Probab. Theory Relat. Fields **113**, 421–441 (1999)

Probability Theory and Related Fields

Statistical approach to some ill-posed problems for linear partial differential equations

Pao-Liu Chow[1,⋆], Ildar A. Ibragimov[2,⋆], Rafail Z. Khasminskii[1,⋆]

[1] Department of Mathematics, Wayne State University, Detroit, MI 48202, USA.
e-mail: rafail@math.wagne.edu
[2] St. Peterburg Branch of Steklov Mathematical Institute, 27 Fontanka, St. Peterburg 191011, Russia

Received: 5 May 1997 / Revised version: 18 June 1998

Abstract. For linear partial differential equations, some inverse source problems are treated statistically based on nonparametric estimation ideas. By observing the solution in a small Gaussian white noise, the kernel type of estimators is used to estimate the unknown source function and its partial derivatives.. It is proved that such estimators are consistent as the noise intensity tends to zero. Depending on the principal part of the differential operator, the optimal asymptotic rate of convergence is ascertained within a wide class of risk functions in a minimax sense.

Mathematics Subject Classification (1991): 60G, 62G, 35R

1. Introduction

According to Hadamard [9], an initial-boundary value problem for partial differential equations is said to be well-posed if the equation has a unique solution which depends continuously on its coefficients and the initial-boundary data. As a reasonable physical model, the well-posedness reflects the desired robustness of the direct problem. However, in many engineering and physical applications, it is of paramount importance to consider the so called inverse problem. Given the usually imprecise information about

⋆ This work was supported in part by the US ONR Grant N00014-95-1-0793 and by the NSF Grant DMS9600245.

the solution, one is required to determine the unknown coefficients and/or the initial-boundary data. Such problems are usually not well posed or ill posed, since a small error in the solutions may result in a large deviation from the true values of the unknown functions to be determined. This lack of continuous dependence is the major source of difficulties for most inverse problems. To alleviate these difficulties, there exist many deterministic approaches to this sort of problems. For instance, the method of quasi-solutions [13] and the Tikhonov regularization method [19]. Alternatively, a statistical approach to such problems by introducing data noise has been proposed by a number of authors, including Sudakov and Khaflin [18], Franklin [5], Wahba [20], O'Sullivan [14], and Ermakov [4]. Some modern nonparametric approach to the inverse problems of deconvolution type can be found in the monographs [1], [8].

The paper is concerned with a statistical approach to the source determination problems governed by general linear partial differential equations. As an example, consider the wave equation for the vibration of an elastic medium in domain $D \in R^3$:

$$\frac{\partial^2 u}{\partial t^2} = c^2 \Delta u + \Theta(t, x), t > 0, \; x \in D \tag{1.1}$$

subject to some given initial-boundary conditions, in which $c > 0$ is a constant, $\Delta = \sum_{i=1}^{3} \frac{\partial^2}{\partial x_i^2}$ is the Laplacian and Θ is the unknown source function. The problem consists of determining the source Θ by observing the solution u in D over a time interval $0 \le t \le T$. Even with very accurate solution data, the determination of Θ from (1.1) faces the well known difficulty in numerical differentiation, which is extremely unstable. In this paper we shall present a statistical approach which generalizes the nonparametric estimation procedure as proposed in [2] for the case of ordinary differential equations. In general, for convenience, we will not distinguish the space and time variables. Consider the problem of estimating the source function $\Theta(x)$, $x \in D \subset R^d$, which is the inhomogeneous term of the equation:

$$\mathscr{L}(x)u = \Theta(x) \; , \tag{1.2}$$

where $\mathscr{L}(x)$ is a linear partial differential operator and u satisfies some appropriate side conditions. As to be seen, in our approach, the knowledge of such conditions is not essential in the estimation of $\Theta(x)$ for $x \in D$ being away from the boundary ∂D. Even the uniqueness of a solution is essential only for the lower bounds. In fact, for given Θ, we assume only that the equation (1.2) has a weak solution u which is observed in the presence of a Gaussian white noise of small intensity ε. The problem is to find a best estimator $\widehat{\Theta}_\varepsilon$ for Θ in some sense and to examine its performance. More precisely the main goal here is to construct a certain type of estimators with the best rate of convergence, as $\varepsilon \to 0$, for the risks in a minimax sense under various conditions on the operator $\mathscr{L}(x)$ and the a priori known smoothness of Θ. In our statistical approach, the nonparametric estimation based on the kernel smoothing technique is adopted. This technique was proposed by Parzen [15] and Rosenblatt [16] for density and regression estimation. It was applied to the estimation of a signal and its derivatives with white noise in one dimension [10], [11]. We shall consider two classes of linear differential operators of order n, depending on whether the principal part $\mathscr{L}_n(x)$ of $\mathscr{L}(x)$ is deficient (for definition, see section 3) or otherwise. In either case we obtain the upper and lower bounds for the risks and prove that the proposed kernel smoothing estimators for Θ and its derivatives converge respectively, as $\varepsilon \to 0$, to their true values at each interior point of D with an optimal rate in the minimax sense. The main results are summarized in Theorems 4.1, 4.2 and 5.1. It will also be shown by several examples that our results can be applied to many well-known equations in mathematical physics of elliptic, hyperbolic and parabolic types to ascertain their respective optimal rates of convergence. To be specific, the paper is organized as follows. In section two, the estimation problem is formulated and stated precisely. Section three contains some technical lemmas and two basic theorems which are essential for proving the main theorems to follow. The source estimation problem for the nondeficient case is treated in section four, where the main results on the upper and lower bounds for estimators are proved in Theorem 4.1 and Theorem 4.2, respectively. Some improved results and examples are also presented there. The analogous results for the deficient case are summarized as Theorem 5.1 in section five with some examples and remarks. In conclusion, several general remarks on the source estimation problems are provided in section six.

2. Statement of problem

Let $\mathscr{L} = \mathscr{L}(x)$ be a linear partial differential operator of order n for $x = \left(x^1, \ldots, x^d\right)$ in domain $D \in R^d$ with smooth boundary ∂D. We assume that the function $u(x)$, $x \in D$, satisfies the equation (1.2) in a weak sense so that, for any $\varphi \in C_K^\infty(D)$,

$$\left(u, \mathscr{L}^*\varphi\right) = (\Theta, \varphi) \tag{2.1}$$

where $C_K^\infty(D)$ denotes the space of infinite-time continuously differentiable functions with compact support $K \subset D$, $\mathscr{L}^* = \mathscr{L}^*(x)$ the formal adjoint operator of $\mathscr{L}$ [6] and $(\cdot, \cdot)$ is the inner product in $L_2\,(D)$.

For a fixed x_0 in D, let $O(x_0) \subset D$ be a ball centered at x_0 of some radius $\rho > 0$. Denote by $\sum(\beta, L, O(x_0)) = \sum(\beta_1, \ldots, \beta_d, L, O(x_0))$ the class of functions g on D having k_i derivatives with respect to the component x^i of x in $O(x_0)$, for $i = 1, \ldots, d$, and satisfying

$$\left|\partial_i^{k_i} g\big(x_h^i\big) - \partial_i^{k_i} g(x)\right| \leq L\, |h|^{\alpha_i}, 0 < \alpha_i \leq 1, \beta_i = k_i + \alpha_i \ ,$$

for all $x, x_h^i \in O(x_0)$ with $i = 1, \ldots, d$ and some constant $L > 0$, where $\partial_i^{k_i} = (\partial/\partial x^i)^{k_i}, x_h^i = x + he_i$, with $h \in R$ and e_i being the unit vector in the x^i- direction.

We suppose that the solution u of (1.2) (or (2.1)) is observed in the presence of a Gaussian white noise of intensity ε. This means specifically that the observed field V_ε is a random measure taking the form

$$V_\varepsilon(dx) = u(x)\, dx + \varepsilon W(dx) \ , \tag{2.2}$$

or, for any Borel set $\Gamma \subset D$,

$$V_\varepsilon(\Gamma) = \int_\Gamma u(x)\, dx + \varepsilon W(\Gamma) \tag{2.3}$$

where $W(\cdot)$ is a Gaussian white noise in R^d: a Gaussian orthogonal random measure on R^d (see [17], [21]) with $EW(\Gamma) = 0$ and

$$EW(\Gamma)W(\Gamma') = m(\Gamma \cap \Gamma')$$

for any Borel sets Γ and Γ', where m denotes the Lebesgue measure. It is known that for any $f \in L_2(R^d)$ the stochastic integral

$$I(f) = \int_{R^d} f(x) W(dx) \tag{2.4}$$

is well defined and has the properties

$$EI(f) = 0 \ ,$$

$$EI(f)I(g) = (f, g), \quad \text{for } \ f, g \in L_2(R^d) \ . \tag{2.5}$$

Here and henceforth (., .) and $\|\cdot\|$ denote the inner product and the L_2 norm in R^d or in D, as the case may be.

Suppose that, for any $\varphi \in L_2(D)$ a statistician can observe

$$\int_D \varphi(x) V_\varepsilon(dx) = \int_D \varphi(x) u(x)\, dx + \varepsilon \int_D \varphi(x) W(dx) \tag{2.6}$$

which is, of course, equivalent to the observation equation (2.3).

Now we state the problem: Given the prior information:

$$\Theta \in \sum (\beta_1, \ldots, \beta_d, L, O(x_0)) \quad ,$$

find the best estimator for $\Theta(x_0)$, in some sense to be made precise later, based on the observation (2.2) or (2.3). This problem will be studied in steps. The first step is to consider in section three, the simplest problem of estimating $u(x)$ and its derivatives in equation (2.2). Then we shall take up the general problem for two distinct cases in section four and section five separately, corresponding to the different type of $\mathscr{L}$ being considered.

3. Preliminary results

Before proceeding to the general problem, let us consider the problem of estimating $u(x_0)$ and its derivatives on the basis of the observation equation (2.2). The preliminary results for this simple problem are essential for the subsequent analysis. This signal-plus-noise model is very close to the one considered in [11], which dealt with the estimation of $u(\cdot)$ and its derivatives in different metrics for the one dimensional case. Here, in contrast, we are interested in the estimation of the functional $u(x_0)$ and its derivatives only.

Clearly the observation (2.2) generates a Gaussian measure $P_u^\varepsilon(\cdot)$ on the space of linear functionals

$$\ell(V_\varepsilon)(\varphi) = \int_D \varphi(x) V_\varepsilon(dx), \varphi \in L_2(D) \quad , \tag{3.1}$$

as defined by (2.6). In view of (2.5) the measure P_u^ε has the mean $E\ell(V_\varepsilon)(\varphi) = (\varphi, u)$ and the identity covariance operator. So the measures P_u^ε and P_0^ε are absolutely continuous (see e.g. Chap. 7 in [7]) if and only if $u \in L_2(D)$ and the Radon-Nikodym derivative is given by

$$\frac{dP_u^\varepsilon}{dP_0^\varepsilon}(V_\varepsilon) = \exp\left\{\frac{1}{\varepsilon}\int_D u(x)W(dx) - \frac{1}{2\varepsilon^2}\|u\|^2\right\} \quad , \tag{3.2}$$

where P_0^ε is the Wiener measure for $\varepsilon W(\cdot)$. Therefore the family of measures P_u^ε on $L_2(D)$ satisfies the local asymptotic normality (LAN) condition in L_2 in the sense of [12] with normalizing factor ε.

To estimate $u(x_0)$ and its derivatives, it is assumed that u is locally smooth such that

$$\partial_x^m u(x) = \partial_1^{m_1}\partial_2^{m_2}\ldots\partial_d^{m_d} u(x) \in \sum (\beta, L, O(x_0)) \tag{3.3}$$

Denote by G_ℓ the set of $(\ell + 1)$-time continuously differentiable kernels g on R with compact support such that the following properties hold:

$$\int_R g(r)\, dr = 1 \quad \text{and} \quad \int_R r^i g(r)\, dr = 0,\ i = 1, \dots, \ell \ . \tag{3.4}$$

Let $\delta = (\delta_1, \dots, \delta_d)$ with $\delta_i > 0$ for each i and $g_i \in G_{k_i}$. Define the kernel

$$g_\delta(x) = \prod_{i=1}^{d} \delta_i^{-1} g_i \left(\frac{x^i}{\delta_i} \right) \ , \tag{3.5}$$

and introduce the following kernel estimator of $\partial^m u\,(x) = \partial_x^m u(x)$:

$$\hat{u}_m^\varepsilon(x) = \int_D \partial_x^m g_\delta\,(x - y)\, V_\varepsilon\,(dy) \ \ . \tag{3.6}$$

Then, noting (2.6) and (3.5), the estimation error can be written as

$$\triangle\hat{u}_m^\varepsilon(x) \doteq \hat{u}_m^\varepsilon(x) - \partial^m u(x) = \tag{3.7}$$

$$\int_D g_\delta\,(x - y) \left[\partial_y^m u\,(y) - \partial_x^m u(x)\right] dy + \varepsilon \int_D \partial_x^m g_\delta\,(x - y)\, W\,(dy) \ \ .$$

In the next two sections, we often need the following two elementary lemmas. The proof of the first one uses the Taylor formula in each variable and (3.4). The proof of the second one is straightforward. Details are therefore omitted (for one-dimensional case see in [11]).

Lemma 3.1. *Let $g_\delta(x)$ be defined by* (3.4) *with $g_i \in G_{k_i}$, and let $f \in \sum(\beta, L, O\,(x_0))$, with $\delta_i \in (0, 1]$ and $\beta_i = k_i + \delta_i$, for $i = 1, \dots, d$. Then there exists $C > 0$ such that*

$$\left| \int_D g_\delta\,(x - y)\,[f\,(y) - f\,(x)]\, dy \right| \le C \sum_{i=1}^{d} \delta_i^{\beta i},\ x \in O\,(x_o) \ \ , \tag{3.8}$$

for some sufficiently small $|\delta|$, where, for any closed set $U \subset O(x_0)$, the constant C can be chosen independent of $x \in U$.

Lemma 3.2. *For the kernel function $g_\delta(x)$ defined by* (3.5), *the following holds*

$$E \left[\int \partial_x^m g_\delta(x - y) W(dy) \right]^2 = \left\| \partial_x^m g_\delta(x - \cdot) \right\|^2$$

$$\le C \prod_{i=1}^{d} \delta_i^{-(2m_i + 1)} \tag{3.9}$$

for some $C > 0$.

With the aid of the basic lemmas, the following theorem can be proved easily.

Theorem 3.1. *Suppose that, for the observation scheme* (2.2) (*or* (2.3)), *the function* $\partial^m u$ *is from the class* $\sum(\beta, L, O(x_0))$ *with* $\overline{O(x_0)} \subset D$. *Then the linear estimator* $\hat{u}_m^\varepsilon$ *defined by* (3.6) *with* $\delta_i = C_i \varepsilon^{\gamma/\beta_i}$ *has the property*

$$\sup_{x \in O(x_0)} E \left| \hat{u}_m^\varepsilon(x) - \partial_x^m u(x) \right|^2 \le C \varepsilon^{2\gamma} \tag{3.10}$$

where

$$\gamma = 2 \left(2 + \sum_{i=1}^{d} \frac{2m_i + 1}{\beta_i} \right)^{-1} \tag{3.11}$$

and C *is independent of* ε.

Proof. By (3.7), we have

$$E \left| \hat{u}_m^\varepsilon(x) - \partial_x^m u(x) \right|^2 = \left| E \Delta \hat{u}_m^\varepsilon(x) \right|^2 + \operatorname{Var} \left\{ \triangle \hat{u}_m^\varepsilon(x) \right\} . \tag{3.12}$$

By assumption, $\partial_x^m u \in \sum(\beta, L, O(x_0))$, apply Lemma 3.1 to get

$$\left| E \Delta \hat{u}_m^\varepsilon(x) \right|^2 = \left| \int_D g_\delta(x - y) \left[\partial_y^m u(y) - \partial_x^m u(x) \right] dy \right|^2 \le C_1 \sum_{i=1}^{d} \delta_i^{2\beta_i} \tag{3.13}$$

for some $C_1 > 0$ independent of δ and $x \in U$. In view of Lemma 3.2,

$$\operatorname{Var} \left\{ \triangle \hat{u}_m^\varepsilon(x) \right\} = \varepsilon^2 E \left[\int \partial_x^m g_\delta(x - y) W(dy) \right]^2 \le C_2 \varepsilon^2 \prod_{i=1}^{d} \delta_i^{-(2m_i+1)} , \tag{3.14}$$

with $C_2 > 0$ independent of x, ε and δ. Making use of the inequalities (3.13) and (3.14), equation (3.12) yields

$$\begin{aligned} E \left| \hat{u}_m^\varepsilon(x) - \partial_x^m u(x) \right|^2 &\le C_1 \sum_{i=1}^{d} \delta_i^{2\beta_i} + C_2 \varepsilon^2 \prod_{i=1}^{d} \delta_i^{-(2m_i+1)} \\ &\le C \varepsilon^{2\gamma} \end{aligned}$$

as asserted, if we choose $\delta_i^{\beta_i} = C_3 \varepsilon^\gamma$ for some $C_3 > 0$ with γ defined by (3.11).

□

Remark 3.1. For $m = 0$, the result (3.10) gives $E\left|\hat{u}^{\varepsilon}(x) - u(x)\right|^2 \le C\,\varepsilon^{4\beta_0/(2\beta_0+1)}$, where $\beta_0^{-1} = \sum_{i=1}^{d} \frac{1}{\beta_i}$. □

Remark 3.2. Since the estimator $\hat{u}_m^{\varepsilon}(x)$ is a Gaussian random variable, it follows from (3.10) that, for any loss function $\ell(r) \le C\exp(\lambda r^2)$, $\lambda \le \lambda_0$, the following inequality holds

$$\sup_{\varepsilon>0} \sup_{x\in O(x_0)} E\ell\left(\left[\hat{u}_m^{\varepsilon}(x) - \partial^m u(x)\right]/\varepsilon^{\gamma}\right) < \infty \ ,$$

if $\partial^m u(x) \in \sum(\beta, L, O(x_0))$. □

In dealing with the lower bounds, consider the estimation of $\theta \subset R^1$ given the observation

$$V_{\varepsilon}(dx) = (\Theta_0(x) + \theta\varphi_{\varepsilon}(x))\ dx + \varepsilon W(dx) \tag{3.15}$$

where both φ_{ε} and Θ_0 are known functions in $L_2(R^d)$. Let T denote the set of all estimators for θ based on the observation $V_{\varepsilon}(\cdot)$. Also introduce the class $\Lambda(\mu)$ of loss functions $\ell : R \to R^+$ which are even, increasing over R^+ with $\ell(0) = 0$ and satisfy $\ell(r) \le Ce^{\mu r^2}$, for some $C > 0$ and $\mu > 0$. The following lemma will be useful for studying lower bounds.

Lemma 3.3. *Consider the estimation of* $\theta \in A \subset R^1$ *based on the observation* (3.15). *Assume that the set* $\{\theta : |\theta| \le b\varepsilon \|\varphi_{\varepsilon}\|^{-1}\} \subset A$. *Then, for any* $\ell \in \Lambda(\mu)$

$$\liminf_{\varepsilon\to 0} \inf_{\theta_{\varepsilon}\in T} \sup_{\theta\in A} E\ell\left(\varepsilon^{-1}\|\varphi_{\varepsilon}\|(\theta_{\varepsilon} - \theta)\right) \ge \frac{1}{2\sqrt{2\pi}} \int_{|r|<b/2} \ell(r)\, e^{-r^2/2}\, dr \tag{3.16}$$

Proof. This follows from the fact that (see (3.2)) the family of measures $P_{\theta}^{\varepsilon}(\cdot)$ for V_{ε} given by (3.15) satisfies the LAN property with the norming factor $\varepsilon\|\varphi_{\varepsilon}\|^{-1}$. The corresponding Fisher information is $I_{\varepsilon} = \varepsilon^{-2}\|\varphi_{\varepsilon}\|^2$. Therefore we can apply the inequality (2.12.19) from [9] to obtain the lower bound (3.16). □

Theorem 3.2. *Consider the observation* (2.3) *and let* $U_{m,\beta}$ *denote the set of functions* $u \in L_2(D)$ *such that*

$$\partial^m u \in \sum(\beta, L, O(x_0)) \ . \tag{3.17}$$

Then, for any $\ell \in \Lambda(u)$,

$$\liminf_{\varepsilon\to 0} \inf_{T_{\varepsilon}\in\mathscr{T}} \sup_{u\in U_{m,\beta}} E\ell\left(\varepsilon^{-\gamma}\left(T_{\varepsilon} - \partial^m u(x_0)\right)\right) > 0 \ , \tag{3.18}$$

where $\mathscr{T}$ *is the set of all estimators based on* (2.3), *and* γ *is defined by* (3.11).

Proof. As in [9] and [10], to prove the lower bound, it is sufficient to restrict the functions u to a one-parameter family linear in $\theta \in (-1, 1)$. We find the family in the form:

$$u_\varepsilon (x, \theta) = \theta \varepsilon^a \prod_{i=1}^{d} \varphi \left(\frac{x^i - x_0^i}{\varepsilon^{bi}} \right) \doteq \theta \varphi_\varepsilon (x - x_0) \quad , \tag{3.19}$$

where φ is a sufficiently smooth function with a compact support and

$$\varphi^{(m_i)}(0) \neq 0,\ i = 1, 2, \ldots, d \quad .$$

For the choice of u_ε by (3.19), the Fisher information for θ given the observation (2.2), can be calculated by noting (3.2) and (3.19) as follows:

$$I_\varepsilon = \varepsilon^{-2} \left\| \varphi_\varepsilon \right\|^2 = \varepsilon^{2(a-1)+\sum_{i=1}^{d} b_i} \left\| \varphi_\varepsilon \right\|^{2d} \quad . \tag{3.20}$$

It is clear from (3.19) that the condition (3.17) is valid if

$$a - \sum_{j=1}^{d} m_j b_j \geq \beta_i b_i,\ i = 1, 2, \ldots d \quad . \tag{3.21}$$

Let

$$\beta_1 b_1 = \beta_2 b_2 = \cdots = \beta_d b_d = \kappa \quad , \tag{3.22}$$

and choose b_i to satisfy the equations

$$\sum_{i=1}^{d} b_i = 2 (1 - a) , a - \sum_{i=1}^{d} m_i b_i = \kappa \quad . \tag{3.23}$$

Then, from (3.22) and (3.23), we obtain

$$a = \kappa \left(\sum_{i=1}^{d} m_i \beta_i^{-1} + 1 \right); \quad b_i = \kappa \beta_i^{-1}, i = 1, 2, \ldots d \quad , \tag{3.24}$$

with

$$\kappa = \gamma = 2 \left[2 + \sum_{i=1}^{d} (2m_i + 1) \beta_i^{-1} \right]^{-1} \quad .$$

For the choice (3.24) of a and b_i, $u_\varepsilon \in U_m$ for any $\varepsilon > 0$ and Fisher's information $I_\varepsilon = I$ beingindependent of ε. For any $\ell \in \Lambda(\mu)$, noting (3.20) and (3.24), we have

$$\begin{aligned}
&\inf_{T_\varepsilon \in \mathscr{T}} \sup_{u \in U_{m,\beta}} E\ell\left((T_\varepsilon - \partial^m u\,(x_0))\,\varepsilon^{-\gamma}\right) \\
&\geq \inf_{T_\varepsilon \in \mathscr{T}} \sup_{\theta \in (-1,1)} E\ell\left(\left(T_\varepsilon - \partial^m_{x_o} u_\varepsilon\,(x_0, \theta)\right)\varepsilon^{-\gamma}\right) \\
&= \inf_{T_\varepsilon \in \mathscr{T}} \sup_{\theta \in (-1,1)} E\ell\left(\left(T_\varepsilon - \theta\varepsilon^\gamma \prod_{i=1}^d \varphi^{(m^i)}(0)\right)\varepsilon^{-\gamma}\right) \\
&\geq \inf_{T_\varepsilon \in \mathscr{T}} \sup_{\theta \in (-1,1)} E\ell\left(\eta\varepsilon^{-1}\,\|\varphi_\varepsilon\|\,(\theta_\varepsilon - \theta)\right) \quad ,
\end{aligned} \tag{3.25}$$

where

$$\eta = \left[|\varphi|^d \prod_{i=1}^d \varphi^{(m_i)}(0)\right]^{-1} \tag{3.26}$$

and $|\varphi|$ denotes the L_2 - norm of φ in $\mathscr{R}^1$. Recall : $\varphi^{(m_i)}(0) \neq 0$, for each i, so that $|\eta| < \infty$. Let us choose $b > 0$ and $b < \varepsilon^{-1}\,\|\varphi_\varepsilon\| = |\varphi|^d$, so that $|\theta| < 1$. Then we can apply Lemma 3.3 to (3.25) to yield the desired result:

$$\begin{aligned}
&\liminf_{\varepsilon \to 0} \inf_{T_\varepsilon \in \mathscr{T}} \sup_{u \in U_{m,\beta}} E\ell\left(\left(T_\varepsilon - \partial^m u(x_0)\right)\varepsilon^{-\gamma}\right) \\
&\geq \frac{1}{2\sqrt{2\pi}} \int\limits_{|r|<b/2} \ell\,(\eta r)\, e^{-r^2/2} dr > 0 \quad ,
\end{aligned}$$

where γ and η are defined by (3.11) and (3.26), respectively. □

In what follows, we say that two real functions $a\,(\varepsilon) \asymp b\,(\varepsilon)$ for $\varepsilon \to 0$ if there exist constants C_1 and C_2 such that

$$0 < C_1 < a(\varepsilon)/b(\varepsilon) < C_2 < \infty$$

for any $\varepsilon \in (0, \varepsilon_0)$, $\varepsilon_0 > 0$.

Remark 3.3. Theorems 3.1 and 3.2 show that, among the class $\Lambda\,(\mu)$ of loss functions ℓ, the optimal rate of convergence for the estimator T_ε to $\partial^m u$ in $\sum(\beta, L, O\,(x_0))$ is of the order of ε^γ, which cannot be improved. So for $\varepsilon \to 0$

$$\liminf_{\varepsilon \to 0} \inf_{T_\varepsilon \in \mathscr{T}} \sup_{u \in U_{m,\beta}} E\ell\left(\left(T_\varepsilon - \partial^m u(x_0)\right)\varepsilon^{-\gamma}\right) \asymp 1 \tag{3.27}$$

□

Remark 3.4. Let $\partial^m \Theta_0 \in \sum (\beta, L_1, O(x_0))$ with $L_1 < L$ and consider the family

$$u_\varepsilon (x, \theta) = \Theta_0(x) + \theta \varphi_\varepsilon (x - x_0) \quad . \tag{3.28}$$

Introduce a metric on $\sum (\beta, L, O\,(x_0))$

$$\rho_{m,\beta} (f_1, f_2) = \sup_{x \in O(x_0)} \sum_{j=1}^{d} \sum_{r_j=0}^{k_j} \sum_{\ell \le m} \left| \partial_j^{r_j} \left(\partial^\ell f_1(x) - \partial^\ell f_2(x) \right) \right| \tag{3.29}$$

where ℓ and m are multi-indices and $\ell \le m$ means $0 \le \ell_i \le m_i, i = 1, 2, \ldots d$.

Then one can conclude from (3.19) that for any $\delta > 0, \theta \in (-1, 1)$

$$\rho_{m,\beta} (u_\varepsilon, \Theta_0) < \delta \quad \text{and} \quad u_\varepsilon (x, \theta) \in \sum (\beta, L, O\,(x_0))$$

if $\varepsilon < \varepsilon_0$ sufficiently small and φ in (3.19) is chosen properly. □

Therefore, by applying Lemma 3.3 again, we arrive at a generalized version of Theorems 3.1, 3.2.

Theorem 3.3. *Consider the observation scheme* (2.2). *Then for any* $\ell \in \Lambda\,(\mu)\,, \partial^m \Theta_0 \in \sum (\beta, L_1, O\,(x_0))$ *with* $L_1 < L$, *and for any* $\delta > 0$, *as* $\varepsilon \to 0$,

$$\inf_{T_\varepsilon \in \mathscr{T}} \sup_{u \in U_{m,\beta} \rho_{m,\beta}(u, \Theta_0) < \delta} E\ell \left(\left(T_\varepsilon - \partial^m u(x_0) \right) \varepsilon^{-\gamma} \right) \asymp 1 \tag{3.30}$$

□

Remark 3.5. From the preliminary results, it has become clear that, inside a bounded domain D, the estimation problem can be localized by the kernel g_δ for small $\delta > 0$. Therefore, in the subsequent analysis, the integration with respect to g_δ over D will often be replaced by that over R^d without further explanation.

4. Source estimation problem

We consider now the estimation problem (1.2), based on the observation (2.2), or more specifically, the estimation of $\Theta(x), x \in D$, in the equation

$$\mathscr{L}(x) u(x) = \Theta(x) \quad , \tag{4.1}$$

where

$$\mathscr{L}(x) = \sum_{k=0}^{n} f_k\ (x, \partial_1, \ldots, \partial_d) \doteq \sum_{k=0}^{n} f_k\ (x, \partial_x)\ ,$$

and, for $x \in D$, $f_k\ (x, \cdot)$ is a homogeneous polynomial of degree k in $\lambda \in R^d$:

$$f_k(x, \lambda_1, \ldots, \lambda_d) = \sum_{|j|=k} a_j^k(x)\lambda^j = \sum_{|j|=k} a^k_{j_1}, \ldots, {}_{j_d}(x)\lambda_1^{j_1} \ldots \lambda_d^{j_d}\ , \tag{4.2}$$

with $\lambda = (\lambda_1, \ldots, \lambda_d)$; $j = (j_1, \ldots, j_d)$ and $|j| = \sum_{i=1}^{d} j_i$.

Let $\mathscr{L}_n(x) = f_n(x, \partial_x)$ be the principal part of $\mathscr{L}(x)$ with the corresponding principal polynomial $f_n\ (x, \lambda)$. As we will show, the rate of convergence for (4.1) depends on the form of $\mathscr{L}_n$. By convention, we will say that $\mathscr{L}\ (x)$ is *nondeficient* or *regular* at x_0 w.r.t. $x = (x^1, \ldots, x^d)$ if all of the coefficients of $\partial_i^n u, i = 1, \ldots, d$, are not equal to 0 at x_0. In other words $\mathscr{L}$ is regular at x_0 if its principal polynomial $f_n\ (x_0, \lambda)$ satisfies the condition:

$$\partial^n_{\lambda_i} f_n\ (x_0, \lambda) \neq 0 \ \text{ for } \ i = 1, 2, \ldots, d\ . \tag{4.3}$$

Otherwise, $\mathscr{L}$ is said to be *deficient* at x_0. In this section we first treat the regular case. A class of problems with deficient $\mathscr{L}$ will be considered in the next section.

To this end, as before, we assume that the solution of (4.1) is observed in a Gaussian white noise, so only $V_\varepsilon(\cdot)$ from (2.3) is available to a statistician. We also assume that

$$\Theta \in \sum (\beta, L, O\ (x_0)) = \sum (\beta_1, \ldots, \beta_d, L, O\ (x_0))\ , x_0 \in D\ , \tag{4.4}$$

and consider the estimation of Θ in (4.1) given the observation (2.3). Here the questions are: How do we construct an estimator for $\Theta\ (x_0)$ at any point $x_0 \in D$, with the optimal rate of convergence for certain risks and what is the optimal rate of convergence as $\varepsilon \to 0$? To answer these questions, let us introduce again the formal adjoint operator $\mathscr{L}^*(x)$ of $\mathscr{L}(x)$. Assume that in (4.2), the coefficients $a_j^k \in C_b^k\ (D)$, the set of all bounded k-times continuously differentiable functions on D, for $k = 0, 1, \ldots, n$. Then the adjoint $\mathscr{L}^*(x)$ exists and is given by [6]

$$\mathscr{L}^*(x)u(x) = \sum_{k=0}^{n} (-1)^k\ \partial_x^j \left[a_j^k(x)u\ (x)\right], u \in C_b^n\ (D)\ . \tag{4.5}$$

The following Green's identity holds for $u, v \in C_b^n(D)$:

$$\int_D \left(v\mathscr{L}u - u\mathscr{L}^* v\right)\, dx = \int_{\partial D} B(u, v)\, dS \ , \tag{4.6}$$

where $B(u, v)$ is a bilinear expression in $u(x)$ and $v(x)$ and their derivatives up to order $(n-1)$ and dS is the surface element on ∂D. In particular let $V(x) = g_\delta(x - x_0)$ for $x_0 \in D$, which is the kernel function as defined in (3.4). Choose $\delta = \delta(\varepsilon) \to 0$ as $\varepsilon \to 0$. Then, for small enough ε, $g_\delta(\cdot - x_0)$ has a compact support in $O(x_0) \subset D$, so that $B(u, g_\delta(\cdot - x_0))$ vanishes on ∂D and the identity (4.6) yields

$$\int g_\delta(y - x_0)\, \mathscr{L}(y)\, u(y)\, dy = \int u(y)\, \mathscr{L}^*(y)\, g_\delta(y - x_0)\, dy \ . \tag{4.7}$$

Now we introduce the estimator

$$\hat{\Theta}_\varepsilon(x_0) = \int \mathscr{L}^*(y)\, g_\delta(y - x_0)\, V_\varepsilon(dy) \ , \tag{4.8}$$

which is a generalization of (3.6). (Here, for convenience, the variable x and y have been interchanged.) Similar to (3.7), by taking the equations (4.1), (4.7) and (2.2) into account, equation (4.8) gives

$$\begin{aligned}\hat{\Theta}_\varepsilon(x_0) - \Theta(x_0) = &\int g_\delta(y - x_0)\, [\Theta(y) - \Theta(x_0)]\, dy \\ &+\varepsilon \int \mathscr{L}^*(y)\, g_\delta(y - x_0)\, W(dy)\, .\end{aligned} \tag{4.9}$$

Note that, by assumption $a_j^k \in C_b^k(D)\, , k \le n$, we have

$$\left\| \mathscr{L}^* g_\delta(\cdot - x_0)\right\|^2 \le \frac{C}{\delta_1 \ldots \delta_d} \sum_{i=1}^{d} \delta_i^{-2n} \tag{4.10}$$

for some $C > 0$.

In view of (4.4) and (4.7), we apply Lemmas 3.1 and 3.2 to get

$$E\left|\hat{\Theta}_\varepsilon(x_0) - \Theta(x_0)\right|^2 \le C\left(\sum_{i=1}^d \delta_i^{2\beta_i} + \frac{\varepsilon^2}{\delta_1 \ldots \delta_d}\sum_{i=1}^d \delta_i^{-2n}\right) \qquad (4.11)$$

It is easy to see that the right hand side of (4.11) is minimal in the order of convergence to 0, as $\varepsilon \to 0$, when

$$\delta_i = C_i \varepsilon^{\mu_i} \quad \text{with} \quad \mu_i = \frac{2\beta_i^{-1}}{2 + \rho + 2n\beta_{\min}^{-1}}\ , \qquad (4.12)$$

where $\rho = \sum_{i=1}^d \beta_i^{-1}$ and $\beta_{\min} = \min\{\beta_1, \beta_2, \ldots, \beta_d\}$.

Upon substituting (4.12) into (4.11), we arrive at the following theorem.

Theorem 4.1. *Let the solution of the equation* (4.1) *be observed in the Gaussian white noise* $W(\cdot)$ *with intensity* ε *and let* $V_\varepsilon(\cdot)$ *be the observed random measure defined by* (2.2). *Assume that the partial differential operator* $\mathscr{L}$ *has its coefficients* $a_j^k \in C_b^k(D)$ *for* $k = 0, 1, \ldots n$; $\Theta \in \sum(\beta, L, O(x_0))$ *and the kernel* g_δ *is defined by* (3.4) *and* (3.5) *with* $g_i \in G_{k_i}, i = 1, 2, \ldots d$. *Then, by choosing the 'bandwidths'* δ_i *in accordance with* (4.12), *the estimator* $\hat{\Theta}_\varepsilon(x_0), x_0 \in D$, *has the property*: *for some constant* $C > 0$ *independent of* ε,

$$E|\hat{\Theta}_\varepsilon(x_0) - \Theta(x_0)|^2 \le C\varepsilon^{2\kappa};\ \kappa = 2(2 + \rho + 2n\beta_{\min}^{-1})^{-1}\ . \qquad (4.13)$$

Remark 4.1. The estimator $\hat{\Theta}_\varepsilon(x_0)$ is a Gaussian random variable, so (see Remark 3.2) for any loss function with $\ell(r) \le C\exp(\lambda r^2), \lambda \le \lambda_0$, the following inequality is valid

$$\limsup_{\varepsilon\to 0} E\ell((\hat{\Theta}_\varepsilon(x_0) - \Theta(x_0))/\varepsilon^\kappa) < \infty\ . \quad \square \qquad (4.14)$$

Remark 4.2. As in the estimation of u, for $\Theta \in \sum(\beta + m, L, O(x_0))$, the mixed partial derivative $\partial^m\Theta_\varepsilon(x_0)$ can be estimated by $\hat{\Theta}_m^\varepsilon(x_0) = \partial^m\hat{\Theta}_\varepsilon(x_0)$ and a similar upper bound for $E|\hat{\Theta}_\varepsilon - \partial^m\Theta_\varepsilon(x_0)|^2$ can be obtained. $\square$

In general the upper bound (4.13) and (4.14) are not tight, but, if the condition (4.3) for $\mathscr{L}$ is valid, they turn out to be tight as we will now show.

Theorem 4.2. *Assume that the equation* (4.1) *has the unique solution for some specified initial/boundary conditions and the assumptions of Theorem*

4.1 are valid. In addition assume that the regularity condition (4.3) *for* $\mathscr{L}$ *holds. Then, for any* $\ell \in \Lambda$,

$$\liminf_{\varepsilon\to 0} \inf_{T_\varepsilon \in \mathscr{T}} \sup_{\Theta \in \Sigma} E\ell((T_\varepsilon - \Theta(x_0))/\varepsilon^\kappa) > 0, \ x_0 \in D \ , \tag{4.15}$$

where $\mathscr{T}$ *is the set of all estimators based on* V_ε *from* (2.2), $\sum = \sum(\beta, L, O(x_0))$ *and* κ *is defined in* (4.13).

Proof. The ideas of proof are similar to that of Theorem 3.2. First of all note that we can restrict ourselves, in the case of lower bound to the zero initial/boundary conditions. This is so because in general the solution of (4.1) can be written as the sum of u_1 and u_2 where u_1 is the solution of (4.1) with zero side conditions and u_2 is the known solution of the homogeneous equation $\mathscr{L}u = 0$ with the given side conditions. So we can consider again the family of functions $u_\varepsilon(x, \theta)$ of the form (3.19) and define

$$\Theta_\varepsilon(x) \doteq \mathscr{L}(x)u_\varepsilon(x, \theta) \tag{4.16}$$

Now choose the parameters a, b_i so that the Fisher's information for the parameter θ is of order one, and $\Theta_\varepsilon(x)$ belongs to $\sum(\beta, L, O(x_0))$ as $\varepsilon \to 0$. Similar to the proof of Theorem 3.2, a proper choice of the parameters leads to the following conditions

$$2a + \sum_{i=1}^{d} b_i = 2; \ a - nb_i \geq b_i\beta_i$$

(in comparison with (3.21) and (3.23)), which are sufficient for these purposes. So, for κ given as in (4.13), the above conditions imply that

$$b_i = \kappa\beta_i^{-1}; \ \ a = (1 + n\beta_{\min}^{-1})\kappa \ .$$

Therefore the family (4.16) of functions $\Theta_\varepsilon \in \sum(\beta, L, O(x_0))$ and Lemma 3.3 is applicable. Now notice that, for this choice of a and b_i,

$$\Theta_\varepsilon(x_0) = \mathscr{L}(x_0)u_\varepsilon(x_0, \theta) = C_1\theta\varepsilon^\kappa(1 + o(1)) \ \text{as} \ \ \varepsilon \to 0 \ ,$$

provided that condition (4.3) is met. So the estimation of θ for this family, as $\varepsilon \to 0$, is equivalent to the estimation of

$$\varepsilon^{-\kappa}\mathscr{L}(x_0)u_\varepsilon(x_0, \theta) = \varepsilon^{-\kappa}\Theta_\varepsilon(x_0) \ .$$

The validity of the theorem as asserted follows from this fact and Lemma 3.3 analogously as in the proof of Theorem 3.2. □

Remark 4.3. It follows from (4.14), (4.15) that for $x_0 \in D$, $\varepsilon \to 0$

$$\inf_{\hat{u}_\varepsilon(x_0)\in\mathscr{T}} \sup_{u\in\sum(\beta,L,O(x_0))} E\ell(\hat{u}_\varepsilon(x_0) - u(x_0)\varepsilon^{-\kappa}) \asymp 1$$

for any $\ell \in \Lambda(\mu)$. In particular, for $n = 0$ we have again (3.27). □

Remark 4.4. As mentioned before (see Remark 3.4), the lower bound can be strengthened. That is, the sup taken in (4.15) over $\sum(\beta, L, O(x_0))$ can be replaced by the sup over its intersection with an arbitrary small $\rho_{m,\beta}$-neighborhood (see (3.29)) of a fixed function $\Theta_0 \in \sum(\beta, L_1, O(x_0))$ with $L_1 < L$, such that there exists a solution of the equation $\mathscr{L}(x)u(x) = \Theta_0(x)$, subject to suitable initial-boundary conditions. □

Remark 4.5. In view of Theorems 4.1 and 4.2 and the previous remarks, we can conclude that the optimal rate of convergence is ε^κ and for $\ell \in \Lambda$,

$$\inf_{T_\varepsilon\in\mathscr{T}} \sup_{\Theta\in\Sigma_1} E\ell((T_\varepsilon - \Theta(x_0))/\varepsilon^\kappa) \asymp 1 \tag{4.17}$$

as $\varepsilon \to 0$, where $\sum_1 = \sum(\beta, L, O(x_0)) \cap \{u : \rho_{m,\beta}(u, \Theta_0) < \delta\}$, $x_0 \in D$ and κ is defined as in (4.13). □

The theorems can be applied to source estimation problems for some well-known equations in mathematical physics. In what follows, we shall give a few examples as special cases of (4.1).

Example 4.1. For $n = 1$, consider the first-order linear equation

$$\mathscr{L}u = \sum_{i=1}^{d} a_i(x)\partial_i u + a_0 u$$

where $a_i \in C_b^1(D)$ with $a_i(x_0) \neq 0$ for $i = 1, 2, \ldots d$, $a_0 \in C_b(D)$, and $C_b^k(D)$ denotes the set of bouned, k-time continuously differential functions on D with $C_b^0 = C_b$. Then

$$\mathscr{L}^* u = -\sum_{i=1}^{d} \partial_i[a_i(x)u] + a_0(x)u \quad ,$$

and the estimator $\hat{\Theta}_\varepsilon(x_0)$ in (4.5) has the convergence properties (4.17) with the rate

$$\kappa = 2(2 + \rho + 2\beta_{\min}^{-1})^{-1};\ \rho = \sum_{i=1}^{d} \beta_i^{-1}$$

Example 4.2. For $n = 2$, consider

$$\mathscr{L}u = \sum_{i,j=1}^{d} a_{ij}(x)\partial_i\partial_j u + \sum_{i=1}^{d} a_i(x)\partial_i u + a_0(x)u \qquad (4.18)$$

where $a_{ij} \in C_b^2(D)$, $a_i \in C_b^1(D)$ for $i \geq 1$ and $a_0 \in C_b(D)$. Then we have

$$\mathscr{L}^* u = \sum_{i,j=1}^{d} \partial_i\partial_j(a_{ij}u) - \sum_{i=1}^{d} \partial_i(a_i u) + a_0 u \ .$$

Assume that $a_{ii}(x_0) \neq 0$, $i = 1, 2, \ldots, d$, so that the regularity condition (4.3) is valid. Thus the estimator has the convergence property (4.17) with the rate

$$\kappa = 2(2 + \rho + 4\beta_{\min}^{-1})^{-1} \ .$$

So we see that the rate of convergence to zero for the best estimator of $\Theta(x)$ is the same for the elliptic and hyperbolic equations of second order.

Example 4.3. For $n = 4$ and $d = 2$, consider the operator $\mathscr{L}$ corresponding the steady-state of a vibrating elastic plate [3],

$$\mathscr{L}u = \Delta^2 u + cu$$

where Δ^2 is the bi-harmonic operator $\Delta^2 u = \partial_1^4 u + 2\partial_1^2\partial_2^2 u + \partial_2^4 u$ and c is a constant. Then $\mathscr{L}^* = \mathscr{L}$ and the condition (4.3) is met. The rate of convergence for this case is given by

$$\kappa = 2(2 + \rho + 8\beta_{\min}^{-1})^{-1}, \quad \text{where} \quad \rho = (\beta_1^{-1} + \beta_2^{-1}).$$

5. Estimation problems with deficiency

When the differential operator $\mathscr{L}$ fails to satisfy the condition (4.3) or is deficient at x_0, the estimator $\hat{\Theta}_\varepsilon$ will have a different rate of convergence as to be shown. To be specific, let us consider a class of operators $\mathscr{L}$ deficient in the x^1this class can be easily generalized. We assume here that the operator $\mathscr{L}$ has the form

$$\mathscr{L} = \sum_{l=1}^{m} b_l(x)\partial_1^l + \sum_{k=0}^{n} \sum_{|j|=k} a_{j_2,\ldots,j_d}(x)\partial_2^{j_2} \ldots \partial_d^{j_d} = \mathscr{L}_1(x) + \mathscr{L}_2(x) \qquad (5.1)$$

For the lower bound we will consider the following condition.

Condition **A.** The operator $\mathscr{L}_2(x)$ is nondeficient at x_0 with respect to variables $x_2, \dots, x_d$, $m < n$ and $b_m(x_0) \neq 0$.

In this case we can proceed as before and use the same kernel type of estimator (4.8) for the upper bounds, with suitably modified choices of $\delta_i = \delta_i(\varepsilon)$. For the lower bounds, the same type of one-parameter family of functions in the form of (4.16), (3.19) can be adopted.

We assume again the coefficients of $\mathscr{L}$ smooth enough so that the identity (4.7) is valid and the source term $\Theta \in \sum(\beta, L, O(x_0))$.Using (4.9) we obtain analogously to (4.11)

$$E|\hat{\Theta}_\varepsilon(x_0) - \Theta(x_0)|^2 \leq C\left[\sum_{i=1}^{d}\delta_i^{2\beta_i} + \frac{\varepsilon^2}{\delta_1 \dots \delta_d}\left(\delta_1^{-2m} + \sum_{i=2}^{d}\delta_i^{-2n}\right)\right] \tag{5.2}$$

It follows from (5.2) that the rate of convergence for the risks depends on the relative magnitudes of δ_i and $\alpha = \min(\beta_2, \dots, \beta_d)$. In fact the following theorem for both the upper and lower bounds holds (compared with Theorems 4.1 and 4.2).

Theorem 5.1. *Let the solution of equation* (4.1) *with the operator* $\mathscr{L}$ *from* (5.1) *be observed in the Gaussian white noise* $W(.)$ *with intensity* ε *and let* $V_\varepsilon(.)$ *be the observed random measure defined by* (2.2). *Assume that* (4.4) *is valid and the smoothness conditions on the coefficients of the operator* $\mathscr{L}$ *from the Section 4 are fulfilled.Consider again the estimator* $\hat{\Theta}_\varepsilon(x_0)$ *given by* (4.8) *with* $\delta_i = \varepsilon^{\kappa/\beta_i}$, $i = 1, 2, \dots, d$, $g_i \in G_{k_i}$, *where*

$$\kappa = 2\left[2 + \rho + \frac{2}{\left(\frac{\beta_1}{m}\right) \wedge \left(\frac{\alpha}{n}\right)}\right]^{-1} \tag{5.3}$$

$\rho = \sum_{i=1}^{d}\beta_i^{-1}$ *and* $\alpha = \min(\beta_2, \dots, \beta_d)$.

Then for any $\ell \in \Lambda$,

$$\sup_{\varepsilon > 0} E\ \ell((\hat{\Theta}_\varepsilon(x_0) - \Theta(x_0))\varepsilon^{-\kappa}) < \infty\ . \tag{5.4}$$

If moreover the condition **A** *is valid, this rate is optimal in the minimax sense: for* $\varepsilon \to 0$

$$\inf_{T_\varepsilon \in \mathscr{T}} \sup_{\Theta \in \Sigma(\beta, L, O(x_0))} E\ell((T_\varepsilon - \Theta(x_0))/\varepsilon^\kappa) \asymp 1 \tag{5.5}$$

Proof. The proof is quite similar to that of Theorems 4.1 and 4.2. We set $\delta_i = \varepsilon^{\kappa/\beta i}$ in (5.2) to get

$$E|\hat{\Theta}_\varepsilon(x_0) - \Theta(x_0)|^2 \leq C\left\{\varepsilon^{2-\kappa\rho}(\varepsilon^{-2\kappa m/\beta_i} + \sum_{i=2}^{d}\varepsilon^{-2\kappa n/\beta_i}) + \varepsilon^{2\kappa}\right\}\ . \tag{5.6}$$

If κ is chosen as in (5.3) then we have from (5.6) that $E|\hat{\Theta}_\varepsilon(x_0) - \Theta(x_0)|^2 \leq C\varepsilon^{2\kappa}$. Taking into account the fact that $\hat{\Theta}_\varepsilon(x_0)$ is a Gaussian random variable, we can deduce that (referring to Remark 4.1) the upper bound (5.4) holds.

For the lower bound we introduce again the family (3.19) in which we choose

$$a = \frac{\zeta}{\rho + \zeta} \quad b_i = \frac{2\beta_i^{-1}}{\rho + \zeta}, \quad \zeta = 2 + [({}^{\beta_1}/_m) \wedge (\alpha/_n)]^{-1} \ . \qquad (5.7)$$

It is an easy calculation to check that for this choice of parameters in (3.19), we have $\hat{\Theta}_\varepsilon(x) = \mathscr{L}(x)u_\varepsilon(x,\theta) \in \sum(\beta, L, O(x_0))$ for all $\theta \in (-1, 1)$. As in (3.19), the Fisher's information for the estimation of θ is of order one as $\varepsilon \to 0$, because (compared with (5.7)) $2a + \sum_{i=1}^{d} b_i = 2$. Further, a straightforward computation implies that for $\varepsilon \to 0$

$$\Theta(x_0) = \theta\mathscr{L}(x)\varphi_\varepsilon(x - x_0)|_{x=x_0} = C\theta\varepsilon^\kappa(1 + o(1)) \ \ (C \neq 0) \ .$$

Therefore the estimation of $\Theta_\varepsilon(x_0)$ is equivalent to that of $\theta\varepsilon^\kappa$ as $\varepsilon \to 0$. Now the lower bound in (5.5) follows from these facts in exactly the same manner as in the proof of Theorem 3.2. □

As an application of this theorem, we will give two examples which involve two prominent equations in physical science. In each case, physically speaking, $x^1 = t$ being the time and $(x_2, \ldots, x_d)$ is the space variable.

Example 5.1. For heat conduction or diffusion, consider the parabolic operator $\mathscr{L}$ in $G = (0, T) \times D$ with $D \subset R^{d-1}$ defined by

$$\mathscr{L}(x)u = -\partial_1 u + \sum_{i,j=2}^{d} a_{ij}(x)\partial_i\partial_j u + \sum_{i=2}^{d} a_i(x)\partial_i u + a_0(x)u \ ,$$

where $a_{ij} \in C_b^2(D)$, $a_i \in C_b^1(D)$ for $i \geq 2$, $a_0 \in C_b(D)$ and $\sum_{i,j=2}^{d} a_{ij}\lambda_i\lambda_j > 0$ for $\lambda \neq 0$.

Then

$$\mathscr{L}^*(x)u = \partial_1 u + \sum_{i,j=2}^{d} \partial_i\partial_j(a_{ij}u) - \sum_{i=2}^{d} \partial_i(a_i u) + a_0 u \ .$$

In this case: $n = 2,\ m = 1$, the rate of convergence in (5.4) and (5.5) is given by

$$\kappa = 2[2 + \rho + \frac{2}{\beta_1 \ \wedge (\alpha/2)}]^{-1}; \ \rho = \sum_{i=1}^{d} \beta_i^{-1}, \ \ \alpha = \min(\beta_2, \ldots, \beta_d)$$

Example 5.2. For the vibration of an elastic plate [3], the operator $\mathscr{L}$ is of the form

$$\mathscr{L}u = (\partial_1^2 - \Delta^2)u \quad ,$$

where $\Delta^2 u = (\partial_2^4 + 2\partial_2^2\partial_3^2 + \partial_3^4)u$.

Here clearly $\mathscr{L} = \mathscr{L}^*$ is formally self-adjoint, and the Condition (A) holds with $n = 4, m = 2$. The assumptions in Theorem 5.1 are fulfilled so that the optimal rate κ of convengence is equal to

$$2[2 + \rho + \frac{4}{\beta_1 \ \wedge (\alpha/2)}]^{-1}, \quad \text{with} \quad \rho = \sum_{i=1}^{3} \beta_j^{-1} \quad .$$

6. Concluding remarks

In this paper a nonparametric estimation approach to inverse source problems for linear partial differential equations is introduced as a viable alternative to the deterministic method. By using the kernel smoothing type of estimators, we proved that, with a proper choice of the kernel and the "bandwidths" δ_i, such estimators for the unknown source functions can achieve the optimal rate of convergence in the minimax sense within a wide class of risks as the noise intensity $\varepsilon \to 0$. Interestingly the optimal rate of convergence depends only on the principal part of the operator $\mathscr{L}(x)$ but not on the type of $\mathscr{L}$, such as elliptic or hyperbolic. The smoothness of solutions turns out to have no effect at all on the convergence rate. In this sense the kernel type of estimators is robust for linear problems. The same type of kernel estimators is applicable to nonlinear inverse problems when either some coefficients $a_j^k(x)$ of $\mathscr{L}$ are to be estimated or $\mathscr{L}(x)$ itself is nonlinear. In contrast with linear problems, it is found that the rate of convergence for a best estimator depends critically on the smoothness of the solution. Results on some nonlinear estimation problems will be discussed in a separate paper, (see [2] for the case $d = 1$).

Although, for brevity, we have only proved theorems concerning the upper and lower bounds for the estimation of the source Θ, it is possible to estimate its derivatives $\partial^m\Theta$ as well. As was noticed in the Introduction we have considered here only one class of inverse problems, the source estimation problem fo PDE. We believe that other interesting and important ill-posed problems can be treated with help of similar nonparametric estimation approach. Note in conclusion that the proposed kernel smoothing method provides a feasible numerical algorithm for implementation.

References

[1] Bickel, P.J., Klaassen, C.A.J., Ritov, Y., Wellner, J.A.: Efficient and adaptive estimation for semiparametric models. Baltimore: John Hopkins University Press, 1993

[2] Chow, P.L., Khasminskii, R.Z.: Statistical approach to dynamical inverse problems. Comm. in Math. Phys. **189**, 521–531 (1997)

[3] Courant, R., Hilbert, D.: Methods of Mathematical Physics, Vol. I. New York: Interscience Pub., Inc. 1953

[4] Ermakov, M.: On optimal solutions of the deconvolution problem. Inverse Problems, **6**, 863–872 (1990)

[5] Franklin, J.: Well-posed extension of ill-posed problems. J. Math. Analys. and Appl. **31**, 682–716 (1970)

[6] Friedman, A.: Partial differential equations. New York: Holt, Reinhart and Winston, 1969

[7] Gikhman, I.I., Skorohod, A.V.: The Theory of Stochastic Processes, Vol. I. New York, Heidelberg, Berlin: Springer-Verlag, 1972

[8] Groneboom, P., Wellner, J.A.:Information Bounds and Nonparametric Maximum Likelihood Estimation. Boston: Birkhäuser, 1992

[9] Hadamard, J.: Le Probléme de Cauchy et les Équations aux Dérivées Partielles Linéaires Hyperboliques. Paris: Herman et Cie, 1932

[10] Ibragimov, I.A., Khasminskii, R.Z.: Statistical Estimation. Asymptotic Theory. New York:Springer, 1981

[11] Ibragimov, I.A., Khasminskii, R.Z.: Estimates of the signal, its derivatives and point of maximum for Gaussian distribution. Theory Probab. and Appl. **25**, 703–730 (1980)

[12] Ibragimov, I.A., Khasminskii, R.Z.: Asymptotically normal families of distributions and efficient estimation. Ann. Statist. **19**, 1681–1724 (1991)

[13] Ivanov, V.K.: On ill-posed problems. Soviet Math., Sbornik. **61**, 211–223 (1963) (in Russian)

[14] O'Sullivan, F.: A statistical perspective on ill-posed inverse problems. Statist. Science **1**, 502–527, 1986

[15] Parzen, E.: On estimation of a probability density function and mode. Ann. Math. Statist. **33**, 1065–1076 (1962)

[16] Rosenblatt, M.: Curve estimation. Ann. Math. Statist. **42**, 1815–1842 (1971)

[17] Skorohod, A.V.: Integration in Hilbert Space. Berlin, New York: Springer-Verlag, 1974

[18] Sudakov, V.N., Halfin, L.A.: Statistical approach to correctness of problems in mathematical physics. Soviet Math. Doklady, **157**, 1094–1096 (1964)

[19] Tikhonov, A.N., Arsenin, V. Ya.: Solutions of Ill-posed Problems. New York: John Wiley and Sons, 1977

[20] Wahba, G.: Smoothing and ill-posed problems, in solution methods for integral equations with applications, ed. by M. Goldberg, New York Plenum, 183–194 (1979)

[21] Walsh, J.B.: An introduction to stochastic partial differential equations. 'Ecole deté Probabilité de Saint Flour XIV-1984, ed by P.L. Hennequin (Lect. Notes in Math. vol 1180). Berlin Heindelberg New York, Springer-Verlag: 265–439 (1984)