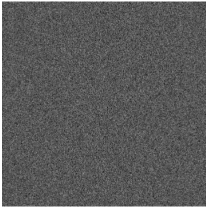
Oracle inequalities and minimax rates for non-local means

Rebecca Willett, Duke

Ery
Arias-Castro,
UCSD

Joseph
Salmon, Duke

 $=$  $+$ 

Observed image $\mathbf{y}$ | Underlying scene $\mathbf{f}$ | Noise ε

$$\mathbf{y} = \mathbf{f} + \varepsilon; \qquad \varepsilon \text{ uncorrelated, mean=0, var=}\sigma^2$$

Estimate f_i as a weighted average of the noisy pixels:

$$\widehat{f}_i = \sum_j w_{i,j} y_j$$

How should we choose the weights?

Kernel-based denoising: $\widehat{f}_i = \sum_j w_{i,j} y_j$

Usual kernel method [a]

$$w_{i,j} = K_h(x_i, x_j)$$

- w has no dependency on y
- K: kernel and h: bandwidth (smoothing parameter)
- Gaussian kernel example : $K_h(x_i, x_j) = e^{-\|x_i - x_j\|_2^2/2h^2}$

[a] Nadaraya '64, Watson '64

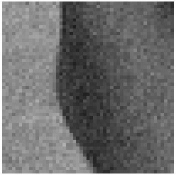
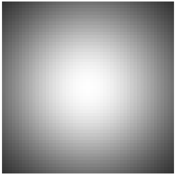


Image Search Zone



Spatial

Kernel-based denoising: $\widehat{f}_i = \sum_j w_{i,j} y_j$

Yaroslavsky/Bilateral Filter [a]

$$w_{i,j} = K_h(x_i, x_j) L_{h_y}(y_i, y_j)$$

- Use spatial *and* photometric proximity
- K, L: kernels; h, h_y: bandwidths (smoothing parameters)

[a] Yaroslavsky '85, Lee '83, Tomasi and Manduchi '98

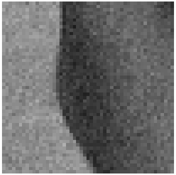
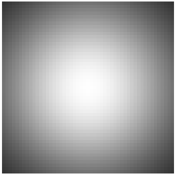
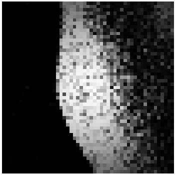


Image Search Zone



Spatial



Yaroslavsky / Bilateral

Kernel-based denoising: $\widehat{f}_i = \sum_j w_{i,j} y_j$

Non-local Means [a]

$$w_{i,j} = K_h(x_i, x_j) L_{h_y}(\mathbf{y}_{P_i}, \mathbf{y}_{P_j})$$

- Use spatial *and* photometric proximity
- K, L: kernels; h, h_y: bandwidths (smoothing parameters)
- P_i is a small patch of pixels centered around pixel i

[a] Buades, Coll & Morel '05

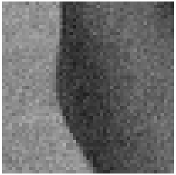
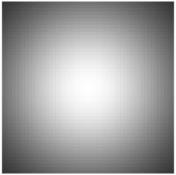
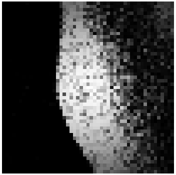


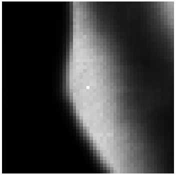
Image Search Zone



Spatial



Yaroslavsky / Bilateral



Non-local means

Problem formulation

We will bound the risk

$$\mathcal{R}_n(\widehat{\mathbf{f}}, \mathcal{F}) := \sup_{\mathbf{f} \in \mathcal{F}} \mathrm{MSE}_{\mathbf{f}}(\widehat{\mathbf{f}}) = \sup_{\mathbf{f} \in \mathcal{F}} \frac{\mathbb{E}\|\widehat{\mathbf{f}} - \mathbf{f}\|_2^2}{n^d}.$$

How do errors scale with

- n (number of pixels),
- d (dimension), and
- σ^2 (noise variance)?

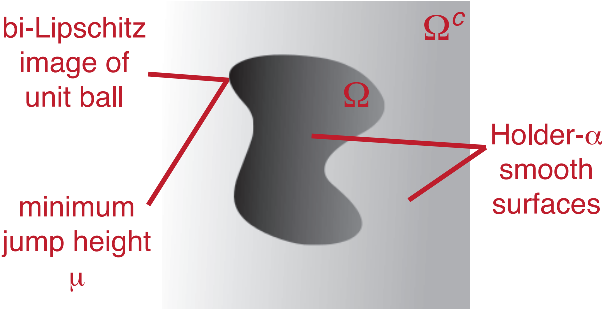
Related theoretical investigations

- Information-theoretic interpretation: Weissman *et al.* '05
- Consistency: Buades *et al.* '05
- Graph diffusion interpretation: Singer *et al.* '09, Taylor & Meyer '11
- Rare patch effect: Duval *et al.* '11
- SURE estimate of parameters: Van De Ville & Kocher '09,'11, Duval *et al.* '11, Deledalle *et al.* '11
- Cramer-Rao bounds: Levin & Nadler '11, Chatterjee & Milanfar '11
- Minimax rates for piecewise constant images: Maleki, Narayan & Baraniuk '11

Cartoon images

$f \in \mathcal{F}^{\text{cartoon}}$ is a "cartoon image" if it is a piecewise smooth (Hölder-α, $\alpha \geq 1$) image with discontinuities along smooth hypersurfaces.[1].

$$f(x) = \mathbb{1}_{\{x \in \Omega\}} f_{\Omega}(x) + \mathbb{1}_{\{x \in \Omega^c\}} f_{\Omega^c}(x),$$



[1] Korostelev and Tsybakov '93

Linear filtering bounds

- If the kernel intersects the boundary, boundary is blurred

$$\mathbb{E}\left((\widehat{f_i} - f_i)^2\right) \asymp 1.$$

$O(n^d h)$ pixels have kernels which intersect the boundary

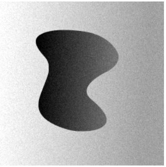
- If the kernel doesn't intersect the boundary,

$$\mathbb{E}\left((\widehat{f_i} - f_i)^2\right) \asymp h^{2\alpha} + \sigma^2 (nh)^{-d}$$

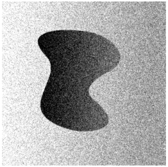
$$\mathcal{R}^{\mathrm{LF}} \asymp (\sigma^2/n^d)^{1/(d+1)}$$

This bound is independent of surface smoothness α!

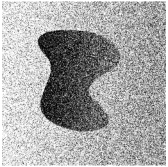
Linear filtering results



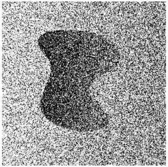
Noisy, MSE = 2.50e+01



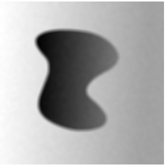
Noisy, MSE = 3.99e+02



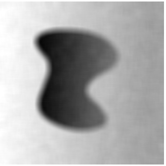
Noisy, MSE = 2.50e+03



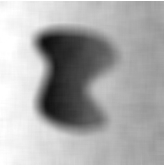
Noisy, MSE = 9.98e+03



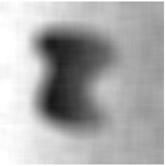
LF0, MSE = 3.52e+01



LF0, MSE = 7.78e+01

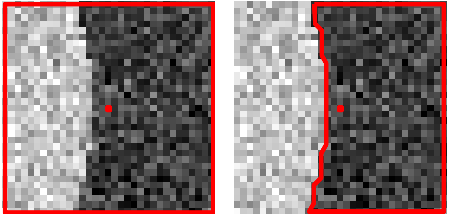


LF0, MSE = 1.51e+02



LF0, MSE = 2.43e+02

Membership oracle (the gold standard)



Kernel smoothing   Membership oracle

We use local polynomial regression[2] of order $r \geq \lfloor \alpha \rfloor$ over the kernel domain.

[2] Fan & Gijbels '96, Hastie, Tibshirani & Friedman '09

Membership oracle bounds

The analysis is very similar to linear filters – only now the kernel never intersects the boundary. This gives

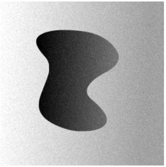
$$\mathcal{R}^{\mathrm{MO}} \asymp (\sigma^2/n^d)^{2\alpha/(d+2\alpha)}$$

Compare with linear filter, which had

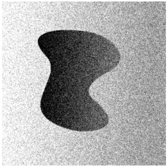
$$\mathsf{MSE} \asymp (\sigma^2/n^d)^{1/(d+1)}$$

for all α.

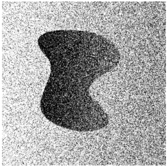
Membership oracle results



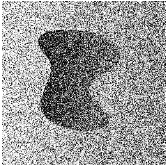
Noisy, MSE = 2.50e+01



Noisy, MSE = 3.99e+02

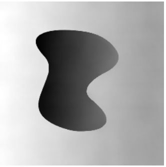


Noisy, MSE = 2.50e+03

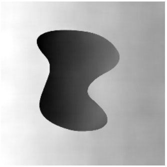


Noisy, MSE = 9.98e+03

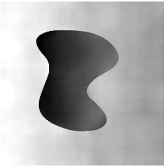
MO2, MSE = 9.57e-01



MO2, MSE = 2.37e+00



MO2, MSE = 6.09e+00

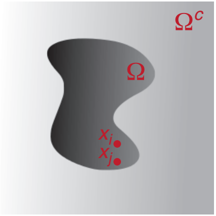


MO2, MSE = 1.96e+01

Yaroslavsky's filter bounds

Basic idea: if noise is small, then Yaroslavsky approximates the Membership Oracle.

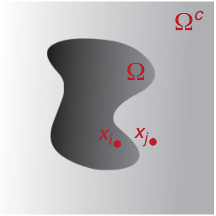
- f varies smoothly within Ω, so if $x_j \in \Omega$, we have an upper bound on $f_i - f_j$ and concentration bounds on $y_i - y_j$.



Yaroslavsky's filter bounds

Basic idea: if noise is small, then Yaroslavsky approximates the Membership Oracle.

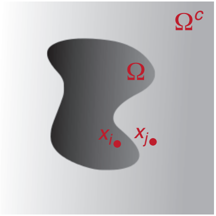
- f varies smoothly within Ω, so if $x_j \in \Omega$, we have an upper bound on $f_i - f_j$ and concentration bounds on $y_i - y_j$.
- We have a jump of height at least μ between Ω and Ω^c, so if $x_j \in \Omega^c$, we have a lower bound on $f_i - f_j$ and concentration bounds on $y_i - y_j$.



Yaroslavsky's filter bounds

Basic idea: if noise is small, then Yaroslavsky approximates the Membership Oracle.

- f varies smoothly within Ω, so if $x_j \in \Omega$, we have an upper bound on $f_i - f_j$ and concentration bounds on $y_i - y_j$.
- We have a jump of height at least μ between Ω and Ω^c, so if $x_j \in \Omega^c$, we have a lower bound on $f_i - f_j$ and concentration bounds on $y_i - y_j$.

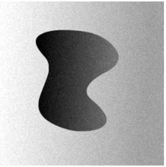


Thus if we choose h_y between these two bounds, we ensure that the y_j we select are in Ω with very high probability (for sufficiently small σ).

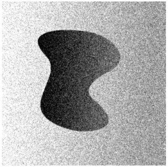
$$\mathcal{R}^{\mathrm{YF}} \leq (1 + o(1))\mathcal{R}^{\mathrm{MO}} \quad \text{for} \quad \sigma = O(1/\sqrt{\log n})$$

Yaroslavsky's filter results

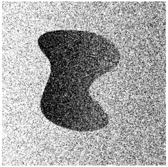
As predicted by theory, performance is very strong for low noise.



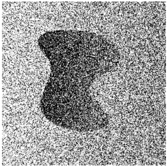
Noisy, MSE = 2.50e+01



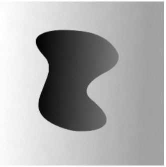
Noisy, MSE = 3.99e+02



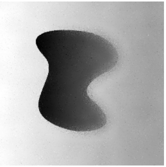
Noisy, MSE = 2.50e+03



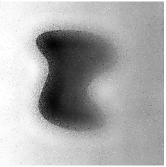
Noisy, MSE = 9.98e+03



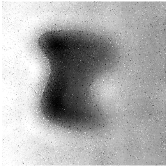
YF2, MSE = 9.37e-01



YF2, MSE = 1.90e+01



YF2, MSE = 1.46e+02

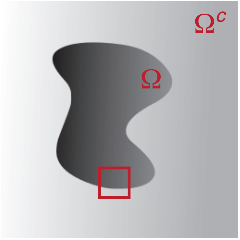


YF2, MSE = 2.98e+02

NLM bounds

Basic idea: patch distance approximates pixel distance, so NLM approximates membership oracle

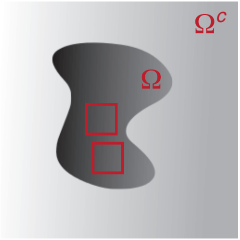
- If x_i is near the boundary, then the error can be $O(1)$, and there are $O(h_{\mathrm{P}} n^d)$ such pixels, where h_{P} is the patch sidelength.



NLM bounds

Basic idea: patch distance approximates pixel distance, so NLM approximates membership oracle

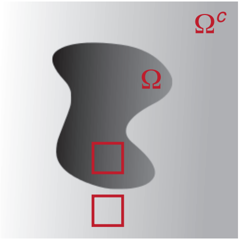
- If x_i is near the boundary, then the error can be $O(1)$, and there are $O(h_{\mathrm{P}} n^d)$ such pixels, where h_{P} is the patch sidelength.
- f varies smoothly within Ω, so if $\mathrm{P}_i, \mathrm{P}_j \subseteq \Omega$, we have an upper bound on $f_i - f_j$ and concentration bounds on $\|\mathbf{y}_{\mathrm{P}_i} - \mathbf{y}_{\mathrm{P}_j}\|_2$.



NLM bounds

Basic idea: patch distance approximates pixel distance, so NLM approximates membership oracle

- If x_i is near the boundary, then the error can be $O(1)$, and there are $O(h_{\mathrm{P}} n^d)$ such pixels, where h_{P} is the patch sidelength.
- f varies smoothly within Ω, so if $\mathrm{P}_i, \mathrm{P}_j \subseteq \Omega$, we have an upper bound on $f_i - f_j$ and concentration bounds on $\|\mathbf{y}_{\mathrm{P}_i} - \mathbf{y}_{\mathrm{P}_j}\|_2$.
- We have a jump of height at least μ between Ω and Ω^c, so if $x_j \in \Omega^c$, we have a lower bound on $f_i - f_j$ and concentration bounds on $\|\mathbf{y}_{\mathrm{P}_i} - \mathbf{y}_{\mathrm{P}_j}\|_2$.



NLM bounds

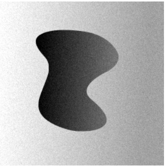
With high probability, NLM

- behaves like the membership oracle away from the boundary and
- behaves like the linear filter for a very small volume near the boundary.

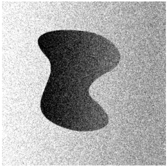
$$\mathcal{R}^{\mathrm{NLM}} \preceq \max\left(\frac{(\sigma^4 \log n)^{1/d}}{n}, (\sigma^2/n^d)^{2\alpha/(d+2\alpha)}\right)$$

NLM results

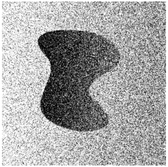
Because NLM uses entire patch to measure similarity between pixels, kernel weights are more robust to noise.



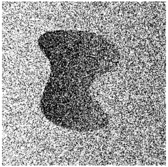
Noisy, MSE = 2.50e+01



Noisy, MSE = 3.99e+02

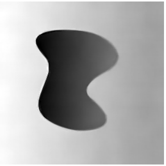


Noisy, MSE = 2.50e+03

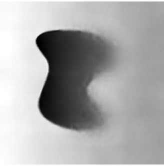


Noisy, MSE = 9.98e+03

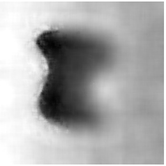
NLM2, MSE = 1.30e+00



NLM2, MSE = 4.92e+00



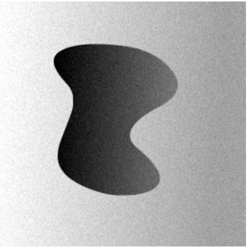
NLM2, MSE = 3.74e+01



NLM2, MSE = 1.37e+02

Examples, $\sigma = 5$

With low noise, all methods perform well.



Noisy, MSE = 2.50e+01

LF2, MSE = 7.21e+01

YF2, MSE = 9.37e-01

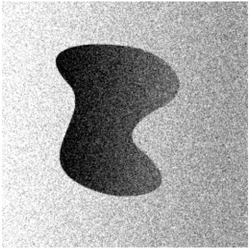
NLM2, MSE = 1.30e+00

MO2, MSE = 9.57e-01

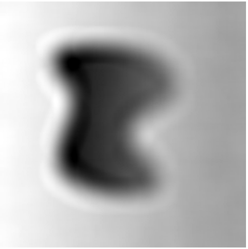
Examples, $\sigma = 5$

Examples, $\sigma = 20$

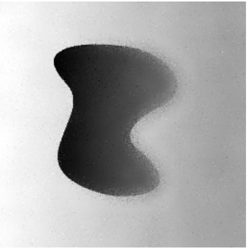
As noise increases, we first see the linear filter start to break down.



Noisy, MSE = 3.99e+02



LF2, MSE = 1.40e+02



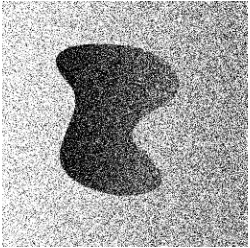
YF2, MSE = 1.90e+01

NLM2, MSE = 4.92e+00

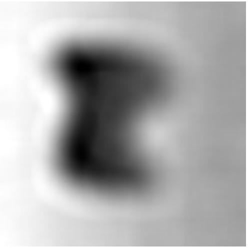
MO2, MSE = 2.37e+00

Examples, $\sigma = 50$

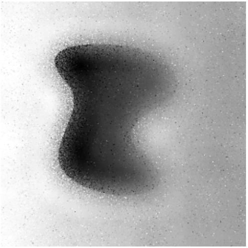
With even more noise, Yaroslavsky's filter starts to perform poorly.



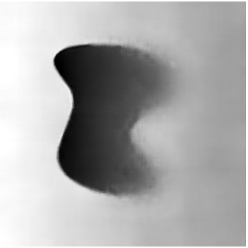
Noisy, MSE = 2.50e+03



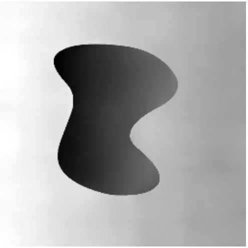
LF2, MSE = 2.11e+02



YF2, MSE = 1.46e+02



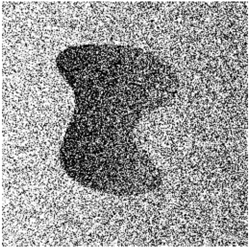
NLM2, MSE = 3.74e+01



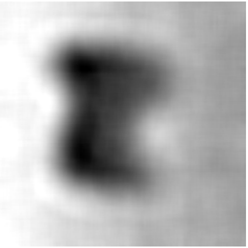
MO2, MSE = 6.09e+00

Examples, $\sigma = 100$

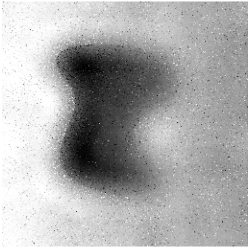
We also see how performance varies with the size of the "jump".



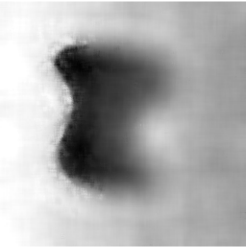
Noisy, MSE $= 9.98e{+}03$



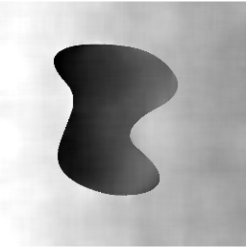
LF2, MSE $= 2.29e{+}02$



YF2, MSE $= 2.98e{+}02$



NLM2, MSE $= 1.37e{+}02$



MO2, MSE $= 1.96e{+}01$

Repeating patterns

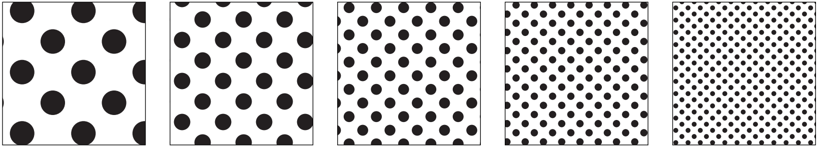
As before,

$$f(x) = \mathbb{1}_{\{x \in \Omega\}} f_\Omega(x) + \mathbb{1}_{\{x \in \Omega^c\}} f_{\Omega^c}(x),$$

but now

$$\Omega = (0,1)^d \cap \bigcup_{v \in a\mathbb{Z}^d} (\Xi + v)$$

where a is the pattern period and $a \to 0$ as $n \to \infty$. This function class is like the cartoon class, but the underlying scene (especially the frequency of repetition) scales with n.



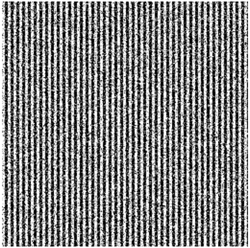
n small $\longrightarrow$ n medium $\longrightarrow$ n large

Performance bounds for patterns

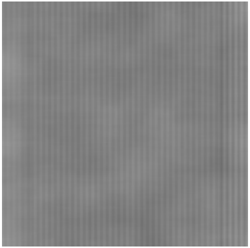
Consider $f \in \mathcal{F}^{\mathrm{pattern}}$. Assume the volumes of Ω and Ω^c are comparable.

- MO with $h = h^{\mathrm{MO}}$ achieves an MSE of order $\mathcal{R}^{\mathrm{MO}}$
- YF with $h^{\mathrm{MO}}, h_y \asymp 1$ achieves an MSE of order $\mathcal{R}^{\mathrm{MO}}$ if the noise is low
- NLM with bandwidths $h = h^{\mathrm{MO}}, h_y = h_y^{\mathrm{NLM}}$ and patch size $h_{\mathrm{P}}^{\mathrm{NLM}}$ achieves an MSE of order $(na)^d\, \mathcal{R}^{\mathrm{MO}}$ if the pattern is sufficiently "strong" (foreground-centered patches must be distinct from background-centered patches).

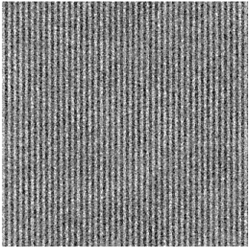
Examples, $\sigma = 100$



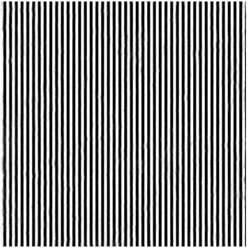
Noisy, MSE = 9.98e+03



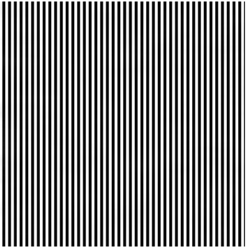
LF2, MSE = 1.71e+04



YF2, MSE = 8.87e+03

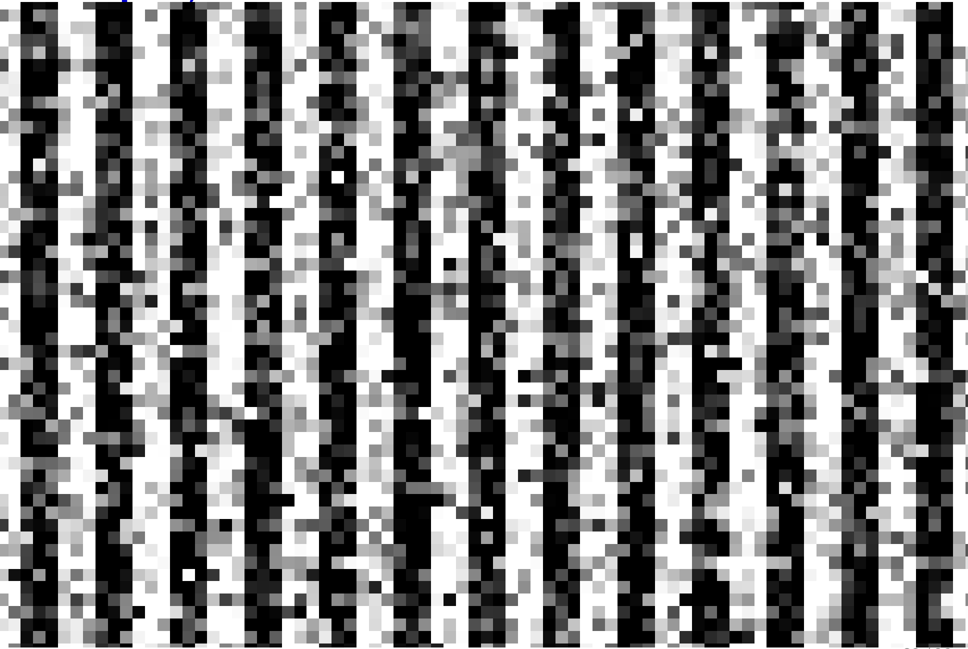


NLM2, MSE = 2.33e+02

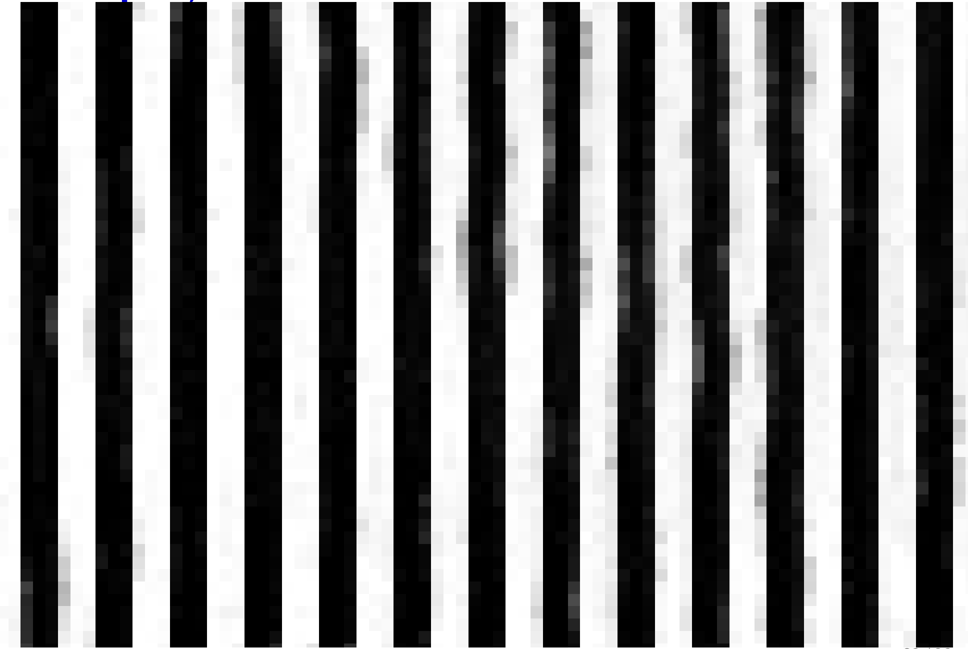


MO2, MSE = 2.28e+01

Examples, $\sigma = 100$



Examples, $\sigma = 100$



Examples, $\sigma = 20$

Noisy, MSE = 3.99e+02

LF2, MSE = 9.01e+02

YF2, MSE = 2.44e+02

NLM2, MSE = 1.31e+02

MO2, MSE = 4.65e+01

Conclusions

- Novel membership oracle gives new insight into key limitations of adaptive filtering methods.
- The classical Yaroslavsky's method behaves optimally at low noise levels.
- NLM mimics Yaroslavsky's filter, but uses patches to robustly determine pixel similarity.
- Novel image class describes repeating patterns and redundancy not present in classical image models and not well-suited to methods like wavelet thresholding – we show how NLM performs well in this setting.

Thank you.

http://arxiv.org/abs/1112.4434