Block Sparse Bayesian Learning: A Diversified Scheme

Yanhao Zhang, Zhihan Zhu

School of Mathematical Sciences, Beihang University

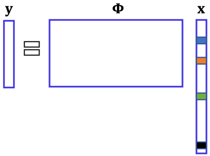
Nov 1, 2024

Joint work with Prof. Yong Xia

Code Paper

Compressed Sensing / Sparse Regression

Compressed Sensing [Donoho, 2006] / Sparse Regression :



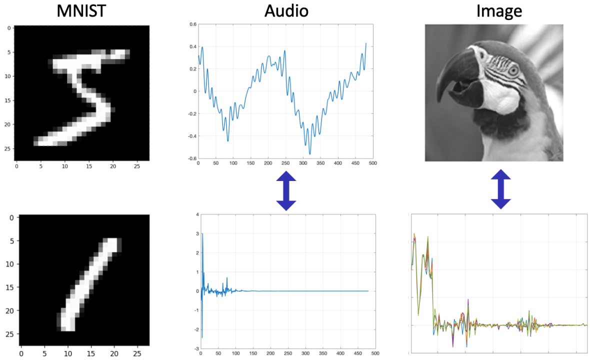
In compressed sensing, **x** often exhibits transform sparsity, becoming sparse in a transform domain such as Wavelet, Fourier, etc.

Classical Methods:

- Orthogonal Matching Pursuit (OMP) [Pati et al., 1993]
- ℓ_1-Minimization: Basis Pursuit [Chen et al., 2001], LASSO [Tibshirani, 1996]
- Replacing $\|\cdot\|_1$ with other non-convex regularization such as $\|\cdot\|_p (0 < p < 1)$ [Frank and Friedman, 1993] or SCAD [Fan and Li, 2001] leads to a non-convex programming.

Block Sparse Phenomenon

Relying solely on the sparsity of **x** *is insufficient, particularly when sample sizes are limited.*[Eldar et al., 2010, Donoho et al., 2013]



Widely encountered real-world data, such as image and audio, often exhibit **block sparsity** or in their transform domain.

Block Sparsity

Block Sparsity: the sparse non-zero entries of $\mathbf{x}$ appear in blocks [Eldar et al., 2010]. Generally, the block structure of $\mathbf{x}$ with g blocks is defined by

$$\mathbf{x} = [\underbrace{x_1 \ldots x_{d_1}}_{\mathbf{x}_1^T} \underbrace{x_{d_1+1} \ldots x_{d_1+d_2}}_{\mathbf{x}_2^T} \cdots \underbrace{x_{N-d_g+1} \ldots x_N}_{\mathbf{x}_g^T}]^T, \tag{1}$$

Suppose only $k (k \ll g)$ blocks are non-zero, indicating that $\mathbf{x}$ is block sparse.

- **Question**: How can block information be used to achieve better accuracy in sparse recovery problems?

Block Sparse Recovery

Literature Review

Algorithms:

- **Block-based**:
  - Group-Lasso [Yuan and Lin, 2006]
  - Group Basis Pursuit [Van den Berg and Friedlander, 2011]
  - Block-OMP [Eldar et al., 2010]
  - Block-SBL (BSBL) [Zhang and Rao, 2013]
- **Pattern-based**:
  - StructOMP [Huang et al., 2009]
  - Pattern-Coupled SBL (PC-SBL) [Fang et al., 2014]
  - Burst PC-SBL [Dai et al., 2018]

Remark 1

Longstanding Issue *(block-based algorithms): Highly dependent on predefined block information, leading to simultaneous learning of block elements as either 0 or* $\sim$*0 based on the predefined blocks.*

Problem Setting

Observation model: Consider block sparse recovery problem as

$$\mathbf{y} = \mathbf{\Phi x} + \mathbf{n}, \tag{2}$$

$\mathbf{x}$ exhibits block sparse structure, yet its partition is unknown.

Model Setting (Block-based): All blocks have equal size L, with total dimension denoted as $N = gL$. Henceforth, $\mathbf{x}$ follows the structure:

$$\mathbf{x} = [\underbrace{x_{11} \ldots x_{1L}}_{\mathbf{x}_1^T} \underbrace{x_{21} \ldots x_{2L}}_{\mathbf{x}_2^T} \cdots \underbrace{x_{g1} \ldots x_{gL}}_{\mathbf{x}_g^T}]^T. \tag{3}$$

- The choice of L is important and sensitive to existing block-based methods!

Previous Works

- **Group Lasso** [Yuan and Lin, 2006]: $\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{\Phi x}-\mathbf{y}\|_2^2+\tau\sum_{i=1}^{g}\|\mathbf{x}_i\|_2$.
- **Group BPDN**: [Van den Berg and Friedlander, 2011]

$$\min_{\mathbf{x}} \sum_{i=1}^{g} \|\mathbf{x}_i\|_2 \quad \text{s. t. } \|\mathbf{\Phi x}-\mathbf{y}\| \leq \sigma.$$

- **Block OMP** [Eldar et al., 2010]: heuristically select blocks instead of elements.
- **Block Sparse Bayesian Learning** [Zhang and Rao, 2013]:

$$p\left(\mathbf{x}_i;\{\gamma_i,\mathbf{B}_i\}\right)=\mathcal{N}\left(\mathbf{0},\gamma_i\mathbf{B}_i\right),\forall i=1,\cdots,g, \tag{4}$$

- All of the block-based algorithms estimate one block to be either zero or non-zero simultaneously!
- Can we overcome the sensitivity issue for block-based methods?

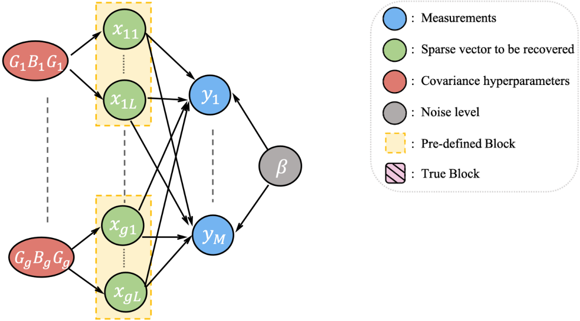
Diversified Block Sparse Prior

Diversified Block Sparse Prior: Each block $\mathbf{x}_i \in \mathbb{R}^{L\times 1}$ is assumed to follow a multivariate Gaussian prior

$$p\left(\mathbf{x}_i; \{\mathbf{G}_i, \mathbf{B}_i\}\right) = \mathcal{N}\left(\mathbf{0}, \mathbf{G}_i\mathbf{B}_i\mathbf{G}_i\right), \forall i = 1, \cdots, g, \tag{5}$$

- $\mathbf{G}_i$: Diversified Variance matrix.
- $\mathbf{B}_i$: Diversified Correlation matrix.
- The formulations of $\mathbf{G}_i$ and $\mathbf{B}_i$ will be detailed later.

Diversified Block Sparse Prior



The prior distribution of the entire signal $\mathbf{x}$ is denoted as

$$p\left(\mathbf{x}; \{\mathbf{G}_i, \mathbf{B}_i\}_{i=1}^{g}\right) = \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_0\right), \tag{6}$$

where $\mathbf{\Sigma}_0 = \text{diag}\left\{\mathbf{G}_1\mathbf{B}_1\mathbf{G}_1, \mathbf{G}_2\mathbf{B}_2\mathbf{G}_2, \cdots, \mathbf{G}_g\mathbf{B}_g\mathbf{G}_g\right\}$.

Diversified Block Sparse Prior

(A) Diversified Intra-Block Variance

Diversified intra-block variance:

- $\mathbf{G}_i$ is defined as

$$\mathbf{G}_i \triangleq \mathrm{diag}\{\sqrt{\gamma_{i1}}, \cdots, \sqrt{\gamma_{iL}}\}, \tag{7}$$

- $\mathbf{B}_i = [\rho^i_{sk}]_{s,k=1\cdots L}$ is a positive definite correlation matrix of the i-th block.

According to the definition of Pearson correlation, the covariance term $\mathbf{G}_i\mathbf{B}_i\mathbf{G}_i$ can be specified as

$$\mathbf{G}_i\mathbf{B}_i\mathbf{G}_i = \begin{bmatrix} \gamma_{i1} & \rho^i_{12}\sqrt{\gamma_{i1}}\sqrt{\gamma_{i2}} & \cdots & \rho^i_{1L}\sqrt{\gamma_{i1}}\sqrt{\gamma_{iL}} \\ \rho^i_{21}\sqrt{\gamma_{i2}}\sqrt{\gamma_{i1}} & \gamma_{i2} & \cdots & \rho^i_{2L}\sqrt{\gamma_{i2}}\sqrt{\gamma_{iL}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^i_{L1}\sqrt{\gamma_{iL}}\sqrt{\gamma_{i1}} & \rho^i_{L2}\sqrt{\gamma_{iL}}\sqrt{\gamma_{i2}} & \cdots & \gamma_{iL} \end{bmatrix},$$

Diversified Block Sparse Prior

(A) Diversified Intra-Block Variance

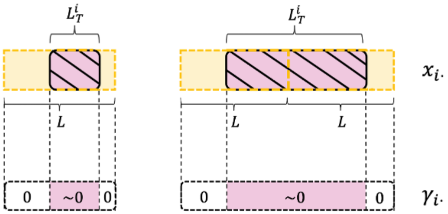
Adaptive blocksize / Insensitivity of preset blocksize L:

Demonstration Rationale

Left: When the true block is located within the preset block.

Right: When the true block is located across the preset block.

The variances $\gamma_{i\cdot}$ corresponding to the non-zero ($\sim$0) or zero positions in $x_{i\cdot}$ will be automatically learned as non-zero or zero values through posterior inference, hence refine to the true block.



- Both size and location of blocks will automatically shrink through posterior inference on the variances term.

Diversified Block Sparse Prior

(B) Diversified inter-block correlation

Here we introduce weak constraints on $\mathbf{B}_i$, specifically,

$$\psi(\mathbf{B}_i) = \psi(\mathbf{B}) \quad \forall i = 1...g, \tag{8}$$

where $\psi : \mathbb{R}^{L^2} \rightarrow \mathbb{R}$ is the weak constraint function and $\mathbf{B}$ is obtained from the strong constraints $\mathbf{B}_i = \mathbf{B}(\forall i)$.

Weak constraints (8) :

- capture the distinct correlation structure but also avoid overfitting issue.
- effectively enhance the convergence rate of the algorithm (Number of constraints: $gL^2 \rightarrow g$) .

Diversified Correlation Matrices by Dual Ascent

How to choose the constraint function ψ ?

- Explicit constraints with complete dual ascent.
- Hidden constraints with one-step dual ascent.

Proposition 1

Define an explicit weak constraint function $\zeta : \mathbb{R}^{n^2} \rightarrow \mathbb{R}$*. For the constrained optimization problem:*

$$\begin{aligned} \min_{\mathbf{B}_i} \quad & Q(\{\mathbf{B}_i\}_{i=1}^{g}, \{\mathbf{G}_i\}_{i=1}^{g}) \\ \text{s. t.} \quad & \zeta(\mathbf{B}_i) = \zeta(\mathbf{B}), \quad \forall i = 1...g, \end{aligned}$$

the stationary point $(\{\mathbf{B}_i^{k+1}\}_{i=1}^{g}, \{\lambda_i^k\}_{i=1}^{g})$ *of the Lagrange function under given multipliers* $\{\lambda_i^k\}_{i=1}^{g}$ *satisfies:*

$$\nabla_{\mathbf{B}_i} Q(\{\mathbf{B}_i^{k+1}\}_{i=1}^{g}, \{\mathbf{G}_i\}_{i=1}^{g}) - \lambda_i^k \nabla \zeta(\mathbf{B}_i^{k+1}) = 0.$$

Then there exists a constrained optimization problem with hidden weak constraint $\psi : \mathbb{R}^{n^2} \rightarrow \mathbb{R}$*:*

$$\begin{aligned} \min_{\mathbf{B}_i} \quad & Q(\{\mathbf{B}_i\}_{i=1}^{g}, \{\mathbf{G}_i\}_{i=1}^{g}) \\ \text{s. t.} \quad & \psi(\mathbf{B}_i) = \psi(\mathbf{B}), \quad \forall i = 1...g, \end{aligned}$$

such that $(\{\mathbf{B}_i^{k+1}\}_{i=1}^{g}, \{\lambda_i^k\}_{i=1}^{g})$ *is a KKT pair of the above optimization problem.*

Posterior Estimation

Diversified **S**parse **B**ayesian **L**earning (DivSBL):

Algorithm 1 DivSBL Algorithm

Input: Measurement matrix $\mathbf{\Phi}$, response $\mathbf{y}$, initialized variance $\boldsymbol{\gamma}$, prior's covariance $\mathbf{\Sigma}_0$, noise's variance β, and multipliers $\boldsymbol{\lambda}^0$.
Output: Posterior mean $\hat{\mathbf{x}}^{MAP}$, posterior covariance $\hat{\mathbf{\Sigma}}$, variance $\hat{\boldsymbol{\gamma}}$, correlation $\hat{\mathbf{B}}_i$, noise $\hat{\beta}$.
repeat
    if mean($\boldsymbol{\gamma}_{l\cdot}$)< threshold **then**
        Prune $\boldsymbol{\gamma}_{l\cdot}$ from the model (set $\boldsymbol{\gamma}_{l\cdot} = \mathbf{0}$). // Zero out small energy for efficiency.
        Set the corresponding $\boldsymbol{\mu}^l = \mathbf{0}, \mathbf{\Sigma}^l = \mathbf{0}_{L\times L}$.
    end if
    Update $\gamma_{ij} \leftarrow \frac{4\mathbf{A}_{ij}^2}{(\sqrt{\mathbf{T}_{ij}^2 + 4\mathbf{A}_{ij}} - \mathbf{T}_{ij})^2}$. // Update diversified variance.
    Update $\mathbf{B} \leftarrow \frac{1}{g}\sum_{i=1}^{g} \mathbf{G}_i^{-1}\left(\mathbf{\Sigma}^i + \boldsymbol{\mu}^i\left(\boldsymbol{\mu}^i\right)^T\right)\mathbf{G}_i^{-1}$. // Avoid overfitting.
    Update $\mathbf{B}_i$ by $\mathbf{B}_i^{k+1} \leftarrow \frac{\mathbf{G}_i^{-1}\left(\mathbf{\Sigma}^i + \boldsymbol{\mu}^i\left(\boldsymbol{\mu}^i\right)^T\right)\mathbf{G}_i^{-1}}{1+2\lambda_i^k}$. // Diversified correlation.
    Update λ_i by $\lambda_i^{k+1} \leftarrow \lambda_i^k + \alpha_i^k(\log\det \mathbf{B}_i^k - \log\det \mathbf{B})$. // Diversified correlation.
    Execute Toeplitz correction for $\mathbf{B}_i$.
    Update $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ by $\boldsymbol{\mu} \leftarrow \beta\mathbf{\Sigma}\mathbf{\Phi}^T\mathbf{y}, \mathbf{\Sigma} \leftarrow \left(\mathbf{\Sigma}_0^{-1} + \beta\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}$.
    Update $\beta \leftarrow \frac{M}{\|\mathbf{y} - \mathbf{\Phi}\boldsymbol{\mu}\|_2^2 + \mathrm{tr}\left(\mathbf{\Sigma}\mathbf{\Phi}^T\mathbf{\Phi}\right)}$.
until convergence criterion met
$\hat{\mathbf{x}}^{MAP} = \boldsymbol{\mu}$. // Use posterior mean as estimate.

Diversified Block Sparse Prior

Connections to classical models

Two classic Sparse Bayesian Learning models, RVM [Tipping, 2001] and BSBL [Zhang and Rao, 2011], are special cases of our model.

- **Connection to RVM**: Taking $\mathbf{B}_i$ as identity matrix, diversified block sparse prior (6) immediately degenerates to RVM model

$$p\left(x_i;\gamma_i\right) = \mathcal{N}\left(0,\gamma_i\right), \forall i = 1,\cdots,N, \tag{9}$$

  which means ignoring the correlation structure.

- **Connection to BSBL**: When $\mathbf{G}_i$ is scalar matrix $\sqrt{\gamma_i}\mathbf{I}$, the formulation (5) becomes

$$p\left(\mathbf{x}_i;\{\gamma_i,\mathbf{B}_i\}\right) = \mathcal{N}\left(\mathbf{0},\gamma_i\mathbf{B}_i\right), \forall i = 1,\cdots,g, \tag{10}$$

  which is exactly BSBL model. In this case, all elements within a block share common variance γ_i.

Global and Local Properties

Property of global minimum:

Theorem 1

As $\beta \to \infty$ and $K_0 < (M+1)/2L$, the unique global minimum $\widehat{\gamma} \triangleq (\widehat{\gamma}_{11}, \ldots, \widehat{\gamma}_{gL})^T$ yields a recovery $\hat{\mathbf{x}}$ that is equal to $\mathbf{x}_{true}$, regardless of the estimated $\widehat{\mathbf{B}}_i$ $(\forall i)$.

Property of local minima:

Theorem 2

Every local minimum of the cost function with respect to γ satisfies $||\hat{\gamma}||_0 \leq \sqrt{M}$, irrespective of noise $(\forall \beta)$ and the estimated $\widehat{\mathbf{B}}_i$ $(\forall i)$.

The above results ensure the sparsity of the final solution obtained.

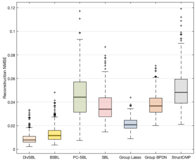
Experiments

- Our algorithm: DivSBL.
- Compare with:
  - Block-based algorithms: (1) BSBL, (2) Group Lasso, (3) Group BPDN.
  - Pattern-based algorithms: (4) PC-SBL, (5) StructOMP.
  - Sparse learning (without structural information): (6) SBL.
- Metrics:
  - Normalized Mean Squared Error (NMSE): $||\hat{x} - x_{\text{true}}||_2^2 / ||x_{\text{true}}||_2^2$.
  - Correlation (Corr)(cosine similarity): $\text{Corr}(x, y) = x'y/(\|x\|\|y\|)$

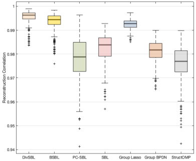
Experiment 1: Synthetic Signal

Table: Reconstruction error (NMSE) and Correlation (mean±std) for synthetic signals.

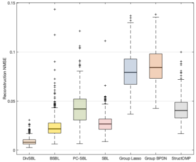
Algorithm	NMSE	Corr
Homoscedastic		
BSBL	0.0132±0.0069	0.9936±0.0034
PC-SBL	0.0450±0.0188	0.9784±0.0090
SBL	0.0263±0.0129	0.9825±0.0062
Group Lasso	0.0215±**0.0052**	0.9925±**0.0020**
Group BPDN	0.0378±0.0087	0.9812±0.0044
StructOMP	0.0508±0.0157	0.9760±0.0073
DivSBL	**0.0094**±0.0053	**0.9955**±0.0026
Heteroscedastic		
BSBL	0.0245±0.0125	0.9883±0.0047
PC-SBL	0.0421±0.0169	0.9798±0.0082
SBL	0.0274±0.0095	0.9873±0.0040
Group Lasso	0.0806±0.0180	0.9642±0.0096
Group BPDN	0.0857±0.0173	0.9608±0.0096
StructOMP	0.0419±0.0123	0.9803±0.0061
DivSBL	**0.0086**±**0.0041**	**0.9958**±**0.0020**



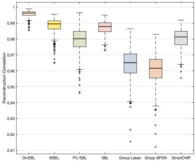
(a)Homo-NMSE



(b)Homo-Corr



(c)Heter-NMSE

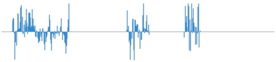


(d)Heter-Corr

The Robustness of Pre-defined Block Sizes

- Resolves the longstanding sensitivity issue of block-based algorithms.
- Exhibits enhanced recovery capability in challenging scenarios.

Environment: Three blocks with sizes of 100, 40, and 30, respectively. The block sizes differ significantly, with some being relatively large, and their positions are random.



Demonstration Rationale

(1) **DivSBL:** The optimal block size for DivSBL is in the range of 20 to 50, which aligns more closely with the actual block size of the signal compared to other block-based algorithms.

(2) **Other block-based algorithms:** They tend to have smaller (more biased) optimal block sizes. Even at their optimal block sizes, the result lags behind DivSBL. Their improvement over element-wise recovery is also limited.

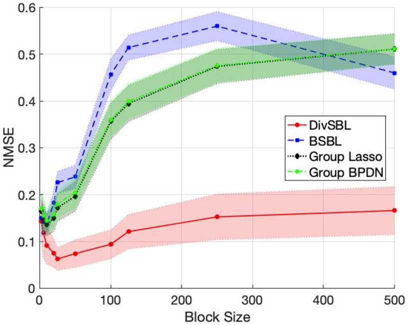


Figure: NMSE variation with changing preset block sizes.

The Robustness of Pre-defined Block Sizes

Variance Learning.

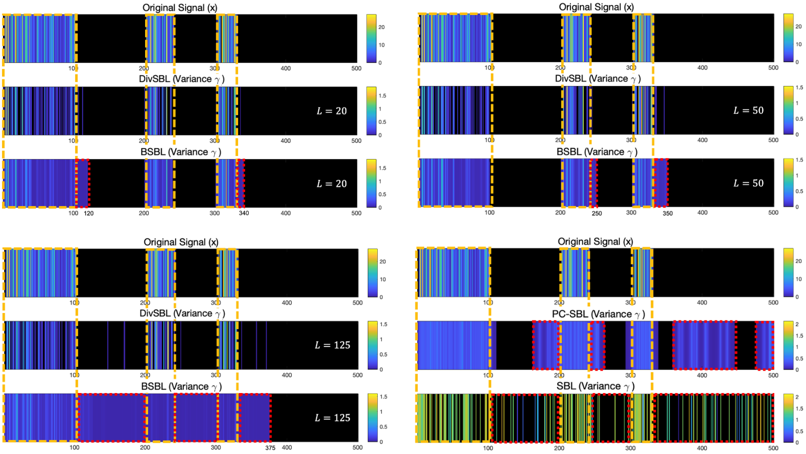


Figure: Posterior variance of DivSBL & BSBL when L=20, 50, 125.

Experiment 2: 1D AudioSet

Audio signals display block sparse structures in the discrete cosine transform (DCT) basis.

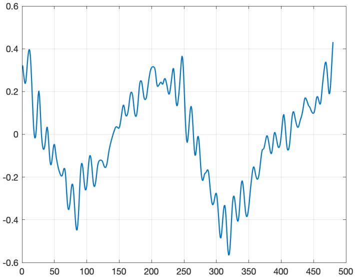


Figure: Original Audio Signal[1]

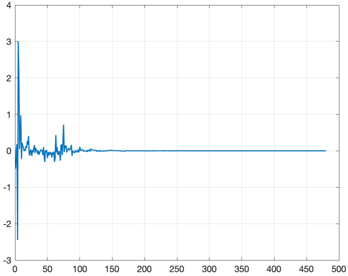


Figure: Sparse Representation

[1] Available at https://research.google.com/audioset/.

1D AudioSet

Phase transition diagram under different SNR (noise) and measurement levels.

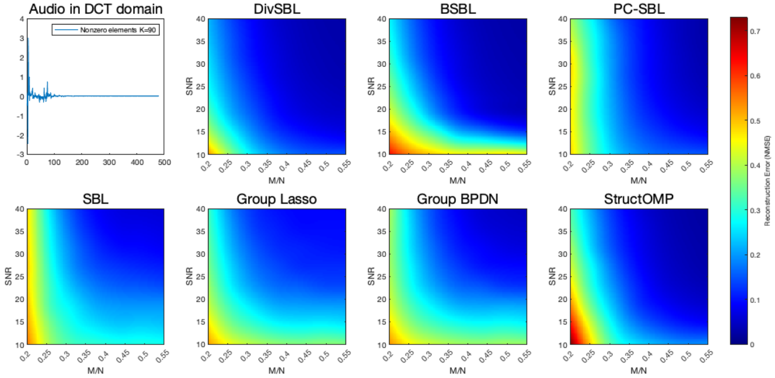


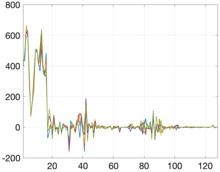
Figure: Phase transition diagram.

Experiment 3: 2D Image Reconstruction[2]

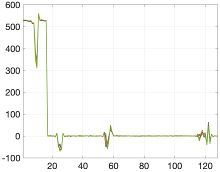
Example:

(a) Parrot Image

(b) House Image



(a) Parrot Image's Sparse Domain



(b) House Image's Sparse Domain

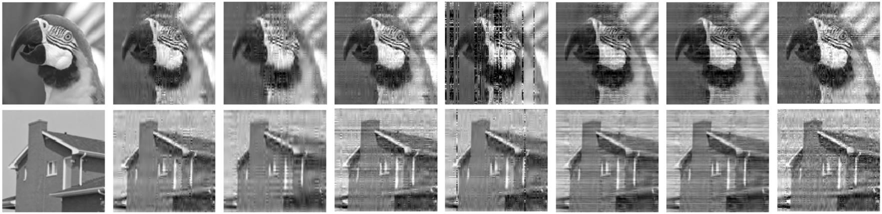
Figure: Parrot and House image data (the first five columns) transformed in discrete wavelet domain.

[2] Available at http://dsp.rice.edu/software/DAMP-toolbox and http://see.xidian.edu.cn/faculty/wsdong/NLR_Exps.htm

Experiment 3: 2D Image Reconstruction

Table: Reconstructed error (NMSE ± std) of the test images. (with an average improvement of **9.8%**)

Algorithm	Parrot	Cameraman	Lena	Boat	House	Barbara	Monarch	Foreman
BSBL	0.139 ± **0.004**	0.156 ± 0.006	0.137 ± **0.004**	0.179 ± **0.007**	0.146 ± 0.007	0.142 ± **0.004**	0.272 ± 0.009	0.125 ± 0.007
PC-SBL	0.133 ± 0.013	0.150 ± 0.012	0.134 ± 0.013	0.159 ± 0.014	0.137 ± 0.013	0.137 ± 0.013	0.208 ± 0.010	0.126 ± 0.014
SBL	0.225 ± 0.121	0.247 ± 0.141	0.223 ± 0.129	0.260 ± 0.114	0.238 ± 0.125	0.228 ± 0.119	0.458 ± 0.106	0.175 ± 0.099
GLasso	0.139 ± 0.017	0.153 ± 0.016	0.134 ± 0.017	0.159 ± 0.018	0.141 ± 0.018	0.135 ± 0.016	0.216 ± 0.020	0.124 ± 0.017
GBPDN	0.138 ± 0.017	0.153 ± 0.017	0.134 ± 0.017	0.159 ± 0.019	0.133 ± 0.019	0.135 ± 0.017	0.218 ± 0.022	0.123 ± 0.017
StructOMP	0.161 ± 0.014	0.184 ± 0.013	0.159 ± 0.013	0.187 ± 0.014	0.162 ± 0.014	0.164 ± 0.013	0.248 ± 0.015	0.149 ± 0.016
DivSBL	**0.117** ± 0.007	**0.142** ± **0.006**	**0.114** ± 0.005	**0.150** ± 0.008	**0.120** ± **0.006**	**0.120** ± 0.005	**0.203** ± **0.008**	**0.101** ± **0.007**



(a) Original (b) DivSBL (c) BSBL (d) PC-SBL (e) SBL (f) Group Lasso (g) Group BPDN (h) StructOMP

Figure: Reconstruction results for Parrot and House images.

Conclusion

- We introduce **Diversified Block Sparse Prior** to characterize block sparsity by allowing diversification on **intra-block variance** and **inter-block correlation matrices**.
- We propose **DivSBL**, utilizing EM algorithm and dual ascent method for hyperparameter estimation.
- We effectively **address the sensitivity issue** of existing block sparse learning methods to pre-defined block information.
- We establish the **optimality theory** and experiments validate its **state-of-the-art performance** on multimodal data.
- Future works include exploration on more effective weak constraints for correlation matrices, and applications on supervised learning tasks such as regression and classification.

References I

[Chen et al., 2001] Chen, S. S., Donoho, D. L., and Saunders, M. A. (2001).
Atomic decomposition by basis pursuit.
SIAM review, 43(1):129–159.

[Dai et al., 2018] Dai, J., Liu, A., and So, H. C. (2018).
Non-uniform burst-sparsity learning for massive MIMO channel estimation.
IEEE Transactions on Signal Processing, 67(4):1075–1087.

[Donoho, 2006] Donoho, D. L. (2006).
Compressed sensing.
IEEE Transactions on information theory, 52(4):1289–1306.

[Donoho et al., 2013] Donoho, D. L., Johnstone, I., and Montanari, A. (2013).
Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising.
IEEE transactions on information theory, 59(6):3396–3433.

[Eldar et al., 2010] Eldar, Y. C., Kuppinger, P., and Bolcskei, H. (2010).
Block-sparse signals: Uncertainty relations and efficient recovery.
IEEE Transactions on Signal Processing, 58(6):3042–3054.

[Fan and Li, 2001] Fan, J. and Li, R. (2001).
Variable selection via nonconcave penalized likelihood and its oracle properties.
Journal of the American statistical Association, 96(456):1348–1360.

[Fang et al., 2014] Fang, J., Shen, Y., Li, H., and Wang, P. (2014).
Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals.
IEEE Transactions on Signal Processing, 63(2):360–372.

[Frank and Friedman, 1993] Frank, L. E. and Friedman, J. H. (1993).
A statistical view of some chemometrics regression tools.
Technometrics, 35(2):109–135.

[Huang et al., 2009] Huang, J., Zhang, T., and Metaxas, D. (2009).
Learning with structured sparsity.
In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 417–424.

References II

[Pati et al., 1993] Pati, Y. C., Rezaiifar, R., and Krishnaprasad, P. S. (1993).
Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition.
In *Proceedings of 27th Asilomar conference on signals, systems and computers*, pages 40–44. IEEE.

[Tibshirani, 1996] Tibshirani, R. (1996).
Regression shrinkage and selection via the lasso.
Journal of the Royal Statistical Society Series B: Statistical Methodology, 58(1):267–288.

[Tipping, 2001] Tipping, M. E. (2001).
Sparse Bayesian learning and the relevance vector machine.
Journal of machine learning research, 1(Jun):211–244.

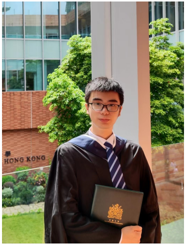
[Van den Berg and Friedlander, 2011] Van den Berg, E. and Friedlander, M. P. (2011).
Sparse optimization with least-squares constraints.
SIAM Journal on Optimization, 21(4):1201–1229.

[Yuan and Lin, 2006] Yuan, M. and Lin, Y. (2006).
Model selection and estimation in regression with grouped variables.
Journal of the Royal Statistical Society Series B: Statistical Methodology, 68(1):49–67.

[Zhang and Rao, 2011] Zhang, Z. and Rao, B. D. (2011).
Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning.
IEEE Journal of Selected Topics in Signal Processing, 5(5):912–926.

[Zhang and Rao, 2013] Zhang, Z. and Rao, B. D. (2013).
Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation.
IEEE Transactions on Signal Processing, 61(8):2009–2015.

Thanks to co-authors



Zhang, Yanhao
Beihang University

Zhu, Zhihan
Beihang University

Prof. Xia, Yong
Beihang University

Thanks for Your Attention!

Code

Paper