Analysis of the Stochastic Inverse Problem

Maria Oprea

SIAM NNP, November 2nd 2024

Joint work with Qin Li (UW Madison), Li Wang (UMN Twin Cities) and Yunan Yang (Cornell)

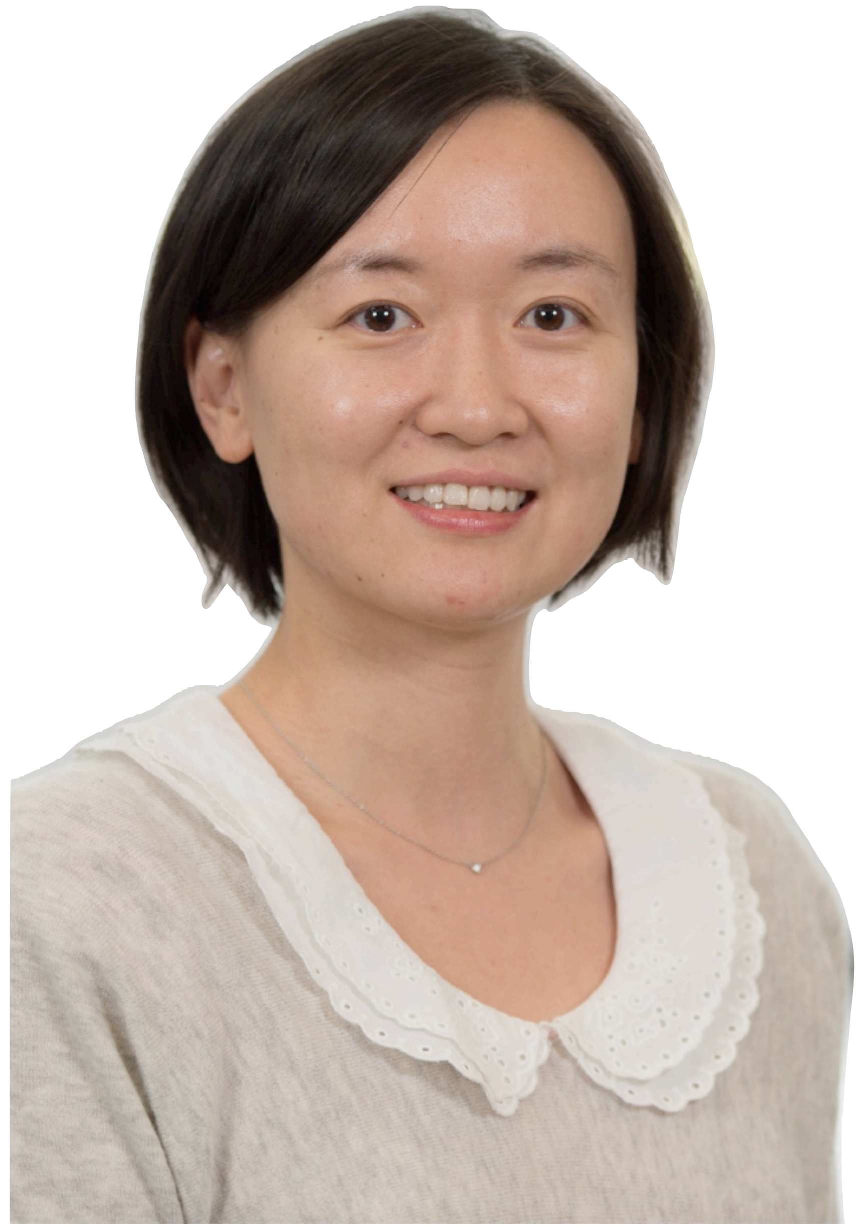
Stochastic Inverse Problem: stability, regularization and Wasserstein gradient flow, Qin Li, O., Li Wang, Yunan Yang, arxiv:2410.00229

Collaborators

Qin Li

UW Madison

Li Wang



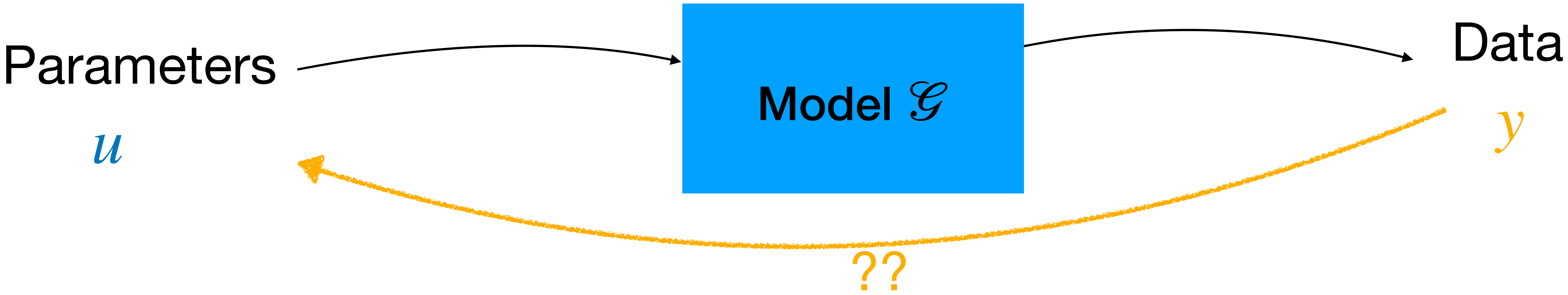
UM Twin Cities

Yunan Yang

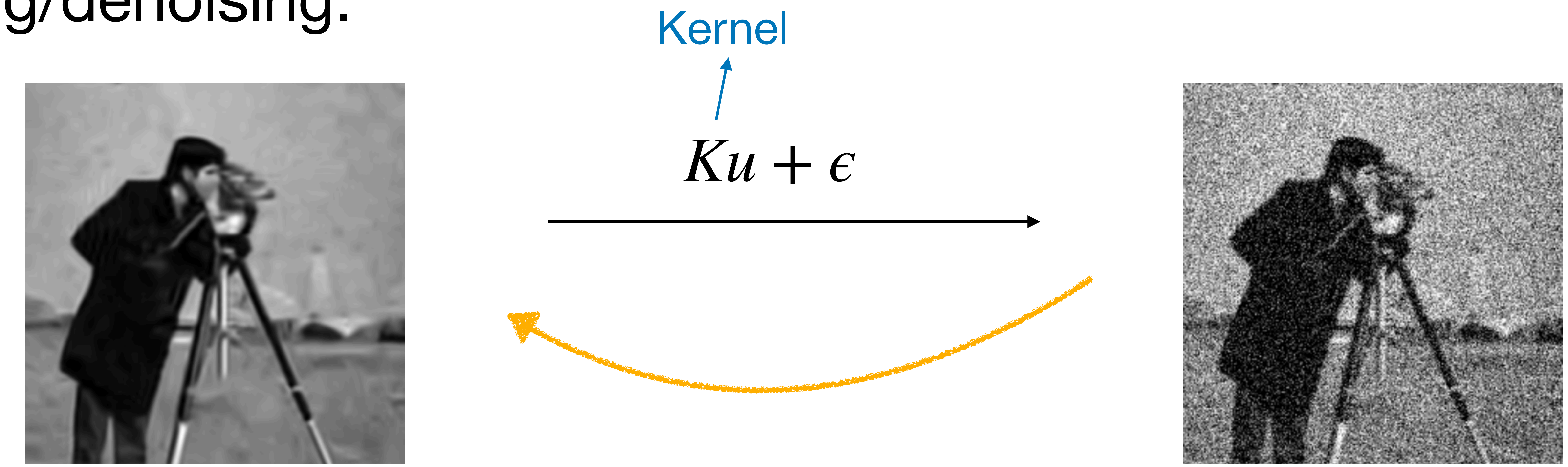
Cornell University

Qin Li, O., Li Wang and Yunan Yang, Stochastic Inverse problem: Stability, Regularization and Wasserstein Gradient Flow, preprint

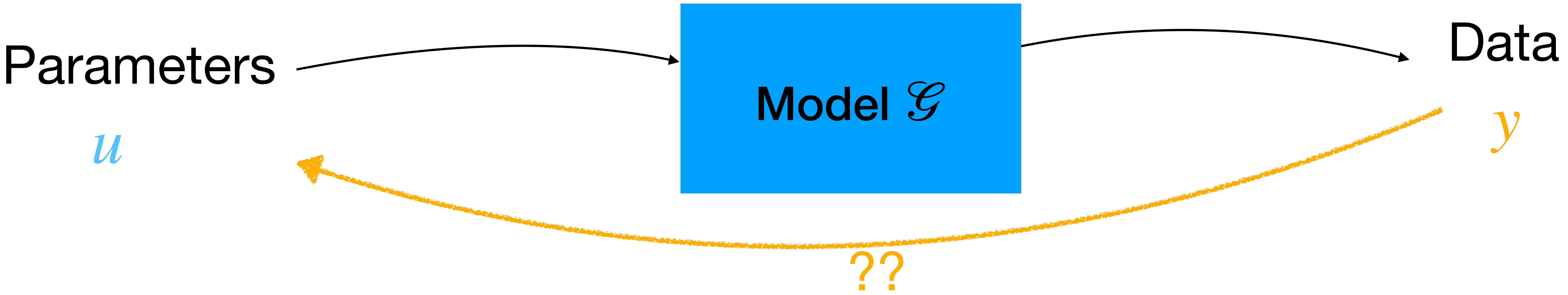
Inverse problems



- Some examples: - training a NN
  - $\mathscr{G}(u)$ = solution to $\nabla \cdot (u(x)\nabla y(x)) = 0, x \in \Omega$ with Dirichlet boundary conditions
  - Image deblurring/denoising:



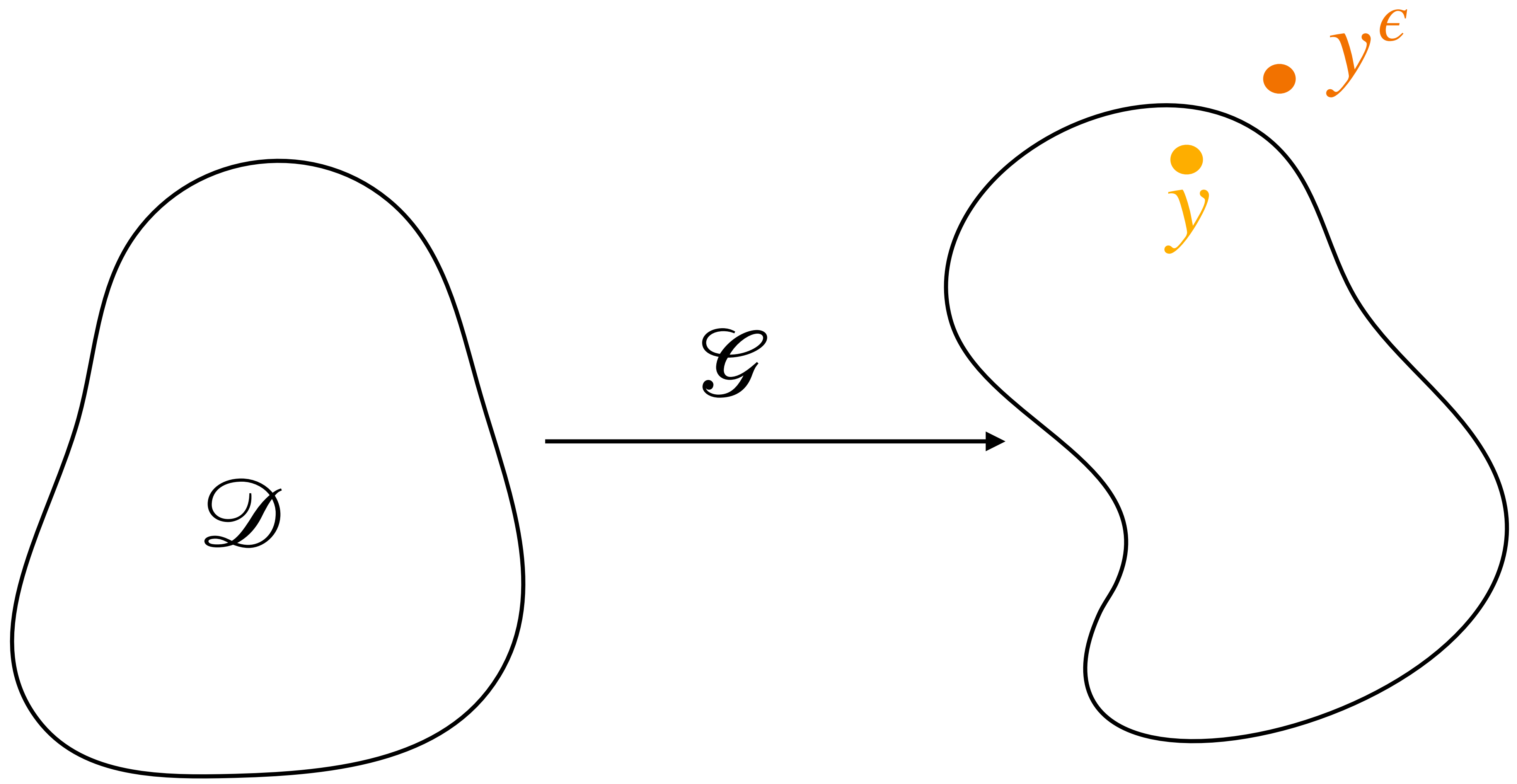
Inverse problems



- Naive idea: $u = \mathscr{G}^{-1}(y)$

  - ✖ $\mathscr{G}$ might not be invertible
  - ✖ Data is noisy $y^{\epsilon} = y + \epsilon$
  - ✖ $y^{\epsilon} \notin \mathscr{G}(\mathscr{D})$
  - ✖ Inversion might not be stable

    $\|\mathscr{G}^{-1}(y) - \mathscr{G}^{-1}(y^{\epsilon})\|_{\mathscr{D}}$ large



Regularization and Gradient Flows

- Reformulate as an optimization problem $\min_{u \in \mathcal{D}} \|\mathcal{G}(u) - y\|_{\mathcal{R}}$

Norm in $\mathcal{R}$

- Add regularization: $\min_{u \in \mathcal{D}} L(u) = \min_{u \in \mathcal{D}} \|\mathcal{G}(u) - y\|_{\mathcal{R}} + \mathrm{R}(u)$

Promote sparsity/smoothnes

- How to solve: gradient descent method $u_{n+1} = u_n - \alpha \nabla L(u_n)$

$$\Big\Downarrow \text{ Continuous time}$$

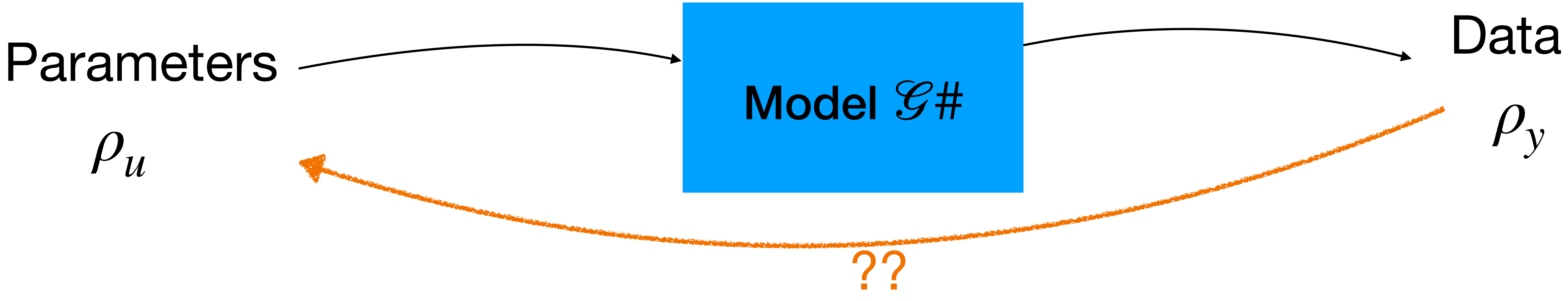
$$\dot{u} = -\nabla_u L(u)$$

Depends on the choice of metric!

- As $t \to \infty$, $u(t) \to u_0$, a minimizer of $L(u)$

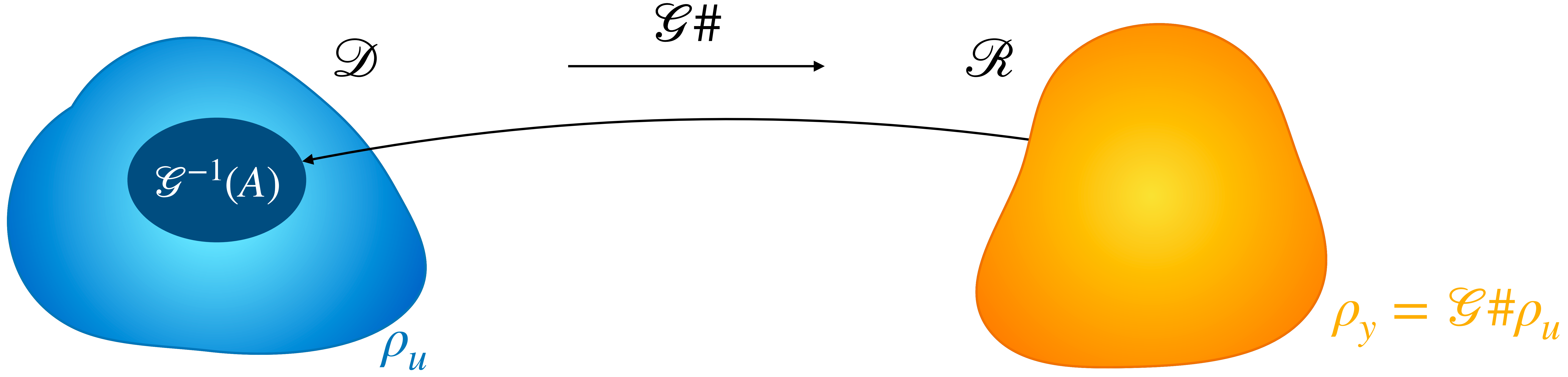
Stochastic inverse problem

- What if the parameter is a distribution?



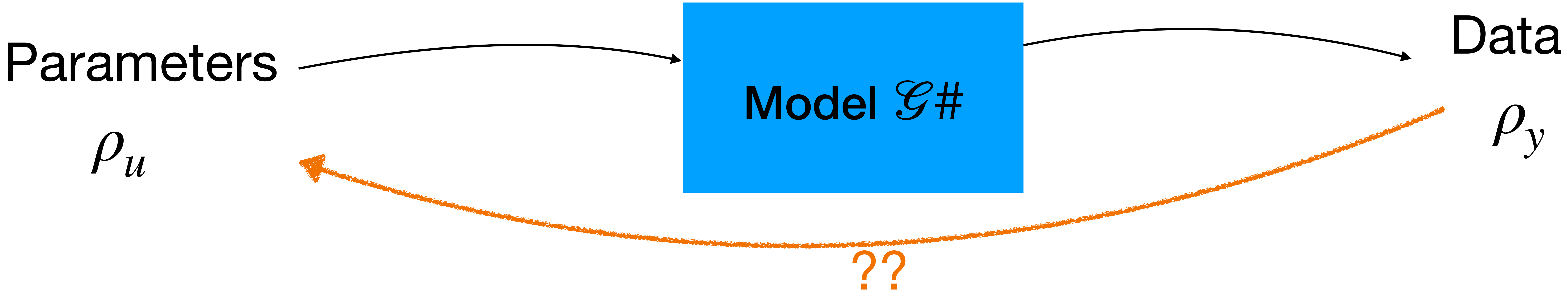
Definition: If $\mathscr{G} : \mathscr{D} \to \mathscr{R}$ then $\mathscr{G}\# : \mathscr{P}(\mathscr{D}) \to \mathscr{P}(\mathscr{R})$ such that $\mathscr{G}\#\mu(A) = \mu(\mathscr{G}^{-1}(A))$

$X \sim \mu \implies \mathscr{G}(X) \sim \mathscr{G}\#\mu$

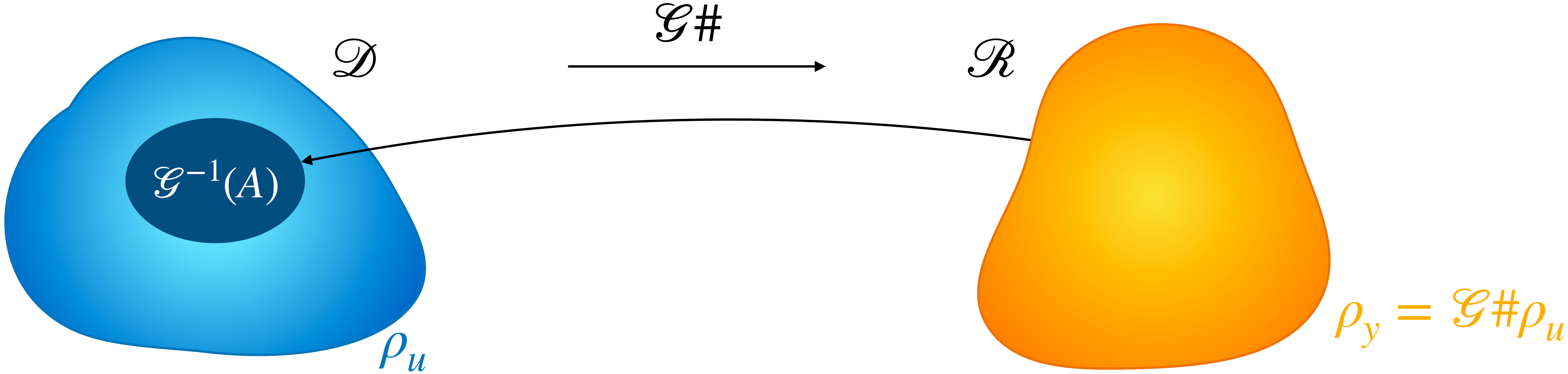


Stochastic inverse problem

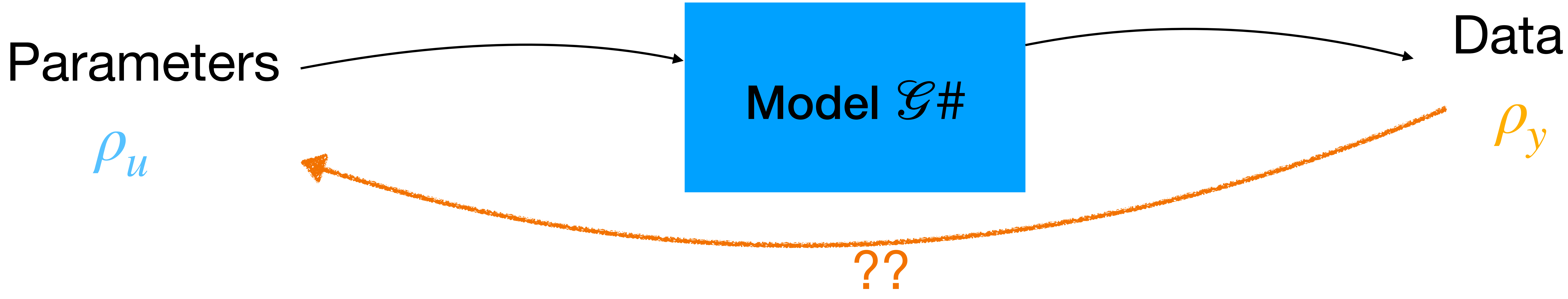
- What if the parameter is a distribution?



$$X \sim \mu \implies \mathscr{G}(X) \sim \mathscr{G}\#\mu$$



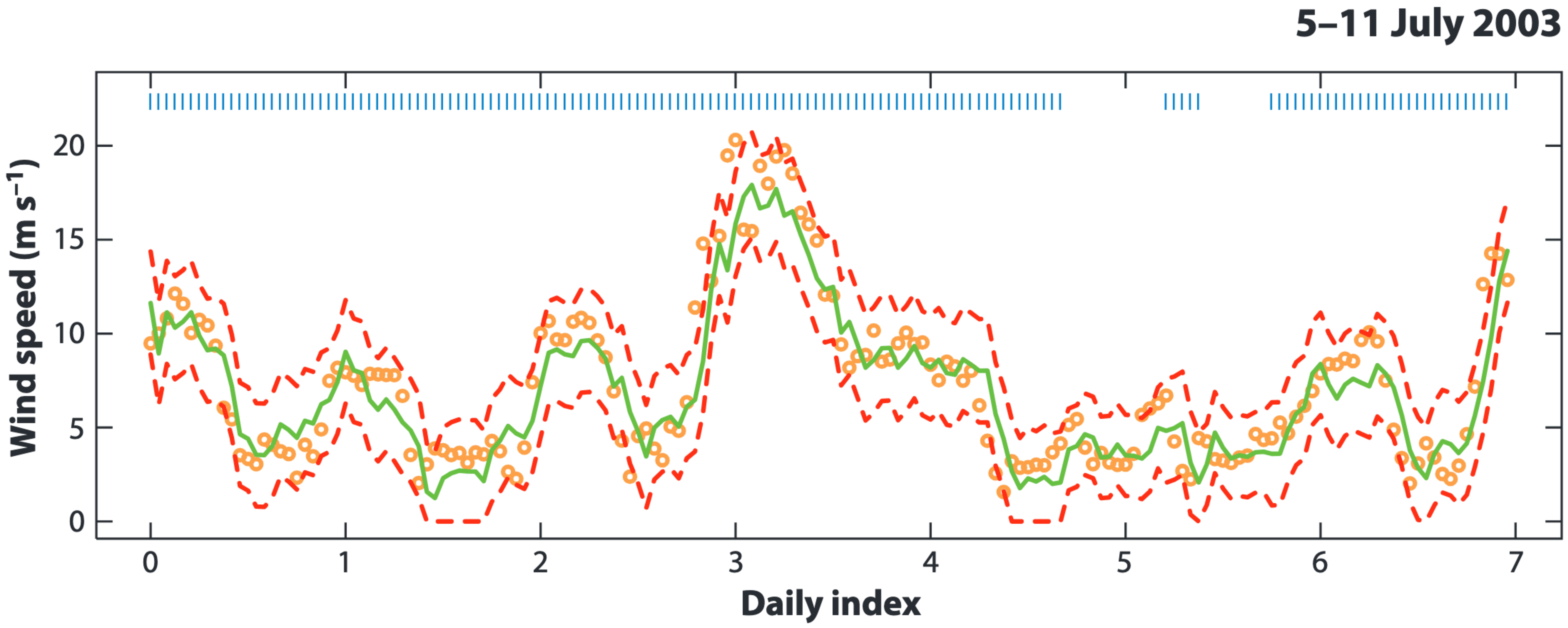
Stochastic inverse problem



- Questions for the rest of the talk:
  - ? How to understand $(\mathscr{G}\#)^{-1}(\rho_y)$? Stability? General $\mathscr{G}$ if invertible, linear under-determined
  - ? What does the variational framework look like? Linear system with W_2 regularizer
  - ? What is the performance of the gradient flow solver? Linear over-determined

Motivation

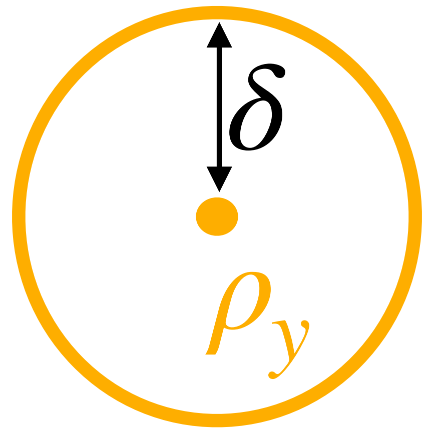
- Repeated measurements $\Longrightarrow$ different results
- Data itself is a distribution: cryoEM, climate modeling, plasma



T. Gneiting and M. Katzfuss, Probabilistic forecasting, Annual Review of Statistics and Its Application, 1 (2014), pp. 125–151

Spaces and norms

- $\rho_u \in \mathscr{P}(\mathscr{D})$; $\rho_y, \rho_y^\delta \in \mathscr{P}(\mathscr{R})$, ρ_y^δ is in the δ ball of ρ_y



measured distribution

- Metrics:

  - Wasserstein distance: $W_p(\mu, \nu) = \left(\inf_{\pi \in \Gamma(\mu,\nu)} \int \|x - y\|_p d\pi \right)^{\frac{1}{p}} \rightarrow \mathscr{P}_p$

  - f divergence $D_f(\mu \| \nu) = \int f\left(\frac{d\mu}{d\nu}\right) d\nu \rightarrow \mathscr{P}_{ac}$

    Ex: KL, χ^2, TV, …

Stochastic Inverse problem

Can be empirical

Goal: Given polluted data ρ_y^δ and the deterministic model, find the parameter distribution ρ_u^δ such that the distance between the true distribution ρ_u and the computed one ρ_u^δ is as small as possible

?

$\longleftarrow$

Direct inversion, linear case

- Assume $\mathcal{G} := A : \mathbb{R}^n \to \mathbb{R}^m$ linear.

- When A is invertible: $(A\#)^{-1} = A^{-1}\#$

Theorem: If $\rho_u = A^{-1}\#\rho_y$ **and** $\rho_u^\delta = A^{-1}\#\rho_y^\delta$ **then:**

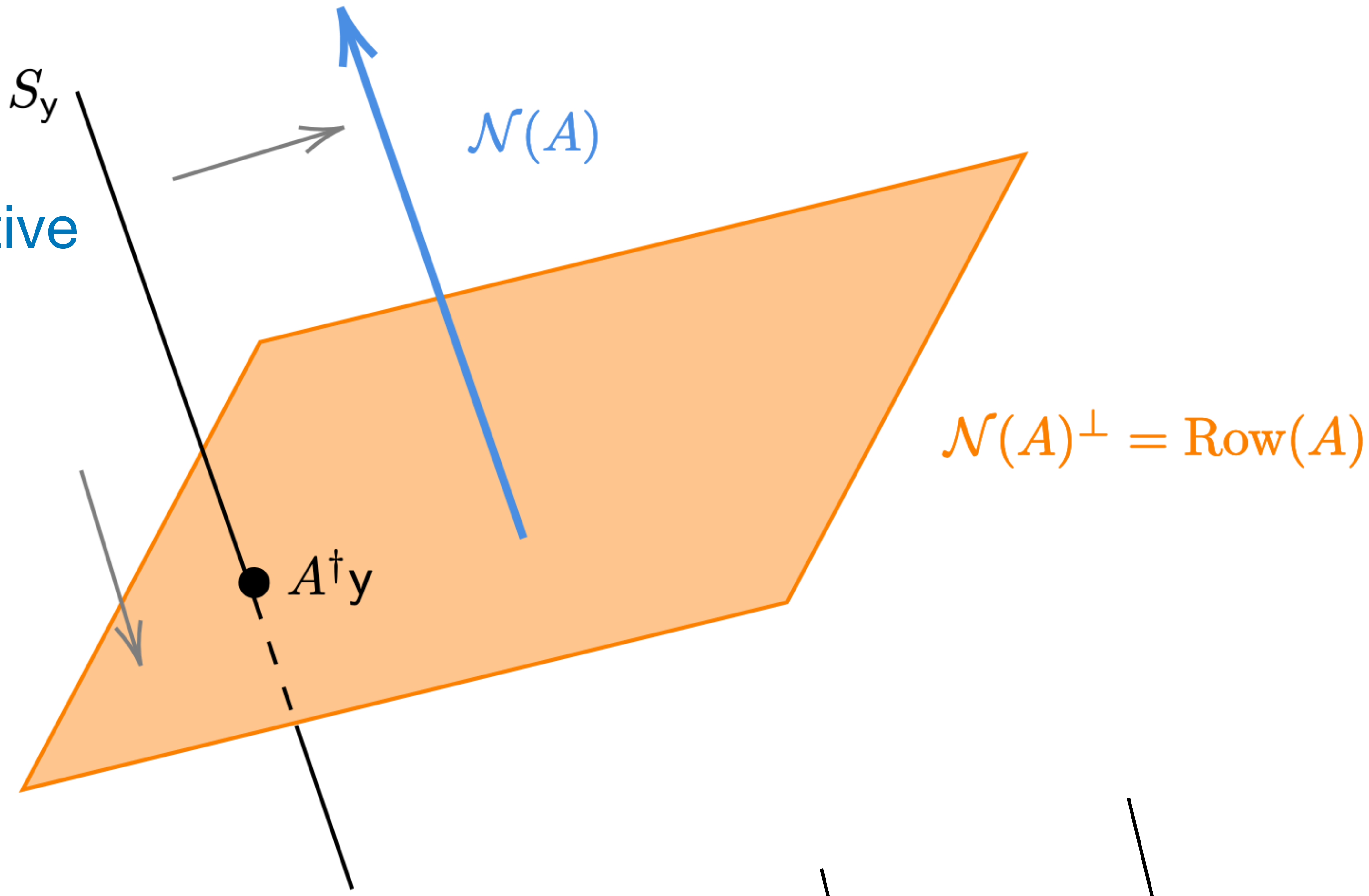
- $W_p(\rho_u, \rho_u^\delta) \leq C_{A^{-1}} W_p(\rho_y, \rho_y^\delta)$ with $C_{A^{-1}} = \dfrac{1}{\sigma_{min}(A)}$
- $D_f(\rho_u \| \rho_u^\delta) = D_f(\rho_y \| \rho_y^\delta)$

Better stability!

- When $\boxed{A}$ is over-determined if $supp(\rho_y) \not\subset Col(A)$ then $\nexists \rho_u$ such that $A\#\rho_u = \rho_y$

Under-determined case

Onto, not injective

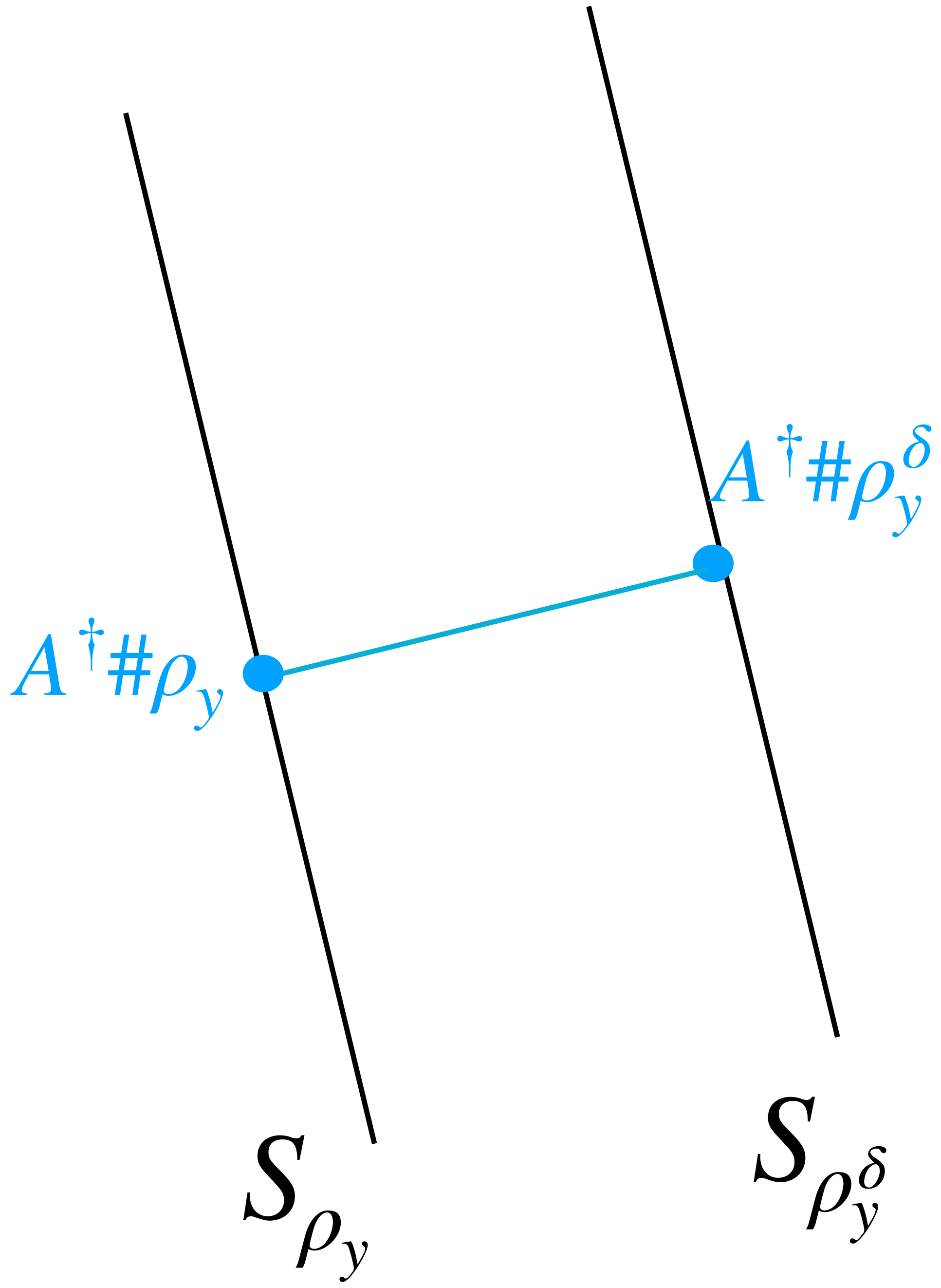


- Solution set $S_{\rho_y} = \{\rho_u \,|\, A\#\rho_u = \rho_y\}$
- Consider the infimum distance

$$d(S_{\rho_y}, S_{\rho_y^\delta}) = \inf_{\substack{\{\mu : A\#\mu = \rho_y^1\} \\ \{\nu : A\#\nu = \rho_y^2\}}} d(\mu, \nu)$$

Theorem: if $\boxed{\textbf{A}}$ has full row rank

- $d^W(S_{\rho_y}, S_{\rho_y^\delta}) = W_2(A^{\dagger}\#\rho_y, A^{\dagger}\#\rho_y^\delta) \le (\sigma_{min}(A))^{-1} W_2(\rho_y, \rho_y^\delta)$ **with**
- $d^f(S_{\rho_y} \| S_{\rho_y^\delta}) = D_f(\rho_y \| \rho_y^\delta)$



- Sup distance $\to \infty$

Variational Framework with regularization

$$\rho_u^\delta = \text{arg} \min_{\rho_u \in \mathscr{P}(\mathscr{D})} E[\rho_u; \rho_y^\delta] := \text{arg} \min_{\rho_u \in \mathscr{P}(\mathscr{D})} D(\mathscr{G}\#\rho_u, \rho_y^\delta) + \alpha \mathrm{R}(\rho_u)$$

- $W_2 - W_2$ pair: $D = W_2$ and $\mathrm{R}(\rho_u) = \underbrace{\int |u|^2 d\rho_u}_{W_2^2(\rho_u, \delta_0)}$

Theorem: If $\mathscr{G} = \boxed{A}$ linear with full column rank then:

- If $\alpha = \delta = 0$ then $\rho_u^* = A^\dagger \# \rho_y$ is the minimizer
- Else, $(\rho_u^\delta)^* = (A^\top A + \alpha^2 \mathrm{I})^{-1} A^\top \# \rho_y^\delta$ is the minimizer

Tikhonov regularization in probability space!

Wasserstein Gradient Flow

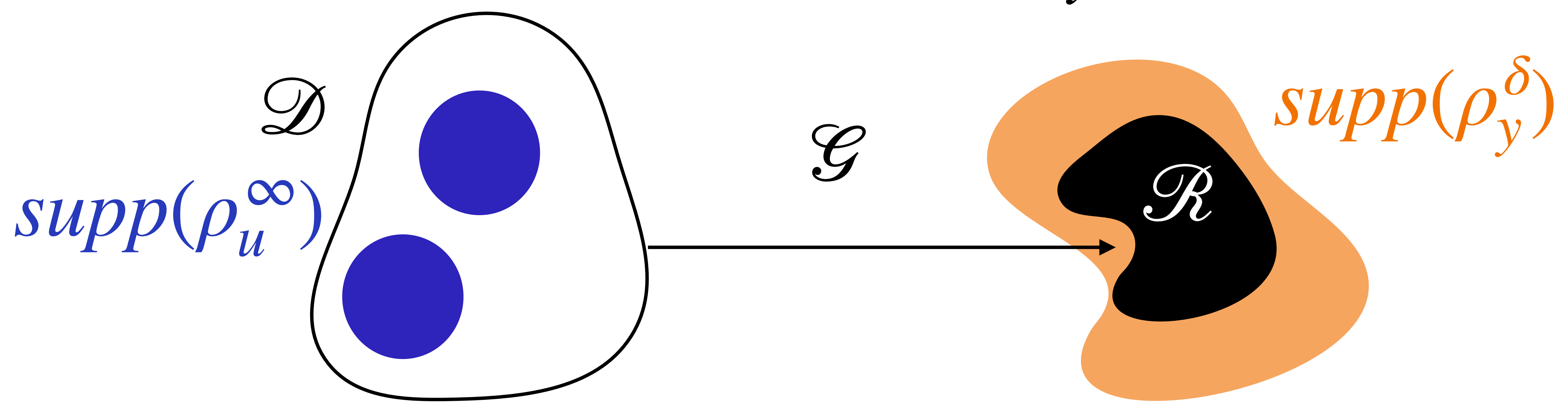
- Treat the case of Wasserstein gradient flow $\partial_t \rho_u = \nabla \cdot \left(\rho_u \nabla \frac{\delta E}{\delta \rho_u} \right)$

- Case 1. Of KL divergence: $D = D_f$

Theorem: Let $\rho_y^\infty = \mathscr{G}\#\rho_u^\infty$ **(equilibrium of the GF), then** $\frac{\rho_y^\infty}{\rho_y^\delta}(\mathscr{G}(u)) = C$ **on connected components of** $supp(\rho_u^\infty)$**. If** $supp(\rho_y^\delta) = \mathscr{R}$ **then** $\rho_y^\infty = \rho_y^\delta$

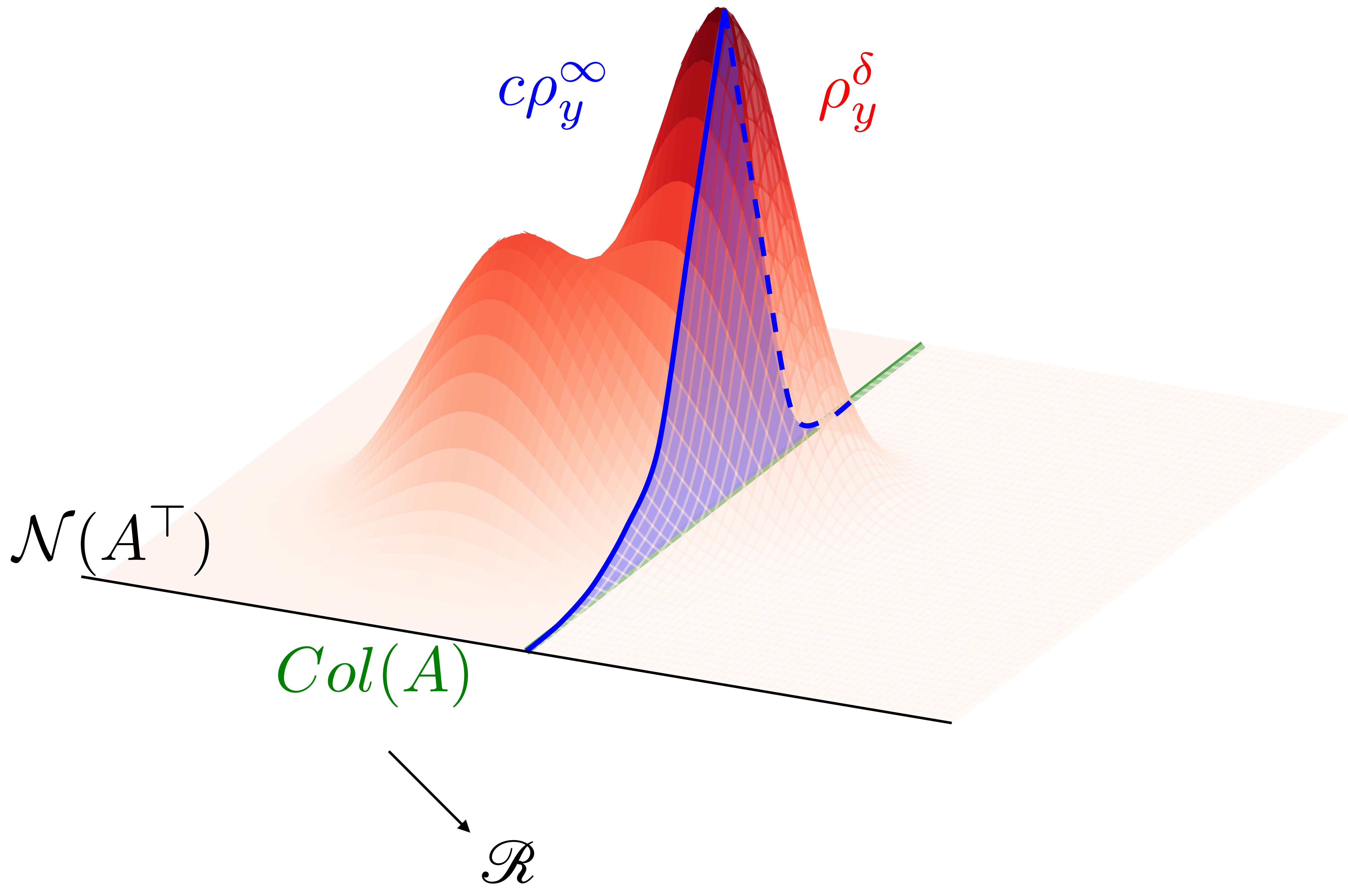
If $\mathscr{R} \subset supp(\rho_y^\delta)$ **then** $\rho_y^\infty = \rho_y^\delta | \mathscr{R}$

- If $\mathscr{G} = A$ linear and over-determined then ρ_y^∞ is the conditional distribution of ρ_y^δ on $Col(A)$



Wasserstein Gradient Flow of KL divergence

- If $\mathscr{G} = A$ linear and over-determined then ρ_y^∞ is the conditional distribution of ρ_y^δ on $Col(A)$

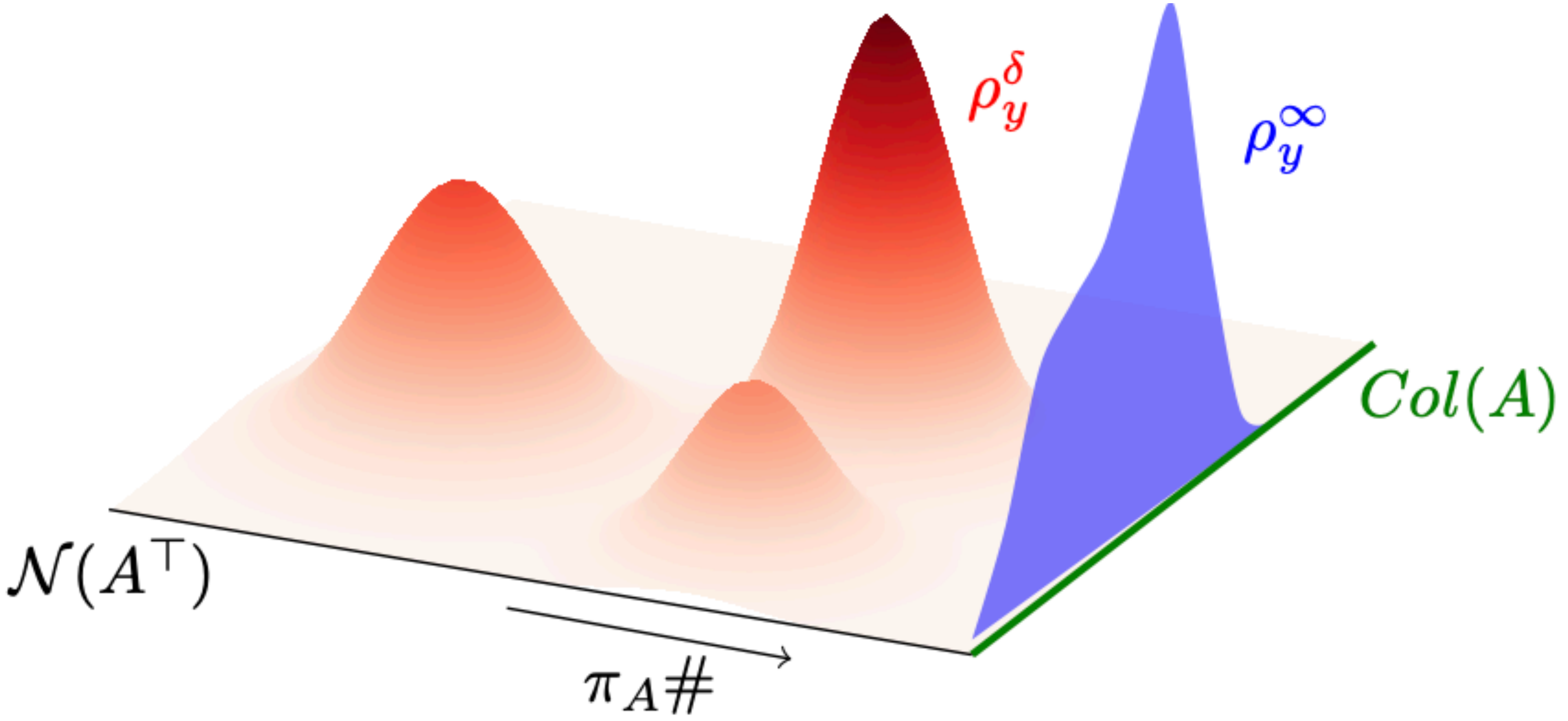


Wasserstein Gradient Flow of W_2

- Case 2. $D = W_2$, consider $\mathscr{G} = A$. Then $\rho_u^\infty = A^\dagger \# \rho_y^\delta$

Projection onto $Col(A)$

- ρ_y^∞ recovers the marginal distribution of ρ_y^δ onto $Col(A)$



Thank you for listening! Questions?

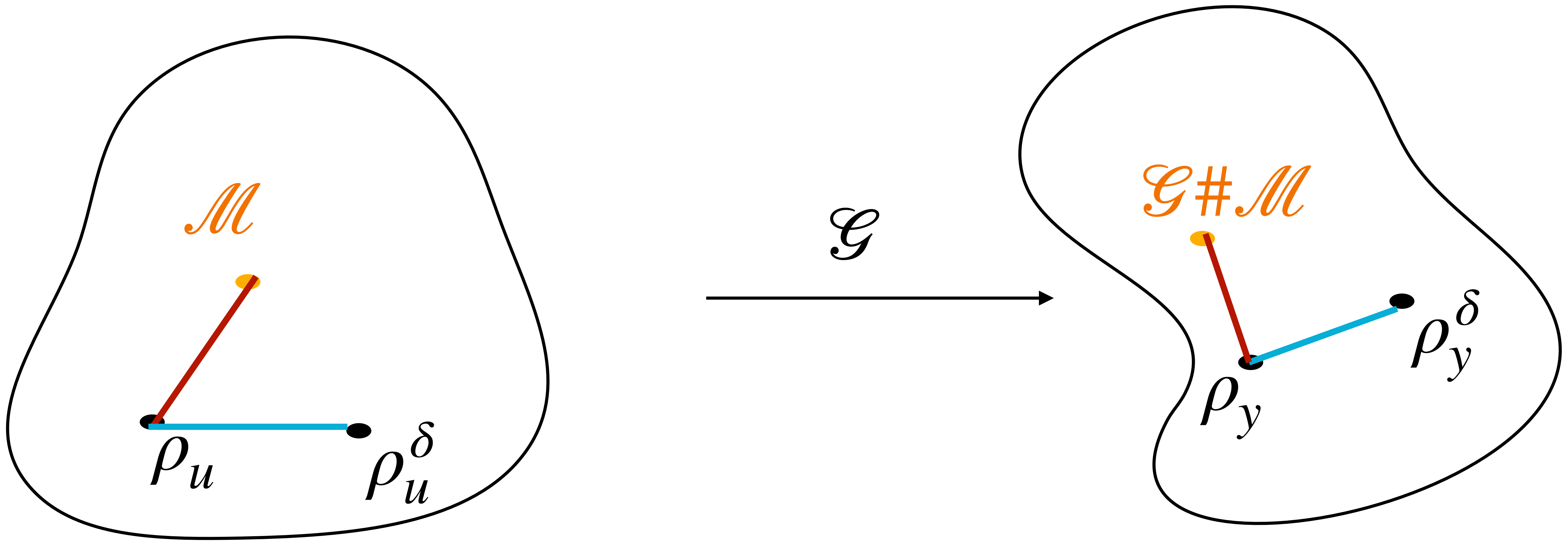
Stochastic Inverse Problem: stability, regularization and Wasserstein gradient flow, Qin Li, O., Li Wang, Yunan Yang, arxiv:2410.00229

Supplemental slides

- Regularization with Entropy-Entropy pair: $D = KL$ and R $= KL(\cdot \| \mathscr{M})$

Theorem: If $\mathscr{G}$ is invertible then:

- $\rho_u^\delta = C\left(\mathscr{G}^{-1}\#\rho_y^\delta \mathscr{M}^\alpha\right)^{\frac{1}{1+\alpha}}$
- $KL(\rho_u \| \rho_u^\delta) = \dfrac{1}{\alpha+1} KL(\rho_y \| \rho_y^\delta) + \dfrac{\alpha}{\alpha+1} KL(\rho_y \| \mathscr{G}\#\mathscr{M}) - \log C$



Variational Framework with regularization

$$\rho_u^\delta = \arg\min_{\rho_u \in \mathscr{P}(\mathscr{D})} E[\rho_u; \rho_y^\delta] := \arg\min_{\rho_u \in \mathscr{P}(\mathscr{D})} D(\mathscr{G}\#\rho_u, \rho_y^\delta) + \alpha \mathrm{R}(\rho_u)$$

- Existence: if E is l.s.c and coercive

Desired output measure

- Entropy-Entropy pair: $D = KL$ and R $= KL(\cdot \| \mathscr{M})$

Theorem: If $\mathscr{G}$ is invertible then:

- $\rho_u^\delta = C\left(\mathscr{G}^{-1}\#\rho_y^\delta \mathscr{M}^\alpha\right)^{\frac{1}{1+\alpha}}$
- $KL(\rho_u \| \rho_u^\delta) = \dfrac{1}{\alpha+1} KL(\rho_y \| \rho_y^\delta) + \dfrac{\alpha}{\alpha+1} KL(\rho_y \| \mathscr{G}^{-1}\#\mathscr{M}) - \log C$

Exponential convergence

Theorem: If $\mathcal{G} = A$ linear, $D = KL$ and ρ_y^δ is $\lambda-$log concave and $KL(\rho_y(0)||\rho_y^\delta) < \infty$ then $\rho_y(t)$ converges to the conditional distribution $\rho_y^\delta \,|\, Col(A) := \rho_{y_A}^\delta$ exponentially fast in terms of the KL divergence i.e:

$$KL(\rho_y(t)||\rho_{y_A}) \leq exp(-2\lambda\sigma_{min}^2 t)KL(\rho_y(0)||\rho_y^\delta)$$

Proof for KL Wasserstein Gradient Flow

- $$\partial_t \rho_u = \nabla \cdot \left(\rho_u \nabla \frac{\delta E}{\delta \rho_u} \right) \implies \partial_t \rho_y = \nabla_y \cdot \left[\rho_y B(y) \nabla_y \log \left(\frac{\rho_y}{\rho_y^\delta} \right) \right], \quad y \in \mathscr{R}$$

- At equilibrium RHS = 0

$$\frac{\rho_y^\infty}{\rho_y^\delta}(\mathscr{G}(u)) = C$$ on simply connected subsets of supp(ρ_u^∞)