FUSION INSPIRED CHANNEL DESIGN

Yuan Wang, Haonan Wang, and Louis L. Scharf, Life Fellow, IEEE

Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA
{wangy, wanghn}@stat.colostate.edu, scharf@engr.colostate.edu

ABSTRACT

This paper is motivated by the problem of integrating multiple sources of measurements. We consider two multiple-input-multiple-output (MIMO) channels, a primary channel and a secondary channel, with dependent input signals. The primary channel carries the signal of interest, and the secondary channel carries a signal that shares a joint distribution with the primary signal. The problem of particular interest is designing the secondary channel matrix, when the primary channel matrix is fixed. We formulate the problem as an optimization problem, in which the optimal secondary channel matrix maximizes an information-based criterion. An analytical solution is provided in a special case. Then an intrinsic search algorithm is proposed to approximate the optimal solutions in general cases. In particular, the intrinsic algorithm exploits the geometry of the unit sphere, a manifold embedded in Euclidean space.

Index Terms— Embedded submanifold, information fusion, MIMO channel design, mutual information, two-channel system.

1. INTRODUCTION

Consider the following two-channel system,

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{F}\boldsymbol{\theta} + \boldsymbol{u} \\ \boldsymbol{y} &= \boldsymbol{G}\boldsymbol{\phi} + \boldsymbol{v}. \end{aligned} \tag{1}$$

The first channel is the primary channel that carries the signal of interest $\boldsymbol{\theta}$. The secondary channel carries a signal $\boldsymbol{\phi}$ that shares a joint distribution with $\boldsymbol{\theta}$. The measurements $\boldsymbol{x}$ and $\boldsymbol{y}$ are linear transformations of the input signals with measurement noises $\boldsymbol{u}$ and $\boldsymbol{v}$, respectively. For example, the elements of the signal of interest $\boldsymbol{\theta}$ might be the complex scattering coefficients of several radar-scattering targets and the elements of the secondary signal $\boldsymbol{\phi}$ might be intensities in an optical map of these same optical-scattering targets. The measurement $\boldsymbol{x}$ is then a range-doppler map and the measurement $\boldsymbol{y}$ is an optical image. We assume a known signal model, i.e., the joint distribution of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$. When the signals $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are correlated, the measurements $\boldsymbol{x}$ and $\boldsymbol{y}$ both contain information about $\boldsymbol{\theta}$ and we can integrate them to estimate $\boldsymbol{\theta}$. The fused estimate is expected to perform better than the estimate from a single source of measurements. In this paper, our objective is to design the channel matrix $\boldsymbol{G}$, with the primary channel fixed, such that the fused estimate achieves the best performance.

For a one-channel system $\boldsymbol{x} = \boldsymbol{F}\boldsymbol{\theta} + \boldsymbol{u}$, designing the channel matrix $\boldsymbol{F}$ exhibits parallels to the linear precoding design problem for multiple-input-multiple-output (MIMO) communications systems by considering $\boldsymbol{F}$ as the precoder into an identity channel matrix. The linear precoding design for MIMO channels has been studied in the literature [1]-[9]. The optimal precoding is designed under various criteria, for example, signal-to-noise ratio (SNR) and signal-to-interference-noise ratio (SINR), [1],[2]. Another criterion that has drawn more attention recently is the mutual information between input signal and output, [3],[4],[6]. This information-based criterion is connected with estimation theory in a Gaussian channel with arbitrary input distribution by linking the mutual information with the minimum mean square error [10],[11]. In [3], an optimal precoding matrix for the MIMO Gaussian channel with arbitrary inputs is expressed as the solution of a fixed point equation. When the input signal is Gaussian distributed, the one-channel design problem can be solved as a singular value decomposition (SVD) problem. The optimal channel matrix has an SVD with the singular vectors allocated to create non-interfering subchannels and the singular values chosen to solve a water-filling problem, [4],[6],[16].

If both $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are of interest, the two-channel system is in fact a one-channel system with a block-diagonal channel matrix. Our two-channel model differs from the previously studied models because of the presence of the "nuisance" signal $\boldsymbol{\phi}$. Moreover, our two-channel system design problem is fundamentally more difficult than the one-channel system design. In this paper, we fix the primary channel and design the secondary channel matrix $\boldsymbol{G}$ that maximizes the information gain by adding the secondary channel, subject to the total power constraint $\mathrm{tr}(\boldsymbol{G}\boldsymbol{G}^T) \leq 1$. We call this a one-channel design problem in a two-channel system. In general, this is not a convex problem. Moreover, this problem cannot be formulated as an SVD problem, in contrast to the one-channel

This work was supported in part by AFOSR contract FA 9550-10-1-0241 and NSF grant CFF-1018472.

system design. Here, we propose an intrinsic gradient algorithm that exploits the geometry of the total power constraint. The unit sphere, an embedded submanifold of $\mathbb{R}^n$, has a well-studied geometry [13],[14].

The rest of the paper is organized as follows. We formulate the design problem in Section 2 and point out the challenges for the design in the two-channel system. An optimal design for the multiple-input-single-output (MISO) model is given in Section 3. In Section 4, we propose the intrinsic gradient search which exploits the geometry of the unit sphere to approximate the optimal channel matrix. Section 5 is a numerical example, and Section 6 concludes the paper.

Notation: The set of length m real vectors is denoted by $\mathbb{R}^m$ and the set of $m \times n$ real matrices is denoted $\mathbb{R}^{m\times n}$. Bold upper case letters denote matrices, boldface lower case letters denote column vectors, and italics denote scalars. The scalar x_i denotes the ith element of vector $\boldsymbol{x}$, and $\boldsymbol{X}_{i,j}$ denotes the element of $\boldsymbol{X}$ at row i and column j. The $n \times n$ identity matrix is denoted by $\boldsymbol{I}_n$. The transpose, inverse, trace and determinant of a matrix are denoted by $(\cdot)^T$, $(\cdot)^{-1}$, $\mathrm{tr}(\cdot)$ and $\det(\cdot)$, respectively.

A covariance matrix is denoted by bold upper case $\boldsymbol{Q}$ with specified subscripts: $\boldsymbol{Q}_{\boldsymbol{zz}}$ denotes the covariance matrix of a random vector $\boldsymbol{z}$; $\boldsymbol{Q}_{\boldsymbol{z}_1\boldsymbol{z}_2}$ is the cross-covariance matrix between $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$; $\boldsymbol{Q}_{\boldsymbol{z}_1\boldsymbol{z}_1|\boldsymbol{z}_2}$ is the conditional covariance matrix of $\boldsymbol{z}_1$ given $\boldsymbol{z}_2$.

2. OVERVIEW

2.1. Problem Formulation

In the two-channel system (1), the signal of interest is $\boldsymbol{\theta}$. When $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are correlated, both channels carry information about $\boldsymbol{\theta}$. Thus, in general, fusing measurements $\boldsymbol{x}$ and $\boldsymbol{y}$ will yield a better estimate of $\boldsymbol{\theta}$ than using measurement $\boldsymbol{x}$ only. The *information gain* by including measurement $\boldsymbol{y}$ is $I(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta}) - I(\boldsymbol{x};\boldsymbol{\theta})$, where $I(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})$ is the information about $\boldsymbol{\theta}$ carried by both channels and $I(\boldsymbol{x};\boldsymbol{\theta})$ is the information about $\boldsymbol{\theta}$ carried by the first channel.

In this paper, we consider maximization of the information gain over the second channel matrix $\boldsymbol{G}$, while the first channel matrix $\boldsymbol{F}$ is fixed. Thus the information gain is a function of $\boldsymbol{G}$. Denote the function by $D(\boldsymbol{G})$. Under the Gaussian distribution assumption, the information gain is

$$D(\boldsymbol{G}) = \frac{1}{2}\log\det \boldsymbol{Q}_{\boldsymbol{\theta\theta}|\boldsymbol{x}} - \frac{1}{2}\log\det \boldsymbol{Q}_{\boldsymbol{\theta\theta}|\boldsymbol{x},\boldsymbol{y}} \qquad (2)$$

where $\boldsymbol{Q}_{\boldsymbol{\theta\theta}|\boldsymbol{x}}$ and $\boldsymbol{Q}_{\boldsymbol{\theta\theta}|\boldsymbol{x},\boldsymbol{y}}$ are the conditional covariance matrices for $\boldsymbol{\theta}$ given $\boldsymbol{x}$ and $\boldsymbol{x},\boldsymbol{y}$, respectively. The function $D(\boldsymbol{G})$ is bounded and nonnegative. In fact, one can show that $D(\boldsymbol{G}) \le I(\boldsymbol{x},\boldsymbol{\phi};\boldsymbol{\theta}) - I(\boldsymbol{x};\boldsymbol{\theta})$, which means the additional information gain the noisy measurement $\boldsymbol{y}$ can bring is no greater than that brought by $\boldsymbol{\phi}$. We further notice that, for any $\boldsymbol{G}$, $D(\lambda\boldsymbol{G})$ is monotone increasing for $\lambda \ge 0$. Therefore, without any constraint, maximization of the information gain will lead to a trivial solution by letting the norm of $\boldsymbol{G}$ go to infinity. Here we maximize the information gain subject to a total power constraint; that is, $\mathrm{tr}(\boldsymbol{G}\boldsymbol{G}^T) \le 1$, or equivalently $\|\boldsymbol{G}\| \le 1$ where $\|\cdot\|$ is Frobenius norm.

The design problem we consider here is a one-channel design problem in a two-channel system. In a one-channel system, the optimal channel matrix that maximizes the mutual information between input and output under the total power constraint can be expressed explicitly. In fact, the optimal channel matrix has a singular value decomposition (SVD) in which its singular values are solutions of a water-filling problem [16] and its singular vectors are allocated to create non-interacting channels, see [4],[6],[15]. In our two-channel system, searching for the optimal secondary channel matrix is more complicated. In general, the optimization problem cannot be reformulated as a standard SVD problem. The difficulty arises due to the non-degenerate joint distribution of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$. In fact, when the conditional covariance matrix $\boldsymbol{Q}_{\boldsymbol{\phi\phi}|\boldsymbol{\theta}}$ is zero, i.e., the value of $\boldsymbol{\phi}$ is fixed given $\boldsymbol{\theta}$, the optimal channel matrix $\boldsymbol{G}$ can be solved from an SVD problem, as in the one-channel system.

As a summary, the problem of interest is:

$$\text{(I)} \qquad \underset{\boldsymbol{G}\in\mathbb{R}^{t\times q}}{\mathrm{maximize}} D(\boldsymbol{G}) \quad s.t. \quad \mathrm{tr}(\boldsymbol{G}\boldsymbol{G}^T) \le 1.$$

2.2. An Insightful Discussion of the Information Gain

To motivate our discussion, we decompose the secondary channel as follows:

$$\begin{aligned} \boldsymbol{y} = &(\boldsymbol{GM}\mathbb{E}[\boldsymbol{\theta}|\boldsymbol{x}]) + (\boldsymbol{GM}(\boldsymbol{\theta} - \mathbb{E}[\boldsymbol{\theta}|\boldsymbol{x}])) \\ &+ (\boldsymbol{G}(\boldsymbol{\phi} - \mathbb{E}[\boldsymbol{\phi}|\boldsymbol{\theta}]) + \boldsymbol{v}), \end{aligned}$$

where $\boldsymbol{M} = \boldsymbol{Q}_{\boldsymbol{\phi\theta}}\boldsymbol{Q}_{\boldsymbol{\theta\theta}}^{-1}$. The secondary channel $\boldsymbol{y}$ is decomposed into three independent components, as illustrated in Fig. 1. The first component $\boldsymbol{GM}\mathbb{E}[\boldsymbol{\theta}|\boldsymbol{x}]$ is completely determined by the first channel $\boldsymbol{x}$ and does not contribute to the information gain brought by $\boldsymbol{y}$. The second component $\boldsymbol{GM}(\boldsymbol{\theta} - \mathbb{E}[\boldsymbol{\theta}|\boldsymbol{x}])$, denoted by $\boldsymbol{\omega}$, is independent of $\boldsymbol{x}$ and it carries the extra information in channel $\boldsymbol{y}$ about $\boldsymbol{\theta}$. The third component $\boldsymbol{G}(\boldsymbol{\phi} - \mathbb{E}[\boldsymbol{\phi}|\boldsymbol{\theta}]) + \boldsymbol{v}$, denoted by $\boldsymbol{\nu}$, is independent of both $\boldsymbol{x}$ and $\boldsymbol{\theta}$, and it can be viewed as noise.

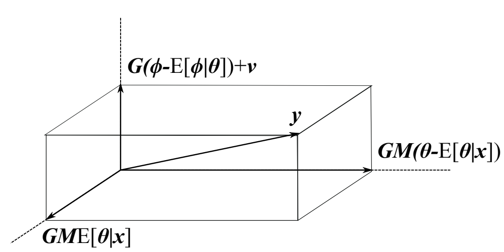

Fig. 1. Decomposition of the secondary channel $\boldsymbol{y}$.

With the covariance matrices of $\boldsymbol{\omega}$ and $\boldsymbol{\nu}$, the information gain $D(\boldsymbol{G})$ can be re-written as

$$D(\boldsymbol{G}) = \frac{1}{2}\log\det[\boldsymbol{I} + \boldsymbol{Q}_{\boldsymbol{\nu}\boldsymbol{\nu}}^{-1/2}\boldsymbol{Q}_{\boldsymbol{\omega}\boldsymbol{\omega}}\boldsymbol{Q}_{\boldsymbol{\nu}\boldsymbol{\nu}}^{-1/2}]. \tag{3}$$

By viewing $\boldsymbol{\omega}$ as a signal and $\boldsymbol{\nu}$ as a noise, $\boldsymbol{Q}_{\boldsymbol{\nu}\boldsymbol{\nu}}^{-1/2}\boldsymbol{Q}_{\boldsymbol{\omega}\boldsymbol{\omega}}\boldsymbol{Q}_{\boldsymbol{\nu}\boldsymbol{\nu}}^{-1/2}$ is a generalized signal-to-noise ratio matrix. Maximizing (3) balances the tradeoff between the noise covariance and the signal covariance. As illustrated in Fig. 1, a good channel design will favor a long parallelepiped with short height. For a MISO channel, maximizing the information gain is simplified to maximizing a scalar signal-to-noise ratio, as will be shown in Section 3. For the multiple-output channel, this is fundamentally more difficult than maximizing the signal-to-noise ratio in the single output case. The difficulty arises because the channel matrix $\boldsymbol{G}$ shapes both $\boldsymbol{Q}_{\boldsymbol{\omega}\boldsymbol{\omega}}$ and $\boldsymbol{Q}_{\boldsymbol{\nu}\boldsymbol{\nu}}$.

3. SINGLE-OUTPUT SOLUTION

Suppose that the second channel has a single output. Then $\boldsymbol{Q}_{\boldsymbol{vv}} = \sigma_v^2 \in \mathbb{R}^1$. The channel matrix $\boldsymbol{G}$ is a row vector, and we denote $\boldsymbol{G} = \boldsymbol{g}^T$ for some vector $\boldsymbol{g} \in \mathbb{R}^q$. The information gain is

$$D(\boldsymbol{g}) = \frac{1}{2}\log\left(1 + \frac{\boldsymbol{g}^T\boldsymbol{M}\boldsymbol{Q}_{\boldsymbol{\theta}\boldsymbol{\theta}|\boldsymbol{x}}\boldsymbol{M}^T\boldsymbol{g}}{\boldsymbol{g}^T\boldsymbol{Q}_{\boldsymbol{\phi}\boldsymbol{\phi}|\boldsymbol{\theta}}\boldsymbol{g} + \sigma_v^2}\right). \tag{4}$$

Notice that $\log(1 + x)$ is strictly increasing for $x \in (0, \infty)$. Therefore, solving problem (I) is equivalent to maximizing the generalized Rayleigh quotient

$$\frac{\boldsymbol{g}^T\boldsymbol{M}\boldsymbol{Q}_{\boldsymbol{\theta}\boldsymbol{\theta}|\boldsymbol{x}}\boldsymbol{M}^T\boldsymbol{g}}{\boldsymbol{g}^T\boldsymbol{Q}_{\boldsymbol{\phi}\boldsymbol{\phi}|\boldsymbol{\theta}}\boldsymbol{g} + \sigma_v^2} \tag{5}$$

subject to $\|\boldsymbol{g}\| \le 1$. Note that (5) is exactly $\sigma_\omega^2/\sigma_\nu^2$, with ω and ν the signal and noise defined in Section 2.2. This means a good channel matrix will maximize the ratio of signal power σ_ω^2 to noise power σ_ν^2. It is easy to see that the maximum is attained when $\|\boldsymbol{g}\| = 1$. Therefore write σ_v^2 as $\sigma_v^2 = \sigma_v^2\boldsymbol{g}^T\boldsymbol{g}$. Let $\widetilde{\boldsymbol{g}} = (\sigma_v^2\boldsymbol{I}_q + \boldsymbol{Q}_{\boldsymbol{\phi}\boldsymbol{\phi}|\boldsymbol{\theta}})^{1/2}\boldsymbol{g}$, we rewrite the Rayleigh quotient as

$$\frac{\boldsymbol{g}^T\boldsymbol{M}\boldsymbol{Q}_{\boldsymbol{\theta}\boldsymbol{\theta}|\boldsymbol{x}}\boldsymbol{M}^T\boldsymbol{g}}{\boldsymbol{g}^T\boldsymbol{Q}_{\boldsymbol{\phi}\boldsymbol{\phi}|\boldsymbol{\theta}}\boldsymbol{g} + \sigma_v^2} = \frac{\widetilde{\boldsymbol{g}}^T\boldsymbol{A}\widetilde{\boldsymbol{g}}}{\widetilde{\boldsymbol{g}}^T\widetilde{\boldsymbol{g}}} \le \lambda_{\max}(\boldsymbol{A})$$

where $\boldsymbol{A} = (\sigma_v^2\boldsymbol{I}_q + \boldsymbol{Q}_{\boldsymbol{\phi}\boldsymbol{\phi}|\boldsymbol{\theta}})^{-1/2}\boldsymbol{M}\boldsymbol{Q}_{\boldsymbol{\theta}\boldsymbol{\theta}|\boldsymbol{x}}\boldsymbol{M}^T(\sigma_v^2\boldsymbol{I}_q + \boldsymbol{Q}_{\boldsymbol{\phi}\boldsymbol{\phi}|\boldsymbol{\theta}})^{-1/2}$. The maximum is attained at $\boldsymbol{g}^* = \alpha(\sigma_v^2\boldsymbol{I}_q + \boldsymbol{Q}_{\boldsymbol{\phi}\boldsymbol{\phi}|\boldsymbol{\theta}})^{-1/2}\boldsymbol{v}_{\max}(\boldsymbol{A})$, with α a scalar such that $\|\boldsymbol{g}^*\| = 1$. The corresponding maximum information gain is

$$D(\boldsymbol{g}^*) = \frac{1}{2}\log\left(1 + \lambda_{\max}(\boldsymbol{A})\right).$$

4. GEOMETRICALLY INSPIRED ALGORITHM

Now we consider the general problem (I). It can be seen that the maximum information gain is obtained at the boundary $\mathrm{tr}(\boldsymbol{G}\boldsymbol{G}^T) = 1$. In fact, for any $\boldsymbol{G}$ such that $\mathrm{tr}(\boldsymbol{G}\boldsymbol{G}^T) < 1$, $\widetilde{\boldsymbol{G}} = \frac{1}{\|\boldsymbol{G}\|}\boldsymbol{G}$ yields a larger information gain than $\boldsymbol{G}$, which is a direct consequence of the fact that $D(\lambda\boldsymbol{G})$ is monotone increasing for $\lambda \ge 0$. In this section, we propose an intrinsic gradient search algorithm to approximate the maximum of $D(\boldsymbol{G})$. We consider $\boldsymbol{G}$ as a point on the unit sphere S^{tq-1}, which is a submanifold of $\mathbb{R}^{tq}$, and the intrinsic gradient is computed by taking the geometry of the manifold S^{tq-1} into consideration.

Let $\boldsymbol{g}$ be the vectorization of matrix $\boldsymbol{G}$, denoted by $\boldsymbol{g} = \mathrm{vec}(\boldsymbol{G})$, i.e.,

$$\boldsymbol{g} = [\boldsymbol{G}_{1,1}, \ldots, \boldsymbol{G}_{t,1}, \boldsymbol{G}_{1,2}, \ldots, \boldsymbol{G}_{t,2}, \ldots, \boldsymbol{G}_{1,q}, \ldots, \boldsymbol{G}_{t,q}]^T.$$

In fact, the vectorization operation is a one-to-one and onto mapping from $\mathbb{R}^{t\times q}$ to $\mathbb{R}^{tq}$; that is, for any $\boldsymbol{g} \in \mathbb{R}^{tq}$, there exists a unique matrix $\boldsymbol{G} \in \mathbb{R}^{t\times q}$ such that $\mathrm{vec}(\boldsymbol{G}) = \boldsymbol{g}$. Therefore without ambiguity we may rewrite $D(\boldsymbol{G})$ as $D(\boldsymbol{g})$ which is a mapping from the vector space $\mathbb{R}^{tq}$ to $\mathbb{R}$. The gradient of D in $\mathbb{R}^{tq}$ is

$$\nabla_{\boldsymbol{g}}D = \mathrm{vec}(\nabla_{\boldsymbol{G}}D) \tag{6}$$

where $\nabla_{\boldsymbol{G}}D$ is the gradient of the function D w.r.t. $\boldsymbol{G}$.

Recall that the solution of the constrained optimization problem (I) satisfies $\mathrm{tr}(\boldsymbol{G}\boldsymbol{G}^T) = 1$. Thus, we only need to consider the unit sphere $S^{tq-1} = \{\boldsymbol{g} \in \mathbb{R}^{tq} : \sum_{i=1}^{tq} g_i^2 = 1\}$. Let D_s be the restriction of D to the unit sphere S^{tq-1}. That is, for any $\boldsymbol{g} \in S^{tq-1}$, $D_S(\boldsymbol{g}) = D(\boldsymbol{g})$. The optimization problem (I) can be expressed as

$$\text{maximize}\, D_S(\boldsymbol{g}), \quad s.t. \quad \boldsymbol{g} \in S^{tq-1}.$$

In general, to solve an optimization problem on a manifold, an intrinsic gradient search should enforce the manifold geometry. In our application, the geometry of S^{tq-1} is enforced by the unit norm constraint. For the vector space $\mathbb{R}^{tq}$, its tangent space $T_{\boldsymbol{x}}\mathbb{R}^{tq}$ is a linear space of dimension tq spanned by $\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_{tq}}$. For any $\boldsymbol{\xi}_{\boldsymbol{x}} \in T_{\boldsymbol{x}}\mathbb{R}^{tq}$, there exists a unique vector $\boldsymbol{\xi} \in \mathbb{R}^{tq}$ such that $\boldsymbol{\xi}_{\boldsymbol{x}} = \sum_{i=1}^{tq}\xi_i\frac{\partial}{\partial x_i}$. Therefore, we can identify $T_{\boldsymbol{x}}\mathbb{R}^{tq}$ by $\mathbb{R}^{tq}$. The unit sphere S^{tq-1} is an embedded submanifold of $\mathbb{R}^{tq}$, and its tangent space $T_{\boldsymbol{g}}S^{tq-1}$ is a subspace of $T_{\boldsymbol{g}}\mathbb{R}^{tq}$. In particular, for any $\boldsymbol{\xi}_{\boldsymbol{g}} \in T_{\boldsymbol{g}}S^{tq-1}$, there exists a unique vector $\boldsymbol{\xi} \in \mathbb{R}^{tq}$ such that $\boldsymbol{g}^T\boldsymbol{\xi} = 0$ and $\boldsymbol{\xi}_{\boldsymbol{g}} = \sum_{i=1}^{tq}\xi_i\frac{\partial}{\partial g_i}$. Therefore, the tangent space $T_{\boldsymbol{g}}S^{tq-1}$ can be identified by

$$T_{\boldsymbol{g}}S^{tq-1} = \{\boldsymbol{\xi} \in \mathbb{R}^{tq} : \boldsymbol{g}^T\boldsymbol{\xi} = 0\}.$$

The orthogonal projection of any $\boldsymbol{h} \in \mathbb{R}^{tq}$ onto the tangent space is

$$P_{T_{\boldsymbol{g}}S^{tq-1}}\boldsymbol{h} = (\boldsymbol{I}_{tq} - \boldsymbol{g}\boldsymbol{g}^T)\boldsymbol{h}.$$

Since S^{tq-1} is a submanifold embedded in $\mathbb{R}^{tq}$, the gradient of D_S on S^{tq-1} is the projection of the Euclidean gradient $\nabla_{\boldsymbol{g}} D$ to the tangent space $T_{\boldsymbol{g}} S^{tq-1}$, i.e.,

$$\boldsymbol{\eta_g} \triangleq \operatorname{grad}(D_S)_{\boldsymbol{g}} = P_{T_{\boldsymbol{g}} S^{tq-1}}(\nabla_{\boldsymbol{g}} D)$$

where $\nabla_{\boldsymbol{g}} D$ is given in (6). Note that here the inner product on the tangent space $T_{\boldsymbol{g}} S^{tq-1}$ is $\langle \boldsymbol{\xi}_1, \boldsymbol{\xi}_2 \rangle_{\boldsymbol{g}} = \boldsymbol{\xi}_1^T \boldsymbol{\xi}_2$. A graphical illustration of the relationship between $\nabla_{\boldsymbol{g}} D$ and $\boldsymbol{\eta_g}$ is given in Fig. 2.

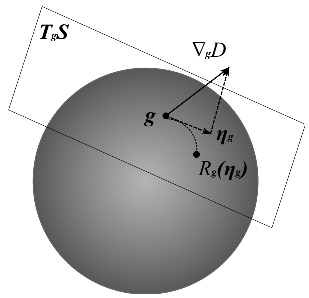

Fig. 2. Projection of the Euclidean gradient to the tangent plane of unit sphere.

Next, we consider a mapping from the tangent space $T_{\boldsymbol{g}} S^{tq-1}$ to the manifold S^{tq-1}, which is defined as

$$R_{\boldsymbol{g}}(\boldsymbol{\xi}) = \boldsymbol{g} \cos(\|\boldsymbol{\xi}\|) + \frac{\boldsymbol{\xi}}{\|\boldsymbol{\xi}\|} \sin(\|\boldsymbol{\xi}\|) \tag{7}$$

for any tangent vector $\boldsymbol{\xi} \in T_{\boldsymbol{g}} S^{tq-1}$. For $t \geq 0$, $\gamma(t) := R_{\boldsymbol{g}}(t\boldsymbol{\xi})$ is a curve on the manifold S^{tq-1} starting from $\boldsymbol{g}$. This curve generalizes the notion of moving on the manifold S^{tq-1} along the direction $\boldsymbol{\xi}$. In fact, $R_{\boldsymbol{g}}$ given in (7) is the exponential map in Riemannian geometry and the resulting curve $\gamma(t)$ generalizes the idea of straight line in Euclidean space. When maximizing the information gain $D_S(\boldsymbol{g})$ over the manifold S^{tq-1}, the intrinsic gradient search algorithm updates $\boldsymbol{g}_{k+1}$ as

$$\boldsymbol{g}_{k+1} = R_{\boldsymbol{g}_k}(\tau_k \boldsymbol{\eta}_{\boldsymbol{g}_k})$$

where $\boldsymbol{\eta}_{\boldsymbol{g}_k}$ is the intrinsic gradient of D_S at $\boldsymbol{g}_k$ and τ_k is a step size.

The following algorithm encodes the intrinsic gradient search which approximates a maximizer of D_S on the manifold S^{tq-1}.

Algorithm: Intrinsic Gradient Search

Input: Initial $\boldsymbol{g}_0 \in S^{tq-1}$
Output: Sequence of iterates $\{\boldsymbol{g}_k\}$.
for $k = 0, 1, 2, \ldots$ **do**
Select $\boldsymbol{g}_{k+1} = R_{\boldsymbol{g}_k}(\tau_k \boldsymbol{\eta}_k)$ where $\boldsymbol{\eta}_k = \operatorname{grad}(D_S)_{\boldsymbol{g}_k}$ and $\tau_k = \arg\max_\tau D_S(R_{\boldsymbol{g}_k}(\tau \boldsymbol{\eta}_k))$.
end for

The step size τ_k can be chosen on the interval $[0, 2\pi/\|\boldsymbol{\eta}_k\|)$ since $R_{\boldsymbol{g}}(\tau \boldsymbol{\eta}_k)$ is a periodic function of τ with period $2\pi/\|\boldsymbol{\eta}_k\|$. By the choice of τ_k, the information gain is non-decreasing, i.e., $D_S(\boldsymbol{g}_{k+1}) \geq D_S(\boldsymbol{g}_k)$ for each k.

It is worth noting that the proposed algorithm works for other optimization criteria. For example, if our goal is to find a channel matrix that minimizes trace or determinant of the conditional covariance matrix $\boldsymbol{Q}_{\boldsymbol{\theta\theta}|\boldsymbol{x},\boldsymbol{y}}$, we only need to replace the gradient correspondingly.

5. SIMULATION

In this section, we give a simple simulation study to demonstrate the performance of the proposed intrinsic search algorithm. We use the intrinsic algorithm to approximate the optimal MISO secondary channel and compare it with the true optimal solution in Section 3. The results suggest that the intrinsic algorithm converges to the global optimal solution.

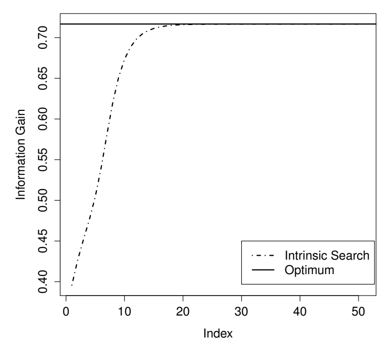

Fig. 3. A numerical study for a MISO secondary channel.

6. CONCLUSIONS

In this paper, we have studied the problem of fusing multiple sources of information. We considered a two-channel system where one of the input signals is of interest and the other is a secondary signal that is jointly distributed with the signal of interest. The objective is to design the secondary channel, with the primary channel fixed. Through an orthogonal decomposition of the secondary channel, we get insight into the problem and find that maximizing the information gain actually maximizes the determinant of a signal-to-noise-ratio matrix. We design the secondary channel to maximize the information gain brought by adding the channel. With the designed secondary channel matrix, combining the measurements of both channels achieves the best information gain. Moreover, the proposed intrinsic algorithm can be used to optimize various other design criteria besides the information rate.

7. REFERENCES

[1] D. P. Palomar, J. Cioffi, and M. A. Lagunas, "Joint TxRx beamforming design for multcarrier MIMO channels: A unifed framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sept. 2003.

[2] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[3] F. Pèrez-Cruz, M. R. D. Rodrigues, and S. Verdù, "MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," in *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1070–1084, Mar. 2010.

[4] W. R. Carson, M. R. D. Rodrigues, M. Chen, L. Carin, and R. Calderbank, "Communications inspired compressive sensing," availabel at arXiv:1206.1973.

[5] A. Vosoughi and A. Scaglione, "Precoding and decoding paradigms for distributed vector data compression," *IEEE Trans. on Signal Process.*, vol. 55, no. 4, pp. 1445–1460, Apr. 2007.

[6] M. Lamarca, "Linear precoding for mutual information maximization in MIMO systems," in*Intl. Symp. on Wireless Communication Systems (ISWCS '09),* pp. 26–30, Sept. 2009.

[7] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Communication*, vol. 49, no.12, pp. 2198–2206, Dec. 2001.

[8] C. Xiao, Y. R. Zheng, and Z. Ding, "Globally optimal linear precoders for finite alphabet signals over complex vector Gaussian channels," *IEEE Trans. Signal Process.,* vol. 59, no. 7, pp. 3301–3314, July 2011.

[9] D. Palomar and Y. Jiang, "MIMO transceiver design via majorization theory," *Now Publishers Inc*, 2007.

[10] D. Guo, S. Shamai, and S. Verdù, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[11] D. Palomar and S. Verdù, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.

[12] D. P. Bertsekas "Projected Newton methods for optimization problems with simple constraints," *SIAM J. Control and Optimization*, vol. 20, no. 2, pp. 221–246, Mar. 1982.

[13] P. A. Absil, R. Mahony and R. Sepulchre, "Optimization algorithms on matrix manifolds," *Princeton University Press*, 2008.

[14] J. M. Lee, "Introduction to Smooth Manifolds," *Springer*, 2000.

[15] H. Witsenhausen, "A determinant maximization problem occuring in the theory of data communication," *Jounal of Applied Mathematics*, vol. 29, no. 3, pp. 515–522, Nov. 1975.

[16] T. M. Cover and J. A. Thomas, "Elements of information theory," 2nd ed. *Wiley-Interscience*, 2005.