HIROSHIMA MATH. J.
29 (1999), 61–87

Minimax estimation of common variance in normal distributions when the mean vector is known to lie in an ellipsoid

Teruo FUJIOKA

(Received December 26, 1997)
(Revised January 20, 1998)

ABSTRACT. This paper is concerned with minimax estimation of variance when n samples $y_1, \dots, y_n$ are independently normally distributed with common variance. Here it is assumed that $(E(y_1), \dots, E(y_n))$ is known to lie in an ellipsoid. A new class of estimators which are quadratic in $y_1, \dots, y_n$ are introduced and the minimax estimators are explicitly given. The case of i.i.d. sample with $N(0, \sigma^2)$ is discussed as a special case where the ellipsoid degenerates to the origin. In this case our minimax estimator provides the minimum mean squared error estimator of σ^2.

1. Introduction

This paper is concerned with minimax estimation of variance in a model which is closely related to a nonparametric regression. We consider a simplified model. Let y_i $(i = 1, \dots, n)$ be independently distributed as $N(\mu_i, \sigma^2)$, where both the mean vector $(\mu_1, \dots, \mu_n)$ and the variance σ^2 are unknown. The mean vector is assumed to lie in an ellipsoid

$$\sum_{i=1}^{n} \lambda_i \mu_i^2 \le r\sigma^2 \tag{1}$$

with fixed constants $0 < \lambda_1 < \cdots < \lambda_n$ and a fixed value $r > 0$. Speckman [21] introduced such a model by considering a simplified formulation of spline smoothing in nonparametric regression. Let the observation y_i be taken at a design point $t_i \in [a, b]$. Suppose that $y_i = f(t_i) + \varepsilon_i$, where f is a smooth function, and ε_i is distributed with mean 0 and unknown variance σ^2. It is assumed that f has a bounded square integrable qth derivative, and a squared norm for f is defined by $\|f^{(q)}\|^2 = \int_a^b |f^{(q)}(t)|^2\, dt$. Let $\mathscr{S}_n^q$ be the space of natural polynomial splines of degree $2q - 1$ with knots $\{t_1, \dots, t_n\}$, and $\{\varphi_1, \dots, \varphi_n\}$ be the basis introduced by Demmler-Reinsch [6]. If $f = \sum \beta_k \varphi_k \in$

1991 *Mathematics Subject Classification*: Primary 62C20; Secondary 62G07.
Key words and phrases: nonparametric regression, quadratic loss function, spline smoothing.

$\mathscr{S}_n^q$, then $\|f^{(q)}\|^2 = \sum \lambda_k \beta_k^2$, where λ_k is the norm for φ_k for $1 \leq k \leq n$. Note that $0 = \lambda_1 = \cdots = \lambda_q < \lambda_{q+1} \leq \cdots \leq \lambda_n$, and that $\{\varphi_1, \dots, \varphi_q\}$ spans the space of polynomials of degree q. Thus a restriction such as $\sum_{k=q+1}^n \lambda_k \beta_k^2 \leq r\sigma^2$ implies that a tradeoff between residual variance and deviation of residuals derived from removing polynomial trend is governed by r. Estimating the function f is referred to as curve estimation. Cubic spline smoothing approach was also discussed by Reinsch [18], Wahba [23], and an excellent survey was given by Silverman [20]. The minimax approach for curve estimation was introduced by Speckman [21]. A study of comparing these two approaches was made by Carter-Eagleson [3]. The estimation problems of the mean when it is assumed to lie in a hyperrectangle or more generally a quadratically convex set were discussed by Donoho-Liu-MacGibbon [7]. One-dimensional problems are referred to as a bounded mean, which were discussed by Casella-Strawderman [4] and Bickel [1].

In a usual regression analysis, researcher's attention is paid mostly on the estimation of the mean, and the variance is usually estimated in terms of the sum of squares of the residuals. On the contrary, if a well-performing estimator of variance is available in advance, we can expect that it often provides more reliable inference on the mean. Actually σ^2 is required to explore smoothing parameter choice for curve estimation; see Craven-Wahba [5], Silverman [20], and Hall-Titterington [13]. In nonparametric regression, difference-based estimators were proposed by Rice [19], Gasser-Sroka-Jennue-Steimmetz [11]. Also the asymptotically minimum mean squared error and optimal convergence rate of estimators were discussed by Buckley-Eagleson-Silverman [2], Hall-Marron [14] and Hall-Kay-Titterington [12]. Ohtaki [17] provided a multivariate extension of the estimators proposed by Gasser-Sroka-Jennue-Steimmetz [11] and Ohtaki [16].

The present paper deals with minimax estimation among various methods of estimating the variance in advance to the estimation of the mean. Minimax estimators are defined by choosing an estimator for which the maximum of expected loss over a given parameter space is as small as possible. We will restrict our consideration on a class of properly selected estimators which satisfies a particular criterion. Nevertheless, such a class may occasionally contain the best estimator over wider class of estimators. Minimax estimators of variance was discussed first by Buckley-Eagleson-Silverman [2] and Fujioka [9, 10] followed them up. In their discussion estimators of σ^2 are assumed to be quadratic in $y_1, \dots, y_n$ and to satisfy a bias condition $\mathrm{E}\{\hat{\sigma}^2\} - \sigma^2 = 0$ for all $\sigma^2 > 0$ at $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) = \mathbf{0}$. The resultant estimator takes the form

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^n q_j y_j^2}{\sum_{j=1}^n q_j}$$

with $q_j \geq 0$ for $1 \leq j \leq n$. Estimators of the above form often appear in a nonparametric regression. For example, the kernel-type estimators by Hall-Marron [14] and the difference-based estimators by Hall-Kay-Titterington [12] take the above form. These estimators have asymptotically optimal properties. However, as is shown in § 2, an estimator of the form

$$\hat{\sigma}^2(x; q_1, \dots, q_n) = \frac{\sum_{j=1}^n q_j y_j^2}{\sum_{j=1}^n q_j(1 + x/\lambda_j)} \tag{2}$$

is superior to $\hat{\sigma}^2(0; q_1, \dots, q_n)$ over all $\boldsymbol{\mu} \in \mathbf{R}^n$ and $\sigma^2 > 0$ in view of the quadratic loss function.

The choice (2) is also reasonable in both Bayesian and non-Bayesian frameworks. Assume that a prior distribution on $\boldsymbol{\mu}$ is $N(\mathbf{0}, x\sigma^2 \Lambda^{-1})$, where Λ is diagonal with diagonal components $0 < \lambda_1 < \cdots < \lambda_n$. Then it holds that $\mathrm{E}\{\hat{\sigma}^2(x; q_1, \dots, q_n)\} = \sigma^2$ for all $\sigma^2 > 0$ which is a property of unbiasedness. Further, the underlying assumption of smoothness (1) may be expressed in terms of the prior distribution on $\boldsymbol{\mu}$; see Lindley-Smith [15], and Efron-Morris [8]. On the other hand, in non-Bayesian frame-work, such a form (2) will remove the bias $\mathrm{E}\{\hat{\sigma}^2(x; q_1, \dots, q_n)\} - \sigma^2 = \sum_{j=1}^n q_j(\mu_j^2 - x\sigma^2/\lambda_j)/(\sum_{j=1}^n q_j(1 + x/\lambda_j))$ for all $\sigma^2 > 0$ at $\boldsymbol{\mu} = (\pm\sqrt{x\sigma^2/\lambda_1}, \dots, \pm\sqrt{x\sigma^2/\lambda_n})$.

In our study, an explicit expression of the minimax estimator of σ^2 based on the form (2) is obtained, and its risk properties are derived. The present paper is organized in the following way. In § 2, the notion of minimaxity is presented, and the superiority of our estimators is demonstrated. In § 3, an alternative proof for $x = 0$ is presented. The argument is also valid for $x > 0$. In § 4, the minimax solution for $0 < x \leq 2/(\sum \lambda_l^{-1})$ is obtained, and it is shown that the minimax risk of minimax estimators is a simple function of minimax solutions. In § 5, five special cases are described in details. Especially in the case $r = +0$, the best minimax estimator among $0 < x \leq 2/(\sum \lambda_l^{-1})$ is obtained when $x = 2/(\sum \lambda_l^{-1})$, and the resultant estimator $\hat{\sigma}^2 = \sum y_i^2/(n+2)$ is the minimum mean squared error estimator. The minimax estimator when the mean is known to lie in a sphere is also discussed. In this case, the minimax solution is obtained for any x, and the best minimax estimator over $0 \leq x < \infty$ is shown to be $\hat{\sigma}^2 = \sum y_i^2/(n + 2 + r/2)$.

2. A minimax approach

In this section, a minimax approach to the estimation problem of σ^2 is introduced. Unknown parameters $(\sigma^2, \boldsymbol{\mu})$ are assumed to be contained in the following subset of $\mathbf{R}_+ \times \mathbf{R}^n$,

$$\Theta := \left\{ (\sigma^2, \boldsymbol{\mu}) \in \mathbf{R}_+ \times \mathbf{R}^n \,\middle|\, \sum \lambda_i \mu_i^2 \le r\sigma^2, \sigma^2 > 0 \right\}. \tag{3}$$

Our interest is to estimate σ^2 when the mean is known to lie in an ellipsoid. We consider a quadratic loss function defined by

$$L(\hat{\sigma}^2, \sigma^2) = \left(\frac{\hat{\sigma}^2}{\sigma^2} - 1 \right)^2.$$

The risk function is denoted as $\mathrm{RL}(\hat{\sigma}^2; \sigma^2, \boldsymbol{\mu}) = \mathrm{E}\{L(\hat{\sigma}^2, \sigma^2) \mid \sigma^2, \boldsymbol{\mu}\}$.

The quantity $\sigma^4 \mathrm{RL}(\hat{\sigma}^2; \sigma^2, \boldsymbol{\mu})$ is the mean squared error of $\hat{\sigma}^2$, and can be decomposed into the sum of the squared bias and the variance of $\hat{\sigma}^2$. The estimator $\hat{\sigma}^2(x; q_1, \dots, q_n)$ is quadratic in in the sample y_i. Therefore, its bias and variance are straightforward. The risk function can be explicitly written as

$$\mathrm{RL}(\hat{\sigma}^2(x; q_1, \dots, q_n); \sigma^2, \boldsymbol{\mu}) = \frac{(\sum q_i \mu_i^2/\sigma^2 - x \sum q_i/\lambda_i)^2 + 4 \sum q_i^2 \mu_i^2/\sigma^2 + 2 \sum q_i^2}{(\sum q_i(1 + x/\lambda_i))^2}. \tag{4}$$

The following proposition shows that the estimator $\hat{\sigma}^2(x; q_1, \dots, q_n)$ has smaller risk than the estimator $\hat{\sigma}^2(0; q_1, \dots, q_n)$.

PROPOSITION 1. (i) *If* $(\sum q_i)^2 - 2 \sum q_i^2 \le 0$, *it holds that*

$$\mathrm{RL}(\hat{\sigma}^2(x; q_1, \dots, q_n); \sigma^2, \boldsymbol{\mu}) < \mathrm{RL}(\hat{\sigma}^2(0; q_1, \dots, q_n); \sigma^2, \boldsymbol{\mu})$$

on the whole space of $\boldsymbol{\mu}$ *and* σ^2 *for any* $x > 0$.

(ii) *If* $(\sum q_i)^2 - 2 \sum q_i^2 > 0$, *there exists an upper bound* $x^*(q_1, \dots, q_n) > 0$ *such that*

$$\mathrm{RL}(\hat{\sigma}^2(x; q_1, \dots, q_n); \sigma^2, \boldsymbol{\mu}) < \mathrm{RL}(\hat{\sigma}^2(0; q_1, \dots, q_n); \sigma^2, \boldsymbol{\mu})$$
$$\textit{for } 0 < x < x^*(q_1, \dots, q_n)$$

on the whole space of $\boldsymbol{\mu}$ *and* σ^2.

(iii) *The above upper bound* $x^*(q_1, \dots, q_n)$ *can be chosen as* $4\lambda_1/(n-2)$ *uniformly in* $(q_1, \dots, q_n) \in \mathbf{R}_+^n$.

PROOF. (i) From the expression (4), we can express the risk difference as

$$\mathrm{RL}(\hat{\sigma}^2(0; q_1, \dots, q_n); \sigma^2, \boldsymbol{\mu}) - \mathrm{RL}(\hat{\sigma}^2(x; q_1, \dots, q_n); \sigma^2, \boldsymbol{\mu})$$
$$= \frac{(\sum q_i/\lambda_i) x (A + Bx)}{\{\sum q_i \sum q_i(1 + x/\lambda_i)\}^2}$$

where

$$A = 2\Big(\sum q_i\Big)\Big\{\Big(\sum q_i\mu_i^2/\sigma^2\Big)^2 + 4\sum q_i^2\mu_i^2/\sigma^2 + \Big(\sum q_i\Big)\Big(\sum q_i\mu_i^2/\sigma^2\Big) + 2\sum q_i^2\Big\},$$

$$B = \Big(\sum q_i/\lambda_i\Big)\Big\{\Big(\sum q_i\mu_i^2/\sigma^2\Big)^2 + 4\sum q_i^2\mu_i^2/\sigma^2 + 2\sum q_i^2 - \Big(\sum q_i\Big)^2\Big\}.$$

Since both A and B are minimized at $\boldsymbol{\mu} = 0$, it holds that $A + xB \geq A_0 + xB_0$ on the whole space of $\boldsymbol{\mu}$ and σ^2, where $A_0 = 4(\sum q_i)(\sum q_i^2)$ and $B_0 = (\sum q_i/\lambda_i)(2\sum q_i^2 - (\sum q_i)^2)$. If $(\sum q_i)^2 - 2\sum q_i^2 \leq 0$, then $A_0 > 0$ and $B_0 \geq 0$. Thus it holds that $A + xB > 0$ for any $x > 0$.

(ii) If $(\sum q_i)^2 - 2\sum q_i^2 > 0$, then $A_0 > 0$ and $B_0 < 0$. Letting $x^*(q_1, \dots, q_n) = -A_0/B_0$, we have the assertion.

(iii) The above upper bound $x^*(q_1, \dots, q_n)$ is expressed as

$$x^*(q_1, \dots, q_n) = \frac{4(\sum q_i)(\sum q_i^2)}{(\sum q_i/\lambda_i)\big((\sum q_i)^2 - 2\sum q_i^2\big)}.$$

The two inequalities $\lambda_1 \sum q_i/\lambda_i \leq \sum q_i \leq \lambda_n \sum q_i/\lambda_i$ and $\sum q_i^2 \leq (\sum q_i)^2 \leq n\sum q_i^2$ for any $(q_1, \dots, q_n) \in \mathbf{R}_+^n$ lead to $x^*(q_1, \dots, q_n) \geq 4\lambda_1/(n-2)$. This completes the proof.

Next, we will discuss the maximum of risk function of $\hat{\sigma}^2(x; q_1, \dots, q_n)$ over the parameter space Θ

$$\overline{M}(x; q_1, \dots, q_n) := \max_{(\sigma^2, \boldsymbol{\mu}) \in \Theta} \mathrm{RL}(\hat{\sigma}^2(x, q_1, \dots, q_n); \sigma^2, \boldsymbol{\mu}).$$

An adequate estimator will be expected to have smaller risk uniformly on the whole parameter space Θ. We introduce an approach of choosing an estimator for which the maximum of its risk function is as small as possible.

Definition 1. The estimator $\hat{\sigma}^2(x; \tilde{q}_1, \dots, \tilde{q}_n)$ is said to be *minimax* if $(\tilde{q}_1, \dots, \tilde{q}_n)$ achieves the minimum of $\overline{M}(x; q_1, \dots, q_n)$ among all estimators of the form $\hat{\sigma}^2(x; q_1, \dots, q_n)$. We call the value of $\overline{M}(x; \tilde{q}_1, \dots, \tilde{q}_n)$ *the minimax risk of* $\hat{\sigma}^2(x; \tilde{q}_1, \dots, \tilde{q}_n)$. In addition, we will call $(\tilde{q}_1, \dots, \tilde{q}_n)$ *the minimax solution.*

For any given σ^2, the parameter space (3) can be regarded as a polyhedron in the space of $\mu_1^2, \dots, \mu_n^2$. Let $\boldsymbol{\mu}^{(0)} := (0, \dots, 0)$, and $\boldsymbol{\mu}^{(i)} := (0, \dots, 0, \mu_i^{(i)}, 0, \dots, 0)$, where $\mu_i^{(i)} = \sqrt{r\sigma^2/\lambda_i}$ for $1 \leq i \leq n$. These points lie on the coordinate axes of $\mathbf{R}^n$, and correspond to extreme points of the above polyhedron. Since the risk function (4) is a convex function of $\mu_1^2, \dots, \mu_n^2$, it is maximized at one of the extreme points of the convex polyhedron. Thus we can rewrite

$$\overline{M}(x; q_1, \dots, q_n) = \max_{0 \le i \le n} \mathrm{RL}(\hat{\sigma}^2(x, q_1, \dots, q_n); \sigma^2, \boldsymbol{\mu}^{(i)}). \tag{5}$$

For $x = 0$, the maximum of the risk function of $\hat{\sigma}^2(x; q_1, \dots, q_n)$ over Θ will be achieved at one of n points $\boldsymbol{\mu}^{(i)}$ for $1 \le i \le n$. In fact, Buckley-Eagleson-Silverman [2] discussed a minimax approach based on

$$\overline{M}(0; q_1, \dots, q_n) = \max_{1 \le i \le n} \mathrm{RL}(\hat{\sigma}^2(0, q_1, \dots, q_n); \sigma^2, \boldsymbol{\mu}^{(i)}). \tag{6}$$

Moreover, from (4) it is seen that $\mathrm{RL}(\hat{\sigma}^2(x, q_1, \dots, q_n); \sigma^2, \boldsymbol{\mu}^{(i)})$ is a continuous function of x. We will restrict our attention on x for which we can express $\overline{M}(x; q_1, \dots, q_n)$ as

$$\overline{M}(x; q_1, \dots, q_n) = \max_{1 \le i \le n} \frac{\left(rq_i/\lambda_i - x \sum_{l=1}^n q_l/\lambda_l\right)^2 + 4rq_i^2/\lambda_i + 2\sum_{l=1}^n q_l^2}{\left(\sum_{l=1}^n q_l(1 + x/\lambda_l)\right)^2}. \tag{7}$$

Actually, this requires that $0 < x \le 2/(\sum \lambda_l^{-1})$, which will be assumed in our results.

Next, following Buckley-Eagleson-Silverman [2], the problem of minimizing $\overline{M}(x; q_1, \dots, q_n)$ will be reduced to that of minimizing a quadratic form of $q_1, \dots, q_n$, in terms of the Lagrange multiplier. Define a function

$$\begin{aligned} \overline{L}(x; q_1, \dots, q_n) := \max_{1 \le i \le n} &\left(\left(\frac{rq_i}{\lambda_i} - x \sum_{l=1}^n \frac{q_l}{\lambda_l} \right)^2 + \frac{4rq_i^2}{\lambda_i} \right) \\ &+ 2 \sum_{l=1}^n q_l^2 - \eta \sum_{l=1}^n q_l \left(1 + \frac{x}{\lambda_l} \right) \end{aligned}$$

with a Lagrange mutiplier η. Note that the problem of minimizing $\overline{M}(x; q_1, \dots, q_n)$ is invariant under multiplying $q_1, \dots, q_n$ by a constant. Replacing q_j by $\eta q_j/(4\sigma^4)$ and multiplying $\overline{L}$ by $16\sigma^4/\eta^2$, we obtain

$$\begin{aligned} &(16\sigma^4/\eta^2)\overline{L}(x; \eta q_1/(4\sigma^4), \dots, \eta q_n/(4\sigma^4)) \\ &= \max_{1 \le i \le n} \left(\left(\frac{rq_i}{\lambda_i} - x \sum_{l=1}^n \frac{q_l}{\lambda_l} \right)^2 + \frac{4rq_i^2}{\lambda_i} \right) + 2 \sum_{l=1}^n q_l^2 - 4 \sum_{l=1}^n q_l \left(1 + \frac{x}{\lambda_l} \right). \end{aligned}$$

Consequently, our problem can be reduced to that of minimizing

$$\overline{H}(x; q_1, \dots, q_n) = \max_{1 \le i \le n} H_i(x; q_1, \dots, q_n)$$

with respect to $(q_1, \dots, q_n) \in \mathbf{R}_+^n$, where

$$H_i(x; q_1, \dots, q_n) = \left(\frac{rq_i}{\lambda_i} - x\sum_{l=1}^{n}\frac{q_l}{\lambda_l}\right)^2 + \frac{4rq_i^2}{\lambda_i} - 4x\sum_{l=1}^{n}\frac{q_l}{\lambda_l} + 2\sum_{l=1}^{n}(q_l^2 - 2q_l).$$

3. Minimax solution for $x = 0$

The minimax estimator of the form $\hat{\sigma}^2(0; q_1, \dots, q_n)$ is obtained by minimizing

$$\overline{H}(0; q_1, \dots, q_n) = r^2 \max_{1 \le i \le n} (q_i/\lambda_i(r))^2 + 2\sum_{l=1}^{n}(q_l^2 - 2q_l)$$

with $\lambda_i(r) = \lambda_i(1 + 4\lambda_i/r)^{-1/2}$ for $i = 1, \dots, n$ and $r > 0$. Buckley-Eagleson-Silverman [2] at first minimized $\overline{H}(0; q_1, \dots, q_n)$ over the space $\{(q_1, \dots, q_n) \mid \max_{1 \le i \le n}(q_i/\lambda_i(r)) = \alpha\}$ for fixed $\alpha > 0$ and then proposed using numerical computations for the remaining minimization over $\alpha > 0$. Thus their expression of minimax estimators involves a redundant quantity α. Fujioka [9] followed up this work, and gave an explicit expression of the minimax estimator. However, the minimax problem for $x > 0$ is so complicated that their technique is no longer applicable.

We give an alternative argument valid for obtaining the minimax solution for $x > 0$. We begin with summarizing the result for $x = 0$.

THEOREM 1. *Let $r^{(i)}$ be a unique solution of an equation $r^2 = 2\sum_{l=1}^{i}\lambda_l(r)$ $(\lambda_i(r) - \lambda_l(r))$ for $2 \le i \le n$, and put $r^{(1)} = 0$ and $r^{(n+1)} = \infty$. Then the minimax solution for $r^{(i)} < r \le r^{(i+1)}$ $(1 \le i \le n)$ is expressed as*

$$\tilde{q}_j(r) = \begin{cases} \dfrac{2\sum_{l=1}^{i}\lambda_l(r)}{r^2 + 2\sum_{l=1}^{i}\lambda_l(r)^2}\lambda_j(r) & \text{for } j \le i \\ 1 & \text{for } j > i. \end{cases}$$

PROOF. Rewrite the function $\overline{H}(0; q_1, \dots, q_n)$ as

$$\overline{H}(0; q_1, \dots, q_n) = r^2\left(\max_{1 \le i \le n}\frac{q_i}{\lambda_i(r)}\right)^2 + 2\sum_{l=1}^{n}(q_l^2 - 2q_l).$$

We consider the problem of minimizing $\overline{H}(0; q_1, \dots, q_n)$ with respect to $(q_1, \dots, q_n) \in \mathbf{R}_+^n$. Note that $H_i(0; q_1, \dots, q_n) \ge H_l(0; q_1, \dots, q_n)$ is equivalent to $q_i/\lambda_i(r) \ge q_j/\lambda_j(r)$ for $i \ne j$. Define subsets of $\mathbf{R}_+^n$ by

$$\begin{aligned} S_i := &\{(q_1, \dots, q_n) \in \mathbf{R}_+^n \mid \overline{H}(0; q_1, \dots, q_n) = H_i(0; q_1, \dots, q_n)\} \\ = &\{(q_1, \dots, q_n) \in \mathbf{R}_+^n \mid q_i/\lambda_i(r) \ge q_l/\lambda_l(r) \text{ for } 1 \le l \le n\}. \end{aligned}$$

From the fact that $\mathbf{R}_+^n = \bigcup_{i=1}^n S_i$ together with the definition of S_i, the problem of minimizing $\overline{H}(0; q_1, \ldots, q_n)$ on $\mathbf{R}_+^n$ can be reduced to that of seeking for $(q_1, \ldots, q_n)$ which attains

$$\min_{1 \le i \le n} \min_{(q_1, \ldots, q_n) \in S_i} \overline{H}(0; q_1, \ldots, q_n)$$

or equivalently,

$$\min_{1 \le i \le n} \min_{(q_1, \ldots, q_n) \in S_i} H_i(0; q_1, \ldots, q_n).$$

It is easily seen that the function $H_i(0; q_1, \ldots, q_n)$ is minimized over $\mathbf{R}_+^n$ when $q_j = q_j^{(i)}(r)$ with

$$q_j^{(i)}(r) = \begin{cases} \dfrac{2\lambda_i(r)^2}{r^2 + 2\lambda_i(r)^2} & \text{for } j = i \\ 1 & \text{for } j \ne i \end{cases}$$

for $1 \le i \le n$. If $(q_1^{(i)}(r), \ldots, q_n^{(i)}(r))$ belongs to S_i, the function $H_i(0; q_1, \ldots, q_n)$ is minimized over S_i at that point. Unless $(q_1^{(i)}(r), \ldots, q_n^{(i)}(r))$ belongs to S_i, the function $H_i(0; q_1, \ldots, q_n)$ is minimized over S_i at one of the boundary points of S_i. Since the contours of $H_i(0; q_1, \ldots, q_n)$ are ellipsoid surfaces with its center at $(q_1^{(i)}(r), \ldots, q_n^{(i)}(r))$, the location of $(q_1^{(i)}(r), \ldots, q_n^{(i)}(r))$ is substantial for the problems of minimizing $H_i(0; q_1, \ldots, q_n)$. It is straightforward that $q_{i-1}^{(i)}(r)/\lambda_{i-1}^{(i)}(r) > q_i^{(i)}(r)/\lambda_i^{(i)}(r)$ or equivalently,

$$H_{i-1}(0; q_1^{(i)}(r), \ldots, q_n^{(i)}(r)) > H_i(0; q_1^{(i)}(r), \ldots, q_n^{(i)}(r)) \quad \text{for } 2 \le i \le n. \tag{8}$$

Therefore, $(q_1^{(i)}(r), \ldots, q_n^{(i)}(r)) \notin S_i$ for $2 \le i \le n$. This implies that the minimum of $H_i(0; q_1, \ldots, q_n)$ over S_i is achieved at a point on the boundary of S_i. By the definition of S_i for $1 \le i \le n$, S_i is the intersection of $n-1$ half-spaces in $\mathbf{R}_+^n$, and can be expressed as

$$S_i = \bigcap_{l \ne i} \{(q_1, \ldots, q_n) \in \mathbf{R}_+^n \mid H_i(0; q_1, \ldots, q_n) \ge H_l(0; q_1, \ldots, q_n)\}.$$

The boundary of S_i is a union of hypersurfaces $\{(q_1, \ldots, q_n) \in \mathbf{R}_+^n \mid H_i(0; q_1, \ldots, q_n) = H_k(0; q_1, \ldots, q_n) \ge H_l(0; q_1, \ldots, q_n) \text{ for } l \ne k, i\}$ for $k \ne i$. Moreover, $(q_1^{(i)}(r), \ldots, q_n^{(i)}(r))$ satisfies inequalities $q_1^{(i)}(r)/\lambda_1(r) > \cdots > q_{i-1}^{(i)}(r)/\lambda_{i-1}(r) > q_{i+1}^{(i)}(r)/\lambda_{i+1}(r) > \cdots > q_n^{(i)}(r)/\lambda_n(r)$ or equivalently,

$$\begin{aligned} H_1(0; q_1^{(i)}(r), \ldots, q_n^{(i)}(r)) > \cdots > H_{i-1}(0; q_1^{(i)}(r), \ldots, q_n^{(i)}(r)) \\ > H_{i+1}(0; q_1^{(i)}(r), \ldots, q_n^{(i)}(r)) > \cdots > H_n(0; q_1^{(i)}(r), \ldots, q_n^{(i)}(r)). \end{aligned} \tag{9}$$

From the inequalities (8) and (9), it follows that $(q_1^{(i)}(r), \ldots, q_n^{(i)}(r)) \in S_1$ for $2 \le i \le n$. Since the contours of $H_i\ (0; q_1, \ldots, q_n)$ are ellipsoid surfaces with center $(q_1^{(i)}(r), \dots, q_n^{(i)}(r))$ in S_1, the minimum of $H_i\ (0; q_1, \dots, q_n)$ over the hypersurface

$$\{(q_1, \dots, q_n) \in \mathbf{R}_+^n \mid H_i(0; q_1, \dots, q_n) = H_k(0; q_1, \dots, q_n) \\ \geq H_l(0; q_1, \dots, q_n) \text{ for } l \neq k, i\}$$

is achieved at a point within S_1 for $2 \leq i \leq n$. Therefore, the minimum of $H_i(0; q_1, \dots, q_n)$ over S_i is achieved at a point on the boundary of S_1 for $2 \leq i \leq n$. Note that values of two functions $H_i(0; q_1, \dots, q_n)$ and $H_1(0; q_1, \dots, q_n)$ coincide with each other on the boundary $S_1 \cap S_i$. Consequently, we have only to seek for $(q_1, \dots, q_n)$ which attains the minimum of $H_1(0; q_1, \dots, q_n)$ over S_1. Again, note that the location of $(q_1^{(1)}(r), \dots, q_n^{(1)}(r))$ is substantial for minimizing $H_1(0; q_1, \dots, q_n)$. Since $(q_1^{(1)}(r), \dots, q_n^{(1)}(r))$ satisfies

$$H_2(0; q_1^{(1)}(r), \dots, q_n^{(1)}(r)) > \cdots > H_n(0; q_1^{(1)}(r), \dots, q_n^{(1)}(r)),$$

it follows that $(q_1^{(1)}(r), \dots, q_n^{(1)}(r))$ lies in $S_1 \cup S_2$. If $r \leq r^{(2)}$, it holds that $q_1^{(1)}(r)/\lambda_1(r) \geq q_2^{(1)}(r)/\lambda_2(r)$. Therefore, $(q_1^{(1)}(r), \dots, q_n^{(1)}(r))$ lies in S_1, and minimizes $H_1(0; q_1, \dots, q_n)$ over S_1. If $r^{(2)} < r$, the point $(q_1^{(1)}(r), \dots, q_n^{(1)}(r))$ satisfies

$$H_2(0; \tilde{q}_1^{(1)}(r), \dots, \tilde{q}_n^{(1)}(r)) > H_1(0; \tilde{q}_1^{(1)}(r), \dots, \tilde{q}_n^{(1)}(r)),$$

and lies outside of S_1. Thus the minimum of $H_1(0; q_1, \dots, q_n)$ over S_1 is achieved at a point on $S_1 \cap S_2$. We can obtain the point achieving the minimum of $H_1(0; q_1, \dots, q_n)$ over the hypersurface $\{(q_1, \dots, q_n) \in \mathbf{R}_+^n \mid H_1(0; q_1, \dots, q_n) = H_2(0; q_1, \dots, q_n)\}$ as

$$\tilde{q}_j^{(2)}(r) = \begin{cases} \dfrac{2\lambda_1(r) + 2\lambda_2(r)}{r^2 + 2\lambda_1(r)^2 + 2\lambda_2(r)^2} \lambda_j(r) & \text{for } j \leq 2 \\ 1 & \text{for } j > 2. \end{cases}$$

If $r^{(2)} \leq r \leq r^{(3)}$, $(\tilde{q}_1^{(2)}(r), \dots, \tilde{q}_n^{(2)}(r))$ lies in S_1, and minimizes $H_1(0; q_1, \dots, q_n)$ over S_1. On the other hand, if $r^{(3)} < r$, it holds that

$$H_1(0; \tilde{q}_1^{(2)}(r), \dots, \tilde{q}_n^{(2)}(r)) = H_2(0; \tilde{q}_1^{(2)}(r), \dots, \tilde{q}_n^{(2)}(r)) < H_3(0; \tilde{q}_1^{(2)}(r), \dots, \tilde{q}_n^{(2)}(r)).$$

Thus the point $(\tilde{q}_1^{(2)}(r), \dots, \tilde{q}_n^{(2)}(r))$ lies outside of $S_1 \cap S_2$, and the minimum of $H_1(0; q_1, \dots, q_n)$ over S_1 is achieved at a point on $S_1 \cap S_2 \cap S_3$. This procedure can be repeated as follows. Let

$$\tilde{q}_j^{(i)}(r) = \begin{cases} \dfrac{2\sum_{l=1}^{i} \lambda_l(r)}{r^2 + 2\sum_{l=1}^{i} \lambda_l(r)^2} \lambda_j(r) & \text{for } j \leq i \\ 1 & \text{for } j > i \end{cases}$$

for $2 \leq i \leq n$. If $r^{(i)} < r \leq r^{(i+1)}$, we have

$$(\tilde{q}_1^{(i)}(r), \ldots, \tilde{q}_n^{(i)}(r)) = \arg \min_{(q_1, \ldots, q_n) \in S_1} H_1(0; q_1, \ldots, q_n).$$

In fact, if $r^{(i)} < r \leq r^{(i+1)}$, then the minimum of H_1 over $\{(q_1, \ldots, q_n) \in \mathbf{R}_+^n \mid H_1(0; q_1, \ldots, q_n) = \cdots = H_i(0; q_1, \ldots, q_n)\}$ is achieved at $(\tilde{q}_1^{(i)}(r), \ldots, \tilde{q}_n^{(i)}(r))$, and $(\tilde{q}_1^{(i)}(r), \ldots, \tilde{q}_n^{(i)}(r))$ lies in $\bigcap_{l=1}^{i} S_l$. Now let S_{1i} be

$$\{(q_1, \ldots, q_n) \in \mathbf{R}_+^n \mid H_1(0; q_1, \ldots, q_n) = \cdots = H_i(0; q_1, \ldots, q_n) \\ \geq \cdots \geq H_n(0; q_1, \ldots, q_n)\}$$

for $1 \leq i \leq n$. It is concluded that if $r^{(i)} < r \leq r^{(i+1)}$, we have

$$\begin{aligned} \arg \min_{(q_1, \ldots, q_n) \in \mathbf{R}_+^n} \overline{H}(0; q_1, \ldots, q_n) &= \arg \min_{(q_1, \ldots, q_n) \in S_1} H_1(0; q_1, \ldots, q_n) \\ &= \arg \min_{(q_1, \ldots, q_n) \in S_{1i}} H_1(0; q_1, \ldots, q_n) \\ &= (\tilde{q}_1^{(i)}(r), \ldots, \tilde{q}_n^{(i)}(r)) \end{aligned}$$

for $1 \leq i \leq n$. The minimax solutions $\tilde{q}_j^{(i-1)}(r)$ and $\tilde{q}_j^{(i)}(r)$ are calculated in adjacent intervals of r, and also they coincide with each other at $r = r^{(i)}$. Thus $r^{(i)}$ can be obtained as a solution of the equation $\tilde{q}_1^{(i-1)}(r) = \tilde{q}_1^{(i)}(r)$. This completes the proof.

The way of obtaining the minimax solution for $x = 0$ is summarized as the following three stages.
Stage 1: To compute the quantities

$$(q_1^{(i)}(r), \ldots, q_n^{(i)}(r)) = \arg \min_{(q_1, \ldots, q_n) \in \mathbf{R}_+^n} H_i(0; q_1, \ldots, q_n)$$

for $1 \leq i \leq n$.
Stage 2: To check the inequalities

$$H_1(0; q_1^{(i)}(r), \ldots, q_n^{(i)}(r)) > \cdots > H_{i-1}(0; q_1^{(i)}(r), \ldots, q_n^{(i)}(r)) \\ > H_{i+1}(0; q_1^{(i)}(r), \ldots, q_n^{(i)}(r)) > \cdots > H_n(q_1^{(i)}(r), \ldots, q_n^{(i)}(r)) \tag{10}$$

and

$$H_{i-1}(q_1^{(i)}(r), \ldots, q_n^{(i)}(r)) > H_i(0; q_1^{(i)}(r), \ldots, q_n^{(i)}(r)) \tag{11}$$

for $1 \leq i \leq n$.

Stage 3: To compute

$$(\tilde{q}_1^{(i)}(r), \dots, \tilde{q}_n^{(i)}(r)) = \arg \min_{(q_1, \dots, q_n) \in S_{1i}} H_1(0; q_1, \dots, q_n)$$

for $1 \leq i \leq n$.

In Stage 1, we specify the centers of ellipsoid surfaces which yield contours of functions $H_i(0; q_1, \dots, q_n)$ for $1 \leq i \leq n$. The inequalities in Stage 2 will give us the locations of the centers obtained in Stage 1. The center $(\tilde{q}_1^{(1)}(r), \dots, \tilde{q}_n^{(1)}(r))$ lies in $S_1 \cup S_2$, and the other centers $(\tilde{q}_1^{(i)}(r), \dots, \tilde{q}_n^{(i)}(r))$ lie in S_1. Thus our problem can be reduced to that of minimizing $H_1(0; q_1, \dots, q_n)$ on S_1, more precisely, that of minimizing $H_1(0; q_1, \dots, q_n)$ on S_{1i} for $1 \leq i \leq n$, which is calculated in Stage 3.

4. Minimax solution for $x > 0$

In this section the minimax solution for $x > 0$ is discussed. In fact our discussion in the proof of Theorem 1 is valid for this general case. The next assumption is required in order to apply the first two stages described in the end of §3.

ASSUMPTION 1.

$$0 < x \leq 2 \Bigg/ \left(\sum_{l=1}^{n} \lambda_l^{-1} \right).$$

The following lemma gives us $(q_1, \dots, q_n)$ which achieves the minimum of $H_i(x; q_1, \dots, q_n)$ over $\mathbf{R}_+^n$ for $1 \leq i \leq n$, which is the first stage for the case $x > 0$.

LEMMA 1. Under Assumption 1, $H_i(x; q_1, \dots, q_n)$ is minimized over $\mathbf{R}_+^n$ when $q_j = q_j^{(i)}(x, r)$ with

$$q_j^{(i)}(x, r) = \begin{cases} \dfrac{\lambda_i + r}{\lambda_i + 2r} + \dfrac{x - r}{\lambda_i + 2r} \xi^{(i)}(x, r) & \text{for } j = i \\ 1 + \dfrac{x}{\lambda_j} \xi^{(i)}(x, r) & \text{for } j \neq i \end{cases}$$

where

$$\xi^{(i)}(x, r) = \frac{(r^2 + 5\lambda_i r + 2\lambda_i^2) w - ((w + 2\lambda_i w) r + \lambda_i w + \lambda_i^2 v) x}{(2\lambda_i r + \lambda_i^2 + w) x^2 - 2wrx + (r^2 + 4\lambda_i r + 2\lambda_i^2) w}$$

with $v = (\sum_{l \neq i} \lambda_l^{-1})(\sum_{l \neq i} \lambda_l^{-2})^{-1}$ and $w = (\sum_{l \neq i} \lambda_l^{-2})^{-1}$.

PROOF. First, let $\delta = \sum_{j=1}^n q_j / \lambda_j$ and $u = (q_i / \lambda_i)(\sum_{j=1}^n q_j / \lambda_j)^{-1}$. Then the quantity q_i is expressed as $q_i = \delta u \lambda_i$, and the other q_j's satisfy the relation

$\sum_{j\neq i} q_j/\lambda_j = \delta(1-u)$. Further, we have

$$\begin{aligned} &H_i(x; q_1, \dots, q_{i-1}, \delta u\lambda_i, q_{i+1}, \dots, q_n) \\ &\quad = \delta^2\{(ru-x)^2 + (4r\lambda_i + 2\lambda_i^2)u^2\} - 4\delta(x+\lambda_i u) + 2\sum_{j\neq i}(q_j^2 - 2q_j). \end{aligned}$$

The minimum of the last term in the above expression of $H_i(x; q_1, \dots, q_n)$ with respect to q_j $(j \neq i)$ under the constraint $\sum_{j\neq i} q_j/\lambda_j = \delta(1-u)$ is attained at $q_j = q_j(u, \delta)$ with

$$q_j(u,\delta) = 1 + (w\delta(1-u) - v)\lambda_j^{-1} \quad \text{for } j \neq i,$$

and the minimum value is $2(w\delta(1-u)-v)^2 w^{-1} - 2(n-1)$. Then we have

$$\begin{aligned} &H_i(x; q_1(u,\delta), \dots, q_{i-1}(u,\delta), \delta u\lambda_i, q_{i+1}(u,\delta), \dots, q_n(u,\delta)) \\ &\quad = \delta^2\{(ru-x)^2 + 4r\lambda_i u^2 + 2\lambda_i^2 u^2 + 2w(1-u)^2\} \\ &\quad\quad - 4\delta(x + \lambda_i u + v(1-u)) + 2v^2 w^{-1} - 2(n-1). \end{aligned}$$

Next, the minimum of $H_i(x; q_1(u,\delta), \dots, q_{i-1}(u,\delta), \delta u\lambda_i, q_{i+1}(u,\delta), \dots, q_n(u,\delta))$ with respect to δ is attained at $\delta = \delta(u)$ with

$$\delta(u) = \frac{2(\lambda_i - v)u + 2x + 2v}{(ru-x)^2 + (4r\lambda_i + 2\lambda_i^2)u^2 + 2w(1-u)^2},$$

and the minimum is given by $-4G(u) + 2v^2 w^{-1} - 2(n-1)$ where

$$G(u) = \frac{((\lambda_i - v)u + x + v)^2}{(ru-x)^2 + (4r\lambda_i + 2\lambda_i^2)u^2 + 2w(1-u)^2}.$$

Let $\tilde{u} = \arg\max_u G(u)$. Then, $q_j^{(i)}(x,r)$ for $1 \leq j \leq n$ will be obtained as

$$q_j^{(i)}(x,r) = \begin{cases} \delta(\tilde{u})\tilde{u}\lambda_i & \text{for } j = i \\ 1 + (w\delta(\tilde{u})(1-\tilde{u}) - v)\lambda_j^{-1} & \text{for } j \neq i. \end{cases}$$

Now we will seek for $\tilde{u}$ and $\delta(\tilde{u})$. Differentiating the logarithm of $G(u)$ and equating the derivative to zero, we have

$$\frac{\lambda_i - v}{(\lambda_i - v)u + x + v} = \frac{u(r^2 + 4r\lambda_i + 2\lambda_i^2 + 2w) - (rx + 2w)}{(ru-x)^2 + 4r\lambda_i u^2 + 2\lambda_i^2 u^2 + 2w(1-u)^2}. \tag{12}$$

Consequently, we obtain

$$\tilde{u} = \frac{(r + \lambda_i - v)x^2 + (rv + 2w)x + 2\lambda_i w}{(r^2 - rv + 5r\lambda_i + 2\lambda_i^2 + 2w)x + r^2 v + 4r\lambda_i v + 2\lambda_i w + 2\lambda_i^2 v}$$

and

$$1-\tilde{u}=\frac{(-r-\lambda_i+v)x^2+(r^2+5r\lambda_i+2\lambda_i^2-2rv)x+(r^2+4r\lambda_i+2\lambda_i^2)v}{(r^2-rv+5r\lambda_i+2\lambda_i^2+2w)x+r^2v+4r\lambda_iv+2\lambda_iw+2\lambda_i^2v}.$$

Now we must check the condition $0<\tilde{u}<1$. Note that Assumption 1 can be written as $0<x\leq 2\lambda_iw/(w+\lambda_iv)$. The denominator of $\tilde{u}$, equivalently, that of $1-\tilde{u}$ is expressed as

$$(r^2+5\lambda_ir+2\lambda_i^2+2w)x+vr^2+\frac{2\lambda_iv(w+2\lambda_iv)}{w+\lambda_iv}r+2\lambda_iw+2\lambda_i^2v+vr\left(\frac{2\lambda_iw}{w+\lambda_iv}-x\right).$$

The numerators of $\tilde{u}$ and $1-\tilde{u}$ is expressed as

$$(r+\lambda_i)x^2+\left(rv+\frac{2w^2}{w+\lambda_iv}\right)x+2\lambda_iw+vx\left(\frac{2\lambda_iw}{w+\lambda_iv}-x\right)$$

and

$$vx^2+\left(r^2+\frac{3\lambda_iw+5\lambda_i^2v}{w+\lambda_iv}r+\frac{2\lambda_i^3v}{w+\lambda_iv}\right)x$$
$$+vr^2+\frac{4\lambda_i^2v^2}{w+\lambda_iv}r+2\lambda_i^2v+((r+\lambda_i)x+2rv)\left(\frac{2\lambda_iw}{w+\lambda_iv}-x\right),$$

respectively. Hence $0<\tilde{u}<1$ by Assumption 1. The equality (12) leads us to

$$\begin{aligned}\delta(\tilde{u})&=\frac{2(\lambda_i-v)}{(r^2+4r\lambda_i+2\lambda_i^2+2w)\tilde{u}-(rx+2w)}\\&=\frac{(r^2+5r\lambda_i+2\lambda_i^2+2w)x+r^2v+4r\lambda_iv+2\lambda_iw+2\lambda_i^2v}{(2\lambda_ir+\lambda_i^2+w)x^2-2wrx+(r^2+4r\lambda_i+2\lambda_i^2)w}.\end{aligned}$$

Similarly it can be verified that $\delta(\tilde{u})>0$ under Assumption 1. Finally we obtain an explicit expression of $q_j^{(i)}(x,r)$ in terms of $\xi^{(i)}(x,r)$. It can be seen that

$$\delta(\tilde{u})\tilde{u}=\frac{r+\lambda_i}{2r\lambda_i+\lambda_i^2}+\frac{x-r}{2r\lambda_i+\lambda_i^2}\xi^{(i)}(x,r)$$

and

$$w\delta(\tilde{u})(1-\tilde{u})-v=x\xi^{(i)}(x,r).$$

This proves the assertion.

REMARK 1. It can be verified that $\xi^{(i)}(x,r) > 0$ under Assumption 1. Let

$$\tilde{x}(r) := \frac{w(r^2 + 5r\lambda_i + 2\lambda_i^2)}{r(w + 2\lambda_i v) + \lambda_1(w + \lambda_i v_i)}.$$

Then, it holds that $\xi^{(i)}(x,r) > 0$ if and only if $x < \tilde{x}(r)$. In addition, from the inequality $2/(\sum \lambda_l^{-1}) < \tilde{x}(r)$, it is seen that that $\xi^{(i)}(x,r) > 0$ if $0 < x \leq 2/(\sum \lambda_l^{-1})$. Such a property will yield convenient inequalities in the following Lemma 2. Taking the limit, we obtain $\tilde{x}(+0) = 2/(\sum \lambda_l^{-1})$. Thus Assumption 1 is inevitable for Lemma 2.

ASSUMPTION 2.

$$\sum_{l=1}^{n} \lambda_l^{-1} \geq \max_{1\leq i\leq n}\left((\lambda_n - \lambda_i)\sum_{l\neq i} \lambda_l^{-2}\right) \text{ and } \left(\sum_{l=1}^{n} \lambda_l^{-1}\right)^2 \geq 4\left(\sum_{l=1}^{n-1} \lambda_l^{-2}\right).$$

Assumption 2 will avoid extreme unbalance in λ_i's, and exclude unrealistic assignments of λ_i's. It is not a severe restriction on λ_i's unless n is small.

LEMMA 2. Under Assumptions 1–2 the quantities $(q_1^{(i)}(x,r),\dots,q_n^{(i)}(x,r))$ satisfy the inequalities

$$\begin{aligned} &H_1(x; q_1^{(i)}(x,r),\dots,q_n^{(i)}(x,r)) > \cdots > H_{i-1}(x; q_1^{(i)}(x,r),\dots,q_n^{(i)}(x,r)) \\ &\quad > H_{i+1}(x; q_1^{(i)}(x,r),\dots,q_n^{(i)}(x,r)) > \cdots > H_n(x; q_1^{(i)}(x,r),\dots,q_n^{(i)}(x,r)) \end{aligned}$$

and

$$H_{i-1}(x; q_1^{(i)}(x,r),\dots,q_n^{(i)}(x,r)) > H_i(x; q_1^{(i)}(x,r),\dots,q_n^{(i)}(x,r)). \tag{13}$$

PROOF. Recall that

$$\begin{aligned} H_j(x; q_1,\dots,q_n) = \frac{rq_j}{\lambda_j}\left((r + 4\lambda_j)\frac{q_j}{\lambda_j} - 2x\sum_{l=1}^{n}\frac{q_l}{\lambda_l}\right) \\ + x^2\left(\sum_{l=1}^{n}\frac{q_l}{\lambda_l}\right)^2 - 4x\sum_{l=1}^{n}\frac{q_l}{\lambda_l} + \sum_{l=1}^{n}(q_l^2 - 2q_l). \end{aligned}$$

The first term is substantial to prove the assertion. Since $\xi^{(i)}(x,r) > 0$ under Assumption 1 by Remark 1, we have

$$q_1^{(i)}(x,r) > \cdots > q_{i-1}^{(i)}(x,r) > 1 + \frac{x}{\lambda_i}\xi^{(i)}(x,r) > q_{i+1}^{(i)}(x,r) > \cdots > q_n^{(i)}(x,r) > 1$$

and

$$1 + \frac{x}{\lambda_i}\xi^{(i)}(x, r) > q_i^{(i)}(x, r). \tag{14}$$

If we can prove that for $1 \leq j \leq n$

$$\frac{r + 4\lambda_j}{\lambda_j}\left(1 + \frac{x}{\lambda_j}\xi^{(i)}(x, r)\right) - 2x\delta(\tilde{u}) > 0 \tag{15}$$

then the inequalities (14) imply (13). Since the left-hand side of (15) is decreasing in j, it is sufficient to prove (15) for $j = n$. Write the denominator of $\xi^{(i)}(x, r)$ as V. Then V is the denominator of $\tilde{u}$, which is positive by the proof of Lemma 1. Setting $W_1 := (r + 4x\xi^{(i)}(x, r))V$ and $W_2 := (4 - 2x\delta(\tilde{u}))V$, we have

$$\begin{aligned} W_1 &= (2\lambda_i r^2 + (\lambda_i^2 - 3w - 8\lambda_i^2 v)r - 4\lambda_i w - 4\lambda_i^2 v)x^2 \\ &\quad + 2w(r^2 + 10\lambda_i r + 4\lambda_i^2)x + w(r^3 + 4\lambda_i r^2 + 2\lambda_i^2 r), \\ W_2 &= 2r(v - \lambda_i - r)x^2 - (2vr^2 + 8(w + \lambda_i v)r + 4\lambda_i w + 4\lambda_i^2 v)x \\ &\quad + 4w(r^2 + 4\lambda_i r + 2\lambda_i^2). \end{aligned}$$

Thus it follows that

$$\begin{aligned} W_1 + \lambda_n W_2 &= (w + 2\lambda_n v)rx^2 + (r - \lambda_i)^2 rw + \frac{4w(\lambda_n w + \lambda_i^2 v)}{w + \lambda_i v} \\ &\quad + \frac{(2\lambda_i^3 r^2 + \lambda_i^4 r)w^3}{(w + \lambda_i v)^2}\left\{\left(\frac{w + \lambda_i v}{\lambda_i w}\right)^2 - \frac{4}{w}\right\} \\ &\quad + \frac{4wrx}{w + \lambda_i v}\{(w + \lambda_i v + 2\lambda_i^2 - \lambda_i\lambda_n)r + \lambda_i(3w + \lambda_i v + \lambda_i^2 - \lambda_i\lambda_n)\} \\ &\quad + Z_1\left(\frac{2\lambda_i w}{w + \lambda_i v} - x\right)^2 + Z_2\left(\frac{2\lambda_i w}{w + \lambda_i v} - x\right) \end{aligned}$$

where

$$\begin{aligned} Z_1 &= 2\lambda_i r^2 + \lambda_i^2 r, \\ Z_2 &= (2\lambda_n r^2 + (2\lambda_i\lambda_n + 4w + 8\lambda_i v)r + 4w\lambda_i + 4\lambda_i^2 v)x \\ &\quad + 2(w + \lambda_n v)r^2 + 8\lambda_n(w + \lambda_i v)r + 4\lambda_i\lambda_n w + 4\lambda_i^2\lambda_n v. \end{aligned}$$

Assumption 2 implies that $w + \lambda_i v + \lambda_i^2 - \lambda_i\lambda_n \geq 0$ for any i, and that $w - 4\lambda_i^2 w^2/(w + \lambda_i v)^2 \geq 0$. Also, Assumption 1 implies that $2\lambda_i/(w + \lambda_i v) - x \geq 0$. The remaining terms are all positive. Consequently, $W_1 + \lambda_n W_2 \geq 0$. This completes the proof.

Stage 3 described in the end of § 3 implies that we have only to minimize $H_1(x; q_1, \dots, q_n)$ with respect to $q_1, \dots, q_n$ under the constraints

$$\left(\frac{rq_1}{\lambda_1} - x\sum_{j=1}^{n}\frac{q_j}{\lambda_j}\right)^2 + \frac{4rq_1^2}{\lambda_1} = \cdots = \left(\frac{rq_k}{\lambda_k} - x\sum_{j=1}^{n}\frac{q_j}{\lambda_j}\right)^2 + \frac{4rq_k^2}{\lambda_k}. \tag{16}$$

Let $u_i = (q_i/\lambda_i)(\sum_{l=1}^{n} q_l/\lambda_l)^{-1}$ for $1 \le i \le n$. Then, the quantities $u_1, \dots, u_k$ satisfy relation

$$(ru_1 - x)^2 + 4r\lambda_1 u_1^2 = \cdots = (ru_k - x)^2 + 4r\lambda_k u_k^2.$$

Let $c^2r^2 + x^2$ be the common value. Then u_i solves the equation $(ru - x)^2 + 4r\lambda_i u^2 = c^2r^2 + x^2$, whose positive solution is given by

$$u_i(c) = \frac{x + \sqrt{x^2 + r(r + 4\lambda_i)c^2}}{r + 4\lambda_i} \quad \text{for } 1 \le i \le k.$$

Before describing a theorem which gives the minimax solution for $x > 0$ we will prepare some notations. Define sequences by $v_k = (\sum_{l=k+1}^{n} \lambda_l^{-1})(\sum_{l=k+1}^{n} \lambda_l^{-2})^{-1}$, $w_k = (\sum_{l=k+1}^{n} \lambda_l^{-2})^{-1}$ for $1 \le k \le n-1$, and $v_n = w_n = 0$. Let $\tilde{c}^{(k)} := \arg\max_c F_k(c)$, where

$$F_k(c) := \frac{\left(x + v_k + \sum_{i=1}^{k}(\lambda_i - v_k)u_i(c)\right)^2}{c^2r^2 + x^2 + 2\sum_{i=1}^{k}\lambda_i^2 u_i(c)^2 + 2w_k\left(1 - \sum_{i=1}^{k} u_i(c)\right)^2}$$

for $1 \le k \le n-1$, and let $\tilde{c}^{(n)}$ be the solution of the equation $\sum_{i=1}^{n} u_i(c) = 1$. Note that $\tilde{c}^{(k)}$ $(k = 1, \dots, n)$ are functions of (x, r). Define

$$\delta_k(c) := \frac{2x + 2\sum_{i=1}^{k}\lambda_i u_i(c) + 2v_k\left(1 - \sum_{i=1}^{k} u_i(c)\right)}{c^2r^2 + x^2 + 2\sum_{i=1}^{k}\lambda_i^2 u_i(c)^2 + 2w_k\left(1 - \sum_{i=1}^{k} u_i(c)\right)^2}$$

and $\tau_k(c) := w_k\delta_k(c)(1 - \sum_{i=1}^{k} u_i(c)) - v_k$ for $1 \le k \le n$. Now define functions $\tilde{q}_j^{(k)}(x, r)$ by

$$\tilde{q}_j^{(k)}(x, r) = \begin{cases} \delta_k(\tilde{c}^{(k)})u_j(\tilde{c}^{(k)})\lambda_j & \text{for } j \le k \\ 1 + \tau_k(\tilde{c}^{(k)})\lambda_j^{-1} & \text{for } j > k \end{cases}$$

for $1 \le j \le n$.

The following theorem gives an explicit expression of the minimax solution for $0 < x \le 2/(\sum_{l=1}^{n} \lambda_l^{-1})$.

THEOREM 2. *Let $r^{(k)}(x)$ be a unique solution of $\tilde{q}_1^{(k-1)}(x, r) = \tilde{q}_1^{(k)}(x, r)$ for $2 \le k \le n$, and put $r^{(1)}(x) = 0$, $r^{(n+1)}(x) = \infty$. Then under Assumptions* 1–2 *the* *minimax solution for* $r^{(k)}(x) < r \leq r^{(k+1)}(x)$ $(1 \leq k \leq n)$ *can be expressed as* $\tilde{q}_j(x,r) = \tilde{q}_j^{(k)}(x,r)$ *for* $1 \leq j \leq n$.

Expressions of the minimax estimators obtained in Theorem 2 have distinct forms on intervals $[r^{(k)}(x), r^{(k+1)}(x)]$ of r. Two expressions corresponding to the adjacent intervals $[r^{k-1}(x), r^{(k)}(x)]$ and $[r^{(k)}(x), r^{(k+1)}(x)]$ coincide at $r = r^{(k)}(x)$. Thus we have $\tilde{q}_j^{(k-1)}(x, r^{(k)}(x)) = \tilde{q}_j^{(k)}(x, r^{(k)}(x))$ for $j \neq 1$, although $r^{(k)}(x)$ is determined by $\tilde{q}_1^{(k-1)}(x, r^{(k)}(x)) = \tilde{q}_1^{(k)}(x, r^{(k)}(x))$.

Proof. Lemmas 1 and 2 complete the first two stages, respectively. We have only to minimize $H_1(x; q_1, \dots, q_n)$ with respect to $(q_1, \dots, q_n)$ under the constraints (16) for $1 \leq k \leq n$. First we consider the case that $1 \leq k \leq n-1$. Under the constraint (16), the quantity q_j for $1 \leq j \leq k$ is a function of c and $\delta = \sum_{l=1}^n q_l/\lambda_l$, and can be written as $q_j = q_j(c,\delta)$, where $q_j(c,\delta) = \delta u_j(c)\lambda_j$. Thus the function H_1 is a function of $(c, \delta, q_{k+1}, q_n)$, and we have

$$\begin{aligned} &H_1(x; q_1(c,\delta), \dots, q_k(c,\delta), q_{k+1}, \dots, q_n) \\ &\quad = \delta^2\left(c^2r^2 + x^2 + 2\sum_{l=1}^k u_l(c)^2\lambda_l^2\right) \\ &\qquad - 4\delta\left(x + \sum_{l=k+1}^n u_l(c)\lambda_l\right) + 2\sum_{l=k+1}^n (q_l^2 - 2q_l). \end{aligned}$$

By definitions of δ and $u_j(c)$ the quantities q_j for $k+1 \leq j \leq n$ satisfy a relation $\sum_{l=k+1}^n q_l/\lambda_l = \delta(1 - \sum_{l=1}^k u_j(c))$. The minimum of the last term in the above expression of $H_1(x; q_1, \dots, q_n)$ with respect to $q_{k+1}, \dots, q_n$ under the constraint $\sum_{j=k+1}^n q_j/\lambda_j = \delta(1 - \sum_{j=1}^k u_j(c))$ is attained at $q_j = q_j(c,\delta)$ with

$$q_j(c,\delta) = 1 + \left(w_k\delta\left(1 - \sum_{l=1}^k u_l(c)\right) - v_k\right)\lambda_j^{-1} \quad \text{for } k+1 \leq j \leq n,$$

and the minimum value is given by $2(w_k\delta(1 - \sum_{j=1}^k u_j(c)) - v_k)^2 w_k^{-1} - 2(n-k)$. Then, we have

$$\begin{aligned} &H_1(x; q_1(c,\delta), \dots, q_k(c,\delta), q_{k+1}(c,\delta), \dots, q_n(c,\delta)) \\ &\quad = \delta^2\left\{c^2r^2 + x^2 + 2\sum_{j=1}^k \lambda_j^2 u_j(c)^2 + 2w_k\left(1 - \sum_{j=1}^k u_j(c)\right)^2\right\} \\ &\qquad - 4\delta\left\{x + \sum_{j=1}^k \lambda_j u_j(c) + v_k\left(1 - \sum_{j=1}^k u_j(c)\right)\right\} + 2v_k^2 w_k^{-1} - 2(n-k). \end{aligned}$$

Next, the function $H_1(x; q_1(c,\delta), \dots, q_k(c,\delta), q_{k+1}(c,\delta), \dots, q_n(c,\delta))$ is minimized with respect to δ when $\delta = \delta_k(c)$, and the minimum is given by

$$\begin{aligned}&\overline{H}(q_1(c,\delta_k(c)), \dots, q_k(c,\delta_k(c)), q_{k+1}(c,\delta_k(c)), \dots, q_n(c,\delta_k(c)))\\ &= \frac{-4\left\{x + \sum_{j=1}^k \lambda_j u_j(c) + v_k\left(1 - \sum_{j=1}^k u_j(c)\right)\right\}^2}{c^2r^2 + x^2 + 2\sum_{j=1}^k \lambda_j^2 u_j(c)^2 + 2w_k\left(1 - \sum_{j=1}^k u_j(c)\right)^2}\\ &\quad + 2v_k^2 w_k^{-1} - 2(n-k). \end{aligned} \tag{17}$$

The above expression can be written as $-4F_k(c) + 2v_k^2 w_k^{-1} - 2(n-k)$. Therefore this minimization problem reduces to the maximization of $F_k(c)$ with respect to c. Next, we consider the case that $k = n$. From the constraint (16) for $k = n$, the q_j's can be expressed as $q_j = \delta u_j(\tilde{c}^{(n)})\lambda_j$ for $1 \le j \le n$, where $\tilde{c}^{(n)}$ is the solution of the equation $\sum_{j=1}^n u_j(c) = 1$. Then, $H_1(x; q_1, \dots, q_n)$ is a function of δ, and is expressed as

$$\begin{aligned}H_1(x; q_1(\delta), \dots, q_n(\delta)) = \delta^2&\left(\tilde{c}^{(n)2}r^2 + x^2 + 2\sum_{j=1}^n \lambda_j^2 u_j(\tilde{c}^{(n)})^2\right)\\ &- 4\delta\left(x + \sum_{j=1}^n \lambda_j u_j(\tilde{c}^{(n)})\right).\end{aligned}$$

The function $H_1(q_1(\delta), \dots, q_n(\delta))$ is minimized when $\delta = \delta_n(c^{(n)})$, and the minimum value is given by $-4\left\{x + \sum_{j=1}^n \lambda_j u_j(c^{(n)})\right\}^2 \{c^{(n)2}r^2 + x^2 + 2\sum_{j=1}^n \lambda_j^2 u_j(c^{(n)})^2\}^{-1}$. Finally replacing c in $q_j(c, \delta_k(c))$ by $\tilde{c}^{(k)}$ for corresponding intervals of r, $[r^{(k)}, r^{(k+1)}]$ for $1 \le k \le n$, we obtain $(q_1, \dots, q_n)$ which provides the minimax estimator of σ^2. This completes the proof.

REMARK 2. Assumption 1 is a sufficient condition for (7). Let

$$H_0(x; q_1, \dots, q_n) = x^2\left(\sum_{l=1}^n \frac{q_l}{\lambda_l}\right)^2 - 4x\sum_{l=1}^n \frac{q_l}{\lambda_l} + 2\sum_{l=1}^n (q_l^2 - 2q_l).$$

Then, it follows that $\overline{H}(x; q_1, \dots, q_n) - H_0(x; q_1, \dots, q_n) = \max_{1 \le i \le n}(rq_i/\lambda_i)$ $(rq_i/\lambda_i - 2x\sum q_l/\lambda_l)$ is a decreasing function of x. We have only to show that for $x = 2/\left(\sum \lambda_l^{-1}\right)$

$$\begin{aligned}&\max_{1 \le i \le n} \mathrm{RL}(\hat{\sigma}^2(x; \tilde{q}_1(x,r), \dots, \tilde{q}_n(x,r)); \sigma^2, \boldsymbol{\mu}^{(i)})\\ &\qquad > \mathrm{RL}(\hat{\sigma}^2(x; \tilde{q}_1(x,r), \dots, \tilde{q}_n(x,r)); \sigma^2, \boldsymbol{\mu}^{(0)}).\end{aligned} \tag{18}$$

The above inequality will be demonstrated in §5.4.

To illustrate the form of minimax solutions given in Theorem 2, we present

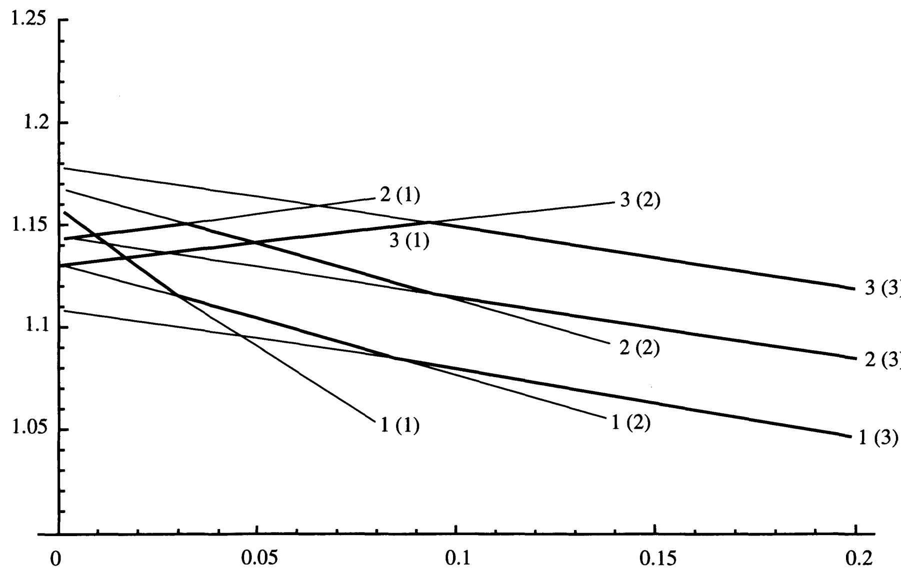

Fig. 1. The components of the minimax solution on the r-interval $(0, 0.2]$ for $x = 1/(\sum \lambda_l^{-1}) = 0.364$ in the case $n = 3$. We set $(\lambda_1, \lambda_2, \lambda_3) = (1, 1.1, 1.2)$. The values of $r^{(i)}(x)$ are given by $r^{(2)}(x) = 0.0307$, $r^{(3)} = 0.0924$. The graphs of $q_j^{(i)}(x, r)$ are described with labels "$j(i)$" at the right, respectively. Three thick curves yield components of the resultant minimax solution.

simple figures of $\tilde{q}_j(x, r)$ as functions of r in the case $n = 3$. We set $(\lambda_1, \lambda_2, \lambda_3) = (1, 1.1, 1.2)$. Three values of x are chosen as one of representative value $x = 1/(\sum \lambda_l^{-1})$, and two boundary values of Assumption 1. Note that the cases of the latter two value of x will be described in the next section.

The minimax risk of our minimax estimator is rather simple as we see in the following Proposition 2, despite of complicated expression for minimax solutions in Theorem 2. This simple form of the minimax risk was also obtained by Fujioka [9], and its properties were discussed by Fujioka [10].

Proposition 2. *The minimax risk of* $\hat{\sigma}^2(x; \tilde{q}_1(x, r), \ldots, \tilde{q}_n(x, r))$ *is given by*

$$\overline{M}(x; \tilde{q}_1(x, r), \ldots, \tilde{q}_n(x, r)) = \frac{2}{\sum_{j=1}^n \tilde{q}_j(x, r)(1 + x/\lambda_j)}.$$

Proof. It can be shown by Theorem 2 that if $r^{(k)}(x) < r \le r^{(k+1)}(x)$

$$\sum_{j=1}^{n} \tilde{q}_j^{(k)}(x, r) = \delta_k(\tilde{c}^{(k)}) \sum_{j=1}^{k} u_j(\tilde{c}^{(k)})\lambda_j + (n - k)$$
$$+ \left(w_k \delta_k(\tilde{c}^{(k)}) \left(1 - \sum_{l=1}^{k} u_l(\tilde{c}^{(k)}) \right) - v_k \right) v_k w_k^{-1}$$

for $1 \le k \le n$. Recalling that $\sum_{j=1}^{n} \tilde{q}_j^{(k)}(x,r)/\lambda_j = \delta_k(\tilde{c}^{(k)})$, we have

$$\sum_{j=1}^{n} \tilde{q}_j^{(k)}(x,r)(1+x/\lambda_j) = \delta_k(\tilde{c}^{(k)})\left(x + v_k + \sum_{j=1}^{k}(\lambda_j - v_k)u_j(c)\right) + 2v_k^2 w_k^{-1} - 2(n-k).$$

On the other hand, from (17), it can be seen that

$$\overline{H}(x; \tilde{q}_1^{(k)}(x,r), \dots, \tilde{q}_n^{(k)}(x,r)) = -2\delta_k(\tilde{c}^{(k)})\left(x + v_k + \sum_{j=1}^{k}(\lambda_j - v_k)u_j(c)\right) + 2v_k^2 w_k^{-1} - 2(n-k).$$

Hence, we obtain

$$\overline{H}(x; \tilde{q}_1^{(k)}(x,r), \dots, \tilde{q}_n^{(k)})(x,r)) = -2\sum_{j=1}^{n} \tilde{q}_j^{(k)}(x,r)(1+x/\lambda_j). \tag{19}$$

A relation

$$\overline{M}(x; \tilde{q}_1^{(k)}(x,r), \dots, \tilde{q}_n^{(k)}(x,r)) = \frac{\overline{H}(x; \tilde{q}_1^{(k)}(x,r), \dots, \tilde{q}_n^{(k)}(x,r)) + 4\sum_{j=1}^{n} \tilde{q}_j^{(k)}(x,r)(1+x/\lambda_j)}{\left\{\sum_{j=1}^{n} \tilde{q}_j^{(k)}(x,r)(1+x/\lambda_j)\right\}^2}$$

together with (19) proves the assertion.

We will fully examine five special cases in the last section, and will give illustrative figures of components of minimax solutions.

5. Special cases

In this section we consider five special cases; (a) $r \le r^{(2)}(x)$, (b) $r \to \infty$, (c) $x = 0$, (d) $x = 2/(\sum \lambda_l^{-1})$, (e) $\lambda_1 = \cdots = \lambda_n$. Among these five cases, the case (a) is very important from a view-point of practical applications. The minimax estimator treated in the present paper was originally introduced for the spline smoothing method in nonparametric regression. Thus the mean function is required to change slowly. This requirement indicates implicitly the assumption that r is small. In the case (b) we will concentrate on sufficiently large r. Then the mean vector is no longer restricted. This situation is somewhat theoretically interesting, even though it may be unrealistic. Note that both cases (c) and (d) are the boundary cases allowed under Assumption

1. In the case (e) the mean is restricted in a sphere. Minimax solutions can be obtained for any $x > 0$.

5.1. The case (a): $r \le r^{(2)}(x)$

The quantity $q_j^{(1)}(x,r)$ given in Lemma 1 itself provides us with the minimax solution, and its explicit form is given by

$$\tilde{q}_j^{(1)}(x,r) = \begin{cases} \dfrac{\lambda_1 + r}{\lambda_1 + 2r} + \dfrac{x - r}{\lambda_1 + 2r}\xi^{(1)}(x,r) & \text{for } j = 1 \\ 1 + \dfrac{x}{\lambda_j}\xi^{(1)}(x,r) & \text{for } j \ne 1 \end{cases}$$

where

$$\xi^{(1)}(x,r) = \frac{(r^2 + 5\lambda_1 r + 2\lambda_1^2)w_1 - ((w_1 + 2\lambda_1 v_1)r + \lambda_1 w_1 + \lambda_1^2 v_1)x}{(2\lambda_1 r + \lambda_1^2 + w_1)x^2 - 2w_1 rx + (r^2 + 4\lambda_1 r + 2\lambda_1^2)w_1}.$$

An interesting special case is $r = +0$. Taking the limit of the above expression, we have

$$\tilde{q}_j^{(1)}(x,+0) = 1 + \frac{x}{\lambda_j}\frac{2 - x\sum\lambda_l^{-1}}{2 + x^2\sum\lambda_l^{-2}} \quad \text{for } 1 \le j \le n.$$

Thus we obtain a variety of minimax estimators $\hat{\sigma}^2(x; \tilde{q}_1(x,+0), \dots, \tilde{q}_n(x,+0))$ for $0 \le x \le 2/(\sum\lambda_l^{-1})$. The minimax risk of the minimax estimator is given by

$$\overline{M}(x; \tilde{q}_1(x,+0), \dots, \tilde{q}_n(x,+0)) = \frac{2(2 + x^2\sum\lambda_l^{-2})}{(n+2)(2 + x^2\sum\lambda_l^{-2}) - (2 - x\sum\lambda_j^{-1})^2},$$

and is minimized over $0 \le x \le 2/(\sum\lambda_l^{-1})$ when $x = 2/(\sum\lambda_l^{-1})$. The resultant estimator of σ^2 is $\hat{\sigma}^2 = \sum y_i^2/(n+2)$, and its minimax risk is given by $2/(n+2)$.

The case $r = +0$ implies that the mean $\boldsymbol{\mu}$ vanishes. In other words, the sample $y_1, \dots, y_n$ is assumed to be independent and identically distributed as $N(0, \sigma^2)$. This case is very similar to the standard theory of statistics. Our result shows that the estimator $\hat{\sigma}^2 = \sum y_i^2/(n+2)$ is the best minimax estimator. The counterpart of the standard theory of statistics is that this estimator is the minimum mean squared error estimator of σ^2. Thus we can expect that the restriction (2) on the form of estimators may not exclude good estimators in a wider class.

5.2. The case (b): $r \to \infty$

When $r \geq r^{(n)}(x)$, the minimax solution is expressed as

$$\tilde{q}_j^{(n)}(x,r) = \delta_n(\tilde{c}^{(n)}) u_j(\tilde{c}^{(n)}) \lambda_j \quad \text{for } 1 \leq j \leq n$$

where $\tilde{c}^{(n)}$ is the solution of $\sum u_j(c) = 1$. For sufficiently large r, the functions $u_j(c)$'s can be approximated by c commonly in j. Thus we have $\tilde{c}^{(n)} = 1/n$ approximately. The common factor in the form of $\tilde{q}_j^{(n)}(x,r)$ can be neglected. When r tends to infinity, the limit of the minimax estimator is given by

$$\hat{\sigma}^2 = \frac{\sum \lambda_j y_j^2}{\sum \lambda_j + nx}$$

for $0 \leq x \leq 2/(\sum \lambda_l^{-1})$. The minimax risk is obviously infinite. Actually, we have

$$\lim_{r\to\infty} \overline{M}(x; \tilde{q}_1^{(n)}(x,r), \ldots, \tilde{q}_n^{(n)}(x,r))/r^2 = \frac{2}{(\sum \lambda_j + nx)^2}.$$

5.3. The case (c): $x = 0$

We substitute $x = 0$ formally for the minimax solution in Theorem 2. Note that $u_j(c) = c(1 + 4\lambda_j/r)^{-1/2} = c\lambda_j(r)/\lambda_j$. We can evaluate $\tilde{c}^{(k)} = \arg\max_c F_k(c)$ explicitly as

$$\tilde{c}^{(k)} = \frac{2w_k \sum_{l=1}^k \lambda_l(r)}{v_k\left(r^2 + 2\sum_{l=1}^k \lambda_l(r)^2\right) + 2w_k\left(\sum_{l=1}^k \lambda_l(r)/\lambda_l\right)\left(\sum_{l=1}^k \lambda_l(r)\right)}.$$

Also, we obtain

$$\delta(\tilde{c}^{(k)}) = \frac{v_k\left(r^2 + 2\sum_{l=1}^k \lambda_l(r)^2\right) + 2w_k\left(\sum_{l=1}^k \lambda_l(r)/\lambda_l\right)\left(\sum_{l=1}^k \lambda_l(r)\right)}{2w_k\left(r^2 + 2\sum_{l=1}^k \lambda_l(r)^2\right)}$$

and $\tau_k(\tilde{c}^{(k)}) = 0$. These calculations yields an explicit expression of the minimax solution for $x = 0$. The resultant expression of $\tilde{q}_j(0,r)$ coincides with that given by Fujioka [9] which is summarized in Theorem 1 in the present paper. A graph of each component of the minimax solution $\tilde{q}_j(0,r)$ is given in Figure 2. The case $n = 3$ is one of the simplest, but it is sufficient to comprehends fundamental features of minimax solutions for a general n.

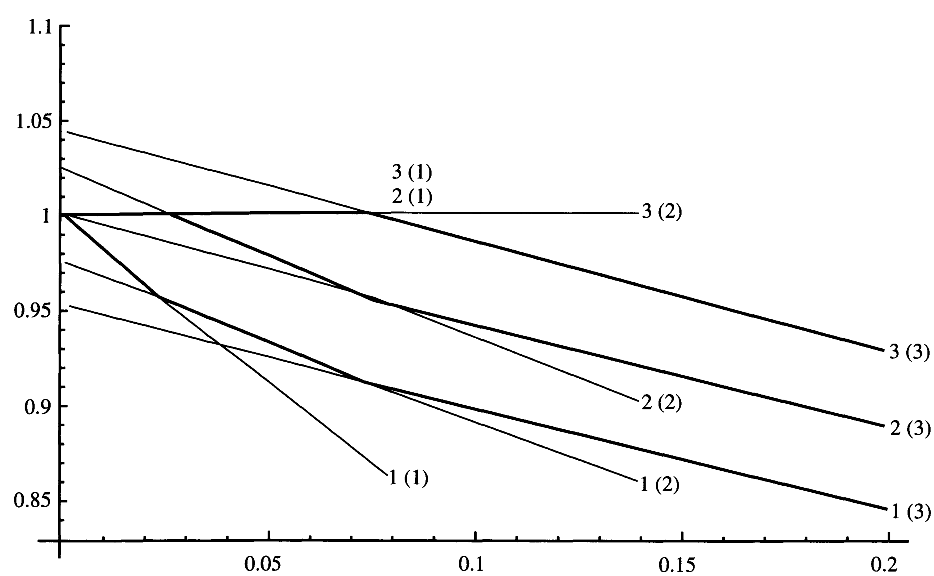

Fig. 2. The components of the minimax solution for $x = 0$. The values of $r^{(i)}(x)$ are given by $r^{(2)}(x) = 0.0248$, $r^{(3)}(x) = 0.0734$.

By Proposition 2, the minimax risk is given by

$$\overline{M}(0; \tilde{q}_1^{(k)}(0, r), \ldots, \tilde{q}_n^{(k)}(0, r)) = \frac{2r^2 + 4\sum_{l=1}^{k} \lambda_l(r)^2}{2\left(\sum \lambda_j(r)\right)^2 + (n-k)\left(r^2 + 2\sum_{l=1}^{k} \lambda_l(r)^2\right)}.$$

Fundamental properties of the minimax risk for $x = 0$ were discussed by Fujioka [10].

5.4. The case (d): $x = 2/(\sum \lambda_l^{-1})$

First we will show that $r^{(1)}(x) = \cdots = r^{(n)}(x) = 0$. From the argument in the case (a), we have $q_j^{(1)}(x, +0) = 1$ for $1 \leq j \leq n$. By the definition of $r^{(n)}(x)$ it suffices to prove that $q_j^{(n)}(x, +0) = 1$ for $1 \leq j \leq n$. The quantity is determined by $\sum u_j(\tilde{c}^{(n)}) = 1$. Approaching r to zero, we have

$$\sum_{j=1}^{n} \frac{x + \sqrt{x^2 + \lim_{r \to +0} r(r + 4\lambda_j)\tilde{c}^{(n)2}}}{4\lambda_j} = 1.$$

In the case $x = 2/(\sum \lambda_l^{-1})$, it holds that $\lim_{r \to +0} r(r + 4\lambda_j)\tilde{c}^{(n)2} = 0$ for $1 \leq j \leq n$. Thus it is verified that $\lim_{r \to +0} u_j(\tilde{c}^{(n)}) = 1/(\lambda_j \sum \lambda_l^{-1})$, and also that $\lim_{r \to +0} \delta_n(\tilde{c}^{(n)}) = \sum \lambda_l^{-1}$. Therefore, we obtain $\tilde{q}_j^{(n)}(x, +0) = 1$ for $1 \leq j \leq n$,

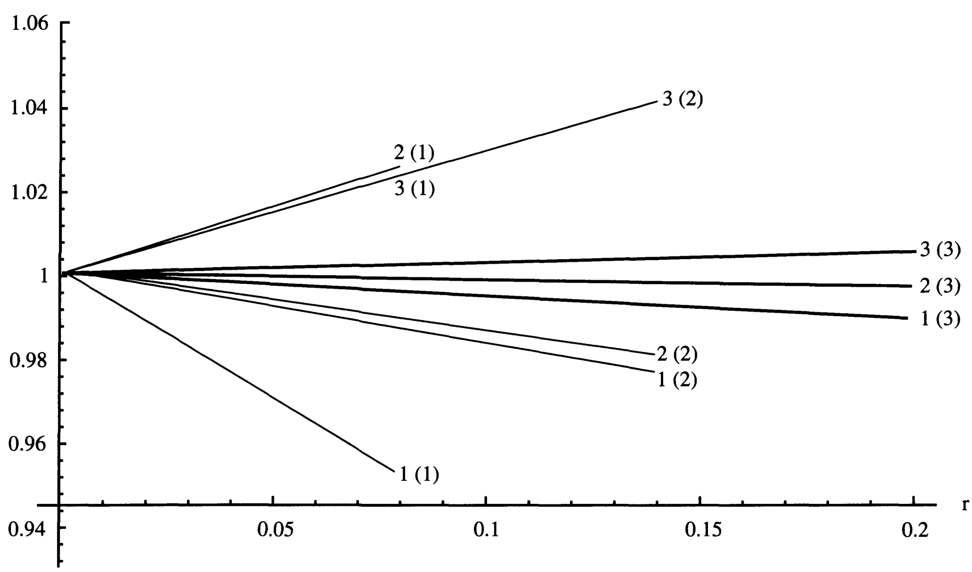

Fig. 3. The components of the minimax solution for $x = 2/(\sum \lambda_l^{-1}) = 0.729$. The values of $r^{(i)}(x)$ are given by $r^{(2)}(x) = r^{(3)}(x) = 0$.

which implies that $r^{(n)}(x) = 0$. By Theorem 2, the minimax solution is given by

$$\tilde{q}_j^{(n)}(x,r) = \delta_n(\tilde{c}^{(n)})u_j(\tilde{c}^{(n)})\lambda_j \quad \text{for } 1 \le j \le n$$

for any $r > 0$. The resultant estimator is given by

$$\hat{\sigma}^2 = \frac{\sum u_j(\tilde{c}^{(n)})\lambda_j y_j^2}{\sum u_j(\tilde{c}^{(n)})\lambda_j + x}.$$

Thus the minimax solution has a single piece of form. Figure 3 is depicted on the same setup as that of Figures 1 and 2 for their comparisons.

By Proposition 2, its minimax risk is given by

$$\overline{M}(x; \tilde{q}_1^{(n)}(x,r), \dots, \tilde{q}_n^{(n)}(x,r)) = \frac{c^{(n)2}r^2 + x^2 + \sum u_j(\tilde{c}^{(n)})^2\lambda_j^2}{\left(\sum u_j(\tilde{c}^{(n)})\lambda_j + x\right)^2}.$$

On the other hand, the risk at $\boldsymbol{\mu} = \boldsymbol{\mu}^{(0)}$ is given by

$$\mathrm{RL}(\hat{\sigma}^2(x; \tilde{q}_1^{(n)}(x,r), \dots, \tilde{q}_n^{(n)}(x,r)); \sigma^2, \boldsymbol{\mu}^{(0)}) = \frac{x^2 + \sum u_j(\tilde{c}^{(n)})^2\lambda_j^2}{\left(\sum u_j(\tilde{c}^{(n)})\lambda_j + x\right)^2}.$$

By Remark 2, the inequality (18) is valid for $0 \le x \le 2/\left(\sum \lambda_l^{-1}\right)$.

5.5. The case (e): $\lambda_1 = \cdots = \lambda_n = 1$

Let λ_i approach to 1 for $1 \le i \le n$ simultaneously. The mean will be restricted within the sphere $\sum \mu_i^2 \le r\sigma^2$. Assumption 1 implies that $0 < x \le 2/n$. Note that $u_1(c) = \cdots = u_n(c)$. By the definition of $\tilde{c}^{(n)}$, we have $u_j(\tilde{c}^{(n)}) = 1/n$ for $1 \le j \le n$. It is straightforward that $\delta_n(\tilde{c}^{(n)}) = 2n^2(x+1)/\{n^2x^2 - 2nrx + r^2 + 4r + 2n\}$. Hence, we have

$$\tilde{q}_j^{(n)}(x, r) = \frac{2n(x+1)}{n^2x^2 - 2nrx + r^2 + 4r + 2n} \quad \text{for } 1 \le j \le n.$$

On the other hand, from the case (a) we have

$$\tilde{q}_j^{(1)}(x, r) = \begin{cases} \dfrac{1+r}{1+2r} + \dfrac{x-r}{1+2r}\xi^{(1)}(x, r) & \text{for } j = 1 \\ 1 + x\xi^{(1)}(x, r) & \text{for } j \ne 1 \end{cases}$$

where

$$\xi^{(1)}(x, r) = \frac{(r^2 + 5r + 2)w_1 - ((w_1 + 2v_1)r + w_1 + v_1)x}{(2r + 1 + w_1)x^2 - 2w_1rx + (r^2 + 4r + 2)w_1}.$$

Taking the limit in the above minimax solutions, we have $\tilde{q}_j^{(1)}(x, +0) = \tilde{q}_j^{(n)}(x, +0) = 2(x+1)/(nx^2+2)$ for $1 \le j \le n$. By Theorem 2 the above equality implies that $r^{(1)} = \cdots = r^{(n)} = 0$, and the minimax solution is given by $\tilde{q}_j(x, r) = \tilde{q}_j^{(n)}(x, r) = \delta(\tilde{c}^{(n)})u_j(\tilde{c}^{(n)})$ for any $r > 0$. The resultant estimator is $\hat{\sigma}^2 = \sum y_i^2/n(1+x)$, and its minimax risk is given by

$$\lim_{1 \le i \le n} \mathrm{RL}(\hat{\sigma}^2; \sigma^2, \boldsymbol{\mu}^{(i)}) = \frac{(r - nx)^2 + 4r + 2n}{n^2(1+x)^2}$$

for $0 \le x \le 2/n$. By Remark 2, the minimax estimator for $2/n < x$ may not be obtained based on (7). Fortunately, the minimax estimator $\hat{\sigma}^2 = \sum y_i^2/n(1+x)$ achieves simultaneously the minimum of the risk at $\boldsymbol{\mu} = \boldsymbol{\mu}^{(0)}$. Note that its risk at $\boldsymbol{\mu} = \boldsymbol{\mu}^{(0)}$ is given by $(nx^2+2)/\{n(1+x)^2\}$. Therefore, the minimax risk is given by

$$\max_{0 \le i \le n} \mathrm{RL}(\hat{\sigma}^2; \sigma^2, \boldsymbol{\mu}^{(i)}) = \max\left\{\frac{(r - nx)^2 + 4r + 2n}{n^2(1+x)^2}, \frac{nx^2 + 2}{n(1+x)^2}\right\}$$

for $x \ge 0$. The best minimax estimator over $x \ge 0$ is obtained when $x = (r+4)/2n$, and the resultant estimator is $\hat{\sigma}^2 = \sum y_i^2/(n + 2 + r/2)$.

Finally we discuss requirements imposed on λ_i's in the present study. The requirement that λ_i's are all distinct is assumed for mathematical convenience, similar to their ordering condition, $0 < \lambda_1 < \cdots < \lambda_n$. As is seen in §5.5, by taking the limit, we can assure that Theorem 2 is valid for general case that some of λ_i's are the same. On the other hand, Assumption 2 can be replaced in a different way by modifying of the present proof of Lemma 2, or by finding alternative proofs.

Acknowledgements

The author expresses his sincere thanks to Professor Y. Fujikoshi of Hiroshima University for his valuable comments and constant encouragement. The author is also grateful to Professor T. Yanagimoto of Institute of Statistical Mathematics for his helpful comments and suggestions. The author wishes to thank the referee that he suggests extensive perspectives of future work.

References

[1] P. J. Bickel, Minimax estimation of the mean of a normal distribution when the parameter space is restricted, *Ann. Statist.* **6** (1981), 1301–1309.

[2] M. J. Buckley, G. K. Eagleson and W. Silverman, The estimation of residual variance in nonparametric regression, *Biometrika* **75** (1988), 189–199.

[3] C. K. Carter, G. K. Eagleson, A comparison of variance estimators in nonparametric regression, *J. Roy. Statist. Soc. B* **54** (1992), 773–780.

[4] G. Casella and W. E. Strawderman, Estimating a bounded mean, *Ann. Statist.* **9** (1981), 970–879.

[5] P. Craven and G. Wahba, Smoothing noisy data with splines, *Numer. Math.* **31** (1979), 377–403.

[6] A. Demmler and C. Reinsch, Oscillation matrices with spline smoothing, *Numer. Math.* **24** (1975), 377–382.

[7] D. L. Donoho, R. C. Liu and B. MacGibbon, Minimax risk over hyperrectangles, and implications, *Ann. Statist.* **18** (1990), 1416–1437.

[8] B. Efron and C. Morris, Empirical Bayes on vector observations: An extension of Stein method, *Biometrika* **59** (1972), 335–347.

[9] T. Fujioka, Notes on minimax approaches in nonparametric regression, *Hiroshima Math. J.* **24** (1994), 55–62.

[10] T. Fujioka, The max-MSE's of minimax estimators of variance in nonparametric regression, *Hiroshima Math. J.* **24** (1994), 573–581.

[11] T. Gasser, L. Sroka and C. Jenneu-Steimmetz, Residual variance and residual pattern in non-linear regression, *Biometrika* **73** (1986), 623–633.

[12] P. Hall, J. W. Kay and D. M. Titterington, Asymptotically optimal difference-based estimation of variance in nonparametric regression, *Biometrika* **77** (1990), 521–528.

[13] P. Hall and D. M. Titterington, Common structure of techniques for choosing smoothing parameters in regression problems, *J. Roy. Statist. Soc. B* **49** (1987), 184–198.

[14] P. Hall and J. S. Marron, On variance estimation in nonparametric regression, *Biometrika* **77** (1990), 415–419.

[15] D. V. Lindley and A. F. M. Smith, Bayes estimation for the linear model (with discussion), *J. Roy. Statist. Soc. B* **30** (1971), 1–41.

[16] M. Ohtaki, A nonparametric estimate of residual variance in one dimensional regression, *Bull. Biometric Soc. Jpn.* **8** (1987), 39–51.

[17] M. Ohtaki, Some estimators of covariance matrix in multivariate nonparametric regression and its applications, *Hiroshima Math. J.* **20** (1990), 63–91.

[18] C. Reinsch, Smoothing by spline functions, *Numer. Math.* **10** (1967), 177–183.

[19] J. Rice, Bandwidth choice for nonparametric kernel regression, *Ann. Statist.* **12** (1984), 1215–1230.

[20] B. W. Silverman, Some aspects of the spline smoothing approach to nonparametric regression curve fitting (with discussion), *J. Roy. Statist. Soc. B* **47** (1985), 1–52.

[21] P. Speckman, Spline smoothing and optimal rates of convergence in nonparametric regression models, *Ann. Statist.* **13** (1985), 970–983.

[22] W. E. Strawderman, Minimax estimation of powers of the variance of a normal population under square error loss, *Ann. Statist.* **2** (1974), 190–198.

[23] G. Wahba, Improper prior, spline smoothing and the problem of guarding against model errors in regression, *J. Roy. Statist. Soc. B* **40** (1978), 364–372.

Department of Mathematics
Faculty of Science
Hiroshima University
Higashi-Hiroshima 739-8526, Japan