L_0-Regularized Parametric Non-negative Factorization for Analyzing Composite Signals

Takumi Kobayashi, Kenji Watanabe and Nobuyuki Otsu
National Institute of Advanced Industrial Science and Technology
Umezono 1-1-1, Central 2, Tsukuba, Japan

***Abstract*—Signal sequences are practically observed as composites in which a few number of factor signals are linearly combined with non-negative weights. Based on prior physical knowledge about the target, the factors can be modeled as parametric functions, and their parameter values benefit further analyses. In this paper, we present a novel factorization method for the composite signals in terms of parametric factor functions. The method optimizes both the factor weights and the parameter values in the factor functions. While the parameter values are simply optimized by gradient descent, we propose L_0-regularized non-negative least squares (L_0-NNLS) for optimizing the factor weights. In L_0-NNLS, both L_0 regularization and non-negativity constraint are imposed on the weights in the least squares to enhance the sparsity as much as possible. Since so regularized least squares is NP-hard, we propose a stepwise forward/backward optimization to efficiently solve it in an approximated manner. Due to the sparsity by the L_0-NNLS, the proposed factorization method can automatically discover the inherent number of factor functions as well as the parametric functions themselves by estimating their parameter values. In the experiments on factorization of simulated signals and practical biological signals, the proposed method exhibits favorable performances.**

***Keywords*-Least squares, non-negativity, sparsity, L_0 regularization, factorization, parametric function**

I. Introduction

In practice, a signal is observed as a composite consisting of several component signals corresponding to the factors inherently contained in the observation target. It is an important process to factorize those composite signals into the components (factors) for analyzing the target; the physical characteristics of the target can be inspected based on the decomposed factors. For example, in the chemical and biological researches, energy dynamics signals such as of chemical particles are observed and then analyzed by fitting physical model functions [1], [2], [3].

Such a composite signal $s(t)$, which possesses an additivity property in nature, is modeled as a linear combination of factors denoted by functions $g_i(t)$: $s(t) \approx \sum_i^N w_i g_i(t)$, where w_i is the factor weight for $g_i(t)$ and N indicates the number of the factors. The task of the factorization problem is to estimate all of them; number of factors N, factor functions $g_i(t)$, and their weights w_i. By using the above linear additive model, the factorization is usually addressed to the least squares [4]. However, without any constraints, it produces unfavorably over-fitted results, affected by irrelevant noises, and thus most methods introduce several appropriate constraints to the factorization model as follows.

The popular method of the factorization is non-negative matrix factorization (NMF) [5]. The method imposes non-negativity constraints on both the weights and the factors. The non-negativity weights are considered to be natural from the physical viewpoint, since the weights w_i can be regarded as the significance and/or the existence (0/1) of factors, both of which are non-negative. Due to such non-negativity of the factor weights, we can easily interpret and analyze the obtained factors. In the NMF, however, the number of factors is required to be manually determined in advance, and only the fixed length (dimensional) signals, i.e., matrix, are dealt with. There are also variants of the NMF such as [6], [7].

On the other hand, sparsity in the factor weights has attracted keen attentions in the field of compressed sensing [8]. The sparsity is an effective criterion to retrieve the essential factors, recovering the signals in disregard of noises. The sparseness is measured by the number of non-zero components in the weights, namely, L_0 norm $||\boldsymbol{w}||_0$. However, it is difficult (NP-hard in general) to seek the global optimum that minimizes the L_0 norm, requiring all combinations of non-zero components to be checked. In the compressed sensing, given the factors, much research effort has been made to establish an efficient approach for finding the optimum factor weights with the minimum L_0 norm; e.g., L_1 relaxation [9], orthogonal matching pursuit (OMP) [10] and regularized OMP (ROMP) [11]. By enhancing the sparseness in the factor weights, the inherent number of factors in the signals would be automatically extracted.

In this paper, we propose a novel factorization method in terms of parametric factor functions. The proposed method comprises two kinds of optimizations for factor weights and parameter values in the factor functions. While the parameter values are optimized based on the gradient descent, we propose L_0-regularized non-negative least squares (L_0-NNLS) for optimizing the factor weights. In the L_0-NNLS, both the L_0 regularization and the non-negativity constraint are imposed on the weights in the least squares so as to enhance sparsity in the weights. Since so regularized least squares is NP-hard, we propose a stepwise forward/backward optimization to efficiently solve it in an approximated manner

978-0-7695-4607-0/11 $26.00 © 2011 IEEE
DOI 10.1109/ICMLA.2011.84

Algorithm 1 : Stepwise forward selection

Input: $\mathcal{I},\ s(t),\ g_i(t)\ (i=1,..,N)$.

```
Initialize w^[0] = 0, I^[0] = φ, J^[0] = inf, l = 1.
repeat
    r(t) = s(t) − Σ_i w_i^[l−1] g_i(t): residual function
    i* = arg max_{i∈I\I^[l−1]} { c_i^F ≜ ∫ r(t) g_i(t) dt / sqrt(∫ g_i(t)^2 dt) }
    I^[l] = I^[l−1] ∪ i*
    Solve (2) only for the weights of I^[l], and set w^[l].
    J^[l] = J(w^[l]), l ← l + 1
until J^[l] − J^[l−1] ⩽ 0
```

Output: $\boldsymbol{w}^F=\boldsymbol{w}^{[l-1]},\ J^F=J^{[l-1]}$

Algorithm 2 : Stepwise backward pruning

Input: $\mathcal{I},\ \boldsymbol{w}^{nnls},\ s(t),\ g_i(t)\ (i=1,..,N)$.

```
Initialize w^[0] = w^nnls, I^[0] = I, J^[0] = J(w^nnls), l = 1.
repeat
    ŝ(t) = Σ_i w_i^[l−1] g_i(t): projection onto the cone
    i* = arg min_{i∈I^[l−1]} { c_i^B ≜ | ∫ ŝ(t) w_i^[l−1] g_i(t) dt | }
    I^[l] = I^[l−1] \ i*
    Solve (2) only for the weights of I^[l], and set w^[l].
    J^[l] = J(w^[l]), l ← l + 1
until J^[l] − J^[l−1] ⩽ 0
```

Output: $\boldsymbol{w}^B=\boldsymbol{w}^{[l-1]},\ J^B=J^{[l-1]}$

from the geometrical viewpoint of the least squares. Due to the sparsity by the L_0-NNLS, the proposed factorization method automatically provides minimal (essential) number of parametric factor functions with sparse positive weights underling the target signals, together with optimizing the parameter values in those functions.

II. L_0-Regularized Nonnegative Least Squares

We give the general formulation of the proposed L_0 regularized non-negative least squares (L_0-NNLS), by using signal "functions", which is also applicable to the "vectors" of the fixed dimensionality.

Given the N factor functions denoted by $g_i(t), (i = 1,..,N)$, the task is to approximate the target function $s(t)$ by their linear combination with the weights w_i, which results in the well-known least squres:

$$\min_{\boldsymbol{w}} \int ||s(t)-\sum_i w_i g_i(t)||^2 dt. \quad (1)$$

Since any constraints such as non-negativity are not imposed on $\boldsymbol{w}$, this can be analytically solved by $\boldsymbol{w}=\boldsymbol{K}^{-1}\boldsymbol{y}$ where $K_{ij}=\int g_i(t)g_j(t)dt$ and $y_i=\int s(t)g_i(t)dt$.

As described in Section I, it is physically natural to impose non-negativity constraints on the weights w_i, which results in the following non-negative least squares (NNLS):

$$\min_{\boldsymbol{w}\geq 0} \int ||s(t)-\sum_i w_i g_i(t)||^2 dt. \quad (2)$$

From geometrical viewpoint, (2) is regarded as the problem to seek the projection of the target $s(t)$ onto the convex cone spanned by the factor functions $g_i(t)$ [12], and only the factors close to the target are assigned with positive weights, which results in sparse non-negative weights w_i [13]. (2) can be efficiently solved by applying the fast non-negative least squares method (fnnls) [14].

In this study, we further introduce the L_0 regularization into the above non-negative least squares (2) to obtain the sparser weights. Such sparse solutions enables us to avoid the interference by irrelevant noises and outliers, while extracting the intrinsic (essential) factors contained in the target signal [8]. The problem is finally defined as

$$\min_{\boldsymbol{w}\geq 0}\left\{ J(\boldsymbol{w}) \triangleq \int ||s(t)-\sum_i w_i g_i(t)||^2 dt + \lambda ||\boldsymbol{w}||_0 \right\}, \quad (3)$$

where $\lambda(>0)$ is the regularization parameter. Since this is an NP-hard problem, we render the way to efficiently solve it in an approximated manner based on the geometrical perspective of the least squares as follows.

A. Efficient Optimization Approach

As described above, the NNLS (2) produces the essential factors in the convex cone which are close to the input, and the L_0 norm of the solution $\boldsymbol{w}^*$ in (3) with $\lambda > 0$ is less than or equal to that of the solution $\boldsymbol{w}^{nnls}$ in (2) which corresponds to (3) with $\lambda=0$. Thus, by assuming the non-zero components of $\boldsymbol{w}^*$ are subset of those of $\boldsymbol{w}^{nnls}$, we solve (3) in the following approximated manner.

We first solve (2) by using fnnls [14] and obtain the initial weights $\boldsymbol{w}^{nnls}$. Then, both stepwise forward selection and backward pruning are applied to $\boldsymbol{w}^{nnls}$ for finding the optimum weight that minimizes J in (3). The selection and pruning are conducted based on the geometrical interpretations of the least squares. Their algorithms are shown in Algorithm 1&2.

Let the index set of the positive weights be denoted by $\mathcal{I}\triangleq\{i|w_i^{nnls}>0\}$.

1) Stepwise forward selection (Algorithm 1): This procedure is motivated by OMP [10]. We begin with empty factors (all weights are zero) and sequentially add the factor from the set $\mathcal{I}$ one-by-one. The selection criterion is the correlation c_i^F between the factor $g_i(t)$ and the residual $r(t)$; the factor of the highest correlation, i.e., the closest to the residual, would significantly reduce the squared error (Fig. 1(a)). At the l-th round, the optimization (3) only for the factors $\mathcal{I}^{[l]}$ results in (2) since the factors $\mathcal{I}^{[l]}$ $(\subset \mathcal{I})$ have all positive weights, producing $||\boldsymbol{w}||_0=l=const$. By adding factors, we seek the optimum factors of the minimum cost value J in (3).

2) Stepwise backward pruning (Algorithm 2): In contrast to the forward selection, we sequentially delete the factor from the whole set $\mathcal{I}$. We consider the contribution of each factor to the projection $\hat{s}(t)$ of the target $s(t)$ onto the

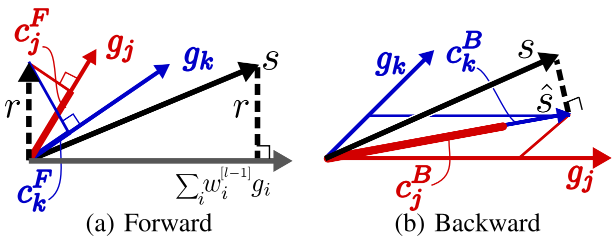


Figure 1. Geometrical illustration for the criterion of forward selection and backward pruning. In forward selection (a), the factor of the highest correlation c^F to the residual r is added, while in backward pruning (b), the factor of the lowest contribution (vector length) c^B to the projected function $\hat{s}$ is removed.

cone spanned by $g_i(t)$ $(i \in \mathcal{I}^{[l-1]})$. The contribution is measured as the (vector) length of the factor in the projection (Fig. 1(b)):

$$\frac{|\int \hat{s}(t) w_i g_i(t) dt|}{\sqrt{\int \hat{s}(t)^2 dt}} \propto |\int \hat{s}(t) w_i g_i(t) dt| \triangleq c_i^B. \quad (4)$$

We prune the factor of the lowest contribution. Such factor less affects the least squares and the error would not be significantly increased even by pruning it. Through the pruning, we seek the optimum factors of the minimum cost value J in (3).

After performing the above procedures independently in an arbitrary order, we choose the weight $\boldsymbol{w}^* \in \{\boldsymbol{w}^F, \boldsymbol{w}^B\}$ that produces the lower cost in those two results J^F, J^B. By searching the minimum cost in both the forward and backward directions, we can avoid the local minimum as much as possible.

3) Practical Issue: When searching the minimum cost in the above two approaches, it is not necessary to go through the whole factors $\mathcal{I}$, but we just look for the convex point by simply checking the differentials $J^{[l]} - J^{[l-1]}$. This is because the cost function would be convex along the number of the factors in almost all cases; in the forward (backward) approach, the increase (decrease) of the regularization cost is constant ($= \lambda$) while the decrease (increase) of the squared error would decline (grow).

The NNLS (2) is repeatedly solved for the small amount of factors in a sequential manner (line 6 in Algorithm 1&2). It requires quite a little computational cost since the non-negative solution is simply obtained by ordinary least squares for those factors in most cases.[1] In addition, such sequential fnnls [14] can be performed in an efficient manner by using 1-rank update for the inverse of the Gram matrix of factors [15].

B. Kernel-based Method

We have considered the linear relationship (inner-product) between functions. Along the recent advances in the kernel-based methods, we can also develop the method of kernel L_0-NNLS via kernel tricks [16]. By substituting the kernel $k(f, g)$ between the functions for the inner-product $\int f(t)g(t)dt$, the above-mentioned procedure is applied.

[1]The angles between the bases are less than 90 degrees.

III. L_0-Regularized Parametric Nonnegative Factorization

We consider the factorization problem using parametric factor functions. Given the family of the factor functions, the task is to estimate the parameter values in the factor functions as well as their weights w_i. Let the parameters be denoted by $\boldsymbol{\tau}_i \in \mathbb{R}^d$ in the i-th factor function, where d is the number of parameters. The factorization problem is defined by

$$\min_{\boldsymbol{w} \geq 0, \boldsymbol{\tau}} \int ||s(t) - \sum_i w_i g_i(t; \boldsymbol{\tau}_i)||^2 dt + \lambda ||\boldsymbol{w}||_0. \quad (5)$$

Assuming a sufficiently large number N of factor functions, some of them are automatically discovered with the positive weights as the essential factors describing the target signal.

In more practical situations, we have multiple target signals, such as for the case that many signal sequences are observed. And, the factor functions are shared by those signals, as in NMF [5], if the observation targets have all the same physical property. Then, the problem is formulated as

$$\min_{\boldsymbol{w} \geq 0, \boldsymbol{\tau}} \left\{ L \triangleq \sum_j \int ||s_j(t) - \sum_i w_{ij} g_i(t; \boldsymbol{\tau}_i)||^2 dt + \lambda \sum_j ||\boldsymbol{w}_j||_0 \right\},$$

where $s_j(t)$ is the j-th signal sequence and w_{ij} is the i-th factor weight for the j-th sequence.

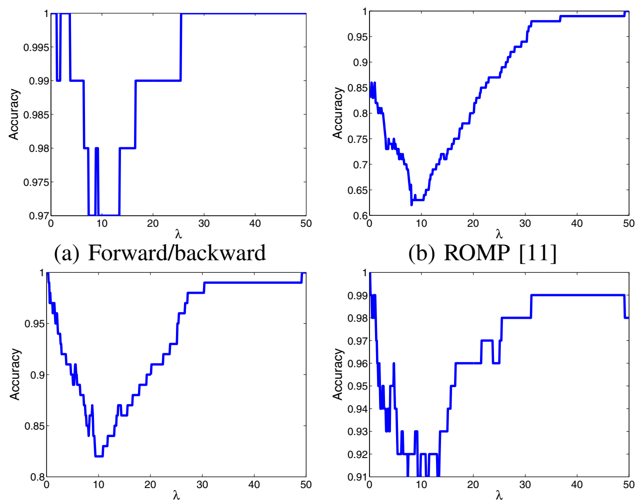
To solve this problem, we take an iterative approach to alternately optimize the factor weights $\boldsymbol{w}_j$ and the parameters $\boldsymbol{\tau}_i$. For the fixed parameter values, the weights are optimized by applying L_0-NNLS with the fixed factor functions to the respective signals. On the other hand, for the fixed weights, the parameter values in the factor functions are optimized based on the gradient descent method, such as conjugate gradient and Newton method [17]. The gradient of the objective cost L with respect to the parameters $\boldsymbol{\tau}_i$ is written by

$$\frac{\partial L}{\partial \boldsymbol{\tau}_i} = -\sum_j 2 \int \left\{ s_j(t) - \sum_i w_{ij} g_i(t; \boldsymbol{\tau}_i) \right\} w_{ij} \frac{\partial g_i(t; \boldsymbol{\tau}_i)}{\partial \boldsymbol{\tau}_i} dt.$$

IV. Experimental Results

A. Optimality in L_0-NNLS

We evaluate the optimality of the proposed stepwise forward/backward optimization (Section II-A) for L_0-NNLS by using the following synthetic data. We sampled 100-dimensional 10 factor vectors $\boldsymbol{G} \in \mathbb{R}^{100 \times 10}$ whose components are randomly drawn from uniform distribution $[0, 1]$, and set the weights for those factors to sparse random positive values $\boldsymbol{w} \in \mathbb{R}_+^{10}$ of which *five* components are randomly assigned with positive random values in $(0, 1]$. The target vector is constructed by $\boldsymbol{s} = \boldsymbol{G}\boldsymbol{w} + \mathcal{N}(\boldsymbol{0}, 0.01)$



(a) Forward/backward (b) ROMP [11]

(c) Forward selection (d) Backward pruning

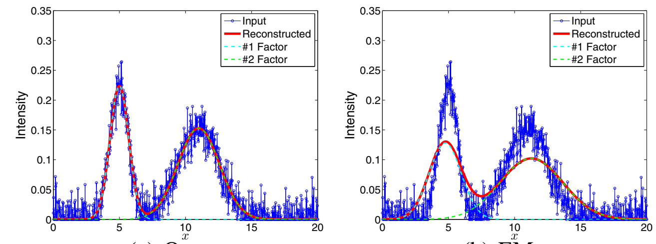
Figure 2. Average accuracy of various optimization approaches in L_0-NNLS with various λ compared to the global optimum on synthetic data. The result of the proposed forward/backward optimization is shown in (a).

where Gaussian noises with 0 mean and 0.01 standard deviation are added. We applied the stepwise forward/backward optimization for L_0-NNLS in (3) to the target vector $\boldsymbol{s}$ by using the given factors $\boldsymbol{G}$. For various parameter values λ, the obtained factor weights were compared to the global optimum ones in (3) by a greedy search which is feasible for such low-dimensional vectors, and then the average accuracy was measured on 100 trials. For comparison, we also applied the method of ROMP [11] which is slightly modified so as to deal with the non-negative least squares, and individually applied the forward selection and the backward pruning. The results are shown in Fig. 2, demonstrating that the proposed forward/backward optimization exhibits superior performances to the others; the accuracy of the proposed method averaged over λ is more than 99% (Fig. 2(a)).

B. Factorization of Simulated Signals

We apply the proposed factorization method using L_0-NNLS to the simulated signals. We drew the sample points from 1-dimensional two-modal Gaussian function with white noise as shown in Fig. 3 where negative values are shifted to zeros by assuming the spectrum data. We set $\lambda = 0.1$ and prepare 10 Gaussian-like factor functions $g_i(t;\mu_i,\tau_i) = \exp(-|t-\mu_i|^2/\tau_i)$, $i = 1,..,10$, whose initial values of μ_i and τ_i are randomly chosen from $[0, 20]$ and $[0, 5]$, respectively. The factorization results are shown in Fig. 3 compared to those by EM algorithm in which the number of components is set to two in advance. The proposed method successfully retrieved the inherent structures (correct number of factors with correct parameter values), while the EM was affected by noises, missing to discover them. It should be noted that by only setting the parameter λ the correct number of factors are automatically obtained in virtue of sparsity induced by L_0 regularization. Even for various parameter

	#1 Factor			#2 Factor		
	μ	σ^2	weight	μ	σ^2	weight
Input	5	0.5	0.4	11	2.5	0.6
Ours	5.00	0.50	0.40	11.00	2.47	0.60
EM	4.75	1.48	0.39	11.27	5.65	0.61



(a) Ours (b) EM

Figure 3. Factorization results of the simulated signal (blue line). Top table shows the estimated parameter values of Gaussian factors. The proposed method automatically discovers the correct factors with correct weights (a), while the EM method failed (b).

values $\lambda \sim 10$, the proposed method produces the same factorization results, which indicates the robustness to the parameter value λ.

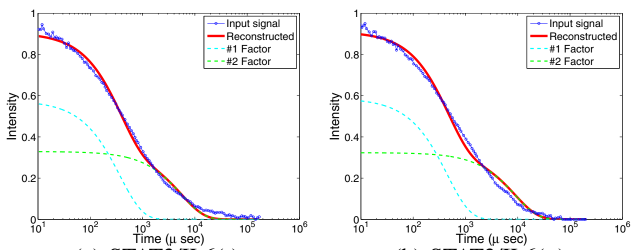
C. Factorization of Biological Signals

Finally, we apply the proposed factorization method to biological signals of functional proteins of signal transducers and activators of transcription 3 (STAT3). We measured the free diffusion of the functional proteins in living cell via fluorescence correlation spectroscopy (FCS) [18], [19], [20] and observed signals by applying the auto-correlations (refer to [2] for details). For analyzing the functional protein's dynamics, it is effective and important to automatically factorize the biological signals, although the free diffusions of proteins are inhibited such as by a protein-protein interaction, making it difficult to correctly decompose the observed signals. We observed 43 sequences before IL-6 stimulation (STAT3/IL6(-)) and 42 sequences after 15 minutes IL-6 stimulation (STAT3/IL6(+)). The number of the sampling time points in STAT3/IL6(-) and STAT3/IL6(+) were 113 ($10.4\mu s \leq t \leq 170393.6\mu s$) and 116 ($9.6\mu s \leq t \leq 196608.0\mu s$), respectively. The factors in this type of signals are physically modeled by the functions $g_i(t;\tau_i) = \exp(-t/\tau_i)$, where t denotes the time index and the parameter τ_i stands for the diffusion time. We prepare 10 factors with random initial parameter values τ_i ($i = 1,..,10$), and set the parameter $\lambda = 0.1$ as in Section IV-B.

Table I shows the parameter values (diffusion times) τ_i of the factors and their (averaged) positive weights w_i, i.e., existence ratio, estimated by the proposed method. The first factor of the dominant existence ratio is regarded as the primary component, since it significantly contributes to compose the signals. The diffusion time of the primary factor is estimated as $369.64\mu s$ in STAT3/IL6(-) and $451.26\mu s$ in STAT3/IL6(+) both of which are quite close to the ideal primary diffusion times, $373\mu s$ and $470\mu s$, computed based on the biological knowledge [2] and the conversion by using

Table I
EXPERIMENTAL RESULTS ON FACTORIZATION OF BIOLOGICAL SIGNALS BY USING THE PROPOSED METHOD

STAT3/IL6(-) primary diffusion time $\tau^* = 373\mu s$		STAT3/IL6(+) primary diffusion time $\tau^* = 470\mu s$	
[obtained 3 factors]		[obtained 2 factors]	
diffusion time τ	weight w	diffusion time τ	weight w
369.64 μs	0.53±0.05	**451.26** μs	0.52±0.04
5537.96 μs	0.39±0.05	8857.79 μs	0.40±0.05
76077.18 μs	0.10		



(a) STAT3/IL6(-) (b) STAT3/IL6(+)

Figure 4. Examples of factorization results. Blue dots show the observed signals, and red solid lines show the reconstructed signals from the obtained factors (cyan and green broken lines).

the Stokes-Einstein equation. Thus, we can say that the proposed method successfully extracts the primary factors from the noisy biological signals. The diffusion time of the second factor is extremely slower than that of the first one in both conditions. From the biological viewpoint [21], [22], [23], we can consider that the second factor captures the inhibition of free diffusion in STAT3 such as by DNA binding and protein-protein interactions. Note that the third factor in STAT3/IL6(-) is used in only one sequence.

Fig. 4 shows the examples of the obtained (weighted) factors as well as the reconstructed signals. The target signals are well reconstructed by using only several factor functions. Although these factors are inherently overlapped, the primary factor is successfully found with the largest weight.

V. CONCLUSION

We have proposed a novel factorization method in terms of parametric factor functions. In the proposed method, while the parameter values in the factor functions are simply optimized based on the gradient descent, we proposed L_0-regularized non-negative least squares (L_0-NNLS) for optimizing the factor weights. The L_0-NNLS produces sparse weights by imposing L_0 regularization and non-negativity constraint on the weights in the least squares. Since the greedy algorithm for it is NP-hard, we proposed the stepwise forward/backward optimization to efficiently solve it in an approximated manner from the geometrical viewpoint. Given sufficient number of factors, the proposed factorization method automatically provides minimal factors with sparse positive weights underling the target signals, together with optimizing the parameter values in the factor functions. In the experiments on factorization of simulated signals and practical biological signals, the proposed method favorably identified the underling factors.

REFERENCES

[1] L. Danos and T. Markvart, “Excitation energy transfer rate from langmuir-blodgett (lb) dye monolayers to silicon: effect of aggregate formation,” *Chemical Physics Letters*, vol. 490, no. 4-6, pp. 194–199, 2010.

[2] K. Watanabe, K. Saito, M. Kinjo, T. Matsuda, M. Tamura, S. Kon, T. Miyazaki, and T. Uede, “Molecular dynamics of stat3 on il-6 signaling pathway in living cells,” *Biochemical and Biophysical Research Communications*, vol. 324, no. 4, pp. 1264–1273, 2004.

[3] A. Kitamura, H. Kubota, C. Pack, G. Matsumoto, S. Hirayama, Y. Takahashi, H. Kimura, M. Kinjo, R. I. Morimoto, and K. Nagata, “Cytosolic chaperonin prevents polyglutamine toxicity with altering the aggregation state,” *Nature Cell Biology*, vol. 8, pp. 1163–1169, 2006.

[4] C. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[5] D. Lee and H. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, pp. 788–791, 1999.

[6] P. O. Hoyer, “Non-negative matrix factorization with sparseness constraints,” *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.

[7] Y. Wang, Y. Jia, C. Hu, and M. Turk, “Fisher non-negative matrix factorization for learning local features,” in *Asian Conference on Computer Vision*, 2004.

[8] D. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[9] E. J. Candès, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.

[10] Y. C. Pati, R. Rezaiifar, Y. C. P. R. Rezaiifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in *The 27th Annual Asilomar Conference on Signals, Systems, and Computers*, 1993, pp. 40–44.

[11] D. Needell and R. Vershynin, “Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 310–316, 2010.

[12] T. Kobayashi and N. Otsu, “Cone-restricted subspace methods,” in *International Conference on Pattern Recognition*, 2008.

[13] ——, “One-class label propagation using local cone based similarity,” in *International Conference on Face and Gesture*, 2011.

[14] R. Bro and S. Jong, “A fast non-negativity-constrained least squares algorithm,” *Journal of Chemometrics*, vol. 11, pp. 393–401, 1997.

[15] T. Kobayashi and N. Otsu, “Efficient reduction of support vectors in kernel-based methods,” in *International Conference on Image Processing*, 2009, pp. 2077–2080.

[16] B. Schölkopf and A. Smola, *Learning with Kernels*. MIT Press, 2001.

[17] J. Nocedal and S. Wright, *Numerical Optimization*. Springer, 1999.

[18] M. Ehrenberg and R. Rigler, “Rotational brownian motion and fluorescence intensify fluctuations,” *Chemical Physics*, vol. 4, no. 3, pp. 390–401, 1974.

[19] D. E. Koppel, “Statistical accuracy in fluorescence correlation spectroscopy,” *Physical Review A.*, vol. 10, no. 6, pp. 1938–1945, 1974.

[20] E. L. Elson and D. Magde, “Fluorescence correlation spectroscopy. i. conceptual basis and theory,” *Biopolymers*, vol. 13, no. 2, pp. 1–27, 2004.

[21] P. Heinrich, I. Behrmann, G. Müller-Newen, F. Schaper, and L. Graeve, “Interleukin-6-type cytokine signalling through the gp130/jak/stat pathway,” *Journal of Biochemistry*, vol. 334, pp. 297–314, 1998.

[22] J. Darnell Jr., “Stats and gene regulation,” *Science*, vol. 277, pp. 1630–1635, 1997.

[23] A. Pranada, S. Metz, A. Herrmann, P. Heinrich, and G. Muller-Newen, “Real time analysis of stat3 nucleocytoplasmic shuttling,” *Journal of Biological Chemistry*, vol. 279, no. 15, pp. 15 114–15 123, 2004.