

Sparsity-Based Direct Location Estimation Based on Two-step Dictionary Learning

Tingting Wang^{1,2,3}, Wei Ke^{2,3}, Gang Liu²

¹The Jiangsu Engineering Center of Meteorological Sensor Network Technology,
Nanjing University of Information Science and Technology, Nanjing, China

²The Jiangsu Key Laboratory on Optoelectronic Technology, School of Physics and Technology,
Nanjing Norm University, Nanjing, China

³Key Laboratory of Disaster Reduction and Emergency Response Engineering of the
Ministry of Civil Affairs, Beijing, China

Email: wkykw@sina.com

Received May 2013

ABSTRACT

This paper proposes an adaptive sparsity-based direct position determination (DPD) approach to locate multiple targets in the case of time-varying channels. The novel feature of this method is to dynamically adjust both the overcomplete basis and the sparse solution based on a two-step dictionary learning (DL) framework. The method first performs supervised offline DL by using the quadratic programming approach, and then the dictionary is continuously updated in an incremental fashion to adapt to the time-varying channel during the online stage. Furthermore, the method does not need the number of emitters *a priori*. Simulation results demonstrate the performance of the proposed algorithm on the location estimation accuracy.

Keywords: Dictionary Learning; Compressive Sensing; Direct Location; Time-Varying Channel; Quadratic Programming

1. Introduction

Wireless localization as a fundamental task in various fields like communications, radar, sonar, seismology and radio astronomy has drawn increasing attention in the past decade. The traditional approach to solve the localization problem consists of two-step procedure. First, the signal parameters such as angle of arrival (AOA) and time of arrival (TOA) are estimated, and second the coordinates of unknown-location targets are calculated by exploiting the parameters estimated in the first step. Although most localization algorithms presented so far concentrate on the two-step method, it is suboptimal in general as explained in references [1]. Recently, a kind of novel DPD methods that directly estimate the results of indeed interest to the end-user, *i.e.*, position coordinates, have been proposed, as a promising positioning technique that is shown to outperform the conventional two-step methods [2]. Weiss *et al* proposed a unique DPD criterion gathering all signals of all stations firstly for a single emitter [3] and then proposed a decoupled approach to treat the multiple emitters case [4]. Since the measurement-to-association step is avoided, the decoupled DPD method can provide superior localization capability

in multi-emitter context. But this method depends on the initial position estimates, and the number of sources needs to be known *a priori*. Other studies have extended the DPD method to handle the global navigation satellite system [5,6]. Basically, the above DPD algorithms generate a set of support points in which the maximum likelihood cost function is evaluated. Therefore, these DPD methods have higher complexity than the two-step approach, which can exploit the explicit geometric relationship.

Compressed sensing (CS), which receives a great deal of attention in recent years, has been successfully applied in the two-step localization method by reducing the dimensions of measurement vectors [7,8]. However, there are few works discussing the CS pattern for more accurate DPD estimation. To the best of our knowledge, only Picard and Weiss treat the DPD problem as a spatial sparsity representation by exploiting the covariance-matrix fitting method in [9]. However, this work makes the premise that the predefined overcomplete basis (a.k.a. dictionary) is ideal and invariable in the localization process. In practice, due to the dynamical change of multipath channels and random noise, the predefined basis may mismatch the actual signals stochastically so that the estimation performance is degraded. In this paper, we pro-

pose an adaptive sparsity-based DPD (ASDPD) algorithm to dynamically adjust both the overcomplete basis and the sparse solution so that the solution can better match the actual scenario. The method first performs supervised offline dictionary training by using the quadratic programming approach. During the online stage, the dictionary is continuously updated in an incremental fashion to adapt to time-varying factors.

The notation used in this paper is according to the convention. Symbols for matrices (upper case) and vectors (lower case) are in boldface. $(\cdot)^H$, $\|\boldsymbol{\theta}\|_0$, $\|\boldsymbol{\theta}\|_1$, $\|\boldsymbol{\theta}\|_2$, \mathbf{I}_N , \otimes and CN denote conjugate transpose (Hermitian), l_0 norm, l_1 norm, l_2 norm, identity matrix with the dimension N , the Kronecker product and complex Gaussian distribution, respectively. For any matrix \mathbf{Y} , $\text{vec}(\mathbf{Y})$ is denoted as the vertical concatenation of the columns of \mathbf{Y} . Finally, $\hat{\mathbf{x}}$ denotes the estimate of the parameter of interest \mathbf{x} .

The remainder of the paper is organized as follows. Section 2 briefly describes the system model assumed throughout this paper and formulates as a sparse recovery problem. In Section 3, we introduce a scheme calibrating the overcomplete basis dynamically and estimating the sparse solution adaptively. Simulation results are given in Section 4. Finally, Section 5 concludes the paper.

2. System Model and Problem Formulation

Consider N base stations (BS) intercepting the narrow-band signals transmitted by L possible sources. Each BS which knows its coordinates is equipped with an antenna array consisting of M elements. Denote the l th unknown target position by the vector of coordinates \mathbf{p}_l . We use the far-field point-target model, which is commonly used for source localization due to its simplicity [3,4,9]. Based on this model, the received signal observed by the n th BS is given by

$$\mathbf{r}_n(t) = \sum_{l=1}^L \mathbf{a}_n(\mathbf{p}_l) s_l(t - \tau_n(\mathbf{p}_l)) + \mathbf{v}_n(t), 0 \leq t \leq T \quad (1)$$

where $s_l(t)$ is the signal waveform considered known. $\mathbf{a}_n(\mathbf{p}_l)$ is the array response at the n th BS from a signal transmitted position, and the propagation delay from the l th transmitter to the n th BS is given by $\tau_n(\mathbf{p}_l)$. The vector $\mathbf{r}_n = \mathbf{H}_n \boldsymbol{\theta} + \mathbf{v}_n, n \in \{1, \dots, N\}$ represents noise terms, which is assumed as the independent and identically distributed (i.i.d.) complex Gaussian process, uncorrelated with the signals.

We divide the area of interest into K grids. In general, $K \gg M > L$. Then, we formulate the location problem as a following CS problem

$$\mathbf{r}_n = \mathbf{H}_n \boldsymbol{\theta} + \mathbf{v}_n, n \in \{1, \dots, N\} \quad (2)$$

where $\mathbf{H}_n = [\mathbf{h}_1^{(n)}, \dots, \mathbf{h}_K^{(n)}]$ is an overcomplete basis matrix at the n th BS, and $\mathbf{h}_i^{(n)}$ corresponds to the noiseless

signal vector between the i th grid and the n th BS. $\boldsymbol{\theta} = [\theta_1, \dots, \theta_K]$ is a sparse vector that having in total L nonzero entries, where the indices of nonzero entries in $\boldsymbol{\theta}$ which represents the actual locations. It should be emphasized that the above matrix \mathbf{H}_n is constructed by ideal signals, where the parameters such as AOA and TOA can be calculated according to the geometric relationship directly. Denote \mathbf{H} the matrix obtained by concatenation of all the matrices $\mathbf{H}_n, i.e.,$

$$\mathbf{H} = [\mathbf{H}_1^T, \dots, \mathbf{H}_N^T]^T. \text{ Similarly, by denoting}$$

$$\mathbf{R} = [\mathbf{r}_1^T, \dots, \mathbf{r}_N^T]^T \text{ and } \mathbf{V} = [\mathbf{v}_1^T, \dots, \mathbf{v}_N^T]^T, \text{ we can obtain}$$

$$\mathbf{R} = \mathbf{H}\boldsymbol{\theta} + \mathbf{V} \quad (3)$$

Note that \mathbf{H} is known under the ideal channel condition, which means that we can estimate the actual coordinates of targets as long as we find the positions of nonzero values in $\boldsymbol{\theta}$. That is, the problem of localization is converted into one of sparse signal recovery from (3). Moreover, the number of these dominant nonzero values gives L .

However, the non-ideal factors are inevitable in a practical localization system. These factors include the channel attenuation, phase error, time-varying fluctuations of the radio channel and so forth. When these happen, the predefined dictionary cannot effectively express the actual signal, which will cause performance degradation in sparse recovery process.

For avoiding the difficulty of estimate all kinds of the time-varying factors, we assume the error dictionary matrix $\boldsymbol{\Gamma}$ which describe the difference between the predefined dictionary and the practical received signals. Note that the error matrix $\boldsymbol{\Gamma}$ is time-varying and cannot be known in advance. In this scenario, the sparse positioning model is correspondingly modified as:

$$\mathbf{R} = \boldsymbol{\Gamma}\mathbf{H}\boldsymbol{\theta} + \mathbf{V} \triangleq \mathbf{D}\boldsymbol{\theta} + \mathbf{V} \quad (4)$$

where $\mathbf{D} = \boldsymbol{\Gamma}\mathbf{H}$ denotes the actual overcomplete basis with the time-varying interference. To prevent \mathbf{D} from having arbitrarily large values (which would lead to arbitrarily small values of $\boldsymbol{\theta}$), it is common to constrain its columns $\mathbf{d}_1, \dots, \mathbf{d}_K$ to have a l_2 norm less than or equal to one. Obviously, the mismatch exists between the columns of \mathbf{D} and the corresponding columns of the predefined basis \mathbf{H} , and thus the performance degradation is inevitable in the sparse recovery process. Focused on this problem, an adaptive sparse recovery algorithm is proposed in this paper, which dynamically calibrate the overcomplete basis so that the sparse solution can better fit the actual scenario.

3. Sparse Representation Based on the Two-stage Dictionary Learning

The key feature of adaptive sparse recovery is the adap-

tive adjustment of the overcomplete basis. This process generally learns the uncertainty of the dictionary, which is not available from the prior knowledge, but rather has to be estimated using a given set of training samples. Several different DL algorithms have been presented recently [10]. However, these methods generally cannot effectively handle very large training sets or dynamic training data changing over time. To overcome these shortcomings, we propose a two-stage DL approach that can adapt to the varied upcoming samples.

So far, the most DL methods are generally based on alternating minimization. In one step, a sparse recovery algorithm finds sparse representations of the training samples with a fixed dictionary. In the other step, the dictionary is updated to decrease the average approximation error while the sparse coefficients remain fixed. The proposed method in this paper also uses this formulation of alternating minimization.

3.1. Sparse Recovery Phase

The above problem of noisy sparse signal recovery can then be converted into a following optimization problem

$$\min \|\mathbf{R} - \mathbf{D}\boldsymbol{\theta}\|_2^2 / 2 + \lambda \|\boldsymbol{\theta}\|_1 \quad (5)$$

where λ is the regularization parameter. However, it should be emphasized that larger coefficients in $\boldsymbol{\theta}$ are penalized more heavily in the l_1 norm than smaller coefficients, unlike the more democratic penalization of the l_0 norm [11]. In practice, large coefficients are usually the entries corresponding to the actual positions of targets, while small coefficients commonly represent the noise entries. The imbalance of the l_1 norm penalty will seriously influence the recovery accuracy, which may result in many false targets. Therefore, in this paper we choose the reweighted l_1 norm minimization algorithm in [11] as our sparse recovery method, which can overcome the mismatch between l_0 norm minimization and l_1 norm minimization while keeping the problem solvable with convex estimation tools.

3.2. Dictionary Learning Phase

In this paper, we propose a two-stage DL framework in which the offline DL method allows to train the dictionary in a supervised manner to integrate the large training sets, and the incremental DL method based on the results in the offline stage handles the unseen online variation to enhance its adaptability.

1) Offline dictionary learning

In this stage, the ideal overcomplete basis \mathbf{H} is optimized to better represent the data of the training sets. Since the sparse coefficients $\boldsymbol{\theta}$ are fixed in the DL stage, the resulting optimization problem becomes:

$$\min \|\mathbf{R} - \mathbf{D}\boldsymbol{\theta}\|_2^2 / 2, \quad s.t. \mathbf{d}_i^H \mathbf{d}_i \leq 1, i = 1, \dots, K \quad (6)$$

in which $\|\mathbf{R} - \mathbf{D}\boldsymbol{\theta}\|_2^2$ can be written as

$$\begin{aligned} \|\mathbf{R} - \mathbf{D}\boldsymbol{\theta}\|_2^2 &= \text{tr}[(\mathbf{R} - \mathbf{D}\boldsymbol{\theta})^H (\mathbf{R} - \mathbf{D}\boldsymbol{\theta})] \\ &= \text{tr}(\mathbf{D}\boldsymbol{\theta}\boldsymbol{\theta}^H \mathbf{D}^H) - 2\text{tr}(\mathbf{R}\boldsymbol{\theta}^H \mathbf{D}^H) + \text{tr}(\mathbf{R}\mathbf{R}^H) \\ &= \text{vec}(\mathbf{D}^H)^H (\mathbf{I} \otimes \boldsymbol{\theta}\boldsymbol{\theta}^H) \text{vec}(\mathbf{D}^H) \\ &\quad - 2\text{vec}(\boldsymbol{\theta}\mathbf{R}^H)^H \text{vec}(\mathbf{D}^H) + \text{tr}(\mathbf{R}\mathbf{R}^H) \end{aligned} \quad (7)$$

Let's introduce several new expressions for clarity of notation

$$\begin{aligned} \mathbf{a} &\triangleq \text{vec}(\mathbf{D}^H) \\ \mathbf{G} &\triangleq \mathbf{I} \otimes \boldsymbol{\theta}\boldsymbol{\theta}^H \\ \boldsymbol{\gamma} &\triangleq \text{vec}(\boldsymbol{\theta}\mathbf{R}^H) \end{aligned}$$

Omitting the terms that do not depend on \mathbf{D} , the objective function in (6) can be equivalent to

$$\min \frac{1}{2} \mathbf{a}^H \mathbf{G} \mathbf{a} - \boldsymbol{\gamma}^H \mathbf{a}, \quad s.t. \mathbf{d}_i^H \mathbf{d}_i \leq 1, i = 1, \dots, K \quad (8)$$

Note that (8) is a standard form of constrained quadratic programming problem which can be solved by any standard optimization method, such as the gradient projection algorithm in [12]. Moreover, the matrix \mathbf{G} is obviously a positively definite matrix, and thus (8) is convex function and can be guaranteed to find a global optimum [13] in this DL phase.

2) Online dictionary learning

Although the offline DL stage has adjust the overcomplete basis according the training data, it is impossible to be fit for all kinds of time-varying interference patterns. Moreover, its computation load is quite large for real-time localization. On the contrary, the online incremental learning is especially applicable when one seeks to find the variation in the sense that the time-varying channels pattern might not be specifically learned offline but can be distinguished from the past online observations. Based on the incremental learning pattern, the online learning algorithm in [14] can use the result of the offline DL stage as a warm restart for computing the next dictionary where the new samples will be fed into the online dictionary learning procedure, and thus a single iteration has empirically been found to be enough [14].

For completeness, a full description of the algorithm is given in Algorithm 1.

Algorithm 1	Two-stage DL algorithm
Initialization: set the training sample set; generate the ideal dictionary \mathbf{H}	
Offline DL stage:	
Input: the training sample set; $\hat{\mathbf{D}}_{\text{off}}^{(0)} = \mathbf{H}$; the number of iterations T ;	
for $j=1$ to T	
1) use the reweighted 1-norm algorithm to compute the vector	

$\hat{\theta}_{\text{off}}^{(j)}$ with $\hat{\mathbf{D}}_{\text{off}}^{(j-1)}$ fixed for each sample;

2) use the gradient projection algorithm to minimize the objective function in (8) with respect to $\hat{\mathbf{D}}_{\text{off}}^{(j)}$ keeping $\hat{\theta}_{\text{off}}^{(j)}$ fixed;

end for

Output: $\hat{\mathbf{D}}_{\text{off}} = \hat{\mathbf{D}}_{\text{off}}^{(T)}$; $\hat{\theta}_{\text{off}} = \hat{\theta}_{\text{off}}^{(T)}$;

Online DL stage:

Input: $\hat{\mathbf{D}}_{\text{on}}^{(0)} = \hat{\mathbf{D}}_{\text{off}}^{(T)}$; $\hat{\theta}_{\text{on}}^{(0)} = \hat{\theta}_{\text{off}}^{(T)}$; the latest observation samples;

1) update the pre-trained dictionary according to the latest observations using the online DL algorithm in [14] with $\hat{\mathbf{D}}_{\text{off}}$ as warm restart, and return the learned dictionary $\hat{\mathbf{D}}_{\text{on}}^{(1)}$;

2) use the reweighted 1-norm algorithm to minimize the objective function in (6) with respect to $\hat{\theta}_{\text{on}}^{(1)}$ keeping $\hat{\mathbf{D}}_{\text{on}}^{(1)}$ fixed;

3) $\hat{\mathbf{D}}_{\text{on}}^{(0)} = \hat{\mathbf{D}}_{\text{on}}^{(1)}$; $\hat{\theta}_{\text{on}}^{(0)} = \hat{\theta}_{\text{on}}^{(1)}$;

Output: $\hat{\mathbf{D}}_{\text{on}} = \hat{\mathbf{D}}_{\text{on}}^{(1)}$; $\hat{\theta}_{\text{on}} = \hat{\theta}_{\text{on}}^{(1)}$;

4. Simulation Results

In order to examine the performance of the proposed ASDPD method, we compare it with the decoupled DPD approach in [4] and covariance-based sparse DPD (CDPD) method [9]. Consider four BSs placed at the corners of a 1 km × 1 km square. Assume the number of grids in the location area is $K = 26 \times 26$, which means yielding a 40m resolution along both x and y axes. The carrier frequency of the simulated signal is assumed to be 900 MHz. Each BS is equipped with a uniform linear array of ten antenna elements with the adjacent elements spacing of half a wavelength. The locations of targets are selected at random, uniformly, within the square. All the simulation results are obtained based on 200 Monte Carlo realizations. In each simulation, we consider the following multipath channel model

$$c(\tau) = \sum_{i=1}^Q \beta_i \delta(\tau - \tau_i) \tag{9}$$

to obtain a set of channel data. $\{\beta_i\}$ is a set of independent and identically distributed (i.i.d.) random variables which satisfy $\beta_i \sim CN(0, e^{-b\tau_i})$. $b = 1/16$ is the exponential power delay profile and τ_i is the delay spread for the i th path [15].

As shown in **Figure 1**, the improvement in location accuracy for the proposed method can be seen in terms of the root mean square error (RMSE), when the number of emitters is two and the SNR is set to 5 dB. We can observe that the location performance of DPD and CDPD algorithms decreases evidently as the number of multipath increases. On the contrary, the variation of RMSE in the ASDPD algorithm is very small due to its adaptive ability through DL technique. This result reveals that our method is very robust to multipath channels and effectively enhances location accuracy.

Figure 2 illustrates the location error with respect to

the number of emitters when the SNR is set to 5 dB. Here, real lines describe the case of single-path channel for three algorithms, while dashed lines represent the case of three paths. With the increase in the number of emitters, the RMSE of DPD algorithm increases quickly due to the high sensitivity to the estimated number of targets. Note that the CDPD method does not rely on a good estimate of the number of emitters in the single-case, but its performance decreases evidently as the number of multipath increases. On the contrary, the ASDPD algorithm is very robust to two scenarios. The importance of the low sensitivity of our algorithm to the number of targets is twofold: first, the number of sources is usually unknown, and second low sensitivity provides robustness against mistakes in estimating the number of targets.

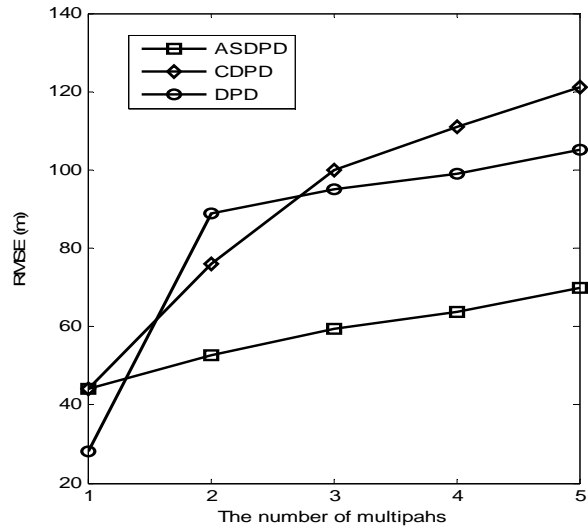


Figure 1. The localization error with respect to the number of multipaths.

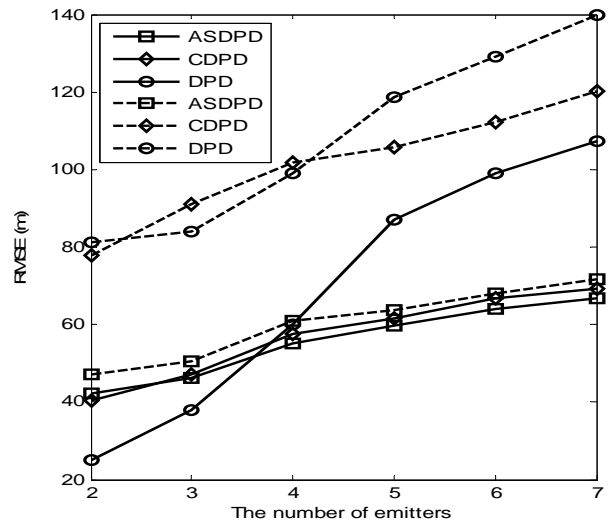


Figure 2. The localization error with respect to the number of emitters.

5. Conclusion

In this paper, we exploit the inherent spatial sparsity to present a novel direct location method by combining the offline training and online learning into a unified DL framework, thereby better matching time-varying scenarios. The effectiveness of the proposed scheme has been demonstrated by simulation results where substantial improvement for localization performance is achieved. Further research will emphasize on the off-grid error analysis and the theoretic bound on the location estimation precision.

6. Acknowledgements

This work is supported in part by the priority academic program development of Jiangsu higher education institutions, and in part by the Open Research Fund of Key Laboratory of Disaster Reduction and Emergency Response Engineering of the Ministry of Civil Affairs under grant No. LDRERE20120303.

REFERENCES

- [1] J. Bosse, A. Ferréol, C. Germond and P. Larzabal, "Passive Geolocalization of Radio Transmitters: Algorithm and Performance in Narrowband Context," *Signal Processing*, Vol. 92, No. 4, 2012, pp. 841-852. <http://dx.doi.org/10.1016/j.sigpro.2011.09.008>
- [2] A. Amar and A. J. Weiss, "New Asymptotic Results on Two Fundamental Approaches to Mobile Terminal Location," *Proceedings of 2008 ISCCSP*, pp. 1320-1323.
- [3] A. J. Weiss, "Direct Position Determination of Narrowband Radio Frequency Transmitters," *IEEE Signal Processing Letters*, Vol. 11, No. 5, 2004, pp. 513-516. <http://dx.doi.org/10.1109/LSP.2004.826501>
- [4] A. Amar and A. J. Weiss, "A Decoupled Algorithm for Geolocation of Multiple Emitters," *Signal Processing*, Vol. 87, No. 10, 2007, pp. 2348-2359. <http://dx.doi.org/10.1016/j.sigpro.2007.03.008>
- [5] P. Closas, C. Fernandez-Prades and J. Fernandez-Rubio, "Maximum Likelihood Estimation of Position in GNSS," *IEEE Signal Processing Letters*, Vol. 14, No. 5, 2007, pp. 359-362. <http://dx.doi.org/10.1109/LSP.2006.888360>
- [6] P. Closas, C. Fernandez-Prades and J. Fernandez-Rubio, "Cramér-Rao Bound Analysis of Position Approaches in GNSS Receivers," *IEEE Transactions on Signal Processing*, Vol. 57, No. 10, 2009, pp. 3775-3786. <http://dx.doi.org/10.1109/TSP.2009.2025083>
- [7] C. Feng, S. Valaee and Z. Tan, "Multiple Target Localization Using Compressive Sensing," *Proceedings of 2009 GLOBECOM*, pp. 1-6.
- [8] B. Zhang, X. Cheng, N. Zhang, Y. Cui, Y. Li and Q. Liang, "Sparse Target Counting and Localization in Sensor Networks Based on Compressive Sensing," *Proceedings of 2011 IEEE INFOCOM*, pp. 2255-2263.
- [9] J. S. Picard and A.J. Weiss, "Localization of Multiple Emitters by Spatial Sparsity Methods in the Presence of Fading Channels," *Proceedings of 2010 WPNC*, pp. 62-67.
- [10] R. Rubinstein, A. M. Bruckstein and M. Elad, "Dictionaries for Sparse Representation Modeling," *Proceedings of IEEE*, Vol. 98, No. 6, Jun. 2010, pp. 1045-1057. <http://dx.doi.org/10.1109/JPROC.2010.2040551>
- [11] E. J. Candes, M. B. Wakin and S. P. Boyd, "Enhancing Sparsity by Reweighted l_1 Minimization," *Journal of Fourier Analysis and Applications*, Vol. 14, No. 5-6, 2008, pp. 877-905. <http://dx.doi.org/10.1007/s00041-008-9045-x>
- [12] J. Nocedal and S. J. Wright, "Numerical Optimization," Springer Verlag, New York, 2006.
- [13] J. Dattorro, "Convex Optimization and Euclidean Distance Geometry," Meboo Publishing, Palo Alto, 2005.
- [14] J. Mairal, F. Bach, J. Ponce and G. Sapiro, "Online Learning for Matrix Factorization and Sparse Coding," *Journal of Machine Learning Research*, Vol. 11, No. 3, 2010, pp. 19-60.
- [15] M. R. Raghavendra and K. Giridhar, "Improving Channel Estimation in OFDM Systems for Sparse Multipath Channels," *IEEE Signal Processing Letters*, Vol. 12, No. 1, 2005, pp. 52-55. <http://dx.doi.org/10.1109/LSP.2004.839702>