PANM 21

Jan Lamač; Miloslav Vlasák

Finding vertex-disjoint cycle cover of undirected graph using the least-squares method

In: Jan Chleboun and Pavel Kůs and Jan Papež and Miroslav Rozložník and Karel Segeth and Jakub Šístek (eds.): Programs and Algorithms of Numerical Mathematics, Proceedings of Seminar. Jablonec nad Nisou, June 19-24, 2022. Institute of Mathematics CAS, Prague, 2023. pp. 97–106.

Persistent URL: `http://dml.cz/dmlcz/703192`

Programs and Algorithms of Numerical Mathematics 21
J. Chleboun, P. Kůs, J. Papež, M. Rozložník, K. Segeth, J. Šístek (Eds.)
Institute of Mathematics CAS, Prague 2023

FINDING VERTEX-DISJOINT CYCLE COVER OF UNDIRECTED GRAPH USING THE LEAST-SQUARES METHOD

Jan Lamač, Miloslav Vlasák

Faculty of Civil Engineering, Czech Technical University in Prague
Thákurova 7, 166 29 Prague 6, Czech Republic
jan.lamac@cvut.cz, miloslav.vlasak@cvut.cz

Abstract: We investigate the properties of the least-squares solution of the system of equations with a matrix being the incidence matrix of a given undirected connected graph G and we propose an algorithm that uses this solution for finding a vertex-disjoint cycle cover (2-factor) of the graph G.

Keywords: cycle cover, 2-factor, Hamiltonian cycle, incidence matrix, least-square method

MSC: 05C50, 05C38, 93E24

1. Introduction

Finding a vertex-disjoint cycle cover (called a 2-factor) of a given undirected graph G consists in finding a set of disjoint cycles which are subgraphs of G and contain all vertices of G (see Figure 1). It is well known that a 2-factor of an undirected 2-factorable graph can be found in polynomial time by finding a perfect matching in some larger graph (cf. [10]). When we prescribe further conditions (e.g. number of components, minimal cycle length) the problem of finding a 2-factor becomes NP-hard (cf. [4]). This includes a 2-factor formed by one component only, i.e. the Hamiltonian cycle of the graph G.

In this paper we investigate the properties of the least-squares solution of the system of equations with a matrix being the incidence matrix of the given undirected connected[1] graph G and propose an algorithm that uses this solution for finding a 2-factor of the graph G. In this algorithm we successively erase the edges from the graph until we obtain the desired 2-factor of the graph. For determination of which edge will be erased we employ and test three strategies: S1, S2 and S3.

DOI: 10.21136/panm.2022.10

[1] All the strategies considered can be easily extended to disconnected graphs.

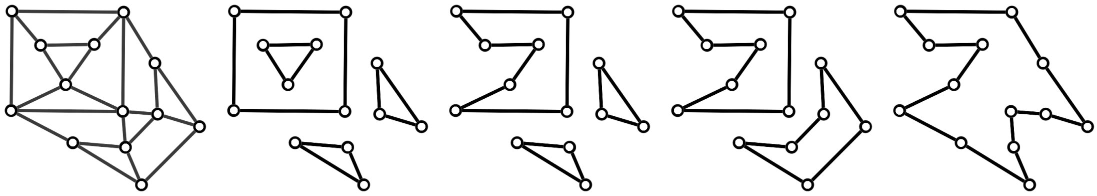

Figure 1: Graph G (left) and its vertex-disjoint cycle covers (2-factors). The last one is the Hamiltonian cycle of the graph G.

2. Graph, its representation and notation

By graph G we consider an ordered pair $G = (V, E)$, where

$$\begin{aligned} V &= V(G) = \{v_1, v_2, \ldots, v_n\} \\ &\quad \text{is a set of vertices of graph } G \text{ and} \\ E &= E(G) = \{e_1, e_2, \ldots, e_m\} \subseteq \binom{V}{2}, \quad e_j = \{v_k, v_l\},\ k \neq l, \\ &\quad \text{is a set of edges of the graph } G. \end{aligned}$$

We denote by $B \in \{0,1\}^{n\times m}$ the incidence matrix of G satisfying $B_{ij} = 1$ if $v_i \in e_j$ and $B_{ij} = 0$ if $v_i \notin e_j$. Arbitrary set of edges can be represented by the vector $x \in \{0,1\}^{m\times 1}$, which is a characteristic vector of the set $X \subseteq E$ satisfying $x_i = 1$ if $e_i \in X$ and $x_i = 0$ otherwise. If we want to refer to a particular edge $e \in X$ we also use a notation $[x]_e$ (instead of x_i). Further, we denote by $B_e \in \{0,1\}^{n\times(m-1)}$ the matrix obtained from B by deleting the column corresponding to the edge e. Similarly, we denote by x_e the vector that we obtain from x by deleting $[x]_e$. Finally, 1_k stands for a column vector formed by k ones.

Using this notation we may define the vertex-disjoint cycle cover x of the graph G being any set of edges satisfying

$$x \in \{0,1\}^{m\times 1} \quad \& \quad 1_m^T x = n \quad \& \quad Bx = 2 \cdot 1_n. \tag{1}$$

While the second condition ensures the cycle cover contains n edges, the third one guarantees that each vertex coincides with exactly 2 edges.

3. Basic properties of the vector x_{LS}

The least-square solution of the system $Bx = 2\cdot 1_n$ is defined using the Moore–Penrose pseudo-inverse of the matrix B (see e.g. [8]) as follows

$$x_{LS} = B^\dagger(2\cdot 1_n) = 2\cdot B^\dagger 1_n. \tag{2}$$

In this section we investigate the properties of the vector x_{LS}.

Lemma 1. *Let the graph G be non-bipartite, then the least-square solution x_{LS} of the system of equations $Bx = 2 \cdot 1_n$ satisfies*

$$1_m^T x_{LS} \;=\; n. \tag{3}$$

For bipartite graph $G = (V_1 \cup V_2, E)$ with $|V_1| = n_1$ and $|V_2| = n_2$ there holds

$$1_m^T x_{LS} \;=\; \frac{4n_1n_2}{n}. \tag{4}$$

Proof. Since the rows of the incidence matrix B are linearly independent for non-bipartite connected graphs (cf. [11]), the pseudo-inverse of the matrix B satisfies $B^\dagger = B^T(BB^T)^{-1}$ and, thus, the least-square solution x_{LS} satisfies $Bx_{LS} = BB^T(BB^T)^{-1}(2 \cdot 1_n) = 2 \cdot 1_n$. Consequently, there holds

$$2 \cdot n \;=\; 2 \cdot 1_n^T 1_n \;=\; 1_n^T(2 \cdot 1_n) \;=\; 1_n^T(Bx_{LS}) \;=\; (B^T 1_n)^T x_{LS} \;=\; 2 \cdot 1_m^T x_{LS}. \tag{5}$$

If G is bipartite (and connected), then the rank of B is $n-1$ (cf. [11]) and its rows are linearly dependent. Hence, one can order columns of B^T (i.e. vertices of G) so that

$$B^T w = 0 \quad \text{for} \quad w = (\underbrace{1, 1, \ldots, 1}_{n_1-\text{times}}, \underbrace{-1, -1, \ldots, -1}_{n_2-\text{times}})^T. \tag{6}$$

Considering the singular value decomposition of B in the form $B = U\Sigma V^T$, the Moore-Penrose inverse of B has a form $B^\dagger = V\Sigma^\dagger U^T$ with $\Sigma_{nn} = \Sigma^\dagger_{nn} = 0$ being the singular value corresponding to the left singular vector[2] $u = \frac{1}{\|w\|} w = \frac{1}{\sqrt{n}} w$, i.e. to the last column of the matrix U. Consequently, for bipartite graphs there holds

$$\begin{aligned}
1_m^T x_{LS} \;&=\; \frac{1}{2}(B^T 1_n)^T x_{LS} \;=\; \frac{1}{2}(B^T 1_n)^T(2B^\dagger 1_n) \;=\; 1_n^T B B^\dagger 1_n \\
&=\; 1_n^T U\Sigma V^T V \Sigma^\dagger U^T 1_n \;=\; 1_n^T U \Sigma\Sigma^\dagger U^T 1_n \;=\; 1_n^T U(I_n - e_n e_n^T)U^T 1_n \\
&=\; 1_n^T 1_n - (1_n^T U e_n)^2 \;=\; n - (1_n^T u)^2 \;=\; n - \frac{1}{n}(1_n^T w)^2 \;=\; n - \frac{(n_1 - n_2)^2}{n} \\
&=\; n - \frac{(n_1+n_2)^2 - 4n_1n_2}{n} \;=\; \frac{4n_1n_2}{n},
\end{aligned} \tag{7}$$

where we applied the equality $B^T 1_n = 2 \cdot 1_m$ resulting from the fact that each row of B^T contains exactly two ones (i.e. each edge connects two vertices). ☐

Lemma 2. *Let $x \in \{0,1\}^m$ be a vertex-disjoint cycle cover, then*

$$\|x - x_{LS}\|^2 \;=\; n - \|x_{LS}\|^2. \tag{8}$$

[2]In the whole paper by the expression $\|x\| = \sqrt{x^T x}$ we denote the standard Euclidean norm of the vector x.

Proof. Since $Bx = 2 \cdot 1_n$ and $x_{LS} = 2B^\dagger 1_n$, there holds

$$\begin{aligned}
\|x - x_{LS}\|^2 &= (x - 2B^\dagger 1_n)^T (x - 2B^\dagger 1_n) = \\
&= \|x\|^2 - 4x^T B^\dagger 1_n + 4 \cdot 1_n^T (B^\dagger)^T B^\dagger 1_n \\
&= n - 4x^T B^\dagger B B^\dagger 1_n + 4 \cdot 1_n^T (B^\dagger)^T B^\dagger 1_n \\
&= n - 4x^T (B^\dagger B)^T B^\dagger 1_n + 4 \cdot 1_n^T (B^\dagger)^T B^\dagger 1_n \\
&= n - 4(B^\dagger B x)^T B^\dagger 1_n + 4 \cdot 1_n^T (B^\dagger)^T B^\dagger 1_n \\
&= n - 8(B^\dagger 1_n)^T B^\dagger 1_n + 4 \cdot 1_n^T (B^\dagger)^T B^\dagger 1_n = n - \|x_{LS}\|^2, \qquad (9)
\end{aligned}$$

where we applied the equalities $B^\dagger = B^\dagger B B^\dagger$ (see e.g. [8]) and $Bx = 2 \cdot 1_n$. □

Corollary 3. *Let the graph G with the incidence matrix B contain a 2-factor. Then the least-square solution to the system $Bx = 2 \cdot 1_n$ satisfies*

$$\|x_{LS}\|^2 \le n. \qquad (10)$$

When $\|x_{LS}\|^2 = n$, then $x = x_{LS}$ is the only 2-factor of the graph G.

Proof. The inequality (10) follows from the inequality $n - \|x_{LS}\|^2 = \|x - x_{LS}\|^2 \ge 0$. When $\|x_{LS}\|^2 = n$, we obtain $\|x - x_{LS}\|^2 = n - \|x_{LS}\|^2 = 0$ for any 2-factor x. This is possible for $x = x_{LS}$ only. □

Corollary 4. *All 2-factors x satisfy*

$$x^T x_{LS} = \|x_{LS}\|^2. \qquad (11)$$

Proof. The equality (11) results from the fact that $\|x\|^2 = n$ and from the relation

$$2 \cdot x^T x_{LS} = \|x\|^2 + \|x_{LS}\|^2 - \|x - x_{LS}\|^2 = \|x\|^2 + \|x_{LS}\|^2 - n + \|x_{LS}\|^2. \qquad (12)$$

□

Remark 5. *From the equality (9) it follows that each 2-factor x lies on the m-dimensional sphere centered in x_{LS} with the radius $\sqrt{n - \|x_{LS}\|^2}$. Thus, assuming the graph G contains k different 2-factors x_i, $i = 1, 2, \ldots, k$, with the mean value $\overline{x}_k = \frac{1}{k}\sum_{i=1}^k x_i$, the multi-dimensional version of the Berry–Esseen theorem gives*

$$\|x_{LS} - \overline{x}_k\| \le C \cdot \frac{\sqrt{n - \|x_{LS}\|^2}}{\sqrt{k}} \xrightarrow{k \to \infty} 0, \qquad (13)$$

providing x_i are independent and identically distributed on the sphere (see e.g. [2] and [3]). However, we expect that this assumption is not fulfilled in this case and the proper formulation for 2-factors needs more investigation. Nevertheless, from the experiments it follows that x_{LS} is, indeed, a good approximation of $\overline{x}$ for large k and that a result similar to (13) really holds (see Figure 2).

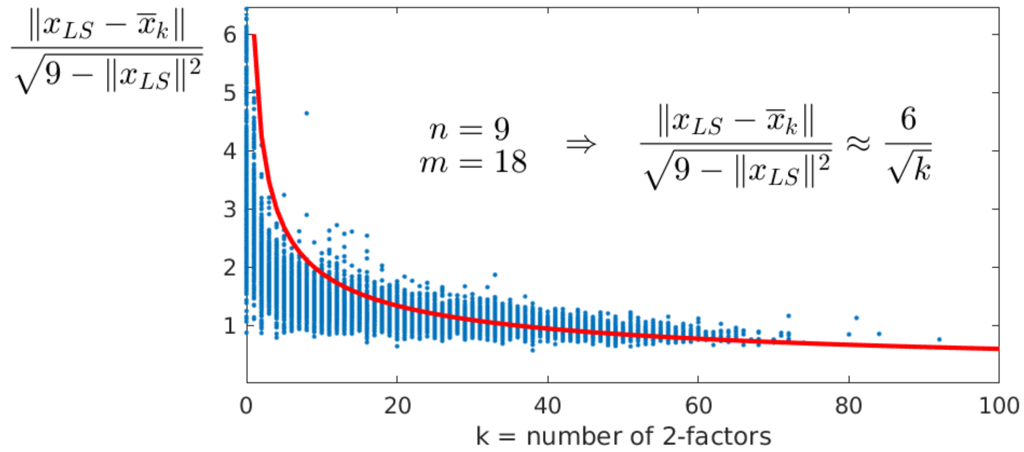

Figure 2: For graphs with a large number of 2-factors the least-square solution x_{LS} is a good approximation of $\overline{x}_k$. Here we considered all non-isomorphic graphs (see [7]) on $n = 9$ vertices and $m = 18$ edges. Each point corresponds to a single graph. The curve is a graph of the function $6/\sqrt{k}$.

Remark 6. *From the equality* (11) *it also follows that finding a* 2*-factor can be interpreted as finding* n *entries of the vector* x_{LS} *that sum up to* $\|x_{LS}\|^2$*. Hence, we obtain the so-called* 0-1 *knapsack problem with the prescribed number of items to include in a collection (for more details about knapsack problems, see e.g.* [5]*).*

Example 7. *Let us consider a graph formed by* 7 *vertices and* 9 *edges depicted on the Figure* 3 *(top left). It contains two* 2*-factors. If we compute the respective vector* x_{LS} *we realize that the values of* x_{LS} *entries are significantly higher for edges belonging to both* 2*-factors. This observation leads us to the strategy (S1) consisting in removing edges with the smallest* x_{LS}*-value.*

4. Sufficient condition

The following theorem provides a useful tool for determining which edge can be removed from the graph. Unfortunately, in most cases, none of the edges satisfy the condition (14) (see Figure 5a). In that situation we remove the edge with the highest value of the left-hand side of (14) (strategy S2).

Theorem 8. *Let* G *be a graph with the incidence matrix* B*, let* $e \in E(G)$ *be any edge such that* $G \setminus e$ *is a connected non-bipartite graph and let* x_{LS} *be the least-square solution to the system* $Bx = 2 \cdot 1_n$*. If there is*

$$\frac{(1 - [x_{LS}]_e)^2}{1 - e^T(BB^T)^{-1}e} \quad > \quad n - \|x_{LS}\|^2, \tag{14}$$

then the following implication holds

$$G \text{ has a 2-factor} \quad \Rightarrow \quad G \setminus e \text{ has a 2-factor.} \tag{15}$$

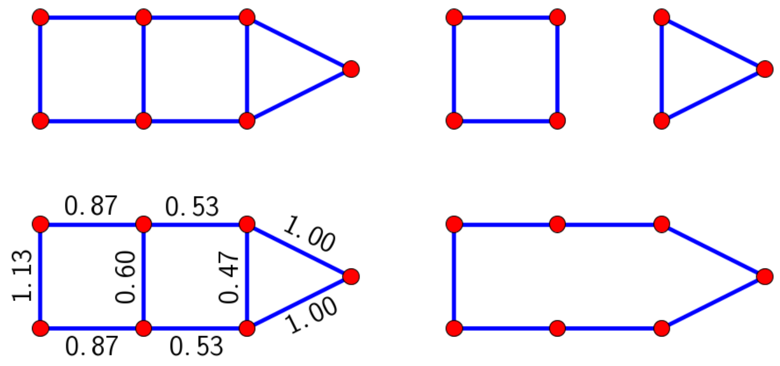

Figure 3: An example of a graph (top left) with two 2-factors (right). The values of x_{LS} entries (bottom left) are significantly higher for edges belonging to both 2-factors.

Proof. For a contradiction, let us suppose that the inequality (14) holds and all 2-factors $x \in \{0,1\}^{m\times 1}$ of the graph G satisfy $[x]_e = 1$. Let us denote by $B_e \in \{0,1\}^{n\times(m-1)}$ the matrix obtained from B by deleting the column corresponding to the edge e. Similarly, let us denote by $x_{LS,e}$ the vector that we obtain from x_{LS} by deleting the entry corresponding to the edge e. If we choose any 2-factor x of the graph G then the following equality holds

$$Bx \;=\; B_e x_e + e \;=\; 2\cdot 1_n \;=\; Bx_{LS} \;=\; B_e x_{LS,e} + [x_{LS}]_e \cdot e, \tag{16}$$

where x_e is obtained from x by deleting the entry corresponding to the edge e.

Hence, $B_e(x_{LS,e} - x_e) = (1 - [x_{LS}]_e)\cdot e$ and for the least-square solution z_{LS} of the system $B_e z = (1 - [x_{LS}]_e)\cdot e$ there holds

$$\|z_{LS}\|^2 \;=\; (1-[x_{LS}]_e)^2\|B_e^\dagger e\|^2 \;\le\; \|x_{LS,e} - x_e\|^2 \;=\; \|x_{LS} - x\|^2 - ([x_{LS}]_e - 1)^2. \tag{17}$$

Thus, using the equality (9) we obtain an estimate

$$\|B_e^\dagger e\|^2 \;\le\; \frac{n - \|x_{LS}\|^2}{([x_{LS}]_e - 1)^2} - 1. \tag{18}$$

It remains to simplify the expression $\|B_e^\dagger e\|^2 = e^T(B_e B_e^T)^{-1}e$. For this purpose we apply the Sherman-Morrison formula (see [9])

$$(B_e B_e^T)^{-1} \;=\; (BB^T - e^T e)^{-1} \;=\; (BB^T)^{-1} + \frac{(BB^T)^{-1}ee^T(BB^T)^{-1}}{1 - e^T(BB^T)^{-1}e} \tag{19}$$

and obtain

$$\begin{aligned} e^T(B_e B_e^T)^{-1}e \;&=\; e^T(BB^T)^{-1}e + \frac{e^T(BB^T)^{-1}ee^T(BB^T)^{-1}e}{1 - e^T(BB^T)^{-1}e} \\ &=\; \frac{e^T(BB^T)^{-1}e}{1 - e^T(BB^T)^{-1}e} \;=\; \frac{1}{1 - e^T(BB^T)^{-1}e} - 1. \end{aligned} \tag{20}$$

Hence

$$\frac{1}{1 - e^T(BB^T)^{-1}e} \leq \frac{n - \|x_{LS}\|^2}{([x_{LS}]_e - 1)^2}, \tag{21}$$

which is in contradiction with the assumption (14). □

Remark 9. *A similar inequality to* (14) *can be derived in the case when* $G \setminus e$ *is a bipartite graph using formulas for the Moore–Penrose inverse of modified matrices, for more details see e.g.* [1] *or* [6].

5. Minimizing the length of x_{LS}

With the aid of [7] we have computed the values of $\|x_{LS}\| = \|x^G_{LS}\|$ for all connected non-isomorphic graphs G on 9 vertices with a minimal vertex degree 2 (see Figure 4) and found out that the value of $\|x^G_{LS}\|$ is significantly smaller for graphs G containing a large number of 2-factors. This has lead us to the strategy (strategy S3) consisting in removing the edge $e \in E(G)$ with a property

$$e = \arg\min_{\widehat{e} \in E(G)} \|x^{G\setminus\widehat{e}}_{LS}\| \tag{22}$$

(see Figure 5*b* for an example of an application of the strategy S3).

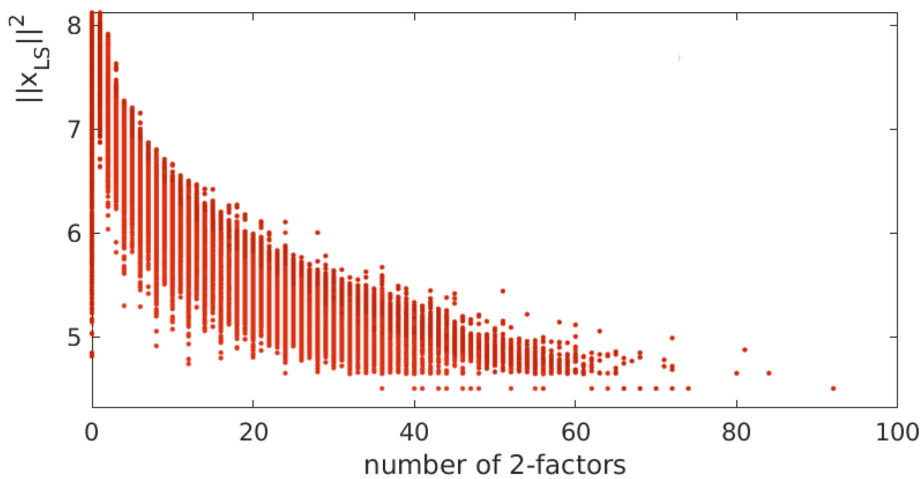

Figure 4: For graphs with a large number of 2-factors the norm of the least-square solution x_{LS} is significantly smaller. Thus, in order to preserve a maximum number of 2-factors in the graph we always try to remove the edge that minimizes $\|x^{G\setminus e}_{LS}\|$ (strategy S3). Here the results for connected non-isomorphic graphs with 9 vertices and minimal vertex degree 2 are shown (each point represents one graph).

For computing $\|x^{G\setminus e}_{LS}\|$ we use the following lemma.

Lemma 10. *Let G be a graph with the incidence matrix B and let $e \in E(G)$ be any edge such that $G \setminus e$ is a connected non-bipartite graph. Further, let x_{LS}^{G} be the least-square solution to the system $Bx = 2 \cdot 1_n$ and let $x_{LS}^{G\setminus e}$ be the least-square solution to the system $B_e x = 2 \cdot 1_n$. Then there holds*

$$\|x_{LS}^{G\setminus e}\|^2 - \|x_{LS}^{G}\|^2 = \frac{[x_{LS}^{G}]_e^2}{1 - e^T(BB^T)^{-1}e}. \tag{23}$$

Proof. We employ the relations from the equalities (9) and (19) and obtain

$$\begin{aligned}\|x_{LS}^{G\setminus e}\|^2 &= \|2B_e^{\dagger}1_n\|^2 = 4 \cdot 1_n^T(B_eB_e^T)^{-1}1_n \\ &= 4 \cdot 1_n^T(BB^T)^{-1}1_n + \frac{4 \cdot 1_n^T(BB^T)^{-1}ee^T(BB^T)^{-1}1_n}{1 - e^T(BB^T)^{-1}e} \\ &= \|x_{LS}^{G}\|^2 + \frac{(2 \cdot e^T(BB^T)^{-1}1_n)^2}{1 - e^T(BB^T)^{-1}e} = \|x_{LS}^{G}\|^2 + \frac{[x_{LS}^{G}]_e^2}{1 - e^T(BB^T)^{-1}e},\end{aligned} \tag{24}$$

where we used the fact that

$$2 \cdot e^T(BB^T)^{-1}1_n = [2 \cdot B^T(BB^T)^{-1}1_n]_e = [2 \cdot B^{\dagger}1_n]_e = [x_{LS}^{G}]_e \tag{25}$$

is the entry of the vector x_{LS}^{G} corresponding to the edge e. □

Remark 11. *As in the case of the inequality (14) similar equality to (23) can be derived when $G \setminus e$ is a bipartite graph using formulas for the Moore–Penrose inverse of modified matrices, for more details see again e.g. [1] or [6].*

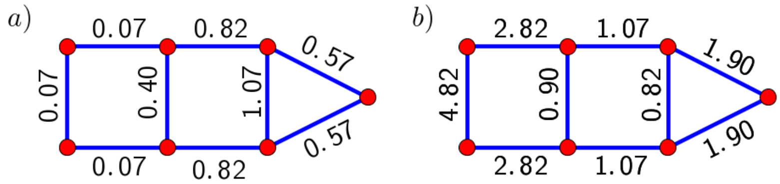

Figure 5: For the edges of the graph from the Example 7 we compute the values of the left-hand side of the inequality (14) (Figure 5a). The edge with the highest value (1.07) will be removed (strategy S2). Unfortunately, $n - \|x_{LS}\|^2 = 1.07$ in this case, hence, the condition (14) is not fulfilled. Analogously, we compute the values of the right-hand side of the equality (23) (Figure 5b). Then the edge with the smallest value (0.82) will be removed (strategy S3) in order to minimize the norm of $x_{LS}^{G\setminus e}$. Thus, for this graph, all three strategies lead to the deletion of the same edge.

6. Numerical experiments

We consider 10 000 randomly generated graphs with 32 vertices and 64 edges containing a Hamiltonian cycle. For each graph we apply all three strategies and successively remove edges. In each row of the Table 1 one can find results for each strategy employed. The numbers of graphs for which the algorithm failed are stored in the second column of the table. In the third to seventh column one can find the numbers of graphs for which the algorithm succeeded and the resulting 2-factor is formed by 1 to 5 components. In the last column an average number of components for each successfully ended strategy is shown.

strategy	failed	1 cmp	2 cmp	3 cmp	4 cmp	5 cmp	avg cmp
S1	372	4498	4297	774	59	0	1.6255
S2	59	5487	3337	930	161	26	1.5818
S3	3582	2096	2816	1232	247	27	1.9550

Table 1: Numerical results for all considered strategies.

7. Conclusion

Numerical experiments show that all three strategies considered were successful in more than 50 percent of all cases and from this point od view we shall say that the considerations from which they were derived were justified. The best result has been achieved by the strategy S2, which succeeded 99.41 percent of the time. The combination of all three strategies, as well as the involvement of some properties of the graph in the edge deletion decision, will be the subject of the future research.

Acknowledgements

The research was partially supported by the Czech Science Foundation grant 20-14736S and by project No. CZ.02.1.01/0.0/0.0/16_019/0000778 financially supported by the European Regional Development Fund (Center of Advanced Applied Sciences – CAAS).

References

[1] Baksalary, J. K , Baksalary, O. M. and Trenkler, G.: A revisitation of formulae for the Moore–Penrose inverse of modified matrices, Linear Algebra and its Applications, 372 (2003), pp. 207-224.

[2] Bentkus, V.: A Lyapunov type bound in R^d, Teor. Veroyatnost. i Primenen., **49** (2), 400–410, 2004.

[3] Esseen, C.-G.: On the Liapounoff limit of error in the theory of probability. Ark. Mat. Astr. Fys. 28A, (1942). no. 9, 19 pp.

[4] Garey, M. and Johnson, D. S.: Computers and Intractability: A Guide to the Theory of NP-Completeness, A Series of Books in the Mathematical Sciences. San Francisco, Calif.: W. H. Freeman and Co., 1979.

[5] Kellerer, H., Pferschy, U. and Pisinger, D.: Knapsack problems, Springer, 2004.

[6] Meyer, Jr. and Carl D.: Generalized Inversion of Modified Matrices, SIAM Journal on Applied Mathematics, **24** (3), 315–323, 1973.

[7] McKay, B.: Collection of combinatorial data - list of simple connected graphs: http://users.cecs.anu.edu.au/~bdm/data/graphs.html

[8] Penrose, R.: A generalized inverse for matrices. Mathematical Proceedings of the Cambridge Philosophical Society, **51**(3), 406-413, 1955.

[9] Sherman, J. and Morrison, W. J.: Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix, Annals of Mathematical Statistics, **21** (1), 124–127, 1950.

[10] Tutte, W.: A Short Proof of the Factor Theorem for Finite Graphs, Canadian Journal of Mathematics, 6 (1954), 347–352.

[11] Van Nuffelen, C.: On the incidence matrix of a graph, IEEE Transactions on Circuits and Systems, **23** (9), pp. 572-572, 1976.