Approximate Quadrature Measures on Data–Defined Spaces

H. N. Mhaskar

Claremont Graduate University

A very general quadrature theorem

(Tchakaloff, 1957)

Let $\mathbb{X}$ be a compact topological space, $\{\phi_j\}_{j=0}^{N-1}$ be continuous real valued functions on $\mathbb{X}$, $\phi_0(x) \equiv 1$, and μ^* be a probability measure on $\mathbb{X}$ (i.e., μ^* is a positive Borel measure with $\mu^*(\mathbb{X}) = 1$). Then there exist N points $x_1, \cdots, x_N$, and non–negative numbers $w_1, \cdots, w_N$ such that

$$\sum_{k=1}^{N} w_k \phi_j(x_k) = \int_{\mathbb{X}} \phi_j(x) d\mu^*(x), \quad j = 0, \cdots, N-1.$$

Exact quadrature

A consequence

Let $V_N = \text{span}\{\phi_0, \cdots, \phi_{N-1}\}$. If $f \in C(\mathbb{X})$ then

$$\left|\int_{\mathbb{X}} f(x)d\mu^*(x) - \sum_{k=1}^{N} w_k f(x_k)\right| \le 2 \min_{P \in V_N} \max_{x \in \mathbb{X}} |f(x) - P(x)|.$$

Approximate quadrature

Important variations

- The weights $w_k = 1/N$, $k = 0, \cdots, N-1$; wish to find points x_k (e.g., Quasi-Monte Carlo (QMC) designs).
- The points x_k are given (e.g., machine learning); wish to find the weights w_k.
- In either case, look for approximate quadrature rather than exact quadrature.

Goal of the current paper: Develop a unified theory.

A prototype : QMC Design sequence

Brauchart-Dick-Saff-Sloan-Wang-Womersley, 2015,
Brauchart-Saff-Sloan-Womersley, 2014
$q \geq 2$ integer, $\mathbb{S}^q$: unit sphere of $\mathbb{R}^{q+1}$, Δ^*: Laplace-Beltrami operator. For $r > 0$, $1 \leq p \leq \infty$,

$$W_r^p = \{f \in L^p(\mathbb{S}^q) : \|f\|_{W_r^p} = \|(1-\Delta^*)^{r/2} f\|_{p,\mathbb{S}^q} < \infty\}.$$

For integer $n \geq 1$, let $\mathcal{C}_n$ be a set of n points on $\mathbb{S}^q$. The sequence $\{\mathcal{C}_n\}$ is a QMC design sequence for W_r^p if

$$\left| \int_{\mathbb{S}^q} f(\mathbf{x}) d\mathbf{x} - \frac{1}{n} \sum_{\mathbf{y} \in \mathcal{C}_n} f(\mathbf{y}) \right| \leq \frac{c}{n^{r/q}} \|f\|_{W_r^p}, \quad f \in W_r^p.$$

A prototype : QMC Design sequence

- Every sequence of n-designs is a QMC design sequence for every W_r^p, $r > q/p$.
- If $\{\mathcal{C}_n\}$ is a QMC design sequence for W_r^1, $r > q$, then

$$\delta_{\mathcal{C}_n} = \max_{\mathbf{x}\in\mathbb{S}^q} \min_{\mathbf{y}\in\mathcal{C}_n} d(\mathbf{x}, \mathbf{y}) \sim n^{-1/q}.$$

(cf. Hesse-Mh.-Sloan, 2007)

- Let $K(t) = \sum_{\ell=1}^{\infty} \hat{K}(\ell) p_\ell^{(q/2-1,q/2-1)}(1) p_\ell^{(q/2-1,q/2-1)}(t)$, $\hat{K}(\ell) \sim (1+\ell)^{-2r}$, and $r > d/2$. If

$$\mathcal{C}_n^* = \{\mathbf{x}_1^*, \cdots, \mathbf{x}_n^*\} = \arg\min \sum_{j,k=1}^{n} K(\mathbf{x}_j \cdot \mathbf{x}_k),$$

then $\{\mathcal{C}_n^*\}$ is a QMC design sequence of W_r^2.

Machine learning problem

Given data of the form $\{(\mathbf{x}_j, y_j)\}_{j=1}^M$ drawn from an unknown probability distribution μ on $\mathbb{R}^D \times \mathbb{R}$, approximate $f(\mathbf{x}) = \mathbb{E}(y|\mathbf{x})$.

Main hurdles:

- D can be very large. (remedy: Try dimension-reduction.)
- The measure μ is not known; in particular, the set $\{\mathbf{x}_j\}$ is typically not dense enough on some cube in $\mathbb{R}^D$. ((Theoretical) remedy: Assume that the marginal distribution μ^* of $\mathbf{x}$ is known; at least assume something about μ^*.)

Diffusion geometry

Belkin-Niyogi, 2003, Coifman group, 2006–
Let μ^* be a probability measure supported on a locally compact, quasi-metric measure space $\mathbb{X}$, (ρ=quasi-metric), $\lambda_k \uparrow \infty$, $\{\phi_k\}$ sequence of orthonormal, continuous functions on $\mathbb{X}$. $\lambda_0 = 0$, $\phi_0 \equiv 1$.
Take the projection on

$$\Pi_n = \text{span}\{\phi_k : \lambda_k < n\},$$

$$\frac{1}{M} \sum_{k:\lambda_k<n} y_j \phi_k(\mathbf{x}_j),$$

as the estimator of $f(\mathbf{x}_j)$.

Diffusion geometry

If $\mathbb{X}$ is a compact manifold, one may take the eigenvectors and eigenvalues of a diffusion matrix as approximations of eigenfunctions and eigenvalues of the Laplace-Beltrami operator on $\mathbb{X}$. Belkin-Niyogi, 2004, 2007, 2008, Lafon, 2004, Singer, 2006, Rosasco-Belkin-De Vito, 2010, $\cdots$

Diffusion geometry

Problem: Develop approximation theory in this context.

Data-defined spaces

The system $\Xi = (\mathbb{X}, \rho, \mu^*, \{\lambda_k\}_{k=0}^\infty, \{\phi_k\}_{k=0}^\infty)$ is called a data-defined space (DDS) if

- For each $x \in \mathbb{X}$ and $r > 0$, the ball $\mathbb{B}(x, r)$ is compact.
- There exists $q > 0$ and $\kappa_2 > 0$ such that

$$\mu^*(\mathbb{B}(x,r)) = \mu^*\left(\{y \in \mathbb{X} : \rho(x,y) < r\}\right) \le \kappa_2 r^q, \quad x \in \mathbb{X},\ r > 0.$$

- For $x, y \in \mathbb{X}$, $0 < t \le 1$,

$$\left|\sum_{k=0}^\infty \exp(-\lambda_k^2 t)\phi_k(x)\phi_k(y)\right| \le \kappa_3 t^{-q/2} \exp\left(-\kappa_4 \frac{\rho(x,y)^2}{t}\right)$$

Remark $\mathbb{S}^q$ is DDS, with $\{\lambda_k\}$, $\{\phi_k\}$ eigensystem of the Laplace-Beltrami operator.

Approximation on DDS

$$
\begin{aligned}
\Pi_n &= \text{span}\{\phi_k : \lambda_k < n\}, \quad (\text{diffusion polynomials})\\
E_{n,p}(f) &= \min_{P\in\Pi_n} \|f-P\|_p = \min_{P\in\Pi_n} \|f-P\|_{\mu^*;p,\mathbb{X}},\\
\hat{f}(k) &= \int_{\mathbb{X}} f(y)\overline{\phi_k(y)}d\mu^*(y),\\
\Delta^r f &\sim \sum_k (\lambda_k+1)^r \hat{f}(k)\phi_k,\\
W_r^p &= \{f\in L^p : \|f\|_{W_r^p} = \|f\|_p + \|\mathcal{D}^r f\|_p < \infty\},\\
\omega_s(p;f,\delta) &= \inf_{f_1\in W_s^p}\{\|f-f_1\|_p + \delta^s\|\Delta^s f_1\|_p\},\\
H_\gamma^p &= \left\{f\in L^p : \|f\|_{H_\gamma^p} = \|f\|_p + \sup_{0<\delta<1}\frac{\omega_s(f,\delta)}{\delta^\gamma} < \infty\right\},
\end{aligned}
$$

$$\text{some } s > \gamma.$$

Approximation on DDS

Maggioni-Mh., 2008, Filbir-Mh., 2010
Let $X^p = \{f \in L^p : \lim_{n\to\infty} E_{n,p}(f) = 0\}$.
For $f \in X^p$

$$\|f\|_{H^p_\gamma} \sim \|f\|_p + \sup_{n\geq 1} n^\gamma E_{n,p}(f).$$

Remark $W^p_r \subsetneq H^p_r$ for all $r > 0$.

Approximate quadrature measures

Goal

For $W \subset L^p$,

$$\mathrm{wor}(W,\nu) = \sup_{f\in W}\left|\int_{\mathbb{X}} fd\mu^* - \int_{\mathbb{X}} fd\nu\right|.$$

ν is typically supported on a finite subset of $\mathbb{X}$.

$$B(\gamma,p) = \{f \in H^p_\gamma : \|f\|_{H^p_\gamma} = 1\}.$$

- Allow ν to be a signed measure.
- Study measures ν_n for which $\mathrm{wor}(B(\gamma,p),\nu_n) = \mathcal{O}(n^{-\gamma})$.

Approximate quadrature measures

Terminology

$\mathcal{M}$=set of all signed (or positive), complete, sigma finite, Borel measures on $\mathbb{X}$.

$|\nu|$=Total variation measure for $\nu \in \mathcal{M}$.

ν is *d*-regular if

$$\|\nu\|_{\mathcal{R}(d)} = \sup_{x \in \mathbb{X}} d^{-q} |\nu|(\mathcal{B}(x, d)) < \infty.$$

ν is quadrature measure of order *n* if

$$\int_{\mathbb{X}} P d\mu^* = \int_{\mathbb{X}} P d\nu, \quad P \in \Pi_n.$$

Approximate quadrature measures

Terminology

A sequence $\aleph = \{\nu_n\} \subset \mathcal{M}$ is a sequence of approximate quadrature measures ($\aleph \in \mathcal{A}(\gamma, p)$) if

- $\sup_{n\geq 1} |\nu_n|(\mathbb{X}) < \infty$.
- $\sup_{x\in\mathbb{X}, n\geq 1} n^q |\nu_n|(\mathbb{B}(x, 1/n)) = \sup_{n\geq 1} \|\nu_n\|_{\mathcal{R}(1/n)} < \infty$.
- $$\sup_{n\geq 1} n^\gamma \text{wor}(B(\gamma, p)\cap\Pi_n, \nu_n) < \infty.$$

A measure ν is an approximate quadrature measure in $\mathcal{A}(\gamma, p, n)$ if it is the n-th term in a sequence of approximate quadrature measures.

Error in quadrature

Let $n \geq 1$, $1 \leq p \leq \infty$, $\gamma > q/p$, ν be a $1/n$-regular measure satisfying $|\nu|(\mathbb{X}) < \infty$ and

$$\mathrm{wor}(B(\gamma, p) \cap \Pi_n, \nu) \leq A n^{-\gamma}.$$

Then for every $f \in H^p_\gamma$,

$$\left| \int_{\mathbb{X}} f d\mu^* - \int_{\mathbb{X}} f d\nu \right| \leq c \left(A + |||\nu|||_{R(1/n)}^{1/p} (|\nu|(\mathbb{X}))^{1/p'} \right) \frac{\|f\|_{H^p_\gamma}}{n^\gamma}.$$

Remarks

- If ν is a quadrature measure of order n, $\|\nu\|_{\mathcal{R}(1/n)} < \infty$, $|\nu|(\mathbb{X}) < \infty$ then $\mathrm{wor}(B(\gamma,p),\nu) = \mathcal{O}(n^{-\gamma})$. (In particular, this holds if ν is a spherical n-design.)
- Under additional conditions, any positive quadrature measure of order n satisfies $|\nu|(\mathbb{X}) < \infty$, and $\|\nu\|_{\mathcal{R}(1/n)} < \infty$, and hence, $\mathrm{wor}(B(\gamma,p),\nu) = \mathcal{O}(n^{-\gamma})$.
- If $\mathbb{X}$ is a compact manifold, and $\{\lambda_k\}$, $\{\phi_k\}$ satisfy some technical conditions then for any finite $\mathcal{C} \subset \mathbb{X}$ with

$$\delta(\mathcal{C}) = \sup_{x\in\mathbb{X}} \min_{y\in\mathcal{C}} \rho(x,y) \le c/n,$$

  there is a positive (respectively, $1/n$-regular) quadrature measure of order n supported on $\mathcal{C}$. Filbir-Mh., 2010

Density

Let $\gamma > 0$ and $\aleph \in \mathcal{A}(\gamma, 1)$. Then there exist constants $c, C_1 > 0$ such that for $n \geq 1$,

$$|\nu|(\mathbb{B}(x, C_1/n)) \geq c/n^q, \qquad x \in \mathbb{X}.$$

In particular, $\delta(\mathrm{supp}(\nu)) \sim 1/n$.

Sparsity

Let $\gamma > 0$, $n > 0$, $1 \le p \le \infty$, and ν be a positive measure satisfying

$$\mathrm{wor}(B(\gamma, p) \cap \Pi_n, \nu) \le An^{-\gamma}.$$

Then (under some additional assumptions)

$$\nu(\mathbb{B}(x, 1/n)) \le cn^{-q/p}, \quad x \in \mathbb{X}.$$

In particular, if $p = 1$, then $\nu \in \mathcal{A}(\gamma, 1, n)$.

Extremal problem

$G(x,y) \sim \sum_j b(\lambda_j)\phi_j(x)\overline{\phi_j(y)}$ is kernel of type β if $b(\lambda_j) \approx \lambda_j^{-\beta}$ in some technical sense. $\tilde{G}(x,y) \sim \sum_{j>0} b(\lambda_j)\phi_j(x)\overline{\phi_j(y)}$.
For $\nu \in \mathcal{M}$, $p' = p/(p-1)$,

$$
\begin{aligned}
M_p(\nu) &= \left\| \int_{\mathbb{X}} G(x,\circ)d\mu^*(x) - \int_{\mathbb{X}} G(x,\circ)d\nu(x) \right\|_{p'} \\
&= \left\| \int_{\mathbb{X}} \tilde{G}(x,\circ)d\nu(\mathbf{x}) \right\|_{p'}.
\end{aligned}
$$

$K_n \subset \mathcal{M}$ is compact, $\sup_{\nu\in K_n} |\nu|(\mathbb{X}) < \infty$, $\sup_{\nu\in K_n} \|\nu\|_{\mathcal{R}(2^{-n})} < \infty$, and K_n contains a quadrature measure of order 2^n.

Extremal problem

Let $n \geq 0$, $\beta > q/p$, $0 < \gamma < \beta$, $\nu^\# \in K_n$, and

$$M_p(\nu^\#) \leq c \inf_{\nu \in K_n} M_p(\nu).$$

Then $\nu^\#$ is an approximate quadrature measure in $\mathcal{A}(\gamma, p, 2^n)$.
Remark $\nu^\#$ need not be an exact minimizer of $M_p(\nu)$.

Consequences

If $p = 2$, $G^*(x, y) = \sum_{j>0} b(\lambda_j)^2 \phi_j(x)\overline{\phi_j(y)}$,

$$M_2(\nu) = \int_{\mathbb{X}} \int_{\mathbb{X}} G^*(x, y) d\nu(x) d\nu(y).$$

If ν associates the mass $1/M$ with each element of a set $\{x_j\}_{j=1}^M$, then

$$M_2(\nu) = \frac{1}{M^2} \sum_{j=1}^{M} \sum_{k=1}^{M} G^*(x_j, x_k).$$

If K_n is the set of all such measures, and K_n contains a quadrature measure ν of order 2^n, then the minimizer of $M_2(\nu)$ is an approximate quadrature measure.
Generalized the result of Brautchart, Saff, Sloan, Womersley, 2014.

Consequences

Let $\mathcal{C}$ be a fixed set of points with $\delta(\mathcal{C}) \leq c/n$ for sufficient small c so that there exists a positive quadrature measure supported on $\mathcal{C}$. We may choose K_n to be the set of all probability measures supported on $\mathcal{C}$. Only approximate minimization is required.

Consequences

If $\mathbb{X}$ is compact, the dimension of Π_{2^n} is M, and K_n denotes the set of all probability measures supported on M-element subsets of $\mathbb{X}$, then we get a construction for an approximate version of Tchakaloff's theorem.

Key ingredients

Let $h : [0, \infty) \to [0, 1]$ be C^∞, $h(t) = 1$ on $[0, 1/2]$, $h(t) = 0$ on $[1, \infty)$, and

$$\Phi_n(x, y) = \sum_k h\left(\frac{\lambda_k}{n}\right) \phi_k(x)\phi_k(y).$$

Then (Maggioni-Mh., 2008, Filbir-Mh. 2010) for $S > q$,

$$|\Phi_n(x, y)| \leq \frac{cn^q}{\max(1, (n\rho(x, y))^S)}$$

Key ingredients

$$\sigma_n(f,x) = \int_{\mathbb{X}} \Phi_n(x,y)f(y)d\mu^*(y),$$

$$\tau_j(f,x) = \begin{cases} \sigma_1(f,x) & \text{if } j=0, \\ \sigma_{2^j}(f,x) - \sigma_{2^{j-1}}(f,x), & \text{if } j = 1,2,3,\cdots \end{cases}$$

If $f \in X^p$, then

$$f = \sum_{j=0}^{\infty} \tau_j(f)$$

and

$$\|f\|_{H^p_\gamma} \sim \|f\|_p + \sup_{j\ge 0} 2^{j\gamma}\|\tau_j(f)\|_p.$$

$$\|\tau_j(f)\|_{\nu;1} \le (1+(2^j d)^q)^{1/p}\|\nu\|_{R(d)}^{1/p}(|\nu|(\mathbb{X}))^{1/p'}\|\tau_j(f)\|_p.$$

Key ingredients

Product assumption (needed only for the sparsity statement) For $A, N > 0$, let

$$\epsilon_{A,N} := \sup_{\lambda_j, \lambda_k \leq N} \mathrm{dist}(\infty; \phi_j \phi_k, \Pi_{AN}).$$

We assume that there exists $A^* \geq 2$ with the following property: for every $R > 0$,

$$\lim_{N \to \infty} N^R \epsilon_{A^*,N} = 0.$$