SCIENCE CHINA
Information Sciences

• **Supplementary File** •

A new optimization scheme for uncertain problems – globally robust solution

Yunmin ZHU[†], Weiyu LI[†], Enbin SONG[*] & Yingting LUO

College of Mathematics, Sichuan University, Chengdu 610064*, China*

Appendix A Formulation of the New Scheme via an Uncertain Hypothesis Testing

To visually demonstrate several drawbacks of the traditional minimax solution and how to formulate the GRS for an uncertain problem, we have the aid of the well-known result of the Bayesian binary hypothesis testing with unknown a priori hypothesis probability [1].

Appendix A.1 Binary Hypothesis Problem With Unknown a Priori Probability

Let us consider the Bayesian binary decision problem with two hypotheses denoted by $\mathbf{H}_0$ and $\mathbf{H}_1$. The prior probabilities P_0 and P_1 of the two hypotheses are unknown and satisfy $P_0 + P_1 = 1$. Let $\mathcal{Y}$ be the entire observation space, and $\mathcal{H}_0/\mathcal{H}_1$ be the decision region. The observation be denoted by $\mathbf{y} \in \mathcal{Y}$, and the conditional densities under the two hypotheses be $p(\mathbf{y}|\mathbf{H}_i)$, $i = 0,\ 1$. For simplicity, there is a cost associated with each decision. Let c_{ij}, $i, j = 0,\ 1$, represent the cost of declaring $\mathbf{H}_i$ true when $\mathbf{H}_j$ is actually true. Hence, for the deterministic decision rule, $\mathcal{H}_0 \cup \mathcal{H}_1 = \mathcal{Y}$ and $\mathcal{H}_0 \cap \mathcal{H}_1 = \emptyset$. In particular, the optimization problem is to minimize the expectation of the decision cost, the following cost functional:

$$\begin{aligned} C(d, P_1) &= \sum_{i=0}^{1}\sum_{j=0}^{1} c_{ij} P(d = \mathbf{H}_i, H = \mathbf{H}_j) \\ &= \sum_{i=0}^{1}\sum_{j=0}^{1} c_{ij} P_j P(d = \mathbf{H}_i | H = \mathbf{H}_j) \\ &= \sum_{i=0}^{1}\sum_{j=0}^{1} c_{ij} P_j \int_{\mathcal{H}_i} p(\mathbf{y}|\mathbf{H}_j) d\mathbf{y}, \end{aligned} \tag{A1}$$

where d is the decision variable. According to likelihood ratio test (LRT) [1], we can get decision region $\mathcal{H}_0$ (to substitute d) corresponding to P_1 and the cost functional $C(\mathcal{H}_0, P_1)$ corresponding to $\mathcal{H}_0$ and P_1. More specifically,

$$\mathcal{H}_0(P_1) := \Big\{\mathbf{y}|\mathbf{p}(\mathbf{y}|\mathbf{H_1}) \cdot \mathbf{P_1}(\mathbf{c_{01}} - \mathbf{c_{11}}) \leqslant \mathbf{p}(\mathbf{y}|\mathbf{H_0}) \cdot (\mathbf{1} - \mathbf{P_1}) \cdot (\mathbf{c_{10}} - \mathbf{c_{00}}), \mathbf{P_1} \in [\mathbf{0}, \mathbf{1}]\Big\} \tag{A2}$$

and

$$\begin{aligned} C(\mathcal{H}_0(P_1), P_1) =&(1 - P_1)c_{10} + P_1 c_{11} + \int_{\mathcal{H}_0} [P_1(c_{01} - c_{11})p(\mathbf{y}|\mathbf{H}_1) \\ &- (1 - P_1)(c_{10} - c_{00})p(\mathbf{y}|\mathbf{H}_0)] d\mathbf{y}, P_1 \in [0, 1]. \end{aligned} \tag{A3}$$

For a parameter P_1^* obtained based on a certain criterion and the true parameter P_1, the cost can be simply expressed as:

$$C_{P_1^*}(P_1) := C(\mathcal{H}_0(P_1^*), P_1). \tag{A4}$$

In particular, when $P_1^* = P_1$, (A4) can be simply as

$$C(P_1) := C_{P_1}(P_1) = C(\mathcal{H}_0(P_1), P_1). \tag{A5}$$

When P_1 is fixed as P_1^*, i.e., the decision regions $\mathcal{H}_0$ and $\mathcal{H}_1$ can be determined by LRT, the right hand side of (A3) represents the cost curve $C_{P_1^*}(P_1)$ with P_1 on [0, 1] which is a tangent line of the above minimum cost curve $C(P_1)$ with $C_{P_1^*}(P_1^*) = C(P_1^*)$ as shown in the following Fig. A1. For details, please refer to [1].

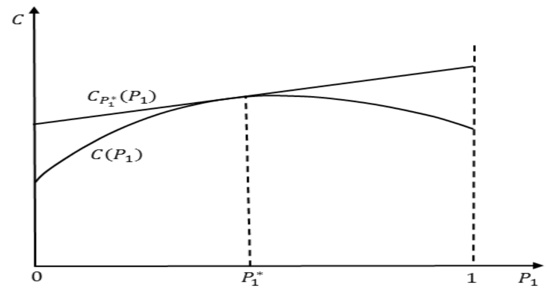

Figure A1 The minimum cost curve $C(P_1)$ and the cost tangent line $P_{P_1^*}(P_1)$.

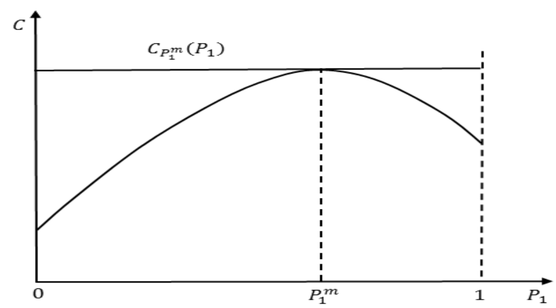

Figure A2 The minimum cost curve $C(P_1)$ and the mnimax cost tangent line $C_{P_1^m}(P_1)$.

Appendix A.2 The Drawbacks of the Traditional Minimax Decision

For simplicity of the above uncertain problem, assume $C(P_1)$ is a smooth curve, and denote P_1^m as the least favorable a priori probability, which makes $C(P_1^m)$ maximum. Hence, the minimax solution should be a horizonal tangent line of $C(P_1)$ at $P_1 = P_1^m$. Denote this tangent line by

$$C_{P_1^m}(P_1),\ P_1 \in [0, 1], \tag{A6}$$

with $C_{P_1^m}(P_1^m) = C(P_1^m)$ and $P_1^m = \arg\max C(P_1)$, as shown in Fig. A2.

Obviously, by definition of the minimax here $C_{P_1^m}(P_1)$ gives a guaranteed minimum cost upper bound no matter which P_1 is true. This is useful if the cost of passing the above upper bound will be unaffordable in practical applications. However, in a lot of uncertain problems without this requirement, for such ones, still using the traditional minimax will bring the following drawbacks:

Drawback 1. Conservativeness/Overcompression of Uncertainty

This drawback is obvious and has been criticized by many researchers (for details, refer to [3] [4] [5] and [6]).

Drawback 2. Wasting Problem Prior Knowledge

As we show in Figs. A3 and A4, once the least favorable problem assumption has been already known, the traditional minimax does not care about other possible problem uncertainty anymore. Hence, a large number of different uncertain problems may have the same solution. In particular, a lot of the uncertain problems in Fig. A4 have the same trivial solution $P_1^m = 0$ ignoring their different problem prior knowledge. **Drawback 3. Less Applicability** Since an actual assumption in all uncertain problem

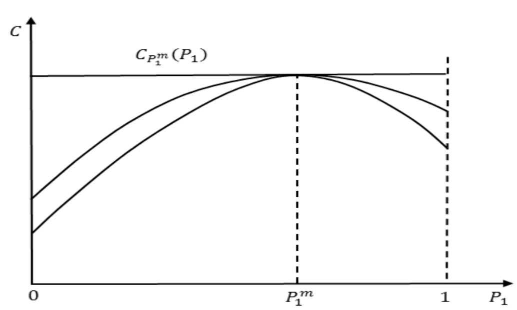

Figure A3 The two different minimum cost curves $C(P_1)$ with the same tangent line $C_{P_1^m}(P_1)$.

* Corresponding author (email: e.b.song@163.com)

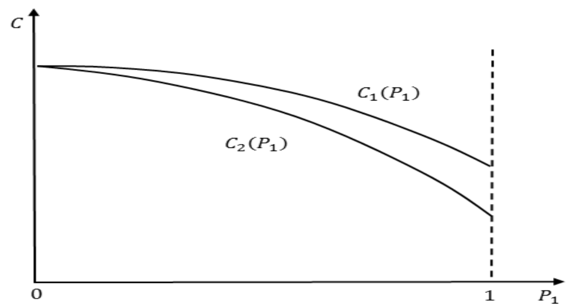

Figure A4 The two different minimum cost curves $C(P_1)$ with the same trivial solution $P_1^m = 0$.

assumptions may happen in any unknown rule, the performance of a traditional minimax solution depends strongly on how often the actual uncertain assumptions are close to the least favorable case. In Observation 3 and Table B3 of Section Appendix B, several such cases will be provided. Therefore, except it has been known that the true P_1s are close to the least favorable P_1 very often, the applicability of the traditional minimax solution is likely to be poor. Besides, there are a lot of uncertain problems that have trivial solutions, i.e., one can easily know the least favorable case in the uncertainty to be on the boundary of the uncertainty, such as the cases in Fig. A4 and the largest noise variance cases in the signal detection example with uncertain noise variances of Section Appendix B. In the previous robust hypothesis testing researches [7] and [8], when the two families of uncertain conditional densities $p(\mathbf{y}|\mathbf{H}_0)$ and $p(\mathbf{y}|\mathbf{H}_1)$ have a non-empty intersection, i.e., the least favorable case is that the two conditional densities are the same one, the traditional minimax robust testing is trivial.

Appendix A.3 Formulation of the GRS Scheme

According to the new scheme, the solution set derived by the GRS will optimally approximate the set of all minimum cost solutions produced by every actual assumption of the uncertain problem in terms of a criterion. Thus its formulation should define the cost tangent line $C_{P_1^G}(P_1)$ as follows:

$$C_{P_1^G}(P_1), P_1 \in [0, 1], \tag{A7}$$

with $C_{P_1^G}(P_1^G) = C(P_1^G)$ and $P_1^G = \arg\min_{P_1^*} \|C_{P_1^*}(\cdot) - C(\cdot)\|$.

where the notation $\|\cdot\|$ stands for a criterion, not necessarily a rigorous norm. Clearly, $C_{P_1^G}(P_1) \geqslant C(P_1)$ and $C_{P_1^G}(P_1^G) = C(P_1^G)$. The $\|\cdot\|$ here is taken to be a norm of function space and it is defined by

$$\|f(\cdot)\| = \sup\{f(x)|x \in \Omega\},$$

where Ω is the definition domain of f (see [9]), the GRS above is just the globally robust minimax solution, as shown in Fig. A5[1)]. Of course, one may use other norms of the function space in (A7) as a criterion to satisfy own favors or practical requirements.

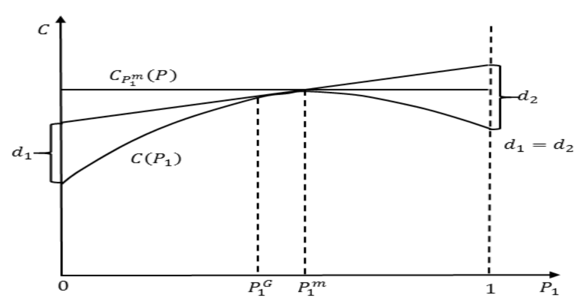

Figure A5 The minimum cost curve $C(P_1)$ and the GRS minimax cost tangent line $C_{P_1^m}(P_1)$, length $d_1 = d_2$.

Appendix A.4 Advantages of the GRS

It is easy from the formulation in Eq. (A7) and Fig. A5 to see that, except the GRS cannot give a guaranteed cost upper bound as the traditional minimax decision, it has the following advantages:

Advantage 1. Global Robustness

The design $\mathcal{H}_0\left(P_1^G\right)$ is a globally robust solution in terms of minimum $\|C_{P_1^G}(\cdot) - C(\cdot)\|$ since $C(P_1)$ is the cost functional of the entire set of all possible actual solutions of the uncertain problem. In other words, the GRS globally suppresses the deviations from all points of the minimum cost curve $C(P_1)$.

1) Clearly, when the two lengths $d_1 = d_2$, the tangent line in Fig. A5 is the GRS, i.e., the globally robust minimax solution.

Advantage 2. Utilizing More Given Problem Knowledge

$C_{P_1^G}(P_1)$ does depend strongly on the entire problem knowledge. Once the latter has a little change, $C_{P_1^G}(P_1)$ will be very likely to change. For example, it is easy to see that in Figs. A3 and A4, the GRS will not be the same as the traditional minimum solution. Even in the case of Fig. A4, all uncertain problems can have different nontrivial GRS, rather than the same trivial solution $P_1 = 0$. Besides, in the following Section **??**, we will show that there are more criteria for the GRS to select so as to sufficiently utilize the given problem knowledge and satisfy the practical requirements.

Advantage 3. More Applicabilities

Naturally, due to the above two advantages, the GRS $C_{P_1^G}(P_1)$ can be applied to more practical scenarios and in general has a better performance than the traditional minimax solutions as shown by the examples in Section Appendix B. In particular, the GRS has no those trivial solution cases presented in Drawback 3 of the traditional minimax. For example, even if the two families of uncertain conditional densities $p(\mathbf{y}|\mathbf{H}_0)$ and $p(\mathbf{y}|\mathbf{H}_1)$ have non-empty intersection in [7] and [8], the GRS can be nontrivial as well.

Appendix B Examples of the GRS

Since the GRS needs to suppress the deviation of the robust solution as small as possible from every optimal solution of all actual problems included in the uncertain problem assumptions, in particular, if $\mathcal{A}$ is not a parameter vector set but a function set, naturally, it is not easy to derive the GRS in general. Certainly, there exist a lot of uncertain problems for people to difficultly derive the GRS, even the traditional minimax solution. For example, suppose the uncertainty of the prior probability P_1 in Subsection Appendix A.1 to be replaced by the uncertainty of the two conditional densities $p(\mathbf{y}|\mathbf{H}_i)$. So far, such uncertain hypothesis testing problems have been solved for a relatively narrow type of uncertain conditional densities [7] and [8] in terms of the traditional minimax, for details, refer to the overview article [10]. However, there are quite a few uncertain problems that can obtain their GRS's, where the uncertainty of problems can be parameterized. In the above, we present the example of hypothesis testing with uncertain prior probabilities to compare the two different schemes: the traditional minimax and the GRS for uncertain systems. Since this uncertain problem is actually parameterized, basically, all the GRS's in those examples can be derived whenever the two conditional probabilities can be given exactly. In the following, we present another type of hypothesis testing examples.

Signal Detection With Uncertain Noise Variances

If the conditional probability densities are uncertain but can be parameterized, the GRS is also derivable in general. In practice, a class of very popular and useful signal detections with uncertain noise statistical knowledge will be solved by the GRS in terms of the aforementioned three criteria. For simplicity of computation, we only show a numerical example to derive the GRS of a signal detection problem with uncertain noises. More generally, when the signal and the noise are both uncertain, or one wants to detect the two random signals with different uncertain energies (covariances), even the two uncertain sets with partial overlap, one can derive the GRS similarly provided that more dimension computations are tractable. Since signal detection problems in many practical situations can be off-line computed in advance, the derived GRS can be used to the uncertain detecting problem. For simplicity of computation and plot presentation, in this subsection, we consider the following signal detection problem: the signal and noises observation model are

$$\mathbf{H}_0 : \mathbf{y} = \nu, \quad \mathbf{H}_1 : \mathbf{y} = s + \nu.$$

where s and ν are mutually independent. In this example we consider the detection of a Gaussian signal with Gaussian noise:

$$s \sim N(m,\ \sigma_s^2),\ \ \nu \sim N(0,\ \sigma_\nu^2).$$

The two conditional probability density functions given $\mathbf{H}_0$ and $\mathbf{H}_1$ are

$$\mathbf{y}|\mathbf{H_0} \sim \mathbf{N}(\mathbf{0},\ \sigma_\nu^{\mathbf{2}}), \quad \mathbf{y}|\mathbf{H_1} \sim \mathbf{N}(\mathbf{m},\ \sigma_\mathbf{s}^\mathbf{2} + \sigma_\nu^{\mathbf{2}}),$$

where $m = 5$, $\sigma_s^2 = 8$, the noise variance σ_ν^2 is not known exactly, but its uncertain interval is [0.25, 9], i.e., $\sigma_\nu \in [0.5, 3]$. Besides, assume $P_0 = P_1 = \frac{1}{2}$, and $c_{00} = c_{11} = 0, c_{10} = c_{01} = 1$. Thus, in this case, the cost functional of the binary hypothesis testing is actually the probability of detection error P_{de}. In this example, the following Figs. B1 – B5 are numerical results in terms of the GRM, MAD, and MED, respectively, and this interval [0.5, 3] of σ_ν is discretized with step 0.01 to compute the optimal cost curve $C(\sigma)$, the traditional minimax curve $C_m(\sigma)$, the GRS minimax curve $C_{GRM}(\sigma)$, the MAD curve $C_{MAD}(\sigma)$, and the MED curve $C_{MED}(\sigma)$.

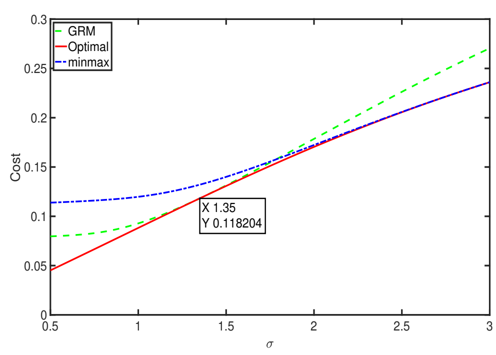

Figure B1 The cost curves of GRM, traditional minimax and optimal decisons.

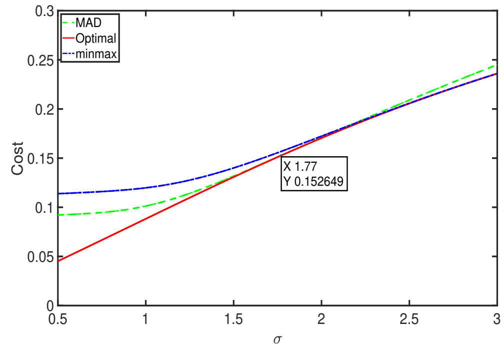

Figure B2 The cost curves of MAD, traditional minimax and optimal decisons.

Table B1 Comparison of the truncated $N(1.5,\ 0.25)$ against minmax.

criterion	sample aver. P_{de}	improve
$\sigma_m = 3$	0.14150	0
$\sigma_{GRM} = 1.35$	0.13236	6.4%
$\sigma_{MAD} = 1.77$	0.13268	6.2%
$\sigma_{MED} = 1.51$	0.13171	6.9%

It is well-known that the most popular and possible distributions are the truncated Gaussian and uniform distribution in practical applications, where the truncated Gaussian densities of σ are possibly in the middle of interval $[0.5, 3]$, or far away from, or close to the least favorable $\sigma = 3$, i.e., σ may follow the three truncated Gaussian densities $N(1.5,\ 0.25)$, $N(1,\ 0.25)$, and $N(2.5,\ 0.25)$ on interval $[0.5, 3]$. Their cost curves are presented respectively in the following Figs.

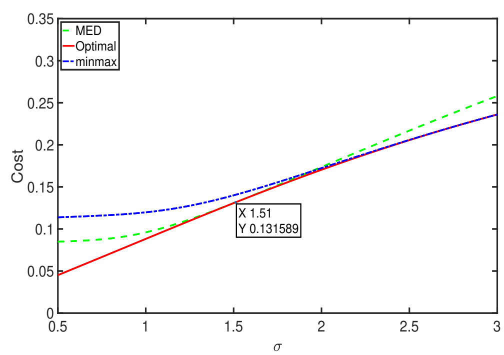

Figure B3 The cost curves of MED $N(1.5,\ 0.25)$, traditional minimax and optimal decisions.

Observation 1. It is easy to see that the cost functional in terms of the traditional minimax does provide a guaranteed upper bound of the decision cost at the maximum noise variance $\sigma_\nu = 3$ that is a trivial solution and can be easily seen intuitively. However, the GRS's in terms of GRM, MAD, and MED are all nontrivial solutions depending on the different global uncertainty of σ.

Observation 2. According to the assumptions of this example, the cost curves are all the detection error probability curves in terms of their criteria. To show the performance differences between with and without the given above three truncated Gaussian distributions and the uniform distribution of the uncertain parameter σ on interval $[0.5, 3]$, we apply Monte-Carlo samples of 10000 times obeying the given four distributions, respectively, to derive the following sample average decision error probabilities of the traditional minimax solution $\sigma_m = 3$, the globally robust solutions $\sigma_{GRM} = 1.35$, $\sigma_{MAD} = 1.77$ and the σ_{MED}s depending on the three corresponding truncated Gaussian distributions and uniform distribution, where the decision error probability of every σ sample has been given in the Figs. B1 – B5. These numerical results are presented in the following tables B1, B2, B3 and B4.

Observation 3. From the above Monte Carlo experiment result tables, it can be seen that leveraging the given distribution of the uncertain parameter σ significantly improves decision performance (with improvements of 6.4%, 6.2%, and 6.9% in Table B1, and 20.7%, 14.3%, and 23.7% in Table B2, among others). In contrast, the traditional minimax approach cannot exploit the information about the uncertainty distribution, even when it is well-known[2)]. Thus, utilizing prior knowledge of the uncertainty is a key advantage of the new scheme.

2) The GRM also shares this limitation due to the minimax definition.

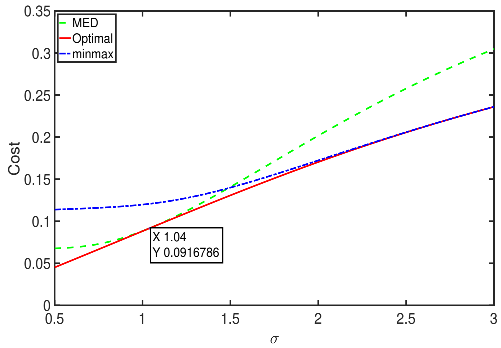

Figure B4 The cost curves of MED $N(1,\ 0.25)$, traditional minimax and optimal decisions.

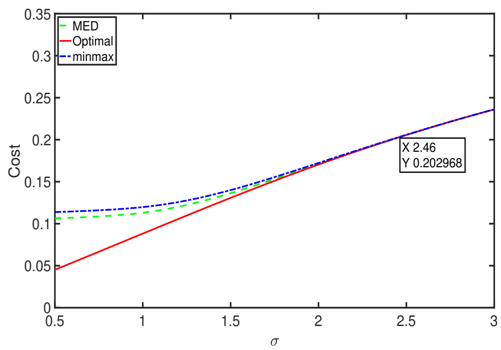

Figure B5 The cost curves of MED $N(2.5,\ 0.25)$, traditional minimax and optimal decisions.

Table B2 Comparison of the truncated $N(1,\ 0.25)$ against minmax.

criterion	sample aver. P_{de}	improve
$\sigma_m = 3$	0.12179	0
$\sigma_{GRM} = 1.35$	0.09645	20.7%
$\sigma_{MAD} = 1.77$	0.10431	14.3%
$\sigma_{MED} = 1.04$	0.09293	23.7%

Table B3 Comparison of the truncated $N(2.5,\ 0.25)$ against minmax.

criterion	sample aver. P_{de}	improve
$\sigma_m = 3$	0.20458	0
$\sigma_{GRM} = 1.35$	0.22461	-0.811%
$\sigma_{MAD} = 1.77$	0.20774	-0.098%
$\sigma_{MED} = 2.46$	0.20427	0.002%

Table B4 Comparison of the uniform distribution against minmax.

criterion	sample aver. P_{de}	improve
$\sigma_m = 3$	0.16211	0
$\sigma_{GRM} = 1.35$	0.16006	1.9%
$\sigma_{MAD} = 1.77$	0.15567	3.9%
$\sigma_{MED} = 1.77$	0.15567	3.9%

Additionally, while the traditional minimax method provides the minimum cost upper bound, this benefit often comes with a performance trade-off, particularly when the least favorable case is located at the boundary of the uncertainty set, whereas the actual uncertain parameters are more likely to deviate significantly from the boundary. As this uncertain signal decision example shows, only when the samples of σ are close to $\sigma = 3$, i.e., σ follows density $N(2.5,\ 0.25)$, the traditional minimax can perform well. However, its performance still can not be better than the MED. In the other two cases: $N(1,\ 0.25)$ and $N(1.5,\ 0.25)$, the GRS's in general outperform the traditional minimax as shown in the above tables.

References

1 Van Trees H L. *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory*. John Wiley & Sons, 2004.

2 Zhu Y M. Multisensor decision and estimation fusion. Kluwer Academic Publishers, 2003.

3 Eldar Y C, Merhav N. A competitive minimax approach to robust estimation of random parameters. IEEE Trans. Signal Process, 2004. 52(7):1931–1946.

4 Bertsekas D, Rhodes I. Recursive state estimation for a set-membership description of uncertainty. IEEE Trans. Automat. Contr., 1971. 16(2):117–128.

5 Sawa T, Hiromatsu T. Minimax regret significance points for a preliminary test in regression analysis. Econometrica, 1973. pages 1093–1101.

6 Ohtani K, Toyoda T. Minimax regret critical values for a preliminary test in pooling variances. Journal of the Japan Statistical Society, Japanese Issue, 1978. 8(1):15–20.

7 Levy B C. Robust hypothesis testing with a relative entropy tolerance. IEEE Trans. Inf. Theory, 2009. 55(1):413–421.

8 Ma T, Song E B, Shi Q J. Globally convergent gradient projection type algorithms for a class of robust hypothesis testings. IEEE Trans. Signal Process, 2021. 69:1828–1841.

9 Stein E M, Shakarchi R. Functional analysis: introduction to further topics in analysis, volume 4. Princeton University Press, 2011.

10 Fauß M, Zoubir A M, Poor H V. Minimax robust detection: Classic results and recent advances. IEEE Trans. Signal Process., 2021. 69:2252–2283.