SOS Boosting of Image Denoising Algorithms*

Michael Elad
The Computer Science Department
The Technion – Israel Institute of technology
Haifa 32000, Israel

January
2015
Switzerland

* Joint work with Yaniv Romano

The research leading to these results has received funding from the European Research Council under European Union's Seventh Framework Program, ERC Grant agreement no. 320649, and by the Intel Collaborative Research Institute for Computational Intelligence

Leading Image Denoising Methods

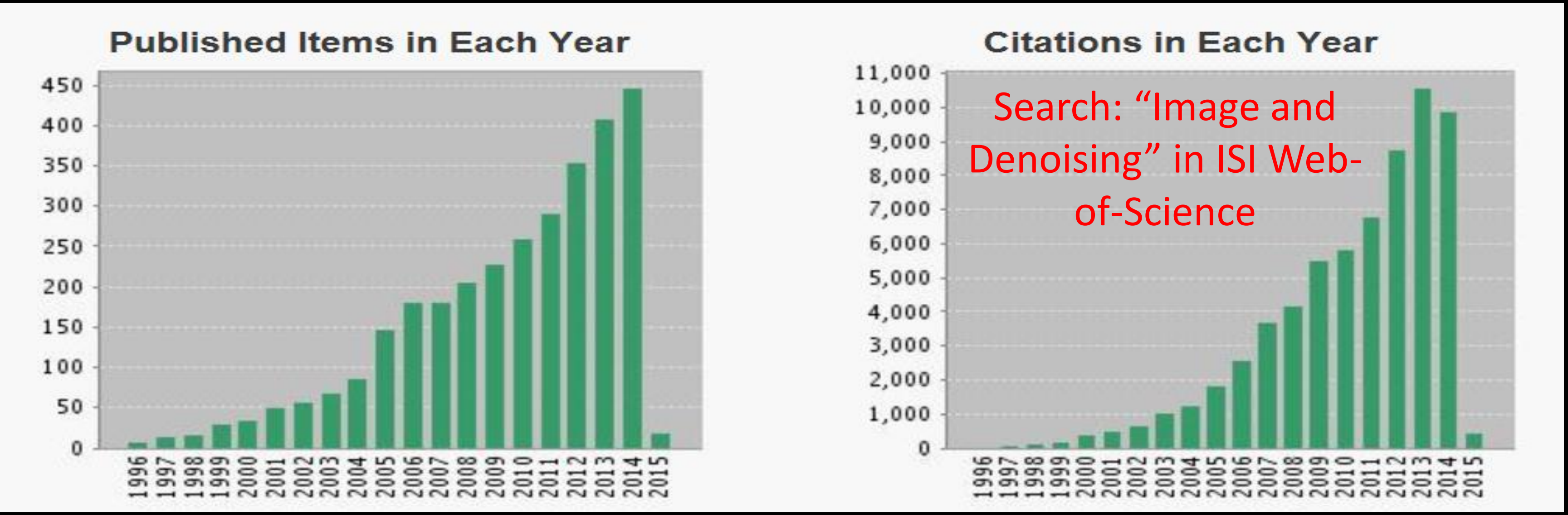
Are built upon powerful patch-based (local) image models:

- Non-Local Means (NLM): self-similarity within natural images
- K-SVD: sparse representation modeling of image patches
- BM3D: combines a sparsity prior and non local self-similarity
- Kernel-regression: offers a local directional filter
- EPLL: exploits a GMM model of the image patches
- …

Today we present a way to improve various such state-of-the-art image denoising methods, simply by applying the original algorithm as a “black-box” several times

Leading Image Denoising Methods

Are built upon powerful patch-based (local) image models:



Today we present a way to improve various such state-of-the-art image denoising methods, simply by applying the original algorithm as a "black-box" several times

Background

Boosting Methods for Denoising

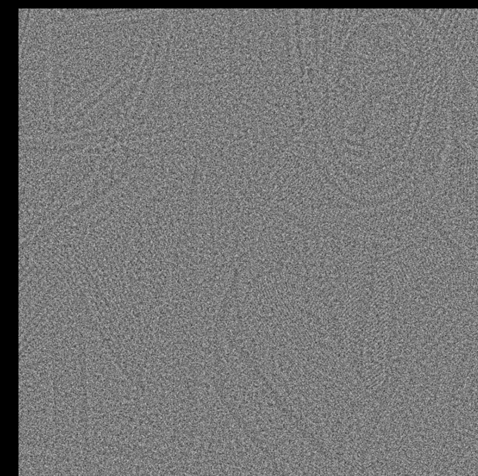
- Improved results can be achieved by processing the residual/method-noise image:

Noisy image

$\mathbf{y}$

Denoised image

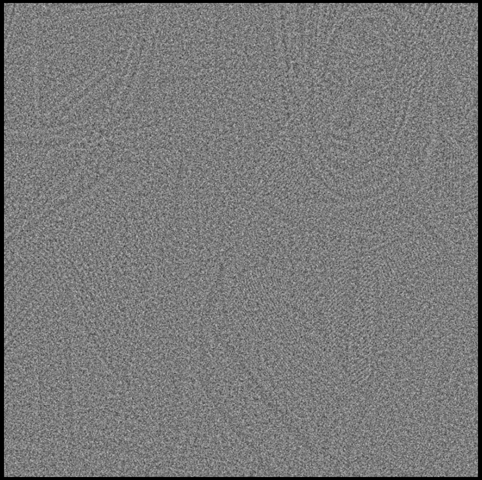
$\hat{\mathbf{x}} = f(\mathbf{y})$



Method Noise

$\mathbf{y} - \hat{\mathbf{x}}$

Processing the Residual Image



- Twicing [Tukey ('77), Charest et al. ('06)]
  - $\hat{\mathbf{x}}^{k+1} = \hat{\mathbf{x}}^k + f\left(\mathbf{y} - \hat{\mathbf{x}}^k\right)$
  - Method noise whitening [Romano & Elad ('13)]

    Recovering the "stolen" content from the method-noise using the same basis elements that were chosen to represent the initially denoised patches

- TV denoising using Bregman distance [Bregman ('67), Osher et al. ('05)]
  - $\hat{\mathbf{x}}^{k+1} = f\left(\hat{\mathbf{x}}^k + \sum_{i=1}^{k}\left(\mathbf{y} - \hat{\mathbf{x}}^i\right)\right)$

Boosting Methods

- Diffusion [Perona-Malik ('90), Coifman et al. ('06), Milanfar ('12)]
  - Removes the noise leftovers that are found in the denoised image
  - $\hat{\mathbf{x}}^{k+1} = f\left(\hat{\mathbf{x}}^{k}\right)$
- SAIF [Talebi et al. ('12)]
  - Chooses automatically the local improvement mechanism:
    - Diffusion
    - Twicing

Reducing the Local/Global Gap

- EPLL [Zoran & Weiss ('09), Sulam & Elad ('14)]
  - Treats a major shortcoming of patch-based methods:
    - The gap between the local patch processing and the global need for a whole restored image
  - By encouraging the patches of the final image (i.e. after patch aggregation) to comply with the local prior
  - In practice – iterated denoising with a diminishing variance

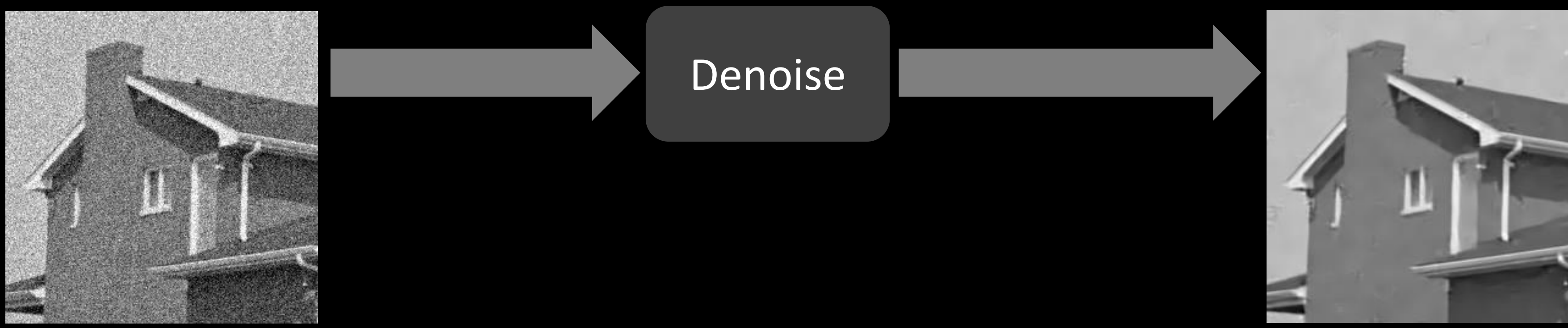
    I. Denoising the patches of $\hat{\mathrm{x}}^k$

    II. Obtain $\hat{\mathrm{x}}^{k+1}$ by averaging the overlapping patches

SOS Boosting

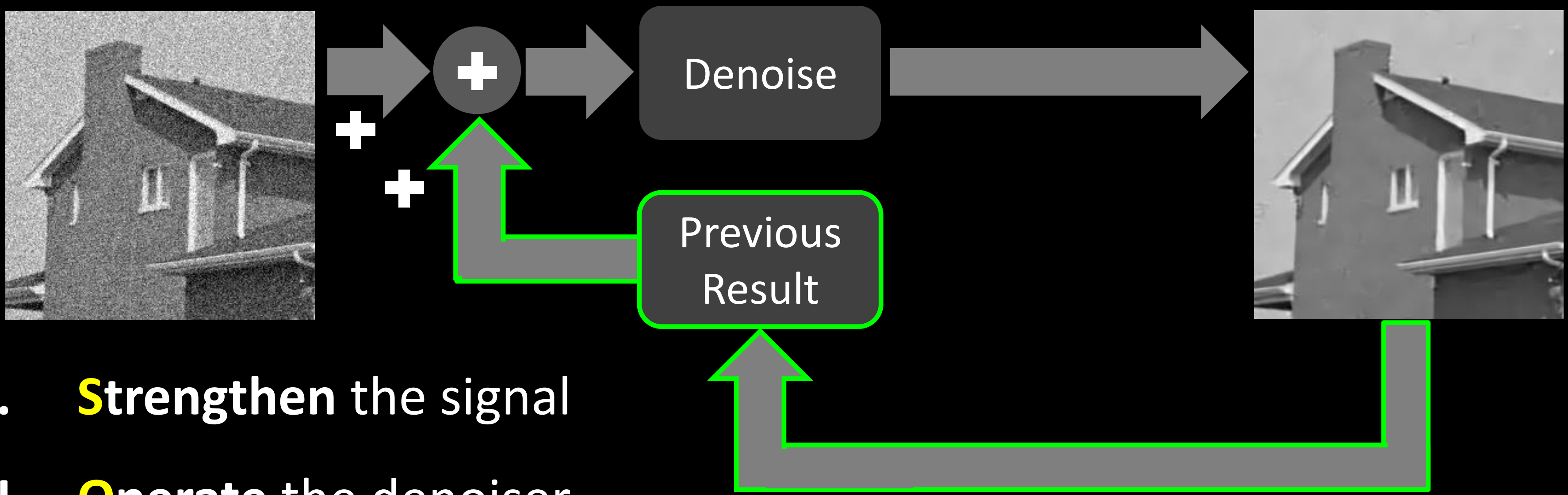
Strengthen - Operate - Subtract Boosting

- Given **any** denoiser, how can we improve its performance?



Strengthen - **O**perate - **S**ubtract Boosting

- Given **any** denoiser, how can we improve its performance?

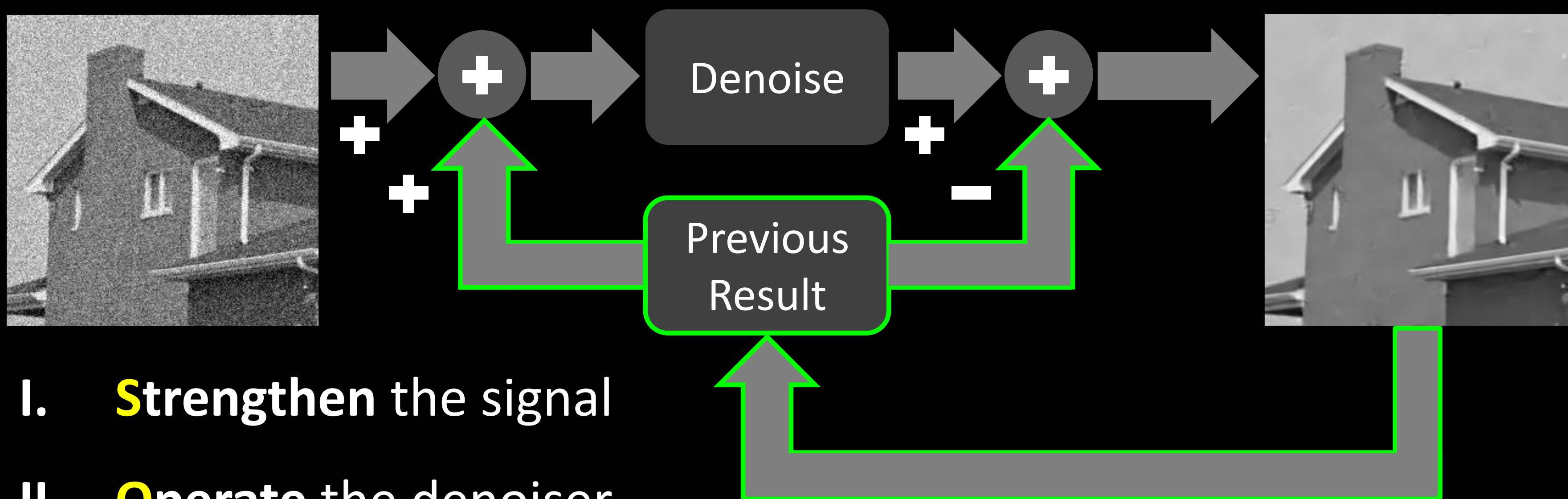


I. **Strengthen** the signal

II. **Operate** the denoiser

Strengthen - Operate - Subtract Boosting

- Given **any** denoiser, how can we improve its performance?



I. **Strengthen** the signal

II. **Operate** the denoiser

III. **Subtract** the previous estimation from the outcome

SOS formulation: $\hat{\mathbf{x}}^{k+1} = f\left(\mathbf{y} + \hat{\mathbf{x}}^{k}\right) - \hat{\mathbf{x}}^{k}$

Strengthen - Operate - Subtract Boosting

- An improvement is obtained since $\text{SNR}\{\mathbf{y}+\hat{\mathbf{x}}\} > \text{SNR}\{\mathbf{y}\}$
  - In the ideal case, where $\hat{\mathbf{x}} = \mathbf{x}$, we get

$$\text{SNR}\{\mathbf{y}+\mathbf{x}\} = 2 \cdot \text{SNR}\{\mathbf{y}\}$$

- We suggest strengthening the underlying signal, rather than
  - Adding the residual back to the noisy image
    - Twicing converges to the noisy image
  - Filtering the previous estimate over and over again
    - Diffusion could lead to over-smoothing, converging to a piece-wise constant image

Image Denoising – A Matrix Formulation

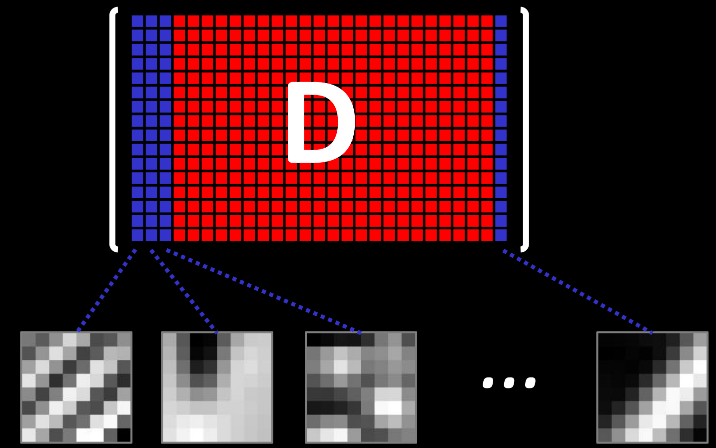
- In order to study the convergence of the SOS function, we represent the denoiser in its matrix form

$$\hat{\mathbf{x}} = f(\mathbf{y}) = \mathbf{W}\,\mathbf{y}$$

- The properties of $\mathbf{W}$:
  - Kernel-based methods (e.g. Bilateral filter, NLM, Kernel Regression) can be approximated as row-stochastic positive definite matrices [Milanfar ('13)]
    - Has eigenvalues in the range [0,…,1]
  - What about sparsity-based methods?

Sparsity Model – The Basics

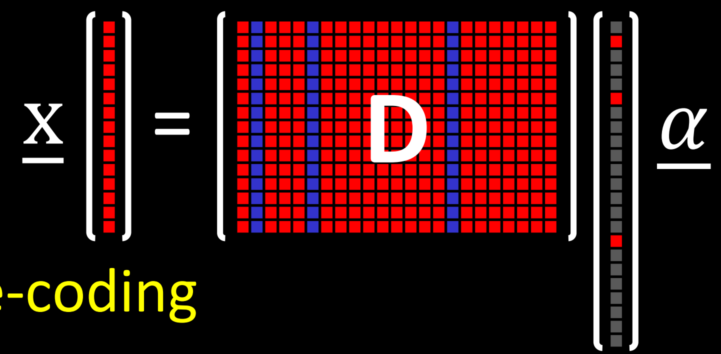
- We assume the existence of a dictionary $\mathbf{D} \in \mathbb{R}^{d \times n}$ whose columns are the atom signals



- Signals are modeled as sparse linear combinations of the dictionary atoms:

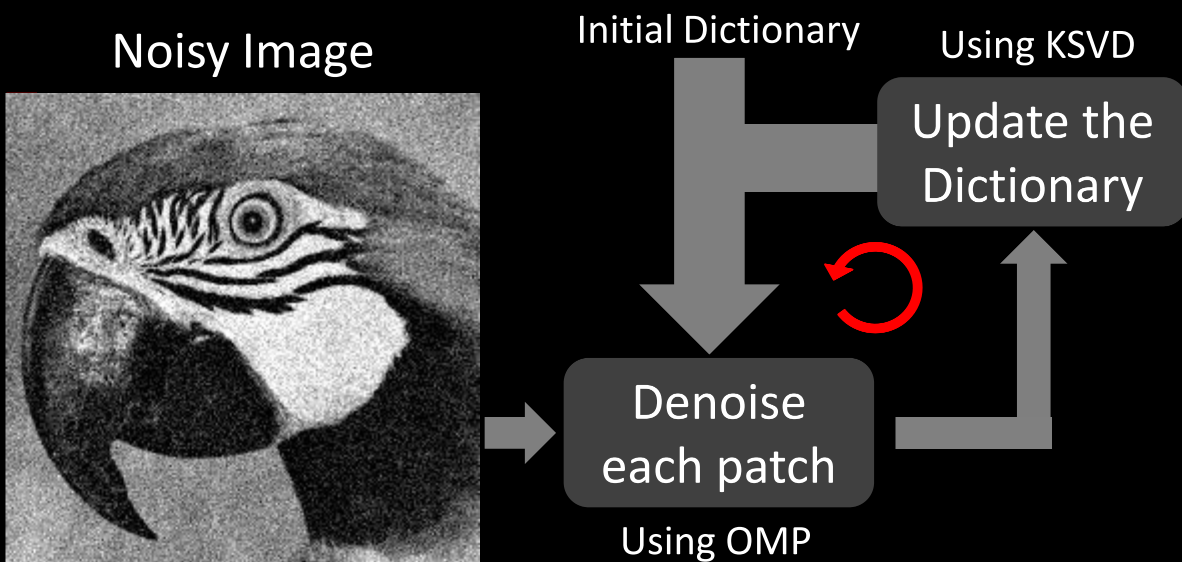
$$\underline{x} = \mathbf{D}\underline{\alpha}$$

where $\underline{\alpha}$ is sparse, meaning that it is assumed to contain mostly zeros



- The computation of $\underline{\alpha}$ from $\underline{x}$ (or its or its noisy version) is called sparse-coding
- The OMP is a popular sparse-coding technique, especially for low dimensional signals

K-SVD Image Denoising [Elad & Aharon ('06)]



$$\hat{\mathbf{p}}_i = \mathbf{D}_{S_i}\alpha_i = \mathbf{D}_{S_i}\left(\mathbf{D}_{S_i}^{\mathrm{T}}\mathbf{D}_{S_i}\right)^{-1}\mathbf{D}_{S_i}^{\mathrm{T}}\mathbf{R}_i\mathbf{y}$$

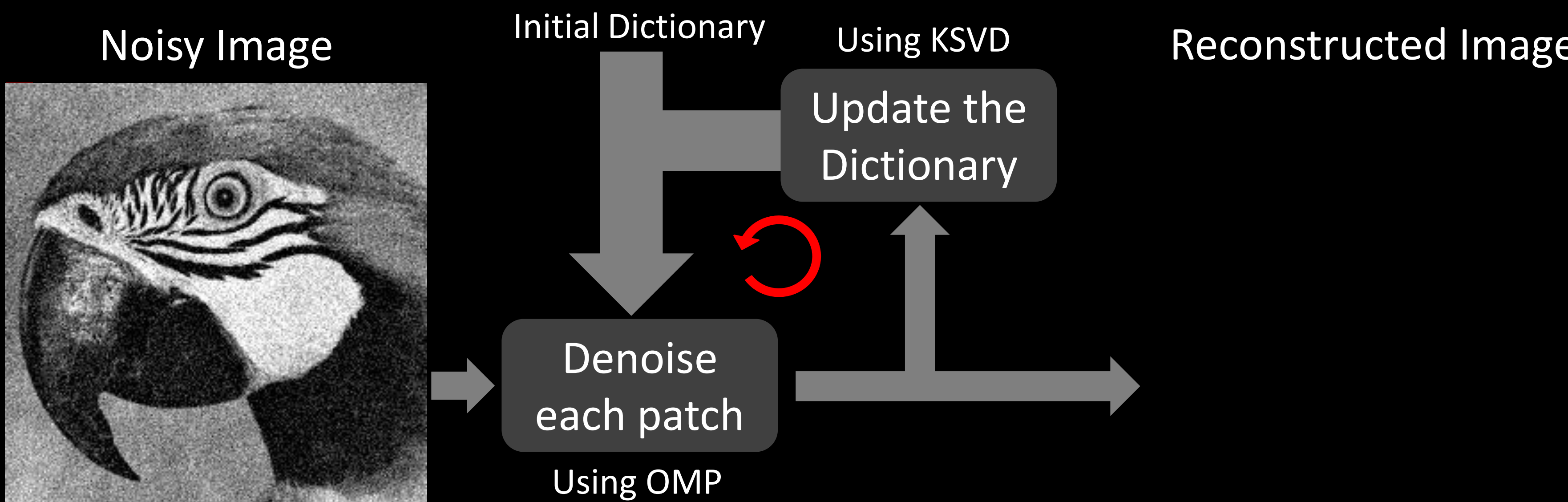
Denoised Patch

A linear combination of few atoms

$$\alpha_i = \min_z \left\|\mathbf{D}_{S_i}z - \mathbf{R}_i\mathbf{y}\right\|_2^2$$

$\mathbf{R}_i$ extracts the i^{th} patch from $\mathbf{y}$

K-SVD Image Denoising [Elad & Aharon ('06)]



$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \mu \|\mathbf{x} - \mathbf{y}\|_2^2 + \sum_i \|\hat{\mathbf{p}}_i - \mathbf{R}_i \mathbf{x}\|_2^2$$

$$= \left(\mu \mathbf{I} + \sum_i \mathbf{R}_i^T \mathbf{R}_i\right)^{-1} \left(\mu \mathbf{I} + \sum_i \mathbf{R}_i^T \mathbf{D}_{S_i} \left(\mathbf{D}_{S_i}^T \mathbf{D}_{S_i}\right)^{-1} \mathbf{D}_{S_i}^T \mathbf{R}_i\right) \mathbf{y}$$

$$= \mathbf{W}\mathbf{y}$$

K-SVD: A Matrix Formulation

- $\mathbf{W}$ is a sum of projection matrices, and has the following properties:
  - Symmetric, $\mathbf{W}^{\mathrm{T}} = \mathbf{W}$ (by assuming periodic boundary condition)
  - Positive definite, $\mathbf{W} \succ \mathbf{0}$
  - Minimal eigenvalue $\lambda_{min}(\mathbf{W}) \geq c$, where $c = \frac{\mu}{\mu+n} > 0$
    - μ is the global averaging constant, and n is the patch size
  - $\mathbf{W}\underline{\mathbf{1}} = \underline{\mathbf{1}}$, thus λ=1 is eigenvalue corresponding to the eigenvector $\underline{\mathbf{1}}$
  - The spectral radius $||\mathbf{W}||_2 = 1$, thus $\lambda_{max}(\mathbf{W}) = 1$
    - $\mathbf{W}$ may have negative entries, which violates the classic definition of row-stochasticness
  - The spectral radius $||\mathbf{W} - \mathbf{I}||_2 \leq 1 - c < 1$

Convergence Study

SOS formulation: $\hat{x}^k = \mathbf{W}_k\left(y + \hat{x}^{k-1}\right) - \hat{x}^{k-1}$

- Given:
  - $\hat{\mathbf{x}}^*$– the denoised image that obtained for a large k
  - $\mathbf{W}_k$, $k = 1,2, \ldots$ – A series of filter-matrices
- The error $e_k = \hat{x}^k - \hat{x}^*$ of the recursive function is expressed by

$$e_k = (\mathbf{W}_k - \mathbf{I})e_{k-1} + (\mathbf{W}_k - \mathbf{W}_*)(\mathbf{y} + \hat{\mathbf{x}}^*)$$

- By assuming $\mathbf{W}_k = \mathrm{W}_*$ from a certain k (which comes up in practice):

$$e_k = (\mathbf{W} - \mathbf{I})e_{k-1}$$

Convergence Study

- The SOS recursive function converges if $\|\mathbf{W}-\mathbf{I}\|_2 < 1$
  - Holds both for kernel-based (Bilateral filter, NLM, Kernel Regression), and sparsity-based methods (K-SVD).

For most denoising algorithms the SOS boosting is "guaranteed" to converge to

$$\hat{x}^* = (2\mathbf{I}-\mathbf{W})^{-1}y$$

SOS formulation: $\hat{x}^k = \mathbf{W}_k\left(y+\hat{x}^{k-1}\right)-\hat{x}^{k-1}$

Generalization

- We introduce 2 parameters that modify
  - The steady-state outcome
  - The requirements for convergence (the eigenvalues range)
  - The rate of convergence
- The parameter ρ, expressed by

$$\hat{\mathbf{x}}^{k+1} = f\left(\mathbf{y} + \rho\hat{\mathbf{x}}^{k}\right) - \rho\hat{\mathbf{x}}^{k}$$

  - Controls the signal emphasis
  - Affects the steady-state outcome

Generalization

- The second parameter, τ, controls the rate-of-convergence
  - The steady state outcome is given by

$$\hat{\mathbf{x}}^* = f(\mathbf{y} + \rho\hat{\mathbf{x}}^*) - \rho\hat{\mathbf{x}}^*$$

  - Multiplying both sides by τ, and adding $(\hat{\mathbf{x}}^* - \hat{\mathbf{x}}^*)$ to the RHS (does not effect the steady state result)

$$\tau\hat{\mathbf{x}}^* = \tau f(\mathbf{y} + \rho\hat{\mathbf{x}}^*) - \tau\rho\hat{\mathbf{x}}^* + \hat{\mathbf{x}}^* - \hat{\mathbf{x}}^*$$

  - Rearranging the equation above, and replacing $\hat{\mathbf{x}}^*$ with $\hat{\mathbf{x}}^k$ leading to the generalized SOS recursive function:

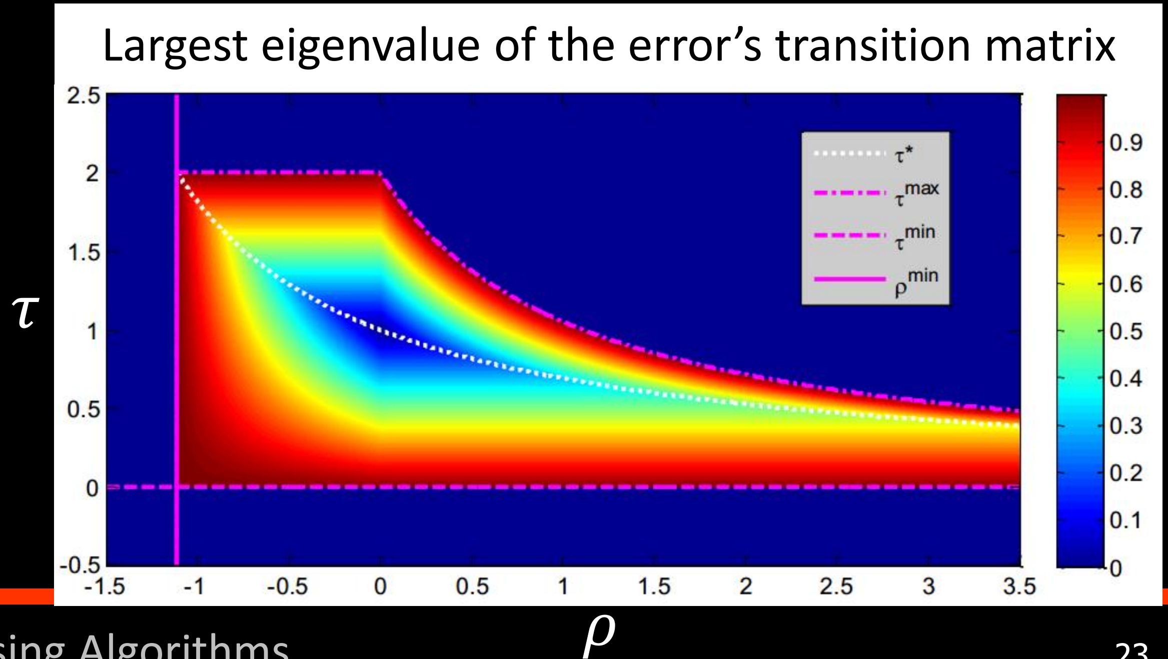
$$\hat{\mathbf{x}}^{k+1} = \tau f(\mathbf{y} + \rho\hat{\mathbf{x}}^k) - (\tau\rho + \tau - 1)\hat{\mathbf{x}}^k$$

Generalization

- By assuming $\mathbf{W}_k = \mathbf{W}_*$, the error $e_k = \hat{\mathrm{x}}^k - \hat{\mathrm{x}}^*$ is given by

$$e_k = (\tau\rho\mathbf{W}_k - (\tau\rho + \tau - 1)\mathbf{I})e_{k-1}$$

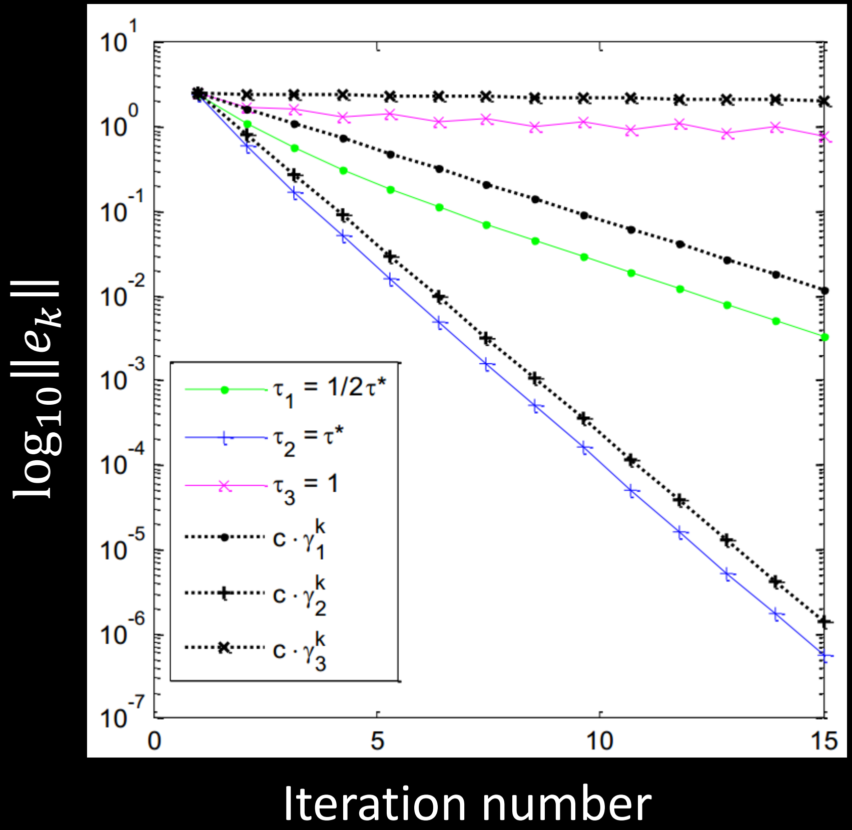
  - As a result, ρ and τ control the rate-of-convergence
- We suggest a closed-form expression for the optimal (ρ, τ) setting
  - Given ρ, what is the best τ, leading to the fastest convergence?



Experiments

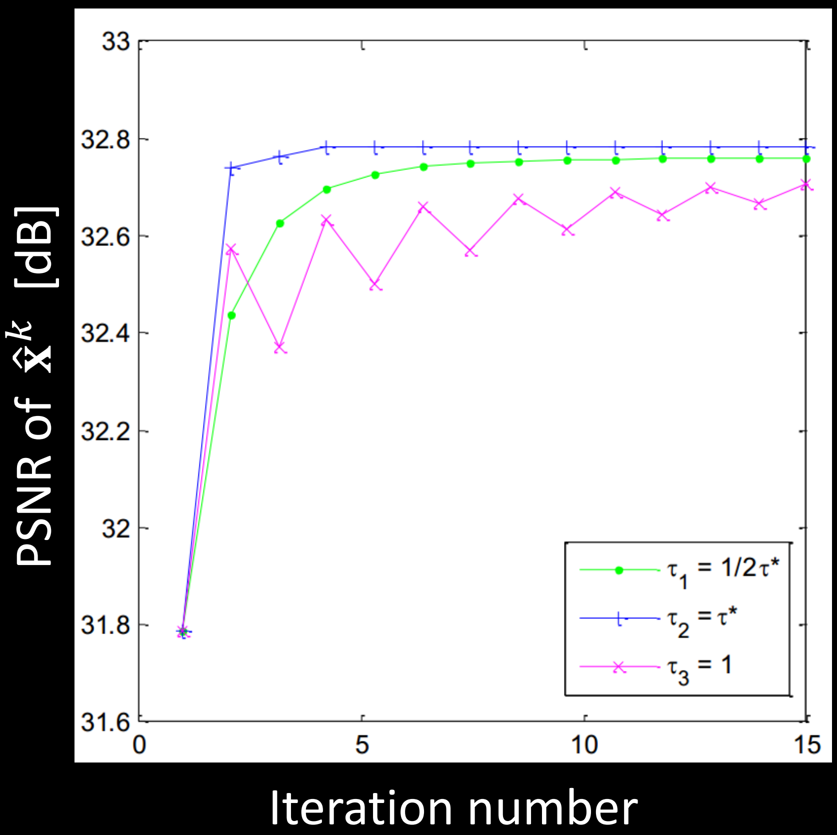
Verifying the Convergence Study

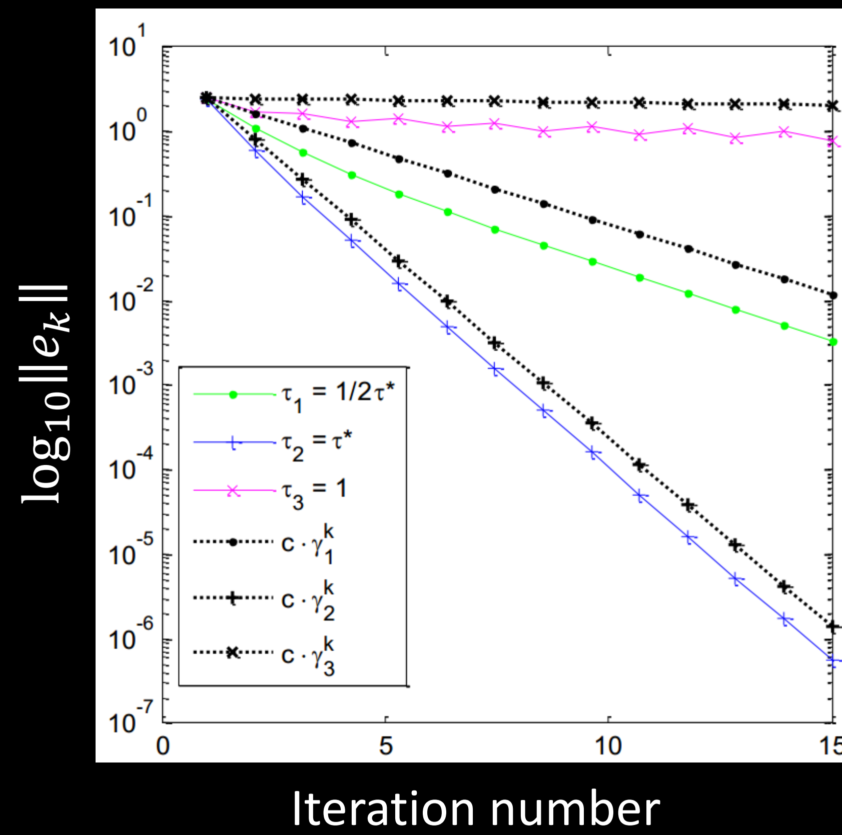
- We apply the K-SVD on the noisy `House` image ($\sigma = 25$), with a fixed $\mathbf{W}$
  - $\boldsymbol{\tau}^*$ is the outcome of our closed-form expression, achieving the fastest convergence
  - $\boldsymbol{\gamma}$ is our analytic expression that bounds the error (largest eigenvalue of the error's transition matrix)



Convergence

- We apply the K-SVD on noisy House image $\sigma = 25$, with a fixed $\mathbf{W}$
  - Converges faster and improves the PSNR





Results

- We successfully boost several state-of-the-art denoising algorithms:
  - K-SVD, NLM, BM3D, and EPLL
  - Without any modifications, simply by applying the original software as a “black-box”
- We manually tuned two parameters
  - ρ – signal emphasis factor
  - σ – noise level, which is an input to the denoiser
    - Since the noise level of $\mathbf{y} + \rho\mathbf{x}^k$ is higher than the one of $\mathbf{y}$

Quantitative Comparison

- Average boosting in PSNR* over 5 images (higher is better):

Noise std	Improved Methods			
σ	**K-SVD**	**NLM**	**BM3D**	**EPLL**
10	**0.13**	**0.44**	**0.01**	**0.13**
20	**0.22**	**0.34**	**0.02**	**0.25**
25	**0.26**	**0.41**	**0.03**	**0.26**
50	**0.77**	**0.30**	**0.07**	**0.30**
75	**1.26**	**0.56**	**0.11**	**0.32**
100	**0.81**	**0.36**	**0.14**	**0.27**

$^*\text{PSNR} = 20\log_{10}\left(255/\sqrt{\text{MSE}}\right)$

Visual Comparison: K-SVD

- Original K-SVD results, $\sigma = 25$

Visual Comparison: K-SVD

- SOS K-SVD results, $\sigma = 25$

Visual Comparison: EPLL

- Original EPLL results, $\sigma = 25$

Forman

32.44dB

House

32.07dB

Visual Comparison: EPLL

- SOS EPLL results, $\sigma = 25$

Forman

32.78dB

House

32.38dB

Reducing the “Local-Global” Gap

Reaching a Consensus

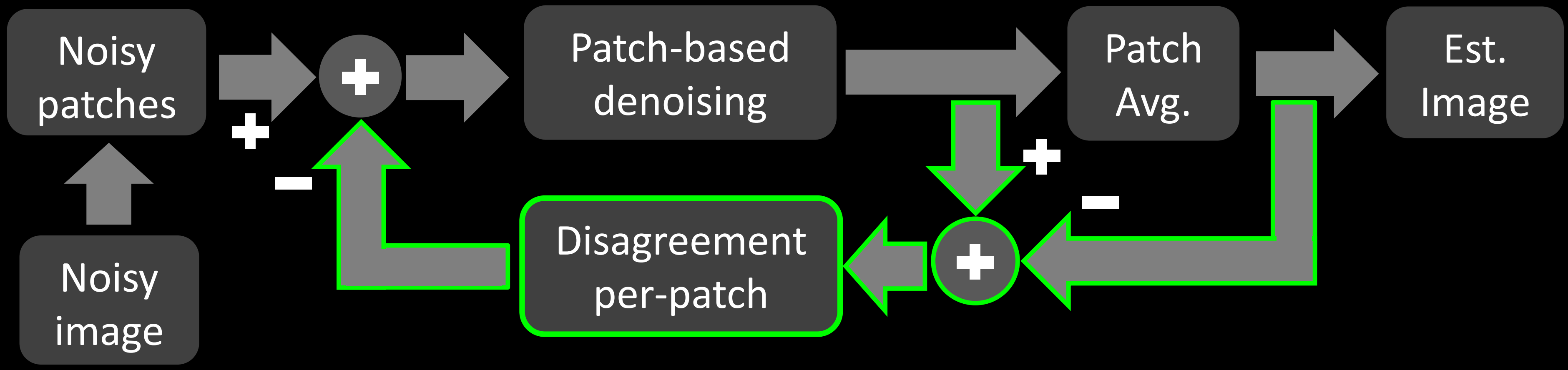
- It turns out that the SOS boosting treats a major shortcoming of many patch-based methods:
  - Local processing of patches VS. the global need in a whole denoised result
- We define the local disagreements by

Disagreement patch = Local independent denoised patch − Globally averaged patch

- The disagreements
  - Naturally exist since each noisy patch is denoised independently
  - Are based on the intermediate results

“Sharing the Disagreement”

- Inspired by the “Consensus and Sharing” problem from game-theory:
  - There are several agents, each one of them aims to minimize its individual cost (i.e., representing the noisy patch sparsely)
  - These agents affect a shared objective term, describing the overall goal (i.e., obtaining the globally denoised image)
- Imitating this concept, we suggest sharing the disagreements



Connection to SOS Boosting

- Interestingly, for a fixed filter matrix $\mathbf{W}$, “sharing the disagreement” and the SOS boosting are equivalent

$$\hat{\mathbf{x}}^{k+1} = \mathbf{W}\left(\mathbf{y} + \hat{\mathbf{x}}^{k}\right) - \hat{\mathbf{x}}^{k}$$

- The connection to the SOS is far from trivial because
  - The SOS is blind to the intermediate results (the independent denoised patches, before patch-averaging)
  - The intermediate results are crucial for “sharing the disagreement” approach

The SOS boosting reduces the Local/Global gap

Conclusions

- The SOS boosting algorithm:
  - ✓ Easy to use
    - In practice, we treat the denoiser $f(\cdot)$ as a “black-box”
  - ✓ Applicable to a wide range of denoising algorithms
  - ✓ Guaranteed to converge for most denoising algorithms
    - Thus, has a straightforward stopping criterion
  - ✓ Reduces the local-global gap
  - ✓ Improves the state-of-the-art methods

Thank You For Your Attention

&

Thanks To the Organizer of BASP Frontiers Yves Wiaux