Outlier-aware Dictionary Learning for Sparse Representation

Sajjad Amini*
Mostafa Sadeghi*
Mohsen Joneidi*
Massoud Babaie-Zadeh*
Christian Jutten**

*Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran.
**GIPSA-Lab, Grenoble, and Institut Universitaire de France, France.

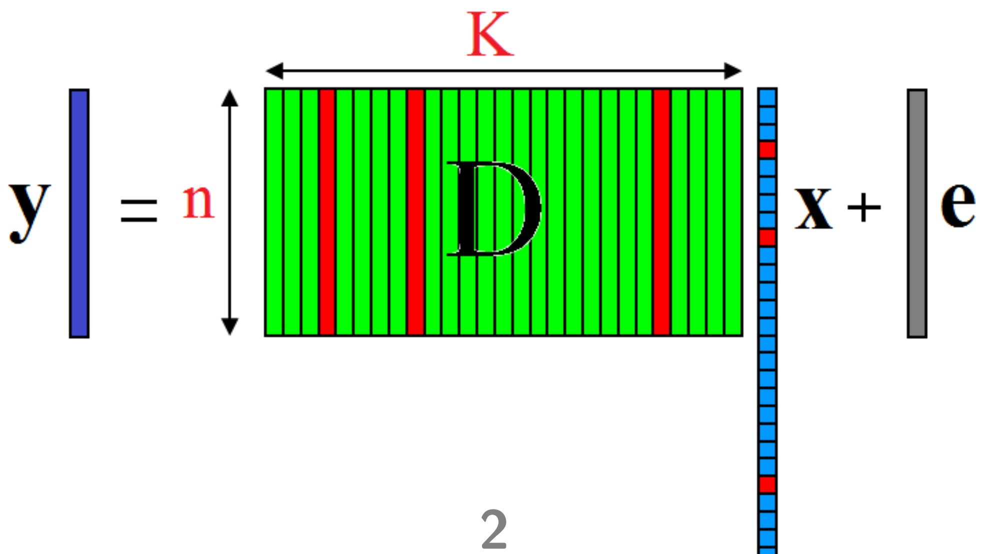
September 2014

Sparse Representation

Underdetermind Linear System of Equations

- $\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{e}$
  - ☞ $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{D} \in \mathbb{R}^{n \times K}$, $\mathbf{x} \in \mathbb{R}^K, \mathbf{e} \in \mathbb{R}^n$
  - ☞ $\mathbf{D} = [\mathbf{d}_1 \ \mathbf{d}_2 \ \ldots \ \mathbf{d}_K]$: dictionary, $\mathbf{d}_i$: atom
  - ☞ The dictionary is usually overcomplete: $K > n$
- Sparse representation problem:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \leq \epsilon$$

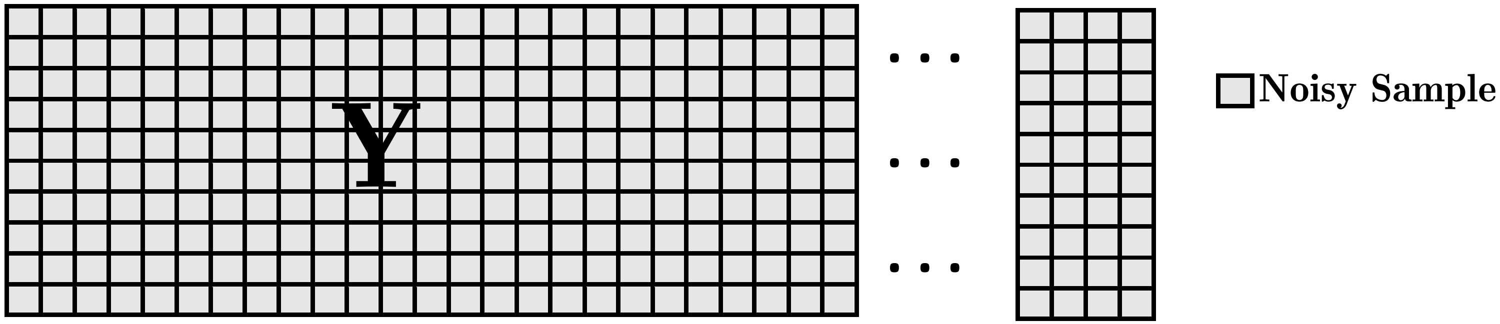


Choosing the Dictionary

- Pre-defined and fixed dictionaries: Fourier, Gabor, DCT, wavelet, ...
  - ✓ Fast computations
  - ✗ Unable to sparsely represent a given signal class

- Learned dictionaries
  - ✓ More efficient for sparse representation
  - ✓ Very promising results in many applications: image enhancement, pattern recognition, ...
  - ✗ High computational load

Dictionary Learning (DL)



- Given a noisy training data matrix, $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_L]$, the goal is to find an over-complete set of basis functions (atoms) over which each data can be sparsely represented

Training Data Model

$$\mathbf{y}_i = \mathbf{D}\mathbf{x}_i + \mathbf{n}_i \quad i = 1, \ldots, L$$

☞ $p(\mathbf{x}) \propto \exp(-\frac{\|\mathbf{x}\|_1}{\beta_1})$

☞ $p(\mathbf{n}) \propto \exp(-\frac{\|\mathbf{n}\|_2^2}{\beta_2})$

Dictionary Learning (DL)

MAP Estimation of Dictionary and Representations

$$\min_{\mathbf{D},\{\mathbf{x}_i\}_{i=1}^{L}} \sum_i \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2 \quad \text{subject to} \quad \begin{cases} \|\mathbf{d}_i\|_2 = 1, \; i = 1, \ldots, K \\ \|\mathbf{x}_j\|_0 \leq T_0, \; j = 1, \ldots, L \end{cases}$$

- $\|\mathbf{x}\|_0 \triangleq |\text{supp}(\mathbf{x}) = \{i : \; x_i \neq 0\}|$

Solution to the Dictionary Learning Problem

- **Alternating Minimization** Starting with an initial dictionary, the following two stages are repeated several times:
  1. Sparse representation:
     $\mathbf{X}^{(k+1)} = \text{argmin}_{\mathbf{X}\in\mathcal{X}} \; \|\mathbf{Y} - \mathbf{D}^{(k)}\mathbf{X}\|_F^2$ subject to $\|\mathbf{x}_j\|_0 \leq T_0, \; j = 1, \ldots, L \Longrightarrow$ OMP
  2. Dictionary update:
     $$\mathbf{D}^{(k+1)} = \underset{\mathbf{D}\in\mathcal{D}}{\text{argmin}} \; \|\mathbf{Y} - \mathbf{D}\mathbf{X}^{(k+1)}\|_F^2 \text{ subject to } \|\mathbf{d}_i\|_2 = 1, \; i = 1, \ldots, K \Longrightarrow \text{Differentiating}$$

Image Denoising Using Dictionary Learning

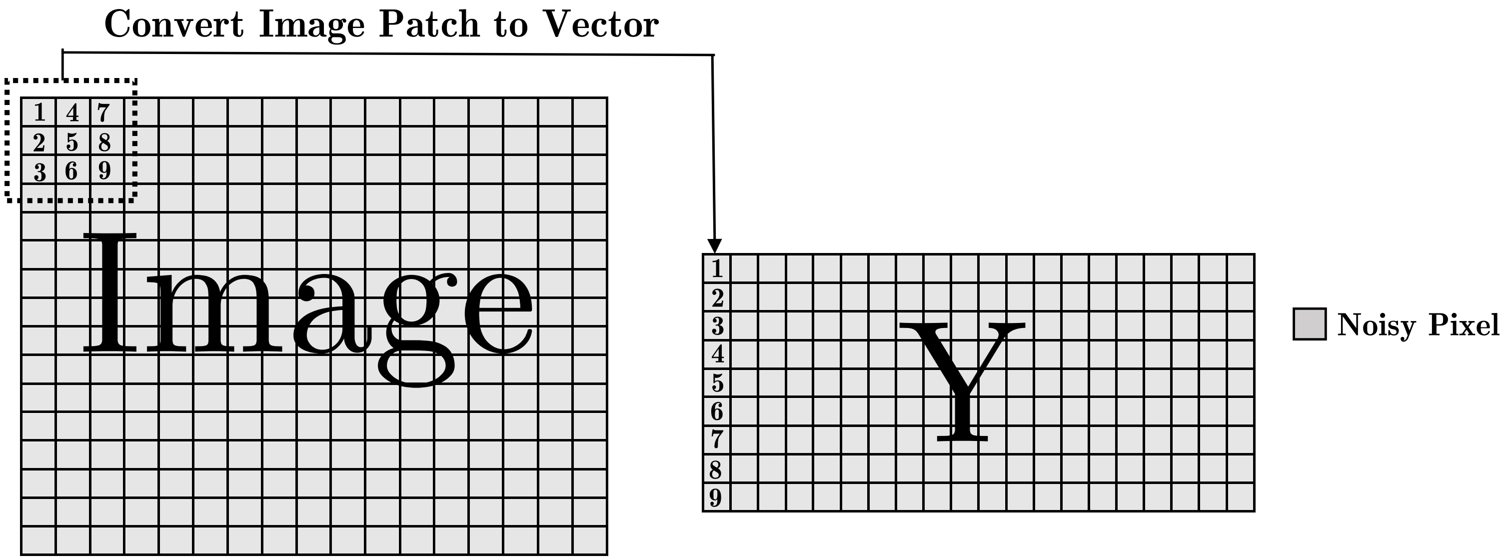
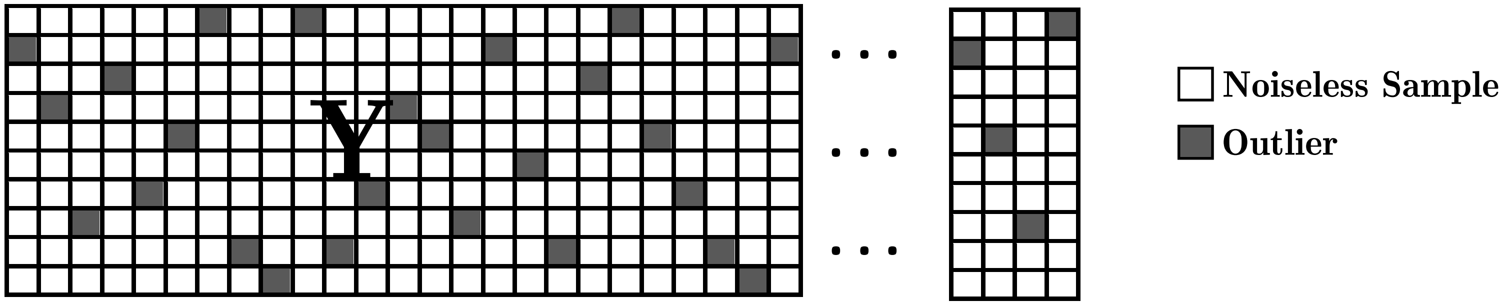


Image Denoising

- Training Image Dictionary ($\mathbf{D}^\star$)
- Estimating Denoised Image Patch Representation
  - ☞ $\hat{\mathbf{x}} = \mathsf{OMP}_{\mathbf{D}^\star}(\mathbf{y})$
- Reconstructing Denoised Image Patch
  - ☞ $\hat{\mathbf{y}} = \mathbf{D}^\star\hat{\mathbf{x}}$

Robust Dictionary Learning



Problem Formulation

- Training Data Model:

$$\mathbf{y}_i = \mathbf{D}\mathbf{x}_i + \mathbf{n}_i \quad i = 1, \dots, L$$

☞ $p(\mathbf{x}) \propto \exp(-\frac{\|\mathbf{x}\|_1}{\beta_1})$

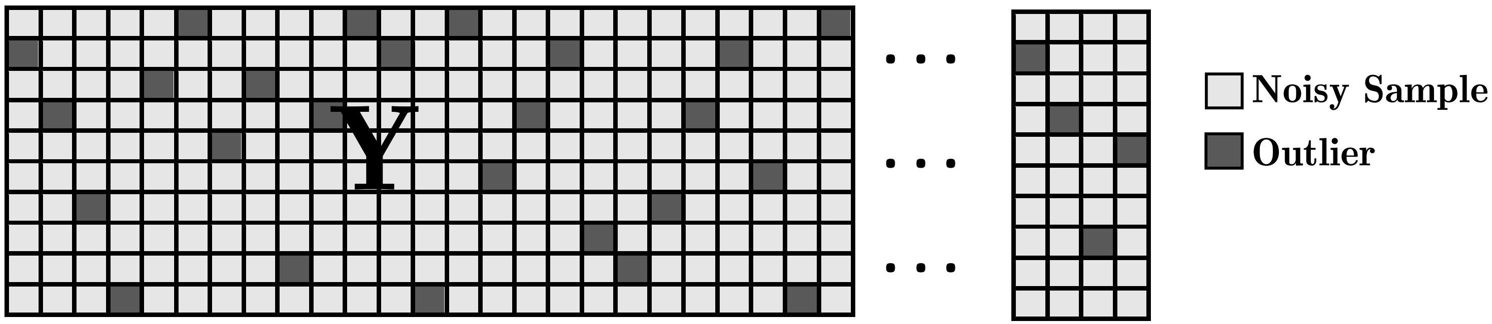
☞ $p(\mathbf{n}) \propto \exp(-\frac{\|\mathbf{n}\|_1}{\beta_2})$

- MAP Estimation of Dictionary and Representations

$$\min_{\mathbf{D}, \{\mathbf{x}_i\}_{i=1}^L} \sum_i (\|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_1 + \lambda \|\mathbf{x}_i\|_1) \text{ subject to } \|\mathbf{d}_i\|_2 = 1, \; i = 1, \dots, K$$

☞ Strategy to Solve$\Longrightarrow$ $\begin{cases} \text{w.r.t. } \mathbf{D} \Rightarrow \text{Iteratively Reweighted Least Squares} \\ \text{w.r.t. } \{\mathbf{x}_i\}_{i=1}^L \Rightarrow \text{Iteratively Reweighted Least Squares} \end{cases}$

Robust Dictionary Learning by Error Source Decomposition



Problem Formulation

- Training Data Model:

$$\mathbf{y}_i = \mathbf{D}\mathbf{x}_i + \mathbf{n}_i + \mathbf{o}_i \quad i = 1, \ldots, L$$

☞ $p(\mathbf{x}) \propto \exp(-\frac{\|\mathbf{x}\|_1}{\beta_1})$

☞ $p(\mathbf{n}) \propto \exp(-\frac{\|\mathbf{n}\|_2^2}{\beta_2})$

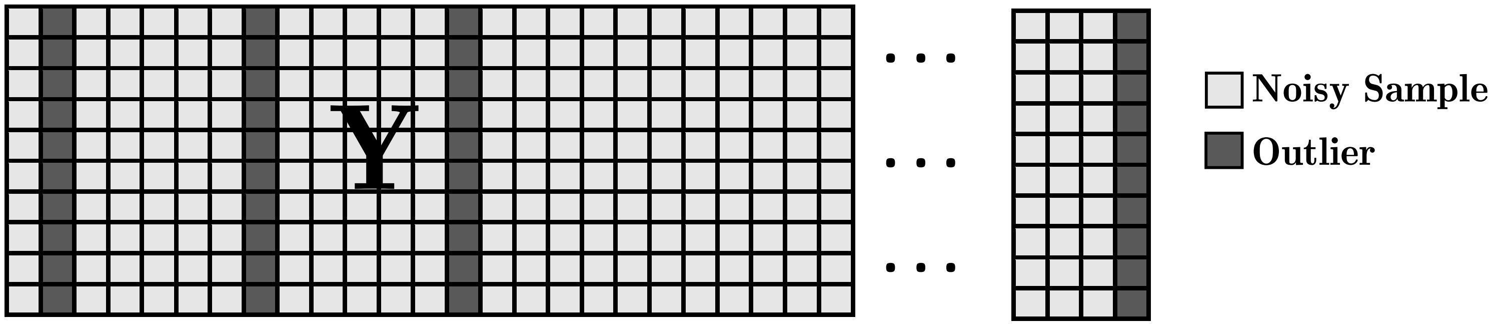
☞ $p(\mathbf{o}) \propto \exp(-\frac{\|\mathbf{o}\|_1}{\beta_3})$

- MAP Estimation of Dictionary and Representations

$$\min_{\mathbf{D},\{\mathbf{x}_i\}_{i=1}^L,\{\mathbf{o}_i\}_{i=1}^L} \sum_i (\|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i - \mathbf{o}_i\|_2^2 + \lambda\|\mathbf{o}_i\|_1) \text{ subject to } \begin{cases} \|\mathbf{d}_i\|_2 = 1, \ i = 1, \ldots, K \\ \|\mathbf{x}_j\|_0 \leq T_0, \ j = 1, \ldots, L \end{cases}$$

☞ Strategy to Solve $\Longrightarrow \begin{cases} \text{w.r.t. } \mathbf{D} \Rightarrow \text{Differentiating} \\ \text{w.r.t. } \{\mathbf{x}_i\}_{i=1}^L, \{\mathbf{o}_i\}_{i=1}^L \Rightarrow \text{Shrinkage} \end{cases}$

Outlier Aware Dictionary Learning (Proposed)



Problem Formulation

- Training Data Model:

  $$\mathbf{y}_i = \mathbf{D}\mathbf{x}_i + \mathbf{n}_i + \mathbf{o}_i \quad i = 1, \ldots, L$$

  ☞ $p(\mathbf{x}) \propto \exp(-\frac{\|\mathbf{x}\|_1}{\beta_1})$

  ☞ $p(\mathbf{n}) \propto \exp(-\frac{\|\mathbf{n}\|_2^2}{\beta_2})$

  ☞ $p(\mathbf{o}) \propto \exp(-\frac{\|\mathbf{o}\|_2}{\beta_3})$

- MAP Estimation of Dictionary and Representations

$$\min_{\mathbf{D}, \{\mathbf{x}_i\}_{i=1}^L, \{\mathbf{o}_i\}_{i=1}^L} \sum_i (\|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i - \mathbf{o}_i\|_2^2 + \lambda \|\mathbf{o}_i\|_2) \text{ subject to } \begin{cases} \|\mathbf{d}_i\|_2 = 1, \; i = 1, \ldots, K \\ \|\mathbf{x}_j\|_0 \leq T_0, \; j = 1, \ldots, L \end{cases}$$

$$\min_{\mathbf{D}, \mathbf{X}, \mathbf{O}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}_i - \mathbf{O}\|_F^2 + \lambda \|\mathbf{O}\|_{21} \text{ subject to } \begin{cases} \|\mathbf{d}_i\|_2 = 1, \; i = 1, \ldots, K \\ \|\mathbf{x}_j\|_0 \leq T_0, \; j = 1, \ldots, L \end{cases}$$

Outlier Aware Dictionary Learning (Proposed)

Solution Strategy

- Alternating Minimization:
  - ☞ w.r.t. $\{\mathbf{x}_i\}_{i=1}^{L}$ $\Rightarrow \mathbf{x}_i = \text{OMP}_{\mathbf{D}}(\mathbf{y}_i - \mathbf{o}_i), i = 1, \ldots, L$
  - ☞ w.r.t. $\{\mathbf{o}_i\}_{i=1}^{L}$ $\Rightarrow \mathbf{o}_i = \begin{cases} (1 - \frac{\lambda}{2\|\mathbf{r}_i\|_2})\mathbf{r}_i, & \text{if } \|\mathbf{r}_i\|_2 > \frac{\lambda}{2} \\ \mathbf{0}, & \text{otherwise} \end{cases}$ $\mathbf{r}_i = \mathbf{y}_i - \mathbf{D}\mathbf{x}_i, i = 1, \ldots, L$
  - ☞ w.r.t. $\mathbf{D}$ $\Rightarrow \mathbf{D} = (\mathbf{Y} - \mathbf{O})\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}$
- Initialization
  - ☞ $\mathbf{D}$ $\Rightarrow$ Overcomplete DCT
  - ☞ $\{\mathbf{o}_i\}_{i=1}^{L}$ $\Rightarrow$ $\mathbf{o}_i = \mathbf{0}, i = 1, \ldots, L$
    - At the beginning, all training signal are considered not to be an outlier.

Image Denoising Based on OADL

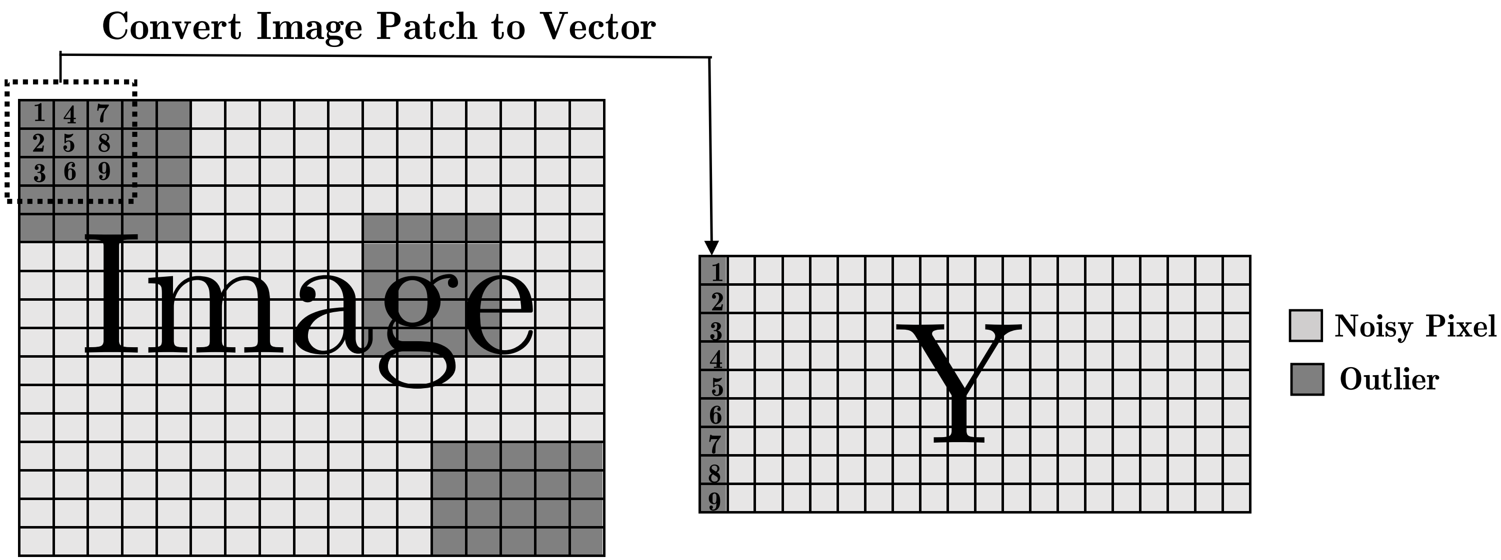


Image Denoising

- Training Image Dictionary Using OADL($\mathbf{D}^\star$)
- Estimating Denoised Image Patch Representation
  - ☞ $(\hat{\mathbf{x}}, \hat{\mathbf{o}}) =$ Alternating Minimization $\begin{cases} \min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ subject to } \|(\mathbf{y}-\mathbf{o}) - \mathbf{D}\mathbf{x}\|_2 \le \epsilon \\ \min_{\mathbf{o}} \|(\mathbf{y}-\mathbf{D}\mathbf{x}) - \mathbf{o}\|_2^2 + \|\mathbf{o}\|_2 \end{cases}$
- Reconstructing Denoised Image Patch
  - ☞ $\hat{\mathbf{y}} = \mathbf{D}^\star \hat{\mathbf{x}}$

Simulation Results

- Synthetic Data
  - Generate a random dictionary ($\mathbf{D}$)
  - Generate 2500 training signals and 500 test signals using 3 atoms of $\mathbf{D}$
  - Add Gaussian noise ($\mathcal{N}(\mathbf{0}, 0.01^2\mathbf{I})$) to each of training signals
  - Add Gaussian noise ($\mathcal{N}(\mathbf{0}, 0.04^2\mathbf{I})$) to $p\%$ of randomly selected training signals (outlier)
  - Train a dictionary using training signals ($\mathbf{D}^\star$)
  - Evaluate the ability of dictionary to code test signals using 3 atoms of $\mathbf{D}^\star$
- Image Denoising
  - Select 6 benchmark images (256×256)
  - Add Gaussian noise ($\mathcal{N}(0, 10^2)$) to each of image pixels
  - Add Gaussian noise ($\mathcal{N}(0, 20^2)$) to B blocks of pixels according to the following pattern
  - Denoise the resultant image

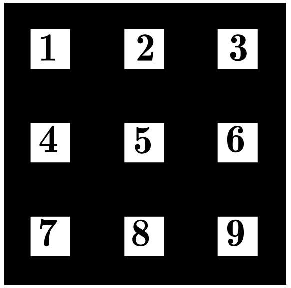
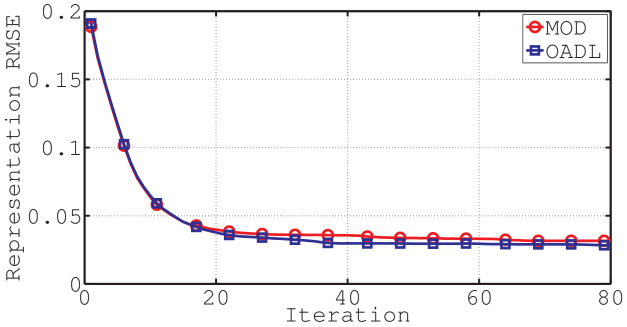
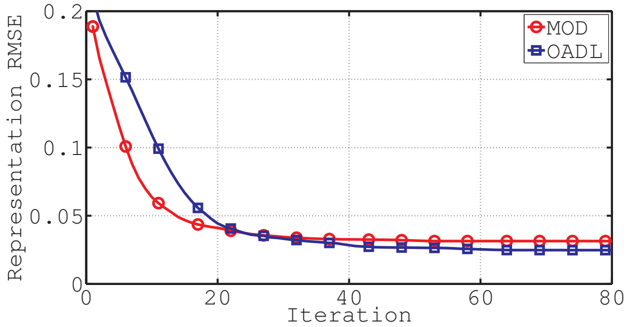


Figure: Outlier block pattern

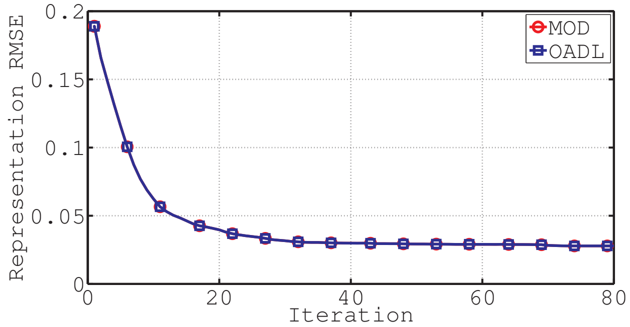
Simulation Results - Synthetic Data



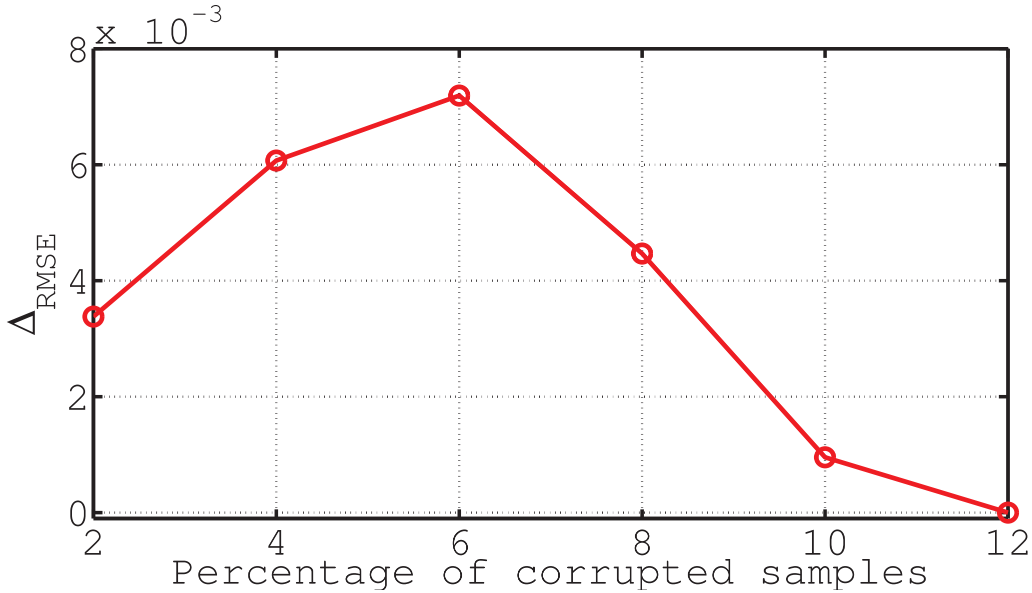
(a) $p = 2\%$



(b) $p = 6\%$



(c) $p = 12\%$



(d) Final RMSE difference

Figure: (a)-(c) Test data representation RMSE along iterations for $p =$ 2, 4 and 6, respctively. (d) Final representation RMSE of test data versus p

Simulation Results - Synthetic Data

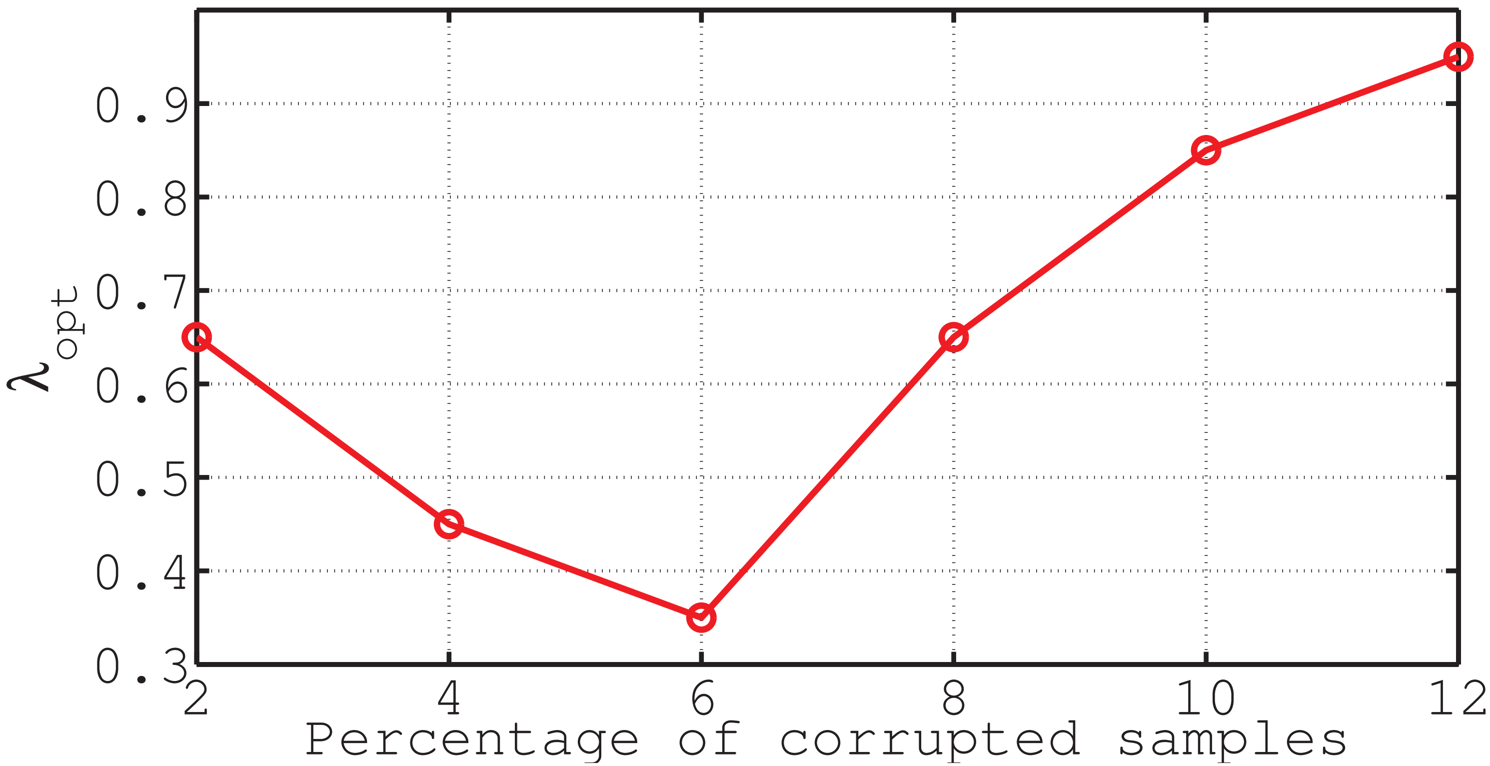


Figure: Best λ, which minimizes test data representation RMSE, versus percentage of outliers.

- $p = 2\%$ to 6% $\begin{cases} p \text{ directly related to } \beta_3 \\ \lambda \text{ inversely related to } \beta_3 \end{cases}$ $\Rightarrow p$ inversely related to λ
- $p = 6\%$ to 12% $\begin{cases} \text{Violation of outlier sparsity assumption} \\ \text{OADL approaches regular Dictionary Learning} \end{cases}$ $\Rightarrow p$ directly related to λ

Simulation Results - Image Denoising

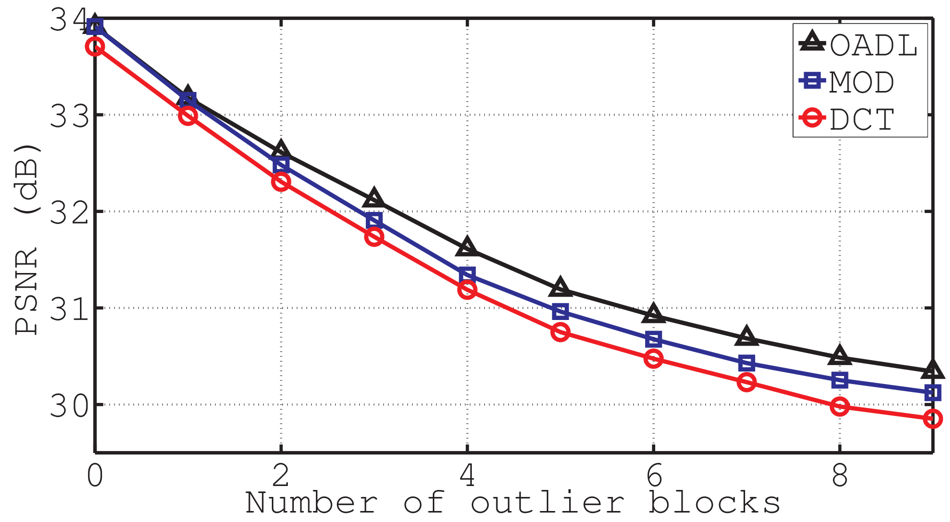


Figure: Averaged PSNR over 6 different test images versus number of outlier blocks.

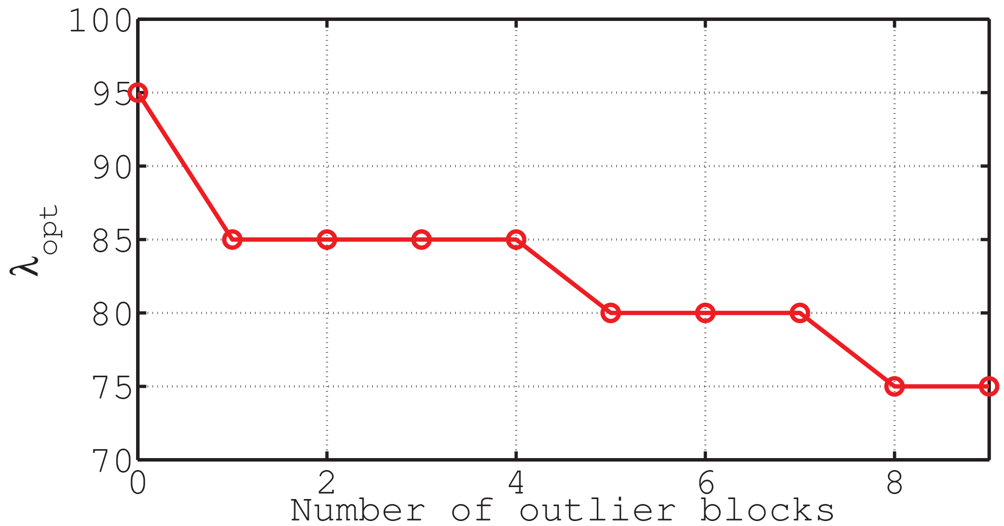


Figure: Best λ, which maximizes PSNR, versus percentage of outliers.

Conclusions

- A new and practical placement of outlier was considered.
- We introduce a new model for training signals based on separating noise and outlier source.
- We formulate DL problem using MAP estimation.
- We introduce a fast and efficient algorithm to solve the proposed robust dictionary learning problem.
- Simulation results showed that our new method leads to considerable improvements over traditional methods

References

1. K. Engan, S. O. Aase, and J. Hakon Husoy, "Method of optimal directions for frame design," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 2443–2446, 1999.
2. M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.
3. M. Elad and M. Aharon, "Image Denoising Via Sparse and Redundant representations over Learned Dictionaries," *IEEE Trans. on Image Processing*, vol. 15, no. 12, pp. 3736-3745, December 2006.
4. C. Lu, J. Shi, and J. Jia, "Online robust dictionary learning," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 415–422. June 2013.
5. Z. Chen, and Y. Wu, "Robust Dictionary Learning by Error Source Decomposition," *IEEE International Conference on Computer Vision (ICCV)*, pp. 2216–2223, December 2013.