Journal of Global Optimization **23:** 1–41, 2002.

Lagrangian bounds in multiextremal polynomial and discrete optimization problems

NAUM Z. SHOR and PETRO I. STETSYUK
Institute of Cybernetics, NAS of Ukraine, Kiev, Ukraine (e-mail: nzshor@d120.icyb.kiev.ua)

Abstract. Many polynomial and discrete optimization problems can be reduced to multiextremal quadratic type models of nonlinear programming. For solving these problems one may use Lagrangian bounds in combination with branch and bound techniques. The Lagrangian bounds may be improved for some important examples by adding in a model the so-called superfluous quadratic constraints which modify Lagrangian bounds. Problems of finding Lagrangian bounds as a rule can be reduced to minimization of nonsmooth convex functions and may be successively solved by modern methods of nondifferentiable optimization. This approach is illustrated by examples of solving polynomial-type problems and some discrete optimization problems on graphs.

Key words: symmetric matrices, eigenvalues, Lagrangian bounds, discrete optimization problems on graphs, superfluous constraints, quadratic type problems, nondifferentiable optimization

1. Lagrangian bounds, nondifferentiable optimization and nonsmooth matrix functions

Consider the problem of nonlinear programming in a general form: to find

$$f^* = \inf_{x \in X \subseteq E^n} f_0(x), \quad \text{subject to} \quad f_i(x) = 0, \quad i = 1, \dots, m, \tag{1}$$

where E^n is the n-dimensional Euclidean space, X is a closed set in this space, $f_0, f_1, \dots, f_m$ are continuous functions, defined on E^n. We set $f^* = +\infty$, if the problem (1) has no feasible solution. Let us form the usual Lagrange function:

$$L(x, u) = f_0(x) + \sum_{i=1}^{m} u_i f_i(x),$$

where $u = (u_1, \dots, u_m)$ is a vector of Lagrange multipliers.

For each $\overline{u} \in R^m$ we obtain the local problem: to find

$$\psi(\overline{u}) = \inf_{x \in X} L(x, \overline{u}). \tag{2}$$

Function $\psi(u)$ is a concave function with respect to u as a result of minimization of the family of linear in u functions $L_x(u) = L(x, u)$.

Let $\psi(u)$ has a nonempty domain of full dimension m and $x(\overline{u})$ is a solution of local problem (2). It is easy to verify that $\psi(\overline{u}) \leqslant f^*$ for an arbitrary $\overline{u} \in dom\, \psi$.

The supergradient of ψ in the point $\overline{u}$ can be calculated by formula:

$$g_\psi(\overline{u}) = \{f_i(x(\overline{u}))\}_{i=1}^m \in R^m. \tag{3}$$

If the set of vectors generated by (3) is not single, then $\overline{u}$ is the point of nondifferentiability of function ψ.

We try to find the best lower bound for f^* in this class of Lagrangian estimates and obtain the coordinating problem: to find

$$\psi^* = \sup_{u \in R^m} \psi(u).$$

Note that the problem of finding the best Lagrangian bounds ψ^* is one of the main sources of generating the nonsmooth optimization models.

Many combinatorial optimization problems can be formulated as Boolean LP problems, and the corresponding dual bounds may be obtained by LP relaxations of such models. But in some cases the nonlinear quadratic-type formulation of a combinatorial problem is more convenient and may give more exact dual bounds. In these cases, as a rule, the problem of obtaining dual bounds may be reduced to the convex programming problems with nonsmooth matrix function (or to the equivalent problems of semidefinite programming).

Now there exist many methods of nondifferentiable optimization, for example, simple subgradient method, ε-subgradient methods, methods with space transformation. One of the most practically effective modern methods is the algorithm with space dilation in the direction of difference of two successive subgradients (the so-called r-algorithm).

This method was proposed by N.Z. Shor in 1970 [32] for acceleration of convergence of subgradient methods. A family of r-algorithms contains different realizations of subgradient-type methods with space dilations in the direction of difference of two successive subgradients. Below we give a general scheme of r-algorithm.

Denote by $R_\alpha(\xi)$ the linear operator of space dilation in the direction ξ, $\|\xi\| = 1$, with coefficient $\alpha,\ \alpha \geqslant 0$, specifying for each $x \in E^n$ the vector $y \in E^n$ due to the formula:

$$y = R_\alpha(\xi)x = x + (\alpha - 1)(x, \xi)\xi.$$

Let $f(x)$ be a minimized convex function defined on E^n, $x_0 \in E^n$ be a given point. Denote by $g_f(\overline{x})$ a subgradient of function f in the point $\overline{x}$. The general scheme of r-algorithm is following:

The first step

$$x_1 = x_0 - h_0 \frac{g_f(x_0)}{\|g_f(x_0)\|},$$

where h_0 is the step multiplier such, that $(g_f(x_0), g_f(x_1)) \leqslant 0$. Fix $g_f(x_0)$, $g_f(x_1)$, $B_0 = I_n$ ($n \times n$ identity matrix). After k steps we have $x_1, x_2, \ldots, x_k$ and fix $x_k,\ g_f(x_{k-1})$ and $n \times n$ matrix B_k.

$(k+1)$**-st step.**
Calculate:
(a) $g_f(x_k)$;
(b) $r_k = g_f(x_k) - g_f(x_{k-1})$;
(c) $\xi_k = \frac{B_{k-1}r_k}{\|B_{k-1}r_{k-1}\|}$;
(d) $B_k = B_{k-1}R_{\beta_k}(\xi_k),\ 0 < \beta_k < 1$;
(e) $x_{k+1} = x_k - h_k B_k \frac{B_k^T g_f(x_k)}{\|B_k^T g_f(x_k)\|}$.
If the stopping criteria is not fulfilled, fix $x_{k+1},\ g_f(x_k)$, B_k and go to the next step.

Comments. After k steps of r-algorithm let A_{k-1} be the resulting matrix of space transformation: $y = A_{k-1}x$, or $x = B_{k-1}y$, where $B_{k-1} = A_{k-1}^{-1}$, and $\varphi_k(y) = f(B_{k-1}y)$. Since $g_{\varphi_k}(y) = B_{k-1}^T g_f(x)$, r_k is the difference of two subgradients of function φ_k taken in the points $y_k = A_{k-1}x_k$ and $\tilde{y}_{k-1} = A_{k-1}x_{k-1}$. So ξ_k is the normalized direction of the difference of two successive subgradients of transformed function $\varphi_k(y)$. In this direction we make a current dilation of transformed space and obtain resulting matrix $A_k = R_{\alpha_k}(\xi_k)A_{k-1}$. The inverse matrix $B_k = B_{k-1}R_{\beta_k}(\xi_k),\ \beta_k = 1/\alpha_k$. Consider $\varphi_{k+1}(y) = f(B_k y)$, and use subgradient step for function $\varphi_{k+1}(y)$ from $\tilde{y}_k = B_k x_k$:

$$y_{k+1} = \tilde{y}_k - h_k \frac{B_k^T g_f(x_k)}{\|B_k^T g_f(x_k)\|}.$$

In original space the point $x_{k+1} = B_k y_{k+1}$ corresponds to the point x_{k+1}, and we obtain

$$x_{k+1} = x_k - h_k B_k \frac{B_k^T g_f(x_k)}{\|B_k^T g_f(x_k)\|}.$$

Thus we calculate $x_{k+1},\ g_f(x_k),\ B_k$, and are ready to make the next iteration.

The family of r-algorithms has two sequences of parameters: $\{\beta_k\}_{k=1}^{\infty}$ and $\{h_k\}_{k=1}^{\infty}$. Naturally, we must think about rational choice of these sequences to obtain 'good' convergence to the optimal point of $f(x)$.

For minimization of nonsmooth convex functions defined on E^n we recommend to use the following specifications of r-algorithm:

The space dilation coefficients α_k equal to α, where $\alpha \in [2, 4]$. To determine a step multiplier h_k we use adaptive technique of step length regulation (see [35]) determined by parameters: $h_0^{(0)}$ (initial step-length), integer number $\overline{m} > 1$, and coefficients $q_1 < 1$ and $q_2 > 1$ for decreasing (increasing) of step-multiplier. After k iterations of r-algorithm we obtain step constant h_k^0. On $(k+1)$-st iteration we choose the direction of descent due to r-algorithm and move in this direction with a step multiplier h_k^0 until the condition of stopping the search along the direction is fulfilled or the number of steps would be equal to $\overline{m}$. In the last case we continue descent along the same direction with a new step constant $h_k^1 = q_2 h_k^0$. If after $\overline{m}$ steps the condition of interrupting of search direction is not fulfilled, we set $h_k^2 = q_2 h_k^1$ and so on.

We suppose that $\lim_{\|x\|\to+\infty} f(x) = +\infty$, so after finite number of steps the stopping condition for directional search will be fulfilled.

The details of such way of regulating step-multiplier one may find in Section 3 of [33].

The results of testing of r-algorithms show that if the errors of rounding are not essential, the objective function values as a rule may be majored by a geometrical progression of the form $Cq^{\frac{k}{n}}$, where k is the current number of steps and $q = \frac{1}{2}$. So, as a rule the convergence of r-algorithm is approximately $2n$ times faster than of well known ellipsoid methods.

The most typical examples of nonsmooth functions are maximal and minimal eigenvalues of symmetric matrices and sums of k largest eigenvalues (for example, see [23]).

Let Σ_n be the class of real $n \times n$ symmetric matrices. Any matrix $A \in \Sigma_n$ has n real eigenvalues (with account of their multiplicity) and a pair of eigenvectors associated with two different eigenvalues are orthogonal. Let $A \in \Sigma_n$, $A = \{a_{ij}\}_{i,j=1}^n$,

$$\lambda_1(A) \geqslant \lambda_2(A) \geqslant \ldots \geqslant \lambda_n(A)$$

be the eigenvalues of A, ordered nonincreasingly.

A symmetric real matrix is called positive definite (semidefinite) if $\lambda_n(A) > 0$ ($\lambda_n(A) \geqslant 0$). We shall write $A \succ 0$ ($A \succeq 0$), if $A \in \Sigma_n$ and A is positive definite (semidefinite).

The Rayleigh–Ritz formula is known for the maximal eigenvalue $\lambda_1(A)$:

$$\lambda_1(A) = \max_{\|y\|=1} (Ay, y) = \max_{\|y\|=1} \sum_{i,j=1}^{n} a_{ij} y_i y_j, \tag{4}$$

where $y = (y_1, \ldots, y_n) \in E^n$. $\lambda_1(A)$ is a convex function defined on Σ_n since formula (4) gives representation of this function as a maximum of a family of linear functions in entries $\{a_{ij}\}_{i,j=1}^n$.

Denote by $Y^*(A)$ a set of normalized vectors y, which give maximum in (4), i.e., $\lambda_1(A) = (Ay^*, y^*)$ for all $y^* \in Y^*(A)$. From (4), one may obtain the subgradient set $G_{\lambda_1}(\overline{A})$ of function $\lambda_1(\cdot)$ in the point $\overline{A}$:

$$G_{\lambda_1}(\overline{A}) = conv\left\{\cup_{y\in Y^*(\overline{A})} yy^T\right\}.$$

The calculation of subgradient $g_{\lambda_1}(\overline{A}) \in G_{\lambda_1}(\overline{A})$ may be reduced to finding an arbitrary $y^*(\overline{A}) \in Y^*(\overline{A})$ and applying the next formula:

$$g_{\lambda_1}(\overline{A}) = \left\{y^*(\overline{A})[y^*(\overline{A})]^T\right\}.$$

(Note that yy^T, $y \in E^n$, is a symmetric matrix of rank 1 with entries $\{y_i y_j\}_{i,j=1}^n$). If $\lambda_1(\overline{A})$ has multiplicity 1, then $g_{\lambda_1}(\overline{A})$ is unique and function $\lambda_1(A)$ is differentiable at the point $\overline{A}$. When multiplicity of $\lambda_1(\overline{A})$ is more than 1, the function $\lambda_1(A)$ is nondifferentiable at $\overline{A}$.

Now we introduce a very interesting class of convex matrix functions defined on symmetric matrices $A \in \Sigma_n$, namely, the sums of the k largest eigenvalues:

$$S_{n,k}(A) = \sum_{r=1}^{k} \lambda_r(A),\ \ 1 \leqslant k \leqslant n.$$

Famous mathematician Fan Ky gave in 1949 variational description of $S_{n,k}(A)$ that is a far going generalization of Rayleigh–Ritz formula (4). Let M_n^k be the class of rectangular $n \times k$ matrices Y, the columns $y_i,\ \ i = 1, \dots, k,\ \ k \leqslant n$ form an orthonormal‘ system of n-dimensional vectors, i.e. $Y^T Y = I_k$ (I_k is $k \times k$ identity matrix).

THEOREM 1. (Fan Ky [17]).

$$S_{n,k}(A) = \max_{Y \in M_n^k} \{tr(AYY^T)\},\ \ \forall A \in \Sigma_n. \tag{5}$$

The maximum in formula (5) is reached at orthonormal system of eigenvectors $y_1^*, \dots, y_k^*$ corresponding to eigenvalues $\lambda_1(A), \dots, \lambda_k(A)$. Indeed,

$$Ay_i^* = \lambda_i(A) y_i^*,\ \ i = 1, \dots, k,$$

$$\mathrm{tr}(AYY^T) = (A, YY^T) = \sum_{i=1}^{k} (\lambda_i(A) y_i^*, y_i^*) =$$

$$= \sum_{i=1}^{k} \lambda_i(A) = S_{n,k}(A).$$

When $k = 1$, formula (5) is reduced to the expression (4).

Consider the class of $n \times n$ matrices

$$C_n^k = \mathrm{conv}\{YY^T : Y \in M_n^k\}.$$

THEOREM 2. *C_n^k coincides with the class of all positive semidefinite matrices C with $\lambda_1(C) \leqslant 1$ and with trace, equal to k [17].*

Due to this theorem, we obtain a new variational formula for $S_{n,k}(A)$:

$$S_{n,k}(A) = \max_{C \in C_n^k} (A, C), \tag{6}$$

where $(A, C) = \sum_{i,j=1}^{n} a_{ij} c_{ij}$ ($A = \{a_{ij}\}_{i,j=1}^{n},\ \ C = \{c_{ij}\}_{i,j=1}^{n}$).

Formulas (5) and (6) give us the representation of the function $S_{n,k}(A)$ as a pointwise maximum function on infinite family of linear (in matrix variable $A \in$

Σ_n) functions. So, $S_{n,k}(A)$ is a convex function for any $n, k \leqslant n$. The structure of subgradient set $G_{S_{n,k}}(A)$ is determined by (6):

$$G_{S_{n,k}}(A) = \{C^* \in C_n^k : S_{n,k}(A) = (A, C^*)\}$$

For almost all A, C^* is unique and gives us the gradient $g_{S_{n,k}}(A)$ of $S_{n,k}$ in the point A.

The most strict way for obtaining a subgradient $g_{S_{n,k}}(A)$ is the following:

(i) solve the eigenvalue problem for matrix A and find the eigenvalues $\lambda_1(A)$, $\dots, \lambda_k(A)$ and the corresponding orthonormal system of eigenvectors $Y_j(A) = \{y_i^j\}_{i=1}^n,\ j = 1, \dots, k$;

(ii) construct the $n \times k$ matrix $Y = \{y_i^j\}_{i=1,\dots,n}^{j=1,\dots,k}$;

(iii) set $\overline{g}_{S_{n,k}}(A) = YY^T$.

In general, when not all k largest eigenvalues of A have multiplicity 1, the subgradient $g_{S_{n,k}}$ is not unique, because the system of eigenvectors $\{Y_i(A)\}_{i=1}^k$ is determined nonuniquely in this case. But if one is interested in calculating any subgradient from $G_{S_{n,k}}(A)$ one can use the procedure described above for arbitrary orthonormal system $Y(A)$ of eigenvectors associated with the k largest eigenvalues $\lambda_1(A) \geqslant \lambda_2(A) \geqslant \dots \geqslant \lambda_k(A)$.

Let A be a diagonal matrix and $a_{11} \geqslant a_{22} \geqslant \dots \geqslant a_{kk} \geqslant \dots \geqslant a_{nn},\ 1 \leqslant k \leqslant n$. Consider two cases:

(I) $a_{kk} > a_{(k+1)(k+1)}$. In this case in formula:

$$S_{n,k}(A) = \max_{C \in C_n^k}(A, C) = (A, C^*).$$

C^* is determined uniquely (C^* is a diagonal matrix with $c_{ii} = 1$ for $i \leqslant k$ and $c_{ii} = 0$ for $i > k$). The subgradient of the function $S_{n,k}$ in the point $A,\ g_{S_{n,k}}(A)$ is equal to C^*, and the function $S_{n,k}$ is differentiable at A;

(II) $a_{kk} = a_{(k+1)(k+1)}$. In this case the subgradient set $G_{S_{n,k}}(A)$ contains more than one extremal point. For example, if $a_{ii} = a_{kk}$ for all $i,\ k-s \leqslant i \leqslant k+p$ ($s \geqslant 0,\ p \geqslant 1$), then an arbitrary diagonal matrix A with properties:

(1) $a_{ii} = 1$ for $i < k - s$;

(2) $a_{ii} = 0$ for $i > k + p$;

(3) the set of values $\{a_{ii}\},\ k - s \leqslant i \leqslant k + p$ contains exactly s ones and p zeroes;

is an extreme point of $G_{S_{n,k}}(A)$. So, in the case (II) the function $S_{n,k}$ is nondifferentiable at A.

In general, if for a symmetric matrix $A,\ \lambda_k(A) > \lambda_{k+1}(A)$ then $S_{n,k}$ is differentiable at A; otherwise (i.e., $\lambda_k(A) = \lambda_{k+1}(A)$) the function $S_{n,k}$ is nondifferentiable at A.

In many applications we meet with a weighted sum of k largest eigenvalues:

$$S_{n,k}(A, w) = \sum_{i=1}^k w_i \lambda_i(A), \text{ where } w = (w_1, \dots, w_k) \geqslant 0.$$

LEMMA 1. *If* $w_1 \geqslant w_2 \geqslant \ldots \geqslant w_k$, *then* $S_{n,k}(A, w)$ *is a convex function defined on* Σ_n.

The weighted sum of the largest eigenvalues $S_{n,k}(A, w)$ can be represented also by the variational formula similar to formula (5):

$$S_{n,k}(A, w) = \max_{Z \in M_n^k(w)} \{\mathrm{tr}(AZZ^T)\},$$

where $M_n^k(w)$ is the class of rectangular $n \times k$ matrices Z, the columns Z_i forming an orthogonal system of n-dimensional vectors, and $\|Z_i\|^2 = w_i$, $i = 1, \ldots, k$, $w_1 \geqslant w_2 \geqslant \ldots \geqslant w_k$.

Due to expression (5) of Theorem 1 the subgradient set for the function $S_{n,k}(\cdot)$ at point X is given by the following expression:

$$G_{S_{n,k}}(X) = conv\left(\sum_{i=1}^{k} y_i y_i^T\right),$$

where y_i, $i = 1, \ldots, k$, form an arbitrary orthonormal system of vectors, associated with eigenvalues $\lambda_1(X), \ldots, \lambda_k(X)$. If multiplicity of all eigenvalues $\lambda_1(X), \ldots, \lambda_k(X)$ is equal to one, then the matrix $\sum_{i=1}^k y_i y_i^T$ is determined uniquely and coincides with gradient $S_{n,k}(X)$ at point X.

Functions $\lambda_m(X)$, $1 < m < n$ are quasi-differentiable functions (in the sense of Demjanov and Rubinov [11]). They may be considered as a difference of two convex functions

$$\lambda_m(X) = S_{n,m-1}(X) - S_{n,m}(X).$$

If $w = \{w_1, \ldots, w_n\} \geqslant 0$ and $w_k \geqslant w_{k+1}$ for $k = 1, \ldots, n-1$, then subgradient set of convex function $S_n^w(X) = \sum_{i=1}^n w_i \lambda_i(X)$ can be represented by the following expression:

$$G_{S_n^w}(X) = conv\left\{\sum_{i=1}^{n} w_i y_i y_i^T\right\},$$

where $\{y_i\}_{i=1}^n$ is any orthonormal system of eigenvectors of matrix X (each y_i is associated with $\lambda_i(X)$). If all eigenvectors $y_i(X)$ with $w_i > 0$ have multiplicity 1, then S_n^w is differentiable at X.

2. Quadratic-type minimization problems, Lagrangian lower bounds and superfluous constraints

Consider now the problems of finding Lagragian lower bounds for quadratic-type optimization models:

$$\text{Find } Q^* = \inf_{x \in E^n} Q_0(x) \quad \text{subject to} \quad Q_i(x) = 0,\ i = 1, \ldots, m, \tag{7}$$

where $Q_\nu(x)$, $\nu = 0, \dots, m$ are quadratic or linear functions, determined on n-dimensional Euclidean space E^n. If the problem (7) has no feasible solution, we set $Q^* = +\infty$. Using usual Lagrange function $L(x,u)$, $u = \{u_1, \dots, u_m\}$, one can obtain Lagrangian lower bounds for such problems by finding $\psi^* = \sup \psi(u)$, where $\psi(u) = \inf\limits_{x \in E^n} L(x,u)$. If $dom\psi$ is nonempty, then $\psi(u)$ is a proper concave function. In opposite case we obtain a trivial bound $\psi^* = -\infty$.

Consider the problem of finding dual (Lagrangian) estimates for quadratic-type problems of the form (7) in more detail. Let quadratic functions $Q_\nu(x)$, $\nu = 0, \dots, m$, have the following description:

$$Q_\nu(x) = (K_\nu x, x) + (c_\nu, x) + d_\nu,$$

where K_ν are symmetric quadratic $n \times n$ matrices, c_ν are n-dimensional vectors, d_ν are numbers. So, $(K_\nu x, x)$ is the quadratic part of $Q_\nu(x)$ and (c_ν, x) is the linear part of $Q_\nu(x)$, $\nu = 0, \dots, m$.

Usual Lagrange function $L(x,u)$ can be represented as

$$L(x,u) = Q_0(x) + \sum_{i=1}^{m} u_i Q_i(x) = (K(u)x, x) + (c(u), x) + d(u),$$

where

$$K(u) = K_0 + \sum_{i=1}^{m} u_i K_i;$$

$$c(u) = c_0 + \sum_{i=1}^{m} u_i c_i;$$

$$d(u) = d_0 + \sum_{i=1}^{m} u_i d_i;$$

(here $u = \{u_1, \dots, u_m\}$ is m-dimensional vector of Lagrange multipliers).

Consider $\psi(u) = \inf\limits_x L(x,u)$. If $K(u)$ is a positive definite matrix then $\psi(u) = L(x(u), u)$, where $x(u)$ is a solution of linear system of equations

$$2K(u)x + c(u) = 0,$$

i.e.

$$x = -\frac{1}{2}\Big(K(u)\Big)^{-1} c(u).$$

If the minimal eigenvalue of $K(u)$ $\lambda_n\Big(K(u)\Big) < 0$, then $\psi(u) = -\infty$.

In the case $\lambda_n\Big(K(u)\Big) = 0$ matrix $K(u)$ is positive semidefinite but singular. Let $s_1, \ldots, s_n$ be the orthonormal basis in E^n, corresponding to eigenvectors ordered in decreasing order of their eigenvalues (with taking in account their multiplicity). Then $c(u)$ can be represented in the form:

$$c(u) = \sum_{i=1}^{n-r} \alpha_i(u)s_i + \sum_{i=n-r}^{r} \alpha_i(u)s;$$

where r is the multiplicity of the minimal eigenvalue equal to zero. If all $\alpha_k(u) = 0$ for $n - r + 1 \leqslant k \leqslant n$, then $x(u)$ exists, i.e. $u \in \mathrm{dom}\psi$. Otherwise $\psi(u) = -\infty$.

Let $\mathrm{dom}\psi$ be non-empty. Then ψ is a proper concave function as a result of minimization with respect to x of functions

$$\psi_u(x) = Q_0(x) + \sum_{i=1}^{m} u_i Q_i(x),$$

which are linear by u for all x.

For each $u \in \mathrm{dom}\psi$ $\psi(u) \leqslant Q^*$ (Q^* is an optimal value of initial problem (7)).

Let $\overline{u} \in E^m$ and $K(\overline{u}) \succ 0$. Then $Q(\overline{u})$ is an interior point of $\mathrm{dom}\psi$. Denote by Ω^+ the set

$$\{u \in E^m \; / \; K(u) \succ 0\}.$$

Boundary $\overline{\Omega}$ of Ω^+ consists of u for which $K(u)$ has the minimal eigenvalue equal to zero. It is easy to prove that all $u \in \overline{\Omega}$ are limit points of the set Ω^+. So if Ω^+ is nonempty then $\mathrm{dom}\psi$ is closed in E^m.

Let $\psi^* = \sup\limits_{u \in E^m} \psi(u)$ and there exists a point u^* such that $\psi^* = \psi(u^*)$. If $u^* \in \Omega^+$ then $g_\psi(u^*) = Q_i(x^*) = 0,\ i = 1, \ldots, m$. In this case $\psi^* = Q^*$. Otherwise, when $x^* \in \overline{\Omega}$, it may be a positive defect of duality:

$$\Delta = Q^* - \psi^* > 0.$$

For many interesting problems we may improve the dual bounds by using the so-called functionally superfluous constraints in the form of quadratic equations (inequalities) which do not change the optimal value of initial polynomial problem but lead to modification of Lagrange function of corresponding quadratic problem. This modification may give substantial increasing of new Lagrangian bound for a modified quadratic type problem in comparison with ψ^* for the old one.

Note that if we add to initial problem (7) new quadratic superfluous constraints $Q_{m+1}(x) = 0, \ldots, Q_{m+r}(x) = 0$, $r \geqslant 1$ and form a longer vector of Lagrange multipliers $U = \Big\{\{u\}, u_{m+1}, \ldots, u_{m+r}\Big\}$ then for the problem:

$$Q^* = \inf_{x \in E^n} Q_0(x) \quad \text{subject to} \quad Q_i(x) = 0,\ i = 1, \ldots, m, m+1, \ldots, m+r$$

the corresponding Lagrange function will be

$$L_1(x, U) = Q_0(x) + \sum_{i=1}^{m+r} u_i Q_i(x) = L(x, u) + \sum_{i=m+1}^{m+r} u_i Q_i(x).$$

So

$$L(x, u) = L_1\Big(x, (\{u\}, 0, \dots, 0)\Big), \psi_1(U) = \inf_x L_1(x, U)$$
$$\geqslant \inf L(x, u) = \psi(u),$$

and

$$\psi_1^* = \sup \psi_1(U) \geqslant \psi^*.$$

We shall demonstrate the possible improving of dual bounds by introducing superfluous constraints on a simple example.

Example. Let $P_6(x_1)$ be a sixth-degree polynomial of one variable x_1:

$$p_6(x_1) = x_1^6 + a_5 x_1^5 + a_4 x_1^4 + a_3 x_1^3 + a_2 x_1^2 + a_1 x_1 + a_0.$$

The problem is to find value p^* of (global) minimum of $p_6(x_1)$. One may transform this problem in a quadratic-type problem by introducing new variables $x_2 = x_1^2$; $x_3 = x_1 x_2$.

Consider a vector of variables $x = \{x_1, x_2, x_3\}$ and obtain the equivalent problem:
to minimize

$$Q_0(x) = x_3^2 + a_5 x_2 x_3 + a_4 x_3 x_1 + a_3 x_3 + a_2 x_2 + a_1 x_1 + a_0 \tag{8}$$

subject to constraints:

$$Q_1(x) = x_1^2 - x_2 = 0, \tag{9}$$

$$Q_2(x) = x_1 x_2 - x_3 = 0. \tag{10}$$

The Lagrange function $L(x, u)$ of this problem has the following form ($u = \{u_1, u_2\}$ is a vector of Lagrange multipliers):

$$L(x, u) = Q_0(x) + u_1 Q_1(x) + u_2 Q_2(x).$$

Consider the matrix $K(u)$ which defines the quadratic part of $L(x, u)$.

$$K(u) = \begin{pmatrix} u_1 & \frac{u_2}{2} & \frac{a_4}{2} \\ \frac{u_2}{2} & 0 & \frac{a_5}{2} \\ \frac{a_4}{2} & \frac{a_5}{2} & 1 \end{pmatrix}.$$

One may see that if $a_5 \neq 0$, for all $u = \{u_1, u_2\}$, $K(u)$ cannot be positive semidefinite, and we obtain a trivial lower bound $\psi^* = -\infty$. So, it seems an attempt to use dual quadratic bounds for our example failed. But if we add to our model a superfluous constraint

$$Q_3(x) = x_2^2 - x_1 x_3 = 0 \tag{11}$$

and modify respectively the Lagrange function, we radically change the situation.

New Lagrange function L_1 has 3 Lagrange multipliers

$$u^{(1)} = \{u_1, u_2, u_3\}$$

and

$$L_1(x, u^{(1)}) = L(x, u) + u_3(x_2^2 - x_1 x_3)$$

and $K(u)$ changes for

$$K_1(u^{(1)}) = \begin{pmatrix} u_1 & \frac{u_2}{2} & \frac{a_4-u_3}{2} \\ \frac{u_2}{2} & u_3 & \frac{a_5}{2} \\ \frac{a_4-u_3}{2} & \frac{a_5}{2} & 1 \end{pmatrix}.$$

It is easy to show that if we choose $u_3 > \frac{a_5^2}{4}$ and u_1 large enough to make $\det\left(K_1(u^{(1)})\right) > 0$, the matrix $K_1(u^{(1)})$ becomes positive definite, so the function $\psi_1(u^{(1)}) = \inf_x L_1(x, u^{(1)})$ has nonempty domain, and we obtain nontrivial Lagrangian bound ψ_1^*. Moreover, we show later that this bound is exact, i.e. $\psi_1^* = p_6^*$.

3. Quadratic-type problems for finding global minimum of polynomials

Consider a more general problem.

Let $P_{2m}(x_1)$ be a polynomial of one variable x_1 of even degree $2m$ with the eldest coefficient 1:

$$P_{2m}(x_1) = x_1^{2m} + \sum_{i=1}^{2m} a_{2m-i}\, x_1^{2m-i}.$$

Introduce variables $x_r = x_1^r$, $r = 1, \ldots, m$, and represent all monomials x_1^k, $k = 1, \ldots, 2m$, as a product of no more than two monomials x_1^r, $0 < r \leqslant m$. We call these representations as feasible. For some k such representations may be nonunique. So we use the so-called 'standard' representation:

$$x_1^k = \begin{cases} x_k & \text{for } k \leqslant m, \\ x_{k-m} x_m & \text{for } m \leqslant k \leqslant 2m. \end{cases} \tag{12}$$

By using (12) we obtain the ‘standard’ representation of polynomial $P_{2m}(x_1)$ as a quadratic function of variables x_r, $r = 1, \dots, m$:

$$P_{2m}(x_1) = K_0(x_1, \dots, x_m) = x_m^2 + \sum_{k=1}^{m-1} a_{2m-k} x_{m-k} x_m + \sum_{k=m}^{2m} a_{2m-k} x_{2m-k}. \tag{13}$$

Consider all possible nonstandard representations of monomials x_1^k, $1 < k \leqslant 2m$. One may obtain the full system of quadratic-type equalities in the form:

$$\pi_{ks} = x_k - x_s x_{k-s} = 0, \quad 1 \leqslant s \leqslant \frac{k}{2}, \quad 2 \leqslant k \leqslant m. \tag{14}$$

$$\pi_{kt} = x_{k-m} x_m - x_t x_{k-t} = 0, \quad t \leqslant k - t < m, \quad 2m - 2 > k > m. \tag{15}$$

For example, if $m = 3$, we obtain the constraints $x_2 - x_1^2 = 0$; $x_3 = x_1 x_2$; $x_1 x_3 - x_2^2 = 0$.

Thus, we have reduced the problem of finding a global minimum value P_{2m}^* of polynomial $P_{2m}(x_1)$ to quadratic-type problem: minimize $K_0(x_1, \dots, x_m)$ subject to constraints of the form (14), (15). It is easy to determine that the number of basic equations necessary to convert the initial problem of finding P_{2m}^* in equivalent quadratic-type problem is $(m - 1)$:

$$x_k = x_1 x_{k-1}; \quad 2 \leqslant k \leqslant m.$$

The other constraints from (14), (15) are superfluous. The number of all constraints from (14), (15) equals to $\frac{m(m-1)}{2}$.

Consider the Lagrange function for the quadratic-type problem (13), (14), (15). Denote by u_{ks} the Lagrange multipliers corresponding to equations $\pi_{ks} = 0$ in (14) and by u_{kt} to the equations $\pi_{kt} = 0$, from (15). Let us form the usual Lagrange function $L_{2m}(x_m, u_m)$ for the quadratic-type problem (13), (14) and (15), where x_m denotes the vector $(x_1, \dots, x_m)$, u_m denotes the vector of Lagrange multipliers corresponding to constraints from (14), (15). Let’s formulate the main Theorem.

THEOREM 3. *The best Lagrangian bound* $\psi_m^* = \sup_{u_m} \psi(u_m)$, *where* $\psi(u_m) = \inf_{x_m} L_m(x_m, u_m)$ *is equal to* $P_2 m^*$, *i.e. this bound is exact.*

To prove this fact we use the following Lemma.

LEMMA 2. *The nonnegative polynomial*

$$\overline{P}_{2m}(x_1) = P_{2m}(x_1) - P_{2m}^*$$

can be represented as a sum of squares of real polynomials of degrees not exceeding m.

Proof. All complex roots of equation $\overline{P}_{2m}(x_1) = 0$, if they exist, form pairs of the form $a_\alpha + ib_\alpha$ and $a_\alpha - ib_\alpha$, here a_α and b_α are real numbers. The real roots β_r, $r = 1, \dots, k$, must have the even multiplicity $2\delta_r$ otherwise $\overline{P}_{2m}(x_1)$ cannot be positive.

Let

$$f_1(x_1) = \prod_\alpha [x_1 - (a_\alpha + ib_\alpha)] = R(x_1) + iQ(x_1),$$
$$f_2(x_1) = \prod_\alpha [x_1 + (a_\alpha - ib_\alpha)] = R(x_1) - iQ(x_1),$$

where R and Q are real polynomials.

Then

$$\overline{P}_{2m}(x_1) = f_1(x_1) f_2(x_1) \left[\prod_r (x_1 - \beta_r)^{2\delta_r}\right] =$$
$$= \left[R^2(x_1) + Q^2(x_1)\right]\left[\prod_r (x_1 - \beta_r)^{\delta_r}\right]^2.$$

So the nonnegative polynomial $\overline{P}_{2m}(x_1)$ can be represented as a sum of squares of real polynomials of degree not exceeding m. The proof is over.

Instead of monomials $x_i = x_1^i$, $i = 1, \dots, m$, consider another system of basic polynomial functions which corresponds to moving of origin of x_1 by constant h:

$$z_i = (x_1 + h)^i; \quad i = 1, \dots, m; \quad z_0 = 1.$$

We want to expose x_1^r, $i = 1, \dots, 2m$, as quadratic (or linear) functions of $z_1, \dots, z_m$. Let us use induction by r. For $r = 2$

$$x_1^2 = (x_1 + h)^2 - 2h(x_1 + h) + h^2 = z_2 - 2hz_1;$$

Denote by $Q_k(z_1, \dots, z_m)$ the representations of x_1^k ($k = 1, \dots, 2m - 1$) as a quadratic function of variables $x_1, \dots, x_m$. Then

$$Q_{k+1}(x_1, \dots, x_m) = x_1 Q_k(z_1, \dots, z_m) = z_1 Q_k(z_1, \dots, z_m) \\ - hQ_k(z_1, \dots, z_m).$$

One may generate the full system of quadratic equalities for monomials z_i^r, $2 \leqslant r \leqslant 2m$, similar to (14), (15):

$$z_k - z_r z_{k-r} = 0 \text{ for } k \leqslant m, \quad r \leqslant \frac{k}{2}; \tag{16}$$

$$z_{k-m} z_m - z_r z_{k-r} = 0 \text{ for } m < k \leqslant 2m; \quad r \leqslant \frac{k}{2}; \quad k - r \neq m. \tag{17}$$

When we use the recurrent formula (16) some monomials after multiplication by z_1 may possess nonstandard form and we must add the equalities of the form (16) or (17) with corresponding multipliers to obtain the expressions for $x_1^{(k)}$, $k = 1, \dots, 2m$, as a quadratic function of variables $z_1, \dots, z_m$ in standard form.

If we substitute these expressions for x_1^k in polynomial $P_{2m}(x_1)$ given in standard form we obtain quadratic function $K_0^{(h)}(z_1, \dots, z_m)$ representing the polynomial $P_{2m}(x_1 + h)$.

Thus $K_0^{(h)}(z_1, \dots, z_m)$ coincides with Lagrange function $L(z_1, \dots, z_m; \overline{U}_h(P_{2m}))$, where $\overline{U}_h(P_{2m})$ is the vector of Lagrange multipliers for equalities (16) and (17), dependent on moving h and coefficients of initial polynomial $P_{2m}(x_1)$ to be minimized. From previous discussion one may obtain the following Lemma.

LEMMA 3. *For an arbitrary fixed moving h and a polynomial $P_{2m}(x_1)$ there exists a vector of Lagrange multipliers $\overline{U}_h(P_{2m})$ such that*

$$L_1\left(x_1, \dots, x_m;\ U - \overline{U}_h(P_{2m})\right) = L\left(x_1, \dots, x_m;\ U)\right),$$

where L_1 is Lagrange function for quadratic representation of polynomial $P_{2m}(x_1 + h)$.

COROLLARY 1. *The best Lagrangian quadratic bounds $\psi^*(h)$ for polynomial $P_{2m}(x_1 + h)$ is equal to ψ^*.*

We say that a polynomial $P_{2m}(x_i)$ possesses E-property if the best Lagrangian quadratic bound of it $\psi^* = \min\limits_{x_1} P_{2m}(x_i) = P^*_{2m}$. It is obvious.

LEMMA 4. *If the polynomial $P_{2m}(x_1)$ of degree $2m$ with the eldest coefficient 1 possesses E-property then for arbitrary $h \in R$ the polynomial $P_{2m}(x_1 + h)$ also possesses E-property.*

Let the point $x_1^* \in E^n$ be a point of global minimum of polynomial $P_{2m}(x_1)$ and $P_{2m}(x_1^*) = P^*_{2m}$. Consider the polynomial $\overline{P}_{2m}(x_1) = P_{2m}(x_1 - x^*) - P^*_{2m}$. The nonnegative polynomial $\overline{P}_{2m}(x_1)$ has its global minimum value 0 in the point $x_1 = 0$. By Lemma 2 polynomial $\overline{P}_{2m}(x_1)$ can be decomposed into a sum of squares of real polynomials of degree not exceeding m.

$$\overline{P}_{2m}(x_1) = \sum_{i=1}^{N} \left[P_m^{(i)}\right]^2. \tag{18}$$

Since $\overline{P}_{2m}(0) = 0$, each of polynomials $P_m^{(i)}(x_1)$ has no constant part.

Let $\overline{L}^*(x_1, \dots, x_m;\ U)$ be the Lagrange function for quadratic representation of $\overline{P}_{2m}(x_1)$. Set $U = 0$. Then $\overline{L}^*(x_1, \dots, x_m;\ 0) = \overline{P}_{2m}(x_1)$ and can be decomposed into a sum of squares (see (18)). Each of polynomials $P_m^{(i)}$ can be represented as a linear form in variables $x_1, \dots, x_m$, so $\min\limits_x \overline{L}^*(x_1, \dots, x_m;\ 0) = \overline{L}^*(0, 0, \dots, 0;\ 0) = 0$, the best Lagrangian quadratic bound is exact.

From Lemma 4 we obtain that the Lagrangian quadratic bound for initial problem: to find a global minimum for polynomial $P_{2m}(x_1)$ is also exact. The proof of Theorem 3 is over.

For polynomials of several variables the situation is more complicated. Further we shall give a review of main results in the theory of dual quadratic bounds (with using of superfluous constraints) for polynomials of several variables.

Let R^n be n-dimensional linear space of real vectors $x = \{x_1, \dots, x_n\}$, $P(x) = P(x_1, \dots, x_n)$ be a polynomial real function defined on R^n. Consider the problem of finding

$$f^* = \inf_{z \in R^n} P(x_1, \dots, x_n).$$

We will be interested in nontrivial case where $f^* > -\infty$, i.e. $P(x)$ is bounded from below. Such polynomials will be called BB-polynomials. It is clear that if $P(x)$ belongs to BB-class, then for any i, $1 \leqslant i \leqslant n$, the highest degrees S_i of variables x_i must be even. Note that the problem of the BB-property is in general similar by its computational complexity to the problem of finding f^*.

Let $S_i = 2l_i$, $i = 1, \dots, n$, and $P(x)$ be recorded in standard form as a sum of monomials with some real nonzero coefficients. For compact record of monomials we use a vector of degree $\alpha = (\alpha_1, \dots, \alpha_n)$ with nonnegative integer entries and symbols $R[\alpha]$ of the corresponding monomials.

$$R[\alpha] = x_1^{\alpha_1}, \dots, x_n^{\alpha_n}; \quad \alpha_i \leqslant S_i, \quad i = 1, \dots, n.$$

So

$$P(x) = \sum_{\alpha} c_\alpha R[\alpha] \text{ in new variables}, \quad 0 \leqslant \alpha_i \leqslant 2l_i, i = 1, \dots, n. \tag{19}$$

Let all monomials of polynomial $P(x) = P(x_1, \dots, x_n)$ have maximal degree on variable x_i equal to $2l_i$ $(i = 1, \dots, n)$. Consider 'feasible' monomials $R[\alpha] = \prod_{i=1}^n x_i^{\alpha_i}$, where $\alpha_i \leqslant l_i$, $i = 1, \dots, n$.

For each monomial $R[\alpha]$ choose the 'standard' representation of monomial $R[\alpha]$ as a product of two feasible monomials:

$$R[\alpha] = R[\alpha_1(\alpha)] \cdot R[\alpha - \alpha_1(\alpha)],$$

$$\alpha_1(\alpha), (\alpha - \alpha_1(\alpha)) \geqslant 0.$$

Moreover, an integer vector

$$\alpha_1(\alpha) \geqslant \alpha - \alpha_1(\alpha) \geqslant 0,$$

or

$$\alpha_1(\alpha) \geqslant \frac{\alpha}{2}.$$

Consider for each monomial $R[\alpha]$, $\alpha_i \leqslant 2l_i$, $i = 1, \ldots, n$ the full set of quadratic-type equalities in feasible R-variables of the form:

$$R[\alpha_1(\alpha)] \cdot R[\alpha - \alpha_1(\alpha)] - R[\beta] \cdot R[\alpha - \beta] = 0, \tag{20}$$

where β runs all possible values of integer feasible vectors, not equal to $\alpha_1(\alpha)$, such that

$$\beta \geqslant \alpha - \beta \geqslant 0, \text{ so } \beta \geqslant \frac{\alpha}{2}.$$

We obtain the 'full' set of constraints if we record the equations of the type (20) for all $\alpha = (\alpha_1, \ldots, \alpha_n)$, $\alpha_i \leqslant 2l_i$, $i = 1, \ldots, n$.

Using the full family of equalities of the type (20) one can get all possible representations of polynomial $P(x_1, \ldots, x_n)$ as quadratic function in feasible R-variables:

$$\begin{aligned} P(x_1, \ldots, x_n) = L(R, u) &= \sum_{\alpha} c_\alpha R[\alpha_1(\alpha)] \cdot R[\alpha - \alpha_1(\alpha)] + \\ &+ \sum_{\alpha,\beta} u_{\alpha\beta}(R[\alpha_1(\alpha)] \cdot R[\alpha - \alpha_1(\alpha)] - R[\beta] \cdot R[\alpha - \beta]), \end{aligned} \tag{21}$$

where $u_{\alpha\beta}$ are arbitrary multipliers in the left part of equalities (20) with corresponding $\{\alpha, \beta\}$, c_α are coefficients in usual representation of $P(x)$ in (19).

On the other hand, we can consider $L(R, u)$ as a Lagrange function of quadratic-type problem in feasible variables $R(\alpha)$:

to minimize

$$P(x_1, \ldots, x_n) = \sum_{\alpha} c_\alpha \prod_{i=1}^{n} x_i^{\alpha_i} = \sum_{\alpha} c_\alpha (R[\alpha_1(\alpha)] \cdot R[\alpha - \alpha_1(\alpha)])$$

subject to the full set of constraints of the form (20), $u = \{u_{\alpha\beta}\}$ are Lagrange multipliers.

Our aim is to find conditions when

$$\psi^* = \sup_u \left[\inf_R L(R, u) \right] = P^* = \min_x P(x).$$

These conditions are formulated in

Main Theorem. *Let polynomial function in n variables $P(x)$ reaches the global minimum at point x^* and $P(x^*) = P^*$.*

Then a dual quadratic bound $\psi^* = P^*$ if and only if the nonnegative polynomial $\overline{P}(x) = P(x) - P^*$ can be represented as a sum of squares of polynomials, which have only feasible monomials.

Proof of the Main Theorem. We say that a polynomial $P(x)$, $\min_{x\in E^n} P(x) = P(x^*) = P^*$, possesses E-property if Lagrangian quadratic bound is exact i.e. $\psi^* = P^*$. To prove the Main Theorem we use the next

Lemma. *If polynomial $P(x)$, $x \in E^n$ possesses E-property then for arbitrary $a \in E^n$ the polynomial $P_a(x) = P(x + a)$ also possesses this property (see [33]).*

Due to the previous Lemma one can suppose without loss of generality that optimal point is $x^* = 0$.

Let BB-polynomial $\overline{P}(x) - P^*$ is represented as a sum of squares of real polynomials $R_i(x)$, $i = 1, \ldots, k$, i.e.

$$\overline{P}(x) = \sum_{i=1}^{k} [R_i(x)]^2.$$

Replace each monomial M_{α^i} contained in polynomial $R_i(x)$ by the corresponding feasible variable $R[\alpha^{(i)}]$. The output of $[R_i(x)]^2$ for each i can be represented as a sum of monomials of the form $C_s^{(i)} C_t^{(t)} R[\alpha^{(it)}] R[\alpha^{(is)}]$. After summing the similar terms, one obtains for each possible $R[\alpha]$ the corresponding coefficient

$$C_\alpha = \sum_i \sum_{(s,t)} C_s^{(i)} C_t^{(i)},$$

where pairs (s, t) are such that $\alpha^{(i,s)} + \alpha^{(i,t)} = \alpha$.

Choose Lagrange multipliers in expression for $L(R, u)$ (21) equal to zero. Then the value of corresponding Lagrange function $L(x, u)$ coincides with the objective function. Since the objective function minus P^* is a sum of squares, the corresponding representation of a quadratic part in $R[\alpha]$ variables is positive semidefinite, so $\overline{u} = 0$ belongs to dom ψ. $x^*(0)$ is a solution of linear system of equations in variables $R[\alpha]$:

$$Q_i[R] = 0, \quad i = 1, \ldots, m.$$

The optimal value of the function $\overline{P}(x)$ equals to zero, i.e. $\overline{P}(0)$. So, due to the Lemma, $P(x)$ can be represented as a sum of squares of real polynomials. $P(x)$ possesses E-property.

Continue the proof of the Main Theorem. Let polynomial $P(x)$ possesses E-property and $\min_{x\in E^n} P(x) = P(x^*) = P^*$. This polynomial $\overline{P}(x) = P(x) - P^*$ takes its minimum at the same point x^* and $\overline{P}(x^*) = 0$. Since $P(x)$ possesses E-property there exists u^* such that for the corresponding quadratic-type problems in variables $R[\alpha]$ Lagrange function $L(R, u)$, when $u = u^*$ is positive-semidefinite in R-variables, so it can be represented as a sum of squares of linear functions in feasible R-variables

$$L(R, u^*) = \sum_{i=1}^{k} \{l_i^{u^*}[R]\}^2.$$

Instead of each term of such defined linear functions in $R[\alpha]$ variables one can substitute corresponding possible feasible monomials in x variables. So we obtain the representation $\overline{P}(x)$ in the form

$$\overline{P}(x) = \sum_{i=1}^{k} \left\{ \left[\sum_{i=1}^{n} c_{ij} M_{ij}(x) \right]^2 \right\},$$

where $\{c_{ij}\}_{i=1}^{k}$ are vectors of coefficients of linear functions $l_i^{u^*}$, $i = 1, \dots, k$, $M_{ij}(x)$ are corresponding monomials. So $\overline{P}(x)$ can be represented as a sum of squares of real polynomials.

The Main Theorem is proved.

A great mathematician D. Hilbert [21] considered the problem of representation of nonnegative polynomial forms as a sum of squares polynomials more than hundred years ago. He proved that if dimension $n = 1$ or 2, arbitrary nonnegative forms (homogeneous polynomials) can be represented as a sum of squares. If $n = 3$ and the number of variables $\leqslant 3$, also the corresponding nonnegative forms can be represented as a sum of squares. But if $n = 3$ and degree $2m \geqslant 4$ there exist nonnegative polynomial forms which cannot be represented as a sum of squares. Our approach gives a possibility not only to determine whether a given nonnegative polynomial can be decomposed into a sum of squares, but to find such decomposition if it exists using the described above algorithm of finding Langrangian quadratic bounds with use of full set of superfluous constraints.

4. Quadratic-type models and upper bounds for the problems of finding the maximum weighted independent set in graphs

Let an undirected graph $G(V, E)$ be given: $V = \{1, \dots, n\}$ is the set of vertices, E is the set of edges; $(i, j) \in E$ is the edge with end points i and j belonging to V, (i, j) and (j, i) are equivalent symbols. The subset $I \subseteq V$ is called independent (stable) if there is no pair i and j such that $i, j \in I$ and $(i, j) \in E$, $i \neq j$.

A subset $K \subseteq V$ is called a clique if for all pairs $i, j \in K$, $(i, j) \in E$, $i \neq j$.

Denote by $A_G = \{a_{ij}\}_{i,j=1}^{n}$ the adjacency matrix of $G(V, E)$:

$$a_{ij} = 1, \text{ if } (i, j) \in E; \quad a_{ij} = 0, \quad \text{if } (i, j) \notin E.$$

The **complement** graph to $G(V, E)$ is $\overline{G}(V, \overline{E})$ with the same vertex set and

$$\overline{E} = \{(i, j), \quad i \neq j/(i, j) \in \overline{E}, \quad \text{iff} \quad (i, j) \notin E\}.$$

The graph $G(V, E)$ may be vertex weighted if for every $i \in V$ the weight $w_i \geqslant 0$ is given.

Let $w = \{w_i\}_{i \in V}$. For a subset $S \subseteq V$ we define $W(S)$ (weight of S) as

$$W(S) = \sum_{i \in S} w_i.$$

We call

$$G(S) = G(S, E \cap S \times S)$$

a subgraph of $G(V, E)$ induced by S. The maximum weight clique problem is to find a clique of maximum weight.

The maximum (weight) independent (stable) set problem is to find an independent set of maximum cardinality (of maximum weight). The size of a maximum independent set is the stability number of G, denoted by $\alpha(G)$. The maximum weight independent set is denoted by $\alpha_w(G)$. It is easy to see that S is a clique of G if and only if S is an independent set of the complement graph $\overline{G}$. So, any result obtained for one of the mentioned problems can be reformulated for another problem. Both of these problems are NP-complete for the class of arbitrary graphs. But for some specific classes of graphs polynomial-time algorithms were constructed.

The weighted maximum stable set problem in graphs can be formulated as the following 0–1 problem: to find

$$\alpha_w(G) = \max(w, x),\ \ x = \{x_1, \dots, x_n\}, \tag{22}$$

$$x_i + x_j \leqslant 1 \text{ for all } (i, j) \in E, \tag{23}$$

$$x_k \in \{0, 1\} \text{ for all } k \in V. \tag{24}$$

We introduce the stable set polytope

$$\text{STAB(G)} := \text{conv}\{x^S \in R^V | S \subseteq V \text{ is a stable set }\}$$

defined as the convex hull of the incidence vectors of all stable sets of vertices of G. $\alpha_w(G)$ is equal to the maximum value of linear function (w, x) on convex polytope STAB(G). Of course, it is very useful to represent the STAB(G) by a system of linear inequalities. Unfortunately, in general case, it is a very hard problem. Therefore we consider some particular cases.

The linear relaxation of the problem (22)–(24) is to find:

$$\alpha_1(G, V) = \max(w, x), \tag{25}$$

$$x_i + x_j \leqslant 1, \forall (i, j) \in E, \tag{26}$$

$$0 \leqslant x_k \leqslant 1,\ \ k = 1, \dots, n. \tag{27}$$

Theorem [20]. *The inequalities (26), (27) give full description of STAB(G) if and only if G is bipartite. Hence, for bipartite graphs the problem (22)–(24) can be solved in polynomial time, as LP problem, if weights are rational.*

The minimal graphs for which inequalities (26), (27) are not sufficient to describe STAB(G) are the odd circuits. In fact, if $G(V, E)$ is an odd circuit then $|V| = 2m + 1$, and the point $x_i = \frac{1}{2}, i \in V$ satisfies the inequalities (26), (27) but does not belong to STAB(G). So we can propose a new class of linear inequalities valid for STAB(G), the so-called **odd circuit constraints**,

$$\sum_{i \in V(C)} x_i \leqslant \frac{|V(C)|-1}{2} \quad \text{for each odd circuit} \quad C, \tag{28}$$

where $V(C)$ is the set of vertices that lie in the circuit C.

Let us call the graph t-perfect if (26), (27) and (28) are enough to describe STAB(G).

In general case we do not know whether the problem of checking the t-perfectness is in NP or in P. Despite this fact a maximum weight stable set in a t-perfect graph can be found in polynomial time by using a slight modification of the ellipsoid method (see [20], p. 276). This problem is reduced to the LP-problem (25)–(27), (28) with possibly exponentially many odd circuits constraints (28). But for obtaining cutting plane in the ellipsoid method it is enough to have one constraint of type (28) that is not satisfied for a given x. The search of such constraint is equivalent to finding the shortest weight odd circuit. For the last problem there exist polynomial time algorithms ([20], p.236).

A partition of V into stable sets (cliques, respectively) is called a **coloring (clique covering**, respectively) of G. The coloring number (respectively, clique covering number) is the smallest number of stable sets in a coloring (respectively, cliques in a clique covering) of G, and is denoted by $\chi(G)$ (respectively, $\overline{\chi}(G)$). It is clear that

$$\chi(G) = \overline{\chi}(\overline{G}).$$

Each stable set in G has no more than one representative in each clique, hence, we have the inequality:

$$\alpha(G) \leqslant \overline{\chi}(G).$$

Similarly we obtain

$$\omega(G) \leqslant \chi(G),$$

where $\omega(G)$ is a clique number of graph G. Berge called a graph G **perfect** if the equality

$$\omega(G') = \chi(G')$$

holds for every induced subgraph G' of G. The first Berge's conjecture about perfect graphs was the following [5, 6]: *The complement of a perfect graph is also perfect.* This was proved by well known Hungarian mathematician Lovász in 1972 [25]. In 1962 Berge also proposed the **Strong Perfect Graph Conjecture.** *The graph is perfect if and only if it or its complement, does not contain an odd circuit of length at least five as an induced subgraph.* This conjecture is still unsolved.

In [24] L. Lóvasz proposed and proved some interesting upper bounds for $\alpha(G)$, which are exact for the class of perfect graphs. Similar results one may find in [26, 31]. Further these results were generalized for a weighted problem of finding $\alpha_w(G)$ (see [20], Chapter 9.3).

Below we shall formulate two general extremal matrix problems studied while finding upper bounds $\nu_w(G)$ for the weighted independent set problem (see [20], §9.3).

I. Let $G = \{V, E\}$ be a graph, $|V| = n$, $w = \{w_i\}_{i=1}^n$ be a vector of vertex weights. Consider F, the class of symmetric matrices $n \times n$ $S(y) = \{s_{ij}(y)\}_{i,j=1}^n$ dependent on parameter vector $y = \{y_1, \dots, y_n\}$, which have following properties for arbitrary $y \in R^n$:

(a) $s_{ij}(y) = 0$, if $(i, j) \in E$;

(b) the matrices $S(y) \in F$ are positive semidefinite;

(c) $\sum\limits_{i=1}^{n} s_{ii}(y) = 1$.

The problem of finding Lagrangian bound $\nu_w(G)$ was reduced to the following extremal problem: to find

$$\max_{\{y:S(y)\in F\}} \sum_{i,j=1}^{n} \sqrt{w_i w_j}\, s_i(y)\, s_j(y).$$

II. The other upper bounds equal to the first one has the form:

$$\vartheta_w(G) = \min_{A\in\Sigma_n} \lambda_{\max}[A + W], \tag{29}$$

where Σ_n is the class of symmetric $n \times n$ matrices, $A = \{a_{ij}\}$, $a_{ii} = 0$ for all $i \in V$, $a_{ij} = 0$ for all i, j nonadjacent in G, $W = \{\sqrt{w_i \cdot w_j}, i, j \in V\}$, $\lambda_{\max}(\cdot)$ denotes the maximal eigenvalue.

Since $\lambda_{\max}(X)$ is a nondifferentiable convex function of entries X, the problem (29) is a typical problem of nondifferentiable optimization and may be solved particularly by r-algorithm. In the case when all weights are 1, the problem (29) is reduced to minimization of maximal eigenvalue on some set of matrices with variable entries. The upper bounds (I) and (II) for $\alpha_w(G)$ were obtained by specific technique of coding theory, namely, by orthonormal representation of graphs (see [24, 20]).

These investigations were stimulated by works of C. Shannon, founder of classical information theory (see [30]). In this paper C. Shannon give the notion of information capacity of graph, which is connected with the problem of speed of transferring long messages, in which some symbols can be mixed. He proposed a linear upper bound for information capacity of graph:

$$\mu(G) = \max \sum_{i=1}^{n} x_i, \tag{30}$$

$$\sum_{i \in S} x_i \leqslant 1, \quad \text{for an arbitrary clique} \quad S \in G, \tag{31}$$

$$0 \leqslant x_k \leqslant 1, \quad k = 1, \dots, n. \tag{32}$$

(Here $G(V, E)$ is a graph, V is a set of symbols and E is a set of edges, corresponding to mixed symbols.) For perfect graph such estimate is exact and coincides with Lóvasz $\nu_w(G)$ bounds.

We can obtain the same bounds using quadratic-type formulation of the problem of finding maximal weighted stable set in graph G and calculating corresponding Lagrangian bounds.

Namely, let the graph $G = (V, E)$, $|V| = n$, $w = \{w_i\}_{i=1}^n$ (the vector of weights) be given. Consider the quadratic-type problem:

to find

$$\alpha_w(G) = \max \sum_{i=1}^{n} w_i x_i \tag{33}$$

subject to constraints:

$$x_k^2 - x_k = 0, \forall k = 1, \dots, n \text{ (Boolean property)} \tag{34}$$

$$x_i x_j = 0, \forall (i, j) \in E. \tag{35}$$

Let $u = (\{u_k\}_{k=1}^n, \{u_{ij}\}_{i,j} \in E)$ be a vector of Lagrange multipliers. Consider the optimal Lagrangian bound

$$\rho_w(G) = \inf_u \sup_x L(x, u).$$

In our case the function $\psi(u) = \sup L(x, u)$ has nonempty domain. As proved in [33], $\rho_w(G)$ has the same value, as Lóvasz's bounds.

These bounds are exact for arbitrary weights if G is a perfect graph. For non-perfect graphs one can find such weights, that corresponding bounds are not exact.

Thus, using a natural quadratic-type formulation of the problem of finding maximal weight stable number in a graph and applying to this problem standard technique of finding Lagrangian bounds we get the same results as obtained by using 'specific' approach of orthonormal representation of graphs.

Consider dual quadratic Lagrangian bounds $\nu_w(G)$ for $\alpha_w(G)$. We want to improve such bounds. For this aim one may use superfluous quadratic inequalities in the problem formulation, for example

$$x_i\, x_j \geqslant 0 \text{ for all (some) nonadjacent pairs } (i, j) \tag{36}$$

or

$$x_k(x_i + x_j) \leqslant x_k, \text{ for any } k \text{ and } (i, j) \in E, \ (i, j, k) \in V. \tag{37}$$

The corresponding modification of Lagrange functions may lead sometimes to more precise dual bound for $\alpha_w(G)$ without essential complication of calculations. So, if we add to the constraints (34), (35) the constraints of the form (36), (37) we then take into attention not only clique constraints (31) but also the odd circuit constraints (28).

Even if we use only one trivial family of superfluous inequalities (36) we can considerably improve the dual upper bounds for $\alpha_w(G)$ in some cases. The corresponding Lagrange function $L_1(x, \lambda)$ can be constructed as follows:

$$L_1(x, \lambda) = \sum_{i=1}^{n} w_i x_i + \sum_{(i,j)\in E} \lambda_{ij} x_i x_j + \sum_{k=1}^{n} \lambda_k (x_k^2 - x_k) - \sum_{(i,j)\in \overline{E}} \lambda_{ij}^{(1)} x_i x_j,$$

where $\lambda = \{\{\lambda_k\}_{k=1}^n,\ \{\lambda_{ij}\}_{i,j\in E},\ \{\lambda_{ij}^{(1)}\}_{i,j\in\overline{E}}$.

Let

$$\varphi_1(\lambda) = \sup_x L_1(x, \lambda), \quad \varphi_1^*(G) = \inf_{\lambda\in\Omega^+} \varphi_1(\lambda)$$

(here Ω^+ is the domain of function φ_1 in λ).

There are some graphs G for which the upper bound φ_1^* for $\alpha(G)$ is much better than $\nu(G)$. For example, let vertices of graph $G_6 = (V, E)$ correspond to the integer numbers from 0 to 63, written in binary codes of length 6, and two vertices $v_1,\ v_2$ are joined by edge if Hamming distance $d(v_1, v_2)$ is no more than 3. It is easy to show that $\alpha(G_6) = 4$ (see [26]) (for instance the maximum stable set is: $\{v_1 = (000000),\ v_2 = (111100),\ v_3 = (110011),\ v_4 = (001111)\}$. It was calculated, that $\nu(G_6) = \frac{16}{3}$, but $\varphi_1^*(G_6) = 4$. It is a bright example of the fact that adding of superfluous constraints to nonconvex problem may considerably improve the dual (Lagrange) bound.

The number of constraints (36) is large, but we may use the fact, that in optimal solution the different dual variables, corresponding to pairs (i, j), which have the same Hamming distance, are equal. Due to this fact in our example the number of dual variables for the constraints (36) can be reduced to 6 variables. The results of numerical experiments one may find in [33, p. 259, 36].

5. Using of the Fan Ky theorem for obtaining dual bounds in some problems of graph theory

The Lovasz's estimate $\nu_w(G)$ may be obtained also in the following way. Let $G = (V, E)$ be a simple undirected graph without loops, $V = \{1, \ldots, n\}$. Denote by $x(S) = \{x_i(S)\}_{i=1}^n$ the indicator vector of the subset $S \subseteq V$. Let $W = \{\sqrt{w_i w_j}\}_{i,j\in V}$ be a $n \times n$- matrix, $u = \{u_{ij}\}_{i,j\in E}$;

$$A_G(u) = \left\{u_{ij}, (i, j) \in E;\ 0,\ \text{otherwise}\right\}$$

be a subclass of symmetric $n \times n$ matrices. For arbitrary $S \subseteq V$ consider an n-dimensional vector

$$w(S) = \left\{ \frac{\sqrt{w_i}}{\sqrt{\sum_{j\in S} w_j}}, i \in S;\ 0,\ \text{otherwise} \right\}.$$

Note that $\|w(S)\| = 1$, and for arbitrary u and a stable set S

$$(W + A_G(u)w(S), w(S)) = (Ww(S), w(S)) = \frac{\sum\limits_{i,j\in S} w_i w_j}{\sum\limits_{k\in S} w_k} = \sum_{i\in S} w_i. \quad (38)$$

Let $\nu_G(u)$ be the maximal eigenvalue of $M(u) = W + A_G(u)$. From (38) one can obtain:

$$\begin{aligned} \nu_w(u) &= \lambda_1[W + A_G(u)] = \max_{\{z:\|z\|=1\}} (M(u)z, z) \\ &\geqslant (M(u)w(S), w(S)) = \sum_{i\in S} w_i \end{aligned}$$

for arbitrary u and a stable set S. So, $\nu_w(u)$ is an upper bound for the maximum weight of stable sets for every u.

Thus, $\nu_w(G) = \min_u \lambda_1(W + A_G(u)) \geqslant \alpha_w(G)$ is a well known estimate of Lovasz.

In a similar way one can obtain the upper bound for $\alpha^{(k)}(G)$, where $\alpha^{(k)}(G)$ is the size of the largest induced k-partite subgraph of $G = (V, E)$, i.e., the maximum number of nodes, that can be covered by k subsets of V so that no edge has both ends in any subset.

Let $S_1, S_2, \ldots, S_k$ be k pairwise nonintersecting stable subsets of V. Introduce the family of n-dimensional vectors $\{Y_r\}_{r=1}^k$, $Y_r = (y_r^{(1)}, \ldots, y_r^{(n)})$,

$$y_r^{(i)} = \begin{cases} \dfrac{1}{\sqrt{|S_i|}} & , i \in S_r, r = 1, \ldots, k; \\ 0 & , \text{ otherwise} \end{cases}$$

Vectors $Y_r,\ r = 1, \ldots, k$, form an orthonormal system. Denote by $Y(k)$ the $n \times k$ matrix with columns $Y_r,\ r = 1, \ldots, k$.

Let $A_G(u),\ u = \{u_{ij}\}_{(i,j)\in E}$, be the parametric family of symmetric $n \times n$ matrices with entries $a_{ij}(u)$

$$a_{ij} = \begin{cases} u_{ij} & , \text{ if } (i, j) \in E, \\ 1 & , \text{ otherwise.} \end{cases}$$

It is easy to verify that for arbitrary u, $\mathrm{tr}\big(A(u)Y(k)[Y(k)]^{(t)}\big) = \sum_{r=1}^k |S_r|$. Due to Fan Ky Theorem (Theorem 1), $\max_{Y\in M_{n,k}} \mathrm{tr}[A(u)YY^{(t)}]$, where $M_{n,k}$ is the class of matrices $n \times k$ with orthonormal system of k columns, is equal to the sum of k largest eigenvalues of symmetric matrix $A(u)$.

So,

$$S_{n,k}(A(u)) = \sum_{r=1}^{k} \lambda_r[A(u)] \geqslant \sum_{r=1} |S_r|$$

for arbitrary u and any k-partite induced subgraph of G with stable sets $S_1, \ldots, S_k$. Thus,

$$\nu^{(k)}(G) = \min_u S_{n,k}(A(u)) \geqslant \alpha^{(k)}(G). \tag{39}$$

Calculation of $\nu^{(k)}(G)$ is the problem of nonsmooth convex optimization and can be solved by subgradient-type methods, particularly, by r-algorithm. Procedure of finding of a subgradient is described in section 1 of this article.

The upper bound (39) for $\alpha^{(k)}(G)$ was first derived by Narasimhan and Manber in [27]. When $k = 1$, one obtains the Lovasz's estimate $\nu(G)$ for the stable number $\alpha(G)$.

Calculation of $\nu^{(k)}(G)$ is reduced to the minimization of $|E|$-dimensional convex nonsmooth function of matrix parameters u, which enter into matrix affinely.

If $\alpha^{(k)}(G) = |V| = n$, it means that graph $G = (V, E)$ has a true vertex colouring by k colours. So if $\nu^{(k)}(G) < n$ the graph G cannot be coloured by k colours. Thus, we can use the upper bound $\nu^{(k)}(G)$ for studying some of the colouring problems.

The problem of minimizing weighted sums of the k largest eigenvalues of a parametric family of symmetric matrices has many combinatorial applications. The best known of them is the graph partitioning problem. The problem is to divide n nodes of a given graph $G(V, E)$ into k disjoint subsets with given cardinalities $m_1 \geqslant \ldots \geqslant m_k$ in a way to minimize the total number of edges connecting different subsets. The problem is NP-hard and therefore we have no hope to get a computationally 'good' algorithm for its precise solution in general case. But Donath and Hoffman proposed in [13, 14] an effective lower bound for this problem that gives us a possibility to obtain good enough approximate solutions or test the accuracy of the solution that can be obtained by other relatively simple heuristic methods.

Let $S_n^k(x) = \sum_{i=1}^k \lambda_i(A(x))$ be a sum of k largest eigenvalues of matrix $A(x)$.We know that $S_n^k(x)$ is a convex function in x. In the case of graph partitioning problem the variables x are contained only in diagonal elements of $A(x)$. The lower bound for the partitioning problem given in [13, 14] is:

$$\rho^*(A, m) = -\frac{1}{2}s^*, \tag{40}$$

where s^* is the solution of the following problem: minimize

$$S_n^k(x, m) = \sum_{i=1}^{k} m_i \lambda_i (A_0 + D(x)) \tag{41}$$

subject to the constraint $\mathrm{tr} D(x) = 0$,

where the nondiagonal elements a_{ij}^0 of the symmetric matrix A_0 are equal to one, if the i-th and j-th nodes are connected, and zero-otherwise; the diagonal elements are defined by

$$a_{ii}^0 = -\sum_{j=1; j\neq i}^{n} (a_{ij}^0), \quad i = 1, \dots, n,$$

$D(x)$ is a diagonal matrix with elements

$$d_{ii} = x_i, \quad , i = 1, \dots, n, \quad \sum_{i=1}^{n} x_i = 0.$$

Note that $m_1 \geqslant m_2 \geqslant \dots \geqslant m_k$, therefore due to Lemma 1 the problem (41) is the problem of convex programming.

The problem (41) is a special case of minimization of the weighted sum of largest eigenvalues of parametric family of matrices (see Section 1).

One can give a more general formulation of the graph partitioning problem (GPP) by introducing nonnegative weights of edges. As earlier we have a set of integers $m_1 \geqslant m_2 \geqslant \dots \geqslant m_k$ with $\sum_{j=1}^{k} m_j = n$. Denote by m the k-vector made up of m_j-s.

Let $G = (V, E)$ be a complete graph with $|V| = n$ and each edge $(i, j), i < j$ has weight $w_{ij} \geqslant 0$. We want to part a set of vertices V into k subsets such that the j-th subset has a prescribed cardinality m_j; and that sum of the weights of those edges whose endpoints are in different subsets is minimized. Let us denote this minimum value by $\pi_m(G)$. Let $W = \{a_{ij}\}$ be a matrix with $a_{ij} = w_{ij}, i < j$, and $a_{ii} = 0, i = 1, \dots, n$.

Donath and Hoffman [13] also proved the following inequality

$$\pi_m(G) \geqslant U_m^W(G) = -\frac{1}{2} \min_x \sum_{j=1}^{k} m_j \lambda_j (W + \mathrm{diag}(x))$$

subject to constraint

$$\sum_{i=1}^{n} x_i = -\sum_{(i,j)\in E} w_{ij}.$$

We can use subgradient-type methods for calculating the lower bounds for $\pi_m(G)$.

The first practically efficient algorithm for obtaining the lower bounds for graph partitioning problems using the minimization of $S_n^k(x, m)$ (see (41)) was proposed in [9] by Cullum et al. for equal $m_i, \ \ i = 1, \dots, k$. They noticed that problem (41) is a problem of nondifferentiable optimization and, moreover as a rule, the optimum point is the point of nondifferentiability of minimized function. Their method is based on idea of smoothing the function $S_n^k(x, m)$ by using information not only about the k largest eigenvalues and corresponding eigenvectors, but also the information about other eigenvalues and eigenvectors, if they are slightly different from k-th eigenvalue.

We used for this purpose one of the modifications of the subgradient-type method with space dilation in the direction of difference of two successive subgradients (r-algorithm) and obtained good results in test experiments (see [33]). The technique for calculation of subgradients is described in the section 1 of this article.

When n is even, $k = 2$, and $m_1 = m_2 = \frac{n}{2}$, the graph partitioning problem can be reduced to the so-called graph bisection problem that can be considered as max-cut problem with one additional constraint.

6. Lagrangian bounds for the maximum cut problem

One of the most bright examples of using quadratic superfluous constraints for improving dual bounds in quadratic-type problems is connected with the max-cut problem.

Let $G(V, E)$ be an ordinary graph with the vertex set $V = \{1, \dots, n\}$ and the edge set $E = \{(i, j) = (j, i)\}$, where (i, j) is the edge, linking vertices (i, j). The weight function W is given by symmetric $n \times n$ matrix

$$W = \begin{cases} 0 & \text{for } (i, j) \notin E, \\ w_{ij} & \text{for } (i, j) \in E. \end{cases}$$

Let the vertex set be divided into two nonempty nonintersecting parts V_1 and V_2: $V = V_1 \bigcup V_2$. We say that the edge (i, j) belongs to the cut $R(V_1, V_2)$ if this edge have its ends in different subsets of subdivision $V = V_1 \bigcup V_2$. We must find such partition $V = V_1 \bigcup V_2$ that the sum of weights of all edges, belonging to the corresponding cut, is maximal.

The max-cut problem is NP-complete, it is proved [37] that it preserves this property even for class of graphs with degrees of vertices not exceeding 3. But for the subclass of the so-called weakly bipartite graphs the max-cut problem with positive weights of edges can be solved by polynomial-time algorithm.

Let $G = (V, E)$ be a graph and $F \subseteq E$ be an edge subset. The vector $y^F \in R^E$ with $y_e^F = 1$ if $e \in F$ and $y_e^F = 0$ if $e \notin F$ is called the incidence vector of F. The polytope $P_B(G) := conv\{y^F \in R^E | (V, F)$ is a bipartite subgraph of $G\}$ is called the **bipartite subgraph polytope** of G. It is clear that for positive edge weights $W(e), \ e \in E$, every optimum basic solution of the linear program

$$\max(W, y), \quad y \in P_B(G),$$

corresponds to a cut.

Consider the trivial inequalities:

$$0 \leqslant y_e \leqslant 1. \tag{42}$$

The inequalities (42) determine $P_B(G)$ completely if and only if G is bipartite. Consider any odd cycle C in G. It is obvious that full edge set of C cannot belong to a cut. So, one can formulate for an incidence vector y of a cut the following odd cut inequalities:

$$y(C) := \sum_{e \in C} y_e \leqslant |C| - 1, \quad C \quad \text{is an odd cycle in} \quad G. \tag{43}$$

DEFINITION 1. A graph $G(V, E)$ that has the property: $P_B(G) = \{y \in R^E | y$ satisfies all inequalities (42) and (43) $\}$ is called **weakly bipartite**.

In [20] (see Chapter 9.3) the polynomial time algorithm is given for checking the feasibility of the vector $\{y_l\}_{l \in E}$; $0 \leqslant y_l \leqslant 1$ for weakly bipartite graphs. This algorithm gives also the odd cycle for which constraint (43) is not fulfilled, if such cycle exists.

Thus, one may use the ellipsoid method for finding a maximal cut for weakly bipartite graph with positive weights (see [20]). But this algorithm is not good for practical calculations, because it is very complex and converges slowly. Therefore we consider below quadratic-type formulation of max-cut problem.

Let $x_k \in \{-1, 1\}$ be a binary variable corresponding to the vertex k ($k \in \{1, 2, \ldots, n\}$),

$$x_k = \begin{cases} -1, & \text{if } x_k \in V_1, \\ +1, & \text{if } x_k \in V_2. \end{cases}$$

Without loss of generality one can assume that the graph $G(V, E)$ is full. In this case the value of cut may be represented in terms of binary variables as a quadratic function:

$$f(x, W) = \frac{1}{8} \sum_{(i,j), i \neq j} w_{ij} (x_i - x_j)^2$$
$$= \frac{1}{4} (\sum_{(i,j)} w_{ij} - \sum_{(i,j)} w_{ij} x_i x_j).$$

Since $\max f(x, W)$ subject to constraints $x_k^2 - 1 = 0, \quad k = 1, \ldots, n$, equal to $1/4 \sum_{(i,j)} w_{ij} - \min_x \sum_{(i,j)} w_{ij} x_i x_j$ subject to the same constraints, the max-cut problem can be reduced to the quadratic-type problem:

to find

$$S(W) = \min_x \sum_{i,j=1}^{n} w_{ij} x_i x_j \tag{44}$$

subject to constraints

$$x_k^2 - 1 = 0, \quad k = 1, \dots, n. \tag{45}$$

Denote by $R^*(W)$ the optimal value of the max-cut. Due to (44), (45)

$$R^*(W) = \frac{1}{4}\left(\sum_{(i,j)} w_{ij} - S(W)\right). \tag{46}$$

Let $u = \{u_1, \dots, u_n\}$ be the vector of Lagrange multipliers corresponding to equalities (45). Then Lagrange function $L(x, u)$ of reduced problem (44)–(45) has the form

$$L(x, u) = (W(u)x, x) - \sum_{k=1}^{n} u_k,$$

where

$$W(u) = W + \operatorname{diag} u,$$

(diag u is a diagonal matrix with the components $d_{ii} = u_i, i = 1, \dots, n$). Let $\lambda_{\min}(A)$ denotes the minimal eigenvalue of matrix A.

The quadratic part of $L(x, u)$ is a homogeneous quadratic in x function, so

$$\inf_x L(x, u) = \begin{cases} -\infty, & \text{if} \quad \lambda_{\min}(W(u)) < 0, \\ -\sum_{k=1}^{n} u_k, & \text{if} \quad W(u) \succeq 0, \ \text{i.e.}\ \lambda_{\min}(W(u)) \geqslant 0. \end{cases}$$

Consider

$$\varphi^* = \min \sum_{k=1}^{n} u_k, \quad W(u) \succeq 0.$$

By using an exact nonsmooth penalty function (see [7], Chapter 4), the problem of finding φ^* is reduced to an unconstrained optimization of nondifferentiable function

$$f(u, s) = \sum_{k=1}^{n} u_k + s\left[\lambda_{\min}^{-}(W(u)\right], \tag{47}$$

where

$$\lambda_{\min}^{-} = \begin{cases} 0, & \text{if } \lambda_{\min} \geqslant 0, \\ \lambda_{\min}, & \text{if } \lambda_{\min} \leqslant 0, \end{cases}$$

s is a penalty multiplier. One can prove, that if $s \geqslant n$, the problem of minimization of $f(u, s)$ is equivalent to the problem (47).

One can solve problem (47) by using r-algorithm. After obtaining an optimal value of (47) $S(W)$ we can obtain the upper bound for the optimal cut value, using (46).

The alternative approaches for finding upper bounds of the max-cut problems are represented in [14, 2, 10]. In articles [1, 22] the bounds similar to ours are obtained by using another technique. In [16] it is shown how to describe the max-cut problem in terms of semidefinite programming.

Consider a weighted graph $G = (V, E)$ with even number of vertices: $|V| = n = 2s$. The bisection problem for G can be formulated as follows:

$$\text{maximize} \quad f(y) = \frac{1}{4}(\sum_{(i,j)\in E} w_{ij} - \sum_{i,j\in E} w_{ij} y_i y_j) \tag{48}$$

subject to constraints:

$$y_i^2 - 1 = 0, \ \forall i \in V = \{1, \dots, n\}, \tag{49}$$

$$\sum_{i=1}^{n} y_i = 0 \tag{50}$$

(for feasible solution $\overline{y}$ the number of $\overline{y}_i$ having value 1 must be equal to number $\overline{y}_i$ having value -1). One may propose several ways for obtaining upper bounds for the problem (48)–(50):

(i) use an estimate of the type (40), (41) with $m_1 = m_2 = s$;

(ii) due to (50), $y_n = \sum_{i=1}^{n-1} y_i$. Set $\tilde{y} = \{y_1, \dots, y_{n-1}\}$.
Consider the problem in variables $\tilde{y}$ which is equivalent to the problem (48)–(50):

$$\begin{aligned} \text{maximize} \quad \overline{f}(\tilde{y}) = \frac{1}{4}\Bigg(\sum_{(i,j)\in E} w_{ij} - \sum_{i,j\in E} \Bigg[w_{ij} y_i y_j \\ - \sum_{i=1}^{n-1} w_{in}(y_i y_n + y_n y_i) \Bigg] \end{aligned} \tag{51}$$

subject to constraints:

$$y_i^2 - 1 = 0, \quad \forall i \in \{1, \dots, n-1\}, \tag{52}$$

$$\left(\sum_{i=1}^{n-1} y_i\right)^2 - 1 = 0. \tag{53}$$

One can construct Lagrange function for the problem (51)–(53) in the form:

$$L(\tilde{y}, u) = -\overline{f}(\tilde{y}) + \sum_{i=1}^{n-1} u_i(y_i^2 - 1) + u_n\left[\left(\sum_{i=1}^{n-1} y_i\right)^2 - 1\right]$$

and consider a marginal function

$$\psi(u) = \inf_{\tilde{y}} L(\tilde{y}, u).$$

$\psi(u)$ is convex function, and $dom\psi \subseteq \overline{\Omega}^+$, where $\overline{\Omega} + (\Omega^+) = \{\overline{u} \in R^n :$ $L(\tilde{y}, \overline{u})$ is convex (strictly convex) function in $\tilde{y}$ for fixed $\overline{u}\}$.
Since $L(\tilde{y}, u)$ has no linear in $\tilde{y}$ terms

$$\psi(u) = -\frac{1}{4}\overline{W} + \sum_{i=1}^{n} u_i, \text{ if } u \in dom\psi,$$

where $\overline{W} = \sum_{i,j=1}^{n} w_{ij}$.
Let $Q(u)$ be a matrix of the quadratic part of $L(x, u)$. Let $\psi^* = \sup_{u \in dom\psi}$ $\psi(u)$

$$\psi^* = \sup_{u \in \Omega^+} \psi(u) = \{\sup \psi(u) : Q(u) \succeq 0\}.$$

Thus the upper dual bound φ^* for the maximum bisection problem (48)–(50) can be calculated by solving the following problem:

$$\text{find } \psi^* = \min\left[\sum_{i=1}^{n} u_i + \frac{1}{4}\overline{w}\right] \tag{54}$$

subject to constraint:

$$\lambda_n[Q(u)] \succeq 0. \tag{55}$$

The convex programming problem (54), (55) can be solved by application of r-algorithm for nonsmooth penalty function

$$\varphi_p(u) = \frac{1}{4}\overline{w} + \sum_{i=1}^{n} u_i + s\lambda_n^+[Q(u)],$$

where

$$\lambda_n^+(u) = \begin{cases} \lambda_n(u), & \text{if } \lambda_n(u) < 0; \\ 0, & \text{if } \lambda_n(u) \geqslant 0; \end{cases}$$

$s > 0$ is a penalty multiplier;

(iii) the bisection problem (48)–(50) can be approximately reduced by use of quadratic-type penalty $s(\sum u_i)^2$ for constraint (50) to usual max-cut problem with other weights:

$$w_{ij}^c = w_{ij} + s,$$

where s is a penalty multiplier.

The max-cut problem may be formulated also as the problem of linear programming in edge Boolean variables $y = \{y_{ij}\}_{i,j=1,i\neq j}^n$, $y_{ij} \in \{0, 1\}(y_{ij} = 1$, if (i, j) belongs to the cut $R(V_1, V_2)$, and 0 in the opposite case) (see [3, 4, 19]).

Let $M(G)$ be the convex hull of all feasible solutions of the max-cut problem in edge variables $\{y_{ij}\}$. So, one may reformulate the max-cut problem as a linear programming problem: to find

$$l^*(G, W) = \max_{y\in M(G)} \sum_{(i,j)} w_{ij} y_{ij}.$$

Edge variables y_{ij} and vertex binary variables are linked by simple formula:

$$y_{ij} = \frac{1 - x_i x_j}{2} \text{ for all } y_{ij}, i \neq j. \tag{56}$$

In a large graph the set $M(G)$ has tremendous number of faces. But for special graphs all its faces can be described in compact form. For example, if graph is planar, for each triangle $\{i, j, k\}$, one may write down the inequality

$$y_{ij} + y_{jk} + y_{ik} \leqslant 2. \tag{57}$$

Barahona and co-workers [3, 4] generalized the algorithm for planar graphs on the family of graphs not contractible to graph K_5 (5-clique).

Denote by $y(N)$ the sum $\sum_{(i,j)\in N} y_{ij}$, where N is a set of edges. Using this notation we can record more general family of inequalities of the following form:

$$y(F) - y(C \setminus F) \leqslant |F| - 1 \text{ for all circuits } C \text{ of } G$$
$$\text{and subsets } F \subseteq C \text{ with } |F| \text{ odd}. \tag{58}$$

These inequalities are valid for $M(G)$.

Barahona and co-workers [3, 4] have shown that the solution set of (58) is equal to $M(G)$ if and only if G has no subgraphs contractible to K_5 (K_5 is 5-clique). There is a simple algorithm (see [20], §8.3, pp. 235–236) for verifying validation of system (58) for given $\overline{y}$ ($0 \leqslant \overline{y}_{ij} \leqslant 1$ for all (i, j)).

We construct a new graph $H = (V' \bigcup V'', E' \bigcup E'' \bigcup E''')$, consisting of disjoint copies $G'(V', E')$ and $G''(V'', E'')$ of G and an additional edge set E''' that contains the edges $u'v''$, $u''v'$ for each $uv \in E$. The edges $u'v' \in E'$ and $u''v'' \in E''$ get the weight $y_{uv'}$ while the edges $u'v''$ and $u''v' \in E'''$ get the weight $1 - y_{uv}$.

For each node $u \in V$ we calculate a shortest path (with respect to weights just defined). Such a path contains an odd number of edges of E''' and corresponds to a closed walk in G, containing u. Clearly, if the shortest of these $[u', u'']$-paths has the length equal to at least one, then y satisfies (58), otherwise there exists a cycle C and a set $F \subseteq C$, $|F|$ odd, such that $\overline{y}$ violates the corresponding inequality. Such linear in $\overline{y}$ inequalities can be converted by substitution (56) in quadratic inequalities in binary variables x which may be added to the set of obtained earlier inequalities in x as superfluous constraints. So we described one of the possible ways to generate the superfluous inequalities for model (44). Further adding of superfluous constraints as a rule essentially improves dual quadratic upper bounds for the max-cut problem. In combination with heuristic methods of constructing feasible cuts such approach allows to obtain an exact solution for graphs with integer weight matrix which subgraphs are not contractible to K_5 ([3, 20]). A more detailed description of corresponding algorithm one can find in [33], §8.1.

Thus, in the case of weakly bipartite graphs the max-cut problem with positive weights of edges reduces to LP problem but the number of inequalities of the type (42) may increase as exponential function of the size of a graph. To overcome this difficulty for obtaining a polynomial time algorithm of solving the max-cut problem in the mentioned above case one can use the ellipsoid method. Certain steps of this method need to solve the problem of finding the odd cycle of minimum weight for checking the feasibility of the inequalities (42). The search of such a cycle can be reduced to at most n applications of the algorithm of finding the shortest even path in a weighted graph (see [20], Chapter 8.4).

The quadratic dual bound for the max-cut problem in the case of weakly bipartite graphs is not always exact, but if all weights are nonnegative, the following estimate is proven.

THEOREM 4. *Let G be a weakly bipartite graph with non-negative weights. Then*

$$\varphi(G) \leqslant \frac{5\sqrt{5}(1+\sqrt{5})}{32} mc(G).$$

In [18, 15] there was proposed a way of generating 'good' feasible solutions for the max-cut problem using the eigenvectors associated with $\lambda_n(\overline{W} + diag(\overline{u}))$ where $\overline{u}$ is an approximation of u^*. Let $s(\overline{u}) = \{s_i(\overline{u})\}_{i=1}^n$ be an eigenvector associated with $\lambda_n(\overline{W} + diag(\overline{u}))$. Write its entries in a nonincreasing order:

$$s_{i_1}(\overline{u}) \geqslant s_{i_2}(\overline{u}) \geqslant \ldots \geqslant s_{i_n}(\overline{u}).$$

For different k, $1 \leqslant k \leqslant n$, construct the partition of V:

$$V = S(k) \cup (V \backslash S(k)),$$

where $S(k) = \{i_1, \ldots, i_k\}$, and find $\max_k c(S(k)) = c(S(\overline{k}))$, where $c(S(k))$ is the value of cut corresponding to two subdivision: $S(k)$ and $V \backslash S(k)$. Choose a feasible vector $y(s(\overline{u}))$ corresponding to this partition:

$$y(s(\overline{u})) = \begin{cases} 1, & \text{for } i \in S(\overline{k}), \\ -1, & \text{for } i \in V \backslash S(\overline{k}). \end{cases}$$

In many cases $y(s(\overline{u}))$ is a good approximation of the optimal solution. So, in the process of minimization of $S_N(u)$ by r-algorithm one may use $u(k)$ at each step k not only for obtaining an upper bound for $f(y^*) = mc(G)$, but also for obtaining a feasible integral solution $y(u(k))$ in the described above way. One can use the record value of the objective function for generated feasible solutions as lower bound for $mc(G)$. The results of numerical experiments are presented in [34].

7. The computational results

(1) We describe the results of test calculations for finding global minimum value of some multiextremal polynomial functions of one variable. For beginning, consider the family of 6-th degree polynomials of the form:

$$P_6(x) = x^6 - 2tx^4 + (t^2 - \varepsilon)x^2 = x^2((x^2 - t)^2 - \varepsilon), \tag{59}$$

where t and ε are parameters of this family.

It is easy to see that if $t > 0$, $\varepsilon > 0$, $t^2 - \varepsilon > 0$, polynomials (59) have two global minima:

$$x_1^* = \sqrt{\frac{2t}{3} + \frac{\sqrt{t^2 + 3\varepsilon}}{3}}; \qquad x_2^* = -\sqrt{\frac{2t}{3} + \frac{\sqrt{t^2 + 3\varepsilon}}{3}}.$$

If $\varepsilon = 0$, polynomials (59) has three global minima: x_1^*, x_2^* and 0.

The results are given in Table 1. Here x_2^* is an approximation of optimal dual bounds obtained for the quadratic-type problems (8)–(11), which correspond to the global minimum of polynomial $P_6(x)$; P^* is an exact value of global minimum of polynomial $P_6(x)$; u_1^*, u_2^*, u_3^* are obtained values of Lagrange multipliers for constraints (9)–(11).

Table 1 shows that in all test experiments ψ_2^* is equal to global minimums of polynomial $P_6(x)$ with good precision. Optimal values of dual variables u_1^*, u_2^*, u_3^* lie at the border of positive definite region of matrix

$$K(u) = K(u_1, u_2, u_3) = \begin{pmatrix} u_1 & u_2/2 & -u_3/2 \\ u_2/2 & u_3 - 2t & 0 \\ -u_3/2 & 0 & 1 \end{pmatrix}.$$

Taking into account that $u_2^* = 0$,

$$K(u_1^*, 0, u_3^*) = \begin{pmatrix} u_1^* & 0 & -u_3^*/2 \\ 0 & u_3^* - 2t & 0 \\ -u_3^*/2 & 0 & 1 \end{pmatrix}$$

Table 1. Results of experiments for $P_6(x)$

t	ε	ψ_2^*	P^*	u_1^*	u_2^*	u_3^*
1.00	1.0000	-1.185185	-1.185185	1.7778	0.0000	2.6667
1.00	0.1000	-0.102387	-0.102387	1.0956	0.0000	2.0935
1.00	0.0100	-0.010025	-0.010025	1.0100	0.0000	2.0099
1.00	0.0010	-0.001000	-0.001000	1.0010	0.0000	2.0010
1.00	0.0001	-0.000100	-0.000100	1.0001	0.0000	2.0001
1.00	0.0000	0.000000	0.000000	1.0000	0.0000	2.0000
0.50	1.0000	-0.758076	-0.758076	0.8728	0.0000	1.8685
0.50	0.1000	-0.054288	-0.054288	0.3370	0.0000	1.1611
0.50	0.0100	-0.005049	-0.005049	0.2598	0.0000	1.0194
0.50	0.0010	-0.000500	-0.000500	0.2510	0.0000	1.0020
0.50	0.0001	-0.000050	-0.000050	0.2501	0.0000	1.0002
0.50	0.0000	0.000000	0.000000	0.2500	0.0000	1.0000
0.25	1.0000	-0.562500	-0.562500	0.5625	0.0000	1.5000
0.25	0.1000	-0.031676	-0.031676	0.1350	0.0000	0.7347
0.25	0.0100	-0.002593	-0.002593	0.0718	0.0000	0.5361
0.25	0.0010	-0.000251	-0.000251	0.0635	0.0000	0.5040
0.25	0.0001	-0.000025	-0.000025	0.0626	0.0000	0.5004
0.25	0.0000	0.000000	0.000000	0.0625	0.0000	0.5000

the eigenvalues of $K(u_1^*, 0, u_3^*)$ will be following

$$\lambda_1 = u_3^* - 2t; \quad \lambda_{2,3} = \frac{(u_1^* + 1) \pm \sqrt{(u_1^* - 1)^2 + (u_3^*)^2}}{2}.$$

For $\varepsilon > 0$ the polynomials $P_6(x)$ have two global minima and eigenvalues of matrices $K(u_1^*, 0, u_3^*)$ satisfy the condition $\lambda_1 > 0$, $\lambda_2 > 0$ and $\lambda_3 = 0$. If $\varepsilon = 0$, $P_6(x)$ has three global minima, and $\lambda_2 > 0$, $\lambda_1 = \lambda_3 = 0$.

Described results show that the method is stable also in the case, when solution is at the border of the set of positive definiteness of matrix $K(u)$.

Some examples of finding Lagrangian bounds were solved for polynomials of degree greater than 6. For polynomial of eighth degree

$$P_8(x) = x^8 - 76/3x^6 + 222x^4 - 756x^2,$$

which has two global minimum points: $x_1^* = \sqrt{3}$ and $x_2^* = -\sqrt{3}$, $P_8(x_1^*) = P_8(x_2^*) = -873$, the following results were obtained: the precision 10^{-6} was reached in 66 iterations of r-algorithms; the precision 10^{-10} was reached in 98 iterations. For the polynomial

$$P_8(x) = x^8 - 8x^7 + 112x^5 - 158x^4 - 392x^3 + 840x^2$$

Table 2. Results for polynomials Chebyshev and Legandre polynomials

	Chebyshev polynomials			Legandre polynomials		
$2m$	$iter$	nf	$\psi_r^* - P^*$	$iter$	nf	$\psi_r^* - T^*$
4	21	31	0.47684E-06	21	31	0.44740E-06
6	35	71	0.17072E-04	47	95	0.72306E-08
8	70	129	0.52532E-04	95	189	0.11723E-09
10	124	268	0.83521E-04	144	257	0.15465E-10
12	183	377	0.83487E-03	432	793	0.43077E-13
14	234	458	0.88431E-03	505	921	0.34885E-12

having one global minimum point: $x_1^* = 5$, $P_8(x_1^*) = -11125$, the following results were obtained: for precision 10^{-11} the 105 iterations were needed. Some experiments for Legandre and Chebyshev polynomials were made. The Legandre polynomials were generated by recurrent formula:

$$P_{n+1}(x) = \frac{2n+1}{n+1} x P_n(x) - \frac{n}{n+1} P_{n-1}(x), \quad P_0(x) = 1, \quad P_1(x) = x$$

the Chebyshev polynomials were generated by formula:

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \quad T_0(x) = 1, \quad T_1(x) = x.$$

The results are given for even degrees $2m = 4, 6, 8, 10, 12$ and 14 in Table 2. Here $iter$ – the number of iterations and nf – the number of calculations of function and subgradients in r-algorithm procedure, necessary for obtaining the defined precision $\varepsilon_x = 1.E-6$. In columns 4 and 7 the deviations of Lagrangian bounds from global minimums are given.

(2) For the max-cut problem some numerical experiments were accomplished. For planar graph in the form of icosahedron (12 vertices and 30 edges) we use the edge weights from Table 3. At first we obtain dual quadratic upper bound 668. Further by heuristic algorithm [18] which use optimal dual variables we obtain a feasible solution with cut value 642 ($V_1 = \{1, 2, 9, 10, 11, 12\}$; ($V_2 = \{3, 4, 5, 6, 7, 8\}$). For obtaining optimal cut we generate successively the triangle constraints (57). In Table 4 the vertices of triangle cycles are represented. After adding ten superfluous constraints in the form of triangles constraints, which were converted by substitution (56) in quadratic inequalities, we found the upper bound φ^* such that $\varphi^* - 642 < 1$. Thus, the found feasible solution gives the optimal cut. Note, that our result corresponds to Theorem 4:

$$(668 - 642)/642 = 26/642 = 23/321 \approx 0.0405 \approx 4\%.$$

Table 3. Edge weights for 'icosahedron'

No. edge	i	j	w_{ij}	No. edge	i	j	w_{ij}	No. edge	i	j	w_{ij}
1	1	2	20.0	11	3	9	18.0	21	7	8	18.0
2	1	3	30.0	12	3	10	15.0	22	7	9	27.0
3	1	4	40.0	13	4	5	32.0	23	7	10	36.0
4	1	5	50.0	14	4	10	24.0	24	7	11	45.0
5	1	6	60.0	15	4	11	20.0	25	7	12	54.0
6	2	3	16.0	16	5	6	40.0	26	8	9	14.0
7	2	6	48.0	17	5	11	30.0	27	8	12	42.0
8	2	8	12.0	18	5	12	25.0	28	9	10	21.0
9	2	9	10.0	19	6	8	30.0	29	10	11	28.0
10	3	4	24.0	20	6	12	36.0	30	11	12	35.0

Table 4. Adding of triangle cycles for 'icosahedron'

No. cycle	i, j, k	φ^*	No. cycle	i, j, k	φ^*
1	6,8,12	665.53	6	5,6,12	647.06
2	1,5,6	661.94	7	2,6,8	646.72
3	1,4,5	657.13	8	5,11,12	646.41
4	7,11,12	652.82	9	7,8,12	645.14
5	7,9,10	650.66	10	7,10,11	642.90

(3) Optimal bisection of graphs (BiSection). BiSection problem for graph $G(V, E)$ is formulated for even $|V|$ and corresponds to MaxCut problem under condition: the number of vertices in subsets V_1 and V_2 are equal. We obtain the quadratic-type problem (44)–(45), with a new constraint

$$\sum_{i=1}^{n} x_i = 0, \tag{60}$$

or its quadratic analogue:

$$\left(\sum_{i=1}^{n} x_i\right)^2 = 0. \tag{61}$$

One may find the upper Lagrangian for BiSection in different ways: (a) using (60) for excluding x_n; (b) using Lagrange multiplier u_{n+1} for account (61); (c) using

Table 5. Edge weights for Petersen graph

(i, j)	w_{ij}								
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9
(1,2)	1	5	2	5	1	5	2	5	1
(1,5)	1	1	3	6	8	2	7	2	3
(1,10)	1	3	4	3	1	2	2	2	3
(2,3)	1	2	7	2	3	1	3	6	8
(2,9)	1	2	6	2	5	5	1	5	9
(3,4)	1	3	1	6	7	3	4	3	1
(3,8)	1	4	3	4	2	1	2	1	2
(4,5)	1	1	2	1	2	2	6	2	5
(4,7)	1	1	5	1	3	4	3	4	2
(5,6)	1	5	1	5	9	3	1	6	7
(6,8)	1	5	5	5	4	1	5	1	3
(6,9)	1	3	1	3	1	5	5	5	4
(7,9)	1	2	2	2	3	3	1	3	1
(7,10)	1	1	2	1	8	1	2	1	8
(8,10)	1	1	1	1	1	1	1	1	1

penalty nonsmooth function with parameter s, where $s > \sum_{(i,j)\in E} |w_{ij}|$. Algorithm for solving 'BiSection' by scheme (c) is similar to scheme of finding dual bounds for max-cut (see Section 6). The difference is that instead of matrix $W(u)$ we use the matrix $\tilde{W}(u) = sJ_n + W(u)$, where J_n is the $n \times n$ matrix with all entries equal to 1.

The results of experiments are given in Table 6. For test problems we use the Petersen graph (10 vertices and 15 edges) with enumeration of vertices as in [34]. The values of edge weights for nine examples $P_1, \dots, P_9$ are given in Table 5.

We use the following parameters of r-algorithm: $\alpha = 2$, $h_0 = 1.0$, $q_1 = 0.9$, $n_h = 3$, $q_2 = 1.1$. For the starting point we use $u_0 = \{1, 1, 1, 1, 1, 0, 0, 0, 0, 0\}$. r-algorithm stops if at the k-th iteration $\|u_k - u_{k-1}\| \leqslant \varepsilon_u = 10^{-3}$ or $\varphi_{up} - \varphi_{lo} < 1$. Here φ_{up} is an upper bound obtained at the current step, φ_{lo} is the value of cut obtained by heuristic procedure of finding 'good' feasible cut (see Section 6). This cut is shown in Table 6 by representation $V_1 \cup V_2$.

From Table 6 we see, that for 5 examples we obtain exact values of the problems after rounding of bounds, and the number of iterations was relatively small.

Table 6. BiSection for Petersen graph

$Test$	$iter$	φ_{up}	φ_{lo}	V_1	V_2
P_1	41	12.50045	11.0	1,2,5,7,8	3,4,6,9,10
P_2	7	34.91024	34.0	2,3,6,7,10	1,4,5,8,9
P_3	2	40.84215	40.0	2,5,7,8,10	1,3,4,6,9
P_4	44	41.63351	40.0	1,3,6,7,10	2,4,5,8,9
P_5	35	52.52589	51.0	1,2,4,6,7	3,5,8,9,10
P_6	32	34.14082	33.0	1,4,5,8,9	2,3,6,7,10
P_7	17	39.99753	39.0	1,2,4,8,9	3,5,6,7,10
P_8	11	40.99564	40.0	2,4,6,8,10	1,3,5,7,9
P_9	7	52.92904	52.0	1,2,4,6,7	3,5,8,9,10

Conclusion

As we have shown in this article the finding of Lagrangian bounds in polynomial multiextremal and Boolean extremal models can be reduced in many cases to the problems of nondifferentiable optimization with specific constraints in the form of semidefinitness of some parametric families of symmetric matrices. These problems can be formulated also in the form of semidefinite programming (SDP) (see [29]) and solved by corresponding algorithms (for example, by interior point method [12]).

Since 1970, we solved many hundreds of sophisticated tests and applied problems using r-algorithms . The results of Nemirovsky–Yudin on information complexity of convex programming algorithms [28] show that in general case one must make $O(n \log \frac{1}{\epsilon})$ measurements of function and subgradients in current points to guarantee the relative accuracy ϵ on minimized function value and discrepancy in constraints.

The results of testing of r-algorithms show that if the errors of rounding are not essential, the objective function values as a rule may be majored by a geometrical progression of the form $Cq^{\frac{k}{n}}$, where k is the number of current step and $q = \frac{1}{2}$. So, as a rule the convergence of r-algorithm is approximately $2n$ times faster than of well known ellipsoid methods.

Our numerical experiments showed that application of nondifferentiable optimization models for obtaining Lagrangian bounds for multiextremal and combinatorial problems have some advantages in comparison with SDP methods: (a) the possibility to take in account specific structure of the problem and the use of decomposition schemes, if the problem has quasi-block structure; (b) simple ways for exchanging models in the case of adding new quadratic constraints; (c) possibility to use exact nonsmooth penalty functions; (d) rather fast rate of convergence and simplicity of r-algorithms.

In [8] the long history of finding good algorithms for solving some extremal mechanical problems posed by Lagrange in 1773 is described. Now these problems were reduced to nonsmooth optimization problems of maximization of the least eigenvalue of a self-adjoint fourth-order differential operator.

Only after more than 200 years good algorithms for solving such problems have been found and implemented. New algorithms were built on the base of last developments of convex analysis and study of the differential properties of nonsmooth functions, particularly, of extremal eigenvalues of differential operators. Note, that similar problems arise when we want to find Lagrangian bounds for a combinatorial problem. As we have shown, the r-algorithms solve such problems successfully.

We want to emphasize that methods of nondifferentiable optimization must become necessary part of the courses of applied mathematics for mathematical and technical education.

References

1. Alizadeh, F. (1995), Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization* 5, 13–51.
2. Anjos, M.F. and Wolkowicz, H. A strengthened SDP relaxation via second lifting for the max-cut problem, University of Waterloo, Department of Combinatorics and Optimization, *Research Report CORR 99-55,* 28 pp.
3. Barahona, F., Grötshel, M. and Mahjoub, A. (1985), Facets of the bipartite subgraph polytope. *Mathematics of Operations Research* 10, 340–358.
4. Barahona, F. and Mahjoub, A. (1986), On the cut polytope. *Mathematics Programming,* 36, 157–173.
5. Berge, C. (1961), Farbung von Graphen, deren samtiche bzv. deren ungerade Kreise starr sind (Zusammenfassung), *Wissenschaftliche Zeitschrift,* Martin Luther Universitat Halle-Wittenberg, Mathematisch-Naturwissenschaftliche Reihe, pp. 114–115.
6. Berge, C. (1962), Sur une conjecture relative au probleme des codes optimaux, *Communication, 13eme assemblee generake de l'URSI,* Tokyo.
7. Bertsekas, D.P. (1995), Nonlinear Programming, *Athena Scientific, Belmont, Massachusetts,* P. 646.
8. Cox, S.J. (1992), The shape of the ideal column, *Mathematical Intelligencer* 14, 16–31.
9. Cullum, J. Donath, W.E. and Wolfe, P. (1975), The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices, *Mathematical Programming Study* 3, 35–55.
10. Delorme, C. and Poljak, S. (1993), Laplacian eigenvalues and the maximum cut problem, *Math. Programming* 62(3), 557–574.
11. Demyanov, V.F. andRubinov, A.M. (1980), On quasidifferentiable functionals, *Soviet Mathematics Doklady 21, 14–17*
12. Dikin, I.I. and Zorkaltsev, V.I. (1980), Iterative solving of mathematical programming problem (interior point method: algoritms), Nauka, Moscow, (in Russian).
13. Donath, W.E. and Hoffman, A.J. (1972), Algorithm for partitioning graphs and computer logic based on eigenvectors of connection matrices, *Math. Prog. Study* 3, 35–55.
14. Donath, W.E. and Hoffman, A.J. (1973), Lower bounds for the partitioning of graphs, *IBM J. Res. Dev.* 17.
15. Frieze, A. and Jerum, M. 1995, Improved approximation algorithms for max k-cut and max bisection, *Reprint, Carnegi Mellon University,* pp. 1-17.

16. Goemans, M.X. 1997, Semidefinite programming in combinatorial optimization, *Math. Programming* 79, 143–162.
17. Fan, K. (1949), On a theorem of Weyl concerning the eigenvalues of linear transformations, *Proceedings of the National Academy of the Sciences of U.S.A.* 35, 652–655.
18. Goemans, M.X. and Williamson, D.P. (1994), 0.878-Approximation algorithms for MAX-CUT and MAX 2SAT, *Proceedings of the 26th Annual ACM Symposium on Theory of Computing*, 422–431.
19. Grötshel, M. and Pulleyblank, W. (1981), Weakly bipartite graphs and the max-cut problem. *Operations Research letters* 1, 23–27.
20. Grötschel, M., Lóvasz, L. and Schrijver, A. (1988), *Geometric Algotithms and Combinatorial Optimization*, Vol. 2 *Combinatorics*. Springer, Berlin.
21. Hilbert, D., Über die Darstellung definiter Formen als Summen von Formen quadraten. *Math. Ann.* Leipzig, 1888, Bd. 22, pp. 342–350.
22. Lawrent, M., Poljak, S. and Rendl, F. (1997), Connections between the semidefinite relaxations of the max-cut and stable set problems, *Math. Programming* 77, 225–246.
23. Lewis, A.S. and Overton, M.L. (1996), Eigenvalue Optimization, *Acta Numerica* 5, 149–190.
24. Lóvasz, L. (1979), On the Shannon-capacity of graph. *IEEE Transactions on information theory.* IT-25 1, 1–7.
25. Lóvasz, L. (1972), Normal hypergraphs and the perfect graph conjecture, *Discrete Mathematics* 2, 253–267.
26. McElice, R., Rodemich, E. and Romsey, Jr. H. (1978), The Lovash bound and some generalization. *Journal of combinatorics, Information and System Science.* 3(3), 134–152.
27. Narasimhan, G. and Manber, R. (1990), A generalization of Lovasz's sandwich theorem, *Polyhedral combinatorics: Proced. of a DIMACS Workshop AMS*.
28. Nemirovsky, A. and Yudin, D.B. (1983), *Problem Complexity and Methods Efficiency in Optimization, J. Wiley, New York.*
29. Nesterov, Yu. and Nemirovskii, A. (1994), Interior point polynomial algorithm in convex programming, *SIAM*, p. 406.
30. Shannon, C. (1956), The zero-error capacity of a noisy channel. *IRe Transactions Information Theory.* IT 2, 8-19.
31. Shrijver, A. (1979), A comparison of the Delsarte and Lovash bounds. *IEEE Transactions on Information Theory*. IT-25, 4.
32. Shor, N.Z. (1970), Methods of non-differentiable optimization and its applications. *Ph.D. Thesis*, Institute of Cybernetics of NASU, Kiev, Ukraine.
33. Shor, N.Z. (1998), *Nondifferentiable Optimization and Polynomial Problems*, Kluwer, Dordrecht.
34. Shor, N.Z. and Berezovski, O.A. (1995). New algorithms for solving weight max-cut problem, *Kibernetika i sistemny analiz* 2, 100–106.
35. Shor, N.Z. Stetsenko, S.I. (1989), *Quadratic extremal problems and nondifferentiable optimization*, Naukova dumka, Kiev, (in Russian).
36. Shor, N.Z. and Stetsyuk, P.I. (1997), Using of r-algorithm modifications for obtaining of global minimum of polynomial functions, *Kibernetika i sistemny analiz* 4, Kiev.
37. Yannakakis, M. (1978), Node and edge deletion NP-complete problems. *Proc. 10th Ann. ACM Symp. on Theory of Computing.* New York, pp. 253–264.