Fast Empirical Scenarios

MICHAEL MULTERER PAUL SCHNEIDER **ROHAN SEN**

USI Lugano, Switzerland

YAMC 2022, Arenzano

September 21, 2022

Table of Contents

INTRODUCTION

Motivation

- Quadrature rules approximate integrals through a small number of *nodes* and *weights* pertaining to a discrete probability measure.
- These nodes parsimoniously describe the important states or *scenarios* that are the best low-dimensional representation of the underlying complicated distribution.
- These scenarios reconcile moment matrices that often feature in many applied situations.
- Our goal is to extract these low-dimensional scenarios and their probabilities from large and high-dimensional datasets.

Contribution

- The extant algorithms do not scale well when solving large and high-dimensional problems and also suffer from numerical instability.
- We propose algorithms that are tractable and computationally efficient at the same and are founded on the intersection of the truncated moment problem from probability theory and reproducing kernel Hilbert spaces.
- We propose a novel approach for the extraction of the scenarios and their probabilities for the specialized case of covariance matrices of high-dimensional random variables.
- We also modify Lasserre's algorithm for multivariate Gaussian quadrature that partially remedies its numerical instability and significantly improves its computational complexity.

Terminologies

Notations

- $\mathcal{S}(\boldsymbol{\mathcal{H}})$: space of symmetric matrices on $\boldsymbol{\mathcal{H}}$
- $\boldsymbol{A} \succeq 0$: $\boldsymbol{A}$ is positive semi-definite matrix
- $\boldsymbol{A} \in \mathcal{S}_+^N$: $\boldsymbol{A} \in \mathbb{R}^{N \times N}$ is symmetric and positive semi-definite
- $\boldsymbol{\mathcal{P}}_t(\boldsymbol{\Omega})$: space of multivariate polynomials on $\boldsymbol{\Omega}$ of maximum degree t
- $s(t)$: $\binom{d+t}{t}$
- $||\cdot||_F$: Frobenius norm
- $||\cdot||_\star$: trace norm

TRUNCATED MOMENT PROBLEM

Moment Sequence and Linear Form

Truncated sequence (in d variables and of degree t):

$$\boldsymbol{y} = (y_{\boldsymbol{\alpha}}) \text{ where } \boldsymbol{\alpha} \in \overline{\mathbb{N}}_t^d := \left\{ (\alpha_1, \cdots, \alpha_d) : \alpha_i \in \overline{\mathbb{N}}, |\alpha_1 + \cdots + \alpha_d| \leq t \right\}$$

Truncated moment sequence (tms):

$$y_{\boldsymbol{\alpha}} = \int_{\Omega} \boldsymbol{x}^{\boldsymbol{\alpha}} \, d\mu(\boldsymbol{x}) \text{ where } \boldsymbol{x}^{\boldsymbol{\alpha}} := x_1^{\alpha_1} \cdots x_d^{\alpha_d}$$

Monomial basis for $\boldsymbol{\mathcal{P}}_t = span \left\{ \boldsymbol{x}^{\alpha} : |\boldsymbol{\alpha}| \leq t \right\}$:

$$\boldsymbol{\tau}_t(\boldsymbol{x}) := \left[1, x_1, \cdots, x_d, \cdots, x_1^t, \cdots, x_d^t \right] \in \mathbb{R}^{s(t)}$$

Riesz functional: Given $\boldsymbol{y} = (y_{\boldsymbol{\alpha}})$ define $\mathscr{L}_{\boldsymbol{y}} \in \left(\mathbb{R}[\boldsymbol{x}] \right)^*$ as:

$$\mathscr{L}_{\boldsymbol{y}}(p) := \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} y_{\boldsymbol{\alpha}} \quad \text{for} \quad p = \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}}$$

Truncated Moment Problem

The truncated moment problem:

Given a truncated sequence y, does there exist a representing measure μ and if so, how to obtain it.

A crucial fact:

Every truncated moment sequence has a representing measure that is a convex combination of at most $s(t) = \binom{d+t}{t}$ many Dirac measures.

Relation with quadrature rules:

Finding measures with a small number of atoms is the equivalent to the problem of finding quadrature rules.

Moment Matrix

Moment matrix: Given $\boldsymbol{y} = \left(y_{\boldsymbol{\gamma}}\right)_{|\boldsymbol{\gamma}| \leq 2t}$, define $\boldsymbol{M}_t(\boldsymbol{y}) \in \mathbb{R}^{s(t) \times s(t)}$ as:

$$\boldsymbol{M}_t(\boldsymbol{y})_{\boldsymbol{\alpha},\boldsymbol{\beta}} := \mathscr{L}_y\left(\boldsymbol{\tau}_t \boldsymbol{\tau}_t^\top\right)_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \mathscr{L}_y\left(\boldsymbol{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}}\right) = y_{\boldsymbol{\alpha}+\boldsymbol{\beta}}$$

Example for $d = t = 2$:

$$\boldsymbol{M}_2(\boldsymbol{y}) = \begin{bmatrix} y_{00} & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{bmatrix}$$

For $\tilde{\boldsymbol{y}} = \left(\tilde{y}_{\boldsymbol{\alpha}}\right)_{\boldsymbol{\alpha} \in \overline{\mathbb{N}}^d}$, we have $\boldsymbol{M}(\tilde{\boldsymbol{y}})_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \tilde{y}_{\boldsymbol{\alpha}+\boldsymbol{\beta}}$ for $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \overline{\mathbb{N}}^d$

Flat extension

Let $\boldsymbol{X}$ be a symmetric matrix with block form

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^\top & \boldsymbol{C} \end{pmatrix}$$

$\boldsymbol{X}$ is called a *flat extension* of $\boldsymbol{A}$ if

$$rank\ \boldsymbol{X} = rank\ \boldsymbol{A}$$

If $\boldsymbol{X}$ is a flat extension of A, then $\boldsymbol{X} \succeq 0 \iff \boldsymbol{A} \succeq 0$.

Flat extension theorem: *For $\boldsymbol{y} = (y_{\boldsymbol{\alpha}})_{|\boldsymbol{\alpha}| \le 2t}$, if $M_t(\boldsymbol{y})$ is a flat extension of $M_{t-1}(\boldsymbol{y})$, then there exists a (unique) sequence $\tilde{\boldsymbol{y}} = (\tilde{y}_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^d}$ for which $M(\tilde{\boldsymbol{y}})$ is flat extension of $M_t(\boldsymbol{y})$.*

Finite Atomic Representing Measures

Theorem 1: *$\tilde{y}$ has a unique representing measure μ which is r-atomic with*

$$supp(\mu) \;=\; \mathcal{V}_{\mathbb{R}}\Big(Ker\, \boldsymbol{M}(\tilde{\boldsymbol{y}})\Big) \;\subseteq\; \mathbb{R}^d$$

$$\Updownarrow$$

$$\boldsymbol{M}(\tilde{\boldsymbol{y}}) \succeq 0 \quad \text{and} \quad rank\, \boldsymbol{M}(\tilde{\boldsymbol{y}}) = r$$

Theorem 2: *y has a unique representing measure μ which is r-atomic with*

$$supp(\mu) \;=\; \mathcal{V}_{\mathbb{R}}\Big(\langle Ker\, \boldsymbol{M}_t(\boldsymbol{y})\rangle\Big) \;\subseteq\; \mathbb{R}^d$$

$$\Updownarrow$$

$$\boldsymbol{M}_t(\boldsymbol{y}) \succeq 0 \quad \text{and} \quad rank\, \boldsymbol{M}_t(\boldsymbol{y}) \;=\; rank\, \boldsymbol{M}_{t-1}(\boldsymbol{y}) \;=\; r$$

Scenario representation:

$$\mu \;=\; \sum_{i=1}^{r} p_i \delta_{\boldsymbol{\xi}_i} \quad \text{where} \quad \boldsymbol{\Xi} \;:=\; \Big\{\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_r\Big\} \;=\; \mathcal{V}_{\mathbb{R}}\Big(\langle Ker\, \boldsymbol{M}_t(\boldsymbol{y})\rangle\Big)$$

Scenarios

Moment matrix representation:

$$\boldsymbol{M}_t(\boldsymbol{y}) = \sum_{i=1}^{r} p_i \boldsymbol{\tau}_t(\boldsymbol{\xi}_i)\boldsymbol{\tau}_t(\boldsymbol{\xi}_i)^\top$$

Vandermonde form:

$$\boldsymbol{M}_t(\boldsymbol{y}) \quad = \quad \boldsymbol{V}_t(\boldsymbol{\Xi}, \boldsymbol{\tau})^\top \boldsymbol{D}\, \boldsymbol{V}_t(\boldsymbol{\Xi}, \boldsymbol{\tau}) \quad \text{with} \quad \boldsymbol{D} \quad := \quad diag(p_1, \cdots, p_r)$$

Vandermonde matrix:

$$\boldsymbol{V}_t(\boldsymbol{\Xi}, \boldsymbol{\tau}) \quad = \quad \left[\boldsymbol{\tau}_t(\boldsymbol{\xi}_1), \cdots, \boldsymbol{\tau}_t(\boldsymbol{\xi}_r)\right]^\top \quad \in \quad \mathbb{R}^{r \times s(t)}$$

Gaussian Quadrature: *Lasserre's algorithm for finding a finite atomic representing measure coincides with that of constructing a quadrature rule with minimal number of nodes, known as Gaussian quadrature.*

Modifications

Flat extension: *Obtained via an SDP that is a trace minimization problem.*

Numerical rank computation: *No guarantee of convergence of the SDP whose size grows exponentially fast, and also necessitates the numerical rank computation of the input moment matrix.*

Pivoted Cholesky decomposition: *Circumvents the computational cost of the usual Cholesky and can be made rank-revealing, also performs Gaussian elimination in a numerically most favorable way.*

Modified Lasserre's Algorithm

Algorithm 3 (extraction algorithm) :

- **Input:** The moment matrix $\boldsymbol{M}_t(\boldsymbol{y})$ with $rank\, \boldsymbol{M}_t(\boldsymbol{y}) = r$
- **Output:** The r nodes $\boldsymbol{\Xi} = \left[\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_r\right]$

1: Perform pivoted Cholesky decomposition to get

$$\boldsymbol{P}^\top \boldsymbol{M}_t(\boldsymbol{y})\boldsymbol{P} = \boldsymbol{P}^\top \boldsymbol{V}_t^\top \boldsymbol{D}\boldsymbol{V}_t\boldsymbol{P} = \widetilde{\boldsymbol{V}}_t^\top \boldsymbol{D}\widetilde{\boldsymbol{V}}_t = \boldsymbol{L}\boldsymbol{L}^\top$$

2: Reduce $\boldsymbol{L}$ to an echelon form $\widetilde{\boldsymbol{L}}$.

3: Extract from $\widetilde{\boldsymbol{L}}$ the multiplication matrices $\boldsymbol{N}_i, i = 1, \cdots, d$.

4: Compute $\boldsymbol{N} := \sum_{i=1}^{d} \rho_i \boldsymbol{N}_i$ with random convex combination.

5: Compute the Schur decomposition $\boldsymbol{N} = \boldsymbol{Q}\boldsymbol{T}\boldsymbol{Q}^\top$ with $\boldsymbol{Q} = \left[\boldsymbol{q}_1 \cdots \boldsymbol{q}_r\right]$.

6: Extract $\boldsymbol{\xi}_j(i) = \boldsymbol{q}_j^\top \boldsymbol{N}_i \boldsymbol{q}_j, i = 1, \cdots, d;\ j = 1, \cdots, r.$

Least-Squares Weights

Algorithm 4 (least-squares weights) :

- **Input:** The moment matrix $\boldsymbol{M}_t(\boldsymbol{y})$ with $rank\, \boldsymbol{M}_t(\boldsymbol{y}) = r$
- **Output:** The r probability weights $\boldsymbol{p} = [p_1, \cdots, p_r]$

1: Compute the generalized inverse (for instance using SVD) $\boldsymbol{V}^\dagger$ of $\boldsymbol{M}_t(\boldsymbol{y})$

2: Compute $\boldsymbol{m}_{\boldsymbol{y}} := diag\Big((\boldsymbol{V}^\dagger)^\top \boldsymbol{M}_t(\boldsymbol{y})\, \boldsymbol{V}^\dagger\Big)$

3: With $\boldsymbol{p} := diag(\boldsymbol{D})$ perform the minimization with :

$$\underset{\boldsymbol{p}\in\mathbb{R}^r_+}{\text{minimize}} \quad \left\|\boldsymbol{m}_{\boldsymbol{y}} - \boldsymbol{p}\right\|_2^2$$

$$\text{subject to:} \quad \mathbf{1}^\top \boldsymbol{p} \;=\; 1$$

EMPIRICAL PROBLEM

Empirical Truncated Moment Problem

Training sample (drawn from μ):

$$\boldsymbol{\mathcal{X}} = \left\{ \tilde{\boldsymbol{x}}_1, \cdots, \tilde{\boldsymbol{x}}_N \right\} \subset \boldsymbol{\Omega} \subseteq \mathbb{R}^d \text{ with } \boldsymbol{\Omega} \text{ Hausdorff and locally compact}$$

Empirical probability measure:

$$\widehat{\mu} = \frac{1}{N} \sum_{i=1}^{N} \delta_{\tilde{\boldsymbol{x}}_i}$$

Empirical moment sequence:

$$\widehat{\boldsymbol{y}}_{\boldsymbol{\alpha}} = \int_{\boldsymbol{\Omega}} \boldsymbol{x}^{\boldsymbol{\alpha}} \, d\widehat{\mu}(\boldsymbol{x}) \text{ where } |\boldsymbol{\alpha}| \leq 2t$$

Empirical moment matrix:

$$\widehat{\boldsymbol{M}}_t = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\tau}_t(\tilde{\boldsymbol{x}}_i) \, \boldsymbol{\tau}_t(\tilde{\boldsymbol{x}}_i)^\top \in \mathbb{R}^{s(t) \times s(t)}$$

Measure Compression

Target compressed measure:

$$\widetilde{\mu} \;=\; \sum_{i=1}^{r} p_i\, \delta_{\boldsymbol{\xi}_i} \;\text{ where }\; \boldsymbol{\Xi} \;:=\; \left\{ \boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_r \right\} \subset \boldsymbol{\mathcal{X}} \;\text{ with }\; r \ll N$$

Model moment matrix:

$$\widetilde{\boldsymbol{M}}_t \;=\; \sum_{i=1}^{r} p_i\, \boldsymbol{\tau}_t(\boldsymbol{\xi}_i)\, \boldsymbol{\tau}_t(\boldsymbol{\xi}_i)^\top \;=\; \boldsymbol{V}_t(\boldsymbol{\Xi}, \boldsymbol{\tau})^\top\, diag(\boldsymbol{p})\, \boldsymbol{V}_t(\boldsymbol{\Xi}, \boldsymbol{\tau}) \;\in\; \mathbb{R}^{s(t)\times s(t)}$$

Optimization Problem:

$$\underset{\substack{\boldsymbol{\xi}_i \in \mathbb{R}^d,\; i=1,\cdots,r \\ \boldsymbol{p}\in\mathbb{R}^r}}{argmin} \left\| \widehat{\boldsymbol{M}}_t \;-\; \boldsymbol{V}_t(\boldsymbol{\Xi}, \boldsymbol{\tau})^\top\, diag(p)\; \boldsymbol{V}_t(\boldsymbol{\Xi}, \boldsymbol{\tau}) \right\|_F \tag{1}$$

subject to: $1^\top \boldsymbol{p} \;=\; 1$

Framework Proposal

Observations

- (1) is a non-convex optimization problem in general (except when $t = 1$) in the nodes $\boldsymbol{\xi}_i$ and the probability weights p_i for $i = 1, \cdots, r$.
- The problem of matching the empirical moments with the model moments is in general, under-determined since we have $r \leq s(t) \ll N$.
- The Vandermonde representation of the positive semi-definite $\widetilde{\boldsymbol{M}}_t$ pertaining to the minimal generating measure is precisely in the optimal form as in the model proposed by Bach, Rudi et. al. within the RKHS framework.
- We develop a relaxation of the optimization problem (1) and reformulate it within the RKHS framework, which makes it convex.

METHODOLOGY

Reproducing Kernel Hilbert Space (RKHS)

Let $(\boldsymbol{\mathcal{H}}, \langle \cdot, \cdot \rangle_{\mathcal{H}}) \subseteq \mathbb{R}^{\boldsymbol{\Omega}}$ be a separable Hilbert space of functions with

$$\boldsymbol{\Omega} \subseteq \mathbb{R}^d.$$

Then $\exists$ a unique reproducing kernel $k : \boldsymbol{\Omega} \times \boldsymbol{\Omega} \longrightarrow \mathbb{R}$ such that:

$$\forall\, \boldsymbol{x} \in \boldsymbol{\Omega}, \quad k_{\boldsymbol{x}} := k(\boldsymbol{x}, \cdot) \in \boldsymbol{\mathcal{H}}$$

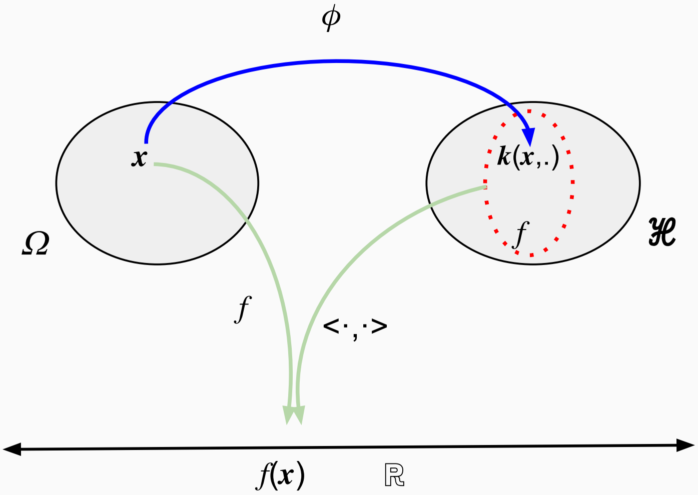
$$\forall\, f \in \boldsymbol{\mathcal{H}}, \quad f(\boldsymbol{x}) = \langle f, k_{\boldsymbol{x}} \rangle_{\mathcal{H}} \;\; \forall\; \boldsymbol{x} \in \boldsymbol{\Omega}$$

k is a *symmetric and positive definite kernel* i.e. for any finite

$$\boldsymbol{\mathcal{X}} = \{\tilde{\boldsymbol{x}}_1, \cdots, \tilde{\boldsymbol{x}}_N\} \subset \boldsymbol{\Omega}, \quad \boldsymbol{K} := \Big[k\big(\tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{x}}_j\big)\Big]_{i,j=1}^{N} \in \mathcal{S}_+^N$$

Corollary: $\forall\; \boldsymbol{x},\; \tilde{\boldsymbol{x}} \in \boldsymbol{\Omega}, \;\; \langle k_{\boldsymbol{x}}, k_{\tilde{\boldsymbol{x}}} \rangle_{\mathcal{H}} = \langle \phi(\boldsymbol{x}), \phi(\tilde{\boldsymbol{x}}) \rangle_{\mathcal{H}}$

Intuition



ϕ
$\boldsymbol{x}$
$k(\boldsymbol{x},.)$
Ω
f
$\mathscr{H}$
f
$<\cdot,\cdot>$
$f(\boldsymbol{x})$
$\mathbb{R}$

Framework

Model:

$$f_A(\boldsymbol{x}) \;=\; \phi(\boldsymbol{x})^\top \boldsymbol{A}\,\phi(\boldsymbol{x}), \qquad \boldsymbol{A} \;\in\; \mathcal{S}_+(\boldsymbol{\mathcal{H}}) \tag{2}$$

Objective Function:

$$\inf_{\boldsymbol{A}\in\mathcal{S}_+(\boldsymbol{\mathcal{H}})} \boldsymbol{L}\Big(f_A(\boldsymbol{\xi}_1),\cdots,f_A(\boldsymbol{\xi}_r)\Big) + \underbrace{\lambda_1||\boldsymbol{A}||_\star + \lambda_2||\boldsymbol{A}||_F^2}_{\widetilde{\Omega}(\boldsymbol{A})}, \qquad \lambda_2 > 0 \tag{3}$$

Representer Theorem: *Let $\boldsymbol{L}$ be lower semi-continuous, bounded below and convex, and $\widetilde{\Omega}(\boldsymbol{A})$ be as above. Then (3) has a unique minimizer*

$$\boldsymbol{A}_* \;=\; \sum_{i=1}^{r}\sum_{j=1}^{r} \boldsymbol{B}_{ij}\,\phi(\boldsymbol{\xi}_i)\,\phi(\boldsymbol{\xi}_j)^\top \qquad \boldsymbol{B}\in\mathbb{R}^{r\times r},\;\; \boldsymbol{B}\succeq 0 \tag{4}$$

Setup

RKHS:

$$(\boldsymbol{\mathcal{H}}, \langle \cdot, \cdot \rangle_{\mathcal{H}}) \;:=\; span\{\boldsymbol{x}^{\boldsymbol{\alpha}} : \boldsymbol{x} \in \boldsymbol{\Omega}, |\boldsymbol{\alpha}| \leq t\} \;=\; \boldsymbol{\mathcal{P}}_t(\boldsymbol{\Omega})$$

with $L^2_\mu(\boldsymbol{\Omega})$ inner product $\langle f, g \rangle_{\mathcal{H}} \;=\; \int_{\boldsymbol{\Omega}} f(\boldsymbol{x})\, g(\boldsymbol{x})\, d\mu(\boldsymbol{x})$

Monomial basis:

$$\boldsymbol{\tau}_t(\boldsymbol{x}) \;:=\; \left[1, x_1, \cdots, x_d, \cdots, x_1^t, \cdots, x_d^t\right]^\top \;\in\; \mathbb{R}^{s(t)}$$

Gram matrix:

$$\left[\boldsymbol{G}_t\right]_{\boldsymbol{\alpha},\boldsymbol{\beta}} \;=\; \langle \boldsymbol{\tau}_t, \boldsymbol{\tau}_t^\top \rangle_{\mathcal{H}} \;=\; \int_{\boldsymbol{\Omega}} \boldsymbol{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}}\, d\mu(\boldsymbol{x}) \qquad |\boldsymbol{\alpha}|\,,\, |\boldsymbol{\beta}| \;\leq t$$

Reproducing Kernel:

$$k_t(\boldsymbol{x}, \tilde{\boldsymbol{x}}) \;=\; \boldsymbol{\tau}_t(\boldsymbol{x})^\top\, \boldsymbol{G}_t^\dagger\, \boldsymbol{\tau}_t(\tilde{\boldsymbol{x}}) \qquad \forall\;\; \boldsymbol{x},\;\; \tilde{\boldsymbol{x}} \in \boldsymbol{\Omega}$$

Numerics

Empirical Gram matrix:

$$\left[\widehat{\boldsymbol{G}}_t\right]_{\boldsymbol{\alpha},\boldsymbol{\beta}} := \int_{\boldsymbol{\Omega}} \tilde{\boldsymbol{x}}^{\boldsymbol{\alpha}+\boldsymbol{\beta}}\, d\hat{\mu}(\boldsymbol{x}) = \frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{x}}_i^{\boldsymbol{\alpha}+\boldsymbol{\beta}} \in \mathbb{R}^{s(t)\times s(t)}$$

Discrete orthonormal basis:

$$\boldsymbol{\psi}_t(\boldsymbol{x}) := \left(\widehat{\boldsymbol{G}}_t^{\dagger}\right)^{1/2} \boldsymbol{\tau}_t(\boldsymbol{x})$$

Kernel matrix:

$$\begin{aligned}\boldsymbol{K}_t = \left[k_t(\tilde{\boldsymbol{x}}_i,\, \tilde{\boldsymbol{x}}_j)\right]_{i,j=1}^{N} &= \boldsymbol{V}_t(\boldsymbol{\mathcal{X}},\boldsymbol{\tau})\; G_t^{\dagger}\; \boldsymbol{V}_t(\boldsymbol{\mathcal{X}},\boldsymbol{\tau})^{\top} \\ &= \boldsymbol{Q}_t(\boldsymbol{\mathcal{X}},\boldsymbol{\tau})\, \boldsymbol{Q}_t(\boldsymbol{\mathcal{X}},\boldsymbol{\tau})^{\top}\end{aligned}$$

$$\boldsymbol{Q}_t(\boldsymbol{\mathcal{X}},\boldsymbol{\tau}) = \left[\boldsymbol{\psi}_t(\tilde{\boldsymbol{x}}_1), \cdots, \boldsymbol{\psi}_t(\tilde{\boldsymbol{x}}_N)\right]^{\top} \in \mathbb{R}^{N\times s(t)}$$

Low-rank approximation

- We would like to extract a suitable subsample of size $r \ll N$ from the original sample such that they correspond to the optimal nodes.
- Computing the spectral decomposition of $\boldsymbol{K}_t$ can be severely prohibitive as the computational cost is $\mathcal{O}(N^3)$.
- Hence, we will use the diagonally pivoted Cholesky decomposition to select r columns which span the dominant subspace generated by the corresponding kernel functions.
- The said algorithm resorts to a greedy strategy that reduces the trace of the kernel matrix in an iterative manner.
- The computational cost of the pivoted Cholesky is $\mathcal{O}(r^2 N)$.

Pivoted Cholesky Decomposition

Algorithm 2 (pivoted Cholesky decomposition) :

- **input:** symmetric and positive semi-definite $\boldsymbol{M} \in \mathbb{R}^{s \times s}$, tolerance $\varepsilon \geq 0$
- **output:** low-rank approximation $\boldsymbol{M} \approx \boldsymbol{L}\boldsymbol{L}^\top$

initialization: set $m := 1$, $\boldsymbol{d} := \operatorname{diag}(\boldsymbol{M})$, $\boldsymbol{L} := [\,]$, $\operatorname{err} := \|\boldsymbol{d}\|_1$
while $\operatorname{err} > \varepsilon$
    determine $j := \arg\max_{1 \leq i \leq s} d_i$
    compute $\hat{\boldsymbol{\ell}}_m := \boldsymbol{M}(:, j) - \boldsymbol{L} * \boldsymbol{L}^\top(:, j)$
    set $\boldsymbol{\ell}_m := \hat{\boldsymbol{\ell}}_m / \sqrt{d_j}$
    set $\boldsymbol{L} := [\boldsymbol{L}, \boldsymbol{\ell}_m]$
    set $\boldsymbol{d} := \boldsymbol{d} - \boldsymbol{\ell}_m \odot \boldsymbol{\ell}_m$
    set $\operatorname{err} := \|\boldsymbol{d}\|_1$
    set $m := m + 1$

Fast Empirical Scenarios

Algorithm 5 (fast empirical scenarios) :

- **Input:** The N samples $\boldsymbol{\mathcal{X}} = \{\tilde{\boldsymbol{x}}_1, \cdots, \tilde{\boldsymbol{x}}_N\}$
- **Output:** The r nodes $\boldsymbol{\Xi} = [\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_r]$ and probability weights $\boldsymbol{p} = [p_1, \cdots, p_r]$

1: Compute the empirical moment matrix $\widehat{\boldsymbol{M}}_t$ and empirical kernel matrix $\boldsymbol{K}_t$

2: Perform the pivoted Cholesky decomposition on $\boldsymbol{K}_t$ to obtain the r nodes $\boldsymbol{\Xi} = [\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_r]$ that generate the dominant subspace

3: Perform the following optimization problem:

$$\underset{\boldsymbol{p}\in\mathbb{R}_+^r}{argmin} \left\| \widehat{\boldsymbol{M}}_t - \boldsymbol{V}_t(\boldsymbol{\Xi}, \boldsymbol{\tau})^\top \, diag(p) \; \boldsymbol{V}_t(\boldsymbol{\Xi}, \boldsymbol{\tau}) \right\|_F \tag{5}$$

$$\text{subject to: } \mathbf{1}^\top \boldsymbol{p} = 1$$

to obtain the probabilities $\boldsymbol{p} = [p_1, \cdots p_r]$

Covariance Scenarios

$$M_1(\boldsymbol{y}) \;=\; \boldsymbol{L}\,\boldsymbol{L}^\top.$$

Let $\boldsymbol{H}_v$ be the Householder reflector, then we have the Vandermonde form:

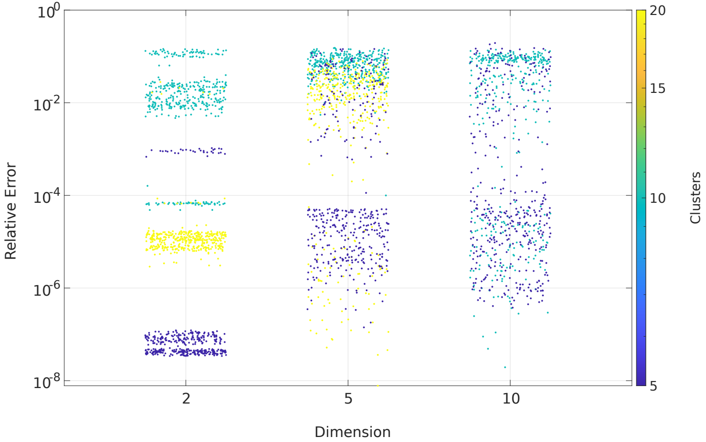
$$\boldsymbol{M}_1(\boldsymbol{y}) \;=\; \boldsymbol{L}\,\boldsymbol{H}_v^\top\,\boldsymbol{H}_v\,\boldsymbol{L}^\top \;=\; \boldsymbol{V}\,\boldsymbol{D}\,\boldsymbol{V}^\top$$

When $t = 1$:

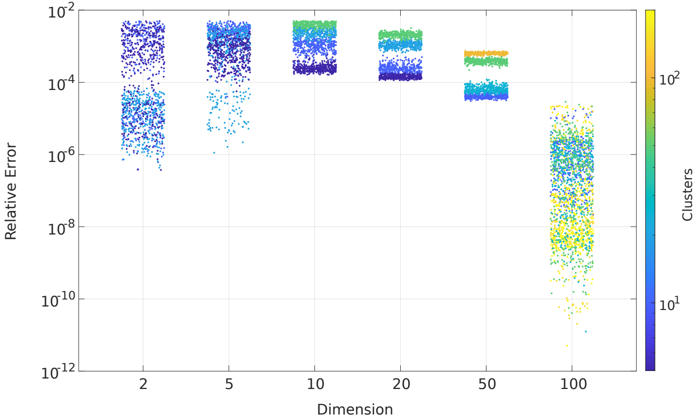
The optimization problem (1) is convex as the Vandermonde matrix is linear in the probability weights and directly solves the interpolation problem

FIGURES

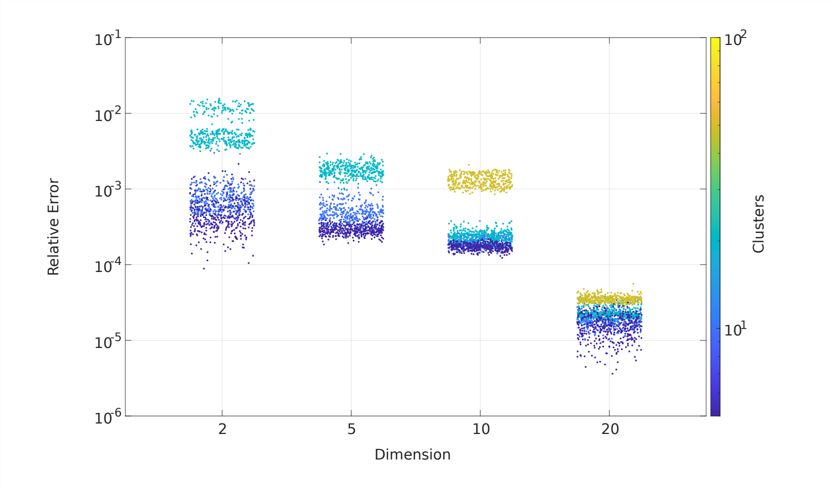
Lasserre's Gaussian Quadrature



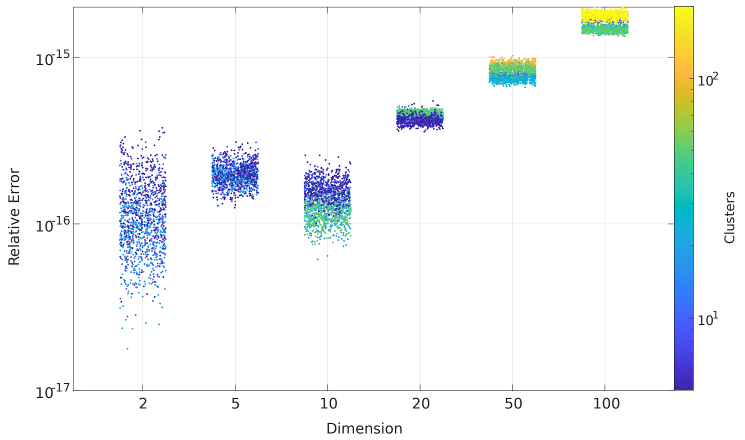
RKHS Scenarios (order 1)



RKHS Scenarios (order 2)

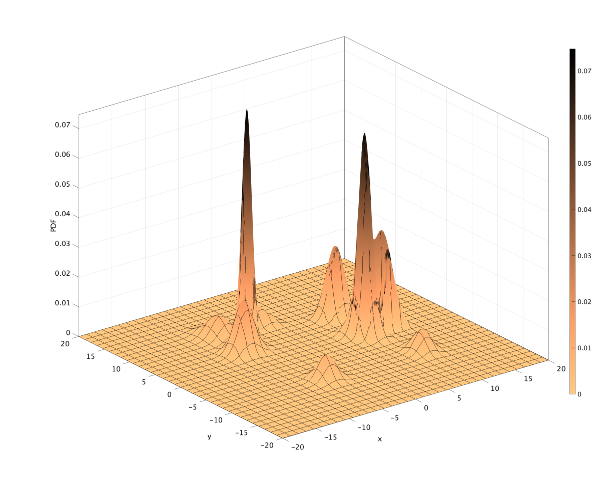


Covariance Scenarios

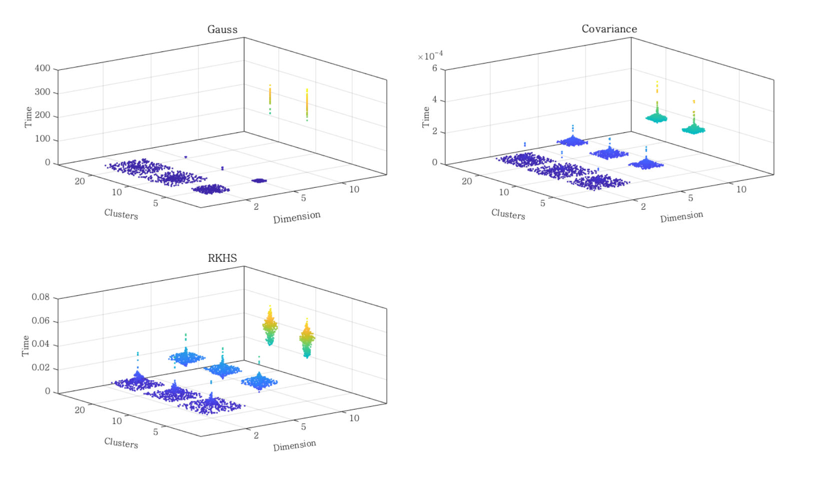


Sample PDF

Dimension $= 2$, Clusters $= 10$



Computation Times



Acknowledgement

I would like to thank the

Swiss National Science Foundation

for the grant of this project.

THANK YOU!

References

- A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces In Probability and Statistics*.
- H. Herbrecht, M. Peters, and R. Schneider, *On the low-rank approximation by the pivoted cholesky decomposition*, (2011).
- J. B. Lasserre, *Moments, Positive Polynomials and Their Applications*.
- U. Marteau-Ferey, F. Bach, and A. Rudi, *Non-parametric models for non-negative functions*, (2020).