Coordinates Bending for Measurement Update

Renato Zanetti*, and Felipe Giraldo-Grueso*
*Department of Aerospace Engineering and Engineering Mechanics
The University of Texas at Austin, Austin, Texas 78712
Email: {renato, fgiraldo}@utexas.edu

***Abstract*—A methodology to assimilate nonlinear measurement data is proposed. The approach relies on a nonlinear coordinate transformation to induce bending of the coordinate axes and skewness of the posterior density function. The proposed algorithm is tested in both single and recursive measurements processing scenarios.**

I. Introduction

In Bayesian statistics, three possible estimates of the value of an unknown state or parameter from available data are Minimum Mean Square Error (MMSE), Linear Minimum Mean Square Error (LMMSE), and Maximum A Posteriori (MAP). It is well known that in the presence of a Gaussian prior and linear/Gaussian measurements the MMSE, LMMSE, and MAP all coincide and can be calculated with the Kalman filter (KF) update equations [1]. In the presence of nonlinearities and/or non-Gaussian distributions, these three estimates no longer coincide and are, typically, not available in closed form. Rather, approximations or numerical solutions are needed. The Extended (EKF), Unscented (UKF), and Quadrature KFs, among others, approximate the LMMSE estimator in the presence of nonlinearities [2], [3], [4]; the particle, point-mass, and Gaussian sum filters, among others, approximate the MMSE estimator [5], [6], [7]. Finally, the Iterated EKF (IEKF), BRUF, and DAMAP are examples of algorithms that approximate the MAP estimator [2], [8], [9].

For nonlinear problems, whether it is the UKF, IEKF, or other KF variant, they do not provide a full parameterization of the posterior distribution, they just provide an estimate $\hat{x}$ and a covariance matrix P. A possible interpretation is that the probability distribution of the state x conditioned on the data y is approximately Gaussian with mean $\hat{x}$ and covariance matrix P; this interpretation leads to the so-called Gaussian-assumed filters [10]. Equivalently, let $T\Sigma T^T = P$ and $\eta \sim \mathcal{N}(0, \Sigma)$, where T is orthogonal and Σ diagonal; the KF approximates the conditional state via a linear (affine) transformation $L(\eta)$ of the Gaussian random variable η

$$x|y = L(\eta) = \hat{x} + T\eta. \quad (1)$$

This approximation comes with the drawback that many distributions are not accurately represented as Gaussian. Eq. (1) is a roto-translation into a Cartesian system aligned with the covariance matrix axes; this work generalizes Eq. (1) to a quadratic to allow for bending of the coordinate axes to capture the skewness of the posterior.

Identify applicable funding agency here. If none, delete this.

II. Background and Algorithm Overview

In the presence of nonlinear measurement functions, the KF update equations are no longer applicable. The EKF makes a first order approximation as it linearizes the nonlinear measurement. For more severe nonlinearities, the EKF might fail, and possible alternatives include second order methods, such as the IEKF and the Gaussian Second Order Filter (GSOF) [11]. Assume the prior distribution is Gaussian, $p(x) = \mathcal{N}(x; \bar{x}, \bar{P})$, and the measurement error is Gaussian and additive, $p(y|x) = \mathcal{N}(y; h(x), R)$, then the MAP coincides with the Least-Squares (LS) estimate and can be calculated with the Gauss-Newton method, which approximates the second order partials to find the LS solution. The IEKF is equivalent to Gauss-Newton [12]. The IEKF calculates the estimate (MAP) arbitrarily well via numerical optimization, the covariance, however, is calculated via linearization and, in the presence of large nonlinearities, can be an inaccurate representation of the true uncertainty.

While the IEKF is a MAP estimator, the GSOF approximates the LMMSE. It does so by also assuming a Gaussian prior and approximating the nonlinear measurement function with a second order Taylor series. Let $h^i(x)$ be the i-th component of the measurement function, then

$$h^i(x) \approx h^i(\bar{x}) + \sum_k h^{i\prime}_k(\bar{x})\ (x^k - \bar{x}^k) + \\ + \frac{1}{2}\sum_{k\ell} h^{i\prime\prime}_{k\ell}(\bar{x})\ (x^k - \bar{x}^k)(x^\ell - \bar{x}^\ell) \quad (2)$$

$$\mathbb{E}\left\{y^i\right\} \approx h^i(\bar{x}) + \frac{1}{2}\sum_{k\ell} h^{i\prime\prime}_{k\ell}(\bar{x})\ \bar{P}^{k\ell} \quad (3)$$

$$\mathrm{Cov}(x^i, y^j) \approx \sum_k h^{j\prime}_k(\bar{x})\ \bar{P}^{ki} \quad (4)$$

$$\mathrm{Cov}(y^i, y^j) \approx \sum_{k\ell} h^{i\prime}_k(\bar{x})\ h^{j\prime}_\ell(\bar{x})\ \bar{P}^{k\ell} + \\ + \frac{1}{2}\sum_{k\ell mn} h^{i\prime\prime}_{k\ell}(\bar{x})\ \bar{P}^{\ell m} h^{j\prime\prime}_{mn}(\bar{x})\ \bar{P}^{nk} \quad (5)$$

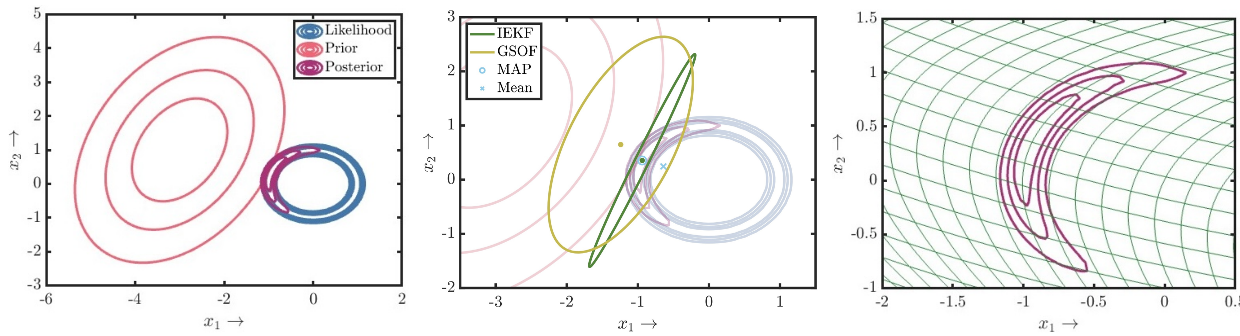

Fig. 1. A nonlinear estimation problem. The left figure shows the prior, likelihood and true posterior. The middle figure (zoomed in) shows the IEKF and GSOF estimates and covariances. The right figure (further zoomed in) shows the proposed bending of the coordinate axes.

where

$$h^{i\prime}_k(\bar{x}) = \left.\frac{\partial h^i(x)}{\partial x_k}\right|_{\hat{x}}, \qquad h^{i\prime\prime}_{k\ell}(\bar{x}) = \left.\frac{\partial^2 h^i(x)}{\partial x_k \partial x_\ell}\right|_{\hat{x}} \tag{6}$$

Fig. 1 depicts a nonlinear estimation problem.[1] The GSOF has an inaccurate estimate but an adequate covariance. The IEKF, on the other hand, produces a very accurate MAP estimate but an inadequate covariance.

As previously mentioned, Gaussian-assumed filters can be interpreted as a linear transformation between a Gaussian with axes aligned with the coordinate system and the state's Probability Density Function (PDF). Let the Gaussian-assumed PDF be $x \sim \mathcal{N}(\hat{x}, P)$, with $T\Sigma T^T = \hat{P}$ and let $\eta \sim \mathcal{N}(0, \Sigma)$, then

$$x = L(\eta) = \hat{x} + T\eta.$$

This work proposes to generalize this approximation to introduce more degrees of freedom in order to better representing a wider variety of distributions. Let $\eta \sim \mathcal{N}(0, \hat{P})$, the state will be approximated with the following transformation

$$x = \phi^{-1}(\eta) \tag{7}$$

$$\phi^i(x) = (x^i - \hat{x}^i) + \frac{1}{2}\sum_{jk} B^i_{jk}\,(x^j - \hat{x}^j)\,(x^k - \hat{x}^k)$$

where the super/subscripts indexes indicate the vector/function/tensor components. In addition to the parameters we had before (the estimate $\hat{x}$ and a symmetric positive definite matrix $\hat{P}$), we have the third order tensor B. The quadratic term allows for the transformed Gaussian to be skewed and when all B^i_{jk} coefficients are equal to zero the transformation is linear and the distribution reduces to a Gaussian. The transformation ϕ^{-1} is a change of variables that allows for bending of the coordinate axes as shown in Fig. 1.

Notice that $\hat{x}$ and $\hat{P}$ no longer represent the mean and covariance matrix of x. The defined transformation ϕ might not be actually invertible. However, we are only interested in state realizations near $\hat{x}$, and we assume the function is locally invertible there. The proposed measurement update is summarized in Algorithm 1.

[1]More information about this example is provided in Section V.

Algorithm 1: Posterior PDF Calculation

Inputs : The prior PDF (parameterized by $\bar{x}$, $\bar{P}$, and $\bar{B}$) and the likelihood function.

Outputs: The posterior PDF (parameterized by $\hat{x}$, $\hat{P}$, and $\hat{B}$).

Step 1. Calculate the un-normalized negative log-posterior $w(x)$ (Eq. (14)) and its first, second, and third order partial derivatives.

Step 2. Compute the MAP estimate $\hat{x}$ (minimum of $w(x)$, e.g., using IEKF) and evaluate partials at $\hat{x}$. Calculate matrix T to diagonalize the Hessian (Eq. (17)).

Step 3. Calculate the partials of the roto-translated negative log-posterior $\gamma(x)$ using Eq. (20), (21), and (22).

Step 4. Obtain matrix A with $A_{ii} = 0$ and with off-diagonal terms from the roots of Eq. (42) or with the approximation in Eq. (43).

Step 5. Compute posterior parameters $\hat{x}$, $\hat{P}$, and $\hat{B}$ using Eq. (50), Eq. (51), and Eq. (52).

III. Algorithm Description

Let $\eta \sim \mathcal{N}(0, \bar{P})$, as stated before we will define the prior distribution of the state as $x = \phi^{-1}(\eta)$, so that

$$p_x(x) = p_\eta(\phi(x))\,\left|\det J(x)\right| \tag{8}$$

where $J(x)$ is the Jacobian of $\phi(x)$. The components of ϕ and its Jacobian are given by

$$\phi^i(x) = (x^i - \bar{x}^i) + \frac{1}{2}\sum_{jk}\bar{B}^i_{jk}(x^j - \bar{x}^j)(x^k - \bar{x}^k) \tag{9}$$

$$\begin{aligned} J_{ij}(x) = \frac{\partial \phi^i}{\partial x^j} &= \delta_{ij} + \frac{1}{2}\sum_k (\bar{B}^i_{jk} + \bar{B}^i_{kj})\,(x^k - \bar{x}^k) \\ &= \delta_{ij} + \sum_k \mathrm{Sym}(\bar{B}^i_{jk})\,(x^k - \bar{x}^k) \end{aligned} \tag{10}$$

The prior distribution of x is defined by three parameters:

1) A vector $\bar{x}$
2) A symmetric positive definite matrix $\bar{P}$ (although carrying $\bar{P}^{-1}$ is more convenient)
3) A third order tensor $\bar{B}$

Define $\varphi(x) = -\log|\det J(x)|$, hence

$$p_x(x) \propto \exp\left\{-\frac{1}{2}\phi(x)^{\mathrm{T}}\bar{P}^{-1}\phi(x) - \varphi(x)\right\} \tag{11}$$

Assume a measurement likelihood of the form

$$p(y|x) \propto \exp\left\{-\frac{1}{2}(y-h(x))^{\mathrm{T}}R^{-1}(y-h(x))\right\} \tag{12}$$

The posterior is given by

$$p(x|y) \propto p(y|x)\ p(x) \propto \exp\big(-w(x)\big) \tag{13}$$

$$w(x) = \frac{1}{2}\phi(x)^{\mathrm{T}}\bar{P}^{-1}\phi(x) + \varphi(x) + \frac{1}{2}(y-h(x))^{\mathrm{T}}R^{-1}(y-h(x)) \tag{14}$$

Constructing a closed-form recursive estimator for this nonlinear measurement is not possible, for instance the prior is not the conjugate prior. When the prior is Gaussian, a common assumption is to approximate the posterior also as Gaussian to enable a recursive approach. We will follow similar steps and approximate the posterior as also having distribution in the form of Eq. (8) and finding the three parameters of it to match Eq. (14) as closely as possible. We will approximate the posterior by matching first, second, and third order derivatives between $w(x)$ and the log-PDF of the assumed posterior. In particular, we will match these derivatives evaluated at the posterior's estimate, which we will select to be the MAP.

$$\hat{x} = \arg\min_x \Big(w(x)\Big) \tag{15}$$

The derivatives of $w(x)$ can be found analytically or via symbolic or numerical differentiation. Notice that for any given likelihood, they have to be derived only once and then evaluated at run time.

Since $\hat{x}$ is the MAP, the Jacobian of $w(x)$ evaluated at $\hat{x}$ is zero and the Hessian evaluated at $\hat{x}$ is symmetric positive definite. The partials of $w(x)$ are derived in the appendix and given by:

$$w'(\hat{x}) = \left.\frac{\partial w(x)}{\partial x}\right|_{\hat{x}} = 0 \tag{16}$$

$$w''(\hat{x}) = \left.\frac{\partial^2 w(x)}{\partial x \partial x^{\mathrm{T}}}\right|_{\hat{x}} = T^{\mathrm{T}}\Sigma T \tag{17}$$

$$w'''_{ijk}(\hat{x}) = \left.\frac{\partial^3 w(x)}{\partial x^i \partial x^j \partial x^k}\right|_{\hat{x}} \tag{18}$$

where T is orthogonal and Σ diagonal with positive elements on the diagonal.

While we could try to calculate the posterior directly by matching the above partials, this will result in a large system of nonlinear algebraic equations to solve. The third order partials are n_x^3 equations, which are likely only solvable in very low dimension. Rather, we will perform a roto-translation $\xi = T(x-\hat{x})$ such that $p(\xi|y)$ has a maximum at the origin and the Hessian evaluated at the origin is diagonal and equal to Σ.

$$\gamma(\xi) = w(T^{\mathrm{T}}\xi + \hat{x}) \tag{19}$$

$$\gamma'(0) = w'(\hat{x})T^{\mathrm{T}} = 0, \tag{20}$$

$$\gamma''(0) = Tw''(\hat{x})T^{\mathrm{T}} = \Sigma \tag{21}$$

$$\gamma'''_{ijk}(0) = \sum_{rs\ell} w'''_{rs\ell}(\hat{x})\ T_{ir}T_{js}T_{k\ell} \tag{22}$$

The partials of γ evaluated at zero can be calculated exactly, and we will choose the parameters $\hat{P}$ and $\hat{B}$ of the posterior PDF to match them. However, for the third order derivatives, we will only match $\gamma'''_{ijj}(0)$, reducing the number of equations/unknowns of the third order tensor from n_x^3 to n_x^2.

Let $\zeta \sim \mathcal{N}(0, W)$, we are again performing a transformation $\xi|y = g^{-1}(\zeta)$, so that

$$p_{\xi|y}(\xi) = p_\zeta(g(\xi))\ \big|\det J(\xi)\big| \tag{23}$$

where $J(\xi)$ is now the Jacobian of $g(\xi)$. Let

$$g^i(\xi) = \xi^i + \frac{1}{2}\sum_j A_{ij}(\xi^j)^2 \tag{24}$$

$$J_{ij}(\xi) = \delta_{ij} + A_{ij}\xi^j \tag{25}$$

We will use the log-det Taylor series expansion:

$$\begin{aligned}-\psi(\xi) &= \log\det J(\xi) = \log\det(I+\Delta)\\ &= \log\det I + \mathrm{tr}\big(\Delta\big) - \tfrac{1}{2}\ \mathrm{tr}\big(\Delta^2\big) + \dots\\ &= \mathrm{tr}\big(\Delta\big) - \tfrac{1}{2}\ \mathrm{tr}(\Delta^2) + \tfrac{1}{3}\ \mathrm{tr}\big(\Delta^3\big) + \dots\\ &= \sum_r A_{rr}\xi^r - \sum_{rs}\tfrac{1}{2}A_{rs}A_{sr}\xi^r\xi^s +\\ &\quad + \sum_{rs\ell}\tfrac{1}{3}A_{rs}A_{s\ell}A_{\ell r}\xi^r\xi^s\xi^\ell + \dots\end{aligned} \tag{26}$$

Taking partials

$$\psi'_i(0) = -A_{ii} \tag{27}$$

$$\psi''_{ij}(0) = A_{ij}A_{ji} \tag{28}$$

$$\psi'''_{ijk}(0) = -\tfrac{1}{3}\Big(A_{ik}A_{kj}A_{ji} + A_{ij}A_{jk}A_{ki} + A_{ki}A_{ij}A_{jk} + A_{kj}A_{ji}A_{ik} + A_{ji}A_{ik}A_{kj} + A_{jk}A_{ki}A_{ij}\Big) \tag{29}$$

When $j = k$, the third order partials are:

$$\psi'''_{ijj}(0) = -2A_{ij}A_{jj}A_{ji} \tag{30}$$

therefore

$$\begin{aligned}\log p_{\xi|y}(\xi) &= \log p_\zeta(g(\xi)) + \log\big|\det J(\xi)\big|\\ &= -v(\xi) - \psi(\xi) + K\end{aligned} \tag{31}$$

The derivatives of $v(\xi) = \frac{1}{2}g(\xi)^{\mathrm{T}}W^{-1}g(\xi)$ are derived in the appendix and given by:

$$v_i'(0) = 0 \tag{32}$$

$$v_{ij}''(0) = W_{ij}^{-1} \tag{33}$$

$$v_{ijj}'''(0) = \sum_s \left(W_{is}^{-1} A_{sj} + 2W_{js}^{-1} A_{si} \delta_{ij} \right) \tag{34}$$

We will choose W and A to match the partials of γ evaluated at zero:

$$v_i'(0) + \psi_i'(0) = \gamma_i'(0) \tag{35}$$

$$v_{ij}''(0) + \psi_{ij}''(0) = \gamma_{ij}''(0) \tag{36}$$

$$v_{ijj}'''(0) + \psi_{ijj}'''(0) = \gamma_{ijj}'''(0) \tag{37}$$

Starting with the first order derivatives in Eq. (35) and noticing that $\gamma_i'(0) = v_i'(0) = 0$, we have that

$$\psi_i'(0) = 0 \tag{38}$$

Combining Eq. (38) and Eq. (27) we have that

$$A_{ii} = 0 \quad \forall i \tag{39}$$

For the second order derivatives, combining Eq. (36), Eq. (33), and Eq. (28) we obtain these conditions:

$$W_{ij}^{-1} = \gamma_{ij}''(0) - A_{ij} A_{ji} \tag{40}$$

For the third order partials, combining Eq. (37), Eq. (34), and Eq. (30) we have

$$\begin{aligned} \gamma_{ijj}'''(0) &= -2A_{ji}A_{ij} + \sum_s \left(W_{is}^{-1} A_{sj} + 2W_{js}^{-1} A_{si} \delta_{ij} A_{jj} \right) \\ &= -2A_{ji}A_{ij} + \sum_s W_{is}^{-1} A_{sj} \end{aligned} \tag{41}$$

the last equality holds because $A_{ii} = 0$.

We have $n^2 - n$ unknowns which are the off-diagonal elements of A. We have $n^2 - n$ equations which are Eq. (41) for $i \neq j$. Plugging Eq. (40) into Eq. (41) we have that for $i \neq j$

$$\gamma_{ijj}'''(0) = -2A_{ji}A_{ij} + \sum_s \left(\gamma_{is}''(0) - A_{is}A_{si} \right) A_{sj}$$

Since γ'' is diagonal and $A_{ii} = 0$, we have that for $i \neq j$

$$\gamma_{ijj}'''(0) = \gamma_{ii}''(0) A_{ij} - 2A_{ji}A_{ij} - \sum_{s \neq i,j} A_{is} A_{si} A_{sj} \tag{42}$$

This is a root finding problem: $r(A) = 0$. It could be solved, for example, using Newton's method. Notice that Eq. (42) is cubic in A, an initial value for Newton's method could be the solution to the linear approximation

$$A_{ij} \approx \frac{\gamma_{ijj}'''(0)}{\gamma_{ii}''(0)} \tag{43}$$

It is also possible to use this approximation to calculate A_{ij} directly and skip the iterations.

Having found A we calculate W with Eq. (40) and

$$\zeta = g(\xi) = g(T(x - \hat{x})) \tag{44}$$

$$\zeta \sim \mathcal{N}(0, W) \tag{45}$$

$$g^i(\xi) = \xi^i + \frac{1}{2} \sum_j A_{ij} (\xi^j)^2 \tag{46}$$

and since $\xi = T(x - \hat{x})$

$$g(T(x - \hat{x})) = T(x - \hat{x}) + b \tag{47}$$

$$\phi(x) = T^{\mathrm{T}} g(T(x - \hat{x})) \tag{48}$$

where the components of vectors b and ϕ are

$$\begin{aligned} b^i(x) &= \frac{1}{2} \sum_\ell A_{i\ell} \left(\sum_k T_{\ell k} (x^k - \hat{x}^k) \right)^2 \\ &= \frac{1}{2} \sum_{jk} \left(\sum_\ell A_{i\ell} T_{\ell j} T_{\ell k} \right) (x^j - \hat{x}^j)(x^k - \hat{x}^k) \end{aligned} \tag{49}$$

$$\begin{aligned} \phi^i(x) &= (x^i - \hat{x}^i) + \frac{1}{2} \sum_{jk} B_{jk}^i (x^j - \hat{x}^j)(x^k - \hat{x}^k) \\ &= (x^i - \hat{x}^i) + \sum_r T_{ri} b^r \end{aligned}$$

Finally, since $\hat{P} = T^{\mathrm{T}} W T$, the parameters of the posterior distribution are

$$\hat{x} = \text{MAP} \tag{50}$$

$$\hat{P}_{ij}^{-1} = w_{ij}''(\hat{x}) - \sum_{pq} T_{pi} A_{pq} A_{qp} T_{qj} \tag{51}$$

$$\hat{B}_{jk}^i = \sum_{r\ell} A_{r\ell} T_{ri} T_{\ell j} T_{\ell k} \tag{52}$$

IV. Inverse function and statistics

To recursively calculate the MAP estimate it is not necessary to invert ϕ. However, if we are interested in the statistics of the conditional distribution, such as the mean or covariance, we do.

Let $\eta \sim \mathcal{N}(0, P)$, and, as before

$$x = \phi^{-1}(\eta)$$

$$\phi^i(x) = (x^i - m^i) + \frac{1}{2} \sum_{jk} B_{jk}^i \, (x^j - m^j) \, (x^k - m^k)$$

therefore

$$\mathbb{E}\{x\} = \mathbb{E}\left\{\phi^{-1}(\eta)\right\} \tag{53}$$

$$\mathrm{Cov}\{x\} = \mathrm{Cov}\{\phi^{-1}(\eta)\} \tag{54}$$

Define $u = x - m$, then

$$\eta = \phi(x) = \phi(u + m) = \rho(u) \tag{55}$$

$$x = \phi^{-1}(\eta) = \rho^{-1}(\eta) + m \tag{56}$$

$$\mathbb{E}\{x\} = \mathbb{E}\{u\} + m = \mathbb{E}\left\{\rho^{-1}(\eta)\right\} + m \tag{57}$$

$$\mathrm{Cov}\{x\} = \mathrm{Cov}\{u\} = \mathrm{Cov}\{\rho^{-1}(\eta)\} \tag{58}$$

We will invert ρ

$$\rho^{-1}(\eta) = \beta(\eta) \tag{59}$$

$$\rho^i(u) = u^i + \frac{1}{2}\sum_{jk} B^i_{jk}\ u^j\ u^k \tag{60}$$

Which is not possible in closed form, so we will write the Maclaurin series of the inverse:

$$\beta^i(\eta) = \sum_{p=1}^{\infty} \pi^i_{(p)}(\eta) \tag{61}$$

where $\pi^i_{(p)}$ is a p-th order polynomial in η. By definition of inverse, $\rho(\beta(\eta)) = \eta$, and

$$\begin{aligned}\rho^i(\beta(\eta)) = &\sum_{p=1}^{\infty}\pi^i_{(p)}(\eta) + \\ &+ \frac{1}{2}\sum_{jk} B^i_{jk}\Big(\sum_{p=1}^{\infty}\pi^j_{(p)}(\eta)\Big)\ \Big(\sum_{p=1}^{\infty}\pi^k_{(p)}(\eta)\Big) \\ = &\ \pi^i_{(1)}(\eta) + \sum_{p=2}^{\infty}\Big\{\pi^i_{(p)}(\eta) + \\ &+ \frac{1}{2}\sum_{jk} B^i_{jk}\sum_{s+r=p}\Big(\pi^j_{(r)}(\eta)\Big)\ \Big(\pi^k_{(s)}(\eta)\Big)\Big\}\end{aligned}$$

It immediately follows that

$$\pi^i_{(1)} = \eta^i \tag{62}$$

$$\pi^i_{(p)} = -\frac{1}{2}\sum_{jk} B^i_{jk}\sum_{s+r=p}\pi^j_{(s)}\ \pi^k_{(r)} \tag{63}$$

the polynomials of order two and three are:

$$\pi^i_{(2)} = -\frac{1}{2}\sum_{jk} B^i_{jk}\pi^j_{(1)}\ \pi^k_{(1)} = -\frac{1}{2}\sum_{jk} B^i_{jk}\eta^j\eta^k \tag{64}$$

$$\begin{aligned}\pi^i_{(3)} &= -\frac{1}{2}\sum_{jk} B^i_{jk}\Big(\pi^j_{(1)}\ \pi^k_{(2)} + \pi^j_{(2)}\ \pi^k_{(1)}\Big) \\ &= \frac{1}{4}\sum_{jkrs} B^i_{jk}\big(B^j_{rs}\ \eta^k + B^k_{rs}\ \eta^j\big)\ \eta^r\eta^s\end{aligned} \tag{65}$$

We have that

$$x^i = m^i + \sum_{p=1}^{\infty}\pi^i_{(p)}(\eta) = m^i + \eta + \sum_{p=2}^{\infty}\pi^i_{(p)}(\eta) \tag{66}$$

$$\mathbb{E}\left\{x^i\right\} = m^i + \sum_{p=1}^{\infty}\mathbb{E}\left\{\pi^i_{(p)}(\eta)\right\} \tag{67}$$

$$\mathbb{E}\left\{(x^i - m^i)(x^k - m^k)\right\} = \sum_{p,q=1}^{\infty}\mathbb{E}\left\{\pi^i_{(p)}(\eta)\ \pi^k_{(q)}(\eta)\right\} \tag{68}$$

Truncating to order two

$$x^i \approx m^i + \eta^i - \frac{1}{2}\sum_{jk} B^i_{jk}\eta^j\eta^k \tag{69}$$

$$\mathbb{E}\left\{x^i\right\} \approx m^i - \frac{1}{2}\sum_{jk} B^i_{jk}\ P_{jk} \tag{70}$$

$$\begin{aligned}\mathbb{E}\left\{(x^i - m^i)(x^\ell - m^\ell)\right\} &\approx \mathbb{E}\left\{\eta^i\eta^\ell\right\} + \\ &+ \frac{1}{4}\sum_{jk}\sum_{rs} B^i_{jk}\ B^\ell_{rs}\ \mathbb{E}\left\{\eta^j\eta^k\eta^r\eta^s\right\}\end{aligned} \tag{71}$$

$$= P_{ij} + \frac{1}{4}\sum_{jkrs} B^i_{jk}\ B^\ell_{rs}\ \big(P_{jk}P_{rs} + P_{jr}P_{ks} + P_{js}P_{kr}\big) \tag{72}$$

where Isserlis theorem was used. Notice that Eq. (72) is not the covariance, because m is not the mean. Rather, if m is our estimate, then Eq. (72) is the Mean Square Error (MSE) which for the MAP estimator is, generally, biased. The covariance is equal to the MSE minus the square of the bias

$$\begin{aligned}\mathrm{Cov}(x^i, x^j) &= \mathbb{E}\left\{(x^i - \mathbb{E}\left\{x^i\right\})(x^\ell - \mathbb{E}\left\{x^\ell\right\})\right\} \\ &= \mathbb{E}\left\{(x^i - m^i)(x^\ell - m^\ell)\right\} + \\ &\quad - (m^i - \mathbb{E}\left\{x^i\right\})(m^\ell - \mathbb{E}\left\{x^\ell\right\})\end{aligned} \tag{73}$$

V. Example

This section tests the proposed algorithm in a non-linear measurement scenario, as shown in Fig. 1. The first part of the example incorporates one measurement, while the second part recursively incorporates 100 measurements. The example starts from a prior distribution and likelihood given by

$$x \sim \mathcal{N}(\bar{x}, \bar{P}), \quad m = \begin{bmatrix}-3\\1\end{bmatrix}, \quad P = \begin{bmatrix}2 & 1\\1 & 4\end{bmatrix},$$

$$p(y|x) = \mathcal{N}(y; \sqrt{x^{\mathrm{T}}x}, R), \quad y = 0.99, \quad R = 0.01.$$

Since the prior is Gaussian, note that $\bar{B} = 0$. Using the proposed algorithm, the posterior PDF is obtained and parametrized by $\hat{x}$, $\hat{P}$, and $\hat{B}$. To obtain the statistics of the posterior PDF, the function ϕ is inverted using the procedure described in Section IV.

Figure 2 shows the mean and covariance of the approximated posterior distribution and Fig. 3 shows the bended axes of the transformed coordinate system. To obtain the mean and covariance the second order approximations of the inverse of ϕ is used. The axes are obtained by transforming the axes of the coordinate system through the inverse of ϕ using the second, third, and fourth order approximations. The figure shows that the approximated posterior PDF captures the skewness of the conditional distribution, due to the quadratic transformation. Additionally, the second, third, and fourth order approximations yield similar results.

Figure 4 shows the first two moments of the approximated posterior distribution, compared with the other filters presented in Fig. 1. To calculate these moments, Eq. (70) and Eq. (73) are used with the obtained $\hat{x}$, $\hat{P}$, and $\hat{B}$. As previously mentioned, for this

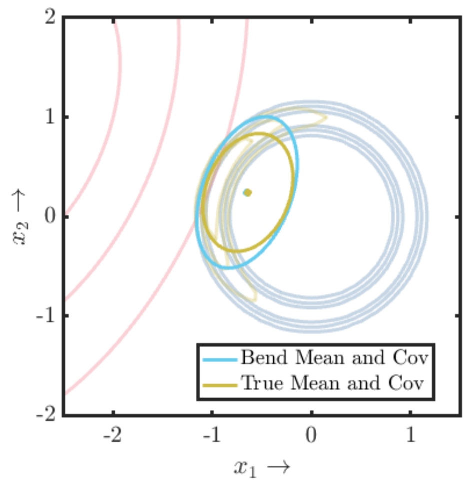

Fig. 2. Mean and Covariance of the approximated posterior.

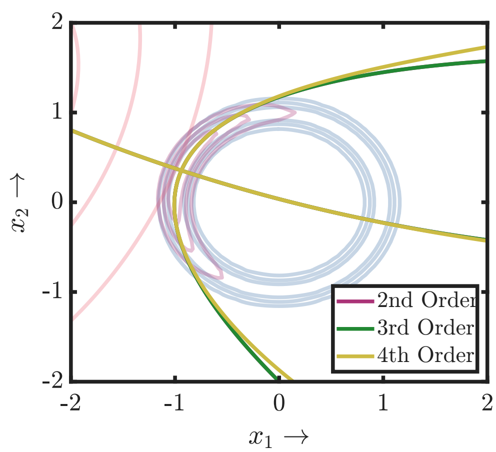

Fig. 3. Eigenaxes of the approximated posterior distribution.

example, the GSOF suffers from an inaccurate estimate but a reasonable covariance, while the IEKF has an accurate MAP estimate but a poor covariance. With the proposed algorithm, the approximated conditional distribution more accurately represent the true posterior distribution compared to the other filters.

The next example incorporates the above measurement 100 times, but now with an inflated variance $100R$, so that the posterior after 100 updates is exactly the same as for the one measurement case discussed above. The first update starts from the Gaussian prior and the Gaussian likelihood, producing a posterior $\hat{x}$, $\hat{P}$, and $\hat{B}$. The posterior becomes the prior for the next measurement which, together with the same likelihood, yields updated parameters. This process is repeated 100 times, and those of the final step are the parametrization of the approximated posterior.

Figure 5 shows the axes of the each posterior as measurement information is introduced. The second order approximation of the inverse of ϕ is shown. After

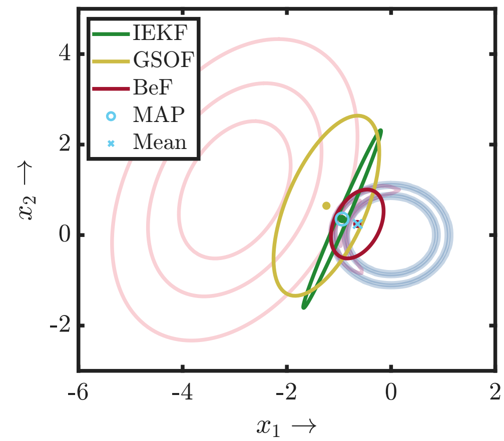

Fig. 4. Comparison of filters including the new quadratic update (BeF).

100 updates, the final axes closely align with the true posterior. Figure 6 shows the evolution of the first two moments of the approximated conditional distribution as measurement information is introduced. From the figure, it is also clear that the final two moments capture the true posterior distribution well.

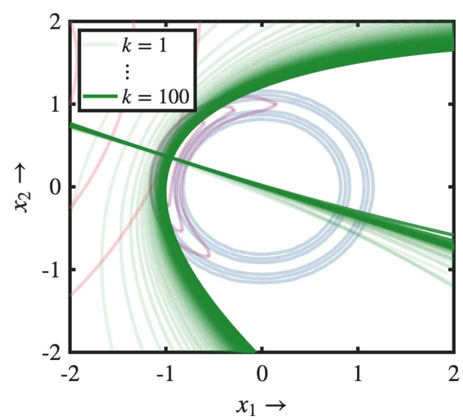

Fig. 5. Recursive shift of the eigenaxes of the approximated posterior distribution.

VI. Conclusions

This paper introduced a nonlinear update method based on a quadratic coordinate transformation, enabling the bending of axes and capturing skewness in the posterior distribution. By extending the Gaussian-assumed framework with a third-order tensor, the proposed approach provides a richer parameterization than traditional Gaussian-assumed filters. This added flexibility allows the algorithm to better approximate the true conditional distribution in nonlinear measurement scenarios.

Numerical experiments demonstrate that the method achieves a more accurate representation of both the mean and covariance compared to IEKF and GSOF, which tend to favor either the estimate or the uncertainty. The

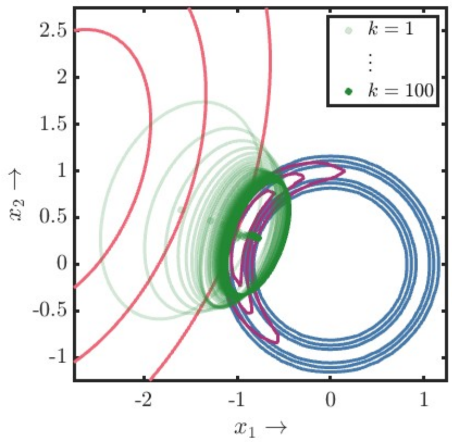

Fig. 6. Recursive shift of the first two moments of the approximated posterior distribution.

recursive implementation further shows that the approach can assimilate measurements recursively while maintaining computational tractability. These results suggest that the new update offers a promising alternative for nonlinear Bayesian estimation.

REFERENCES

[1] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*. Wiley, 2001.

[2] A. Gelb, Ed., *Applied Optimal Estimation*. Cambridge, MA: The MIT press, 1974.

[3] S. J. Julier and J. K. Uhlmann, "A New Extension of the Kalman Filter to Nonlinear Systems," in *Int. Symp. Aerospace/Defense Sensing, Simul. and Controls, Orlando, FL*, 1997.

[4] I. Arasaratnam, S. Haykin, and R. J. Elliot, "Discrete-time nonlinear filtering algorithms using gauss-hermite quadrature," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 953–977, May 2007.

[5] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, ser. Information Science and Statistics. New York: Springer–Verlag, 2001.

[6] R. S. Bucy and K. D. Senne, "Digital synthesis of non-linear filters," *Automatica*, vol. 7, no. 3, pp. 287–298, May 1971.

[7] D. Alspach and H. Sorenson, "Nonlinear bayesian estimation using gaussian sum approximations," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, August 1972, doi: 10.1109/SAP.1970.270017.

[8] K. Michaelson, A. A. Popov, and R. Zanetti, "Multiple data assimilation as an approximate maximum a posteriori estimator," *Computational Geosciences*, vol. 29, no. 17, 2025, https://doi.org/10.1007/s10596-025-10355-9.

[9] S. Servadio, R. Zanetti, and R. Armellin, "Maximum a posteriori estimation of hamiltonian systems with high order taylor polynomials," *The Journal of the Astronautical Sciences*, vol. 69, no. 2, pp. 511–536, April 2022, https://doi.org/10.1007/s40295-022-00304-4.

[10] S. Sarkka, *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.

[11] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, ser. Mathematics in Sciences and Engineering. New York, New York 10003: Academic Press, 1970, vol. 64.

[12] B. M. Bell and F. W. Cathey, "The Iterated Kalman FIlter Update as a Gauss-Newton Method," *IEEE Transactions on Automatic Control*, vol. 38, no. 2, pp. 294–297, February 1993, doi:10.1109/9.250476 .

APPENDIX

We have the following function

$$q(x) = \frac{1}{2} f(x)^T S f(x) = \frac{1}{2} \sum_{rs} S_{rs} f^r(x) f^s(x) \quad (74)$$

where $S = S^{\mathrm{T}}$. Dropping functional dependencies

$$q = \frac{1}{2} \sum_{rs} S_{rs} f^r f^s \quad (75)$$

We now take partials

$$q'_i = \frac{\partial q}{\partial x^i} = \sum_{rs} S_{rs} f_i^{r\prime} f^s$$

$$q''_{ij} = \frac{\partial^2 q}{\partial x^i \partial x^j} = \sum_{rs} S_{rs} \left(f_i^{r\prime} f_j^{s\prime} + f_{ij}^{r\prime\prime} f^s \right)$$

$$q'''_{ijk} = \sum_{rs} S_{rs} \left(f_{ik}^{r\prime\prime} f_j^{s\prime} + f_i^{r\prime} f_{jk}^{s\prime\prime} + f_{ij}^{r\prime\prime} f_k^{s\prime} + f_{ijk}^{r\prime\prime\prime} f^s \right)$$

$$= \sum_{rs} S_{rs} \left(f_i^{r\prime} f_{jk}^{s\prime\prime} + f_j^{r\prime} f_{ik}^{s\prime\prime} + f_k^{r\prime} f_{ij}^{s\prime\prime} + f_{ijk}^{r\prime\prime\prime} f^s \right)$$

where $S_{rs} = S_{sr}$ was used. From these equations, we will calculate the partials of $w(x)$ and of $v(\xi)$.

A. *Partials of $w(x)$*

We start by specializing the above result to Eq. (14):

$$w(x) = q(x) + \varphi(x) = \frac{1}{2} f(x)^T S f(x) + \varphi(x) \quad (76)$$

$$f(x) = \begin{bmatrix} \phi(x) \\ h(x) - y \end{bmatrix}, \qquad S = \begin{bmatrix} \bar{P}^{-1} & 0 \\ 0 & R^{-1} \end{bmatrix} \quad (77)$$

$$\varphi(x) = -\log |\det J(x)| \quad (78)$$

We will do it in three parts: partial of the prior, partials of the likelihood, and partials of the determinant.

1) Partials of Prior: Let's only consider the prior:

$$f^r = \phi^r = (x^r - \bar{x}^r) + \frac{1}{2} \sum_{\ell m} \bar{B}^r_{\ell m} (x^\ell - \bar{x}^\ell)\ (x^m - \bar{x}^m)$$

hence

$$f_i^{r\prime} = \delta_{ri} + \frac{1}{2} \sum_{\ell} \left(\bar{B}^r_{\ell i} + \bar{B}^r_{i \ell} \right) (x^\ell - \bar{x}^\ell)$$

$$= \delta_{ri} + \sum_{\ell} \mathrm{Sym}(\bar{B}^r_{\ell i})(x^\ell - \bar{x}^\ell) \quad (79)$$

$$f_{ij}^{r\prime\prime} = \frac{1}{2} \left(\bar{B}^r_{ji} + \bar{B}^r_{ij} \right) = \mathrm{Sym}(\bar{B}^r_{ij}) \quad (80)$$

$$f_{ijk}^{r\prime\prime\prime} = 0 \quad (81)$$

resulting in Eq. (82)–Eq. (84).

$$q'_i = \sum_{rs} \bar{P}^{-1}_{rs} \Big(\delta_{ri} + \sum_{\ell} \mathrm{Sym}(\bar{B}^r_{\ell i})(x^\ell - \bar{x}^\ell)\Big) \ \Big((x^s - \bar{x}^s) + \frac{1}{2}\sum_{\ell m} B^s_{\ell m}(x^\ell - \bar{x}^\ell)\ (x^m - \bar{x}^m)\Big) \tag{82}$$

$$q''_{ij} = \sum_{sr} \bar{P}^{-1}_{rs} \Bigg\{\Big(\delta_{ri} + \sum_{\ell} \mathrm{Sym}(\bar{B}^r_{\ell i})(x^\ell - \bar{x}^\ell)\Big)\Big(\delta_{sj} + \sum_{\ell} \mathrm{Sym}(\bar{B}^s_{\ell j})(x^\ell - \bar{x}^\ell)\Big) + \tag{83}$$

$$+ \mathrm{Sym}(\bar{B}^r_{ji})\ \Big((x^s - \bar{x}^s) + \frac{1}{2}\sum_{\ell m} \bar{B}^s_{\ell m}(x^\ell - \bar{x}^\ell)\ (x^m - \bar{x}^m)\Big)\Bigg\}$$

$$q'''_{ijk} = \sum_{sr} \bar{P}^{-1}_{rs} \Bigg\{\mathrm{Sym}(\bar{B}^s_{jk})\ \Bigg(\delta_{ri} + \sum_{\ell} \mathrm{Sym}(\bar{B}^r_{\ell i})(x^\ell - \bar{x}^\ell)\Bigg) + \tag{84}$$

$$+ \mathrm{Sym}(\bar{B}^s_{ki})\ \Bigg(\delta_{rj} + \sum_{\ell} \mathrm{Sym}(\bar{B}^r_{\ell j})(x^\ell - \bar{x}^\ell)\Bigg) + \mathrm{Sym}(\bar{B}^s_{ji})\ \Bigg(\delta_{rk} + \sum_{\ell} \mathrm{Sym}(\bar{B}^r_{\ell k})(x^\ell - \bar{x}^\ell)\Bigg)\Bigg\}$$

2) Partials of the Likelihood: Let's now consider the likelihood, so that

$$f^r = h^r - y^r \qquad f^{r\prime}_i = h^{r\prime}_i$$
$$f^{r\prime\prime}_{ij} = h^{r\prime\prime}_{ij} \qquad f^{r\prime\prime\prime}_{ijk} = h^{r\prime\prime\prime}_{ijk}$$

plugging these values in, we obtain

$$q'_i = \sum_{rs} R^{-1}_{rs} h^{r\prime}_i (h^s - y^s) \tag{85}$$

$$q''_{ij} = \sum_{rs} R^{-1}_{rs} \Big(h^{r\prime}_i h^{s\prime}_j + h^{r\prime\prime}_{ij}(h^s - y^s)\Big) \tag{86}$$

$$q'''_{ijk} = \sum_{rs} R^{-1}_{rs} \Big(h^{r\prime}_i h^{s\prime\prime}_{jk} + h^{r\prime}_j h^{s\prime\prime}_{ik} + h^{r\prime}_k h^{s\prime\prime}_{ij} + h^{r\prime\prime\prime}_{ijk}(h^s - y^s)\Big) \tag{87}$$

3) Partials of the Determinant: We have that

$$-\varphi(x) = \log|\det J(x)| \tag{88}$$

$$J_{ij}(x) = \delta_{ij} + \sum_r \mathrm{Sym}(\bar{B}^i_{jr})(x^r - \bar{x}^r) \tag{89}$$

$$\frac{\partial J_{ij}(x)}{\partial x^r} = \mathrm{Sym}(\bar{B}^i_{jr}) \tag{90}$$

$$\frac{\partial^2 J_{ij}(x)}{\partial x^r \partial x^s} = 0 \tag{91}$$

We will use the chain rule for matrices and the derivative of the inverse

$$\frac{\partial f(M(x))}{\partial x^i} = \mathrm{trace}\left(\frac{\partial f(M)}{\partial M}^{\mathrm{T}} \frac{\partial M(x)}{\partial x^i}\right) = \sum_{jk}\left(\frac{\partial f}{\partial M}^{\mathrm{T}}\right)_{jk} \frac{\partial M_{kj}}{\partial x^i} \tag{92}$$

$$\frac{\partial M(x)^{-1}}{\partial x^i} = -M^{-1}\frac{\partial M(x)}{\partial x^i} M^{-1} \tag{93}$$

noticing that

$$\frac{\partial \log|\det M|}{\partial M} = (M^{-1})^{\mathrm{T}} \tag{94}$$

we can apply Eq. (92) to $\varphi(J(x))$

$$\varphi'_i = -\sum_{m\ell} J^{-1}_{m\ell}\ \mathrm{Sym}(\bar{B}^m_{\ell i}) \tag{95}$$

$$\begin{aligned}\varphi''_{ij} &= -\sum_{m\ell} \frac{\partial J^{-1}_{m\ell}}{\partial x^j} \mathrm{Sym}(\bar{B}^m_{\ell i}) \\ &= \sum_{m\ell}\left(\sum_{rs} J^{-1}_{\ell r} \frac{\partial J_{rs}}{\partial x^j} J^{-1}_{sm}\right)\ \mathrm{Sym}(\bar{B}^m_{\ell i}) \\ &= \sum_{m\ell rs} J^{-1}_{\ell r}\ J^{-1}_{sm}\ \mathrm{Sym}(\bar{B}^r_{sj})\ \mathrm{Sym}(\bar{B}^m_{\ell i})\end{aligned} \tag{96}$$

$$\varphi'''_{ijk} = -\sum_{m\ell rsnp} \left(J^{-1}_{\ell n}\ J^{-1}_{pr}\ J^{-1}_{sm} + J^{-1}_{\ell r}\ J^{-1}_{sn}\ J^{-1}_{pm}\right) \mathrm{Sym}(\bar{B}^n_{pk})\ \mathrm{Sym}(\bar{B}^r_{sj})\ \mathrm{Sym}(\bar{B}^m_{\ell i}) \tag{97}$$

B. Partials of $v(\xi)$

The derivatives of $v = \frac{1}{2} g^{\mathrm{T}} W^{-1} g$ are:

$$v'_i = \sum_{rs} W^{-1}_{rs} g'_{ri} g_s \tag{98}$$

$$v''_{ij} = \sum_{rs} W^{-1}_{rs}\Big(g'_{ri}g'_{sj} + g''_{rij}g_s\Big) \tag{99}$$

$$v'''_{ijk} = \sum_{rs} W^{-1}_{rs}\Big(g'_{ri}g''_{sjk} + g'_{rj}g''_{sik} + g'_{rk}g''_{sij} + g'''_{rijk}g_s\Big) \tag{100}$$

since $W^{-1}_{rs} = W^{-1}_{sr}$. Using the definition of $g(\xi)$ from Eq. (24) we obtain

$$g'_{ri} = \delta_{ri} + A_{ri}\xi^i, \quad g''_{rij} = A_{ri}\delta_{ij}, \quad g'''_{rijk} = 0 \tag{101}$$

finally, evaluating them at 0

$$g_i(0) = 0 \qquad v(0) = 0 \tag{102}$$

$$g'_{ri}(0) = \delta_{ri} \qquad v'_i(0) = 0 \tag{103}$$

$$g''_{rij}(0) = A_{ri}\delta_{ij} \qquad v''_{ij}(0) = W^{-1}_{ij} \tag{104}$$

$$g'''_{rijk}(0) = 0 \tag{105}$$

$$v'''_{ijk}(0) = W^{-1}_{is}A_{sj}\delta_{jk} + W^{-1}_{js}A_{si}\delta_{ik} + W^{-1}_{ks}A_{si}\delta_{ij} \tag{106}$$