Computationally Efficient Active Inference in High-Dimensional State Spaces

Dmitry Kopitkov† and Vadim Indelman*

***Abstract*— We develop a novel approach for decision making under uncertainty in high-dimensional state spaces, considering both active unfocused and focused inference, where in the latter case reducing the uncertainty of only a subset of variables is of interest. State of the art approaches typically first calculate the posterior information (or covariance) matrix, followed by determinant calculation of thereof, and do so separately for each candidate action. In contrast, using the generalized matrix determinant lemma, we avoid calculating these posteriors and determinants of large matrices. Furthermore, as our key contribution we introduce the concept of calculation re-use, performing a one-time computation that depends on state dimensionality and system sparsity, after which evaluating the impact of each candidate action no longer depends on state dimensionality. Such a concept is derived for both active focused and unfocused inference, leading to general, non-myopic and exact approaches that are faster by orders of magnitude compared to the state of the art.**

I. Introduction

Decision making under uncertainty is a fundamental problem in robotics and artificial intelligence, with applications including autonomous driving, sensor deployment and active SLAM. The problem, also sometimes referred to as active inference, can be formulated as selecting optimal action from a set of candidates, based on some cost function. In information-based decision making the cost function typically contains terms that evaluate the expected posterior uncertainty upon action execution, with commonly used costs including entropy and mutual information. Thus, the corresponding calculations typically involve calculating a determinant of a posteriori covariance (information) matrices.

Decision making under uncertainty becomes an even more challenging problem when considering *high* dimensional state spaces. Such a setup is common in robotics, for example in the context of active SLAM, sensor deployment and graph sparsification. In particular, calculating a determinant of information (covariance) matrix for an n-dimensional state is in general $O(n^3)$, and is smaller for sparse matrices as in SLAM problems [1]. Moreover, state of the art approaches typically perform these calculations from scratch for *each* candidate action (for example in active SLAM [3], [9] and in sensor deployment [13], [14]).

In this paper we develop a computationally efficient and exact approach for decision making in high-dimensional state spaces that addresses the aforementioned challenges. The *key idea* is to use the general matrix determinant lemma to calculate action impact with complexity *independent* of state dimensionality n, while *re-using* calculations between evaluating impact for different candidate actions. Our approach supports general observation and motion models, and nonmyopic planning, and is thus applicable to a wide range of applications such as those mentioned above, where fast decision making in high-dimensional state spaces is required.

Moreover, we show the proposed concept is applicable also to active *focused* inference. Unlike the *unfocused* case discussed thus far, active *focused* inference approaches aim to reduce the uncertainty over only a predefined set of the variables. While the set of *focused* variables can be small, exact state of the art approaches calculate the marginal posterior covariance (information) matrix, for each action [11], which involves a computationally expensive Schur compliment operation.

In contrast, we provide a novel way to calculate posterior entropy of *focused* variables, which is fast, simple and general, yet, it does not require calculation of a posterior covariance matrix. In combination with our *re-use* algorithm, it provides *focused* decision making solver which is significantly faster (and exact) compared to standard approaches.

Finally, there is also a relation to the recently introduced concept of decision making in a conservative sparse information space [5]. While our approach confirms the concept from [5] considering the same assumptions, it addresses a general non-myopic decision making problem, with arbitrary observation and motion models.

To summarize, our contributions are: (a) we develop an approach for a nonmyopic decision making in high-dimensional state spaces that uses the matrix determinant lemma to avoid calculating determinants of large matrices, with per-candidate complexity independent of state dimensionality; (b) we show calculations can be re-used when evaluating impact of different candidate actions; (c) we develop a corresponding approach also for active *focused* inference.

II. Notations and Problem Definition

Consider the joint probability distribution function (pdf) $p(X_k|Z_{0:k}, u_{0:k-1})$ at time t_k over a high-dimensional problem-dependent state vector $X_k \in \mathbb{R}^n$. For example, in a SLAM problem X_k could represent robot poses and mapped landmarks, while in a sensor deployment problem X_k would represent an uncertainty field to be measured or monitored by adequately deploying sensors. Here, $Z_{0:k}$ and $u_{0:k-1}$ represent, respectively, all the observations and controls until time t_k. The joint pdf can be written as

†Technion Autonomous Systems Program (TASP), Technion - Israel Institute of Technology, Haifa 32000, Israel.

*Department of Aerospace Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel.

$$p(X_k|Z_{0:k}, u_{0:k-1}) = \eta p(x_0) \prod_{i=1}^{k} p(x_i|x_{i-1}, u_{i-1}) p(Z_i|X_i^o), \tag{1}$$

where η is a normalization constant and $p(x_0)$ is a prior on the first pose. The motion and observation models $p(x_i|x_{i-1}, u_{i-1})$ and $p(Z_i|X_i^o)$ are defined by $x_{i+1} = f(x_i, u_i) + \omega_i$ and $Z_i = h(X_i^o) + \upsilon_i$, respectively. Here, $\omega_i \sim N(0, \Sigma_{\omega,i})$ and $\upsilon_{i,j} \sim N(0, \Sigma_{\upsilon,i})$, with corresponding covariance matrices $\Sigma_{\omega,i}$ and $\Sigma_{\upsilon,i}$.

We use the notation $X_i^o \subseteq X_i$ to indicate the involved subset of states in the observation function $h(\cdot)$. In particular, such a formulation can be used to describe a SLAM problem: at any given time step t_i, the robot may acquire multiple landmark observations, denoted by $Z_i = \{z_{i,1}, ..., z_{i,n_i}\}$, where $z_{i,j}$ is a single observation of the jth landmark and n_i is the number of such observations. In such a case, the measurement likelihood term in Eq. (1) becomes $p(Z_i|X_i^o) = \prod_{j=1}^{n_i} p(z_{i,j}|x_i, l_j)$.

Following standard maximum a posteriori (MAP) inference, it is possible to efficiently infer, while exploiting sparsity and re-using calculations, the mean $\hat{X}_k$ and covariance Σ_k of the multivariate Gaussian $b[X_k] \doteq p(X_k|Z_{0:k}, u_{0:k-1}) = \mathcal{N}(\hat{X}_k, \Sigma_k)$, see e.g. [8].

In the context of decision making under uncertainty, one can now reason how the pdf (1), or the *belief* $b[X_k]$, will evolve as a result of some candidate action. Considering a planning horizon of L look ahead steps and a sequence of actions $u_{k+1:k+L-1}$, the belief $b[X_{k+L}] \doteq p(X_{k+L}|Z_{0:k+L}, u_{0:k+L-1})$ can be written as :

$$b[X_{k+L}] = \eta b[X_k] \prod_{l=k+1}^{k+L} p(x_l|x_{l-1}, u_{l-1}) p(Z_l|X_l^o) \tag{2}$$

It is not difficult to show (see e.g. [6]) that the posterior information matrix of the belief $b[X_{k+L}]$ is given by:

$$\Lambda_{k+L} = \Lambda_k + A^T A, \tag{3}$$

where $A \in \mathbb{R}^{m \times n}$ represents Jacobians combined in one single matrix of all new factor terms in Eq. (2) (motion and observation terms all together), linearized about the current estimate of X_k.

For notational convenience, we define the set of candidate actions by $\mathcal{A} = \{a_1, a_2, .., a_k\}$ with appropriate Jacobian matrices $\Phi_A = \{A_1, A_2, .., A_k\}$. While the planning horizon is not explicitly shown, each $a \in \mathcal{A}$ can represent a sequence of actions, e.g. $a = u_{k:k+L-1}$ for L look ahead steps.

In this paper we focus on information-theoretic decision-making and consider differential entropy $\mathcal{H}$ (further referred to just as entropy) as the cost function. Thus, we re-define the objective function as $J_{\mathcal{H}}(a) \doteq \mathcal{H}(b[X_{k+L}])$, where the belief $b[X_{k+L}]$ is a function of the controls $a = u_{k:k+L-1}$.

In particular, for Gaussian distributions, entropy is a function of the determinant of a posterior information (covariance) matrix, i.e. $\mathcal{H}(b[X_{k+L}]) \equiv \mathcal{H}(\Lambda_{k+L})$:

$$\mathcal{H}(b[X_{k+L}]) = \frac{n}{2} \cdot (1 + \ln(2\pi)) - \frac{1}{2} \ln |\Lambda_{k+L}|, \tag{4}$$

where $\Lambda_{k+L} = \Lambda_k + A^T A$, according to Eq. (3). Thus, evaluating $J_{\mathcal{H}}(a)$ requires determinant calculation of an $n \times n$ matrix, which is in general $O(n^3)$, per candidate action $a \in \mathcal{A}$. The optimal action is then given by $a^* = \arg\min_{a \in \mathcal{A}} J_{\mathcal{H}}(a)$.

Information gain (IG) is another common information-theoretic cost (e.g. [4], [11]) that we will use in this paper:

$$J_{IG}(a) \doteq \mathcal{H}(b[X_k]) - \mathcal{H}(b[X_{k+L}]). \tag{5}$$

The optimal action is defined for this cost as $a^* = \arg\max_{a \in \mathcal{A}} J_{IG}(a)$. Note that both objective functions $J_{\mathcal{H}}(a)$ and $J_{IG}(a)$ yield the same result, yet the latter will be computationally beneficial (see Section III).

Thus far, the exposition referred to active *unfocused* inference. However, as will be shown in the sequel, our approach is applicable also to active *focused* inference.

Active *focused* inference is another important problem, where in contrast to the former case, only a subset of variables is of interest (see, e.g., [10], [11]). Considering posterior entropy over the *focused* variables $X_{k+L}^F \subseteq X_{k+L}$ we can write:

$$J_{\mathcal{H}}^F(a) = \mathcal{H}(X_{k+L}^F) = \frac{n_F}{2} \cdot (1 + \ln(2\pi)) + \frac{1}{2} \ln \left| \Sigma_{k+L}^{M,F} \right|, \tag{6}$$

where n_F is the dimension of X_{k+L}^F, and $\Sigma_{k+L}^{M,F}$ is the posterior marginal covariance of X_{k+L}^F (suffix M for marginal), calculated by simply retrieving appropriate parts of posterior covariance matrix $\Sigma_{k+L} = \Lambda_{k+L}^{-1}$.

In the following section we develop a computationally efficient approach that addresses both active *unfocused* and *focused* inference. As will be seen, this approach naturally supports non-myopic planning with arbitrary motion and observation models, and it is in particular attractive for decision making in high-dimensional state spaces.

III. Approach

A. *Unfocused Active Inference*

Information theoretic decision making involves evaluating the cost (4) or (5), an operation that requires calculating the determinant of a large $n \times n$ matrix (posterior information matrix), with n being the dimensionality of the state X. State of the art approaches typically perform these calculations from scratch for each candidate action.

In contrast, our approach contains a one-time calculation that will be reused afterwards to calculate impact of each candidate action. As will be seen below, the latter depends only on the number of new factor terms in the Jacobian matrix $A \in \mathbb{R}^{m \times n}$, which is a function of L.

We first consider the IG as the utility function. It is not difficult to show that Eq. (5) can be written as $J_{IG}(a) = \frac{1}{2} \ln \frac{|\Lambda_k + A^T A|}{|\Lambda_k|}$. Using the generalized matrix determinant lemma [2], this equation can be written as $J_{IG}(a) = \frac{1}{2} \ln \left| I_m + A \cdot \Sigma_k \cdot A^T \right|$ where $\Sigma_k \equiv \Lambda_k^{-1}$, as previously suggested in [4], [11] in the context of compact pose-SLAM and *focused* active inference.

The expression for $J_{IG}(a)$ provides an exact and general solution for information-based decision making, where each action candidate can produce any number of new factors (nonmyopic planning) and where factors themselves can be of any measurement model (unary, pairwise, etc.).

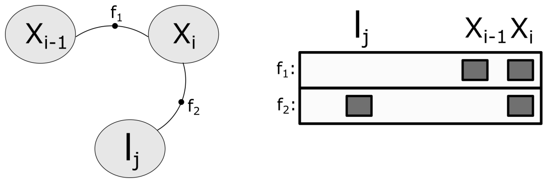

Fig. 1: Concept illustration of A's structure. Each column represents some variable from state vector. Each row represents some factor from Eq. (1). Factor f_1 of motion model that involves two poses x_i and x_{i-1} will have non-zero values only at columns of x_i and x_{i-1}. Factor f_2 of observation model that involves together variables x_i and l_j will have non-zero values only at columns of x_i and l_j.

It is important to note that A is a *sparse* $m \times n$ matrix, see Figure 1. Denote by C the set of all variables that are involved in A, i.e. these are the variables that are involved in at least one factor among the new factors generated due to the currently considered candidate action $a \equiv u_{k:k+L-1}$, see Eq. (2). Clearly, the only non-zero columns in A will be of variables in C. Consequently, the only required entries from the covariance matrix Σ_k are also those related to variables in C, and $J_{IG}(a)$ can be re-written as

$$J_{IG}(a) = \frac{1}{2} \ln \left| I_m + A_C \cdot \Sigma_k^{M,C} \cdot A_C^T \right| \tag{7}$$

where A_C is constructed from A by removing all zero columns, and $\Sigma_k^{M,C}$ is a prior joint marginal covariance of variables in C, which should be calculated from the (square root) information matrix Λ_k.

In particular, in case of myopic decision making with unary observation models (that involve only a single state variable), calculation of $IG(a)$ for different candidate actions only requires recovering the diagonal entries of Σ_k, regardless of the actual correlations between the states, as recently shown in [5]. However, while in the mentioned papers the per-action calculation takes $O(n)$, the $IG(a)$ calculation is $O(1)$ as will be shown below.

Given covariance matrix, the calculation in Eq. (7) is bounded by calculating determinant of an $m \times m$ matrix which is in general $O(m^3)$, where m is the number of constraints due to new factors (for a given candidate action a). This calculation should be performed for each candidate action in the set $\mathcal{A}$. Furthermore, in many problems it is logical to assume that $m << n$, as m depends mostly on the planning horizon L, which is typically defined and constant, while n (state dimensionality) can be huge and grow with time in real systems (e.g. SLAM). Consequently, given the prior covariance our complexity for selecting best action is $O(|\mathcal{A}|)$, i.e. *independent* of state dimensionality n.

Next, we make a *key observation* that covariance calculation for all candidate actions can be united into one computational block, calculated at the beginning of decision making phase and then later be *re-used* upon calculating IG for each candidate action.

In more detail, because only the joint covariance of involved variables is needed in order to calculate IG (7), we define the mutual set C_{All} as the state variables that are involved in at least one candidate action in $\mathcal{A}$. Before evaluating Eq. (7) for each candidate, we perform a *one-time calculation* to retrieve the joint covariance for variables C_{All}, i.e. $\Sigma_k^{M,C_{All}}$. Then Eq. (7) can be evaluated for each candidate action $a \in \mathcal{A}$ by simply retrieving appropriate blocks from $\Sigma_k^{M,C_{All}}$ in order to get $\Sigma_k^{M,C}$ for action a.

Intuitively, in most cases the candidates will have many mutual variables as they all are related to the current robot's location, in one way or another, making it even more reasonable to compute C_{All} in one calculation.

The complexity of this one-time calculation is different in different applications. When we use information filter, the system is represented by information matrix Λ_k, and in general the inverse of Schur compliment of C_{All} variables should be calculated. Yet, there are techniques that use sparse matrix nature of SLAM in order to efficiently recover marginal covariances [7].

In particular, in iSAM [8] the (linearized) system is represented by a squared root information matrix R_k, which is encoded, while exploiting sparsity, by the Bayes tree data structure. Decision making then can be performed by calculating, for each candidate action, the posterior matrix R_{k+L} (e.g. via Givens rotations [8] or another incremental factorization update method), and then calculating the determinant $|\Lambda_{k+L}| = \prod_{i=1}^{n} r_{ii}^2$, with r_{ii} being the ith entry on the diagonal of R_{k+L}. Yet, calculating R_{k+L} for each action can be expensive, particularly in loop closures, and requires copy/clone of the original matrix R_k. In contrast, per candidate calculation in Eq. (7) is constant in general.

The combination of calculation re-use and IG, Eq. (7), was evaluated in scenario of *unfocused* sensor deployment, where subset of locations should be selected which reduces the most the whole field uncertainty. The substantial reduction in running time of our approach, compared to the Standard approach, can be clearly seen in Figure 2a, which considers the entire decision making problem, i.e. evaluation of all candidate actions $\mathcal{A}$. The figure shows running time for sequential decision making, where at each time instant we choose the best locations of 2 sensors, with around $|\mathcal{A}| = 10^5$ candidate actions. The number of all sensor locations is $n = 625$ in this example. Overall, 15 sequential decisions were made. As seen, decision making using our approach requires only about 3 seconds, while the the Standard approach requires about 400 seconds.

B. Extension to Active Focused Inference

In this section we present a novel approach to calculate entropy of a *focused* set of variables, and then combine it with the ideas from the previous section (generalized matrix determinant lemma, IG cost function and calculation re-use) to develop a computationally efficient algorithm for *focused* information-based decision making.

First we recall definitions from Section II and introduce additional notations: $X_k^F \in \mathbb{R}^{n_F}$ denotes the set of *focused* variables, $X_k^R \doteq X_k / X_k^F \in \mathbb{R}^{n_R}$ is a set of the remaining variables, with $n = n_F + n_R$. The $n_F \times n_F$ marginal covariance and information matrices of X_k^F are denoted, respectively, by $\Sigma_k^{M,F}$ (suffix M for marginal) and $\Lambda_k^{M,F} \equiv (\Sigma_k^{M,F})^{-1}$. Furthermore, we partition the joint information matrix Λ_k as

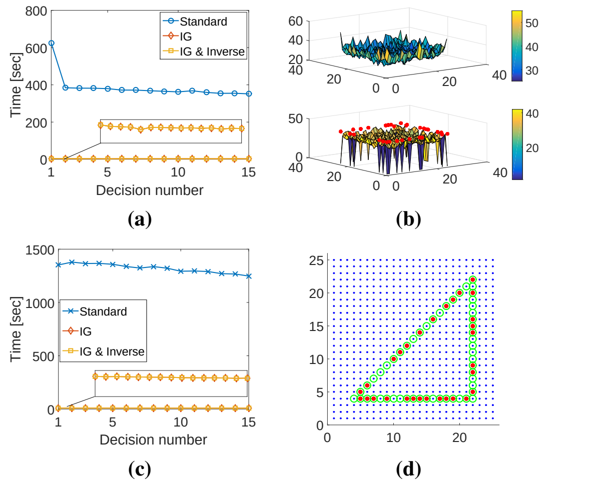

Fig. 2: *Unfocused* sensor deployment scenario: (a) running time for decision making, i.e. evaluating impact of all candidate actions, each representing candidate locations of 2 sensors, (b) prior and final uncertainty, with red dots marking selected locations. *Focused* sensor deployment scenario: (c) running time for decision making, (d) *focused* set of variables (green circles) and locations selected by algorithm (red dots).

$$\Sigma_k = \begin{bmatrix} \Sigma_k^{M,R} & \Sigma_k^{M,RF} \\ (\Sigma_k^{M,RF})^T & \Sigma_k^{M,F} \end{bmatrix}, \Lambda_k = \begin{bmatrix} \Lambda_k^{R} & \Lambda_k^{R,F} \\ (\Lambda_k^{R,F})^T & \Lambda_k^{F} \end{bmatrix}. \quad (8)$$

where $\Lambda_k^F \in \mathbb{R}^{n_F \times n_F}$ is constructed by retrieving from Λ_k only the rows and the columns related to F, $\Lambda_k^R \in \mathbb{R}^{n_R \times n_R}$ is the same thing but related to X_k^R, and $\Lambda_k^{R,F} \in \mathbb{R}^{n_R \times n_F}$ contains remaining blocks of Λ_k as shown in Eq. (8).

The marginal information matrix of X_k^F, i.e. $\Lambda_k^{M,F}$, can be calculated via Schur complement $\Lambda_k^{M,F} = \Lambda_k^F - (\Lambda_k^{RF})^T \cdot (\Lambda_k^R)^{-1} \cdot \Lambda_k^{RF}$. However, it can be shown [12] that $\left|\Lambda_k^{M,F}\right| = \frac{1}{\left|\Sigma_k^{M,F}\right|} = \frac{\left|\Lambda_k\right|}{\left|\Lambda_k^R\right|}$. Therefore, the posterior entropy of X_{k+L}^F is a function of the posterior Λ_{k+L} and its partition Λ_{k+L}^R:

$$J_{\mathcal{H}}^F(a) = \mathcal{H}(X_{k+L}^F) = \frac{n_F}{2} \cdot (1 + \ln(2\pi)) - \frac{1}{2} \ln \frac{\left|\Lambda_{k+L}\right|}{\left|\Lambda_{k+L}^R\right|}. \quad (9)$$

From Eq. (3) one can observe that $\Lambda_{k+L}^R = \Lambda_k^R + A_R^T A_R$, where $A_R \in \mathbb{R}^{m \times n_R}$ is constructed from Jacobian A by taking only the columns that are related to variables in X_k^R.

The next step is to use IG instead of entropy, with the same motivation and benefits as in the *unfocused* case (Section III-A). The optimal action $a^* = \arg\max_{a \in \mathcal{A}} J_{IG}^F(a)$ will maximize $J_{IG}^F(a) = \mathcal{H}(X_k^F) - \mathcal{H}(X_{k+L}^F)$, and using the generalized matrix determinant lemma we can write:

$$J_{IG}^F(a) = \frac{1}{2} \ln \frac{\left|I_m + A \cdot \Sigma_k \cdot A^T\right|}{\left|I_m + A_R \cdot \Sigma_k^{R|F} \cdot A_R^T\right|}, \quad (10)$$

where $\Sigma_k^{R|F} \in \mathbb{R}^{n_R \times n_R}$ is a prior covariance matrix of X_k^R conditioned on X_k^F, and it is actually the inverse of Λ_k^R.

The same concepts of calculation re-use are valid also here. The only required entries of $\Sigma_k^{R|F}$ are those related to variables involved in one of candidate actions. Thus their retrieval can be combined into one computational block.

We now consider the *focused* version of the sensor deployment problem (Eq. 6). In other words, the goal is to find sensor locations that maximally reduce uncertainty about chosen *focused* variables X^F. We have 54 such variables, which are shown in Figure 2d, while the rest of the problem setup remains identical to the *unfocused* case. In Figure 2c we show the corresponding results of our approach, compared to the Standard approach. The latter first calculates, for each candidate action, the posterior $\Lambda^+ = \Lambda + A^T A$, followed by calculation of Schur complement $\Lambda^{M,F}$ of the *focused* set X^F, and its determinant $\left|\Lambda^{M,F}\right|$ in order to get $J_{\mathcal{H}}^F(a)$ (Eq. 6). Our *focused* approach makes it possible to drastically reduce running time as shown in Figure 2c (9 seconds versus about 1400 in Standard approach).

IV. Conclusions

We developed a novel non-myopic and exact approach for information theoretic decision making in high dimensional state spaces, considering both *unfocused* and *focused* active inference problems. The key idea is to use the generalized matrix determinant lemma and re-use of calculations to efficiently evaluate the impact of each candidate action on posterior entropy. Our approach drastically reduces running time compared to the state of the art, especially when set of candidate actions is large, with running time being independent of state dimensionality. The approach was examined in problem of sensor deployment, exhibiting superior performance compared to the state of the art, and reducing running time by several orders of magnitude.

Acknowledgments

This work was supported by the Israel Science Foundation.

References

[1] Zhaojun Bai, Gark Fahey, and Gene Golub. Some large-scale matrix computation problems. *Journal of Computational and Applied Mathematics*, 74(1):71–89, 1996.

[2] David A Harville. Matrix algebra from a statistician's perspective. *Technometrics*, 40(2):164–164, 1998.

[3] S. Huang, N. Kwok, G. Dissanayake, Q. Ha, and G. Fang. Multi-step look-ahead trajectory planning in SLAM: Possibility and necessity. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 1091–1096, 2005.

[4] V. Ila, J. M. Porta, and J. Andrade-Cetto. Information-based compact Pose SLAM. *IEEE Trans. Robotics*, 26(1), 2010. In press.

[5] V. Indelman. No correlations involved: Decision making under uncertainty in a conservative sparse information space. *IEEE Robotics and Automation Letters (RA-L)*, 1(1):1–8, 2016.

[6] V. Indelman, L. Carlone, and F. Dellaert. Planning in the continuous domain: a generalized belief space approach for autonomous navigation in unknown environments. *Intl. J. of Robotics Research*, 34(7):849–882, 2015.

[7] M. Kaess and F. Dellaert. Covariance recovery from a square root information matrix for data association. *Robotics and Autonomous Systems*, 57(12):1198–1210, 2009.

[8] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert. iSAM2: Incremental smoothing and mapping using the Bayes tree. *Intl. J. of Robotics Research*, 31:217–236, Feb 2012.

[9] A. Kim and R. M. Eustice. Active visual slam for robotic area coverage: Theory and experiment. *Intl. J. of Robotics Research*, 2014.

[10] D. Levine and J. P. How. Sensor selection in high-dimensional gaussian trees with nuisances. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2211–2219, 2013.

[11] Beipeng Mu, Ali-akbar Agha-mohammadi, Liam Paull, Mathew Graham, Jonathan How, and John Leonard. Two-stage focused inference for resource-constrained collision-free navigation. In *Robotics: Science and Systems (RSS)*, 2015.

[12] Diane Valerie Ouellette. Schur complements and statistics. *Linear Algebra and its Applications*, 36:187–295, 1981.

[13] Zhengyuan Zhu and Michael L Stein. Spatial sampling design for prediction with estimated parameters. *Journal of Agricultural, Biological, and Environmental Statistics*, 11(1):24–44, 2006.

[14] Dale L Zimmerman. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics*, 17(6):635–652, 2006.