L_P-REGULARIZED OPTIMIZATION BY USING ORTHANT-WISE APPROACH FOR INDUCING SPARSITY

Takumi Kobayashi

National Institute of Advanced Industrial Science and Technology
1-1-1 Umezono, Tsukuba, Japan

ABSTRACT

Sparsity induced in the optimized weights effectively works for factorization with robustness to noises and for classification with feature selection. For enhancing the sparsity, L_1 regularization is introduced into the objective cost function to be minimized. In general, however, L_p ($p<1$) regularization leads to more sparse solutions than L_1, though L_p regularized problem is difficult to be effectively optimized. In this paper, we propose a method to efficiently optimize the L_p regularized problem. The method reduces the L_p problem into L_1 regularized one via transforming target variables by the mapping based on L_p, and optimizes it by using orthant-wise approach. In the proposed method, the L_p problem is directly optimized for computational efficiency without reformulating it into iteratively reweighting scheme. The proposed method is generally applicable to various problems with L_p regularization, such as factorization and classification. In the experiments on the classification using logistic regression and factorization based on least squares, the proposed method produces favorable sparse results.

Index Terms— Optimization, L_p regularization, sparsity induced model, orthant-wise optimization

1. INTRODUCTION

Sparsity induced models have attracted keen attention in the fields of signal processing, such as for factorization [1], pattern classification [2] and computer vision [3]. The sparsity is crucial for retrieving essential factors from noisy signals, known as sparse coding [4] and compressed sensing [5], and for selecting essential (discriminative) features from a plenty of feature components [6].

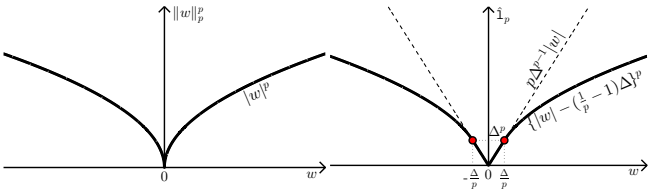
In general, the sparsity is induced via the regularization incorporated into the objective cost function to be minimized, such as reconstruction errors in factorization and classification errors. The sparseness is measured by the number of non-zero components in the target variable, e.g., weights $\boldsymbol{w}$, namely L_0 norm $\|\boldsymbol{w}\|_0$. However, it is difficult to find the global optimum that minimizes the L_0 norm since all combinations of nonzero components are required to be checked. In the compressed sensing [5], 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, such as orthogonal matching pursuit (OMP) [7] and regularized OMP (ROMP) [8]. It is also theoretically shown that there is the condition to obtain the L_0 optimum factors by solving the L_1 regularization [9]. Therefore, the sparsity is usually address in the form of L_1 regularization in most studies.

On the other hand, there is a work to optimize L_p-regularized problem ($0 \leq p \leq 1$) mainly based on least-squares for factorization [1, 10]. The method reformulates the L_p problem into the iteratively reweighted L_1 optimizations [11]. It, however, requires large computation cost and is applicable only to small-scale problems since the L_1 optimization computed at every iteration is generally time consuming. Nonetheless, the L_p regularization ($0 \leq p < 1$) is useful for leading to more sparse solutions than L_1. The property of L_p regularization is also theoretically investigated in the framework of compressed sensing [12].

In this paper, we propose a method to effectively optimize the L_p regularized problem. The proposed method reduces the L_p regularized problem into the L_1 problem via transforming the target variables by the mapping based on L_p norm. Then, the transformed L_1 problem is efficiently optimized by using the orthant-wise approach [13] which was recently proposed to optimize L_1 regularized logistic regression. Our contributions are; 1) we can directly optimize the L_p regularized problem for computational efficiency without reformulating it into iterative (exhaustive) L_1 optimizations unlike the other methods [1, 10]. And, 2) the L_p regularization ($p < 1$) that we address in this study contributes to favorably provide higher sparsity than L_1. 3) The proposed method is generally applicable to various tasks on not only factorization but also classification, with L_p regularization for inducing high sparsity.

2. PROPOSED METHOD

Let $\boldsymbol{w} \in \mathbb{R}^d$ be the d-dimensional target variables (weights), and $\mathtt{f}(\boldsymbol{w})$ be the objective cost function which depends on the task, such as classification and factorization. The opti-



(a) L_p ($p = 0.5$) (b) $\hat{\mathtt{l}}_p$ ($p = 0.5$)

Fig. 1. Regularization

mization problem that we address in this paper is generally formulated using L_p regularization by

$$\min_{\boldsymbol{w}} \mathtt{f}(\boldsymbol{w}) + \lambda \|\boldsymbol{w}\|_p^p, \tag{1}$$

where λ is the balancing parameter for the L_p regularization term. The L_p regularization term, $\|\boldsymbol{w}\|_p^p = \sum_{i=1}^d |w_i|^p$, induces sparsity of $\boldsymbol{w}$ for $0 \leq p \leq 1$; especially, L_1 is closely related to Laplacian priors [14], while L_2 is the commonly used regularization called weight decay [15]. Fig. 1a shows the form of such regularization of $p = 0.5$. The standard optimization approach such as gradient descent is not directly applicable since L_p regularized cost function ($0 \leq p \leq 1$) in (1) has discontinuous first derivatives as shown in Fig. 1a; meanwhile, only for $p = 1$, some effective optimization approaches have been proposed so far [14]. In this study, we first reduce (1) into L_1 regularized problem by transforming the variable $\boldsymbol{w}$ via the mapping based on L_p norm, and the resultant L_1 problem is effectively optimized by the orthant-wise approach [13].

2.1. Variable transformation based on L_p

For constructing smooth mapping of the variables, we first consider to slightly modify the L_p regularization term. Since the first derivative of L_p reaches $\pm\infty$ as $w_i \to \pm 0$, the regularization is relaxed into

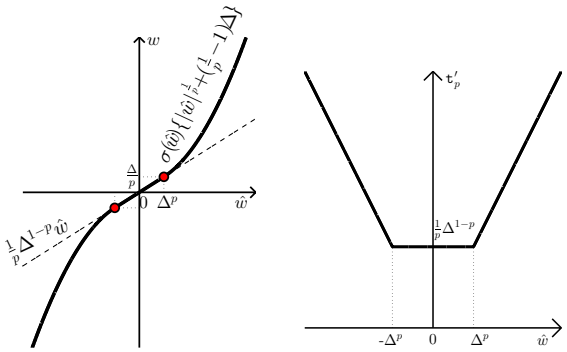
$$\hat{\mathtt{l}}_p(w_i) = \begin{cases} p\Delta^{p-1}|w_i| & (|w_i| \leq \frac{\Delta}{p}) \\ \left\{|w_i| - \left(\frac{1}{p} - 1\right)\Delta\right\}^p & (|w_i| > \frac{\Delta}{p}) \end{cases}, \tag{2}$$

where Δ is the small positive value, say $\Delta = 10^{-5}$ in this study, for the relaxation and the form of $\hat{\mathtt{l}}_p$ is shown in Fig. 1b. This is designed so as to linearly approximate p root function and to smoothly connect the linear and p root functions at $\frac{\Delta}{p}$. By using the function $\hat{\mathtt{l}}_p$, the optimization problem (1) is represented by $\min_{\boldsymbol{w}} \mathtt{f}(\boldsymbol{w}) + \lambda \sum_{i=1}^d \hat{\mathtt{l}}_p(w_i)$.

Next, we introduce the following 1-to-1 mapping of the variable w_i; let $\sigma(w) \triangleq \mathrm{sign}(w) \in \{-1, 0, 1\}$,

$$\hat{w}_i = \sigma(w_i)\hat{\mathtt{l}}_p(w_i) = \mathtt{t}_p^{-1}(w_i), \tag{3}$$

$$w_i = \mathtt{t}_p(\hat{w}_i) = \begin{cases} \frac{1}{p}\Delta^{1-p}\hat{w}_i & (|\hat{w}_i| \leq \Delta^p) \\ \sigma(\hat{w}_i)\left\{|\hat{w}_i|^{\frac{1}{p}} + \left(\frac{1}{p} - 1\right)\Delta\right\} & (|\hat{w}_i| > \Delta^p) \end{cases}, \tag{4}$$



(a) $\mathtt{t}_p$ (b) $\mathtt{t}'_p$

Fig. 2. Transform function of the variable ($p = 0.5$)

where the transformation $\mathtt{t}_p$ is the inverse function of $\sigma(w)\hat{\mathtt{l}}_p(w)$ and it is shown in Fig. 2a. The first derivative of $\mathtt{t}_p$ is given by

$$\mathtt{t}'_p(\hat{w}_i) = \begin{cases} \frac{1}{p}\Delta^{1-p} & (|\hat{w}_i| \leq \Delta^p) \\ \frac{1}{p}|\hat{w}_i|^{\frac{1}{p}-1} & (|\hat{w}_i| > \Delta^p) \end{cases}. \tag{5}$$

Note that $\mathtt{t}_p$ is smooth continuous transformation, also resulting in the continuous derivative.

By transforming the variables $\boldsymbol{w}$ into $\hat{\boldsymbol{w}}$ via $\mathtt{t}_p^{-1}$, the optimization problem (1) finally results in

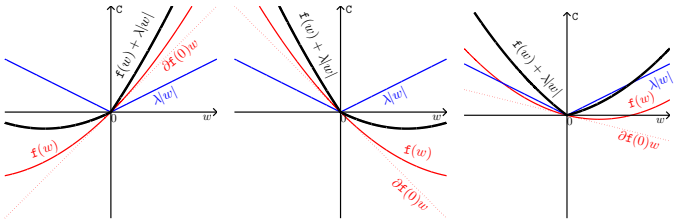
$$\min_{\hat{\boldsymbol{w}}} \mathtt{f}(\mathtt{t}_p(\hat{\boldsymbol{w}})) + \lambda \sum_i^d |\hat{w}_i| = \mathtt{f} \circ \mathtt{t}_p(\hat{\boldsymbol{w}}) + \lambda \|\hat{\boldsymbol{w}}\|_1. \tag{6}$$

This is a standard L_1 regularized optimization problem w.r.t. $\hat{\boldsymbol{w}}$, to which any types of L_1 optimizer can be applicable; in this study, we employ the orthant-wise approach [13] which is effectively applied to L_1 regularized logistic regression.

2.2. Orthant-wise Optimization

Andrew and Gao proposed the orthant-wise limited-memory quasi-Newton (OWLQN) method that effectively minimizes the cost function including L_1 regularization [13]; in that paper, OWLQN is applied to L_1 regularized logistic regression. The OWLQN is based on the L-BFGS, an efficient general minimizer applicable to differentiable problems [16], and it is experimentally shown to be effective [14]. Without doubling the variables nor relaxing the regularization, the OWLQN can directly minimize (6) by effectively determining the gradients around zeros; we briefly describe the OWLQN in the followings, and for further details, refer to [13].

Orthant is defined as the region on which variables never change their signs along in each coordinate, as an extension of quadrants in two dimensions. The OWLQN basically works on the orthant of the current $\boldsymbol{w}$ and goes through the boundary of the orthant according to the gradients (derivatives) which are defined especially on $w_i = 0$ as follows. Let $\mathtt{C}(\boldsymbol{w}) \triangleq \mathtt{f}(\boldsymbol{w}) + \lambda\|\boldsymbol{w}\|_1$ be the regularized cost function to be minimized, and the regularization term takes the derivative either



(a) $\partial_i \mathtt{f} < -\lambda$ (b) $\partial_i \mathtt{f} > \lambda$ (c) $-\lambda \leq \partial_i \mathtt{f} \leq \lambda$

Fig. 3. Three situations at $w_i = 0$ in OWLQN

of λ or $-\lambda$ at $w = 0$, which results in the three situations (Fig. 3):

$$\text{at } w_i = 0, \quad \begin{array}{lll} a) & \partial_i \mathtt{f} < -\lambda & \Rightarrow \partial_i \mathtt{C} = \partial_i \mathtt{f} + \lambda \\ b) & \partial_i \mathtt{f} > \lambda & \Rightarrow \partial_i \mathtt{C} = \partial_i \mathtt{f} - \lambda \\ c) & -\lambda \leq \partial_i \mathtt{f} \leq \lambda & \Rightarrow \partial_i \mathtt{C} = 0 \end{array} \tag{7}$$

$$\text{at } w_i \neq 0, \quad \partial_i \mathtt{C} = \partial_i \mathtt{f} + \sigma(w_i)\lambda.$$

The first two are the cases that the cost $\mathtt{C}(\boldsymbol{w})$ can be decreased toward positive or negative domain (orthant) of w_i, while for the third case, $\mathtt{C}(\boldsymbol{w})$ has the hollow at $w_i = 0$ by considering the regularization term, which thus makes the gradient zero. L-BFGS is applied while retaining the consistency of the orthant, and based on (7) the target orthant is changed, going through or staying at zeros.

Note that the OWLQN is applied to (6) in the proposed method. The derivative of $\mathtt{f} \circ \mathtt{t}_p(\hat{\boldsymbol{w}})$ is simply given by using (5) as

$$\partial_i(\mathtt{f} \circ \mathtt{t}_p) = (\partial_i \mathtt{f})\mathtt{t}_p'(\hat{w}_i). \tag{8}$$

3. EXPERIMENTAL RESULTS

We apply the proposed method to the tasks on classification and factorization. In these experiments, the classification is formulated by using logistic regression model and the factorization is based on the least squares. The sparsity is induced into the weights via L_p regularization for both tasks. In the proposed method, we employ the regularization of $p = 0.5$ ($L_{0.5}$) which produces highly sparse weights more than L_1 regularization.

3.1. Classification

For classification, we employ the logistic regression (LR) model [15]. Let $\boldsymbol{x}_i \in \mathbb{R}^d, i = 1, \cdots, N$ be the i-th sample (feature) vector and $y_{ic} \in \{+1, -1\}, c = 1, \cdots, C$ be its label of the c-th class; ± 1 indicates whether the i-th sample belongs to the c-th class or not. Thus, the cost function in the logistic regression is defined by

$$\mathtt{f}(\{\boldsymbol{w}_c\}_c^{C-1}) = -\sum_i^N \sum_c^C y_{ic} \log \eta_c(\boldsymbol{x}_i; \{\boldsymbol{w}_c\}_c^{C-1}), \tag{9}$$

Table 1. Classification performances by logistic regression on benchmark datasets. 'nnz' indicates the number of non-zero weights, and lower nnz means sparser weights.

		L_1-LR		L_p-LR	
Method	λ	acc. (%)	nnz	acc. (%)	nnz
ARCENE	1	71.97 (±6.31)	87.0 (±6.6)	64.99 (±7.16)	19.3 (±1.2)
ARCENE	0.1	68.96 (±7.39)	100.7 (±3.5)	77.48 (±6.60)	22.7 (±1.2)
DEXTER	1	92.0 (±2.65)	229.0 (±61.5)	88.50 (±2.60)	41.7 (±2.5)
DEXTER	0.1	91.5 (±3.12)	288.3 (±61.2)	86.33 (±1.44)	48.0 (±1.7)
DOROTHEA	1	92.69 (±0.95)	152.0 (±78.1)	92.52 (±0.76)	38.3 (±1.2)
DOROTHEA	0.1	92.43 (±0.95)	153.7 (±50.6)	92.87 (±0.80)	43.0 (±6.2)
GISETTE	1	97.61 (±0.07)	579.3 (±16.1)	96.47 (±0.22)	142.7 (±3.1)
GISETTE	0.1	97.53 (±0.16)	656.3 (±16.8)	96.36 (±0.45)	159.3 (±15.8)
MADELON	1	56.08 (±1.89)	470.0 (±3.6)	55.38 (±1.08)	159.7 (±2.3)
MADELON	0.1	55.61 (±1.06)	497.7 (±2.1)	54.61 (±1.18)	342.0 (±11.4)

where $\boldsymbol{w}_c, c = 1, \cdots, C-1$ are the classifier weights to be optimized, and the (multi-nominal) logistic function η_c is defined by

$$\eta_c(\boldsymbol{x}; \{\boldsymbol{w}_c\}_c^{C-1}) = \begin{cases} \frac{\exp(\boldsymbol{w}_c^\top \boldsymbol{x})}{1+\sum_c^{C-1} \exp(\boldsymbol{w}_c^\top \boldsymbol{x})} & (c < C) \\ \frac{1}{1+\sum_c^{C-1} \exp(\boldsymbol{w}_c^\top \boldsymbol{x})} & (c = C) \end{cases}. \tag{10}$$

The first derivative is also given by

$$\partial_{\boldsymbol{w}_c} \mathtt{f} = \sum_i^N \{\eta_c(\boldsymbol{x}_i; \{\boldsymbol{w}_c\}_c^{C-1}) - y_{ic}\}\boldsymbol{x}_i. \tag{11}$$

By introducing the L_p regularization into (9), we can obtain the *sparse* classifier weights $\boldsymbol{w}$, which contributes to not only classification itself but also feature selection; the feature components that the non-zero weight is assigned with is regarded as discriminatively selected features for the classification.

We test the proposed method for the L_p-regularized LR on the several benchmark datasets[1], ARCENE (10000), DEXTER (20000), DOROTHEA (100000), GISETTE (5000) and MADELON (500) where the numbers in the parentheses indicate the number of features, i.e., the feature dimensionality d. The performance is evaluated on 3-fold cross validations and averaged accuracy is reported. The results are shown in Table 1 with comparison to L_1-regularized LR [2]. While the classification accuracies are comparable to L_1-LR, the proposed L_p-LR produces highly sparse classification weights, which also enables faster classification. The L_p regularization contributes to dig out important features from a large number of features.

3.2. Factorization

We next apply the proposed method to the task on factorization of biological signals of protein dynamics in living cells (EGFP) and chemical particle dynamics in an aqueous solution (Rh6G) [17]. Those signals were measured by using fluorescence correlation spectroscopy (FCS) [18]. Automatic factorization of the signals facilitates the biological analysis; the

[1]UCI-repository http://archive.ics.uci.edu/ml/datasets.html

signals are difficult to be decomposed into factors by hand due to heavily overlapped factors (see Fig. 4). We observed 44 sequences of EGFP and 53 sequences of Rh6G, and the number of the sampling time points were 92 ($1.6\mu s \leq t \leq 4505.6\mu s$) and 120 ($2.2\mu s \leq t \leq 65536\mu s$), respectively. The factors in this type of signals are physically modeled by the functions $g_i(t;\tau_i) = \exp(-t/\tau_i)$, where the parameter τ_i stands for the diffusion time.

The input signal $s(t)$ is a linear composite of several factor signals with weights, $s(t) \approx \sum_j^d w_j g_j(t;\tau_j)$, and the cost for the factorization is determined based on least squares;

$$\mathtt{f}(\boldsymbol{w}) = \sum_i^N \int \|s_i(t) - \sum_j^d w_{ij} g_j(t;\tau_j)\|^2 dt. \qquad (12)$$

where N is the number of signal sequence. The L_p regularization is imposed on the weights $\boldsymbol{w}$ so as to be sparse factorization, and the factorization task is to estimate both the parameters τ_j and the weights $\boldsymbol{w}$. For that purpose, those τ_j and $\boldsymbol{w}$ are optimized alternately; given fixed weights $\boldsymbol{w}$, the parameters τ_j are optimized by simply applying gradient descent approach to minimize (12), and for the fixed parameters τ_i, the weights $\boldsymbol{w}$ are optimized by using the proposed method with the derivatives given by

$$\partial_{w_{ij}} \mathtt{f} = 2 \int g_j(t;\tau_j) \Big\{ \sum_j^d w_{ij} g_j(t;\tau_j) - s_i(t) \Big\} dt. \qquad (13)$$

We initially prepare 20 factors ($d = 20$) with random initial parameter values τ_j ($j = 1, .., 20$) with $\lambda = 0.1$; however, most of those factors will be assigned with zero weight at the optimum, and thus the factorization results are non-zero weights and their corresponding factors (i.e., τ).

The factorization results produced by the proposed method are shown in Table 2, and the example of the reconstructed signal with the retrieved factors is also shown in Fig. 4. The parameter value τ of the primary factor is estimated as $196.91\mu s$ in EGFP and $27.60\mu s$ in Rh6G, both of which are close to the ideal primary values, 244 (± 53)μs and 25 (± 11)μs based on the biological knowledge [19]. The second factors which are slower than the primary ones can be caused by such as protein-protein interactions, and the third and the latter factors are trivial since they are active only at two sequences in the datasets. The L_1 optimization is also applied for comparison, though decomposing the input signals *incorrectly* with wrong factors; the primary factor has $\tau = 126.47\mu s$ among the obtained 20 factors in EGFP and the primary factor of $\tau = 19.0\mu s$ among 18 factors in Rh6G. Thus, it can be said that, in this case that the factors are heavily overlapped, the higher sparsity is required to retrieve the underlying essential factors from the noisy input (composite) signals.

Table 2. Factorization results of biological signals by using the proposed method (L_p regularization)

EGFP primary diffusion time $\tau^* = 244(\pm 53)\mu s$		Rh6G primary diffusion time $\tau^* = 25(\pm 11)\mu s$	
[obtained 4 factors]		[obtained 3 factors]	
diffusion time τ	weight w	diffusion time τ	weight w
196.91 μs	0.77 ± 0.05	**27.60** μs	0.83 ± 0.03
3768.22 μs	0.16 ± 0.06	460.89 μs	0.13 ± 0.03
17466.16 μs	0.0062 ± 0.029	2174.27 μs	0.0028 ± 0.01
36832.84 μs	0.0044 ± 0.020		

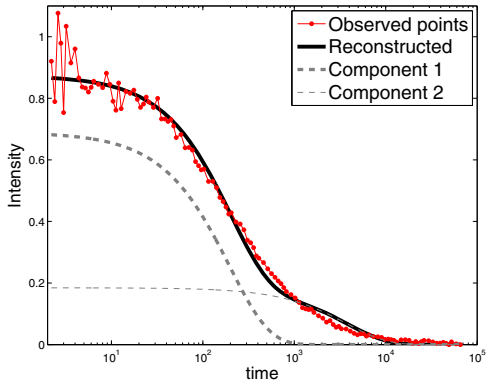


Fig. 4. Example of the reconstructed signal with the obtained factors in EGFP. The dotted lines show retrieved components.

4. CONCLUSION

We have proposed a method to efficiently optimize the L_p regularized problem. In the proposed method, the target variables, e.g., weights, are first transformed so as to reduce the L_p regularized problem into the L_1 problem. Then, so transformed L_1 problem is efficiently optimized by using the orthant-wise approach [13]. The proposed method directly optimizes the L_p problem in a computationally efficient way without reformulating the primal L_p problem into iteratively reweighting schemes which are commonly utilized in optimizing L_p regularized problems [1, 10]. The experimental results on classification using logistic regression and on factorization by least squares demonstrate that the proposed method produces favorable sparse results.

Acknowledgement. The author thanks Dr. Kenji Watanabe for kindly providing the biological signal data and helping with biological analysis.

5. REFERENCES

[1] M.-H. Wei, J.H. McClellan, and W.R. Scott, "Application of l_p-regularized least squares for $0 \leq p \leq 1$ in estimating discrete spectrum models from sparse frequency measurements," in *International Conference on Acoustics, Speech, and Signal Processing*, 2010, pp. 4010–4013.

[2] S.-I. Lee, H. Lee, P. Abbeel, and A. Y. Ng, "Efficient l_1 regularized logistic regression," in *AAAI*, 2006.

[3] Deng Cai, Hujun Bao, and Xiaofei He, "Sparse concept coding for visual analysis," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 2905–2910.

[4] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y. Ng, "Efficient sparse coding algorithms," in *Advances in Neural Information Processing Systems 19*, 2007, pp. 801–808.

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

[6] Andrew Y. Ng, "Feature selection, l1 vs. l2 regularization, and rotational invariance," in *International Conference on Machine Learning*, 2004.

[7] 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.

[8] 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.

[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] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *International Conference on Acoustics, Speech, and Signal Processing*, 2008, pp. 3869–3872.

[11] E. J. Candès, B. Wakin, and S.P. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.

[12] M. E. Davies and R. Gribonval, "Restricted isometry constants where l^p sparse recovery can fail for $0 < p \leq 1$," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2203–2214, 2009.

[13] G. Andrew and J. Gao, "Scalable training of l_1-regularized log-linear models," in *International Conference on Machine Learning*, 2007, pp. 33–40.

[14] Mark Schmidt, Glenn Fung, and Romer Rosales, "Optimization methods for l1-regularization," Tech. Rep. TR-2009-19, University of British Columbia, 2009.

[15] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2007.

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

[17] K. Watanabe, A. Hidaka, N. Otsu, and T. Kurita, "Automatic analysis of composite physical signals using non-negative factorization and information criterion," *PLoS One*, vol. 7, no. 3, pp. e32352, 03 2012.

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

[19] 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.