

Anisotropic Local Likelihood Approximations: Theory, Algorithms, Applications

Vladimir Katkovnik, Alessandro Foi, Karen Egiazarian, Jaakko Astola
Institute of Signal Processing, Tampere University of Technology, FIN-33101, Tampere,
Finland.

ABSTRACT

We consider a signal restoration from observations corrupted by random noise. The local maximum likelihood technique allows to deal with quite general statistical models of signal dependent observations, relaxes the standard parametric modelling of the standard maximum likelihood, and results in flexible nonparametric regression estimation of the signal. We deal with the anisotropy of the signal using multi-window directional sectorial local polynomial approximation. The data-driven sizes of the sectorial windows, obtained by the intersection of confidence interval (*ICI*) algorithm, allow to form starshaped adaptive neighborhoods used for the pointwise estimation. The developed approach is quite general and is applicable for multivariable data. A fast adaptive algorithm implementation is proposed. It is applied for photon-limited imaging with the Poisson distribution of data. Simulation experiments and comparison with some of the best results in the field demonstrate an advanced performance of the developed algorithms.

Keywords: nonparametric regression, local maximum likelihood, adaptive window size, adaptive scale, photon-limited imaging, Poisson regression.

1. INTRODUCTION

Although continuous models of the discrete image intensity are widely used in image processing, the corresponding estimates are too rough in order to be useful for those applications where the sharpness and details are of first priority. For discrete images lacking global differentiability or continuity the only reliable way to obtain an accurate restoration is to process y locally.

The nonparametric local regression originated in mathematical statistics offers an original approach to signal processing problems (e.g.^{1,9}). It basically results in linear filtering with the linear filters designed using some moving window local approximations.⁴⁻⁶ Adaptive versions of these algorithms are able to produce efficient filtering with the varying window size (scale, bandwidth) which is pointwise adaptive (see⁸ and references therein). This pointwise adaptive scale selection is based on the following idea known as Lepski's approach. The algorithm searches for a largest local vicinity of the point of estimation where the estimate fits well to the data. The estimates $\hat{y}_h(x)$ are calculated for a set of window sizes (scales) $h \in H$ and compared. The adaptive scale is defined as the largest of those windows in the grid which estimate does not differ significantly from the estimators corresponding to the smaller window sizes. The intersection of confidence intervals (*ICI*) rule³ being one of the versions of this approach has appeared to be quite efficient for the adaptive scale image restoration.⁴⁻⁶

Cited above papers on the adaptive scale kernel estimation concern a scalar scale parameter and assume that the estimators can be ordered by their variances. Vector scale parameter kernels in d -dimensional space, $x \in \mathbb{R}^d$, $h \in \mathbb{R}^+{}^d$, are of special interest for anisotropic functions with highly varying properties at different directions. Imaging is one of the typical examples of such problems. A direct generalization of the Lepski's approach to adaptive smoothing with a vector scale parameter $h \in \mathbb{R}^+{}^d$ faces a principal obstacle as the variance of the estimates calculated for different h cannot be ordered and can be only semi-ordered as there could be many estimators with the same or similar variance.

The main intention of the new approach to anisotropic estimation introduced recently in⁷ is to obtain in a data-driven way a largest local vicinity of the estimation point in which the underlying model fits the data. It is assumed that this vicinity is a starshaped set, which can be approximated by some sectorial segmentation with, say, K sectors. These estimators are equipped with univariate scale parameters defining the size of the supports in the sector. The *ICI* rule is exploited K times, once for each sector, in order to find the optimal pointwise

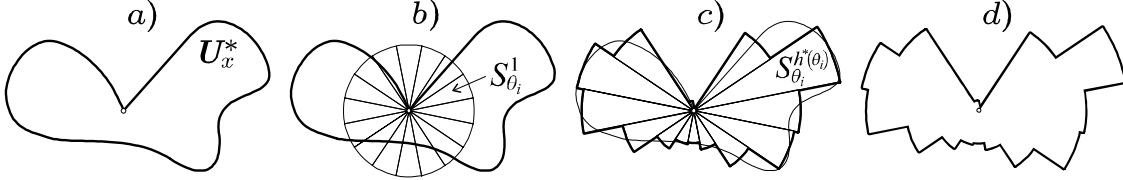


Figure 1. A neighborhood of the estimation point x : a) the best starshaped estimation set U_x^* , b) the unit ball segmentation, c) sectors with adaptive lengths, d) sectorial approximation of U_x^* .

adaptive scales for each sector's estimates which are then combined into the final one. In this way, we reduce the d -dimensional scale selection problem to a multiple univariate one. Figure 1 illustrates this concept and shows sequentially: a local best estimation neighborhood U^* , a sectorial segmentation of the unit ball, and the sectorial approximation of U^* using the adaptive scales $h^*(\theta_i)$ defining the length of the corresponding sectors. Varying size sectors enable one to get a good approximation of any neighborhood of the estimation point x provided that it is a starshaped body.

All previously obtained techniques^{2,5-7} based on the above strategy concerned the Gaussian observation model and linear estimators.

The main contribution of this paper concerns two problems. First, we generalize the directional anisotropic image processing to a broad class of non-Gaussian observation models and nonlinear filters based on the local maximum likelihood methods. Second, the local polynomial approximation technique and the *ICI* rule are modified for these nonlinear estimation problems. As an example of these developments, a novel adaptive image restoration algorithm for Poissonian data is presented.

2. LOCAL LIKELIHOOD

2.1. Basic concept

The observations given by pairs $\{z_s, X_s\}$, where $X_s \in \mathbb{R}^d$ form the regular or irregular discrete d -dimensional grids and z is random with a conditional probability density function (*pdf*) $f(z, y(x))$. Here y is a parameter of this *pdf* depending on x . In this model the random z_s depends on x through $y(x)$ and corresponds to $x = X_s$. For the Gaussian *pdf* and $y(x)$ being the expectation of z we arrive to the standard additive observation model with an invariant standard deviation of the noise

$$z_s = y(X) + \varepsilon_s, \quad s = 1, \dots, n, \quad (1)$$

where $\varepsilon_s \sim \mathcal{N}(0, \sigma)$. Depending on the *pdf* and the parameter y the considered models may allow or may not allow this additive noise representation. Even if this representation is valid the noise ε , in general, is signal dependent. As examples of this sort of imaging problems, we wish to mention the impulse noise observations in radar and remote sensing applications as well as the Poisson random observations typical for many medical imaging problems. For independent (w.r.t. s) observations $\{z_s\}$ with fixed X_s , the conventional log-likelihood is defined as

$$L = \sum_s \log f(z_s, y(X_s)).$$

Various models and methods are used for $y(x)$ as a function of x in order to obtain a reasonable reconstruction of this function even provided that we have only single observation for each X_s .

The standard likelihood approach starts from a global parametrical model $y_C(x)$ with the estimation of the parameters C according to the maximum likelihood (*ML*) method

$$\hat{C} = \arg \max_C (L(C)), \quad L(C) = \sum_s \log f(z_s, y_C(X_s)).$$

Then $y(x)$ can be estimated, for any x , as $y_{\hat{C}}(x)$. The *local maximum likelihood (LML)* is a nonparametric counterpart of this widely used parametric *ML* technique. It extends the scope of the parametric *ML* to a much

wider class of functions. The local likelihood model does no longer assume a global parametric form of $y_C(x)$, but fits this model locally within a window. The local likelihood and local log-likelihood use the window function w for localization of the likelihood function in a neighborhood of the point of the interest x and replace the global parametric model $y_C(x)$ by a local one.

Let us use linear on C models for y in the log-likelihoods $L(C)$. These linear models are known in statistics as the *generalized linear models*. The standard linear model for y assumes a quadratic loss function and results in estimates which are linear with respect to observations. ‘‘Generalized’’ means that the linear model is exploited for the general type of distributions and for not necessarily quadratic loss functions. The generalized linear models result in estimates which are nonlinear with respect to the observations.

For the polynomial linear model this local log-likelihood can be introduced in the form,^{1,9}

$$L_h(C) = \sum_s \log f(z_s, y_h(x, X_s)) w_h(x - X_s), \quad (2)$$

$$y_h(x, v) = C^T \phi_h(x - v), \quad \phi_h(x) = \phi(x/h), \quad w_h(x) = w(x/h)/h^d, \quad x \in \mathbb{R}^d \quad (3)$$

where $\phi(x)$ is a vector of the polynomials, $w(x)$ is the window function and h stays for a scalar scale (window size) parameter.

A function $y_h(x, v)$ of the form (3) is called a local polynomial approximation (*LPA*) of y at the point x . This x is termed *centre* of the *LPA*, since the polynomials ϕ_h are centered in x and the window w_h has x as its reference point. The scale h is used both in the window function $w(x)$, $x \in \mathbb{R}^d$, and for the scaling of the polynomials $\phi(x)$, $x \in \mathbb{R}^d$.

What distinguish the localized likelihood $L_h(C)$ from the original non-local likelihood $L(C)$ is that the latter uses a global parametric model assumed to be valid for all x . The local likelihood relaxes this assumption and assumes instead that the used linear model is valid only locally, in some neighborhood of the estimation point. According to the idea of the *LPA*, the function estimate \hat{y}_h for the point x is calculated as follows:

$$\hat{y}_h = \hat{y}_h(x, v)|_{x=v}, \quad \hat{y}_h(x, v) = \hat{C}^T \phi_h(x - v), \quad \hat{C} = \arg \max_C (L_h(C)). \quad (4)$$

It means that the parameter C is selected by maximization of the log likelihood and the corresponding function estimate is obtained using the local *ML* fit $\hat{y}_h(x, v)$ only for $x = v$.

It follows immediately that

$$\hat{y}_h = \hat{C}^T \phi_h(0). \quad (5)$$

It is well known that linear estimates can degrade significantly if the random observations obey a non-Gaussian distribution. For a non-Gaussian noise the linear *LPA* should be replaced by relevant nonlinear operations.

The *local maximum likelihood*^{1,9,16} approach is introduced as a generalization of the linear *LPA*. It provides flexible tools for design of the nonlinear methods and algorithms for variety of stochastic models different from standard Gaussian. This approach combines two different ideas: generalized linear model and localization of parametric fitting.

In the generalized linear models the parameters of the distributions of random observations are assumed to be varying and the linear regression is used to fit these varying parameters. The *ML* approach is used to estimate these parameters.¹⁵

The localization is introduced in order to relax restrictions imposed by parametric modeling in the standard *ML*. In what follows we restrict our attention to the *2D* case, where the directionality can be presented in a very transparent way. The generalization to a higher number of dimensions is straightforward.

2.2. Polynomials

Let us introduce the *2D* polynomials in the form $x_1^{k_1} x_2^{k_2} / k_1! k_2!$, $k_1 = 0, \dots, m_1$, $k_2 = 0, \dots, m_2$, where the powers m_1 and m_2 can be different. These polynomials are linearly independent with the total number of the polynomials $(m_1 + 1)(m_2 + 1)$ and the maximum power equal to $(m_1 + m_2)$. We will use the power m as a multi-index $m = [m_1, m_2]$ with elements showing the powers on the corresponding variables.

The *LPA* estimates are defined according to the model (3)-(4) where the vector $\phi(x)$ is composed from the above polynomials, and (3) with a given multi-index m explicitly defines the *LPA* model of y .

2.3. Windows

The finite support window w in (2) is a key tool in design of the directional estimate. We call the window *directional* if it is narrow and oriented along the axis and pointed to one of the directions along this axis. The sectors in Figure 1 are clear examples of such directional supports.

We consider a rotated sectorial support window w_θ with the angle θ as the main direction. The axis u_1, u_2 are bound with the rotated sector. The axis u_1 points in the direction θ , and u_2 is orthogonal to u_1 . The variable $x = (x_1, x_2)$ and $u = (u_1, u_2)$ are linked by the rotation transform equation

$$u = U(\theta)x, \quad x = U^T(\theta)u, \quad U(\theta) = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}. \quad (6)$$

The rotated window is defined as $w_\theta(x) = w(U(\theta)x)$. The scaled rotated window is $w_{\theta,h}(x) = w(U(\theta)x/h)/h^2$.

The above polynomials are used for the *LPA* of the powers m_1 and m_2 with respect variables u_1 and u_2 . Then, the directional estimates based on the observation in the sector θ are defined from the directional local log-likelihood

$$L_{h,\theta}(C) = \sum_s \log f(z_s, y_{h,\theta}(x, X_s))w_{h,\theta}(x - X_s), \quad (7)$$

$$y_{h,\theta}(x, v) = C^T \phi_h(U(\theta)(x - v)), \quad w_{h,\theta}(x) = w_h(U(\theta)x)$$

and

$$\hat{y}_{h,\theta} = \hat{C}^T \phi_h(0), \quad \hat{C} = \arg \max_C (L_{h,\theta}(C)).$$

The window w has to satisfy to the standard *LPA* assumptions. In particular, the origin $(0, 0)$ is always corresponds to the maximum value of w . The directionality of the estimate is guaranteed by the support and shape of the window function. The only requirement restricting the window and the scale selection is that the likelihood $L_{h,\theta}(C)$ is convex on C and the maximization of $L_{h,\theta}(C)$ on C has a unique solution. The window and the scale are called admissible if this condition holds. When we use the zero order model, $m = [0, 0]$, any nonempty window $w_{h,\theta}$ is admissible.

3. PARTICULAR DISTRIBUTIONS

The models of observations considered further on allow to cover a lot of practical scenarios where the noise variance is signal dependent. In more general terms these models allow to take into consideration even a more delicate case when the *distribution* is signal dependent.

In order to simply the notation we assume $\theta = 0$ and drop the index θ in the estimates and in the distributions. In order to obtain the estimate for the direction θ the rotation transform $U(\theta)$ should be used as it is (7).

3.1. Additive i.i.d. noise

First of all, the observation model (1) is changed from global $y_C(x)$ to the local one:

$$z_s = y_h(x, X_s) + \varepsilon_s, \quad y_h(x, X_s) = C^T \phi_h(x - X_s), \quad s = 1, \dots, n, \quad (8)$$

where the random ε_s are i.i.d. with the probability density f .

The deterministic part $y_h(x, X_s)$ of the model (8) is local depending on x . It is a local model valid for a single fixed x . Remind that in the *LPA* the argument x is treated as a fixed parameter of the model. Further, the local observation model depends on the scale parameter h .

The probability density of z is $f(z - y_C)$ and for the set of observations we have the local log-likelihood (2) in the forms:

$$L_h(C) = \sum_s \ln(f(e_s))w_h(x - X_s), \quad e_s = z_s - y_h(x, X_s).$$

Denote $\psi(v) = -\ln f(v)$. Then, $\sum_s \ln(f(e_s)) = -\sum_s \psi(e_s)$, and the local log-likelihood estimates are found from the problem (2) as

$$\hat{C} = \arg \min_{C \in \mathbb{R}^M} \sum_s \psi(e_s) w_h(x - X_s), \quad e_s = z_s - y_h(x, X_s).$$

where ψ is a loss function of the residuals.

In this way we derive the local nonparametric *LPA* robust *M*-estimates with the loss function ψ defined through the probability density f .¹¹ This link with the *ML* guarantees for the *M*-estimates all good properties of the *ML* provided that the loss function and the noise probability density agree.

3.2. Gaussian noise with a signal dependent variance

Let the observation noise in (8) be Gaussian with the variance depending on y , $\varepsilon \sim \mathcal{N}(0, \sigma^2(y))$. Omitting the invariant part $\ln(\sqrt{2\pi}\sigma(y))$ in $\ln f = -\ln(\sqrt{2\pi}\sigma(y)) - v^2/(2\sigma^2(y))$, we can define the loss function ψ as $\psi(v) = v^2/\sigma^2(y)$ with the weighted mean squared residual criterion

$$\sum_s e_s^2/\sigma^2(y_h(x, X_s)) w_h(x - X_s), \quad e_s = z_s - y_h(x, X_s). \quad (9)$$

and the weighted *LSM* and the *ML* estimate in the form

$$\hat{C} = \arg \min_{C \in \mathbb{R}^M} \sum_s e_s^2/\sigma^2(y_h(x, X_s)) w_h(x - X_s). \quad (10)$$

Note that $y_h(x, X_s)$ as the estimate of the signal is used for calculation of the variance $\sigma^2(y)$.

3.3. Gaussian noise with constant variance

If σ^2 is constant, it gives the standard linear *LPA* estimate

$$\hat{C} = \arg \min_{C \in \mathbb{R}^M} \sum_s e_s^2 w_h(x - X_s). \quad (11)$$

In this case, the analytical solution can be found as

$$\hat{C} = \Phi_h^{-1} \sum_s w_h(x - X_s) \phi_h(x - X_s) z_s, \quad \Phi_h = \sum_s w_h(x - X_s) \phi_h(x - X_s) \phi_h^T(x - X_s). \quad (12)$$

Inserting \hat{C} in (5), we obtain the estimate in the kernel form

$$\hat{y}_h = \sum_s g_h(x - X_s) z_s, \quad g_h(x) = w_h(x) \phi_h^T(0) \Phi_h^{-1} \phi_h(x). \quad (13)$$

For the unrestricted regular grid of the observation points X_s the matrix Φ_h is constant and the kernel $g_h(x)$ is shift invariant. Then the estimate (13) can be calculated as the convolution

$$\hat{y}_h(x) = (g_h \otimes z_s)(x), \quad x \in X. \quad (14)$$

For zero order *LPA*, with $m = 0$, the local estimate is a constant, $y_h(x, X_s) = C_1$, and

$$g_h(x) = w_h(x) / \sum_s w_h(X_s). \quad (15)$$

Then, the estimate (13) is the weighted sample mean known as the Nadaraya-Watson nonparametric regression estimate.^{1,9}

3.4. Laplacian noise

Let the observation noise in the additive model (8) be independent Laplacian, i.e. $f(v) = \frac{1}{2a} \exp(-|v|)$. Then $\ln f(v) = \ln(1/(2a)) - |v|$ and the loss function ψ can be defined as $\psi(v) = |v|$ with the absolute value residual criterion

$$\sum_s w_h(x - X_s) |e_s|, \quad e_s = z_s - y_h(x, X_s). \quad (16)$$

In this way we derive the *M*-estimates with the loss function $\psi(x) = |x|$ as the *ML* estimate. The absolute error loss function ψ means that the estimate is the weighted median.¹²

3.5. Poisson observations

In many imaging systems the recorded observations have the physical meaning of numbers of detected photons. The photons are counted at different spatial locations and in this way form an image of an object. The Poisson distribution is a conventional probabilistic model used for a random number of photons detected during an exposure time. When the number of photons is relatively small and the level of the noise is high the problem is referred as the *photon-limited* signal processing .

For a fixed x the observed random z has the discrete Poisson distribution $z \sim \mathcal{P}(y)$:

$$P\{z = k\} = e^{-y} y^k / k!, \quad k = 0, 1, \dots, \quad E\{z\} = y, \quad var\{z\} = y, \quad (17)$$

where y stays for the intensity or the mean value of the Poisson distribution. It is emphasized that y is non-negative, being the mean value of the integer counts.

For the model where the argument x is essential $E\{z(x)\} = y(x)$, $var\{z\} = y(x)$, we can see that the variance of the observation is signal dependent. The problem is to reconstruct the function $y(x)$ from the observations $z_s = z(X_s)$.

The random variable z is nonnegative and integer. The corresponding likelihood for the observations $\{z_s, X_s\}$ is defined as $l = \prod_s e^{-y(X_s)} y^{z_s} / z_s!$, with the log-likelihood

$$L = \ln(l) = \sum_s [-y(X_s) + z_s \ln(y(X_s)) - \ln(z_s!)]. \quad (18)$$

The local version (2) of this log-likelihood is

$$L_h(x, C) = \sum_s w_h(x - X_s) [-y_h(x, X_s) + z_s \ln(y_h(x, X_s)) - \ln(z_s!)], \quad (19)$$

$$\hat{y}_h(x, v) = \hat{C}^T \phi_h(x - v), \quad \hat{y}_h(x) = \hat{C}^T \phi_h(0).$$

3.5.1. Zero-order model

The zero order model ($m = 0$) is of special interest since the local estimate $y_h(x, X_s)$ is constant and

$$L_h(x, C) = \sum_s (-C + z_s \ln C) w_h(x - X_s),$$

$$\hat{C} = \arg(\max_C (L_h(x, C))), \quad \hat{y}_h(x) = \hat{C}.$$

The equation $\partial_C L_h(x, C) = 0$ gives as a solution

$$\hat{y}_h(x) = \sum_s g_h(x - X_s) z_s, \quad g_h(x) = w_h(x) / \sum_s w_h(X_s). \quad (20)$$

Thus, we arrive again to the Nadaraya-Watson linear nonparametric estimate.

3.5.2. Higher-order model

For the higher order model there are some principal problems. The parameters C in (19) found by maximizing the log-likelihood

$$\hat{C} = \arg(\max_C L_h(x, C)), \quad \hat{y}_h(x) = \hat{C}^T \phi_h(0), \quad (21)$$

gives the estimate $\hat{y}_h(x)$ which can be negative. It is in a contradiction with the meaning of $y \geq 0$.

There are two ways how to overcome this problem. In statistics, the canonical variable $\eta = \log y$ is recommended instead of y .^{1,9} The *LPA* is exploited for this new variable η and the estimate of y is found by inversion of the log function as $\hat{y}_h(x) = \exp(\hat{\eta}_h)$, where $\hat{\eta}_h$ is the local *ML* estimate of η . Then this estimate is always non-negative $\hat{y}_h(x) \geq 0$.

Another way, more in tradition with the engineering signal processing, is to project the estimate (21) on the half-line $[0, \infty)$.

4. ADAPTIVE SCALE SELECTION

The properties of the introduced nonlinear nonparametric regression estimates are quite similar to their well known linear counterparts.⁴⁻⁶ We do not go into details in this paper, however, general features of this similarity will be noted.

Firstly, if the estimated $y(x)$ is polynomials of the powers m_1 and m_2 used in the *LPA* the *ML* estimates are asymptotically unbiased. This asymptotic mainly assumes that the number of observations $n \rightarrow \infty$ and the scale parameter $h \rightarrow 0$ in appropriate way (details can be seen in^{1,4,5}). It ensures the convergence and the polynomial smoothness of the estimates.

Secondly, the bias and the variance of the estimate have asymptotically the form ch^a and dh^{-b} , respectively. The bias is small for small h and large for large h . The order of the bias $\mathcal{O}(h^a)$ depends on the smoothness of the estimated regression and the order of the *LPA*. The variance is large for small scales and small for large ones. We obtain for the mean squared error of estimation the expression $l(h) = c^2h^{a^2} + dh^{-b}$. This mean squared error is concave on h and the optimal scale can be found from the equation $\partial_h l(h) = 0$. This optimal scale gives the optimal variance-bias trade-off typical for the nonparametric estimation.

The found optimal value of h is called ideal as its calculation requires an accurate knowledge of derivatives of $y(x)$. This information is not available in any practical scenario.

A similarity between this sort of accuracy results for general *ML* estimates and the results known for the linear estimates⁴⁻⁶ allows to conclude that the intersection of confidence intervals (*ICI*)^{3-6,10} rule is applicable for the *ML* estimates and it gives the adaptive scale estimates close to the ideal ones.

Let us remind the *ICI* rule technique.

Given a finite set of the ordered scales $H = \{h_1 < h_2 < \dots < h_J\}$ and their corresponding varying scale kernel estimates $\{\hat{y}_{h_j}(x)\}_{j=1}^J$, with decreasing standard deviations $\sigma_{\hat{y}_{h_1}} > \dots > \sigma_{\hat{y}_{h_J}}$, we determine a sequence of confidence intervals

$$\mathcal{D}_j = \left[\hat{y}_{h_j}(x) - \Gamma \sigma_{\hat{y}_{h_j}}, \hat{y}_{h_j}(x) + \Gamma \sigma_{\hat{y}_{h_j}} \right], \quad (22)$$

where $\Gamma > 0$ is a threshold parameter. The *ICI* rule can be stated as follows:

Consider the intersection of confidence intervals $\mathcal{I}_j = \bigcap_{i=1}^j \mathcal{D}_i$ and let j^+ be the largest of the indexes j for which \mathcal{I}_j is non-empty, $\mathcal{I}_{j^+} \neq \emptyset$ and $\mathcal{I}_{j^++1} = \emptyset$. The optimal scale h^+ is defined as $h^+ = h_{j^+}$ and the optimal scale kernel estimate is $\hat{y}_{h^+}(x)$.

This is a procedure for a fixed x . It is produced for all $x \in X$ and in this way we obtain varying adaptive scale $h^+(x)$. The *ICI* procedure requires to know only the estimates for different scales as well as the corresponding variances of these estimates.

5. ANISOTROPIC ESTIMATOR BASED ON DIRECTIONAL OPTIMIZATIONS

The basic strategy to deal with the anisotropic y is follows. For each x and each θ_i , we calculate a collection $\{\hat{y}_{h,\theta_i}(x)\}_{h \in H}$ of varying-scale sectorial directional estimates, $\hat{y}_{h,\theta_i} = g_{h,\theta_i} \otimes z$. The *ICI* is used in order to find the best scale for each direction θ_i and each point x . It gives the pointwise adaptive-scale directional estimates $\hat{y}_{\theta_i}^+(x) = \hat{y}_{h^+,\theta_i}(x)$. With K sectors θ_i we have K adaptive-scale estimates $\hat{y}_{\theta_i}^+(x)$ for each x . These many adaptive estimates have to be fused (combined) into the final one.

We use as the final estimate the convex combination of the adaptive directional estimates $\hat{y}_{\theta_i}^+$ with their inverse variances as the weights:

$$\hat{y}(x) = \sum_{i=1}^K \lambda(\theta_i) \hat{y}_{\theta_i}^+(x), \quad \lambda(\theta_i) = \sigma_{\theta_i}^{-2}(x) / \sum_{j=1}^K \sigma_{\theta_j}^{-2}(x), \quad (23)$$

where $\sigma_{\theta_i}^2(x) = \sigma_{\hat{y}_{h^+,\theta_i}}^2$ is the variance of the adaptive estimate, and

$$\sigma_{\hat{y}_{h,\theta_i}}^2(x) = \sum_s g_{h,\theta_i}^2(x - X_s) \sigma_{z_s}^2, \quad g_{h,\theta_i}(x) = w_{h,\theta_i}(x) / \sum_s w_{h,\theta_i}(X_s). \quad (24)$$

Note that the weights $\lambda(\theta_i)$ in (23) are data-driven adaptive as $\sigma_{\theta_i}^{-2}(x)$ depend on the adaptive point-wise h^+ .

6. RECURSIVE *LPA-ICI* ADAPTIVE VARIANCE (*AV*) ALGORITHM

This algorithm is developed for Poisson-distributed data and it assumes that the observation variance $\sigma_z^2(x) = E\{z(x)\} = y(x)$. Nevertheless, by specifying another dependence of the variance σ_z^2 on the signal y , the algorithm can be used also for many other signal dependent variance problems.

6.1. *LPA-ICI* adaptive variance filter

For each step the *LPA-ICI* filter is multidirectional with narrow conical supports of the estimation kernels. The linear *LPA* estimates of the zero-order $m = 0$ are calculated for all scales H and for all K directions:

$$\hat{y}_{h,\theta_i}(x) = (g_{h,\theta_i} \otimes z)(x), \quad i = 1, \dots, K, \quad h \in H, \quad x = X. \quad (25)$$

The variance of these estimates can be calculated according to the formula

$$\hat{\sigma}_{\hat{y}_{h,\theta_i}}^2(x) = \sum_s \hat{\sigma}_{z(X_s)}^2 g_{h,\theta_i}^2(x - X_s), \quad \hat{\sigma}_{z(X_s)}^2 = \hat{y}(X_s), \quad (26)$$

and applied in the *ICI*. Here $\hat{y}(X_s)$ is some estimate of the signal.

6.2. Steps of the recursive *LPA-ICI* adaptive variance algorithm

The algorithm steps are different by: input data, methods used for calculation of the variances for these input data and the parameters of the algorithms.

Step 1: initialization

The input data are the observations z . The variances of the observations used for the *ICI* (24) and fusing (23) are $\hat{\sigma}_z^2 = z$. This is the simplest “initial guess” for the observations’ variance. Nevertheless, experiments show that this choice works quite well. In particular, it is much better than a constant-variance assumption.

Output data are the estimates $\hat{y}^{[1]}(x)$, where the superscript “[1]” shows the step number.

Step 2: The input data are again the original observations z_s . The variances of the observations for *ICI* and fusing are now $\hat{\sigma}_z^2(x) = \hat{y}^{[1]}(x)$.

In fact, the first initialization step is used only in order to obtain a refined variance for the observations. These two steps correspond to the *LPA-ICI* algorithm with one iteration for the variance update.

Outputs are the estimates $\hat{y}^{[2]}(x)$. The filtered estimate $\hat{y}^{[2]}$ is no longer Poissonian, and its variance is not equal to the pointwise mean value of the observation.

Step 3: The input data are the estimates $\hat{y}^{[2]}(x)$ used instead of the original data. The variances required for *ICI* are calculated using both $\hat{y}^{[2]}(x)$ and $\sigma_{\hat{y}_{h^+, \theta_i}^{[2]}}^2(x) = \sigma_{\theta_i}^{-2}(x)$ as

$$\sigma_{\hat{y}}^2(x) = \frac{2}{3} \left(\sum_i \sigma_{\theta_i}^{-2}(x) - \sum_i \left(\frac{\hat{\sigma}_z(x) g_{h^+(x, \theta_i), \theta_i}(0)}{\sigma_{\theta_i}^{-2}(x)} \right)^2 + \left(\sum_i \frac{\hat{\sigma}_z(x) g_{h^+(x, \theta_i), \theta_i}(0)}{\sigma_{\theta_i}^{-2}(x)} \right)^2 \right) \left(\sum_j \sigma_{\theta_j}^{-2}(x) \right)^{-2}$$

In this empirical formula we try to take into consideration also the correlation of the estimates, which are not independent after the filtering of the previous step, since the directional kernels always have the origin pixel as the only common point. Note that the variance of the observation $\hat{\sigma}_z(x)$ is calculated using the latest estimate of the signal (i.e. $\hat{\sigma}_z = \hat{y}^{[2]}$).

Outputs are the estimates $\hat{y}^{[3]}(x)$.

Step 4 and further are identical to Step 3, with input signal obtained from the previous steps.

Comment: The particular form of recursive filtering exploited from the third step onwards is a refined Poissonian version of the recursive *LPA-ICI* filtering for Gaussian-distributed observations.²

7. NUMERICAL EXPERIMENTS

7.1. Optimization of the algorithm

Some work has been done in order to optimize the design parameters of the above algorithm. After this optimization, the algorithm with these parameter values was used for multiple experiments, part of which is presented in what follows.

The directional kernels $g_{h,\theta}$ are defined as a linear combination of zero and first order kernels:

$$g_{h,\theta} = \alpha \tilde{g}_{h,\theta}|_{m=(0,0)} + (1 - \alpha) \tilde{g}_{h,\theta}|_{m=(1,0)}. \quad (27)$$

These $\tilde{g}_{h,\theta}|_{m=(0,0)}$ and $\tilde{g}_{h,\theta}|_{m=(1,0)}$ are directional *LPA* kernels designed from a set of window functions w_h constant on their sectorial support. The scale parameters h , which define the length of the support, were taken from the following set:

$$H = \{1, 2, 3, 4, 6, 8, 10, 12\}. \quad (28)$$

The *ICI* rule is applied for the selection of the length h of the kernel $g_{h,\theta}$.

It was found that significant improvement is achieved in the first three steps described above. The maximum number of steps has been thus restricted to four.

The parameter α in the combined kernel $g_{h,\theta}$ (27) is taken with different values for the different steps of the algorithm, starting from $\alpha = 1$ (zero order), so to be consistent with the considerations from section 3.5.1, and increasing then the importance of the first order component as the algorithm progress.

The performance benefits if the observations are not oversmoothed at each single step, as further smoothing will be provided by the subsequent iterations. Thus, the selected threshold parameter for *ICI* is rather small, $\Gamma = 0.7$. The same Γ is used for all four steps.

7.2. Poisson observations

In our simulations for the Poissonian case, in order to achieve different level of randomness (i.e. different *SNR*) in the noisy observations, we first multiply the true signal y^{TRUE} (which has range $[0,1]$) by a scaling factor χ :

$$y = \chi \cdot y^{\text{TRUE}}, \quad z \sim \mathcal{P}(y).$$

Thus, $E\{z\} = \text{var}\{z\} = y = \chi \cdot y^{\text{TRUE}}$, and $y/\text{std}\{z\} = \sqrt{\chi} \sqrt{y^{\text{TRUE}}}$, i.e. better *SNR* corresponds to larger χ .

This modelling of Poisson data allows to produce a comparison with the similar simulation scenarios appeared in a number of publications.^{13,14,17,18} We make a comparison with the wavelet based methods recently developed for the Poisson data and demonstrating a good performance.

Only to visualise the data we divide back by the factor χ , so that the expected range of the signal intensity is again $[0, 1]$. Figure 2 illustrates the effect of this scaling factor in modelling Poisson observations. Comparing the images in this figure, we can see that the noise level for $\chi=120$ is much lower than it is for $\chi=30$. From the cross-section we can note that the level of this random disturbance is clearly signal-dependent. Large value of the signal means larger level of the noise.

7.3. Simulation results

Images shown in Figures 3 and 4 show the noisy and original images and the estimates obtained at the first (initialization) and the last (fourth) iteration. The *MSE* values demonstrate a fast performance improvement in the successive iterations. The quality of the final estimates is quite good visually and numerically. In particular, for *Cameraman* we achieve: *ISNR*=9.34dB for $\chi=30$, *ISNR*=8.05dB for $\chi=60$, *ISNR*=7.45dB for $\chi=90$, *ISNR*=6.82dB for $\chi=120$.

Some numerical results and comparison with other methods for the *Cameraman* and *Lena* images are presented in Table 1. The results in the table are the values of the *MSE*, calculated as the mean value of $|\hat{y} - y|^2$. This table includes and extends the results shown in.¹⁴

Comparing the *MSE* values obtained for the successive steps we can note that the main improvement is achieved in first three steps. Starting from the second step of the recursive procedure, the *LPA-ICI* shows superior results which are essentially better than those from the other methods.^{13,14}

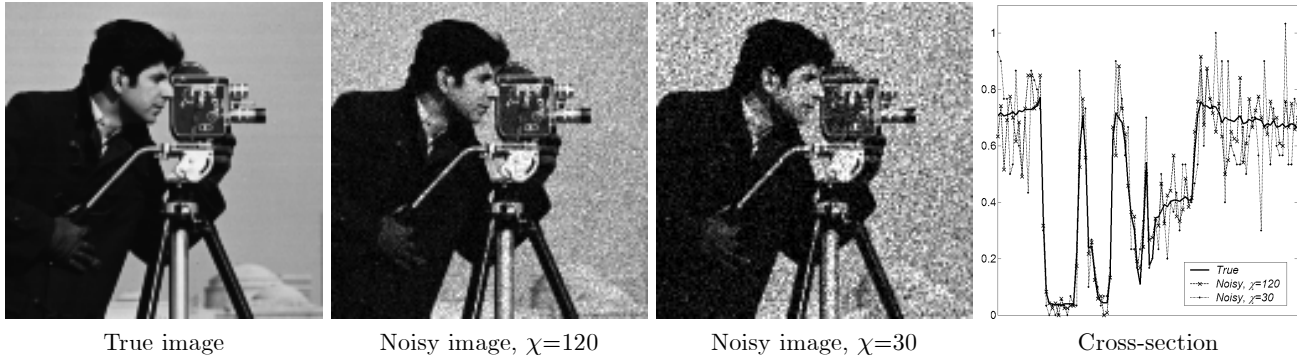


Figure 2. Cameraman fragment: true and Poisson noisy images.



Figure 3. Filtering a fragment of the *Cameraman* image: (from left to right) noisy data, $\chi=60$ ($MSE=23.9$), estimate from the first iteration of the *LPA-ICI-AV* algorithm ($MSE=7.90$, $ISNR=4.81$ dB), estimate after the fourth iteration ($MSE=4.36$, $ISNR=7.40$ dB), and original image.

Table 1. *MSE* for the *Cameraman* and *Lena* images for different algorithms and different levels of noise

	<i>Cameraman</i>				<i>Lena</i>			
	$\chi=30$	$\chi=60$	$\chi=90$	$\chi=120$	$\chi=30$	$\chi=60$	$\chi=90$	$\chi=120$
<i>Noisy image</i>	13.9	27.5	42.1	56.0	14.5	29.0	43.6	58.0
<i>TN method</i> ¹³	2.76	7.73	14.11	21.59	2.33	6.46	11.62	17.89
<i>Improved TN method</i> ¹⁴	2.13	5.37	9.22	13.59	1.98	5.32	9.35	14.03
<i>LPA-ICI</i> , 1 st step	3.75	8.04	13.4	18.8	3.29	7.07	11.3	15.9
<i>LPA-ICI</i> , 2 nd step	2.28	5.30	8.96	13.2	1.76	4.07	6.80	9.84
<i>LPA-ICI</i> , 3 rd step	1.79	4.50	7.79	11.8	1.20	3.05	5.33	7.99
<i>LPA-ICI</i> , 4 th step	1.62	4.30	7.58	11.6	0.99	2.62	4.72	7.19

8. ACKNOWLEDGMENTS

The work of the first author was supported in part by a *Visiting Fellow* grant from Nokia Foundation.

REFERENCES

1. Fan J. and I. Gijbels, *Local polynomial modelling and its application*. London: Chapman and Hall, 1996.
2. Foi, A., Katkovnik, V., Egiazarian, K. & Astola, J., "A novel anisotropic local polynomial estimator based on directional multiscale optimizations", *Proc. of the 6th IMA Int. Conf. Math. in Signal Processing*, Cirencester (UK), pp. 79-82 (2004).
3. Katkovnik V., "A new method for varying adaptive bandwidth selection", *IEEE Trans. on Signal Proc.*, vol. 47, no. 9, pp. 2567-2571, 1999.



Figure 4. Filtering the *Lena* image: top row, left - noisy data $\chi=60$ ($MSE=29.0$, $SNR=15.3$ dB); top row, right - first iteration of the *LPA-ICI-AV* algorithm ($MSE=7.07$, $ISNR=6.13$ dB); second row, left - fourth iteration ($MSE=2.62$, $ISNR=10.4$ dB); second row, right - original image.

4. Katkovnik V, "Multiresolution kernel nonparametric regression: a new approach to pointwise spatial adaptation," *Discrete Signal Processing*, 2004, in print.
5. Katkovnik V., K. Egiazarian, and J. Astola, "Adaptive window size image de-noising based on intersection of confidence intervals (ICI) rule". *J. of Math. Imaging and Vision*, vol. 16, N. 3, pp. 223-235, 2002.
6. Katkovnik V., K. Egiazarian, and J. Astola. *Adaptive varying scale methods in image processing*. Tampere International Center for Signal Processing, TICSP Series, N.19, Tampere, TTY, Monistamo, 2003.
7. Katkovnik V., Foi A., K. Egiazarian, and J. Astola, "Directional varying scale approximations for anisotropic signal processing," *EUSIPCO 2004*, Wien, 2004.
8. Lepski, O., E. Mammen and V. Spokoiny, "Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selection". *Annals of Statistics*, vol. 25, no. 3, 929-947, 1997.
9. Loader C., *Local regression and likelihood*, Series Statistics and Computing, Springer-Verlag New York, 1999.
10. Goldenshluger, A. and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression," *Math. Meth. Statistics*, vol. 6, pp. 135-170, 1997.
11. Huber P.J., *Robust statistics*, John Wiley & Sons Inc., 1981.
12. Astola J. and P. Kuosmanen, *Fundamentals of Nonlinear Digital Filtering*, New York, CRC Press, 1997.
13. Timmermann K. E. and R. Nowak, "Multiscale modeling and estimation of Poisson processes with application to photon-limited imaging" *IEEE Trans. Information Theory*, Vol. 45, no. 3, pp. 846-862, 1999.
14. Lu H., Kim Y. and J. M. M. Anderson, "Improved Poisson intensity estimation: denoising application using poisson data," *IEEE Trans. Image Processing*, vol. 13, no. 8, pp 1128-1135, 2004.
15. Hastie T. J. and R. J. Tibshirani. *Generalized linear models*. London: Chapman and Hall, 1990.
16. Tibshirani, R. J. and T. J. Hastie, "Local likelihood estimation", *Journal of American Statistical Association*, vol. 82, pp. 559-567, 1987.
17. Nowak R. D. and R. G. Baraniuk, "Wavelet-domain filtering for photon imaging systems," *IEEE Trans. Image Processing*, vol. 8, no. 5, pp 666-678, 1999.
18. Willett, R. M. and R. D. Nowak, "Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging", *IEEE Trans Medical Imaging*, vol. 22, no. 3, pp. 332-350, 2003.