $\mathbf{s}_c \in \mathbb{O}^{N_t}$, satisfies $\mathbb{E}(\mathbf{s}_c\mathbf{s}_c^H) = \xi_s \mathbf{I}_{N_t}$, where $\mathbb{E}(\cdot)$ denotes expectation operator, $(\cdot)^H$ denotes Hermitian transpose, ξ_s denotes the average power of each component in $\mathbf{s}_c$, $\mathbf{I}_{N_t} \in \mathbb{C}^{N_t \times N_t}$ is an identity matrix. $\mathbf{n}_c \in \mathbb{C}^{N_t \times 1}$ denotes i.i.d additive white Gaussian noise with zeros mean and satisfies $\mathbb{E}(\mathbf{n}_c\mathbf{n}_c^H) = N_0 \mathbf{I}_{N_r}$. N_0 denotes the power spectrum density of each component in $\mathbf{n}_c$. The signal to noise ratio (SNR) is defined as $\frac{\xi_s}{N_0}$.

The discrete time model in (1) is defined in complex domain, one can transfer this model into an equivalent real-valued model. Let $\Re(\cdot)$ denotes the real part and $\Im(\cdot)$ denotes the imaginary part, $\mathbf{y}_c, \mathbf{H}_c, \mathbf{s}_c$ and $\mathbf{n}_c$ can be decomposed by

$$\begin{aligned} &\mathbf{y}_c = \Re(\mathbf{y}_c) + \jmath\Im(\mathbf{y}_c), \mathbf{H}_c = \Re(\mathbf{H}_c) + \jmath\Im(\mathbf{H}_c), \\ &\mathbf{s}_c = \Re(\mathbf{s}_c) + \jmath\Im(\mathbf{s}_c), \mathbf{n}_c = \Re(\mathbf{n}_c) + \jmath\Im(\mathbf{n}_c), \end{aligned} \tag{2}$$

define

$$\begin{aligned} &\mathbf{y} = [\Re(\mathbf{y}_c)^T, \Im(\mathbf{y}_c)^T]^T, \mathbf{H} = \begin{bmatrix} \Re(\mathbf{H}_c) & -\Im(\mathbf{H}_c) \\ \Im(\mathbf{H}_c) & \Re(\mathbf{H}_c) \end{bmatrix} \\ &\mathbf{s} = [\Re(\mathbf{s}_c)^T, \Im(\mathbf{s}_c)^T]^T, \mathbf{n} = [\Re(\mathbf{n}_c)^T, \Im(\mathbf{n}_c)^T]^T, \end{aligned} \tag{3}$$

The complex model in (1) can be transferred into an equivalent real-valued model, which is given by

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{4}$$

B. MIMO Detection in view of Statistical Learning

Generally speaking, the task of LS-MIMO detection is to estimate the transmit symbol vector $\mathbf{s}$ based on the knowledge of receive symbol vector $\mathbf{y}$ and channel state information (CSI) $\mathbf{H}$. In view of statistical learning, LS-MIMO detection problem is a multidimensional (N_t) regression estimation problem with N_r training data samples, which is given by

$$\mathbf{y}_i = f_{LS-MIMO}(\mathbf{h}_i) + \mathbf{n}_i, i = 1, 2, \ldots, N_r, \tag{5}$$

where $\mathbf{h}_i$ denotes the ith row of $\mathbf{H}$, is the ith input data vector, $f_{LS-MIMO}(\mathbf{h}_i) = \mathbf{h}_i\mathbf{s}$, is a N_t-dimensional linear function, $\mathbf{s}$ is the regression coefficient vector, $\mathbf{y}_i$ is the ith observation disrupted by additive noise (error) $\mathbf{n}_i$.

The loss function $L(\mathbf{y}_i, f_c(\mathbf{h}_i))$ is a measure of the loss between the observations and the output of a candidate function $f_c\cdot$, the actual risk of a regression estimate $f_c(\cdot)$ is defined by [1]

$$R(f_c) = \int L(y, f_c(h))dP(y, \mathbf{h}), f_c \in \Lambda \tag{6}$$

where Λ is the set of candidate functions, $P(y, \mathbf{h})$ is the joint probability density function (pdf) of input-output data set $(y, \mathbf{h})$ where y is a real value and $\mathbf{h}$ is a vector, (6) is the expectation of loss function. The goal of regression estimation is to find the best function $f_0(\cdot)$ (e.g., the most accurate approximation to the regression coefficient vector $\mathbf{s}$ in (5)) that minimize the actual risk, based on the training data set.

The loss function for regression estimation, based on Huber's robust estimator principle [2] is given by

$$L(y, f_c(\mathbf{h})) = -\ln(Pr(y - f_c(\mathbf{h}))), \tag{7}$$

where $Pr(\cdot)$ denotes the pdf of additive noise, when the additive noise is Gaussian distributed, the robust loss function is given by

$$L(y, f_c(h)) = (y - f_c(h))^2, \tag{8}$$

II. Empirical Risk Minimization Principle and Its Correlation with Linear Detectors

The actual risk is a theoretical measure of the regression estimation performance with respect to all the data set $(y, \mathbf{h})$. However, in practical the training data set is a finite set of the samples in the data set space and the probability measure $P(y, \mathbf{h})$ is unknown, thus empirical risk is used as an alternative measure of the performance of regression estimation performance. Given a training data set $(y_1, \mathbf{h}_1), (y_2, \mathbf{h}_2), \ldots, (y_l, \mathbf{h}_l)$, where l is the number of training data, empirical risk is given by

$$R_{emp}(f_c) = \frac{1}{l}\sum_{i=1}^{l} L(y_i, f_c(\mathbf{h}_i)), f_c \in \Lambda \tag{9}$$

For regression estimation problems in LS-MIMO detection, use Huber's robust loss function in (8), the empirical risk in LS-MIMO detection is given by

$$R_{emp}^{LS-MIMO}(f_c) = \frac{1}{N_r}\sum_{i=1}^{N_r}(\mathbf{y}_i - f_c(\mathbf{h}_i))^2, \tag{10}$$

where $f_c(\mathbf{h}_i) = \mathbf{h}_i\hat{\mathbf{s}}$, $\hat{\mathbf{s}}$ is the estimation of the regression coefficient vector. The empirical risk minimization principle (ERM) is to approximate the minimizer f_0 of actual risk $R(f_c)$ by the minimizer f_{emp} of the corresponding empirical risk $R_{emp}(f_c)$.

In MIMO detection, linear detectors (LD) (zeros forcing (ZF) and minimum mean square error (MMSE)) from the view of regression estimation, are the estimators designed based on ERM principle. We elaborate a little further here.

ZF is derived from the principle of solving an unconstrained least square problem in a continuous vector space [3], then slicing the results by symbol constellation alphabet (e.g., 4-QAM, 16-QAM or 64-QAM), the unconstrained least square problem is given by

$$\min_{\hat{\mathbf{s}}} ||\mathbf{y} - \mathbf{H}\hat{\mathbf{s}}||^2 = \min_{\hat{\mathbf{s}}} \sum_{i=1}^{N_r} (\mathbf{y}_i - \mathbf{h}_i\hat{\mathbf{s}})^2, \tag{11}$$

where $||\cdot||$ denotes 2-norm operator. The solution of (11) is given by

$$\tilde{\mathbf{s}} = \mathbf{H}^{\dagger}\mathbf{y}, \tag{12}$$

where $\mathbf{H}^{\dagger} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$ is the pseudo-inverse of $\mathbf{H}$.

MMSE detector is designed based on the principle that minimizaing the mean square error between the transmit symbol vector and the estimated symbol vector, that is [4]

$$\min_{\hat{\mathbf{s}}} \mathbb{E}(||\mathbf{s} - \hat{\mathbf{s}}||^2), \tag{13}$$

where $\hat{\mathbf{s}} = \mathbf{G}\mathbf{y}$, $\mathbf{G}$ is the equalization matrix, given by $\mathbf{G} = (\mathbf{H}^T\mathbf{H} + \rho^{-1}\mathbf{I})^{-1}\mathbf{H}^T$, $\rho = \frac{\xi_s}{N_0}$ denotes the SNR. Although ZF and MMSE are based on the different principles, the latter one can also be interpreted as solving the similar unconstrained least square problem as in (11) for an augmented linear equation system. The augmented linear equation system takes into account the noise.

$$\begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{H} \\ \rho^{-1/2}\mathbf{I}_{\mathbf{N_t}} \end{bmatrix} \mathbf{s} + \begin{bmatrix} \mathbf{n} \\ -\rho^{-1/2}\mathbf{I}_{\mathbf{N_t}}\mathbf{s} \end{bmatrix}$$

define $\bar{\mathbf{y}} = [\mathbf{y}^T, 0]^T$ and $\bar{\mathbf{H}} = [\mathbf{H}^T, \sqrt{N_0}\mathbf{I}_{N_t}]^T$, $\mathbf{I}_{N_t}$ denotes $N_t \times N_t$ identity matrix, $\bar{\mathbf{h}}_i$ denotes the ith row of $\bar{\mathbf{H}}$. The equivalent unconstrained least square problem is given by

$$\min_{\hat{\mathbf{s}}} ||\bar{\mathbf{y}} - \bar{\mathbf{H}}\hat{\mathbf{s}}||^2 = \min_{\hat{\mathbf{s}}} \sum_{i=1}^{N_r+N_t} (\bar{\mathbf{y}}_i - \bar{\mathbf{h}}_i\hat{\mathbf{s}})^2, \tag{14}$$

Similar to (12), the solution of (14) is given by [5] [6]

$$\tilde{\mathbf{s}} = \bar{\mathbf{H}}^{\dagger}\bar{\mathbf{y}} = \mathbf{G}\mathbf{y}, \tag{15}$$

From (10), (11) and (14), we can conclude ZF and MMSE are based on ERM principle. In view of statistical learning, when the VC-dimension of the regression function is finite [1], with the number of training data (in LS-MIMO it refers to N_r) increasing, empirical risk converges to actual risk rapidly, independent to the probability measure ($P(y, \mathbf{h})$) [1]. The asymptotical consistence of empirical risk is achieved when the number of training data is large, this explains why the ERM based estimators (ZF and MMSE) can achieve near optimal performance in the LS-MIMO with low loading factor ($\alpha = \frac{N_t}{N_r} \ll 1$, $N_t \ll N_r$, the diversity gain of ZF/MMSE is $N_r - N_t + 1$ [4]). This demonstrates the effectiveness of LDs (ZF/MMSE) for low loading LS-MIMO systems. However, in medium and full loading LS-MIMO, the diversity of ZF/MMSE is close to or equal to 1, this can be explained by the fact that ERM principle is not effective when the size of the training data set is small.

III. Structural Risk Minimization: why support vector regression

In this section, we present the explanation why support vector regression (SVR) are promising in the medium and full loading LS-MIMO. Based on the theoretical justifications of SVR based on the structural risk minimization (SRM) principle, we demonstrate the fitness of SVR based detector to be used as a preliminary estimator that can replace the traditional ZF/MMSE detectors.

Given the situation that ZF/MMSE have inferior performance in full loading LS-MIMO systems because of the asymptotical consistence of ERM principle. We consider another principle in statistical learning theory that takes into account the small training data set size. Based on statistical learning theory, the actual risk is upper bounded in probability by [1]

$$R(f_c) \leq R_{emp}(f_c) + \frac{B\epsilon}{2}(1 + \sqrt{1 + \frac{4R_{emp}(f_c)}{B\epsilon}}), \tag{16}$$

if the loss function is bounded by

$$0 \leq L(y, f_c(\mathbf{h})) \leq B \quad f_c \in \Lambda, B > 0 \tag{17}$$

ϵ in (16) is

$$\epsilon = 4\frac{m(\ln(\frac{2l}{m}) + 1)}{l} \tag{18}$$

where l is the number of training data and m is the VC-dimension of candidate function f_c.

The second summand of the right hand side of (16) is called confidence interval, which is determined by the complexity of the candidate function (VC-dimension) f_c [7]. From (16) we can conclude that

1. If $\frac{l}{m}$ is large (typically that is $\frac{l}{m} \geq 20$), the confidence interval in (16) tends to be vanished, therefore a small empirical risk can guarantee a small actual risk, ERM principle is effective.

2. If $\frac{l}{m}$ is small, a small empirical risk can not guarantee a small actual risk.

In the medium or full loading LS-MIMO, where the number of training data N_r is not very large comparing to the dimensionality of the regression function N_t, structural risk minimization (SRM) principle, which minimize the two terms in the right hand side of (16) simultaneously is more effective to find a solution that guarantee a small actual risk. SRM principle aims to find a simplest function (small confidence interval) that minimize the training error (empirical risk).

Based on the constructive SRM principle, SVR is designed by making a trade off between minimizing the complexity of function (regularization) and minimizing training error (risk functional). As to the full loading LS-MIMO systems, SVR based detector can guarantee a better regression performance (lower error probability of detection) than ZF/MMSE LDs in a feasible SNR region. Furthermore, By exploiting ”channel hardening” phenomenon [8], a fast training algorithm can be developed in dual optimization process of SVR.

IV. Conclusion

In conclusion, SVR is a suitable algorithmic framework for designing the preliminary detector in full loading LS-MIMO systems, which can replace traditional LD (ZF/MMSE) in the preprocessing stage of some searching algorithms such as likelihood ascend searching (LAS) [9] [10] and genetic algorithm (GA) [11] [12]. SVR based detectors have the following advantages:

1. Better performance in the feasible SNR region (more accurate estimate of the regression coefficient vector $\mathbf{s}$ in (5)).

2. Lower computational complexity by exploiting the characteristic of high dimensional random channel matrix $\mathbf{H}$.

3. Comparing with MMSE, SVR based detector do not require the information of noise variance.

References

[1] V. N. Vapnik and V. Vapnik, *Statistical learning theory*. Wiley New York, 1998, vol. 1.

[2] P. J. Huber, *Robust statistics*. Springer, 2011.

[3] T. Kailath, H. Vikalo, and B. Hassibi, “MIMO receive algorithms,” *Space-Time Wireless Systems: From Array Processing to MIMO Communications*, 2005.

[4] A. Paulraj, R. Nabar, and D. Gore, *Introduction to space-time wireless communications*. Cambridge university press, 2003.

[5] D. Gore, R. W. Heath Jr, and A. Paulraj, “On performance of the zero forcing receiver in presence of transmit correlation,” in *Information Theory, 2002. Proceedings. 2002 IEEE International Symposium on*. IEEE, 2002, p. 159.

[6] D. Wübben, R. Böhnke, V. Kühn, and K.-D. Kammeyer, “MMSE extension of V-BLAST based on sorted QR decomposition,” in *Vehicular technology conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, vol. 1. IEEE, 2003, pp. 508–512.

[7] V. N. Vapnik, “An overview of statistical learning theory,” *Neural Networks, IEEE Transactions on*, vol. 10, no. 5, pp. 988–999, 1999.

[8] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, “Multiple-antenna channel hardening and its implications for rate feedback and scheduling,” *Information Theory, IEEE Transactions on*, vol. 50, no. 9, pp. 1893–1909, 2004.

[9] K. V. Vardhan, S. K. Mohammed, A. Chockalingam, and B. S. Rajan, “A Low-Complexity Detector for large MIMO systems and multicarrier CDMA systems,” *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 3, pp. 473–485, 2008.

[10] P. Li and R. D. Murch, “Multiple output selection-LAS algorithm in large MIMO systems,” *Communications Letters, IEEE*, vol. 14, no. 5, pp. 399–401, 2010.

[11] M. J. Juntti, T. Schlösser, and J. O. Lilleberg, “Genetic algorithms for multiuser detection in synchronous CDMA,” in *Information Theory. 1997. Proceedings., 1997 IEEE International Symposium on*. IEEE, 1997, p. 492.

[12] C. Ergün and K. Hacioglu, “Multiuser detection using a genetic algorithm in CDMA communications systems,” *Communications, IEEE Transactions on*, vol. 48, no. 8, pp. 1374–1383, 2000.