

2. A Trilateral Weighted Sparse Coding Scheme

Topics

01. Real-world (RW) noisy image vs synthetic noisy image with additive white Gaussian noise(AWGN)
02. Sparse coding (SC) and Trilateral weighted sparse coding (TWSC)
03. Proposed denoising algorithm
 - 3.1. The Trilateral Weighted Sparse Coding Model
 - 3.2. The Setting of Weight Matrices
 - 3.3. Model Optimization
 - 3.4. The Denoising Algorithm
04. Visualization of The Weight Matrices

1. RW noisy image vs noisy image with AWGN

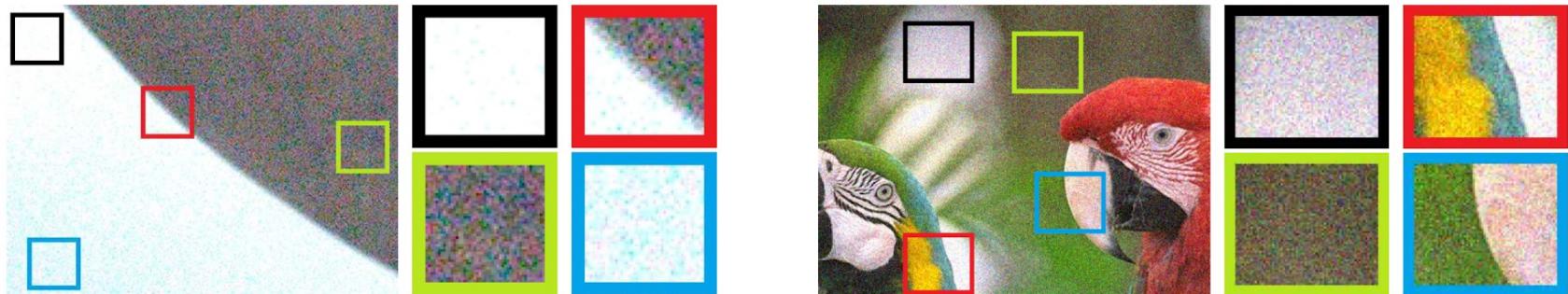


Fig. 1: Comparison of noisy image patches in real-world noisy image (left) and synthetic noisy image with additive white Gaussian noise (right).

1. RW noisy image vs noisy image with AWGN

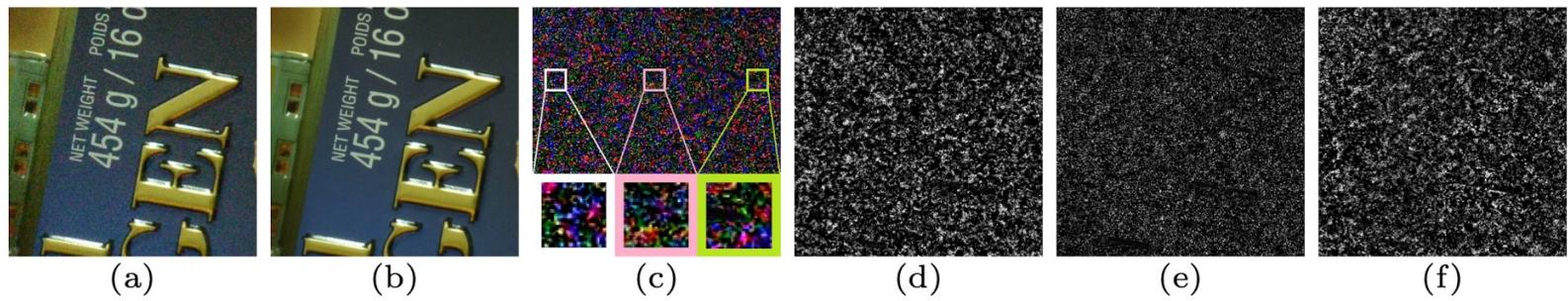


Fig. 2: An example of realistic noise. (a) A real-world noisy image captured by a Nikon D800 camera with ISO = 6400; (b) the “Ground Truth” image (please refer to Section 4.3) of (a); (c) difference between (a) and (b) (amplified for better illustration); (d)-(f) red, green, and blue channel of (c), respectively. The standard deviations (stds) of noise in the three boxes (white, pink, and green) plotted in (c) are 5.2, 6.5, and 3.3, respectively, while the stds of noise in each channel (d), (e), and (f) are 5.8, 4.4, and 5.5, respectively.

2. SC and TWSC

Denoising By Energy Minimization


$$f(\underline{x}) = \frac{1}{2} \|\underline{x} - \underline{y}\|_2^2 + G(\underline{x})$$

\underline{y} : Given measurements
 \underline{x} : Unknown to be recovered

Relation to measurements Prior or regularization

- This is in-fact a Bayesian point of view, adopting the Maximum-A-posteriori Probability (MAP) estimation.



Thomas Bayes
1702 - 1761

4

[3] <https://www.coursera.org/learn/image-processing>

2. SC and TWSC

Denoising By Energy Minimization


$$f(\underline{x}) = \frac{1}{2} \|\underline{x} - \underline{y}\|_2^2 + G(\underline{x})$$

\underline{y} : Given measurements
 \underline{x} : Unknown to be recovered

- This is in-fact a Bayesian point of view, adopting the Maximum-A-posteriori Probability (MAP) estimation.
- Clearly, the wisdom in such an approach is within the choice of the prior – **modeling the images** of interest.



Thomas Bayes
1702 - 1761

4

[3] <https://www.coursera.org/learn/image-processing>

2. SC and TWSC

The Evolution of $G(\underline{x})$

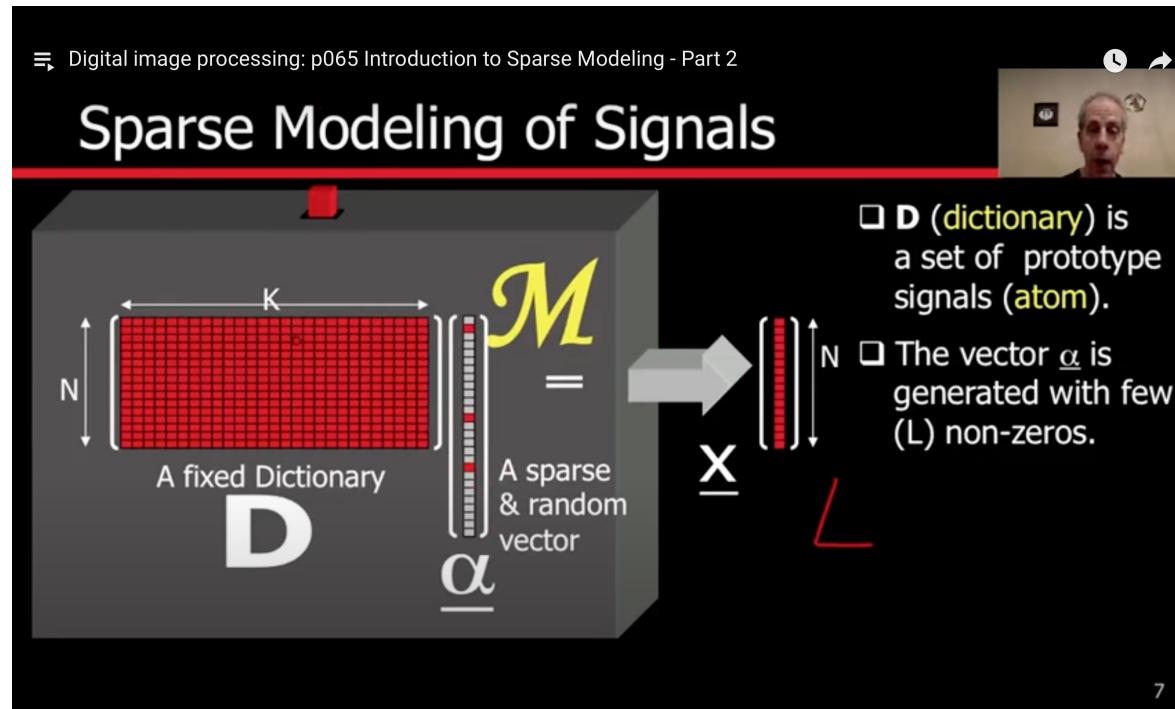
During the past several decades we have made all sort of guesses about the prior $G(\underline{x})$ for images:

$G(\underline{x}) = \lambda \ \underline{x}\ _2^2$	$G(\underline{x}) = \lambda \ \mathbf{L}\underline{x}\ _2^2$	$G(\underline{x}) = \lambda \ \mathbf{L}\underline{x}\ _{\mathbf{w}}^2$	$G(\underline{x}) = \lambda \rho \{\mathbf{L}\underline{x}\}$
Energy	Smoothness	Adapt+Smooth	Robust Statistics
$G(\underline{x}) = \lambda \ \nabla \underline{x}\ _1$	$G(\underline{x}) = \lambda \ \mathbf{W}\underline{x}\ _1$	$G(\underline{x}) = \lambda \ \underline{\alpha}\ _0$ for $\underline{x} = \mathbf{D}\underline{\alpha}$	<ul style="list-style-type: none">• Hidden Markov Models,• Compression algorithms as priors,• ...
Total-Variation	Wavelet Sparsity	Sparse & Redundant	

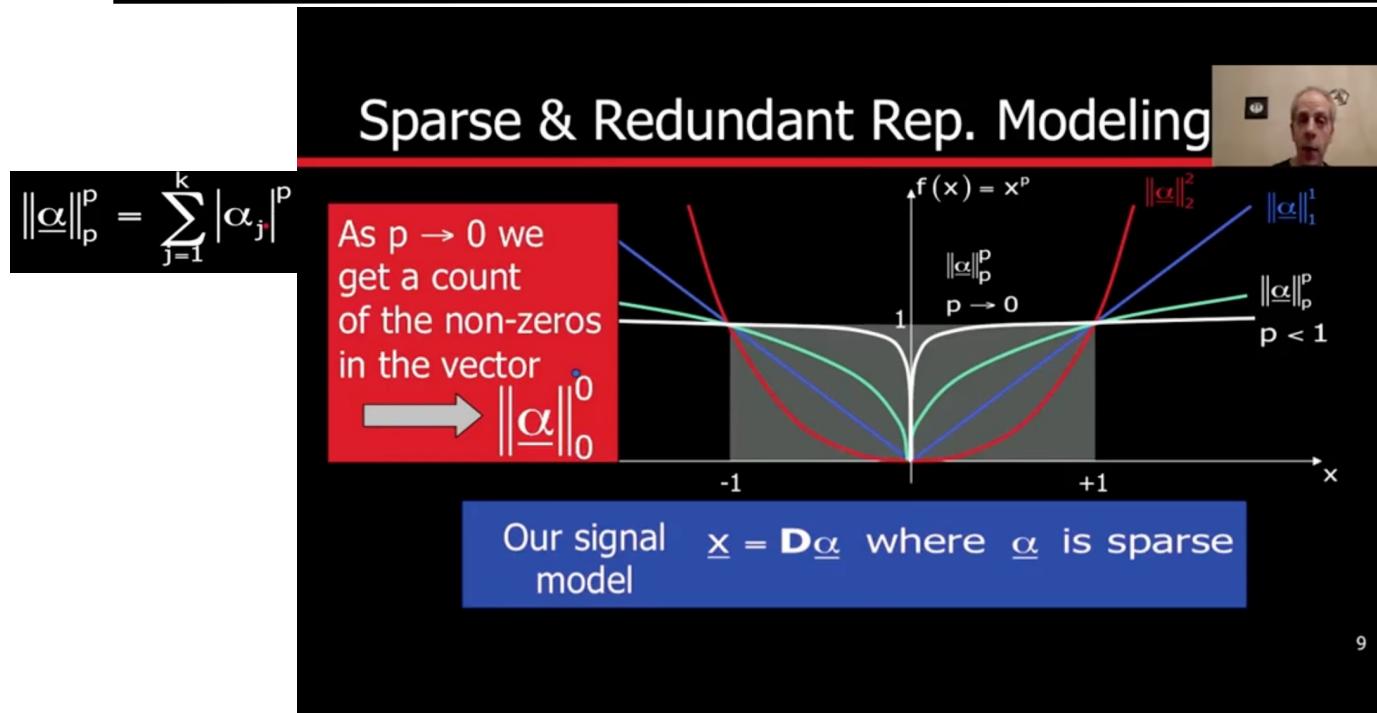
5

[3] <https://www.coursera.org/learn/image-processing>

2. SC and TWSC



2. SC and TWSC



[3] <https://www.coursera.org/learn/image-processing>

2. SC and TWSC

Back to Our MAP Energy Function



- L_0 "norm" is effectively counting the number of non-zeros in $\underline{\alpha}$.

$$\hat{\underline{\alpha}} = \arg \min_{\underline{\alpha}} \frac{1}{2} \|\mathbf{D}\underline{\alpha} - \mathbf{y}\|_2^2 \text{ s.t. } \|\underline{\alpha}\|_0 \leq L$$
$$\hat{\mathbf{x}} = \mathbf{D}\hat{\underline{\alpha}}$$

- The vector $\underline{\alpha}$ is the representation signal x .

$$\mathbf{D}\underline{\alpha} - \mathbf{y} =$$

- Few (L out of K) atoms can be combined to form the true signal, the noise cannot be fitted well. We obtain an effective projection of the noise onto a very low-dimensional space:
Denoising



10

2. SC and TWSC

Wait! There are Some Issues 

- **Numerical Problems:** How should we solve or approximate the solution of the problem

$$\min_{\underline{\alpha}} \|\mathbf{D}\underline{\alpha} - \mathbf{y}\|_2^2 \text{ s.t. } \|\underline{\alpha}\|_0^0 \leq L$$
$$\min_{\underline{\alpha}} \|\underline{\alpha}\|_0^0 \text{ s.t. } \|\mathbf{D}\underline{\alpha} - \mathbf{y}\|_2^2 \leq \varepsilon^2$$
$$\min_{\underline{\alpha}} \lambda \cdot \|\underline{\alpha}\|_0^0 + \|\mathbf{D}\underline{\alpha} - \mathbf{y}\|_2^2$$

- **Theoretical Problems:** Is there a unique sparse representation? If we are to approximate the solution somehow, how close will we get?
- **Practical Problems:** What dictionary \mathbf{D} should we use, such that all this leads to effective denoising? Will all this work in applications?

11

2. SC and TWSC

Wait! There are Some Issues 

- **Numerical Problems:** How should we solve or approximate the solution of the problem

$$\min_{\underline{\alpha}} \|\mathbf{D}\underline{\alpha} - \mathbf{y}\|_2^2 \text{ s.t. } \|\underline{\alpha}\|_0^0 \leq L$$
$$\min_{\underline{\alpha}} \|\underline{\alpha}\|_0^0 \text{ s.t. } \|\mathbf{D}\underline{\alpha} - \mathbf{y}\|_2^2 \leq \varepsilon^2$$
$$\min_{\underline{\alpha}} \lambda \cdot \|\underline{\alpha}\|_0^0 + \|\mathbf{D}\underline{\alpha} - \mathbf{y}\|_2^2$$

- **Theoretical Problems:** Is there a unique sparse representation? If we are to approximate the solution somehow, how close will we get?
- **Practical Problems:** What dictionary \mathbf{D} should we use, such that all this leads to effective denoising? Will all this work in applications?

11

30

2. SC and TWSC

Lets Start with the Noiseless Problem

Suppose we build a signal by the relation $\mathbf{D}\underline{\alpha} = \underline{x}$

We aim to find the signal's representation:

$$\hat{\underline{\alpha}} = \underset{\underline{\alpha}}{\text{Arg Min}} \|\underline{\alpha}\|_0^0 \text{ s.t. } \underline{x} = \mathbf{D}\underline{\alpha}$$

Why should we necessarily get $\hat{\underline{\alpha}} = \underline{\alpha}$?
It might happen that eventually $\|\hat{\underline{\alpha}}\|_0^0 < \|\underline{\alpha}\|_0^0$.

Uniqueness



14

2. SC and TWSC

Our Goal


$$\min_{\underline{\alpha}} \|\underline{\alpha}\|_0^0 \text{ s.t. } \|\mathbf{D}_{\underline{\alpha}} - \mathbf{y}\|_2^2 \leq \varepsilon^2$$

This is a combinatorial problem, proven to be NP-Hard!

Recipe for solving this problem:

```
graph LR; A[Set L=1] --> B[Gather all the supports {S_i}_i of cardinality L]; B --> C[Solve the LS problem  
for each support  
 $\min_{\underline{\alpha}} \|\mathbf{D}_{\underline{\alpha}} - \mathbf{y}\|_2^2 \text{ s.t. } \text{supp}(\underline{\alpha}) = S_i$ ]; C --> D[LS error  $\leq \varepsilon^2$ ?]; D -- No --> E[Set L=L+1]; D -- Yes --> F[Done];  
B -- "There are  $\binom{K}{L}$  such supports" --> C;
```

Assume: K=1000, L=10 (known!), 1 nano-sec per each LS
We shall need $\sim 8e+6$ years to solve this problem !!!!!

15

2. SC and TWSC

Lets Approximate


$$\min_{\underline{\alpha}} \|\underline{\alpha}\|_0^0 \text{ s.t. } \|\mathbf{D}\underline{\alpha} - \mathbf{y}\|_2^2 \leq \varepsilon^2$$

Relaxation methods

Smooth the L_0 and use continuous optimization techniques

Greedy methods

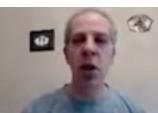
Build the solution one non-zero element at a time

16

33

2. SC and TWSC

Relaxation – The Basis Pursuit (BP)



Instead of solving
 $\underset{\underline{\alpha}}{\text{Min}} \|\underline{\alpha}\|_0^0 \text{ s.t. } \|\mathbf{D}\underline{\alpha} - \underline{y}\|_2 \leq \varepsilon$

Solve Instead
 $\underset{\underline{\alpha}}{\text{Min}} \|\underline{\alpha}\|_1 \text{ s.t. } \|\mathbf{D}\underline{\alpha} - \underline{y}\|_2 \leq \varepsilon$

- This is known as the Basis-Pursuit (BP) [Chen, Donoho & Saunders ('95)].
- The newly defined problem is convex (quad. programming).
- Very efficient solvers can be deployed:
 - Interior point methods [Chen, Donoho, & Saunders ('95)] [Kim, Koh, Lustig, Boyd, & D. Gorinevsky ('07)].
 - Sequential shrinkage for union of ortho-bases [Bruce et.al. ('98)].
 - Iterative shrinkage [Figueiredo & Nowak ('03)] [Daubechies, Defrise, & De-Mole ('04)] [E. ('05)] [E., Matalon, & Zibulevsky ('06)] [Beck & Teboulle ('09)] ...

17

2. SC and TWSC

Go Greedy: Matching Pursuit (MP)



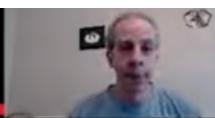
- The MP is one of the greedy algorithms that finds one atom at a time [Mallat & Zhang ('93)].
- Step 1: find the one atom that **best matches** the signal.
- Next steps: given the previously found atoms, find the next **one** to **best fit** the residual.
- The algorithm stops when the error $\|\mathbf{D}\underline{\alpha} - \mathbf{y}\|_2$ is below the destination threshold.
- The Orthogonal MP (OMP) is an improved version that re-evaluates the coefficients by Least-Squares after each round.

$$\left[\begin{array}{c|c|c|c} \text{Red} & \text{White} & \text{Red} & \text{White} \\ \text{Matrix A} & & & \end{array} \right] \left[\begin{array}{c} \text{Red} \\ \vdots \\ \text{Vector x} \end{array} \right] \approx \left[\begin{array}{c} \text{Red} \\ \vdots \\ \text{Vector b} \end{array} \right]$$

A diagram illustrating the Matching Pursuit (MP) algorithm. On the left, a matrix \mathbf{D} is shown with red and white columns, representing the dictionary. To its right is the signal vector \mathbf{y} , consisting of vertical bars of varying heights. An arrow points from the matrix towards the signal, indicating the process of finding the best matching atom. The symbol \approx is placed between the matrix and the signal, signifying the approximation of the signal by the selected atoms.

2. SC and TWSC

What Should \mathbf{D} Be?


$$\hat{\underline{\alpha}} = \arg \min_{\underline{\alpha}} \|\underline{\alpha}\|_0^0 \text{ s.t. } \frac{1}{2} \|\mathbf{D}\underline{\alpha} - \underline{y}\|_2^2 \leq \varepsilon^2 \rightarrow \hat{\underline{x}} = \mathbf{D}\hat{\underline{\alpha}}$$

Our Assumption: Good-behaved Images
have a sparse representation

\mathbf{D} should be chosen such that it sparsifies the representations

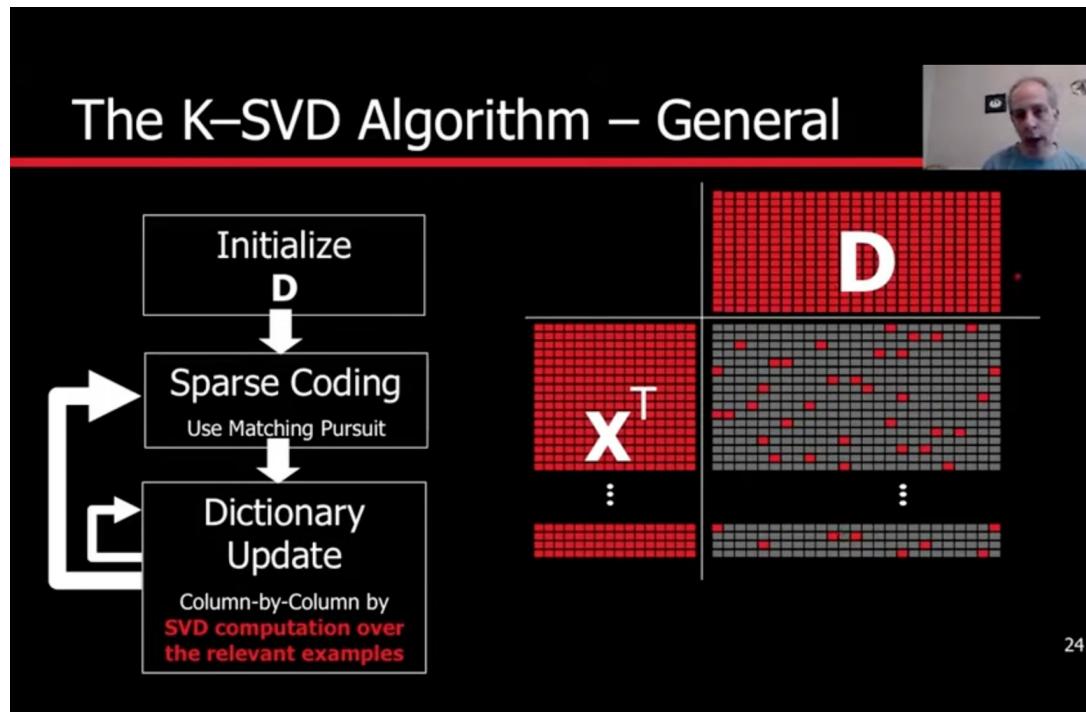
One approach to choose \mathbf{D} is from a known set of transforms (Steerable wavelet, Curvelet, Contourlets, Bandlets, Shearlets ...)



Training

Learning from Examples

2. SC and TWSC



[3] <https://www.coursera.org/learn/image-processing>

2. SC and TWSC

K-SVD: Dictionary Update Stage



We refer only to the examples that use the column \underline{d}_k

We should solve:

$$\text{Min}_{\underline{d}_k, \alpha_k} \left\| \alpha_k \underline{d}_k^T - \underline{\mathbf{E}} \right\|_F^2$$

Fixing all \mathbf{A} and \mathbf{D} apart from the k^{th} column, and seek both \underline{d}_k and the k^{th} column in \mathbf{A} to better fit the **residual!**

26

[3] <https://www.coursera.org/learn/image-processing>

2. SC and TWSC

$$\hat{\mathbf{C}} = \arg \min_{\mathbf{C}} \|\mathbf{Y} - \mathbf{DC}\|_F^2 + \lambda \|\mathbf{C}\|_1 \quad \text{Eq. (2)}$$

Data Fidelity term Regularization term

- SC based denoising methods are essentially limited by the data fidelity term described by L₂ norm which actually assumes White Gaussian Noise (WGN) and is not able to characterize the signal dependent & realistic noise.

$$\tilde{\mathbf{Y}} = \mathbf{X} + \mathbf{N} \in \mathbb{R}^{3p^2 \times M}$$

$$\begin{aligned}\mathbf{Y} &= [\mathbf{Y}_r^\top \ \mathbf{Y}_g^\top \ \mathbf{Y}_b^\top]^\top \\ \mathbf{D} &= [\mathbf{D}_r^\top \ \mathbf{D}_g^\top \ \mathbf{D}_b^\top]^\top\end{aligned}$$

$$\hat{\mathbf{X}} = \mathbf{D}\hat{\mathbf{C}}$$

[2]: Xu, J., Zhang, L. and Zhang, D., 2018. A trilateral weighted sparse coding scheme for real-world image denoising. In Proceedings of the European conference on computer vision (ECCV) (pp. 20-36).

3.1 The Trilateral Weighted Sparse Coding Model

- We introduce two weight matrices $W_1 \in \mathbb{R}^{3p^2 \times 3p^2}$ and $W_2 \in \mathbb{R}^{M \times M}$ to characterize the SC residual ($Y - DC$) in the data-fidelity term of Eq. (2). Besides, to better characterize the sparsity priors of the natural images, we introduce a third weight matrix W_3 , which is related to the distribution of the sparse coefficients matrix C , into the regularization term of Eq. (2)
- For the dictionary D , we learn it adaptively by applying the SVD to the given data matrix Y as:

$$Y = DSV^\top$$

- Note that the aim of this paper wasn't proposing a new dictionary. Once obtained from SVD, the dictionary D is fixed and not updated iteratively.

3.1 The Trilateral Weighted Sparse Coding Model

- Finally, the proposed trilateral weighted sparse coding (TWSC) model is formulated as:

$$\min_{\mathbf{C}} \|\mathbf{W}_1(\mathbf{Y} - \mathbf{DC})\mathbf{W}_2\|_F^2 + \|\mathbf{W}_3^{-1}\mathbf{C}\|_1$$

- λ has been implicitly incorporated into the weight matrix \mathbf{W}_3 .

$\mathbf{W}_1, \mathbf{W}_2$  Realistic Noise

\mathbf{W}_3  Sparsity priors of natural images

3.2. The Setting of Weight Matrices

- In this paper, we set the three weight matrices W_1 , W_2 , and W_3 as diagonal matrices and grant clear physical meanings to them.

describing noise Properties related to each channel

$$W_1 = \begin{bmatrix} R\text{-related} \\ G\text{-related} \\ B\text{-related} \end{bmatrix}$$

block diagonal matrix

$3p^2 \times 3p^2$

3.2. The Setting of Weight Matrices

$$w_2 = \begin{bmatrix} \text{elliptical shape} \end{bmatrix}_{M \times M}$$

Describing noise variance
in the corresponding
patch y .

3.2. The Setting of Weight Matrices

$w_1 \rightarrow$ regularizes row discrepancy of $(Y - DC)$
 $w_2 \rightarrow$ regularizes column discrepancy of $(Y - DC)$
 $w_3 \rightarrow$ each diagonal element is set based
on sparsity priors on C

3.2. The Setting of Weight Matrices

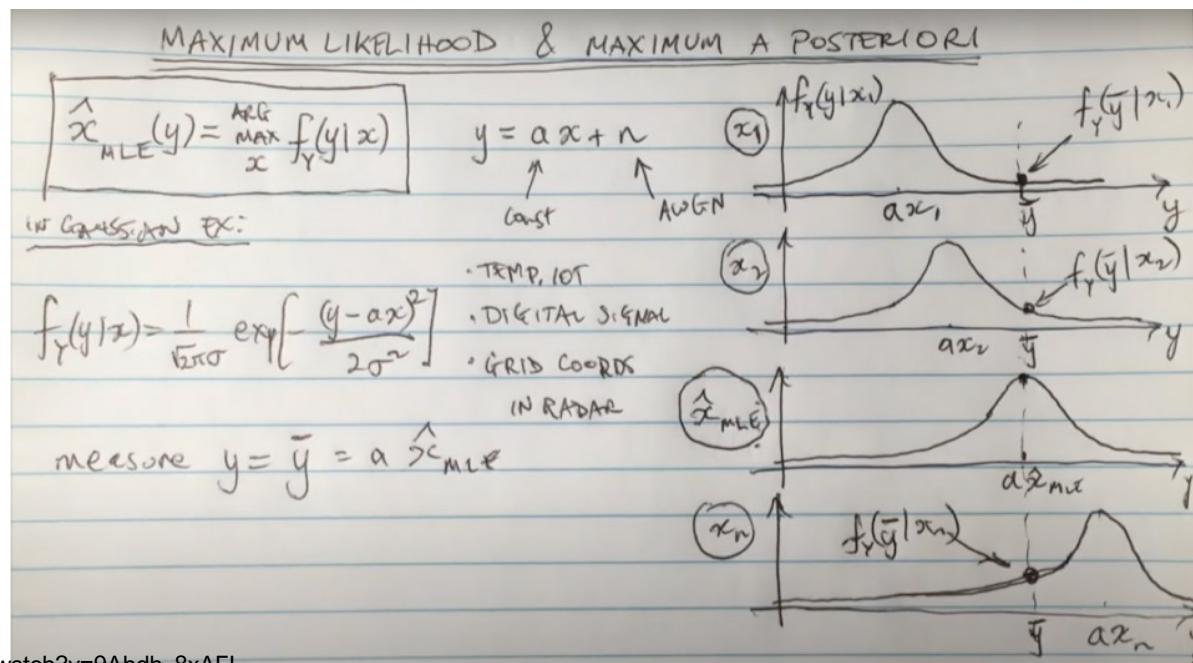
We determine the three weight matrices W_1 , W_2 , and W_3 by employing the Maximum A-Posterior (MAP) estimation technique:

$$\hat{\mathbf{C}} = \arg \max_{\mathbf{C}} \ln P(\mathbf{C}|\mathbf{Y}) = \arg \max_{\mathbf{C}} \{\ln P(\mathbf{Y}|\mathbf{C}) + \ln P(\mathbf{C})\}. \quad \text{Eq. (5)}$$

The log-likelihood term $\ln(P(\mathbf{Y}|\mathbf{C}))$ is characterized by the statistics of noise. According to, it can be assumed that the noise is independently and identically distributed (i.i.d.) in each channel and each patch with Gaussian distribution.

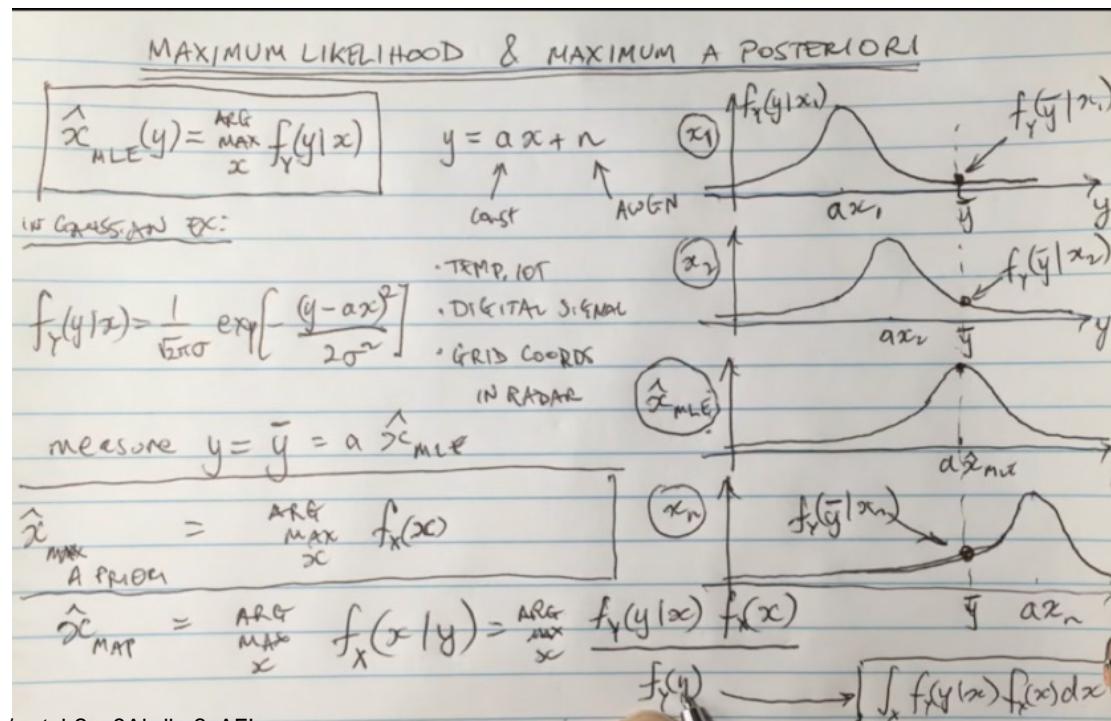
3.2. The Setting of Weight Matrices

- Maximum Likelihood & Maximum A Posteriori



[4] https://www.youtube.com/watch?v=9Ahdh_8xAEI

3.2. The Setting of Weight Matrices



[4] https://www.youtube.com/watch?v=9Ahdh_8xAEI

3.2. The Setting of Weight Matrices

- Denote by y_{cm} and c_m the m th column of the matrices \mathbf{Y}_c and C , respectively, and denote by σ_{cm} the noise std of y_{cm} . We have:

$$P(\mathbf{Y}|\mathbf{C}) = \prod_{c \in \{r,g,b\}} \prod_{m=1}^M (\pi \sigma_{cm})^{-p^2} e^{-\sigma_{cm}^{-2} \|\mathbf{y}_{cm} - \mathbf{D}_c \mathbf{c}_m\|_2^2}, \quad \text{Eq. (6)}$$

y_{cm} : mth column of \mathbf{Y}_c
 c_m : mth column of C
 $\tilde{\sigma}_{cm}$: Std of y_{cm}

$\{\tilde{\sigma}_{cm}\}_{3 \times M}$ can be estimated $\rightarrow \sigma_c, \sigma_m$

$\tilde{\sigma}_{cm} = \sigma_c^{l_1} \sigma_m^{l_2}$
 $l_1 + l_2 = 1$ (we consider: $l_1 = l_2 = 0.5$)

3.2. The Setting of Weight Matrices

- The sparsity prior is imposed on the coefficients matrix C, we assume that each column c_m of C follows i.i.d. Laplacian distribution. Specifically, for each entry c_{mi} , which is the coding coefficient of the m th patch y_m over the i th atom of dictionary D, we assume that it follows distribution of:

$$\left[\frac{1}{2S_i} e^{\left(-\frac{|c_{mi}|}{S_i} \right)} \right]$$

- Where S_i is the i th diagonal element of the singular value matrix S in :

$$\mathbf{Y} = \mathbf{DSV}^\top$$

3.2. The Setting of Weight Matrices

- Note that we set the scale factor of the distribution as the inverse of the i th singular value S_i . This is because the larger the singular value S_i is, the more important the i th atom (i.e., singular vector) in D should be, and hence the distribution of the coding coefficients over this singular vector should have stronger regularization with weaker sparsity. The prior term in Eq. (5) becomes:

$$P(\mathbf{C}) = \prod_{m=1}^M \prod_{i=1}^{3p^2} (2S_i)^{-1} e^{-S_i^{-1} |\mathbf{c}_m^i|} \quad \text{Eq. (7)}$$

3.2. The Setting of Weight Matrices

- Put (7) and (6) into (5) and consider the log-linear model $\sigma_{cm} = \sigma_c^{1/2} \sigma_m^{1/2}$, we have:

$$\begin{aligned}\hat{\mathbf{C}} &= \arg \min_{\mathbf{C}} \sum_{c \in \{r, g, b\}} \sum_{m=1}^M \sigma_{cm}^{-2} \|\mathbf{y}_{cm} - \mathbf{D}_c \mathbf{c}_m\|_2^2 + \sum_{m=1}^M \|\mathbf{S}^{-1} \mathbf{c}_m\|_1 \\ &= \arg \min_{\mathbf{C}} \sum_{c \in \{r, g, b\}} \sigma_c^{-1} \|(\mathbf{Y}_c - \mathbf{D}_c \mathbf{C}) \mathbf{W}_2\|_F^2 + \|\mathbf{S}^{-1} \mathbf{C}\|_1 \\ &= \arg \min_{\mathbf{C}} \|\mathbf{W}_1 (\mathbf{Y} - \mathbf{DC}) \mathbf{W}_2\|_F^2 + \|\mathbf{W}_3^{-1} \mathbf{C}\|_1,\end{aligned}$$

- Where:

$$\begin{aligned}\mathbf{W}_1 &= \text{diag}(\sigma_r^{-1/2} \mathbf{I}_{p^2}, \sigma_g^{-1/2} \mathbf{I}_{p^2}, \sigma_b^{-1/2} \mathbf{I}_{p^2}), \\ \mathbf{W}_2 &= \text{diag}(\sigma_1^{-1/2}, \dots, \sigma_M^{-1/2}), \mathbf{W}_3 = \mathbf{S},\end{aligned}$$

[2]: Xu, J., Zhang, L. and Zhang, D., 2018. A trilateral weighted sparse coding scheme for real-world image denoising. In Proceedings of the European conference on computer vision (ECCV) (pp. 20-36).

3.3. Model Optimization

- Letting $C^* = W_3^{-1}C$, we can transfer the weight matrix W_3 into the data fidelity term of (4). Thus, the TWSC scheme (4) is reformulated as:

$$\min_{C^*} \|W_1(Y - DW_3C^*)W_2\|_F^2 + \|C^*\|_1$$

- By introducing an augmented variable Z , the problem (10) is reformulated as a linear equality-constrained problem with two variables C and Z :

$$\min_{C, Z} \|W_1(Y - DW_3C)W_2\|_F^2 + \|Z\|_1 \text{ s.t. } C = Z.$$

- Since the objective function is separable w.r.t. the two variables, the problem can be solved under the alternating direction method of multipliers (ADMM) framework.

[2]: Xu, J., Zhang, L. and Zhang, D., 2018. A trilateral weighted sparse coding scheme for real-world image denoising. In Proceedings of the European conference on computer vision (ECCV) (pp. 20-36).

3.3. Model Optimization

- Since the objective function is separable w.r.t. the two variables, the problem can be solved under the alternating direction method of multipliers (ADMM) framework.
 - The *alternating direction method of multipliers* (ADMM) is an algorithm that solves convex optimization problems by breaking them into smaller pieces, each of which are then easier to handle. It has recently found wide application in a number of areas.
- The augmented Lagrangian function is:

$$\mathcal{L}(\mathbf{C}, \mathbf{Z}, \Delta, \rho) = \|\mathbf{W}_1(\mathbf{Y} - \mathbf{D}\mathbf{W}_3\mathbf{C})\mathbf{W}_2\|_F^2 + \|\mathbf{Z}\|_1 + \langle \Delta, \mathbf{C} - \mathbf{Z} \rangle + \frac{\rho}{2} \|\mathbf{C} - \mathbf{Z}\|_F^2,$$

- Where Δ is the augmented Lagrangian multiplier and $\rho > 0$ is the penalty parameter.
- We initialize the matrix variables \mathbf{C}_0 , \mathbf{Z}_0 , and Δ_0 to be comfortable zero matrices and $\rho_0 > 0$. Denote by $(\mathbf{C}_k, \mathbf{Z}_k)$ and Δ_k the optimization variables and Lagrange multiplier at iteration k ($k = 0, 1, 2, \dots$), respectively.

[2]: Xu, J., Zhang, L. and Zhang, D., 2018. A trilateral weighted sparse coding scheme for real-world image denoising. In Proceedings of the European conference on computer vision (ECCV) (pp. 20-36).

3.3. Model Optimization

- By taking derivatives of the Lagrangian function L w.r.t. C and Z, and setting the derivatives to be zeros, we can alternatively update the variables as follows:

1. Update C by fixing Z and Δ :

$$\mathbf{C}_{k+1} = \arg \min_{\mathbf{C}} \|\mathbf{W}_1(\mathbf{Y} - \mathbf{D}\mathbf{W}_3\mathbf{C})\mathbf{W}_2\|_F^2 + \frac{\rho_k}{2} \|\mathbf{C} - \mathbf{Z}_k + \rho_k^{-1} \boldsymbol{\Delta}_k\|_F^2$$

- This is a two-sided weighted least squares regression problem with the solution satisfying that:

$$\mathbf{A}\mathbf{C}_{k+1} + \mathbf{C}_{k+1}\mathbf{B}_k = \mathbf{E}_k, \quad \text{Eq. (14)}$$

Where:

$$\mathbf{A} = \mathbf{W}_3^\top \mathbf{D}^\top \mathbf{W}_1^\top \mathbf{W}_1 \mathbf{D} \mathbf{W}_3, \mathbf{B}_k = \frac{\rho_k}{2} (\mathbf{W}_2 \mathbf{W}_2^\top)^{-1},$$

$$\mathbf{E}_k = \mathbf{W}_3^\top \mathbf{D}^\top \mathbf{W}_1^\top \mathbf{W}_1 \mathbf{Y} + \left(\frac{\rho_k}{2} \mathbf{Z}_k - \frac{1}{2} \boldsymbol{\Delta}_k \right) (\mathbf{W}_2 \mathbf{W}_2^\top)^{-1}.$$

[2]: Xu, J., Zhang, L. and Zhang, D., 2018. A trilateral weighted sparse coding scheme for real-world image denoising. In Proceedings of the European conference on computer vision (ECCV) (pp. 20-36).

3.3. Model Optimization

1. Update \mathbf{C} by fixing \mathbf{Z} and Δ :

- Eq. (14) is a standard Sylvester equation (SE) which has a unique solution if and only if $\sigma(\mathbf{A}) \cap \sigma(-\mathbf{B}_k) = \emptyset$, where $\sigma(F)$ denotes the spectrum, i.e., the set of eigenvalues, of the matrix F . We can rewrite the SE (14) as:

$$(\mathbf{I}_M \otimes \mathbf{A} + \mathbf{B}_k^\top \otimes \mathbf{I}_{3p^2}) \text{vec}(\mathbf{C}_{k+1}) = \text{vec}(\mathbf{E}_k)$$

- and the solution \mathbf{C}_{k+1} (if existed) can be obtained via $\mathbf{C}_{k+1} = \text{vec}^{-1}(\text{vec}(\mathbf{C}_{k+1}))$, where $\text{vec}^{-1}(\bullet)$ is the inverse of the vec-operator $\text{vec}(\bullet)$.

3.3. Model Optimization

2. Update \mathbf{Z} by fixing \mathbf{C} and Δ :

$$\mathbf{Z}_{k+1} = \arg \min_{\mathbf{Z}} \frac{\rho_k}{2} \|\mathbf{Z} - (\mathbf{C}_{k+1} + \rho_k^{-1} \Delta_k)\|_F^2 + \|\mathbf{Z}\|_1.$$

- This problem has a closed-form solution as:

$$\mathbf{Z}_{k+1} = \mathcal{S}_{\rho_k^{-1}}(\mathbf{C}_{k+1} + \rho_k^{-1} \Delta_k)$$

- Where $\mathcal{S}_\lambda(x) = \text{sign}(x) \cdot \max(x - \lambda, 0)$ is the soft-thresholding operator.

3.3. Model Optimization

3. Update Δ by fixing X and Z :

$$\Delta_{k+1} = \Delta_k + \rho_k (\mathbf{C}_{k+1} - \mathbf{Z}_{k+1}).$$

[2]: Xu, J., Zhang, L. and Zhang, D., 2018. A trilateral weighted sparse coding scheme for real-world image denoising. In Proceedings of the European conference on computer vision (ECCV) (pp. 20-36).

3.3. Model Optimization

(4) Update ρ :

$$\rho_{k+1} = \mu \rho_k, \text{ where } \mu \geq 1.$$

- The above alternative updating steps are repeated until the convergence condition is satisfied or the number of iterations exceeds a preset threshold K . The ADMM algorithm converges when $\|C_{k+1} - Z_{k+1}\|_F \leq Tol$, $\|C_{k+1} - C_k\|_F \leq Tol$, and $\|Z_{k+1} - Z_k\|_F \leq Tol$ are simultaneously satisfied, where $Tol > 0$ is a small tolerance number.

3.3. Model Optimization

- **Convergence Analysis.**
 - The convergence of Algorithm 1 can be guaranteed since the overall objective function (11) is convex with a global optimal solution. In Fig. 3, we can see that the maximal values in $|\mathbf{C}_{k+1} - \mathbf{Z}_{k+1}|$, $|\mathbf{C}_{k+1} - \mathbf{C}_k|$, $|\mathbf{Z}_{k+1} - \mathbf{Z}_k|$ approach to 0 simultaneously in 50 iterations.

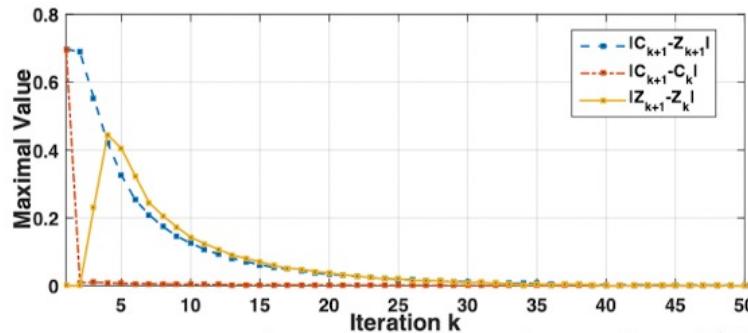


Fig. 3: The convergence curves of maximal values in entries of $|\mathbf{C}_{k+1} - \mathbf{Z}_{k+1}|$ (blue line), $|\mathbf{C}_{k+1} - \mathbf{C}_k|$ (red line), and $|\mathbf{Z}_{k+1} - \mathbf{Z}_k|$ (yellow line). The test image is the image in Fig. 2 (a).

3.3. Model Optimization

- **Summary**

Algorithm 1: Solve the TWSC Model (4) via ADMM

Input: $\mathbf{Y}, \mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \mu, \text{Tol}, K_1$;
Initialization: $\mathbf{C}_0 = \mathbf{Z}_0 = \Delta_0 = \mathbf{0}$, $\rho_0 > 0$, $k = 0$, $T = \text{False}$;
While ($T == \text{false}$) **do**
 1. Update \mathbf{C}_{k+1} by solving Eq. (13);
 2. Update \mathbf{Z}_{k+1} by soft thresholding (18);
 3. Update Δ_{k+1} by Eq. (19);
 4. Update ρ_{k+1} by $\rho_{k+1} = \mu\rho_k$, where $\mu \geq 1$;
 5. $k \leftarrow k + 1$;
 if (Converged) or ($k \geq K_1$)
 6. $T \leftarrow \text{True}$;
 end if
 end while
Output: Matrices \mathbf{C} and \mathbf{Z} .

[2]: Xu, J., Zhang, L. and Zhang, D., 2018. A trilateral weighted sparse coding scheme for real-world image denoising. In Proceedings of the European conference on computer vision (ECCV) (pp. 20-36).

3.4. The Denoising Algorithm

- Given a noisy color image, suppose that we have extracted N local patches $\{y_j\}_{j=1}^N$ and their similar patches. Then N noisy patch matrices $\{Y_j\}_{j=1}^N$ can be formed to estimate the clean patch matrices $\{X_j\}_{j=1}^N$. The patches in matrices $\{X_j\}_{j=1}^N$ are aggregated to form the denoised image $\hat{x_c}$. To obtain better denoising results, we perform the above denoising procedures for several (e.g., K 2) iterations.

3.4. The Denoising Algorithm

- Summary

Algorithm 2: Image Denoising by TWSC

Input: Noisy image \mathbf{y}_c , $\{\sigma_r, \sigma_g, \sigma_b\}$, K_2 ;

Initialization: $\hat{\mathbf{x}}_c^{(0)} = \mathbf{y}_c$, $\mathbf{y}_c^{(0)} = \mathbf{y}_c$;

for $k = 1 : K_2$ **do**

 1. Set $\mathbf{y}_c^{(k)} = \hat{\mathbf{x}}_c^{(k-1)}$;

 2. Extract local patches $\{\mathbf{y}_j\}_{j=1}^N$ from $\mathbf{y}_c^{(k)}$;

for each patch \mathbf{y}_j **do**

 3. Search nonlocal similar patches \mathbf{Y}_j ;

 4. Apply the TWSC scheme (4) to \mathbf{Y}_j and obtain the estimated $\mathbf{X}_j = \mathbf{DC}$;

end for

 5. Aggregate $\{\mathbf{X}_j\}_{j=1}^N$ to form the image $\hat{\mathbf{x}}_c^{(k)}$;

end for

Output: Denoised image $\hat{\mathbf{x}}_c^{(K_2)}$.

4.1 Visualization of The Weight Matrices

- The three diagonal weight matrices in the proposed TWSC model (4) have clear physical meanings, and it is interesting to analyze how the matrices actually relate to the input image by visualizing the resulting matrices.
- One can see that the matrix W_1 reflects well the noise levels in the images. Though matrix W_2 is initialized as an identity matrix, it is changed in iterations since noise in different patches are removed differently. For real-world noisy images, the noise levels of different patches in Y are different, hence the elements of W_2 vary a lot. In contrast, the noise levels of patches in the synthetic noisy image are similar, thus the elements of W_2 are similar. The weight matrix W_3 is basically determined by the patch structure but not noise, and we do not plot it here.

4.1 Visualization of The Weight Matrices

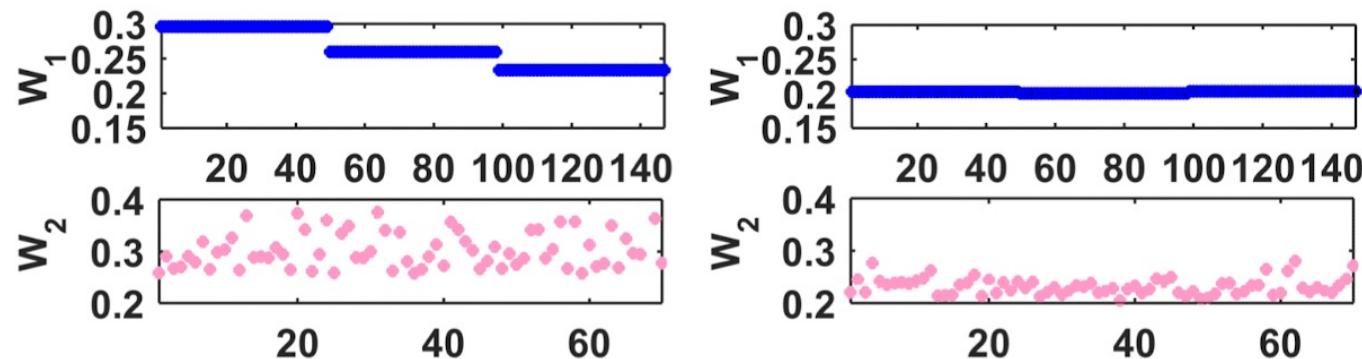


Fig. 7: Visualization of weight matrices \mathbf{W}_1 and \mathbf{W}_2 on the real-world noisy image (left) and the synthetic noisy image (right) shown in Fig. 1.

References:

- [1]: Zhu, L., Fu, C.W., Brown, M.S. and Heng, P.A., 2017. A non-local low-rank framework for ultrasound speckle reduction. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5650-5658).
- [2]: Xu, J., Zhang, L. and Zhang, D., 2018. A trilateral weighted sparse coding scheme for real-world image denoising. In Proceedings of the European conference on computer vision (ECCV) (pp. 20-36).
- [3] <https://www.coursera.org/learn/image-processing>
- [4] https://www.youtube.com/watch?v=9Ahdh_8xAEI