# Sampling from arbitrary distributions using PSD Models

Ulysse Marteau-Ferey, Francis Bach, Alessandro Rudi

## Modelling probability distributions

### Main properties needed and issues

**Goal :** Find model $f_\theta$ for un-normalized distributions

**(P) Positivity :** $f_\theta \geq 0$

**(I) Integrability :** $I_\theta = \int_{\mathcal{X}} f_\theta(x)\ dx$ computable

**(A) Good approximation power.** $\exists \theta,\ \frac{f_\theta}{I_\theta} \approx p$ for many $p$

**Models in the literature:**

1) Generalised linear models $g_\theta(x) = \exp(\theta^\top \varphi(x))$ **(I)**
2) Linear models $g_\theta(x) = \theta^\top \varphi(x)$ **(P)**
3) Nadarawa-Watson : $f_\alpha(x) = \sum_{i=1}^{n} \alpha_i k(x_i, x),\ k \geq 0,\ \alpha \geq 0$ **(A)**

### Definition of PSD models

**Intuition** : linear model

$$g_a(x; (x_j), \tau) = \sum_{j=1}^{m} a_j k_\tau(x, x_j),\ k_\tau(x, x') = e^{-\tau \|x - x'\|^2}$$

**Rank-1 PSD model :** square of a linear model

$$f_A(x) = g_a(x)^2 = \sum_{i,j=1}^{m} A_{ij} k_\tau(x, x_j) k_\tau(x, x_i),\ A = aa^\top$$

**General PSD model :**

$$f_A(x) = \sum_{i,j=1}^{m} A_{ij} k_\tau(x, x_j) k_\tau(x, x_i),\ A \succeq 0$$

### Properties

1) (P), (I), and (A) satisfied
2) **Convexity** : if $L$ is a convex loss, the following is convex:

$$\min_A L(f_A(\tilde{x}_1), \ldots, f_A(\tilde{x}_n)) + Tr(A)$$

3) **Integrals on hyper-rectangles** can be computed easily

$$I(Q) = \int_Q f_A(x)\ dx = \sum_{ij} A_{ij} K_{ij} G_{ij} \text{ where } K_{ij} = k_{\tau/2}(x_i, x_j)$$

$[G_Q]_{ij}$ is computable using the $erf$ function $erf(t) = \int_0^t e^{-u^2} du$

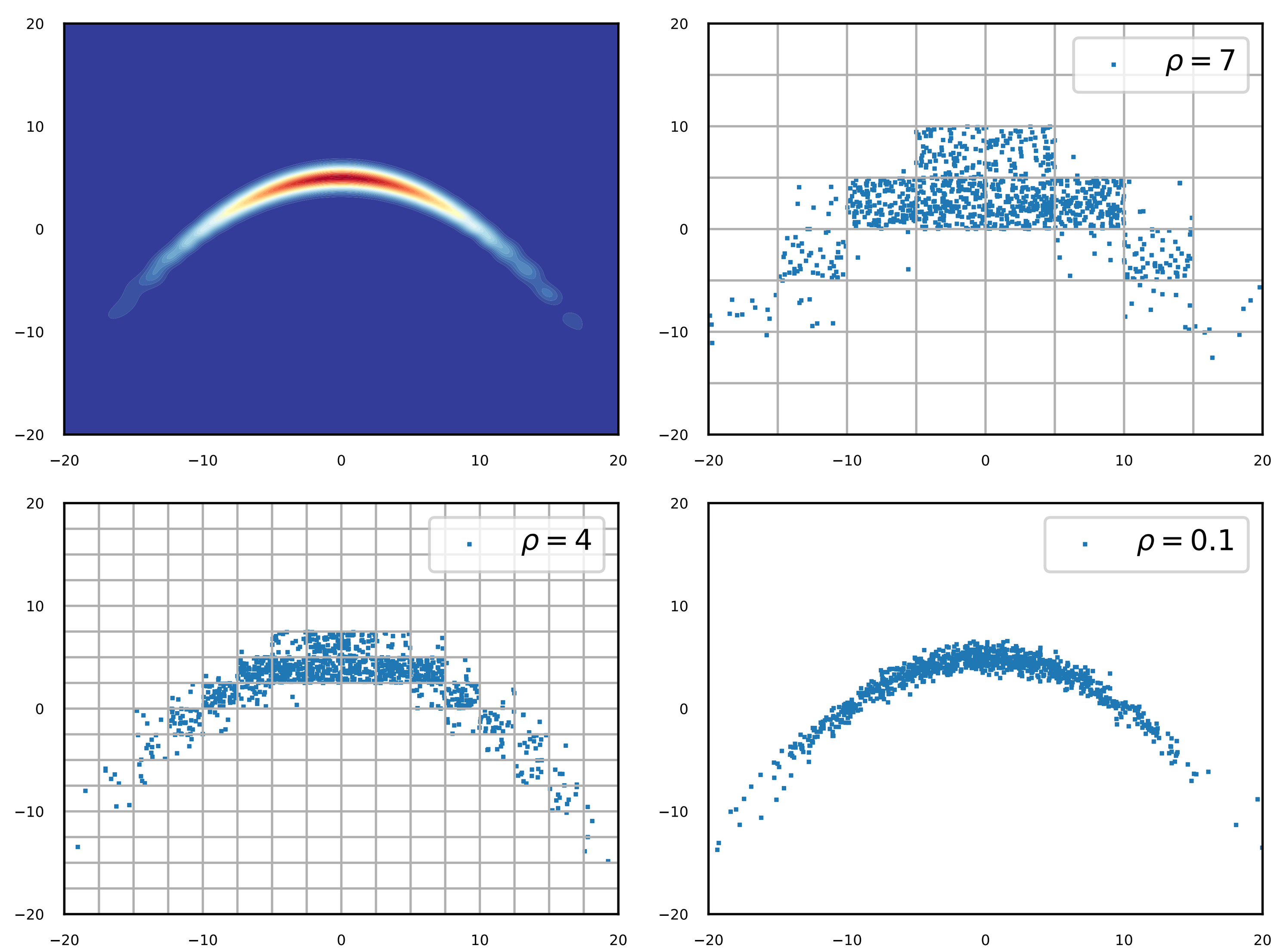
## Sampling from PSD models

**Algorithm 1** Approximately sampling from $p_Q$

**function** SampleRec($Q, N, \rho$)
  **if** $N = 0$ **then**
    **return** EmptyList
  **else if** MaxLen($Q$) $\leq \rho$ **then**
    **return** SampleUniform($Q, N$)
  **else**
    $Q_1, Q_2 =$ SplitLargestSide($Q$)
    $q = I(Q_1)/I(Q)$
    $k =$ SampleBinomial($N, q$)
    $L_1 =$ SampleRec($Q_1, k, \rho$)
    $L_2 =$ SampleRec($Q_2, N - k, \rho$)
    **return** Concatenate($L_1, L_2$)
  **end if**
**end function**

**function** Sample($Q, N, \rho$)
  $L =$ SampleRec($Q, n, \rho$)
  **return** RandomPerm($L$)
**end function**

Based on integrals : We can compute I(Q) on hyper-rectangles




### Theoretical grantees

**Samples :** $N$ i.i.d. samples from

$$p_\rho = \frac{1}{|I(Q)|} \sum_{Q_\rho \in \mathcal{D}_\rho(Q)} \frac{I(Q_\rho)}{|Q_\rho|} 1_{Q_\rho}$$

**Guarantees :**

$$d_{TV}(p_\rho, p) \leq d\ Lip(p)\rho \qquad \mathbb{W}_1(p, p_\rho) \leq \sqrt{d}\rho$$

**Adaptivity :** Possibility to choose $\rho$ in adaptive way :

$$Lip(p) \leq \sqrt{8\tau}\ \|K^{1/2} A K^{1/2}\|_F, \quad K_{ij} = k_\tau(x_i, x_j)$$

## Sampling from arbitrary distributions

### Step 1 : Learning a PSD model from a an un-normalized distribution

**Target :** distribution $p \propto g^2$ on $[-1,1]^d$

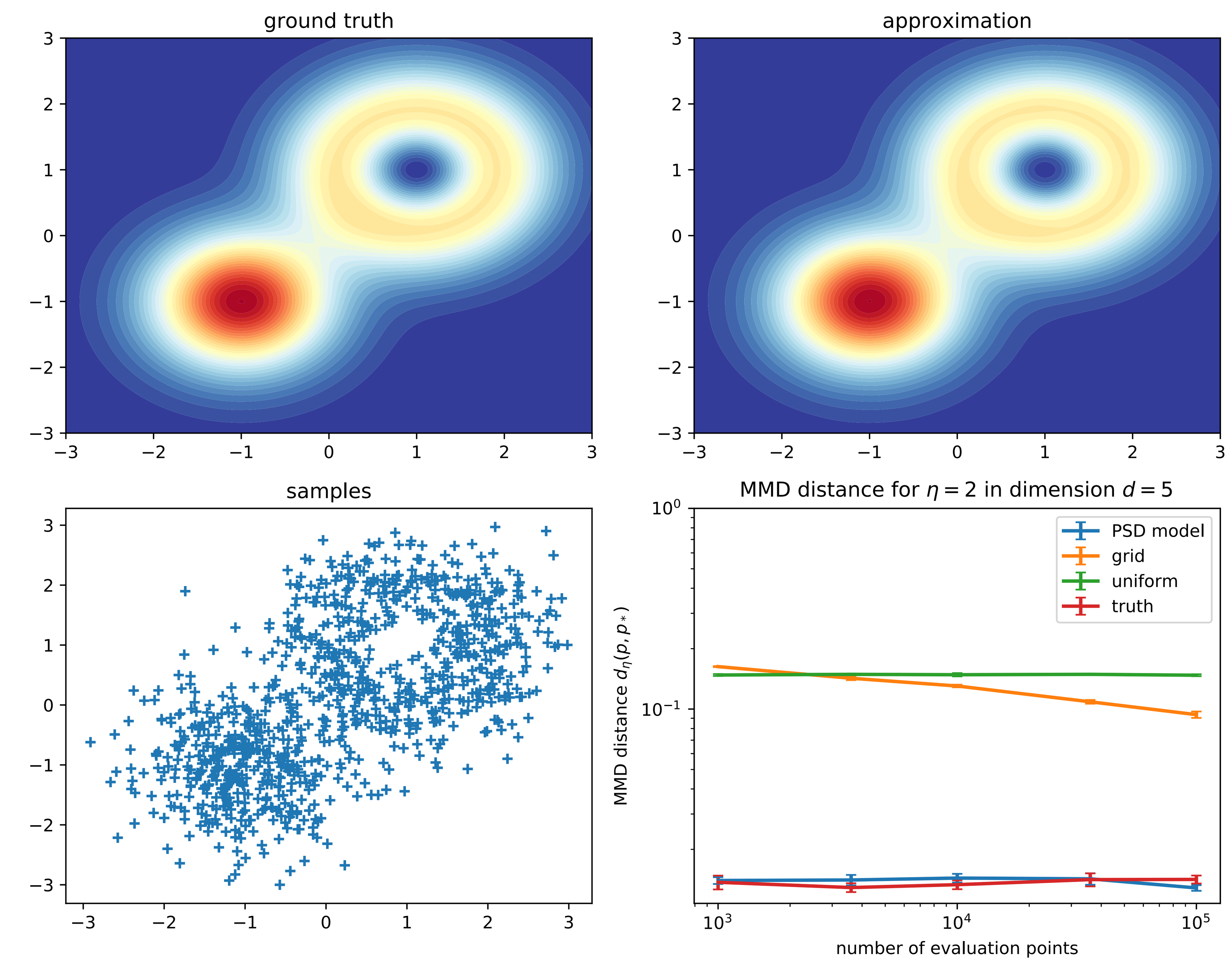
**Samples :** $(\tilde{x}_i)_{1 \leq i \leq n}$ uniformly on $[-1,1]^d$

**Approximate g :**

$$a \in \arg\min_{a \in \mathbb{R}^n} \frac{1}{n} \sum_{i=1}^{n} |g_a(x_i) - g(x_i)|^2 + \lambda \|a\|_K^2$$

**Approximate p :** $\hat{p} \propto g_a^2$

### Step 2 : Apply algorithm with $\rho$ on $\hat{p}$




### Theoretical guarantees

**Assumption :** $p = q^2,\ \partial_\alpha q \in L^2(\mathcal{X}),\ |\alpha| \leq \beta$

**Conciseness :** $m \geq C_{p,d,\beta}\ \epsilon^{-d/\beta}$

**Number of samples :** $n \geq C_{p,d,\beta}\ \epsilon^{-2d/\beta}$

$$\exists \lambda,\ \sigma,\ d_{TV}(p, \hat{p}),\ d(p, p_\rho) \leq \epsilon$$