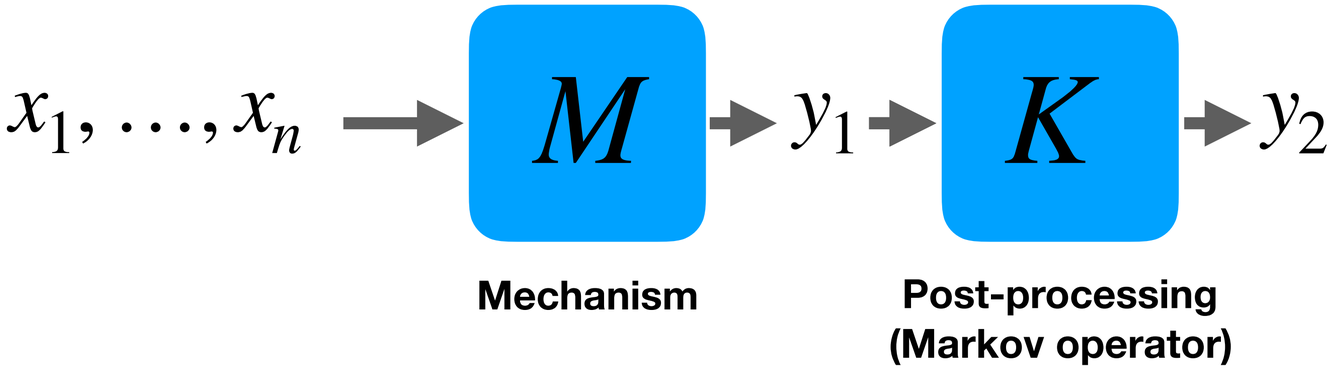
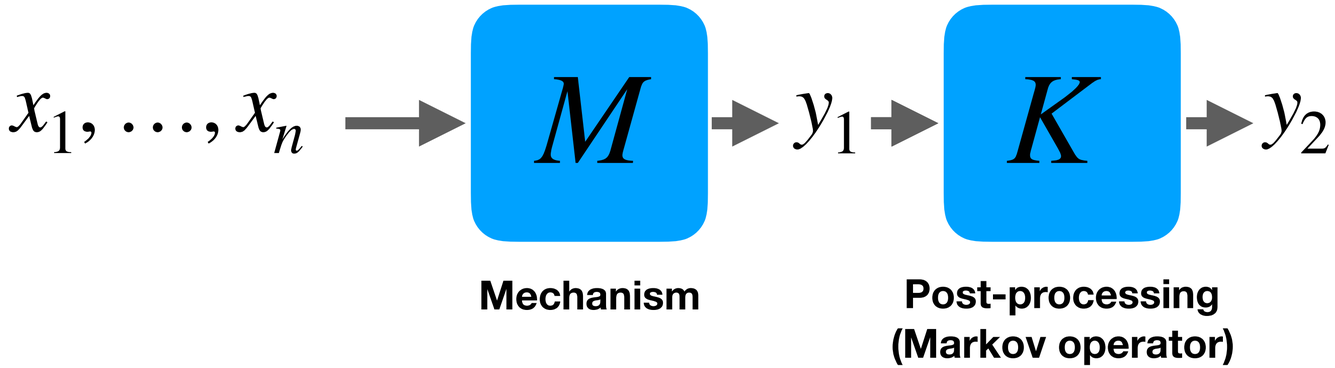
Privacy Amplification by Mixing and Diffusion Mechanisms

Borja Balle, Gilles Barthe, Marco Gaboardi, Joseph Geumlek

Amplification by Postprocessing

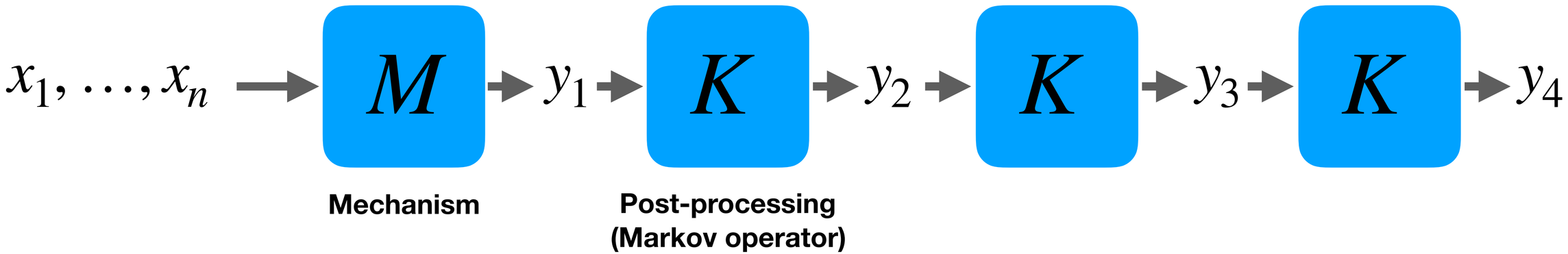


Amplification by Postprocessing



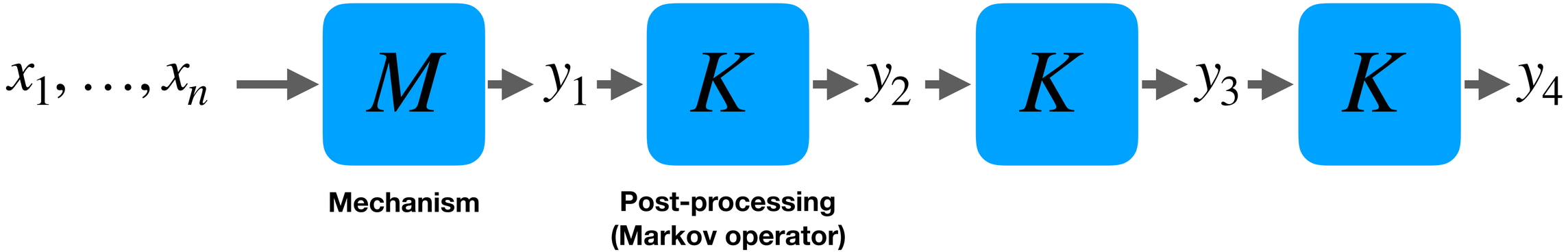
- When is K∘M more private than M?

Amplification by Postprocessing



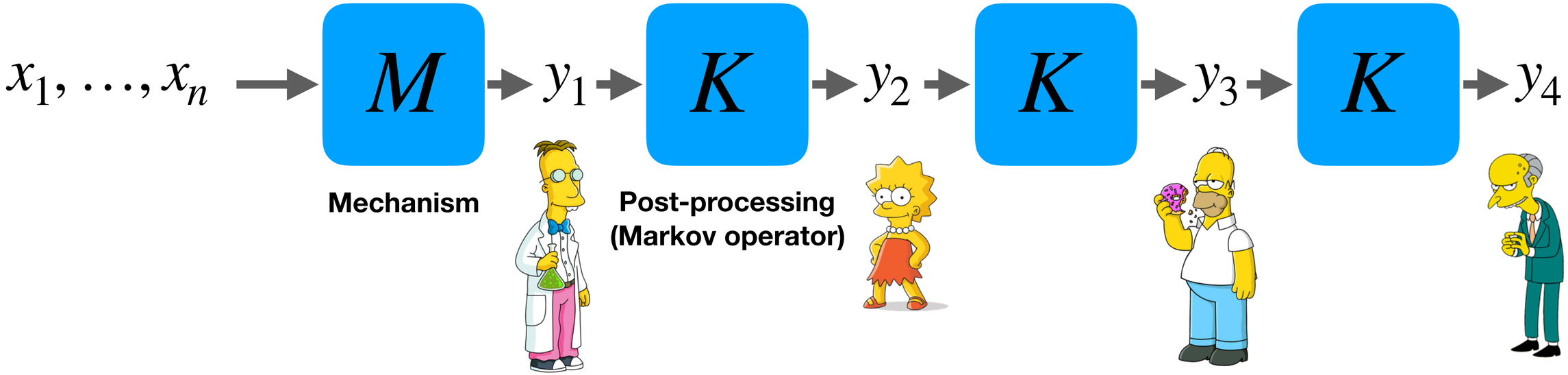
- When is K○M more private than M?

Amplification by Postprocessing



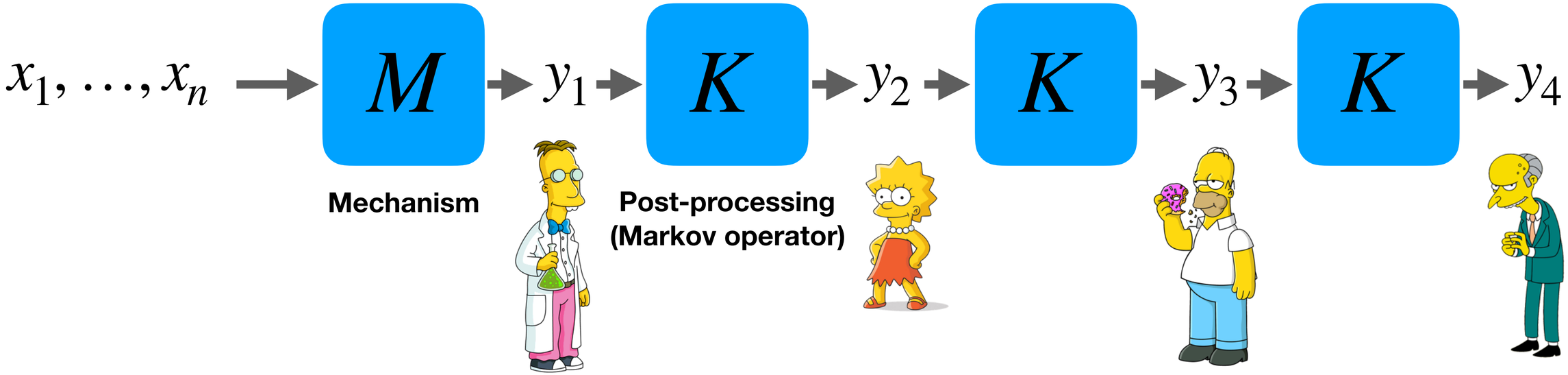
- When is K○M more private than M?
- How does privacy relate to mixing in the Markov chain?

Amplification by Postprocessing



- When is K∘M more private than M?
- How does privacy relate to mixing in the Markov chain?

Amplification by Postprocessing



- When is K○M more private than M?
- How does privacy relate to mixing in the Markov chain?
- Starting point for “Hierarchical DP”

Our Results

- **Amplification under uniform mixing**
  - Relates to classical mixing conditions (eg. Dobrushin, Doeblin) and local DP properties of K
  - Eg. if M is ε-DP and K is $\log \frac{1}{1-\gamma}$-LDP, them K∘M is $\log(1+\gamma(e^{\varepsilon}-1))$-DP

Our Results

- **Amplification under uniform mixing**
  - Relates to classical mixing conditions (eg. Dobrushin, Doeblin) and local DP properties of K
  - Eg. if M is ε-DP and K is $\log \frac{1}{1-\gamma}$-LDP, them K∘M is $\log(1+\gamma(e^{\varepsilon}-1))$-DP
- **Amplification from couplings**
  - Generalizes amplification by iteration [Feldman et al. 2018]
  - Applied to SGD: exponential amplification in the strongly convex case

Our Results

- **Amplification under uniform mixing**
  - Relates to classical mixing conditions (eg. Dobrushin, Doeblin) and local DP properties of K
  - Eg. if M is ε-DP and K is $\log \frac{1}{1-\gamma}$-LDP, them K$\circ$M is $\log(1+\gamma(e^{\varepsilon}-1))$-DP
- **Amplification from couplings**
  - Generalizes amplification by iteration [Feldman et al. 2018]
  - Applied to SGD: exponential amplification in the strongly convex case
- **The continuous time limit: diffusion mechanisms**
  - General RDP analysis via heat-flow argument
  - New Ornstein-Uhlenbeck mechanism with better MSE than Gaussian mechanism

Amplification by Iteration in NoisySGD

Algorithm 1: Noisy Projected Stochastic Gradient Descent — NoisyProjSGD$(D, \ell, \eta, \sigma, \xi_0)$

Input: Dataset $D = (z_1, \ldots, z_n)$, loss function $\ell : \mathbb{K} \times \mathbb{D} \to \mathbb{R}$, learning rate η, noise parameter σ, initial distribution $\xi_0 \in \mathcal{P}(\mathbb{K})$

Sample $x_0 \sim \xi_0$

for $i \in [n]$ **do**

  $x_i \leftarrow \Pi_{\mathbb{K}}\left(x_{i-1} - \eta(\nabla_x \ell(x_{i-1}, z_i) + Z)\right)$ with $Z \sim \mathcal{N}(0, \sigma^2 I)$

return x_n

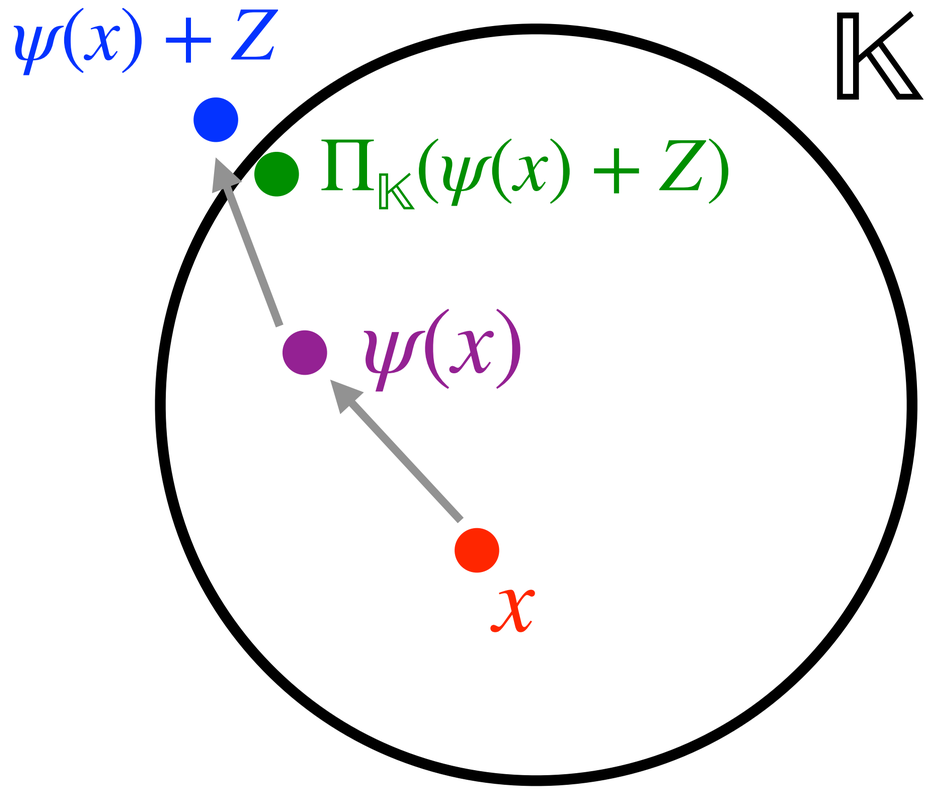
- If D and D' differ in position j, then the last n-j iterations are postprocessing
  - Can also use public data for the last r iterations
- Start from a coupling between x_j and x_j' and propagate it through
  - Keep all the mass as close to the diagonal as possible

[FMTT'18]

Projected Generalized Gaussian Mechanism

$$K(x) = \Pi_{\mathbb{K}}(\mathcal{N}(\psi(x), \sigma^2 I))$$

$$\psi : \mathbb{R}^d \to \mathbb{R}^d$$



Amplification by Coupling

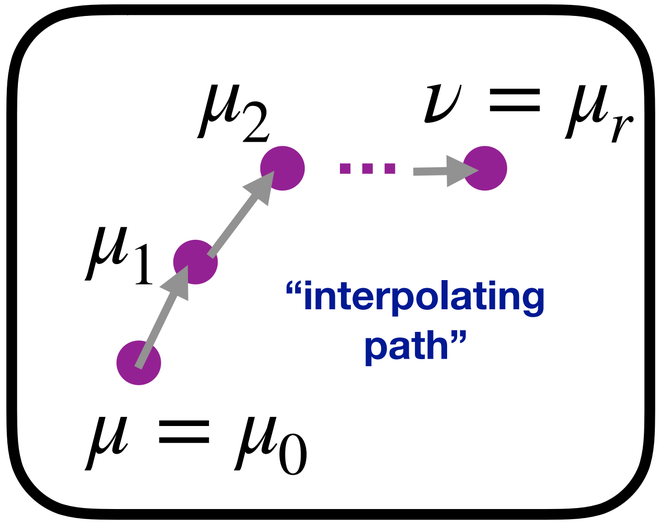
Suppose $\psi_1, \ldots, \psi_r$ are L-Lipschitz

$$K_i(x) = \Pi_{\mathbb{K}}(\mathcal{N}(\psi_i(x), \sigma^2 I))$$

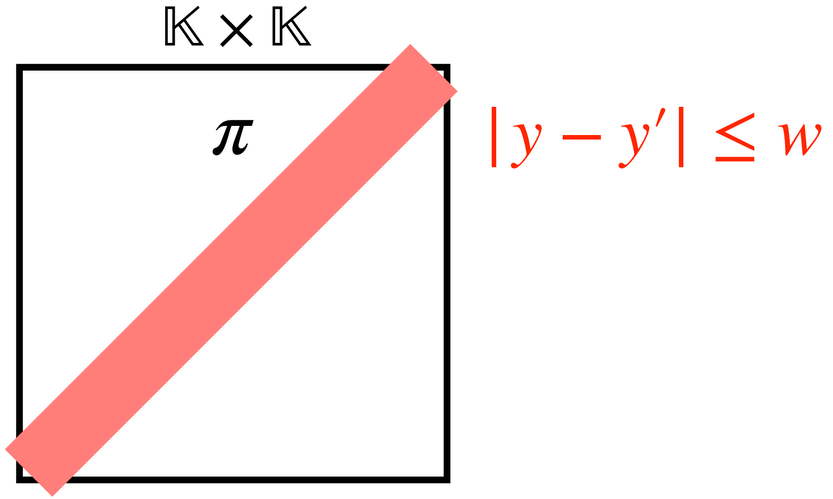
$$\mathsf{R}_\alpha(\mu K_1 \cdots K_r \| \nu K_1 \cdots K_r) \leq \frac{\alpha L^2}{2\sigma^2} \sum_{i=1}^{r} L^{2(r-i)} \mathsf{W}_\infty(\mu_i, \mu_{i-1})^2$$

Rényi Divergence **Wasserstein Distance**

$\mathscr{P}(\mathbb{K})$



Applications:
- Bound L
- Optimize path



Per-index RDP in NoisySGD

Suppose the loss is Lipschitz and smooth

*If loss is **convex** can take L=1. Then i-th person receives $\varepsilon_i(\alpha)$-RDP with*

$$\epsilon_i(\alpha) = O\left(\frac{\alpha}{(n-i)\sigma^2}\right)$$

[FMTT'18]

*If loss is **strongly convex** can take L< 1. Then i-th person receives $\varepsilon_i(\alpha)$-RDP with*

$$\epsilon_i(\alpha) = O\left(\frac{\alpha L^{(n-i)/2}}{(n-i)\sigma^2}\right)$$

Summary

- Couplings (including overlapping mixtures) provide a powerful methodology to study privacy amplification in many settings
  - Including: subsampling, postprocessing, shuffling and iteration
- Properties of divergences related to (R)DP (eg. advanced joint convexity) are “necessary” to get tight amplification bounds
- Different types of couplings are useful (eg. maximal and small distance)