In this chapter, a support vector regression (SVR) based detector is presented for detection in large-MIMO systems. The main motivation for the current work is that the existing large-MIMO detectors do not offer a significant performance-complexity tradeoff for BPSK constellations.

The binary signaling scheme in BPSK constellations need to be exploited to reduce the complexity of the detectors. There is a need for a detection technique for large-MIMO systems employing BPSK constellations that can offer a better performance-complexity tradeoff.

The main objective is to design an efficient and low complex detector for BPSK constellations, and the choice is support vector regression. This is motivated from the applications of support vector regression available in the literature. The efficiency of the SVR detector is determined by the BER performance which needs to be superior
to the lattice reduction aided detectors.

The existing detection algorithms for large-MIMO systems do not offer performance-complexity trade-off as they have a non-linear formulation that can detect symbols from any general square constellation. The solution to the non-linear formulation renders computational overhead. On the other hand, a simple solution methodology will result in poor performance. Hence those detectors fail to achieve a performance-complexity trade-off in the case of BPSK constellations in which, the symbols are linearly separable in the classification point of view. This forms the basic premise on which, the SVR detector is constructed.

The outline of the work in this chapter is as follows.

- The detection problem in a large-MIMO system employing BPSK constellations is formulated using a support vector regression model and subsequently solved.
- The performance of SVR detectors in $16 \times 16$ and $32 \times 32$ MIMO systems are studied.
- The BER and complexity of SVR detector is compared with lattice reduction aided detectors and its efficiency is substantiated.

Support vector machines belong to the class of artificial intelligence and have been used to solve large scale classification and regression problems. Support vector machines were initially developed as classifiers but their use later extended to the problem of regression. Support vector machines has found its application in many fields. The recent applications include pedestrian classification in [Aly (2014)], vehicle type and color
classification in Benkedjouh et al. (2013), fault detection in Gryllias and Antoniadis (2012), credit scoring in Han et al. (2013) and digital watermarking in Yang et al. (2013). These applications confirm that support vector machines are low complex and can learn large scale data. Further linear circuit implementation of support vector machines available in Decherchi et al. (2008) suggest the possibility of hardware implementation. Improvised version of SVM through combination with particle swarm optimization is available and ant colony optimization are also available in Liao et al. (2011) and Zhou et al. (2012) respectively.

Support vector regression (SVR) is the natural extension of large margin kernel methods used for classification to regression analysis. It retains all the properties that characterize maximal margin algorithms of support vector machines such as duality, sparseness, kernel and convexity. As a difference, support vector regression formulation introduces the concept of loss function which ignores errors occurring within a distance of the true value. Several variants of support vector regression are available and its solution can be obtained by using any standard quadratic programming solver. SVR has become a powerful technique for predictive data analysis with many applications in varied areas of study. A wide range of applications solved using support vector regression is available in the literature which include image denoising by Santamaría (2003), linear phase FIR filter design by Cheng et al. (2004) and voice conversion by Song et al. (2011).

Recently wireless communication has attracted support vector concept to address problems like joint transmit and receive antenna selection in Naeem and Lee (2011),
decision making in cognitive radios by Wu et al. (2012), pilot design for downlink large scale MIMO systems in Gao et al. (2014) and large-MIMO beamforming by Wang et al. (2015). This literature recommends the support vector regression as a low complex approach which can learn large data thereby confirming the suitability of SVR for large scale systems. These studies have motivated the use of SVR for detection in large-MIMO systems.

The organization of the chapter is as follows. In section 5.1 a brief review of the support vector regression is presented with the required mathematical insights. The support vector regression formulation for large-MIMO detection is presented in section 5.2. The performance evaluation of the SVR detector is presented in section 5.3 and the summary is presented in section 5.4.

### 5.1 Support Vector Regression

In this section, a brief review of support vector regression is presented. Support vector theory is firmly grounded in the framework of statistical learning theory, or Vapnik-Chervonenkis (VC) theory, which has been developed over the last decades. In a nutshell, VC theory characterizes properties of learning machines which enable them to generalize to unknown data. There is plethora of materials available for study on support vector regression. However Smola and Scholkopf (2003) provides a good tutorial on support vector regression and better insight into the algorithms used to solve it. In an \( \varepsilon \)-SV regression, the goal is to find a function \( f(x) \) that has at most \( \varepsilon \) deviation from the actually obtained targets \( y_i \) for all the training data, and at the same time should
be as flat as possible. In other words, errors less than \( \varepsilon \) are ignored, but any deviation larger than this is not allowed. In the case of linear functions \( f \), taking the form

\[
f(x) = \langle w, x \rangle + b
\]

(5.1)

where \( \langle ., . \rangle \) refers to the standard inner product, flatness in the above case means that one seeks a small \( w \). One way to ensure this is to minimize the norm, i.e \( \|w\|^2 = \langle w, w \rangle \).

This problem can be cast as a convex optimization problem:

\[
\min \frac{1}{2} \|x\|^2
\]

(5.2)

\[
s.t. \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon \\ -y_i + \langle w, x_i \rangle + b \leq \varepsilon \end{cases}
\]

(5.3)

The assumption in the above equation is that a function \( f \) actually exists and it approximates all pairs \((x_i, y_i)\) with \( \varepsilon \) precision. This implies that the convex optimization problem is feasible. However, this may not be the case always. Similar to the “soft margin” loss function, the slack variables \( \xi_i, \xi_i^* \) are introduced to cope with infeasible constraints of the above optimization problem. The following formulation is obtained.

\[
\min \frac{1}{2} \|x\|^2 + C \sum_{i=1}^{t} (\xi_i + \xi_i^*)
\]

(5.4)

\[
s.t. \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ -y_i + \langle w, x_i \rangle + b \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}
\]

(5.5)

The constant \( C > 0 \) determines the trade-off between the flatness of \( f \) and the tolerance in deviations larger than \( \varepsilon \). This \( \varepsilon \)-insensitive loss function \( |\xi|_\varepsilon \) is described by,

\[
|\xi|_\varepsilon := \begin{cases} 0 & \text{if } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{otherwise} \end{cases}
\]

(5.6)
The soft margin loss setting for a linear SVM is depicted in Fig. 5.1. Only the points outside the shaded region contribute to the cost as the deviations are penalized in a linear fashion. In most of the cases the above optimization problem can be solved easily in its dual formulation.

Since we are interested in linear and simple formulation, the above formulation also called as the primal formulation is sufficient. There is also a dual formulation. The key idea in dual formulation is to construct a Lagrange function from the primal function and the corresponding constraints, by introducing a dual set of variables. It can be shown that this function has a saddle point with respect to the primal and dual variables at the solution point. In the Support Vector expansion, $w$ can be completely described as a linear combination of the training patterns $x_i$. In a sense, the complexity of a function’s representation by support vectors (SVs) is independent of the dimensionality.
of the input space $X$, and depends only on the number of SVs.

The complete algorithm is defined in terms of dot products between the data. Computation of $b$ is done by exploiting the Karush-Kuhn-Tucker (KKT) Conditions. The KKT conditions state that at the point of the solution, the product between dual variables and constraints vanishes. The cases where the dual variables are non-zeros are called the support vectors. It is assumed that the training set is drawn from some i.i.d (independent and identically distributed) distribution $P(x,y)$. The goal is to find a function $f$ that minimizes the expected risk.

$$R[f] = \int c(x,y,f(x))dP(x,y)$$

(5.7)

Here $(c(x,y,f(x)))$ denotes a cost function. Since, the distribution $P(x,y)$ is not known, only data set $X$ can be used. A possible approximation is to replace the integration by the empirical estimate to get the empirical risk functional given by,

$$R_{emp}[f] = \frac{1}{l} \sum_{i=1}^{l} c(x_i, y_i, f(x_i))$$

(5.8)

To achieve good generalization properties, a control term is added to this.

$$R_{reg}[f] = R_{emp}[f] + \frac{\lambda}{2} ||w||^2$$

(5.9)

where $\lambda > 0$ is a regularization constant. The standard setting in the support vector regression is the $\varepsilon$-insensitive loss,

$$c(x,y,f(x,y)) = |y - f(x)|_\varepsilon$$

(5.10)

Minimizing (5.9) with the particular loss function of (5.10) is equivalent to minimizing (5.4). The cost function in (5.9) need to be chosen properly. A particular cost function
that suits the problem best and which is low complex should be chosen. In addition, the samples were assumed to be generated by an underlying functional dependency plus additive noise, i.e. \( y_i = f_{true}(x_i) + \xi_i \) with density \( p(\xi) \). The optimal cost function in a maximum likelihood sense is given by,

\[
c(x, y, f(x, y)) = -\log(p(y - f(x)))
\] (5.11)

The likelihood of an estimate \( X_f = ((x_1, f(x_1)), ..., (x_l, f(x_l))) \) for additive noise and i.i.d data is,

\[
p(X_f/X) = \Pi_{i=1}^l p(f(x_i)|(x_i, y_i)) = \Pi_{i=1}^l p(y_i - f(x_i))
\] (5.12)

Maximizing \( p(X_f/X) \) is equivalent to minimizing \(-\log p(X_f/X)\), and hence we get as,

\[
-\log p(X_f/X) = \sum_{i=1}^l c(x_i, y_i, f(x_i, y_i))
\] (5.13)

However, the cost function thus obtained might not be convex. In that case, it is necessary to find a convex approximation to deal with the situation efficiently. For the sake of simplicity it is assumed that \( c \) is symmetric and has discontinuities at \( \pm \varepsilon, \varepsilon \geq 0 \) in the first derivative, and to be zero in the interval \([ -\varepsilon, \varepsilon ]\). Hence \( c \) will be of the form \( c(x, y, f(x, y)) = \tilde{c}(|y - f(x)|_\varepsilon) \). This is similar to the Vapnik’s \( \varepsilon \)-insensitive loss. This special choice can be extended to more general convex cost functions. For nonzero cost functions in the interval \([ -\varepsilon, \varepsilon ]\), an additional pair of slack variables need to be assumed.
5.2 Large-MIMO Detection via SVR

In this section, the formulation of the detection problem in large-MIMO system using the support vector regression is presented. Basically, the problem of detection is viewed as a problem of estimating the weight vectors in this SVR algorithm. The intention is not to estimate the data but to get the weights, which is the transmitted vector in this case. Considering the following model,

\[
\begin{bmatrix}
\Re\{y_c\} \\
\Im\{y_c\}
\end{bmatrix}
\approx
\begin{bmatrix}
\Re\{H_c\} & -\Im\{H_c\} \\
\Im\{H_c\} & \Re\{H_c\}
\end{bmatrix}
\begin{bmatrix}
\Re\{x_c\} \\
\Im\{x_c\}
\end{bmatrix}
+ \begin{bmatrix}
\Re\{n_c\} \\
\Im\{n_c\}
\end{bmatrix}
\] (5.14)

where \(\Re\{\cdot\}\) and \(\Im\{\cdot\}\) represents real and imaginary operations respectively. The linear regression model for each symbol is written as,

\[y_i = h_i^T x + n_i\] (5.15)

where \(y_i\) is the symbol received at the \(i^{th}\) antenna, \(h_i^T\) is the \(i^{th}\) row of the channel matrix and \(x\) is the transmit symbol vector. This is because, in MIMO, each received symbol is a linear combination of the transmitted symbols. Alternatively, we can treat the received symbols to be drawn from an affine space whose basis are the columns of the respective channel matrix. Support vector regression performs linear regression using a \(\varepsilon\)-insensitive loss function and attempts to reduce the model complexity by minimizing norm square of the weight vector. The loss function is given by,

\[L_\varepsilon(y_i, h_i^T x) = \begin{cases} 
0 & \text{if } |y_i - h_i^T x| \leq \varepsilon \\
|y_i - h_i^T x| - \varepsilon & \text{otherwise}
\end{cases}\] (5.16)
The non-negative slack variables are introduced to measure the deviation of the samples outside the $\varepsilon$-insensitive region. The support vector regression formulation is written as,

$$\min \frac{1}{2} \|x\|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*)$$

$$\text{s.t.} \begin{cases}
h_i^T x - y_i \leq \varepsilon + \xi_i \\
y_i - h_i^T x \leq \varepsilon + \xi_i^* \\
\xi_i, \xi_i^* \geq 0, i = 1, ..., 2N_r
\end{cases}$$

If the data point falls within the zone, then both the slack variables takes the value zero. If it falls above the zone, then $\xi_i > 0$ and if it falls below the zone $\xi_i^* > 0$. It should be noted that only either of the slack variables are non-zero at any point. The quadratic term in the objective function helps to avoid over-fitting and renders better generalization capability. The above problem employs a linear kernel and hence can be solved in primal itself. There is no requirement to employ a non-linear kernel or bias functions in this problem. This is because, the detection problem is a identical to solve a linear system of equations with the data points equally spaced with a minimum Euclidean distance. Employing a non-linear kernel is not expected to improve the results significantly. Nevertheless, it increases the computational complexity rather.

Letting $z = \begin{bmatrix} x \\ \xi \\ \xi^* \end{bmatrix}$ of dimension $(2N_T + 4N_R) \times 1$ and the constraints rewritten as,

$$\begin{cases}
h_i^T x - \xi_i \leq \varepsilon + y_i \\
-h_i^T x - \xi_i^* \leq \varepsilon - y_i \\
\xi_i, \xi_i^* \geq 0, i = 1, ..., 2N_r
\end{cases}$$

The quadratic problem in standard form is given by,
\[
\min \frac{1}{2} z^T Q z + q^T z \quad (5.20)
\]

\[
\begin{align*}
\text{s.t.} \quad & \quad \quad \quad B_{ineq} z \leq b_{ineq} \\
& \quad \quad \quad \quad B_{eq} z = b_{eq} \\
& \quad \quad \quad \quad z_{LB} \leq z \leq z_{UB}
\end{align*}
\]

(5.21)

In this formulation, we set

\[
Q = \begin{bmatrix}
I_{2NT} & 0_{2NT \times 4NR} \\
0_{4NR \times 2NT} & 0_{4NR \times 4NR}
\end{bmatrix}
\]

(5.22)

thus forcing the quadratic term to be purely a \(l_2\)-norm of the vector \(z\). The vector \(q\) of dimension \((2NT + 4NR) \times 1\) consists of \(C\) in the last \(4NR\) positions. The inequality constraints in the formulation is defined by setting

\[
B_{ineq} = \begin{bmatrix}
H_{2NR \times 2NT} & -I_{2NR} & 0_{2NR} \\
-I_{2NR} & H_{2NR \times 2NT} & 0_{2NR}
\end{bmatrix};
\quad
b_{ineq} = \begin{bmatrix}
\varepsilon + y \\
\varepsilon + y
\end{bmatrix}_{4NR \times 1}
\]

(5.23)

The matrices characterizing the objective function and the inequality constraints are sparse and hence facilitates faster computation and lowers the computational complexity. Since there are no equality constraints in the detection problem, they are ignored in the formulation. The lower and upper bounds are suitably chosen according to the constellation mapping. For example, in the case of BPSK constellation, the lower and upper bounds are fixed as \(-1\) and \(+1\) respectively. The value of \(C\) and \(\varepsilon\) is empirically fixed to be 100 and 0.2 respectively. The formulation proposed in this section can be extended to higher order constellations. However, the regression model should be for multiclass instead of binary class. An alternative way to tackle this problem is to construct multiple binary support vector regression models and fuse the decision of these models to detect symbols of higher order constellations.
5.3 Results and discussions

In this section, the performance of the support vector regression (SVR) detector is compared with lattice reduction aided detector (LR-MMSE-SIC) in Zhou and Ma (2013a). The comparison is for various MIMO configurations employing BPSK constellation. The results are presented for both the uncorrelated and spatially correlated channels as well.

5.3.1 Performance in uncorrelated channels

![Figure 5.2: BER performance comparison of SVR & LR detectors in 16 × 16 MIMO with BPSK in uncorrelated channels.](image)

In Fig. 5.2 the BER performance of SVR detector is compared with LR-MMSE-SIC in 16 × 16 MIMO system. It is observed that SVR detector offers an SNR gain of 1.5 dB in low SNR regime and around 0.5 dB for high SNRs. The performance
in $32 \times 32$ system is depicted in Fig. 5.3. It is observed that the SNR gain offered by SVR detector over LR-MMSE-SIC detector is improved to 2.5 dB for SNR greater than 10 dB. Although the SVR detector shows better performance than LR-MMSE-SIC detectors, the gain is more prominent in $32 \times 32$ MIMO than $16 \times 16$ MIMO. The results indicate an increase in system dimension increases the gain offered by the SVR detector.

![Figure 5.3: BER performance comparison of SVR & LR detectors in $32 \times 32$ MIMO with BPSK in uncorrelated channels.](image)

5.3.2 Performance in correlated channels

In this section, the performance comparison in spatially correlated MIMO channels are discussed. Figs. 5.4 to 5.6 presents the performance in $16 \times 16$ MIMO systems. From Fig. 5.4 in low correlated scenarios, it is observed that the SVR detector offers a higher gain in the low to medium SNR region rather than high SNRs. The gain is around 2
dB in medium SNRs while the gain is around 1 dB for high SNRs. From Fig. 5.5 in medium correlated cases, the gain offered by SVR detector is increased and is around 3 to 4 dB in low to medium correlated scenarios and around 2 dB for high SNRs. In high correlated scenarios, the SVR detector consistently outperforms the LR-MMSE-SIC detector with a gain close to 1 dB for any value of SNR greater than 6 dB.

![Figure 5.4: BER performance comparison of SVR & LR detectors in 16 × 16 MIMO with BPSK in low correlated channels.](image)

In general, the performance of SVR detector starts deviating from the LR-MMSE-SIC curve at 4 dB for low ($p = 0.3$) and medium correlation ($p = 0.5$) and at 6 dB for high correlation ($p = 0.7$). The performance gain is minimum 1 dB for low and high correlation cases, while the gain is still higher for medium correlation. For medium correlation, the gain is around 3 - 4 dB for low SNR and 2 dB for high SNRs. In high correlated channels the LR-MMSE-SIC detector achieves a BER of $10^{-3}$ at 25 dB, while
Figure 5.5: BER performance comparison of SVR & LR detectors in $16 \times 16$ MIMO with BPSK in medium correlated channels.

Figure 5.6: BER performance comparison of SVR & LR detectors in $16 \times 16$ MIMO with BPSK in high correlated channels.
the SVR detector requires 1 dB less compared to LR-MMSE-SIC detectors.

Figs. 5.7 to 5.9 depict the performance in $32 \times 32$ MIMO system. In low correlation conditions, it is observed that after an SNR of 6 dB, the SVR detector starts to offer a gain which increases with increase in SNR. For instance, at BER of $10^{-3}$, the SNR gain offered by SVR detector over the LR-MMSE-SIC detectors is close to 3 dB. From Fig. 5.8, it is observed that the SVR detector shows similar performance in medium and low correlation conditions. The performance does not degrade in the case of medium correlation. Nevertheless, the LR-MMSE-SIC detector shows a degradation upto 2.5 dB. Hence the gain offered by the SVR detector over the LR-MMSE-SIC detector is increased in medium correlation channels.

![Graph](image.png)

Figure 5.7: BER performance comparison of SVR & LR detectors in $32 \times 32$ MIMO with BPSK in low correlated channels.

In channels with high correlation, the performance of SVR detector shows signifi-
Figure 5.8: BER performance comparison of SVR & LR detectors in $32 \times 32$ MIMO with BPSK in medium correlated channels.

Figure 5.9: BER performance comparison of SVR & LR detectors in $32 \times 32$ MIMO with BPSK in high correlated channels.
cant performance gain over LR-MMSE-SIC detector compared to the low and medium correlated channels. For instance, to achieve a BER of $10^{-3}$, SVR detector needs SNR of only 18 dB, whereas the LR-MMSE-SIC detector needs 23 dB. In addition, the SVR detector starts performing better than the LR-MMSE-SIC detector even at low SNR of nearly 5 dB.

Table 5.1: Performance and Average running Time (ART) comparison of the SVR & LR-MMSE-SIC detectors in 16 × 16 MIMO systems.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>For BER of $10^{-2}$, BPSK system, SNR in dB &amp; Average Running Time (ART) in sec</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p=0</td>
</tr>
<tr>
<td>LR-MMSE-SIC</td>
<td>SNR</td>
</tr>
<tr>
<td>13</td>
<td>5E-6</td>
</tr>
<tr>
<td>SVR</td>
<td>12.5</td>
</tr>
</tbody>
</table>

Table 5.2: BER Performance and Average running Time (ART) comparison of SVR & LR detectors in 32 × 32 MIMO systems.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>For BER of $10^{-3}$, BPSK system, SNR in dB &amp; Average Running Time (ART) in sec</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p=0</td>
</tr>
<tr>
<td>LR-MMSE-SIC</td>
<td>SNR</td>
</tr>
<tr>
<td>15</td>
<td>7.8E-6</td>
</tr>
<tr>
<td>SVR</td>
<td>12</td>
</tr>
</tbody>
</table>

The performance and average running time comparison of SVR and LR-MMSE-SIC detectors in 16 × 16 and 32 × 32 MIMO systems are presented in the Table 5.1 and 5.2 respectively. The major observations are as follows.

1. It is observed that the complexity of SVR detector is slightly higher than the LR-MMSE-SIC detectors. But this justified with the performance gain of the SVR detectors, which is superior to LR-MMSE-SIC detectors in all the correlation and SNR conditions.
2. The degradation in performance from uncorrelated to correlated channels follows similar pattern in $16 \times 16$ MIMO system. In $32 \times 32$ MIMO systems, the degradation rate is slower in SVR compared to LR-MMSE-SIC detectors.

3. The advantage of SVR detector is prominent in large-MIMO systems of order of 32 antennas compared to a less number. Since, the detection strategy rely on the learning capability of the support vector machine, more the number of data points better is the learning accuracy. This is the probable reason for better performance in increased system dimensions i.e. large-MIMO.

The results demonstrate the efficiency of SVR detector and its advantage over the LR detectors in spatially correlated channels. In addition, the linear kernel employed ensures a better generalization capability. The formulation of SVR is tailored with binary constellations. Although the formulation is specific for binary symbols (BPSK), it can be employed for 4-QAM constellation, as the real-valued representation of 4-QAM constellation is identical to BPSK constellation. With even higher constellations like 16-QAM and 64-QAM the constellations are likely to take more values and hence the SVR detector needs to be customized to handle those constellations better and which needs further investigations.

Although the SVR detector has an increased ART requirement, the overhead is around 10% which is compromised for an SNR gain of around 3 - 5 dB in large-MIMO systems, where performance is more at stake due to correlated channels. The increased ART required is because of the quadratic programming involved. However, the ART of
the SVR detector can be reduced by employing an appropriate solver for this purpose.

5.4 Summary

A support vector regression (SVR) detector for large-MIMO system has been presented and the simulation results substantiate the performance improvement of the SVR detector over LR aided detectors. The SVR detector guarantees a minimum gain of 2 dB in low SNR and 1 dB in high SNR regime. The performance improves as the size of MIMO increases, thus more suitable for large-MIMO systems. Further, the performance gain in correlated channels is also encouraging. The SVR detector guarantees a gain of 2 - 3 dB gain in $16 \times 16$ systems and around 4 - 5 dB gain in $32 \times 32$ systems. Corroborating the results, the SVR detector is a viable alternative for LR detectors in large-MIMO systems especially with BPSK constellations.

With the SVR detector in place, the receiver is expected to be less complex and more compact with reduced antenna spacing, and can accommodate more antennas in the fixed space for better capacity. The support vector regression is best suited for binary constellations, the possibility of extension to higher order needs to be exploited.