STRUCTURE ADAPTIVE METHODS IN IMAGE DENOISING

Vladimir Spokoiny

Weierstrass Institute for Applied Analysis and Stochastics,
Mohrenstrasse 39, D - 10117 Berlin, Germany
e-mail: spokoiny@wias-berlin.de

1. ABSTRACT

The nonparametric regression originated in mathematical statistics offers an original approach to signal processing problems (e.g., [1], [2]). It basically results in linear filtering with the linear filters designed using some moving window local approximations. In many applications like speech recognition or image denoising, nonlinear or locally adaptive methods have been shown to be more efficient than the linear ones. The typical examples are given by non-linear wavelet thresholding, [3], and pointwise adaptive kernel smoothing, [4], [5]. The first local pointwise (varying window size) adaptive nonparametric regression statistical procedure was suggested by Lepski [6] (see also [4], [5], and [7]). This approach has been further developed in application to various signal and image processing problems [8]–[12]. Particularly, [12] offered another view on the problem of local adaptive estimation based on the link between adaptive estimation and multiple testing. This allows to treat in a unified way different types of images, including Gaussian and Poissonian. Another important feature is that the problem of choosing the tuning parameters of the procedure is carefully addressed, leading to an efficient automatic procedure. The presentation extends these ideas to a general approach to spatially adaptive local parametric estimation.

Suppose we have *independent* observations $\{Z_i\}_{i=1}^n$ of the form $Z_i = (X_i, Y_i)$. Here X_i denotes a vector of "features" or explanatory variables which determines the distribution of the random "observation " Y_i. The d-dimensional vector $X_i \in \mathbb{R}^d$ can be viewed as a location in time or space and Y_i as the "observation at X_i". Our model assumes that the distribution of each Y_i is determined by a parameter f_i which may depend on the location X_i, $f_i = f(X_i)$. In many cases the *natural* parametrization is chosen which provides the relation $f_i = E\{Y_i\}$. The estimation problem is to reconstruct $f(x)$ from the data $\{Z_i\}_{i=1,\ldots,n}$. This set-up includes *Gaussian images* $Y_i = f(X_i) + \varepsilon_i$ with a regression function $f(\cdot)$ and i.i.d. Gaussian errors $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$; *Poisson images* with $P(Y_i = k | X_i = x) = f^k(x) \exp(-f(x))/k!$; *Bernoulli (binary response) images* with $P(Y_i = 1) = f(X_i)$, $P(Y_i = 0) = 1 - f(X_i)$. The joint distribution of the samples $Y_1, \ldots, Y_n$ is given by the log-likelihood $L = \sum_{i=1}^n \log p(Y_i, f(X_i))$.

In the parametric setup, the whole function $f(\cdot)$ is determined by a parameter vector $\boldsymbol{\theta}$: $f(\cdot) = f(\cdot, \boldsymbol{\theta})$. This reduces the problem of estimating f to the problem of estimating $\boldsymbol{\theta} \in \Theta \in \mathbb{R}^p$. The maximum likelihood approach yields the estimates

$$\tilde{\boldsymbol{\theta}} = \operatorname*{argmax}_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) = \operatorname*{argmax}_{\boldsymbol{\theta}} \sum_i \log p(Y_i, f(X_i, \boldsymbol{\theta})).$$

However, a parametric assumption $f(\cdot) \equiv f(\cdot, \boldsymbol{\theta})$ can be too restrictive, especially if the family $f(\cdot, \boldsymbol{\theta})$ is not very rich. *Local parametric approach* supposes that for every x there is a vicinity U of x such that $f(X_i) \approx f(X_i, \boldsymbol{\theta})$ for all $X_i \in U$. Given $U = U(x)$, one can estimate the local parameter $\boldsymbol{\theta} = \boldsymbol{\theta}(x)$ by maximizing the localized log-likelihood $L_U(\boldsymbol{\theta})$:

$$\begin{aligned} \tilde{\boldsymbol{\theta}}(x) &= \operatorname*{argmax}_{\boldsymbol{\theta}} L_U(\boldsymbol{\theta}) \\ &= \operatorname*{argmax}_{\boldsymbol{\theta}} \sum_{X_i \in U} \log p(Y_i, f(X_i, \boldsymbol{\theta})). \end{aligned}$$

leading to the estimate $\tilde{f}(x) = f(x, \tilde{\boldsymbol{\theta}}(x))$. More generally, one can define locality around x by a collection of weights $W(x) = \{w_i(x)\}$ leading to the local estimate

$$\begin{aligned} \tilde{\boldsymbol{\theta}}(x) &= \operatorname*{argmax}_{\boldsymbol{\theta}} L(W(x), \boldsymbol{\theta}) \\ &= \operatorname*{argmax}_{\boldsymbol{\theta}} \sum_i \log p(Y_i, f(X_i, \boldsymbol{\theta})) w_i(x). \end{aligned}$$

The focus in applying the local parametric approach is choosing the collection of localizing weights $w_i(x)$. We discuss two different approaches for selecting such weights adaptively from the observed data. One is based on the pointwise adaptive choosing of directional bandwidths [12]. The other one explores the idea of structural adaptive estimation, so called Adaptive Weights Smoothing procedure [13]-[15].

REFERENCES

[1] J. Fan J. and I. Gijbels, *Local polynomial modelling and its application.* London: Chapman and Hall, 1996.

[2] C. Loader, *Local regression and likelihood,* Series Statistics and Computing, Springer-Verlag New York, 1999.

[3] D. Donoho and I. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika,* vol. 81, pp. 425–455, 1994.

[4] O. Lepski, E. Mammen and V. Spokoiny, "Ideal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selection," *The Annals of Statistic*s, vol. 25, no. 3, 929–947, 1997.

[5] A. Goldenshluger and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression", *Math. Meth. Statistics*, vol.6, pp.135–170, 1997.

[6] O.V. Lepski, "One problem of adaptive estimation in Gaussian white noise," *Theory Probab. Appl.*, vol. 35, no. 3, pp. 459–470, 1990.

[7] O. Lepski and V. Spokoiny, "Optimal pointwise adaptive methods in nonparametric estimation," *The Annals of Statistics,* vol. 25, no. 6, pp. 2512–2546, 1997.

[8] V. Katkovnik, "A new method for varying adaptive bandwidth selection," *IEEE Trans. Sig. Proc.*, vol. 47, no. 9, pp. 2567–2571, 1999.

[9] V. Katkovnik, K. Egiazarian and J. Astola, "Adaptive window size image de-noising based on intersection of confidence intervals (ICI) rule," *Journal of Math. Imaging and Vision*, vol. 16, no. 3, pp. 223–235, 2002.

[10] A. Foi, *Anisotropic nonparametric image processing: theory, algorithms and applications,* Ph.D. Thesis, Dip. di Matematica, Politecnico di Milano, ERLTDD-D01290, April 2005. Available: *www.cs.tut.fi/~lasip*.

[11] V. Katkovnik, K. Egiazarian, J. Astola, *Local Approximation Techniques in Signal and Image Processing*. SPIE PRESS, Bellingham, Washington, 2006.

[12] V. Katkovnik and V. Spokoiny. Spatially adaptive estimation via fitted local likelihood techniques. *IEEE Transactions on Signal Processing*, 56(3), 873–886, 2008.

[13] J. Polzehl and V. Spokoiny. Adaptive weights smoothing with applications to image restoration. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 62(2):335–354, 2000.

[14] J. Polzehl and V. Spokoiny, "Propagation-separation approach for local likelihood estimation, *Probab. Theory Related Fields*, vol. 135, no. 3, 335–362, 2005.

[15] V. Spokoiny, *Local parametric methods in nonparametric estimation*, Springer, 2008. To appear.