Dr. Michael Clark
Associate Editor
Motorola
Mail Drop 1J
5555 N. Beach St.
Ft. Worth, TX 76137-2794

Dear Dr. Clark:

We have prepared the enclosed revision of paper SP-7811 which was submitted for publication in the IEEE Transactions on Signal Processing. Note the minor change in title and the change in ordering of authors which was necessary given the nature of our extensive revisions. The comments of the reviewers were very helpful and we would like to thank the first three for their thorough reviews. In addition to a point-by-point response to each of the four reviews, we enclose a general description of the revisions.

We look forward to hearing from you soon on the status of the revision.

Respectfully,

Alfred Hero Associate Prof. of EECS

Bias-Variance tradeoffs for parametric estimation problems...

General Response to Reviewers

We regret that the presentation in the original manuscript contained some confusing statements, notations, and applications and thank the reviewers for pointing many of these out to us. The revised manuscript has been completely rewritten so as to more clearly describe our contribution. The comments of all four reviewers have contributed to what we think is a much improved presentation and their comments have motivated us to considerably sharpen results in this revision.

An outline of some of the more important changes in our revision follows below. We respond point-by-point to each of the reviews on the attached sheets.

1. The introduction has been rewritten to emphasize that our paper addresses fundamental bias-variance tradeoffs as opposed to MSE. We give several examples and bibliographic citations from a wide literature, including linear and non-linear regression, time series analysis, and image and signal processing, to support our use of the bias-variance plane for estimator design and performance evaluation.

2. We have demonstrated the usefulness of the uniform bound analysis by establishing the following:

   (a) *usefulness for validation of the correctness of simulations of biased estimators*: we can only be confident of simulations that give results lying above the fundamental bias-variance tradeoff curve.

   (b) *usefulness to prove estimator optimality in the bias-variance tradeoff plane*:for the inverse problems considered in the paper we display estimators that achieve the bound.

   (c) *usefulness of bound structure to suggest structure of new optimal estimators*: for ill-posed (rank deficient) linear inverse problems the mathematical form of the optimal bias-variance tradeoff curve suggested a way to construct a new estimator which can be made to achieve the optimal curve.

3. Motivated by a suggestion of reviewer 1, we have generalized our setting to cover estimation of arbitrary differentiable functionals of the parameter vector. Two functional estimation examples have been included in the Applications section.

4. We have included a derivation which relates the norm of the bias gradient directly to the maximal variation of the bias over a region of parameter space. All the theorems and applications have been modified to account for an arbitary weighted norm constraint on the bias gradient.

5. Using the relation between bias gradient norm and maximal bias variation we have derive a tight lower bound on the bias gradient norm which specifies a region of the bias variance plane where the uniform CR bound is definitely not achievable. This lower bound is greater than zero only for singular Fisher information matrices.

6. We have shown for the inverse problems considered that the bias gradient is essentially a measure of spatial resolution for linear and non-linear penalized maximum likelihood estimators of a point source.

7. We have eliminated the third example (edge detection), where the interpretation of the bias gradient is not yet clear, in order to shorten and improve the paper.

8. We have given concise computational recipes for evaluating the uniform bound and for placing an estimator on the bias-variance plane via simulation. For the latter, we introduce a bootstrap estimator of bias gradient norm to improve accuracy. The advantage of the bootstrap is illustrated in our new simulations for the Poisson inverse problem.

9. For the Poisson inverse problem we give asymptotic formulas for the bias and variance of penalized maximum likelihood estimators. Simple extensions of these formulas can be used to compute asymptotic bias and variance approximations for other maximum likelihood estimation problems.

10. To minimize possible confusion we have renamed the bias-variance tradeoff plane the delta-sigma tradeoff plane.

Bias-Variance tradeoffs for parametric estimation problems...

Author's response to reviewer #1

We would like to thank this reviewer for a very careful review. Responses to his numbered comments follow.

1.,2.,3. We agree with the author's three criteria for usefulness of a bound (actually we would add a fourth criterion, see 2.a of "General comments"). However, we disagree with the reviewer's comment that $B(\theta, \delta)$ is not useful in the sense of any of the three criteria. The reviewer's statement is true only if one adopts MSE as the criterion of interest. As discussed in the revision there are many excellent motivations (and precedents) for using bias and variance as performance criteria instead of, or in addition to, MSE.

As shown in the revision, the bias (actually the more relevant variation in bias) can be directly related to the bias gradient norm. Furthermore, we have demonstrated that a well known class of estimators (quadratically penalized maximum likelihood) are optimal in terms of attaining the best possible bias-variance tradeoff. In our opinion this provides sufficient evidence of the usefulness of our bound.

4.,5. We agree that $B(\theta, b)$ is not a bound. In the original m.s. the plots containing $B(\theta, b)$ were included to indicate the behavior of the bias - all plots of $B(\theta, b)$ have been removed in the revision and the discussion of $B(\theta, b)$ also has been eliminated.

6. We presume the reviewer is referring to the statement about achievability of the uniform bound on the top of p. 8 of the original m.s. There it was emphasized that only local achievability is guaranteed where the definition "local" is described fully in reference [11] of the revised m.s.. Since this is a technical point which would lengthen an already lengthy paper we elected to refer the interested reader to [11] (in revised m.s.) for the details. In the paper we show the stronger property that the bound is theoretically achievable (globally) in the linear Gaussian inverse problem and experimentally achievable (for large δ) in the non-linear Poisson inverse problem.

7. In the revision we have removed all references to "achievable regions" except for the case for which we can prove that the region above the bound is truly achievable.

8. Moving the definitions out of Theorem 1 would lengthen the paper. We have reworded the statement of the theorem to make it more readable.

9. Theorem 1 has been restructured so that $B(\theta, 0)$ no longer appears in this equation (in the original m.s. we wrote $B(\theta, \delta)$ in this way only to make it obvious that $B(\theta, \delta) \leq B(\theta, 0)$ with equality at $\delta = 0$)

10. In the proof of Theorem 1 (moved to Appendix A) it is shown that a solution $\lambda \in [0, \infty)$ to the constraint always exists and is unique. This has been spelled out in the Theorem.

11. Corrected.

12. We did not spell this out in the m.s. because of space limitations; in the revision we have added two sentences of explanation. The reader is still referred to the report [11] (in revised m.s.) for details – note that [11] is an easily accessible publication which can be obtained by writing to M.I.T. Lincoln Laboratory. Many IEEE Transactions paper refer the reader to in house technical reports for technical details.

13. As pointed out above, all references to $B(\theta, b)$ have been eliminated.

14. Done

15. Convexity of $g(\lambda)$ is only important for root finding, which we do not discuss in this paper. The monotonicity of $g(\lambda)$ is important for establishing existence and uniqueness of a solution λ to $g(\lambda) = \delta^2$. In the revision we have eliminated these comments in the statement of the theorem – monotonicity properties appear only as part of the proof given in Appendix A.

16. Sections 6.1 and 6.2 have been eliminated. The sections on unbiased and biased CR bounds have been combined as suggested.

17. The Appendix on the variance of the bias gradient for the LSE has been eliminated.

18. We appreciate the reviewers comment about the application of the bound to estimation of a functional of many pixels. We initially chose to stick to the single pixel application because of simplicity - we did not want to overly complicate the paper. We have revised the paper to accomodate arbitrary functions of the parameters and have included a bias variance study of contrast estimation in our applications section.

Bias-Variance tradeoffs for parametric estimation problems...

Author's Response to Reviewer #2

We thank this reviewer for the helpful comments and for pointing out the typos. As discussed in response to reviewer #1, we have eliminated Sec. 6.1 and 6.2 entirely.

We appreciate the reviewers comments on relating our results to MSE but emphasize that it is not MSE that we are after (See *General comments to reviewers*), rather it is the bias and variance tradeoffs which are of interest. We agree with this reviewer's general comments (paragraph 4 and 5 of the review) it would be worthwhile to investigate the relation between the local elements going into the CR bound and global MSE. However, we feel that this type of result falls outside of the scope of our paper. It may be of interest to reviewer #2 that we have shown in the MIT Technical Report ([11] of new m.s.) that using small size bias gradients has only a limiting effect on bias but can have a significant effect on (local) MSE. In particular, we showed that one can use the form of the optimal bias gradient which follows from our uniform bound to construct *locally* superefficient estimators which have significantly reduced MSE (locally). This may provide a good starting point for undertaking the type of study suggested by reviewer#2 but would take us too far afield for the present paper.

Bias-Variance tradeoffs for parametric estimation problems...

Author's Response to Reviewer #3

We thank this reviewer for very helpful recommendations. We have responded to the comment at the bottom of para. 2 by clarifying the meaning of bias gradient in two directions

1. In Section 2 we have derived a general relation between the bias gradient and the maximal variation of the bias, which is more relevant to us than the bias itself since constant bias components are removable by a trivial modification of the estimator.

2. In Section 5 (Applications to inverse problems) we have related bias gradient directly to the spatial resolution of the estimators under study.

We have responded to the main concern (para. 3) of this reviewer by simultaneously displaying several estimators on the bias-variance tradeoff plane for most of the examples presented. This was very easy to do and we agree that it will make the usefulness of the uniform bound more apparent.

Bias-Variance tradeoffs for parametric estimation problems...

Author's Response to Reviewer #4

While we feel that this reviewers assesment of our paper is overly harsh, we accept some of the blame for not explaining our contribution with sufficient clarity in the original m.s. While the MSE may be the sole "interesting quantity" to reviewer #4, there are many applications in statistics, image processing, and signal processing where bias and variance have been adopted as useful criteria in their own right (examples and citations are given in the paper). We emphasize in the revision that our objective is to study fundamental tradeoffs between bias and variance functions and not to study the MSE.

Responses to this reviewers specific comments follow:

1. Contrary to this reviewers opinion, we think that the SP Transactions is of high quality and an appropriate journal for our submission. We chose not to submit our work to the Trans. on IT due to the more practical interests of the SP readership who would be more likely to actually use and benefit from the methodology presented here.

2. We never stated in the Introduction or anywhere else that we "wanted a bound on the MSE of any estimator regardless of its bias." The reviewer may have erroneously inferred this from our opening paragraph which mentions the MSE. We regret this misunderstanding and have completely rewritten the Introduction to more clearly state our objective. With the exception of the unpublished reference by VanTrees, we are familiar with all references provided by the reviewer. These latter references all are Bayesian type bounds, i.e. they assume random parameters, and cannot be used for bias-variance tradeoff analysis since they are independent of bias. They are therefore not relevant to the focus of this paper.

3. In the revision the bias gradient norm has been related to the maximal variation of the bias function and also to the spatial resolution in our application. The usefulness of the bound has been demonstrated for several applications (see *General response to reviewers* item 2). Therefore, in so far as bias-variance tradeoffs are concerned, for the applications studied we think our results demonstrate that one should be interested in the bias gradient norm.