Gradient Flows in Uncertainty Propagation and Filtering of Linear Gaussian Systems

Abhishek Halder

Department of Mechanical and Aerospace Engineering
University of California, Irvine
Irvine, CA 92697-3975

Joint work with Tryphon T. Georgiou
Motivation: Uncertainty Propagation

Initial conditions → Process model → Process noise → State density

Parameters

\[ \rho(x(t), t) \]

Dynamics

\[ \mathbf{r}(x(t), t) \]

Need to compute evolution of density \( \rho(x(t), t) \)

Trajectory flow:

\[ \frac{d}{dt} \mathbf{x}(t) = f(\mathbf{x}, t) + g(\mathbf{x}, t) \mathbf{w}(t), \mathbf{w}(t) \sim \mathcal{N}(0, Q dt) \]

Density flow:

\[ \frac{\partial \rho}{\partial t} = \mathbf{L}_{FP} \rho = -\nabla \cdot (\rho f) + \frac{1}{2} \sum_{i, j=1}^{n} (g Q g^\top)_{ij} \rho, \rho(x(0), 0) = \rho_0(x) \]
Motivation: Uncertainty Propagation

Trajectory flow:
\[ d\mathbf{x}(t) = f(\mathbf{x}, t) \, dt + g(\mathbf{x}, t) \, dw(t), \quad dw(t) \sim \mathcal{N}(0, Q \, dt) \]
**Motivation: Uncertainty Propagation**

Trajectory flow:

\[ \mathbf{d} \mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t) \, dt + \mathbf{g}(\mathbf{x}, t) \, d\mathbf{w}(t), \quad d\mathbf{w}(t) \sim \mathcal{N}(0, \mathbf{Q} \, dt) \]

Density flow:

\[ \frac{\partial \rho}{\partial t} = \mathcal{L}_{FP}(\rho) := -\nabla \cdot (\rho \mathbf{f}) + \frac{1}{2} \sum_{i,j=1}^{n} \left( (\mathbf{g} \mathbf{Q} \mathbf{g}^\top)_{ij} \rho \right), \quad \rho(\mathbf{x}(0), 0) = \rho_0(\mathbf{x}) \]
Motivation: Filtering

Initial conditions → Process model → Prior density

Parameters → Process model → Prior density

Process noise → Prior density

Measurement model → Posterior density

Sensor noise → Measurement model → Posterior density

\[ \rho^+ (x(t), t) \]

\[ \rho^- (x(t), t) \]
**Motivation: Filtering**

- **Initial conditions**
- **Parameters**
- **Process model**

**Process noise**

**Prior density**

**Measurement model**

**Sensor noise**

**Posterior density**

**Trajectory flow:**

\[
\begin{align*}
d\mathbf{x}(t) &= f(\mathbf{x}, t) \, dt + g(\mathbf{x}, t) \, d\mathbf{w}(t), \quad d\mathbf{w}(t) \sim \mathcal{N}(0, Q dt) \\
d\mathbf{z}(t) &= h(\mathbf{x}, t) \, dt + d\mathbf{v}(t), \quad d\mathbf{v}(t) \sim \mathcal{N}(0, R dt)
\end{align*}
\]
**Motivation: Filtering**

**Trajectory flow:**

\[
\begin{align*}
\mathrm{d}x(t) &= f(x, t) \, \mathrm{d}t + g(x, t) \, \mathrm{d}w(t), \\
\mathrm{d}z(t) &= h(x, t) \, \mathrm{d}t + \mathrm{d}v(t),
\end{align*}
\]

**Density flow:**

\[
\mathrm{d}\rho^+ = \left[ \mathcal{L}_{FP} \, \mathrm{d}t + (h(x, t) - \mathbb{E}_{\rho^+} \{h(x, t)\})^\top \, R^{-1} \, (dz(t) - \mathbb{E}_{\rho^+} \{h(x, t)\} \, \mathrm{d}t) \right] \rho^+
\]
Research Scope

Density Flow

PDE formulation ↔ Variational formulation

Numerically approximate solution of PDE

Recursively evaluate proximal operators
Research Scope

Density Flow

- PDE formulation
- Variational formulation

Numerically approximate solution of PDE
Recursively evaluate proximal operators

Density flow $\sim$ gradient descent in infinite dimensions
Gradient Descent

**Finite dimensions**

\[
\frac{dx}{dt} = -\nabla \phi(x), \quad x \in \mathbb{R}^n
\]

\[x_k(h) = x_{k-1} - h \nabla \phi(x_{k-1})\]

\[= \arg\min_x \left\{ \frac{1}{2} \| x - x_{k-1} \|^2 + h\phi(x) \right\}\]

\[= \text{proximal}_{h\phi} (x_{k-1})\]

\[x_k(h) \to x(t=kh), \text{ as } h \downarrow 0\]

**Infinite dimensions**

\[
\frac{\partial \rho}{\partial t} = \mathcal{L}(x, \rho), \quad x \in \mathbb{R}^n, \rho \in \mathcal{D}
\]

\[\rho_k(x, h)\]

\[= \arg\min_{\rho} \left\{ \frac{1}{2} d(\rho, \rho_{k-1})^2 + h\Phi(\rho) \right\}\]

\[= \text{proximal}_{h\Phi}^d (\rho_{k-1})\]

\[\rho_k(x, h) \to \rho(x, t=kh), \text{ as } h \downarrow 0\]
# Two Important Results from Literature


Trajectory dynamics is gradient flow:

\[
\frac{dx(t)}{dt} = -\nabla U(x) \, dt + \sqrt{2\beta^{-1}} \, dw(t), \quad x \in \mathbb{R}^n, \, U(x) \geq 0, \, \beta > 0
\]

Fokker-Planck PDE for density flow:

\[
\frac{\partial \rho}{\partial t} = \nabla \cdot (\nabla U(x) \rho) + \beta^{-1} \Delta \rho, \quad \rho(x, 0) = \rho_0(x), \, \rho_\infty(x) \propto e^{-\beta U(x)}
\]
Two Important Results from Literature


Trajectory dynamics is gradient flow:

\[
dx(t) = -\nabla U(x) \, dt + \sqrt{2\beta^{-1}} \, dw(t), \quad x \in \mathbb{R}^n, U(x) \geq 0, \beta > 0
\]

Fokker-Planck PDE for density flow:

\[
\frac{\partial \rho}{\partial t} = \nabla \cdot (\nabla U(x) \rho) + \beta^{-1} \Delta \rho, \quad \rho(x, 0) = \rho_0(x), \rho_\infty(x) \propto e^{-\beta U(x)}
\]

Gradient descent in \(D_2 \coloneqq \{\rho \in D : \int x^\top x \rho(x) \, dx < \infty\}:

\[
\rho_k(x, h) = \arg\inf_{\rho \in D_2} \left\{ \frac{1}{2} W_2^2(\rho, \rho_{k-1}) + h \mathcal{F}(\rho) \right\}, \quad k = 1, 2, \ldots
\]

where \(\mathcal{F}(\rho) \coloneqq \mathcal{E}(\rho) + \beta^{-1} S(\rho)\)

\[
= \int U(x) \rho(x) \, dx + \beta^{-1} \int \rho(x) \log \rho(x) \, dx
\]
Two Important Results from Literature


Trajectory dynamics is gradient flow:
\[ \text{d}x(t) = -\nabla U(x) \, \text{d}t + \sqrt{2\beta^{-1}} \text{d}w(t), \quad x \in \mathbb{R}^n, \, U(x) \geq 0, \, \beta > 0 \]

Fokker-Planck PDE for density flow:
\[ \frac{\partial \rho}{\partial t} = \nabla \cdot (\nabla U(x) \rho) + \beta^{-1} \Delta \rho, \quad \rho(x,0) = \rho_0(x), \, \rho_\infty(x) \propto e^{-\beta U(x)} \]

Gradient descent in \( \mathcal{D}_2 := \{ \rho \in \mathcal{D} : \int x^\top x \rho(x) \, dx < \infty \} \):

\[ \rho_k(x,h) = \underset{\rho \in \mathcal{D}_2}{\text{argmin}} \left\{ \frac{1}{2} W_2^2(\rho, \rho_{k-1}) + h \mathcal{F}(\rho) \right\}, \quad k = 1, 2, \ldots \]

where \( \mathcal{F}(\rho) := \mathcal{E}(\rho) + \beta^{-1} \mathcal{S}(\rho) \)
\[ = \int U(x) \rho(x) \, dx + \beta^{-1} \int \rho(x) \log \rho(x) \, dx \]
Two Important Results from Literature

#2. LMMR scheme (SIAM J. Control Optim., 2015)

No process dynamics, only measurement update:

\[ dx(t) = 0, \quad dz(t) = h(x, t) \, dt + dv(t), \quad dv(t) \sim \mathcal{N}(0, R dt) \]

Kushner-Stratonovich SPDE for density flow:

\[ d\rho^+ = \left[ (h(x, t) - \mathbb{E}_{\rho^+} \{ h(x, t) \})^\top R^{-1} (dz(t) - \mathbb{E}_{\rho^+} \{ h(x, t) \} dt) \right] \rho^+ \]
Two Important Results from Literature

#2. LMMR scheme (*SIAM J. Control Optim.*, 2015)

No process dynamics, only measurement update:

\[ dx(t) = 0, \quad dz(t) = h(x, t)\, dt + dv(t), \quad dv(t) \sim \mathcal{N}(0, R dt) \]

Kushner-Stratonovich SPDE for density flow:

\[ d\rho^+ = \left[ (h(x, t) - \mathbb{E}_{\rho^+}\{h(x, t)\})^\top R^{-1}(dz(t) - \mathbb{E}_{\rho^+}\{h(x, t)\}dt) \right] \rho^+ \]

Gradient descent in \( \mathcal{D}_2 := \{ \rho \in \mathcal{D} : \int x^\top x \rho(x) dx < \infty \} : \)

\[ \rho_k^+(x, h) = \operatorname{arg inf}_{\rho \in \mathcal{D}_2} \left\{ D_{KL}(\rho \parallel \rho_k^-) + h \Phi(\rho) \right\}, \quad k = 1, 2, \ldots \]

where \( \Phi(\rho) := \frac{1}{2} \mathbb{E}_\rho \left[ (y_k - h(x, t))^\top R^{-1}(y_k - h(x, t)) \right] , \)

and \( y_k := \frac{1}{h}(z_k - z_{k-1}) \)
The Case for Linear Gaussian Systems

Model:
\[
dx(t) = Ax(t)dt + Bdw(t), \quad dw(t) \sim \mathcal{N}(0, Qdt)
\]
\[
dz(t) = Cx(t)dt + dv(t), \quad dv(t) \sim \mathcal{N}(0, Rdtdt)
\]

Assumptions: A Hurwitz, \((A, B)\) controllable pair

Given \(x(0) \sim \mathcal{N}(\mu_0, P_0)\), want to recover:

For uncertainty propagation:
\[
\dot{\mu} = A\mu, \quad \mu(0) = \mu_0; \quad \dot{P} = AP + PA^\top + BQB^\top, \quad P(0) = P_0.
\]

For filtering:
\[
d\mu^+(t) = A\mu^+(t)dt + \underbrace{K(t)}_{P^+CR^{-1}} (dz(t) - C\mu^+(t)dt),
\]
\[
\dot{P}^+(t) = AP^+(t) + P^+(t)A^\top + BQB^\top - K(t)RK(t)^\top.
\]
Challenge 1:
How to actually perform the infinite dimensional optimization over $\mathcal{D}_2$?

Challenge 2:
If and how one can apply the variational schemes for generic linear system with Hurwitz $A$ and controllable $(A, B)$?
Addressing Challenge 1: How to Compute

Two Step Optimization Strategy

- Notice that the objective is a sum:

\[
\arg\inf_{\rho \in \mathcal{D}_2} \left\{ \frac{1}{2} d(\rho, \rho_{k-1})^2 + h\Phi(\rho) \right\}
\]

- Choose a parametrized subspace of \( \mathcal{D}_2 \) such that the individual minimizers over that subspace match

- Then optimize over parameters

- \( \mathcal{D}_{\mu,P} \subset \mathcal{D}_2 \) works!
Addressing Challenge 2: Generic \((A, \sqrt{2}B)\)

Two Successive Coordinate Transformations

#1. Equipartition of energy:

- Define *thermodynamic temperature* \(\theta := \frac{1}{n} \text{tr}(P_{\infty})\), and *inverse temperature* \(\beta := \theta^{-1}\)
- State vector: \(x \mapsto x_{ep} := \sqrt{\theta}P_{\infty}^{-\frac{1}{2}}x\)
- System matrices:

\[
A_{ep}, \sqrt{2}B \mapsto P_{\infty}^{-\frac{1}{2}}AP_{\infty}^{\frac{1}{2}}, \sqrt{2}\theta \quad B_{ep} \mapsto P_{\infty}^{-\frac{1}{2}}B
\]

- Stationary covariance: \(P_{\infty} \mapsto \theta I\)
Addressing Challenge 2: Generic \((A, \sqrt{2}B)\)

Two Successive Coordinate Transformations

#2. Symmetrization:

- State vector: \(x_{\text{ep}} \mapsto x_{\text{sym}} := e^{-A_{\text{ep}}^{\text{skew}} t} x_{\text{ep}}\)
- System matrices:

\[
A_{\text{ep}}, \sqrt{2\theta}B_{\text{ep}} \mapsto e^{-A_{\text{ep}}^{\text{skew}} t} A_{\text{ep}}^{\text{sym}} e^{A_{\text{ep}}^{\text{skew}} t}, \sqrt{2\theta} e^{-A_{\text{ep}}^{\text{skew}} t} B_{\text{ep}}
\]

- Stationary covariance:
  \(\theta I \mapsto \theta I\)
- Potential: \(U(x_{\text{sym}}) := -\frac{1}{2} x_{\text{sym}}^\top F(t) x_{\text{sym}} \geq 0\)
Summary of Results

- Two successive coordinate transformations bring generic linear system to JKO canonical form

- Can apply two step optimization strategy in $x_{\text{sym}}$ coordinate

- Recovers mean-covariance propagation, and Kalman-Bucy filter in $h \downarrow 0$ limit

- Changing the distance in LMMR from $D_{\text{KL}}$ to $\frac{1}{2} W_2^2$ gives Luenberger-type observers

- **Future work:** computation for nonlinear filtering
Our preprint:

JKO scheme:

LMMR scheme:
Thank You