HHS Public Access

Author manuscript

Biomed Signal Process Control. Author manuscript; available in PMC 2018 March 01.

Published in final edited form as:

Biomed Signal Process Control. 2017 March ; 33: 392–399. doi:10.1016/j.bspc.2016.12.003.

Prostate cancer recognition based on mass spectrometry sensing data and data fingerprint recovery

Khalfalla Awedat[a], **Ikhlas Abdel-Qader**[b], and **James R. Springstead**[c]

[a]Computer Science, Pacific Luthran University, Tacoma, WA, USA

[b]Electrical and Computer Engineering, Western Michigan University, Kalamazoo, MI, USA

[c]Chemical and Paper Engineering, Western Michigan University, Kalamazoo, MI, USA

Abstract

The high dimensionality and noisy spectra of Mass Spectrometry (MS) data are two of the main challenges to achieving high accuracy recognition. The objective of this work is to produce an accurate prediction of class content by employing compressive sensing (CS). Not only can CS significantly reduce MS data dimensionality, but it will also allow for full reconstruction of original data. We are proposing a weighted mixing of L1- and L2-norms via a regularization term as a classifier within compressive sensing framework. Using performance measures such as OSR, PPV, NPV, Sen and Spec, we show that the L2-algorithm with regularization terms outperforms the L1-algorithm and Q5 under all applicable assumptions. We also aimed to use Block Sparse Bayesian Learning (BSBL) to reconstruct the MS data fingerprint which has also shown better performance results that those of L1-norm. These techniques were successfully applied to MS data to determine patient risk of prostate cancer by tracking Prostate-specific antigen (PSA) protein, and this analysis resulted in better performance when compared to currently used algorithms such as L1 minimization. This proposed work will be particularly useful in MS data reduction for assessing disease risk in patients and in future personalized medicine applications.

Keywords

Compressive sensing; Mass spectrometry; BSBL; MS-classification; Confusion matrix

1. Introduction

Mass Spectrometry (MS) is often used to identify and quantify protein peptides and has the potential to be clinically used to differentiate between healthy and diseased patient samples. It has gained significant importance over the past years and of paramount challenge is the fact that MS data comes with high dimensionality. Being of such high dimensionality, MS data classification is computationally complex. Data reduction algorithms will be of critical importance in medicine going forward, having extensive application in the areas of disease risk assessment and personalized medicine. Major efforts are focused on improving classification while reducing computation [1]. Many algorithms have been proposed to

Correspondence to: Khalfalla Awedat.

classify MS data. In some methods the classification utilizes the whole MS data where all peak intensities are considered. In other studies, [2,3], the linear discriminant analysis (LDA) and continuous wavelet (CWT) space have been used for MS classification. Furthermore, the Q5 algorithm has also been proposed for the probabilistic classification of a serum sample using mass spectrometry [4]. They enforced a dimensionality reduction via PCA, projecting the spectra-space into a lower dimension, where the cross class variance is maximized. Then, LDA is applied to classify the projecting data. Other Partial features are candidates for classification where some peaks or ranges of spectra, such as alignments or filters, are excluded during the preprocessing procedures. Guyon et. al [5] propose a Recursive Feature Elimination (SVMRFE) algorithm that selects important genes/ biomarkers for the classification of noisy data. The sparse proteomics analysis (SPA) is another way to complete feature selection based on the compressive sensing concept [6]. Sparse features are a small subset of features that can be used to accurately predict unknown proteomic data. Huang et. al propose sparse signal representation to be used for classification among multiple linear regressions [7]. In using this method, the test sample is linearly represented of all training samples. Coefficients entries are all zeros except for those associated with a particular class or category.

In this paper, we used regularization of least squares with L1 and L2-norm methods to recover and classify within data sparse representation. Furthermore, we verified our proposed method using a prostate cancer database. Finally, accuracy and precision of our results were compared to those using the L2-norm method or the Q5 method.

2. Material and methods

2.1. Compressive sensing framework

In compressive sensing, most efforts target an optimum solution for the linear system equation

$$y=\phi x \quad (1)$$

where $x \in \mathbb{R}^N$ is a sparse signal, $\boldsymbol{\phi} \in \mathbb{R}^{dxN}$ is the measurement or sensing matrix, $y \in \mathbb{R}^d$ is a measurement vector, and d is the number of measurements retained from the original length N. Choosing $\boldsymbol{d} \ll \mathrm{N}$ immediately gives a compressed measurement vector y of length $\boldsymbol{d}$ instead of N. The $\boldsymbol{\phi}$ rows are incoherent and the columns are linearly independent [8]. The encoding phase is non-adaptive and does not need analysis in order to find the final encoding. Retained measurements $\boldsymbol{d}$ should always satisfy:

$$\mathrm{d} \geq \mathrm{C_o} klog(\mathrm{N}) \quad (2)$$

where C_0 is a constant and is the k number of non-zero entries in x. Therefore, CS is based on the assumption of a severely undersampled signal but reconstruction is secured using methods of convex optimization [9], as given in Eq. (3).

$$min\|x\|_1 \ \ subject\,to \quad \phi x = y \quad (3)$$

2.2. CS-based MS classification

MS data has very high dimensionality and the classification process is computationally expensive. A main objective of this study is to propose an accurate MS data classifier while reducing dimensionality. By modeling the MS data using CS technique, the sensing data does not only include lower dimensionality than the original data, but also the original information is preserved. This will allow us to go through the classification process with lower data dimensionality, leading to faster processes without losing classification accuracy. We are particularly focused on producing optimal and robust solutions from MS data where the following assumptions are considered:

1. The MS data is noisy
2. The collected data (MS sample) is of a high dimension [typically 10^5 to 10^8].
3. The number of samples in the database is relatively small [typically 10^2 to 10^4].

Each sample is represented by a vector pair $\{m/z, I\} \in \mathbb{R}^N$ where m/z is the mass to charge ratio $x_i = \{I_{i,1}, I_{i,2}, \ldots\ldots I_{i,ni}\} \in \mathbb{R}^{Nxni}$ and I is the spectral intensity. Then we stack ni columns of i^{th} class as $x_i = \{I_{i,1}, I_{i,2}, \ldots\ldots I_{i,ni}\} \in \mathbb{R}^{Nxni}$. Then the training set containing the n samples belonging to K classes can be represented as $X = [x_1, x_2, \ldots\ldots x_K] \in \mathbb{R}^{Nxn}$, thus $n = \sum_{i=1}^{K} n_i$. In sparse representation, any test sample, $x \in \mathbb{R}^N$, can be represented as a linear combination of the entire training samples [10].

$$x = Xr, \ \ x \in \mathbb{R}^N \quad (4)$$

where $r \in \mathbb{R}^n$ represents the coefficient vector that needs to be estimated. When $N < n$, the system is an underdetermined and would have an infinite number of solutions leading to a non-unique r. While the sparsest solution can be found using L1 norm, others chose to use nonlinear methods to find the nearest solution, such as convex optimization [8] and Newton methods [11]. It is proposed to reduce original high dimensionality of the data much using a sensing matrix and taking advantage of CS framework as also utilized by Liu et. al [12]. Instead of dealing with the X matrix, our MS data set, a new sensing data is generated by

$$y = \phi x = \phi X r = Y r \quad (5)$$

where $Y = [y_1, y_2, \ldots. y_K] \in R^{dxn}$ and $\boldsymbol{\phi} \in \mathbb{R}^{dxn}$ is the transformation matrix($\mathbb{R}^N \rightarrow \mathbb{R}^d$). In general, ***d*** has to be much smaller ***than N***, to satisfy the underdetermined condition. Due to high dimensionality of MS features and especially in comparison with the number database

samples, we still have an overdetermined system. In contrast to the other study and their proposed solution via L1 [12], it is possible for us to estimate r using L2 norm by solving:

$$\underset{r\in\mathbb{R}^n}{argmin}\|y-Yr\|_2^2 \quad (6)$$

However, to overcome the limitation of L1 and L2 overfitting, the regularized regression method that linearly combines the L1 and L2 penalties has been suggested by Zou et. al [13]. Therefore, Eq. (6) is replaced with Eq. (7) in our solution:

$$argmin\|y-Yr\|_2^2+\lambda_1\|r\|_1+\lambda_2\|r\|_2^2 \quad (7)$$

where the term $\lambda_1 r_1+\lambda_2 r_2^2$ is known as the Elastic net penalty, and both of the trade-off parameters λ_1 and$\lambda_2 \geq 0$. Both represent the compromise between model complexity and results accuracy. Eq. (7) is equivalent to the optimization problem:

$$\underset{r}{argmin}\|y-Yr\|_2^2\, s.t\, \lambda_1\|r\|_1+\lambda_2\|r\|_2^2 \quad (8)$$

Once r coefficients are estimated, the identity of test sample y can be determined based on how well the coefficients from each category are assigned to the object by calculating the residuals between the sensing test sample and all categories. The class is assigned based on minimum residual as:

$$\min_i r_i(y)=\|y-Y_i\delta r_i\|_2 i=1,2,....K \quad (9)$$

where δr_i is the regularization subvector coefficient of class i with dimension n_i consisting of components of r and Y_i is a dxn_i submatrix of Y, both corresponding to the class i samples [14]. The procedure for this proposed work is shown in Fig. 1.

3. MS recovery using CS framework

Although MS data is not naturally sparse, the difference between any two samples can be assumed as relatively sparse [12]. In Fig. 2, we show two diseased samples (D1 and D2) along with a healthy sample (C1) all taken from prostate cancer MS dataset for tracking PSA. This database is routinely used in assessment of patient prostate cancer risk [15]. We also show the difference between samples from patients with prostate cancer (D1–D2) is a sparse signal while the difference between the samples from healthy patients and patients with prostate cancer (C1–D2) is much less sparse. Consequently, we can use sparsity for reconstruction of the sample to its original size if needed, such as when abnormalities necessitate further analysis of original data.

Using regulated L2 classification results to identify the nearest sample y^* to a test sample y_t, we can create the fingerprint signal as:

$$y_t = \phi x_t + \varepsilon \& y^* = \phi x^* + \varepsilon \quad (10)$$

$$y_{FP} = y_t - y^\star = \phi\,(x_t - x^*) + \varepsilon = \phi x_{FP} + \varepsilon \quad (11)$$

where $y_{FP} \in \mathbb{R}^M$ is the measurements vector which has been taken from the original signal fingerprint $\boldsymbol{x_{FP}}$ (that is D1–D2). x_{FP} is in a high dimensional space and can be recovered using compressed data y_{FP} while ε is noise. The L1 minimization has been used to recover the data fingerprint in the study by Liu et. al [12]. However, due to MS instrument error, which causes a shift of features around m/z locations [16], the number of non-zero coefficients in the diseased sample fingerprint will still be large. Therefore, we propose using a technique that will have better results using less sparse data.

Recent algorithms such as Block Sparse Bayesian Learning (BSBL) were proposed as new methods to compress/reconstruct a non-sparse data set [17]. The BSBL family has been applied for non-sparse signals such as EEG and ECG by exploiting the intracorrelation of a block itself. In this paper, the BSBL is proposed to reconstruct the fingerprint MS data. The BSBL framework exploits the temporal correlation to improve its performance with the first basic model assumption:

$$x_{FP} = [x_1^T \underbrace{x_1, x_2 \ldots\ldots x_{\mathrm{bi}}}_{x_g^T} \ldots\ldots\ldots, \underbrace{x_{g-1}+1, \ldots\ldots x_{bg}}_{x_g^T}]^T \quad (12)$$

where x_{FP} can be viewed as non-overlapping blocks $\left\{x_1^T, \ldots\ldots\ldots x_g^T\right\}$, b is the size of the blocks and $b_i\,(\forall i)$ are not necessarily identical, with the locations of non-zero blocks unknown. The second assumption is to model the noise signal ε, for sparse Bayesian framework, as a Gaussian distribution $p(\varepsilon, \sigma^2) = N(0, \sigma^2)$ with zero-mean and a variance of σ^2. One advantage of the multivariate normal distribution stems from the fact that it is mathematically tractable and quality results can be and have been obtained [18,19]. We propose to apply Bayesian inference on MS signal fingerprint

Each block $x_i \in \mathbb{R}^{bix1}$, $i = 1, 2, \ldots.g$ is assumed to satisfy a parameterized multivariate Gaussian distribution given by:

$$p(x_i; \alpha_i, B_i) \sim N(0, \alpha_i B_i) \quad (13)$$

where $\alpha_i\,(\forall i)$ are hyperparameters controlling the block-sparsity of x. Few blocks have a nonzero value, and $B_i \in \mathbb{R}^{dixdi}\,(\forall i)$ is a positive definite matrix which captures the

correlation structure among the blocks which are assumed to be mutually uncorrelated. The αB terms are represented in the diagonal matrix Σ_0:

$$\sum_0 = \begin{bmatrix} \alpha_i B_i & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \alpha_g B_g \end{bmatrix} \quad (14)$$

Having defined this matrix with Bayes' rule implementation, we can express the posterior distribution of all unknowns overweight by:

$$p(x_{FP}/y_{FP}, \sigma^2, \{\alpha_i B_i\}_{i=1}^g = N\left(m_{x_{FP}}, \sum_{x_{FP}}\right) \quad (15)$$

where the posterior covariance and mean are given by:

$$m_{x_{FP}} = \sigma^{-2} \sum_{x_{FP}} \phi^T y_{FP} \quad (16)$$

$$\sum_{x_{FP}} = \left(\frac{1}{\sigma^2}\phi^T\phi + \sum_0^{-1}\right)^{-1} \quad (17)$$

Once the hyper parameters $\sigma^2, \{\alpha_i B_i\}_{i=1}^g$ are estimated, usually carried out using the type-II Maximum Likelihood (ML) [20], we apply the maximum a posterior (MAP) to estimate x_{FP}. [21]. The estimated data $\hat{x}_{FP}$ can be reconstructed directly from the posterior mean:

$$\hat{x}_{FP} \triangleq \sigma^{-2} \sum_x \phi^T y_{FP} = \sum_0 \phi^T \left(\sigma^2 I + \phi \sum_0 \phi^T\right)^{-1} \quad (18)$$

The block sparsity is controlled by two parameters $\alpha_i' s$ and Σ_0; by setting $\alpha_k = 0$ for kth block, block k will be pruned and due to the presence of noise, α_k will never be zero. Thus a threshold δ will be used to prune out small α_k ($\forall i$). The smaller value of threshold means fewer $\alpha_i' s$ are pruned out, and thus few blocks will be zeros [22]. Once x_{FP} has been estimated, it can be used to reconstruct the original data as:

$$x = x^\star + x_{FP} \quad (19)$$

where x^* is the training sample that is the corresponding to y^* signal which is derived from the classification process.

4. Results and discussion

To verify our proposed framework, we used prostate cancer SELD–TOF mass spectra datasets from the NIH and FDA Clinical Proteomic Program (http://home.ccr.cancer.gov/ncifdapromics-/ppatterns.asp) [15]. Table 1 shows this data content with respect to Prostate Specific Antigen (PSA) levels. PSA is a protein produced by cells of the prostate gland, levels at which routinely are measured to assess patient risk of prostate cancer. PSA levels were measured in blood samples, and diagnosis is performed on these measurements as listed in Table 1.

Due to high variation of amplitude peaks, the normalization process is used to set the intensities to new values in the range [0,1]. For normalization purposes we used:

$$x_i^{normal} = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (20)$$

where max (x_i) and min (x_i) are the maximum and minimum intensity peaks.

4.1. Classification performance

The database matrix has been arranged with the dimension of (15200 × 237) and contains four categories. To assess the L2-norm method, one sample has been selected randomly as a test sample. The dimensionality reduction step has been applied using the sensing matrix ϕ with the number of rows (0.2*N where N is number of features). Each column contained random (0.125*N) entries equal to one, while other entries equal to zero [23]. The classifier can achieve the same accuracy even with a compression ratio less than 0.2. Importantly, original data can also be recovered in our system in cases where further analysis is critical. Therefore, in this work, 0.2 compression ratio has been selected in all classification and recovery procedures. In Fig. 3, the test sample chosen belongs to subject 2, so one can assign the test sample to the category, which can give the best approximation $minr_i(y)$.

Performance of this suggested scheme was also compared with two most effective classification algorithms PCA/LDA [24] and compressed sensing recognition (CSR) [12].

For performance assessment of the predictor, the confusion matrix has been used. Table 2 shows the result of applying the L2, the L2-Regulator, and PCA/LDA (Q5) algorithms for classification. The data set was divided with 60% (143 samples) of the data as a training set and the remaining data used for testing. Furthermore, to make sure that the algorithm performance is not dependent on a specific set, the results are obtained from the average of 10- fold cross-validation. For L2-Regularization, we let $\eta = \frac{\lambda_2}{(\lambda_1+\lambda_2)}$ and put $0 \le \eta \le 1$ as a condition to select those parameters. The regularization parameters (λ_1, λ_2) are selected randomly 150, 0.5 (η = 0.003322) respectively.

Overall Success Rate (OSR), sensitivity (Sen.), Positive Predictive Value (PPV), Negative Predictive Value (NPV) and Specificity (Spec.) have been calculated for all simulations. The results are listed in Table 3.

To assess the performance of L2 with regularization under a very small features selection or measurements for underdetermined matrix, the same number of features (M = 140) has been selected for L1, Q5, L2-norm and L2-regularization. Table 4 shows that the L1-algorithm has a better performance than Q5 and L2 without regularization terms; however, overall, the L2-regularization has the best performance. The average accuracies (OSR) are 0.9321, 0.5106, 0.9149 and 0.9621. Of critical importance, PPV is significantly higher when L2-regularization is applied, as it is of particular clinical importance to identify patients with high risk of prostate cancer.

A training data set is determined in order to make sure that the L2-norm algorithm is robust under a variety of conditions such as the size of training samples. Table 5 shows the L2-regulation is still able to achieve a high performance in all assessment parameters.

To show the validity of L2-regularization, we implemented the algorithm using different values of η based on λ_1 and λ_2 for the range of $0 \leq \eta \leq 1$. For $\eta = 0$, the *L2* penalty term will be removed while the *L1* term is null for $\eta = 1$. Fig. 4 shows the effect of the regulation algorithms *L1* and *L2* in the all assessed performance parameters. All parameters have been affected by L1 regulator term with different values of λ_2. However, the best performance for the all parameters based on parameter λ_1 was in the range $140 \leq \lambda_1 \leq 180$. Improved algorithm performance, as indicated by increased measures of OSR, Spec, and PPV, will allow for higher data throughput and accurate measurements in applications such as disease biomarker detection in a laboratory or clinical setting.

Overall, results imply that the proposed method has more practicability than Q5. One can conclude from the results that by adding the regularization terms, the recognition MS data performance are improved for classification analysis.

4.2. Recovery procedure

The features used for recognition will also be used to reconstruct the original MS data if it has been classified as a sample from a diseased patient by taking advantage of the MS sensing data that was used to recover the signal fingerprint. We use block-sparse model by dividing the signal difference (SD) into groups/blocks where just a few have non-zero elements. For evaluation purposes, the recovery error has been calculated as $\|x - x_R\|_2$ where x is the original data and x_R is the reconstruction data. The BSBL-BO (the Bound-Optimization Method) package [25] was applied to recover the data fingerprint. For simplicity, the size of all blocks was selected to be $b_1 = b_2 = \ldots b_g = 100$. To evaluate the effect of the pruning threshold (δ), we show in Table 6 the Normalized Mean Square Error's (NMSE) which have been calculated using $\|\mathrm{x} - x_R\|_2^2 / \|\mathrm{x}\|_2^2$ where x and x_R are original and recovered signals with respect to change δ values respectively.

We chose $\delta = 10^{-3}$ as a pruning threshold, and the number of measurements ratio M/N varies from 0.01 to 0.5. The reconstruction performance of x_{FP} has compared with performance of spectral projected gradient (SPGL1) [26,27]. Fig. 5 shows that the BSBL-BO algorithm has a lower recovery error in all ranges of compressing ratios. Fig. 6 provides an example of recovering data using L1-minimization and BSBL-BO using the same number of measurements and sensing matrix. As clearly demonstrated in these figures, the BSBL-

BO algorithm offers superior data recovery over L1-minimization, a currently preferred algorithm of choice for sparse signal recovery [25].

5. Conclusion

The high dimensionality of Mass Spectrometry (MS) data is the main challenge facing high accuracy sample classification, including clinical application. In this work, a new method has been presented for accurate classification by employing compressive sensing (CS). Not only can CS significantly reduce MS data dimensionality, but also will allow for the full reconstruction of original data. The classification framework is capable of solving an overdetermined system with significantly reduced dimensionality, and without any loss of accuracy. Classification is established using the dimensionally reduced MS data using L2-norm and mixed L2–L1-norms regularization.

A prostate cancer database has been used to validate the proposed method. Results demonstrate that L2-algorithm with regularization performed better than both the L1-algorithm and Q5 under all applicable conditions. Regularization terms were used as design parameters, and by selecting $0 \leq \eta \leq 1$, the algorithm performance was improved. In addition, a signal difference was used to sparsify the MS signals and implement a reconstruction scheme for any disease signal. Specifically, L1-minimization and BSBL algorithms were used to reconstruct MS data, and we have found that BSBL outperforms L1. Using this method, all patients with prostate cancer were predicted as high risk patients. These results demonstrate an improvement in algorithm performance for the analysis of complex MS data, very important in eventual utilization of MS data in determining patient risk for diseases, including prostate cancer.

Acknowledgments

This work was supported by National Institutes of Health Grant R15HL121770-01A1 (J.R.S.).

References

1. Conrad, Tim, et al. Sparse Proteomics Analysis-A compressed sensing-based approach for feature selection and classification of high-dimensional proteomics mass spectrometry data. 2015 arXiv preprint arXiv:1506.03620.
2. Hong, Yan-jun, et al. Discrimination analysis of mass spectrometry proteomics for ovarian cancer detection. Acta Pharmacol Sin. 2008; 29(10):1240–1246. [PubMed: 18817630]
3. Du, Pan, Kibbe Warren, A., Lin, Simon M. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. Bioinformatics. 2006; 22(17): 2059–2065. [PubMed: 16820428]
4. Lilien, Ryan H., Farid, Hany, Donald, Bruce R. Probabilistic disease classification of expression-dependent proteomic data from mass spectrometry of human serum. J Comput Biol. 2003; 10(6): 925–946. [PubMed: 14980018]
5. Guyon, Isabelle, et al. Gene selection for cancer classification using support vector machines. Mach Learn. 2016; 46(1–3):389–422.
6. Conrad T, Genzel M, Cvetkovic N, Wulkow N, Vybiral J, Kutyniok G, Schütte C. Sparse Proteomics Analysis-a compressed sensing-based approach for feature selection and classification of high-dimensional proteomics mass spectrometry data. 2015 eprint arXiv:1506.03620.
7. Huang, Ke, Aviyente, Selin. Sparse representation for signal classification. Advances in Neural Information Processing Systems. 2006

8. van den Berg, Ewout. Convex optimization for generalized sparse recovery. Vancouver: 2009.
9. Stephen, B., Lieven, V. Convex Optimization. 7. Cambridge University Press; 2007.
10. Liu, Ji-xin, Sun, Quan-Sen. Mass spectrum data processing based on compressed sensing recognition and sparse difference recovery, Fuzzy Systems and Knowledge Discovery (FSKD). 2012 9th International Conference On; IEEE; 2012.
11. Simonis, Joseph P. Inexact Newton Methods Applied to Under–Determined Systems. Sandia National Laboratories; 2006.
12. Liu, Ji-xin, Sun, Quan-Sen. Mass spectrum data processing based on compressed sensing recognition and sparse difference recovery, Fuzzy Systems and Knowledge Discovery (FSKD). 2012 9th International Conference On; IEEE; 2012.
13. Zou H, Trevor H. Regularization and variable selection via the elastic net. J R Stat Soc B. 2005; 267(2):301–320.
14. Shi, Qinfeng, Eriksson, Anders, Van Den Hengel, Anton, Shen, Chunhua. Is face recognition really a compressive sensing problem? Computer Vision and Pattern Recognition (CVPR). 2011 IEEE Conference; 2011.
15. Petricoin, Emanuel F., et al. Serum proteomic patterns for detection of prostate cancer. J Natl Cancer Inst. 2002; 94(20):1576–1578. [PubMed: 12381711]
16. Hong, Yan-jun, et al. Discrimination analysis of mass spectrometry proteomics for ovarian cancer detection. Acta Pharmacol Sin. 2008; 29(10):1240–1246. [PubMed: 18817630]
17. Zhang Z, Rao Bhaskar D. Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning, Selected Topics in Signal Processing. IEEE Journal of. 2011; 5(5):912–926.
18. Tipping, Michael E. Sparse Bayesian learning and the relevance vector machine. J Mach Learn Res. 2001; 1:211–244.
19. Tipping ACFME, Faul A. Analysis of sparse Bayesian learning. Adv Neural Inf Process Syst. 2002; 14:383–389.
20. Tipping, Michael E. Sparse Bayesian learning and the relevance vector machine. J Mach Learn Res. 2001; 1:211–244.
21. Zhang Z, Rao B. Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation. Signal Process IEEE Trans. 2013; 61(8):2009–2015.
22. Wipf, David P., Rao, Bhaskar D., Nagarajan, Srikantan. Latent variable Bayesian models for promoting sparsity. Inf Theory IEEE Trans. 2011; 57(9):6236–6255.
23. Amini, Arash, Farokh, M. Deterministic construction of binary, bipolar, and ternary compressed sensing matrices. Inf Theory IEEE Trans. 2011; 57(4):2360–2370.
24. Lilien, Ryan H., Hany, F., Donald, Bruce R. Probabilistic disease classification of expression-dependent proteomic data from mass spectrometry of human serum. J Comput Biol. 2003; 10(6): 925–946. [PubMed: 14980018]
25. Zhang Z, Rao Bhaskar D. Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation. Signal Processing, IEEE Transactions. 2013; 61(8):2009–2015.
26. Van Den Berg, Ewout, Friedlander, Michael P. Probing the Pareto frontier for basis pursuit solutions. SIAM Journal on Scientific Computing. 2008; 31:890–912.
27. Friedlander E, van den Berg MP. {SPGL1}: A solver for large-scale sparse reconstruction. Jun. 2007 [Online].

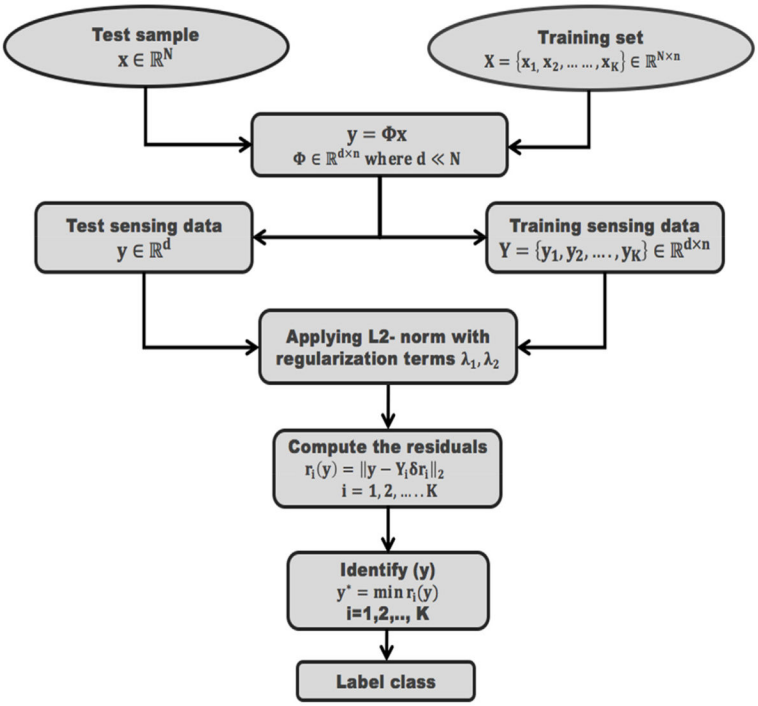

Fig. 1.
An overview of the proposed framework for MS data classification.

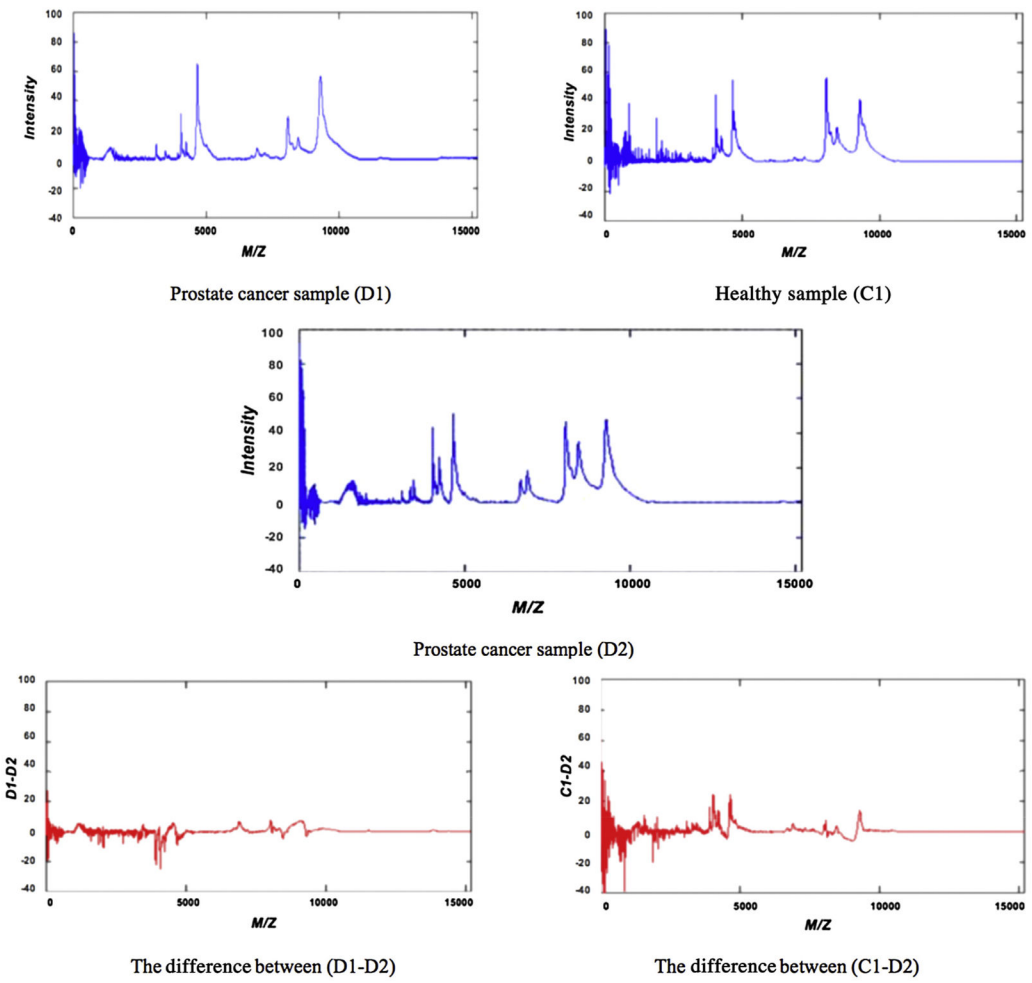

Fig. 2.

The difference between two diseased samples and a diseased sample with a healthy sample in prostate cancer MS dataset.

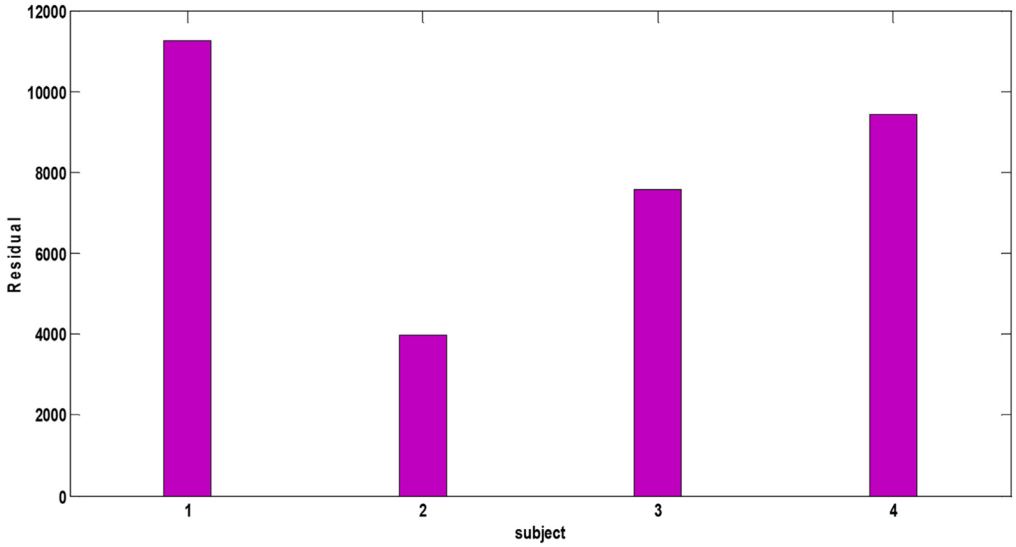

Fig. 3.

Histogram showing residuals $r_i(y)$ of the test sample with respect to the projection of sparse representation computed δr_i by L2-norm.

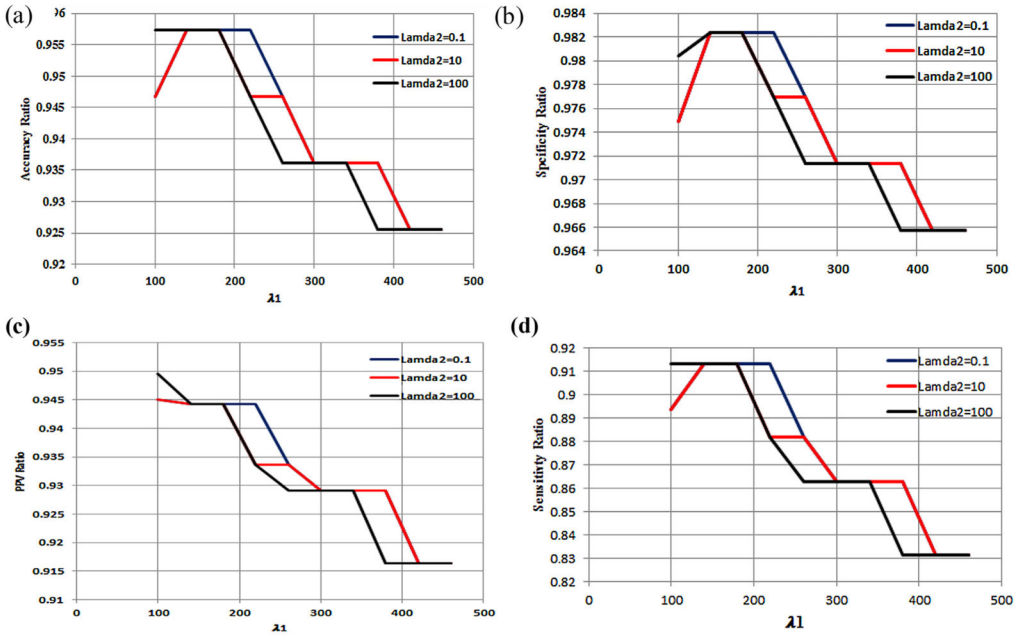

Fig. 4.

The regulation parameters versus the performance parameters (a) accuracy, (b) Specificity, (c) PPV and (d) Sensitivity.

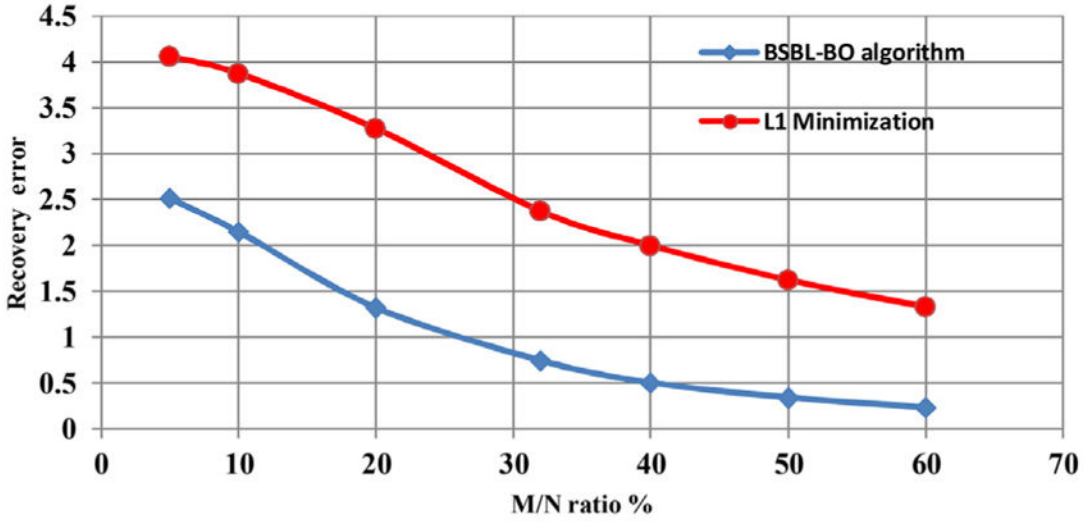

Fig. 5.

The average recovery error of L1-minimization and BSBL-BO under different measurement rates.

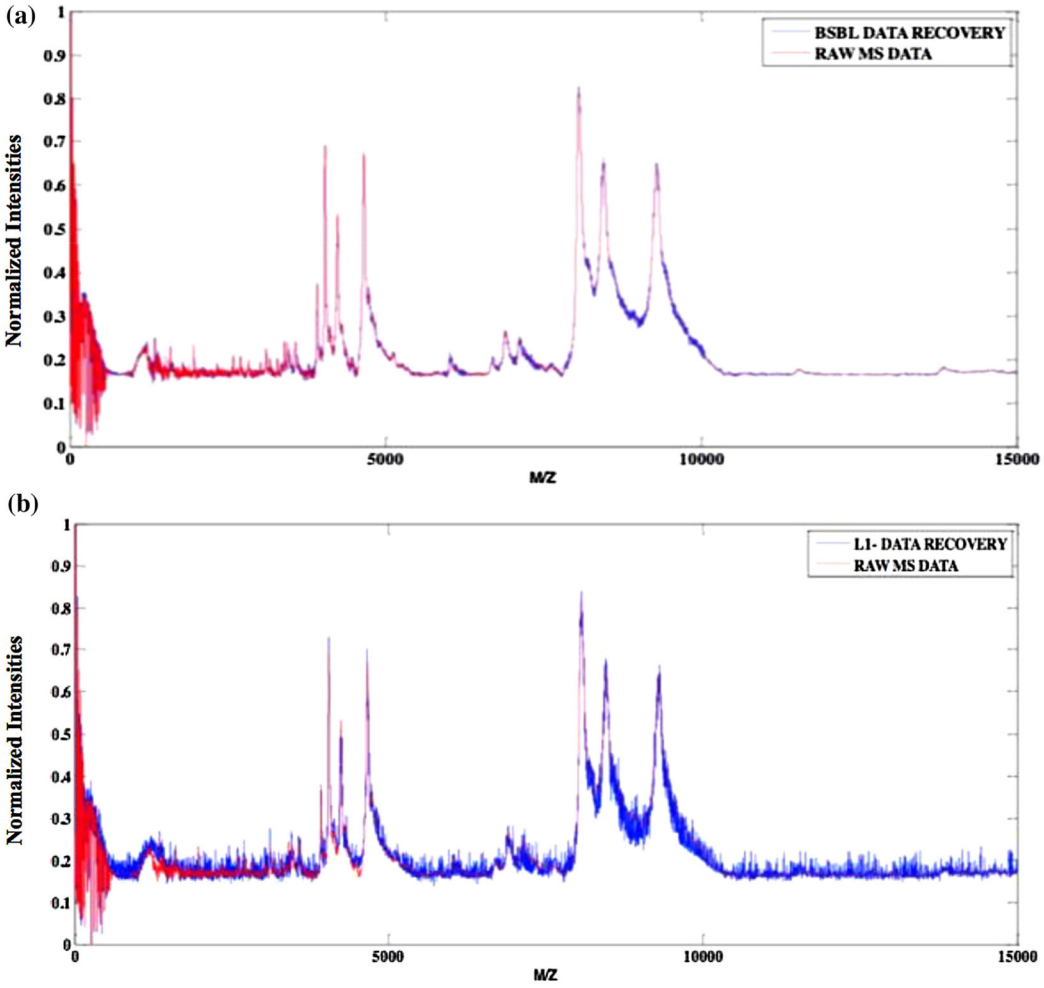

Fig. 6.

An example of recovering an MS data sample using two scenarios: (a) BSBL-BO technique and (b) L1-minminization.

Table 1

The Database of Prostate Cancer According to PSA Level.

Disease status	N
No evidence of disease and PSA level < 1 ng/mL [CLASS A]	60
Prostate cancer with PSA level > 10 ng/mL [CLASS B]	34
Benign and PSA level > [CLASS C]	120
Prostate cancer with PSA level 4–10 ng/mL [CLASS D]	23
Total	237

Table 2

The Confusion Matrix for Four Classes in L2, L2-Regulation and Q5 Algorithms.

Known Class	PCA/LDA A	B	C	D	L2- algorithm A	B	C	D	L2- Regularization $\lambda1=150, \lambda2=0.5$ A	B	C	D
CLASS A	20	1	3	0	24	0	0	0	24	0	0	0
CLASS B	0	11	2	0	0	10	3	0	0	11	2	0
CLASS C	0	3	43	3	0	1	45	3	0	0	48	1
CLASS D	1	0	2	5	0	0	1	7	0	0	2	6

Table 3

The performance evaluation has been estimated according to confusion matrix values.

Known class	PCA/LDA				L2- algorithm				L2- Regularization λ1=150, λ2=0.5			
	Sen	Spec	PPV	NPV	Sen	Spec	PPV	NPV	Sen	Spec	PPV	NPV
CLASS A	1	1	1	0.94	1	1	1	1	1	1	1	1
CLASS B	0.98	0.93	0.87	0.97	0.97	0.83	0.77	0.96	1	1	0.84	0.97
CLASS C	0.97	0.64	0.84	0.86	0.93	0.93	0.92	0.91	0.91	0.92	0.98	0.97
CLASS D	0.98	0.57	0.5	0.96	0.96	0.7	0.87	0.98	0.98	0.75	0.85	0.98
OSR_AVG=0.906					OSR_AVG=0.9149				OSR_AVG=0.9468			

Table 4

Results of Classification Comparison between L1-algorithm, L2- algorithm and Q5.

ALGORITHM	M	TRAINING SET	TEST SET	OSR	Sen	Spec	PPV	NPV
L1-algorithm	140	143	94	0.9321	0.9775	0.6868	0.705	0.750
PCA/LDA	140	143	94	0.5106	0.8295	0.4717	0.4401	0.567
L2-norm	140	143	94	0.9149	0.9579	0.750	0.840	0.879
L2-Regularization	140	143	94	0.9621	0.9588	0.761	0.841	0.910

Biomed Signal Process Control. Author manuscript; available in PMC 2018 March 01.

Table 5

Results of Comparison between the Three Algorithms in Terms of OSR, SPEC, PPV and SEN with Different Training Set Percentages.

Training%	PCA/LDA					L2- NORM					L2- REGULARIZATION λ1 = 150, λ2 = 0.5				
	OSR	Sen	Spec	PPV	NPV	OSR	Sen	Spec	PPV	NPV	OSR	Sen	Spec	PPV	NPV
30%	0.820	0.707	0.950	0.730	0.756	0.890	0.847	0.964	0.816	0.786	0.930	0.865	0.978	0.860	0.890
60%	0.920	0.805	0.980	0.870	0.860	0.920	0.890	0.968	0.787	0.810	0.950	0.893	0.974	0.950	0.900
75%	0.850	0.788	0.950	0.750	0.780	0.880	0.837	0.956	0.820	0.831	0.950	0.893	0.974	0.950	0.932

Biomed Signal Process Control. Author manuscript; available in PMC 2018 March 01.

Table 6

The recovery performance under prune threshold parameter.

δ	0	10^{-4}	10^{-3}	10^{-2}	10^{-1}
$NMSE$ ($10\times^{-4}$)	9.9584	9.9584	7.7820	7.7820	7.7820