Adaptive estimation for inverse problems: a logarithmic effect in $L_2$

A.B. Tsybakov

Université Paris VI

March 27, 2000

Abstract

This note presents a method of adaptive estimation for a class of inverse problems with random noise. We show that the proposed method attains the best rate of adaptive estimation in $L_2$, but this rate is slower than the "non-adaptive" minimax rate by a logarithmic factor.

Mathematics Subject Classifications: 62G05, 62G20
Key Words: Adaptive curve estimation, Statistical inverse problems, Heavily ill-posed problems, Optimal rate of adaptation

1 Introduction

Assume that one observes a sequence of random variables $y = (y_1, y_2, \ldots)$ where

$$y_k = b_k \theta_k + \varepsilon \xi_k, \ k = 1, 2, \ldots,$$

(1)

$b_k > 0$ are known numbers, $\xi_k$ are i.i.d. standard gaussian random variables, $0 < \varepsilon < 1$ is the noise level and $\theta = (\theta_1, \theta_2, \ldots) \in l_2$ is an unknown sequence to be estimated.

The model (1) is used as a sequence space framework for statistical linear inverse problems where one wants to estimate the solution $f$ of the operator equation $g = Af$ from noisy observations of $g$ (see e.g. Korostelev and Tsybakov (1993), Donoho (1995), Mair and Ruymgaart (1996)). Here $A$ is a linear operator defined on a Hilbert space $X$ with scalar
product $(\cdot, \cdot)$ and $f \in X$. Then (1) is related to the singular value decomposition: $b_k^2$ are interpreted as the eigenvalues of the operator $A^* A$, while $b_k = (f, \varphi_k)$, where $A^*$ is the adjoint of $A$ and $\{ \varphi_k \}$ are the orthonormalized eigenfunctions of $A^* A$.

We assume that there exist $b_{\min} > 0, b_{\max} > 0, \beta > 0, q \in \mathbb{R}$, such that

$$b_{\min} k^2 e^{-\beta k} \leq b_k \leq b_{\max} k^2 e^{-\beta k}, \quad \forall k. \tag{2}$$

Condition (2) on the eigenvalues holds for several well-known problems of mathematical physics that can be called heavily ill-posed. We mention here the Cauchy problem for the Laplace equation in a rectangle, the inverse Dirichlet problem for the Laplace equation in the disk, the problem of satellite geodesy (see Golubev and Khasminskii (1999a, b), Pereverzev and Schock (1999) and the references therein).

The model (1) can be written in the equivalent form

$$\overline{y}_k = \theta_k + \varepsilon \frac{1}{b_k} \xi_k, \quad k = 1, 2, \ldots, \tag{3}$$

where $\overline{y}_k = y_k / b_k$, and can be also interpreted as the model with direct observations and correlated data (cf. Johnstone (1999)).

Assume that the sequence $\theta$ belongs to the following ellipsoid in $l_2$:

$$\Theta(\alpha, r, Q) = \left\{ \theta : \sum_{k=1}^{\infty} a_k^2 \theta_k^2 \leq Q \right\},$$

where $a_k = a_k(\alpha) = k^\alpha e^{\beta k}$ with $\alpha > 0, Q > 0$ and $r \in \mathbb{R}$.

Pinsker (1980) obtains linear minimax estimators of $\theta$ on $\Theta(\alpha, r, Q)$ in the model (3) for general $b_k$, in particular, for those satisfying (2). Golubev and Khasminskii (1999b) study exact asymptotic behavior on $\Theta(\alpha, r, Q)$ of the minimax risk among all estimators for $q = r = 1$. They show that linear minimax estimates as in Pinsker (1980) are not sharp minimax in this model. Ermakov (1989), Efroimovich (1997), Efroimovich and Kolchinskii (1998) study minimax estimation in statistical inverse problems for some related setups.

Here we consider the problem of adaptive estimation of $\theta$. We assume that the statistician does not know the values $\alpha, r, Q$. It is only known that there exist some (unknown) positive constants $\alpha_{\min}, \alpha_{\max}, r_{\max}, Q_{\max}$ such that $\alpha_{\min} \leq \alpha \leq \alpha_{\max}, |r| \leq r_{\max}$ and $0 < Q \leq Q_{\max}$.

We construct an estimator (called adaptive) such that it attains the fastest possible rate of convergence uniformly on the scale of classes $\{ \Theta(\alpha, r, Q), \alpha_{\min} \leq \alpha \leq \alpha_{\max}, |r| \leq r_{\max}, 0 < Q \leq Q_{\max} \}$. The most interesting fact is that the optimal rate of adaptation for this problem is slower (by a logarithmic factor) than the minimax "non-adaptive" rate. Such a logarithmic phenomenon in adaptation was known before only for the problems of estimation at a fixed point. It was first found by Leptskii (1990), and studied in the context of statistical inverse problems by Goldenshluger (1998), Goldenshluger and Pereverzev (1999). On the other hand, it was believed that such effects do not occur in $L_2$, where not only the rates but also
exact constants are often the same in adaptive and non-adaptive settings (cf. Efroimovich and Pinsker (1985), Cavalier and Tsybakov (1999)). The results below may be viewed as a counter-example showing that the logarithmic deterioration of adaptive rate may happen in $L_2$. This effect is based on the exponentiality of both $a_k$ and $b_k$. The key assumption is that the exponents of $a_k$ and $b_k$ contain the same powers of $k$. If this is not the case, or at least one of $a_k$ or $b_k$ has a power behavior rather than exponential one, the logarithmic factor in the adaptive rate disappears and the optimal rate of adaptation coincides with the usual minimax rate.

2 Results

For an arbitrary estimator $\hat{\theta}$ define the maximal risk

$$R_\varepsilon(\hat{\theta}, \alpha, r, Q) = \sup_{\hat{\theta} \in \Theta(\alpha, r, Q)} \mathbb{E}_\theta \|\hat{\theta} - \theta\|^2$$

where $\mathbb{E}_\theta$ is the expectation w.r.t. the distribution of $y$ satisfying (1), and $\|\cdot\|$ is the $l_2$-norm.

Let $\overline{\gamma}_k = y_k/b_k$. For any integer $N$ define the projection estimator

$$\hat{\theta}(N) = (\overline{\gamma}_1, \ldots, \overline{\gamma}_N, 0, 0, \ldots).$$

Set

$$\psi_\alpha(\varepsilon) = \left(\log \frac{1}{\varepsilon}\right)^{-\frac{2\alpha+\alpha}{\beta+3}} \varepsilon^{\beta/(\beta+\alpha)}$$

**Theorem 1** Let $\alpha > 0$, $r \in \mathbb{R}$, $Q > 0$, and let $N = N(\varepsilon) = \left[\frac{1}{\beta+3} \log \frac{1}{\varepsilon} + \frac{a^2}{\beta+3} \log \log \frac{1}{\varepsilon}\right]$ where $[\cdot]$ denotes the integer part. Then

$$\limsup_{\varepsilon \to 0} R_\varepsilon(\hat{\theta}(N(\varepsilon)), \alpha, r, Q) \psi_\alpha^{-2}(\varepsilon) \leq C < \infty, \quad (4)$$

and

$$\liminf_{\varepsilon \to 0} \inf_{\hat{\theta}} R_\varepsilon(\hat{\theta}, \alpha, r, Q) \psi_\alpha^{-2}(\varepsilon) \geq c > 0 \quad (5)$$

where $\inf_{\hat{\theta}}$ denotes the infimum over all estimators.

This theorem entails that $\psi_\alpha(\varepsilon)$ is the minimax rate of convergence for estimation of $\theta$ in the model (1) on the class $\Theta(\alpha, r, Q)$.

For the upper bound (4) note that, for $\varepsilon$ small enough, $a_k^2[N(\varepsilon)] \leq a_k^2$, $\forall k > N(\varepsilon)$, and thus

$$\mathbb{E}_\theta \|\hat{\theta}(N) - \theta\|^2 = \varepsilon^2 \sum_{k=1}^{N} b_k^2 + \sum_{k=N+1}^{\infty} \theta_k^2 \leq K \varepsilon^2 N^{-2\alpha} e^{2\beta N} + Q a_N^{-2} = O \left(\left(\log \frac{1}{\varepsilon}\right)^{-\frac{2\alpha+\alpha}{\beta+3}} \varepsilon^{2\beta/(\beta+\alpha)}\right).$$
as $\varepsilon \to 0$, where $K > 0$ is a constant that does not depend on $\varepsilon$ and $\theta$. The lower bound (5) is proved by reduction to the problem of testing 2 simple hypotheses. We omit the proof, since it is not far from the argument in Golubev and Khasminskii (1999b) where the case $q = r = 1$ is treated.

Note that the parameter $Q$ was not used in the construction of the rate optimal estimator $\hat{\theta}(N(\varepsilon))$. However, the knowledge of the parameters $\alpha$ and $r$ was crucial. If these parameters are not available consider an adaptive estimator of $\theta$. Denote

$$N_{\text{max}}(\varepsilon) = \left[ \frac{1}{\beta} \log \frac{1}{\varepsilon} \right].$$

(6)

It is clear that $N_{\text{max}}(\varepsilon) > N(\varepsilon)$ for $\varepsilon$ small enough.

Define the estimator:

$$\tilde{\theta} = \hat{\theta}(\hat{N}),$$

where $\hat{N}$ is given by a thresholding of Lepski type:

$$\hat{N} = \min \{1 \leq N \leq N_{\text{max}} : \|\hat{\theta}(N) - \hat{\theta}(N')\| \leq 10 \eta_N(\varepsilon), \forall N' \geq N, 1 \leq N' \leq N_{\text{max}}\},$$

with $\eta_N(\varepsilon) = \left(\varepsilon^2 \log \frac{1}{\varepsilon} \sqrt{\sum_{k=1}^{N} b_k^{-1}}\right)^{1/2}$. Denote

$$\tilde{\psi}_{\alpha}(\varepsilon) = \left(\log \frac{1}{\varepsilon}\right)^{-\frac{\alpha + \log \varepsilon}{2 \beta + \alpha}} \left(\varepsilon \sqrt{\log \frac{1}{\varepsilon}}\right)^{\alpha/(\beta + \alpha)}.$$

**Theorem 2** Let $W = \{(\alpha, r, Q) : \alpha_{\min} \leq \alpha \leq \alpha_{\max}, |r| \leq r_{\max}, 0 < Q < Q_{\max}\}$. Then

$$\limsup_{\varepsilon \to 0} \sup_{(\alpha, r, Q) \in W} R_\varepsilon(\tilde{\theta}, \alpha, r, Q)\tilde{\psi}_{\alpha}(\varepsilon) \leq C_{\text{max}} < \infty.$$

Thus we have the upper bound $\tilde{\psi}_{\alpha}(\varepsilon)$ on the rate of convergence of the adaptive estimator $\tilde{\theta}$ on $\Theta(\alpha, r, Q)$ which is worse than the minimax rate $\psi_{\alpha}(\varepsilon)$ by a logarithmic factor. The next theorem shows that $\tilde{\psi}_{\alpha}(\varepsilon)$ is, in fact, the optimal rate of adaptation: it cannot be improved even in the simplest case where $\alpha$ takes only two possible values.

**Theorem 3** For any $r > 0$, $Q > 0$ and any $\alpha_2 > \alpha_1 > 0$ we have

$$\liminf_{\varepsilon \to 0} \inf_{\tilde{\theta}} \sup_{\alpha \in \{\alpha_1, \alpha_2\}} \max_{r, Q} R_\varepsilon(\tilde{\theta}, \alpha, r, Q)\tilde{\psi}_{\alpha}(\varepsilon) \geq C_{\text{min}} > 0.$$
3 Proofs of the theorems

Note that
\[ \sum_{k=1}^{N} k^{-4q} e^{4\beta k} = \frac{e^{4\beta}}{e^{4\beta} - 1} N^{-4q} e^{4\beta N}(1 + o(1)), \quad N \to \infty. \]

Thus, there exist \( C_{b1} > 0, C_{b2} > 0 \) depending only on \( b_{\text{max}}, b_{\text{min}}, q, \beta \) such that
\[ C_{b1} N^{-4q} e^{4\beta N} \leq \sum_{k=1}^{N} b_k^{-4} \leq C_{b2} N^{-4q} e^{4\beta N}, \quad \forall N. \quad (7) \]

Similarly, there exists \( C_{b3} > 0 \) depending only on \( b_{\text{max}}, b_{\text{min}}, q, \beta \) such that
\[ \sum_{k=1}^{N} b_k^{-2} \leq C_{b3} N^{-2q} e^{2\beta N}, \quad \forall N. \quad (8) \]

**Proof of Theorem 2.** Consider the random event
\[ \mathcal{A} = \{ N_{1, \ldots, N_{\text{max}}} : \epsilon^2 \sum_{k=1}^{N} b_k^{-2} \xi_k^2 / \eta_{N, \epsilon}^2(\xi) \leq 17 \} \]

It is shown in the Appendix that
\[ \mathbb{P}(\overline{\mathcal{A}}) \leq N_{\text{max}} \epsilon^{1/4} \quad (9) \]

for \( \epsilon \) small enough. Fix \( \alpha \in [\alpha_{\text{min}}, \alpha_{\text{max}}] \) and denote
\[ N_* = \min\{ 1 \leq N \leq N_{\text{max}} : N^{-2q} e^{-2\alpha N} \leq C_0 \eta_N^2(\epsilon) \} \]

where \( C_0 > 0 \) is a constant that will be chosen later. Using (7) it is easy to show that, for \( \epsilon \) small enough,
\[ C_1 \log(1/\epsilon) \leq N_* \leq C_2 \log(1/\epsilon) \]

where \( C_1 > 0, C_2 > 0 \) depend only on \( b_{\text{max}}, b_{\text{min}}, q, \beta, r_{\text{max}}, \alpha_{\text{min}}, \alpha_{\text{max}}, Q_{\text{max}} \).

Assume that \( \theta \in \Theta(\alpha, r, Q) \) and that \( \mathcal{A} \) holds. Then, for any \( N \geq N_* \), we get
\[ ||\hat{\theta}(N) - \theta||^2 \leq \epsilon^2 \sum_{k=1}^{N} b_k^{-2} \xi_k^2 + Q_{\text{max}} \max_{k \geq N} a_k^{-2}(\alpha) \]
\[ \leq 17\eta_{N, \epsilon}^2(\xi) + Q_{\text{max}} N^{-2} e^{-2\alpha N} \]

where we assumed that \( \epsilon \) is small enough to have monotone increasing \( a_k \) for \( k \geq N_* \). Since, in addition, \( \eta_N(\epsilon) \) is monotone increasing in \( N \), and using the definition of \( N_* \), we find
\[ N^{-2q} e^{-2\alpha N} \leq C_0 \eta_N^2(\epsilon), \quad \forall N \geq N_. \]
This and (11) yield, for \( \varepsilon \) small enough,
\[
||\hat{\theta}(N) - \theta||^2 \leq (17 + Q_{\max}C_0)\eta_{\tilde{N}}^2(\varepsilon), \quad \forall N \geq N_*,
\]
provided \( \theta \in \Theta(\alpha, r, Q) \) and \( \mathcal{A} \) holds. Choose \( C_0 = 8/Q_{\max} \). Then
\[
||\hat{\theta}(N) - \theta||^2 \leq 25\eta_{\tilde{N}}^2(\varepsilon), \quad \forall N \geq N_*.
\]
Consequently,
\[
||\hat{\theta}(N) - \hat{\theta}(N_*)|| \leq ||\hat{\theta}(N) - \theta|| + ||\hat{\theta}(N_*) - \theta|| \leq 10\eta_{\tilde{N}}(\varepsilon), \quad \forall N \geq N_*,
\]
and thus, by definition of \( \hat{N} \), we have \( N_* \geq \hat{N} \). This entails, again by definition of \( \hat{N} \), that
\[
||\hat{\theta}(\hat{N}) - \hat{\theta}(N_*)|| \leq 10\eta_{\varepsilon}(\varepsilon) \quad \text{where} \quad \eta_{\varepsilon}(\varepsilon) = \eta_{N_*}(\varepsilon).
\]
Now, applying (13) with \( N = N_* \) and the last inequality we conclude that
\[
||\hat{\theta} - \theta|| = ||\hat{\theta}(\hat{N}) - \theta|| \leq ||\hat{\theta}(\hat{N}) - \hat{\theta}(N_*)|| + ||\hat{\theta}(N_*) - \theta|| \leq 15\eta_{\varepsilon}(\varepsilon),
\]
provided \( \theta \in \Theta(\alpha, r, Q) \) and \( \mathcal{A} \) holds. Using this, (9), (6) and Lemmas 1, 2 from the Appendix we get, for \( \varepsilon \) small enough,
\[
\sup_{\hat{\theta} \in \Theta(\alpha, r, Q)} E_{\theta}||\hat{\theta} - \theta||^2 \leq (15\eta_{\varepsilon}(\varepsilon))^2 + \sup_{\hat{\theta} \in \Theta(\alpha, r, Q)} \left( E_{\theta}||\hat{\theta} - \theta||^4 \right)^{1/2} P^{1/2}(\mathcal{A})
\leq 225\varepsilon^2 \log \frac{N_*}{\varepsilon} + \sqrt{N_{\max}} \varepsilon^2 \sup_{(\alpha, r, Q) \in \mathcal{W}} \sup_{\hat{\theta} \in \Theta(\alpha, r, Q)} \left( \sum_{N=1}^{N_{\max}} E_{\theta}||\hat{\theta}(N) - \theta||^4 \right)^{1/2}
\leq C_3(\varepsilon^2 + \varepsilon^2 \log C_4(1/\varepsilon))
\]
where the constants \( C_3 > 0, C_4 > 0 \) do not depend on \( \alpha, r, Q \). Since
\[
\sup_{(\alpha, r, Q) \in \mathcal{W}} \frac{\varepsilon^2 \log C_4(1/\varepsilon)}{\psi_{\alpha}^2(\varepsilon)} = O(1), \quad \varepsilon \to 0,
\]
Theorem 2 is proved.

**Proof of Theorem 3.** Denote \( \theta^{(0)} = (0, 0, \ldots) \) the zero sequence and \( \theta^{(1)} = (0, \ldots, 0, v, 0, \ldots) \) the sequence that has all elements 0 except for the \( m \)th element which equals \( v = v_0 \psi_{\alpha_1}(\varepsilon) \).

Here
\[
m = \left[ \frac{1}{\beta + \alpha_1} \log \frac{1}{\varepsilon} + \frac{1}{\beta + \alpha_1} \left( q - r - \frac{1}{2} \right) \log \log \frac{1}{\varepsilon} \right]
\]
and \( v_0 \) is a positive number. We choose \( v_0 \) small enough to guarantee that \( \theta^{(1)} \in \Theta(\alpha_1, r, Q) \).
In fact, \( \theta^{(1)} \in \Theta(\alpha_1, r, Q) \) if
\[
v^2 m^2 \varepsilon^{2m} \leq Q.
\]
Let us check this inequality. We have, for $\varepsilon$ small enough
\[
m \leq \frac{1}{\beta + \alpha_1} \log \frac{1}{\varepsilon} + \frac{1}{\beta + \alpha_1} \left( q - r - \frac{1}{2} \right) \log \log \frac{1}{\varepsilon} \leq K_1 \log \frac{1}{\varepsilon}
\]
(14)
where $K_1 = (3/2 + |q| + |r|)/\beta$. Hence
\[
v^2 m^2 \varepsilon^{2m} v_{\alpha\alpha}(\varepsilon) \left( K_1 \log \frac{1}{\varepsilon} \right)^{2\varepsilon} \exp \left( \frac{2\alpha_1}{\beta + \alpha_1} \log \frac{1}{\varepsilon} + \frac{2\alpha_1}{\beta + \alpha_1} \left( q - r - \frac{1}{2} \right) \log \log \frac{1}{\varepsilon} \right)
\]
\[
= v_0^2 K_1^{2\varepsilon} \leq Q
\]
for $v_0$ small enough. This shows that $\theta^{(1)} \in \Theta(\alpha_1, r, Q)$ for $v_0$ small enough. Denote $P_0$, $P_1$ the probability measures generated by observations (1) with $\theta = \theta^{(0)}$ and $\theta = \theta^{(1)}$ respectively and let $E_j = E_{\theta(\delta)}$, $j = 0, 1$. Then, for any estimator $\hat{\theta}$,
\[
\max_{\alpha \in (\alpha_1, \alpha_2)} \sup_{\delta \in \Theta(\alpha_1, \alpha_2)} E_0[||\hat{\theta} - \theta||^2 \psi_{\alpha}(\varepsilon)] \geq \max \left\{ E_0[||\hat{\theta} - \theta^{(0)}||^2 \psi_{\alpha}(\varepsilon)], E_1[||\hat{\theta} - \theta^{(1)}||^2 \psi_{\alpha}(\varepsilon)] \right\}
\]
\[
= v_0^2 \max \left\{ E_0(\lambda^2 d^2(\hat{\theta}, \theta^{(0)})), E_1(\lambda^2 d^2(\hat{\theta}, \theta^{(1)})) \right\}
\]
(15)
where $d(\theta, \theta') = ||\theta - \theta'|| v_0^{-1} \psi_{\alpha}^{-1}(\varepsilon)$ and $\lambda = \varepsilon^{-\gamma} \left( \sqrt{\log \frac{1}{\varepsilon}} \right)^{\gamma_1}$, with $\gamma = \frac{\beta(\alpha_2 - \alpha_1)}{\beta + \alpha_1} > 0$, and $\gamma_1 = \frac{\beta(\alpha_2 - \alpha_1)}{(\beta + \alpha_1)(\beta + \alpha_2)} (r - q + \frac{1}{2})$. Apply now Theorem 6(i) of Tsybakov (1998), with $M = 1$, $u(t) = t^2$, $\delta = 1/4$. Note that $d(\theta^{(0)}, \theta^{(1)}) = 1$, which yields (A.1) of Tsybakov (1998). Next, check (A.2) of Tsybakov (1998). Set $\tau = \varepsilon^\gamma$. Then, denoting $\zeta$ a standard normal random variable and assuming that $v_0$ is small enough, we get
\[
P_1 \left( \frac{dP_0}{dP_1} \geq \tau \right) = P \left( \exp \left( \zeta \varepsilon^{-1} b_m v - \frac{\varepsilon^{-2} b_m^2 v^2}{2} \right) \geq \tau \right) \geq P \left( \zeta \geq -\sqrt{\frac{\gamma}{2} \log \frac{1}{\varepsilon}} \right) \rightarrow 1,
\]
as $\varepsilon \to 0$. To prove the last inequality note first that
\[
m \geq \frac{1}{\beta + \alpha_1} \log \frac{1}{\varepsilon} + \frac{1}{\beta + \alpha_1} \left( q - r - \frac{1}{2} \right) \log \log \frac{1}{\varepsilon} - 1.
\]
This and (14) yield
\[
\varepsilon^{-1} b_m v \leq b_{\text{max}} v \varepsilon^{-1} m^q \varepsilon^{-\beta m} \psi_{\alpha}(\varepsilon)
\]
\[
\leq b_{\text{max}} v_0 \varepsilon^{-1} \left( K_1 \log \frac{1}{\varepsilon} \right)^{\gamma} \exp \left( -\frac{\beta}{\beta + \alpha_1} \log \frac{1}{\varepsilon} - \frac{\beta}{\beta + \alpha_1} \left( q - r - \frac{1}{2} \right) \log \log \frac{1}{\varepsilon} + \beta \right) \psi_{\alpha}(\varepsilon)
\]
\[
= K_2 v_0 \sqrt{\log \frac{1}{\varepsilon}}
\]
where $K_2 = b_{\max} e^{\beta} K_1^\gamma$. Hence, for $v_0$ small enough, so that $K_2^2 v_0^2 < \gamma/2$, we obtain
\[
P \left( \exp \left( \frac{\zeta \varepsilon^{-1} b_m v - \varepsilon^{-2} b_m^2 v^2}{2} \right) \geq \tau \right) = P \left( \zeta \varepsilon^{-1} b_m v - \frac{\varepsilon^{-2} b_m^2 v^2}{2} \geq -\gamma \log \frac{1}{\varepsilon} \right)
\]
\[
\geq P \left( \zeta \varepsilon^{-1} b_m v \geq -\frac{\gamma}{2} \log \frac{1}{\varepsilon} \right) \geq P \left( \zeta \geq -\frac{\gamma}{2K_2 v_0} \sqrt{\log \frac{1}{\varepsilon}} \right)
\]
\[
\geq P \left( \zeta \geq -\sqrt{\frac{\gamma}{2} \log \frac{1}{\varepsilon}} \right).
\]

Hence, for $\varepsilon$ small enough, $P_1 \left( \frac{d\hat{\theta}}{d\hat{A}} \geq \tau \right) \geq 1/2$, which yields (A.2) of Tsybakov (1998). Thus, we can apply (A.3) of Tsybakov (1998), and we get, for $\varepsilon$ small enough,
\[
\max \{ E_0(\lambda_2 d^2(\hat{\theta}, \theta^{(0)})), E_1(d^2(\hat{\theta}, \theta^{(1)})) \} \geq \tau \lambda_2^2 / (32 + 8 \tau \lambda^2) \rightarrow 1/8,
\]
as $\varepsilon \to 0$, since $\tau \lambda^2 \to \infty$. This and (15) prove Theorem 3.

APPENDIX

Proof of (9).
We have
\[
P(\mathcal{A}) \leq N_{\max} \max_{N=1, \ldots, N_{\max}} \mathbb{P} \left\{ \varepsilon^2 \sum_{k=1}^{N} b_k^{-2} \xi_k^2 > 17n_{\lambda}^2(\varepsilon) \right\}
\]

Note that for all $N$ and all $t > 0$
\[
P \left\{ \sum_{k=1}^{N} b_k^{-2} (\xi_k^2 - 1) > t \right\} \leq \exp(-ht) \prod_{k=1}^{N} \mathbb{E} \exp(hb_k^{-2}(\xi_k^2 - 1))
\]
\[
= \exp(-ht) \prod_{k=1}^{N} g(hb_k^{-2})
\]

where $g(x) = e^{-x}(1 - 2x)^{-1/2}$ and it is assumed that $h \max_{1 \leq k \leq N} b_k^{-2} < 1/4$. Note that $g(x) \leq \exp(4x^2)$, $\forall 0 < x < 1/4$, and thus
\[
P \left\{ \sum_{k=1}^{N} b_k^{-2} (\xi_k^2 - 1) > t \right\} \leq \exp(-ht + 4h^2 \sum_{k=1}^{N} b_k^{-4} t).
\]
Let \( h = \left(4\sqrt{\sum_{k=1}^{N} b_k^{-4}}\right)^{-1}. \) Then \( h \max_{1 \leq k \leq N} b_k^{-2} < 1/4 \) and we have

\[
P\{\sum_{k=1}^{N} b_k^{-2}(\xi_k^2 - 1) > t\} \leq \exp\left(-\frac{t}{4\sqrt{\sum_{k=1}^{N} b_k^{-4}}} + \frac{1}{4}\right).
\]

(17)

Now, (7) entails that for \( \varepsilon \) small enough

\[
\log(1/\varepsilon) \sqrt{\sum_{k=1}^{N} b_k^{-4}} \geq \sum_{k=1}^{N} b_k^{-2}, \quad \forall N.
\]

Using this and (17) we find

\[
P\{\varepsilon^2 \sum_{k=1}^{N} b_k^{-2} \xi_k^2 > 17r_N^2(\varepsilon)\}
\]

\[
= P\left\{\sum_{k=1}^{N} b_k^{-2}(\xi_k^2 - 1) > 17 \log(1/\varepsilon) \sqrt{\sum_{k=1}^{N} b_k^{-4} - \sum_{k=1}^{N} b_k^{-2}}\right\}
\]

\[
\leq P\left\{\sum_{k=1}^{N} b_k^{-2}(\xi_k^2 - 1) > 16 \log(1/\varepsilon) \sqrt{\sum_{k=1}^{N} b_k^{-4}}\right\}
\]

\[
\leq e^{1/4} \exp(-4 \log(1/\varepsilon)) = e^{1/4} \varepsilon^4.
\]

This proves (9).

**Lemma 1** There exist constants \( C_5 > 0, C_6 > 0 \) such that

\[
\sup_{(\omega, Q) \in W} \sup_{\theta \in \Theta(\omega, Q)} \left(\sum_{N=1}^{N_{\max}} \mathbb{E}_{\theta}||\hat{\theta}(N) - \theta||^4\right)^{1/2} \leq C_5 \log^{C_6}(1/\varepsilon).
\]

**Proof.**

\[
||\hat{\theta}(N) - \theta||^4 = (\varepsilon^2 \sum_{k=1}^{N} b_k^{-2} \xi_k^2 + \sum_{k=N+1}^{\infty} \theta_k^2)^2
\]

\[
\leq 2((\varepsilon^2 \sum_{k=1}^{N} b_k^{-2} \xi_k^2)^2 + (\sum_{k=N+1}^{\infty} \theta_k^2)^2)
\]

\[
\leq 2((\varepsilon^2 \sum_{k=1}^{N} b_k^{-2} \xi_k^2)^2 + Q_{\max}^2 \max_{k > N} k^{-2r} e^{-2ok}).
\]
Let $\varepsilon$ be small enough to have monotonicity of $k^{-2r} e^{-2\alpha N}$ for $k \geq N_{\text{max}}$, whatever are $\alpha \in [\alpha_{\text{min}}, \alpha_{\text{max}}]$, $|r| \leq r_{\text{max}}$. Then

$$\max_{k > N} k^{-2r} e^{-2\alpha N} \leq N_{\text{max}}^{-2r} e^{-2\alpha N_{\text{max}}} + \max_{k < N_{\text{max}}} k^{-2r} e^{-2\alpha k} \leq N_{\text{max}}^{-2r} + \max_{k < N_{\text{max}}} k^{-2r} \leq 2N_{\text{max}}^{2r_{\text{max}}}.$$  

Using this we get

$$||\hat{\theta}(N) - \theta||^4 \leq 2(\varepsilon^2 \sum_{k=1}^{N} b_k^{-2} \xi_k^2)^2 + 2Q_{\text{max}}^2 N_{\text{max}}^{2r_{\text{max}}}. \quad (18)$$

It is easy to see that

$$\mathbb{E}\{(\varepsilon^2 \sum_{k=1}^{N} b_k^{-2} \xi_k^2)^2\} = \varepsilon^4 \sum_{k,j=1}^{N} b_k^{-2} b_j^{-2} \mathbb{E}(\xi_k^2 \xi_j^2)$$

$$\leq 3\varepsilon^4 \sum_{k,j=1}^{N} b_k^{-2} b_j^{-2} = 3\varepsilon^4 \sum_{k=1}^{N} b_k^{-2} \leq 3C_{b_3}^2 \varepsilon^4 N^{-4} \leq 3C_{b_3}^2 \varepsilon^4 N_{\text{max}}^{4r_{\text{max}}},$$

where we used (8) in the last inequality. This and (6) yield

$$\mathbb{E}\{(\varepsilon^2 \sum_{k=1}^{N} b_k^{-2} \xi_k^2)^2\} \leq 3C_{b_3}^2 N^{4r_{\text{max}}} \leq 3C_{b_3}^2 N_{\text{max}}^{4r_{\text{max}}}. \quad (19)$$

This, together with (18) and (6) proves the lemma.

**Lemma 2** There exists a constant $C_7 > 0$ such that for any $(\alpha, r, Q) \in W$ we have

$$\varepsilon^2 \log \frac{1}{\varepsilon} \sqrt{\sum_{k=1}^{N_*} b_k^{-4}} \leq C_7 \psi_0^2(\varepsilon).$$

**Proof.** By definition of $N_*$,

$$(N_* - 1)^{-2r} e^{-2\alpha (N_* - 1)} > C_0 \varepsilon^2 \log \frac{1}{\varepsilon} \sqrt{\sum_{k=1}^{N_* - 1} b_k^{-4}}.$$

This and (7) entail

$$(N_* - 1)^{-2r - q} e^{-2(\alpha + \beta) (N_* - 1)} > C_0 \sqrt{C_{b_1} \varepsilon^2 \log \frac{1}{\varepsilon}}. \quad (20)$$
Using (7) and then (20) we get
\[
\sqrt{\sum_{k=1}^{N_h} b_k^{-4}} \leq \sqrt{C_{b_2} N_*^{-2q_2} e^{2\beta N_*}} = \sqrt{C_{b_2} e^{2\beta N_* - 2q_2} e^{2\beta (N_*-1)}}
\leq \sqrt{C_{b_2} e^{2\beta (C_0\sqrt{C_{b_1}})^{-3/(\alpha+3)} N_*^{-2q_2} (N_*-1)^{-2(\alpha-1)/\alpha}} (e^{2 \log \frac{1}{\varepsilon}})^{-\beta/(\alpha+3)}
\]
Combining this and (10) we finally get
\[
\varepsilon^2 \log \frac{1}{\varepsilon} \sqrt{\sum_{k=1}^{N_h} b_k^{-4}} \leq C_8 (\varepsilon^2 \log \frac{1}{\varepsilon})^{2/(\alpha+3)} N_*^{-2(\beta+2\gamma)/\beta+3} \leq C_7 \psi^2_0(\varepsilon)
\]
where $C_7 > 0, C_8 > 0$ depend only on $b_{\max}, b_{\min}, q, \beta, r_{\max}, \alpha_{\min}, \alpha_{\max}, Q_{\max}$. This proves the lemma.

**Acknowledgement.** This research was supported by Centre for Mathematics and its Applications of Australian National University.

**References**


ALEXANDRE TSYBAKOV
LABORATOIRE DE PROBABILITÉS ET MODÈLES ALÉATOIRES
UNIVERSITÉ PARIS 6
4, pl. Jussieu, BP 188, 75252 PARIS, FRANCE
EMAIL: TSYBAKOV@CCR.JUSSIEU.FR