ETHZ, May 2006

Polynomial technique in combinatorial optimization

Yu. Nesterov

CORE-FSA

Catholic University of Louvain

Louvain-la-Neuve, Belgium

Combinatorial problems

Problem:

Check that the boolean knapsack polytope is nonempty:

$$\mathcal{B}_a^1 = \{x \in \{0,1\}^n : \langle a, x \rangle = b\} \neq \emptyset,$$

where $b \in Z$ and $a \in Z^n$.

Different approaches:

1. Check positivity:

$$(\langle a, x \rangle - b)^2 + \sum_{i=1}^{n} [x^{(i)}(1 - x^{(i)})]^2 \geq 1 \quad \forall x \in R^n.$$

- Representation as a sum of squares?
- Computational complexity?

2. Direct methods:

Complete enumeration related to a certain grid.

Advantage: Simple complexity bounds.

Approximation by points from a grid

Problem:

$$f_d(x) \quad \to \ \min: \ x \in \Delta_n(1),$$

$$\Delta_n(k) = \{x \in R^n : \ \sum_{i=1}^{n} x^{(i)} = k,$$

$$x^{(i)} \geq 0, \ i = 1, \ldots, n\},$$

where $f(x)$ is a homogeneous polynomial of degree $d \geq 2$.

Case $d = 2$:

1. Nesterov 1999. Take $\hat{f} = \min_{x \in \mathcal{T}_n} f_2(x)$ with

$$\mathcal{T}_n = \{x = \tfrac{e_i + e_j}{2}, \ i, j = 1, \ldots, n\}.$$

Then $\hat{f} - f_* \leq \frac{1}{2} \cdot (\max_{1 \leq i \leq n} f_2(e_i) - f_*)$.

(Based on Linear Matrix Inequalities.)

2. Bomze, de Klerk 2002. Take

$$\mathcal{T}_n(k) = \{x = \tfrac{1}{k} y, \ y \in \Delta_n(k) \cap Z^n\}.$$

Then $\hat{f} - f_* \leq \frac{1}{k} \cdot (\max_{x \in \Delta_n(1)} f_2(x) - f_*)$.

(Based on representations of sums of squares.)

Random walk in a simplex

Let us fix some vector of probabilities $p \in \Delta_n(1)$.
Denote by $\zeta(p)$ a discrete random variable distributed as

$$\text{Prob}\{\zeta(p) = i\} = p^{(i)}, \quad i = 1, \ldots, n. \qquad (1)$$

Consider the following process:

$$\begin{aligned} x_0(p) &= 0 \in R^n, \\ x_{k+1}(p) &= x_k(p) + e_{\zeta_k(p)}, \; k \geq 0, \end{aligned} \qquad (2)$$

where all $\zeta_k(p)$ are distributed as (1).

Note: all $x_k(p) \in \Delta_n(k)$. Hence, the process

$$y_k(p) = \tfrac{1}{k} x_k(p), \quad k \geq 1, \qquad (3)$$

is a random walk in the simplex $\Delta_n(1)$.

A direct implementation of $y_k(p)$ is as follows:

$$\begin{aligned} y_0(p) &= 0 \in R^n, \\ y_{k+1}(p) &= \tfrac{k}{k+1} y_k(p) + \tfrac{1}{k+1} e_{\zeta_k(p)}, \; k \geq 0. \end{aligned} \qquad (4)$$

Expectations

Note: $x_k(p) \in \Delta_n(k) \cap Z^n$ and for any $\alpha \in \Delta_n(k) \cap Z^n$ we have

$$\textbf{Prob}\,(x_k(p) = \alpha) = \tfrac{k!}{\alpha!} \cdot p^\alpha,$$

where

$$\alpha! = \prod_{i=1}^{n} \alpha^{(i)}!, \quad p^\alpha = \prod_{i=1}^{n} (p^{(i)})^{\alpha^{(i)}}.$$

Therefore, for $i, j = 1, \dots, n$, we have

$$\begin{aligned} E(x_k(p)^{(i)}) &= kp^{(i)}, \\ E([x_k(p)^{(i)}]^2) &= kp^{(i)} + k(k-1)\left(p^{(i)}\right)^2, \\ E(x_k(p)^{(i)} x_k(p)^{(j)}) &= k(k-1)p^{(i)}p^{(j)}. \end{aligned}$$

Version for $y_k(p)$:

$$\begin{aligned} E(y_k(p)^{(i)}) &= p^{(i)}, \\ E([y_k(p)^{(i)}]^2) &= \tfrac{1}{k}p^{(i)} + \left(1 - \tfrac{1}{k}\right)\left(p^{(i)}\right)^2, \\ E(y_k(p)^{(i)} y_k(p)^{(j)}) &= \left(1 - \tfrac{1}{k}\right) p^{(i)}p^{(j)}. \end{aligned}$$

Quadratic optimization

Problem:

Find $f_* = \min\limits_x \{ f_2(x) \equiv \langle Qx, x \rangle : \ x \in \Delta_n(1) \}, \quad (5)$

where Q is a symmetric $n \times n$-matrix. Define

$$\hat{f}_k = \min_\alpha \left\{ \tfrac{1}{k^2} f_2(\alpha) : \ \alpha \in \Delta_n(k) \cap Z^n \right\}, \quad k \geq 1.$$

Theorem 1 *For any $k \geq 1$ we have*

$$0 \leq \hat{f}_k - f_* \leq \tfrac{1}{k} \left[\max_{1 \leq i \leq n} Q^{(i,i)} - f_* \right]. \tag{6}$$

Proof: Indeed, $f_* \leq \hat{f}_k$. Let us choose $p = x^*$. Then

$$\begin{aligned}
\hat{f}_k &= \min_\alpha \{ f_2\left(\tfrac{\alpha}{k}\right) : \ \alpha \in \Delta_n(k) \cap Z^n \} \\
&\leq E[f(y_k(p))] = E[\langle Q y_k(p), y_k(p) \rangle] \\
&= \sum_{i,j=1}^n Q^{(i,j)} E\left[y_k^{(i)}(p) \cdot y_k^{(j)}(p) \right] \\
&= \sum_{i=1}^n Q^{(i,i)} E\left[\left(y_k^{(i)}(p) \right)^2 \right] + \sum_{i \neq j} Q^{(i,j)} E\left[y_k^{(i)}(p) y_k^{(j)}(p) \right] \\
&= \tfrac{1}{k} \sum_{i=1}^n Q^{(i,i)} p^{(i)} + \left(1 - \tfrac{1}{k}\right) \sum_{i,j=1}^n Q^{(i,j)} p^{(i)} p^{(j)} \\
&\leq \tfrac{1}{k} \max_{1 \leq i \leq n} Q^{(i,i)} + \left(1 - \tfrac{1}{k}\right) f_*. \qquad \Box
\end{aligned}$$

Corollary 1 *If $f_2(e_i) \leq 0 \quad \forall i$, then $f_* \leq 0$ and*

$$\hat{f}_k - f_* \leq \tfrac{1}{k}(-f_*). \tag{7}$$

Complexity:

1. We need to compute $f_2(x)$ in all nodes of $\Delta_n(k) \cap Z^n$ with

$$|\Delta_n(k) \cap Z^n| = \binom{n+k-1}{n-1}.$$

2. For $x \in \Delta_n(k) \cap Z^n$, $k \le n$, the number of non-zero elements is at most k. Thus

$$\text{Computation of } f(x) \quad \Leftrightarrow \quad \tfrac{k(k+1)}{2} \text{ operations.}$$

Examples:

$$k = 3: \qquad n(n+1)(n+2) \qquad \text{operations,}$$

$$k = 4: \; \tfrac{5}{12} n(n+1)(n+2)(n+3) \; \text{operations.}$$

Problems on a polytope

Consider the problem

$$\text{Find } f_* = \min_x \{ f_2(x) \equiv \langle \hat{Q}x, x \rangle : \; x \in \mathcal{P} \}, \qquad (8)$$

where $\mathcal{P} = \text{Conv}\,\{u_i \in R^n, \; i = 1, \ldots, N\}$.

Denote $U = (u_1, \ldots, u_N)$. Define

$$\begin{aligned} \hat{f}_k &= \min_\alpha \left\{ \tfrac{1}{k^2} f_2(U\alpha) : \; \alpha \in \Delta_N(k) \cap Z^n \right\}, \quad k \geq 1, \\ f^* &= \max_{1 \leq i \leq N} f_2(u_i). \end{aligned}$$

Theorem 2 *For any $k \geq 1$ we have*

$$0 \leq \hat{f}_k - f_* \leq \tfrac{1}{k}[f^* - f_*]. \qquad (9)$$

If $f^* \leq 0$, then the relative accuracy of approximation f_k^* is at least $\frac{1}{k}$.

Proof: The problem (8) is equivalent to (5) with

$$Q = U^T \hat{Q} U.$$

□

Random walk in polytope $\mathcal{P}$: $\quad p \in \Delta_N(1)$,

$$\begin{aligned} y_0(p) &= 0 \in R^n, \\ y_{k+1}(p) &= \tfrac{k}{k+1} y_k(p) + \tfrac{1}{k+1} u_{\zeta_k(p)}, \; k \geq 0, \end{aligned} \qquad (10)$$

where $\text{Prob}[\zeta_k(p) = i] = p^{(i)}$, $i = 1, \ldots, N$.

Computing the integer volumes

Denote by $\mathcal{N}(Q)$ the number of integer points in the set Q ($\mathcal{N}(\emptyset) = 0$).

Def. Consider a finite parametric family of discrete sets

$$\widehat{\mathcal{R}} \equiv \{\mathcal{R}(b)\}_{b \in Z_+}.$$

We assume that $\mathcal{R}(b) = \emptyset$ for all b large enough.

The *generating function* of this family is defined as

$$f_{\widehat{\mathcal{R}}}(t) = \sum_{b=0}^{\infty} \mathcal{N}(\mathcal{R}(b)) \cdot t^b, \quad t \in R.$$

Note: In many cases $f_{\widehat{\mathcal{R}}}(t)$ has a *short* form.

Knapsack problems

Main object:

$$B_a^u(b) = \left\{ x \in \prod_{i=1}^{n} \{0, \ldots, u^{(i)}\} : \ \langle a, x \rangle = b \right\},$$

the bounded knapsack polytope.

For $\mathcal{B}_a^u = \{B_a^u(b)\}_{b \in Z_+}$. Its generating function is

$$f_{\mathcal{B}_a^u}(t) = \sum_{b=0}^{\infty} \mathcal{N}(B_a^u(b)) \cdot t^b, \quad t \in R. \tag{11}$$

That is a polynomial of degree $\langle a, u \rangle$.

Lemma 1

$$f_{\mathcal{B}_a^u}(t) = \prod_{i=1}^{n} \left(\sum_{k=0}^{u^{(i)}} t^{ka^{(i)}} \right). \tag{12}$$

Multiplication of polynomials

Lemma 2 *Let polynomial $f(t)$ be represented as a product of several polynomials:*

$$f(t) = \prod_{i=1}^{n} p_i(t).$$

Then its coefficients can be computed by FFT in

$$O(D(f)\ \ln D(f)\ \ln n)$$

arithmetic operations.

Proof: Multiplication by pairs. □

(Consecutive multiplication gives $O(nD)$.)

Theorem 3 *All $\langle a, u\rangle$ coefficients of the polynomial $f_{\mathcal{B}_a^u}(t)$ can be computed by FFT in*

$$O(\langle a, u\rangle\ \ln\langle a, u\rangle\ \ln n)$$

arithmetic operations. □

Unbounded knapsack

Consider now the generating function

$$f_{\mathcal{B}_a^\infty}(t) = \sum_{b=0}^{\infty} \mathcal{N}(B_a^\infty(b)) \cdot t^b. \tag{13}$$

It can be represented in a short form:

$$f_{\mathcal{B}_a^\infty}(t) \equiv \prod_{i=1}^{n} \frac{1}{1 - t^{a^{(i)}}}\ , \quad |t| < 1. \tag{14}$$

Theorem 4 *The coefficients of polynomial*

$$g(t) = \prod_{i=1}^{n} (1 - t^{a^{(i)}})$$

can be computed by FFT in

$$O(\|a\|_1 \ \ln \|a\|_1 \ \ln n) \quad a.o. \tag{15}$$

Then, the first $b+1$ *coefficients of* $f_{\mathcal{B}_a^\infty}(t)$ *can be computed in*

$$O(b \ \min\{\ln^2 b, \ln^2 n\}) \quad a.o.$$

Note:

The standard (Dynamic Programming) approach needs

$$O(nb) \quad \text{a.o.}$$

Characteristic functions

Let us fix a cost vector $c \in R^n$.

For a finite set of $\mathcal{R} \subset R^n$, the *characteristic function* is defined by:

$$g_{\mathcal{R}}(c) = \sum_{x \in \mathcal{R}} e^{\langle c,x \rangle},$$

If $\mathcal{R} = \emptyset$, we set $g_{\mathcal{R}}(c) \equiv 0$.

Note: for $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$ we have

$$g_{\mathcal{R}}(c) = g_{\mathcal{R}_1}(c) + g_{\mathcal{R}_2}(c).$$

The *potential function* of $\mathcal{R}$ is given by

$$\psi_{\mathcal{R}}(c) = \ln g_{\mathcal{R}}(c).$$

Define the *support function* of the set $\mathcal{R}$:

$$\xi_{\mathcal{R}}(c) \equiv \max_{x \in \mathcal{R}} \langle c, x \rangle \leq \psi_{\mathcal{R}}(c) \leq \xi_{\mathcal{R}}(c) + \ln \mathcal{N}(\mathcal{R}).$$

Potential and support functions can be as close as needed:

$$\xi_{\mathcal{R}}(c) \leq \mu \psi_{\mathcal{R}}(c/\mu) \leq \xi_{\mathcal{R}}(c) + \mu \ln \mathcal{N}(\mathcal{R}),$$

where $\mu > 0$.

Augmented generating functions

Definition:

$$F_{\widehat{\mathcal{R}}}(c,t) = \sum_{b=0}^{\infty} g_{\mathcal{R}(b)}(c) \cdot t^b, \quad t \in R.$$

Note that $F_{\widehat{\mathcal{R}}}(0,t) \equiv f_{\widehat{\mathcal{R}}}(t)$.

Short form:

Bounded knapsack: $F_{\mathcal{B}_a^u}(c,t) = \prod_{i=1}^{n} \left(\sum_{k=0}^{u^{(i)}} e^{kc^{(i)}} t^{ka^{(i)}} \right)$.

Unbounded knapsack:

$$F_{\mathcal{B}_a^\infty}(c,t) = \left[\prod_{i=1}^{n} (1 - e^{c^{(i)}} t^{a^{(i)}}) \right]^{-1},$$

with $|t| < \min_{1 \le i \le n} e^{-c^{(i)}/a^{(i)}}$.

Optimizing the knapsack

Problem:

$$\text{Find } f^* = \max_{x \in Z_+^n} \{ \langle c, x \rangle : \ \langle a, x \rangle = b \}, \qquad (16)$$

Strategy:

1. Choose μ small enough.

2. Compute coefficients of $f(t) = \prod_{i=1}^{n} (1 - e^{c^{(i)}/\mu} \cdot t^{a^{(i)}})$.

3. Compute the first $b+1$ coefficients of $g(t) = \frac{1}{f(t)}$.

(17)

Theorem 5 *The optimal value of problem (16) can be found by (17) in*

$$O(\|a\|_1 \cdot \ln \|a\|_1 \cdot \ln n \ + \ b \cdot \ln^2 n)$$

operations of <u>*exact real arithmetics*</u>.

Multidimensional parameters

Consider a parametric family of sets in Z^n:

$$\mathcal{X} = \{X(y),\ y \in \Delta\} \subset Z^n,$$

where Δ is a finite subset of Z^m. As before,

$$\psi_{X(y)}(c) = \begin{cases} \sum\limits_{x \in X(y)} e^{\langle c,x \rangle}, & \text{if } X(y) \neq \emptyset, \\ 0, & \text{otherwise}, \end{cases} \qquad (c \in R^n).$$

Hence, $\psi_{X(y)}(0) = \mathcal{N}(X(y))$.

Define (augmented) generating function as

$$g_{\mathcal{X},c}(v) = \sum_{y \in \Delta} \psi_{X(y)}(c) \cdot v^y, \quad v \in C^m, \qquad (18)$$

where $v^y = \prod\limits_{i=1}^{m} (v^{(i)})^{y^{(i)}}$.

Note: all numerical computations with polynomials are very unstable.

Suggestion: restrict the argument onto the unit circle

$$\mathcal{S}_m = \{v \in C^m, \quad |v^{(i)}| = 1,\ i = 1, \ldots, m\}.$$

Then $g_{\mathcal{X},c}(v)$ becomes a *trigonometric polynomial.*

Main advantage:

the system of monomials $\{v^y\}_{y\in Z^m}$, $v \in \mathcal{S}_m$, becomes *orthogonal*.

Lemma 3 *Denote* $\mathbf{j} = \sqrt{-1}$, *and for* $\varphi \in R^m$ *denote*

$$e^{\mathbf{j}\,\varphi} = (e^{\mathbf{j}\,\varphi^{(1)}}, \dots e^{\mathbf{j}\,\varphi^{(m)}})^T,$$

and $d\varphi = d\varphi^{(1)} \dots d\varphi^{(m)}$. *Then*

$$\psi_{X(y)}(c) = \tfrac{1}{(2\pi)^m} \int_0^{2\pi} \dots \int_0^{2\pi} e^{-\mathbf{j}\,\langle y,\varphi\rangle} g_{\mathcal{X},c}(e^{\mathbf{j}\,\varphi})\, d\varphi. \qquad (19)$$

Note: In (19) we need to integrate a *polynomial*.

The value of this integral can be computed by *exact* cubature formulaes.

Theorem 6 *For* $L \in Z^m_+$ *define the following grid*

$$\mathcal{G}_L \;=\; \{\varphi \in R^m :\; \varphi^{(i)} = \tfrac{2\pi}{L^{(i)}} k_i,\; k_i \in Z,$$

$$0 \le k_i \le L^{(i)} - 1,\; i = 1, \dots, m\},$$

$$|\mathcal{G}_L| \;=\; \prod_{i=1}^{m} L^{(i)}.$$

Let $L^{(i)} > |y^{(i)}|$, $i = 1, \dots, m$, *for any* $y \in \Delta$. *Then*

$$\psi_{X(y)}(c) \;=\; \tfrac{1}{|\mathcal{G}_L|} \sum_{\varphi\in\mathcal{G}_L} g_{\mathcal{X},c}(e^{\mathbf{j}\,\varphi}) e^{-\mathbf{j}\,\langle y,\varphi\rangle}, \quad y \in \Delta. \qquad (20)$$

Application example

For $u \in Z_+^n$ and $y \in Z^m$ denote

$$B(u) \;=\; \{x \in Z^n : \; 0 \le x \le u\}$$

$$X_u(y) \;=\; \{x \in B(u) : \; Ax = y\},$$

$$\mathcal{X} \;=\; \{X_u(y), \; y \in \Delta \stackrel{\text{def}}{=} AB(u)\}.$$

where A is an $m \times n$-matrix with integer coefficients.

Let us introduce the trigonometric generating function:

$$g_{\mathcal{X},c}(v) = \sum_{y \in \Delta} \psi_{X_u(y)}(c) \cdot v^y, \quad v \in \mathcal{S}_m. \qquad (21)$$

Lemma 4

$$g_{\mathcal{X},c}(e^{\mathbf{j}\varphi}) \;=\; \prod_{j=1}^{n} \left[1 + \sum_{k=1}^{u^{(j)}} e^{k\cdot(c^{(j)} + \mathbf{j}\,\langle a_j, \varphi\rangle)}\right], \quad \varphi \in R^m,$$

$$= \prod_{j=1}^{n} \frac{e^{(u^{(j)}+1)\cdot(c^{(j)} + \mathbf{j}\,\langle a_j,\varphi\rangle)} - 1}{e^{c^{(j)} + \mathbf{j}\,\langle a_j,\varphi\rangle} - 1}. \qquad (22)$$

where a_j is the jth column of matrix A.

Thus, the value $g_{\mathcal{X},c}(e^{\mathbf{j}\varphi})$ can be computed in

$$O(mn) \quad \text{a.o.}$$

Complexity analysis

1. Size of $\Delta = AB(u)$. Assume that

$$|A^{(i,j)}| \le \alpha, \quad i = 1, \ldots, m, \; j = 1, \ldots, n,$$

and that the box $B(u)$ is uniform:

$$u^{(i)} = \beta, \quad i = 1, \ldots, n.$$

Then, for any $x \in B(u)$ we have

$$| \sum_{j=1}^{n} A^{(i,j)} x^{(j)} | \le \alpha\beta \cdot n. \quad i = 1, \ldots, m.$$

Hence, we can take

$$L^{(i)} = 1 + \alpha\beta \cdot n, \quad i = 1, \ldots, m,$$

and computation of value $\psi_{X(y)}(c)$ by (20) takes

$$O\left(mn \cdot (1 + \alpha\beta \cdot n)^m\right) \quad \text{a.o.} \tag{23}$$

For fixed m, this dependence is polynomial in n.

Alternative: A direct inspection of all $x \in B(u)$, and checking $Ax = b$ takes

$$O\left(mn \cdot (1 + \beta)^n\right) \quad \text{a.o.}$$

(Exponential in n.)

Note: we can solve optimization problems by bisection.