

Simplex Embedding Method in Decomposition of Large Sparse Convex Nondifferentiable Optimization Problems ^{*}

Anton Kolosnitsyn

Melentiev Energy Systems Institute SB RAS,
130 Lermontov Str., Irkutsk, Russia
ankolos25@mail.ru

Abstract. We consider an adaptation of the simplex embedding method to the decomposition of large-scale convex problem with sparse block-wise constraint matrix. According to the Lagrangean relaxation technique such problem is reduced to the maximization of nondifferentiable concave function with subgradient that can be easily calculated at each feasible point. Simplex embedding method with modifications gives us the appropriate performance to optimize nondifferentiable function that will be demonstrated on the numerical tests.

Keywords: Decomposition · Lagrangean relaxation · Simplex embedding method · Nondifferentiable optimization

1 Introduction

Mathematical models are widely spread in different areas of our life and appear in economics, business, engineering and social sciences. To get proper vision of real object functioning and obtain reliable optimizing model parameters one can use the mathematical programming technique in which a system is represented by an objective function and several constraints. There are lots of mathematical programming problems that have huge number of variables and constraints due to the system they are related to. As the examples we can notice multidivisional problems with large number of subsystems, combinatorial problems containing large amount of model alternatives, dynamic problems with many replicating constraints or variables and stochastic problems with large amount of possibilities involved [10].

When we come across the large or too complicated mathematical programming problems it is sometimes necessary or more convenient to apply decomposition approaches. As a rule large-scale problems have a block-wise structure of

^{*} The reported study was supported by RFBR, research project No. 18-07-01432.

Copyright © by the paper's authors. Copying permitted for private and academic purposes.
In: S. Belim et al. (eds.): OPTA-SCL 2018, Omsk, Russia, published at <http://ceur-ws.org>

matrix constraints and can be split into less dimension problems using different decomposition methods.

Dantzig-Wolfe decomposition [8] and Benders decomposition [5, 22] are the most representative methods of all decomposition techniques. A comprehensive survey up to 1979 year of developing these methods is given in [18]. Decomposition procedures in terms of algorithmic applications and questions about convergence of decomposition methods are discussed in [9].

Our interest within this article is concentrated on the Lagrangean relaxation technique for the decomposition problems. Such approach is used for reducing the initial large-scale problem to the maximizing of the nondifferentiable concave function with the subgradient that can be easily calculated at each feasible point. Detailed description of Lagrangean relaxation technique is given in [15, 6, 16]. In considering mixed integer nonlinear programming we refer to [17] where Lagrangean relaxation method is applied. Once we apply the Lagrangean relaxation technique the nondifferentiable optimization method must be chosen. In [6, 11] different methods of nondifferentiable optimization problems (subgradient methods [24], cutting plane methods [20], bundle methods [3]) are represented in terms of their applying to the decomposition problems.

One of the significant source of the large-scale optimization problems is a stochastic nature of the complex systems under study. In [23] it is represented the applying of decomposition methods to the convex stochastic programming problems.

Notice at last that decomposition is commonly used in practical applications as network design problems [4, 7], network utility problems [21] and producing problems [6].

The current work continues the series of articles [2, 12–14] devoted to the simplex embedding method modifications and applications of the method to solving convex optimization problems. We present the adaptation of simplex embedding method to solving large sparse convex optimization problems by means of decomposition technique. This technique will be introduced in the section 3. We use the simplex embedding method modification with cutting plane shift that demonstrated quite good performance according to the numerical experiments implemented in [14]. The simplex embedding method with corresponding algorithm of implementation will be presented in the section 4. We will give the algorithm of combining the method modification with the decomposition procedure in the section 5. The results of numerical experiment will be demonstrated in the section 6. We conclude the paper with a short discussion of results in the section 7.

2 Problem Statement

We aim to solve the following convex optimization problem

$$\begin{aligned} & \text{minimize} && f(x) \\ & \text{subject to} && Ax \leq b, \\ & && Dx \leq d, \\ & && x \geq 0, \end{aligned} \tag{1}$$

where matrix A is the $m_1 \times n$ -dimensional matrix, D is the $m_2 \times n$ -dimensional matrix that has block-wise structure with K blocks

$$D = \begin{bmatrix} D_1 & & & 0 \\ & D_2 & & \\ & & \ddots & \\ 0 & & & D_K \end{bmatrix},$$

$x \in R^n, b \in R^{m_1}, d \in R^{m_2}$. Function $f : R^n \rightarrow R$ is convex function that can be represented as a sum of separable items

$$f(x) = f_1(x_1) + f_2(x_2) + \dots + f_K(x_K).$$

Vectors x, d can be split into subvectors that are corresponded to one of the k -th block, $k = 1, \dots, K$:

$$\begin{aligned} x &= (x_1, \dots, x_K), \\ d &= (d_1, \dots, d_K), \end{aligned}$$

matrix A is split into submatrixes (A_1, \dots, A_K) .

Let us introduce polytopes $X_k, k = 1, \dots, K$ for notation convenient:

$$X_k = \{x_k \mid D_k x_k \leq d_k; x_k \geq 0\}, \quad k = 1, \dots, K.$$

Then the problem (1) can be represented in the following form:

$$\begin{aligned} & \text{minimize} && \sum_{k=1}^K f_k(x_k), \\ & \text{subject to} && \sum_{k=1}^K A_k x_k \leq b, \\ & && x_k \in X_k, \quad k = 1, \dots, K. \end{aligned} \tag{2}$$

The problem (2) corresponds to so called complicating constraints problem. It means that the structure of constraints is represented not only by the separable blocks $D_k x_k \leq d_k, k = 1, \dots, K$ but also it is represented by constraints $Ax \leq b$ joining these blocks. We will consider only problems with this constraint structure. Figure 1 reflects schematically the structure of such problems.

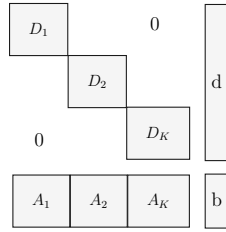


Fig. 1. Block-wise structure of the optimization problem

3 Lagrangean Relaxation

Let us describe the Lagrangean relaxation technique for solving the problem (2).

We associate the Lagrangean multiplier $\lambda_i \geq 0, i = 1, \dots, m_1$ with each complicating constraint and form the Lagrangean function:

$$L(x, \lambda) = f(x) + \lambda^T (Ax - b) = \sum_{k=1}^K (f_k(x_k) + \lambda^T A_k x_k) - \lambda b.$$

A dual function is defined in the following way

$$\omega(\lambda) = \min_{x \in X} \{L(x, \lambda)\} = \sum_{k=1}^K \min_{x_k \in X_K} \{f_k(x_k) + \lambda^T A_k x_k\} - \lambda b,$$

$$X = X_1 \times X_2 \times \dots \times X_K.$$

Notice that the calculation of the dual function value is split into solving the optimization subproblems

$$\begin{aligned} &\text{minimize} && f_k(x_k) + \lambda^T A_k x_k, \\ &\text{subject to} && D_k x_k \leq d_k, \\ &&& x_k \geq 0. \end{aligned} \tag{3}$$

Each of subproblem from (3) is the less dimensional convex problem in comparison with initial problem (2). Using performance of multiprocessor systems one can get an appropriate accelerate in calculations. To find the optimal solution of the problem

$$\begin{aligned} &\text{maximize} && \omega(\lambda), \\ &\text{subject to} && \lambda \geq 0, \end{aligned} \tag{4}$$

we need to consider several important properties of the function ω [15, 16].

Let us first give the definition of a saddle point for the function L .

Definition 1. A point (x^0, λ^0) is said to be a constrained saddle point for $L(x, \lambda)$ if it satisfies

$$\begin{aligned} L(x^0, \lambda^0) &\leq L(x, \lambda^0), \quad \text{for all } x \in X, \\ L(x^0, \lambda^0) &\geq L(x^0, \lambda), \quad \text{for all } \lambda \geq 0. \end{aligned}$$

The main properties of the function ω is given below.

1. Function ω is concave and nondifferentiable.
2. If the problem (2) has the solution x^* with finite value then there exists a saddle point and for all $\lambda \geq 0$ and for all solution x of the problem (2) we have the following expression

$$\omega(\lambda) \leq \omega(\lambda^*) = f(x^*) \leq f(x),$$

where λ^* is the optimal solution of the dual problem.

3. Let the vector $\hat{x}_k, k = 1, \dots, K$ be the optimal solution of the problem (3), then the vector

$$\gamma = \sum_{k=1}^K A_k \hat{x}_k - b \tag{5}$$

defines the subgradient of the function ω at the point λ .

A correspondence between the solution λ^* for the dual problem and the solution x^* for the initial problem (2) is set by the following theorems [15].

Theorem 1. Let $\lambda^0 \geq 0$. A point (x^0, λ^0) is a constrained saddle point for $L(x, \lambda)$ if and only if

$$\begin{aligned} x^0 &\text{ minimizes } L(x, \lambda^0) \text{ over } X, \\ Dx^0 &\leq d, \\ \lambda^0 (Ax^0 - b) &= 0. \end{aligned} \tag{6}$$

Theorem 2. If (x^0, λ^0) is a saddle point for $L(x, \lambda)$, then x^0 solves the primal problem (2).

Since a Subgradient of the Dual Function ω at Each Point $\lambda \geq 0$ Can Be Calculated It is Possible to Apply One of the Nondifferentiable Convex Optimization Method (concave in Our Case). In the Section Below We Introduce the Simplex Embedding Method Adapted for the Decomposition Technique.

4 Simplex Embedding Method

Simplex embedding method [1] is a centered cutting plane method that uses n -dimensional simplexes to localize the problem solution. The main idea of the method is similar to the well-known ellipsoid method [19]. To find the optimization problem solution one need to embed the feasible set of the problem to initial simplex. Then it is drawn the cutting plane through a simplex center.

This cutting plane divides simplex into two parts. The key step of the method is in embedding the truncated simplex containing the problem solution into a new minimal volume simplex. Then all the procedures repeat one more time for obtained minimal volume simplex. Implementation of such steps continues up to getting a simplex with a quite small volume. Then the algorithm is terminated.

To start the detailed method description we will give some important definitions.

Definition 2. Let S be the subset of R^n . We call S the simplex if it has the following form

$$S = \left\{ x \in R^n : x = x^0 + \sum_{i=1}^n \sigma_i (x^i - x^0), \sigma_i \geq 0, \sum_{i=1}^n \sigma_i \leq 1 \right\} \quad (7)$$

where x^0 is the simplex vertex, $(x^1 - x^0, \dots, x^n - x^0)$ are the simplex edges that form basis in R^n .

Definition 3. The center of the simplex S is the point $x^c \in S$ defined by the expression

$$x_c = \frac{1}{n+1} (x^1 + \dots + x^n). \quad (8)$$

Definition 4. The volume of the simplex S is defined by the next formula

$$V(S) = \frac{1}{n!} |\det(\bar{X})|,$$

where \bar{X} is the $n \times n$ dimension matrix. The columns of this matrix are represented by the vectors x^1, x^2, \dots, x^n .

Definition 5. Each hyperplane of the form

$$L = \{x : g^T (x - x^c) = 0\}$$

is called the cutting hyperplane passing through the center x^c in the simplex S .

Definition 6. The vertex x^i of the simplex S is said not to be cut off by the cutting hyperplane L if

$$\alpha_i = g^T (x^i - x^c) < 0, \quad (9)$$

and this vertex is cut off if $\alpha_i \geq 0$, $i = 0, \dots, n$.

The immersion procedure of truncated simplex into a new minimal volume simplex is the important principle of the method. It provides the convergence to the optimal problem solution. According to (7) let us define a new minimal volume simplex containing a truncated simplex S_G in the following form:

$$S(\tau) = \left\{ x : x = x^0 + \tau_1 \sigma_1 (x^1 - x^0) + \dots + \tau_n \sigma_n (x_n - x^0), \right. \\ \left. \sum_{i=1}^n \sigma_i \leq 1, \sigma_i \geq 0, i = 1, \dots, n. \right\}$$

The theorem below contains the description of minimal volume simplex constructing [1].

Theorem 3. *The minimal volume simplex $S(\tau^*)$ containing a simplex S_G is defined by the parameters*

$$\tau^* = 1 + \alpha_i t^*, \quad i = 1, \dots, n,$$

where t^* is the solution of one-dimensional optimization problem

$$\min_{0 \leq t \leq -1/\alpha_0} \prod_{i=1}^n (1 + \alpha_i t)^{-1}, \quad \alpha_0 = \min_{0 \leq i \leq n} \alpha_i.$$

The following theorem gives the estimation of the method convergence [1].

Theorem 4. *Let $S \subset R^n$ be the n -dimensional simplex, x^c is the center of the simplex, $S_G = \{x \in S, g^T(x - x^c) \leq 0\}$ is truncated simplex. Simplex S_G can be immersed into a simplex S^* so that the following relation between the volumes $V(S)$ and $V(S^*)$ of the simplexes S and S^* is fulfilled:*

$$\frac{V(S^*)}{V(S)} \leq \begin{cases} \frac{1}{2}, & k = 1, \\ \left(\frac{k}{k+1}\right)^k \left(\frac{k}{k-1}\right)^{k-1}, & 2 \leq k \leq n, \end{cases} \quad (10)$$

where k is the amount of saved vertices after drawing a cutting hyperplane through the simplex center, S^* is the minimal volume simplex constructed according to the theorem 3.

Convergence estimation obtained in (10) depends only on the amount of saved simplex vertices. The less vertices are saved the higher convergence rate of the method is obtained. As soon as only one simplex vertex is saved the analogue of dichotomy method can be obtain.

Let us introduce the general algorithm for the simplex embedding method implementation to solve the optimization problem

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m, \end{aligned} \quad (11)$$

where $f_i(x)$, $i = 0, \dots, m$ are convex functions, $x \in R^n$.

Step 0. Choose an initial simplex S^0 with matrix X^0 so that S^0 contains a feasible set of the problem (11). Set the accuracy parameter ε and put $\rho^0 = 1$. Iteration k , $k = 0, 1, 2, \dots$, proceeds as follows.

Step 1. Find the simplex center x_c^k using the formula (8).

Step 2. Compute maximal residual

$$h_k = \max_{1 \leq i \leq m} \{0, f_i(x_c^k)\} = f_s(x_c^k).$$

Step 3. Find the cutting plane normal

$$a_k = \begin{cases} \partial f_0(x_c^k), & \text{if } h_k = 0, \\ \partial f_i(x_c^k), & \text{if } h_k > 0, \end{cases}$$

where $\partial f_i(x_c^k)$ is the subgradient of the convex function f_i at the point x_c^k .

Step 4. Construct the truncated simplex

$$S_G^k = \{x : x \in S_k, \langle a_k, x - x_c^k \rangle \leq 0\}.$$

Step 5. Find the parameters α^k using (9) and compute the stretch coefficients τ^k following the directions of the theorem 3. Define $p = \min_{0 \leq i \leq n} \alpha_i$.

Step 6. Find simplex S^{k+1} with matrix X^{k+1}

$$X_{ij}^{k+1} = \begin{cases} X_{pj}^k + \tau_i^k (X_{ij}^k - X_{pj}^k), & i \neq p, \quad i = 0, 1, \dots, n, \\ X_{ij}^k, & i = p, \quad i = 0, 1, \dots, n, \end{cases}$$

Step 7. Define the maximum edge of the simplex S^{k+1} :

$$\rho^{k+1} = \max_{i,j} \|v_i^k - v_j^k\| \quad i \neq j,$$

where v_i^k , $i = 0, \dots, n$ is the coordinate of the i -th vertex of the simplex S^k .

Step 8. If $\rho^{k+1} \leq \varepsilon$ then terminate the algorithm with ε -optimal solution x_c^k . Otherwise increment $k \rightarrow k + 1$ and move to the *Step 1*.

According to the results of the theorem 4 several modifications of the simplex embedding method were developed. The simultaneous introduction of several cutting planes into the simplex was considered in [2]. The technique of subdifferential description for a special class of convex nondifferentiable functions was represented in [12]. Such description was used to form a resulting cutting plane to cut off as much simplex vertices as possible. Constructing the minimal volume simplex around the certain set of points is considered in [13]. The inactive constraints identification by means of simplex embedding method and numerical comparison of the simplex embedding method modifications with cutting plane shift were considered in [14]. These modifications reduce the volume of truncated simplex containing the optimal solution and improves the method convergence.

To give an explanation of the cutting plane shift modification suppose we have a current simplex center x_c and the set of linear constraints $a_i x \leq b_i$, where $a_i, x \in R^n$, $i = 1, \dots, m$, $b \in R^m$. The modified algorithm of simplex embedding method has the following form.

Step 1. Calculate the residuals at the point x_c : $\delta_i = a_i x_c - b_i$, $i = 1, \dots, m$.

Step 2. Define the maximal residual and corresponding residual number:

$$\delta_{max} = \max_{i=1, \dots, m} \delta_i, \quad i_\delta = i \left| \max_{i=1, \dots, m} \delta_i = \delta_{max} \right.$$

Step 3. Define a new cutting plane

$$L_k = \{x : a_{i_\delta} x - b_{i_\delta} = 0\}.$$

Numerical tests represented in [14] showed better performance of the modified algorithm for solving the set of nondifferentiable optimization problems. The way of obtaining a subgradient for the problem (4) includes an additional computational procedure that we will describe in the next section in details. Such procedure prevents us from direct comparison with the results from the work [14]. Nevertheless simplex embedding method modification with cutting plane shift is applicable for solving the decomposition problem.

5 Decomposition Algorithm

To solve the problem (2) we can implement a two-level decomposition technique with the first level solving subproblems (3) and the second level adjusting the multipliers λ by means of simplex embedding method. The detailed algorithm of such technique is given below.

Step 0. Choose initial values $\lambda_0 \geq 0$. Iteration i , $i = 0, 1, 2, \dots$, proceeds as follows.

Step 1. Solve the subproblems (3) with $\lambda = \lambda_i$ obtaining the solution $x(\lambda_i)$.

Step 2. Calculate the subgradient of the function $\omega(\lambda)$:

$$\gamma_i = \sum_{k=1}^K A_k x_k(\lambda_i) - b.$$

Step 3. Input the obtained subgradient γ_i to the simplex embedding method to get a new point λ_i^* .

Step 4. If the simplex embedding method is terminated due to its inner stopping criterion then we terminate the decomposition algorithm with the solution λ_i^* of dual problem (4). Otherwise increment $i \rightarrow i + 1$, define $\lambda_{i+1} = \lambda_i^*$ and move to the *Step 1*.

6 Preliminary Numerical Experiment

To test the two-level decomposition technique we used simplex embedding method modification with cutting plane shift [14]. Preliminary numerical testing was carried out on the decomposition problems that were generated automatically in the form (2) with random matrixes A and D and objective function $\sum_{i=1}^n x_i^2$. Solution x^* was set beforehand to obtain vectors $b = Ax^*$, $d = Dx^*$. The average results of numerical test series are represented in the table below where we give a comparison with the subgradient method. Each test series consisted of 5 problems of corresponding dimension. We accepted the following notation: *SGM* – subgradient method, *SEM* – simplex embedding method with cutting plane shift, n – number of variables, m – number of constraints, k – number of separable constraint blocks, *iter* – number of iterations, *time* is given in seconds. We took the accuracy parameter for both algorithms $\varepsilon = 10^{-3}$. All calculations were implemented in Matlab system on the computer fitted with 8-core processor AMD FX-8350 that has 4 GHz clock speed per each core. RAM of the computer was 8 Gb.

m, n, k	<i>SGM</i>		<i>SEM</i>	
	<i>iter</i>	<i>time</i>	<i>iter</i>	<i>time</i>
10, 10, 5	> 2000	> 80	21	1
50, 50, 5	> 2000	> 250	108	15
100, 100, 10	> 2000	> 1000	355	195
150, 150, 15	> 2000	> 1000	771	978

7 Conclusions

We implemented a preliminary numerical testing of simplex embedding method with one of the possible modification that demonstrated better performance than subgradient method. Flexible and adaptive structure of the simplex embedding method makes it a perspective method for solving decomposition problems by means of Lagrangean relaxation technique.

References

1. Antsiferov, E.G., Bulatov, V.P.: An algorithm of simplex imbeddings in convex programming. *Zh. Vychisl. Mat. Mat. Fiz.* 27(3), 377–384 (1987), (in Russian)
2. Apekina, Y.V., Khamisov, O.V.: A modified simplex immersions method with simultaneous introduction of several intersecting planes. *Russian Mathematics (Iz. VUZ)* 41(12), 14–22 (1997)
3. Bagirov, A., Karmitsa, N., Makela, M.M.: *Introduction to Nonsmooth Optimization. Theory, Practice and Software.* Springer International Publishing Switzerland (2014)
4. Barmann, A.: *Solving Network Design Problems via Decomposition, Aggregation and Approximation.* Springer Fachmedien Wiesbaden (2016)
5. Benders, J.F.: Partitioning procedures for solving mixed-variables programming problems. *Numer. Math.* 4(3), 238–252 (1962)
6. Conejo, A.J., Castillo, E., Minguez, R., Garcia-Bertrand, R.: *Decomposition Techniques in Mathematical Programming. Engineering and Science Applications.* Springer-Verlag, Berlin Heidelberg (2006)
7. Costa, A.M.: A survey on benders decomposition applied to fixed-charge network design problems. *Computers & Operations Research* 32, 1429–1450 (2005)
8. Dantzig, G.B., Wolfe, P.: Decomposition principle for linear programs. *Oper. Res.* 8(1), 101–111 (1961)
9. Flippo, O.E., Rinnooy Kan, A.H.G.: Decomposition in general mathematical programming. *Mathematical Programming* 60, 361–382 (1993)
10. Geoffrion, A.M.: *Perspectives on Optimization.* Addison-Wesley Publ. Co, Reading, Mass (1972)
11. Kiwiel, K.C.: *Approximations in Decomposition of Large-Scale Convex Programs via a Nondifferentiable Optimization Method.* IFAC 11th Triennial World Congress, Tallinn, Estonia, USSR (1990)
12. Kolosnitsyn, A.V.: Using of modified simplex imbeddings method for solving special class of convex non-differentiable optimization problems. *IIGU Ser. Matematika* 11, 54–68 (2015)
13. Kolosnitsyn, A.: Modified simplex imbeddings method in convex non-differentiable optimization. In: *DOOR-2016, CEUR-WS.* vol. 1623. pp. 218–255 (2016)
14. Kolosnitsyn, A.V.: Computational efficiency of the simplex embedding method in convex nondifferentiable optimization. *Computational Mathematics and Mathematical Physics* 58(2), 215–222 (2018)
15. Lasdon, L.S.: *Duality and decomposition in mathematical programming.* IEEE Transactions on System Science and Cybernetics ssc-4(2), 86–100 (1968)
16. Minoux, M.: *Mathematical Programming: Theory and Algorithms.* John Wiley & Sons Ltd (1986)

17. Nowak, I.: Relaxation and Decomposition Methods for Mixed Integer Nonlinear Programming. ISNM vol. 152, Birkhauser Verlag, Switzerland (2005)
18. Molina, F.W.: A Survey of resource directive decomposition in mathematical programming. *Computing Surveys* 11(2), 95–104 (1979)
19. Nemirovsky, A., Yudin, D.: Informational complexity and efficient methods for solution of convex extremal problems. J. Wiley & Sons, New York (1983)
20. Nesterov, Yu.E.: Introductory Lectures on Convex Programming: A Basic Course. Springer Science + Business Media, New York (2004)
21. Palomar, D.P., Chiang, M.: A tutorial on decomposition methods for network utility maximization. *IEEE Journal on Selected Areas in Communications* 24(8), 1439–1451 (2006)
22. Rahmaniani, R., Crainic, T.G., Gendreau, M., Rei, W.: The Benders decomposition algorithm: A literature review. *European Journal of Operational Research* 259, 801–817 (2017)
23. Ruszczyński, A.: Decomposition Methods. *Handbooks in OR & MS*, vol. 10 (2003)
24. Shor, N.Z.: Minimization Methods for Non-differentiable Functions. Springer-Verlag, Berlin (1985)