Geometric Formulation of Uncertainties and Estimation for Three-Dimensional Rotations

Weixin Wang
Advisor: Prof. Taeyoung Lee

Department of Mechanical and Aerospace Engineering
George Washington University, Washington DC

July 21, 2022

Overview

Attitude Estimation in Engineering

- 3D Rigid Body Attitude
  - The orientation of a reference frame fixed to the rigid body.
  - The space of 3D attitude: three dimensional special orthogonal group.

$$\mathrm{SO}(3) = \{R \in \mathbb{R}^{3\times 3} \mid RR^T = I_{3\times 3},\ \det(R) = 1\}.$$

  - Rotation matrix $R \in \mathrm{SO}(3)$: transform the coordinates of a vector from the body-fixed frame x^B to inertial frame x^I.

$$x^I = Rx^B$$

- Applications in Engineering
  - Alignment of two satellites: laser communication.
  - Attitude control for UAVs.
  - Inertial navigation.

Uncertainty for 3D Attitude

- State Estimation of a Dynamical System
  - Uncertainty Propagation: propagete the mean and covariance matrix of the state through the kinematic equations.
  - Measurement Update: use new measurements to correct propagation errors.
  - Covariance matrices are used to weigh the propagation and measurement noises.
- What is Covariance Matrix for 3D Rotation?
  - Covariance for lower dimensional parameterizations:
    - Three dimensional: Euler angles, (modified) Rodrigues parameters, etc.
    - Four dimensional: Unit quaternions.
  - Problems with three dimensional parameterizations:
    - Singularities: such as gimbal lock for Euler angles.
  - Problems with quaternions:
    - The unit norm constraint makes the covariance matrix singular.

Multiplicative Extended Kalman Filter

- Covariance Matrix for Attitude Error

$$q_t = q \otimes q\{\delta v\}.$$

  - $q_t, q \in \mathbb{S}^3$: true and estimated attitude represented by unit quaternions.
  - $q\{\delta v\} \in \mathbb{S}^3$: error attitude expressed as a rotation vector $\delta v \in \mathbb{R}^3$.
  - $\delta v \sim \mathcal{N}(0, \Sigma)$, where $\Sigma \in \mathbb{R}^{3\times 3}$ is the covariance matrix, representing the uncertainty of the attitude.
    - δv is three dimensional.
    - δv is usually small, so singularity is avoided.

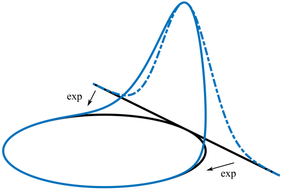
- Multiplicative Extended Kalman Filter (MEKF)
  - Using Gaussian distribution to model the uncertainty of attitude error.
  - Problems: not suitable for large uncertainty
    - Linearization error.
    - Wrapping error.

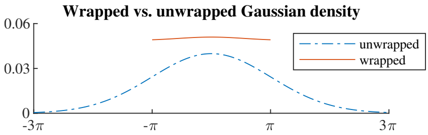
Multiplicative Extended Kalman Filter

- Covariance Matrix for Attitude Error

$$q_t = q \otimes q\{\delta v\}.$$

  - $q_t, q \in \mathbb{S}^3$: true and estimated attitude represented by unit quaternions.
  - $q\{\delta v\} \in \mathbb{S}^3$: error attitude expressed as a rotation vector $\delta v \in \mathbb{R}^3$.
  - $\delta v \sim \mathcal{N}(0, \Sigma)$, where $\Sigma \in \mathbb{R}^{3\times 3}$ is the covariance matrix, representing the uncertainty of the attitude.
    - δv is three dimensional.
    - δv is usually small, so singularity is avoided.





Large Attitude Uncertainty

- Large Attitude Uncertainty in Estimation Problems
  - Unknown initial conditions, e.g., unknown heading direction in a building.
  - Not properly observed degree of freedom, e.g., single reference direction that is slowly varying in the inertial frame.
  - Sensor failure, e.g., no GPS signal in a tunnel.
- Goal of This Dissertation
  - Deal with large attitude uncertainty using probability distribution defined intrinsically on $\mathrm{SO}(3)$ for attitude.
    - Matrix Fisher distribution on $\mathrm{SO}(3)$.
    - Bingham distribution on $\mathbb{S}^3$ for unit quaternions.
    - No singularities, no geometric constraints, able to model arbitrarily large attitude uncertainty.
  - Design Bayesian filters for classical estimation problems involving the attitude/pose of a rigid body.

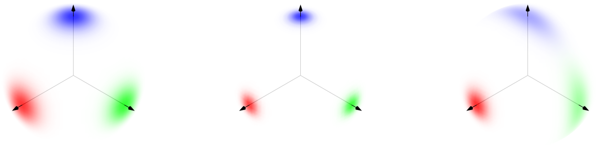
Matrix Fisher Distribution

- Matrix Fisher Distribution: Defined on $\mathrm{SO}(3)$
  - Construction: condition a Gaussian distribution from $\mathbb{R}^9$ to $\mathrm{SO}(3)$.
  - Density function for $R \sim \mathcal{M}(F)$:

$$p(R) = \frac{1}{c(F)} \exp\left\{ \mathrm{tr}\left(FR^T\right) \right\}.$$

    - $F \in \mathbb{R}^{3\times 3}$ is the parameter, $c(F) \in \mathbb{R}$ is the normalizing constant.

- Bingham Distribution: Defined on $\mathbb{S}^3$
  - Equivalent to the matrix Fisher distribution under the homomorphism from $\mathrm{SO}(3)$ to $\mathbb{S}^3$.



Properties

- Shape of the Density Function
  - proper singular value decomposition (pSVD) of the parameter: $F = USV^T$.
  - Mean attitude: $M = UV^T$ (uni-modal).
  - Principal axes:
    - Columns of U in the inertial frame.
    - Columns of V in the body-fixed frame of the mean attitude M.
  - Dispersion: $s_i + s_j$ specifies the dispersion along the k-th principal axis, for $i \neq j \neq k$.
  - Analogous to a Gaussian distribution.
- Maximum Likelihood Estimation (MLE) for Parameters
  - Moments

$$\mathrm{E}[R] = UDV^T, \qquad d_i = \frac{1}{c(S)} \frac{\partial c(S)}{\partial s_i}.$$

  - MLE for $U, V \in \mathrm{SO}(3)$: the pSVD of $\mathrm{E}[R] = USV^T$.
  - MLE for S: solving $d_i = \frac{1}{c(S)} \frac{\partial c(S)}{\partial s_i}$ from D.

Central Moments

- Motivation
  - Designing Bayesian filters using matrix Fisher distribution sometimes needs to evaluate its higher order moments.
- Central Moments for Matrix Fisher Distribution: $Q = U^T RV$

$$\mathrm{E}\left[Q_{i_1 j_1} \cdots Q_{i_n j_n}\right] = \frac{1}{c(S)} \left. \frac{\partial c(S+T)}{\partial T_{i_1 j_1} \cdots \partial T_{i_n j_n}} \right|_{T=0}.$$

  - $\mathrm{E}\left[Q_{i_1 j_1} \cdots Q_{i_n j_n}\right] = 0$ if $\{i_k, j_k\}_{k=1}^n$ has odd number of 1, 2, or 3.
  - $\mathrm{E}\left[Q_{i_1 j_1} \cdots Q_{i_n j_n}\right] = \mathrm{E}\left[Q_{j_1 i_1} \cdots Q_{j_n i_n}\right]$.
- Computation
  - $\mathrm{E}\left[Q_{i_1 j_1} \cdots Q_{i_n j_n}\right]$ is a linear combination of the derivatives of $c(S)$ up to n-th order.
  - The coefficients and derivatives of $c(S)$ can be calculated recursively.
  - The recursion starts from $c(S)$ and its first order derivatives $\frac{\partial c(S)}{\partial s_i}$.

Highly Concentrated Approximations

- Motivation
  - Evaluating the normalizing constant $c(S)$ and its derivatives is hard.
  - The MLE $d_i = \frac{1}{c(S)} \frac{\partial c(S)}{\partial s_i}$ is very hard to solve.
- Highly Concentrated in Three Degrees of Freedom [1]

Theorem

Let $R \sim \mathcal{M}(F)$, where $F = USV^T$ is the pSVD of F. Suppose $s_2 + s_3 \gg 0$. Let $Q = U^T R V = \exp(\hat{\eta})$, then $\eta \mathrel{\dot\sim} \mathcal{N}\left(0, (\operatorname{tr}(S) I_{3\times 3} - S)^{-1}\right)$.

  - The matrix Fisher distribution is approximated by a Gaussian distribution in $\mathbb{R}^3$.

[1] T. Lee, "Bayesian attitude estimation with approximate matrix Fisher distributions on SO(3)," in *Conference on Decision and Control*, 2018, pp. 5319–5325.

Highly Concentrated Approximations

- Motivation
  - Evaluating the normalizing constant $c(S)$ and its derivatives is hard.
  - The MLE $d_i = \frac{1}{c(S)}\frac{\partial c(S)}{\partial s_i}$ is very hard to solve.
- Highly Concentrated in Three Degrees of Freedom
  - Approximations of $c(S)$ and its derivatives ($i \neq j \neq k$):

$$c(S) \approx \frac{\operatorname{etr}(S)}{\sqrt{8\pi(s_1+s_2)(s_1+s_3)(s_2+s_3)}}$$

$$\frac{1}{c(S)}\frac{\partial c(S)}{\partial s_i} \approx 1 - \frac{1}{2}\left(\frac{1}{s_i+s_j} + \frac{1}{s_i+s_k}\right),$$

Highly Concentrated Approximations

- Motivation
  - Evaluating the normalizing constant $c(S)$ and its derivatives is hard.
  - The MLE $d_i = \frac{1}{c(S)} \frac{\partial c(S)}{\partial s_i}$ is very hard to solve.
- Highly Concentrated in Two Degrees of Freedom

Theorem

Let $R \sim \mathcal{M}(F)$, where $F = USV^T$ is the pSVD of F. Suppose $s_1 + s_3 \gg s_2 + s_3 \geq 0$. Let $Q = U^T RV = \exp(\hat{\eta}) \exp(\hat{\eta}')$, where $\eta = [0, \eta_2, \eta_3]^T$, and $\eta' = [\eta_1, 0, 0]^T$. Then $\eta_3 \approx \mathcal{VM}(0, s_2 + s_3)$, and $[\eta_2, \eta_3]^T \approx \mathcal{N}\left(0, \operatorname{diag}\left(\frac{1}{s_1+s_3}, \frac{1}{s_1+s_2}\right)\right)$, and they are approximately independent.

  - The matrix Fisher distribution is approximated by a combination of Gaussian distribution in $\mathbb{R}^2$, and a von Mises distribution on $\mathbb{S}^1$.

Highly Concentrated Approximations

- Motivation
  - Evaluating the normalizing constant $c(S)$ and its derivatives is hard.
  - The MLE $d_i = \frac{1}{c(S)}\frac{\partial c(S)}{\partial s_i}$ is very hard to solve.
- Highly Concentrated in Two Degrees of Freedom
  - Approximations of $c(S)$ and its derivatives ($j \in \{2,3\}$):

$$c(S) \approx \frac{\exp(s_1) I_0(s_2+s_3)}{2\sqrt{(s_1+s_2)(s_1+s_3)}},$$

$$\frac{1}{c(S)}\frac{\partial c(S)}{\partial s_1} \approx 1 - \frac{1}{2}\left(\frac{1}{s_1+s_2} + \frac{1}{s_1+s_3}\right),$$

$$\frac{1}{c(S)}\frac{\partial c(S)}{\partial s_j} \approx \frac{I_1(s_2+s_3)}{I_0(s_2+s_3)} - \frac{1}{2}\frac{1}{s_1+s_j},$$

Correlation Between $\mathrm{SO}(3)$ and $\mathbb{R}^n$

- Matrix Fisher Distribution
  - Defined on $\mathrm{SO}(3)$.
  - Cannot model the correlation between $\mathrm{SO}(3)$ and $\mathbb{R}^n$.
  - Correlation: transfer information from observed state to unobserved state.
- Examples
  - Attitude estimation: correlation between attitude and gyroscope bias.
  - Inertial navigation: correlation between attitude and position.
  - SLAM: correlation between attitude and landmark locations.

Correlation Between $\mathrm{SO}(3)$ and $\mathbb{R}^n$

- Existing Models
  - In Markley 2006[1], the matrix Fisher distribution on $\mathrm{SO}(3)$ is combined with Gaussian distribution in $\mathbb{R}^n$.
  - In Darling 2016[2], the Bingham distribution on $\mathbb{S}^3$ is combined with Gaussian distribution in $\mathbb{R}^n$.
  - Problems:
    - These models do not have a generic geometric construction.
    - Their MLEs are complicated and need numerical optimizations, so they are not suitable for real time implementations.
- Objective: a New Model on $\mathrm{SO}(3) \times \mathbb{R}^n$
  - Generic construction from directional statistics.
  - Closed form maximum likelihood estimation.

[1] F. L. Markley, "Attitude filtering on $\mathrm{SO}(3)$," *The Journal of the Astronautical Sciences*, vol. 54, no. 3-4, pp. 391–413, 2006.

[2] J. E. Darling and K. J. DeMars, "Uncertainty propagation of correlated quaternion and Euclidean states using the Gauss-Bingham density," *Journal of Advances in Information Fusion*, vol. 11, no. 2, pp. 1–20, 2016

Matrix Fisher–Gaussian Distribution

- Matrix Fisher–Gaussian Distribution (MFG)
  - Density: $\mathrm{SO}(3) \times \mathbb{R}^n \ni (R, x) \sim \mathcal{MG}(\mu, \Sigma, P, U, S, V)$ if

$$f(R, x) = \frac{1}{c(S)\sqrt{(2\pi)^n \det(\Sigma_c)}} \exp\left\{-\frac{1}{2}(x-\mu_c)^T \Sigma_c^{-1} (x-\mu_c)\right\} \exp\left\{\mathrm{tr}\left(FR^T\right)\right\}$$

    - $F = USV^T$;
    - $\Sigma_c = \Sigma - P(\mathrm{tr}\,(S)\,I - S)P^T$;
    - Two definitions: $\mu_c = \mu + P(QS - SQ^T)^\vee$ (MFGI), or $\mu_c = \mu + P(SQ - Q^T S)^\vee$ (MFGB), where $Q = U^T RV$.
  - The correlation between $\mathrm{SO}(3)$ and $\mathbb{R}^n$ is quantified by P.

- Bingham-Gaussian Distribution
  - Density: $\mathbb{S}^3 \times \mathbb{R}^n \ni (q, x) \sim \mathcal{MG}(\mu, \Sigma, P, M, Z)$ if

$$f(q, x) = \frac{1}{c(Z)\sqrt{(2\pi)^n \det(\Sigma_c)}} \exp\left\{-\frac{1}{2}(x-\mu_c)^T \Sigma_c^{-1} (x-\mu_c)\right\} \exp\left\{q^T MZM^T q\right\}$$

Construction

- Conditioning from a $(9+n)$-variate Gaussian distribution (MFGI).

$$\begin{bmatrix} x \\ \mathrm{vec}(R^T) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_R \\ \mu \end{bmatrix}, \begin{bmatrix} \Sigma_R & P_R^T \\ P_R & \Sigma \end{bmatrix} \right)$$

  - $\mu_R = \mathrm{vec}(M^T) = \mathrm{VU}^{\mathrm{T}} \in \mathbb{R}^9$.
  - $\Sigma_R^{-1} = I_{3\times 3} \otimes K \in \mathbb{R}^{9\times 9}$, with $K = VSV^T$.
  - MFGI uses row vectorization $\mathrm{vec}(R^T)$, MFGB uses column vectorization $\mathrm{vec}(R)$.

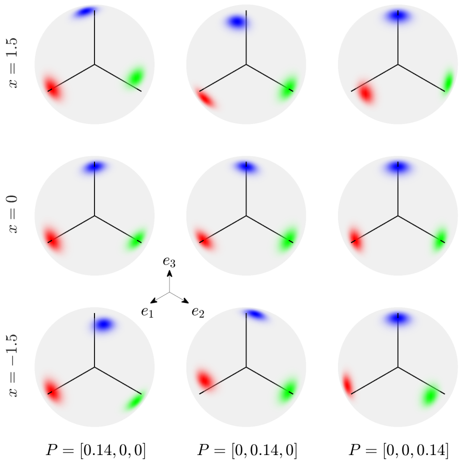
- Correlation: only in the tangent space of the mean attitude $M = UV^T$.
  - Basis of the tangent space at M: $t_i = \mathrm{vec}[(M\widehat{Ve_i})^T]$ for $i \in \{1,2,3\}$.
  - Construct correlation: $P_R = P \begin{bmatrix} t_1 & t_2 & t_3 \end{bmatrix}^T \in \mathbb{R}^{n\times 9}$.
  - Avoids over-parameterization.

Theorem

The conditional distribution $(R, x) \mid RR^T = I_{3\times 3},\ \det(R) = 1$ is a matrix Fisher–Gaussian distribution $\mathcal{MG}(\mu, \Sigma, P, U, S, V)$.

Geometric Interpretation of the Correlation

Conditional distribution of attitude in MFG

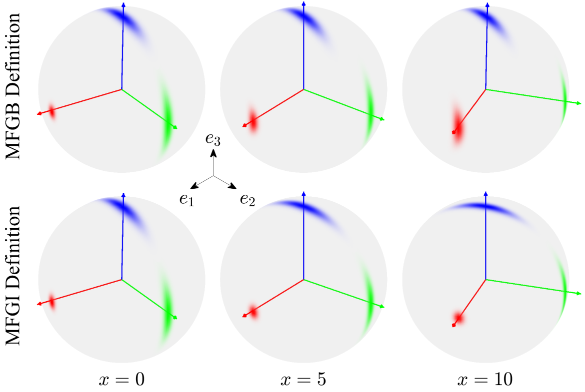


Difference Between MFGI and MFGB

- Difference between MFGI and MFGB
  - MFGI: x is correlated with rotations expressed in the inertial frame.
    - $R|x \overset{\cdot}{\sim} \mathcal{MG}\left(\exp(\widehat{Uv(x)})USV^T\right)$.
  - MFGB: x is correlated with rotations expressed in the body-fixed frame.
    - $R|x \overset{\cdot}{\sim} \mathcal{MG}\left(USV^T\exp(\widehat{Vv(x)})\right)$.

Difference Between MFGI and MFGB

- Difference between MFGI and MFGB



Properties

- Moments
  - Attitude: $\mathrm{E}[R] = UDV^T$, $\mathrm{E}[\nu_R] = 0$,
    $\mathrm{E}\left[\nu_R \nu_R^T\right]_{ii} = s_j \mathrm{E}[Q_{jj}] + s_k \mathrm{E}[Q_{kk}]$ for $i \neq j \neq k$.
  - Linear: $\mathrm{E}[x] = \mu$,
    $\mathrm{E}\left[xx^T\right] = \mu\mu^T + \Sigma - P(\mathrm{tr}(S) I - S)P^T + P\mathrm{E}\left[\nu_R \nu_R^T\right] P^T$.
  - Correlation: $\mathrm{E}\left[x\nu_R^T\right] = P\mathrm{E}\left[\nu_R \nu_R^T\right]$.
- Maximum Likelihood Estimation for Parameters
  - Marginal Likelihood for R: U, S, V.
    - $UDV^T = \mathrm{E}[R]$ is the proper SVD.
    - Solve S from $d_i = \frac{1}{c(S)} \frac{\partial c(S)}{\partial s_i}$ for $i = 1, 2, 3$.
  - Conditional Likelihood for $x|R$: μ, Σ, P
    - $P = \mathrm{cov}(x, \nu_R)\mathrm{cov}(\nu_R, \nu_R)^{-1}$.
    - $\mu = \mathrm{E}[x] - P\mathrm{E}[\nu_R]$.
    - $\Sigma = \mathrm{cov}(x, x) - P\mathrm{cov}(x, \nu_R)^T + P(\mathrm{tr}(S) I_{3\times 3} - S)P^T$.
  - The marginal-conditional MLE is an approximation to the full MLE.
  - This closed form solution makes the calculation efficient, thus suitable for real time implementations.

Attitude Estimation

- Matrix Fisher Gaussian Distribution
  - Defined on $\mathrm{SO}(3) \times \mathbb{R}^n$.
  - Extends the matrix Fisher distribution to deal with the correlation between $\mathrm{SO}(3)$ and $\mathbb{R}^n$.
  - Applications: simultaneous estimation of attitude and other Euclidean quantities.
    - Sensor biases.
    - Velocity, position.
    - Landmark locations.
- Application of MFG in Attitude Estimation
  - Attitude Observability with Single Direction Measurements
  - Attitude Estimation Based on MFG

- Attitude kinematics

$$\frac{\mathrm{d}R(t)}{\mathrm{d}t} = R(t)\hat{\Omega}(t) \tag{1a}$$
$$\frac{\mathrm{d}R(t)}{\mathrm{d}t} = \hat{\omega}(t)R(t) \tag{1b}$$

  - $\Omega(t)$: angular velocity measured in body-fixed frame.
  - $\omega(t)$: angular velocity measured in inertial frame.

- Single direction measurements

$$x(t) = R(t)^T a \tag{2a}$$
$$y(t) = R(t)b \tag{2b}$$

  - a: reference vector in inertial frame, e.g., the direction of magnetic field, or towards a star. $x(t)$: measurement in body-fixed frame.
  - b: reference vector in body-fixed frame, e.g., the direction towards solar panel from the center. $y(t)$: measurement in inertial frame.

Observable Conditions

$$\frac{\mathrm{d}R(t)}{\mathrm{d}t} = R(t)\hat{\Omega}(t) \tag{1a}$$

$$\frac{\mathrm{d}R(t)}{\mathrm{d}t} = \hat{\omega}(t)R(t) \tag{1b}$$

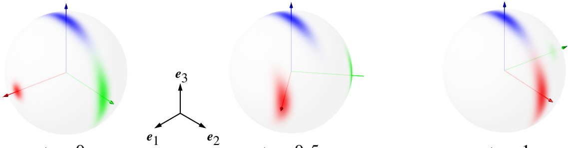
$$x(t) = R(t)^T a \tag{2a}$$

$$y(t) = R(t)b \tag{2b}$$

ang. vel. / ref. vec.	body-fixed frame (1a)	inertial frame (1b)
body-fixed frame (2b)	observable	unobservable
inertial frame (2a)	unobservable	observable

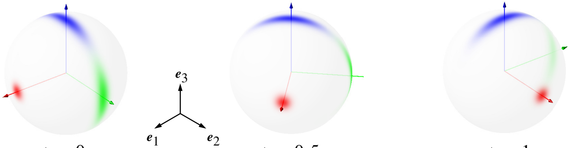
Uncertainty Propagation

- Angular velocity in body-fixed frame: $\frac{\mathrm{d}R(t)}{\mathrm{d}t} = R(t)\hat{\Omega}(t)$
  - The uncertainty is rotated in the body-fixed frame, but is unchanged in the inertial frame.



$t = 0$s $\quad t = 0.5$s $\quad t = 1$s

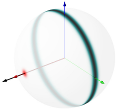
- Angular velocity in inertial frame: $\frac{\mathrm{d}R(t)}{\mathrm{d}t} = \hat{\omega}(t)R(t)$
  - The uncertainty is unchanged in the body-fixed frame, but is rotated in the inertial frame.



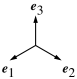
$t = 0$s $\quad t = 0.5$s $\quad t = 1$s

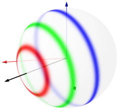
Correction From Direction Measurements

- Reference direction in inertial frame: $x(t) = R(t)^T a$
  - Non-informative along direction a; a is fixed in inertial frame.

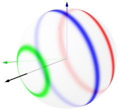


$x = e_1$



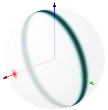


$x = \frac{\sqrt{3}}{2} e_1 + \frac{1}{2} e_2$

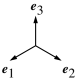


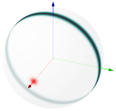
$x = -\frac{1}{2} e_1 + \frac{\sqrt{3}}{2} e_2$

- Reference direction in body-fixed frame: $y(t) = R(t) b$
  - Non-informative along direction b; b is fixed in body-fixed frame.

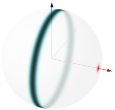


$y = e_1$





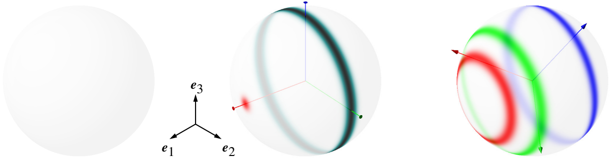
$y = \frac{\sqrt{3}}{2} e_1 + \frac{1}{2} e_2$



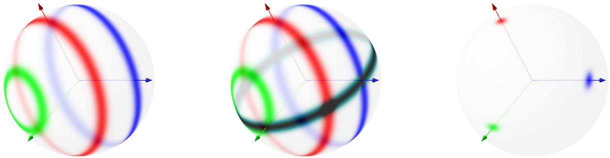
$y = -\frac{1}{2} e_1 + \frac{\sqrt{3}}{2} e_2$

Observability

- Observable:
  - Angular velocity in body-fixed frame: $\frac{\mathrm{d}R(t)}{\mathrm{d}t} = R(t)\hat{\Omega}(t)$
  - Reference direction in body-fixed frame: $y(t) = R(t)b$



Prior dist. at $t = 0$ | Posterior dist. at $t = 0$ | Propagated dist. at $t = 0.5$

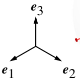


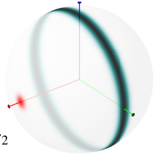
Propagated dist. at $t = 1$ | Propagated dist. overlapped with measured dist. at $t = 1$ | Posterior dist. at $t = 1$

Unobservability

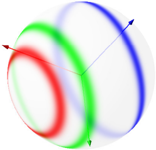
- Unobservable:
  - Angular velocity in body-fixed frame: $\frac{\mathrm{d}R(t)}{\mathrm{d}t} = R(t)\hat{\Omega}(t)$
  - Reference direction in inertial frame: $x(t) = R(t)^T a$

Prior dist. at $t = 0$

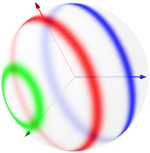




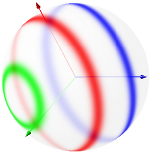
Posterior dist. at $t = 0$



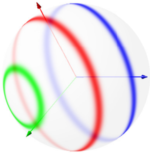
Propagated dist. at $t = 0.5$



Propagated dist. at $t = 1$



Propagated dist. overlapped with measured dist. at $t = 1$

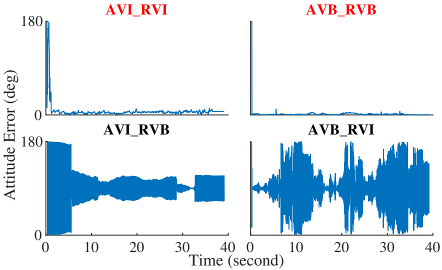


Posterior dist. at $t = 1$

Experimental Verification

ang. vel. / ref. vec.	body-fixed frame (1a)	inertial frame (1b)
body-fixed frame (2a)	AVB_RVB	AVI_RVB
inertial frame (2b)	AVB_RVI	AVI_RVI

- Attitude error [3]



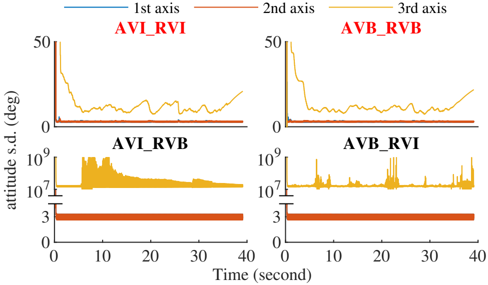
- AVB_RVB and AVI_RVI : attitude error converges to around zero.
- AVB_RVI and AVI_RVB: attitude error does not converge.

[3] I would like to thank Kanishke Gamagedara for conducting the experiment.

Experimental Verification

ang. vel. / ref. vec.	body-fixed frame (1a)	inertial frame (1b)
body-fixed frame (2a)	AVB_RVB	AVI_RVB
inertial frame (2b)	AVB_RVI	AVI_RVI

- Attitude uncertainty



- AVB_RVB and AVI_RVI : standard deviations for all axes are small.
- AVB_RVI and AVI_RVB: standard deviation for one axis is large.

- Gyroscope Model

$$R_{k+1} = R_k \exp\left\{h(\hat{\Omega}_k + \hat{x}_k) + (H_u \Delta W_u)^\wedge\right\},$$
$$x_{k+1} = x_k + H_v \Delta W_v.$$

  - Angle random walk noise: $H_u \Delta W_u$.
  - Bias random walk noise: $H_v \Delta W_v$.

- Measurement Model
  - Attitude measurement: $Z_i = R_t \delta R$, $\delta R \sim \mathcal{M}(F_{Z_i})$, $i = 1, \ldots, N_a$.
  - Vector measurement: $z_j \in \mathbb{S}^2 \sim \mathcal{VM}(R^T B_j a_j, \kappa_j)$, $j = 1, \ldots, N_v$.
    - $\mathcal{VM}$: von Mises–Fisher distribution on $\mathbb{S}^2$.
    - a_j: a reference vector in the inertial frame.

- Bayesian Assumed Density Filter
  - Uncertainty propagation: $(R, x)_k^+ \sim \mathcal{MG}(\mu, \Sigma, P, U, S, V)_k^+$ $\Rightarrow (R, x)_{k+1}^- \sim \mathcal{MG}(\mu, \Sigma, P, U, S, V)_{k+1}^-$.
  - Measurement update: $(R, x)_{k+1}^- \sim \mathcal{MG}(\mu, \Sigma, P, U, S, V)_{k+1}^-$ $\Rightarrow (R, x)_{k+1}^+ \sim \mathcal{MG}(\mu, \Sigma, P, U, S, V)_{k+1}^+$.

Uncertainty Propagation

- Analytical Propagation
  - Calculate analytical approximations $O(h^2)$ of moments for MLE.
  - Marginal MLE: $\mathrm{E}\left[R_{k+1}^-\right] \to U_{k+1}^-, S_{k+1}^-, V_{k+1}^-$.
  - Conditional MLE: $\mathrm{E}\left[x_{k+1}^-\right], \mathrm{E}\left[\nu_{R_{k+1}}^-\right], \mathrm{E}\left[xx^T\right]_{k+1}^-, \mathrm{E}\left[x\nu_R^T\right]_{k+1}^-,$ $\mathrm{E}\left[\nu_R\nu_R^T\right]_{k+1}^- \Rightarrow \mu_{k+1}^-, \Sigma_{k+1}^-, P_{k+1}^-$.
- Unscented Propagation
  - Select sigma points $(R_i, x_i, w_i)_{i=1}^{13}$ from $\mathcal{MG}(\mu, \Sigma, P, U, S, V)_k^+$.
  - Select sigma points $(\eta_j, w_j)_{j=1}^{7}$ from noise $H_u \Delta W_u$.
  - Propagate sigma points through the kinematic equations.

$$R_{ij} = R_i \exp(h(\hat{\Omega}_k + \hat{x}_k) + \hat{\eta}_j), \qquad x_{ij} = x_i$$

  - Update weights: $w_{ij} = w_i w_j$.
  - Recover $\mathcal{MG}_{k+1}^-$ from the sigma points $(R_{ij}, x_{ij}, w_{ij})_{i=1,j=1}^{i=13,j=7}$ using MLE.
  - Set $\Sigma_{k+1}^- = \Sigma_{k+1}^- + hH_vH_v^T$.

Measurement Update

- Measurement Model
  - Attitude measurement: $Z_i = R_t \delta R$, $\delta R \sim \mathcal{M}(F_{Z_i})$, $i = 1, \ldots, N_a$.
  - Vector measurement: $z_j \in \mathbb{S}^2 \sim \mathcal{VM}(R_j^T B_j a_j, \kappa_j)$, $j = 1, \ldots, N_v$.
- Bayes's formula

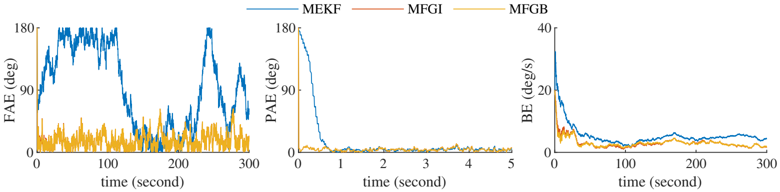
$$p(x, R \mid Z_{1:N_a}, z_{1:N_v}) \propto p(x, R) \cdot p(Z_{1:N_a}, z_{1:N_v} \mid x, R)$$

$$= \operatorname{etr}\left\{\left(F + \sum_{i=1}^{N_a} Z_i F_i^T + \sum_{j=1}^{N_v} \kappa_j B_j a_j z_j^T\right) R^T\right\} \cdot \exp\left\{-\tfrac{1}{2}(x - \mu_c)^T \Sigma_c^{-1} (x - \mu_c)\right\}$$

- Match to a New MFG
  - Marginal MLE:
    - $U_{k+1}^+ S_{k+1}^+ (V_{k+1}^+)^T = F + \sum_{i=1}^{N_a} Z_i F_i^T + \sum_{j=1}^{N_v} \kappa_j B_j a_j z_j^T$.
  - Conditional MLE:
    - Calculate $\mathrm{E}\left[x_{k+1}^+\right]$, $\mathrm{E}\left[xx^T\right]_{k+1}^+$, $\mathrm{E}\left[x\nu_R^T\right]_{k+1}^+$, $\mathrm{E}\left[\nu_R \nu_R^T\right]_{k+1}^+$.
    - $\Rightarrow \mu_{k+1}^+, \Sigma_{k+1}^+, P_{k+1}^+$.

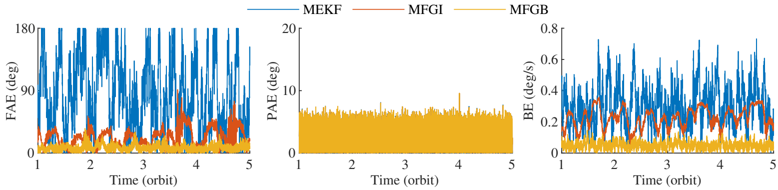
Simulations with Two Direction Measurements

- Settings
  - False initial conditions.
  - Two direction measurements: accurate and inaccurate.
    - Uncertainty around the first reference direction is very large.
    - Uncertainties in the two other axes are small.
- Estimation Errors
  - Compared with MEKF: the state of art method.
  - MFG has better accuracy around the first reference direction.
  - MFG has faster convergence with false initial values.



Simulations with Single Direction Measurements

- Settings
  - A spacecraft orbiting around the Earth.
  - False initial conditions.
  - Time-varying magnetic field direction measurements.
    - Uncertainty around the magnetic field is very large.
    - Uncertainties in the two other axes are small.
- Estimation Errors
  - MFG has better accuracy around the magnetic field.
  - MFG has better accuracy of gyroscope bias.
  - MFGB is more accurate than MFGI.



- Application of MFG in Attitude Estimation
  - Loosely Coupled IMU-GNSS Integration
  - Visual-Inertial Navigation

- IMU Kinematics

$$
\begin{aligned}
R_{k+1} &= R_k \exp\left\{ h(\hat{\Omega}_k + \hat{b}_{g,k}) + H_{gu}\Delta W_{gu} \right\}, \\
b_{g,k+1} &= b_{g,k} + H_{gv}\Delta W_{gv}, \\
p_{k+1} &= p_k + h v_k, \\
v_{k+1} &= v_k - gh + R_k \left(h(a_k + b_{a,k}) + H_{au}\Delta W_{au} \right), \\
b_{a,k+1} &= b_{a,k} + H_{av}\Delta W_{av}.
\end{aligned}
$$

  - $x = [b_g, p, v, b_a] \in \mathbb{R}^{12}$, where b_g, b_a are gyroscope and accelerometer biases, and v, p are velocity and position.

- Measurements
  - Attitude and direction measurements (optional).
  - GNSS receiver: position measurements.

$$
y_k | p_k = p_k + \mathcal{N}(0, \Sigma_y).
$$

Bayesian Assumed Density Filter

- Uncertainty Propagation
  - Calculate analytical approximations $O(h^2)$ of moments.
  - Use MLE to match the propagated moments to an MFG.
- Measurement Update
  - Posterior probability density

$$p(R,x|\mathcal{Z}) \propto \mathrm{etr}\left\{\left(F + \sum_{i=1}^{N_a} Z_i F_i^T + \sum_{j=1}^{N_v} \kappa_j B_j a_j z_j^T\right) R^T\right\}$$
$$\cdot \exp\left\{-\tfrac{1}{2}(x-\mu_c)^T \Sigma_c^{-1}(x-\mu_c)\right\} \cdot \exp\left\{-\tfrac{1}{2}(Hx-y)^T \Sigma_y^{-1}(Hx-y)\right\}.$$

  - Factorization of posterior probability density

$$p(R,x|\mathcal{Z}) \propto p_R(R) \cdot p_{x|R}(x|R),$$
$$p_R(R) = \mathrm{etr}\left(\tilde{F}R^T\right)\exp\left\{-\tfrac{1}{2}(H\mu_c - y)^T(\Sigma_y + H\Sigma_c H^T)^{-1}(H\mu_c - y)\right\},$$
$$p_{x|R}(x|R) = \exp\left\{-\tfrac{1}{2}\left(x - K_p\mu_c - K_m y\right)^T\left((I - K_m H)\Sigma_c\right)^{-1}\left(x - K_p\mu_c - K_m y\right)\right\}.$$

  - Marginal likelihood for $p_R(R)$, conditional likelihood for $p_{x|R}(x|R)$.

Progressive Unscented Measurement Update

- Marginal Likelihood for $p_R(R)$.

$$p_R(R) = \operatorname{etr}\left(\tilde{F}R^T\right)\exp\left\{-\frac{1}{2}(H\mu_c - y)^T(\Sigma_y + H\Sigma_c H^T)^{-1}(H\mu_c - y)\right\}.$$

  - Unscented update
    - Select sigma points $\{R_i, w_i\}_{i=1}^{7}$ from $\mathcal{M}(\tilde{F})$.
    - Reweigh the weights by $w_i^+ = w_i f_m(R_i)$, where $f_m(R)$ is the second term on the right hand side.
    - Calculate $\mathrm{E}[R]^+$: $w_i^+ = w_i^+ / \sum_{j=1}^{7} w_j^+$, $\mathrm{E}[R]^+ = \sum_{i=1} w_i^+ R_i$.
    - Use the marginal MLE: $U^+ D^+ (V^+)^T = \mathrm{E}[R]^+$, $d_i^+ = \frac{1}{c(S^+)}\frac{\partial c(S^+)}{\partial s_i^+}$.
    - Sample degeneration: w_i^+ can be close to zero.
  - Progressive unscented update [3]

$$f_m(R) = f_m(R)^{\lambda_1} f_m(R)^{\lambda_2}\cdots f_m(R)^{\lambda_l}$$

    - Update the weights progressively: $w_i^k = w_i f_m(R_i)^{\lambda_k}$.
    - Avoid sample degeneration: $\frac{\min_i\{w_i^k\}}{\max_i\{w_i^k\}} < \tau$, where $0 < \tau < 1$.

[3] U. D. Hanebeck, "PGF 42: Progressive Gaussian filtering with a twist," in *International Conference on Information Fusion*, 2013, pp. 1103–1110.

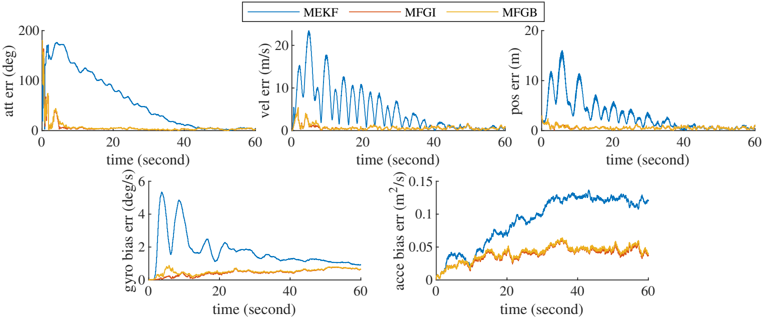
Conditional MLE

$$p(R, x|\mathcal{Z}) \propto \operatorname{etr}\left(F^{+}R^{T}\right) \cdot p_{x|R}(x|R),$$
$$p_{x|R}(x|R) = \exp\left\{-\frac{1}{2}\left(x - K_p\mu_c - K_m y\right)^T \left((I - K_m H)\Sigma_c\right)^{-1}\left(x - K_p\mu_c - K_m y\right)\right\}.$$

- Conditional MLE for $p_{x|R}(x|R)$.
  - Calculate the moments $\mathrm{E}[x]$, $\mathrm{E}\left[xx^T\right]$, $\mathrm{E}\left[x\nu_R^T\right]$, $\mathrm{E}[\nu_R]$, $\mathrm{E}\left[\nu_R\nu_R^T\right]$, with respect to the above density function.
  - $\Rightarrow \mu^{+}, \Sigma^{+}, P^{+}$.

Simulations

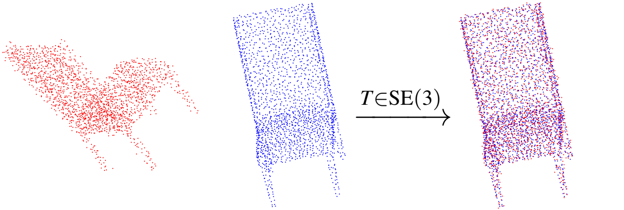
- Initial conditions: highly accurate pitch and roll, reversed yaw.
  - For example from indoors to outdoors.
- Estimation Errors
  - MFG has Faster convergence of yaw.
  - MFG has Better accuracy of position and velocity.



Visual-Inertial Navigation

- 3D Feature Location Measurements
  - Camera types: RGB-D camera, stereo camera.
  - Measures the coordinates of landmark or feature locations in the body-fixed frame.

- Two Point Sets Alignment



$\{p_i \in \mathbb{R}^3\}$ $\{p_i' \in \mathbb{R}^3\}$ Align $\{p_i\}$ and $\{p_i'\}$ using $T \in \mathrm{SE}(3)$

- 3D Feature Location Measurements
  - Camera types: RGB-D camera, stereo camera.
  - Measures the coordinates of landmark or feature locations in the body-fixed frame.
- Two Point Sets Alignment
  - map point set: landmark/feature coordinates $\{p_i\}_{i=1}^N$ in inertial frame.
    - Without noise: $\{p_i\}_{i=1}^N$ are deterministic.
    - With noise: $p_i \sim \mathcal{N}(\bar{p}_i, B_i)$.
  - measurement point set: coordinates in body-fixed frame.

$$p_i' = R^T(p_i - t) + \mathcal{N}(0, A_i).$$

  - The correspondences between two point sets are known.
  - Objective: find the distribution of $(R, t) \in \mathrm{SE}(3)$, and match it to an MFG.

Deterministic Map Points

- Likelihood for (R, t)

$$p(p_1', \ldots, p_N' | R, t) = \exp\left\{ -\frac{1}{2} \sum_{i=1}^{N} \left(R^T(p_i - t) - p_i'\right)^T A_i^{-1} \left(R^T(p_i - t) - p_i'\right) \right\}.$$

- Factorization of Likelihood

$$\begin{aligned} p(p_1', \ldots, p_N' | R, t) &\propto p_R(R) p_{t|R}(t|R), \\ p_R(R) &= \exp\left\{ -\tfrac{1}{2} (\mathbf{T}\boldsymbol{p}' - \mathbf{T}\boldsymbol{p}_m)^T \left(\mathbf{TAT}^T\right)^{-1} (\mathbf{T}\boldsymbol{p}' - \mathbf{T}\boldsymbol{p}_m) \right\}, \\ p_{t|R}(t|R) &= \exp\left\{ -\tfrac{1}{2} (t - \mu_{t|R})^T \Sigma_{t|R}^{-1} (t - \mu_{t|R}) \right\}. \end{aligned}$$

  - Motivated by the conventional SVD method.
  - $\mathbf{T}\boldsymbol{p}_m$ is independent of t.
  - If $A_i = \sigma_i^2 I_{3\times 3}$, $\Sigma_{t|R}$ is independent of R, $\mu_{t|R}$ is linear in R.
  - Use marginal MLE to $p_R(R)$.
  - Use conditional MLE to $p_{x|R}(x|R)$.

Marginal MLE

- Marginal MLE for $p_R(R)$

$$p_R(R) = \exp\left\{-\frac{1}{2}(\mathbf{T}\boldsymbol{p}' - \mathbf{T}\boldsymbol{p}_m)^T\left(\mathbf{T}\mathbf{A}\mathbf{T}^T\right)^{-1}(\mathbf{T}\boldsymbol{p}' - \mathbf{T}\boldsymbol{p}_m)\right\}$$

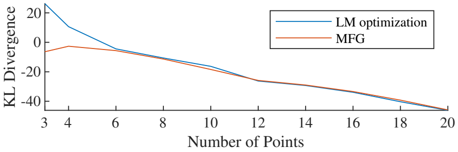
- $\mathbf{T}$ is used to center $\{p_i\}$ and $\{p_i'\}$, $i = 1, \ldots, N-1$.

$$q_i = p_i - \frac{1}{N}\sum_{i=1}^{N} p_i, \qquad q_i' = p_i' - \frac{1}{N}\sum_{i=1}^{N} p_i' = R^T q_i + \text{noise},$$

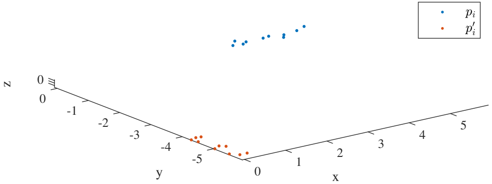
- Wahba's problem:
  - $\{q_i\}_{i=1}^{N-1}$ are reference vectors in the inertial frame.
  - $\{q_i'\}_{i=1}^{N-1}$ are measurements in the body-fixed frame.
  - However, $\{q_i'\}$ are correlated.
- Importance sampling
  - Find a matrix Fisher distribution $\mathcal{M}(\tilde{F})$ through Wahba's problem.
  - Select sigma points $\{R_i, w_i\}_{i=1}^{7}$ from $\mathcal{M}(\tilde{F})$.
  - Reweigh the weights as $w_i^+ = w_i p_R(R_i)/p_{\text{prop}}(R_i)$.
  - Use MLE to the reweighed sigma points $\{R_i, w_i^+\}_{i=1}^{7}$.
  - Progressive update can also be applied to improve accuracy.

Simulation: KL-Divergence

- Comparison with LM-Optimization
  - Find the best $(R, t) \in \mathrm{SE}(3)$ that maximizes the likelihood $p(p'_i, \ldots, p'_N | R, t)$.
  - Gradient based Gauss-Newton method for nonlinear least square.
  - The Jacobian $(J_r^T J_r)^{-1} \in \mathbb{R}^{6\times 6}$ represents the covariance matrix.
- KL-Divergence
  - A metric for the difference between two probability distributions.
  - Simulation results with varying number of map points.

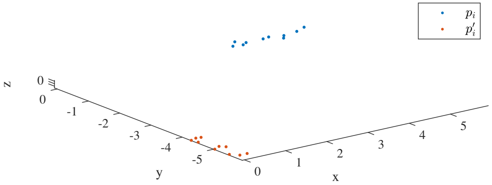


  - The MFG is a much better approximation of the true likelihood when the number of map points is small, i.e., the uncertainty is large.



- Two Point Sets Alignment
  - Ten map points scattered around x-axis.
  - The uncertainty in x-axis in inertial frame is large.
  - The uncertainties in two other axes are small.
  - Concentration of map points around x-axis, $a \in \{1, 0.1, 0.01, 0\}$.

Simulation: Pose Accuracy



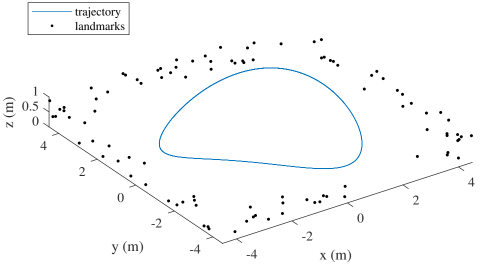
- Estimation Errors

estimator	MFG			
a	1	0.1	0.01	0
att err (deg)	3.93±1.79	24.0±21.7	82.4±51.4	90.1±0.2
pos err	0.28±0.15	0.36±0.20	0.14±0.06	0.11±0.08

estimator	LM optimization			
a	1	0.1	0.01	0
att err (deg)	3.93±1.79	24.0±21.8	80.7±51.2	90.1±0.1
pos err	0.28±0.15	0.39±0.22	0.39±0.22	0.38±0.20

  - MFG has better accuracy in translation when the uncertainty about x-axis is large.

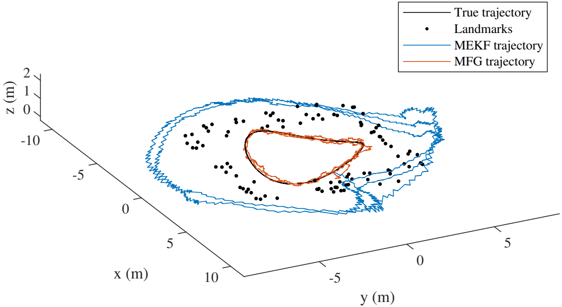
- Bayesian Assumed Density Filter
  - Uncertainty propagation: IMU kinematics.
  - Measurement update: landmark coordinate measurements combined with pose prior.
- Settings
  - A circular trajectory with landmarks scattered around.



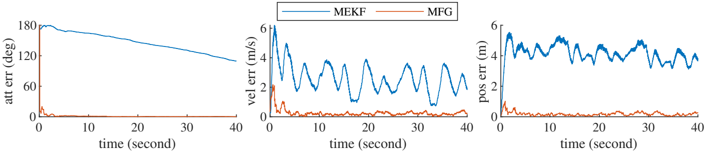
  - Initial conditions: reversed yaw direction.

Simulation: Visual-Inertial Navigation

- Bayesian Assumed Density Filter
  - Uncertainty propagation: IMU kinematics.
  - Measurement update: landmark coordinate measurements combined with pose prior.
- Estimation Errors



- Bayesian Assumed Density Filter
  - Uncertainty propagation: IMU kinematics.
  - Measurement update: landmark coordinate measurements combined with pose prior.
- Estimation Errors
  - MFG has Faster convergence of yaw.
  - MFG has Better accuracy of position and velocity.



Noisy Map Points

- Posterior Density for Pose (R, t) and Map Points $\{p_i\}$

$$p\left(R, t, \{p_i\} | \{p_i'\}\right) \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\left(R^T(p_i - t) - p_i'\right)^T A_i^{-1}\left(R^T(p_i - t) - p_i'\right)\right\} \cdot \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}(p_i - \bar{p}_i)^T B_i^{-1}(p_i - \bar{p}_i)\right\}.$$

- Factorization of Posterior Density

$$p\left(R, t, \{p_i\} | \{p_i'\}\right) \propto p_R(R | \{p_i'\}) \cdot p_{x|R}(x|R),$$
$$p_R(R) = \exp\left\{-\tfrac{1}{2}(\mathbf{T}\boldsymbol{p}' - \mathbf{T}\boldsymbol{p}_m)^T\left(\mathbf{TAT}^T\right)^{-1}(\mathbf{T}\boldsymbol{p}' - \mathbf{T}\boldsymbol{p}_m)\right\},$$
$$p_{x|R}(x|R) = \exp\left\{-\tfrac{1}{2}(x - \mu_{x|R})^T \Sigma_{x|R}^{-1}(x - \mu_{x|R})\right\}.$$

  - $x = [t, p_1, \ldots, p_N]$ is the translation and all map points.
  - Use marginal MLE to $p_R(R)$.
  - Use conditional MLE to $p_{x|R}(x|R)$ with $\{p_i\}$ marginalized.
  - Calculate $\mathrm{E}[p_i]$, $\mathrm{E}\left[p_i p_i^T\right]$, and match them to Gaussian distributions.

Simulation: Pose and Map Accuracy

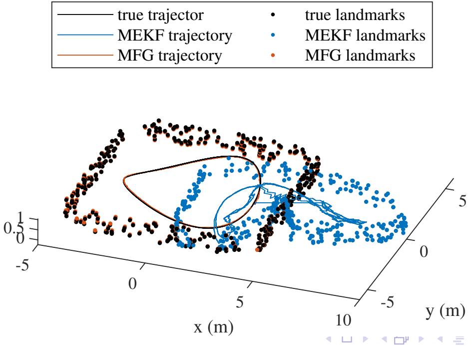
- Two Point Sets Alignment
  - Ten map points:
    - N_1 of them have small uncertainty.
    - The rest of $10 - N_1$ have large uncertainty.
- Estimation Accuracy

estimator	MFG		
N_1	4	2	0
att err (deg)	16.5±7.9	50.6±30.9	77.1±41.1
pos err	0.305±0.147	0.755±0.372	1.05±0.38
map err (N_1)	0.138±0.030	0.149±0.046	-
map err ($10 - N_1$)	0.225±0.056	0.501±0.189	0.661±0.173

estimator	LM optimization		
N_1	4	2	0
att err (deg)	16.0±7.5	48.6±29.5	76.4±40.8
pos err	0.294±0.149	0.772±0.455	1.17±0.55
map err (N_1)	0.138±0.030	0.153±0.048	-
map err ($10 - N_1$)	0.230±0.056	0.518±0.216	0.708±0.196

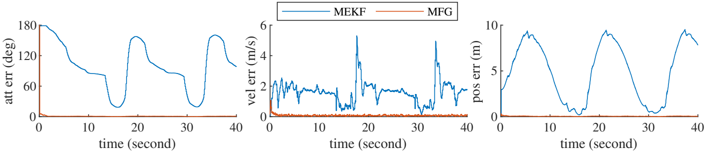
Simulation: Visual-Inertial Navigation

- Bayesian Assumed Density Filter
  - Uncertainty propagation: IMU kinematics, static map points.
  - Measurement update: landmark coordinate measurements combined with pose and map priors.
- Initial conditions: reversed yaw direction.
- Estimation Errors



Simulation: Visual-Inertial Navigation

- Bayesian Assumed Density Filter
  - Uncertainty propagation: IMU kinematics, static map points.
  - Measurement update: landmark coordinate measurements combined with pose and map priors.
- Initial conditions: reversed yaw direction.
- Estimation Errors
  - MFG converges very quickly from wrong initial attitude
  - MEKF does not converge.



Visual-Inertial Odometry

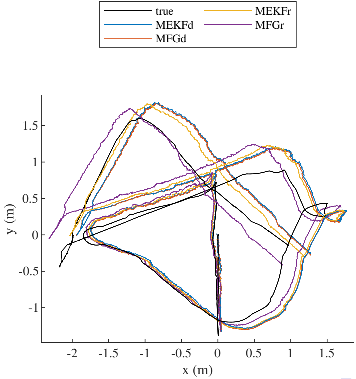
- No Prior Information for the Map
  - The features captured by cameras are usually geometric corners.
  - The map must be built incrementally from previously captured features.
- Build Map From Feature Coordinate Measurements

$$p_{p_i}(p_i) = \int_{R \in \mathrm{SO}(3)} \int_{t \in \mathbb{R}^3} p_{R,t}(R,t) p_{p_i'}\left(R^T(p_i - t)\right) \mathrm{d}t \mathrm{d}R.$$

  - $p_{R,t}(R,t)$ is the probability density for pose.
  - $p_{p_i'}(p_i')$ is the probability density for measurements.
  - Calculate $\mathrm{E}[p_i]$, $\mathrm{E}\left[p_i p_i^T\right]$, and match them to Gaussian distributions.
  - When R has large uncertainty, $p_{p_i}(p_i)$ is far from a Gaussian distribution, so this method is not suitable for large uncertainty.

VCU-RVI Dataset

- VCU-RVI Dataset: Visual-Inertial Odometry
  - Inertial and RGB-D camera measurements.
  - Ground truth provided by a motion capture system.
- Preliminary Results (Proof of Concept)



Summary of Contributions

- Objective: Deal With Large Attitude Uncertainty
  - Unknown initial conditions.
  - Not properly observed degree of freedom for attitude.
  - Sensor failure.
- Contributions
  - Use probability distribution intrinsically defined on $\mathrm{SO}(3)$.
  - Matrix Fisher distribution
    - Higher order moments and highly concentrated approximations.
  - Matrix Fisher–Gaussian distribution (MFG)
    - Correlation between attitude and Euclidean quantities.
  - MFG in classical estimation problems
    - Attitude estimation with gyroscope and direction sensors.
    - Loosely coupled IMU-GNSS navigation.
    - Visual-inertial navigation/odometry.
    - Improved accuracy and convergence speed when the attitude uncertainty is large.