Advanced Image Processing

Part VI:

Image Denoising

S. Voloshynovskiy

Recommended books

- A. K. Jain, Fundamentals of Digital Image Processing, Prentice-Hall, 1989.
- R. Lagendijk and J. Biemond, Iterative Identification and restoration of Images, Kluwer Academic Publishers, 1991.
- M. Bertero and P. Boccacci, Introduction to Inverse Problems in Imaging, IOP Publishing LTD, 1998.
- A.N. Tikhonov and V.Y. Arsenin, Solutions of ill-posed problems, Washington: Winston/Willey, 1977.
- V.A. Morozov, Methods for Solving Incorrectly Posed Problems, Springer, 1984.

Recommended books

Robust estimation:

- P.J. Huber. *Robust Statistics*. John Wiley & Sons, New York, 1981.
- F.R. Hampel. Robust estimation: A condensed partial survey. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, 27:87-104, 1973
- P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*. John Wiley & Sons, New York, 1987.

Roadmap:

1. Introduction. Elements of Estimation Theory:
    - Maximum-Likelihood (ML) Estimate
    - Properties of Estimators
    - Maximum a Posteriori (MAP) Estimate: Role of Prior Information
2. ML-estimators:
    - Removal of additive noise
    - Robust M-estimators
3. MAP-estimators: Removal of Gaussian noise
   (Wiener, soft-shrinkage and hard-thresholding)
4. Penalized Maximum Likelihood (PML) Estimators
5. Impulse noise removal using prediction models
6. Removal of speckle

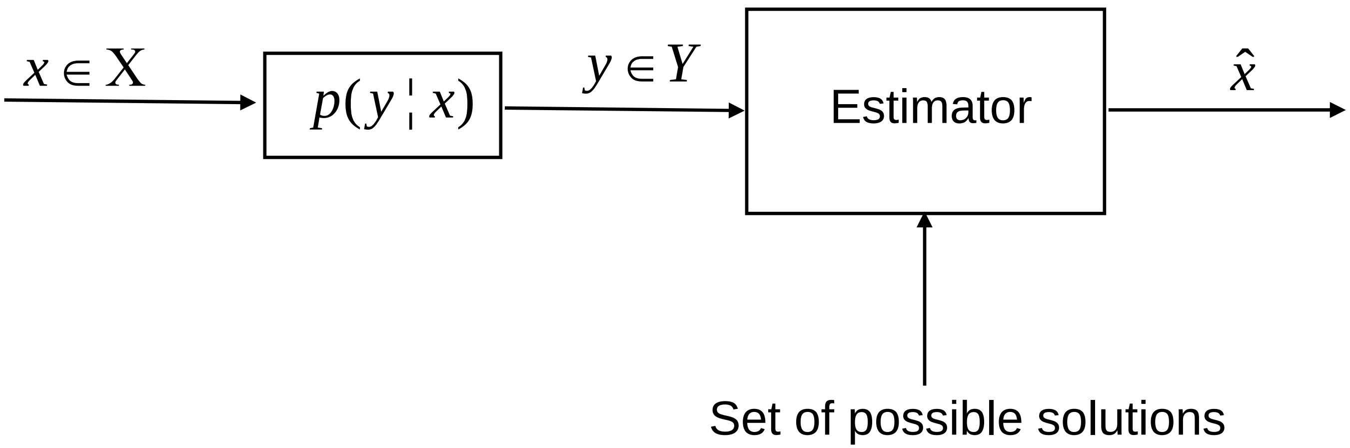
1. General Model of Stochastic Image Denoising

$$x \in X \longrightarrow \boxed{p(y|x)} \longrightarrow y \in Y$$

- Sets *X* and *Y*: continuous or discrete-domains

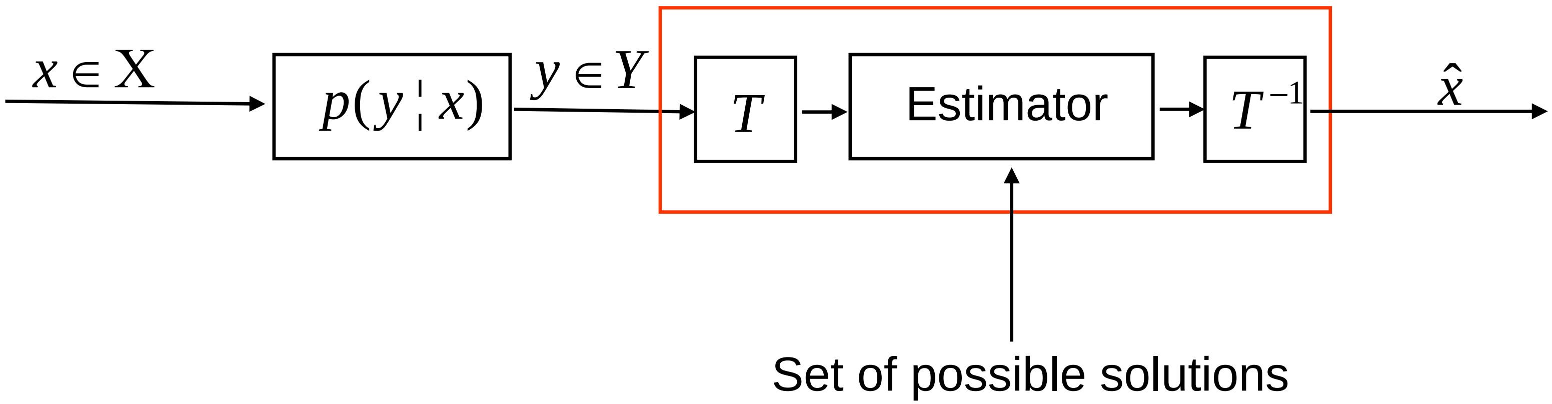
  For mathematical convenience we usually work in discrete domains
- Conditional pdf $p(y|x)$ models the degradation process.

1. General Model of Stochastic Image Denoising



1. General Model of Stochastic Image Denoising

Transform domain



T and T^{-1} direct and inverse transforms:

- Fourier
- DCT
- wavelet

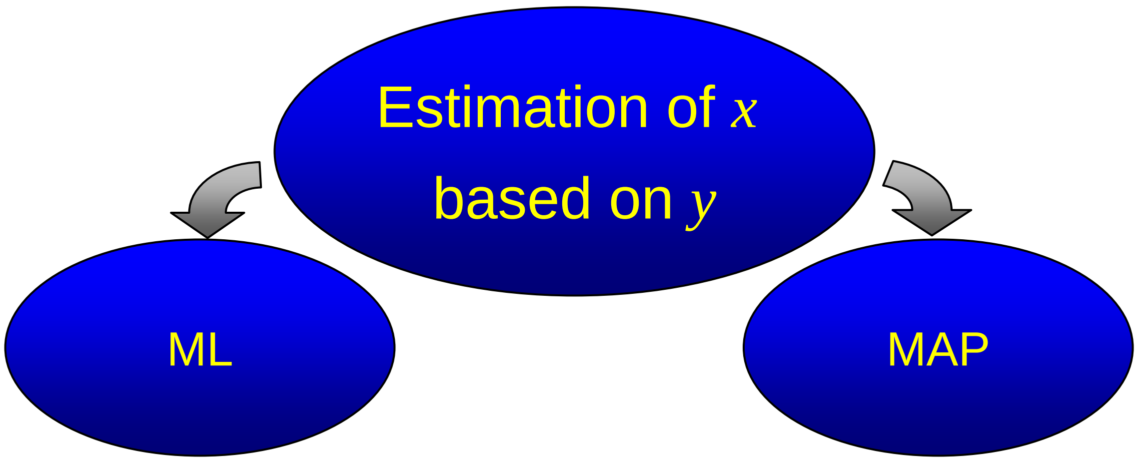
1. General Model of Stochastic Image Denoising

Advantages of estimation in the transform domain:

- partial image decorrelation (approximation of KLT);
- energy compaction;
- possibility to incorporate Human Visual System.

1. Elements of Estimation Theory

$$x \in X \longrightarrow \boxed{p(y|x)} \longrightarrow y \in Y$$



x :

- deterministic, unknown image
- constant on the observation interval
- no reliable statistics

x :

- random image that is completely described by pdf p(x)

1. Maximum-Likelihood (ML) Estimate

$$x \in X \longrightarrow \boxed{p(y \mid x)} \longrightarrow y \in Y$$

- ML is an estimation method which is applicable to arbitrary degradation models $p(y \mid x)$.
- Assume that y is a N-vector and x is a deterministic, unknown image that is completely described by some K-dimensional parameter θ.
- Then estimating x from the data y is equivalent to estimating θ from y.

Ex.:

- constant image: $x(n) = \theta, \ \forall n \ (K = 1)$
- planar patch : $x(n) = \theta_1 + \theta_2 n_1 + \theta_2 n_2, \ (K = 3)$

1. Performance Measures of Estimators

Any estimator:

$$\boxed{\hat{\theta} = f\left(y, N, Model\right)}$$

The performance measures of any estimator:

- Expected value of estimate: $E\left[\hat{\theta}\right]$
- Bias of estimate: $E\left[\hat{\theta} - \theta\right] = E\left[\hat{\theta}\right] - \theta$
- Covariance of estimate: $Cov\left[\hat{\theta}\right] = E\left[\left(\hat{\theta} - E\left[\hat{\theta}\right]\right)\left(\hat{\theta} - E\left[\hat{\theta}\right]\right)^T\right]$

Optimal estimators aim at zero bias and minimum estimation error covariance.

1. Estimate: Desirable properties

Desirable properties of any estimator $\hat{\theta}(y)$ of θ are the following:

(a) Unbiasedness: $\boxed{E\left[\hat{\theta}\right]=\theta}$

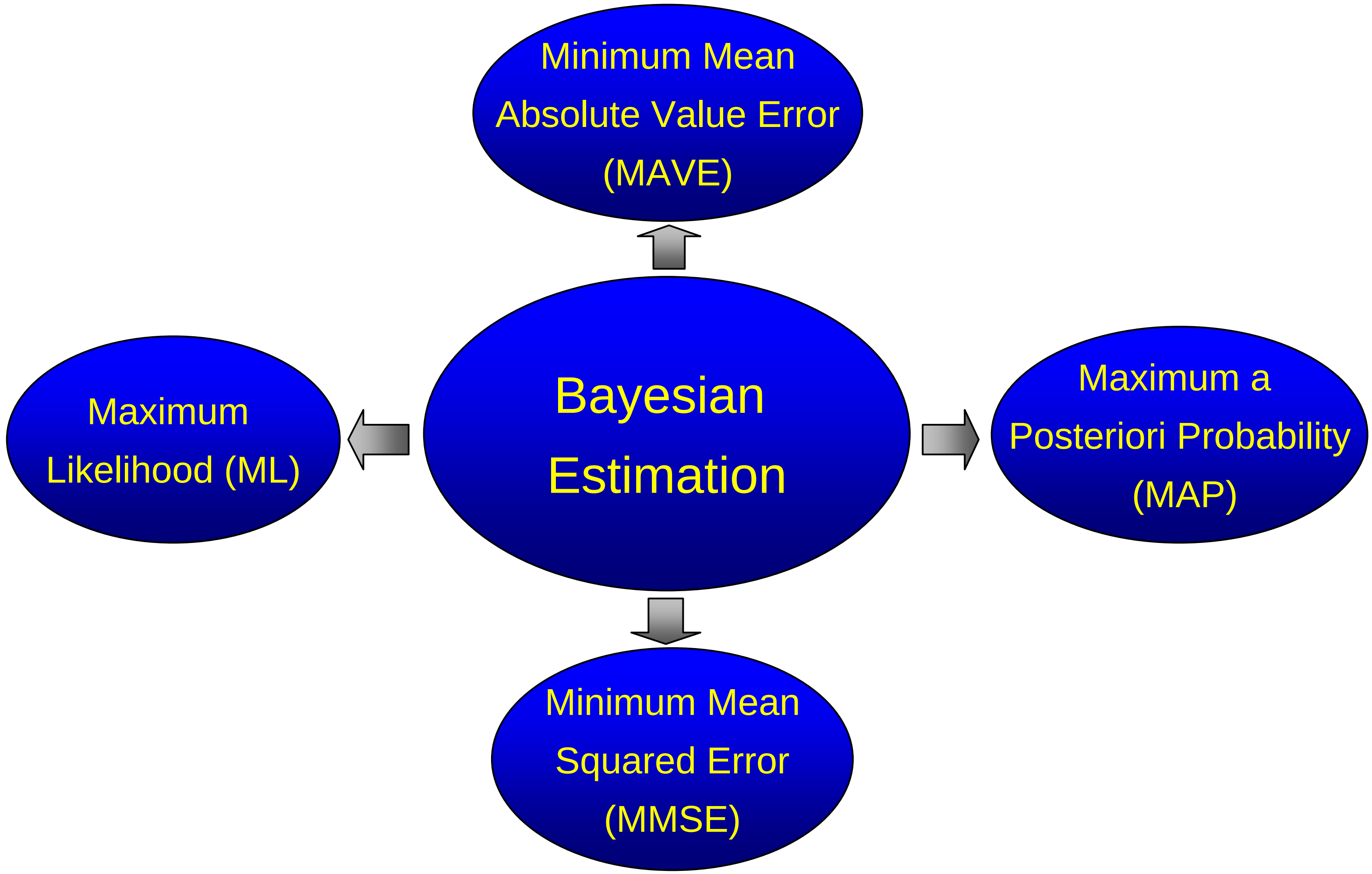
An estimator is *asymptotically unbiased* if for increasing length of observations *N* we have: $\boxed{\lim_{N\to\infty} E\left[\hat{\theta}\right]=\theta}$

(b) Efficient estimator: An unbiased estimator of θ is an efficient estimator if it has the smallest covariance matrix compared to all other unbiased estimators of θ: $\boxed{\operatorname{cov}\left[\hat{\theta}_{Efficient}\right]\le \operatorname{cov}\left[\hat{\theta}\right]}$

(c) Consistent estimator in probability: $\hat{\theta}\to\theta$ in probability, as $N\to\infty$

$$\boxed{\lim_{N\to\infty} P\left[\left|\hat{\theta}-\theta\right|>\varepsilon\right]=0}$$

1. Bayesian Estimation: general framework

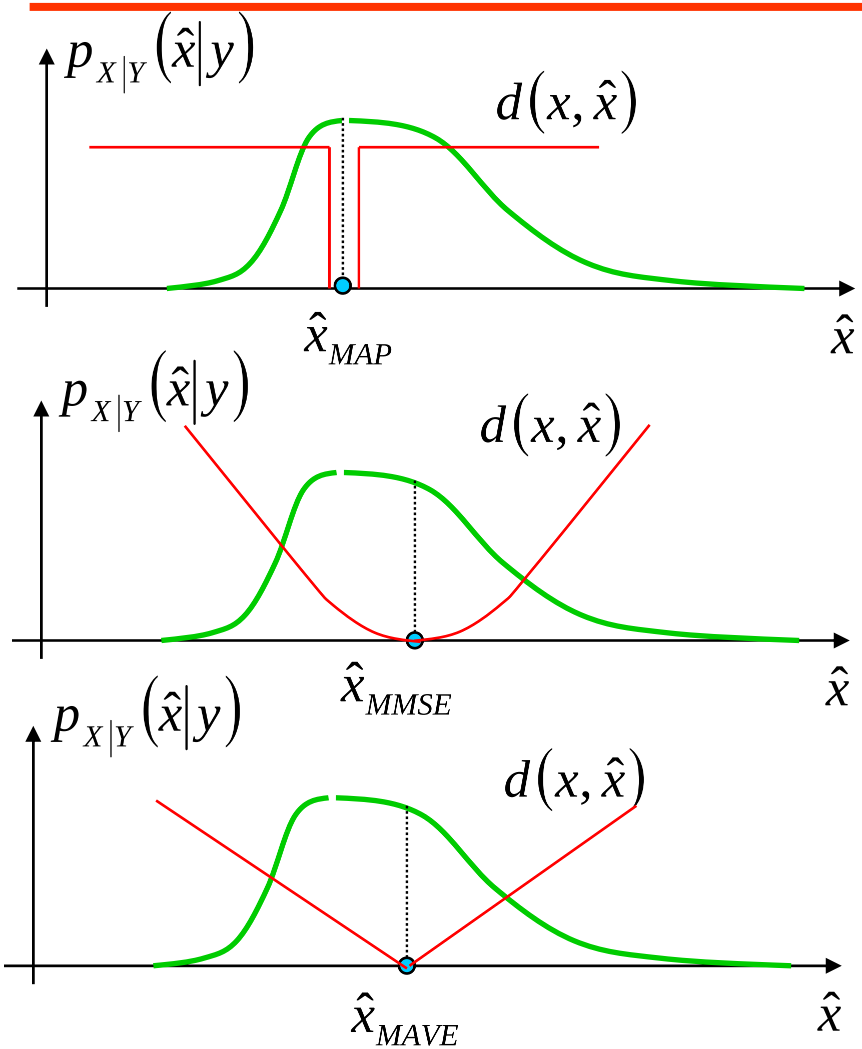


1. Bayesian Estimation: general framework (Poor, ch.IV.B)

- Unlike the ML and the MAP estimation methods, which find a theoretical justification in asymptotic setup, Bayesian estimation methods yield estimates that are optimal for arbitrary sample size.

- The key ingredient of this estimation technique is the definition of a cost function $d(x,\hat{x})$ which quantifies the quality of an estimate $\hat{x}$ of x.

1. Bayesian Estimation: cost-of-error function



$$d(x,\hat{x}) = \begin{cases} 0, \ if \ |x-\hat{x}| < 0, \\ 1, \ else. \end{cases}$$

MAP

$$d(x,\hat{x}) = \|x-\hat{x}\|_2^2$$

MMSE

$$d(x,\hat{x}) = |x-\hat{x}|_1$$

MAVE

1. Bayesian Estimation

- The Bayesian estimation of a parameter vector x is based on the minimization of a Bayesian risk function defined as an average cost-of-error function:

$$\boxed{\begin{aligned}\Re(\hat{x}) &= E\left[d(x,\hat{x})\right] = \int_X \int_Y d(x,\hat{x})\, p_{Y,X}(y,x)\, dy dx \\ &= \int_X \int_Y d(x,\hat{x})\, p_{X|Y}(x|y)\, p_Y(y)\, dy dx\end{aligned}}$$

- Since $p_Y(y)$ is constant for a given observation vector y and has no effect on the risk minimization, we can rewrite:

$$\boxed{\Re(\hat{x}|y) = E\left[d(x,\hat{x})\right] = \int_X d(x,\hat{x})\, p_{X|Y}(x|y)\, dx}$$

1. Bayesian Estimation

- The Bayesian estimate is obtained as the minimum-risk parameter vector:

$$\hat{x}_{Bayesian} = \arg\min_{\hat{x}} \Re\left(\hat{x} \middle| y\right) = \arg\min_{\hat{x}} \left[\int_X d\left(x, \hat{x}\right) p_{X|Y}\left(x \middle| y\right) dx\right]$$

- Using Bayes' rule:

$$\hat{x}_{Bayesian} = \arg\min_{\hat{x}} \Re\left(\hat{x} \middle| y\right) = \arg\min_{\hat{x}} \left[\int_X d\left(x, \hat{x}\right) p_{Y|X}\left(y \middle| x\right) p_X\left(x\right) dx\right]$$

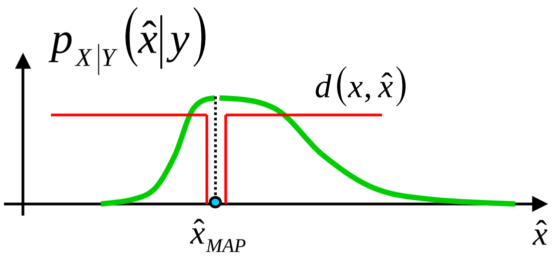
- Solution:

$$\hat{x}_{Bayesian} = \arg zero_{\hat{x}} \nabla_{\hat{x}} \Re\left(\hat{x} \middle| y\right)$$

1. Bayesian Estimation: Maximum a Posteriori (MAP)

- The cost function (so-called uniform cost):

$$\boxed{d(x,\hat{x}) = \begin{cases} 0, \ if \ \left|x-\hat{x}\right| < \Delta_{|\Delta\to 0}, \\ 1, \ else. \end{cases} = 1-\delta(x,\hat{x})}$$



- The Bayesian risk:

$$\Re_{MAP}(\hat{x}|y) = \int_X [1-\delta(x,\hat{x})] p_{X|Y}(x|y)dx = 1 - p_{X|Y}(\hat{x}|y)$$

- Therefore, the minimum is achieved for the maximum of the posterior function (mode of estimate)

$$\boxed{\hat{x}_{MAP} = \arg\max_{\hat{x}} p_{X|Y}(\hat{x}|y) = \arg\max_{\hat{x}} \left[p_{Y|X}(y|x) p_X(x) \right]}$$

1. Bayesian Estimation: Maximum-Likelihood (ML)

- The cost function is uniform and a uniform parameter prior pdf:
- The Bayesian risk:

$$\Re_{ML}(\hat{x}|y) = \int_X [1-\delta(x,\hat{x})] p_{Y|X}(y|x) p_X(x) dx = const\left[1 - p_{Y|X}(y|\hat{x})\right]$$

- Therefore, the ML estimator either does not use prior at all or assumes the uniform (non-informative) prior.

$$\boxed{\hat{x}_{ML} = \arg\max_{\hat{x}} \left[p_{Y|X}(y|x)\right]}$$

- In practice it is convenient to maximize the log-likelihood function instead of the likelihood:

$$\hat{x}_{ML} = \arg\max_{\hat{x}} \log\left[p_{Y|X}(y|x)\right]$$

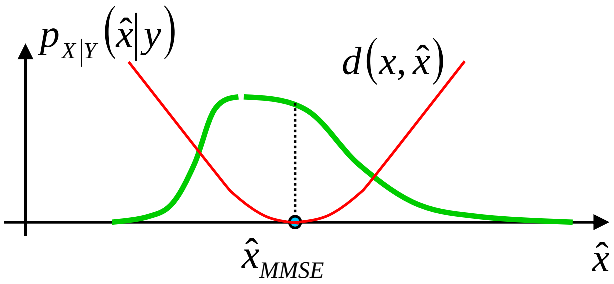
1. Bayesian Estimation: Maximum-Likelihood (ML)

- The log-likelihood is usually chosen in practice because:
  - the logarithm is a monotonic function, and hence the log-likelihood has the same turning points as the likelihood function;
  - the joint log-likelihood of a set of independent variables is the sum of the log-likelihoods of individual elements; and
  - unlike the likelihood function, the log-likelihood has a dynamic range that does not cause the computational under-flow.

1. Bayesian Estimation: Minimum Mean Square Error

- The cost function L_2:

$$\boxed{d(x,\hat{x}) = \|x-\hat{x}\|_2^2}$$



- The Bayesian risk:

$$\Re_{MMSE}(\hat{x}|y) = E\left[(x-\hat{x})^2 \,|\, y\right] = \int_X (x-\hat{x})^2 p_{X|Y}(x|y)dx$$

- The solution:

$$\boxed{\hat{x}_{MMSE} = \arg zero_{\hat{x}} \nabla_{\hat{x}} \Re(\hat{x}|y) = 2\int_x x p_{X|Y}(x|y)dx - 2\hat{x}\underbrace{\int_x p_{X|Y}(x|y)dx}_{1}}$$

1. Bayesian Estimation: Minimum Mean Square Error

$$\hat{x}_{MMSE} = \arg zero_{\hat{x}} \nabla_{\hat{x}} \Re\left(\hat{x}\middle|y\right) = 2\int_{x} x p_{X|Y}\left(x\middle|y\right)dx - 2\hat{x}$$

$$\boxed{\hat{x}_{MMSE} = \int_{x} x p_{X|Y}\left(x\middle|y\right)dx}$$

The Bayesian MMSE estimator is the conditional mean of the posterior pdf.

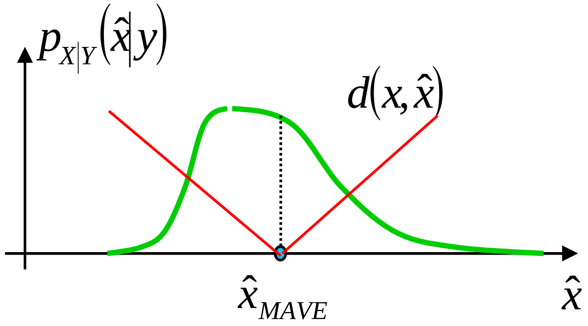
- For cases where we do not have a pdf model of x, the MMSE is reduced to least square error (LSE) estimator:

$$\boxed{\hat{x}_{LSE} = \arg\min_{\hat{x}} E\left[e^{2}\left(\hat{x}\middle|y\right)\right]}$$

1. Bayesian Estimation: Minimum Absolute Value of Error (MAVE)

- The cost function L_1 :

$$\boxed{d(x,\hat{x}) = \|x-\hat{x}\|_1}$$



- The Bayesian risk:

$$\Re_{MAVE}(\hat{x}|y) = E[|x-\hat{x}||y] = \int_X |x-\hat{x}|\, p_{X|Y}(x|y)dx$$

- or:

$$\Re_{MAVE}(\hat{x}|y) = \int_{-\infty}^{\hat{x}} [\hat{x}-x]\, p_{X|Y}(x|y)\,dx + \int_{\hat{x}}^{\infty} [x-\hat{x}]\, p_{X|Y}(x|y)\,dx$$

- Taking derivative:

$$\nabla_{\hat{x}} \Re_{MAVE}(\hat{x}|y) = \int_{-\infty}^{\hat{x}} p_{X|Y}(x|y)dx - \int_{\hat{x}}^{\infty} p_{X|Y}(x|y)dx$$

1. Bayesian Estimation: Minimum Absolute Value of Error (MAVE)

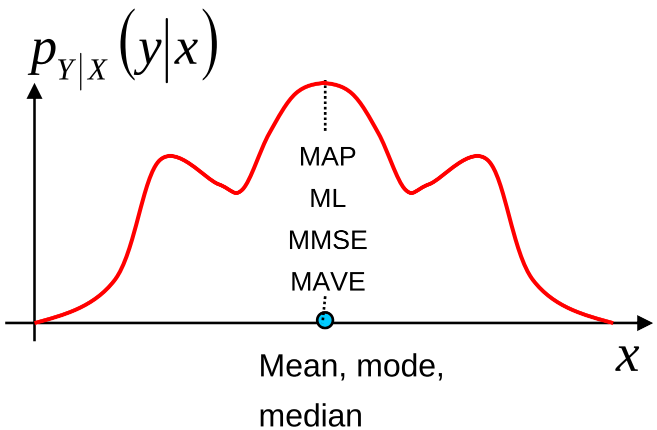
$$\nabla_{\hat{x}} \Re_{MAVE}\left(\hat{x} \mid y\right) = \int_{-\infty}^{\hat{x}} p_{X|Y}\left(x \mid y\right) dx - \int_{\hat{x}}^{\infty} p_{X|Y}\left(x \mid y\right) dx = 0$$

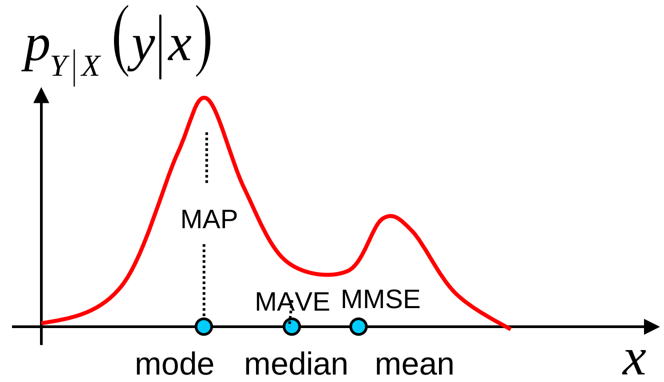
$$\boxed{\int_{-\infty}^{\hat{x}_{MAVE}} p_{X|Y}\left(x \mid y\right) dx = \int_{\hat{x}_{MAVE}}^{\infty} p_{X|Y}\left(x \mid y\right) dx}$$

The Bayesian MAVE estimator is the median of the posterior pdf.

There are several fast implementations to find the median in Matlab and C.

1. Relationships between MAP, MAVE and MMSE

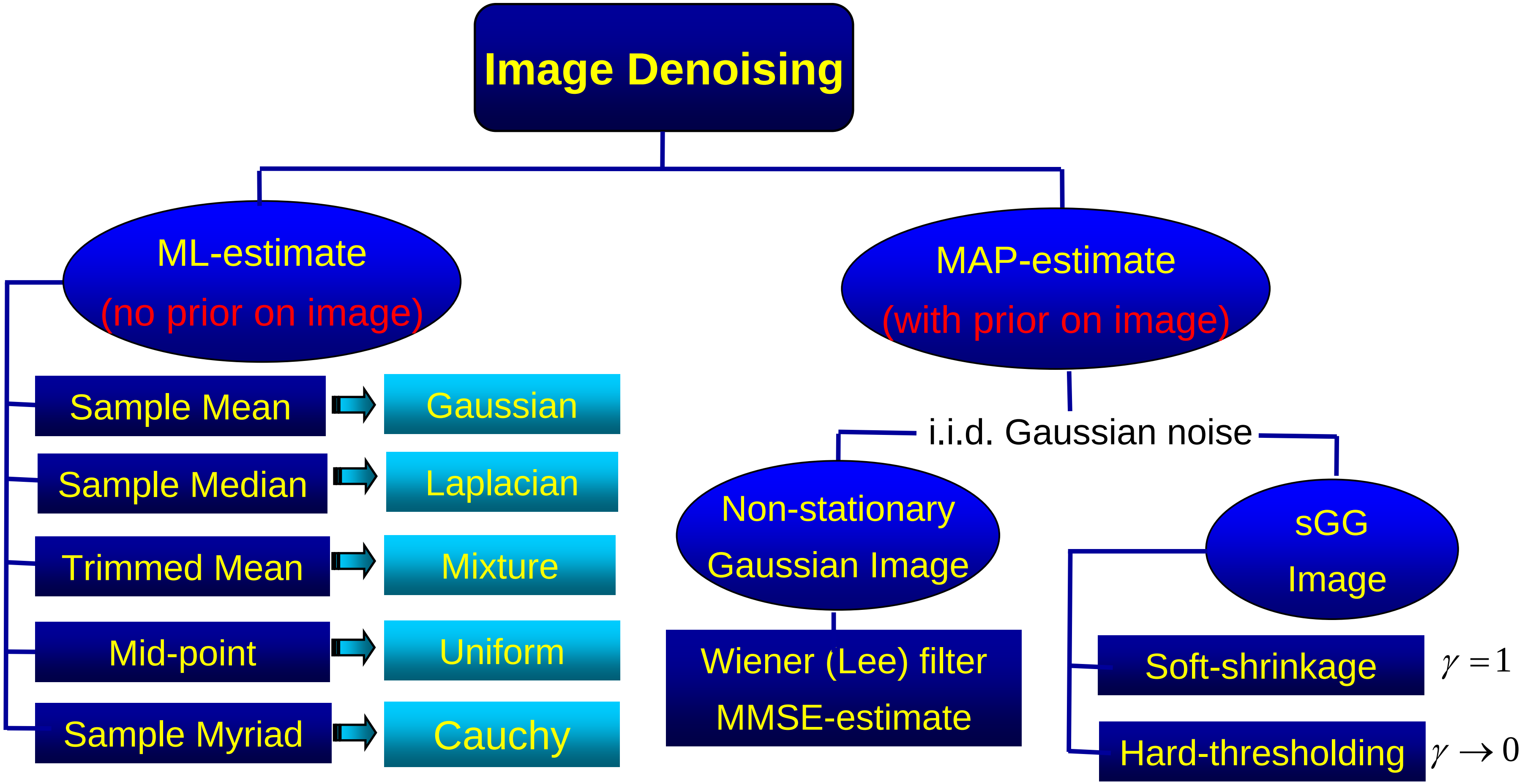




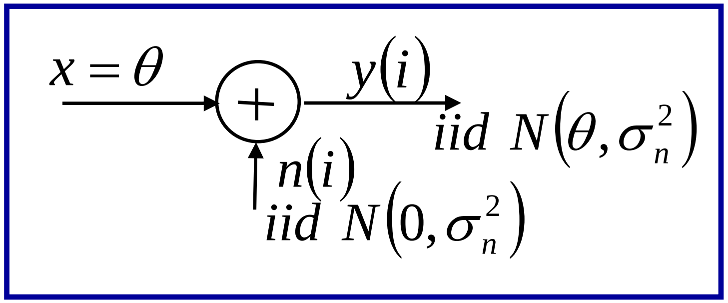
Properties of estimators:

- For a Gaussian a posteriori pdf: ML and LSE are identical;
- The MAP estimate of a Gaussian parameter tends to the ML and LSE estimates, if the parameter variance increases or equivalently as the parameter prior pdf tends to a uniform distribution;
- In general, for any symmetric distribution, centered round the maximum, the mode, the mean and the median are identical (MAP, ML, MMSE and MAVE are identical).

2. ML and MAP-Estimators: Additive Noise



2. ML-Estimate: Additive White Gaussian Noise



$$\boxed{y(i)=\theta+n(i)}$$ Constant image in AWGN

- Estimate θ using ML-estimate.
- Likelihood function:

$$L(\theta)=\prod_{i=0}^{N-1}\left(\frac{1}{\sqrt{2\pi\sigma_n^2}}e^{-\frac{|y(i)-\theta|^2}{2\sigma_n^2}}\right)$$

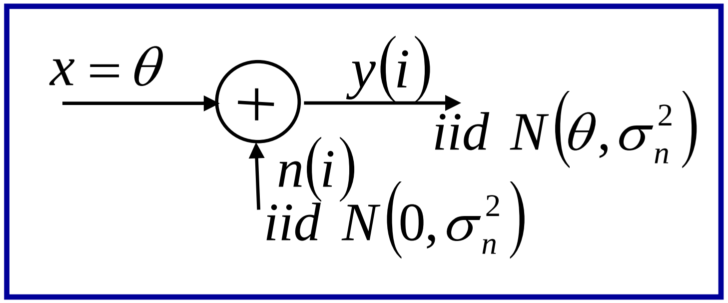
- Log-Likelihood function:

$$\ell(\theta)=-\frac{N}{2}\ln\left(2\pi\sigma_n^2\right)-\frac{1}{2\sigma_n^2}\sum_{i=0}^{N-1}|y(i)-\theta|^2$$

- Setting derivative to zero yields:

$$0=\frac{d\ell(\theta)}{d\theta}=\sum_{i=0}^{N-1}y(i)-N\theta\Rightarrow\boxed{\hat{\theta}_{ML}=\frac{1}{N}\sum_{i=0}^{N-1}y(i)}$$ Sample mean

2. ML-Estimate: Additive White Gaussian Noise



$$\boxed{y(i)=\theta+n(i)}$$ Constant image in AWGN

- Unbiased estimate:

$$\boxed{E\left[\hat{\theta}_{ML}\right]=E\left[\frac{1}{N}\sum_{i=0}^{N-1}\left|\theta+n(i)\right|\right]=\theta\Big|_{N\to\infty}}$$

- The variance of estimate:

$$\boxed{Var\left[\hat{\theta}_{ML}\right]=E\left[\left(\hat{\theta}_{ML}-\theta\right)^2\right]=E\left[\left(\frac{1}{N}\sum_{i=0}^{N-1}\underbrace{y(i)}_{iid\ N(0,\sigma_n^2)}-\theta\right)^2\right]=\frac{\sigma_n^2}{N}}$$

Note: the variance of the ML-estimate decreases with the increasing length of the observation.

2. Properties of ML-Estimate

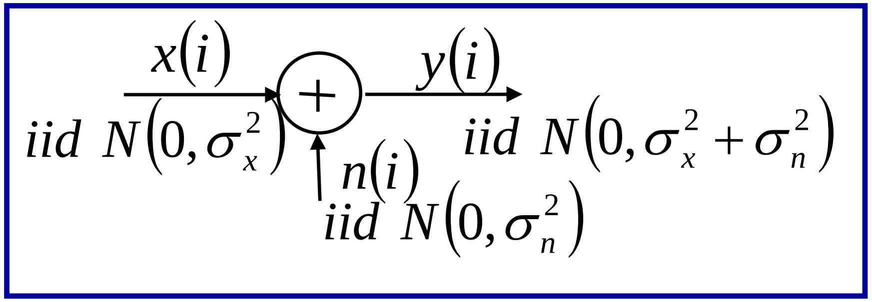
- Properties of ML- Estimate:

  (1) $\hat{\theta}_{ML}$ is asymptotically unbiased: $\lim_{N\to\infty} E\left[\hat{\theta}\right] = \theta$

  (2) $\hat{\theta}_{ML}$ is consistent in probability: $\hat{\theta} \to \theta$ in probability as $N \to \infty$

- Despite its attractive asymptotic properties, $\hat{\theta}_{ML}$ may not be the best estimator for finite N!
- In fact, there is no guarantee that $\hat{\theta}_{ML}$ is good at all for small N.

2. ML-Estimation of Variance



- Estimate σ_x^2 using ML-estimate.
- Likelihood function:

$$L(\sigma_x^2) = \prod_{i=0}^{N-1}\left(\frac{1}{\sqrt{2\pi(\sigma_x^2+\sigma_n^2)}} e^{-\frac{|y(i)|^2}{2(\sigma_x^2+\sigma_n^2)}}\right)$$

- Log-Likelihood function:

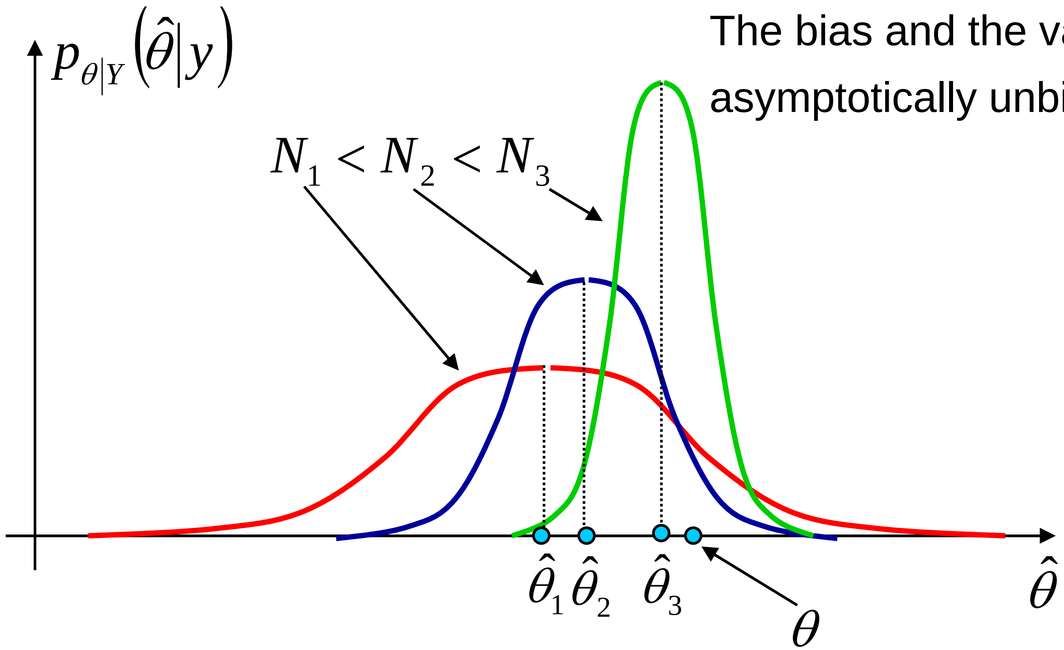
$$\ell(\sigma_x^2) = -\frac{N}{2}\ln(\sigma_x^2+\sigma_n^2) - \frac{1}{2(\sigma_x^2+\sigma_n^2)}\sum_{i=0}^{N-1}|y(i)|^2 - \frac{N}{2}\ln(2\pi)$$

- Setting derivative to zero yields:

$$0 = 2\frac{d\ell(\sigma_x^2)}{d\sigma_x^2} = \frac{N}{\sigma_x^2+\sigma_n^2} - \frac{1}{(\sigma_x^2+\sigma_n^2)^2}\sum_{i=0}^{N-1}|y(i)|^2$$

$$\Rightarrow \boxed{\sigma_x^2 = \max\left(0, \frac{1}{N}\sum_{i=0}^{N-1}|y(i)|^2 - \sigma_n^2\right)}$$

2. ML-Estimate: Desirable properties



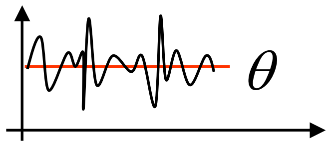
The bias and the variance of an asymptotically unbiased estimator.

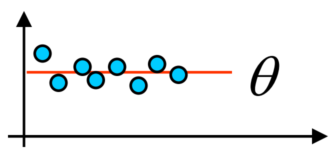
In general the bias and the variance of an estimate decrease with the increasing number of observation samples N.

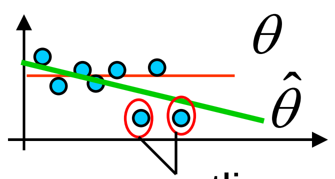
2. Robust M-Estimators

$$\boxed{y(i) = \theta + n(i),\ i = 0, .., N-1}$$

It is supposed to be constant on the observation interval







$$\boxed{\hat{\theta} = \arg\min_{\theta} \left\{ \sum_{i=0}^{N-1} \rho\left[y(i) - \theta \right] \right\}}$$

$$iid \ \ \rho(r) = -\ln p_{Y|\theta}(r)$$

$$\boxed{\nabla_{\theta}[.] = \sum_{i=0}^{N-1} \psi(r_i) \frac{\partial r_i}{\partial \theta} = 0}$$

$\psi(r) = d\rho(r)/dr$ is influence function

Let $w(r) = \dfrac{\psi(r_i)}{r}$ is weight function

2. Robust M-Estimators

$$\boxed{\sum_{i=0}^{N-1} w(r_i) r_i \frac{\partial r_i}{\partial \theta} = 0}$$ or it is equivalent to minimization of: $$\boxed{\min\left\{\sum_{i=0}^{N-1} w(r_i) r_i^{\,2}\right\}}$$

The above problem is solved using Reweighted Least-Squares (RLS) method

$$\boxed{\min\left\{\sum_{i=0}^{N-1} w\left(r_i^{(k-1)}\right) r_i^{\,2}\right\}}$$

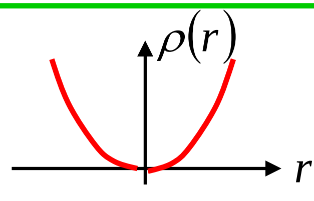
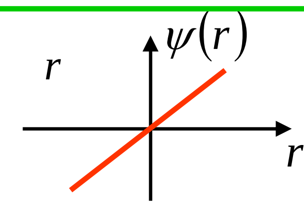
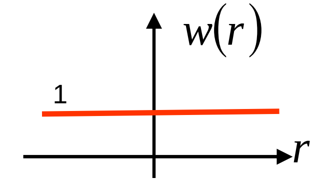
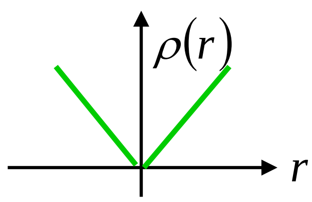
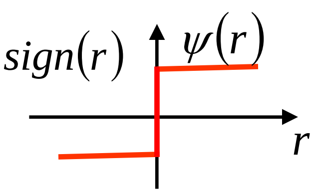
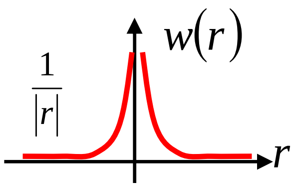
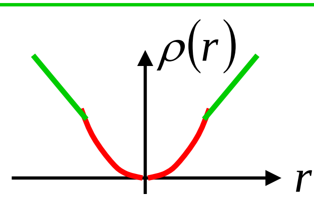
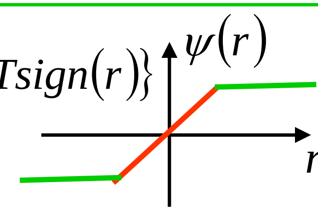
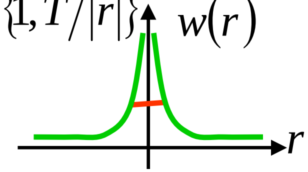
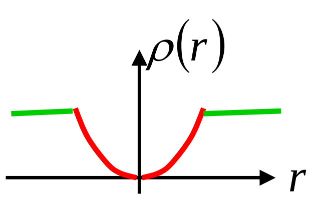
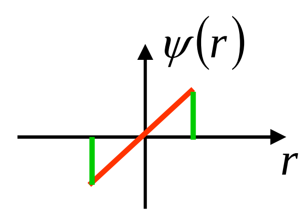
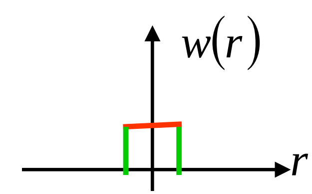
The weight $w\left(r_i^{(k-1)}\right)$ should be recomputed after each iteration in order to be used in the next iteration.

2. Robust M-Estimators

- The influence function measures the influence of a datum on the value of the parameter estimate.

- There are several constraints that a robust *M*-estimator should meet:
  - The first is of course to have a bounded influence function.
  - The second is naturally the requirement of the robust estimator to be unique. This implies that the objective function of parameter vector to be minimized should have a unique minimum.
    This requires that *the individual* ρ*-function is convex in variable* θ.
    This is necessary because only requiring a ρ-function to have a unique minimum is not sufficient.
    The convexity constraint is equivalent to imposing that $\frac{\partial^2 \rho(.)}{\partial \theta^2}$ is non-negative definite.
  - The third one is a practical requirement. Whenever $\frac{\partial^2 \rho(.)}{\partial \theta^2}$ is singular, the objective should have a gradient, $\frac{\partial \rho(.)}{\partial \theta} \neq 0$. This avoids having to search through the complete parameter space.

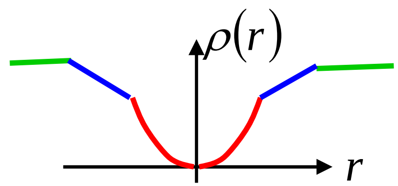
2. Robust M-Estimators

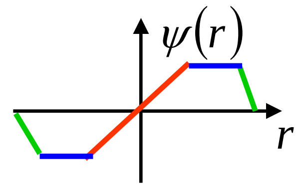
Type of noise	Penalty function	Influence function	Weight
Gaussian noise $L_2 \qquad \rho(r)=\frac{1}{2}r^2$			
Laplacian noise $L_1 \qquad \rho(r)=\lvert r\rvert$		$sign(r)$ 	$\frac{1}{\lvert r\rvert}$ 
ε - contaminated noise Huber $\rho(r)=\min\left\{\frac{r^2}{2},T\left(\lvert x\rvert-T/2\right)\right\}$		$\min\{r,Tsign(r)\}$ 	$\min\{1,T/\lvert r\rvert\}$ 
Talvar			

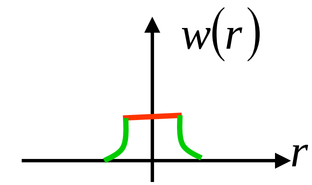
2. Robust M-Estimators

Type of noise | **Penalty function** | **Influence function** | **Weight**

Hampel



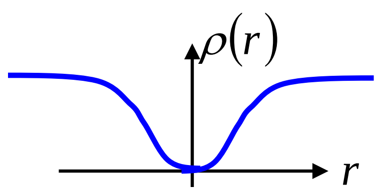


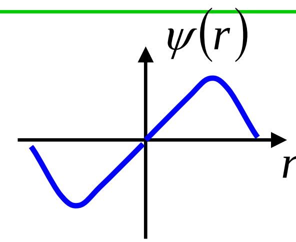


continuos case:

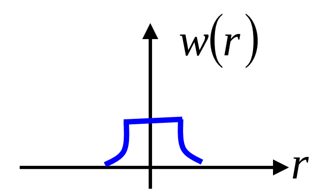
- Tukeys bi-weight
- Cauchy noise

$$\rho(r)=\frac{K^2}{2}\log\left(1+\left(\frac{r}{K}\right)^2\right)$$





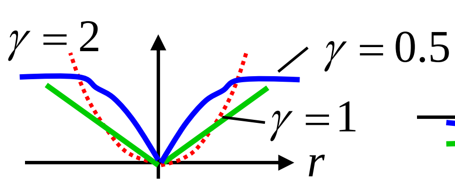
$$\frac{r}{1+(r/K)^2}$$

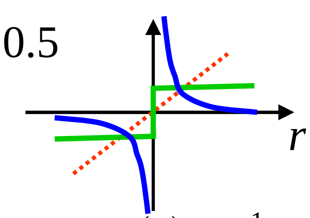


Generalized Gaussian

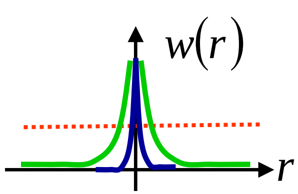
L_p

$$\rho(r)=|r|^{\gamma}/\gamma$$





$$sign(r)|r|^{\gamma-1}$$



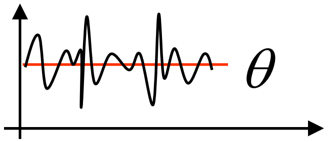
$$|r|^{\gamma-2}$$

2. Order Statistic Filters: L-Estimators

$$y(i)=\theta+n(i),\ i=0,..,N-1$$

It is supposed to be constant on the observation interval

ML-type estimation



$$\hat{\theta}=\arg\min_{\theta}\left\{\sum_{i=0}^{N-1}\rho\left[y(i)-\theta\right]\right\}$$

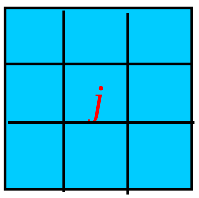
Difference with M-estimation is given by the constraint on the close form solution:

Ordered sequence

$$\hat{\theta}=\frac{\sum_{i=0}^{N-1}y'(i)w(i)}{\sum_{i=0}^{N-1}w(i)}$$

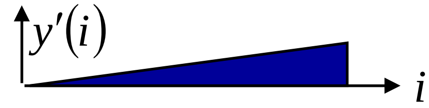
2-D case

$$\hat{\theta}_j=\frac{\sum_{i\in N_j}y'(i)w(i)}{\sum_{i\in N_j}w(i)}$$



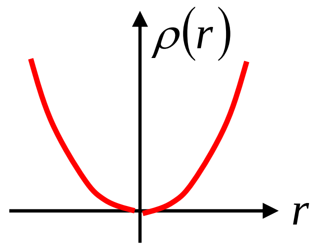
N_j

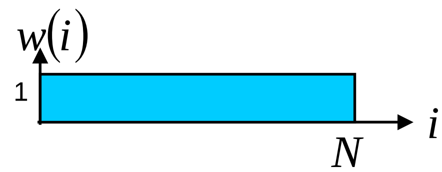
2. Order Statistic Filters: L-Estimators



Gaussian noise

$$\rho(r) = r^2$$

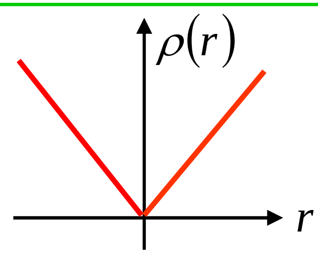


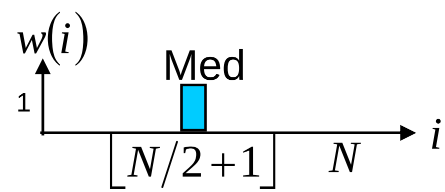


Sample mean

Laplacian noise

$$\rho(r) = |r|$$

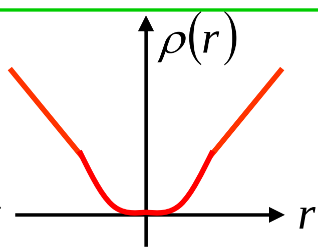


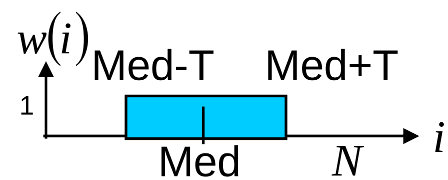


Sample median

ε - contaminated noise

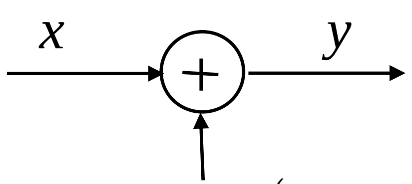
$$\rho(r) = \min\left\{\frac{r^2}{2}, T\left(|x| - T/2\right)\right\}$$





Sample trimmed mean

3. MAP: AWGN and stationary Gaussian prior



$$y = x + n$$

$n \sim$ **i.i.d.** $N\left(0, \sigma_n^2 I\right)$

$x \sim$ **i.i.d.** $N\left(\bar{x}, \sigma_x^2 I\right)$

$$\hat{x}_{MAP} = \arg\max_{\hat{x}\in\aleph}\left[\ln p_{Y|X}\left(y|x\right) + \ln p_X\left(x\right)\right] =$$

$$= \arg\max_{\hat{x}\in\aleph}\left[-\frac{1}{2\sigma_n^2}\left\|y - x\right\|_2^2 - \frac{1}{2\sigma_x^2}\left\|x - \bar{x}\right\|_2^2\right]$$

$$0 = \nabla_x[.] = \left[\frac{1}{\sigma_n^2}\left\|y - x\right\|_2 + \frac{1}{\sigma_x^2}\left\|x - \bar{x}\right\|_2\right]$$

Wiener (Lee) filter

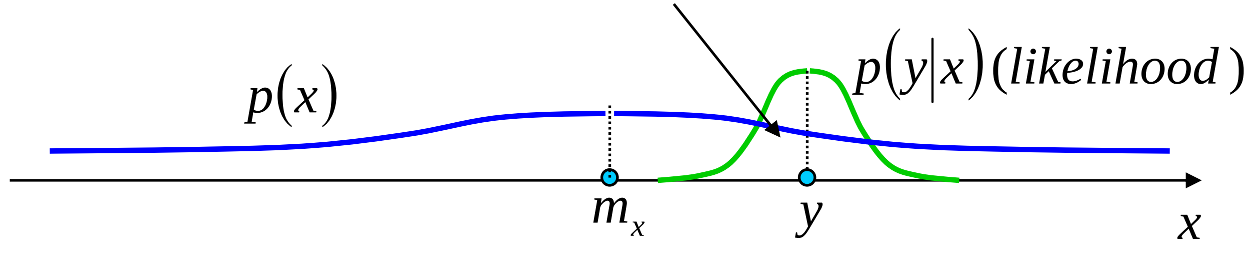
$$\boxed{\hat{x}_{MAP} = \left(\frac{1}{\sigma_n^2}I + \frac{1}{\sigma_x^2}I\right)^{-1}\left(\frac{1}{\sigma_n^2}y + \frac{1}{\sigma_x^2}\bar{x}\right) = \bar{x} + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_n^2}\left(y - \bar{x}\right)}$$

3. MAP-Estimate: Desirable properties

Consider the following two extreme cases.

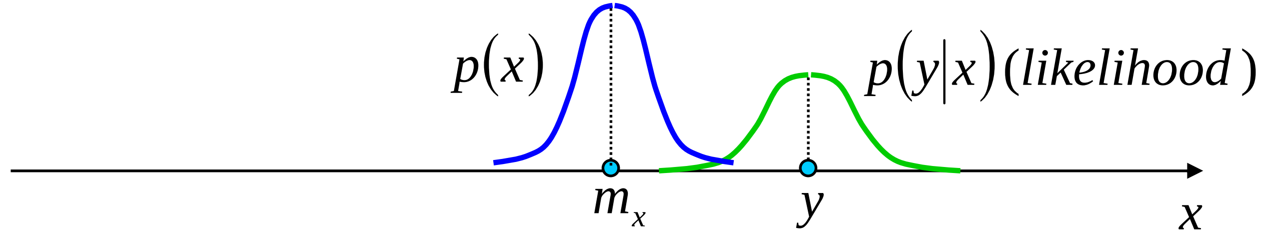
- As $\sigma_x^2 \to \infty$, we obtain $\hat{x}_{MAP} \to y$

In this case, the prior is noniformative (flat in region of support of likelihood function)

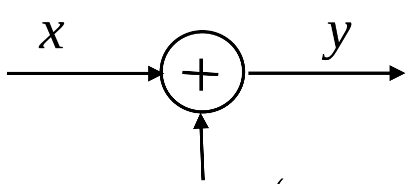


- As $\sigma_x^2 \to 0$, we obtain $\hat{x}_{MAP} \to m_x = \bar{x}$

In this case, the prior dominates:



3. MAP: AWGN and stationary SGG prior



$$n \sim \textbf{i.i.d. } N\left(0, \sigma_n^2 I\right)$$

$$x \sim \textbf{i.i.d. } sGG\left(\bar{x}, \sigma_x^2, \gamma\right)$$

$$\boxed{\begin{aligned}\hat{x}_{MAP} &= \arg\max_{\hat{x}\in\aleph}\left[\ln p_{Y|X}\left(y|x\right) + \ln p_X\left(x\right)\right] = \\ &= \arg\min_{\hat{x}\in\aleph}\left[\frac{1}{2\sigma_n^2}\left\|y-x\right\|_2^2 + \phi\left(\underbrace{x-\bar{x}}_{Cx}\right)\right]\end{aligned}}$$

High-pass

$$p_X\left(x\right) = \left(\frac{\gamma\eta\left(\gamma\right)}{2\Gamma\left(\frac{1}{\gamma}\right)}\right)\cdot\frac{1}{\sigma_n}\cdot\exp\left\{-\left[\eta\left(\gamma\right)\left|\frac{x}{\sigma_n}\right|\right]^{\gamma}\right\}$$

- $\gamma = 2$ Gaussian
- $\gamma = 1$ Laplacian
- $\gamma \to \infty$ Uniform

$$\eta\left(\gamma\right) = \sqrt{\frac{\Gamma\left(3/\gamma\right)}{\Gamma\left(1/\gamma\right)}}$$

$$p_X\left(x\right) = A\cdot\exp\left\{-\phi\left(x\right)\right\}$$

$$\boxed{\phi\left(x\right) = \left[\eta\left(\gamma\right)\left|\frac{x}{\sigma_n}\right|\right]^{\gamma}}$$

$$\ln p_X\left(x\right) = \ln A + \ln\exp\left\{-\phi\left(x\right)\right\} = \ln A - \phi\left(x\right)$$

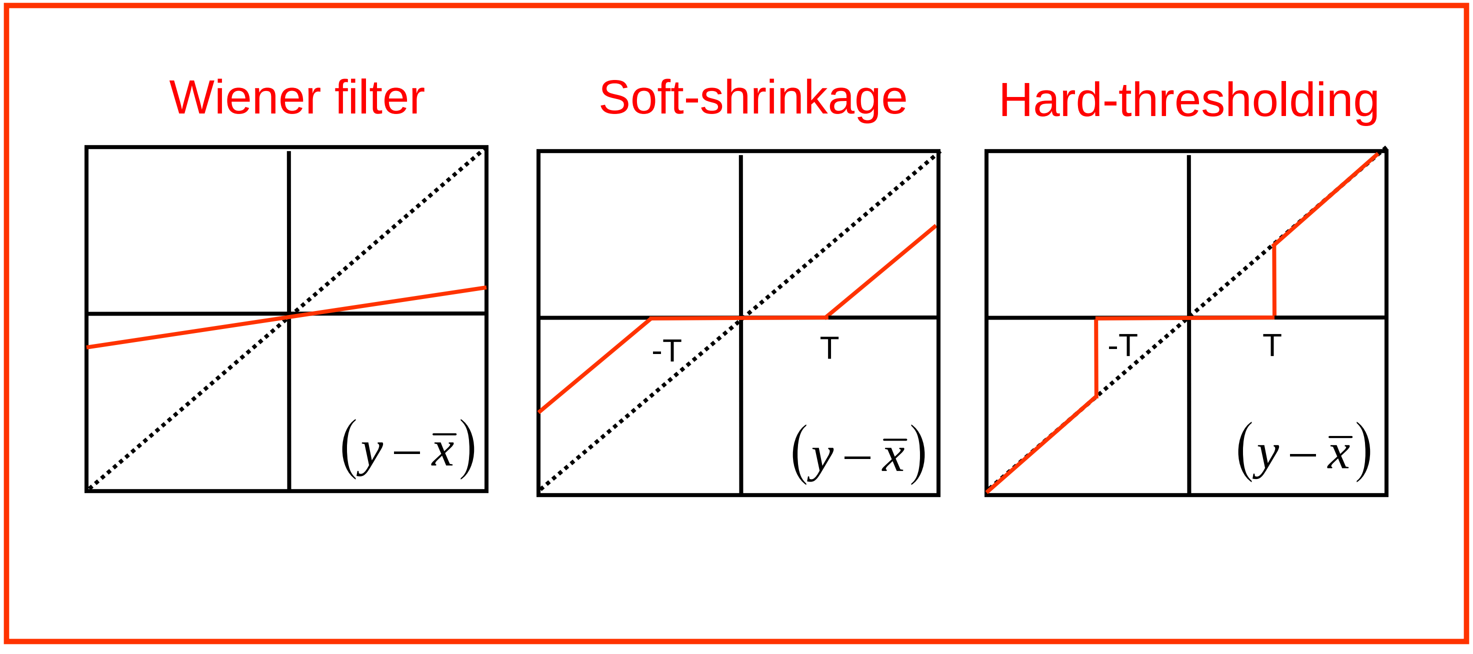
3. MAP: AWGN and stationary SGG prior

$$\hat{x}_{MAP} = \nabla_x[.] = \left[-\frac{1}{\sigma_n^2}\|y-x\|_2 + \phi'(x-\bar{x}) \right] = 0$$

$\gamma = 2$	$\hat{x}_{MAP} = \bar{x} + \frac{\sigma_x^2}{\sigma_x^2+\sigma_n^2}(y-\bar{x})$	Wiener filter
$\gamma = 1$	$\hat{x}_{MAP} = \bar{x} + \max(0, \lvert y-\bar{x}\rvert - T)\,sign(y-\bar{x})$ $T = \sqrt{2}\frac{\sigma_n^2}{\sigma_x}$	Soft-shrinkage
$\gamma \to 0$	$\hat{x}_{MAP} = \bar{x} + (y-\bar{x})1(\lvert y-\bar{x}\rvert > T)$ $T = \sigma_x\sqrt{6\sqrt{3}/e\gamma^{-1/2}}as \quad \gamma \to 0$	Hard-thresholding

3. MAP: AWGN and stationary SGG prior

Scaling functions of denoisers



3. AWGN: Comparison

Original | Noisy (Var=25) | ML 3x3 | ML 5x5

MAP: Laplacian (soft-shrinkage) | MAP: non-stationary Gaussian (Wiener)

4. Penalized Maximum-Likelihood (PML) Estimator

$$y(i) = x(i) + n(i),\ i = 0,..,N-1 \Rightarrow \hat{x} = \arg\min_{x \in \aleph}\{-\ln p(y|x) + \alpha\Phi(x)\}$$

Regularization parameter

$\Phi(x)$ penalty function (regularization):

- MAP as particular case:
  (exponential prior family)

$$p(x) \propto e^{-\alpha\Phi(x)}$$

sGG: $\ell_2 \rightarrow \|\cdot\|_2^2$, $\ell_1 \rightarrow |\cdot|$

GGMRF

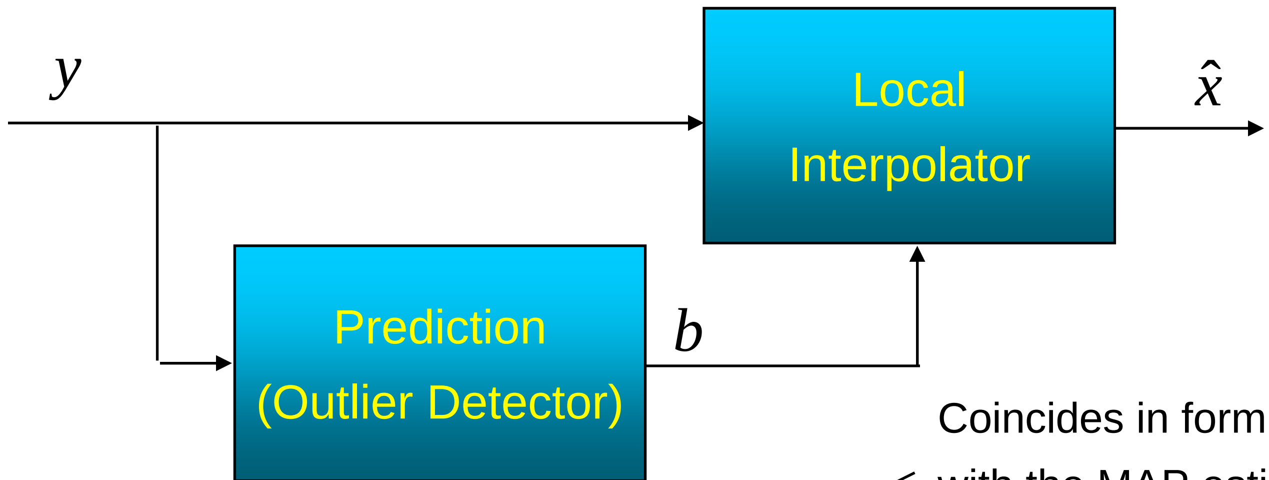
- maximum entropy:

$$-\Phi(x) = -\sum_{i=1}^{N} x(i)\log_2(x(i))$$

- maximum divergence:

$$-\Phi(x) = -\sum_{i=1}^{N} \log_2(x(i))$$

5. Impulse Noise Removal Using Prediction Models



Coincides in form of solution
with the MAP estimators for AWGN

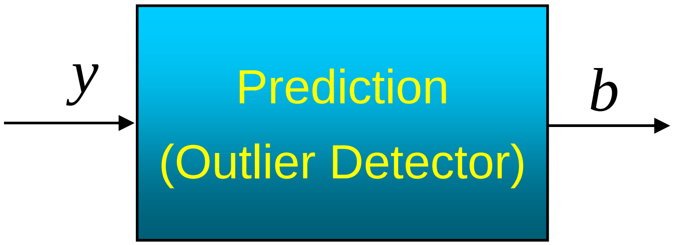
$$\boxed{\hat{x} = \bar{x} + b\left(y - \bar{x}\right)}$$

b - indicator function

$\bar{x}$ - local predicted value

$$b = \begin{cases} 0, & \textit{pixel is corrupted}, \\ 1, & \textit{pixel is not corrupted}. \end{cases}$$

5. Impulse Noise Removal Using Prediction Models

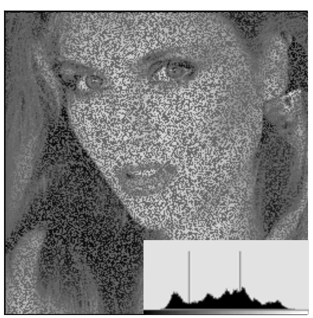


Prediction of outliers:

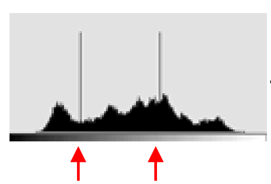
- Robust statistics

$$b=\begin{cases}1, & \left|y-med(y)\right|>T,\\ 0, & otherwise.\end{cases}$$

- Histogram detection



Histogram analysis



z -set of corrupted values

$$b=\begin{cases}0, & if\ \ y\in z,\\ 1, & otherwise.\end{cases}$$

5. Impulse Noise Removal Using Prediction Models

$$\bar{x} = E[y]$$

$$\bar{x}_i = \frac{\sum_{j \in N_i} y_j b_j}{\sum_{j \in N_i} b_j}$$

Modified mean (excludes outliers)

L-filters (median or trimmed mean)

5. Impulse Noise Removal Using Prediction Models

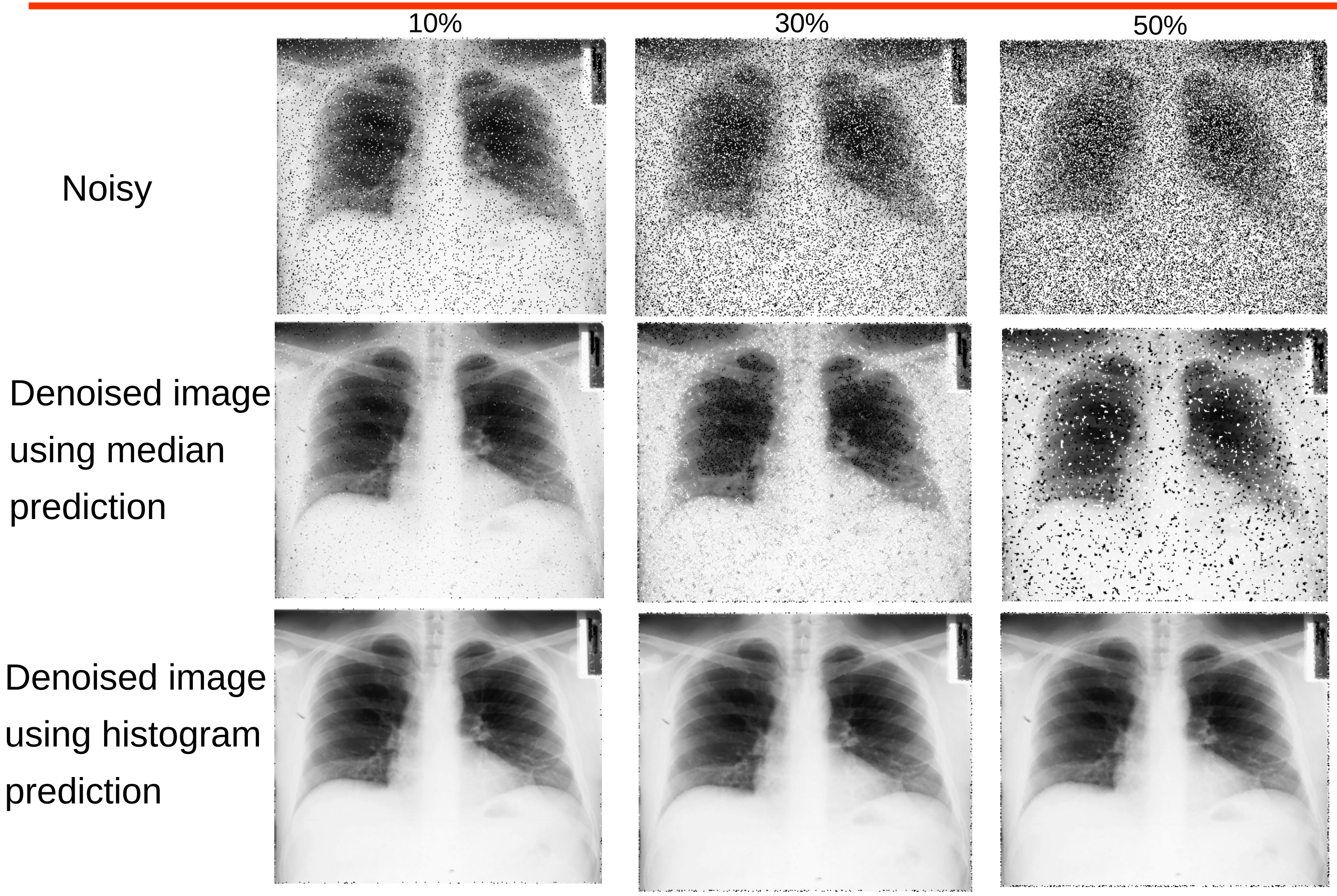

Original


Bernoulli (50%)


Median filter 3x3


Histogram detector
and local interpolator

5. Impulse Noise Removal Using Prediction Models



6. Speckle noise

$$\boxed{v(x,y) \cong |g(x,y)|^2 |h(x,y)|^2 + \eta(x,y) = u(x,y)s(x,y) + \eta(x,y)}$$

Multiplicative noise

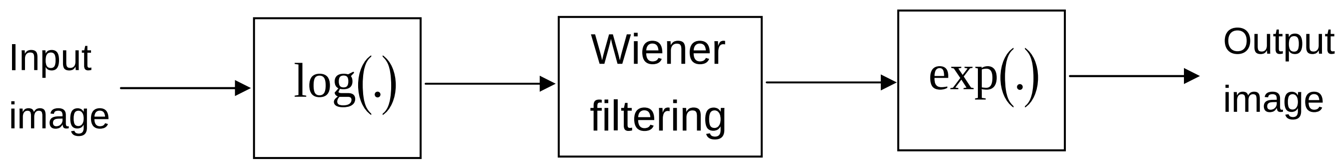
Image intensity

$$u(x,y) \equiv |g(x,y)|^2$$

$$h(x,y) = \iint H(x,y;x',y') e^{j\phi(x',y')} dx'dy'$$

6. Speckle Reduction

Homomorphic filtering



Multiplicative noise is transformed into additive noise:

$$\log v(x,y)=\log u(x,y)+\log s(x,y) \quad \Longrightarrow \quad w(x,y)=z(x,y)+\xi(x,y)$$

Stationary white noise

Wiener filter is used to filter noise