Risk-aversive optimal planning of sensing

Mikko Lauri, Aino Ropponen, Risto Ritala

Tampere University of Technology, Department of Automation Science and Engineering, P.O. Box 692, FIN-33101 Tampere, FINLAND

risto.ritala@tut.fi

Abstract. Operable sensing is studied as a means for improving system performance. Both single and sequential planning are analysed, and the complexity due to risk-averseness in various formulations discussed. Three classes of sensing operation are defined and analysed.

1. Introduction

Autonomous machines and automated processes are increasingly equipped with sensing systems that are operable. Field of view in machine vision systems can be varied by zooming and/or turning the cameras, and the systems can switch between stereo vision mode and optical flow mode. In chemical process industries automated process analyzers are able to measure many chemical parameters – one at a time – and to handle samples from many locations of the process. Future autonomous mobile machines (AMMs) are expected to have separate measurement machines, such as unmanned Micro Aerial Vehicles (MAV) that will collect data about the working environment in order to improve the performance of the AMM. Similarly, in social or biological systems, before deciding of an operational action, there may be limited time available or cost associated to obtain new data about the system and it must be decided, which data source provided should be accessed by some observation action. Recently theories relevant to artificial systems have been proposed about human vision behavior [1-2].

Sensing actions add value only when the system state is not known completely. The incomplete information about the state is represented probabilistically. Sensing can be optimized either to maximize the quality of information, i.e. the prior expectation of posterior entropy, or to maximize the system performance for the current task. When deciding about the sensing actions, the data resulting from such actions is obviously not known. Thus even if the system were deterministic the performance resulting from sensing and action on the system will be subject to uncertainty. In decision about sensing the probability distribution of interest is the prior distribution of posterior optimal performance [3]. In such decisions under uncertainty the attitude towards risk may affect the choices but is rarely considered in control/sensing optimization.

For dynamic systems a sequence of decisions is to be planned instead of a single decision. Optimal planning takes into account that in all the future time instants spanning the sequence, there will be an opportunity to make further sensing actions. The receding horizon principle is applied: a sequence is planned, the first action is then implemented and at the next time instant the planned sensing action is carried out, and based on data obtained a new sequence is planned [4].

This paper discusses the formulation and solutions of the optimal sensing problem through problem classes of practical interest. In particular, the attitude towards uncertainty is discussed and problems arising in different formulations of the risk attitude in sequential decision tasks are analyzed. Section 2

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

formulates optimal sensing. Section 3 presents three widely applicable problem classes of optimal sensing. Section 4 concludes the challenges when planning optimal sensing.

2. Objectives for operation of sensing: performance and attitude towards risk

Contrary to the design of experiments, in which the objective is to sense the system in order to minimize the uncertainty – in the most general form the entropy – about its state, here we consider sensing in the context of providing optimal support for a task given for the system. The sensing actions and the actions on the system to carry out the task are chosen to maximize a performance measure. The actions on the system are committed only after the data from sensing actions has been received.

Two typical performance measures are the reward (to be maximized) and the regulated end result (absolute difference form a desired end result to be minimized). These can be considered as linear, respective quadratic functions of the system state which is to be estimated based on sensing data.

According to utility theory optimal choices between distributions can be reduced to finding the distribution that maximizes the expectation of a utility function [5]. However, construction of a utility function is rather heuristic. There are also several other heuristic approaches for decisions between probability distributions. The uncertainty about a performance Q in sensing optimization can be taken into account, e.g. as follows:

- risk neutral: maximize the expected performance μ_Q without considering uncertainty
- risk constrained: maximize the expected performance μ_Q subject to that the probability of unsatisfactory performance Q_B is constrained, $P(Q < Q_B) < p_{constr}$
- risk aversive: concave utility function, e.g. $U_0(Q) = 1 - \exp(-\alpha(Q - Q_0))$ so that for Gaussian information about performance the expression $\mu_Q - \alpha\sigma_Q^2/2$ is maximized
- risk aversive: heuristic risk premium, $\mu_Q - \beta\sigma_Q$ is maximized
- risk conscious: maximize the probability of satisfactory performance, $P(Q > Q_S)$ [6]; this corresponds to a utility function which is constant above Q_S and negative infinite below.

The uncertainty may result from two main sources: either the predicted system state is uncertain, or the parameters of the performance function are uncertain. The latter case can be turned into state uncertainty by augmenting the uncertain parameters to the state. The decision about the sensing is based on the prior distribution of the posterior optimal performance. Let $s \in S$ be the system state, $a \in A$ an action on the system, $m \in M$ the observation action, and $y^{(m)}$ the data resulting from the observation action. The system develops from state s to state s' as a result of the action a with probability (density) $f(s'|s,a)$. Furthermore, let $Q(s',a,m)$ be the system performance, to be maximized, and $U(Q)$ be the utility of performance expressing the attitude towards risk. Then the posterior optimal action is

$$\begin{aligned} a_{opt}(m, y^{(m)}) &= \arg\max_{a \in A} \int_S ds'\, f(s'|y^{(m)}) U(Q(s',a,m)) \\ &= \arg\max_{a \in A} \int_S ds' ds f(s'|s,a) f(s|y^{(m)}) U(Q(s',a,m)) \end{aligned} \quad (1).$$

The posterior optimal performance is then

$$\tilde{Q}(m, y^{(m)}) = \int_S ds' ds f(s'|s,a) f(s|y^{(m)}) Q(s', a_{opt}(m, y^{(m)}), m) \quad (2),$$

This is a random variable before the measurement data has been received. Hence the measurement to be chosen is the one that maximizes the prior expectation of the utility of the posterior optimal performance:

$$m_{opt} = \arg\max_{m \in M} \int_{Y^{(m)}} dy^{(m)} f^{(ap)}\left(y^{(m)}\right) U\left(Q\left(m, y^{(m)}\right)\right) \quad (3),$$

where the prior probility (density) of data is obtained from measurement description $f\left(y^{(m)} \mid s\right)$ and prior state information $f^{(ap)}(s)$ as

$$f^{(ap)}(y^{(m)}) = \int_S ds f\left(y^{(m)} \mid s\right) f^{(ap)}(s) \quad (4).$$

A common class of optimal sensing problems is sequential sensing planning: the overall performance is a sum of instantaneous performances over a time horizon. As the expected performance is cumulative, the optimal sequence of decisions can be solved with dynamic programming, i.e. with the Bellman equation. However, if the decision considers the risk, e.g. through that the decision objective is a nonlinear utility of the performance or the risk premium is included, the objective function is in general not cumulative and Bellman's principle of optimality, the cornerstone of dynamic programming, is not valid. The utility function $U_0(Q)$ above is special in that if the performance is a sum of independent stage performances, the variances are additive. Hence the objective function is cumulative and Bellman's principle of optimality applies.

3. Case studies on optimal operation of sensing

This section discusses three classes of optimal sensing tasks, relevant in many real-life applications.

3.1. Discrete action options with Gaussian performances – path planning

Consider an AMM whose task is to optimize its route to a final destination. The map of routes is a directed acyclic graph. The information about the performance (e.g. time) of each leg is incomplete. AMM is equipped with an MAV as a measurement machine capable of improving the information about the cost of the next leg. When AMM enters a junction, it must decide if the MAV should be sent to scout the condition of the optional adjacent legs before choosing along which to proceed. Decision about measurements may be sequential or at one stage. Repeating the same measurement may improve information (regular measurement) or it may be non-informative (e.g. image-based measurement in which the uncertainty is in the interpretation of the image, not in capturing the image).

Let the prior information about the performances Q_i of the legs be Gaussian and independent on one another, $N\left(\mu_i, \sigma_i^2\right)$. The measurement of performance of the route is unbiased but uncertain, described by a Gaussian with measurement uncertainty $\sigma_{meas,i}^2$. There is a cost c_i of making the measurement. E.g. when repeated measurements improve information, the optimal value of utility function U_0 at a junction with two parallel legs satisfies the following functional equation:

$$V\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = \max\left\{ \begin{array}{c} \mu_1 - \alpha\sigma_1^2/2 \\ \mu_2 - \alpha\sigma_2^2/2 \\ -c_1 + \int_{-\infty}^{\infty} dy_1 N\left(y_1; \mu_1, \sigma_1^2 + \sigma_{meas,1}^2\right) V\left(\mu_1 + \frac{\sigma_1^2(\mu_1 - y_1)}{\sigma_1^2 + \sigma_{meas,1}^2}, \mu_2, \frac{\sigma_1^2 \sigma_{meas,1}^2}{\sigma_1^2 + \sigma_{meas,1}^2}, \sigma_2^2\right) \\ -c_2 + \int_{-\infty}^{\infty} dy_2 N\left(y_2; \mu_2, \sigma_2^2 + \sigma_{meas,2}^2\right) V\left(\mu_1, \mu_2 + \frac{\sigma_2^2(\mu_2 - y_2)}{\sigma_2^2 + \sigma_{meas,2}^2}, \sigma_1^2, \frac{\sigma_2^2 \sigma_{meas,2}^2}{\sigma_2^2 + \sigma_{meas,2}^2}\right) \end{array} \right. \quad (5)$$

If repeated measurements are noninformative the decision tree is finite and junction problems can be solved straight forwardly from leaves to root. The decision tree at a node is shown in Figure 1.

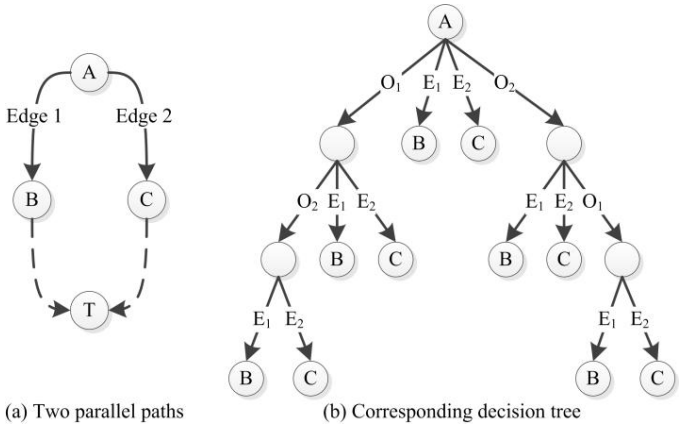


Figure 1. Sequential decisions at a junction of a directed acyclic graph when repeated measurements are noninformative. Prior optimal decisions from nodes B and C to target node T have been solved. In the case of additive objective function, decisions at A are made taking into account the measurement actions and choices of edges A-B and A-C and the optimal objectives B-T and C-T.

If the utility is cumulative – either the expected value of performance is maximized, or the utility U_0 is maximized and the performance terms are independent – a sequential route optimization on any directed acyclic graph can be solved by applying Bellman's principle of optimality both for informative and noninformative repeated measurements. This deterministic solution provides decision rules at each junction: to make measurement or not, and if measurement is made, how the next decision about measurement/advancing to next node depends on data obtained. However, the junction problems are computationally rather expensive as the prior expectations with respect to the measurement data of posterior optimal performances needs to be computed numerically. This limits practical solutions to maps where junctions have at most two exiting legs.

If the objective is to maximize the probability that the performance exceeds some given aspired value Q_S, $P(Q > Q_S)$, the problem can be solved backwards recursively as in the case of additive objective. However, the decisions at a junction obviously depend on how much performance has been accumulated when entering the junction. As this is a random variable not known when path is being planned, the optimal decisions must be solved for all values Q_S. Figure 2 shows an example of decisions at a node. First when entering the node there are four options in the root of Figure 1b. The probability that performance in these options exceeds Q_S (here: cost C is below a given aspired cost C_s) is given in Fig 2a. For example if $C_S = 33.6$, the optimal action is to measure the leg 2. Fig 2b shows $P(C < C_S)$ as a function of data $y^{(2)}$ obtained when measuring leg 2. If data obtained is $y^{(2)} = 33$, the optimal action is to measure also the leg 1. Correspondingly a decision rule between advancing along the route 1, respectively 2 can be given as a function of $y^{(1)}$. Fig 2c shows the entire system in which this decision takes place. the node considered is 5 after optimal decisions of its descendants 8 and 16 (trivial, being the target node) have been solved.

3.2. Optimal observation for occupancies in path planning

Consider an AMM traversing from a start to a goal location on a grid, where each traversed link between two adjacent locations results in a cost. Many real life planning problems can be formulated

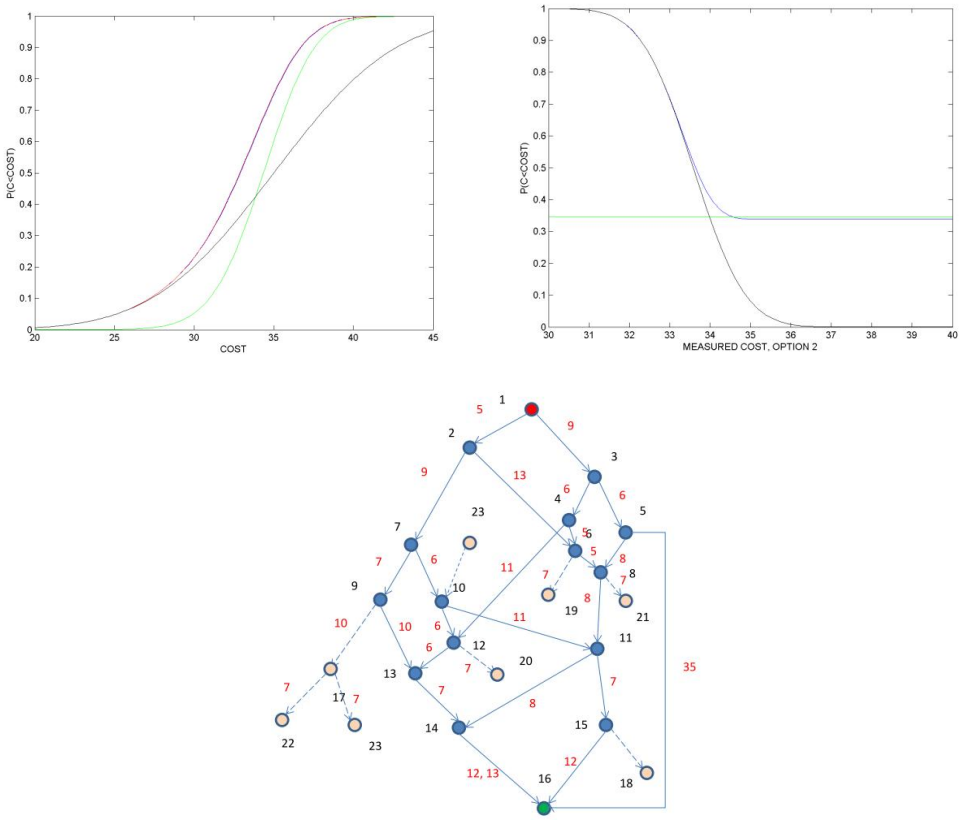


Figure 2. Maximizing the probability that the cost will not exceed a given aspired level. Top left: decision at a junction as a function of aspired cost; four options: measure leg 1 (blue), measure 2 (red), advance 1 (green), advance 2 (black). If aspired level of cost is 33.6, optimal is to measure 2. Top right: decision after measuring leg 2 as a function of measurement results; three options (colors as in top left). If measurement result is 33, next it is optimal to measure leg 1. A similar rule is established after obtaining measurement data from leg 1. Bottom: the entire path planning problem; initial site is 1 and target is 16. Red numbers indicate prior expected leg costs. The analysis presented in top figures is at node 5, after prvious optimal decisions have led to the route 1->3->5.

as path planning on an occupancy map with cost accumulation. We study the case in which the location rewards consist of known costs and unknown dynamic costs, dependent on Markovian binary occupancy states of the locations (high cost to penalize collisions). The location of the AMM at the grid is assumed known exactly. The AMM can choose between of moving to adjacent locations or observing the occupancies of some nearby locations. The npsotive and negative false observations are possible.

An approximate solution is generated as follows. First the paths from each location to goal location and the corresponding value functions are solved offline with the only the known rewards. Then the movement and observation actions are planned online over a short time horizon, typically 3-4 decision intervals, with the optimal value function from the first solution as the end reward. Typical result is shown in Fig. 3. When the agent is uncertain about obstacles not in the map, it chooses routes for which an observed obstacle can be by-passed from both sides. The computational complexity prevents longer on-line horizons, which in turn prevents the AMM to carry out complicated maneuvers to avoid observed obstacles.

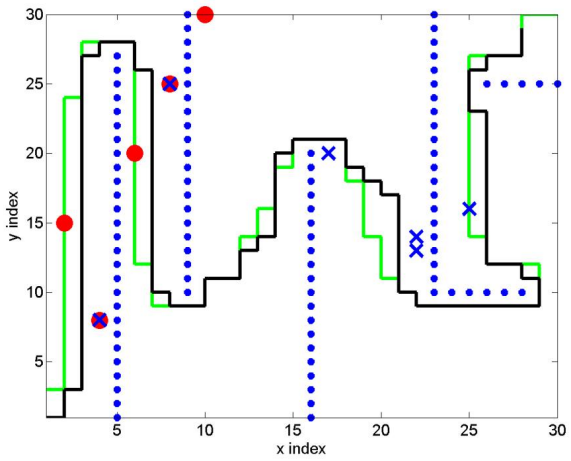


Figure 3. Path planning with a map in the presence of unmapped obstacles. Blue dots: obstacles in the map. Red dots: unmapped obstacles. Blue crosses: observations (true or false) about unmapped obstacles. Green line: optimal path according to known map. Black line: optimal route with optimal sensing for unmapped obstacles.

3.3. Optimal sensing of Linear-Quadratic-Gaussian systems

Quadratic objectives arise commonly in control problems. If the response of the system is linear and the initial state information is Gaussian, then the prior information about the end state is Gaussian. Furthermore, in risk neutral decisions, the optimal measurement action is independent of the estimate of the initial state and the optimal action on the system is independent on the initial state uncertainty [7]. Thus the optimal action and optimal sensing problems are decoupled and the optimal sensing can be solved as a policy offline. When the attitude is risk-aversive, the optimization of actions on the system and about sensing are coupled. In all risk-aversive forms the objective of sequential decisions is non-additive and Bellman's principle of optimality does not apply.

4. Conclusions

Operable sensing subsystem is a means of optimizing system performance. The formulation of risk-aversive planning of optimal sensing was discussed. Three classes of problems were given as examples. The non-neutral attitude towards risk was shown to lead radically increased complexity in sensing planning. In general the Bellman's principle of optimality in sequential decision making is no longer valid, except under exceptional circumstances. In LQG risk-aversive optimization sensing and optimal control are coupled. Some solutions to optimal sensing problems were outlined. This work is funded by Academy of Finland (project O3-SAM, 268152).

References

[1] Najemnik, J., & Geisler, W. S., Optimal eye movement strategies in visual search, Nature, 434(7031), 387-391 (2005).
[2] Johnson L., Sullivan B., Hayhoe M., Ballard D., Predicting human visuomotor behavior in driving task, Phil. Trans. R. Soc. B 369 20130044 (2014).
[3] Raiffa H., Schlaifer R., Applied Statistical Decision theory, Ch. 4.5, Wiley Classics (2000).
[4] see e.g. Thrun S., Burgard W., Fox D., Probabilistic Robotics, Ch. 17, MIT Press (2005).
[5] French S., Rios Insua D., Statistical Decision Theory, Kendall's Library of Statistics, Ch 2, Arnold Publishing (2000).
[6] Lim S., Sommer C., Nikolova E., Rus D., Practical Route Planning Under Delay Uncertainty: Stochastic Shortest Path Queries, Robotics, Science and Systems VIII (2012)
[7] Meier III L., Peschon J., Dressler R. M., Optimal Control of Measurement Subsystems, IEEE on Automatic Control, AC-12, 528-536 (1967).