Report

Tianpei Chen
Department of Electrical and Computer Engineering
McGill University

August 11, 2015

I. Introduction

Decoder is one of the key components of Multiple-Input Multiple-Output (MIMO) systems. Designing of high performance and low complexity detector has become a bottleneck of Large MIMO systems.

Firmly grounded in framework of statistical learning theory, Support Vector Machine (SVM) is proposed in 1960s [ref vapnik], and of immense research and industry interest since 1990s. SVM is a powerful tool for supervised learning tasks such as classification, regression and prediction. Moreover, the kernel trick [ref learning with kernels SVM regularization] makes it possible to map data samples into higher dimensional feature space. Therefore SVM can deal with non-linear learning tasks. This makes SVM become a promising tool for complex real-world problems. Based on the similar principle, ϵ-Support Vector Regression (epsilon-SVR) [vapnik 1995, smola 2003], is developed.

Like SVM, epsilon-SVR first change primal objective function into dual optimization task, then solving the dual quadratic optimization problem. Typically this kind of problem can be solved by numerical quadratic optimization (QP) methods, however, they are computational costly. Decomposition methods, denotes a set of algorithms that divide the optimization variables (Lagrange multipliers) into two sets W and N, W is the work set and N contains the rest optimization variables. In each iteration, only work set is updated for optimization while the other variables are fixed. Sequential Minimal Optimization (SMO) [ref A fast algorithm sequential minimal optimization] is an extreme case of decomposition methods which chooses dual Lagrange multiplier to optimize in each iteration. In each iteration, decomposition method can find an analytic optimal solution for work set, which makes the solver works much more faster than numerical QP algorithms. Decomposition methods can be employed to epsilon SVR by the similar manner.

Bouloulis et employ Wirtingers calculus into Reproducing Kernel Hilbert Space (RKHS) so that expands real-SVM to pure complex SVM by exploiting complex kernel [ref complex support vector machine]. Based on this work, we construct a prototype of a complexity \$ performance controllable detector for large MIMO based on dual channel complex SVR. The detector can be divided into two parallel real SVR optimization problem which can be solved independently. Moreover, only real part of kernel matrix is needed in both channel. This means a large amount of computation can be reduced.

Steinwart et[ref SVM without offset] shows with a proper designed work set selection strategy, the approach that choosing double Lagrange multipliers can be much more faster than choosing single Lagrange multiplier without performance loss.

Based on the discrete time MIMO channel model, In our regression model, this CSVR-detector is constructed without offset, The offset in SVR imposes an additional linear quality constraint, which makes it necessary for decomposition methods such as Sequential Minimal Optimization to update more than one Lagrange multipliers in each iteration.

Therefore, for each real SVR without offset, in principle, only one offset is needed to be updated in each iteration, In our prototype, we propose a sequential single Lagrange multiplier search strategy that find two Lagrange multiplier sequentially, which can approximate the optimal dual Lagrange multiplier searching strategy. The former one only requires O(n) searches in one iteration, while the optimal dual Lagrange multiplier strategy requires O(n2) searches per iteration.

II. System Model

We consider a complex uncoded spatial multiplexing MIMO system with N_r receive and N_t transmit antennas, $N_r \geq N_t$, over a flat fading channel. Using a discrete time model, $\mathbf{y} \in \mathbb{C}^{N_r \times 1}$

is the received symbol vector written as:

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

where $\mathbf{s} \in \mathbb{C}^{N_t}$ is the transmitted symbol vector, with components that are mutually independent and taken from a finite signal constellation alphabet $\mathbb{O}$ (e.g. 4-QAM, 16-QAM, 64-QAM) of size M. The possible transmitted symbol vectors $\mathbf{s} \in \mathbb{O}^{N_t}$, satisfy $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_{N_t} E_s$, where E_s denotes the symbol average energy, and $\mathbb{E}[\cdot]$ denotes the expectation operation. Furthermore $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ denotes the Rayleigh fading channel propagation matrix with independent identically distributed (i.i.d) circularly symmetric complex Gaussian zero mean components with unit variance. Finally, $\mathbf{n} \in \mathbb{C}^{N_r}$ is the additive white Gaussian noise (AWGN) vector with zero mean components and $\mathbb{E}[\mathbf{n}\mathbf{n}^H] = \mathbf{I}_{N_r} N_0$, where N_0 denotes the noise power spectrum density, and hence $\frac{E_s}{N_0}$ is the signal to noise ratio (SNR).

Assume the receiver has perfect channel state information (CSI), meaning that $\mathbf{H}$ is known, as well as the SNR. The task of the MIMO decoder is to recover $\mathbf{s}$ based on $\mathbf{y}$ and $\mathbf{H}$.

III. Preliminaries

Orthogonality deficiency measures the how orthogonal a matrix is [13], which is defined by

$$\phi_{od} = 1 - \frac{det(\mathbf{W})}{\prod_{i=1}^{N_t} ||\mathbf{h}_i||^2}, \quad (2)$$

where $\mathbf{W} = \mathbf{H}^H\mathbf{H}$ denotes Wishart matrix, $\mathbf{h}_i$ denotes the i th column of $\mathbf{H}$, $det(\cdot)$ denotes determinant operation, $||\cdot||^2$ denotes 2-norm operation. In (2), $||\mathbf{h}_i||^2 = \sum_{j=1}^{N_r} |\mathbf{H}_{ji}|^2$, $\mathbf{H}_{ji}$ denotes the component of $\mathbf{H}$ at j th row and i th column. $|\mathbf{H}_{ji}| \sim Rayleigh(1/\sqrt{2})$, therefore $||\mathbf{h}_i||^2 \sim Gamma(N_r, 1)$ [14]. $Gamma(k, \theta)$ denotes Gamma distribution, with k degrees of freedom and scale θ. Furthermore, we have:

$$2||\mathbf{h}_i||^2 \sim Gamma(N_r, 2) \sim \chi^2_{2N_r}, \quad (3)$$

χ^2_k denotes chi-square distribution with k degrees of freedom. For the sake of simplicity, (2) can be changed to:

$$\phi_{om} = \frac{det(\mathbf{W})}{\prod_{i=1}^{N_t} ||\mathbf{h}_i||^2} = \frac{2^{N_t} det(\mathbf{W})}{\prod_{i=1}^{N_t} 2||\mathbf{h}_i||^2}. \quad (4)$$

Taking logarithmic operation to ϕ_{om} we have

$$\ln(\phi_{om}) = N_t \ln(2) + \ln(det(\mathbf{W})) - \sum_{i=1}^{N_t} \ln(2||\mathbf{h}_i||^2), \quad (5)$$

ϕ_{om} in (4) is defined as Orthogonality Measure. Based on Hadamard's inequality ($\prod_{i=1}^{N_t} ||\mathbf{h}_i|| \geq det(\mathbf{H})$). $\phi_{om} \in [0, 1]$. If ϕ_{om} is more closer to 1, $\mathbf{H}$ is closer to orthogonal matrix.

Because $\mathbf{W} = \mathbf{H}^H\mathbf{H}$, do QR factorization to $\mathbf{H}$

$$\mathbf{H} = \mathbf{Q}\mathbf{R}, \tag{6}$$

where $\mathbf{Q} \in \mathbb{C}^{N_r \times N_t}$ is a unitary matrix and $\mathbf{R} \in \mathbb{C}^{N_t \times N_t}$ is the upper triangular matrix. Using (6), we have $\mathbf{W} = \mathbf{R}^H\mathbf{R}$. r_{ii} denotes the i th diagonal component of $\mathbf{R}$, thus $det(\mathbf{W})$ can be rewritten as:

$$det(\mathbf{W}) = det(\mathbf{R}^H\mathbf{R}) = det(\mathbf{R}^H)det(\mathbf{R}) = \prod_{i=1}^{N_t} r_{ii}^H \prod_{i=1}^{N_t} r_{ii} = \prod_{i=1}^{N_t} |r_{ii}|^2. \tag{7}$$

Notice that $\mathbf{R}$ can be viewed as the Cholesky factorization of $\mathbf{W}$. Therefore, we have

$$||\mathbf{h}_i||^2 = \mathbf{W}_{ii} = \sum_{j=1}^{i-1} |r_{ji}|^2 + |r_{ii}|^2, \tag{8}$$

where $\mathbf{W}_{ii}$ denotes the i th diagonal element of $\mathbf{W}$. Thus based on (7) and (8), (4) can be rewritten as:

$$\phi_{om} = \prod_{i=1}^{N_t} \frac{|r_{ii}|^2}{|r_{ii}|^2 + \sum_{j=1}^{i-1} |r_{ji}|^2}. \tag{9}$$

IV. Brief Introduction to ϵ-Support Vector Regression

Suppose we are given training data set $((\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_l, y_l))$, l denotes the number of training samples, $\mathbf{x} \in \mathbb{R}^v$ denotes input data vector, v is the number of features in $\mathbf{x}$. y denotes output. The regression model (either linear or non-linear regression) is given by

$$y_i = \mathbf{w}^T \Phi(\mathbf{x}_i) + b \quad i \in 1 \cdots l \tag{10}$$

where $\mathbf{w}$ denotes regression coefficient vector, $\Phi(x)$ denotes the mapping of $\mathbf{x}$ to higher dimensional feature space. 10.

Here we give the primal optimization problem directly

$$\begin{aligned} & \frac{1}{2}||\mathbf{w}||^2 + \sum_{j=1}^{l} C_i(R(\xi_i) + R(\hat{\xi}_i)) \\ s.t. & \begin{cases} y_i - \mathbf{w}^T \Phi(\mathbf{x}_i) - b & \leq \epsilon + \xi_i \\ \mathbf{w}^T \Phi(\mathbf{x}_i) + b - y_i & \leq \epsilon + \hat{\xi}_i \\ \epsilon, \xi, \hat{\xi} & \geq 0 \end{cases} \end{aligned} \tag{11}$$

In 11, $\frac{1}{2}||\mathbf{w}||^2$ is the regularization term in order to ensure the flatness of regression model. ϵ denotes the precision, if the error between estimation and real output is less than ϵ, As shown in Fig 1, only those data points outside the shadow part, which is called ϵ tube, contribute to cost

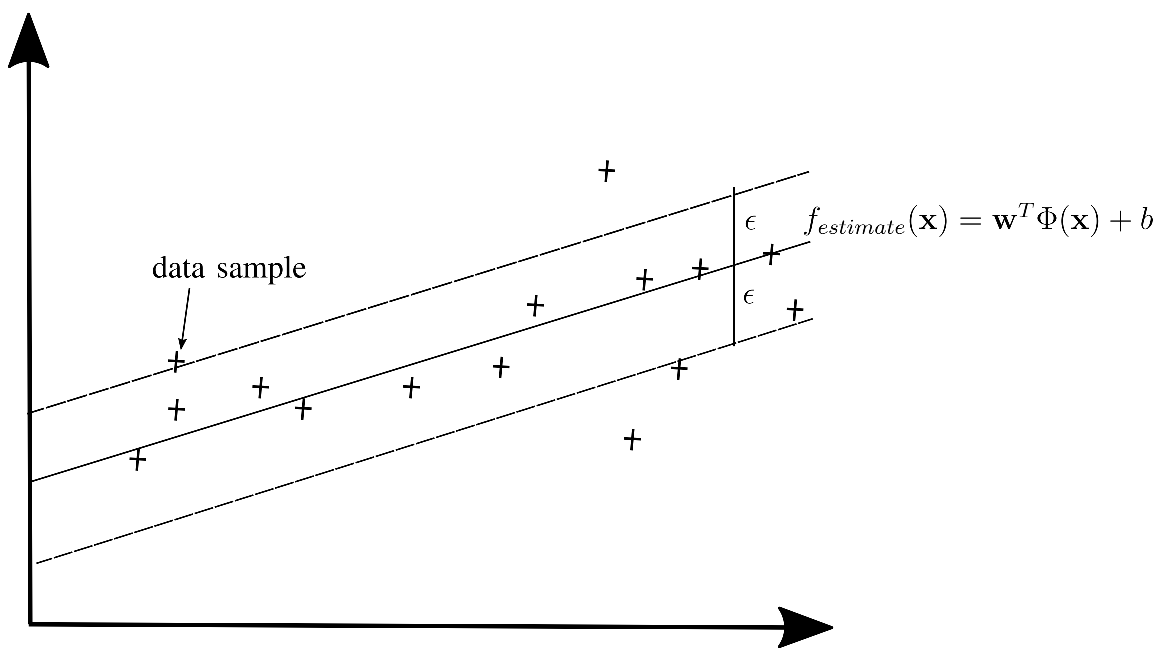
data sample
ϵ
ϵ
$f_{estimate}(\mathbf{x}) = \mathbf{w}^T\Phi(\mathbf{x}) + b$

Fig. 1

function. ξ and $\hat{\xi}$ denote slack variables that cope with noise of input data set, $R(x)$ denotes cost function, the simplest cost function is $R(x) = x$, risk function is determined by the statistical distribution of noise [**?**], for example if the noise subject to Gaussian distribution, the optimal cost function is $R(x) = \frac{1}{2}x^2$. $C\sum_{i=1}^{l}(\xi_i + \hat{\xi}_i)$ denotes the penalty of noise, $C \in \mathbb{R}$ and $C \geq 0$ controls the trade off between regularization term and noise penalty term.

From the rationale of risk function, let $f_{true}(\mathbf{x})$ denotes true regression function and $f_{estimate}(\mathbf{x})$, $c(\mathbf{x}, y, f_{estimate}(\mathbf{x}))$ denotes the risk function, the regression model can be written as $y = f_{true}(\mathbf{x}) + \xi$, ξ denotes additive noise. Assume the data samples are i.i.d. Based on Maximum Likelihood (ML) principle we want to

$$maximize \quad \prod_{i=1}^{l} P(y_i | f_{estimation}(\mathbf{x}_i)) \qquad = \prod_{i=1}^{l} P(\xi_i) \tag{12}$$

$$= \prod_{i=1}^{l} P(y_i - f(\mathbf{x}_i)), \tag{13}$$

Take the logarithm of (13), we have

$$maximize \quad \sum_{i=1}^{l} log(P(y_i - f_{estimation}(\mathbf{x}_i))), \tag{14}$$

Therefore the ith risk function of $(\mathbf{x_i}, y_i)$ can be written as

$$c(\mathbf{x}_i, y_i, f_{estimation}(\mathbf{x}_i)) = -log(P(y_i - f_{estimation}(\mathbf{x}_i))). \tag{15}$$

In ϵ-SVR, Vapnik's ϵ-insensitive function, as shown in (16), is applied to (15).

$$|x|_\epsilon = \begin{cases} 0 & if \quad |x| < \epsilon \\ \\ |x| - \epsilon & otherwise \end{cases} \tag{16}$$

Thus the cost function in ϵ-SVR can be written as

$$\tilde{c}(\mathbf{x}, y, f_{estimation}(\mathbf{x})) = \frac{1}{l}\sum_{i=1}^{l} m_i(-log(P(|y_i - f_{estimation}(\mathbf{x}_i)|_\epsilon))), \tag{17}$$

where $m_i \in \mathbb{R}$, $m_i > 0$ denotes the weight parameter, if $y_i > f_{estiamtion}(\mathbf{x})$, $m_i = m_{positive}$, else $m_i = m_{negative}$, Therefore the cost function with regularization term can written as

$$minimize \quad \lambda||w||^2 + \tilde{c}(\mathbf{x}, y, f_{estimation}(\mathbf{x})), \tag{18}$$

where λ denotes the weight of regularization term, divide (18) by $\frac{1}{2\lambda}$, we have the optimization problem

$$minimize \quad \frac{1}{2}||w||^2 + \sum_{i=1}^{l} C_i(-log(P(|y_i - f_{estimation}(\mathbf{x}_i)|_\epsilon))), \tag{19}$$

where $C_i = \frac{m_i}{2\lambda l}$, based on (19), by introducing slack variables, we can easily derive the equivalent optimization problem as same as (11):

$$\frac{1}{2}||\mathbf{w}||^2 + \sum_{j=1}^{l} C_i(R(\xi_i) + R(\hat{\xi}_i))$$
$$s.t. \begin{cases} y_i - \mathbf{w}^T\Phi(\mathbf{x}_i) - b & \leq \epsilon + \xi_i \\ \mathbf{w}^T\Phi(\mathbf{x}_i) + b - y_i & \leq \epsilon + \hat{\xi}_i \\ \epsilon, \xi, \hat{\xi} & \geq 0 \end{cases} \tag{20}$$

where $R(x) = -log(P(x))$, by this way, the discontinuity of ϵ-insensitive function is conquered, we arrive to at a convex minimization problem [**?**].

we construct dual form of (11) by introducing Lagrange multiplier, we have dual optimization problem

$$\min_{\mathbf{w},\xi,\hat{\xi}} \max_{\alpha,\hat{\alpha},\eta,\hat{\eta}} \theta = \frac{1}{2}||\mathbf{w}||^2 + \sum_{j=1}^{l} C_i(R(\xi_i) + R(\hat{\xi}_i)) - \sum_{i=1}^{l}(\eta_i\xi_i + \hat{\eta}_i\hat{\xi}_i)$$
$$- \sum_{i=1}^{l} \alpha_i(\epsilon + \xi_i - y_i + \mathbf{w}^T\Phi(\mathbf{x}_i)) - \sum_{i=1}^{l} \hat{\alpha}_i(\epsilon + \hat{\xi}_i + y_i - \mathbf{w}^T\Phi(\mathbf{x}_i))$$
$$s.t. \begin{cases} \eta, \hat{\eta}, \alpha, \hat{\alpha} \geq 0 \\ \xi, \hat{\xi} \geq 0 \end{cases} \tag{21}$$

where η,$\hat{\eta}$, α, $\hat{\alpha}$ are Lagrange multipliers, the solution of dual optimization problem is saddle point, from [?], (21) is equivalent to

$$\max_{\alpha,\hat{\alpha},\eta,\hat{\eta}} \min_{\mathbf{w},\xi,\hat{\xi}} \Theta = \frac{1}{2}||\mathbf{w}||^2 + \sum_{j=1}^{l} C_i(R(\xi_i) + R(\hat{\xi}_i)) - \sum_{i=1}^{l}(\eta_i\xi_i + \hat{\eta}_i\hat{\xi}_i)$$
$$- \sum_{i=1}^{l} \alpha_i(\epsilon + \xi_i - y_i + \mathbf{w}^T\Phi(\mathbf{x}_i)) - \sum_{i=1}^{l} \hat{\alpha}_i(\epsilon + \hat{\xi}_i + y_i - \mathbf{w}^T\Phi(\mathbf{x}_i))$$
$$s.t. \begin{cases} \eta, \hat{\eta}, \alpha, \hat{\alpha} \geq 0 \\ \xi, \hat{\xi} \geq 0 \end{cases} \tag{22}$$

Take the partial derivative of Θ with respect to $\mathbf{w}$, ξ, $\hat{\xi}$ and b, and find the minimum.

$$\frac{\partial\theta}{\partial\mathbf{w}} = w - \sum_{i=1}^{l}(\alpha_i - \hat{\alpha}_i) \tag{23}$$

$$\frac{\partial\theta}{\partial\xi} = C_iR^{'}(\xi_i) - \eta_i - \alpha_i = 0 \tag{24}$$

$$\frac{\partial\theta}{\partial\hat{\xi}} = C_iR^{'}(\hat{\xi}_i) - \hat{\eta}_i - \hat{\alpha}_i = 0 \tag{25}$$

$$\frac{\partial\theta}{\partial b} = \sum_{i=1}^{l}(\alpha_i - \hat{\alpha}_i) = 0 \tag{26}$$

Then substitute (23)-(26) to (22), yields the final dual optimization objective function, for sake of brevity, we make C_i uniform to all data samples

$$\begin{aligned} maximize \quad \Theta &= \frac{1}{2}\sum_{i=1}^{l}\sum_{i=1}^{l}(\alpha_i-\hat{\alpha}_i)(\alpha_j-\hat{\alpha}_j)\Phi(\mathbf{x}_j)^T\Phi(\mathbf{x}_i)+C\sum_{i=1}^{l}[(R(\xi_i)-\xi_iR'(\xi_i)) \\ &+ (R(\hat{\xi}_i)-\hat{\xi}_iR'(\hat{\xi}_i))]+\sum_{i=1}^{l}[(\alpha_i-\hat{\alpha}_i)y_i-(\alpha_i+\hat{\alpha}_i)\epsilon] \\ &- \sum_{i=1}^{l}\sum_{i=1}^{l}(\alpha_i-\hat{\alpha}_i)(\alpha_j-\hat{\alpha}_j)\Phi(\mathbf{x}_j)^T\Phi(\mathbf{x}_i), \end{aligned} \tag{27}$$

$$s.t. \begin{cases} \sum_{i=1}^{l}(\alpha_i-\hat{\alpha}_i)=0 \\ 0<\alpha<C\tilde{R}(\alpha) \\ 0<\hat{\alpha}<C\tilde{R}(\hat{\alpha}) \end{cases}$$

(28) can be simplified to

$$\begin{aligned} maximize \quad \Theta &= -\frac{1}{2}\sum_{i=1}^{l}\sum_{i=1}^{l}(\alpha_i-\hat{\alpha}_i)(\alpha_j-\hat{\alpha}_j)\Phi(\mathbf{x}_j)^T\Phi(\mathbf{x}_i)+\sum_{i=i}^{l}[(\alpha_i-\hat{\alpha}_i)y_i-(\alpha_i+\hat{\alpha}_i)\epsilon] \\ &+ C\sum_{i=1}^{l}[\tilde{R}(\xi_i)+\tilde{R}(\hat{\xi}_i)] \\ &= -\frac{1}{2}(\mathbf{a}-\hat{\mathbf{a}})^T\mathbf{K}(\mathbf{a}-\hat{\mathbf{a}})+(\mathbf{y}-\epsilon)^T\mathbf{a}+(-\mathbf{y}-\epsilon)^T\hat{\mathbf{a}}+\mathbf{e}^TC(\tilde{R}(\xi)+\tilde{R}(\hat{\xi})), \end{aligned} \tag{28}$$

where $\mathbf{a}=[\alpha_1,\alpha_2,\ldots,\alpha_l]^T$, $\hat{\mathbf{a}}=[\hat{\alpha}_1,\hat{\alpha}_2,\ldots,\hat{\alpha}_l]^T$, $\mathbf{y}=[y_1,y_2,\ldots y_l]^T$, $\mathbf{e}=[1,1,\ldots,1]^T\in\mathbb{R}^l$, $\mathbf{e}_i$ denotes the vector that only ith component is 1 while the rest are all 0, $\tilde{R}(\xi)=R(\xi)-\xi R'(\xi)\in\mathbb{R}^l$, $\mathbf{K}_{ij}=\Phi(\mathbf{x}_j)^T\Phi(\mathbf{x}_i)$ denotes data kernel matrix. We define the following $2l$ vectors $\mathbf{a}^{(*)}$=

$[\begin{smallmatrix}\mathbf{a}\\ \hat{\mathbf{a}}\end{smallmatrix}]$, $\mathbf{v} \in \mathbb{R}^{2l}$,

$$\mathbf{v}_i = \begin{cases} 1 & i = 1, \ldots, l \\ \\ -1 & i = l+1, \ldots, 2l \end{cases} \tag{29}$$

(28) can be reformulate as

$$maximize \quad \Theta = -\frac{1}{2}(\mathbf{a}^*)^T \begin{bmatrix} \mathbf{K} & -\mathbf{K} \\ \\ -\mathbf{K} & \mathbf{K} \end{bmatrix} \mathbf{a}^{(*)} + [(\mathbf{y}-\epsilon)^T, (-\mathbf{y}-\epsilon)^T]\mathbf{a}^{(*)} + \mathbf{e}^T C(\tilde{R}(\xi) + \tilde{R}(\hat{\xi})), \tag{30}$$

V. Dual Channel Complex Kernel Support Vector Regression for Large MIMO System

Taking expectation of (5), we have

$$\mathbb{E}[\ln(\phi_{om})] = N_t \ln(2) + \mathbb{E}[\ln(det(\mathbf{W}))] - \sum_{i=1}^{N_t} \mathbb{E}[\ln(2||\mathbf{h}_i||^2)]. \tag{31}$$

Consider $\mathbf{H} = [\mathbf{h}_1^{'}, \mathbf{h}_2^{'}, \cdots \mathbf{h}_{N_r}^{'}]^{'}$, where $\mathbf{h}_i$ denotes the i th row of $\mathbf{H}$, because each component of $\mathbf{H}$ is mutually independent and subject to circularly symmetric complex Gaussian distribution, i.e. $\mathbf{h}_i \sim \mathbb{C}N(\mathbf{0}, \mathbf{I}_{N_t})$. Therefore, $\mathbf{W} = \mathbf{H}^H\mathbf{H} \sim \mathbb{C}W(N_r, \mathbf{I}_{N_t})$, $\mathbb{C}W(n, \mathbf{\Sigma})$ denotes complex Wishart distribution with n degrees of freedom and covariance matrix $\mathbf{\Sigma}$. The logarithmic expectation of $\mathbf{W}$ can be rewritten as

$$\mathbb{E}[\ln(det(\mathbf{W}))] = \frac{\tilde{\Gamma}'_{N_t}(N_r)}{\tilde{\Gamma}_{N_t}(N_r)} = \sum_{i=1}^{N_t} \psi(N_r - i + 1), \tag{32}$$

where $\tilde{\Gamma}_m(n)$ denotes the multivariate Gamma function and $\psi(n)$ denotes Digamma function. Proof: see Appendix A.

Because the logarithmic expectation of a Gamma distribution variable $\eth \sim Gamma(n, \theta)$ can be written as:

$$\mathbb{E}[\ln(\eth)] = \psi(n) + \ln(\theta), \tag{33}$$

where $\psi(n)$ denotes Digamma function. Thus according to (3), we have:

$$\mathbb{E}[\ln(2||\mathbf{h}_i||^2)] = \psi(N_r) + \ln(2). \tag{34}$$

Proof: see Appendix B.

Based on (31)(32)(34), The logarithmic expectation of ϕ_{om} can be written as:

$$\begin{aligned} \mathbb{E}[\ln(\phi_{om})] &= N_t \ln(2) + \sum_{i=1}^{N_t} \psi(N_r - i + 1) - N_t \psi(N_r) - N_t \ln(2) \\ &= \sum_{i=1}^{N_t} \psi(N_r - i + 1) - N_t \psi(N_r) \end{aligned} \tag{35}$$

VI. Work Set Selection

A. Single Direction Solver

B. Double Direction Solver

C. Approximation to Optimal Double Direction Solver based on Single Direction Solver

Recall (9)

$$\phi_{om} = \prod_{i=1}^{N_t} \frac{|r_{ii}|^2}{|r_{ii}|^2 + \sum_{j<i} |r_{ji}|^2}. \tag{36}$$

All the components in $\mathbf{R}$ are independently distributed and $r_{ji} \sim \mathbb{C}N(0,1)$, $|r_{ii}|^2 \sim Gamma(N_r - i + 1, 1)$ [15]. Because $|r_{ji}| \sim Rayleigh(1/\sqrt{2})$, $\sum_{j<i} |r_{ji}|^2 \sim Gamma(i-1,1)$. Defining $\alpha_i = \sum_{j<i} |r_{ji}|^2$ and $\beta_i = |r_{ii}|^2$, α_i and β_i are mutually independent, therefore (9) can be rewritten as

$$\phi_{om} = \prod_{i=1}^{N_t} \frac{\beta_i}{\beta_i + \alpha_i}, \tag{37}$$

From [16], if $X \sim Gamma(k_1, \theta)$ and $Y \sim Gamma(k_2, \theta)$, then $\frac{X}{X+Y} \sim B(k_1, k_2)$, where B denotes Beta distribution. Therefore $\frac{\beta_i}{\beta_i + \alpha_i} \sim B(k_1^i, k_2^i)$, where $k_1^i = N_r - i + 1$, $k_2^i = i - 1$. we define $\eta_i = \frac{\beta_i}{\beta_i + \alpha_i}$, it is obvious that η_i are independently distributed. Based on (37), we have

$$\phi_{om} = \prod_{i=1}^{N_t} \eta_i. \tag{38}$$

Therefore the density function of ϕ_{om} can be defined as

$$f_{\phi_{om}}(x) = \frac{1}{x}\sum_{\mathbf{j}}(\prod_{i=1}^{N_t} c(k_1^i k, k_2^i, j^i)) f(-\ln(x)|\mathbf{k_1}+\mathbf{j}), \tag{39}$$

where $\sum_{\mathbf{j}} = \sum_{j^1}\sum_{j^2}\cdots\sum_{j^{N_t}}$, the range of $j^i \in [0, k_2^i - 1]$, $c(k_1^i, k_2^i, j_i) = (-1)^{j^i}\binom{k_2^i-1}{j^i}[(k_1^i + k_2^i)\mathbb{B}(k_1^i, k_2^i)]^{-1}$, $\mathbb{B}(\alpha,\beta)$ denotes beta function. $f(-\ln(x)|\mathbf{k_1}+\mathbf{j}) = (\prod_{i=1}^{N_t}(k_1^i+j^i))\sum_{i=1}^{N_t}[exp((k_1^i+j^i)\ln(x))/\prod_{j=1 j\neq i}^{N_t}(k_1^j + j^j - k_1^i - j^i)]$. $\mathbf{k_1}+\mathbf{j} = [k_1^1 + j^1, \cdots k_1^{N_t-1} + j^{N_t-1}, k_1^{N_t} + j^{N_t}]$. Proof: see Appendix C.

Consider logarithmic expectation of ϕ_{om}, we have

$$E[\ln(\phi_{om})] = \sum_{i=1}^{N_t} E[\ln(\eta_i)], \tag{40}$$

where $E[\ln(\eta_i)] = \psi(k_1^i) - \psi(k_1^i + k_2^i)$, thus we have

$$E[\ln(\phi_{om})] = \sum_{i=1}^{N_t}\psi(N_r - i + 1) - N_t\psi(N_r). \tag{41}$$

we can find (41) is consistent with (35).

VII. Initialization

Computer simulations are made for different sizes of V-BLAST MIMO systems, with $5 \leq N_r \leq 100, 5 \leq N_t \leq N_r$, the empirical estimation of logarithmic expectation of ϕ_{om}, $E[\ln(\phi_{om})]_{em}$, is calculated by taking average over $1e4$ channel realizations for each size of MIMO systems, as shown in Fig.**??**, the Theoretical logarithmic expectation of ϕ_{om} $E[\ln(\phi_{om})]_t$ in (41) is plotted in Fig.**??**. Average deviation between $E[\ln(\phi_{om})]_{em}$ and $E[\ln(\phi_{om})]_t$ is also calculated, $V_{em-t} = 7.3043e - 04$.

Fig.**??** demonstrates the relation between the number of users (N_t) and $E[\ln(\phi_{om})]_t$ under cases of different numbers of antennas at base station (N_r). From Fig.**??**, we can see, on the one hand, with N_r fixed, $E[\ln(\phi_{om})]$ decreases while N_t increases, however the gradient of each curve becomes more and more gentle. On the other hand, when N_r becomes larger $E[\ln(\phi_{om})]$ becomes more insensitive to variation of N_t.

VIII. Stopping Criteria

IX. Hyperparameter Setting

X. Computer Simulations

XI. Conclusion

The conclusion goes here.

Appendix A

Proof of the First Zonklar Equation

Appendix one text goes here.

Appendix B

Appendix two text goes here.

Appendix C

Let $\mathbf{A} \in \mathbb{C}^{m \times m}$, $A \sim \mathbb{C}W(n, \mathbf{\Sigma})$, $\mathbb{C}W(n, \mathbf{\Sigma})$ denotes complex Wishart distribution with n degrees of freedom and covariance matrix $\mathbf{\Sigma}$. It is obvious $\mathbf{A}$ is Hermition positive definite matrix, $\mathbf{A} = \mathbf{A}^H > 0$.

The pdf of $\mathbf{A}$ can be written as [15]:

$$f(\mathbf{A}) = \{\tilde{\Gamma}_m(n) det(\mathbf{\Sigma})^n\}^{-1} det(\mathbf{A})^{n-m} etr(-\mathbf{\Sigma}^{-1}\mathbf{A}), \tag{42}$$

where $\tilde{\Gamma}_m(\beta)$ denotes multivariate complex Gamma function defined by:

$$\tilde{\Gamma}_m(\beta) = \pi^{\frac{m(m-1)}{2}} \prod_{i=1}^{m} \Gamma(\beta - i + 1) \quad Re(\beta) > m - 1. \tag{43}$$

Furthermore, from [15], we have

$$\tilde{\Gamma}_m(\beta) = \int_{\mathbf{X}=\mathbf{X}^H>0} etr(-\mathbf{X}) det(\mathbf{X})^{\beta-m} d\mathbf{X} \quad Re(\beta) > m - 1. \tag{44}$$

We derive logarithmic expectation of $det(\mathbf{A})$

$$\begin{aligned} E[\ln(det(\mathbf{A}))] &= \int_{\mathbf{A}=\mathbf{A}^H>0} \ln(det(\mathbf{A}))f(\mathbf{A})d\mathbf{A} \\ &= \int_{\mathbf{A}=\mathbf{A}^H>0} \ln(det(\mathbf{A}))\{\tilde{\Gamma}_m(n)det(\mathbf{\Sigma})^n\}^{-1}det(\mathbf{A})^{n-m}etr(-\mathbf{\Sigma}^{-1}\mathbf{A})d\mathbf{A} \\ &= \frac{det(\mathbf{\Sigma})^{-n}}{\tilde{\Gamma}_m(n)}\int_{\mathbf{A}=\mathbf{A}^H>0} \ln(det(\mathbf{A}))det(\mathbf{A})^{n-m}etr(-\mathbf{\Sigma}^{-1}\mathbf{A})d\mathbf{A}, \end{aligned} \tag{45}$$

if $\mathbf{\Sigma} = \mathbf{I}$, (45) can be written as

$$E[\ln(det(\mathbf{A}))] = \frac{1}{\tilde{\Gamma}_m(n)}\int_{\mathbf{A}=\mathbf{A}^H>0} \ln(det(\mathbf{A}))det(\mathbf{A})^{n-m}etr(-\mathbf{A})d\mathbf{A}. \tag{46}$$

Because $\frac{d}{dn}[det(\mathbf{A})]^{n-m} = \ln(det(\mathbf{A}))det(\mathbf{A})^{n-m}$, (46) can be rewritten as

$$E[\ln(det(\mathbf{A}))] = \frac{1}{\tilde{\Gamma}_m(n)}\frac{d}{dn}\int_{\mathbf{A}=\mathbf{A}^H>0} etr(-\mathbf{A})det(\mathbf{A})^{n-m}d\mathbf{A}, \tag{47}$$

using (44), (47) can be rewritten as

$$E[\ln(\mathbf{A})] = \frac{\tilde{\Gamma}'_m(n)}{\tilde{\Gamma}_m(n)}. \tag{48}$$

Based on (43), we have

$$\tilde{\Gamma}'_m(n) = \pi^{\frac{m(m-1)}{2}}\sum_{i=1}^{m}[\Gamma'(n-i+1)\prod_{j=1,j\neq i}^{m}\Gamma(n-j+1)], \tag{49}$$

Thus we have

$$E[\ln(det(\mathbf{A}))] = \frac{\tilde{\Gamma}'_m(n)}{\tilde{\Gamma}_m(n)} = \sum_{i=1}^{m} \frac{\Gamma'(n-i+1)}{\Gamma(n-i+1)} = \sum_{i=1}^{m} \psi(n-i+1), \tag{50}$$

where ψ denotes Digamma function.

APPENDIX D

If $x \sim Gamma(n, \theta)$, with shape parameter k and scale parameter θ, $x > 0$, $\Gamma(k)$ denotes Gamma function, the density function of Gamma distribution is

$$f(x, k, \theta) = \frac{x^{k-1} e^{-x/\theta}}{\Gamma(k)\theta^k}. \tag{51}$$

Thus we have

$$E[\ln(x)] = \frac{1}{\Gamma(k)} \int_0^{\infty} \ln(x) x^{k-1} e^{-x/\theta} \theta^{-k} dx, \tag{52}$$

define $z = x/\theta$ and since $\Gamma(k) = \int_0^{\infty} x^{k-1} e^{-x} dx$, (52) can be rewritten as

$$E[\ln(x)] = \ln(\theta) + \frac{1}{\Gamma(k)} \int_0^{\infty} \ln(z) z^{k-1} e^{-z} dz. \tag{53}$$

Because $\frac{d(z^{k-1})}{dk} = \ln(z)z^{k-1}$, (53) can be rewritten as

$$
\begin{aligned}
E[\ln(z)] &= \ln(\theta) + \frac{1}{\Gamma(k)}\frac{d}{dk}\int_0^\infty z^{k-1}e^{-z}dz \\
&= \ln(\theta) + \frac{\Gamma'(k)}{\Gamma(k)} \\
&= \ln(\theta) + \psi(k),
\end{aligned}
$$

where $\psi(k)$ denotes Digamma function.

APPENDIX E

$x_1, x_2, \cdots x_{N_t}$ are independent beta variables, the probability density function (pdf) can be written as:

$$
f(x_i) = \frac{1}{\mathbb{B}(k_1^i, k_2^i)} x_i^{k_1^i - 1}(1 - x_i)^{k_2^i - 1}, \tag{54}
$$

define $y_i = -\ln(x_i) = g(x_i)$, Based on Jacobian transformation, we have

$$
f_{y_i}(\rho) = |\frac{dy_i}{dx_i}|^{-1} f_{x_i}(g^{-1}(\rho)) = \frac{1}{\mathbb{B}(k_1^i, k_2^i)} e^{-k_1^i \rho}(1 - e^{-\rho})^{k_2^i - 1}. \tag{55}
$$

where (55) can be alternatively expressed as [17]

$$
f_{y_i}(\rho) = \sum_{j^i=0}^{k_2^i - 1} c(k_1^i, k_2^i, j^i)(k_1^i + j^i) exp(-(k_1^i + j^i)\rho), \tag{56}
$$

where $c(k_1^i, k_2^i, j_i) = (-1)^{j^i} \binom{k_2^i - 1}{j^i} [(k_1^i + k_2^i)\mathbb{B}(k_1^i, k_2^i)]^{-1}$, $\mathbb{B}(\alpha, \beta)$ denotes beta function. Based on the lemma 1 of [17], if $a_1, a_2, \cdots a_n$ are independent exponentially distributed random variables, with pdf given by

$$t_i exp(-t_i a_i) \tag{57}$$

then pdf of $a = \sum_{i=1}^{n} a_i$ can be written as

$$f(a|\mathbf{t}) = \prod_{i=1}^{n} t_i \sum_{i=1}^{n} [exp(-t_i a) / \prod_{j=1 j\neq i}^{j=n} (t_j - t_i)], \tag{58}$$

where $t = [t_1, t_2, \cdots t_n]$. The pdf of y_i can be viewed as the weighting summation of exponential distribution functions, define $y = \sum_{i=1}^{n} y_i$, based on (58), the pdf of y is given by

$$f_y(m) = \sum_{\mathbf{j}} \{[\prod_{i=1}^{n} c(k_1^i, k_2^i, j^i)] f(m|\mathbf{k_1} + \mathbf{j})\}, \tag{59}$$

where $\sum_{\mathbf{j}} = \sum_{j^1} \sum_{j^2} \cdots \sum_{j^n}$, the range of j^i is defined by $j^i \in [0, k_2^i]$, $f(m|\mathbf{k_1} + \mathbf{j}) = (\prod_{i=1}^{N_t} (k_1^i + j^i)) \sum_{i=1}^{N_t} [exp(-(k_1^i + j^i)m) / \prod_{j=1 j\neq i}^{N_t} (k_1^j + j^j - k_1^i - j^i)]$, $\mathbf{k_1} + \mathbf{j} = [k_1^1 + j^1, k_1^2 + j^2 \cdots k_1^n + j^n]$. we define $U = exp(-y) = \prod_{i=1}^{n} x_i$, using Jacobian transformation, the pdf of U is given by

$$f_U(u) = |\frac{du}{dy}|^{-1} f_y(-\ln(u)) = \frac{1}{u} \sum_{\mathbf{j}} \{[\prod_{i=1}^{n} c(k_1^i, k_2^i, j^i)] f(-\ln(u)|\mathbf{k_1} + \mathbf{j})\}. \tag{60}$$

References

[1] 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.

[2] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[3] V. A. Marčenko and L. A. Pastur, “Distribution of eigenvalues for some sets of random matrices,” *Sbornik: Mathematics*, vol. 1, no. 4, pp. 457–483, 1967.

[4] A. M. Tulino and S. Verdú, “Random matrix theory and wireless communications,” *Communications and Information theory*, vol. 1, no. 1, pp. 1–182, 2004.

[5] M. Wu, B. Yin, A. Vosoughi, C. Studer, J. R. Cavallaro, and C. Dick, “Approximate matrix inversion for high-throughput data detection in the large-scale MIMO uplink,” in *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*. IEEE, 2013, pp. 2155–2158.

[6] J. C. Fricke, M. Sandell, J. Mietzner, and P. A. Hoeher, “Impact of the Gaussian approximation on the performance of the probabilistic data association MIMO decoder,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2005, no. 5, pp. 796–800, 1900.

[7] D. Pham, K. R. Pattipati, P. K. Willett, and J. Luo, “A generalized probabilistic data association detector for multiple antenna systems,” *IEEE Communications Letters*, vol. 8, no. 4, pp. 205–207, 2004.

[8] P. Som, T. Datta, N. Srinidhi, A. Chockalingam, and B. S. Rajan, “Low-complexity detection in large-dimension MIMO-ISI channels using graphical models,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 8, pp. 1497–1511, 2011.

[9] J. Goldberger and A. Leshem, “MIMO detection for high-order QAM based on a Gaussian tree approximation,” *Information Theory, IEEE Transactions on*, vol. 57, no. 8, pp. 4973–4982, 2011.

[10] T. Lakshmi Narasimhan and A. Chockalingam, “Channel hardening-exploiting message passing (CHEMP) receiver in large-scale MIMO systems,” 2013.

[11] B. Farhang-Boroujeny, H. Zhu, and Z. Shi, “Markov chain Monte Carlo algorithms for CDMA and MIMO communication systems,” *Signal Processing, IEEE Transactions on*, vol. 54, no. 5, pp. 1896–1909, 2006.

[12] T. Datta, N. Ashok Kumar, A. Chockalingam, and B. S. Rajan, “A novel MCMC algorithm for near-optimal detection in large-scale uplink mulituser MIMO systems,” in *Information Theory and Applications Workshop (ITA), 2012*. IEEE, 2012, pp. 69–77.

[13] X. Ma and W. Zhang, “Performance analysis for MIMO systems with lattice-reduction aided linear equalization,” *Communications, IEEE Transactions on*, vol. 56, no. 2, pp. 309–318, 2008.

[14] A. Papoulis, “Stochastic processes,” *McGra. w*, 1996.

[15] D. K. Nagar and A. K. Gupta, “Expectations of functions of complex Wishart matrix,” *Acta applicandae mathematicae*, vol. 113, no. 3, pp. 265–288, 2011.

[16] A. K. Gupta and S. Nadarajah, *Handbook of beta distribution and its applications*. CRC Press, 2004.

[17] R. Bhargava and C. Khatri, “The distribution of product of independent beta random variables with application to multivariate analysis,” *Annals of the Institute of Statistical Mathematics*, vol. 33, no. 1, pp. 287–296, 1981.