

Information based complexity in stochastic optimization

Teemu Pennanen
Department of Mathematics
King's College London

Stochastic optimization

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

Let $(\Omega, \mathcal{F}, P)$ be a probability space with $\Omega \subset \mathbb{R}^d$ and consider the **stochastic optimization** problem

$$\text{minimize } E^P f(x, \omega) \quad \text{over } x \in X, \quad (\text{SP})$$

where $X \subset \mathbb{R}^n$ is compact and $f : \mathbb{R}^n \times \Omega \rightarrow \mathbb{R}$ is continuous in x and integrable in ω .

- Fundamental problem in risk management, finance, statistics, machine learning, ...
- In decision making, P is often **infinitely** supported.
- Our aim is to find sub-classes of (SP) whose **information based complexity** depends polynomially on the dimension n of the decision variable x and the dimension d of the random variable ω .

Information based complexity

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

- **Information based complexity** (IBC) studies the solution accuracy obtainable with a given number of function/gradient evaluations.
- Many notions of IBC: worst case, randomized, average, ...
- Well-studied problems: integration, linear equations, optimization, ...
- IBC of (SP) in the **randomized** setting (stochastic gradient descent) has been extensively studied by e.g. Nemirovski, Juditsky, Lan, ...
- We will derive an upper-bound on the IBC of (SP) in the **worst case** setting by combining existing complexity results for **numerical integration** and **convex optimization**.

IBC of convex optimization

Convex optimization

Numerical integration

Stochastic optimization

Dynamic SP

Galerkin method

An application

“Approximate” solutions of **convex** optimization problems

$$\text{minimize } F(x) \quad \text{over } x \in X$$

can be constructed by “adaptive” algorithms where each iteration requires the evaluation of a **subgradient** of F and a **projection** on the set X .

- An $x \in X$ is ε -**optimal** if

$$F(x) \leq \inf\{F(z) \mid z \in X\} + \varepsilon.$$

- A vector $v \in \mathbb{R}^n$ is a **subgradient** of F at a point $x \in \mathbb{R}^n$ if

$$f(z) \geq f(x) + \langle v, z - x \rangle \quad \forall z \in \mathbb{R}^n.$$

IBC of convex optimization

Convex optimization

Numerical integration

Stochastic optimization

Dynamic SP

Galerkin method

An application

Theorem 1 *Assume that $X \subset \mathbb{B}$ is convex and that F is convex and Lipschitz continuous. Then the gradient projection algorithm of [Shor, 1967] finds an ε -optimal solution in*

$$k(n, \varepsilon) = O \left(\frac{\text{lip } F}{\varepsilon} \right)^2$$

iterations. This is the best (up to a constant) dimension independent bound.

- The constants as well as the projection operator depend on the chosen “metric” so that should be chosen carefully.
- The variant obtained with a **Bregman distance** is called the **mirror prox**.

IBC of convex optimization

Convex optimization

Numerical integration

Stochastic optimization

Dynamic SP

Galerkin method

An application

Theorem 2 *Assume that $X \subset \mathbb{B}$ is convex and that F is convex with a Lipschitz continuous gradient. Then the algorithm of [Nesterov, 1983] finds an ε -optimal solution in*

$$k(n, \varepsilon) = O \left(\frac{\text{lip } \nabla F}{\varepsilon} \right)^{\frac{1}{2}}$$

iterations. This is the best (up to a constant) dimension independent bound.

IBC of convex optimization

Convex optimization

Numerical integration

Stochastic optimization

Dynamic SP

Galerkin method

An application

Theorem 3 *Assume that $X, Y \subset \mathbb{B}$ are convex and that*

$$F(x) = \sup_{y \in Y} L(x, y),$$

*where L is convex-concave with Lipschitz continuous gradient. [Nesterov, 2004] applies the previous algorithm to a **smooth approximation** of F and finds an ε -optimal solution in*

$$k(n, \varepsilon) = O \left(\frac{\text{lip } \nabla L}{\varepsilon} \right)$$

iterations. This is the best (up to a constant) dimension independent bound.

IBC of numerical integration

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

Multivariate integrals

$$E^P f = \int_{\Omega} f(\omega) P(d\omega)$$

are often approximated by **quadratures** of the form

$$E^{P^\nu} f = \sum_{i=1}^{\nu} p_i^\nu f(\omega_i^\nu).$$

In order to bound the approximation error, we have to restrict the class of integrands f .

IBC of numerical integration

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

- Let $\mathcal{X} \subset L^1(\Omega, \mathcal{F}, P)$ be a function space where $f \mapsto E^P f$ and $f \mapsto f(\omega)$ for $\omega \in \Omega$ are **continuous**
- Measuring the “regularity” of an integrand by a **seminorm** $V : \mathcal{X} \rightarrow \overline{\mathbb{R}}$, we have the **error bound**

$$|E^P f - E^{P^\nu} f| \leq V(f) V^\circ(P - P^\nu),$$

where V° is the **polar** of V , defined for each $\mu \in \mathcal{X}^*$ by

$$V^\circ(\mu) = \sup_{f \in \mathcal{X}} \{\mu(f) \mid V(f) \leq 1\}.$$

The bound is tight if $\mathcal{X}$ is locally convex and V is lsc.

- The error bound is useful if we can
 1. bound $V(f)$ for the given f ,
 2. construct quadratures P^ν so that $V^\circ(P - P^\nu)$ is small.

IBC of numerical integration

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

Given the dimension d of Ω and an $\varepsilon > 0$, let $\nu(d, \varepsilon)$ be the number of quadrature points required to achieve

$$V^\circ(P - P^\nu) \leq \varepsilon V^\circ(P).$$

The problem (V, P) of evaluating $E^P f$ for functions $f \in \mathcal{X}$ with finite $V(f)$ is

- **tractable** if there exist constants a , b and c such that

$$\nu(d, \varepsilon) \leq cd^a \varepsilon^{-b},$$

- **strongly tractable** if $a = 0$, i.e. the complexity is independent of the dimension.

IBC of numerical integration

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

Example 4 (Modulus of Lipschitz continuity) *Let $\Omega \subset \mathbb{R}^d$ be compact and $\mathcal{X} = C(\Omega)$ the Banach space of continuous functions with the supremum norm. Let $V(f)$ be the **modulus of Lipschitz continuity** of f .*

- *In this case, $\mathcal{X}^*$ is the space of regular Borel measures and $V^\circ(P - P^\nu)$ is the **Wasserstein metric**.*
- *The construction of discrete measures with small $V^\circ(P - P^\nu)$ and a given number of quadrature points is known as **quantization**.*
- *In general, the integration problem (V, P) for the Lipschitz modulus V is **not tractable**.*

IBC of numerical integration

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

Example 5 (Variation) Let P be the Lebesgue measure on $\Omega := [0, 1]^d$ and $\mathcal{X}$ the Banach space of bounded measurable functions with the supremum norm. Let V be the **variation in the sense of Hardy and Krause**. If P^ν is a quadrature with equal weights $p^{\nu,i} = 1/\nu$, then

$$V^\circ(P - P^\nu) = \sup_{\omega \in (0,1)^d} |P([0, \omega]) - P^\nu([0, \omega])| \quad (\text{star-discrepancy}).$$

- There exist various constructions with

$$V^\circ(P - P^\nu) \leq c_d \frac{(\ln \nu)^{d-1}}{\nu}$$

- [Heinrich, Novak, Wasilkowski and Woźniakowski, 2001] proved the existence of a quadrature with

$$\nu(d, \varepsilon) \leq c \frac{d}{\varepsilon^2}.$$

IBC of numerical integration

Convex optimization

Numerical integration

Stochastic optimization

Dynamic SP

Galerkin method

An application

Example 6 (Weighted RKHS) *In certain reproducing kernel Hilbert spaces, tractability and even strong tractability is achieved by **lattice rules** and certain **low discrepancy sequences**.*

Sloan and Wozniakowski, When are Quasi-Monte Carlo algorithms efficient for high dimensional integrals?, 1998.

—, Tractability of Multivariate Integration for Weighted Korobov Classes, 2001.

Novak and Wozniakowski, Tractability of Multivariate Problems. Vols. I & II, 2008 & 2010.

IBC of stochastic optimization

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

Given a quadrature P^ν , we replace the objective in

$$\text{minimize } F(x) := E f(x, \omega) \quad \text{over } x \in X, \quad (\text{SP})$$

by

$$F^\nu(x) := E^\nu f(x, \omega) = \sum_{i=1}^{\nu} p^{\nu,i} f(x, \omega^{\nu,i}).$$

- F^ν is convex/continuous/smooth when $f(\cdot, \omega)$ is so.
- If the complexity of minimizing F^ν over X can be bounded by a function $k(n, \varepsilon)$ independently of P^ν , then F^ν can be minimized with $\nu k(n, \varepsilon)$ gradient evaluations of f .

IBC of stochastic optimization

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

Theorem 7 *Assume that*

1. *the integration problem (V, P) has complexity $\nu(d, \varepsilon)$ and*

$$V(f) := \sup_{x \in X} V(f(x, \cdot)) < \infty,$$

2. *minimization of F^ν over X has complexity $k(n, \varepsilon)$ independent of P^ν ,*

Then an ε -solution of (SP) can be found with

$$l(n, d, \varepsilon) = \inf_{\varepsilon_1, \varepsilon_2} \{ \nu(d, \varepsilon_1) k(n, \varepsilon_2) \mid 2V(f)\varepsilon_1 + \varepsilon_2 \leq \varepsilon \},$$

gradient evaluations of f .

IBC of stochastic optimization

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

Corollary 8 *Assume that the optimization and integration problems are tractable with $\nu(d, \varepsilon) = c_1 d^{a_1} / \varepsilon^{b_1}$ and $k(n, \varepsilon) = c_2 n^{a_2} / \varepsilon^{b_2}$, respectively. Then (SP) is tractable with*

$$l(n, d, \varepsilon) = c_3 \frac{d^{a_1} n^{a_2}}{\varepsilon^{b_1 + b_2}},$$

where c_3 is independent of the dimensions. This is achieved by a quadrature with worst case error

$$\varepsilon_1 = \frac{b_1}{b_1 + b_2} \frac{\varepsilon}{2V(f)}$$

and by solving the discretized problem with accuracy

$$\varepsilon_2 = \frac{b_2}{b_1 + b_2} \varepsilon.$$

Smoothing in numerical integration?

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

Given any $\tilde{f} \in \mathcal{L}^1$,

$$\begin{aligned} |E^P f - E^{P^\nu} f| &\leq |E^P \tilde{f} - E^{P^\nu} \tilde{f}| + E^P |\tilde{f} - f| + E^{P^\nu} |\tilde{f} - f| \\ &\leq V(\tilde{f})V^\circ(P - P^\nu) + 2\|\tilde{f} - f\|_\infty, \end{aligned}$$

so

$$|E^P f - E^{P^\nu} f| \leq \inf_{\tilde{f} \in \mathcal{L}^1} \{V(\tilde{f})V^\circ(P - P^\nu) + 2\|\tilde{f} - f\|_\infty\}.$$

- This can be much smaller than $V(f)V^\circ(P - P^\nu)$: if $f(\omega) = \epsilon \sin(\omega/\epsilon^2)$ and $\tilde{f} = 0$, we have $V(f) = 1/\epsilon$ while the above gives $|E^P f - E^{P^\nu} f| \leq 2\epsilon$.
- When/how can we evaluate/overestimate the above?
- Mollification in RKHS?

Dynamic stochastic optimization

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

Let $(\mathcal{F}_t)_{t=0}^T$ be an increasing sequence of sigma-algebras in the probability space $(\Omega, \mathcal{F}, P)$ and consider the **dynamic stochastic optimization** problem

$$\begin{aligned} & \text{minimize} && E f(x(\omega), \omega) && \text{over } x \in \mathcal{N}, \\ & \text{subject to} && x(\omega) \in X(\omega) && P\text{-a.s.} \end{aligned} \tag{DSP}$$

where

- $\mathcal{N} = \{(x_t)_{t=0}^T \mid x_t \in L^\infty(\Omega, \mathcal{F}_t, P; \mathbb{R}^{n_t})\}$ for given $n_t \in \mathbb{N}$,
- f is a convex Carathéodory function on $\mathbb{R}^n \times \Omega$, where $n = n_0 + \cdots + n_T$,
- $X : \Omega \rightrightarrows \mathbb{R}^n$ is $\mathcal{F}$ -measurable and convex-valued.

Galerkin method

- Let $\{x^i \in \mathcal{N} \mid i \in I\}$ be a finite collection of feasible **basis strategies** and $\Delta := \{\alpha \in \mathbb{R}_+^I \mid \sum_{i \in I} \alpha^i = 1\}$.
- Every strategy in the **finite-dimensional** set

$$\mathcal{N}_\Delta := \left\{ \sum_{i \in I} \alpha^i x^i \mid \alpha \in \Delta \right\}$$

is adapted and feasible.

- We could also take
 - $\Delta = \mathbb{R}^I$ when X is linear-valued
 - $\Delta = \mathbb{R}_+^I$ when X is cone-valued,

Galerkin method

The corresponding **Galerkin approximation** of (DSP) is

$$\text{minimize} \quad E f(x(\omega), \omega) \quad \text{over} \quad x \in \mathcal{N}_\Delta.$$

This can be written in the **static** form

$$\text{minimize} \quad E \varphi(\alpha, \omega) \quad \text{over} \quad \alpha \in \Delta,$$

where

$$\varphi(\alpha, \omega) := f \left(\sum_{i \in I} \alpha^i x^i(\omega), \omega \right)$$

is convex in α and integrable in ω .

- In a quadrature approximation, we only need to evaluate each $x^i(\omega^j)$ once.

Galerkin method

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

Lemma 9 *We have*

$$\inf_{\mathcal{N}_\Delta} E f(x) - \inf_{\mathcal{N}_X} E f(x) \leq (\text{lip } E f) \text{gap}(\mathcal{N}_\Delta, \varepsilon\text{-argmin}_{x \in \mathcal{N}_X} E f) + \varepsilon,$$

where $\varepsilon \geq 0$ and $\text{gap}(C^1, C^2) := \min\{\|x^1 - x^2\| \mid x^i \in C^i\}$

- Compare with Céa's lemma for elliptic PDEs.

Lemma 10 *If there is an $L \in \mathbb{R}$ such that, for every $\omega \in \Omega$,*

$$|f(x_1, \omega) - f(x_2, \omega)| \leq L|x_1 - x_2| \quad \forall x_1, x_2 \in X(\omega)$$

then $\text{lip } E f \leq L$ wrt the L^∞ -norm..

An application

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

Consider the **dynamic portfolio optimization problem** of a trader who has **initial capital** V_0 and has to pay out **claims** c_t , $t = 1, \dots, T$.

- Assume a finite set J of investment classes (bonds, equities, real estate, ...).
- Let $R_{t,j}$ be the total **return** on class $j \in J$ over $[t-1, t]$.
- $x_{t,j}$ is the amount of **money invested** in class $j \in J$ at the beginning of period t .
- $(\mathcal{F}_t)_{t=0}^T$ is the filtration generated by $c = (c_t)_{t=1}^T$, $R = (R_t)_{t=1}^T$ and possibly by other observable quantities.

An application

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

Consider the problem

$$\begin{aligned} & \text{minimize} && CV@R_\delta(-V_T) && \text{over} && h \in \mathcal{N} \\ & \text{subject to} && \sum_{j \in J} x_{0,j} \leq V_0, \\ & && \sum_{j \in J} x_{t,j} + c_t \leq \sum_{j \in J} R_{t,j} x_{t-1,j} \quad t = 1, \dots, T, \\ & && V_T \leq \sum_{j \in J} x_{T,j}, \\ & && x_{t,j} \geq 0 \quad j \in J \setminus \{0\}. \end{aligned}$$

Here

$$CV@R_\delta(c) = \inf_{z \in \mathbb{R}} E \left[z + (1 - \delta)^{-1} [c(\omega) - z]_+ \right].$$

An application

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

- We will employ the Galerkin method in an instance with $T = 82$ and five asset classes.
- The claims c_t are the yearly pensions of the Finnish private sector occupational pension system.
- We use 529 basis strategies which are variations of
 - buy and hold
 - constant proportions
 - portfolio insurance
 - “target date fund”strategies with varying parameters. The strategies are modified in order to allow for the claim payments.
- These strategies are all adapted to $(\mathcal{F}_t)_{t=0}^T$.

An application

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

The first two columns give the optimal weights and the types of the active basis strategies. The last column gives the corresponding objective value.

Weight	Type	$CV@R_{97.5\%}$ (billion €)
0.665	BH	1569
0.029	BH	6567
0.104	BH	5041
0.022	CP	3324
0.039	PI	1420
0.099	PI	1907
0.042	PI	2417
Best basis		1020
Galerkin		251

An application

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

Initial capital V_0 required for $CV@R_\delta(-V_T) \leq 0$

	Confidence level δ				
	95%	90%	85%	80%	66%
Best basis	296	284	273	261	239
Optimized	288	271	254	236	202

Solvency ratio: $72.1/V_0$

	Confidence level δ				
	95%	90%	85%	80%	66%
Best basis	24.3	25.4	26.4	27.6	30.1
Optimized	25.0	26.6	28.3	30.5	35.6

Conclusions

Convex optimization
Numerical integration
Stochastic optimization
Dynamic SP
Galerkin method
An application

- Stochastic optimization is fundamental in risk management.
- Numerical techniques are necessary for most realistic models. Requires both numerical integration and optimization.
- Promising numerical results.
- Open problems:
 - Systematic/adaptive construction of basis strategies
 - Employing problem structure and dynamic programming
 - Computable error bounds for nonsmooth integrands
 - Average case complexity.