Randomized methods based on new Monte Carlo schemes for control and optimization

B.T. Polyak

Institute for Control Sciences, Russian Academy of Sciences, Moscow, Russia

July 2008

with E. Gryazina, F. Dabbene, P. Scherbakov

Efficient Monte Carlo: From Variance Reduction to Combinatorial Optimization — A Conference on the Occasion of R.Y.Rubinstein's 70th Birthday (Soenderborg)

Outline

- "Ideal" Monte Carlo and convergence estimates
- Implementable random algorithms: boundary oracle
  - Hit-and-Run
  - Shake-and-Bake
- Numerical simulation
- Conclusions

Historical background

- First random search methods
  Rastrigin (1960-ies)
- Pessimism on effectiveness of randomized algorithms
  Nemirovski and Yudin (1983)
- Revival of randomized approaches for optimization
  Bertsimas and Vempala (2004), Polyak and Shcherbakov (2006) and Dabbene (2008), Campi (2008)

Historical background

- First random search methods
  Rastrigin (1960-ies)
- Pessimism on effectiveness of randomized algorithms
  Nemirovski and Yudin (1983)
- Revival of randomized approaches for optimization
  Bertsimas and Vempala (2004), Polyak and Shcherbakov (2006) and Dabbene (2008), Campi (2008)

Historical background

- First random search methods
  Rastrigin (1960-ies)
- Pessimism on effectiveness of randomized algorithms
  Nemirovski and Yudin (1983)
- Revival of randomized approaches for optimization
  Bertsimas and Vempala (2004), Polyak and Shcherbakov (2006) and Dabbene (2008), Campi (2008)

Problem statement

$$\min \; c^T x$$
$$\text{s.t. } x \in X$$

X is a convex bounded closed set in $\mathbb{R}^n$ with nonempty interior

Remark: Any convex optimization problem

$$\min \; f(x)$$
$$\text{s.t. } x \in Q,$$
$$Q \text{ and } f \text{ are convex}$$

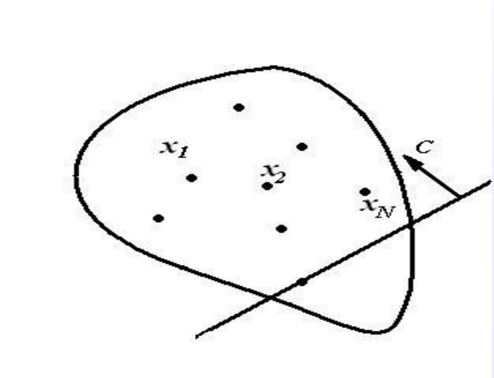
can be converted to this format.

"Ideal" Monte Carlo

$x_1, x_2, \ldots x_N$ independent uniformly distributed points in X

Cutting plane method

1. Set $X_1 = X$
2. Generate uniform $x_1, x_2, \ldots x_N \in X_k$
3. Find $f_k = \min c^T x_i$
4. Set $X_{k+1} = X_k \bigcap \{x : c^T x \leq f_k\}$
   go to Step 2.



Convergence estimates

$$f^* = \max_{x \in X} \; c^T x, \quad f_* = \min_{x \in X} \; c^T x, h = f^* - f_*$$

Theorem

After k iterations of the algorithm

$$E\left[f_k\right] - f_* \leq q^k, \quad q = \frac{h}{n} B\left(N+1, \frac{1}{n}\right),$$

where $B(a,b)$ is Euler beta-function.

Case of a special interest: $N = 1, \; k = 1$

Theorem

Let x_1 be a random point uniformly distributed in X. Then

$$E\left[c^T x_1\right] - f_* \leq h\left(1 - \frac{1}{n+1}\right).$$

Radon theorem and center of gravity method

Remark:

$$E\left[x_1\right] = g \text{ (center of gravity of } X\text{)}, \quad \Rightarrow \quad c^T g - f_* \leq h\left(1 - \frac{1}{n+1}\right)$$

[Radon theorem (1916)]

Deterministic version:

center of gravity method

$$x^k = g^k, \quad X_{k+1} = X_k \bigcap \{x : c^T x \leq c^T g^k\}$$

Implementable random algorithms: boundary oracle

uniform random samples $\Rightarrow$ *asymptotically* uniformly distributed via Markov-chain Monte Carlo schemes

Given $x_0 \in X$, d — vector specifying the direction in $\mathbb{R}^n$

Boundary oracle

$L = \{t \in \mathbb{R} : x^0 + td \in X\}$

For convex sets $L = (\underline{t}, \bar{t})$,
where $\underline{t} = \inf\{t : x^0 + td \in X\}$, $\bar{t} = \sup\{t : x^0 + td \in X\}$

Complete boundary oracle

$L = \{t \in \mathbb{R} : x^0 + td \in X\}$ + inner normals to X at the boundary points

Implementable random algorithms: boundary oracle

uniform random samples $\Rightarrow$ *asymptotically* uniformly distributed via Markov-chain Monte Carlo schemes

Given $x_0 \in X$, d — vector specifying the direction in $\mathbb{R}^n$

Boundary oracle

$L = \{t \in \mathbb{R} : x^0 + td \in X\}$

For convex sets $L = (\underline{t}, \bar{t})$,
where $\underline{t} = \inf\{t : x^0 + td \in X\}$, $\bar{t} = \sup\{t : x^0 + td \in X\}$

Complete boundary oracle

$L = \{t \in \mathbb{R} : x^0 + td \in X\}$ + inner normals to X at the boundary points

Implementable random algorithms: boundary oracle

uniform random samples $\Rightarrow$ *asymptotically* uniformly distributed via Markov-chain Monte Carlo schemes

Given $x_0 \in X$, d — vector specifying the direction in $\mathbb{R}^n$

Boundary oracle

$L = \{t \in \mathbb{R} : x^0 + td \in X\}$

For convex sets $L = (\underline{t}, \bar{t})$,
where $\underline{t} = \inf\{t : x^0 + td \in X\}$, $\bar{t} = \sup\{t : x^0 + td \in X\}$

Complete boundary oracle

$L = \{t \in \mathbb{R} : x^0 + td \in X\}$ + inner normals to X at the boundary points

Boundary oracle is available for numerous sets

- LMI set

$$X = \left\{ x \in \mathbb{R}^n : A_0 + \sum_{i=1}^{n} x_i A_i \leq 0 \right\}$$

- LMI constrained set of symmetric matrices P

$$X = \left\{ P : AP + PA^T + C \leq 0,\ P \geq 0 \right\}$$

- Quadratic matrix inequalities set

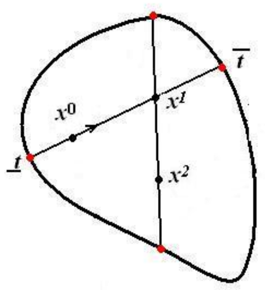
$$X = \left\{ P : AP + PA^T + PBB^T P + C \leq 0,\ P \geq 0 \right\}$$

- Linear algebraic inequalities set

$$X = \left\{ x \in \mathbb{R}^n : c_i^T x \leq a_i,\ i = 1, \ldots, m \right\}$$

Hit-and-Run

1. $i = 0$, $x^0 \in X$
2. Choose random direction d uniformly distributed on the unit sphere
3. $x^{i+1} = x^i + t_1 d$,
   t_1 is uniformly distributed on $L = (\underline{t}, \overline{t})$
4. L is updated with respect to x^{i+1}, go to Step 2.



Theorem (Smith (1984))

Let X be bounded open or coincides with the closure of interior points of X. Then for any measurable set $A \subset X$ probability $P_i(A) = P(x^i \in A | x^0)$ can be estimated as $|P_i(A) - P(A)| \leq q^i$, where $P(A) = \dfrac{\mathrm{Vol(A)}}{\mathrm{Vol(X)}}$ and $q < 1$ does not depend on x^0.

Shake-and-Bake: an alternative way for generating points

Points are *asymptotically* uniformly distributed in the boundary of X.
Complete boundary oracle is exploited.

- LMI set

$$X = \left\{ x \in \mathbb{R}^n : A_0 + \sum_{i=1}^{n} x_i A_i \leq 0 \right\}$$

$n_i = -(A_i e, e),$ where e is the eigenvector corresponding to zero eigenvalue of the matrix $A_0 + \sum_{i=1}^{n} x_i^0 A_i.$

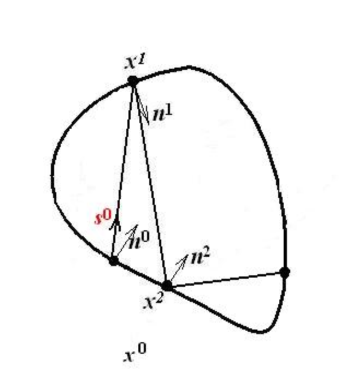
- LMI constrained set of symmetric matrices P

$$X = \left\{ P : AP + PA^T + C \leq 0 \right\}$$

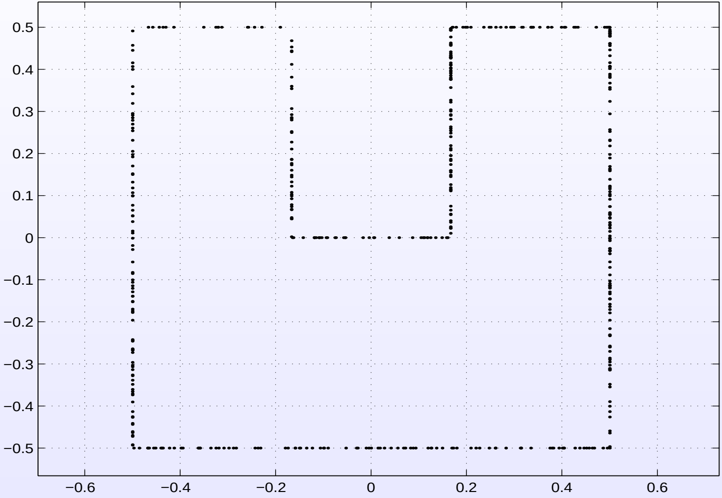
$N = -(ee^T A + A^T ee^T),$ where e is the eigenvector corresponding to zero eigenvalue of the matrix $AP_0 + P_0 A^T + C.$

Shake-and-Bake: the algorithm

1. $i = 0$, $x^0 \in \partial X$, n^0 is the normal.
2. Choose random direction s^i,
   $s^i = \sqrt{1 - \xi^{\frac{2}{n-1}}} n^0 + r$,
   ξ uniform random in $(0, 1)$,
   r is random unit uniform direction $(n^0, r) = 0$.
3. $x^{i+1} = x^i + \bar{t}s$,
   $\bar{t}$ is given by the boundary oracle for the direction s.
4. L is updated with respect to x^{i+1}, go to Step 2.



Shake-and-Bake for nonconvex sets



Numerical simulation

Standard SDP of the form

$$\min \ c^T x$$

$$\text{s.t. } A_0 + \sum_{i=1}^{n} x_i A_i \leq 0$$

A_i, $i = 0, 1, \dots n$ — symmetric real matrices $m \times m$; $c = [0, \dots, 0, 1]$
We applied modified HR where $\min \ x_i$ was replaced with averaged X_i $+$ various heuristic acceleration methods (scaling, projecting, accelerating step)
Open problem: number of HR points in every step.

Numerical simulation

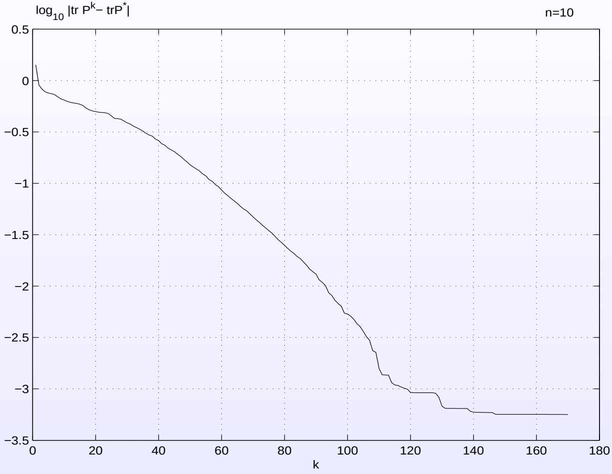
Standard SDP of the form

$$\min \; trP$$
$$\text{s.t. } AP + PA^T \leq 0$$
$$P \geq I$$

A is stable matrix $n \times n$

We applied SB for various dimensions
$+$ extension for the nonsmooth boundary

Simulation results:small size problem



Applications to control

Sets with available boundary oracle

- Stability set for polynomials

$$\mathcal{K} = \{k \in \mathbb{R}^n : p(s,k) = p_0(s) + \sum_{i=1}^{n} k_i p_i(s) \text{ is stable}\}$$

- Stability set for matrices

$$A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}, C \in \mathbb{R}^{l\times n}$$

$$\mathcal{K} = \{K \in \mathbb{R}^{m\times l} : A + BKC \text{ is stable}\}$$

- Robust stability set for polynomials

$$\mathcal{K} = \{k : P_0(s,q) + \sum_{i=1}^{n} k_i P_i(s,q) \text{ is stable } \forall q \in Q\}, \quad Q \subset \mathbb{R}^m$$

- Quadratic stability set

$$\dot{x} = Ax$$

$$\mathcal{K} = \{P > 0 : AP + PA^T \leq 0\}$$

Stability set for polynomials

$$\mathcal{K} = \{k \in \mathbb{R}^n : p(s,k) = p_0(s) + \sum_{i=1}^{n} k_i p_i(s) \text{ is stable}\}$$

$k^0 \in \mathcal{K}$ i.e. $p(s,k^0)$ is stable,

$d = s/||s||, s =$ `randn(n,1)` — random direction

Boundary oracle: $L = \{t \in \mathbb{R} : k^0 + td \in \mathcal{K}\}$,

i.e. $\{t \in \mathbb{R} : p(s,k^0) + t \sum d_i p_i(s) \text{ is stable}\}$.

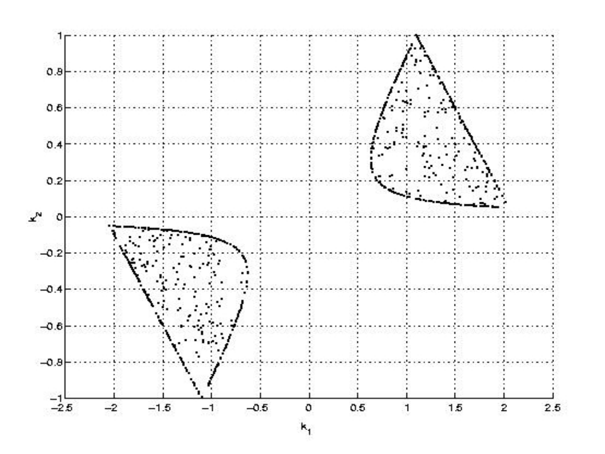
D**-decomposition problem for real scalar parameter** t**!**

Gryazina E. N., Polyak B. T. Stability regions in the parameter space: D-decomposition revisited //Automatica. 2006. Vol. 42, No. 1, P. 13–26.

Example: Generating points in the disconnected set

$$\mathcal{K} = \{k \in \mathbb{R}^n : p(s,k) = p_0(s) + \sum_{i=1}^{n} k_i p_i(s) \text{ is stable}\},$$

$$p(s,k) = 2.2s^3 + 1.9s^2 + 1.9s + 2.2 + k_1(s^3 + s^2 - s - 1) + k_2(s^3 - 3s^2 + 3s - 1)$$



Stability set for matrices

$$\dot{x} = Ax + Bu, \quad y = Cx, \quad u = Ky$$

$$A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}, C \in \mathbb{R}^{l\times n}; \quad \mathcal{K} = \{K \in \mathbb{R}^{m\times l} : A + BKC \text{ is stable}\}$$

$K^0 \in \mathcal{K}$, i.e. $A + BK^0C$ is stable
$D = Y/||Y||, Y =$ `randn`(m, l) — random direction in the matrix space K

$A + B(K^0 + tD)C = F + tG$, where $F = A + BK^0C$, $G = BDC$

Boundary oracle: $L = \{t \in \mathbb{R} : F + tG \text{ is stable}\}$
Total description of L is hard:
$f(t) = \max Re \,\text{eig}(F + tG)$
numerical solution of the equation $f(t) = 0$, $t > 0$ (MatLab command `fsolve`)

Quadratic stability

$$\dot{x} = Ax + Bu, \quad u = Kx$$

$$\mathcal{K} = \{K : \exists P > 0, A_c^T P + P A_c \le 0\}, \quad A_c = A + BK$$

$\mathcal{K}$ is convex and bounded.

$$Q = P^{-1} > 0, \quad QA^T + AQ + BY + Y^T B^T < 0, \quad Y = KQ.$$

$k^0 \in \mathcal{K}$, $Q_0 = P_0^{-1}$, $Y_0 = K_0 Q_0$ — starting points

$Q = Q_0 + tJ$, $Y = Y_0 + tG$, where J and G are random directions in the matrix space.

initial inequality $\iff F + tR < 0$

Boundary oracle: $L = (-\underline{t}, \overline{t})$,

where $\overline{t} = \min \lambda_i$, $\underline{t} = \min \mu_i$;

λ_i — real positive eigenvalues for the pair of matrices

$F = Q_0 A^T + AQ_0 + BY_0 + Y_0^T B^T$ and $-R = JA^T + AJ + BG + G^T B^T$;

μ_i correspondingly for matrices F, R.

Polyak B.T., Shcherbakov P.S. The D-decomposition technique for linear matrix inequalities // Automation and Remote Control. 2006. No. 11. P. 1847–1861

Conclusions

- Randomized approaches for optimization are promising.
- Proposed methods are simple in implementation and give an opportunity to solve large-dimensional problems.