Cooperative Multi-Robot Localization under Communication Constraints

Nikolas Trawny and Stergios I. Roumeliotis
Department of Computer Science & Engineering
University of Minnesota

Multiple Autonomous
Robotic Systems Laboratory

Technical Report
Number 2008-001, Rev. 34
May 2008

Dept. of Computer Science & Engineering
University of Minnesota
4-192 EE/CS Building
200 Union St. S.E.
Minneapolis, MN 55455
Tel: (612) 625-2217
Fax: (612) 625-0572
URL: http://www.cs.umn.edu/~trawny

Cooperative Multi-Robot Localization under Communication Constraints

Nikolas Trawny and Stergios I. Roumeliotis
Department of Computer Science & Engineering
University of Minnesota

Multiple Autonomous Robotic Systems Laboratory, TR-2008-001, Rev. 34
May 2008

1 Problem formulation and Modeling

In this techreport, we will derive algorithms for quantized batch estimation for linear systems with Gaussian noise. We will start by defining the system model and the problem of maximum a posteriori (MAP) estimation.

Let us assume a discrete-time, linear dynamic system with scalar measurements

$$\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{w}_{k-1} \tag{1}$$

$$z_k = \mathbf{h}_k^T\mathbf{x}_k + v_k \tag{2}$$

with initial condition

$$\mathbf{x}_0 \sim \mathcal{N}(\mathbf{x}(0), \mathbf{P}_0) \tag{3}$$

and zero-mean, white Gaussian noise

$$E[\mathbf{w}_k] = E[v_k] = 0, \quad E[\mathbf{w}_k\mathbf{w}_l^T] = \delta_{kl}\mathbf{Q}_k, \quad E[v_k v_l^T] = \delta_{kl}\sigma_k^2, \quad E[v_k\mathbf{w}_l^T] = 0 \tag{4}$$

For Bayesian estimation, the estimator is a function of the posterior pdf $p(\mathbf{x}|\mathbf{z})$, i.e., the pdf of the state conditioned on the measurements.

The Kalman Filter recursively computes the MMSE estimate, i.e. the conditional *mean* or expected value of the posterior pdf

$$\hat{\mathbf{x}}_{\text{MMSE}} = E[\mathbf{x}|\mathbf{z}] \tag{5}$$

The maximum a posteriori, or MAP estimator, however, computes its estimate as the *maximum* or the mode of the posterior pdf

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg\max p(\mathbf{x}|\mathbf{z}) \tag{6}$$

In a random process with linear system and measurement model and Gaussian noise, the posterior pdf is known to be Gaussian. For Gaussian pdfs, the mean and the mode coincide, making both estimates equivalent.

Once nonlinearities enter the system, either through nonlinear system or measurement models, or, as in our case, through quantization, the posterior pdf is not Gaussian anymore, and the two estimates are generally different.

1.1 MAP estimation

In contrast to filtering, the batch estimate is the best estimate of the *entire* state history (denoted as $\mathbf{x}_{0:K}$) given *all* measurements up to time K (denoted as $\mathbf{z}_{0:K}$). Using Bayes' rule, we can write the batch MAP estimate as

$$\begin{aligned}
\hat{\mathbf{x}}_{0:K} &= \arg\max p(\mathbf{x}_{0:K}|\mathbf{z}_{0:K}) && (7)\\
&= \arg\max \frac{1}{p(\mathbf{z}_{0:K})} p(\mathbf{z}_{0:K}|\mathbf{x}_{0:K}) p(\mathbf{x}_{0:K}) && (8)\\
&= \arg\max \frac{1}{p(\mathbf{z}_{0:K})} \prod_{k=0}^{K} p(z_k|\mathbf{x}_k) \prod_{k=0}^{K-1} p(\mathbf{x}_{k+1}|\mathbf{x}_k) \cdot p(\mathbf{x}_0) && (9)\\
&= \arg\min \left(-\sum_{k=0}^{K} \log p(z_k|\mathbf{x}_k) - \sum_{k=0}^{K-1} \log p(\mathbf{x}_{k+1}|\mathbf{x}_k) - \log p(\mathbf{x}_0) \right) && (10)
\end{aligned}$$

The above derivation relies on the Gauss-Markov property of the dynamic system, which renders measurements and states at different times conditionally independent, and on the monotonicity of the logarithm. Also, the leading term $\frac{1}{p(\mathbf{z}_{0:K})}$ is considered a normalizing constant independent of $\mathbf{x}$, which can be neglected during optimization.

For the linear system with Gaussian noise (1)-(4), the conditional pdfs in (10) are also Gaussian, and their logarithm is a quadratic form (up to a constant), so that the MAP estimate can be formulated as a weighted Least Squares (WLS) problem.

$$\hat{\mathbf{x}}_{0:K} = \arg\min \left(\sum_{k=0}^{K} \frac{1}{2} \|z_k - \mathbf{h}_k^T \mathbf{x}_k\|^2_{\sigma_k^{-2}} + \sum_{k=0}^{K-1} \frac{1}{2} \|\mathbf{x}_{k+1} - \mathbf{F}_k \mathbf{x}_k\|^2_{(\mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^T)^{-1}} + \frac{1}{2} \|\mathbf{x}_0 - \mathbf{x}(0)\|^2_{\mathbf{P}_0^{-1}} \right) \quad (11)$$

The notation $\|\mathbf{x}\|_{\mathbf{W}}$ denotes the weighted L_2-norm, which can be converted to a regular L_2-norm by pre-whitening

$$\|\mathbf{x}\|_{\mathbf{W}} = \sqrt{\mathbf{x}^T \mathbf{W} \mathbf{x}} = \|\mathbf{W}^{1/2} \mathbf{x}\|_2 \quad (12)$$

where $\mathbf{W}^{1/2}$ denotes the matrix square-root so that $(\mathbf{W}^{1/2})^T \mathbf{W}^{1/2} = \mathbf{W}$.

Following the appropriate pre-whitening step, we can stack the summands of (11) to form a large regular Least Squares system with the closed-form solution given by the normal equations

$$\begin{aligned}
\hat{\mathbf{x}}_{0:K} &= \arg\min \frac{1}{2} \|\mathbf{A}\mathbf{x}_{0:K} - \mathbf{b}\|_2^2 && (13)\\
&= (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} && (14)
\end{aligned}$$

where $\mathbf{A}$, $\mathbf{b}$ are functions of the quantities $\mathbf{F}_k, \mathbf{G}_k, \mathbf{Q}_k, z_k, \mathbf{h}_k, \sigma_k^2, \mathbf{x}(0)$, and $\mathbf{P}_0$. The covariance $\mathbf{P}$ of this estimate is given by $\mathbf{P} = (\mathbf{A}^T \mathbf{A})^{-1}$.

We will solve the WLS problem using QR-decomposition, which according to Golub and van Loan [1] is numerically more accurate than the method of normal equations for problems with small residual $\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}$. Let $\mathbf{A} = \mathbf{Q}\mathbf{R}$ denote the thin QR decomposition [1] of $\mathbf{A}$, such that $\mathbf{Q}$ has the same dimension as $\mathbf{A}$ and $\mathbf{R}$ is a square, upper triangular matrix. Then, $\|\mathbf{A}\mathbf{x}_{0:K} - \mathbf{b}\|_2^2 = \|\mathbf{R}\mathbf{x}_{0:K} - \mathbf{Q}^T\mathbf{b}\|_2^2 + \|\mathbf{d}\|_2^2$, where $\mathbf{d}$ is a constant independent of $\mathbf{x}_{0:K}$, and we can efficiently compute the WLS estimate by solving $\mathbf{R}\hat{\mathbf{x}}_{0:K} = \mathbf{Q}^T\mathbf{b}$ via backsubstitution. The matrix $\mathbf{R}$ is the Cholesky factor of the information matrix (the inverse covariance), hence $\mathbf{P} = (\mathbf{R}^T\mathbf{R})^{-1}$.

For nonlinear system or measurement models, we linearize the above system around its current estimate. The optimization problem is then solved using iterative WLS, with the p-th iteration given by

$$\begin{aligned}
\boldsymbol{\delta}\mathbf{x}^* &= \arg\min \frac{1}{2} \|\mathbf{A}(\hat{\mathbf{x}}_{0:K}^{(p)}) \boldsymbol{\delta}\mathbf{x} - \mathbf{b}(\hat{\mathbf{x}}_{0:K}^{(p)})\|_2^2 && (15)\\
\hat{\mathbf{x}}_{0:K}^{(p+1)} &= \hat{\mathbf{x}}_{0:K}^{(p)} + \boldsymbol{\delta}\mathbf{x}^* && (16)
\end{aligned}$$

The notation is chosen to explicitly indicate the dependence of $\mathbf{A}$ and $\mathbf{b}$ on the current linearization point $\hat{\mathbf{x}}_{0:K}^{(p)}$.

2 Single-bit quantized MAP estimation (QMAP)

We will first analyze the case where each measurement is quantized by exactly one bit. In particular, we will use the following quantization rule

$$b_k = \begin{cases} +1 & \text{if } z_k - \mathbf{h}_k^T \mathbf{x}_{\tau_k} > 0 \\ -1 & \text{else} \end{cases} \tag{17}$$

The quantity $\mathbf{x}_{\tau_k}$ is a design parameter, which we will for now choose to equal the currently available state estimate just before the quantized measurement b_k is computed, similarly as in [2]. Choice of $\mathbf{x}_{\tau_k}$ has an important impact on performance [3]. Although the actual quantization *threshold* which we compare our measurement against is the scalar $\mathbf{h}_k^T \mathbf{x}_{\tau_k}$, in the following we will for convenience refer to $\mathbf{x}_{\tau_k}$ as the threshold and design parameter. Since $\mathbf{h}_k$ is fixed, specifying $\mathbf{x}_{\tau_k}$ uniquely determines the scalar quantization threshold. This threshold can be determined by all sensors independently of the analog measurement. In what follows, we compute the conditional probability $p(b_k|\mathbf{x}_k)$. Using properties (176) and (177) of the Q-Function, this conditional probability can be written as

$$p(b_k = 1|\mathbf{x}_k) = \Pr\left\{z_k - \mathbf{h}_k^T \mathbf{x}_{\tau_k} > 0 \,|\mathbf{x}_k\right\} = \Pr\left\{v_k > -\mathbf{h}_k^T(\mathbf{x}_k - \mathbf{x}_{\tau_k}) \,|\mathbf{x}_k\right\} \tag{18}$$

$$= Q\left(\frac{-\mathbf{h}_k^T(\mathbf{x}_k - \mathbf{x}_{\tau_k}) - 0}{\sigma_k}\right) = Q\left(\frac{-b_k \mathbf{h}_k^T(\mathbf{x}_k - \mathbf{x}_{\tau_k})}{\sigma_k}\right) \tag{19}$$

$$p(b_k = -1|\mathbf{x}_k) = \Pr\left\{z_k - \mathbf{h}_k^T \mathbf{x}_{\tau_k} \leq 0 \,|\mathbf{x}_k\right\} = \Pr\left\{v_k \leq -\mathbf{h}_k^T(\mathbf{x}_k - \mathbf{x}_{\tau_k}) \,|x\right\} \tag{20}$$

$$= Q\left(-\frac{-\mathbf{h}_k^T(\mathbf{x}_k - \mathbf{x}_{\tau_k}) - 0}{\sigma_k}\right) = Q\left(\frac{-b_k \mathbf{h}_k^T(\mathbf{x}_k - \mathbf{x}_{\tau_k})}{\sigma_k}\right) \tag{21}$$

In order to find the MAP estimator, we replace the terms $p(z_k|\mathbf{x}_k)$ in the analog posterior pdf (11) by the expressions for $p(b_k|\mathbf{x}_k)$ to obtain the single-bit MAP posterior pdf

Proposition 1. Single-Bit Quantized MAP (QMAP). *Assume the linear model of* (1)-(4)*. If a single bit is allocated per measurement, with a quantization rule as in* (17)*, the posterior pdf is given by*

$$p(\mathbf{x}_{0:K}|\mathbf{b}_{0:K}) = \frac{1}{p(\mathbf{b}_{0:K})} \prod_{k=0}^{K} p(b_k|\mathbf{x}_k) \prod_{k=0}^{K-1} p(\mathbf{x}_{k+1}|\mathbf{x}_k) \cdot p(\mathbf{x}_0) \tag{22}$$

$$= \frac{1}{p(\mathbf{b}_{0:K})} \prod_{k=0}^{K} Q\left(\frac{-b_k \mathbf{h}_k^T(\mathbf{x}_k - \mathbf{x}_{\tau_k})}{\sigma_k}\right) \prod_{k=0}^{K-1} \mathcal{N}(\mathbf{F}_k \mathbf{x}_k, \mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^T) \cdot p(\mathbf{x}_0) \tag{23}$$

where $Q(.)$ denotes the Gaussian tail probability. The MAP estimate is given by

$$\hat{\mathbf{x}}_{0:K} = \arg\max p(\mathbf{x}_{0:K}|\mathbf{b}_{0:K}) \tag{24}$$

$$= \arg\min -\sum_{k=0}^{K} Q\left(\frac{-b_k \mathbf{h}_k^T(\mathbf{x}_k - \mathbf{x}_{\tau_k})}{\sigma_k}\right) + \sum_{k=0}^{K-1} \frac{1}{2}\|\mathbf{x}_{k+1} - \mathbf{F}_k \mathbf{x}_k\|^2_{(\mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^T)^{-1}} + \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}(0)\|^2_{\mathbf{P}_0^{-1}} \tag{25}$$

An important consequence is the following Lemma

Lemma 1. *The posterior pdf of the QMAP given in* (23) *is log-concave in* $\mathbf{x}$*.*

Proof. Both the Gaussian pdf and the Q-function are log-concave [4, Ex. 3.39], log-concavity is closed under multiplication, and is preserved under linear transformation of the argument. □

Due to this log-concavity, finding the MAP estimate is a nonlinear but convex optimization problem, ensuring that the MAP estimate is unique for every choice of $\mathbf{x}_{\tau_k}$. Efficient methods, such as sequential quadratic programming, Newton methods or interior point methods are therefore guaranteed to converge to the global optimum.

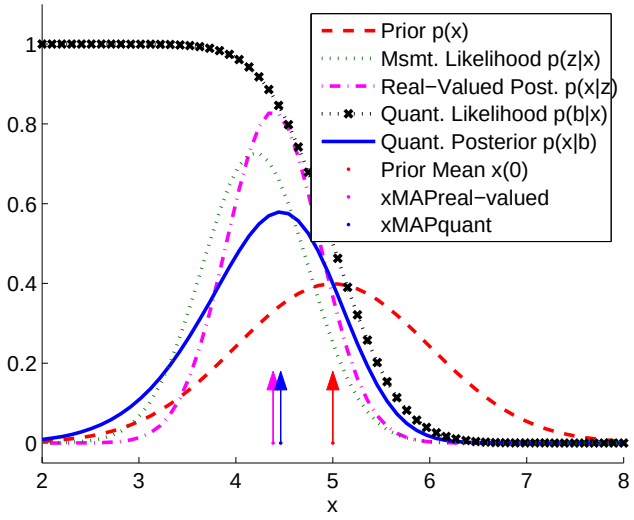


Figure 1: Comparison of real-valued and quantized measurement-based posterior pdfs for scalar prior and one scalar measurement $z = x + v$, $x \sim \mathcal{N}(5, 1)$, $v \sim \mathcal{N}(0, 0.3)$ quantized by a single bit. The posterior for the quantized measurement is skewed and heavy-tailed, compared to the real-valued (Gaussian) posterior.

Example: To illustrate the difference between real-valued and single-bit MAP, let us consider the simple case of scalar prior plus one scalar measurement $z = x + v,\ v \sim \mathcal{N}(0, \sigma^2)$ quantized by one bit.

$$\hat{x}_r = \arg\min \left(\frac{1}{2} \|z - x\|^2_{\sigma^{-2}} + \frac{1}{2} \|x - x_0\|^2_{\mathbf{P}_0^{-1}} \right) \tag{26}$$

$$\hat{x}_q = \arg\min \left(-\log Q \left(\frac{-b(x - x_\tau)}{\sigma} \right) + \frac{1}{2} \|x - x_0\|^2_{\mathbf{P}_0^{-1}} \right) \tag{27}$$

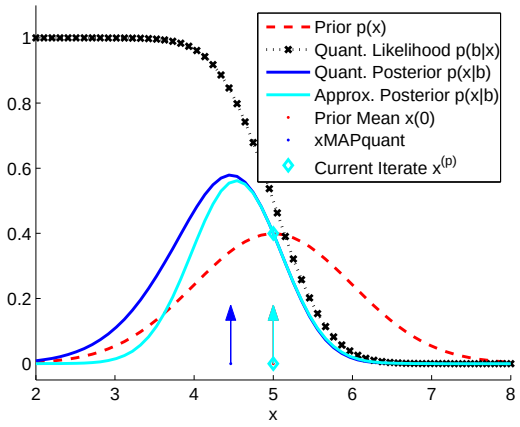
with the threshold $x_\tau = x_0$ chosen as discussed above. The posteriors are shown in Fig. 1. While the analog posterior is Gaussian, the quantized posterior is skewed and heavy-tailed. However, its log-concavity ensures that the MAP estimate can be found by sequential quadratic programming, where the negative logarithm of the posterior is approximated by its second-order Taylor series expansion. This iterative process is shown in Fig. 2.

2.1 Newton update

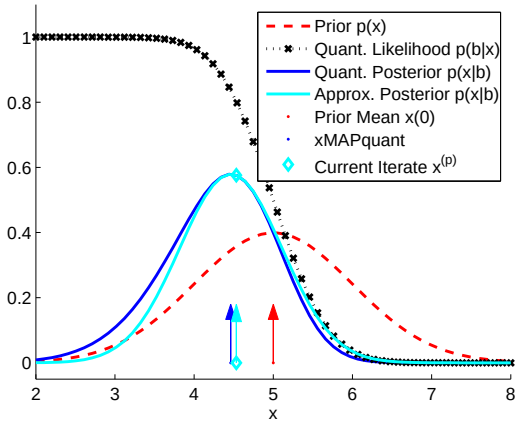
In what follows, we will present the derivation of the Newton step for the single-bit MAP cost function (25). In general, the p-th Newton iteration is given by

$$\boldsymbol{\delta}\mathbf{x} = -(\nabla^2 f(\hat{\mathbf{x}}^{(p)}))^{-1} \cdot \nabla f(\hat{\mathbf{x}}^{(p)}) \tag{28}$$

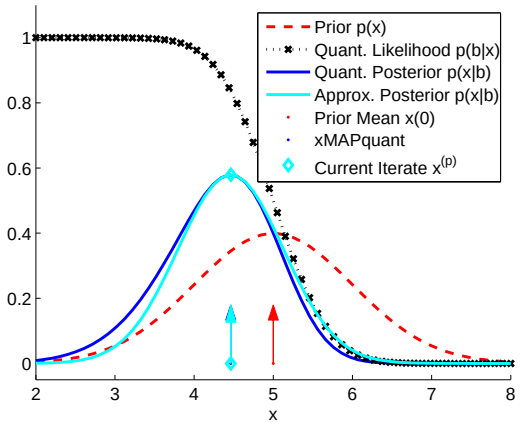
$$\hat{\mathbf{x}}^{(p+1)} = \hat{\mathbf{x}}^{(p)} + \boldsymbol{\delta}\mathbf{x} \tag{29}$$



(a) Initial approximation. $x_k = x(0)$



(b) After first iteration.



(c) After second iteration.

Figure 2: Sequential quadratic approximation of the negative log-likelihood yields an approximate posterior distribution and the MAP estimate. Only three iterations are shown, since afterwards the estimates become visually indistinguishable.

In order to facilitate notation, we will define the quantities

$$\mathbf{x}_{0:K} = \begin{bmatrix}\mathbf{x}_0^T & \mathbf{x}_1^T & \dots & \mathbf{x}_K^T\end{bmatrix}^T \quad \text{the stacked state vector} \tag{30}$$

$$\mathfrak{x}_0 = \begin{bmatrix}\mathbf{x}(0)^T & 0 & \dots & 0\end{bmatrix}^T \quad \text{the zero-padded initial condition} \tag{31}$$

$$\mathfrak{x}_{\tau_k} = \begin{bmatrix}0 & \dots & 0 & x_{\tau_k}^T & 0 & \dots & 0\end{bmatrix}^T \quad \text{the zero-padded quantization thresholds} \tag{32}$$

$$\mathfrak{F}_k = \begin{bmatrix}0 & \dots & 0 & -\mathbf{F}_k & \mathbf{I} & 0 & \dots & 0\end{bmatrix} \quad \text{so that } \mathbf{F}_k\mathbf{x}_{0:K} = \mathbf{x}_{k+1} - \mathbf{F}_k\mathbf{x}_k \tag{33}$$

$$\mathfrak{F}_{-1} = \begin{bmatrix}\mathbf{I} & 0 & \dots & 0\end{bmatrix} \quad \text{so that } \mathfrak{F}_{-1}\mathbf{x}_{0:K} = \mathbf{x}_0 \tag{34}$$

$$\mathfrak{h}_k^T = \begin{bmatrix}0 & \dots & 0 & \mathbf{h}_k^T & 0 & \dots & 0\end{bmatrix} \quad \text{so that } \mathfrak{h}_k^T\mathbf{x}_{0:K} = \mathbf{h}_k^T\mathbf{x}_k \tag{35}$$

$$u_k = -\frac{b_k}{\sigma_k}\mathfrak{h}_k^T(\hat{\mathbf{x}}^{(p)} - \mathfrak{x}_{\tau_k}) \tag{36}$$

The gradient and hessian of the cost function (25) are now computed as (cf. Appendix B)

$$\nabla f(\hat{\mathbf{x}}^{(p)}) = -\sum \frac{\mathcal{N}(u_k,1)}{Q(u_k)} \cdot \frac{b_k}{\sigma_k}\mathfrak{h}_k + \sum \mathfrak{F}_k^T(\mathbf{G}_k\mathbf{Q}_k\mathbf{G}_k^T)^{-1}\mathfrak{F}_k\hat{\mathbf{x}}^{(p)} + \mathfrak{F}_{-1}^T\mathbf{P}_0^{-1}\mathfrak{F}_{-1}(\hat{\mathbf{x}}^{(p)} - \mathfrak{x}_0) \tag{37}$$

$$\nabla^2 f(\hat{\mathbf{x}}^{(p)}) = \sum \frac{1}{\sigma_k^2}\frac{-u_k\mathcal{N}(u_k,1)\cdot Q(u_k) + \mathcal{N}^2(u_k,1)}{Q^2(u_k)}\mathfrak{h}_k\mathfrak{h}_k^T + \sum \mathfrak{F}_k^T(\mathbf{G}_k\mathbf{Q}_k\mathbf{G}_k^T)^{-1}\mathfrak{F}_k + \mathfrak{F}_{-1}^T\mathbf{P}_0^{-1}\mathfrak{F}_{-1} \tag{38}$$

2.2 Measurement matrix form

We can also use the second order Taylor series approximation to bring the cost function in a form reminiscent of iterative WLS, amenable to solving via QR-decomposition, as follows. Starting from the original cost function

$$\hat{\mathbf{x}} = \arg\min -\sum \log Q\left(-\frac{b_k}{\sigma_k}\mathfrak{h}_k^T(\mathbf{x}_{0:K} - \mathfrak{x}_{\tau_k})\right) + \sum \frac{1}{2}\|\mathfrak{F}_k\mathbf{x}_{0:K}\|^2_{(\mathbf{G}_k\mathbf{Q}_k\mathbf{G}_k^T)^{-1}} + \frac{1}{2}\|\mathfrak{F}_{-1}(\mathbf{x}_{0:K} - \mathfrak{x}_0)\|^2_{\mathbf{P}_0^{-1}} \tag{39}$$

we obtain the WLS iteration by considering a second order Taylor series expansion around the current estimate $\hat{\mathbf{x}}^{(p)}$, noting that $\mathbf{x}_{0:K} = \hat{\mathbf{x}}^{(p)} + \boldsymbol{\delta}\mathbf{x}$. Since the second order Taylor series expansion will result in a quadratic form, we can write it (up to a constant, which we can neglect for optimization purposes) as

$$\boldsymbol{\delta}\mathbf{x}^* = \arg\min \sum \frac{1}{2}||\mathbf{c}_k^T(\hat{\mathbf{x}}^{(p)})\boldsymbol{\delta}\mathbf{x} - d_k(\hat{\mathbf{x}}^{(p)})||_2^2 + \sum \frac{1}{2}\|\mathfrak{F}_k\boldsymbol{\delta}\mathbf{x} - \mathfrak{F}_k(-\hat{\mathbf{x}}^{(p)})\|^2_{(\mathbf{G}_k\mathbf{Q}_k\mathbf{G}_k^T)^{-1}} + \frac{1}{2}\|\mathfrak{F}_{-1}\boldsymbol{\delta}\mathbf{x} - \mathfrak{F}_{-1}(\mathfrak{x}_0 - \hat{\mathbf{x}}^{(p)})\|^2_{\mathbf{P}_0^{-1}} \tag{40}$$

In order to find the vector $\mathbf{c}_k(\hat{\mathbf{x}}^{(p)})$ and scalar $d_k(\hat{\mathbf{x}}^{(p)})$, we expand the above 2-norm and compare coefficients with the second order Taylor series expansion of $Q\left(-\frac{b_k}{\sigma_k}\mathfrak{h}_k^T(\mathbf{x}_{0:K} - \mathfrak{x}_{\tau_k})\right)$. We start by

$$\frac{1}{2}||\mathbf{c}_k^T(\hat{\mathbf{x}}^{(p)})\boldsymbol{\delta}\mathbf{x} - d_k(\hat{\mathbf{x}}^{(p)})||_2^2 = \frac{1}{2}\boldsymbol{\delta}\mathbf{x}^T c_k(\hat{\mathbf{x}}^{(p)})c_k(\hat{\mathbf{x}}^{(p)})^T\boldsymbol{\delta}\mathbf{x} - d_k(\hat{\mathbf{x}}^{(p)})c_k(\hat{\mathbf{x}}^{(p)})^T\boldsymbol{\delta}\mathbf{x} + \frac{1}{2}d_k(\hat{\mathbf{x}}^{(p)})^2 \tag{41}$$

On the other hand,

$$\begin{aligned}-\log Q\left(-\frac{b_k}{\sigma_k}\mathfrak{h}_k^T(\mathbf{x}_{0:K} - \mathfrak{x}_{\tau_k})\right) \approx &-\log Q(u_k) - \frac{\mathcal{N}(u_k,1)}{Q(u_k)}\cdot\frac{b_k}{\sigma_k}\mathfrak{h}_k^T\boldsymbol{\delta}\mathbf{x} \\ &+ \frac{1}{2\sigma_k^2}\frac{-u_k\mathcal{N}(u_k,1)\cdot Q(u_k) + \mathcal{N}^2(u_k,1)}{Q^2(u_k)}\boldsymbol{\delta}\mathbf{x}^T\mathfrak{h}_k\mathfrak{h}_k^T\boldsymbol{\delta}\mathbf{x}\end{aligned} \tag{42}$$

We find that

$$\mathbf{c}_k(\hat{\mathbf{x}}^{(p)}) = \frac{1}{\sigma_k}\frac{\sqrt{-u_k\mathcal{N}(u_k,1)\cdot Q(u_k) + \mathcal{N}^2(u_k,1)}}{Q(u_k)}\mathfrak{h}_k \tag{43}$$

and

$$-d_k(\hat{\mathbf{x}}^{(p)})\mathbf{c}_k(\hat{\mathbf{x}}^{(p)}) = -\frac{\mathcal{N}(u_k,1)}{Q(u_k)} \cdot \frac{b_k}{\sigma_k}\mathfrak{h}_k \tag{44}$$

$$\Leftrightarrow d_k(\hat{\mathbf{x}}^{(p)}) = \frac{\mathcal{N}(u_k,1)}{Q(u_k)} \cdot \frac{b_k}{\sigma_k} \cdot \frac{\sigma_k \cdot Q(u_k)}{\sqrt{-u_k\mathcal{N}(u_k,1)\cdot Q(u_k) + \mathcal{N}^2(u_k,1)}} \tag{45}$$

$$= \frac{b_k \cdot \mathcal{N}(u_k,1)}{\sqrt{-u_k\mathcal{N}(u_k,1)\cdot Q(u_k) + \mathcal{N}^2(u_k,1)}} \tag{46}$$

Note that for this choice of parameters, the value of the additive constant is not equal, i.e.,

$$\frac{1}{2}d_k(\hat{\mathbf{x}}^{(p)})^2 \neq -\log Q(u_k) \tag{47}$$

but this is irrelevant for optimization.

2.3 Approximations for small and large arguments

These expressions for $\mathbf{c}_k(\hat{\mathbf{x}}^{(p)})$ and $d_k(\hat{\mathbf{x}}^{(p)})$ suffer from numerical instability for larger values of $|u_k|$, even for the range of values regularly appearing during normal optimization. It is therefore necessary to develop asymptotic expressions. Fig. 3 shows the scalar coefficients appearing in $\mathbf{c}_k(\hat{\mathbf{x}}^{(p)})$ and $d_k(\hat{\mathbf{x}}^{(p)})$ as functions of the argument u_k and the impact of approximation.

Consider the identity

$$\mathbf{c}_k(\hat{\mathbf{x}}^{(p)}) = \frac{1}{\sigma_k}\frac{\sqrt{-u_k\mathcal{N}(u_k,1)\cdot Q(u_k) + \mathcal{N}^2(u_k,1)}}{Q(u_k)}\mathfrak{h}_k \tag{48}$$

$$= \frac{1}{\sigma_k}\sqrt{\nabla_{u_k}\Big(\frac{\mathcal{N}(u_k,1)}{Q(u_k)}\Big)}\mathfrak{h}_k \tag{49}$$

We note that

$$\lim_{u_k\to\infty}\frac{\mathcal{N}(u_k,1)}{Q(u_k)} \tag{50}$$

$$= \lim_{u_k\to\infty}\frac{-u_k\mathcal{N}(u_k,1)}{-\mathcal{N}(u_k,1)} = u_k \tag{51}$$

by l'Hospital's rule, and further, that $\sqrt{\nabla_{u_k}u_k} = \sqrt{1} = 1$. Also,

$$\lim_{u_k\to-\infty}\frac{\mathcal{N}(u_k,1)}{Q(u_k)} = \frac{0}{1} = 0 \tag{52}$$

and therefore $\sqrt{\nabla_{u_k}0} = \sqrt{0} = 0$.

For $d_k(\hat{\mathbf{x}}^{(p)})$, we write

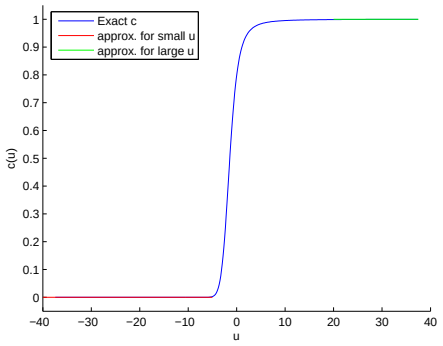
$$d_k(\hat{\mathbf{x}}^{(p)}) = \frac{b_k \cdot \mathcal{N}(u_k,1)}{\sqrt{-u_k\mathcal{N}(u_k,1)\cdot Q(u_k) + \mathcal{N}^2(u_k,1)}} \tag{53}$$

$$= b_k \cdot \sqrt{\frac{\mathcal{N}(u_k,1)}{\mathcal{N}(u_k,1) - u_kQ(u_k)}} \tag{54}$$

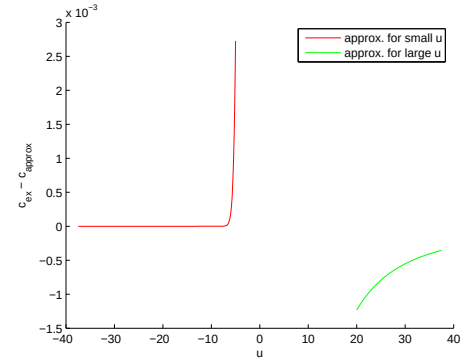
$$= b_k \cdot \sqrt{\frac{\frac{\mathcal{N}(u_k,1)}{u_k}}{\frac{\mathcal{N}(u_k,1)}{u_k} - Q(u_k)}} \tag{55}$$

We see that

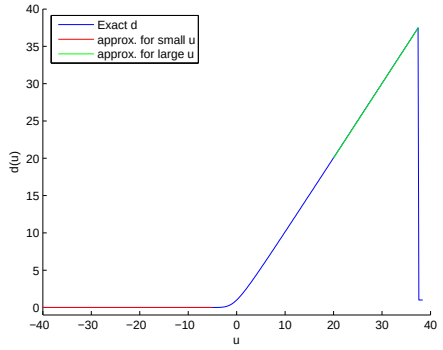
$$\lim_{u_k\to-\infty} d_k(\hat{\mathbf{x}}^{(p)}) = b_k\sqrt{\frac{0}{0-1}} = 0 \tag{56}$$



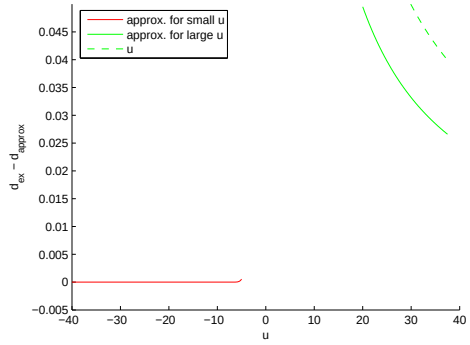
(a) Evolution of the coefficient of $\mathbf{c}_k(\hat{\mathbf{x}}^{(p)})$, $\sqrt{\nabla_{u_k}\left(\frac{\mathcal{N}(u_k,1)}{Q(u_k)}\right)}$, and its small and large approximations



(b) Approximation error for the coefficient of $c_k(\hat{\mathbf{x}}^{(p)})$



(c) Evolution of the coefficient of $d_k(\hat{\mathbf{x}}^{(p)})$, $\sqrt{\frac{\mathcal{N}(u_k,1)}{\mathcal{N}(u_k,1)-u_kQ(u_k)}}$, and its small and large approximations. Notice how the numerical instability causes erroneous values for $u_k > 37$.



(d) Approximation error for the coefficient of $d_k(\hat{\mathbf{x}}^{(p)})$. Notice that the approximation $d_k(\hat{\mathbf{x}}^{(p)}) = b_k u_k$ is slightly worse than $b_k\sqrt{1+u_k^2}$

Figure 3: The coefficients $c_k(\hat{\mathbf{x}}^{(p)})$ and $d_k(\hat{\mathbf{x}}^{(p)})$ in the measurement matrix form, and their approximations.

On the other hand, l'Hospital's rule helps us to write

$$\lim_{u_k\to\infty} d_k(\hat{\mathbf{x}}^{(p)}) = b_k \cdot \sqrt{\frac{\mathcal{N}(u_k,1)(-1-\frac{1}{u_k^2})}{-\mathcal{N}(u_k,1)\frac{1}{u_k^2}}} = b_k\sqrt{1+u_k^2} \approx b_k u_k \tag{57}$$

In conclusion, we will use the following approximations

$$\mathbf{c}_k(\hat{\mathbf{x}}^{(p)}) \approx \begin{cases} \frac{\mathfrak{h}_k}{\sigma_k} & \text{if } u_k \geq 37 \quad (\mathbf{c}_k(\hat{\mathbf{x}}^{(p)})(37) - \mathbf{c}_k(\hat{\mathbf{x}}^{(p)})_{\text{approx}} = -3.6\cdot 10^{-4}\cdot\frac{\mathfrak{h}_k}{\sigma_k}) \\ 0 & \text{if } u_k \leq -8 \quad (c_k(\hat{\mathbf{x}}^{(p)})(-8) - c_k(\hat{\mathbf{x}}^{(p)})_{\text{approx}} = 2.0\cdot 10^{-7}\cdot\frac{\mathfrak{h}_k}{\sigma_k}) \end{cases} \tag{58}$$

$$d_k(\hat{\mathbf{x}}^{(p)}) \approx \begin{cases} b_k(u_k) = -\frac{\mathfrak{h}_k^T(\hat{\mathbf{x}}^{(p)}-\mathfrak{x}_{\tau_k})}{\sigma_k} & \text{if } u_k \geq 37 \quad (d_k(\hat{\mathbf{x}}^{(p)})(37) - d_k(\hat{\mathbf{x}}^{(p)})_{\text{approx}} = 0.0405) \\ 0 & \text{if } u_k \leq -8 \quad (d_k(\hat{\mathbf{x}}^{(p)})(-8) - d_k(\hat{\mathbf{x}}^{(p)})_{\text{approx}} = 2.5\cdot 10^{-8}) \end{cases} \tag{59}$$

3 Batch-quantized multibit MAP

In this section, we assume we can transmit not one, but m bits per measurement. m bits enumerate 2^m non-overlapping partitions of $\mathbb{R}$, and the transmitted bit sequence specifies the particular interval the current measurement residual

2^m	$\Delta(2)$	$\Delta(3)$
2	0	∞

2^m	$\Delta(3)$	$\Delta(4)$	$\Delta(5)$
4	0	0.982	∞

2^m	$\Delta(5)$	$\Delta(6)$	$\Delta(7)$	$\Delta(8)$	$\Delta(9)$
8	0	0.501	1.050	1.748	∞

2^m	$\Delta(9)$	$\Delta(10)$	$\Delta(11)$	$\Delta(12)$	$\Delta(13)$	$\Delta(14)$	$\Delta(15)$	$\Delta(16)$	$\Delta(17)$
16	0	0.258	0.522	0.800	1.099	1.437	1.844	2.401	∞

Table 1: Quantization Thresholds for Gaussian PDF, $\Delta(i) = -\Delta(2^m + 2 - i), \forall i \in \{1, 2, \ldots, 2^m\}$

belongs to. The quantization rule we choose is

$$b_k = i \quad \text{if } \tau(i) < z_k - \mathbf{h}_k^T \mathbf{x}_{\tau_k} \leq \tau(i+1), \quad i = 1 \ldots 2^m \tag{60}$$

and b_k is transmitted in its binary representation.

The thresholds $\tau(i)$ are defined as

$$\tau(i) = \Delta(i) \cdot \sqrt{\mathbf{h}_k^T \mathbf{P} \mathbf{h}_k + \sigma^2} \tag{61}$$

where $\mathbf{P}$ denotes the covariance of the estimate which we choose as threshold x_{τ_k} at the time of quantization.

The values $\Delta(i)$ are taken from [5], and correspond to the thresholds of the LLoyd-Max quantizer [6, 7] for quantization with minimum MSE distortion.

For this quantization rule, the conditional probability $p(b = i|\mathbf{x}_k)$ is given by

$$p(b_k = i|\mathbf{x}_k) = \Pr\left\{\tau(i) < \mathbf{h}_k^T(\mathbf{x}_k - x_{\tau_k}) + n_k \leq \tau(i+1)|\mathbf{x}_k\right\} \tag{62}$$

$$= Q\left(\frac{\tau(i) - \mathbf{h}_k^T(\mathbf{x}_k - \mathbf{x}_{\tau_k})}{\sigma_k}\right) - Q\left(\frac{\tau(i+1) - \mathbf{h}_k^T(\mathbf{x}_k - \mathbf{x}_{\tau_k})}{\sigma_k}\right) \tag{63}$$

Note that similarly to the single-bit case, we have the important result

Lemma 2. *$p(b_k = i|\mathbf{x}_k)$ is log-concave.*

Proof. Define the residual $r_k = z_k - \mathbf{h}_k^T \mathbf{x}_{\tau_k}$. We have

$$p(b_k = i|\mathbf{x}_k) = \int_{\tau(i)}^{\tau(i+1)} p(r_k|\mathbf{x}_k) dr_k, \text{ where } p(r_k|\mathbf{x}_k) = \mathcal{N}(\mathbf{h}_k^T(\mathbf{x}_k - \mathbf{x}_{\tau_k}), \sigma^2) \tag{64}$$

This is an integration over a convex set of a conditional pdf that is log-concave in both arguments, r_k and $\mathbf{x}_k$. According to [8, p. 12] it follows that $p(b_k = i|\mathbf{x}_k)$ is also log-concave. □

We can now compute the posterior pdf.

Proposition 2 (Batch-quantized MAP (BQMAP))**.** *Assume the linear model of* (1)-(4)*. If m bits are allocated per measurement, with a quantization rule as in* (60)*, the posterior pdf is given by*

$$p(\mathbf{x}|\mathbf{b}_{0:N}) = \frac{1}{p(\mathbf{b}_{0:N})} \prod p(b_k|\mathbf{x}_k) \prod p(\mathbf{x}_k|\mathbf{x}_{k-1}) \cdot p(\mathbf{x}_0) \tag{65}$$

$$= \frac{1}{p(\mathbf{b}_{0:N})} \prod \left[Q\left(\frac{\tau(i) - \mathbf{h}_k^T(\mathbf{x}_k - \mathbf{x}_{\tau_k})}{\sigma_k}\right) - Q\left(\frac{\tau(i+1) - \mathbf{h}_k^T(\mathbf{x}_k - \mathbf{x}_{\tau_k})}{\sigma_k}\right) \right]$$
$$\cdot \prod \mathcal{N}(\mathbf{x}_{k+1} - \mathbf{F}_k \mathbf{x}_k, \mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^T) \cdot p(\mathbf{x}_0) \tag{66}$$

and the MAP estimate by

$$\hat{\mathbf{x}} = \arg\max p(\mathbf{x}|\mathbf{b}_{0:N}) \tag{67}$$

$$= \arg\min - \sum \log \left[Q\left(\frac{\tau(i) - \mathbf{h}_k^T(\mathbf{x}_k - \mathbf{x}_{\tau_k})}{\sigma_k}\right) - Q\left(\frac{\tau(i+1) - \mathbf{h}_k^T(\mathbf{x}_k - \mathbf{x}_{\tau_k})}{\sigma_k}\right) \right]$$
$$+ \sum \frac{1}{2} \|\mathbf{x}_{k+1} - F_k \mathbf{x}_k\|^2_{(\mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^T)^{-1}} + \frac{1}{2} \|\mathbf{x}_0 - \mathbf{x}(0)\|^2_{\mathbf{P}_0^{-1}} \tag{68}$$

Due to the log-concavity of $p(b_k|\mathbf{x}_k)$, we also have

Lemma 3. *The posterior pdf of the BQMAP given in Eq.* (66) *is log-concave in* $\mathbf{x}$.

Similarly as before, we define the auxiliary quantities

$$u_{k,i} = \frac{\tau(i) - \mathfrak{h}_k^T(\hat{x}_p - \mathfrak{x}_{\tau_k})}{\sigma_k}, \quad u_{k,i+1} = \frac{\tau(i+1) - \mathfrak{h}_k^T(\hat{x}_p - \mathfrak{x}_{\tau_k})}{\sigma_k} \tag{69}$$

The gradient and hessian of the cost function (68) are now computed as (cf. Appendix C)

$$\nabla f(\hat{\mathbf{x}}_p) = \sum \frac{\mathcal{N}(u_{k,i+1},1) - \mathcal{N}(u_{k,i},1)}{Q(u_{k,i}) - Q(u_{k,i+1})} \cdot \frac{\mathfrak{h}_k}{\sigma_k} + \sum \mathfrak{F}_k^T(\mathbf{G}_k\mathbf{Q}_k\mathbf{G}_k^T)^{-1}\mathfrak{F}_k\hat{\mathbf{x}}_p + \mathfrak{F}_{-1}^T\mathbf{P}_0^{-1}\mathfrak{F}_{-1}(\hat{\mathbf{x}}_p - \mathfrak{x}_0) \tag{70}$$

$$\nabla^2 f(\hat{\mathbf{x}}_p) = \sum \mathfrak{F}_k^T(\mathbf{G}_k\mathbf{Q}_k\mathbf{G}_k^T)^{-1}\mathfrak{F}_k + \mathfrak{F}_{-1}^T\mathbf{P}_0^{-1}\mathfrak{F}_{-1}$$
$$+ \sum \frac{1}{\sigma_k^2} \frac{\Big(\mathcal{N}(u_{k,i+1},1)\cdot u_{k,i+1} - \mathcal{N}(u_{k,i},1)\cdot u_{k,i}\Big)\cdot\Big(Q(u_{k,i}) - Q(u_{k,i+1})\Big) + \Big(\mathcal{N}(u_{k,i+1},1) - \mathcal{N}(u_{k,i},1)\Big)^2}{\Big(Q(u_{k,i}) - Q(u_{k,i+1})\Big)^2}\mathfrak{h}_k\mathfrak{h}_k^T \tag{71}$$

3.1 Measurement matrix form

As for the single-bit MAP, we use the second order Taylor series expansion of the cost function to obtain the iteration of nonlinear WLS, by rewriting the cost function as

$$\boldsymbol{\delta}\mathbf{x}^* = \arg\min \sum \frac{1}{2}||\mathbf{c}_k(\hat{x}_p)^T\boldsymbol{\delta}\mathbf{x} - d_k(\hat{\mathbf{x}}_p)||_2^2 + \sum \frac{1}{2}\|\mathfrak{F}_k\boldsymbol{\delta}\mathbf{x} - \mathfrak{F}_k(-\hat{\mathbf{x}}_p)\|^2_{(\mathbf{G}_k\mathbf{Q}_k\mathbf{G}_k^T)^{-1}} + \frac{1}{2}\|\mathfrak{F}_{-1}\boldsymbol{\delta}\mathbf{x} - \mathfrak{F}_{-1}(\mathfrak{x}_0 - \hat{\mathbf{x}}_p)\|^2_{\mathbf{P}_0^{-1}} \tag{72}$$

Following the same arguments as in the single bit case, we obtain

$$\mathbf{c}_k(\hat{\mathbf{x}}_p) = \frac{\sqrt{\Big(\mathcal{N}(u_{k,i+1},1)\cdot u_{k,i+1} - \mathcal{N}(u_{k,i},1)\cdot u_{k,i}\Big)\Big(Q(u_{k,i}) - Q(u_{k,i+1})\Big) + \Big(\mathcal{N}(u_{k,i+1},1) - \mathcal{N}(u_{k,i},1)\Big)^2}}{Q(u_{k,i}) - Q(u_{k,i+1})} \cdot \frac{\mathfrak{h}_k^T}{\sigma_k} \tag{73}$$

$$d_k(\hat{\mathbf{x}}_p) = -\frac{\mathcal{N}(u_{k,i+1},1) - \mathcal{N}(u_{k,i},1)}{\sqrt{\Big(\mathcal{N}(u_{k,i+1},1)\cdot u_{k,i+1} - \mathcal{N}(u_{k,i},1)\cdot u_{k,i}\Big)\Big(Q(u_{k,i}) - Q(u_{k,i+1})\Big) + \Big(\mathcal{N}(u_{k,i+1},1) - \mathcal{N}(u_{k,i},1)\Big)^2}} \tag{74}$$

3.2 Approximations for small and large arguments

For the cases where $u_{k,i} = -\infty$ or $u_{k,i+1} = \infty$, we will need to use the limit

$$\lim_{x\to\pm\infty} \mathcal{N}(0,1)\cdot x = 0 \tag{75}$$

Case $u_{k,i} = -\infty$ If $u_{k,i} = -\infty$ and $u_{k,i+1}$ is small, we can employ the following approximations

$$\mathbf{c}_k(\hat{\mathbf{x}}_p) = \lim_{u_{k,i+1}\to-\infty} \frac{\sqrt{\Big(\mathcal{N}(u_{k,i+1},1)\cdot u_{k,i+1}\Big)\Big(1 - Q(u_{k,i+1})\Big) + \Big(\mathcal{N}(u_{k,i+1},1)\Big)^2}}{1 - Q(u_{k,i+1})} \cdot \frac{h^T}{\sigma_k} \tag{76}$$

$$= \lim_{u_{k,i+1}\to-\infty} \sqrt{-\nabla_{u_{k,i+1}} \frac{\mathcal{N}(u_{k,i+1},1)}{Q(-u_{k,i+1})}} \cdot \frac{h^T}{\sigma} \tag{77}$$

$$= \sqrt{-\nabla_{u_{k,i+1}}(-u_{k,i+1})} = 1 \cdot \frac{h^T}{\sigma_k} \tag{78}$$

$$d_k(\hat{\mathbf{x}}_p) = \lim_{u_{k,i+1}\to-\infty} -\frac{\mathcal{N}(u_{k,i+1},1)}{\Big(1-Q(u_{k,i+1})\Big)} \frac{\Big(1-Q(u_{k,i+1})\Big)}{\sqrt{\Big(\mathcal{N}(u_{k,i+1},1)\cdot u_{k,i+1}\Big)\Big(1-Q(u_{k,i+1})\Big)+\Big(\mathcal{N}(u_{k,i+1},1)\Big)^2}} \tag{79}$$

$$= \lim_{u_{k,i+1}\to-\infty} \frac{-\mathcal{N}(u_{k,i+1},1)}{\Big(1-Q(u_{k,i+1})\Big)} \frac{1}{\sqrt{1}} \tag{80}$$

$$= u_{k,i+1} \tag{81}$$

Case $u_{k,i+1}=\infty$ If $u_{k,i+1}=\infty$ and $u_{k,i}$ and is large, we can employ the following approximations

$$\mathbf{c}_k(\hat{\mathbf{x}}_p) = \lim_{u_{k,i}\to\infty} \frac{\sqrt{\Big(-\mathcal{N}(u_{k,i},1)\cdot u_{k,i}\Big)\Big(Q(u_{k,i}))\Big)+\Big(-\mathcal{N}(u_{k,i},1)\Big)^2}}{Q(u_{k,i}))} \cdot \frac{\mathfrak{h}^T}{\sigma_k} \tag{82}$$

$$= \lim_{u_{k,i}\to\infty} \sqrt{\nabla_{u_{k,i}} \frac{\mathcal{N}(u_{k,i},1)}{Q(u_{k,i})}} \cdot \frac{\mathfrak{h}^T}{\sigma_k} \tag{83}$$

$$= \sqrt{\nabla_{u_{k,i}}(u_{k,i})} = 1 \cdot \frac{\mathfrak{h}^T}{\sigma_k} \tag{84}$$

$$d_k(\hat{\mathbf{x}}_p) = \lim_{u_{k,i}\to\infty} -\frac{-\mathcal{N}(u_{k,i},1)}{Q(u_{k,i})} \frac{Q(u_{k,i}))}{\sqrt{\Big(-\mathcal{N}(u_{k,i},1)\cdot u_{k,i}\Big)\Big(Q(u_{k,i}))\Big)+\Big(-\mathcal{N}(u_{k,i},1)\Big)^2}} \tag{85}$$

$$= \lim_{u_{k,i}\to\infty} -\frac{-\mathcal{N}(u_{k,i},1)}{Q(u_{k,i})} \frac{1}{\sqrt{1}} \tag{86}$$

$$= u_{k,i} \tag{87}$$

4 Iteratively quantized multibit MAP (IQMAP)

Batch-quantized MAP requires that the number of bits used to quantize a measurement is known and fixed at the time of quantization. There is no principled method to refine this quantization once an additional bit becomes available later. To overcome this limitations of batch-quantized MAP, we introduce the iteratively quantized MAP. The aim is to find a quantization rule that allows to refine the quantization of a given measurement iteratively once an additional bit becomes available. To this end, we take inspiration from the iteratively-quantized KF [9]. First, we investigate augmenting the state vector by the measurement noise. For illustration, we consider the case of a single measurement z of a variable x with a Gaussian prior.

We first define an augmented state vector containing the original state and the measurement noise.

$$\breve{\mathbf{x}}_k = \begin{bmatrix}\mathbf{x}_k \\ n_k\end{bmatrix} \quad \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{x}_k \\ 0\end{bmatrix}, \begin{bmatrix}\mathbf{P}_k & 0 \\ 0 & \sigma_k^2\end{bmatrix}\right) \tag{88}$$

We also define a modified measurement

$$\breve{z}_k = \breve{\mathbf{h}}_k^T \breve{\mathbf{x}}_k + v_k \tag{89}$$

where

$$\breve{\mathbf{h}}^T = \begin{bmatrix}\mathbf{h}^T & 1\end{bmatrix}, \quad v_k \sim \mathcal{N}(0,0) \tag{90}$$

This measurement can be thought of as a perfect measurement, since the new measurement noise v_k has zero covariance.

We define our new quantization rule as

$$b_k = \begin{cases} +1 & \text{if } z_k - \breve{\mathbf{h}}_k^T \breve{\mathbf{x}}_{\tau_k} \geq 0 \\ -1 & \text{else} \end{cases} \tag{91}$$

and note that now the conditional probability $p(b_k = i|\breve{\mathbf{x}}_k)$ is given by

$$p(b_k = i|\breve{\mathbf{x}}_k) = Q\left(\frac{-b_k \breve{\mathbf{h}}_k^T(\breve{\mathbf{x}}_k - \breve{\mathbf{x}}_{\tau_k})}{0}\right) \tag{92}$$

$$= \begin{cases} +1 & \text{if } -b_k \breve{\mathbf{h}}_k^T(\breve{\mathbf{x}} - \breve{\mathbf{x}}_{\tau_k}) < 0 \\ 0 & \text{else} \end{cases} \tag{93}$$

which is essentially an indicator function.

Example: Let us consider a simple example, with only a prior and a single measurement. Assume we have received two bits, b_1 and b_2, for the same measurement z. The posterior pdf is then given by

$$p(\breve{\mathbf{x}}|b_1, b_2) = \frac{p(b_2|\breve{\mathbf{x}}, b_1)p(b_1|\breve{\mathbf{x}})p(\breve{\mathbf{x}})}{p(b_1, b_2)} \tag{94}$$

$$= \frac{1}{p(b_1, b_2)} Q\left(\frac{-b_2 \breve{\mathbf{h}}^T(\breve{\mathbf{x}} - \breve{\mathbf{x}}_{\tau_2})}{0}\right) Q\left(\frac{-b_1 \breve{\mathbf{h}}^T(\breve{\mathbf{x}} - \breve{\mathbf{x}}_{\tau_1})}{0}\right) \mathcal{N}\left(\begin{bmatrix} x_0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{P}_0 & 0 \\ 0 & \sigma^2 \end{bmatrix}\right) \tag{95}$$

Note that, since the Q-function acts as an indicator function, solving for the MAP estimate of $\breve{\mathbf{x}}$ corresponds to solving a least squares problem with linear inequality constraints, i.e.,

$$\hat{\breve{\mathbf{x}}} = \arg\min \frac{1}{2}\left(\breve{\mathbf{x}} - \begin{bmatrix} x_0 \\ 0 \end{bmatrix}\right)^T \begin{bmatrix} \mathbf{P}_0 & 0 \\ 0 & \sigma^2 \end{bmatrix}^{-1} \left(\breve{\mathbf{x}} - \begin{bmatrix} x_0 \\ 0 \end{bmatrix}\right) \tag{96}$$

$$\text{s.t.} - b_1 \breve{\mathbf{h}}^T(\breve{\mathbf{x}} - \breve{\mathbf{x}}_{\tau_1}) < 0$$

$$- b_2 \breve{\mathbf{h}}^T(\breve{\mathbf{x}} - \breve{\mathbf{x}}_{\tau_2}) < 0$$

4.1 Equivalence of Batch-quantized multibit MAP with marginalized sequential MAP

However, we are not so much interested in $\hat{\breve{\mathbf{x}}}$ but rather in $\hat{\mathbf{x}}$. The latter can be obtained by marginalizing out the noise terms by which we had augmented the original state vector. Using the prior plus single measurement example, we will now show that this process results in a form equivalent to the batch-quantized multibit MAP estimator.

$$p(\mathbf{x}|b_1, b_2) = \frac{1}{p(b_1, b_2)} \int_{-\infty}^{\infty} Q\left(\frac{-b_2 \breve{\mathbf{h}}^T(\breve{\mathbf{x}} - \breve{\mathbf{x}}_{\tau_2})}{0}\right) Q\left(\frac{-b_1 \breve{\mathbf{h}}^T(\breve{\mathbf{x}} - \breve{\mathbf{x}}_{\tau_1})}{0}\right) \mathcal{N}\left(\begin{bmatrix} \mathbf{x}_0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{P}_0 & 0 \\ 0 & \sigma^2 \end{bmatrix}\right) dn \tag{97}$$

$$= \frac{1}{p(b_1, b_2)} \int_{-\infty}^{\infty} Q\left(\frac{-b_2 \breve{\mathbf{h}}^T(\breve{\mathbf{x}} - \breve{\mathbf{x}}_{\tau_2})}{0}\right) Q\left(\frac{-b_1 \breve{\mathbf{h}}^T(\breve{\mathbf{x}} - \breve{\mathbf{x}}_{\tau_1})}{0}\right) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{n^2}{\sigma^2}} dn \cdot \mathcal{N}(\mathbf{x}_0, \mathbf{P}_0) \tag{98}$$

Let us slightly rewrite the linear inequality constraints resulting from the Q-functions

$$- b \breve{\mathbf{h}}^T(\breve{\mathbf{x}} - \breve{\mathbf{x}}_{\tau}) < 0 \tag{99}$$

$$\Leftrightarrow b \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau}) + b(n - n_{\tau}) > 0 \tag{100}$$

$$\Leftrightarrow bn > b\big(n_{\tau} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau})\big) \tag{101}$$

We will now assume specific cases for the signs of the bits.

Case $b_1 > 0$**,** $b_2 < 0$ The constraints are now

$$n > n_{\tau_1} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_1}) \tag{102}$$
$$n < n_{\tau_2} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_2}) \tag{103}$$

The posterior can be computed as

$$p(\mathbf{x}|b_1, b_2) = \frac{1}{p(b_1, b_2)} \int_{n_{\tau_1} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_1})}^{n_{\tau_2} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_2})} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{n^2}{\sigma^2}} dn \cdot \mathcal{N}(\mathbf{x}_0, \mathbf{P}_0) \tag{104}$$

$$= \frac{1}{p(b_1, b_2)} \left[\int_{n_{\tau_1} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_1})}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{n^2}{\sigma^2}} dn - \int_{n_{\tau_2} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_2})}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{n^2}{\sigma^2}} dn \right] \cdot \mathcal{N}(\mathbf{x}_0, \mathbf{P}_0) \tag{105}$$

$$= \frac{1}{p(b_1, b_2)} \left[Q\left(\frac{n_{\tau_1} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_1})}{\sigma} \right) - Q\left(\frac{n_{\tau_2} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_2})}{\sigma} \right) \right] \cdot \mathcal{N}(\mathbf{x}_0, \mathbf{P}_0) \tag{106}$$

Case $b_1 < 0$**,** $b_2 > 0$ The constraints are now

$$n < n_{\tau_1} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_1}) \tag{107}$$
$$n > n_{\tau_2} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_2}) \tag{108}$$

The posterior can be computed as

$$p(\mathbf{x}|b_1, b_2) = \frac{1}{p(b_1, b_2)} \left[Q\left(\frac{n_{\tau_2} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_2})}{\sigma} \right) - Q\left(\frac{n_{\tau_1} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_1})}{\sigma} \right) \right] \cdot \mathcal{N}(\mathbf{x}_0, \mathbf{P}_0) \tag{109}$$

Obviously, this is only valid if $n_{\tau_1} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_1}) > n_{\tau_2} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_2})$, otherwise the difference of Q-functions will become negative and not yield a proper pdf. In that sense, this case is the same as the one before, with differently ordered thresholds.

Case $b_1 > 0$**,** $b_2 > 0$ The constraints are now

$$n > n_{\tau_1} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_1}) \tag{110}$$
$$n > n_{\tau_2} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_2}) \tag{111}$$

The posterior is

$$p(\mathbf{x}|b_1, b_2) = \frac{1}{p(b_1, b_2)} \left[Q\left(\frac{\max\Big(n_{\tau_1} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_1}), n_{\tau_2} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_2})\Big)}{\sigma} \right) \right] \cdot \mathcal{N}(\mathbf{x}_0, \mathbf{P}_0) \tag{112}$$

$$= \frac{1}{p(b_1, b_2)} \left[Q\left(\frac{\max\Big(n_{\tau_1} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_1}), n_{\tau_2} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_2})\Big)}{\sigma} \right) - Q(\infty) \right] \cdot \mathcal{N}(\mathbf{x}_0, \mathbf{P}_0) \tag{113}$$

Case $b_1 < 0$**,** $b_2 < 0$ The constraints are now

$$n < n_{\tau_1} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_1}) \tag{114}$$
$$n < n_{\tau_2} - \mathbf{h}^T(\mathbf{x} - \mathbf{x}_{\tau_2}) \tag{115}$$

The posterior is

$$p(\mathbf{x}|b_1,b_2) = \frac{1}{p(b_1,b_2)} \int_{-\infty}^{\min\left(n_{\tau_1}-\mathbf{h}^T(\mathbf{x}-\mathbf{x}_{\tau_1}),n_{\tau_2}-\mathbf{h}^T(\mathbf{x}-\mathbf{x}_{\tau_2})\right)} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{n^2}{\sigma^2}} dn \cdot \mathcal{N}(\mathbf{x}_0,\mathbf{P}_0) \tag{116}$$

$$= \frac{1}{p(b_1,b_2)} \left[1 - Q\left(\frac{\min\left(n_{\tau_1}-\mathbf{h}^T(\mathbf{x}-\mathbf{x}_{\tau_1}),n_{\tau_2}-\mathbf{h}^T(\mathbf{x}-\mathbf{x}_{\tau_2})\right)}{\sigma}\right)\right] \cdot \mathcal{N}(\mathbf{x}_0,\mathbf{P}_0) \tag{117}$$

$$= \frac{1}{p(b_1,b_2)} \left[Q(-\infty) - Q\left(\frac{\min\left(n_{\tau_1}-\mathbf{h}^T(\mathbf{x}-\mathbf{x}_{\tau_1}),n_{\tau_2}-\mathbf{h}^T(\mathbf{x}-\mathbf{x}_{\tau_2})\right)}{\sigma}\right)\right] \cdot \mathcal{N}(\mathbf{x}_0,\mathbf{P}_0) \tag{118}$$

We see that we can always express the information from one analog measurement as the difference of two Q-functions, which essentially bound the noise from below and above. This is a generalization of the batch-quantized estimate, where we had

$$p(b=i|\mathbf{x}) = Q\left(\frac{\tau(i)-\mathbf{h}^T(\mathbf{x}-\mathbf{x}_\tau)}{\sigma}\right) - Q\left(\frac{\tau(i+1)-\mathbf{h}^T(\mathbf{x}-\mathbf{x}_\tau)}{\sigma}\right) \tag{119}$$

We now have, after receiving k bits,

$$p(b_1,b_2,\ldots,b_k|\mathbf{x}) = \left[Q\left(\frac{n_{\tau_l}-\mathbf{h}^T(\mathbf{x}-\mathbf{x}_{\tau_l})}{\sigma}\right) - Q\left(\frac{n_{\tau_u}-\mathbf{h}^T(\mathbf{x}-\mathbf{x}_{\tau_u})}{\sigma}\right)\right] \tag{120}$$

$$= \left[Q\left(\frac{\breve{h}^T\breve{\mathbf{x}}_{\tau_l}-\mathbf{h}^T\mathbf{x}}{\sigma}\right) - Q\left(\frac{\breve{h}^T\breve{\mathbf{x}}_{\tau_u}-\mathbf{h}^T\mathbf{x}}{\sigma}\right)\right] \tag{121}$$

This important insight will guide us in our choice of thresholds. Moreover, we have seen that the above Q-functions become indicator functions when the measurement noise is assumed zero. This will provide a means to incorporate uniform noise pdfs, or set-membership approaches, which are essentially based on linear inequality constraints. Moreover, we see that to solve for the MAP estimate, we can use the same Taylor series developments and measurement matrix forms developed for the batch-quantized multibit MAP (cf. Section 3), but now with a general, iterative, flexible threshold selection scheme.

We can generalize this result, developed for the simple one measurement, no time-evolution case.

For each measurement, we use the following binary quantization rule for the i-th bit used on the k-th measurement

$$b_{k,i} = \begin{cases} +1 & \text{if } z_k - \breve{\mathbf{h}}_k^T\breve{\mathbf{x}}_{\tau_{ki}} \geq 0 \\ -1 & \text{else} \end{cases} \tag{122}$$

where $\breve{\mathbf{h}}_k = \begin{bmatrix}\mathbf{h}_k^T & 1\end{bmatrix}^T$ is an augmented measurement matrix. The threshold $\breve{\mathbf{x}}_{\tau_{ki}}$ is adjusted every time before a new bit becomes available. At any time, there exist upper and lower thresholds, $\breve{\mathbf{x}}_{\tau_{kl}}$ and $\breve{\mathbf{x}}_{\tau_{ku}}$, so that $\breve{\mathbf{h}}_k^T\breve{\mathbf{x}}_{\tau_{ku}} > z > \breve{\mathbf{h}}_k^T\breve{\mathbf{x}}_{\tau_{kl}}$. Similar to bisection, the new threshold $\breve{\mathbf{x}}_{\tau_{ki}}$ divides the region between $\breve{\mathbf{h}}_k^T\breve{\mathbf{x}}_{\tau_{ku}}$ and $\breve{\mathbf{h}}_k^T\breve{\mathbf{x}}_{\tau_{kl}}$, and, depending on the outcome of the quantization (122), becomes the new upper or lower threshold. Notice that, in contrast to the batch-quantized MAP, there are principled methods to adjust these thresholds *any time* that additional bits become available, not only when the measurement is quantized for the first time. The exact procedure how to choose the new threshold $\breve{\mathbf{x}}_{\tau_{ki}}$ will be discussed shortly.

Proposition 3 (Iteratively quantized MAP (IQMAP))**.** *Assume the linear model of* (1)-(4)*. If multiple bits are allo-*

cated per measurement according to the quantization rule (122)*, the posterior pdf is given by*

$$p(\mathbf{x}_{0:K}|\mathbf{b}_{0:K}) = \frac{1}{p(\mathbf{b}_{0:K})} \prod_{k=0}^{K} p(b(k)|\mathbf{x}_k) \prod_{k=0}^{K-1} p(\mathbf{x}_{k+1}|\mathbf{x}_k) \cdot p(\mathbf{x}_0) \tag{123}$$

$$= \frac{1}{p(\mathbf{b}_{0:K})} \prod_{k=0}^{K} \left[Q\left(\frac{\breve{\mathbf{h}}_k^T \breve{\mathbf{x}}_{\tau_{kl}} - \mathbf{h}_k^T \mathbf{x}_k}{\sigma_k} \right) - Q\left(\frac{\breve{\mathbf{h}}_k^T \breve{\mathbf{x}}_{\tau_{ku}} - \mathbf{h}_k^T \mathbf{x}_k}{\sigma_k} \right) \right]$$

$$\cdot \prod_{k=0}^{K-1} \mathcal{N}(\mathbf{F}_k \mathbf{x}_k, \mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^T) \cdot p(\mathbf{x}_0) \tag{124}$$

and the MAP estimate by

$$\hat{\mathbf{x}}_{0:K} = \arg\max p(\mathbf{x}_{0:K}|\mathbf{b}_{0:K}) \tag{125}$$

$$= \arg\min - \sum \log \left[Q\left(\frac{\breve{\mathbf{h}}_k^T \breve{\mathbf{x}}_{\tau_{kl}} - \mathbf{h}_k^T \mathbf{x}_k}{\sigma_k} \right) - Q\left(\frac{\breve{\mathbf{h}}^T \breve{\mathbf{x}}_{\tau_{ku}} - \mathbf{h}_k^T \mathbf{x}_k}{\sigma_k} \right) \right]$$

$$+ \sum \frac{1}{2} \|\mathbf{x}_{k+1} - \mathbf{F}_k \mathbf{x}_k\|^2_{(\mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^T)^{-1}} + \frac{1}{2} \|\mathbf{x}_0 - \mathbf{x}(0)\|^2_{\mathbf{P}_0^{-1}} \tag{126}$$

For the same reasons as in the BQMAP, we have

Lemma 4. *The posterior pdf of the IQMAP given in Eq.* (124) *is log-concave in* $\mathbf{x}$*.*

4.2 Threshold selection

Once an additional bit can be used to refine a given measurement z_k, a new quantization threshold has to be computed. This threshold $\breve{\mathbf{x}}_{\tau_{ki}}$ is chosen so that, conditioned on all previous quantized observations, the measurement z_k could lie on each side of the new threshold with equal probability. More precisely, we desire that

$$\Pr\left\{ \breve{\mathbf{h}}_k^T \breve{\mathbf{x}}_{\tau_{kl}} < z_k \leq \breve{\mathbf{h}}_k^T \breve{\mathbf{x}}_{\tau_{ki}} \middle| \mathbf{b}_{0:K} \right\} = \Pr\left\{ \breve{\mathbf{h}}_k^T \breve{\mathbf{x}}_{\tau_{ki}} < z_k \leq \breve{\mathbf{h}}_k^T \breve{\mathbf{x}}_{\tau_{ku}} \middle| \mathbf{b}_{0:K} \right\} \tag{127}$$

We can express these two probabilities in terms of the Q-function, if we approximate the conditional pdf $p(z_k|\mathbf{b}_{0:K})$ by a Gaussian with mean $\mathbf{h_k}^T \hat{\mathbf{x}}_k$ and covariance matrix $\mathbf{h}_k^T \mathbf{P} \mathbf{h}_k + \sigma_k^2$, where $\hat{\mathbf{x}}_k$ denote the current MAP estimate of the state at time step k, and $\mathbf{P} = (\mathbf{A}^T \mathbf{A})^{-1}$ its covariance.

We want

$$\Pr\left\{ \breve{\mathbf{h}}_k^T \breve{\mathbf{x}}_{\tau_{kl}} < z_k \leq \breve{\mathbf{h}}_k^T \breve{\mathbf{x}}_{\tau_{ki}} \middle| \mathbf{b}_{0:K} \right\} = \Pr\left\{ \breve{\mathbf{h}}_k^T \breve{\mathbf{x}}_{\tau_{ki}} < z_k \leq \breve{\mathbf{h}}_k^T \breve{\mathbf{x}}_{\tau_{ku}} \middle| \mathbf{b}_{0:K} \right\} \tag{128}$$

$$\Leftrightarrow Q\left(\frac{\breve{\mathbf{h}}^T \breve{\mathbf{x}}_{\tau_{kl}} - \breve{\mathbf{h}}^T \hat{\breve{\mathbf{x}}}}{\sqrt{\mathbf{h}^T \mathbf{P} \mathbf{h} + \sigma_k^2}} \right) - Q\left(\frac{\breve{\mathbf{h}}^T \breve{\mathbf{x}}_{\tau_{ki}} - \breve{\mathbf{h}}^T \hat{\breve{\mathbf{x}}}}{\sqrt{\mathbf{h}^T \mathbf{P} \mathbf{h} + \sigma_k^2}} \right) = Q\left(\frac{\breve{\mathbf{h}}^T \breve{\mathbf{x}}_{\tau_{ki}} - \breve{\mathbf{h}}^T \hat{\breve{\mathbf{x}}}}{\sqrt{\mathbf{h}^T \mathbf{P} \mathbf{h} + \sigma_k^2}} \right) - Q\left(\frac{\breve{\mathbf{h}}^T \breve{\mathbf{x}}_{\tau_{ku}} - \breve{\mathbf{h}}^T \hat{\breve{\mathbf{x}}}}{\sqrt{\mathbf{h}^T \mathbf{P} \mathbf{h} + \sigma_k^2}} \right) \tag{129}$$

$$\Leftrightarrow 2Q\left(\frac{\breve{\mathbf{h}}^T \breve{\mathbf{x}}_{\tau_{ki}} - \breve{\mathbf{h}}^T \hat{\breve{\mathbf{x}}}}{\sqrt{\mathbf{h}^T \mathbf{P} \mathbf{h} + \sigma_k^2}} \right) = Q\left(\frac{\breve{\mathbf{h}}^T \breve{\mathbf{x}}_{\tau_{kl}} - \breve{\mathbf{h}}^T \hat{\breve{\mathbf{x}}}}{\sqrt{\mathbf{h}^T \mathbf{P} \mathbf{h} + \sigma_k^2}} \right) + Q\left(\frac{\breve{\mathbf{h}}^T \breve{\mathbf{x}}_{\tau_{ku}} - \breve{\mathbf{h}}^T \hat{\breve{\mathbf{x}}}}{\sqrt{\mathbf{h}^T \mathbf{P} \mathbf{h} + \sigma_k^2}} \right) \tag{130}$$

which yields the new threshold as

$$\breve{\mathbf{h}}^T \breve{\mathbf{x}}_{\tau_{ki}} = \sqrt{\mathbf{h}^T \mathbf{P} \mathbf{h} + \sigma_k^2} \cdot Q^{\text{inv}}\left(\frac{1}{2} \left(Q\left(\frac{\breve{\mathbf{h}}^T \breve{\mathbf{x}}_{\tau_{kl}} - \breve{\mathbf{h}}^T \hat{\breve{\mathbf{x}}}}{\sqrt{\mathbf{h}^T \mathbf{P} \mathbf{h} + \sigma_k^2}} \right) + Q\left(\frac{\breve{\mathbf{h}}^T \breve{\mathbf{x}}_{\tau_{ku}} - \breve{\mathbf{h}}^T \hat{\breve{\mathbf{x}}}}{\sqrt{\mathbf{h}^T \mathbf{P} \mathbf{h} + \sigma_k^2}} \right) \right) \right) + \breve{\mathbf{h}}^T \hat{\breve{\mathbf{x}}} \tag{131}$$

5 Optimal bit allocation

MAP estimation, contrary to filtering, allows to include information from asynchronous, i.e., time-delayed measurements. In the context of quantized distributed estimation, it also allows to utilize unexpectedly available bandwidth resources to transmit additional information about earlier measurements. This section will address the question of how to optimally allocate those additional bits.

5.1 Single bit refinement of one previous measurement

The first approach is use one additional bit to refine one selected measurement using the iterative threshold selection scheme. We will select the measurement in question by its anticipated contribution to reducing the uncertainty in terms of the trace of the covariance matrix. To this end, we compute for each measurement the new threshold $\breve{x}_{T_n}$ according to the threshold selection scheme described in Section 4.2. Depending on the outcome of the quantization, there are two possible new likelihoods for this measurement, either $\left[Q\left(\frac{\breve{h}^T\breve{x}_{T_l}-h^Tx}{\sigma}\right)-Q\left(\frac{\breve{h}^T\breve{x}_{T_n}-h^Tx}{\sigma}\right)\right]$ or $\left[Q\left(\frac{\breve{h}^T\breve{x}_{T_n}-h^Tx}{\sigma}\right)-Q\left(\frac{\breve{h}^T\breve{x}_{T_u}-h^Tx}{\sigma}\right)\right]$

For both cases, we can compute the new vector c_n from Eq. (73). Notice that the particular measurement in question corresponds to exactly one line (row vector c) in the measurement matrix, which would now be replaced by this new c_n.

Let us assume that at the current estimate, the least squares problem can be written as

$$\min \|A_\ominus \cdot \boldsymbol{\delta}\mathbf{x} - b\|_2^2 \tag{132}$$

and from solving this system we also have the current QR-factorization

$$A_\ominus = \begin{bmatrix} A_1 \\ c^T \\ A_2 \end{bmatrix} = Q \cdot R \tag{133}$$

We also know that the covariance of the current estimation error is given by

$$P_\ominus = (A_\ominus^T A_\ominus)^{-1} = (R^T R)^{-1} = R^{-1}R^{-T} \tag{134}$$

.

The new system will be given by

$$A_\oplus = \begin{bmatrix} A_1 \\ c_n^T \\ A_2 \end{bmatrix} \tag{135}$$

Our goal is to compute the new covariance matrix

$$\begin{aligned} P_\oplus &= (A_\oplus^T A_\oplus)^{-1} && (136)\\ &= (A_1^T A_1 + A_2^T A_2 + c_n c_n^T)^{-1} && (137)\\ &= (A_1^T A_1 + A_2^T A_2 + cc^T - cc^T + c_n c_n^T)^{-1} && (138)\\ &= (A_\ominus^T A_\ominus - cc^T + c_n c_n^T)^{-1} && (139) \end{aligned}$$

Let us denote the SVD of $c_n c_n^T - cc^T$ as

$$c_n c_n^T - cc^T = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} u_1^T \\ u_2^T \end{bmatrix} = U\Sigma U^T \tag{140}$$

Notice that, as a sum of two vector outer products, the rank of $c_n c_n^T - cc^T$ can be at most two. Further, both vectors c and c_n are extremely sparse and have their non-zero entries at the same positions. We continue with the computation of $P_\oplus$.

$$\begin{aligned} P_\oplus &= (A_\ominus^T A_\ominus + U\Sigma U^T)^{-1} && (141)\\ &= P_\ominus - P_\ominus U(U^T P_\ominus U + \Sigma^{-1})^{-1} U^T P_\ominus && (142) \end{aligned}$$

The two matrix inversions appearing in the interior are both of 2×2 matrices which can be done in closed form.

We are interested in minimizing the trace of the updated covariance matrix

$$\begin{aligned}\operatorname{tr} P_{\oplus} &= \operatorname{tr}(P_{\ominus} - P_{\ominus}U(U^T P_{\ominus} U + \Sigma^{-1})^{-1} U^T P_{\ominus}) && (143)\\ &= \operatorname{tr} P_{\ominus} - \operatorname{tr}(P_{\ominus}U(U^T P_{\ominus} U + \Sigma^{-1})^{-1} U^T P_{\ominus}) && (144)\\ &= \operatorname{tr} P_{\ominus} - \operatorname{tr}((U^T P_{\ominus} U + \Sigma^{-1})^{-1} U^T P_{\ominus} P_{\ominus} U) && (145)\\ &= \operatorname{tr} P_{\ominus} - \operatorname{tr}((U^T R^{-1} R^{-T} U + \Sigma^{-1})^{-1} U^T R^{-1} R^{-T} R^{-1} R^{-T} U) && (146)\end{aligned}$$

Minimizing the trace of the updated covariance corresponds to

$$\max \operatorname{tr}((U^T R^{-1} R^{-T} U + \Sigma^{-1})^{-1} U^T R^{-1} R^{-T} R^{-1} R^{-T} U) \quad (147)$$

which is a product of two 2×2 matrices. These can be efficiently computed by

1. Compute R^{-1} which is upper triangular using backsubstitution
2. Compute $v_1 = R^{-T}U$ ($n \times 2$ matrix)
3. Compute $v_2 = U^T R^{-1} R^{-T} U = v_1^T v_1$ (2×2 matrix)
4. Compute $v_3 = R^{-1} R^{-T} U = R^{-1} v_1$ ($n \times 2$ matrix)
5. Compute $v_4 = U^T R^{-1} R^{-T} R^{-1} R^{-T} U = v_3^T v_3$ (2×2 matrix)
6. Finally, compute $v_5 = \operatorname{tr}((U^T R^{-1} R^{-T} U + \Sigma^{-1})^{-1} U^T R^{-1} R^{-T} R^{-1} R^{-T} U) = \operatorname{tr}((v_2 + \Sigma^{-1})^{-1} v_4)$

Notice that the most costly operation, namely the inversion of R, needs to be done only once, the remaining operations have to be carried out for each measurement. We assume that the new bit from the measurement will be $+1$ or -1,i.e., that the two new likelihoods mentioned above will occur, with equal probability. We then compute the expected reduction of covariance by refining this measurement as the average of the two quantities v_5 obtained for the two cases. Finally, we choose that measurement for refinement that results in the maximum expected covariance reduction, i.e., the one that obtained maximal average v_5.

Notice that this approach is approximate, since computation of the final covariance matrix would require iteration to convergence which we disregard due to excessive computational complexity. We expect that the approach outlined above will yield sufficient results for practical purposes.

6 Simulation results: 1-D multirobot cooperative localization

We have tested the above methods for a cooperative localization scenario of two robots moving in 1D. Contrary to the more realistic 2D scenario, 1D robot localization is a *linear* problem, and is therefore often used for feasibility studies, and to provide mathematical insight into problems [10, 11].

6.1 System and Measurement Model

In 1D cooperative localization, the system model for each robot is given by

$$x_i(k) = x_i(k-1) + u_i(k-1) \quad (148)$$

The control input $u_i(k-1)$ is measured, i.e., from odometric sensors, and therefore corrupted by white, Gaussian noise $\mu \sim \mathcal{N}(0, \sigma_\mu^2)$

$$u_{i,m}(k-1) = u_i(k-1) + \mu_i(k-1) \quad (149)$$

Notice that, unlike our system model description in the estimator derivation of the previous sections, the system model contains control inputs given in the form of real-valued, noisy measurements. In the current form, the derived estimator is unable to handle this system model. We therefore reformulate the 1D cooperative localization problem, using a statistical, zero acceleration tracking model for the system dynamics, and treating the odometry measurements as another form of sensor update. This transformation requires including the robot velocity in the state. We now have

$$\begin{bmatrix} x_i(k) \\ v_i(k) \end{bmatrix} = \begin{bmatrix} 1 & \delta T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_i(k-1) \\ v_i(k-1) \end{bmatrix} + \mu_i(k-1) \tag{150}$$

where the noise $\mu_i(k-1)$ is now a 2×1 vector with Gaussian distribution, $\mu_i \sim \mathcal{N}(0, Q_i)$, and characterizes how much the robot deviates statistically from moving with constant velocity or zero acceleration. Ideally, the noise would be one-dimensional (or Q_i of rank 1) and only impacting the velocity term. For easier mathematical treatment, however, we keep Q_i of rank 2, with the variance of the position noise set to a very small value. δT denotes the sampling time. We stack the tracking models of all N robots into one big state vector $x = \begin{bmatrix} x_1 & v_1 & \dots & x_N & v_N \end{bmatrix}^T$, obeying the system dynamics

$$x_k = Ax_{k-1} + \mu(k-1) \tag{151}$$

where A is blockdiagonal and has N blocks $\left[\begin{smallmatrix} 1 & \delta T \\ 0 & 1 \end{smallmatrix}\right]$ on its diagonal.

The odometry measurements are now given by

$$\begin{aligned} z_{v_i}(k) &= v_i(k) + \nu_{v_i}(k) && (152) \\ &= h_{v_i}^T x_k + \nu_{v_i}(k) && (153) \end{aligned}$$

where

$$h_{v_i}^T = \begin{bmatrix} 0 & 0 & \dots & \underbrace{0}_{i} & 1 & \dots & 0 & 0 \end{bmatrix} \tag{154}$$

and $\nu_{v_i} \sim \mathcal{N}(0, \sigma_{\nu_v}^2)$ is white Gaussian measurement noise, independent of the process noise μ.

The robots can also process relative measurements of their positions, e.g.,

$$\begin{aligned} z_{ij}(k) &= x_j(k) - x_i(k) + \nu_{ij}(k) && (155) \\ &= h_{ij}^T x_k + \nu_{ij}(k) && (156) \end{aligned}$$

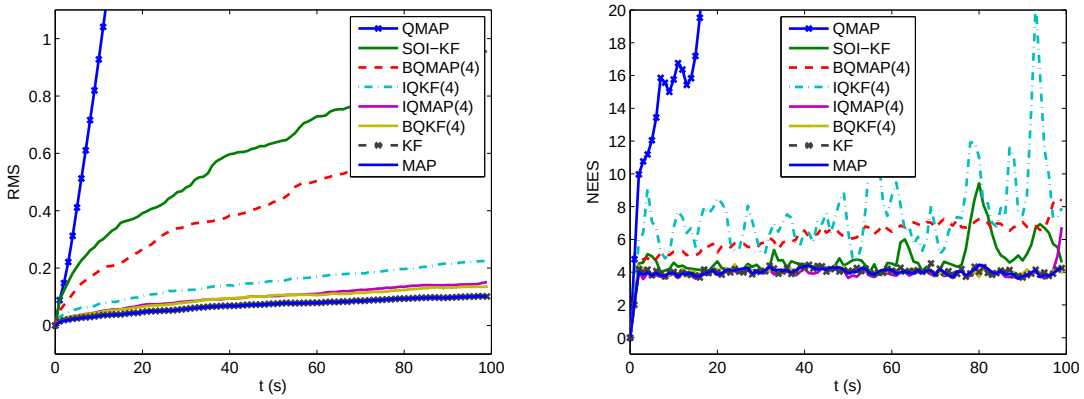
where

$$h_{ij}^T = \begin{bmatrix} 0 & 0 & \dots & \underbrace{-1}_{i} & \underbrace{0}_{} & \dots & \underbrace{1}_{j} & \underbrace{0}_{} & \dots & 0 & 0 \end{bmatrix} \tag{157}$$

is a very sparse row vector, and $\nu_{ij} \sim \mathcal{N}(0, \sigma_{\nu_{ij}}^2)$ is white Gaussian measurement noise, again independent of the process noise μ.

We have compared single-bitMAP, BQMAP and IQMAP against single-bit KF, BQKF and IQKF, i.e., their filter counterparts. Moreover, we have also tested the performance loss compared to real-valued KF and MAP. As performance criterion we have used the root-mean-square error (RMS) and the normalized estimation error squared (NEES). Notice that the RMS will increase over time since cooperative localization without absolute position measurements is unobservable. We also remark that the NEES, defined as $\tilde{x}P^{-1}\tilde{x}$ follows a Chi-square distribution if the estimator is consistent, with expected value equal to the number of states, in our case four. The simulation results are promising and raise some interesting questions.

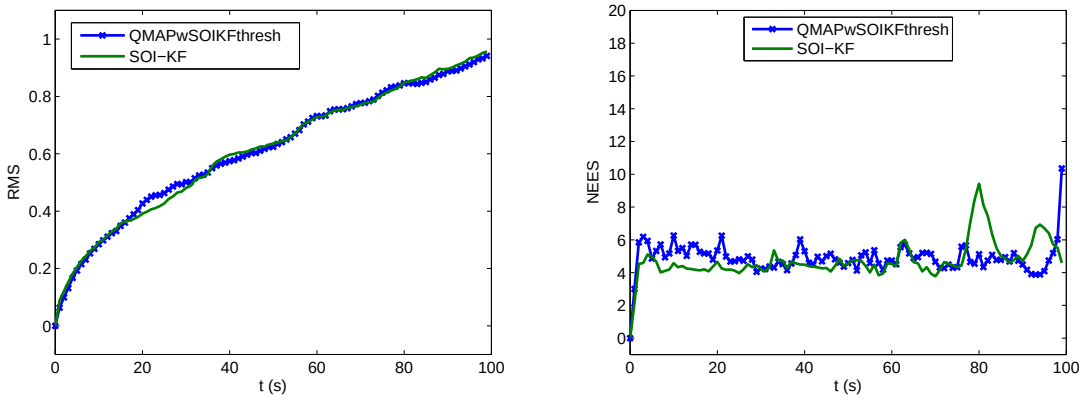
In Fig. 4, we compare the RMS and NEES for analog and quantized estimators, using a single bit and four bits per measurement, as well as the current best estimates as quantization thresholds as discussed above. We observe, as expected, an increase in performance for increasing the number of bits from one to four, to analog measurements. Moreover, we observe that the BQKF performs significantly better than the IQKF, even though both use the same number of bits. This is also expected due to the recurring Gaussian approximation of the posterior, as discussed



(a) RMS error. Single bitMAP with the thresholds selected as the current MAP estimate is performing worst, analog KF and IQbitMAP and BQKF performing as well as analog KF and MAP, MAP best (almost indistinguishable). IQKF is performing worse than BQKF due to its repeated approximation of the posterior. IQbitMAP and BQKF are performing similarly, even though the filter algorithms are optimized for minimum RMS (they are the LMMSE estimators).

(b) NEES. Again, single-bit MAP is performing worst, but reaching the expected value of four. Notice that the IQKF is obviously less consistent than the BQKF.

Figure 4: Simulation Results for 1D Cooperative Localization for different estimators. Iteratively- and batch-quantized estimators were operated with four bits, and results are averaged over 200 Monte Carlo realizations.



(a) RMS

(b) NEES

Figure 5: Comparison of single-bit MAP and SOIKF, if the MAP estimator uses the quantization thresholds obtained from the SOIKF. These thresholds yield a significant performance increase compared to the previous case where the current MAP estimates were used as thresholds for quantization.

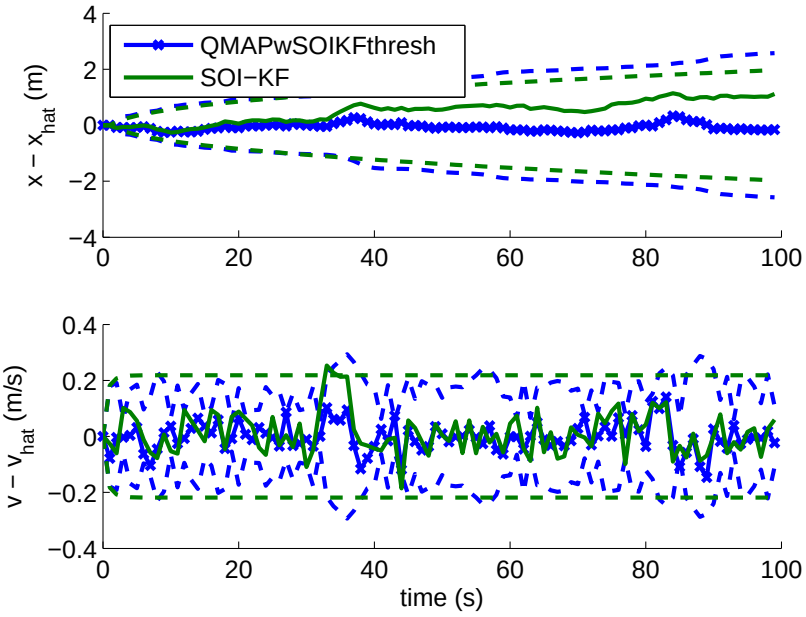


Figure 6: Typical example of robot position and velocity estimation errors (solid lines) and their 3σ-bounds (dashed lines) for SOIKF and single-bit MAP when using the same quantization thresholds. We see that the MAP covariance estimate follows the velocity error more closely than the SOIKF, and that the position covariance estimate is slightly more conservative after high values of velocity uncertainty, compared to the SOIKF.

in [12]. The NEES clearly shows that it is less consistent than the BQKF. Finally, we observe that IQMAP and BQKF have comparable performance, close to that of the analog estimators.

A surprising observation is the relatively weak performance of single-bit MAP and BQMAP compared to their filtering counterparts. An additional study identifies the thresholds as the critical factor. Fig. 5 shows the performance for the same datasets if the single-bit MAP uses the quantization thresholds computed by the SOIKF. We observe that its performance in terms of RMS and NEES is tremendously improved and becomes as good as that of the SOIKF.

Moreover, if we compare the actual errors and their 3σ-bounds, as shown in Fig. 6, we see that the MAP estimator seems to have a covariance estimate that bounds the velocity error more tightly than the SOIKF, whereas the position error is bounded more conservatively, especially after spikes in the velocity covariance.

Notice that the thresholds are used only for quantization in the MAP estimation, not as starting point for the optimization. One could assume that the performance of single-bit MAP and SOIKF have become so similar only because the maximization of the posteriori pdf was started using the SOIKF estimate as an initial guess, and then stopped at a close-by local optimum. This is not the case, since the posterior pdf is convex as previously shown. The final estimate is therefore unique and does not depend on the initial guess, only on the quantized measurements and their thresholds.

We believe the reason for the weak performance of the single-bit MAP when using the current MAP estimate as quantization threshold to be caused by the skew of the posterior pdf. We believe that this skew, observed already in the simple example in Fig. 1, can be exacerbated by poor threshold selection. Since in this case, the MAP estimate does not divide the probability mass evenly, the binary observations contains less and less information, reflected in growing errors and growing covariance. For more bits, especially when allocated with the aim to divide into equiprobable regions, the posterior pdf becomes more peaked, thus alleviating the problem and explaining the better performance of the IQMAP.

The above finding underline impressively the need for optimal threshold selection in MAP estimation with quantized estimation, which confirms similar findings in parameter estimation [3]. We believe that the work of Shah [2] offers a good starting point to determine such optimal thresholds, which is going to be subject of our proposed work.

We want to further note that the presented case of 1D cooperative localization is an example of linear process and measurement dynamics. We believe that the capacity of the MAP estimator to better handle nonlinearities will become

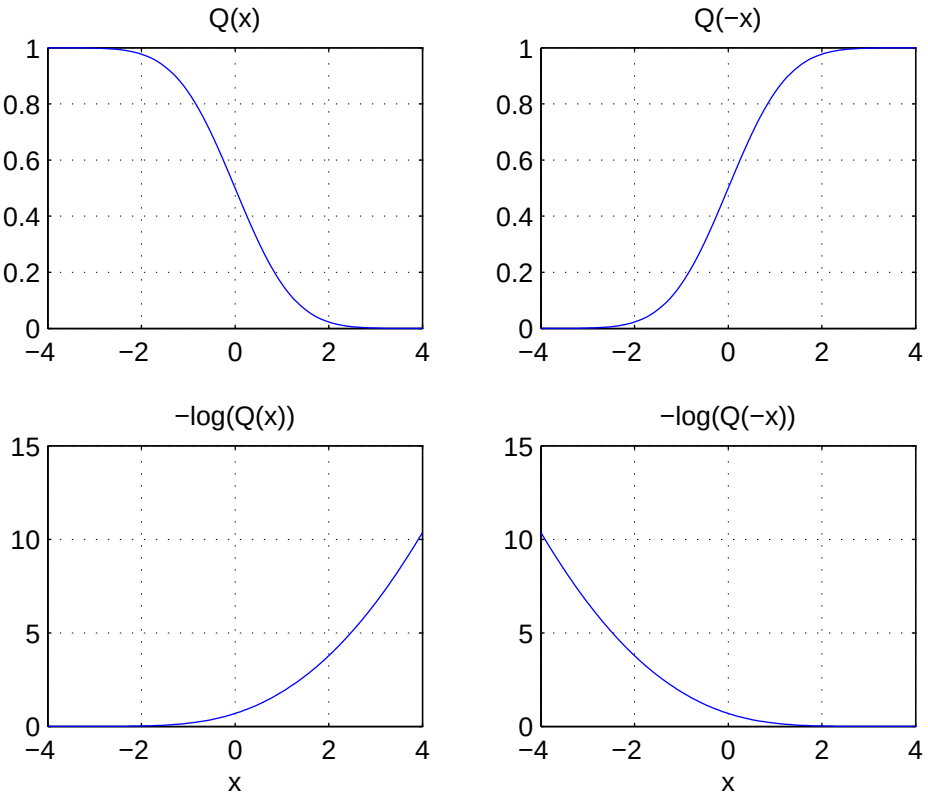


Figure 7: The Q-function and its negative logarithm. The Q-function is monotonically decreasing and log-concave.

much more apparent in nonlinear settings, such as 2D cooperative localization, SLAM, or target tracking.

Finally, the main reason for preferring bitMAP over the quantized filter variants is its ability to use additional bandwidth to retroactively transmit information about earlier measurements. This important ability was not demonstrated by the simulations shown above, where a constant bandwidth or bitrate was assumed.

A Properties of the Q-Function (Gaussian Tail Function)

Definition

$$Q(x) := \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \tag{158}$$

$$= \Pr\{X > x\} \quad \text{if } X \sim \mathcal{N}(0,1) \tag{159}$$

The Q function is monotonically decreasing and log-concave (cf. Fig. 7).

Example values

$$Q(-\infty) = 1 \qquad Q(0) = \frac{1}{2} \qquad Q(\infty) = 0 \tag{160}$$

Q(-x)

$$Q(-x) = \int_{-x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \tag{161}$$

$$= 1 - \int_{-\infty}^{-x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \qquad |v = -u \tag{162}$$

$$= 1 + \int_{\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} dv \tag{163}$$

$$= 1 - \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} dv \tag{164}$$

$$= 1 - Q(x) \tag{165}$$

$$= \Pr\{X \leq x\} \quad \text{if } X \sim \mathcal{N}(0,1) \tag{166}$$

Relation to erf **and** erfc

$$\operatorname{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \tag{167}$$

$$= 1 - 2Q(\sqrt{2}\, x) \tag{168}$$

$$\operatorname{erfc}(x) := \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt \tag{169}$$

$$= 1 - \operatorname{erf}(x) \tag{170}$$

$$= 2Q(\sqrt{2}\, x) \tag{171}$$

$$Q(x) = \frac{\operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right)}{2} \tag{172}$$

$$= \frac{1}{2}\left(1 - \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right) \tag{173}$$

General normal distribution If $X \sim \mathcal{N}(\mu, \sigma^2)$ then

$$\Pr\{X > x\} = \frac{1}{\sqrt{2\pi}\,\sigma} \int_x^{\infty} e^{-\frac{1}{2}\frac{(v-\mu)^2}{\sigma^2}} dv \quad |u = \frac{v-\mu}{\sigma},\ du = \frac{1}{\sigma} dv \tag{174}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\frac{x-\mu}{\sigma}}^{\infty} e^{-\frac{u^2}{2}} du \tag{175}$$

$$= Q\left(\frac{x-\mu}{\sigma}\right) \tag{176}$$

$$\Pr\{X \leq x\} = Q\left(-\frac{x-\mu}{\sigma}\right) \tag{177}$$

Inverse Function

$$Q^{-1}(y) = \sqrt{(2)}\, \operatorname{erfinv}(1-2y) \tag{178}$$

Derivatives

$$\frac{dQ(x)}{dx} = \frac{d}{dx} \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \tag{179}$$

$$= -\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{180}$$

$$= -\mathcal{N}(x, 0, 1) \tag{181}$$

where we made use of the Leibniz-rule for parameter integrals.

$$\frac{d^2Q(x)}{dx^2} = -(-x)\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \tag{182}$$

$$= x\mathcal{N}(x,0,1) \tag{183}$$

B Quadratic Taylor approximation of $-\log Q(-b\frac{h^T(x-x_T)}{\sigma})$

In order to derive the second order Taylor approximation of $-\log Q(-b\frac{h^T(x-x_T)}{\sigma})$, we start by writing the first and second derivatives (gradient and Hessian) as

$$f(x) = -\log Q\Big(-b\frac{h^T(x-x_T)}{\sigma}\Big) \tag{184}$$

$$\frac{\partial f}{\partial x_i} = -\frac{\mathcal{N}\Big(-b\frac{h^T(x-x_T)}{\sigma},0,1\Big)}{Q\Big(-b\frac{h^T(x-x_T)}{\sigma}\Big)}\cdot b\frac{h^T}{\sigma} \tag{185}$$

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = -b\frac{h}{\sigma}\nabla_x \frac{\mathcal{N}\Big(-b\frac{h^T(x-x_T)}{\sigma},0,1\Big)}{Q\Big(-b\frac{h^T(x-x_T)}{\sigma}\Big)} \tag{186}$$

$$= \frac{1}{\sigma^2}\frac{b\frac{h^T(x-x_T)}{\sigma}\mathcal{N}\Big(-b\frac{h^T(x-x_T)}{\sigma},0,1\Big)\cdot Q\Big(-b\frac{h^T(x-x_T)}{\sigma}\Big)+\mathcal{N}^2\Big(-b\frac{h^T(x-x_T)}{\sigma},0,1\Big)}{Q^2\Big(-b\frac{h^T(x-x_T)}{\sigma}\Big)}hh^T \tag{187}$$

where we used $b^2 = 1$. Developing around $\hat{x}$, we find that

$$f(x) \approx f(\hat{x}) + \frac{\partial f}{\partial x_i}\Big|_{x=\hat{x}}(x-\hat{x}) + \frac{1}{2}(x-\hat{x})^T\frac{\partial^2 f}{\partial x_i \partial x_j}\Big|_{x=\hat{x}}(x-\hat{x}) \tag{188}$$

$$\begin{aligned}-\log Q\Big(-b\frac{h^T(x-x_T)}{\sigma}\Big) \approx& -\log Q\Big(-b\frac{h^T(\hat{x}-x_T)}{\sigma}\Big)\\ &-\frac{\mathcal{N}\Big(-b\frac{h^T(\hat{x}-x_T)}{\sigma},0,1\Big)}{Q\Big(-b\frac{h^T(\hat{x}-x_T)}{\sigma}\Big)}\cdot\frac{b}{\sigma}h^T(x-\hat{x})\\ &+\frac{1}{2\sigma^2}\frac{b\frac{h^T(\hat{x}-x_T)}{\sigma}\mathcal{N}\Big(-b\frac{h^T(\hat{x}-x_T)}{\sigma},0,1\Big)\cdot Q\Big(-b\frac{h^T(\hat{x}-x_T)}{\sigma}\Big)+\mathcal{N}^2\Big(-b\frac{h^T(\hat{x}-x_T)}{\sigma},0,1\Big)}{Q^2\Big(-b\frac{h^T(\hat{x}-x_T)}{\sigma}\Big)}(x-\hat{x})^T hh^T(x-\hat{x})\end{aligned} \tag{189}$$

For the special case $\hat{x} = x_T$, we obtain

$$-\log Q\Big(-b\frac{h^T(x-x_T)}{\sigma}\Big) \approx \log(2) - \frac{2b}{\sqrt{2\pi}\sigma}h^T(x-x_T) + \frac{1}{\pi\sigma^2}(x-\hat{x})^T hh^T(x-\hat{x}) \tag{190}$$

C Quadratic Taylor approximation of $-\log\left[Q\left(\frac{\tau(i)-h^T(x-x_T)}{\sigma}\right)-Q\left(\frac{\tau(i+1)-h^T(x-x_T)}{\sigma}\right)\right]$

We start by defining the abbreviations

$$u_i = \frac{\tau(i)-h^T(x-x_T)}{\sigma}, \quad u_{i+1} = \frac{\tau(i+1)-h^T(x-x_T)}{\sigma} \tag{191}$$

and note that

$$\frac{\partial u_i}{\partial x_k} = \frac{\partial u_{i+1}}{\partial x_k} = -\frac{h^T}{\sigma} \tag{192}$$

We then write the first and second derivatives (gradient and Hessian) as

$$f(x) = -\log[Q(u_i) - Q(u_{i+1})] \tag{193}$$

$$\frac{\partial f}{\partial x_k} = -\frac{1}{Q(u_i) - Q(u_{i+1})}\left(-\mathcal{N}(u_i,0,1)(-\frac{h^T}{\sigma}) - \left(-\mathcal{N}(u_{i+1},0,1)(-\frac{h^T}{\sigma})\right)\right) \tag{194}$$

$$= \frac{\mathcal{N}(u_{i+1},0,1) - \mathcal{N}(u_i,0,1)}{Q(u_i) - Q(u_{i+1})} \cdot \frac{h^T}{\sigma} \tag{195}$$

$$\frac{\partial^2 f}{\partial x_k \partial x_l} = \frac{h}{\sigma} \cdot \Bigg[\Big(\mathcal{N}(u_{i+1},0,1)(-u_{i+1}) - \mathcal{N}(u_i,0,1)(-u_i)\Big)\Big(Q(u_i) - Q(u_{i+1})\Big) - \Big(\mathcal{N}(u_{i+1},0,1) - \mathcal{N}(u_i,0,1)\Big) \cdot \Big(\mathcal{N}(u_{i+1},0,1) - \mathcal{N}(u_i,0,1)\Big)\Bigg] \Big/ \Big(Q(u_i) - Q(u_{i+1})\Big)^2 \cdot \Big(-\frac{h^T}{\sigma}\Big) \tag{196}$$

$$= \frac{\Big(\mathcal{N}(u_{i+1},0,1)\cdot u_{i+1} - \mathcal{N}(u_i,0,1)\cdot u_i\Big)\Big(Q(u_i) - Q(u_{i+1})\Big) + \Big(\mathcal{N}(u_{i+1},0,1) - \mathcal{N}(u_i,0,1)\Big)^2}{\Big(Q(u_i) - Q(u_{i+1})\Big)^2} \cdot \frac{hh^T}{\sigma^2} \tag{197}$$

Developing around $\hat{x}$, we find that

$$f(x) \approx f(\hat{x}) + \frac{\partial f}{\partial x_k}\Big|_{x=\hat{x}}(x-\hat{x}) + \frac{1}{2}(x-\hat{x})^T \frac{\partial^2 f}{\partial x_k \partial x_l}\Big|_{x=\hat{x}}(x-\hat{x}) \tag{198}$$

$$\begin{aligned} -\log[Q(u_i) - Q(u_{i+1})] \approx & -\log[Q(\hat{u}_i) - Q(\hat{u}_{i+1})] \\ & + \frac{\mathcal{N}(\hat{u}_{i+1},0,1) - \mathcal{N}(\hat{u}_i,0,1)}{Q(\hat{u}_i) - Q(\hat{u}_{i+1})} \cdot \frac{h^T}{\sigma}(x-\hat{x}) \\ & + \frac{1}{2\sigma^2}\frac{\Big(\mathcal{N}(u_{i+1},0,1)\cdot u_{i+1} - \mathcal{N}(u_i,0,1)\cdot u_i\Big)\Big(Q(u_i) - Q(u_{i+1})\Big) + \Big(\mathcal{N}(u_{i+1},0,1) - \mathcal{N}(u_i,0,1)\Big)^2}{\Big(Q(u_i) - Q(u_{i+1})\Big)^2} \\ & \cdot (x-\hat{x})^T hh^T (x-\hat{x}) \end{aligned} \tag{199}$$

with

$$\hat{u}_i = \frac{\tau(i) - h^T(\hat{x} - x_T)}{\sigma}, \quad \hat{u}_{i+1} = \frac{\tau(i+1) - h^T(\hat{x} - x_T)}{\sigma} \tag{200}$$

as implied.

References

[1] G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed. Baltimore and London: Johns Hopkins University Press, 1996.

[2] F. A. Shah, A. Ribeiro, and G. B. Giannakis, "Bandwidth-constrained MAP estimation for wireless sensor networks," in *Proceedings of the 39th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Oct. 30 – Nov. 2, 2005, pp. 215–219.

[3] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks – Part I: Gaussian case," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1131–1143, Mar. 2006.

[4] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[5] E. J. Msechu, S. I. Roumeliotis, A. Ribeiro, and G. B. Giannakis, "Decentralized quantized kalman filtering with scalable communication cost," *IEEE Transactions on Signal Processing*, vol. 56, pp. 3727–3741, Aug. 2008.

[6] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.

[7] J. Max, "Quantizing for minimum distortion," *IEEE Transactions on Information Theory*, vol. 6, no. 1, pp. 7–12, Mar. 1960.

[8] A. Prékopa, "Logarithmic concave measures and related topics," in *Stochastic Programming*, M. A. H. Dempster, Ed. Academic Press, 1980, pp. 63–82.

[9] E. J. Msechu, S. I. Roumeliotis, A. Ribeiro, and G. B. Giannakis, "Distributed iteratively quantized Kalman filtering for wireless sensor networks," in *Conference Record of the 41st Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 4–7, 2007, pp. 646–650.

[10] A. Martinelli, "Improving the precision on multi robot localization by using a series of filters hierarchically distributed," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Diego, CA, Oct. 29 – Nov. 2, 2007, pp. 1053–1058.

[11] P. W. Gibbens, G. M. W. M. Dissanayake, and H. F. Durrant-Whyte, "A closed form solution to the single degree of freedom simultaneous localisation and map building (SLAM) problem," in *Proceedings of the 39th IEEE Conference on Decision and Control*, vol. 1, Sydney, Dec. 2000, pp. 191–196.

[12] E. J. Msechu, A. Ribeiro, S. I. Roumeliotis, and G. B. Giannakis, "Distributed Kalman filtering based on quantized innovations," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, NV, USA, Mar./Apr. 2008, pp. 3293–3296.