REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
03-03-2022	Final Report	14-Dec-2016 - 13-Nov-2021

4. TITLE AND SUBTITLE	
Final Report: Almost-Smooth Nonparametric Regression and Pattern Recognition (W911NF-12-R-0011-03; 3.2.2 Statistical Analysis and Methods; d. Bayesian and Non-parametric Statistics)	5a. CONTRACT NUMBER W911NF-17-1-0040
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102
6. AUTHORS	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
University of Kentucky 500 South Limestone Street 109 Kinkead Hall Lexington, KY 40526 -0001	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	ARO
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 68125-MA.3

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UU	UU	UU	UU		Richard Charnigo
					19b. TELEPHONE NUMBER 859-218-2072

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

RPPR Final Report
as of 04-Mar-2022

Agency Code: 21XD

Proposal Number: 68125MA
Agreement Number: W911NF-17-1-0040

INVESTIGATOR(S):

Name: Richard Charnigo
Email: richard.charnigo@uky.edu
Phone Number: 8592182072
Principal: Y

Organization: **University of Kentucky**
Address: 500 South Limestone Street, Lexington, KY 405260001
Country: USA

DUNS Number: 939017877
EIN: 616033693
Report Date: 13-Feb-2022
Date Received: 03-Mar-2022
Final Report for Period Beginning 14-Dec-2016 and Ending 13-Nov-2021
Title: Almost-Smooth Nonparametric Regression and Pattern Recognition (W911NF-12-R-0011-03; 3.2.2 Statistical Analysis and Methods; d. Bayesian and Non-parametric Statistics)
Begin Performance Period: 14-Dec-2016
End Performance Period: 13-Nov-2021
Report Term: 0-Other

Submitted By: Richard Charnigo
Email: richard.charnigo@uky.edu
Phone: (859) 218-2072

Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 1
STEM Participants: 4

Major Goals: Our project has two major thrusts:

(1) We seek to establish methodology for nonparametric regression modeling with an almost smooth mean response function. This includes detection of jump discontinuities in the derivative of a mean response function of one variable (Subproject 1.1), in a mean response function of two variables (Subproject 1.2), and in a mean response function of two variables when a polar coordinate system can effectively reduce dimension (Subproject 1.3).

(2) We seek to establish methodology for identifying the mean response function which describes the data generating mechanism in a nonparametric regression model. This includes study of convergence rates of misclassification probabilities (Subproject 2.1), examination of sampling schemes in this context (Subproject 2.2), and embedding the classification problem into an estimation problem via convex combinations of candidate functions (Subproject 2.3).

Accomplishments: Our accomplishments and post-grant plans are described in the uploaded PDF document.

Training Opportunities: Four graduate student research assistants from the University of Kentucky were employed during the life of the grant:

In 2017, Mr. Sisheng Liu (at the time, a Ph.D. candidate in Statistics) worked on an approach to estimate the number of jump discontinuities in an almost smooth mean response function. This approach was included in his Ph.D. dissertation which was supervised by Professor Charnigo. Mr. Liu graduated with a Ph.D. in Statistics later that year.

In 2021, Ms. Shaowli Kabir (a Ph.D. candidate in Epidemiology and Biostatistics) began to work on what we have referred to (in the Goals section) as Subproject 1.1. She reviewed some relevant literature. However, her director of graduate studies re-assigned her to another assistantship before Ms. Kabir had an opportunity to make further progress. Although the grant has concluded, Subproject 1.1 will still be pursued.

In 2021, Mr. Jiacheng Xu (a Ph.D. student in Statistics) worked on Subproject 1.2. Mr. Xu reviewed some relevant literature, wrote R code, and conducted simulation studies in R. Although the grant has concluded, Mr. Xu plans to continue working on Subproject 1.2 as part of his Ph.D. dissertation under the supervision of Professor Charnigo.

In 2021, Mr. Pengyuan Chen (an M.S. student in Statistics) worked on Subproject 1.3. Mr. Chen reviewed some relevant literature and wrote R code. Although the grant has concluded, Mr. Chen plans to continue working on Subproject 1.3 under the supervision of Professor Charnigo.

Results Dissemination: In 2017, Professor Charnigo presented “Inference for a Nonparametric Model without Nonparametric Estimation” (based on joint work with Professor Srinivasan) at the Joint Statistical Meetings in Baltimore MD. Professor Srinivasan presented “Detecting Discontinuities in a Regression Curve” (based on joint work with Professor Charnigo and Mr. Liu) at the Joint Statistical Meetings in Baltimore MD.

In 2017, Professors Charnigo and Srinivasan engaged Professor Qiu from the University of Florida as a consultant. Professor Qiu visited the University of Kentucky and offered comments on the research in progress. Later (in 2021), Professor Qiu joined the research team by subcontract.

In 2018, Professor Srinivasan presented "Inference for a Nonparametric Model without Nonparametric Estimation" (invited and based on joint work with Professor Charnigo) at the International Symposium on Optimization and Game Theory: Modeling and Computation at the Indian Statistical Institute.

In 2020, Professor Charnigo was scheduled to give a presentation at the ENAR conference in Nashville TN. Unfortunately, the conference was disrupted by the initial outbreak of COVID-19 in the United States.

A manuscript “Classifying Nonparametric Regression Curves” (by Professor Charnigo and Professor Srinivasan) was submitted for journal publication but not accepted. This manuscript will be revised and re-submitted. Ultimately we hope to submit several manuscripts for publication (ideally one for each of the six subprojects).

Honors and Awards: In 2017, Professor Charnigo was a recipient of the “Teacher Who Made a Difference” award from the
University of Kentucky’s College of Education.

In 2017, 2018, 2019, and 2020, Professor Charnigo received Wethington awards from the University of Kentucky for engagement in extramurally funded research.

In 2019 and 2021, Professor Charnigo was elected to the University (of Kentucky) Senate, a body of faculty and students that participates in shared governance of the University of Kentucky.

In 2019, Professor Charnigo was elected to the University of Kentucky’s Senate Council, which is the executive committee for the University Senate.

In 2020, Professor Qiu was the plenary speaker at the 10th International Conference and Workshop on High-Dimensional Data Analysis (online due to COVID-19).

In 2021, Professor Cheng received a Wethington Award from the University of Kentucky for engagement in extramurally funded research.

In 2021, Professor Qiu was the plenary speaker at the 37th ASA Quality and Productivity Research Conference in Tallahassee FL.

In 2021, Professor Charnigo was appointed to and became the chairperson of the Health Care and Clinical Sciences Area Committee at the University of Kentucky.

In 2021 (after the grant concluded but before this report was submitted), Professor Qiu was elected a fellow of the American Society for Quality.

Protocol Activity Status:

Technology Transfer: Nothing to Report

PARTICIPANTS:

Participant Type: PD/PI
Participant: Richard Charnigo
Person Months Worked: 9.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Co PD/PI
Participant: Cidambi Srinivasan
Person Months Worked: 7.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Co PD/PI
Participant: Qiang Cheng
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Faculty

Participant: Peihua Qiu
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Sisheng Liu
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Shaowli Kabir
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Jiacheng Xu
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Pengyuan Chen
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Partners

,

I certify that the information in the report is complete and accurate:
Signature: Richard Charnigo
Signature Date: 3/3/22 8:49PM

ACCOMPLISHMENTS AND POST-GRANT PLANS

We thank the Army Research Office – including Dr. Joseph Myers, Dr. Andrew Vlasic, and Dr. Michael Lavine – for support of our project. Although the grant has now concluded, we do not intend to abandon our work. We hope to eventually submit several papers to reputable peer-review journals.

Two main objectives for our project are as follows:

First, we seek to establish methodology for nonparametric regression modeling with an "almost smooth" mean response function. (This is a mean response function that is multiple-times differentiable on most of its domain but for which jump discontinuities may be present in either the function itself or its derivatives.)

Second, we seek to establish methodology for identifying which of finitely many candidate mean response functions describes the actual data generating mechanism in nonparametric regression modeling.

For each of the two main objectives, there are three subprojects, described in detail below. Now, a few months after conclusion of the grant, the subprojects are at various stages of progress, as indicated below.

Virtually all of the effort since the preceding interim progress report has been in Subprojects 1.1, 1.2, and 1.3. The most substantial updates to descriptions from our preceding interim progress report are also with respect to Subprojects 1.1, 1.2, and 1.3. In addition, the description of Subproject 2.2 has been updated to include an illustrative graphic.

SUBPROJECT 1.1

Here we seek to formulate an information-type criterion for identifying the number of discontinuities in the first derivative of a mean response function of one design variable. The mean response function itself will be assumed continuous. (This is a special case of "almost smooth", in which discontinuities manifest only in derivatives.) The criterion will consist of a normal log likelihood plus a penalty.

Our aim is to specify the penalty so that, as the sample size increases: (a) the number of discontinuities is correctly identified with probability approaching one; (b) the errors in estimating the locations of the discontinuities tend to zero in a probabilistic sense; and, (c) the errors in estimating the magnitudes of the discontinuities also tend to zero in a probabilistic sense.

This will not be as simple as adopting a penalty like that in the Bayesian Information Criterion (BIC). While the BIC is capable of consistently identifying the order of a statistical model in standard problems, Xia and Qiu (2015) (please see References section at the end of this document) employed a different kind of penalty when identifying the number of discontinuities in a mean response function of one design variable.

Although the present problem differs from that considered by Xia and Qiu (2015), in that we are seeking discontinuities in the first derivative rather than in the mean response function itself, we anticipate that a penalty resembling theirs may be suitable for the present problem. However, instead of a penalty based on magnitudes of estimated discontinuities in the mean response function, we will have a penalty based on magnitudes of estimated discontinuities in the first derivative.

In fact, the jump information criterion of Xia and Qiu (2015) was able to accomplish goals (a), (b), and (c) (from three paragraphs above) with reference to the discontinuities in the mean response function. As such, we anticipated that the expertise of Prof. Qiu would be valuable in the pursuit of Subproject 1.1. Accordingly, we sought his participation via a subcontract from the grant.

As a small footnote, we mention that Prof. Charnigo (principal investigator on the grant) has some expertise with model selection criteria. In particular, he was an invited written discussant to the work of Drton and Plummer (2017) on model selection when overparameterization induces singularity of the Fisher information matrix.

Since the previous interim progress report, we undertook a detailed review of Xia and Qiu (2015), and its theoretical supplement, to gain insight into how the above goals (particularly, goal (a)) may be pursued in our context (i.e., with a continuous mean response function but a potentially discontinuous first derivative).

To begin with, the relevant statistical model is

$$Y_i = f(x_i) + \varepsilon_i$$

for $i = 1, 2, \ldots, n$, where n is the sample size, x_i is the i^{th} value of the design variable (which we will assume equals i / n), $f(x)$ is the mean response function whose domain is the closed interval $[0, 1]$, ε_i is a random term with zero mean, and Y_i is the i^{th} value of the outcome variable.

In our context, and here there is a difference from the model of Xia and Qiu (2015), we have

$$f(x) = f_C(x) + \sum_{j=1}^{m_0} d_j\left(x - s_j\right)_+$$

where $f_C(x)$ is a function with continuous first derivative, m_0 denotes the true number of jumps in the first derivative of the mean response function, s_j denotes the location of the j^{th} jump, d_j denotes the signed jump at s_j, and $(x - s_j)_+$ denotes the positive part of $(x - s_j)$.

Making the provisional (and possibly incorrect) assumption that $m_0 = m$ (think of m as an integer-valued variable, whereas m_0 is a fixed truth), we can estimate s_1 and d_1 by maximizing (over x) the absolute difference between local quadratic estimators of $f'(x)^-$ and $f'(x)^+$, where the superscripts identify left-handed and right-handed derivatives. More specifically, the estimator of s_1 is the argument maximum, and the estimator of d_1 is the appropriately signed maximum.

To clarify the preceding, note that, wherever $f'(x)$ is continuous, the difference between estimators of $f'(x)^-$ and $f'(x)^+$ should be close to 0. So, a large absolute difference may signal a jump discontinuity in $f'(x)$. Also, this signaling device differs from that of Xia and Qiu (2015), who examined absolute differences between local linear estimators of $f(x)^-$ and $f(x)^+$ rather than of their derivatives.

Next, we can estimate s_2 and d_2 by maximizing (over x) the absolute difference between local quadratic estimators of $f'(x)^-$ and $f'(x)^+$ outside a neighborhood of the estimated s_1. After that, we can estimate s_3 and d_3 by maximizing the absolute difference between local quadratic estimators outside the union of neighborhoods of the estimated s_1 and s_2.

This process will continue until we have estimated $s_1, s_2, \ldots, s_m$ and $d_1, d_2, \ldots, d_m$. Then we can define altered outcomes by $Y_{i,m} := Y_i - \sum_{j=1}^{m} \widehat{d_j}\left(x_i - \widehat{s_j}\right)_+$ for $i = 1, 2, \ldots, n$, where quantities with circumflexes ("hats") represent estimators. To the altered outcomes, we will apply a conventional smoothing method (in particular, one which assumes a continuous first derivative). This will yield an estimator of $f_C(x)$, which we can denote by $\widehat{f_{C,m}(x)}$.

Then we can define an estimator of $f(x)$ by $\widehat{f_m(x)} := \widehat{f_{C,m}(x)} + \sum_{j=1}^{m} \widehat{d_j}\left(x - \widehat{s_j}\right)_+$. By construction, this estimator has exactly m discontinuities in its first derivative.

Next, we can define a criterion by a log residual sum of squares plus a penalty,

$$\log\left(\sum_{i=1}^{n} \left[Y_i - \widehat{f_m(x_i)}\right]^2\right) + P(n) \sum_{j=1}^{m} \left|\,\widehat{d_j}\right|^{-1}$$

where $P(n)$ is a penalty coefficient which is largely driven by the sample size n. The exact form of $P(n)$ remains to be determined, but it is not expected to match the penalty coefficient of Xia and Qiu (2015).

The criterion balances fidelity to the data against model complexity. More specifically, as m becomes larger, we anticipate that $\widehat{f_m(x)}$ will become closer to $f(x)$ and, therefore, that the log residual sum of squares will decrease. This will be the case especially over the range $1 \leq m \leq m_0$ but still, we believe, to some degree when $m > m_0$. (The latter reflects an "overfitting" phenomenon; see, e.g., Hastie, Tibshirani, and Friedman, 2009.)

On the other hand, when $m > m_0$ (i.e., the assumed number of jumps is greater than the true number of jumps), we anticipate that $\widehat{d_j}$ will be close to 0 for $j > m_0$, because there are no nonzero jumps remaining to be estimated. Thus, since the penalty depends on $|\widehat{d_J}|^{-1}$, the criterion may be greatly inflated when $m > m_0$.

The preceding two paragraphs are the heuristic basis for our expectation that, when n is large, the criterion will very likely be minimized when $m = m_0$. However, we will need to formally prove, for appropriately specified P(n) and under other appropriate assumptions, that the probability of the criterion being minimized at $m = m_0$ tends to 1 as the sample size n tends to infinity. In statistical parlance, we will need to formally prove that the criterion is consistent.

Before concluding this portion of the report, let us mention a couple of issues which emerged during the detailed review of Xia and Qiu (2015).

First, consider the following proposition: *Let k be an arbitrary but fixed positive integer other than m_0. Then, with probability approaching 1 as the sample size tends to infinity, the criterion will be smaller at $m = m_0$ than at $m = k$.*

The proposition above seems equivalent to consistency of the criterion (as described three paragraphs above) but is actually weaker. In essence, that's because the proposition does not rule out the possibility that the sequence of minimizers of the criterion may drift to infinity as the sample size tends to infinity. We do not believe that will be the case. However, the detailed review of Xia and Qiu (2015) heightened our awareness of the distinction between the proposition above and consistency of the criterion.

If nothing else, we can resolve this issue by restricting the domain of the criterion to a finite subset of the positive integers (i.e., by assuming *a priori* that there exists a known upper bound M for m_0). An analogous restriction has been imposed by, e.g., Drton and Plummer (2017) when they developed their singular Bayesian information criterion.

Second, an appropriate choice of penalty coefficient P(n) will depend in large part on how the residual sum of squares changes with m. Intuition does not provide an immediate answer here. However, the residual sum of squares may be less sensitive to m in our context of detecting discontinuities in a first derivative, versus the problem addressed by Xia and Qiu (2015), namely detecting discontinuities in the mean response function itself. That's because the residual sum of squares evaluates the fit of $\widehat{f_m(x)}$ to the data, and $\widehat{f_m(x)}$ is continuous by construction.

An alternative approach, considered by us at an earlier stage but now deemed beyond the scope of Subproject 1.1, would have been to create a criterion that more directly evaluated the fit of the derivative of $\widehat{f_m(x)}$. This would have required us to identify appropriate analogues to the Y_i in the residual sum of squares. A possibility for that would have been to use the empirical derivatives defined by Charnigo, Hall, and Srinivasan (2011).

SUBPROJECT 1.2

Here we wish to formulate an information-type criterion for recovering a discontinuous mean response function of two design variables, in a way that balances control over squared error with accuracy of discontinuity detection.

Squared error is a traditional evaluation metric and is substantially driven by accuracy in estimating the mean response function over subdomains on which it is continuous. Accuracy of discontinuity detection is not as often considered, because allowing discontinuities is not always part of the statistical modeling strategy, even in nonparametric settings (e.g., Charnigo and Srinivasan, 2011; Charnigo, Feng, and Srinivasan, 2015). Yet, discontinuity detection can be important. For example, detecting discontinuities in a two-dimensional image helps to distinguish physical structures (either normal or pathological).

At first, Subproject 1.2 may seem to be a two-dimensional generalization of Xia and Qiu (2015). However, the multi-dimensional character of Subproject 1.2 introduces a challenge, namely that the presence of a single discontinuity will typically imply the existence of infinitely many discontinuities. This is because discontinuities are not localized to isolated points as in one dimension but rather present along jump location curves (JLC's). For example, if a mean response function corresponds to a two-dimensional image, then the edge of an object shown in the image defines a JLC.

One possible solution to the aforementioned challenge may focus on inferring the number of JLC's rather than the number of discontinuities (which, as noted above, will typically be either zero or infinite). However, with input from Prof. Qiu, we have decided to take a different approach. This will allow us to use our criterion with an underlying method (for recovering a discontinuous mean response function of two design variables) that flags design points likely to be close to JLC's, without requiring that the flagged design points themselves define a curve. (However, see also Subproject 1.3.) The method of Qiu (1998) is one such; others are documented by Chu and colleagues (2012).

The penalty (to the log likelihood in our criterion) is then to be defined not based on an estimated number of JLC's but based on an index of "edginess" for the estimated mean response function. This "edginess" index, originally proposed by Hall and Qiu (2007), can be calculated for any underlying method; all that is required is an estimated mean response function on a two-dimensional grid of design points. We mention here that the context for Hall and Qiu (2007) was image deblurring and also that Prof. Charnigo has had previous experience with imaging (Charnigo, Sun, and Muzic, 2006).

Regarding progress on Subproject 1.2, we have translated relevant Fortran code (courtesy of Prof. Qiu) from Qiu (1998) and from Hall and Qiu (2007) into R, our statistical programming language of choice. We have also initiated simulation studies to show how the log likelihood and the "edginess" index behave for different tuning parameters in the method of Qiu (1998). These simulation studies are also intended to reveal the influence of tuning parameters on squared error and accuracy of discontinuity detection, as quantified by Hausdorff distance between the set of points on a JLC and the set of flagged design points.

Since the previous interim progress report, there has been debugging of the R code, a necessary if not glamorous task, and preliminary investigation of the behavior of our criterion.

The relevant tuning parameters from Qiu (1998) may be denoted k and k′. They are integer-valued and represent the amount of smoothing both in the edge detection itself and in subsequent estimation of the mean response function (i.e., obtaining a fitted surface).

Tentatively, our criterion has the form

$$(1 - \omega)\ \mathrm{RSS}(k, k') + \omega\ A_1(k, k'),$$

where RSS denotes a residual sum of squares and A_1 is the aforementioned "edginess" index from Hall and Qiu (2007). The arguments k and k′ emphasize the dependence of both these quantities on the tuning parameters from Qiu (1998). Above, ω is a number between 0 and 1, which controls the balance between fitting the data and "edginess" (which is like a surrogate for model complexity).

For fixed k and k′, our simulation studies show that the criterion is minimized as ω approaches 1. However, what we're really interested in is the following:

- For fixed ω, which k and k′ will minimize the criterion ?
- Which k and k′ will yield best performance, as judged by squared error and accuracy of discontinuity detection ? (These can be ascertained in simulation studies but not for most real-world data sets, because the mean response function is generally unknown in real-world data sets.)
- For which ω, if any, is there a match between the k and k′ identified in the previous two bullet points ? (The answer may depend on the mean response function itself as well as the sample size and the underlying noise variance.)

A theoretical justification of our criterion for this two-dimensional scenario will be more subtle than for the one-dimensional scenario of Xia and Qiu (2015). We do not promise that we will accomplish it in the post-grant setting (a thorough empirical treatment of the above issues already seems sufficient for a journal paper), but here are some relevant considerations.

Seeking analogues to properties (a), (b), and (c) (mentioned above for Subproject 1.1) does not seem like an appropriate strategy here. Rather, an appropriate strategy may be to define a <u>loss functional</u> that quantifies deviations of an <u>estimated</u> mean response function from the <u>actual</u> mean response function we are trying to recover.

For a mean response function which may have discontinuities, a suitable loss functional will involve both squared error and accuracy of discontinuity detection. The hope is that the loss functional is likely to be nearly minimized when our criterion is minimized. Here, minimization is with respect to tuning parameters of an underlying method (for discontinuity detection and mean response function estimation), such as that of Qiu (1998). Thus, a theoretical justification of our criterion may be specific to the underlying method.

If minimization of our criterion translates into near-minimization of the loss functional, then our criterion may be a helpful proxy for the loss functional in the analysis of real-world data. The loss functional itself will not generally be computable with real-world data, as it will be with simulation studies. This is because the loss functional will depend on the actual mean response function, which is generally unknown with real-world data.

SUBPROJECT 1.3

Here we are again concerned with a discontinuous mean response function of two design variables. Please recall, from the description of Subproject 1.2, that a jump location curve (JLC) describes the positions of discontinuities (or edges) in the two-dimensional plane. The work in Subproject 1.2 allowed design points near JLC's to be flagged without requiring that the flagged points themselves should define a curve. However, there may be situations in which it is desirable to sacrifice this flexibility and insist that the flagged points should define a curve.

For example, Qiu and Sun (2007, 2009) present such a situation in microarray image analysis. Here, the relevant JLC is a boundary that separates the image into a foreground and a background. Gene expression is quantified based on the intensity of image pixels in the foreground. Therefore, defining a curve which estimates the JLC will help determine which pixels should be used to quantify gene expression. We also mention here that Prof. Charnigo has previously done statistical work with microarray data (Dai and Charnigo, 2008 and 2010), albeit at a point downstream from the image analysis described above.

The basic idea of Qiu and Sun (2007, 2009) is that, for a suitably identified central location (or "pole"), one can parametrize a JLC using polar coordinates and then estimate the radius from the central location as a function of the polar angle. This essentially reduces a two-dimensional problem to a one-dimensional problem. However, three concerns may be raised:

First, not every JLC is parametrizable using polar coordinates.

Second, by virtue of the method used to construct it, the estimated JLC is a smooth curve; that is, the estimated JLC is differentiable with respect to the polar angle.

Third, the theory of Qiu and Sun (2007) assumes that the actual JLC is smooth.

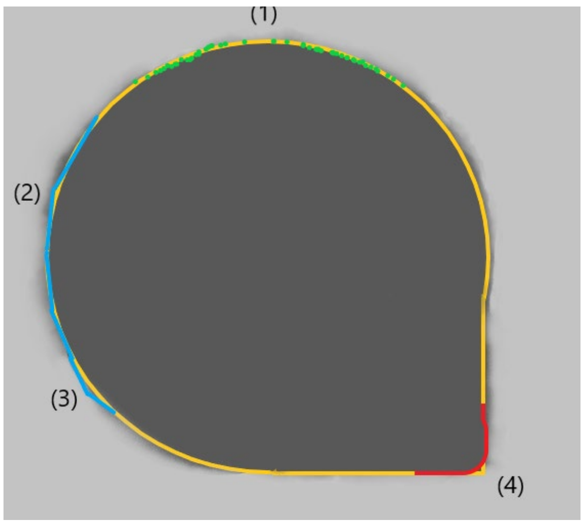
The first concern is perhaps not acute for microarray image analysis, in which the foreground is anticipated to be convex or nearly so; however, this concern does call for future research.

The second concern, on the other hand, falls within the remit of Subproject 1.3. We do not wish for an estimated JLC to be constrained to be a smooth curve, because the actual JLC may not be smooth. A JLC will have corners if, in a two-dimensional image, it represents the boundary of a physical structure that has corners.

In contrast, Chu and colleagues (2012) offer a method for estimating a JLC which does not result in a smooth curve. A limitation of their method, however, is that corners appear as an artifact of the method; they are not necessarily indicative of corners in the actual JLC.

Subproject 1.3 aims to provide methodology for estimating a JLC which can be parametrized using polar coordinates and which may have corners. In particular, corners in the estimated curve will be deliberately placed to identify plausible locations for corners in the actual JLC.

A sketch of our intended methodology is as follows. Given a set of design points flagged as being near a JLC (but not necessarily defining a curve), we can convert these design points to polar coordinate space and perform nonparametric regression of the radius with respect to the polar angle. Where we will depart from Qiu and Sun (2007, 2009) is that this nonparametric regression will allow for discontinuities in the first derivative of the radius (as a function of the polar angle). Translating the results of the nonparametric regression back into the original coordinates, we will then have an estimated JLC. Further, the discontinuities in the first derivative will translate into corners of the estimated JLC.



The graphic at left provides a visual illustration of what we are trying to do. Suppose that grayscale intensity corresponds to the value of a response variable in relation to two spatial coordinates (design variables).

The yellow band represents the locations of discontinuities in the underlying mean response function. In other words, the yellow band is a JLC. The JLC is mostly smooth but has a corner at lower right.

Subproject 1.3 is concerned with estimating the JLC. Some existing approaches have one of the following weaknesses:

- Individual points are flagged but do not themselves constitute a curve, illustrated by the green dots near the top (1).
- The yellow band is approximated by piecewise linear segments, introducing corners as artifacts, illustrated by the blue line segments at left and most visible at locations (2) and (3).
- The yellow band is approximated by a smooth curve, which rounds off corners that actually do exist, illustrated by the red curve at lower right (4).

Regarding progress on Subproject 1.3, we are well underway in writing R code that implements the nonparametric regression scheme of Hoo and Qiu (2009). This nonparametric regression scheme allows for discontinuities in a mean response function and in its first derivative. The latter is of relevance to Subproject 1.3; as mentioned earlier, discontinuities in the first derivative of the radius (as a function of the polar angle) will translate into corners of an estimated JLC in the original coordinates.

Writing this R code has been taking more time than anticipated. (There was not existing code available in another programming language that we could translate to R.)

On the other hand, we were able to obtain existing GAUSS code for the method of Chu and colleagues (2012) (courtesy of Prof. Deng), and we have also started to translate that code to R. The relevance of this latter code is that we wish to have a competing method available when we begin numerical evaluation of our methodology for Subproject 1.3.

Since the previous interim progress report, there has been further work on and debugging of the R code implementing the scheme of Hoo and Qiu (2009), a necessary if not glamorous task.

Finally, regarding the third concern mentioned several paragraphs above, a theoretical justification for our methodology would also be desirable. Such a theoretical justification would entail showing that, with probability approaching one as the sample size increases without bound, the estimated JLC has the same number of corners as the actual JLC, the estimated locations of the corners converge to the actual locations, and (in some appropriate metric, such as Hausdorff distance) the estimated JLC converges to the actual JLC. However, this theoretical work is beyond what we are likely to accomplish in the post-grant setting.

SUBPROJECT 2.1

Suppose outcome data are generated, at values of a single design variable along a grid (in the unit interval), by a mean response function of that design variable with additive errors. The error distribution is assumed to be symmetric (about zero) with finite moment generating function. Neither the error variance nor even the parametric family of the error distribution is necessarily known to the data analyst.

However, the data analyst knows (or reasonably believes) that the mean response function belongs to a finite list of candidate functions to which he or she has access. The candidate functions, or their derivatives, may have jump discontinuities in the interior of the design space. (Thus, if not smooth, the candidate functions are at least "almost smooth".) The data analyst will infer that the actual mean response function is that candidate function with minimal sum of squared deviations from the outcome data at the grid points.

Such a scenario may arise in, for instance, analytic chemistry. A scientist may have Raman spectra (e.g., Vandenabeele, 2013) from two samples of materials which are visually similar to each other but known to be of different compositions, such as diamond and cubic zirconia (cf. Jenkins and Larsen, 2004). (When we speak of samples of materials here, we are referring to the physical sense of having a small quantity of a material, not to the statistical sense of data acquisition.)

The scientist may also have a Raman spectrum from a third sample of material whose composition is uncertain but strongly suspected to match one of the two previous samples. By calculating the sum of squared deviations of vertical coordinates of the third Raman spectrum from those of the first two Raman spectra, the scientist may arrive at a conclusion about the composition of the third sample of material.

This analytic chemistry example is simplified in some ways. For instance, what if the horizontal coordinates of the three Raman spectra are not aligned ?

However, this example still motivates the inferential problem in the second paragraph of this section. And this brings us to the following statistical question: how quickly does the probability of an incorrect inference – or, let us say, a misclassification – converge to zero as the density of the grid (and, therefore, the size of the data set) increases ?

This question can be posed more specifically: since the logarithm of such a probability should tend to negative infinity as the size of the data set increases, can we describe the convergence in terms of limiting behavior for the logarithm of the probability divided by the size of the data set ? Hereafter, we refer to this as the "scaled log probability".

We have ascertained that the scaled log probability of a misclassification may have various limiting behaviors, relating to the parametric family of the error distribution and/or the proximity of the candidate mean response functions. A finite limiting behavior for the scaled log probability corresponds to exponential convergence (to zero) of the raw probability of an incorrect inference, while an infinite limiting behavior corresponds to super-exponential convergence.

If the error distribution is normal, then the scaled log probability converges to a negative multiple of the integrated squared difference between candidate mean response functions, which can also be interpreted as a negative multiple of Kullback-Leibler divergence.

If the error distribution is compactly supported, and if the candidate mean response functions are not too close to each other (based on relative sizes of integrated absolute difference and integrated squared difference), then the scaled log probability diverges to negative infinity.

Limiting behaviors for other error distributions are less succinctly described but can in some cases be bounded, in one or the other direction, by the limiting behavior in the normal case.

If there are more than two candidate mean response functions, then two further considerations are as follows.

First, the probability of incorrectly inferring the mean response function to be candidate "B" when it is actually candidate "A", let us say, will not generally equal that of incorrectly inferring the mean response function to be "C" when it is actually "A". Relatedly, the probability of any incorrect inference whatsoever when the mean response function is actually "A" will be greater than the probability of a particular incorrect inference (such as the mean response function being "B").

Second, the limiting behaviors of which we have spoken rely on two regularity conditions which are vacuously satisfied when there are two candidate mean response functions but which have nontrivial content otherwise; more will be said about regularity conditions in the material on Subproject 2.2.

We made multiple attempts to have a manuscript based on this work published in a reputable peer review journal. Unfortunately, we have encountered quite a bit of resistance from editorial boards. This resistance has seemed not entirely warranted but has led us to broaden our work.

In particular, we wondered what would happen if, supposing the parametric family of the error distribution to be known, we selected that candidate mean response function which maximized the log likelihood for the observed data ? Of course, a normal error distribution returns us to an inferential framework based on minimizing a sum of squared deviations. However, other error distributions may yield different inferences.

We have discovered that the scaled log probability converges to the negative integral of Bhattacharyya distances between the error distribution and other distributions which are location-shifted from it by differences between candidate mean response functions. Some by-products of this computation are upper and lower bounds for the limiting behavior, given by negative multiples of integrated squared Hellinger distances and integrated Kullback-Leibler divergences, respectively.

In addition, there are some non-normal error distributions for which the value of an unknown nuisance parameter (related to dispersion) is irrelevant to inference, thereby obviating necessity of its estimation.

The piece that we are still missing is how to cope with non-normal error distributions for which the value of an unknown nuisance parameter may affect inference based on a log likelihood. Will estimation of such a nuisance parameter disturb the findings about scaled log probability (two paragraphs above) ? If we can answer this question, then we will have a manuscript superior to the one previously submitted.

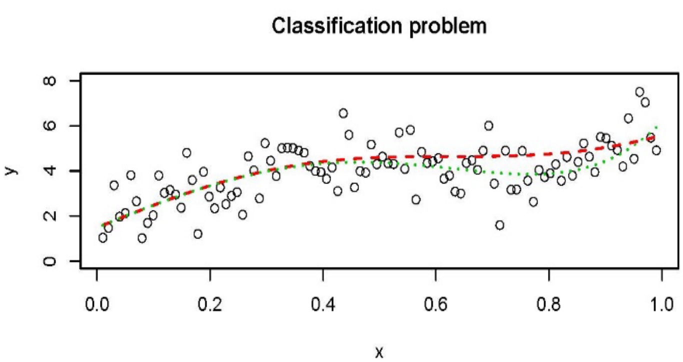
SUBPROJECT 2.2

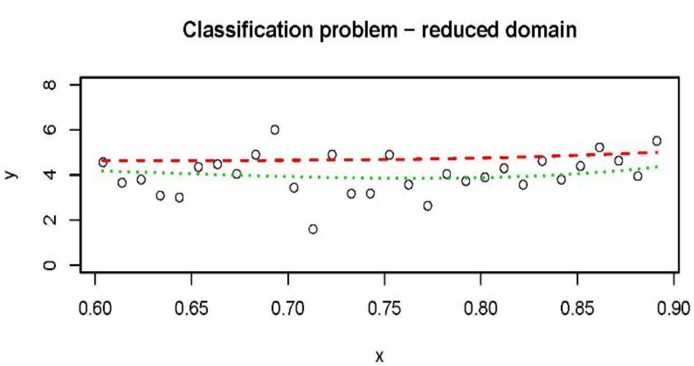
In Subproject 2.1, we were concerned with identifying which of finitely many candidate mean response functions generated outcome data, corresponding to a grid of values for a single design variable. Inspired by a suggestion from the previous program officer, Prof. Michael Lavine, we wondered whether replacing the grid of values by an alternate sampling scheme might improve the limiting behavior of the scaled log probability (i.e., hasten the rate at which the misclassification probability tends to zero). This is, in effect, a question of experimental design. In what follows, we assume a normal error distribution.

To begin with, suppose that there are two candidate mean response functions. Assuming there is a non-empty set of argument maxima for the absolute difference between the two candidate functions (for which a sufficient, but not necessary, condition is continuity of the two candidate functions), the limit of the log scaled probability can be made proportional to the negative of the maximum squared difference between the two candidates. Such a limit will be attained when the design variable is confined to the aforementioned set of argument maxima.

Moreover, except when the two candidate functions differ by a constant, this limit will be greater in magnitude than that obtained when the design variable is sampled from a grid. This is because the maximum squared difference between candidate functions exceeds the integrated squared distance.

Here we also mention a practical caveat, namely that confining the design variable to the set of argument maxima may not be advisable, if there is any reasonable likelihood that the data analyst is mistaken in his or her belief that the actual mean response function is one of the two candidates.





A practical compromise may be to confine the design variable to a subinterval on which the two candidates are fairly well separated. This is illustrated in the graphic at left.

The data (open circles) come from one of two candidate mean response functions: the one in red dashes, or the one in green dots. The least squares method of Subproject 2.1 suggests that the data more plausibly originated from the mean response function in green dots.

However, because the two candidates are nearly coincident over much of their domain, we may wish to confine attention to the reduced domain determined by a subinterval on which the two candidates are fairly well separated. A clearer signal from the data points may then be obtained.

In the present illustrative example, the bottom panel of the graphic suggests that the candidate function in red dashes is implausible; compare the number of data points above versus below.

On the other hand, if we really are confident that the actual mean response function is one of the two candidates, then we can reasonably restrict attention to a very limited number of distinct x values (providing that we can obtain repeated measurements at each), at which the candidate mean response functions are most separated.

Next, suppose that there are D candidate mean response functions. Even if the $D(D-1)/2$ sets of argument maxima for the absolute differences between pairs of candidate functions are all non-empty, their intersection may be empty. Intuitively, good values of the design variable for distinguishing between candidate "A" and candidate "B" may not be useful for distinguishing between candidate "A" and candidate "C". One may suspect that allocating values of the design variable to the union of the $D(D-1)/2$ sets of argument maxima should be optimal. However, for reasons given below, such an allocation strategy may not be advisable.

Extending the results from Subproject 2.1, we have, under some <u>regularity conditions</u>, that the scaled log probability of incorrectly inferring the mean response function to be "B" when it is actually "A" converges to a negative multiple of an integrated squared distance between "B" and "A", the integral now taken with respect to a probability measure describing the limiting distribution of the design variable. (The results from Subproject 2.1 correspond to Lebesgue measure on the unit interval and a uniform distribution of the design variable.)

As such, we are interested in identifying a probability measure which maximizes the smallest integrated squared distance between candidate functions. (This "maximin" idea is a sort of reversal of the "minimax" concept seen elsewhere in statistical inference.)

If the candidate mean response functions are continuous, then we have a very clean result, namely that there exists a "maximin" probability measure which is <u>finitely supported</u>. This is a bit reminiscent of Lindsay's (1983) result on nonparametric maximum likelihood estimation of a mixing distribution, which Lindsay himself noted was related to the theory of optimal design.

In fact, there exists a "maximin" probability measure which is supported on at most $D(D-1)/2$ points, where D is the number of candidate functions. Interestingly, the support points of a "maximin" probability measure need <u>not</u> be among the $D(D-1)/2$ sets of argument maxima mentioned a few paragraphs above. This is why such an allocation strategy may not be advisable.

For further perspective, a theorem in the monograph by Silvey (1980) can be made to apply to our problem, under an additional assumption concerning the candidate mean response functions. This theorem from Silvey (1980) yields an upper bound for the number of support points of a "maximin" probability measure. This upper bound is on the order of D^4. In contrast, our result provides an upper bound on the order of D^2.

Continuous candidate mean response functions may be smooth, or they may be "almost smooth" in that their derivatives may have jump discontinuities. To accommodate "almost smooth" candidate mean response functions which themselves have discontinuities, some restriction is needed on the class of probability measures over which we optimize. (This seems unavoidable, unless, perhaps, we consider some notion of "near optimality". In fact, a counterexample can be provided even with $D = 2$, showing that, without continuity of the candidate functions, no “maximin” probability measure exists in the class of all probability measures on the design space.)

One possibility is to restrict attention to a class of probability measures whose support sets are contained in an arbitrary finite subset of the design space. In this class, a “maximin” probability measure exists which is supported on at most $D(D-1)/2$ points. The above restriction is technically inelegant but does not render the theorem vacuous. The arbitrary finite subset can consist of dozens, hundreds, or thousands of points. Moreover, in a scientific application (like the analytic chemistry example mentioned for Subproject 2.1), limited precision of a measuring instrument may effectively reduce the design space to a discrete set of points anyway.

We have ascertained (pending verification of one technical detail) that removing the regularity conditions (mentioned several paragraphs above) will not adversely affect limiting behavior. That is, the scaled log probability of incorrectly inferring the mean response function to be “B” when it is actually “A” is asymptotically <u>less than or equal</u> to a negative multiple of the integrated squared distance between “B” and “A”. As above, the integral is taken with respect to a probability measure describing the limiting distribution of the design variable.

This finding is remarkable, because a scaled log probability being <u>less than or equal</u> to a negative number provides a <u>minimum guaranteed asymptotic performance</u> of the classification method. One may think it counterintuitive that removing the regularity conditions could yield a better performance, but such a phenomenon exists elsewhere in statistical inference. An example which may be familiar is the Cramer-Rao lower bound for the variance of an unbiased estimator from a simple random sample. The Cramer-Rao lower bound is on the order of the inverse sample size. However, some unbiased estimators for which regularity conditions are not met have smaller variance, on the order of the squared inverse sample size.

This finding is also useful. For given candidate mean response functions, the regularity conditions may hold for some probability measures but not for others. Thus, even if the regularity conditions do not hold for a “maximin” probability measure, the aforementioned <u>minimum guaranteed asymptotic performance</u> means that we can still safely use this measure to select values of the design variable.

We now seem to have enough theoretical content to prepare a manuscript on Subproject 2.2. A convincing illustration with real data would be desirable, but there is a sort of “catch 22”. To illustrate how to select design points optimally in our context (i.e., in accord with a “maximin” probability measure), we would need the candidate functions to be defined on a grid. Once design points were selected, we would need repeated measurements of a test function (i.e., that which is to be classified as one of the candidate functions) at the selected design points. Moreover, the numbers of repetitions at the selected design points would need to correspond to the “maximin” probability measure.

However, real data in a scientific application (like the analytic chemistry example) may only have one main format: either one measurement at each of many grid points, or repeated measurements at each of a few design points.

One way to resolve the “catch 22” might be to use data in the first format but mimic repetition by considering measurements corresponding to successive values of the design variable. To illustrate, suppose the design variable runs from 0 to 1 in increments of 0.01, and suppose we have only one measurement at each value of the design variable. Moreover, suppose that a "maximin" probability measure is found which places 50% mass at 0.25 and 50% mass at 0.75.

If we want to approximate the result of having five measurements at 0.25 and five measurements at 0.75, we can consider the measurements at 0.23, 0.24, 0.25, 0.26, 0.27, 0.73, 0.74, 0.75, 0.76, and 0.77 respectively. A comparator with the same total number of measurements can be based on measurements at 0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, and 0.95 respectively.

In the preceding paragraph, the first scheme approximates a design variable governed by the "maximin" probability measure, while the second scheme approximates a design variable governed by a uniform probability measure. A data analyst can then see whether each set of ten measurements yields a correct classification. If the former does, and if the latter does not, then there is at least a "proof of concept" for adoption of the "maximin" probability measure. If a correct classification is obtained either way, then the data analyst can still examine the strength of the inference using, say, Royall's (1997) evidentiary framework.

Another way to resolve the "catch 22" may be to begin with data in the first format (one measurement at each of many grid points) but simulate additional observations from the test function using a pseudo-random number generator. This resolution may be met with skepticism by some reviewers, but it offers three benefits.

First, because the native format of the data will entail measurement at each of many grid points, a comparator to the "maximin" probability measure (namely, the uniform probability measure) is readily available.

Second, the pseudo-random numbers could be chosen to reflect different error distributions, providing some indication of robustness (or lack thereof) to departures from normality.

Third, multiple simulations could be run to empirically estimate a probability of correct classification, which is more informative than seeing whether a correct classification is obtained with a single realized data set.

SUBPROJECT 2.3

In Subprojects 2.1 and 2.2, we wanted to identify the mean response function that generated outcome data. We assumed that this mean response function belonged to a finite set of candidate functions.

However, another modeling scenario – and the subject of Subproject 2.3 – is that the mean response function may be a convex combination of the candidate functions. In this case, we may more properly refer to the candidate functions as basis functions.

Regardless of terminology, this modeling scenario may be seen as containing the classification problem of Subprojects 2.1 and 2.2. That's because equality of the mean response function to one of the candidates can be viewed as a degenerate convex combination, in which all weights except one are equal to zero. Put differently, we can embed the classification problem into an estimation problem, as suggested to us by our former program officer, Prof. Michael Lavine.

Indeed, the estimation problem of Subproject 2.3 may have physical meaning in some scientific applications. Suppose an engineer has a collection of heterogeneous nanoparticles, perhaps including 30% of one variety and 70% of another. The engineer, not knowing *a priori* that the proportions are 30% and 70% in the collection of interest (or even that only two varieties are present), wishes to characterize the nanoparticles. The engineer may decide to conduct, e.g., a surface wave scattering experiment, as proposed by Videen and colleagues (2005) and Aslan and colleagues (2005).

The surface wave scattering experiment will yield a "scattering profile", which we may think of as a noise-contaminated realization of a mean response function for the nanoparticle collection of interest. Because the nanoparticles consist of two varieties in the proportions of 30% and 70%, the experimental scattering profile will be approximately a convex combination (with 30% and 70% weights) of the scattering profiles that would be obtained from collections of homogeneous nanoparticles (one of each variety). We will refer to those as "reference" scattering profiles.

If the engineer actually has available the reference scattering profiles (from previous laboratory work under known, controlled conditions), then the engineer can estimate the weights of the convex combination that will most closely reproduce the experimental scattering profile. As indicated above, these weights will correspond approximately to the proportions of the two varieties of nanoparticles in the collection of interest.

A while ago, we spent a little time on Subproject 2.3. However, we have not revisited Subproject 2.3 recently. So, we reiterate here some possible avenues of advancement suggested in earlier progress reports.

First, if we estimate weights of a convex combination of candidate (basis) functions, what will be the theoretical properties of the estimators, assuming that the mean response function really is such a convex combination ?

In fact, we can ask this question regarding both constrained least squares estimators of the weights and constrained maximum likelihood estimators. (Recall our efforts in Subproject 2.1 to accommodate non-normal error distributions.) Here, "constrained" refers to consequences of convexity, namely that the estimated weights should be nonnegative and sum to 100%.

Constrained least squares and constrained maximum likelihood may yield very similar results to their unconstrained analogues, if all of the actual weights are nonzero. On the other hand, the most scientifically interesting or useful applications of the work in Subproject 2.3 may be in situations where some of the actual weights are zero. Such situations entail parameter values (here, the weights of the convex combination) being on the boundary of the parameter space. In other statistical areas such as mixture modeling (e.g., Chen, Chen, and Kalbfleisch, 2001), similar phenomena have complicated inference.

Second, even though Subproject 2.3 concerns estimation, the idea of experimental design from Subproject 2.2 may have relevance here. In particular, the variances of the constrained least squares or constrained maximum likelihood estimators of the weights will depend on how values for the design variable are specified. If we specify values for the design variable according to some probability measure on the design space, then how can we choose a probability measure optimally ?

A probability measure yielding smaller variances of parameter estimators will be preferred to a probability measure yielding larger variances. However, we may not have uniform superiority of one probability measure regarding variances of all parameter estimators. Moreover, such variances may depend on the actual parameter values (again, the weights of the convex combination), which are unknown before data acquisition and analysis.

Perhaps, as in Silvey (1980), the logarithm of the determinant of the Fisher information matrix might be taken as an overall index of how well variation is controlled, and we might seek a probability measure for which this index remains tolerable over most of the parameter space.

Third, if we estimate a weight to be almost zero, under what conditions might we reasonably infer that the actual weight is zero and adjust the estimate accordingly ? In regression problems without a convexity constraint on the parameters of interest, this issue has been handled by variable selection and shrinkage methods such as the lasso (cf. Hastie, Tibshirani, and Friedman, 2009). Perhaps such strategies could be adapted to the present problem with a convexity constraint.

This last avenue of research re-connects the estimation problem with the embedded classification problem. Although the mean response function we are trying to describe may not be equal to one of the candidate (basis) functions, we may still ask which of the candidate (basis) functions contribute non-trivially to the mean response function. The probability of incorrectly asserting that a non-contributing function is relevant is loosely analogous to a misclassification probability in the sense of Subprojects 2.1 and 2.2. Likewise, we may be interested in the probability of incorrectly asserting that a contributing function is irrelevant.

In the post-grant setting, we probably won't pursue all of the avenues indicated above. However, we do hope to eventually publish a paper related to Subproject 2.3.

REFERENCES

Aslan M., Menguc, M. P., and Videen G. (2005). Characterization of metallic nano-particles via surface wave scattering: B. Physical concept and numerical experiments. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **93**: 207 – 217.

Charnigo, R., Feng, L., and Srinivasan C. (2015). Nonparametric and semiparametric compound estimation in multiple covariates. *Journal of Multivariate Analysis*, **141**: 179 – 196.

Charnigo, R., Hall, B., and Srinivasan, C. (2011). A generalized C_p criterion for derivative estimation. *Technometrics*, **53**: 238-253.

Charnigo, R. and Srinivasan, C. (2011). Self-consistent estimation of mean response functions and their derivatives. *Canadian Journal of Statistics*, **39**: 280 – 299.

Charnigo, R., Sun, J., and Muzic, R. (2006). A semi-local paradigm for wavelet denoising. *IEEE Transactions on Image Processing*, **15**: 666-677.

Chen, H., Chen, J., and Kalbfleisch, J. (2001). A modified likelihood ratio test for homogeneity in finite mixture models. *Journal of the Royal Statistical Society B*, **63**: 19 – 29.

Chu, C., Siao, J., Wang L., and Deng., W. (2012). Estimation of 2D jump location curve and 3D jump location surface in nonparametric regression. *Statistical Computing*, **22**: 17 – 31.

Dai, H. and Charnigo, R. (2008). Omnibus testing and gene filtration in microarray data analysis. *Journal of Applied Statistics*, **35**: 31 – 47.

Dai, H. and Charnigo, R. (2010). Contaminated normal modeling with application to microarray data analysis. *Canadian Journal of Statistics*, **38**: 315 – 332.

Drton, M. and Plummer, M. (2017). A Bayesian information criterion for singular models. *Journal of the Royal Statistical Society Series B*, **79**: 323 – 380.

Hall, P. and Qiu, P. (2007). Blind deconvolution and deblurring in image analysis. *Statistica Sinica*, **17**: 1483 – 1509.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edition. Springer, New York.

Hoo, J. and Qiu, P. (2009). Jump regression in a regression curve and its derivative. *Technometrics*, **51**: 289 – 305.

Jenkins, A. and Larsen, R. (2004). Gemstone identification using Raman spectroscopy. *Spectroscopy*, **19**: 20 – 25.

Lindsay, B. (1983). The geometry of mixture likelihoods: a general theory. *Annals of Statistics*, **11**: 86 – 94.

Qiu, P. (1998). Discontinuous regression surfaces fitting. *Annals of Statistics*, **26**: 2218 – 2245.

Qiu, P. and Sun, J. (2007). Local smoothing image segmentation for spotted microarray images. *Journal of the American Statistical Association*, **102**: 1129 – 1144.

Qiu, P. and Sun, J. (2009). Using conventional edge detectors and postsmoothing for segmentation of spotted microarray images. *Journal of Computational and Graphical Statistics*, **18**: 147 – 164.

Royall, R. (1997). *Statistical Evidence: A Likelihood Paradigm*. Chapman & Hall/CRC Press, Boca Raton.

Silvey, S. (1980). *An Introduction to the Theory of Parameter Estimation*. Springer, Netherlands.

Vandenabeele, P. (2013). *Practical Raman Spectroscopy: An Introduction*. John Wiley & Sons, Hoboken.

Videen, G., Aslan, M., and Menguc, M. (2005). Characterization of metallic nano-particles via surface wave scattering: A. Theoretical framework and formulation. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **93**: 195 – 206.

Xia, Z. and Qiu, P. (2015). Jump information criterion for statistical inference in estimating discontinuous curves. *Biometrika*, **102**: 397-408.