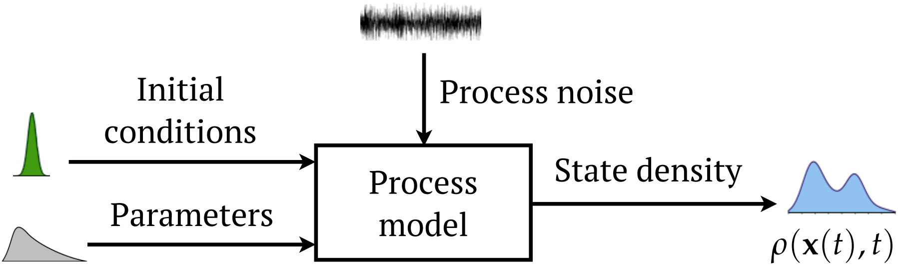
Gradient Flows in Uncertainty Propagation and Filtering of Linear Gaussian Systems

Abhishek Halder

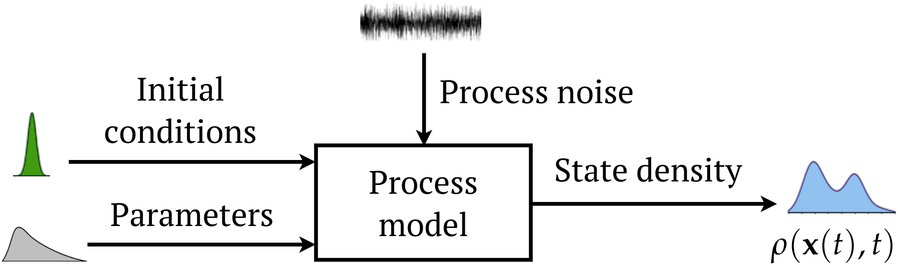
Department of Mechanical and Aerospace Engineering
University of California, Irvine
Irvine, CA 92697-3975

Joint work with Tryphon T. Georgiou

Motivation: Uncertainty Propagation



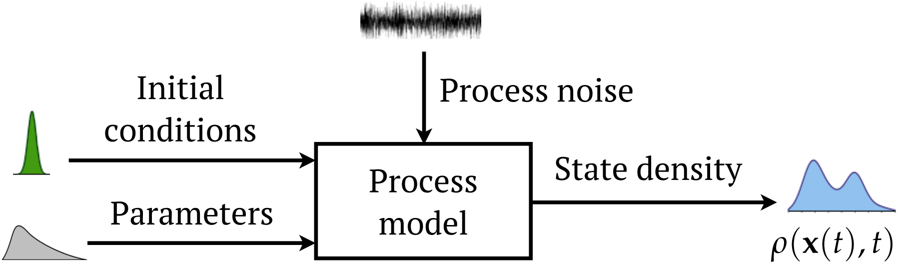
Motivation: Uncertainty Propagation



Trajectory flow:

$$\mathrm{d}\mathbf{X}(t) = \mathbf{f}(\mathbf{X}, t)\,\mathrm{d}t + \mathbf{g}(\mathbf{X}, t)\,\mathrm{d}\mathbf{w}(t), \quad \mathrm{d}\mathbf{w}(t) \sim \mathcal{N}(0, \mathbf{Q}\mathrm{d}t)$$

Motivation: Uncertainty Propagation



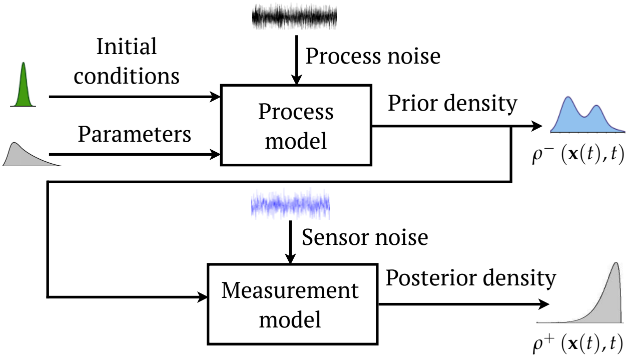
Trajectory flow:

$$\mathrm{d}\mathbf{X}(t) = \mathbf{f}(\mathbf{X}, t)\,\mathrm{d}t + \mathbf{g}(\mathbf{X}, t)\,\mathrm{d}\mathbf{w}(t), \quad \mathrm{d}\mathbf{w}(t) \sim \mathcal{N}(0, \mathbf{Q}\mathrm{d}t)$$

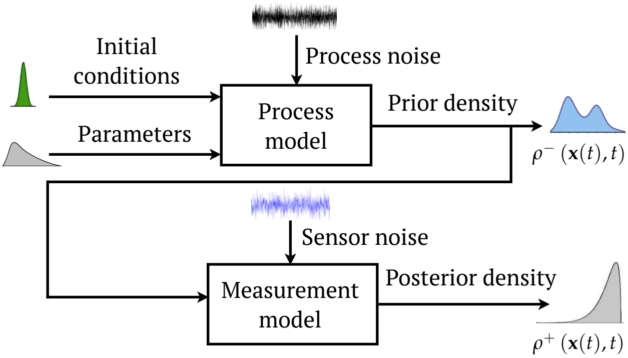
Density flow:

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

Motivation: Filtering



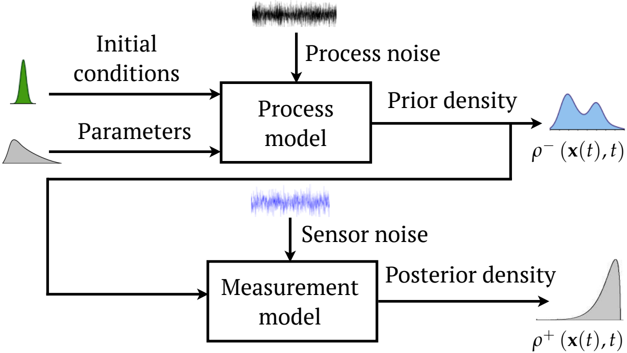
Motivation: Filtering



Trajectory flow:

$$\begin{aligned} \mathrm{d}\mathbf{X}(t) &= \mathbf{f}(\mathbf{X}, t)\,\mathrm{d}t + \mathbf{g}(\mathbf{X}, t)\,\mathrm{d}\mathbf{w}(t), \qquad && \mathrm{d}\mathbf{w}(t) \sim \mathcal{N}(0, \mathbf{Q}\mathrm{d}t) \\ \mathrm{d}\mathbf{Z}(t) &= \mathbf{h}(\mathbf{X}, t)\,\mathrm{d}t + \mathrm{d}\mathbf{v}(t), && \mathrm{d}\mathbf{v}(t) \sim \mathcal{N}(0, \mathbf{R}\mathrm{d}t) \end{aligned}$$

Motivation: Filtering



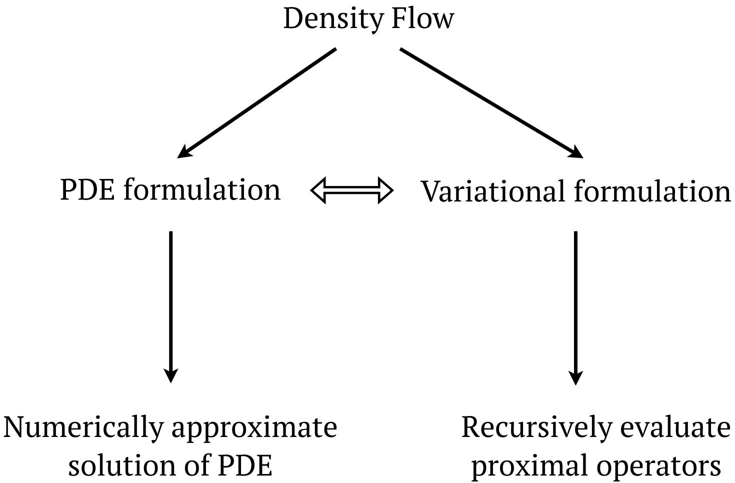
Trajectory flow:

$$
\begin{aligned}
\mathrm{d}\mathbf{X}(t) &= \mathbf{f}(\mathbf{X}, t)\,\mathrm{d}t + \mathbf{g}(\mathbf{X}, t)\,\mathrm{d}\mathbf{w}(t), \qquad \mathrm{d}\mathbf{w}(t) \sim \mathcal{N}(0, \mathbf{Q}\mathrm{d}t)\\
\mathrm{d}\mathbf{Z}(t) &= \mathbf{h}(\mathbf{X}, t)\,\mathrm{d}t + \mathrm{d}\mathbf{v}(t), \qquad \mathrm{d}\mathbf{v}(t) \sim \mathcal{N}(0, \mathbf{R}\mathrm{d}t)
\end{aligned}
$$

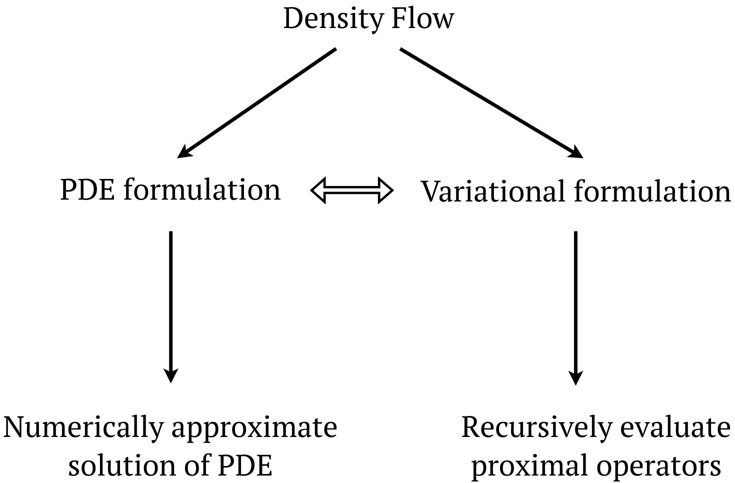
Density flow:

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

Research Scope



Research Scope



Density flow $\rightsquigarrow$ gradient descent in infinite dimensions

Gradient Descent

Finite dimensions

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = -\nabla\phi(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^n$$

$$\mathbf{x}_k(h) = \mathbf{x}_{k-1} - h\nabla\phi(\mathbf{x}_{k-1})$$

$$= \underset{\mathbf{x}}{\operatorname{argmin}}\{\tfrac{1}{2}\|\mathbf{x} - \mathbf{x}_{k-1}\|^2 + h\phi(\mathbf{x})\}$$

$$= \operatorname{proximal}_{h\phi}^{\|\cdot\|}(\mathbf{x}_{k-1})$$

$$\mathbf{x}_k(h) \rightarrow \mathbf{x}(t = kh), \text{ as } h \downarrow 0$$

Infinite dimensions

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

$$\rho_k(\mathbf{x},h)$$

$$= \underset{\rho}{\operatorname{argmin}}\{\tfrac{1}{2}d(\rho,\rho_{k-1})^2 + h\Phi(\rho)\}$$

$$= \operatorname{proximal}_{h\Phi}^{d(\cdot,\cdot)}(\rho_{k-1})$$

$$\rho_k(\mathbf{x},h) \rightharpoonup \rho(\mathbf{x}, t = kh), \text{ as } h \downarrow 0$$

Two Important Results from Literature

#1. JKO scheme (*SIAM J. Math. Analysis*, 1998)

Trajectory dynamics is gradient flow:

$$\mathrm{d}\mathbf{x}(t) = -\nabla U(\mathbf{x})\,\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}\mathbf{w}(t), \quad \mathbf{x} \in \mathbb{R}^n,\ U(\mathbf{x}) \geq 0,\ \beta > 0$$

Fokker-Planck PDE for density flow:

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

Two Important Results from Literature

#1. JKO scheme (*SIAM J. Math. Analysis*, 1998)

Trajectory dynamics is gradient flow:

$$\mathrm{d}\mathbf{x}(t) = -\nabla U(\mathbf{x})\,\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}\mathbf{w}(t), \quad \mathbf{x}\in\mathbb{R}^n,\, U(\mathbf{x})\geq 0,\, \beta>0$$

Fokker-Planck PDE for density flow:

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

Gradient descent in $\mathscr{D}_2 := \{\rho\in\mathscr{D} : \int \mathbf{x}^\top\mathbf{x}\,\rho(\mathbf{x})\mathrm{d}\mathbf{x} < \infty\}$:

$$\rho_k(\mathbf{x},h) = \underset{\rho\in\mathscr{D}_2}{\arg\inf}\left\{\tfrac{1}{2}W_2^2(\rho,\rho_{k-1}) + h\,\mathcal{F}(\rho)\right\}, \quad k=1,2,\ldots$$

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

$$= \int U(\mathbf{x})\rho(\mathbf{x})\mathrm{d}\mathbf{x} + \beta^{-1}\int\rho(\mathbf{x})\log\rho(\mathbf{x})\mathrm{d}\mathbf{x}$$

Two Important Results from Literature

#1. JKO scheme (*SIAM J. Math. Analysis*, 1998)

Trajectory dynamics is gradient flow:

$$\mathrm{d}\mathbf{x}(t) = -\nabla U(\mathbf{x})\,\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}\mathbf{w}(t), \quad \mathbf{x}\in\mathbb{R}^n,\, U(\mathbf{x})\geq 0,\, \beta > 0$$

Fokker-Planck PDE for density flow:

Gibbs density

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

Gradient descent in $\mathscr{D}_2 := \{\rho\in\mathscr{D} : \int \mathbf{x}^\top\mathbf{x}\,\rho(\mathbf{x})\mathrm{d}\mathbf{x} < \infty\}$:

$$\rho_k(\mathbf{x},h) = \underset{\rho\in\mathscr{D}_2}{\operatorname{arginf}}\left\{\tfrac{1}{2}W_2^2(\rho,\rho_{k-1}) + h\,\mathcal{F}(\rho)\right\}, \quad k = 1,2,\ldots$$

thermodynamic temperature

$$\text{where } \mathcal{F}(\rho) := \mathcal{E}(\rho) + \beta^{-1}\,\mathcal{S}(\rho)$$
$$= \int U(\mathbf{x})\rho(\mathbf{x})\mathrm{d}\mathbf{x} + \beta^{-1}\int \rho(\mathbf{x})\log\rho(\mathbf{x})\mathrm{d}\mathbf{x}$$

Two Important Results from Literature

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

No process dynamics, only measurement update:

$$\mathrm{d}\mathbf{x}(t) = 0, \quad \mathrm{d}\mathbf{z}(t) = \mathbf{h}(\mathbf{x}, t)\,\mathrm{d}t + \mathrm{d}\mathbf{v}(t), \quad \mathrm{d}\mathbf{v}(t) \sim \mathcal{N}(0, \mathbf{R}\mathrm{d}t)$$

Kushner-Stratonovich SPDE for density flow:

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

Two Important Results from Literature

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

No process dynamics, only measurement update:

$$\mathrm{d}\mathbf{x}(t) = 0, \quad \mathrm{d}\mathbf{z}(t) = \mathbf{h}(\mathbf{x},t)\,\mathrm{d}t + \mathrm{d}\mathbf{v}(t), \;\; \mathrm{d}\mathbf{v}(t) \sim \mathcal{N}(0, \mathbf{R}\mathrm{d}t)$$

Kushner-Stratonovich SPDE for density flow:

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

Gradient descent in $\mathscr{D}_2 := \{\rho \in \mathscr{D} : \int \mathbf{x}^{\top}\mathbf{x}\,\rho(\mathbf{x})\mathrm{d}\mathbf{x} < \infty\}$:

$$\rho_k^{+}(\mathbf{x},h) = \underset{\rho\in\mathscr{D}_2}{\operatorname{arginf}}\left\{D_{\mathrm{KL}}\left(\rho || \rho_k^{-}\right) + h\,\Phi(\rho)\right\}, \quad k = 1, 2, \ldots$$

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

and $\mathbf{y}_k := \frac{1}{h}(\mathbf{z}_k - \mathbf{z}_{k-1})$

The Case for Linear Gaussian Systems

Model:

$$\mathrm{d}\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t)\mathrm{d}t + \mathbf{B}\mathrm{d}\mathbf{w}(t), \quad \mathrm{d}\mathbf{w}(t) \sim \mathcal{N}(0, \mathbf{Q}\mathrm{d}t)$$

$$\mathrm{d}\mathbf{z}(t) = \mathbf{C}\mathbf{x}(t)\mathrm{d}t + \mathrm{d}\mathbf{v}(t), \quad \mathrm{d}\mathbf{v}(t) \sim \mathcal{N}(0, \mathbf{R}\mathrm{d}t)$$

Assumptions: $\mathbf{A}$ Hurwitz, $(\mathbf{A}, \mathbf{B})$ controllable pair

Given $\mathbf{x}(0) \sim \mathcal{N}(\mu_0, \mathbf{P}_0)$, want to recover:

For uncertainty propagation:

$$\dot{\mu} = \mathbf{A}\mu,\ \mu(0) = \mu_0; \quad \dot{\mathbf{P}} = \mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^\top + \mathbf{B}\mathbf{Q}\mathbf{B}^\top,\ \mathbf{P}(0) = \mathbf{P}_0.$$

For filtering:

$\mathbf{P}^+\mathbf{C}\mathbf{R}^{-1}$

$$\mathrm{d}\mu^+(t) = \mathbf{A}\mu^+(t)\mathrm{d}t + \mathbf{K}(t) \left(\mathrm{d}\mathbf{z}(t) - \mathbf{C}\mu^+(t)\mathrm{d}t\right),$$

$$\dot{\mathbf{P}}^+(t) = \mathbf{A}\mathbf{P}^+(t) + \mathbf{P}^+(t)\mathbf{A}^\top + \mathbf{B}\mathbf{Q}\mathbf{B}^\top - \mathbf{K}(t)\mathbf{R}\mathbf{K}(t)^\top.$$

The Case for Linear Gaussian Systems

Challenge 1:

How to actually perform the infinite dimensional optimization over $\mathscr{D}_2$?

Challenge 2:

If and how one can apply the variational schemes for generic linear system with Hurwitz $\mathbf{A}$ and controllable $(\mathbf{A}, \mathbf{B})$?

Addressing Challenge 1: How to Compute

Two Step Optimization Strategy

- Notice that the objective is a *sum*:

first functional

second functional

$$\operatorname*{arginf}_{\rho \in \mathscr{D}_2} \left\{ \frac{1}{2} d(\rho, \rho_{k-1})^2 + h\Phi(\rho) \right\}$$

- Choose a parametrized subspace of $\mathscr{D}_2$ such that the individual minimizers over that subspace match

- Then optimize over parameters

- $\mathscr{D}_{\mu,\mathbf{P}} \subset \mathscr{D}_2$ works!

Addressing Challenge 2: Generic $(\mathbf{A}, \sqrt{2}\mathbf{B})$

Two Successive Coordinate Transformations

#1. Equipartition of energy:

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

$$\mathbf{A}, \sqrt{2}\mathbf{B} \mapsto \underbrace{\mathbf{P}_\infty^{-\frac{1}{2}}\mathbf{A}\mathbf{P}_\infty^{\frac{1}{2}}}_{\mathbf{A}_{\mathrm{ep}}}, \sqrt{2\theta}\ \underbrace{\mathbf{P}_\infty^{-\frac{1}{2}}\mathbf{B}}_{\mathbf{B}_{\mathrm{ep}}}$$

- Stationary covariance: $\mathbf{P}_\infty \mapsto \theta\mathbf{I}$

Addressing Challenge 2: Generic $(\mathbf{A}, \sqrt{2}\mathbf{B})$

Two Successive Coordinate Transformations

#2. Symmetrization:

- State vector: $\mathbf{x}_{\text{ep}} \mapsto \mathbf{x}_{\text{sym}} := e^{-\mathbf{A}_{\text{ep}}^{\text{skew}} t}\mathbf{x}_{\text{ep}}$
- System matrices:

$$\mathbf{A}_{\text{ep}}, \sqrt{2\theta}\mathbf{B}_{\text{ep}} \mapsto \underbrace{e^{-\mathbf{A}_{\text{ep}}^{\text{skew}} t}\mathbf{A}_{\text{ep}}^{\text{sym}} e^{\mathbf{A}_{\text{ep}}^{\text{skew}} t}}_{\mathbf{F}(t)}, \sqrt{2\theta}\, \underbrace{e^{-\mathbf{A}_{\text{ep}}^{\text{skew}} t}\mathbf{B}_{\text{ep}}}_{\mathbf{G}(t)}$$

- Stationary covariance:
  $\theta\mathbf{I} \mapsto \theta\mathbf{I}$
- Potential: $U(\mathbf{x}_{\text{sym}}) := -\frac{1}{2}\mathbf{x}_{\text{sym}}^\top \mathbf{F}(t)\mathbf{x}_{\text{sym}} \geq 0$

Summary of Results

- Two successive coordinate transformations bring generic linear system to JKO canonical form
- Can apply two step optimization strategy in $\mathbf{x}_{\text{sym}}$ coordinate
- Recovers mean-covariance propagation, and Kalman-Bucy filter in $h \downarrow 0$ limit
- Changing the distance in LMMR from D_{KL} to $\frac{1}{2}W_2^2$ gives Luenberger-type observers
- **Future work:** computation for nonlinear filtering

Details

Our preprint:
A. Halder, and T.T. Georgiou, "Gradient Flows in Uncertainty Propagation and Filtering of Linear Gaussian Systems". arXiv:1704.00102, 2017.

JKO scheme:
R. Jordan, D. Kinderlehrer, and F. Otto, "The Variational Formulation of the Fokker–Planck Equation". *SIAM J. Math. Analysis*, Vol. 29, no. 1, pp. 1–17, 1998.

LMMR scheme:
R.S. Laugesen, P.G. Mehta, S.P. Meyn, and M. Raginsky, "Poisson's Equation in Nonlinear Filtering". *SIAM J. Control Optim.*, Vol. 53, no. 1, pp. 501–525, 2015.

Thank You