# Research Notes

Deepak E. Gopinath

## Quantifying Fisher Information for a Control Dimension

We are given a goal distribution. We also assume that the user has a one and only one intended goal.
We are interested in using *Fisher Information* to quantify the amount of information a constrained motion along specific control dimension contains about the user's intended goal. Once the control dimension is specified, then the only motion that the user can initiate is along the specified control dimension. This is equivalent to using a 1D control interface for controlling the robot.
The formalism is inspired by how Todd Murphey's group have gone about utilizing Fisher information in the context of active sensing. In the problem they tackle, they have a mobile robot equipped with a sensor. The "unknown" parameter they try to estimate is the 2D position of a goal. The measurement model is specified as well. The sensor reads voltages and is a function of the current position of the robot and the intended goal position and is corrupted by Gaussian white noise.
Once Fisher Information Matrix (FIM) is computed, the Expected Information Matrix (EIM) is computed by taking the expectation wrt to the unknown parameter. The determinant of the EIM, called the Expected Information Density (EID) is a measure of the goodness of being at a world position in determining the goal position. EID can then be used within the context of optimal control framework to synthesize control that would help in maximizing information gain regarding the unknown goal.
Equivalently, in our domain the unknown parameters is the goal itself. That is, the unknown parameter is **discrete**. There is no physical sensor attached to te robot that gives an estimate of the intended goal, instead the sensor itself is virtual and is a mathematical function that maps the user control command, robot position, goal position into a "sensor reading" between 0 and 1.
The uncertainty in the measurement model can be formulated in different ways:

1. The uncertainty is directly in the formulation itself. When the system designer chooses to use a particular form of confidence function, he/she is artifically creating a bias. Therefore, one way to create a "distribution" over confidence functions is to sample from different confidence functions. However, it is unclear that how many different formulations can be used and what they are. Furthermore, the distribution may be drastically different from Gaussian and might have a nonparametric form which would make the computation of Fisher information intractable.

2. Fix on a particular form of confidence function. Artificially make it noisy by

adding Gaussian white noise that has very low variance. However, confidence functions need to be within 0 and 1. Converting the confidence values into a Gaussian RV changes the range to -inf to +inf. Another option is to TRANSFORM the confidence function into a range (-inf, inf) and then model the transformed variable as a Gaussian distribution.

3. Model the confidence function as a beta distribution. This will impose bounds on the values confidence can take. Given the mode (the most likely value of confidence for a goal) and a user-specified variance, the shape parameters can possibly be computed. Fisher information can be computed for Beta distribution, although the result may not be as pretty as for Gaussian distributions.

Each one of these approaches involve some sort of intervention from the system designer. This will be inevitable.

**Math for computing Fisher Information**

**Measurement Model**
$c \in [0, 1]$ and $c \sim Beta(\alpha, \beta)$, where $\alpha$ and $\beta$ can be computed given the mode of the distribution $m_B$ and the user specified variance $\sigma_B$.

We assume that the "directedness based" confidence function. This is given by

$$m_B = c_g(\boldsymbol{x_r}, \boldsymbol{x_g}, \boldsymbol{u_h}) = \frac{(\boldsymbol{u_h} \cdot (\boldsymbol{x_g} - \boldsymbol{x_r}) + 1)}{2} = \frac{cos(\theta_0) + 1}{2}$$

Being in a control mode $m$ is equivalent to constraining what the value of $\boldsymbol{u_h}$ can be. So, if we are interested in characterizing the information density in for each control mode $m$ then it is equivalent to characterizing the ID for each $\boldsymbol{u_h}$.

In Todd's work, for *every* possible goal position (denoted by $\theta$) there is a distribution for what the voltage measurement could be. Analogously, for every possible intended goal the confidence measurement will have a distribution.

In Todd's work, the "unknown parameter", goal position is denoted as $\boldsymbol{\theta}$ and the "other" parameter on which the voltage measurement depends is denoted as $x$. Therefore, the measurement function is given as $\Upsilon(\boldsymbol{\theta}, x)$ and $p(v|\theta, x) = \mathcal{N}(\Upsilon(\boldsymbol{\theta}, x), \sigma^2)$. The Fisher information is then denoted as

$$\mathcal{I}(\boldsymbol{\theta}, x) = \left( \frac{\partial \Upsilon(\boldsymbol{\theta}, x)}{\partial \boldsymbol{\theta}} \cdot \frac{1}{\sigma} \right)^2$$

and EID is computed by marginalizing everything but $x$. That is,

$$\Phi(x) = \int_{\theta} \mathcal{I}(\boldsymbol{\theta}, x) p(\theta) d\theta$$

In our work, the "unknown parameter" is the intended goal itself and can be denoted by $\boldsymbol{g}$. For every intended goal position $\boldsymbol{g}_i$ there is a distribution for the "sensor" measurement (confidence). The "other" parameters on which the confidence depends on are $\boldsymbol{u_h}, \boldsymbol{x_r}$. The EID is on $\boldsymbol{u_h}$, therefore the marginalization will be on $\boldsymbol{x_r}$ and over the goals ($\boldsymbol{g}$).

For every goal position $\boldsymbol{x_g}$, for control command $\boldsymbol{u_h}$ (which is characteristic of the control mode $m$), and for robot position $\boldsymbol{x_r}$,

$$p(c_g|\boldsymbol{x_g}, \boldsymbol{u_h}, \boldsymbol{x_r}) = Beta(\alpha, \beta) \quad \text{with} \quad c_g \in [0, 1]$$

where the shape parameters $\alpha$ and $\beta$ are solved for given the mode and the user-defined variance (refer to my script BetaDistribution.m). The mode of the distribution is given by (for example, the directedness based confidence function)

$$m_B = \frac{(\boldsymbol{u_h} \cdot (\boldsymbol{x_g} - \boldsymbol{x_r}) + 1)}{2} = \frac{cos(\theta_0) + 1}{2}$$

The shape parameters $\alpha$ and $\beta$ depend on $\boldsymbol{x_g}$ and therefore will show up in the exponents and the Gamma functions in Beta distribution and this can complicate the computation of Fisher information wrt $\boldsymbol{x_g}$.

Another option is to transform $c \in [0, 1]$ to another variable whose range is $[-\infty, \infty]$. Such a transformation can be defined as

$$t(c) = \frac{c}{1-c} - \frac{1-c}{c}$$

The Fisher Information is then defined as

$$\mathcal{I}(\boldsymbol{x_g}, \boldsymbol{u_h}, \boldsymbol{x_r}) = \int\limits_{c_g} \left( \frac{\partial p(c_g|\boldsymbol{x_g}, \boldsymbol{u_h}, \boldsymbol{x_r})}{\partial \boldsymbol{x_g}} \right)^2 \frac{1}{p(c_g|\boldsymbol{x_g}, \boldsymbol{u_h}, \boldsymbol{x_r})} dc_g$$

**Note: The notion of a derivative of pdf wrt to discrete goals is ill-defined. We need to have a point wise estimate for the quantity of interest. However, for argument-sake we will continue to write this symbolically as above.**

Fisher information for beta distributions are described in `http://www.math.bas.bg/serdica/2004/2004-513-526.pdf`.

Marginalizing over $\boldsymbol{x_g}$ and $\boldsymbol{x_r}$(entire workspace) we will get Expected Information Density for wrt $\boldsymbol{u_h}$

$$\Phi(\boldsymbol{u_h}) = \int_{\boldsymbol{x_r}} \left[ \int_{\boldsymbol{x_g}} \left[ \int_{c_g} \left( \frac{\partial p(c_g|\boldsymbol{x_g}, \boldsymbol{u_h}, \boldsymbol{x_r})}{\partial \boldsymbol{x_g}} \right)^2 \frac{1}{p(c_g|\boldsymbol{x_g}, \boldsymbol{u_h}, \boldsymbol{x_r})} dc_g \right] p(\boldsymbol{x_g})\, d\boldsymbol{x_g} \right] p(\boldsymbol{x_r})\, d\boldsymbol{x_r}$$

**Note: The expectation integral wrt to $g$ will actually be computed as a weighted summation, since $p(\boldsymbol{x_g})$ is a a pmf. $p(\boldsymbol{x_r})$ is the prior over the robot's position. We can assume that the robot position is known perfectly in which case that outermost integral collapses to a single point.**

The issue of computing the derivative of the confidence distribution wrt the discrete goal parameter is agnostic to how we specify the confidence distribution.

## Some new thoughts

The trouble with the earlier formulation was the claim that the derivative that need to be computed for the Fisher Information computation is ill-defined. This is true if the "unknown parameter" is denoted as $\boldsymbol{g}$. However, if goals are represented using their positions wrt to a global frame, then the domain changes to $\boldsymbol{x_g}$. $\boldsymbol{x_g}$ lives in a continuous domain and a function of $\boldsymbol{x_g}$ is also continuous and therefore the derivative is well-defined.

In our case, eventually we are interested in the "Expected Information Density" as a function of $\boldsymbol{u_h}$ and as a result $\boldsymbol{x_r}$ and $\boldsymbol{x_g}$ marginalizes out. The probability density function $p(\boldsymbol{x_r})$ is a delta function since the position of the robot is known fully. The pdf $p(\boldsymbol{x_g})$ is only defined at discrete locations in the domain and therefore is a point-wise estimate.

### Math

### Miller et al

$$\mu_v = \Upsilon(\alpha, x) = \chi \frac{r^3 E(\alpha) \cdot (x - \alpha)}{\|x - \alpha\|^2}$$

However, the measurements are normalized to be always between [0,1]. Therefore $\mu_v \in [0, 1]$.

$$p(v|\alpha, x) = \mathcal{N}(\Upsilon(\alpha, x), \sigma^2)$$

### Our work

For simplicity sake assume that the world we are considering is the 2D position. Assuming directedness based confidence function, we can have

$$\mu_{c_g} = \Upsilon(\boldsymbol{x_g}, \boldsymbol{x_r}, \boldsymbol{u_h}) = \frac{1 + cos(\theta)}{2}, \quad \mu_{c_g} \in [0, 1]$$

where $\mu_{c_g} = 1$ indicates maximum confidence that the intended goal is $g$ and $\mu_{c_g} = 0$ indicates least confidence. In the above equation $cos(\theta)$ is given by

$$cos(\theta) = \frac{\boldsymbol{u_h} \cdot (\boldsymbol{x_g} - \boldsymbol{x_g})}{\|\boldsymbol{u_h}\| \cdot \|\boldsymbol{x_g} - \boldsymbol{x_r}\|}$$

The measurement model is then given by

$$c_g = \Upsilon(\boldsymbol{x_g}, \boldsymbol{x_r}, \boldsymbol{u_h}) + \delta, \quad \delta \sim \mathcal{N}(0, \sigma^2)$$

Note that $c_g$ is a continuous function and $\sigma^2$ is a pre-defined variance. Then,

$$p(c_g|\boldsymbol{x_g}, \boldsymbol{x_r}, \boldsymbol{u_h}) = \mathcal{N}(\Upsilon(\boldsymbol{x_g}, \boldsymbol{x_r}, \boldsymbol{u_h}), \sigma^2)$$

Since $c_g$ is distributed as a Gaussian RV, the Fisher information (the amount of information a "measurement" has regarding $\boldsymbol{x_g}$) is given by

$$\mathcal{I}(\boldsymbol{x_g}, \boldsymbol{u_h}, \boldsymbol{x_r}) = \left( \frac{\partial \Upsilon(\boldsymbol{x_g}, \boldsymbol{u_h}, \boldsymbol{x_r})}{\partial \boldsymbol{x_g}} \cdot \frac{1}{\sigma} \right)^2$$

Since $\boldsymbol{x_g}$ is a 2D parameter, the FI is a matrix and not a scalar.

$$\mathcal{I}_{i,j}(\boldsymbol{x_g}, \boldsymbol{u_h}, \boldsymbol{x_r}) = \frac{1}{\sigma^2}\frac{\partial^2 \Upsilon(\boldsymbol{x_g}, \boldsymbol{x_r}, \boldsymbol{u_h})}{\partial x_g^i x_g^j}$$

Marginalizing over $\boldsymbol{x_g}$

$$\Phi_{i,j}(\boldsymbol{u_h}, \boldsymbol{x_r}) = \frac{1}{\sigma^2}\int_{x_g^i}\int_{x_g^j}\frac{\partial^2 \Upsilon(\boldsymbol{x_g}, \boldsymbol{x_r}, \boldsymbol{u_h})}{\partial x_g^i x_g^j} p(x_g^i, x_g^j) dx_g^j dx_g^i$$

Marginalizing over $\boldsymbol{x_r}$

$$\Phi(\boldsymbol{u_h}) = \int_{\boldsymbol{x_r}}\left[\frac{1}{\sigma^2}\int_{x_g^i}\int_{x_g^j}\frac{\partial^2 \Upsilon(\boldsymbol{x_g}, \boldsymbol{x_r}, \boldsymbol{u_h})}{\partial x_g^i x_g^j} p(x_g^i, x_g^j) dx_g^j dx_g^i\right] p(\boldsymbol{x_r}) d\boldsymbol{x_r}$$

The pdf's are given by

$$p(\boldsymbol{x_r}) = \begin{cases} 1 & \boldsymbol{x_r} = \boldsymbol{x_{true}} \\ 0 & otherwise \end{cases}$$

and

$$p(\boldsymbol{x_g}) = \begin{cases} p_i & for\ \boldsymbol{x_g} = \boldsymbol{x_{g_i}} \\ 0 & otherwise \end{cases}$$

such that $\sum_{i=1}^{n_g} p_i = 1$