Birge and Qi method for three-stage stochastic programs using IPM

Georg Ch. Pflug Ladislav Halada *

1 Introduction

Stochastic programming provides an effective framework for addressing decision problems under uncertainty in diverse fields. It can be used to model a wide range of practical applications with uncertain input data. When the input data are discretely distributed - represented, for example, by set of scenario - the stochastic program can be formulated as a deterministic equivalent linear program with a dual, block angular constraint matrix. Solving the deterministic equivalent formulation of two-stage stochastic programs using interior point method requires the solution of linear systems of the form

$$(ADA^t)\;dy = b. \tag{1}$$

Solving of this problem requires more then $90-95\%$ of total programming time [1]. Birge and Holmes [2] compared different methods for the solution of this system. They found that the factorization technique based on the work of Birge and Qi (BQ) [3] is more efficient and stable than other methods. They also suggested BQ for parallel computation. A parallel version of BQ for two-stage stochastic programs was implemented on an Intel iPSC/860 hypercube and a Connection Machine CM-5 with nearly perfect speedup [4].

According to our knowledge, this method has not been used for a three-stage stochastic program so far. The aim of this report is analysis of the BQ method and a suggestion how the BQ method can be used for three-stage stochastic programming. Section 2 very briefly describes the problem formulation. Review of the BQ matrix decomposition is given in Section 3. Section 4 contains analysis of the BQ matrix factorization from a computational point of view. An application of the BQ method to a three-stage stochastic model is in the last section.

*Department of Statistics and Decision Support Systems,Vienna University, Universitaesstr. 5, A-1010 Wien, Austria, e-mail:georg.pflug@univie.ac.at

2 The problem formulation

Let we have a primal linear program

$$\begin{aligned} min\ c^t x \\ Ax = b \\ x \geq 0 \end{aligned} \tag{2}$$

and its dual counterpart

$$\begin{aligned} max\ b^t y \\ A^t y + z = c \\ z \geq 0. \end{aligned} \tag{3}$$

We can state the Kuhn-Karush-Tucker (KKT) first order optimality conditions

$$\begin{aligned} Ax &= b \\ A^t y + z &= c \\ XZe &= 0 \\ x, z &\geq 0, \end{aligned} \tag{4}$$

where A is $m \times n$ matrix and the remaining vectors and matrices are conforming. X and Z are diagonal matrices whose diagonal entries come from vectors x and z, respectively. A nonnegative solution of the above equation (if it exists) is the optimal solution of the optimization problems (2) and (3). The Primal-Dual method is one a way how to find it. But besides of the KKT equations in this method the perturbed KKT equations is used, where μ is a perturbation parameter

$$\begin{aligned} Ax &= b \\ A^t y + z &= c \\ XZe &= \mu\, e \\ x, z &\geq 0. \end{aligned} \tag{5}$$

As $\mu \downarrow 0$ the solution of the system (5) converge to the solution of the original KKT equations. These equations (5) can be written as a block matrix equation

$$\begin{pmatrix} & A^t & I \\ A & & \\ Z & & X \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} c \\ b \\ \mu e \end{pmatrix}. \tag{6}$$

The standard approach to finding zero points of such equation is to use the iterative Newton method. Given a strictly positive x, z and an arbitrary y this method iteratively updates the current approximate solution $(x, y, z,)$ by the formula

$$\begin{array}{rcl} x: & = & x + \alpha \Delta x \\ y: & = & y + \alpha \Delta y \\ z: & = & z + \alpha \Delta z, \end{array} \tag{7}$$

where $\alpha \in (0, 1)$ and the vectors $\Delta x,\ \Delta y,\ \Delta z$ are solutions of the equation

$$J(x, y, z) \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = -F(x, y, z). \tag{8}$$

$J(x, y, z)$ is Jacobian matrix and $F(x, y, z)$ is defined as

$$F(x, y, z) = \begin{pmatrix} A^t y + z - c \\ Ax - b \\ XZe - \mu e \end{pmatrix}. \tag{9}$$

Thus, the linear system we need to solve in each iteration has the form

$$\begin{pmatrix} & A^t & I \\ A & & \\ Z & & X \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r_d \\ r_p \\ r_c \end{pmatrix}, \tag{10}$$

where r_p, r_d and r_c are residual vectors of perturbed KKT system of equations. By an elimination process we can derive the following expressions

$$\begin{array}{rcl} \Delta z & = & X^{-1} r_c - X^{-1} Z \Delta x \\ \Delta x & = & Z^{-1}(X A^t \Delta y + r_c - X r_d) \\ (ADA^t)\ \Delta y & = & r_p + AZ^{-1}(X r_d - r_c), \end{array} \tag{11}$$

where $D = Z^{-1}X$ is a diagonal matrix with all entries strictly positive. We note, if matrix A has full row rank then ADA^t is a symmetric positive definite and thus the system has a well defined solution.

3 The BQ matrix factorization

Let A be a constraint matrix of the two-stage stochastic model and D a positive definite and diagonal matrix

$$A = \begin{pmatrix} A_0 & & & & \\ T_1 & A_1 & & & \\ T_2 & & A_2 & & \\ \vdots & & & \ddots & \\ T_k & & & & A_k \end{pmatrix} \qquad D = \begin{pmatrix} D_0 & & & & \\ & D_1 & & & \\ & & D_2 & & \\ & & & \ddots & \\ & & & & D_k \end{pmatrix}, \tag{12}$$

where A_i are $m_i \times n_i$ and D_i are positive definite diagonal $n_i \times n_i$ matrices, $i = 0, 1, 2, \ldots, k$. T_i are $m_i \times n_0$ matrices, $i = 1, 2, \ldots, k$. Then (ADA^t) can be expressed as sum of two matrices

$$\begin{pmatrix} A_0D_0A_0^t & A_0D_0T_1^t & \ldots & A_0D_0T_k^t \\ T_1D_0A_0^t & T_1D_0T_1^t & \ldots & T_1D_0T_k^t \\ \vdots & \vdots & & \vdots \\ T_kD_0A_0^t & T_kD_0T_1^t & \ldots & T_kD_0T_k^t \end{pmatrix} + \begin{pmatrix} 0 & & & \\ & A_1D_1A_1^t & & \\ & & \ddots & \\ & & & A_kD_kA_k^t \end{pmatrix}. \tag{13}$$

Because the second matrix is singular matrix let us change its zero entry by Identity matrix I_{m_0}. Then we will have

$$\begin{pmatrix} A_0D_0A_0^t - I_{m_0} & A_0D_0T_1^t & \ldots & A_0D_0T_k^t \\ T_1D_0A_0^t & T_1D_0T_1^t & \ldots & T_1D_0T_k^t \\ \vdots & \vdots & & \vdots \\ T_kD_0A_0^t & T_kD_0T_1^t & \ldots & T_kD_0T_k^t \end{pmatrix} + \begin{pmatrix} I_{m_0} & & & \\ & A_1D_1A_1^t & & \\ & & \ddots & \\ & & & A_kD_kA_k^t \end{pmatrix}. \tag{14}$$

It is easy to verify, that the first matrix of (14) can be expressed as a product of three matrices

$$\begin{pmatrix} A_0 & I_{m_0} \\ T_1 & \\ T_2 & \\ \vdots & \\ T_k & \end{pmatrix} \times \begin{pmatrix} D_0 & \\ & I_{m_0} \end{pmatrix} \times \begin{pmatrix} A_0^t & T_1^t & T_2^t & \ldots & T_k^t \\ -I_{m_0} & & & & \end{pmatrix} = U\mathcal{D}V^t \tag{15}$$

and the second matrix is a diagonal matrix with positive definite matrices on the diagonal entries

$$\mathcal{R} = Diag(I_{m_0}, A_1D_1A_1^t, A_2D_2A_2^t, \ldots, A_kD_kA_k^t) = Diag(I_{m_0}, R_1, R_2, \ldots, R_k). \tag{16}$$

Thus ADA^t can be decomposed as

$$ADA^t = \mathcal{R} + U(\mathcal{D}V^t) = \mathcal{R} + UW^t. \tag{17}$$

Now, if we need the inverse of (ADA^t), we can use Sherman-Morrison-Woodbury formula. It holds [5]

$$(ADA^t)^{-1} = (\mathcal{R} + UW^t)^{-1} = \mathcal{R}^{-1} - \mathcal{R}^{-1}U(I_{n_0+m_0} + W^t\mathcal{R}^{-1}U)^{-1}W^t\mathcal{R}^{-1}. \tag{18}$$

Because $W^t = \mathcal{D}V^t$, (18) can be expressed as

$$(ADA^t)^{-1} = \mathcal{R}^{-1} - \mathcal{R}^{-1}U(I_{n_0+m_0} + \mathcal{D}V^t\mathcal{R}^{-1}U)^{-1}\mathcal{D}V^t\mathcal{R}^{-1} = \mathcal{R}^{-1} - \mathcal{R}^{-1}UG^{-1}V^t\mathcal{R}^{-1}, \tag{19}$$

where

$$\begin{aligned} G^{-1} &= (I_{n_0+m_0} + \mathcal{D}V^t\mathcal{R}^{-1}U)^{-1}\mathcal{D} \\ G &= \mathcal{D}^{-1}(I_{n_0+m_0} + \mathcal{D}V^t\mathcal{R}^{-1}U) = \mathcal{D}^{-1} + V^t\mathcal{R}^{-1}U. \end{aligned} \tag{20}$$

The structure and the sub-block matrices of G can be easily obtained by the multiplication of matrices $V^t\mathcal{R}^{-1}U$.

$$G = \begin{pmatrix} D_0^{-1} + A_0^tA_0 + \sum_{i=1}^k T_i^tR_i^{-1}T_i & A_0^t \\ -A_0 & 0 \end{pmatrix} = \begin{pmatrix} \hat{G} & A_0^t \\ -A_0 & 0 \end{pmatrix}. \tag{21}$$

Thus, we can rewrite the solution of the system $(ADA^t)dy = b$ by relations (18)-(20) as follows:

$$\begin{aligned} dy &= (ADA^t)^{-1}b \\ dy &= (\mathcal{R}^{-1} - \mathcal{R}^{-1}UG^{-1}V^t\mathcal{R}^{-1})b \\ dy &= \mathcal{R}^{-1}b - \mathcal{R}^{-1}UG^{-1}V^t\mathcal{R}^{-1}b. \end{aligned} \tag{22}$$

Now, if we use the substitutions

$$\begin{aligned} \mathcal{R}^{-1}b &= p \\ \mathcal{R}^{-1}UG^{-1}V^tp &= s \\ G^{-1}V^tp &= q \end{aligned}$$

then $dy = p - s$, where

$$\mathcal{R}p = b \tag{23}$$

$$Gq = V^tp \tag{24}$$

$$\mathcal{R}s = Uq. \tag{25}$$

We state that by this way the solution of $(ADA^t)dy = b$ has been decomposed on the solution of three sequential sub-problems. The advantage of this decomposition is that the matrix $\mathcal{R}$ is block diagonal and therefore equations (23) and (25) can be solved by matrix sub-block computations independently of one another. More precisely,

$$\mathcal{R}p = \begin{pmatrix} I_{m_0} & & & \\ & R_1 & & \\ & & \ddots & \\ & & & R_k \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ \vdots \\ p_k \end{pmatrix} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix}. \tag{26}$$

Hence, the vector p can be computed component-wise by solving sub-block systems

$$\begin{aligned} p_0 &= b_0 \\ R_i p_i &= b_i, \qquad i = 1, 2, ...k. \end{aligned} \tag{27}$$

We can proceed similarly in the case of computation of the vector s

$$\mathcal{R}s = \begin{pmatrix} I_{m_0} & & & \\ & R_1 & & \\ & & \ddots & \\ & & & R_k \end{pmatrix} \begin{pmatrix} s_0 \\ s_1 \\ \vdots \\ s_k \end{pmatrix} = Uq = \begin{pmatrix} A_0 & I_{m_0} \\ T_1 & \\ \vdots & \\ T_k & \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} A_0 q_1 + q_2 \\ T_1 q_1 \\ \vdots \\ T_k q_1 \end{pmatrix}. \tag{28}$$

It means

$$\begin{aligned} s_0 &= A_0\, q_1 + q_2 \\ R_i s_i &= T_i\, q_1, \qquad i = 1, 2, ...k. \end{aligned} \tag{29}$$

Because $R_i = A_i D_i A_i^t, \quad i = 1, 2, ...k$ is symmetric positive definite matrix, its Cholesky decomposition can be used for the solution of both equations (27) and (29). The solution of the system (24) can be found by a two ways. If we exploit the structure of matrix G and the vector $V^t p$ we will have

$$Gq = \begin{pmatrix} \hat{G} & A_0^t \\ -A_0 & 0 \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} \hat{v}_1 \\ \hat{v}_2 \end{pmatrix}, \tag{30}$$

where

$$\begin{pmatrix} \hat{v}_1 \\ \hat{v}_2 \end{pmatrix} = V^t p = \begin{pmatrix} A_0^t & T_1^t & \dots & T_k^t \\ -I_{m_0} & & & \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ \vdots \\ p_k \end{pmatrix} = \begin{pmatrix} A_0^t p_0 + \sum_{i=1}^k T_i^t p_i \\ -p_0 \end{pmatrix}. \tag{31}$$

The vectors q_i and $\hat{v}_i$, i=1,2 have the size corresponding to the matrix structure G. By an elimination process we get for the unknowns of the system (30)

$$\begin{aligned} (A_0 \hat{G}^{-1} A_0^t) q_2 &= A_0 \hat{G}^{-1} \hat{v}_1 + \hat{v}_2 \\ \hat{G} q_1 &= \hat{v}_1 - A_0^t q_2. \end{aligned} \tag{32}$$

Because both matrices $(A_0 \hat{G}^{-1} A_0^t)$ and $\hat{G}$ are symmetric positive definite, one can used their Cholesky decomposition for solving these systems. This method requires

$$\frac{n_0^3}{3} + n_0^2(m_0 + 3) + O(m_0^2 n_0) \quad flops.$$

Another possibility is to solve the system (30) directly by LDL^t decomposition with using triangular solver. This method requires

$$\frac{n_0^3}{3} + \frac{n_0^2(4m_0 + 9)}{2} + O(m_0^2 n_0) \quad flops.$$

From the comparison we can see, that the first method requires about

$$\frac{n_0^2(2m_0+3)}{2} \quad flops$$

more then the second method. However, the Cholesky decomposition is in the solution of such problem preferred owing to the numerical stability of this process.

The procedure for sequential computing of the vector dy by (23)-(25) has been named Findy in [2]. The parallel version of Findy has been formulated in [4]. They steps are as follows:

PROCEDURE Parallel Findy($\mathcal{R},\ A_0,\ D_0, T_1, \ldots, T_k, b, dy$)

1. Parallel solution of $\mathcal{R}p = b$.
   In parallel, on processors $i = 1, 2, \ldots, k$ form $R_i = A_i D_i A_i^t$ and solve $R_i p_i = b_i$, $p_0 = b_0$.

2. Solution of $Gq = V^t p$.

   (a) In parallel, on processors $i = 1, 2, \ldots, k$ solve $R_i \hat{T}_i = T_i$.

   (b) In parallel, on processor $i = 1, 2, \ldots, k$ multiply $T_i^t \hat{T}_i$ and $T_i^t p_i$.

   (c) Communicate, form $\hat{G}$,$\hat{v}_1$, $\hat{v}_2$ on 0-processor.

   (d) Solve, on 0-processor $\hat{G}B_0 = A_0^t$, $\hat{G}\hat{v}_1^{(s)} = \hat{v}_1$.

   (e) Form, on 0-processor $G = A_0 B_0$ and $r_1 = A_0 \hat{v}_1^{(s)} + \hat{v}_2$.

   (f) Solve, on 0-processor $Gq_2 = r_1$.

   (g) Solve, on 0-processor $\hat{G}q_1 = \hat{v}_1 - A_0^t q_2$.
   Broadcast to all processors $i = 1, 2, \ldots, k$ the vector q_1.

3. Parallel solution of $\mathcal{R}s = Uq$.
   In parallel, on processors $i = 1, 2, \ldots, k$ solve $R_i s_i = (Uq)_i$.

4. In parallel form $(dy)_i = p_i - s_i$.
   Communicate, form dy.

In [4] this procedure has been implemented on a distributed-memory multiple-instruction multiple-date (MIMD) message-passing parallel computers an Intel iPSC/860 hypercube and a Connection Machine CM-5. Results are reported with the solution of the linear systems arising when solving stochastic programs with $98,304$ scenarios, which correspond to deterministic equivalent linear programs with up to $1,966,090$ constraints and $13,762,630$ variables. From the timing data presented in this paper it is evident that the speed- up and the efficiency is most influenced by the percentage of time spent in communication and also by the ratio m_i/n_i.

4 Analysis of the BQ matrix factorization

The matrix factorization ADA^t starts with a convenient expression of this matrix as a sum of two nonsingular matrices. This property is inevitable as it has been seen. Therefore the authors of [3] have expressed the element on the position $(1,1)$ as

$$A_0D_0A_0^t = A_0D_0A_0^t - I_{m_0} + I_{m_0}. \tag{33}$$

But generally, one can use instead of I_{m_0} an arbitrary regular matrix or product of matrices XY. Naturally, the choice of such matrices has influence on matrix G and on the computation of the system $Gq = V^tp$, too.

Let us use instead of expression (33) the following one:

$$A_0D_0A_0^t = A_0D_0A_0^t - XY + XY, \tag{34}$$

where X and Y are $m_0 \times k$ and $k \times m_0$ matrices, respectively, $k \geq m_0$. Then we can write

$$ADA^t = \bar{\mathcal{R}} + \bar{U}(\bar{\mathcal{D}}\bar{V}^t) = \bar{\mathcal{R}} + \bar{U}\bar{W}^t, \tag{35}$$

where

$$\bar{\mathcal{R}} = Diag(XY,\ A_1D_1A_1^t, \ldots, A_kD_kA_k^t) = Diag(XY,\ R_1, \ldots, R_k), \tag{36}$$

$$\bar{U} = \begin{pmatrix} A_0 & X \\ T_1 & \\ T_2 & \\ \vdots & \\ T_k & \end{pmatrix}, \quad \bar{\mathcal{D}} = \begin{pmatrix} D_0 & \\ & I_k \end{pmatrix}, \quad \bar{V}^t = \begin{pmatrix} A_0^t & T_1^t & T_2^t & \ldots & T_k^t \\ -Y & & & & \end{pmatrix}. \tag{37}$$

Now, if we use again the Sherman-Morrison-Woodbury formula for the computation $(ADA^t)^{-1}$ we obtain by the same way that the solution $(ADA^t)dy = b$ fulfils
$dy = p - s$, while

$$\begin{aligned} \bar{\mathcal{R}}p &= b && (38)\\ \bar{G}q &= \bar{V}^tp && (39)\\ \bar{\mathcal{R}}s &= \bar{U}q. && (40) \end{aligned}$$

The matrix $\bar{G}$ and the equation (39) has the form

$$\begin{pmatrix} D_0^{-1} + A_0^t(XY)^{-1}A_0 + \sum_{i=1}^k T_i^tR_i^{-1}T_i & A_0^t(XY)^{-1}X \\ -Y(XY)^{-1}A_0 & I_k - Y(XY)^{-1}X \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} \bar{v}_1 \\ \bar{v}_2 \end{pmatrix}, \tag{41}$$

where

$$\begin{pmatrix} \bar{v}_1 \\ \bar{v}_2 \end{pmatrix} = \begin{pmatrix} A_0^t p_0 + \sum_{i=1}^k T_i^t p_i \\ -Y p_0 \end{pmatrix}. \tag{42}$$

It is clear, if $X = Y = I_{m_0}$ then the relations (38)- (40) represent BQ method.

However, it remains an open question, if there exist such X and Y which make possible to solve (39) more suitable then (24) in BQ method. An answer on this question could be interesting from the point of view of multistage stochastic model application, too.

Thus, a problem is to find matrices X, Y such that the matrix

$$\begin{pmatrix} \hat{G} & A_0^t (XY)^{-1} X \\ -Y(XY)^{-1} A_0 & I_k - Y(XY)^{-1} X \end{pmatrix}$$

is easily invertible (or a matrix equation is easily solvable). In our case it means, that $P = Y(XY)^{-1}X$ is diagonal matrix. We remark that $P = Y(XY)^{-1}X$ is idempotent (i.e. $P^2 = P$). Therefore, if P is diagonal, it may only contain 0 or 1 as diagonal elements. Moreover, since $k \geq m_0$ and $rank(X) = rank(Y) = m_0$, we have $rank(P) = m_0$. It follows from the fact that $PY = Y$. Because, the rank of P cannot be smaller than m_0. On the other hand, it cannot be larger than m_0, therefore it must be m_0. Then $rank(I_k - P) = k - m_0$. Thus $I_k - P$ cannot be invertible, unless $m_0 = 0$. There is no hope to find X, Y such that $I_k - P$ is invertible.

From this analysis one can see that the decomposition (14) with $\pm I_{m_0}$ in the position $(1, 1)$ is the best solution from the computational point of view.

5 The BQ method applied to three-stage stochastic model

Let A be here a constraint matrix of a three-stage stochastic model

$$A = \begin{pmatrix} A_0 & & & & & & & & \\ T_{10} & A_{10} & & & & & & & \\ & T_{11} & A_{11} & & & & & & \\ & T_{12} & & A_{12} & & & & & \\ & T_{13} & & & A_{13} & & & & \\ T_{20} & & & & & A_{20} & & & \\ & & & & & T_{21} & A_{21} & & \\ & & & & & T_{22} & & A_{22} & \\ & & & & & T_{23} & & & A_{23} \end{pmatrix} = \begin{pmatrix} A_0 & & \\ T_1 & A_1 & \\ T_2 & & A_2 \end{pmatrix}, \tag{43}$$

and D again diagonal and positive definite matrix.

$$D = Diag(D_0;\ D_{10}, D_{11}, D_{12}, D_{13};\ D_{20}, D_{21}, D_{22}, D_{23}) = Diag(D_0, D_1, D_2), \quad (44)$$

where A_0 is $m_0 \times n_0$ and A_{ij} are $m_{ij} \times n_{ij}$ matrices. T_{ij} has the size conformable to the matrices A_0 and A_{ij}. D_0 and D_{ij} are diagonal $n_0 \times n_0$ and $n_{ij} \times n_{ij}$ matrices with positive entries, $i = 1, 2;\ j = 0, 1, 2, 3$..
Such a constraint matrix can be obtained for example in a case of a stochastic problem with the following tree structure.

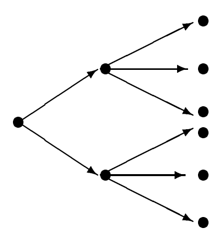

Figure 1. A tree structure problem

The matrix product ADA^t in the denotation of the largesized blocks matrix A can be decomposed as follows:

$$\begin{pmatrix} I_{m_0} & & \\ & A_1D_1A_1^t & \\ & & A_2D_2A_2^t \end{pmatrix} + \begin{pmatrix} A_0 & I_{m_0} \\ T_1 & \\ T_2 & \end{pmatrix} \begin{pmatrix} D_0 & \\ & I_{m_0} \end{pmatrix} \begin{pmatrix} A_0^t & T_1^t & T_2^t \\ -I_{m_0} & & \end{pmatrix}. \quad (45)$$

Let us use for the matrices of this factorization the same characters $\mathcal{R}$, U, $\mathcal{D}$ and V, respectively as in the foregoing paragraphs. Moreover let the upper case index represent a level of the factorization of the matrix ADA^t. It means for the initial factorization with the largesized blocks in A the index 1 will be used. Thus, on the basis of Section 3 we can write

$$ADA^t = \mathcal{R}^{(1)} + U^{(1)}[\mathcal{D}^{(1)}\ (V^{(1)})^t] = \mathcal{R}^{(1)} + U^{(1)}(W^{(1)})^t. \quad (46)$$

By the Sherman-Morrison-Woodbury formula we obtain

$$(ADA^t)^{-1} = (\mathcal{R}^{(1)})^{-1} - (\mathcal{R}^{(1)})^{-1}U^{(1)}\ (G^{(1)})^{-1}\ (V^{(1)})^t\ (\mathcal{R}^{(1)})^{-1}, \quad (47)$$

where

$$\mathcal{R}^{(1)} = Diag(I_{m_0}, A_1D_1A_1^t, A_2D_2A_2^t) = Diag(I_{m_0}, R_1^{(1)}, R_2^{(1)}) \quad (48)$$

and

$$\begin{aligned} (G^{(1)})^{-1} &= [I_{n_0+m_0} + \mathcal{D}^{(1)}\, (V^{(1)})^t\, (\mathcal{R}^{(1)})^{-1}\, U^{(1)}]^{-1}\, \mathcal{D}^{(1)} \\ G^{(1)} &= (\mathcal{D}^{(1)})^{-1} + (V^{(1)})^t\, (\mathcal{R}^{(1)})^{-1}\, U^{(1)}. \end{aligned} \tag{49}$$

In the matrix form

$$G^{(1)} = \begin{pmatrix} D_0^{-1} + A_0^t A_0 + \sum_{i=1}^2 T_i^t (R_i^{(1)})^{-1} T_i & A_0^t \\ -A_0 & 0 \end{pmatrix} = \begin{pmatrix} \hat{G}^{(1)} & A_0^t \\ -A_0 & 0 \end{pmatrix}. \tag{50}$$

Thus, the solution of linear system equations $ADA^t dy = b$ can be expressed by the inversion of ADA^t as $dy = p^{(1)} - s^{(1)}$ while

$$\mathcal{R}^{(1)} p^{(1)} = b, \tag{51}$$

$$G^{(1)} q^{(1)} = (V^{(1)})^t p^{(1)}, \tag{52}$$

$$\mathcal{R}^{(1)} s^{(1)} = U^{(1)} q^{(1)}. \tag{53}$$

We may state, the equations (51)- (53) represent the decomposition of the original problem into three sub-problems. An advantage of such decomposition is that $\mathcal{R}^{(1)}$ is the block-diagonal matrix amenable to further decomposition.

5.1 Solving of the equation $\mathcal{R}^{(1)} p^{(1)} = b$

It is easy to see from the equation

$$\mathcal{R}^{(1)} p^{(1)} = \begin{pmatrix} I_{m_0} & & \\ & R_1^{(1)} & \\ & & R_2^{(1)} \end{pmatrix} \begin{pmatrix} p_0^{(1)} \\ p_1^{(1)} \\ p_2^{(1)} \end{pmatrix} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} \tag{54}$$

that this system represents the following independent systems

$$p_0^{(1)} = b_0 \tag{55}$$

$$R_i^{(1)} p_i^{(1)} = b_i, \quad i = 1, 2, \tag{56}$$

where $R_i^{(1)} = A_i D_i A_i^t$, $i = 1, 2$ represent the matrix of the two-stage model problem, which has been described in Section 3. Therefore, the systems (56) can be solved by the procedure Findy or Parallel Findy, depending on the number of processors. Its input parameters are readable from the entries of matrix A_i, D_i defined in (43)-(44). The right-hand side and the solution vector are b_i and $p_i^{(1)}$, $i = 1, 2$, respectively. It is clear that in our case the parameters are

$$Findy(\mathcal{R}_i,\ A_{i0},\ D_{i0}, T_{i1}, \ldots, T_{ik}, b_i, p_i^{(1)}), \quad i = 1, 2,$$

where $\mathcal{R}_i$ is the diagonal matrix in the decomposition $R_i^{(1)}$, i.e.

$$R_i^{(1)} = \mathcal{R}_i + U_i W_i^t \quad i = 1, 2 \tag{57}$$

and

$$\mathcal{R}_i = Diag(I_{m_{i0}},\ R_{i1}, R_{i2}, R_{i3}), \quad R_{ij} = A_{ij} D_{ij} A_{ij}^t, \quad i = 1, 2 \quad j = 1, 2, 3.$$

$$U_i = \begin{pmatrix} A_{i0} & I_{m_{i0}} \\ T_{i1} & \\ T_{i2} & \\ T_{i3} & \end{pmatrix}, \quad W_i^t = (\mathcal{D}_i V_i^t) = \begin{pmatrix} D_{i0} & \\ & I_{m_{i0}} \end{pmatrix} \begin{pmatrix} A_{i0}^t & T_{i1}^t & T_{i2}^t & T_{i3}^t \\ -I_{m_{i0}} & & & \end{pmatrix}. \tag{58}$$

If this procedure is applied for the given values, having in the mind the relations (55)-(56) we are able to compose the vector $p^{(1)}$.

5.2 Solving of the equation $G^{(1)} q^{(1)} = (V^{(1)})^t p^{(1)}$

Solving of this equation requires to have available the entries of the right-hand side vector and the sub-block matrix $\hat{G}^{(1)}$.With this aim, we denote the elements of the vector $(V^{(1)})^t p^{(1)}$ as $(\hat{v}_1^{(1)}, \hat{v}_2^{(1)})^t$. Then we have

$$\begin{pmatrix} \hat{v}_1^{(1)} \\ \hat{v}_2^{(1)} \end{pmatrix} = (V^{(1)})^t p^{(1)} = \begin{pmatrix} A_0^t & T_1^t & T_2^t \\ -I_{m_0} & & \end{pmatrix} \begin{pmatrix} p_0^{(1)} \\ p_1^{(1)} \\ p_2^{(1)} \end{pmatrix} = \begin{pmatrix} A_0^t p_0^{(1)} + \sum_{i=1}^2 T_{i0}^t p_{i0} \\ -p_0^{(1)} \end{pmatrix}, \tag{59}$$

where p_{i0}, $i = 1, 2$ is the vector of the first m_{i0}-elements of $p_i^{(1)}$. We know from (50), that $\hat{G}^{(1)} = D_0^{-1} + A_0^t A_0 + \sum_{i=1}^2 T_i^t (R_i^{(1)})^{-1} T_i$. For the relatively complicated expression $T_i^t (R_i^{(1)})^{-1} T_i$ we have proved (the proof is in the Appendix), that

$$T_i^t (R_i^{(1)})^{-1} T_i = T_{i0}^t (\hat{T}_{i0} - T_{i0}), \quad i = 1, 2 \tag{60}$$

where $\hat{T}_{i0}$ is the solution of the equation

$$(A_{i0} \hat{G}_i^{(-1)} A_{i0}^t) \hat{T}_{i0} = T_{i0}, \quad i = 1, 2. \tag{61}$$

Thus,

$$\hat{G}^{(1)} = D_0^{-1} + A_0^t A_0 + \sum_{i=1}^{2} T_{i0}^t (\hat{T}_{i0} - T_{i0}). \tag{62}$$

We remember that the Cholesky decomposition of matrix $A_{i0}\hat{G}_i^{(-1)}A_{i0}^t$ has been performed during the procedure Findy applied on matrix A_i, $i = 1, 2$. Thus, this decomposition was available already and only triangular solver is used for the computation of $\hat{T}_{i0}$, $i = 1, 2$ in this step.

Having the values of $\hat{G}^{(1)}$ and $(\hat{v}_1^{(1)}, \hat{v}_2^{(1)})^t$ we can solve the system

$$\begin{pmatrix} \hat{G}^{(1)} & A_0^t \\ -A_0 & 0 \end{pmatrix} \begin{pmatrix} q_1^{(1)} \\ q_2^{(1)} \end{pmatrix} = \begin{pmatrix} \hat{v}_1^{(1)} \\ \hat{v}_2^{(1)} \end{pmatrix}. \tag{63}$$

The standard elimination process applied on this system yields

$$[(A_0\,(\hat{G}^{(1)})^{-1}\,A_0^t]\,q_2^{(1)} = A_0\,(\hat{G}^{(1)})^{-1}\,\hat{v}_1^{(1)} + \hat{v}_2^{(1)} \tag{64}$$

$$\hat{G}^{(1)}\,q_1^{(1)} = \hat{v}_1^{(1)} - A_0^t\,q_2^{(1)}. \tag{65}$$

Thus, to solve (63) the following is required:
PROCEDURE Updy($\hat{G}^{(1)}$, A_0, $\hat{v}_1^{(1)}, \hat{v}_2^{(1)}$)
(a) The Cholesky decomposition of $\hat{G}^{(1)}$,
(b) The solution of $\hat{G}^{(1)}B_0 = A_0^t$, $\quad \hat{G}^{(1)}(\hat{v}_1^{(1)})^{(s)} = \hat{v}_1^{(1)}$,
(c) The Cholesky decomposition of A_0B_0,
(d) The solution of the systems (64) and (65) by triangular solver.

5.3 Solving of the equation $\mathcal{R}^{(1)}s^{(1)} = U^{(1)}q^{(1)}$

Now, having the vector $q^{(1)}$ we can solve the system $\mathcal{R}^{(1)}s^{(1)} = U^{(1)}q^{(1)}$ by the similar way as in Section 5.1. The right-hand side equals

$$U^{(1)}q^{(1)} = \begin{pmatrix} A_0 & I_{m_0} \\ T_1 & 0 \\ T_2 & 0 \end{pmatrix} \begin{pmatrix} q_1^{(1)} \\ q_2^{(1)} \end{pmatrix} = \begin{pmatrix} A_0\,q_1^{(1)} + q_2^{(1)} \\ T_1\,q_1^{(1)} \\ T_2\,q_1^{(1)} \end{pmatrix}. \tag{66}$$

Thus, the system has the form

$$\mathcal{R}^{(1)}p^{(1)} = \begin{pmatrix} I_{m_0} & & \\ & R_1^{(1)} & \\ & & R_2^{(1)} \end{pmatrix} \begin{pmatrix} s_0^{(1)} \\ s_1^{(1)} \\ s_2^{(1)} \end{pmatrix} = \begin{pmatrix} A_0\,q_1^{(1)} + q_2^{(1)} \\ T_1\,q_1^{(1)} \\ T_2\,q_1^{(1)} \end{pmatrix}, \tag{67}$$

from which we obtain independent equations

$$s_0^{(1)} = A_0\,q_1^{(1)} + q_2^{(1)} \tag{68}$$

$$R_i^{(1)}s_i^{(1)} = T_i\,q_1^{(1)},\ i = 1, 2. \tag{69}$$

The last two equations are again solvable by the procedure Findy as in Section 5.1 with the right-hand side $T_iq_1^{(1)}$, $i = 1, 2$. But, owing to the structure of this vector, where only the first m_{i0}- entries are nonzero, we suggest for its computation a modification the already mentioned procedure Findy.

Because (57) holds, the solution $s_i^{(1)}$ of (69) can be expressed as $s_i^{(1)} = \hat{p}_i - \hat{s}_i$, where $\hat{p}_i$ and $\hat{s}_i$, $i = 1,2$ fulfil

$$\mathcal{R}_i \hat{p}_i = T_i q_1^{(1)} \tag{70}$$
$$G_i \hat{q}_i = V_i^t \hat{p}_i \tag{71}$$
$$\mathcal{R}_i \hat{s}_i = U_i \hat{q}_i. \tag{72}$$

Equation (70) represent the system

$$\mathcal{R}_i \hat{p}_i = \begin{pmatrix} I_{m_{i0}} & & & \\ & R_{i1} & & \\ & & R_{i2} & \\ & & & R_{i3} \end{pmatrix} \begin{pmatrix} \hat{p}_{i0} \\ \hat{p}_{i1} \\ \hat{p}_{i2} \\ \hat{p}_{i3} \end{pmatrix} = \begin{pmatrix} T_{i0} q_1^{(1)} \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{73}$$

from which we have immediately

$$\hat{p}_{i0} = T_{i0} q_1^{(1)}, \quad i = 1,2 \tag{74}$$
$$\hat{p}_{ij} = 0, \quad i = 1,2 \quad j = 1,2,3. \tag{75}$$

To find the solution of equation (71) means to solve the following matrix equation

$$\begin{pmatrix} \hat{G}_i & A_{i0}^t \\ -A_{i0} & 0 \end{pmatrix} \begin{pmatrix} \hat{q}_{i1} \\ \hat{q}_{i2} \end{pmatrix} = \begin{pmatrix} A_{i0}^t \, \hat{p}_{i0} \\ -\hat{p}_{i0} \end{pmatrix}. \tag{76}$$

A standard block elimination of this matrix leads to the systems

$$(A_{i0} \hat{G}_i^{-1} A_{i0}^t) \, \hat{q}_{i2} = (A_{i0} \hat{G}_i^{-1} A_{i0}^t) \hat{p}_{i0} - \hat{p}_{i0} \tag{77}$$
$$\hat{G}_i \, \hat{q}_{i1} = A_{i0}^t (\hat{p}_{i0} - \hat{q}_{i2}). \tag{78}$$

The last equation (72) represent the system

$$\mathcal{R}_i \hat{s}_i = \begin{pmatrix} I_{m_{i0}} & & & \\ & R_{i1} & & \\ & & R_{i2} & \\ & & & R_{i3} \end{pmatrix} \begin{pmatrix} \hat{s}_{i0} \\ \hat{s}_{i1} \\ \hat{s}_{i2} \\ \hat{s}_{i3} \end{pmatrix} = \begin{pmatrix} A_{i0} \hat{q}_{i1} + \hat{q}_{i2} \\ T_{i1} \hat{q}_{i1} \\ T_{i2} \hat{q}_{i1} \\ T_{i3} \hat{q}_{i1} \end{pmatrix}, \tag{79}$$

from which we have

$$\hat{s}_{i0} = A_{i0} \hat{q}_{i1} + \hat{q}_{i2}, \quad i = 1,2 \tag{80}$$
$$R_{ij} \hat{s}_{ij} = T_{ij} \hat{q}_{i1}, \quad i = 1,2 \quad j = 1,2,3. \tag{81}$$

With the vector $\hat{s}_{ij}$ available, we have the result

$$s_i^{(1)} = \hat{p}_i - \hat{s}_i = \begin{pmatrix} \hat{p}_{i0} - \hat{s}_{i0} \\ -\hat{s}_{i1} \\ -\hat{s}_{i2} \\ -\hat{s}_{i3} \end{pmatrix}, \quad i = 1, 2. \tag{82}$$

Finally, the computing of $s_i^{(1)}$, $i = 1, 2$ consist of the following steps:

PROCEDURE $Findysparse(\mathcal{R}_i, A_{i0}, T_{i1}, T_{i2}, T_{i3}, q^{(1)}, s_i^{(1)})$

1. Set $\hat{p}_{i0} = T_{i0} q_1^{(1)}, \quad i = 1, 2$
2. Solve $(A_{i0} \hat{G}_i^{-1} A_{i0}^t)(\hat{p}_{i0} - \hat{q}_{i2}) = \hat{p}_{i0}$ and $\hat{G}_i \, \hat{q}_{i1} = A_{i0}^t (\hat{p}_{i0} - \hat{q}_{i2}), \quad i = 1, 2$
3. (a) Set $\hat{s}_{i0} = A_{i0} \hat{q}_{i1} + \hat{q}_{i2}, \quad i = 1, 2$
   (b) Solve $R_{ij} \hat{s}_{ij} = T_{ij} \hat{q}_{i1}, \quad i = 1, 2;\ j = 1, 2, 3$
4. Set $s_i^{(1)} = (\hat{p}_{i0} - \hat{s}_{i0}, -\hat{s}_{i1}, -\hat{s}_{i2}, -\hat{s}_{i3})^t, \quad i = 1, 2.$

Note we have available the Cholesky decomposition of the system matrices in steps 2 and 3b. These decomposition has been computed by the procedure Findy.

In the end, the result of a three-stage stochastic model problem $(ADA^t)dy = b$ equals $dy = p^{(1)} - s^{(1)}$. This difference is obtained by the following computational process :

1. (a) Call Findy($\mathcal{R}_1, ..., s_1^{(1)}$)
   (b) Call Findy($\mathcal{R}_2, ..., s_2^{(1)}$)
2. Call Updy($\hat{G}^{(1)}$, A_0, $\hat{v}_1^{(1)}$, $\hat{v}_2^{(1)}$)
3. (a) Call Findysparse($\mathcal{R}_1, ..., p_1^{(1)}$)
   (b) Call Findysparse($\mathcal{R}_2, ..., p_2^{(1)}$).

Naturally, the procedures in steps 1 and 3 are independent and can be computed at the same time. Step 2 represents a binding of existing two-stage models and allows to calculate $s_0^{(1)}$ and the parameter $q^{(1)}$ for Findysparse(). Roughly, it may be symbolically written as:

$$dy = \begin{pmatrix} p_0^{(1)} \\ p_1^{(1)} \\ p_2^{(1)} \end{pmatrix} - \begin{pmatrix} s_0^{(1)} \\ s_1^{(1)} \\ s_2^{(1)} \end{pmatrix} = \begin{pmatrix} b_0 \\ Findy(\mathcal{R}_1, ..., p_1^{(1)}) \\ Findy(\mathcal{R}_2, ..., p_2^{(1)}) \end{pmatrix} - \begin{pmatrix} A_0 q_1^{(1)} + q_2^{(1)} \\ Findysparse(\mathcal{R}_1, ..., s_1^{(1)}) \\ Findysparse(\mathcal{R}_2, ..., s_2^{(1)}) \end{pmatrix}. \tag{83}$$

6 Conclusion

The aim of our paper has been to use BQ factorization technique for three-stage stochastic program in a framework of an interior point method. As we can see this technique leads to the solution of independent sub-problems. More over, these sub-problems are again scalable into smaller linear system of equations. The whole process contains a serial coordination step, but the range of a sequential computation is not critical for large-scale stochastic program.

References

[1] Vladimirou, H. Zenios, S.A.: Scalable parallel computations for large-scale stochastic programming,Annals of Oper. Res. 90,1999, pp.87-129.

[2] Birge, J.R., Holmes, D.F.: Efficient solution of two-stage stochastic linear programs using interior point methods, Comput. Optim. Appl., 1, 1992,pp. 245-276.

[3] Birge, J.R., Qi, L.: Computing block-angular Karmarkar projections with applications to stachastic programming, Management Sci., 34,1988, pp.1472-1479.

[4] Jessup, E.R.,Yang, D., Zenios, S.A.: Parallel factorization of structured matrices arising in stochastic programming, SIAM J.Optimization, Vol. 4., No.4,1994, pp. 833-846.

[5] Golub,G.H.,Van Loan, Ch.F.: Matrix Computations, The Johns Hopkins University Press, 1996.

[6] A.Ralston : A first course in numerical analysis, McGraw-Hill, inc., New York, 1965.

APPENDIX

According to (57) and Sherman-Morrison-Woodbury formula we have

$$T_i^t(R_i^{(1)})^{(-1)}T_i = T_i^t(\mathcal{R}_i^{-1} - \mathcal{R}_i^{-1}U_iG_i^{-1}V_i^t\mathcal{R}_i^{-1})T_i = T_i^t\mathcal{R}_i^{-1}T_i - T_i^t\mathcal{R}_i^{-1}U_iG_i^{-1}V_i^t\mathcal{R}_i^{-1}T_i. \tag{84}$$

It holds

$$T_i^t\mathcal{R}_i^{-1}T_i = T_{i0}^tT_{i0}, \quad T_i^t\mathcal{R}_i^{-1}U_i = (T_{i0}^tA_{i0}, T_{i0}^t), \quad V_i^t\mathcal{R}_i^{-1}T_i = (A_{i0}^tT_{i0}, -T_{i0})^t. \tag{85}$$

Therefore

$$T_i^t(R_i^{(1)})^{(-1)}T_i = T_{i0}^t\, T_{i0} - (T_{i0}^tA_{i0},\ T_{i0}^t)\begin{pmatrix} \hat{G}_i & A_{i0}^t \\ -A_{i0} & 0 \end{pmatrix}^{-1}\begin{pmatrix} A_{i0}^tT_{i0} \\ -T_{i0} \end{pmatrix}. \tag{86}$$

Now let

$$\begin{pmatrix} K & L \\ M & N \end{pmatrix} = \begin{pmatrix} \hat{G}_i & A_{i0}^t \\ -A_{i0} & 0 \end{pmatrix}^{-1}. \tag{87}$$

According to [6]

$$N = (A_{i0}\hat{G}_i^{-1}A_{i0}^t)^{-1} \tag{88}$$

$$L = -\hat{G}_i^{-1}A_{i0}^t\,(A_{i0}\hat{G}_i^{-1}A_{i0}^t)^{-1} \tag{89}$$

$$M = (A_{i0}\hat{G}_i^{-1}A_{i0}^t)^{-1}\,A_{i0}\hat{G}_i^{-1} \tag{90}$$

$$K = \hat{G}_i^{-1} - \hat{G}_i^{-1}A_{i0}^t(A_{i0}\hat{G}_i^{-1}A_{i0}^t)^{-1}A_{i0}\hat{G}_i^{-1}. \tag{91}$$

Now, if we use (88)-(91) in (86) we obtain

$$T_i^t(R_i^{(1)})^{(-1)}T_i = T_{i0}^t(A_{i0}\hat{G}_i^{-1}A_{i0}^t)^{-1}T_{i0} - T_{i0}^t\,T_{i0}. \tag{92}$$