Naval Research Laboratory

Washington, DC 20375-5320

NRL/MR/6180--16-9711

Using Fisher Information Criteria for Chemical Sensor Selection via Convex Optimization Methods

ADAM C. KNAPP

National Research Council Postdoctoral Associate
Navy Technology Center for Safety and Survivability
Chemistry Division

KEVIN J. JOHNSON

Navy Technology Center for Safety and Survivability
Chemistry Division

November 16, 2016

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
16-11-2016	Memorandum Report	October 2015 – September 2016

4. TITLE AND SUBTITLE

Using Fisher Information Criteria for Chemical Sensor Selection via Convex Optimization Methods

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
0601153N

6. AUTHOR(S)

Adam C. Knapp* and Kevin J. Johnson

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
61-1E39-06-5

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5320

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/MR/6180--16-9711

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5320

10. SPONSOR / MONITOR'S ACRONYM(S)

NRL

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

*National Research Council Postdoctoral Associate

14. ABSTRACT

This report presents methodology for the near optimal selection of chemical sensors in a chemical sensing array. While the sensing criteria are task specific, generally one may consider a criterion which maximizes the signal strength or conversely minimizes global error to be best. The quantification of this criteria proceeds from the determinant of the inverse Fisher information matrix which is proportional to the global error volume. If a practitioner has a suitable probabilistic noise model for his or her chemical sensing array and pool of available sensors, the Fisher information matrix may be parametrized to select the best sensors after an optimization procedure. Due to the positive definite nature of the Fisher information matrix, convex optimization may be used to accomplish this task. This report presents the derivation of the supporting set-up, expressions, and constraints for this procedure.

15. SUBJECT TERMS

Sensor array optimization
Sensor selection
Fisher information
Fisher information matrix
Convex optimization
Elliptically contoured distributions

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Kevin J. Johnson
Unclassified Unlimited	Unclassified Unlimited	Unclassified Unlimited	Unclassified Unlimited	20	**19b. TELEPHONE NUMBER** *(include area code)* 202-404-5407

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

CONTENTS

EXECUTIVE SUMMARY E-1

1. BACKGROUND AND OVERVIEW 1

2. FISHER INFORMATION IN CHEMICAL SENSOR ARRAY DESIGN 2
   - 2.1 General Applicability of Fisher Information to Sensor Systems 2
   - 2.2 Fisher Information and the Cramér-Rao Inequality 3
   - 2.3 Derivation of a Lower Bound to the Fisher Information Matrix 5

3. CONVEX OPTIMIZATION OF THE FISHER INFORMATION MATRIX 6
   - 3.1 Background on Convex Optimization 6
   - 3.2 Elliptically Contoured Distributions: A Correlated Noise Model for Chemical Sensor Arrays 8
   - 3.3 Gradients and Hessians for the Fisher Information Matrices of Elliptically Contoured Distributions 10
   - 3.4 Defining the Mean Response Vector, ECD Scale Matrix, Slack Variables and their Constraints for Convex Optimization 12
   - 3.5 Applying Convex Optimization: Interpreting the Results 14

4. CONCLUSIONS 14

ACKNOWLEDGMENTS 15

REFERENCES 15

EXECUTIVE SUMMARY

This report presents methodology for the near optimal selection of chemical sensors in a chemical sensing array. While the sensing criteria are task specific, generally one may consider a criterion which maximizes the signal strength or conversely minimizes global error to be best. The quantification of this criteria proceeds from the determinant of the inverse Fisher information matrix which is proportional to the global error volume. If a practitioner has a suitable probabilistic noise model for his or her chemical sensing array and pool of available sensors, the Fisher information matrix may be parametrized to select the best sensors after an optimization procedure. Due to the positive definite nature of the Fisher information matrix, convex optimization may be used to accomplish this task. This report presents the derivation of the supporting set-up, expressions, and constraints for this procedure.

USING FISHER INFORMATION CRITERIA FOR CHEMICAL SENSOR SELECTION VIA CONVEX OPTIMIZATION METHODS

1. BACKGROUND AND OVERVIEW

The design of chemical sensor arrays from the standpoint of chemical sensor selection and error quantification has historically proceeded as an *ad hoc* process. Frequently, chemical sensors are developed not as general purpose sensing devices, but as analyte or chemical class specific detectors. When such single purpose devices are integrated together as a chemical sensor array, it is unclear *a priori* how well they will function in concert with each other to provide expanded capabilities, an observation that is true of the integration of analytical instruments as well [1]. Further complicating the combination and optimization of these devices is that it is semantically unclear precisely what the combined device or array ought to do. Defining what a combined sensing device ought to do is difficult and highly dependent upon the analytical task the array will be intended to support as well as the specific goals of the array designer.

In the face of an otherwise unspecified sensing task, it is reasonable to assume that the practitioner ought to attempt to minimize the global error of the array, or conversely, to maximize the signal. This is the approach is taken by the authors within this report. The question remains, however, as to how to best fulfill this objective. While a hypothetical practitioner may be able to take an exhaustive approach to sensor array design by experimentally evaluating all possible sensor combinations, this method quickly becomes infeasible as the number of sensors relative to array slots becomes coequal or large. In the rare cases when a sensor array optimization has been attempted (as opposed to using whatever sensors were immediately available), it is this aforementioned approach of combinatorial experimentation which historically has typified chemical sensor array design, and thus, severely limited the optimization of sensor arrays. Alternative approaches to array design based on neural networks and machine learning have also been tried e.g. [1–4]. However, due to their opacity, these methods fail to provide significant insight into the chemical detection problem or to suggest subsequent ways to further improve the array design. Consequently, an explicit, precise, and mathematically rigorous approach to chemical sensor array design and optimization is greatly desired.

Given its wide range of applications it is surprising that the literature centered on chemical sensor array optimization strategies is rather sparse, despite the relative frequency of reports describing specific sensor arrays and applications. A notable exception is the Fisher information matrix-based approach proposed by Pearce and Sánchez-Montaes and theoretically applied to simple linear sensor systems with uncorrelated noise [5–7]. Unfortunately, this methodology has not been greatly developed since its inaugural set of papers. In the view of this reports authors, this is most likely due to the mathematical complexities and difficulties presented by implementing this program as well as the accompanying change in mentality this forces upon the typical practitioner in the chemical sensing field.

This report further develops the use of the Fisher information matrix as a quantitative descriptor for hypothetical chemical sensor array scenarios in which a collection of co-located sensors respond to chemical mixtures resulting from a pool of possible analytes. It assumes that the underlying sensors provide additive

Manuscript approved September 22, 2016.

linear responses with respect to the system of analytes and that they may exhibit statistically correlated noise. The latter is important as correlated measurement error is realistic, yet frequently unacknowledged in the literature. The former is generally a reasonable assumption in low concentration regimes, which typify the bulk of analytical sensing applications, and present the greatest challenges regarding desired sensitivity and selectivity.

This work describes how the positive (semi)definite nature of the Fisher information matrix enables algorithmic chemical sensor array design via convex optimization techniques. This property is a rare bit of mathematical good fortune as the general case of global optimization is generally computationally intractable. The use of elliptically contoured distributions as a general-purpose means of modeling correlated sensor noise is introduced and developed for convex optimization of sensor arrays. Ultimately, this report presents a theoretical summary description of this approach to chemical sensor array design and optimization by showing how to (nearly) best select a subset of sensors for a sensor array from a much larger collection while assuming correlated noise and the specific challenges of a chemical environment. This effort was conducted in support of a basic research program seeking to develop better methodology for the design of chemical sensor arrays using techniques from information theory.

2. FISHER INFORMATION IN CHEMICAL SENSOR ARRAY DESIGN

2.1 General Applicability of Fisher Information to Sensor Systems

Both Fisher information (FI) and its generalization to multi-parameter estimation, the Fisher information matrix (FIM), are relevant to the design of statistical estimators (i.e. sensors) as their respective inverses act as lower bounds to the (co)variances of the subject estimator, a property which is referred to as the Cramér-Rao lower bound [8].

To more concretely motivate this assertion, consider a chemical sensor array response, $\boldsymbol{\mu}(\boldsymbol{\theta})+\boldsymbol{\delta}(\boldsymbol{\theta})$, where $\boldsymbol{\mu}(\boldsymbol{\theta})$and $\boldsymbol{\delta}(\boldsymbol{\theta})$are the idealized sensor response vector and noise vector respectively. $\boldsymbol{\theta}$ denotes an external parameter vector which is environmentally dependent. For chemical sensors and sensor arrays, this is typically the analyte concentration vector. Such a sensor array response may then be modeled with a probability density function, $\rho(\boldsymbol{X};\boldsymbol{\mu}(\boldsymbol{\theta}))$ [5], as follows, $\boldsymbol{\mu}(\boldsymbol{\theta}) = \int d\boldsymbol{X}\, \boldsymbol{X}\rho(\boldsymbol{X};\boldsymbol{\mu}(\boldsymbol{\theta}))$, with a covariance matrix given by,

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \int d\boldsymbol{X}\,(\boldsymbol{X}-\mu(\boldsymbol{\theta}))(\boldsymbol{X}-\mu(\boldsymbol{\theta}))^T\rho(\boldsymbol{X};\boldsymbol{\mu}(\boldsymbol{\theta})) \tag{1}$$

A typical goal of sensor array optimization is to minimize the global error of the sensor array. This quantity is captured by $\det(\boldsymbol{\Sigma}(\boldsymbol{\theta}))$, the determinant of the covariance matrix. Since $\boldsymbol{\Sigma}(\boldsymbol{\theta})$is a positive definite matrix, its determinant describes a strictly positive volume that may act as a score or metric for the global error [9]. Thus, from the standpoint of global noise, the goal of the sensor array designer is to minimize this determinant. Unfortunately, it is often either impractical or computationally intensive to directly calculate $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ in a way that allows for the analytic optimization and design of arrays, particularly if a complicated estimator is used. It is also worth considering that many different physical sensor system setups or statistical estimators may be constructed for the same system i.e. sensor response probability distribution. This multitude of specific estimator possibilities forces the practitioner to seek a design criterion that is robust in the face of many potentially similar but varying covariance matrices or array responses.

Fortunately, via the Cramér-Rao inequality, FI/FIM provide a lower bound in the positive definite sense for the covariance matrix of such a sensor array that is independent of the actual estimator being used. This provides a useful expression of the fundamental analytical potential of the device. Importantly, if the practitioner tunes or re-tunes their setup, this quantity will never change. Thus, we conclude that the FI/FIM provides a robust metric to optimize in the design of chemical sensor arrays.

2.2 Fisher Information and the Cramér-Rao Inequality

Before showing how to utilize the FI/FIM in the context Further manipulation of the integrand yields optimization for sensor selection, it is informative to first derive the FI/FIM relation to the Cramér-Rao lower bound. As a prelude, the FI is defined as [10],

$$f(\mu;\theta) = \int dx \rho(\mu;\theta|x) \left(\frac{\partial \ln(\rho)}{\partial \theta} \right)^2 \tag{2}$$

and each element of the FIM itself is defined as,

$$\boldsymbol{F}(\mu;\boldsymbol{\theta})_{ij} = \int d\boldsymbol{X} \rho(\mu;\boldsymbol{\theta}|\boldsymbol{X}) \left(\frac{\partial \ln(\rho)}{\partial \theta_i} \right) \left(\frac{\partial \ln(\rho)}{\partial \theta_j} \right) \tag{3}$$

with the FIM reducing to the FI in the univariate case. Informally, the FI/FIM may be thought of as conveying how much information an observed random variable, x, or set of random variables, $\boldsymbol{X}$, carry about a parameter(s), θ or $\boldsymbol{\theta}$.

In the event those deterministic parameters $\boldsymbol{\theta}$ are being statistically estimated, the FI/FIM provides a lower bound to their (co)variance independent of the employed statistical estimator(s). Beginning with the univariate case, the FI may be derived [11] by first considering the following expectation value,

$$\boldsymbol{E}[\hat{\theta}(x) - \theta] = \int \left(\hat{\theta}(x) - \theta \right) \rho(x;\theta) = 0 \tag{4}$$

where $\hat{\theta}(x)$ is an unbiased statistical estimator for θ. Next, differentiating by the deterministic parameter yields,

$$\frac{\partial}{\partial \theta} \int dx \left(\hat{\theta}(x) - \theta \right) \rho(x;\theta) = \int dx \left(\hat{\theta}(x) - \theta \right) \frac{\partial \rho}{\partial \theta} - \int dx \rho = 0 \tag{5}$$

Recognizing that since ρ is a probability distribution,

$$\int dx \rho(x;\theta) = 1 \tag{6}$$

and

$$\frac{\partial \rho}{\partial \theta} = \rho \frac{\partial \ln(\rho)}{\partial \theta} \tag{7}$$

which implies

$$\int dx \left(\hat{\theta}(x) - \theta\right) \rho(x;\theta) \frac{\partial \ln(\rho)}{\partial \theta} = 1 \tag{8}$$

Further manipulation of the integrand yields,

$$\int dx \left(\left(\hat{\theta}(x) - \theta\right) \sqrt{\rho} \right) \left(\frac{\partial \ln(\rho)}{\partial \theta} \sqrt{(\rho)} \right) = 1 \tag{9}$$

Applying the Cauchy-Schwartz inequality[1] to this manipulated integrand gives,

$$\begin{aligned} \left(\int dx \left(\left(\hat{\theta}(x) - \theta\right) \sqrt{\rho} \right) \left(\frac{\partial \ln(\rho)}{\partial \theta} \sqrt{(\rho)} \right) \right)^2 &= 1^2 \\ &= 1 \leq \left(\int dx \rho \left(\hat{\theta}(x) - \theta\right)^2 \right) \left(\int dx \rho \left(\frac{\partial \ln(\rho)}{\partial \theta} \right)^2 \right) \end{aligned} \tag{10}$$

After some manipulation of the preceding integrand, the expression resolves itself as,

$$\frac{1}{\left(\int dx \rho \left(\frac{\partial \ln(\rho)}{\partial \theta} \right)^2 \right)} \leq \left(\int dx \rho \left(\hat{\theta}(x) - \theta\right)^2 \right) \rightarrow \frac{1}{f(\theta)} \leq \left(\mathrm{Var}(\hat{\theta}) \right) \tag{11}$$

which is the Cramér-Rao lower bound for the univariate case.

The derivation of the Fisher information matrix (FIM) for the multivariate case [10] is performed in a similar fashion to the univariate case by first considering,

$$\boldsymbol{E}[\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta}] = \int d\boldsymbol{X} \rho(\boldsymbol{X}|\boldsymbol{\theta}) (\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta}) = \boldsymbol{0} \tag{12}$$

And then differentiating this equation so that,

$$\partial_{\boldsymbol{\theta}} \int d\boldsymbol{X} \rho(\boldsymbol{X}|\boldsymbol{\theta}) \left(\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta} \right) = \int d\boldsymbol{X} \left(\partial_{\boldsymbol{\theta}} \rho(\boldsymbol{X}|\boldsymbol{\theta}) \right) \left(\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta} \right) - \underbrace{(\partial_{\boldsymbol{\theta}} \boldsymbol{\theta})}_{\boldsymbol{I}} \underbrace{\int d\boldsymbol{X} \rho(\boldsymbol{X}|\boldsymbol{\theta})}_{1} = 0 \tag{13}$$

where $\partial_{\boldsymbol{\theta}}$ indicates a derivative with respect to the vector $\boldsymbol{\theta}$. Rearranging terms as before, it becomes,

$$\int d\boldsymbol{X} \rho \partial_{\boldsymbol{\theta}} \ln(\rho) \left(\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta} \right) = \int d\boldsymbol{X} \left((\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta}) \sqrt{\rho} \right) (\sqrt{\rho} \partial_{\boldsymbol{\theta}} \ln(\rho)) = \boldsymbol{I} \tag{14}$$

[1] $|\langle \boldsymbol{u}, \boldsymbol{v} \rangle|^2 \leq \langle \boldsymbol{u}, \boldsymbol{u} \rangle \cdot \langle \boldsymbol{v}, \boldsymbol{v} \rangle$ where $\boldsymbol{u}$ and $\boldsymbol{b}$ are vectors with the inner product $\langle \cdot, \cdot \rangle$

and applying the Cauchy-Schwartz inequality gives,

$$\boldsymbol{I} \leq \int d\boldsymbol{X}\rho \left(\left(\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta} \right) \left(\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta} \right)^T \right) \cdot \int d\boldsymbol{X}\rho \left(\left(\partial_{\boldsymbol{\theta}} \ln(\rho) \right) \left(\partial_{\boldsymbol{\theta}} \ln(\rho) \right)^T \right) \tag{15}$$

so that the FIM provides a lower bound to the covariance matrix,

$$\boldsymbol{F}(\boldsymbol{\mu};\boldsymbol{\theta})^{-1} \leq \boldsymbol{\Sigma}(\boldsymbol{\theta}) \tag{16}$$

with the $\leq$ relation is in the sense of a positive definite matrix and the FIM and covariance matrix ($\boldsymbol{F}$ and $\boldsymbol{\Sigma}$ respectively) defined as,

$$\boldsymbol{F}(\boldsymbol{\mu};\boldsymbol{\theta}) = \int d\boldsymbol{X}\rho \left(\left(\partial_{\boldsymbol{\theta}} \ln(\rho) \right) \left(\partial_{\boldsymbol{\theta}} \ln(\rho) \right)^T \right) \tag{17}$$

and

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \int d\boldsymbol{X}\rho \left(\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta} \right) \left(\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta} \right)^T \tag{18}$$

Clearly, the so-derived FIM also implies the univariate case.

2.3 Derivation of a Lower Bound to the Fisher Information Matrix

While the FI and the FIM derived in the prior section are potentially useful for optimizing a sensor array, they nonetheless require a specific noise model for the sensor array which may not be forthcoming in practice. Nonetheless, it is highly desirable for a practitioner to be able to select relevant sensors for an array in a fashion which provides some reasonable assurance of being optimal to some degree. The following lower bound to the FI and FIM provides such an assurance while satisfying the need to be a metric defined by experimentally accessible parameters. Moreover, because it in essence defines a FI/FIM for a Gaussian model, it is amenable to the convex optimization framework which will be subsequently developed.

For purposes of expediency, first consider generic vector functions [12], $\boldsymbol{h}(\boldsymbol{y})$ and $\boldsymbol{f}(\boldsymbol{x})$, in the vector expression,

$$\hat{\boldsymbol{f}}(\boldsymbol{x}) = \boldsymbol{E}_{xy}[\boldsymbol{h}(\boldsymbol{y})\boldsymbol{f}(\boldsymbol{x})^T]\boldsymbol{E}_x[\boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})^T]^{-1}\boldsymbol{f}(\boldsymbol{x}) \tag{19}$$

and the positive semidefinite matrix expectation value,

$$\boldsymbol{E}_{xy}[\left(\boldsymbol{h}(\boldsymbol{y}) - \hat{\boldsymbol{f}}(\boldsymbol{x}) \right) \left(\boldsymbol{h}(\boldsymbol{y}) - \hat{\boldsymbol{f}}(\boldsymbol{x}) \right)^T] \geq \boldsymbol{0} \tag{20}$$

After expansion and rearrangement of the prior expression, this yields the following matrix, expression assuming a joint probability distribution, $p(\boldsymbol{x},\boldsymbol{y})$,

$$\boldsymbol{E}_y[\boldsymbol{h}(\boldsymbol{y})\boldsymbol{h}(\boldsymbol{y})^T] \geq \boldsymbol{E}_{xy}[\boldsymbol{h}(\boldsymbol{y})\boldsymbol{f}(\boldsymbol{x})^T]\boldsymbol{E}_x[\boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})^T]^{-1}\boldsymbol{E}_{xy}[\boldsymbol{f}(\boldsymbol{x})\boldsymbol{h}(\boldsymbol{y})^T] \tag{21}$$

Setting $\boldsymbol{h}(\boldsymbol{y}) = \frac{\partial \ln p_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ and $\boldsymbol{f}(\boldsymbol{x}) = (\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\theta}))$ and then integrating appropriately yields the following matrix inequality,

$$\boldsymbol{F}(\mu;\boldsymbol{\theta}) \geq \left(\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \left(\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) \tag{22}$$

which provides a lower bound to the Fisher information matrix [12].

It is noteworthy that this lower bound is the Fisher information matrix for a system with parameter independent Gaussian noise. This result suggests two possible strategies for analyzing and optimizing a sensor array based on the knowledge available regarding the noise characteristics of the sensor. First, it suggests that if one only has experimentally derived sensor responses and covariances available for a sensor system one should initially optimize assuming Gaussian noise as this represents a worst case for such a system and is thus a conservative optimization strategy. Conversely, if one has knowledge of a specific noise model for a chemical sensor system that is not Gaussian with constant noise, it suggests that it would be beneficial to use that model instead to optimize the array since one should always be able to do better than a comparable Gaussian system.

3. CONVEX OPTIMIZATION OF THE FISHER INFORMATION MATRIX

3.1 Background on Convex Optimization

Recall the definition of the FIM,

$$\boldsymbol{F}(\boldsymbol{\mu};\boldsymbol{\theta}) = \int d\boldsymbol{X} \rho(\boldsymbol{\mu};\boldsymbol{\theta}|\boldsymbol{X}) \left(\frac{\partial \ln(\rho)}{\partial \boldsymbol{\theta}}\right) \left(\frac{\partial \ln(\rho)}{\partial \boldsymbol{\theta}}\right)^T \tag{23}$$

Due to its structure as matrix defined by an integral over an exterior vector product, the FIM is a positive semidefinite matrix, i.e. $\boldsymbol{a}^T \boldsymbol{F} \boldsymbol{a} \geq 0$, where $\boldsymbol{a}$ is an arbitrary real-valued vector of appropriate dimension. Positive semidefiniteness is crucial as this property allows for the so-described sensor array to be optimized with respect to sensor configuration via convex optimization techniques.

In order to properly implement this idea for sensor array optimization, specifications for of an appropriate objective function as well as a set of constraints are required. To setup this problem, first the objective function will be defined and then the relevant constraints detailed. In the process of setting up the constraints and detailing the supporting mathematical elements for the convex optimization, appropriate sensor responses and noise models for the chemical sensor array will be proposed.

Barring other priorities or specific knowledge of the analytical task, a reasonable design goal for a general purpose chemical sensor array is to minimize the global error (maximize the signal) of the chemical sensor array. A useful measure for this global error is the volume of the ellipsoid cast by the covariance matrix of the relevant estimators since this provides a reasonable metric for the global uncertainty of estimated chemical concentrations and thus for the discernability of similar chemical mixtures. The volume of this error ellipsoid is given by the following expression [9],

$$\boldsymbol{vol}(\boldsymbol{\Sigma}) = \frac{2\pi^{d/2}}{d\Gamma\left(\frac{d}{2}\right)} \det(\boldsymbol{\Sigma})^{1/2} \tag{24}$$

where $\mathbf{\Sigma}$ is the covariance matrix and d is the dimension of the volume. Minimizing this volume term, $\boldsymbol{vol}(\mathbf{\Sigma})$, ultimately requires the minimization of $\det(\mathbf{\Sigma})^{1/2}$ as all other terms for the volume expression are related to the system dimension, which is not subject to optimization.

Since the composition of convex functions are themselves convex and both the square root function of $x > 0$ and the determinant of a positive semidefinite matrix like $\mathbf{\Sigma}$ are convex functions themselves, the objective function may be further simplified to$\det(\mathbf{\Sigma})$. For reasons of subsequent numerical convenience, this objective function is composed with the natural logarithm to give $\ln(\det(\boldsymbol{X}))$ as the final objective function. It is proven below that this function is concave (convex up) for all positive semidefinite matrices, $\boldsymbol{X}$, by showing that this function satisfies concavity [13].

First consider the following,

$$\begin{aligned} g(t) &= \ln(\det(\boldsymbol{X})) \\ &= \ln(\det(\boldsymbol{Z}+t\boldsymbol{V})) \end{aligned} \tag{25}$$

where $\boldsymbol{X} = \boldsymbol{Z}+t\boldsymbol{V} > 0$. $\boldsymbol{X}$, $\boldsymbol{Z}$, and $\boldsymbol{V}$ are positive definite matrices and $t \geq 0$ is a scalar parameter. Manipulating the matrix function in question to ensure positive definite matrices yields,

$$\begin{aligned} g(t) &= \ln(\det(\boldsymbol{Z}+t\boldsymbol{V})) \\ &= \ln(\det(\boldsymbol{Z}^{1/2}(\boldsymbol{I}+t\boldsymbol{Z}^{-1/2}\boldsymbol{V}\boldsymbol{Z}^{-1/2})\boldsymbol{Z}^{1/2})) \\ &= \ln(\det(\boldsymbol{I}+t\boldsymbol{V})) + \ln(\det(\boldsymbol{Z})) \end{aligned} \tag{26}$$

so that the first and second derivatives of $g(t)$ may be taken as,

$$g\prime(t) = \sum_{i=1}^{n} \frac{\lambda_i}{1+t\lambda_i} \tag{27}$$

and

$$g\prime\prime(t) = -\sum_{i=1}^{n} \frac{\lambda_i^2}{(1+t\lambda_i)^2} \tag{28}$$

Since $\lambda_i > 0$ due to the definition of positive definite matrices, it follows that $g\prime(t) > 0$ and $g\prime\prime(t) < 0$ for $t \geq 0$. This implies that $\ln(\det(\boldsymbol{X}))$ is a convex function for positive definite $\boldsymbol{X}$ [13].

Having shown that this objective function is valid for the optimization problem, it is now important to consider what variables to actually use to optimize the $\ln(\det(\boldsymbol{X}))$ objective. Specifically, it is not the covariance matrix that is being input into the objective function, but the inverse Fisher information matrix. This substitution is justified due to the Cramér-Rao lower bound (just as the Gaussian FIM substitution would be in the case of the upper bound).

Consequently, it is the FIM of a probability distribution and not its covariance matrix which is parametrized for optimization and the objective function thus becomes,

$$\ln(\det(\boldsymbol{C}(\boldsymbol{\theta}))) \geq \ln(\det(\boldsymbol{F}^{-1}(\boldsymbol{\theta};\boldsymbol{s}))) = -\ln(\det(\boldsymbol{F}(\boldsymbol{\theta};\boldsymbol{s}))) \tag{29}$$

where $\boldsymbol{s}$ are the slack variables subject to the optimization. For a given convex optimization in addition to the inequality and equality constraints, the practitioner must supply the convex optimization routine with gradient and Hessian routines for the objective function in the slack variables as well [13].

The gradient for $-\ln(\det(\boldsymbol{F}(\boldsymbol{\theta};\boldsymbol{s})))$, the objective function, is given by,

$$-\nabla \ln(\det(\boldsymbol{F}(\boldsymbol{\theta};\boldsymbol{s}))) = -\sum_{i=1}^{\#(\boldsymbol{s})} \widehat{\boldsymbol{e}}_i \mathrm{Tr}(\boldsymbol{F}^{-1}\frac{\partial \boldsymbol{F}}{\partial s_i}) \tag{30}$$

where $\#(\boldsymbol{s})$ is the cardinality of the slack variables, $\boldsymbol{s}$, and $\hat{\boldsymbol{e}}_i$ denote the relevant vector basis set. The matrix elements for the Hessian, $\boldsymbol{h}(\boldsymbol{\theta};\boldsymbol{s})$, are defined by,

$$h_{ij}(\boldsymbol{\theta};\boldsymbol{s}) = -\frac{\partial^2}{\partial s_i \partial s_j}\ln(\det(\boldsymbol{F}(\boldsymbol{\theta};\boldsymbol{s}))) = \mathrm{Tr}(\boldsymbol{F}^{-1}\frac{\partial \boldsymbol{F}}{\partial s_i}\boldsymbol{F}^{-1}\frac{\partial \boldsymbol{F}}{\partial s_j}) - \mathrm{Tr}(\boldsymbol{F}^{-1}\frac{\partial^2 \boldsymbol{F}}{\partial s_i \partial s_j}) \tag{31}$$

To evaluate these quantities it is necessary to first choose a noise model, so that the FIM may be properly parametrized for optimization. This matter is discussed in the following section.

3.2 Elliptically Contoured Distributions: A Correlated Noise Model for Chemical Sensor Arrays

Elliptically contoured distributions (ECDs) [14] are a class of statistical model which generalize the multivariate Gaussian and includes many standard statistical models like the multivariate Students t-distribution. They are defined as follows

$$\frac{g((\boldsymbol{x}-\boldsymbol{\mu}(\boldsymbol{\theta}))^T\boldsymbol{\Sigma}(\boldsymbol{\theta})(\boldsymbol{x}-\boldsymbol{\mu}(\boldsymbol{\theta})))}{N(\boldsymbol{\theta})} \tag{32}$$

where $g(\cdot)$ is an arbitrary univariate probability distribution, $\boldsymbol{\theta}$ are the external deterministic parameters being bounded by the FIM, $\boldsymbol{\mu}(\boldsymbol{\theta})$ is the mean response function, is a positive definite scale matrix which reduces to the covariance matrix if $g(\cdot) = \exp(-(\cdot))$, and $N(\boldsymbol{\theta})$ is the normalization constant for the probability density function. These distributions are chosen to model the correlated noise of chemical sensor arrays as they can model correlation among sensor responses while remaining both analytically tractable and relatively general.

Examples of FIMs for various well-known probability distributions are given by following: The FIM for the multivariate Gaussian is given by the so-called Slepian-Bangs formula as [15],

$$F_{ij}(\boldsymbol{\theta}) = 2\left(\frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i}\right)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\left(\frac{\partial \boldsymbol{\mu}}{\partial \theta_j}\right) + \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j) \tag{33}$$

where $\mathbf{\Sigma}_i = \frac{\partial \mathbf{\Sigma}}{\partial \theta_i}$ and the FIM for the multivariate Student-t distribution [15] is

$$F_{ij}(\boldsymbol{\theta}) = 2\frac{d+M}{d+M+1}\left(\frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i}\right)\mathbf{\Sigma}^{-1}(\boldsymbol{\theta})\left(\frac{\partial \boldsymbol{\mu}}{\partial \theta_j}\right) - \frac{1}{d+M+1}\mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_i)\mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_j)$$

$$+ \frac{d+M}{d+M+1}\mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_i\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_j) \tag{34}$$

where d is the degrees of freedom, a distribution specific quantity, of the Student-t distribution and M is the rank of the scale matrix $\mathbf{\Sigma}$.

The FIM for ECDs [15] has been recently derived as a generalization of the Slepian-Bangs formula as

$$F_{ij}(\boldsymbol{\theta}) = 2\frac{\boldsymbol{E}_p[q\phi^2(q)]}{M}\left(\frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i}\right)\mathbf{\Sigma}^{-1}(\boldsymbol{\theta})\left(\frac{\partial \boldsymbol{\mu}}{\partial \theta_j}\right) + \left[\frac{\boldsymbol{E}_p[q^2\phi^2(q)]}{M(M+1)} - 1\right]\mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_i)\mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_j)$$

$$+ \frac{\boldsymbol{E}_p[q^2\phi^2(q)]}{M(M+1)}\mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_i\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_j) \tag{35}$$

where M is the scalar dimensionality or rank of the scale matrix, $\mathbf{\Sigma} \in \mathbb{R}^{M\times M}$, $\boldsymbol{E}_p[\cdot]$ denotes an expectation value with regard to a probability density,

$$p(q) = \frac{1}{\delta_{M,g}} q^{M-1} g(q) \tag{36}$$

so that,

$$\boldsymbol{E}_p[\cdot] = \frac{1}{\delta_{M,g}} \int_0^\infty dq\,(\cdot)\, q^{M-1} g(q) \tag{37}$$

where

$$\delta_{M,g} = \int_0^\infty dt\, t^{M-1} g(t) \tag{38}$$

and

$$\phi(t) = \frac{g\prime(t)}{g(t)} \tag{39}$$

Using ECDs and their corresponding FIMs as reasonable models for correlated chemical sensor arrays allows the practitioner to propose a specific model for convex optimization.

3.3 Gradients and Hessians for the Fisher Information Matrices of Elliptically Contoured Distributions

Recall from the prior section that the model dependent components of the gradient and Hessian matrix for the convex optimization of the FIM are the matrices $\frac{\partial \boldsymbol{F}}{\partial s_p}$and $\frac{\partial^2 \boldsymbol{F}}{\partial s_p \partial s_q}$. The expressions for the matrix elements of these matrices are developed in the following subsection.

First, express the ECD FIM elements as follows,

$$\boldsymbol{F}_{ECD}(i,j) = \alpha \underbrace{\frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \theta_j}}_{G} + \beta \underbrace{\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i)\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j)}_{J} + \gamma \underbrace{\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j)}_{K} \tag{40}$$

where α, β, and γ are distribution specific constants that are not dependent upon the slack variables, s_p and s_q, and are given by,

$$\alpha = \frac{2\boldsymbol{E}_p[q\phi^2(q)]}{M} \tag{41}$$

$$\beta = \frac{\boldsymbol{E}_p[q^2\phi^2(q)]}{M(M+1)} - 1 \tag{42}$$

$$\gamma = \frac{\boldsymbol{E}_p[q^2\phi^2(q)]}{M(M+1)} \tag{43}$$

where the subscript p denotes an average with respect to $p(q)$ and $\boldsymbol{\Sigma}_i = \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}$, and G, J, and K are so defined to simplify the derivation and presentation.

The derivatives of each of these sub-expressions are given as follows,

$$\frac{\partial G}{\partial s_p} = \frac{\partial^2 \boldsymbol{\mu}^T}{\partial s_p \partial \theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \theta_j} + \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \theta_j} + \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial^2 \boldsymbol{\mu}}{\partial s_p \partial \theta_j} \tag{44}$$

$$\frac{\partial J}{\partial s_p} = \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p i})\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j) + \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i)\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p j}) \tag{45}$$

$$\frac{\partial K}{\partial s_p} = \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p i}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p j}) \tag{46}$$

where $\boldsymbol{\Sigma}_{s_p} = \frac{\partial \boldsymbol{\Sigma}}{\partial s_p}$ and $\boldsymbol{\Sigma}_{s_p i} = \frac{\partial^2 \boldsymbol{\Sigma}}{\partial s_p \partial \theta_i}$.

The Hessian elements for the G and J terms are given by,

$$\frac{\partial^2 G}{\partial s_p \partial s_q} = \frac{\partial^3 \boldsymbol{\mu}^T}{\partial s_p \partial s_q \partial \theta_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_j} + \frac{\partial^2 \boldsymbol{\mu}^T}{\partial s_p \partial \theta_i} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_j} + \frac{\partial^2 \boldsymbol{\mu}^T}{\partial s_p \partial \theta_i} \boldsymbol{\Sigma}^{-1} \frac{\partial^2 \boldsymbol{\mu}}{\partial s_q \partial \theta_j} + \frac{\partial^2 \boldsymbol{\mu}^T}{\partial s_q \partial \theta_i} \boldsymbol{\Sigma}^{-1} \frac{\partial^2 \boldsymbol{\mu}}{\partial s_p \partial \theta_j}$$

$$+ \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q} \boldsymbol{\Sigma}^{-1} \frac{\partial^2 \boldsymbol{\mu}}{\partial s_p \partial \theta_j} + \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \frac{\partial^3 \boldsymbol{\mu}}{\partial s_q \partial s_p \partial \theta_j} + \frac{\partial^2 \boldsymbol{\mu}^T}{\partial s_q \partial \theta_i} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_j}$$

$$+ \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_j} + \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q s_p} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_j} + \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_j}$$

$$+ \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p} \boldsymbol{\Sigma}^{-1} \frac{\partial^2 \boldsymbol{\mu}}{\partial s_q \partial \theta_j} \tag{47}$$

$$\frac{\partial^2 J}{\partial s_p \partial s_q} = \mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_j) \mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q s_p} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_i)$$

$$+ \mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_j) \mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q i} + \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p i} + \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q s_p i})$$

$$+ \mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p i}) \mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q j})$$

$$+ \mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q i}) \mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q j})$$

$$+ \mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_i) \mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q s_p} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_j)$$

$$+ \mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_i) \mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q j} + \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p j} + \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q s_p j}) \tag{48}$$

and the Hessian element for K is setup as follows,

$$\frac{\partial^2 K}{\partial s_p \partial s_q} = \frac{\partial}{\partial s_q} \underbrace{\mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_j)}_{A_1} + \frac{\partial}{\partial s_q} \underbrace{\mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p i} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_j)}_{A_2}$$
$$+ \frac{\partial}{\partial s_q} \underbrace{\mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_j)}_{A_3} + \frac{\partial}{\partial s_q} \underbrace{\mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p j})}_{A_4} \tag{49}$$

where the derivatives of each of the sub-terms of the expression are given by the following:

$$\frac{\partial A_1}{\partial s_q} = \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_ps_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j) + \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_qi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_qj}) \tag{50}$$

$$\frac{\partial A_2}{\partial s_q} = \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_pi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_ps_qi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_pi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_pi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_qj}) \tag{51}$$

$$\frac{\partial A_3}{\partial s_q} = \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_pi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j) + \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_qs_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_qj}) \tag{52}$$

$$\frac{\partial A_4}{\partial s_q} = \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_pj} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_qi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_pj} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_pj} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_ps_qj}) \tag{53}$$

where $\boldsymbol{\Sigma}_{ab} = \frac{\partial^2\boldsymbol{\Sigma}}{\partial a\partial b}$ and $\boldsymbol{\Sigma}_{abc} = \frac{\partial^3\boldsymbol{\Sigma}}{\partial a\partial b\partial c}$.

3.4 Defining the Mean Response Vector, ECD Scale Matrix, Slack Variables and their Constraints for Convex Optimization

This leaves two system specific quantities, $\boldsymbol{\mu}(\boldsymbol{\theta})$, the sensor response vector, and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, the scale matrix, still to be defined for the convex optimization as well as the slack variables themselves for the optimization. Setting up both of these quantities to both remain true to the goal of sensor selection while defining a properly positive definite argument as required by the definition of an ECD requires careful consideration.

Among the first things to define are how many slack variables are required for this problem and thus what their uses and constraints might be. Since the goal is to select sensors to fill an array from sensor choices and because any sensor can fill any slot, nm, variables are needed so that any sensor can be put in any slot of the array.

This decision for the sensors yields the following constraints,

$$\forall i : 0 \leq \sum_{j=1}^{m} s_{ij} \leq 1 \tag{54}$$

or

$$\forall i : \| \boldsymbol{s}_i \|_2 \leq 1 \tag{55}$$

since each sensor can only be used at most once in the sensor array.

$$\forall j : \sum_{i=1}^{n} s_{ij} = 1 \tag{56}$$

because each sensor slot must be filled where the subscripts i and j denote the placement of a sensor i in an array slot j.

The slack variables themselves are also used in the context of the ECDs since sensor selection implicitly redefines the underlying noise model. Supporting this observation is the parametrized mean response vector $\boldsymbol{\mu}$ as well as the gradient and Hessian,

$$\mu_j(\boldsymbol{\theta}) = \sum_{i=1}^{n} s_{ij}\,\mu_i(\boldsymbol{\theta}) \tag{57}$$

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \sum_{j=1}^{m} \hat{\boldsymbol{e}}_j\,\mu_j(\boldsymbol{\theta}) \tag{58}$$

$$\frac{\partial \boldsymbol{\mu}}{\partial s_{ij}} = \hat{\boldsymbol{e}}_j \frac{\partial \mu_j}{\partial s_{ij}} = \hat{\boldsymbol{e}}_j\,\mu_i \tag{59}$$

$$\frac{\partial^2 \boldsymbol{\mu}}{\partial s_p \partial s_q} = 0 \tag{60}$$

The ECD specific scale matrix, $\boldsymbol{\Sigma}(\boldsymbol{\theta};\boldsymbol{s})$, as well as its gradient and Hessian terms are parametrized and constructed in analogous way to the response vector $\boldsymbol{\mu}$ as follows,

$$\boldsymbol{\Sigma}(\boldsymbol{\theta};\boldsymbol{s}) = \boldsymbol{\sigma}(\boldsymbol{\theta};\boldsymbol{s}) \otimes \boldsymbol{\sigma}(\boldsymbol{\theta};\boldsymbol{s})^T \tag{61}$$

$$\sigma_j(\boldsymbol{\theta};\boldsymbol{s}) = \sum_{i}^{n} s_{ij}\sigma_i(\boldsymbol{\theta}) \tag{62}$$

$$\boldsymbol{\sigma}(\boldsymbol{\theta};\boldsymbol{s}) = \sum_{j=1}^{m} \hat{\boldsymbol{e}}_j \sigma_j(\boldsymbol{\theta};\boldsymbol{s}) \tag{63}$$

$$\frac{\partial \boldsymbol{\sigma}}{\partial s_{ij}} = \hat{\boldsymbol{e}}_j \frac{\partial \sigma_j}{\partial s_{ij}} = \hat{\boldsymbol{e}}_j\,\sigma_i \tag{64}$$

$$\frac{\partial^2 \boldsymbol{\sigma}}{\partial s_p \partial s_q} = 0 \tag{65}$$

$$\frac{\partial \boldsymbol{\Sigma}}{\partial s_{ij}} = \frac{\partial \boldsymbol{\sigma}}{\partial s_{ij}} \otimes \boldsymbol{\sigma}(\boldsymbol{\theta};\boldsymbol{s})^T + \boldsymbol{\sigma}(\boldsymbol{\theta};\boldsymbol{s}) \otimes \frac{\partial \boldsymbol{\sigma}}{\partial s_{ij}}^T \tag{66}$$

$$\begin{aligned}\frac{\partial^2 \boldsymbol{\Sigma}}{\partial s_p \partial s_q} &= \frac{\partial^2 \boldsymbol{\sigma}}{\partial s_p \partial s_q} \otimes \boldsymbol{\sigma}(\boldsymbol{\theta};\boldsymbol{s})^T + \frac{\partial \boldsymbol{\sigma}}{\partial s_p} \otimes \frac{\partial \boldsymbol{\sigma}^T}{\partial s_q} + \frac{\partial \boldsymbol{\sigma}}{\partial s_q} \otimes \frac{\partial \boldsymbol{\sigma}}{\partial s_p}^T + \boldsymbol{\sigma}(\boldsymbol{\theta};\boldsymbol{s}) \otimes \frac{\partial^2 \boldsymbol{\sigma}^T}{\partial s_p \partial s_q} \\ &= \frac{\partial \boldsymbol{\sigma}}{\partial s_p} \otimes \frac{\partial \boldsymbol{\sigma}^T}{\partial s_q} + \frac{\partial \boldsymbol{\sigma}}{\partial s_q} \otimes \frac{\partial \boldsymbol{\sigma}}{\partial s_p}^T \end{aligned} \tag{67}$$

for the scale matrix $\boldsymbol{\Sigma}$, where $\otimes$ denotes the outer product.

3.5 Applying Convex Optimization: Interpreting the Results

Before running the convex optimization, a test point for the optimization needs to be found which satisfies all of the constraint values. This point implicitly defines the subset region of the positive definite matrices available for the optimization. This is necessary as there is an inherent degeneracy in assigning sensors to slots since in an optimized setting one could permute sensors with slots and have the same answer. However, in order continuously reach another region where the sensors and slots are permuted, one would have to travel through an area where the FIM becomes singular by continuity. Consequently each point exists in a subregion defined by these singular bounds, which incidentally have an infinite determinant since the objective function is the determinant of the inverse FIM. Since all of these regions are identical by relabeling, optimizing in one subregion is as good as optimizing in any of the others. Due to the now defined subregion being described by positive definite matrices this remains a problem in convex optimization. Thus, by selecting a specific starting point one breaks the permutation symmetry of this problem while still allowing for the usage of convex optimization in sensor selection.

On a more practical note, a unique starting point which satisfies this problems constraints may be defined as follows:

> Consider the sensor collection ordered in a list. Take the first m sensors and assign each one to a unique array slot j. Set the corresponding slack variables for this sensor-slot selection equal, s_{ij}, to 1. Set all other slack variables equal to 0.

This defines a unique starting point for the optimization which obeys the system constraints.

After the appropriate convex optimization has been performed, a vector corresponding to numeric values for the slack variable vector $\boldsymbol{s}$ will be output. Assuming that the optimization has preceded correctly, the vector should have m slack variables close to 1 (Otherwise sort the slack variable vector in numeric order and choose the top m). Round these up to 1 and all other variables down to 0. This resulting vector tells the practitioner which sensors have been selected for which slots by whether or not a slack variable is 1 or 0. If it is 1 then the corresponding sensor and slot has been selected; otherwise it has not. This completes the sensor selection process.

4. CONCLUSIONS

This memo report has considered sensor selection for a nonspecific chemical array under the influence of correlated noise. It has used global error minimization or conversely signal maximization as a criteria for optimization by considering the determinant of the covariance matrix as idealized by the Fisher information

matrix as a scalar criterion for this optimization. Using these definitions as well as the mathematical properties of this underlying matrix, it has been able to set up this optimization problem in the context of convex optimization and has developed this scenario along with the supporting mathematics for this methodology.

It has presented two distinct approaches to this optimization problem. First it has considered and presented the optimization methodology for a family of solved non-trivial correlated noise models, the elliptically contoured distributions, which include many standard distributions such as the multivariate Gaussian and Student-t distributions. It has also taken a more practical approach and considered a lower bound to the Fisher information matrix which would allow a working practitioner to select sensors with only knowledge of a correlation matrix and the sensor response using the same framework and methodology. While the later of these two approaches would not necessarily be optimal it could be significantly better than what a practitioner might develop through trial and error.

ACKNOWLEDGMENTS

The authors would like to thank the Office of Naval Research for funding. Adam C. Knapp would like to gratefully acknowledge support through the NRC Research Associate Program.

REFERENCES

1. K. J. Johnson and S. L. Rose-Pehrsson, "Sensor array design for complex sensing tasks*," *Annual Review of Analytical Chemistry* **8**, 287–310 (2015).
2. K. E. Kramer, S. L. Rose-Pehrsson, K. J. Johnson, and C. P. Minor, "Hybrid arrays for chemical sensing," in *Computational Methods for Sensor Material Selection*, pp. 265–298 (Springer, 2009).
3. M. Pardo and G. Sberveglieri, "Classification of electronic nose data with support vector machines," *Sensors and Actuators B: Chemical* **107**(2), 730–737 (2005).
4. R. E. Shaffer, S. L. Rose-Pehrsson, and R. A. McGill, "A comparison study of chemical sensor array pattern recognition algorithms," *Analytica Chimica Acta* **384**(3), 305–317 (1999).
5. T. C. Pearce and M. Sanchez-Montanes, "Chemical sensor array optimization: geometric and information theoretic approaches," *Handbook of Artificial Olfaction Machines* pp. 347–376 (2003).
6. M. A. Sánchez-Montañés and T. C. Pearce, "Fisher information and optimal odor sensors," *Neurocomputing* **38**, 335–341 (2001).
7. M. A. Sanchez-Montanes, J. W. Gardner, and T. C. Pearce, "Spatio-temporal information in an artificial olfactory mucosa," in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 464, pp. 1057–1077 (The Royal Society, 2008).
8. T. M. Cover and J. A. Thomas, *Elements of information theory* (John Wiley & Sons, 2012).
9. S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Transactions on Signal Processing* **57**(2), 451–462 (2009).
10. C. Arndt, *Information measures: information and its description in science and engineering* (Springer Science & Business Media, 2012).

11. B. R. Frieden, *Science from Fisher information: a unification* (Cambridge University Press, 2004).

12. M. Stein, A. Mezghani, and J. A. Nossek, "A lower bound for the fisher information measure," *IEEE Signal Processing Letters* **21**(7), 796–799 (2014).

13. S. Boyd and L. Vandenberghe, *Convex optimization* (Cambridge university press, 2004).

14. F. Kai-Tai and Z. Yao-Ting, *Generalized multivariate analysis* (Science Press, 1990).

15. O. Besson and Y. I. Abramovich, "On the fisher information matrix for multivariate elliptically contoured distributions," *IEEE Signal Processing Letters* **20**(11), 1130–1133 (2013).