The work of Oleg Lepski: beyond a "discourse on method"

Marc Hoffmann

Celebrating the 60th birthday of Oleg Lepski and Sacha Tsybakov

CIRM, Luminy, 21 December 2017

In this talk

- This is a slightly biased – but hopefully honest – account of *some* of Oleg's many contributions to mathematical statistics over the last thirty years.
- This is also a personal approach, a tentative grasp of the part of his work and his scientific life that I like best or that I know best.
- Disclaimer: I am solely responsible for this talk and apologise for the (likely many) inaccuracies and oversimplifications that may follow.

Content

- A brief overview of Oleg's *vitae* and career
- Part I: The Moscow years
- Part II: The Berlin years
- Part III: The Marseille years, up to now
- A final word of conclusion

Content

Part I: The Moscow years (*ca* 1983 – 1991)

- Early works
- A seminal paper
- Adaptive estimation: a theory in the making

Part II: The Berlin years (*ca* 1993 – 1998)

- Asymptotic adaptive efficiency in estimation and testing, more adaptive estimation
- Mixing estimation and testing: the dawn of adaptive confidence sets?

Content

Part III: The Marseille years (1998 - today)

- Anisotropy
- The third adaptation period: the oracle approach and Anisotropy again

A brief overview of Oleg's career and *vitae*

Figure: Oleg (courtesy of the Oberwolfach collection).

What do we know about Oleg?

We know him as

- A mathematician, a statistician

What do we know about Oleg?

We know him as

- A mathematician, a statistician
- An academic, with a rich career – slightly thwarted by history – over several countries

What do we know about Oleg?

We know him as

- A mathematician, a statistician
- An academic, with a rich career – slightly thwarted by history – over several countries
- A poet, a political pamphleteer, a friend

What do we know about Oleg?

Figure: Oleg *ca* 1975. A poet in the making?

What do we know about Oleg?

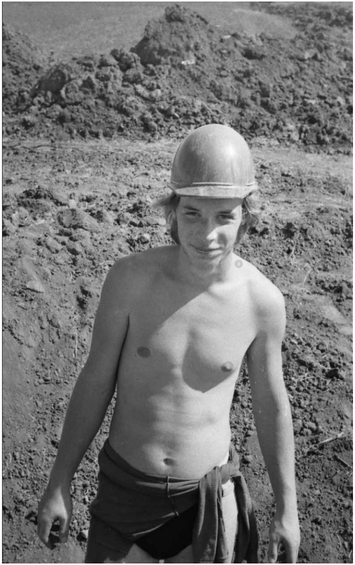


Figure: Oleg *ca* 1975. Or a builder, a future theory maker?

What do we know about Oleg?

Figure: Oleg *ca* 1975. He wouldn't be alone... (The "Pulsar" construction crew)

What do we know about Oleg?

Figure: Oleg *ca* 1975. And would work in an orderly way... .

What do we know about Oleg?

Figure: Oleg *ca* 2010. Apparently, a lot has been achieved...

What do we know about Oleg?

Figure: Oleg *ca* 1975. Simply a youthful character full of promises.

Oleg's brief *vitae*

- Born in 1957 in Moscow, son of an architect.
- PhD on 12 November 1984, "supervised" by A. P. Korostelev.
- Affiliated to the *Institute for Systems Analysis*, Russian Academy of Sciences.
- Participant of the Khasminski seminar in Moscow.
- 1992: Post-doc in Louvain-la-Neuve.
- Humboldt-Universität zu Berlin from 1993 to 1998. Stayed in the WIAS.
- 1998 up to now: Professor in LATP at the university of Aix-Marseille.
- Medallion lecture at the 2005 JSM, Minneapolis.

Oleg's brief *vitae*

- PhD students: A. Yode, N. Klutchnikoff, F. Chiabrando, M. Chichignoud, N.B. Nguyen, G. Rebelles.
- Co-authors:

  L. Cavalier, V.V. Fedorov, D. Feldmann, A. Goldenshluger, G.K. Golubev, C. Hafner, M. Hoffmann, W. Härdle, Wolfgang, Y.I. Ingster, A. Juditsky, G. Kerkyacharian, A.P. Korostelev, F. Leblanc, B. Levit, E. Mammen, A.S. Nemirovski, D. Picard, C. Pouet, N. Serdyukova, V. Spokoiny, A.B. Tsybakov, T. Willer.
- MathSciNet records 46 published papers between 1983 and 2017. (There are more to come!)

Oleg's brief *vitae*

MathSciNet records 46 published papers between 1983 and 2017.
(There are more to come!)

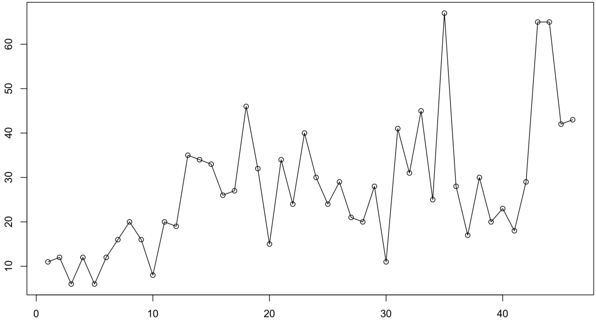


Figure: *x*-axis: papers (in paper time). *y*-axis: number of pages.

Oleg's brief *vitae*

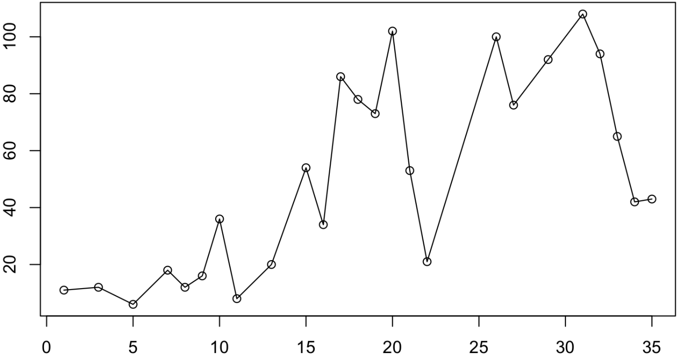


Figure: x-axis: time (in years since first publication). y-axis: number of pages.

Oleg's brief *vitae*

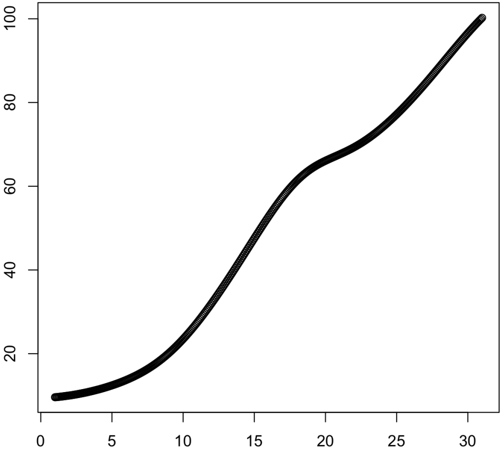


Figure: x-axis: time (in years since first publication). y-axis: number of pages. Smoothing by adaptive choice of bandwidth $h = 4$ years on $[0, 31]$, `KernSmooth` package in `R`, `locpoly` function.

Part I: The Moscow years

Figure: Oleg around 1985.

Early works

- The first (recorded) published paper by Oleg was actually devoted to solving a statistical problem...

Early works

- The first (recorded) published paper by Oleg was actually devoted to solving a statistical problem...
- ... for *stochastic processes in discrete time!*

Early works

- Lepskii, O.V. Asymptotic properties of estimates of parameters of a generalised autoregression scheme in the unstable case. (Russian). *Vsesoyuz. Naucho-Issled. Instit. Sistem. Issled., Moscow* (1983).
- Problem: estimate ϑ (and σ) from data $(X_t)_{t=1,\dots,n}$ in the model

$$X_{t+1} = \ell_t(\vartheta)X_t + \sigma_t(\vartheta,\sigma)\xi_t, \;\; \mathbb{E}[\xi_t] = 0, \mathbb{E}[\xi_t^2] = 1.$$

- Under suitable assumptions, the problem is soved in the unstable case (something like $|\ell_t(\vartheta)| > 1$), a minimum contrast estimator is built and its limiting distribution is studied.

Early works

- Another paper is published in the same area two years later
- Korostelev, A.P., Lepskii, O.V., Fedorov, V.V. Analysis of time series generated by stochastic differential equations. (Russian). *Uchen. Zap. Statist., 49, "Nauka", Moscow* (1985)

Early works

- Probably around 1984-85, Oleg leaves (presumably forever) the subject of statistics for stochastic processes and moves to *minimax parametric estimation*. He still works closely with A.P. Korostelev.
- Lepskii, O.V. Asymptotically minimax estimation of a parameter under asymmetric loss functions. (Russian) *Teor. Veroyatnost. i Primenen.* (1987).
- Lepskii, O.V. Asymptotic minimax estimation with prescribed properties. (Russian) *Teor. Veroyatnost. i Primenen.* (1989).
- Korostelëv, A. P., Lepskii, O. V. Asymptotically minimax estimation in the change-point problem. (Russian) *Statistics and control of random processes* (Russian) (Preila, 1987) "Nauka", Moscow, 1989.

Early works

- In the first paper above, Oleg studies *"situations [...] in which, apart from the natural requirement that an estimator is close to the parameter, the estimator is also subject to additional prescribed properties..."*
- This is reminiscent of what Mark Low will call a few years later *constrained estimation*, in a problem of adaptive estimation, precisely already solved some years before by... Oleg!
- The work is inspired by D. Anbar (AOS, 1977) and allows one to consider *e.g.* in dimension one, a loss of the type

$$w(x) = (1 + |x|)\mathbf{1}_{\{x<0\}} + x\mathbf{1}_{\{x\geq 0\}}$$

  resulting in a minimax risk of the form

$$\inf_{\widehat{\vartheta}} \sup_{\vartheta \in \Theta} \left(\mathbb{P}_\vartheta(\widehat{\vartheta} < \vartheta) + \mathbb{E}_\vartheta\left[|\widehat{\vartheta} - \vartheta|\right]\right).$$

Early works

- In an asymptotic setting, a minimax lower bound is proved under a general LAN-type assumption on the model.
- If there exists an estimator $\widehat{\vartheta}_n$ with a certain concentration property (namely $\mathbb{P}_\vartheta(v_n^{-1}|\widehat{\vartheta} - \vartheta| \geq t) \lesssim e^{-Ct^{C'}}$) the LB is achieved.
- The paper combines estimation and testing in a way that will prove essential in the later works of Oleg.
- I&K book's techniques in parametric estimation are everywhere, but limiting distributions are not the main focus for Oleg.

A seminal paper: the no-adaptation result

- Around 1987, Oleg presents at the Khasminski seminar an unexpected result on *nonparametric adaptation*, quite a new topic at that time.
- Oleg shows for the first time that there exist reasonable settings where rate adaptation is simply *not feasible*!
- He also proposes a notion of *weaker adaptation* and builds an adaptive estimator in that new sense.
- The construction lays the foundations of what will later be coined *the Lepski's method*.

A seminal paper: the no-adaptation result

Том 35 | ТЕОРИЯ ВЕРОЯТНОСТЕЙ И ЕЕ ПРИМЕНЕНИЯ 1990 | Выпуск 3

ЛЕПСКИЙ О. В.

ОБ ОДНОЙ ЗАДАЧЕ АДАПТИВНОГО ОЦЕНИВАНИЯ В ГАУССОВСКОМ БЕЛОМ ШУМЕ

§ 1. Введение

На отрезке [0, 1] наблюдается случайный процесс, удовлетворяющий стохастическому дифференциальному уравнению

$$dX_\varepsilon(t) = S(t)\,dt + \varepsilon\,dw(t), \qquad (1)$$

где $w(\cdot)$ — стандартный винеровский процесс, $\varepsilon \to 0$ — малый параметр. По наблюдению за траекторией $X_\varepsilon(t)$, $0 \leqslant t \leqslant 1$, требуется оценить величину $S(t_0)$, где t_0 — известная фиксированная точка из интервала (0, 1).

Обозначим через $\mathscr{Y}_\varepsilon$ совокупность измеримых относительно $X_\varepsilon(t)$, $0 \leqslant t \leqslant 1$, функций (оценок), а через $\mathbf{P}_{S(\cdot)}$ и $\mathbf{E}_{S(\cdot)}$ — меры и усреднения, соответствующие процессу $X_\varepsilon(t)$, $0 \leqslant t \leqslant 1$, при условии, что уравнение (1) порождалось функцией $S(\cdot)$.

Функцию $l\colon R^1 \to R^1$ будем называть функцией потерь (ф.п.), если она неотрицательна, симметрична, монотонно не убывает на положительной полуоси, непрерывна в нуле и $l(0) = 0$.

Пусть Σ — произвольное множество функций. Рассмотрим минимаксный риск вида

$$R_\varepsilon(\tilde\theta_\varepsilon, \Sigma, \varphi_\Sigma(\varepsilon)) = \sup_{S(\cdot)\in\Sigma} \mathbf{E}_{S(\cdot)} l(\varphi_\Sigma^{-1}(\varepsilon)(\tilde\theta_\varepsilon - S(t_0))),$$

где $\tilde\theta_\varepsilon \in \mathscr{Y}_\varepsilon$, а $\varphi_\Sigma(\varepsilon) > 0$ — нормирующая функция.

О п р е д е л е н и е 1. Функцию $\varphi_\Sigma(\varepsilon) > 0$ будем называть минимаксным порядком точности оценивания (МПТ) величины $S(t_0)$ н а множестве Σ относительно ф.п. $l(\cdot)$, если

$$\lim\ \inf\ R_\varepsilon(\tilde\theta_\varepsilon, \Sigma, \varphi_\Sigma(\varepsilon)) > 0$$

Figure: A problem of adaptive estimation in Gaussian white noise, *Teor. Veroyatnost. i Primenen.*, 1990, Volume 35, Issue 3, 459–470.

A seminal paper: the no-adaptation result

Отсюда

$$\overline{\lim_{\varepsilon \to 0}}\, u_b(\varepsilon)\,(\varepsilon^2 \ln(1/\varepsilon))^{-b/(2b+1)} = 0. \qquad (31)$$

В силу теоремы 2 выполнение соотношений (30) и (31) противоречит принадлежности семейства $\{U\}$ множеству $\mathfrak{A}$. Полученное противоречие завершает доказательство теоремы.

Теорема доказана.

СПИСОК ЛИТЕРАТУРЫ

1. *Ибрагимов И. А., Хасьминский Р. З.* Асимптотическая теория оценивания. М.: Наука, 1979, 528 с.
2. *Ибрагимов И. А., Хасьминский Р. З.* Об оценках сигнала, его производных и точки максимума для гауссовских наблюдений.— Теория вероятн. и ее примен., 1980, т. 25, в. 4, с. 718—733.
3. *Ефроймович С. Ю., Пинскер М. С.* Самообучающийся алгоритм непараметрической фильтрации.— Автомат. и телемех., 1984. т. 11, с. 58—65.

Поступила в редакцию
21.XII.1987

Figure: A problem of adaptive estimation in Gaussian white noise, *Teor. Veroyatnost. i Primenen.*, 1990, Volume 35, Issue 3, 459–470.

A seminal paper: the no-adaptation result

:РАТУРЫ

тотическая теория оценивания. М.: На -

энках сигнала, его производных и точки
– Теория вероятн. и ее примен., 1980,

чающийся алгоритм непараметрической
. т. 11, с. 58—65.

Поступила в редакцию
21.XII.1987

Figure: A problem of adaptive estimation in Gaussian white noise, *Teor. Veroyatnost. i Primenen.*, 1990, Volume 35, Issue 3, 459–470.

A seminal paper: the no-adaptation result

$$Y_\varepsilon = f + \varepsilon \dot{W} \ \text{ on } \ L^2([0,1]) \ \text{ over Hölder classes } \ \Sigma_\beta, \beta > 0.$$

Result 1: For any $0 < \beta_1 < \beta_2, t_0 \in [0,1], q > 0$, one has

$$\liminf_{\varepsilon \to 0} \inf_{F(\cdot)} \sum_{i=1}^{2} \left(\varepsilon^2 \log(1/\varepsilon)\right)^{-\frac{\beta_i q}{2\beta_i+1}} \sup_{f \in \Sigma_{\beta_i}} \mathbb{E}_f \left[\left| F(Y_\varepsilon) - f(t_0) \right|^q \right] > 0.$$

Result 2: There exists an estimator $\widehat{f}_\varepsilon(t_0)$, based on a certain comparison scheme, such that for any (closed and bounded) $\mathcal{I} \subset (0, \infty)$,

$$\sup_{\beta \in \mathcal{I}} \left(\varepsilon^2 \log(1/\varepsilon)\right)^{-\frac{\beta q}{2\beta+1}} \sup_{f \in \Sigma_{\beta_i}} \mathbb{E}_f \left[\left| \widehat{f}_\varepsilon(t_0) - f(t_0) \right|^q \right] \lesssim 1.$$

A seminal paper: the no-adaptation result

- The construction looks very simple if $\mathcal{I} = \{\beta_{\min}, \beta_{\max}\}$. Let $h_\varepsilon(\beta) \approx \big(\varepsilon^2 \log(1/\varepsilon)\big)^{2/(2\beta+1)}$. One can take

$$\widehat{f}_\varepsilon(t_0) = \begin{cases} K_{h_\varepsilon(\beta_{\min})} \star Y_\varepsilon & \text{if } \left|(K_{h_\varepsilon(\beta_{\min})} - K_{h_\varepsilon(\beta_{\max})}) \star Y_\varepsilon\right| \lesssim \varepsilon^{\frac{2\beta_{\min}}{2\beta_{\min}+1}} \\ K_{h_\varepsilon(\beta_{\max})} \star Y_\varepsilon & \text{otherwise} \end{cases}$$

- But Oleg looks further, and introduces a notion of *optimal adaptation* in this context.

A seminal paper: the no-adaptation result

- A family of normalising factors $\mathcal{F} = \{r_\varepsilon(\beta), \beta \in \mathcal{I}\}$ is *admissible* if $r_\varepsilon(\beta)$ is an adaptively achievable: there exists $\widehat{f}_\varepsilon(t_0)$ such that

$$\sup_{\beta \in \mathcal{I}} r_\varepsilon(\beta)^{-q} \sup_{f \in \Sigma_\beta} \mathbb{E}_f \left[\left| \widehat{f}_\varepsilon(t_0) - f(t_0) \right|^q \right] \lesssim 1.$$

Let

$$\Psi_\varepsilon(\mathcal{F}) = \sup_{\beta \in \mathcal{I}} r_\varepsilon(\beta) \varepsilon^{-2\beta/(2\beta+1)} \text{ be the } \textit{characteristic} \text{ of } \mathcal{F}.$$

The family $\mathcal{F}^\star$ is then *Lepski adaptive optimal* if

i) $\mathcal{F}^\star$ is admissible

ii) There is no admissible $\mathcal{F}$ such that $\Psi_\varepsilon(\mathcal{F}) \ll \Psi_\varepsilon(\mathcal{F}^\star)$.

A seminal paper: the no-adaptation result

Two new features depart from previous existing (rigorous) adaptation results (Efromovich & Pinsker 1984, Golubev 1987):

- The Oleg's scheme is extremely flexible w.r.t. the loss function.
- It is a procedure selection among a given family of estimators that separately attain their minimax target.

This is the beginning of a vast program about a *theory of adaptive estimation* (still ongoing!)

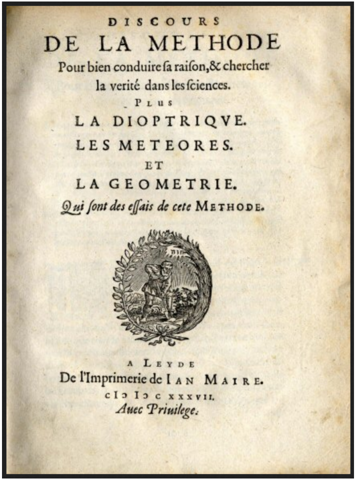
A seminal paper: the no-adaptation result

- Oleg subsequently produced three papers developping a systematic approach for solving adaptive estimation (or possibly the lack thereof)
- Lepskii, O. V. Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. (Russian) *Teor. Veroyatnost. i Primenen.* (1991).
- Lepskii, O. V. On problems of adaptive estimation in white Gaussian noise. Topics in nonparametric estimation, *Adv. Soviet Math.*, Amer. Math. Soc., Providence, RI, (1992).
- Lepskii, O. V. Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation. Adaptive estimates. (Russian) *Teor. Veroyatnost. i Primenen.* (1992).

A seminal paper: the no-adaptation result

- In these papers, Oleg refines the *Lepski's method* by addressing several problems, including
  i) Finding sufficient conditions for existence of adaptive estimation with a generic construction,
  ii) Estimating adaptively Hölder signals in white noise under L^p-norm error and related functionals,
  iii) Finding sufficient conditions for the nonexistence of adaptive estimators,
  iv) Estimating adaptively functionals of Hölder signals in white noise.
- But, as claimed in the title, this talk shall not be a "discourse on method" so let us move on!

René Descartes: Discours de la méthode


DISCOURS
DE LA METHODE
Pour bien conduire ſa raiſon, & chercher
la verité dans les ſciences.
PLUS
LA DIOPTRIQVE.
LES METEORES.
ET
LA GEOMETRIE.
Qui ſont des eſſais de cete METHODE.

A LEYDE
De l'Imprimerie de IAN MAIRE.
CIↃ IↃ C XXXVII.
Auec Priuilege.

Figure: Discours de la méthode, pour bien conduire sa raison & chercher la vérité dans les sciences, 1668.

Part II: The Berlin years

Figure: Weierstraß Institut, Mohrenstraße 39, Berlin Stadtmitte.

Uncertain times

- Uncertain times followed the collapse of Soviet Union for the participants of the now former Khasminski seminar.
- Oleg was invited in Louvain-la-Neuve in 1992, thanks to Sacha Tsybakov's influence – Sacha had preceded him there, before moving to France himself in 1993.
- In 1993, Oleg was invited to Berlin by M. Nußbaum, head of the Statistics research group in the *Weierstraß Institute* (WIAS) at that time.
- Affiliated to *Humboldt University*, Oleg would stay in WIAS for a few years. Volodia Spokoiny was there too.

The WIAS seminar

Figure: Celebration of the 50th anniversary of the WIAS seminar, less than a month ago!

Adaptive efficient estimation and testing

- Besides the *WIAS seminar*, the *Paris-Berlin Seminar* was launched *ca* 1994. It would last until 1998 and would be followed by... *Luminy!*
- In Berlin, a fruitful period of collaboration between Oleg and Volodia (and others, including Sacha) started.
- It would lead to *some astonishingly original* results in nonparametric estimation.

Adaptive efficient estimation and testing

- Lepskii, O. V., Spokoiny, V. G. Local adaptation to inhomogeneous smoothness: resolution level. *Math. Methods Statist.* (1995)
- Lepskii, O. V., Mammen, E., Spokoiny, V. G. Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* (1997).
- Lepski, O. V., Spokoiny, V. G. Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* (1997).
- Lepski, O., Nemirovski, A., Spokoiny, V. On estimation of the L^r-norm of a regression function. *PTRF* (1999).
- Lepski, Oleg V., Spokoiny, V. G. Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. *Bernoulli* (1999).

On estimation of the L^r-norm of a regression function

$$Y_\varepsilon = f + n^{-1/2}\dot{W} \ \text{ on } \ L^2([0,1]) \ \text{ over Hölder classes } \ \Sigma_\beta, \beta > 0.$$

- Estimate: $\|f\|_r = \big(\int_0^1 |f(t)|^r dt\big)^{1/r}$ for a given $r \geq 1$.
- Minimax risk:

$$\mathcal{R}_r(\widehat{f}_n) = \sup_{f \in \Sigma_\beta \cap \{\|f\|_\infty \leq \varrho\}} \mathbb{E}_f\Big[\big|\widehat{f}_n - \|f\|_r\big|\Big], \ \ \mathcal{R}_r^\star = \inf_{\widehat{f}_n} \mathcal{R}(\widehat{f}_n).$$

- Several results in connection with this problem were known for $r = 1, 2, 3, \infty$ at that time (an asymptotically exact constant result has also been proved by Oleg during the Moscow years). Also reminiscent of hypothesis testing.
- The optimal rate lies somewhere between $n^{-1/2}$ and $n^{-\beta/(2\beta+1)}$. It depends on *the oddity* of the exponent r!

On estimation of the L^r-norm of a regression function

- Result 1: For $r = 1$ one can construct $\widehat{f}_n$ such that

$$\mathcal{R}_1(\widehat{f}_n) \lesssim (n \log n)^{-\beta/(2\beta+1)}$$

which is better than the "nonparametric rate" $n^{-\beta/(2\beta+1)}$.

- Result 2: If r is *not* an even integer

$$\mathcal{R}_r^\star \gtrsim (n \log n)^{-\beta/(2\beta+1)} (\log n)^{-r}$$

- Result 3: If r *is* an even integer

$$n^{-\beta/(2\beta+1-1/r)} \lesssim \mathcal{R}_r^\star \lesssim n^{-\beta/(2\beta+1-1/r)}.$$

- The construction uses previous (and tricky!) works by Ibragimov, Nemirovski and Khasminskii.

Random normalising factors

- Lepski, O. V. How to improve the accuracy of estimation. *Math. Methods Statist.* (1999)
- Hoffmann, M., Lepski, O. Random rates in anisotropic regression. *Ann. Statist.* (2002)
- General principle for two hypotheses: we are given

$$\left\{\mathbb{P}_f^n, f \in \Sigma_{\beta_{\min}}\right\}, \quad \Sigma_{\beta_{\max}} \subset \Sigma_{\beta_{\min}},$$

with r_n^β, $\beta \in \{\beta_{\min}, \beta_{\max}\}$ minimax rates. Also, there exists an adaptive estimator $\widehat{f}_n$:

$$\sup_{f \in \Sigma_{\beta_{\min}}} \mathbb{E}_f^n\left[r_n(f)^{-q}\|\widehat{f}_n - f\|_{L^2}^q\right] \lesssim 1, \quad q > 0,$$

where

$$r_n(f) = \begin{cases} r_n^{\beta_{\max}} & f \in \Sigma_{\beta_{\max}}, \\ r_n^{\beta_{\min}} & f \in \Sigma_{\beta_{\min}} \setminus \Sigma_{\beta_{\max}}. \end{cases}$$

Random normalising factors

- In particular, w.o.p.,

$$f \in \mathrm{Ball}_{L^2}\big(\widehat{f}_n, Cr_n(f)\big).$$

  Not a confidence statement! (or *accuracy of estimation*).
- Oleg's program: $r_n(f) \rightsquigarrow \widehat{r}_n$ (data-dependent) such that
  i) Convergence over $\Sigma_{\min}$ is achievable for $\widehat{r}_n$ in place of $r_n(f)$,
  ii) Adaptation is still in force,
  iii) If $f \in \Sigma_{\beta_{\max}}$, w.o.p.,

$$\widehat{r}_n \ll r_n^{\beta_{\min}}$$

  i.e. we have *improvement* of the accuracy of estimation.

Random normalising factors

- Typical result: for Hölder smoothness classes in Gaussian white noise:

$$\widehat{r}_n = \begin{cases} n^{-\frac{\beta_{\min}}{2\beta_{\min}+1/2} \wedge \frac{\beta_{\max}}{2\beta_{\max}+1}} & \text{w.o.p. if } \quad f \in \Sigma_{\beta_{\max}}, \\ n^{-\frac{\beta_{\min}}{2\beta_{\min}+1}} & \text{w.o.p. if } \quad f \in \Sigma_{\beta_{\min}} \setminus \Sigma_{\beta_{\max}}. \end{cases}$$

- Yields honest adaptive confidence L^2-balls for $\beta_{\min} \le \beta_{\max} \le 2\beta_{\min}$ and nontrivial improvement otherwise.
- The construction relies on the fact that the *rate of testing* in L^2 is faster than the *rate of estimation*.
- Generalisation to several hypotheses and other situations.
- This somehow anticipates the development of *adaptive confidence balls* (Baraud, Van der Vaart, Cai & Low, Nickl and co-authors).

Part III: The Marseille years

Figure: Notre-Dame de la Garde, Marseille.

Moving to France

- In the mid-nineties, thanks in part to the *Paris-Berlin Seminar*, Oleg and his work are now well introduced in France, Germany and beyond.
- A active group, driven by Dominique, Gérard and Sacha, prepares the next move for Oleg: France!
- Under the auspices of Etienne Pardoux, a probabilist, head of the *Analysis and Probability Laboratory* (LATP) in Marseille, a professorship is offered to Oleg in 1998.
- This is the beginning of the third part of Oleg's carreer.

The “Pulsar” Oleg’s crew by the end of the 1990’s

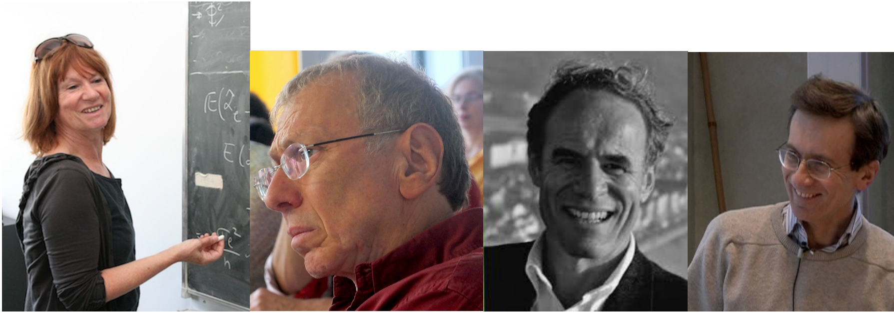


Figure: Dominique, Gérard, Etienne and Sacha: friends, but also strong scientific and academic allies.

Moving to France

- Yuri Golubev soon joins Oleg in Marseille as a senior researcher in CNRS.
- In a very few years, the *Marseille school of mathematical statistics* is raised, practically out of nothing!
- The group is alimented in part by students of Dominique and Sacha (Laurent, Christophe, Florent, Thomas).
- Starting in 2000, it organises a *yearly Mathematical Statistics* conference that will shape our community: the *Luminy* as we know it today!

The Marseille school

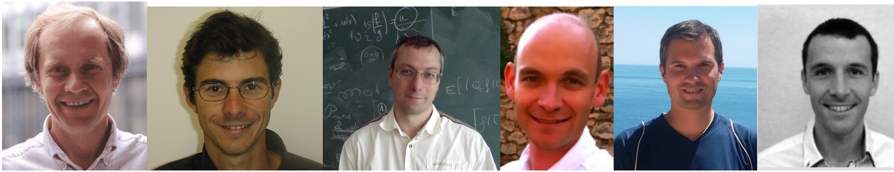


Figure: Yuri, Laurent, Christophe, Florent, Thomas, Thibault... and many of their other PhD students who moved to other academic institutions...

The Marseille school

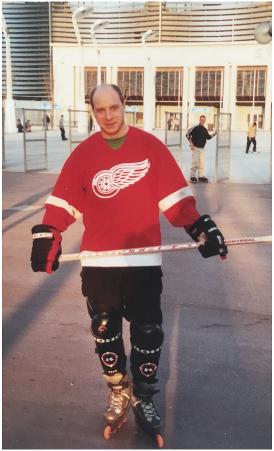


Figure: and Oleg of course... (Marseille, February 2002)

Anisotropy I

Figure: Dominique and Oleg, *ca* 2010.

Anisotropy I

$$Y_\varepsilon = f + n^{-1/2}\dot{W} \ \text{ on } \ L^2([0,1]^d), \ \ d \geq 1.$$

- Goal: estimate f in L^p-norm (minimax optimally, possibly adaptively) when the smoothness of f is non-homogeneous!
- $f \in B^{(s_1,\dots,s_d)}_{(p_1,\dots,p_d),\infty}$ if

$$\exists \ell \in \mathbb{N}, \forall i, s_i < \ell, \ \ \left\|\Delta^\ell_{h\boldsymbol{e}_i} f\right\|_{L^{p_i}} \leq C(f)|h|^{s_i}.$$

- It will take two papers to Dominique, Gérard and Oleg to solve the problem, with a *formidable* combination of approximation analysis and modifications of *Lepski's method*.

Anisotropy I

- Kerkyacharian, G., Lepski, O., Picard D. Nonlinear estimation in anisotropic multi-index denoising. *PTRF* (2001)
- Kerkyacharian, G., Lepski, O., Picard, D. Nonlinear estimation in anisotropic multiindex denoising. Sparse case. *Teor. Veroyatn. Primen.* (2007), translation in *Theory Probab. Appl.* (2008)
- An extension of dense and sparse zones are defined, according to the sign of

$$\frac{2}{p} - \sum_{i=1}^{d} s_i^{-1}(p_i^{-1} - \tfrac{1}{2}).$$

- For the dense case, we retrieve the expected rates:

$$\inf_{\widehat{f}_n} \sup_{\|f\|_{B^{(s_1,\dots,s_d)}_{(p_1,\dots,p_d),\infty}} \lesssim 1} \mathbb{E}_f\left[\|\widehat{f}_n - f\|^p_{L^p}\right] \approx n^{-sp/(2s+1)},$$

where $s^{-1} = \sum_{i=1}^d s_i^{-1}$ is the effective smoothness of the problem.

Anisotropy I

- In the sparse zone, the situation is more intricate, but similar phenomenae appear as in the homogeneous case. Minimaxity and adaptation (up to log terms in some cases) are fully covered.
- In some sense, the difficulties encountered in these papers (for adjusting *Lepski's method*, a cumbersome system of nonlinear equations for the bandwidths) will lead Oleg to revisit adaptive estimation.
- This somehow anticipates the *third period of adaptive estimation* for Oleg, with a new Sasha!

From anisotropy to the third adaptation period

Figure: Dominique, Oleg, Sasha *ca* 2010.

Anisotropy II

- Around 2005, Oleg has already started to collaborate with Sasha Goldenshluger on a new approach to *adaptive estimation*.
- Sacha Tsybakov and Oleg reunite to start their fourth common project. Anatoli Juditsky joins the party.
- They address an apparently innocuous problem of adaptive estimation, but it is probably a *new gateway to anisotropy*.
- The result is presented at the IMS Medallion Lecture of Oleg in Minneapolis, 2005.

Anisotropy II

Figure: Oleg (with S. Leonov and S. Efromovich) at the 2005 JSM, Minneapolis, where he presented *Nonparametric estimation of composite functions.*

Anisotropy II: intrinsic geometric smoothness

$$Y_\varepsilon = f + n^{-1/2}\dot{W} \ \text{ on } \ L^2([0,1]^d), \ \ d \geq 1.$$

- The problem of estimating f is now fully understood if $f \in B^{(s_1,\dots,s_d)}_{(p_1,\dots,p_d),\infty}$.
- For $f : [0,1]^d \to \mathbb{R}$ viewed as a (smooth) graph manifold, approximation properties heavily depend on the choice of coordinates.
- Typical examples include smooth image domain boundaries or certain solutions of nonlinear PDE's (recommended ref. is De Vore *et al.* Anisotropic smoothness spaces via level sets, *Comm. Pure Appl. Math.* 2008).

Anisotropy II

- Another example is given by the large population limiting processes $g(t,a)$ in age-structured populations:

$$\begin{cases} \frac{\partial}{\partial t} g(t,a) + \frac{\partial}{\partial a}\big(v(a) g(t,a)\big) + \mu(t,a) g(t,a) = 0, \\ g(0,a) = \phi(a), \ \ g(t,0) = \int_{\mathbb{R}_+} b(t,a) g(t,a) da. \end{cases}$$

- $\mu \in B_\infty^{s_1,s_2}$, $b \in B_\infty^{s_2,s_3}$ and g is observed in small noise.
- We have $g \in B_\infty^{\alpha,\beta}$ for some $\alpha(s_i)$ and $\beta(s_i)$.
- However, for some $G : \mathbb{R}^2 \to \mathbb{R}^2$ related to $v(\cdot)$ we have

$$g(t,a) = f \circ G(t,a)$$

where $f \in B_\infty^{\alpha_+,\beta_+}$ is such that $\alpha_+ \geq \alpha$ and $\beta_+ \geq \beta$.

Anisotropy II

- Juditsky, A. B., Lepski, O. V., Tsybakov, A. B. Nonparametric estimation of composite functions. *Ann. Statist.* (2009)
- Under the structure $g = f \circ G$ with $f : \mathbb{R} \to \mathbb{R}$ and $G : \mathbb{R}^d \to \mathbb{R}$, an estimator is constructed over isotropic Hölder classes.
- The structure encompasses many situations, including single-index and additive models.
- For $f \in B_\infty^\gamma$ and $G \in B_\infty^{\beta,\dots,\beta}$ with $\gamma, \beta > 0$, define

$$r_n^{\gamma,\beta} = \begin{cases} (n^{-1}\sqrt{\log n})^{\gamma/(2\gamma+1+(d-1)/\beta)}, & \beta > 1, \beta \geq d(\gamma-1)+1, \\ (n^{-1}\sqrt{\log n})^{1/(2+d/\beta)}, & \gamma > 1, \beta < d(\gamma-1)+1, \\ (n^{-1}\sqrt{\log n})^{2/(2+d/(\gamma\beta))}, & 0 < \gamma, \beta \leq 1. \end{cases}$$

Anisotropy II

- A *minimax lower bound* is obtained: for every $q, \gamma, \beta > 0$:

$$\inf_{\widehat{g}_n} \sup_{g \in \mathcal{S}(\gamma,\beta)} \mathbb{E}_f \left[\|\widehat{g}_n - g\|_{L^\infty}^q \right] \gtrsim (r_n^{\gamma,\beta})^{-q}.$$

- An *adaptive estimator* is constructed, that achieves the lower bound up to $\log\log n$ terms in some cases (over the second region in the definition of $r_n^{\gamma,\beta}$ if moreover $\gamma \leq \beta \leq 2$).
- We have $\mathcal{S}(\gamma,\beta) \subset B_{\infty,\infty}^{\gamma\beta}$ if $0 < \gamma, \beta \leq 1$ and $\gamma \wedge \beta$ otherwise.
- However, several regions allows for rate improvements thanks to the *underlying composite structure* of the model.
- Adaptation is done w.r.t. the local structure, *not* the smoothness indices (γ, β).

The third adaptation period: the oracle approach

- Goldenshluger, A., Lepski, O. Structural adaptation via $\mathbb{L}^p$-norm oracle inequalities. *PTRF* (2009),
- Goldenshluger, A. Lepski, O. Universal pointwise selection rule in multivariate function estimation. *Bernoulli* (2008),
- Goldenshluger, A., Lepski, O. Uniform bounds for norms of sums of independent random functions. *Ann. Probab.* (2011),
- Goldenshluger, A., Lepski, O. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.* (2011),
- Goldenshluger, A., Lepski, O. On adaptive minimax density estimation on $\mathbb{R}^d$. *PTRF* (2014).

The third adaptation period: the oracle approach

- Sasha and Oleg meet for the first time here in Luminy in 2000.
- In spring 2005, Oleg visits Sasha in Haïfa.
- They embark on an ambitious project about generalising and reformulating *adaptive estimation techniques*, motivated in part by the work with Dominique and Gérard on anisotropy.
- Around 2006, continuing ideas previously developed by Oleg with B. Levit, they progressively realise that they can base procedures *à la Lepski* by solving appropriate optimisation problems instead of comparisons of pseudo-estimators.
- This leads to the *Goldenshluger-Lepski procedure*.

Towards an end

Figure: A last tale... A. Pouchkine "Boris Godounov".

Happy birthday Oleg and Sacha!

THANK YOU FOR YOUR ATTENTION!

Acknowledgements: Thanks to Y.G, S.G., S.G., D.P. (and other anonymous sources) for helping me with pictures and memories.
Special thanks to Evgenia Magnien for helping me in Russian!