Natural Image Denoising: Optimality and Inherent Bounds

Boaz Nadler

Department of Computer Science and Applied Math

The Weizmann Institute of Science

Currently on Sabbatical at UC Berkeley

Joint work with Anat Levin and our cluster.

Natural image denoising

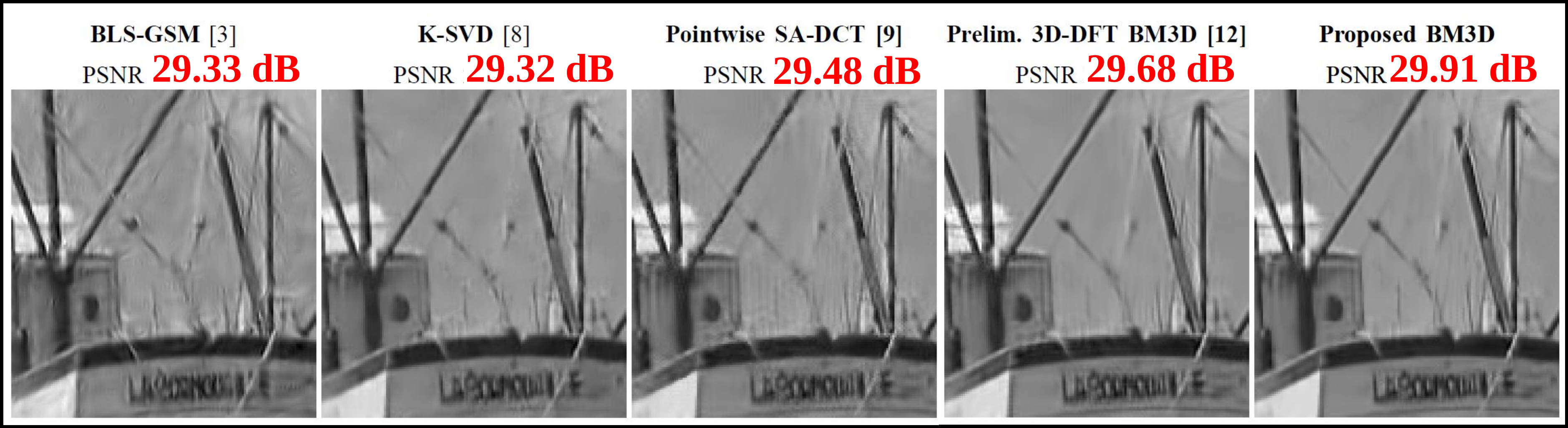
Input: noisy image

$$y = x + \sigma\xi$$

Goal: estimate noise free image

Main idea: Use additional knowledge on natural images

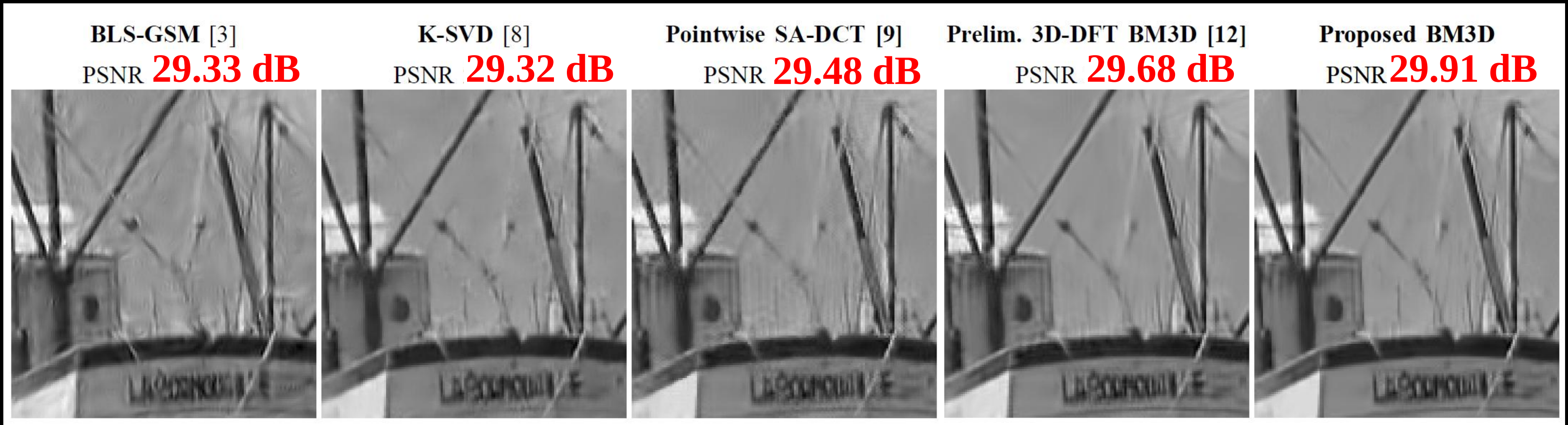
Natural image denoising



Key questions:

- How much can we improve current denoising algorithms?
- *Optimal* denoising?
- What are the limits of natural image statistics?

Natural image denoising



Should you do your Ph.D. on natural image denoising?

Denoising limits- prior work

- **Signal processing assumptions (Wiener filter, Gaussian priors)**

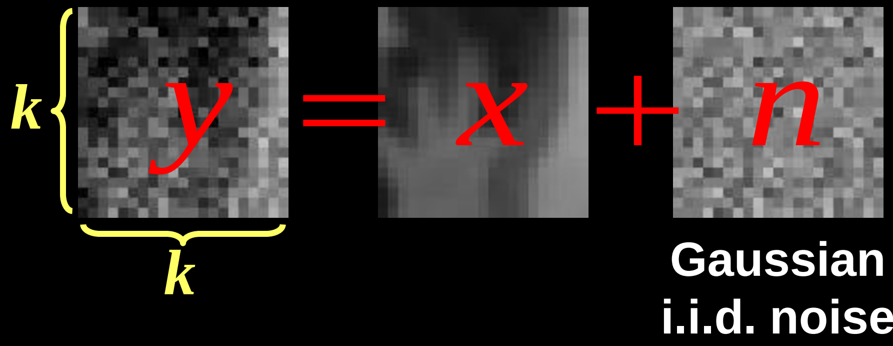
- **Limits on super resolution- numerical arguments, no prior** *[Baker&Kanade 02]*

- **Sharp bounds for perfectly piecewise constant images** *[Korostelev&Tsybakov 93, Polzehl&Spokoiny 03]*

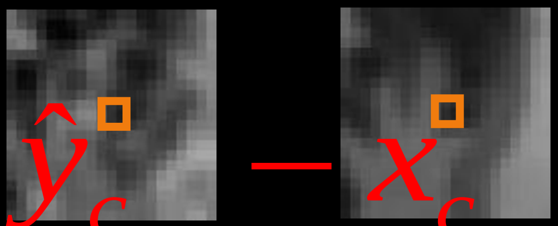
- **Non-local means- asymptotically optimal for infinitely large images. No analysis of finite size images.** *[Buades,Coll&Morel. 05]*

- **Natural image denoising limits, but many assumptions which may not hold in practice and effect conclusions.** *[Chatterjee and Milanfar 10]*

Statistical framework



Quality measure: Mean Squared Error (MSE):



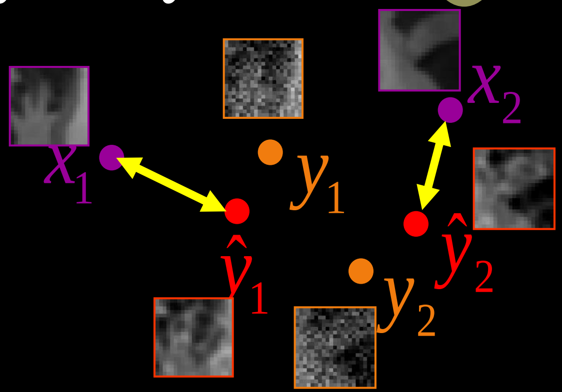
$$\mathrm{MSE} = \mathrm{E}\left[\left(\hat{y}_c - x_c\right)^2\right]$$

- **Estimate central pixel x_c using a k x k support**
- ***Generic* natural image prior**

Two views on MSE

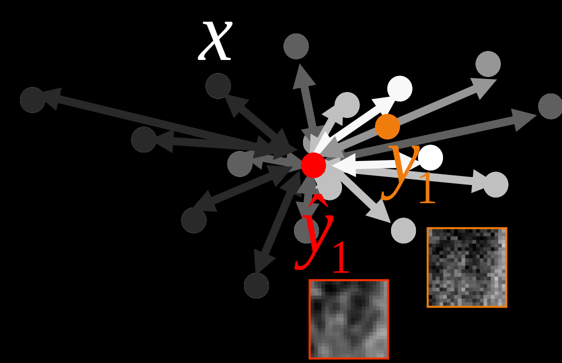
Error interpretation

$$\int p(x)\int p(y\,|\,x)(\hat{y}_c - x_c)^2\,dydx$$



Variance interpretation

$$\int p(y)\int p(x\,|\,y)(\hat{y}_c - x_c)^2\,dxdy$$

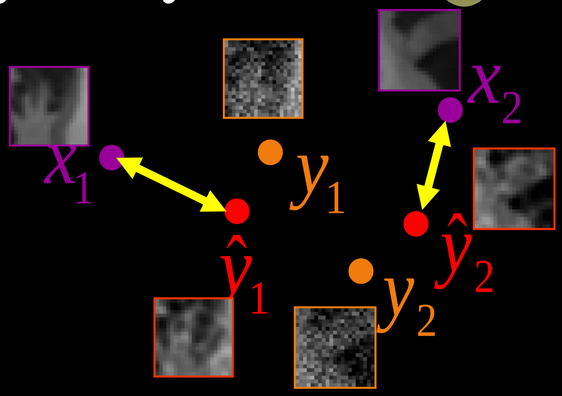


Optimal estimator?

Two views on MSE

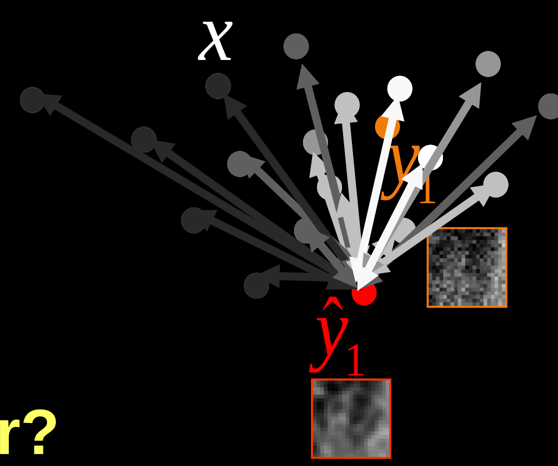
Error interpretation

$$\int p(x)\int p(y\mid x)(\,?\,-x_c)^2\,dy\,dx$$



Variance interpretation

$$\int p(y)\int p(x\mid y)(\,?\,-x_c)^2\,dx\,dy$$

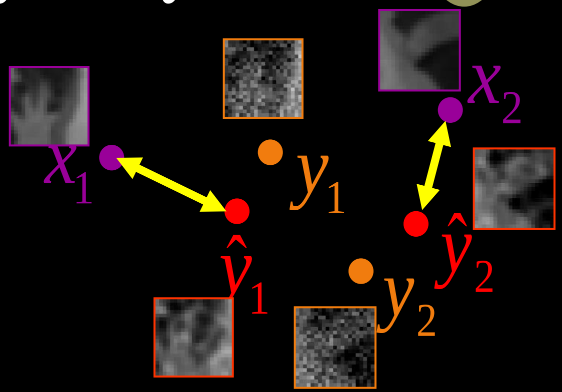


Optimal estimator?

Two views on MSE

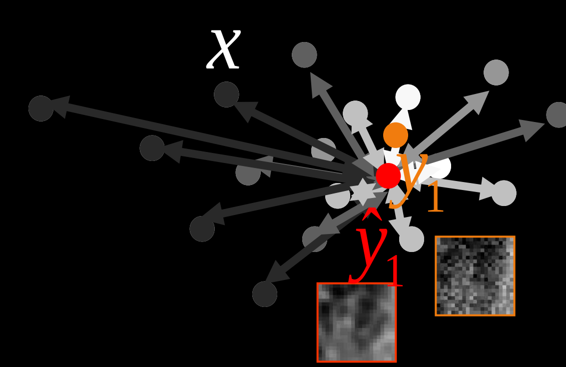
Error interpretation

$$\int p(x)\int p(y\,|\,x)(?_c - x_c)^2\,dy\,dx$$



Variance interpretation

$$\int p(y)\int p(x\,|\,y)(?_c - x_c)^2\,dx\,dy$$



Optimal estimator? - the mean

$$\mu(y) = \mathrm{E}[x_c\,|\,y] = \int p(x\,|\,y)\,x_c\,dx$$

Also known as

Bayesian minimum mean squared error (MMSE) estimator

Estimating image denoising bounds

Achieved: **formula for best possible minimum mean square error (MMSE)**

$$\text{MMSE} = \int p(y) \int p(x \mid y)(x_c - \mu(y))^2 \, dx\, dy$$

Challenges: **Intractable integral**

We don't know p(x)

Estimating image denoising bounds

$$\text{MMSE} = \int p(y) \int p(x \mid y)(x_c - \mu(y))^2 \, dx dy$$

Challenge: **Compute MMSE without knowing p(x)?**

The trick:

We don't know p(x) but we can sample from it

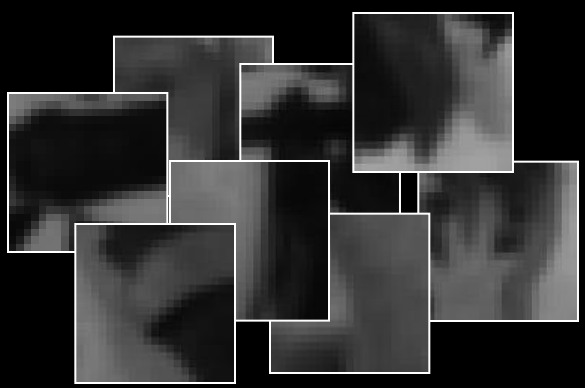
Approximate MMSE non parametrically

Statistical measure to judge if samples are dense enough:
lower **and** ***upper*** **bounds on MMSE, measure gap**

Data

N=10^{10} generic clean patches

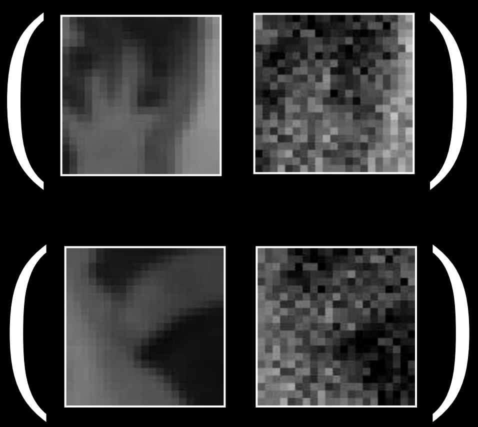
$$\{x_i\}_{i=1}^{N}$$



20K images from LabelMe dataset

M=2000 clean + noisy pairs

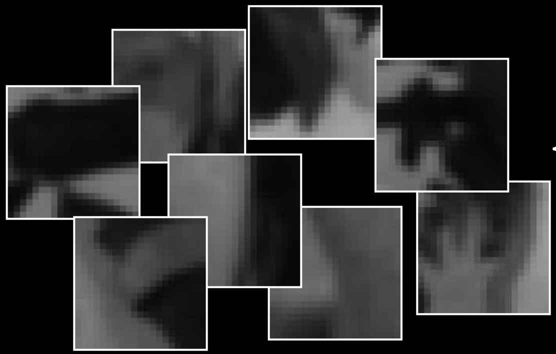
$$\{(\tilde{x}_j, \tilde{y}_j)\}_{j=1}^{M}$$



Non parametric estimation

MMSE estimator- analytic mean

$$\mu(y) = \int p(x \mid y) x_c dx = \frac{\int p(y \mid x) p(x) x_c dx}{\int p(y \mid x) p(x) dx}$$



$$\{x_i\} \sim p(x)$$

Approximated mean

$$\hat{\mu}(y) = \frac{\frac{1}{N}\sum_i p(y \mid x_i) x_{i,c}}{\frac{1}{N}\sum_i p(y \mid x_i)}$$

For Gaussian i.i.d. noise: $p(y \mid x) \propto e^{-0.5\|y-x\|^2/\sigma^2}$

Non parametric MMSE bounds

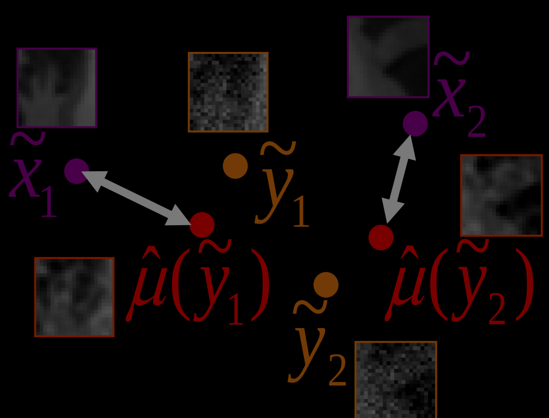
Every estimator upper bounds the MMSE

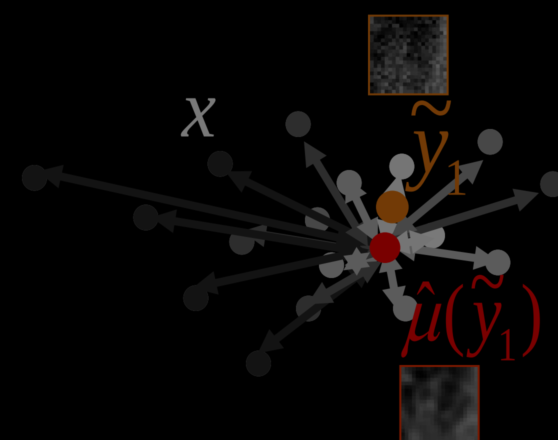
Proof required

$$E_N\left[MMSE^U\right] \geqslant MMSE \geqslant E_N\left[MMSE^L\right]$$

$$\frac{1}{M}\sum_j \left(\tilde{x}_j - \hat{\mu}(\tilde{y}_j)\right)^2 \qquad\qquad \frac{1}{M}\sum_j \frac{\frac{1}{N}\sum_i p(\tilde{y}_j \mid x_i)(x_{i,c} - \hat{\mu}(\tilde{y}_j))^2}{\frac{1}{N}\sum_i p(\tilde{y}_j \mid x_i)}$$

Non parametric error **Non parametric variance**





MMSE bounds

Two statistical measures:

$MMSE^{U}$ **MSE of yet another (not so fast) algorithm-**
Upper **bounds the** $MMSE$

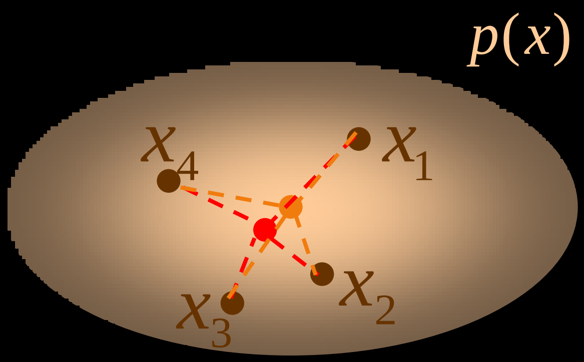
$MMSE^{L}$ **average variance of the samples.**
Lower **bounds the** $MMSE$

If number of samples N is sufficiently large

$$E_N[MMSE^{L}] \sim E_N[MMSE^{U}]$$

and we have an estimate of the best possible $MMSE$

Intuition on sample variance



Estimated mean

$$\hat{\mu} = \frac{1}{N}\sum_i x_i$$

Exact mean

$$\mu = E[x]$$

$$\hat{V} = \boldsymbol{E}\left[\frac{1}{N}\sum_i (x_i - \hat{\mu})^2\right] \leq \boldsymbol{E}\left[\frac{1}{N}\sum_i (x_i - \mu)^2\right] = V$$

Approximated variance from samples

Exact variance

Sample variance- theoretical results

Claim 1:

Let $$\hat{\mathrm{V}}(y) = \frac{\frac{1}{N}\sum_i p(y \mid x_i)(x_{i,c} - \mu(y))^2}{\frac{1}{N}\sum_i p(y \mid x_i)}$$

Denote the sample variance around y. In expectation:

$$\mathrm{E}[\hat{\mathrm{V}}(y)] = \mathrm{V}(y) + C(y)B(y) + o\left(\tfrac{1}{N}\right)$$

With:

$$B(y) = \mathrm{E}[x_c^2 \mid y] - 3\mathrm{E}[x_c \mid y]^2 - 2\mathrm{E}_{\sigma*}[x_c^2 \mid y] + 4\mathrm{E}[x_c \mid y]\mathrm{E}_{\sigma*}[x_c \mid y]$$

$$C(y) = \frac{1}{N}\frac{p_{\sigma*}(y)}{(4\pi\sigma^2)^{k^2/2}\, p(y)^2}$$

Where $p_{\sigma*}(\cdot), \mathrm{E}_{\sigma*}(\cdot)$ **denote probability and expectation of random variables with reduced noise variance** $\sigma*^2 = \sigma^2/2$

Sample variance- theoretical results

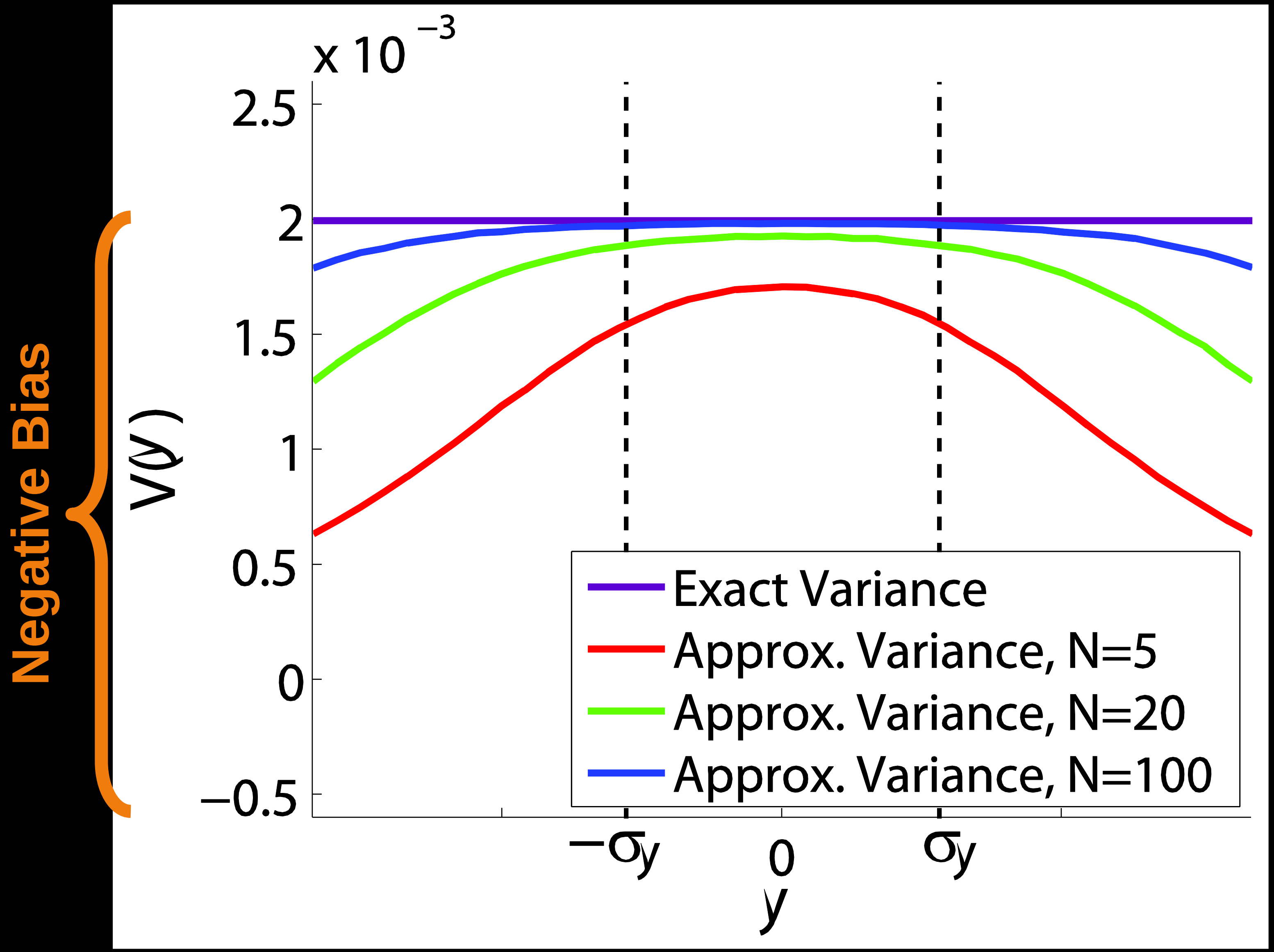
We consider a local Gaussian approximation for p(x) around x=y.

For a Gaussian distribution the bias can be computed analytically

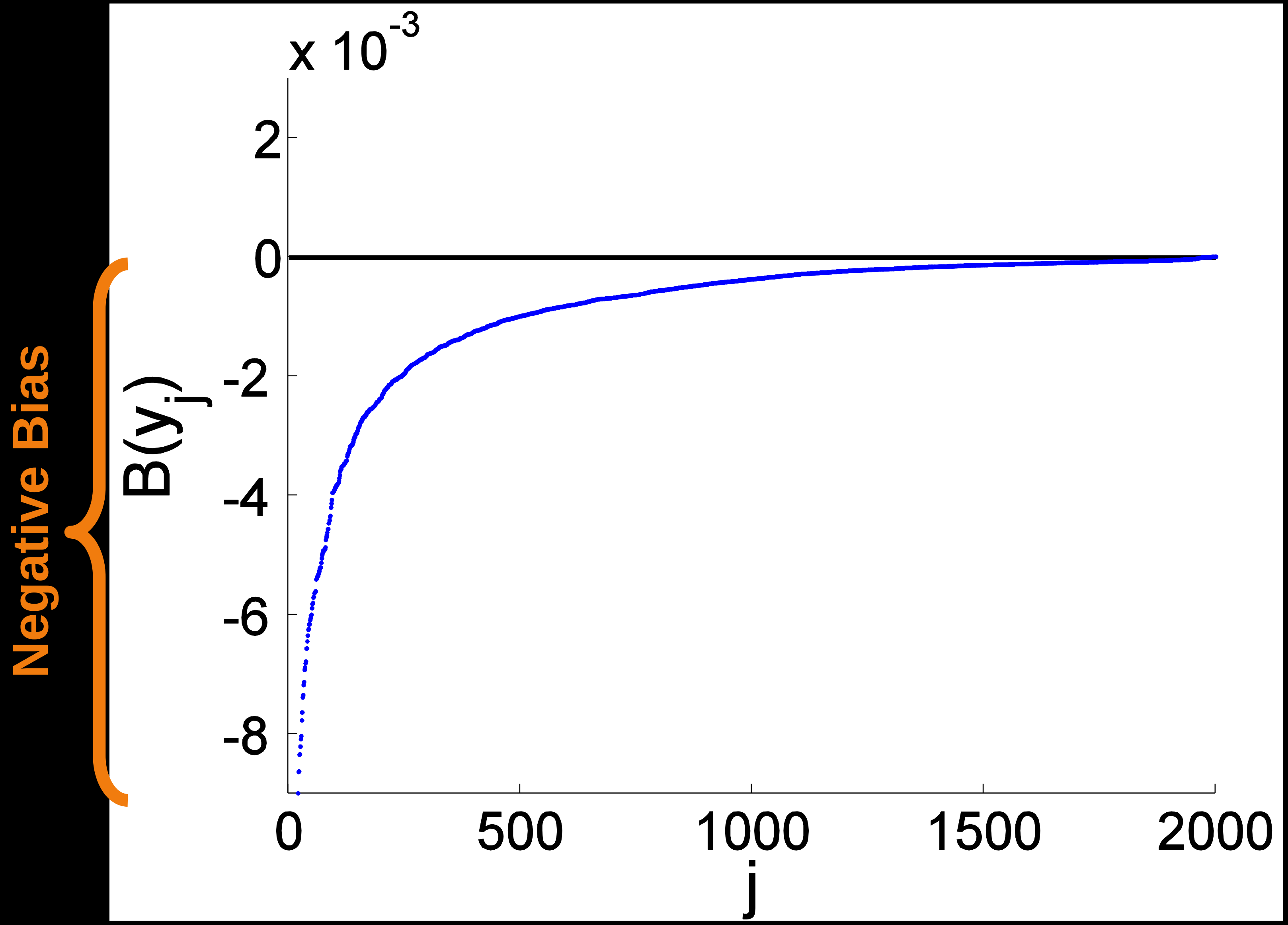
Claim 2:

For a Gaussian distribution $B(y) \leq 0$ **for all y**

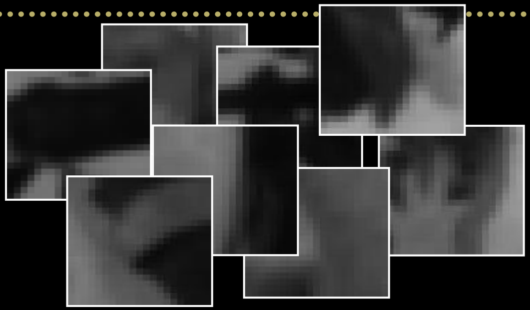
Numerical bias evaluation- 1D Gaussian



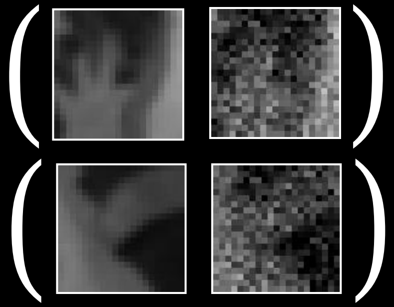
Plug-in bias estimates, 3x3 patches



Experimental setup



Set of $N=10^{10}$ natural image patches out of 20K images from the LabelMe dataset



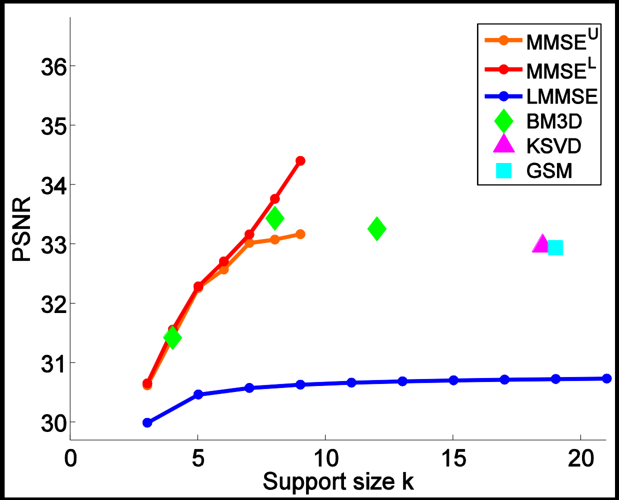
Independent set of M=2000 noisy+clean patches

Added noise at different levels σ =18,55,170 and with different window sizes k=3,4,…,21

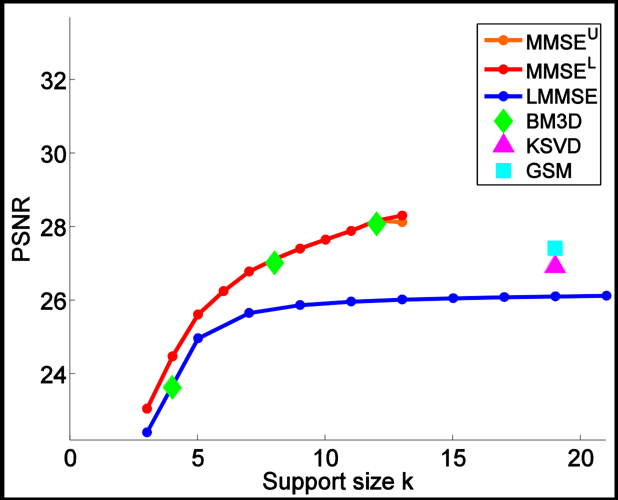
Compare $MMSE^{L}$, $MMSE^{U}$, as well as MSE of various denoising algorithms

*Acknowledgments: the cluster

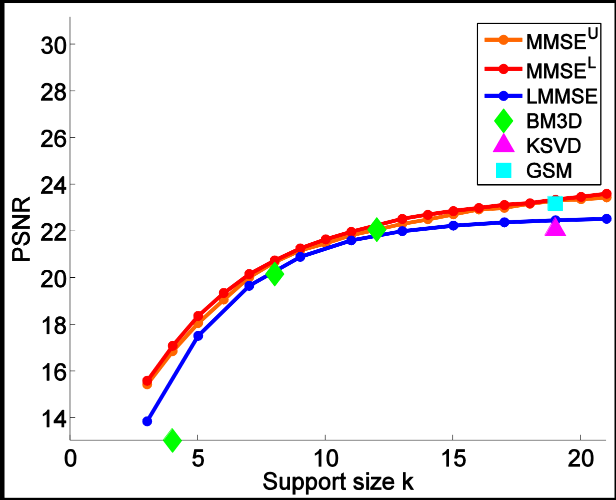
MMSE and MSE Evaluation



$\sigma = 18$

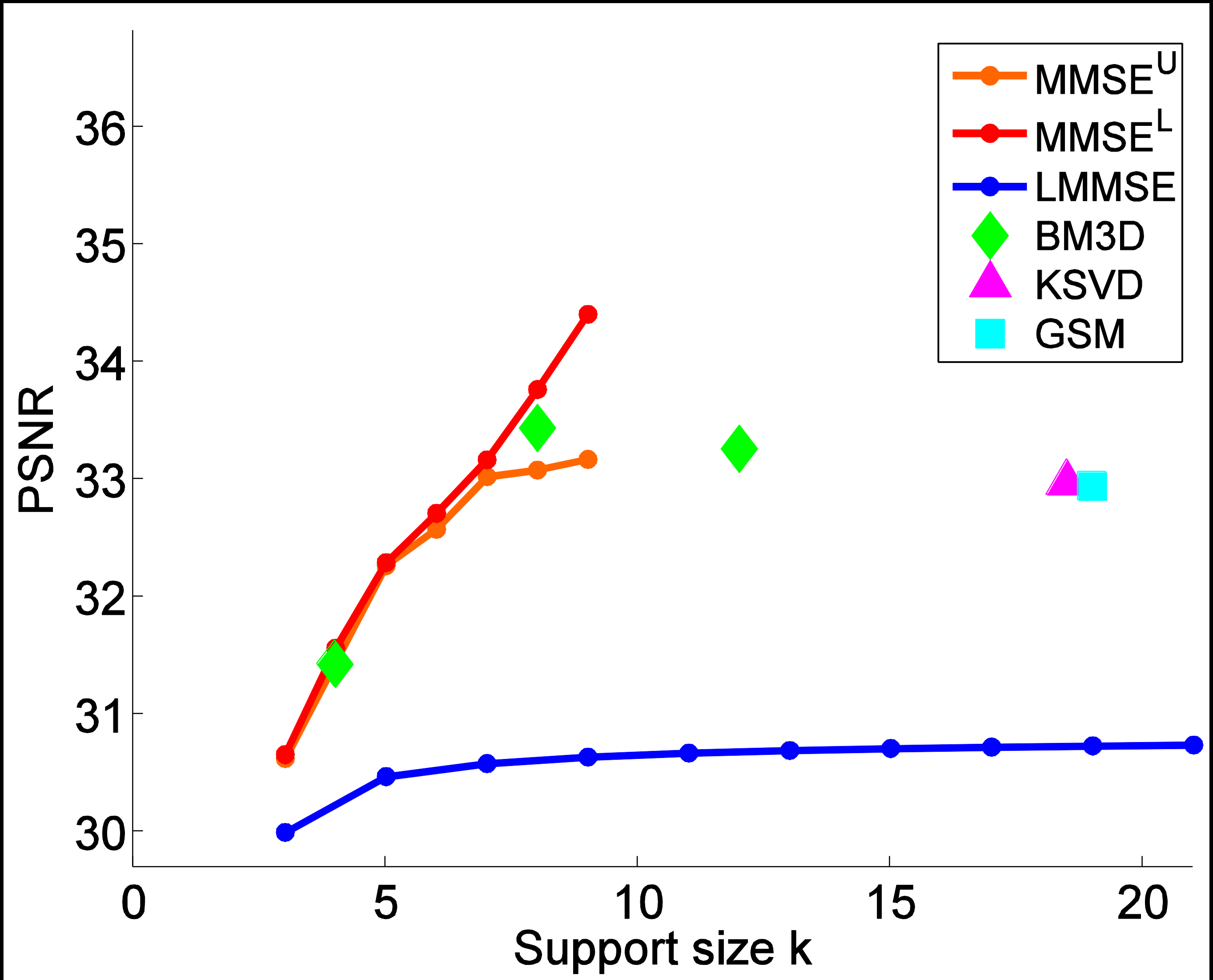


$\sigma = 55$



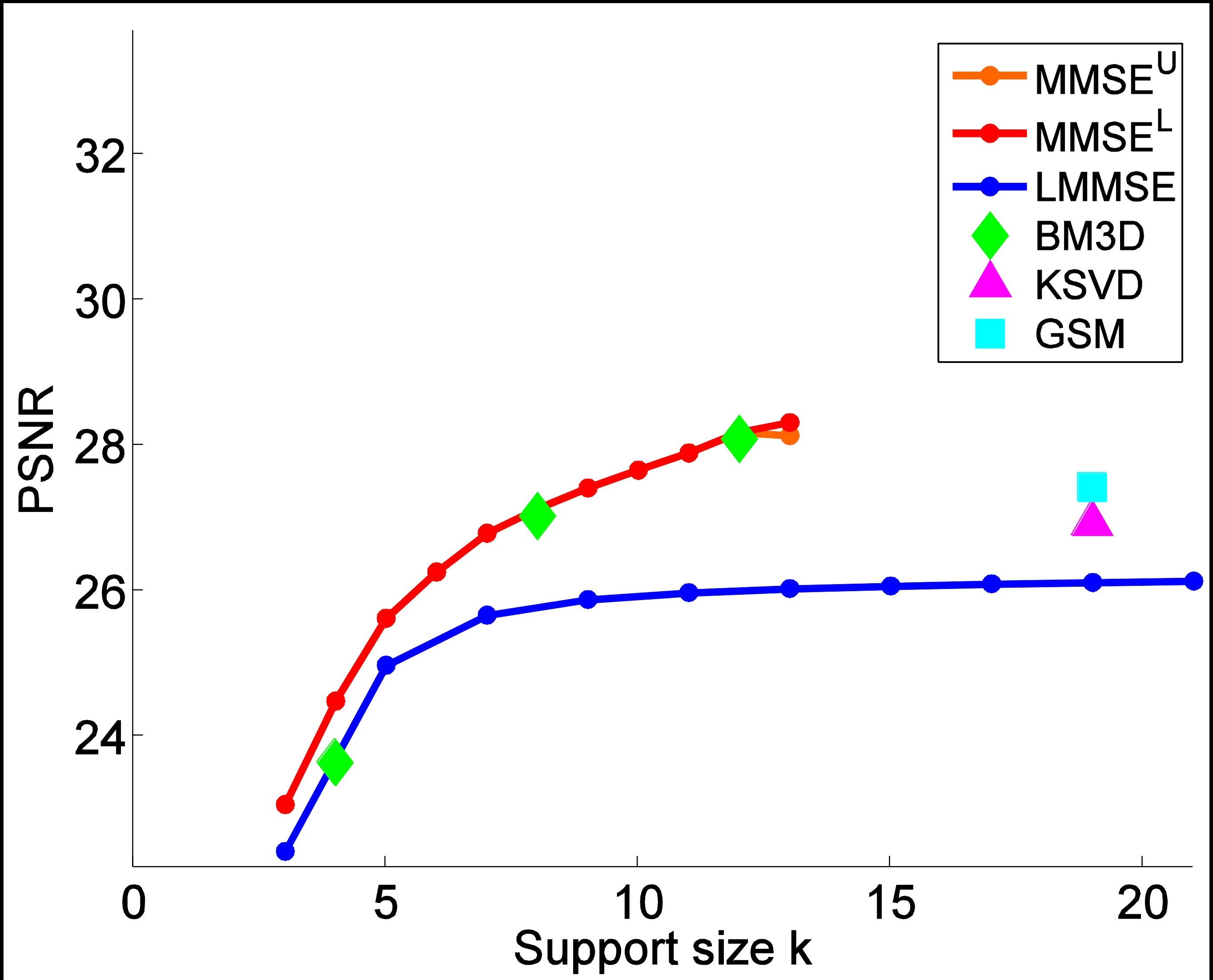
$\sigma = 170$

MMSE and MSE Evaluation



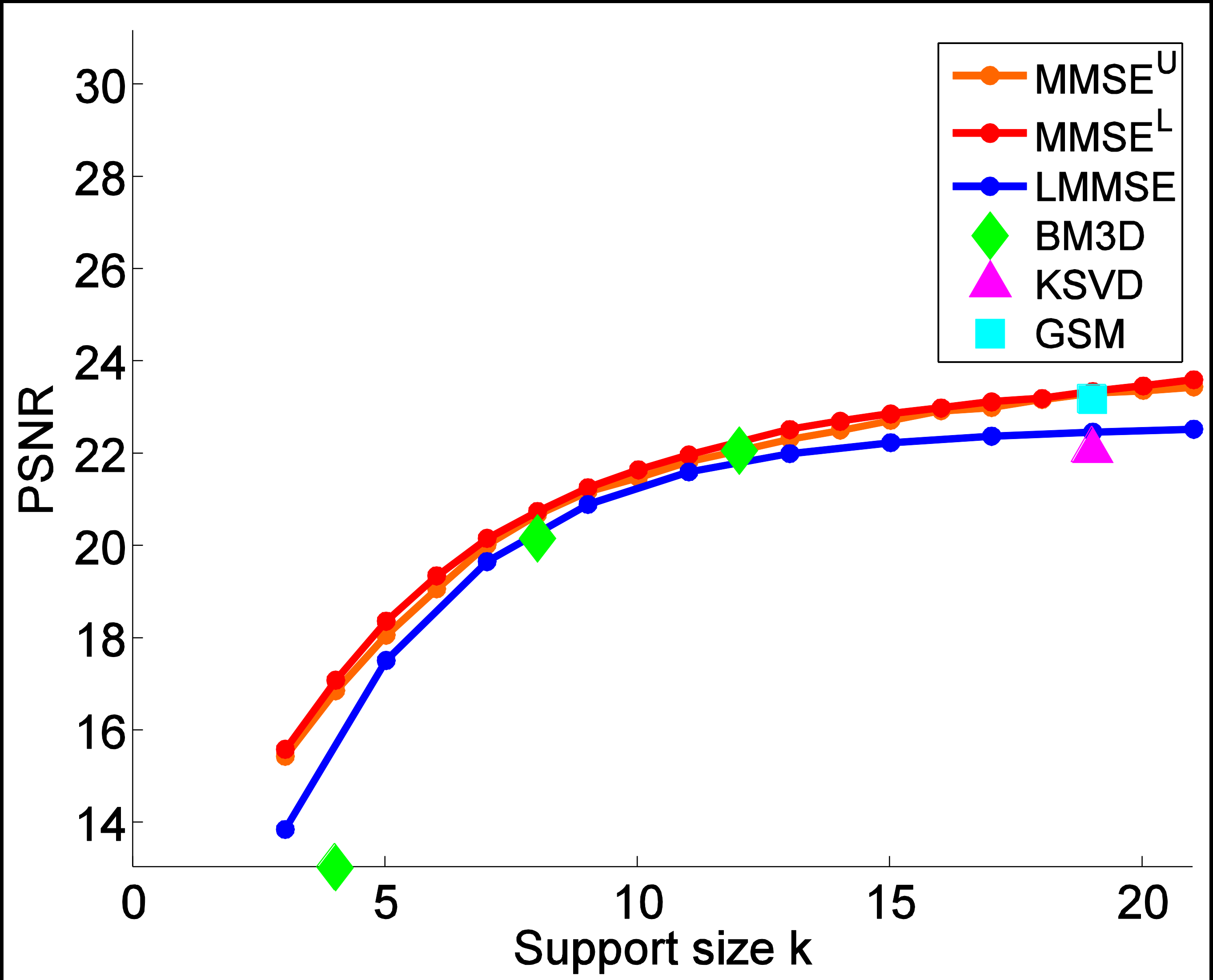
$$\sigma = 18$$

MMSE and MSE Evaluation



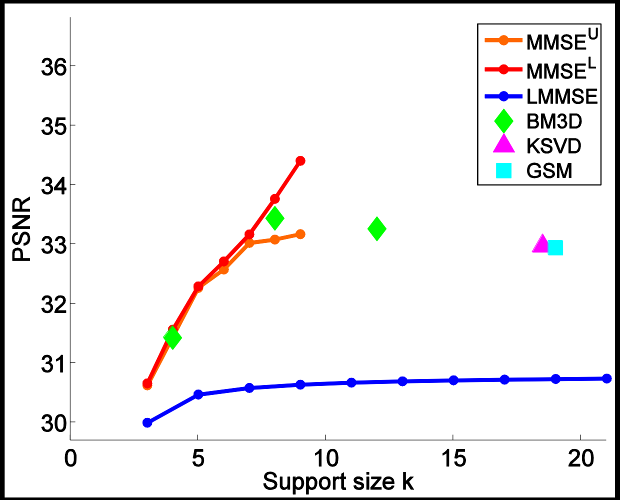
$$\sigma = 55$$

MMSE and MSE Evaluation

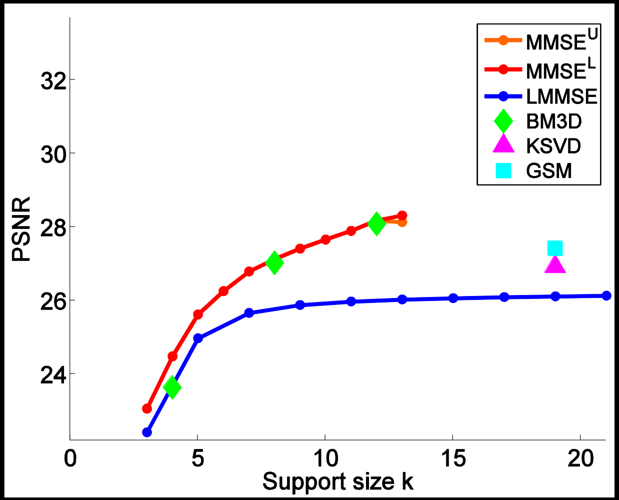


$$\sigma = 170$$

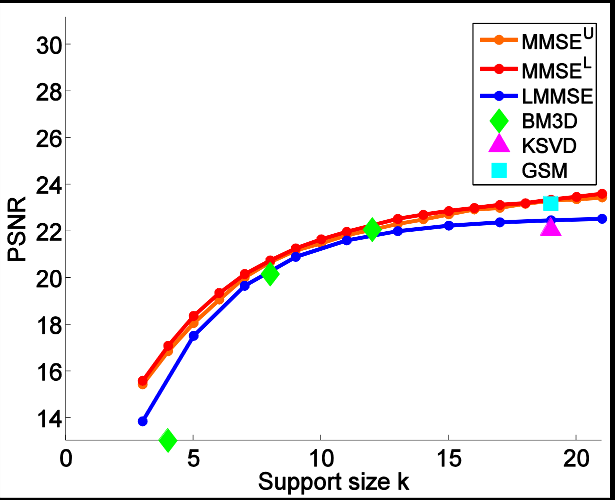
MMSE and MSE Evaluation



$\sigma = 18$



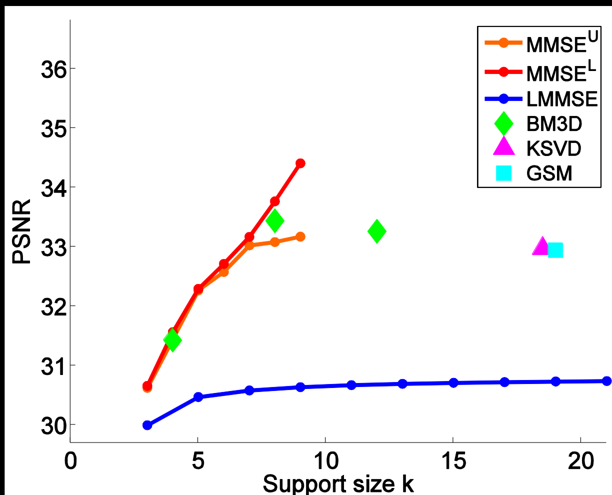
$\sigma = 55$



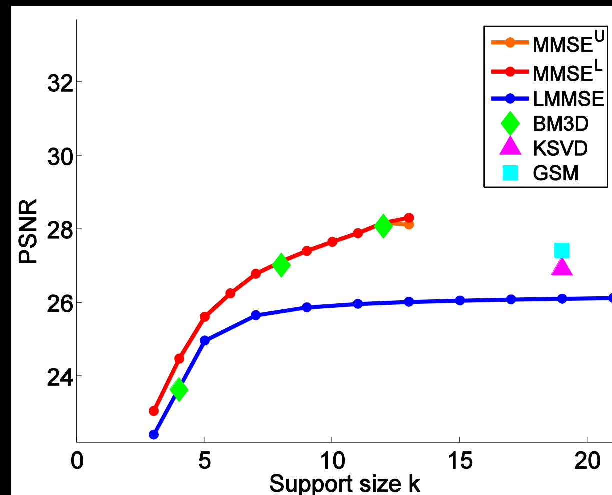
$\sigma = 170$

Observation 1: **As predicted by the theory, all algorithms perform below the bound**

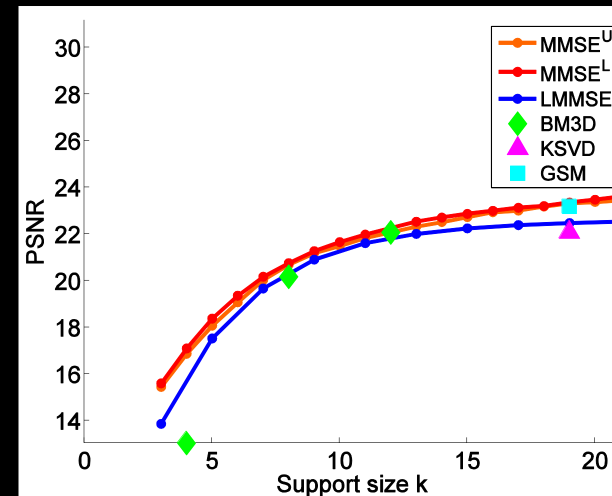
MMSE and MSE Evaluation



$\sigma = 18$



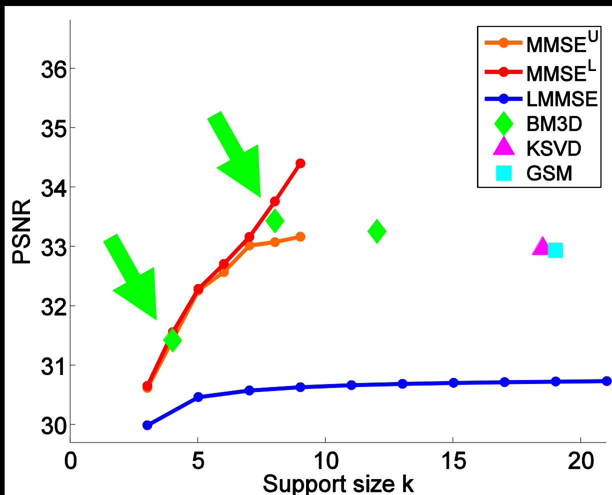
$\sigma = 55$



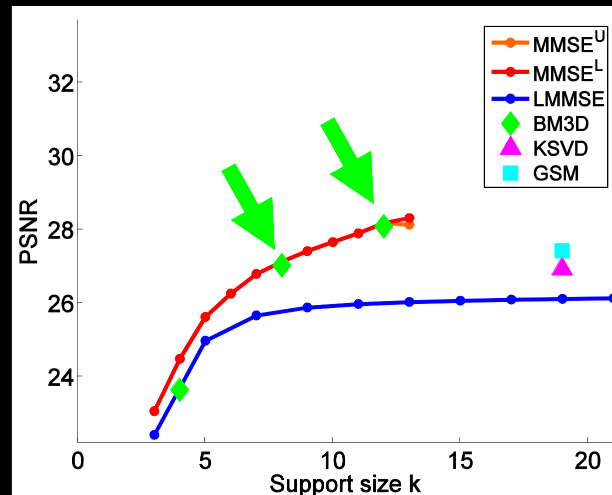
$\sigma = 170$

Observation 2: **For small patches/ large noise we have enough observations to compute the MMSE accurately. (lower and upper bounds coincide)**

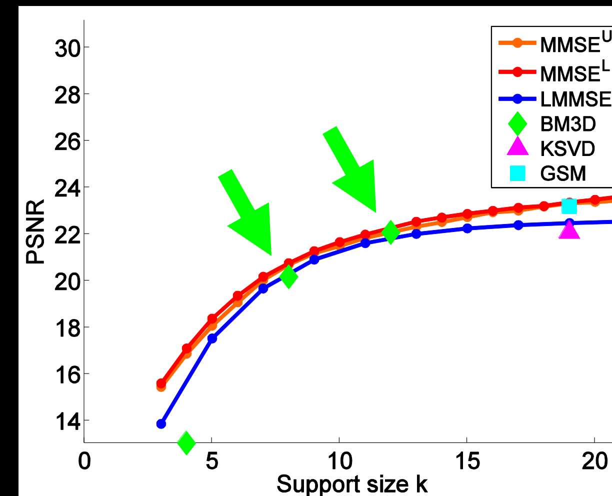
MMSE and MSE Evaluation



$\sigma = 18$



$\sigma = 55$

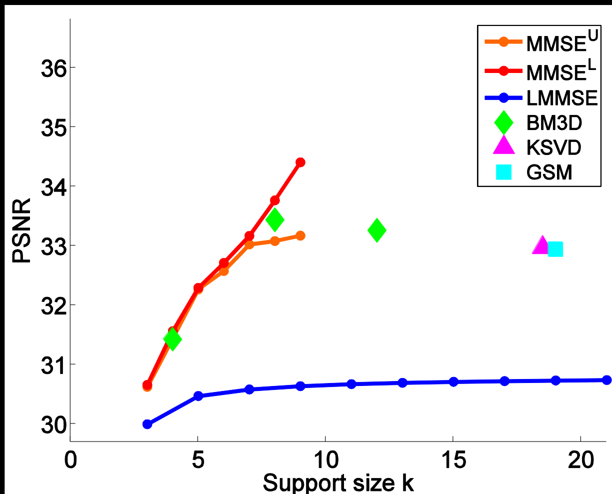


$\sigma = 170$

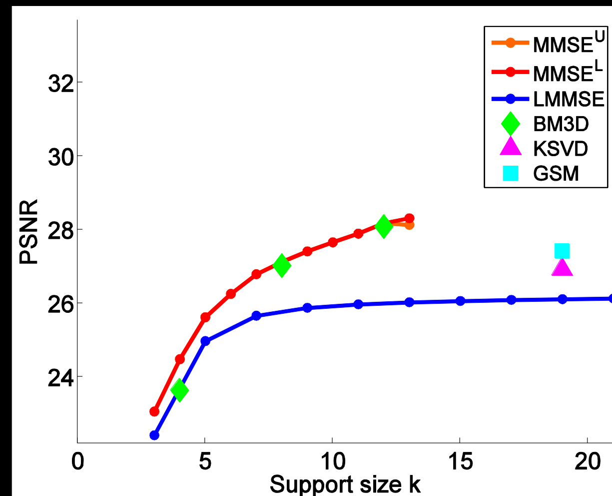
Observation 3: BM3D less than *0.1dB* from bound.

For small windows, state of the art denoising cannot be improved beyond *0.1dB* values

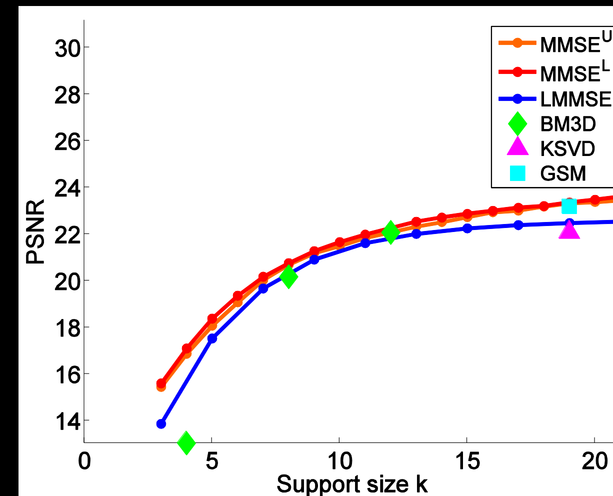
MMSE and MSE Evaluation



$\sigma = 18$



$\sigma = 55$

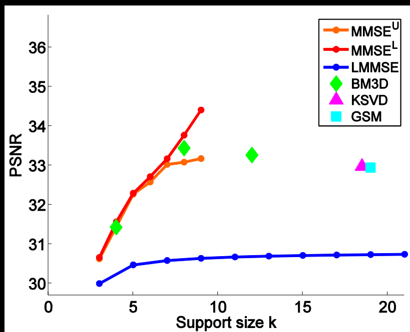


$\sigma = 170$

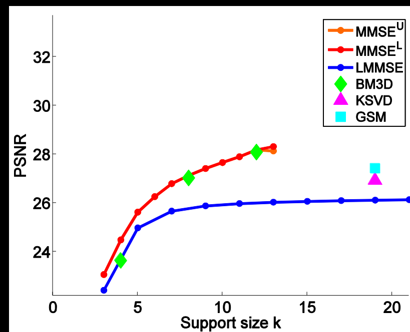
Observation 4: **wide windows can improve denoising quality, but non parametric techniques suffer from curse of dimensionality.**

Increasing window size probably requires parametric approaches

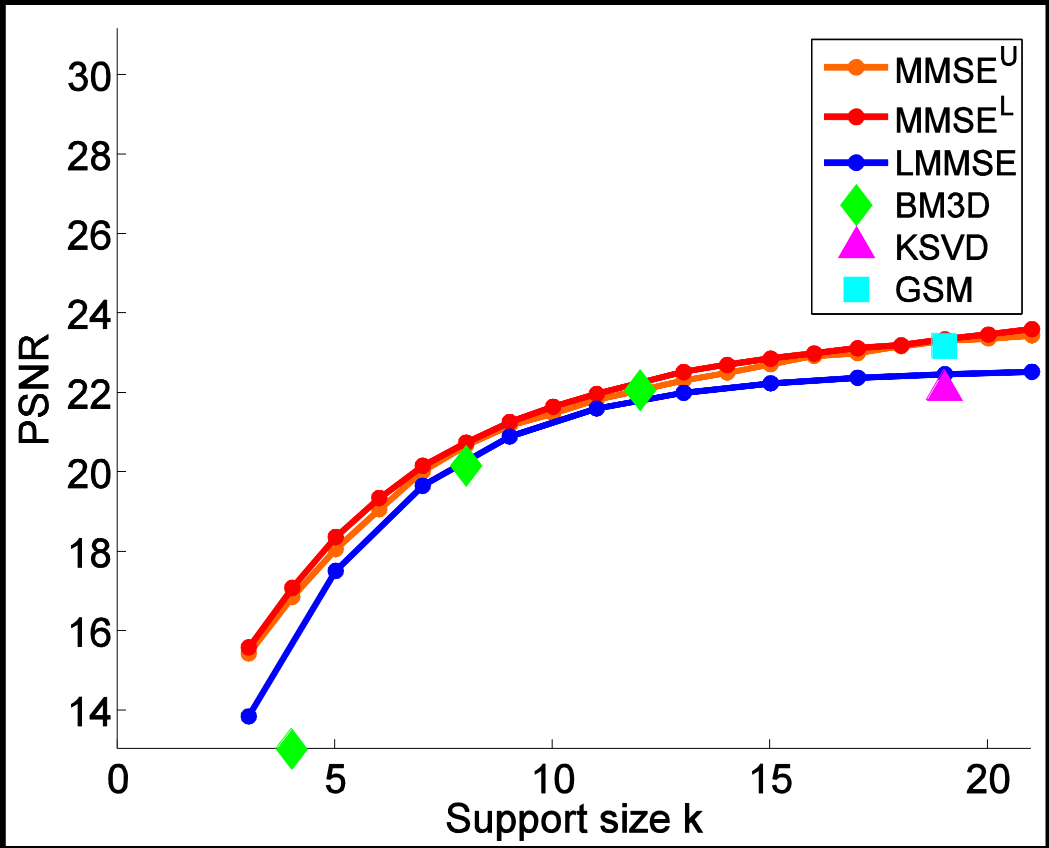
MMSE and MSE Evaluation



$\sigma = 18$



$\sigma = 55$



$\sigma = 170$

Observation 5: **Natural image priors are of interest at medium noise level. At high noise the error is solely determined by second order statistics- LMMSE (Wiener filter) does equally good.**

Visualizing optimal denoising

Original

Our Opt. MMSE
PSNR 23.93dB

KSVD
PSNR 22.41dB

Noisy input

BM3D
PSNR 23.86dB

GSM
PSNR 23.28dB

Gap: *0.07dB !*

Summary

- Statistical framework for optimal natural image denoising.

- Small windows / moderate to large noise, current methods are nearly optimal.

- Larger window size can yield better results - but need parametric approach!

- Implications to other low level vision tasks - deconvolution, super resolution.