No Correlations Involved: Decision Making Under Uncertainty in a Conservative Sparse Information Space

Vadim Indelman

***Abstract*—This letter is concerned with decision making under uncertainty in problems involving high dimensional state spaces. Inspired by conservative information fusion techniques, we propose a novel paradigm where decision making is performed over a conservative rather than the original information space. The key idea is that regardless of the sparsity pattern of the latter, one can always calculate a *sparse* conservative information space, which admits computationally efficient decision making. In this letter, we take this concept to the extreme and consider a conservative approximation that decouples the state variables, leading to a conservative *diagonal* information matrix. As a result, the computational complexity involved with evaluating impact of a candidate action is reduced to $O(n)$, for an n-dimensional state, as the calculations do not involve any correlations. Importantly, we show that for measurement observation models involving arbitrary single state variables, this concept yields *exactly* the same results compared to using the original information matrix. We demonstrate applicability of this concept to a sensor deployment problem.**

***Index Terms*—Autonomous Agents, AI Reasoning Methods, Optimization and Optimal Control, SLAM.**

I. Introduction

Decision making under uncertainty is essential in numerous problems involving robot autonomy and artificial intelligence, including informative planning, active sensing, sensor deployment and active simultaneous localization and mapping (SLAM). These problems are typically addressed using information-theoretic decision making and belief-space planning approaches, where one looks for the most informative action(s) according to a given objective function.

Different information-theoretic objective functions have been proposed, including entropy and mutual information. These functions are used to evaluate the impact of a candidate action, in the discrete case, or calculating the appropriate gradient, in the continuous case. In the more general problem setting of planning under uncertainty, the objective function may include also additional terms, expressing, for example in an autonomous navigation scenario distance to goal and control usage. Calculating an optimal solution involves identifying a sequence of actions that minimizes the objective function, a problem that was shown to be NP-complete.

The problem becomes even more computationally challenging when considering high dimensional state spaces.In particular, evaluating the impact of a *single* candidate action, e.g. calculating the conditional entropy, is by itself an expensive operation: considering an n dimensional state and without making any assumptions on the sparsity of the information (covariance) matrix, the computational complexity is $O\left(n^3\right)$. This complexity can be significantly reduced in the case the involved matrices are sparse, e.g. as in SLAM problems. In the limit, if the information matrix was diagonal, i.e. no correlations between states, the complexity would be $O(n)$. Deciding which action is the best, in a discrete case with m possible actions, is therefore $O\left(mn^3\right)$ in general and, hypothetically, could be reduced to $O(mn)$ in case states are not correlated.

In this letter we suggest a new paradigm for decision making under uncertainty, where the complexity of choosing a single best action is $O(mn)$ without assuming sparse matrices. Specifically, we consider decision making with no dynamics and unary measurement models that possibly involve different states, a formulation relevant to problems such as active sensing and sensor deployment. We show that for such problems the proposed paradigm reduces significantly computational complexity *without* sacrificing performance.

Inspired by conservative information fusion techniques, we propose to perform decision making while considering a conservative approximation of the underlying probability distribution function (pdf), which corresponds in the Gaussian case to decision making in the conservative rather than the original information space.

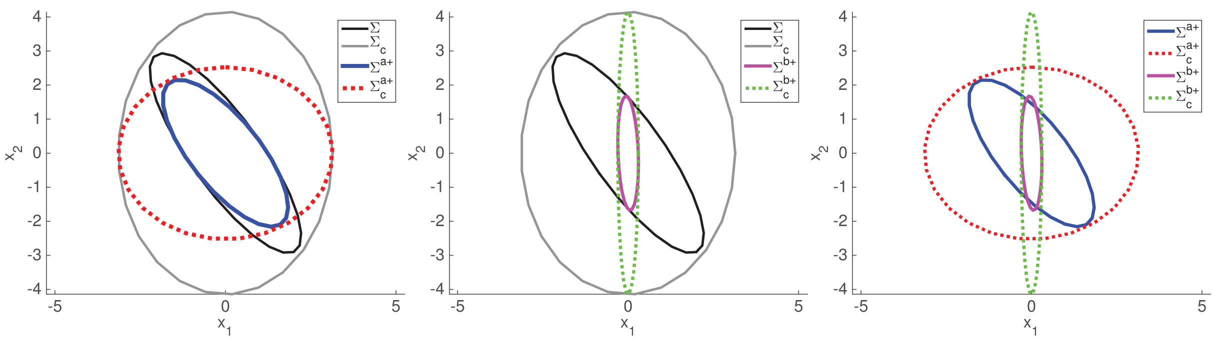
Such a concept is motivated by the fact that conservative information fusion techniques, e.g. covariance intersection [1], allow the correlations between different states to be unknown, while yielding consistent (but sub-optimal) state estimates. What happens if similar ideas are used for information-theoretic decision making?

In general, regardless of the sparsity pattern of the original information space, one can always resort to a *sparse* conservative information space, where some of the correlations are appropriately dropped, and perform decision making considering that information space. Clearly, this would result in a greatly reduced computational complexity. The question, however, is whether the performance is sacrificed with respect to using the original information space, i.e. whether the same decisions

Manuscript received August 25, 2015; accepted January 9, 2016. Date of publication January 14, 2016; date of current version February 12, 2016. This paper was recommended for publication by Associate Editor S. Chakravorty and Editor N. Amato upon evaluation of the reviewers' comments. This work was supported by the Israel Science Foundation under Grant 153535.

The author is with Department of Aerospace Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: vadim.indelman@technion.ac.il).

Digital Object Identifier 10.1109/LRA.2016.2518224



(a) (b) (c)

Fig. 1. Concept illustration for a two dimensional case, $X = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$, considering two candidate actions corresponding to measuring either the first or the second state. (a)–(b) Figures show an *a priori* covariance Σ and the conservative covariance Σ_c with no correlations between states, and the *a posteriori* covariances corresponding to two actions a and b that are equivalent to choosing the appropriate measurement model (2). (c) A posteriori covariances for each action using original and conservative *a priori* covariances: $\Sigma^{a+}, \Sigma^{b+}, \Sigma_c^{a+}, \Sigma_c^{b+}$. Conjecture 3 states that if the area enclosed by Σ^{a+} is smaller (larger) than the area enclosed by Σ^{b+}, then necessarily, the same relation holds also between Σ_c^{a+} and Σ_c^{b+}.

would be made in both cases. The best one could hope for, is that such a concept would allow evaluating a candidate action *without accounting for any* correlations while guaranteeing no sacrifice in performance.

In this letter we show, for the first time, this is indeed possible. We take the mentioned concept to the extreme, and consider a conservative pdf that *decouples* the correlations between the states, yielding a *diagonal* information matrix. Importantly, considering decision making with unary observation models over different state variables (as mentioned above), we prove decision making over this decoupled conservative information space yields *identical* results to those that would be obtained by using an original information space, see Figure 1. As a result, the complexity of evaluating a single candidate action is reduced from $O\left(n^3\right)$ to $O\left(n\right)$, regardless if the original information space is sparse or not.

II. Related Work

Information-theoretic decision making approaches have been investigated in different contexts for several decades. These include classical sensor selection and sensor management problems (e.g. [2]–[4]), informative path planning and active sensing [5], [6], sensor deployment [7], active vision [8], and active SLAM and belief space planning [9], [10], [11], [12], [13], [14] approaches. In the latter case, the objective function is not purely information-theoretic, as it also includes additional terms, such as distance to goal.

Since calculating a globally optimal solution is computationally intractable (at least NP-complete, see e.g. [7], [4]), different approaches have been developed that trade-off computational complexity with performance. These include greedy approaches that make a single decision at a time given the decisions made thus far (e.g. [7], [15]), branch and bound approaches (e.g. [6]), and direct optimization methods that calculate a locally-optimal solution given a nominal solution [10], [12], [14]. In particular, Krause et al. [7] presented a polynomial-time approximation for the sensor placement problem considering submodular objective functions, such as mutual information. Importantly, in that work the authors prove the performance of their greedy approach is within a constant factor from the optimal performance.

In this letter we develop a greedy algorithm that uses the concept of decision making in a conservative, sparse, information space to substantially reduce computational complexity. To the best of our knowledge, this concept is novel and has been only recently introduced in our work [16], [17]. With respect to the latter, in this letter we make a number of contributions. First, we formulate the concept within sequential decision making framework and consider its application to sensor deployment problems. Second, we make significant progress in rigorously proving the concept for high dimensional state spaces and show numerically the remaining presumed relations indeed hold. Due to space limitation, the letter is accompanied with supplementary material [18], that provides a detailed exposition of the proofs and additional results. Third, we analyze the concept's computational complexity and compare to computational complexity of state of the art approaches.

III. Problem Formulation

Let $p\left(X\right)$ denote a probability distribution function (pdf) over variables of interest $X \in \mathbb{R}^n$, which could represent, for example, robot poses, observed environment or an uncertainty field, such as temperature in different locations. In this letter, we model $p\left(X\right)$ by a multivariate Gaussian pdf

$$p\left(X\right) = N(\hat{X}, \Lambda^{-1}) \tag{1}$$

with estimate $\hat{X}$ and information matrix Λ. These parameters are conisidered to be known at the current time, e.g. estimated via maximum a posteriori (MAP) inference based on all information available thus far.

Suppose we have access to different sensor modalities and are to decide what parts of the state space to measure. Specifically, consider different measurement likelihoods $p_i\left(z_i|x_i\right)$ involving arbitrary state variables x_i from X, with $i \in [1, n]$, and the corresponding observation models of the form

$$z_i = h_i\left(x_i\right) + v_i, \quad v_i \sim N\left(0, \Sigma_{vi}\right), \tag{2}$$

where h_i and Σ_{vi} are, respectively, a nonlinear observation model and measurement noise covariance. Both h_i and Σ_{vi} can vary with i.

In the considered problem setup, we assume a determinstic control and no dynamics. This is a fairly typical assumption in sensor deployment and additional related problems (see, e.g. [4]). In other words, given a decision what state variable to measure (e.g. by deploying a sensor to the corresponding location), we assume this state variable will indeed be measured upon execution of the command. Future research will examine the more general case with stochastic control, for example, in the context of belief space planning.

A single decision involves evaluating the impact of each candidate measurement likelihood $p_i(z_i|x_i)$ according to a given information-theoretic utility measure J. In this letter we will consider entropy as the utility function J. For Gaussian distribution (1), entropy can be expressed as

$$H(p(X)) = \frac{1}{2}\log\left[(2\pi e)^n |\Lambda|^{-1}\right]. \quad (3)$$

The *a posteriori* pdf incorporating the dth candidate measurement likelihood $p_d(z_d|x_d)$ is given by

$$p(X|d, z_d) \propto p(X)\, p_d(z_d|x_d), \quad (4)$$

where we included conditioning on the candidate decision d to explicitly state the considered measurement likelihood.

One can now evaluate J for all possible decisions d and choose the best decision as

$$d^\star = \arg\min_d J(d). \quad (5)$$

It is often the case, however, that we are to make k best decisions at each time. Such a problem is encountered in numerous application domains, e.g. in the context of sensor selection, sensor deployment, and informative planning. This problem was shown to be computationally intractable, and is thus commonly addressed by approximate approaches; in particular, these include *greedy* methods that aim to make the best single decision given the decisions made thus far.

Assume l decisions have been already made and let the set $\mathcal{D}_l = \{d_1, \ldots, d_l\}$ represent these decisions. The corresponding pdf is thus

$$p(X|\mathcal{D}_l, \mathcal{Z}_l) \propto p(X)\prod_{i=1}^{l} p_{d_i}(z_{d_i}|x_{d_i}), \quad (6)$$

The measurements set $\mathcal{Z}_l = \{z_{d_i}\}$ represents the measurements corresponding to the l decisions $\mathcal{D}_l$. We note that at decision time these measurements are typically unknown, e.g. in sensor deployment we reason about where to position the sensors such that these would yield most informative observations. Nevertheless, assuming Gaussian distributions, the second moment of the pdf (6), which is involved in information-theoretic utility measures, can be typically approximated well by performing a single Gauss Newton iteration [19] and thus is not a function of the (unknown) measurements $\mathcal{Z}_l$.

To make the next decision (decision number $l+1$) with the considered greedy approach, one has to evaluate $J(d)$ for each candidate action d with the following *a posteriori* pdf (similarly to Eq. (4))

$$p(X|\mathcal{D}_l, \mathcal{Z}_l, d, z_d) \propto p(X|\mathcal{D}_l, \mathcal{Z}_l)\, p_d(z_d|x_d), \quad (7)$$

and then choose d_{l+1} according to (5). This process continues until k decisions are made.

The computational complexity of this process depends on several factors, including the objective function J and the involved correlations between the different states in X. Because of the latter, although the observation model is unary (involves only a single, arbitrary, state variable), additional states in X are impacted. In particular, in the fully-correlated case, all the states will be updated for any unary observation model. Moreover, the objective function J typically involves calculating the determinant of the *a posteriori* information (covariance) matrix, a computationally expensive operation in the general case.

In the following we develop an alternative approach that attains the *same* results, i.e. the same decisions, while drastically reducing computational complexity.

At this point it is beneficial to summarize the key assumptions made thus far: (i) the pdf $p(X)$ is assumed to be a Gaussian distribution; (ii) unary observation models (2) on arbitrary state elements from X are considered; (iii) in this letter we do not consider dynamics, and assume deterministic control as mentioned above.

IV. Approach Overview

Our approach is based on the concept of performing decision making in the conservative information space [16]. Consider, as before, the first l decisions have been made, with $l \in [0, k)$, and we are to make the next decision.

The *key idea* is to replace Eq. (7) with

$$p_c(X|\mathcal{D}_l, \mathcal{Z}_l, d, z_d) \propto p_c(X|\mathcal{D}_l, \mathcal{Z}_l)\, p_d(z_d|x_d), \quad (8)$$

where $p_c(X|\mathcal{D}_l, \mathcal{Z}_l)$ represents a conservative approximation of the pdf $p(X|\mathcal{D}_l, \mathcal{Z}_l)$ from Eq. (7).

The conservative pdf $p_c(X|\mathcal{D}_l, \mathcal{Z}_l)$ can be constructed such that its information matrix Λ_c is sparse regardless of the sparsity pattern of the information matrix Λ corresponding to $p(X|\mathcal{D}_l, \mathcal{Z}_l)$. By doing so, the computational complexity of evaluating the information-theoretic objective $J(d)$ with the *a posteriori* pdf from Eq. (8) can be significantly reduced.

In this letter we take this concept into the extreme and consider a conservative pdf $p_c(X|\mathcal{D}_l, \mathcal{Z}_l)$ where all the states are *independent*, with a corresponding *diagonal* information matrix Λ_c. We refer to this specific pdf as a *decoupling* conservative pdf. An illustration for a 2D case is given in Figures 1a–1b, which are further explained in Section V-C.

As we show next, this concept allows to determine the best action according to Eq. (5) by evaluating the objective function J using the conservative pdf (8) instead of the original pdf (7). Importantly, we prove that for unary observation models (2) this concept yields *identical* decisions as would be obtained by using the original pdf (7).

Because Λ_c is diagonal, calculating the *a posteriori* pdf (8) is a trivial operation for the considered unary observation models (2). Moreover, the corresponding *a posteriori* information

Algorithm 1. Decision making in the conservative information space

Inputs:
 k: number of decisions to make
 $p(X) = N(\hat{X}, \Lambda)$: *a priori* pdf over X
 $p_i(z_i|x_i)$: available measurement likelihoods $i \in [1, N]$
Outputs:
 Decisions set $\mathcal{D}$
Initialize: $\mathcal{D} = \phi$
for $l = 1 : k$ **do**
 Calculate Λ_c from Λ via Eq. (13)
 /* Evaluate H for each candidate */
 for $d = 1 : N$ **do**
  $A_d = \frac{\partial h_d}{\partial X}|_{\hat{X}}$ /* Calculate Jacobian A_d */
  $\Lambda_c^+ = \Lambda_c + A_d^T \Sigma_{vd}^{-1} A_d$ /* Calculate Λ_c^+ */
  $|\Lambda_c^+| = \prod_{i=1}^n (r_{c,ii}^+)^2$ /* Calculate $|\Lambda_c^+|$ */
  Calculate H from $|\Lambda_c^+|$ using Eq. (3)
 end
 Choose $d^\star$ according to Eq. (5)
 $\mathcal{D} \leftarrow \mathcal{D} \cup \{d^\star\}$
 $\Lambda = \Lambda + A_{d^\star}^T \Sigma_{vd^\star}^{-1} A_{d^\star}$
end
return $\mathcal{D}$

matrix Λ_c^+, given by

$$\Lambda_c^+ = \Lambda_c + A^T \Sigma_v^{-1} A, \tag{9}$$

where $A = \frac{\partial h_d}{\partial X}$ is an appropriate measurement Jacobian, linearized about the estimate $\hat{X}$ from Eq. (1), remains to be diagonal. As a result, calculating the determinant $|\Lambda_c^+|$, as required by most information-theoretic measures (e.g. entropy, mutual information), has *linear* computational complexity $O(n)$ in the number of states:

$$|\Lambda_c^+| = \prod_{i=1}^{n} \left(r_{c,ii}^+\right)^2, \tag{10}$$

where $\left(r_{c,ii}^+\right)^2$ is the ith entryon the diagonal of Λ_c^+. We note that this linear complexity is obtained *regardless of the actual correlations* between the states.

Algorithm 1 summarizes the proposed greedy approach for decision making in the conservative information space.

V. Decision Making in the Conservative Information Space

In this section we present the concept of decision making in the conservative information space, first discussing a single dimensional case, and then focusing on high dimensional state spaces.

A. 1D Case

The concept is formulated in the following theorem considering a one-dimensional state $X \in \mathbb{R}$ and some two candidate actions a and b.

Theorem 1:

$$\left|\Lambda^{a+}\right| \leq \left|\Lambda^{b+}\right| \text{ iff } \left|\Lambda_c^{a+}\right| \leq \left|\Lambda_c^{b+}\right|. \tag{11}$$

Here, Λ^{a+} and Λ_c^{a+} represent the *a posteriori* information matrices after performing action a

$$\Lambda^{a+} = \Lambda + A^T \Sigma_v^{-1} A, \quad \Lambda_c^{a+} = \Lambda_c + A^T \Sigma_v^{-1} A, \tag{12}$$

where $A \doteq \frac{\partial h_a}{\partial x}$ is the measurement Jacobian corresponding to the measurement model (2) of action a. The information matrices Λ^{b+} and Λ_c^{b+} are similarly defined for action b with a measurement Jacobian $B \doteq \frac{\partial h_b}{\partial x}$.

Theorem 1 states the impact of any two candidate actions on entropy, or other information-theoretic utility measures involving the determinant operator, has the same trend regardless if it is calculated based on the original pdf $p(x)$ or based on the conservative pdf $p_c(x)$. The proof of this theorem for a single dimensional state is trivial [16].

B. High-Dimensional State Space

Theorem 1 provides a mechanism to calculate the impact of candidate actions using *some* conservative information matrix. While formulated for a single dimensional state, the question is whether it is valid also for high-dimensional states $X \in \mathbb{R}^n$. Intuitively, if Theorem 1 is indeed valid, then one may construct a *sparse* conservative approximation of the information matrix Λ, which will yield significant reduction in computational complexity (e.g. when calculating the determinant), especially in problems that involve many states and a dense Λ. Importantly, Theorem 1 states this will be obtained for free, without any sacrifice in performance.

Here, we take this idea to the extreme and consider a conservative approximation that decouples the states in X, leading to a *diagonal* matrix Λ_c. This decoupled conservative approximation can be obtained as follows

$$p_c(X) \doteq \eta \prod_i p^{w_i}(x_i), \tag{13}$$

with x_i being the ith component in X, η a normalization constant, and w_i are weights such that $\sum_i w_i = 1$. Here, $p(x_i)$ is the marginal distribution over $x_i : p(x_i) = \int_{\neg x_i} p(X)$.

For Gaussian distributions, this is equivalent to scaling the covariance of each component of Σ by a factor of $1/w$, and can be considered as a special instance of covariance intersection [1], i.e. $\Sigma_{c,ii} = \Sigma_{ii}/w_i$ where Σ_{ii} is the ith diagonal entry in Σ. Thus, if $p(x_i) = N(\hat{x}_i, \Sigma_{ii})$, we have

$$\eta_i p^{w_i}(x_i) = N(\hat{x}_i, \Sigma_{ii}/w_i), \tag{14}$$

with an appropriate normalization constant η_i (see, e.g., [20]).

As a result, the covariance Σ_c and the (square root) information matrix $\Lambda_c = \Sigma_c^{-1}$, representing the second moment of the pdf (13), become *diagonal* matrices.

In practice, one can directly calculate these diagonal entries from non-zero elements in the square root information matrix R, with $\Lambda = R^T R$, thereby avoiding calculating an inverse of a large matrix (information matrix Λ). The corresponding equations are [21] (see also [15])

$$\Sigma_{ll} = \frac{1}{r_{ll}}\left(\frac{1}{r_{ll}} - \sum_{j=l+1}^{n} r_{lj}\Sigma_{jl}\right) \tag{15}$$

$$\Sigma_{il} = \frac{1}{r_{ii}}\left(- \sum_{j=i+1}^{l} r_{ij}\Sigma_{jl} - \sum_{j=l+1}^{n} r_{ij}\Sigma_{lj}\right) \tag{16}$$

for $l = n, \ldots, 1$ and $i = l-1, \ldots, 1$. Here, r_{ij} and Σ_{lj} correspond respectively to the element from row i and column j in the matrices R and Σ.

Remark: Observe the difference with conservative information fusion formulation, that can be written for any two pdfs $p_a(X)$ and $p_b(X)$ as $p_c(X) = \eta p_a^w(X)\, p_b^{1-w}(X)$, where η is a normalization constant [20]. Referring to Eqs. (7) and (8), the formulation considered herein differs in two respects [16]: (a) it calculates a conservative approximation of one of the pdfs via Eq. (13), leaving the other unchanged; (b) in the context of decision making, only the trend in entropy (or other measure) of a posteriori distributions for different candidate actions is of interest.

Let Λ^+ and Λ_c^+ represent the *a posteriori* information matrices defined for an arbitrary Jacobian $A = \frac{\partial h_i}{\partial X}$, according to measurement model (2), as given by Eq. (12).

The following Lemmas and Conjectures are necessary for stating our main result (Conjecture 3).

Lemma 1: The determinant Λ^+ can always be written as

$$\left|\Lambda^+\right| = \eta_n + a^2\gamma_n \tag{17}$$

where $\gamma_n > 0$ is only a function of entries of R, and η_n is defined as

$$\eta_n \doteq \prod_{i=1}^{n} r_{ii}^2. \tag{18}$$

The proof of Lemma 1 is given in Appendix A in [18].

Proof sketch: We use mathematical induction to prove Eq. (17) holds for any n. To do so, we recall that $|\Lambda^+|$ can be calculated from the diagonal of the corresponding square root information matrix R^+. The relation in Eq. (17) then follows by analyzing, using Givens rotations, how an a priori square root information matrix R is updated, due to observation model (2).

Lemma 2: The determinant Λ_c^+ can always be written as

$$\left|\Lambda_c^+\right| = \frac{\alpha_n}{\beta_n}\left[\eta_n + na^2\gamma_{c,n}\right] \tag{19}$$

where $\gamma_{c,n} > 0$ is only a function of entries of R,

$$\alpha_n \doteq n^{-n}\prod_{i=2}^{n} r_{ii}^{2(i-1)}, \tag{20}$$

and β_n is given by $\beta_n = \prod_{i=1}^{n}\gamma_{c,n-i+1}$.

The proof of Lemma 2 is given in Appendix B in [18].

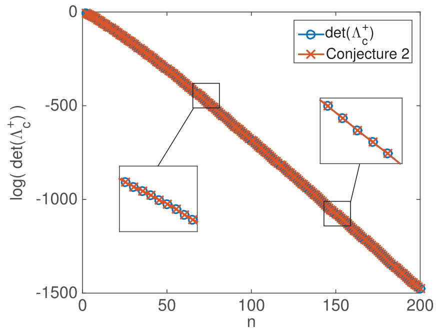


Fig. 2. Numerical evaluation comparing $\left|\Lambda_c^+\right|$ calculated by taking the determinant of Λ_c^+, with the same quantity calculated according to Conjectures 1–2, i.e. via Eq. (21). For numerical reasons, calculations are performed in log space.

Proof sketch: Recall Eq. (10): $|\Lambda_c^+| = \prod_{i=1}^{n}\left(r_{c,ii}^+\right)^2$. According to Eq. (13), $r_{c,ii}^2 = w_i\Sigma_{ii}^{-1}$ where Σ_{ii} is the corresponding entry on the diagonal of the covariance matrix $\Sigma \equiv \Lambda^{-1}$. Writing the Jacobian explicitly as $A \doteq \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}$, and considering some ith state variable is measured, we have $a_i \equiv a$ and $a_j = 0$ for $j \neq i$. Since R_c is diagonal, one can verify that $r_{c,jj}^+ = r_{c,jj}$ for $j \neq i$. Without loss of generality, we now arbitrarily assume the first state is measured, i.e. $i = 1$. Using Givens rotations we then show [18] that $|\Lambda_c^+| = \left(\Sigma_{11}^{-1} + na^2\right)n^{-n}\prod_{i=2}^{n}\Sigma_{ii}^{-1}$. The covariance entries Σ_{ii} can be efficiently calculated directly from the non-zero entries in matrix R [21], as shown in Eqs. (15)–(16). Manipulating algebraically the resulting expressions and resorting to recursive formulation yields Eq. (19).

We note that α_n, η_n and β_n are only functions of entries of R and do not involve any entries of the Jacobian A.

Conjecture 1: $\forall n: \ \ \gamma_n \equiv \gamma_{c,n}$.

Conjecture 2: The determinants of Λ_c^+ and Λ^+ are related according to

$$\left|\Lambda_c^+\right| = \frac{\alpha_n}{\beta_n}\left[n\left|\Lambda^+\right| - (n-1)\,\eta_n\right], \tag{21}$$

Proof: The relation immediately follows from Lemmas 1 and 2 and Conjecture 1. ■

In Appendix C in [18], we prove Conjecture 1 for the cases $n = 2$ and $n = 3$. Figure 2 provides numerical evidence that Eq. (21), and therefore Conjecture 1, holds within the range $n \in [1, 200]$, and suggests the relation is valid for any n: In that figure the a priori information matrix Λ and the Jacobian A describing the unary observation model (2) were randomly generated for different considered dimensions (parameter n). In each case, the determinant $|\Lambda_c^+|$ was evaluated directly and via Eq. (21). As seen, these calculations yielded identical results (also when repeated numerous times). Of course, while encouraging, these empirical results cannot replace a formal proof and for that reason we use the term 'Conjecture' and not 'Lemma' (see also Remark at the end of this section).

We can now formulate the *main result* of this letter.

Conjecture 3: Theorem 1 holds for high dimensional state spaces for unary measurement models (2) and the decoupled conservative approximation (13).

Proof: Consider the Jacobians A and B that correspond to two arbitrary measurement models (2). According to Conjecture 2, the following holds:

$$\left|\Lambda_c^{a+}\right| = \frac{\alpha_n}{\beta_n}\left[n\left|\Lambda^{a+}\right| - (n-1)\,\eta_n\right], \tag{22}$$

$$\left|\Lambda_c^{b+}\right| = \frac{\alpha_n}{\beta_n}\left[n\left|\Lambda^{b+}\right| - (n-1)\,\eta_n\right]. \tag{23}$$

Therefore:

$$\left|\Lambda_c^{a+}\right| - \left|\Lambda_c^{b+}\right| = n\frac{\alpha_n}{\beta_n}\left[\left|\Lambda^{a+}\right| - \left|\Lambda^{b+}\right|\right]. \tag{24}$$

However, according to Eq. (20) and Lemma 1, both α_n and β_n are always positive and are *not* functions of a. Therefore, the relation given in Eq. (11), and thus also Conjecture 3, directly follows from Eq. (24). ■

Remark: We note that Conjectures 2-3 rely on Conjecture 1. Proving Conjecture 1 for general n will thus allow reformulating these Conjectures into Theorems - this endeavor is left to future research.

C. Discussion

According to Conjecture 3, decision making involving unary observation models (2) can be performed over the decoupled conservative information space, represented by the diagonal information matrix Λ_c. This framework admits decision making without accounting for any correlations between the states while guaranteeing to *yield exactly the same decisions* as would be obtained by using the original information space, and without making any assumptions regarding the latter.

The practical implication is a major reduction in computational complexity: evaluating the impact of a candidate action typically involves calculating the determinant $|\Lambda^+|$, which is $O\left(n^3\right)$ in the general case. Our approach instead operates over the conservative information space, thus replacing this calculation with $|\Lambda_c^+|$. Since Λ_c^+ is diagonal, computational complexity is reduced to $O\left(n\right)$. We note this reduced computational complexity is obtained regardless if $\Lambda^+ \in \mathbb{R}^{n\times n}$ is sparse or not.

However, there is also a *conceptual* implication: decision making involving unary measurement models (2) can be performed *without accounting for the correlations* between the states.

We illustrate this statement in Figure 1 in a simple example involving two states, i.e. $X \in \mathbb{R}^2$. Figures 1a and 1b show the *a priori* covariance Σ and the calculated conservative covariance Σ_c according to Eq. (13). One can observe the two states are not correlated in the latter case. These two figures also show the *a posteriori* covariances that correspond to two actions a and b with the measurement models (2) involving different states:

$$A^T\Sigma_{va}^{-1}A = \begin{bmatrix}0 & 0\\ 0 & 0.2\end{bmatrix},\; B^T\Sigma_{vb}^{-1}B = \begin{bmatrix}24.7 & 0\\ 0 & 0\end{bmatrix}. \tag{25}$$

Note that each observation model involves a different state. However, because of correlation terms, each such observation model effects *both* states, as indicated by Σ^{a+} and Σ^{b+} in the figures. In contrast, since Λ_c is diagonal, each unary observation model involving state x_i, has impact only on the state x_i and not on the rest of the states, regardless of the actual correlations - this is shown by Σ_c^{a+} and Σ_c^{b+}.

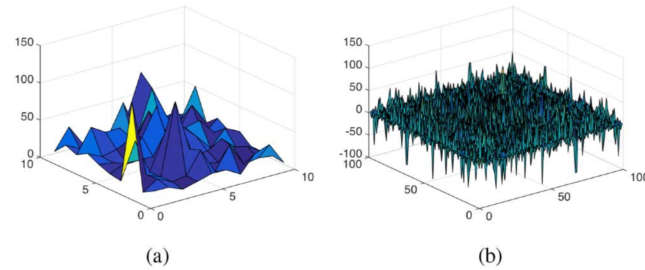


Fig. 3. Uncertainty field synthetic example: (a) A priori variance in each cell of the $N \times N$ grid; (b) A priori covariance $\Sigma \in \mathbb{R}^{N^2\times N^2}$.

Nevertheless, Conjecture 3 states that if the area covered by Σ^{a+} is larger (smaller) than the area covered by Σ^{b+}, then necessarily, the area covered by Σ_c^{a+} is larger (smaller) than the area covered by Σ_c^{b+}; See Figure 1c. Thus, decision making can be performed based on Σ_c^+ instead of Σ^+.

In the next section we consider an application of Conjecture 3 to sensor deployment and other closely related problems, developing a greedy approach with a significantly reduced computational complexity.

VI. Application to Sensor Deployment Problems

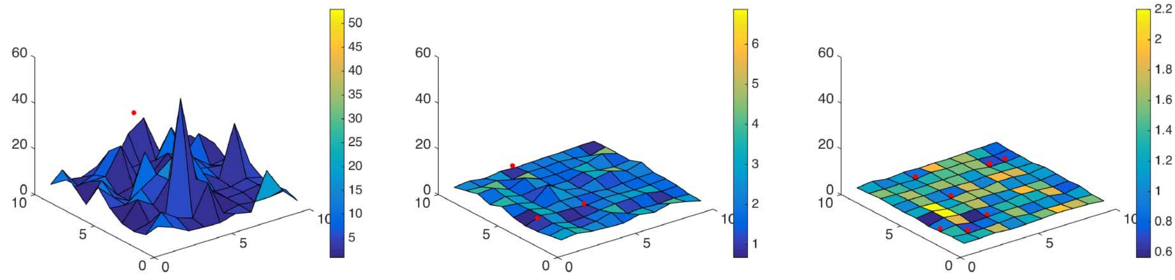
In this section we show how our approach can be applied to sensor deployment, sensor selection and other similar problems. The basic problem can be described as follows. There are k sensors that should be scattered in a much larger area such that these can, for example, monitor best a physical phenomena (e.g. temperature, wind), or provide localization for robots operating in different regions within the area.

For the sake of simplicity, we discretize the area into an $N \times N$ grid and thus would like to identify k cells for deployment of our k sensors. We note that, alternatively, we could avoid this discretization with a pre-defined resolution and resort to Gaussian Processes, as in [7].

Similar to [7], we assume there is some *a priori* covariance $\Sigma \in \mathbb{R}^{n\times n}$, with $n \doteq N^2$, that either describes prior knowledge on the uncertainty field we are to monitor, or is empirically determined from data.

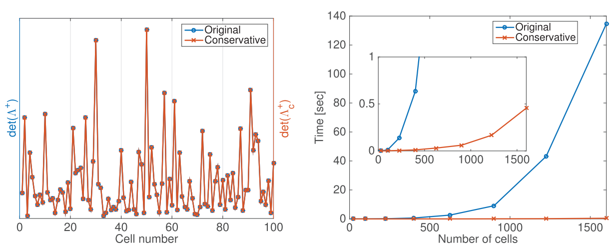
As an optimal solution to this problem is NP-hard, we consider a greedy approach, as discussed in Sections III and IV, and use entropy over the entire $N \times N$ area as the utility measure function.

Applying our approach for selecting the best k sensor locations, considering entropy as the objective function, is straightforward, as summarized in Algorithm 1. For the lth decision ($l \in [1, k]$), we first calculate the diagonal conservative information matrix Λ_c from the information matrix Λ that accounts for all the decisions made thus far (line 9). Then, we should evaluate the impact of each candidate sensor location by calculating the entropy over the *a posteriori* information matrix Λ_c^+. To do so, we compute the determinant $|\Lambda_c^+|$ (line 13) which is $O\left(n\right)$ as the matrix Λ_c^+ is diagonal. Given $|\Lambda_c^+|$, entropy calculation is trivial. Next, the best location is chosen according



(a) 1st sensor (b) 3 sensors (c) 7 sensors

Fig. 4. Variance of each cell after placing k sensors (cross-covariances between cells are not shown). Sensor chosen locations are denoted by red points. Colors indicate different uncertainty levels, with high and low uncertainty represented by yellow and blue colors, respectively.



(a) (b)

Fig. 5. (a) Impact of each candidate decision (sensor location) using the original and conservative information matrices (Λ and Λ_c). Although values are different, the trend is *identical* in both cases for any two candidate actions, as stated by Conjecture 3. (b) Processing time as a function of number of cells $n = N \times N$ in each case. Each run corresponds to solving the sensor deployment problem for $k = 10$ sensors.

to minimum entropy (line 16), and the information matrix Λ is updated accordingly (line 18). This procedure is repeated k times.

Next we demonstrate the method by presenting typical results in a synthetic scenario, and then in Section VI-B analyze the computational complexity of our approach.

A. *Typical Results*

Figure 3 shows the *a priori* covariance $\Sigma = \Lambda^{-1} \in \mathbb{R}^{N \times N}$, which was randomly generated, while enforcing a positive definite symmetric matrix. A relatively small problem is considered, with $N = 10$ and therefore $X \in \mathbb{R}^{100}$, to clearly illustrate the key points; results for a larger scenario with $N = 40$ (i.e. $X \in \mathbb{R}^{1600}$) are provided in Appendix D in [18]. The variance in each cell in the grid is shown in Figure 3a, while Figure 3b provides the entire joint covariance Σ. From the latter figure, one can clearly see the non-zero correlations between different cells.

Figure 4 shows several snapshots of the evolution of the *a posteriori* covariance after deploying a number of sensors that are chosen using our approach (Algorithm 1). As expected, one can observe uncertainty is greatly reduced upon deploying additional sensors. Identical sensor deployment results were obtained when using the conservative and the original information space. This is demonstrated in Figure 5 which shows the impact of each candidate action, sensor locations in our case, in terms of $|\Lambda^+|$ and $|\Lambda_c^+|$. The shown results refer to a single decision making event given candidate actions. As seen, while the actual values are different, the trend is identical. A similar behavior is also obtained while considering a larger scenario and with sequential greedy decision making, as shown in Appendix D in [18]. Thus, information-theoretic decision making can be performed using the conservative information space.

Figure 5b shows processing time for the sensor deployment problem with $k = 10$ sensors as a function of number of cells n. The shown timing results were obtained in a Matlab implementation of both methods. One can clearly observe decision making in a conservative information space requires much less processing time compared to using the original information space, especially as n increases.

B. *Computational Complexity*

Evaluating the impact of a single sensor location using our approach is $O(n)$ and doing so for all n location candidates is therefore $O\left(n^2\right)$, after which the best sensor location can be determined. Making k such decisions using this greedy approach is therefore $O\left(kn^2\right)$.

At this point, it is beneficial to compare this complexity to other greedy approaches. Using the *original* information space and without making any assumptions on the sparsity pattern of Λ, the computational complexity is $O\left(kn^4\right)$, since a single determinant calculation $|\Lambda^+|$ is $O\left(n^3\right)$. Robertazzi and Schwartz [22] and Krause et al. [7] propose the so called lazy evaluation technique that reduces, under certain conditions, the complexity to $O\left(kn^3\right)$. Exploiting the problem structure and using local kernels, i.e. assuming sparsity of the information matrix Λ, Krause et al. [7] are able to reduce the complexity to $O(kn)$.

Note that while the complexity of our approach is higher ($O\left(kn^2\right)$ vs $O(kn)$), no assumption regarding the sparsity pattern of Λ was made. We envision, however, that using similar concepts as in [7], computational complexity of our approach could be reduced even further.

VII. Limitations and Potential Extensions

While the approach developed herein addresses a specific family of problems, i.e. sensor deployment and related problems, an interesting question is whether it can be extended to more general settings, for example in the context of belief space planning. Addressing this question requires relaxing (some of) the assumptions that are mentioned towards the end of Section III (e.g. incorporating stochastic control and pairwise observation models) and to consider more general objective functions. The latter could include besides the information theoretic measure (e.g. entropy) also additional terms such as distance to goal and penalty on control usage (see e.g. [12], [14]). These aspects are outside the scope of this letter and should be investigated in future research.

We note the proposed concept of sparsifying the information space without, or with minimal, sacrifice in decision making performance is quite general. The decoupled conservative information space (13) considered in this letter is an extreme case for sparsification (since all correlations are appropriately dropped). Thus, other sparisfication approaches, including those discussed in [23], [24] in the context of long-term autonomy, could be appropriate to handle more general cases mentioned above.

VIII. Conclusions

We introduced a new paradigm where information-theoretic decision making is performed over a *conservative* information space. The concept is motivated by the fact that, regardless of the sparsity pattern of the original information space, one can always calculate a *sparse* conservative approximation of the latter, which admits computationally efficient decision making over high dimensional state spaces. Considering Gaussian distributions and unary measurement models possibly involving different state variables, we took this concept to the extreme and proved that using a *decoupled* conservative information space (diagonal information matrix) for decision making yields exactly the same results as would be obtained by the original, possibly highly-correlated, information space. As a consequence, regardless whether the latter is sparse or not, computational complexity for evaluating the impact of a single candidate action becomes *linear* in the number of states. We applied this concept to a sensor deployment problem, developing a greedy algorithm that uses a decoupled conservative information space for determining sensor locations.

Future research aims to extend the proposed concept to more general cases, including non Gaussian probability distributions, non-myopic planning and pairwise potentials.

References

[1] S. Julier and J. Uhlmann, “A non-divergent estimation algorithm in the presence of unknown correlations,” in *Proc. Amer. Control Conf.*, 1997, pp. 2369–73.

[2] L. Meier, J. Peschon, and R. M. Dressler, “Optimal control of measurement subsystems,” *IEEE Trans. Autom. Control*, vol. 12, no. 5, pp. 528–536, Oct. 1967.

[3] D. Ucinski, *Optimal Measurement Methods for Distributed Parameter System Identification*. New York, NY, USA: Springer, 2005.

[4] S. Joshi and S. Boyd, “Sensor selection via convex optimization,” *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009.

[5] A. Singh, A. Krause, C. Guestrin, and W. Kaiser, “Efficient informative sensing using multiple robots,” *J. Artif. Intell. Res.*, vol. 34, pp. 707–755, 2009.

[6] J. Binney and G. S. Sukhatme, “Branch and bound for informative path planning,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2012, pp. 2147–2154.

[7] A. Krause, A. Singh, and C. Guestrin, “Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies,” *J. Mach. Learn. Res.*, vol. 9, pp. 235–284, 2008.

[8] S. Wenhardt, B. Deutsch, J. Hornegger, H. Niemann, and J. Denzler, “An information theoretic approach for next best view planning in 3-D reconstruction,” in *Proc. Int. Conf. Pattern Recog. (ICPR)*, 2006, vol. 1, pp. 103–106.

[9] C. Leung, S. Huang, and G. Dissanayake, “Active SLAM using model predictive control and attractor based exploration,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2006, pp. 5026–5031.

[10] R. Platt, R. Tedrake, L. Kaelbling, and T. Lozano-Pérez, “Belief space planning assuming maximum likelihood observations,” in *Robotics: Science and Systems (RSS)*, Zaragoza, Spain, 2010, pp. 587–593.

[11] R. Valencia, J. V. Miró, G. Dissanayake, and J. Andrade-Cetto, “Active pose SLAM,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2011, pp. 1885–1891.

[12] J. Van Den Berg, S. Patil, and R. Alterovitz, “Motion planning under uncertainty using iterative local optimization in belief space,” *Int. J. Robot. Res.*, vol. 31, no. 11, pp. 1263–1278, 2012.

[13] A. Kim and R. Eustice, “Perception-driven navigation: Active visual SLAM for robotic area coverage,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2013, pp. 3196–3203.

[14] V. Indelman, L. Carlone, and F. Dellaert, “Planning in the continuous domain: A generalized belief space approach for autonomous navigation in unknown environments,” *Int. J. Robot. Res.*, vol. 34, no. 7, pp. 849–882, 2015.

[15] M. Kaess and F. Dellaert, “Covariance recovery from a square root information matrix for data association,” *Robot. Autonom. Syst.*, vol. 57, no. 12, pp. 1198–1210, 2009.

[16] V. Indelman, “Towards information-theoretic decision making in a conservative information space,” in *Proc. Amer. Control Conf.*, July 2015, pp. 2420–2426.

[17] V. Indelman, “On decision making and planning in the conservative information space—Is the concept applicable to active slam?” in *Proc. Prob. Mobile Sens.: Setting Future Goals Indic. Prog. SLAM; Robot. Sci. Syst. (RSS) Conf.*, July 2015, pp. 1–2.

[18] V. Indelman, “No correlations involved: Decision making under uncertainty in the conservative information space—Supplementary material,” Technion—Israel Inst. Technol., Tech. Rep. ANPL-2016-01. [Online]. Available: http://vindelman.technion.ac.il/Publications/Indelman16ral_Supplementary.pdf

[19] V. Indelman, L. Carlone, and F. Dellaert, “Towards planning in generalized belief space,” in *Proc. 16th Int. Symp. Robot. Res.*, Singapore, Dec. 2013.

[20] T. Bailey, S. Julier, and G. Agamennoni, “On conservative fusion of information with unknown non-gaussian dependence,” in *Proc. Int. Conf. Inf. Fusion (FUSION)*, 2012, pp. 1876–1883.

[21] G. Golub and R. Plemmons, “Large-scale geodetic least-squares adjustment by dissection and orthogonal decomposition,” *Linear Algebra Appl.*, vol. 34, pp. 3–28, Dec 1980.

[22] T.-G. Robertazzi and S.-C. Schwartz, “An accelerated sequential algorithm for producing D-optimal designs,” *SIAM J. Sci. Statist. Comput.*, vol. 10, no. 2, pp. 341–358, 1989.

[23] G. Huang, M. Kaess, and J. Leonard, “Consistent sparsification for graph optimization,” in *Proc. Eur. Conf. Mobile Robots (ECMR)*, 2012, pp. 150–157.

[24] N. Carlevaris-Bianco, M. Kaess, and R. M. Eustice, “Generic node removal for factor-graph SLAM,” *IEEE Trans. Robot.*, vol. 30, no. 6, pp. 1371–1385, Dec. 2014.