On Solving the Problem of Optimal Probability Distribution Quantization

Dmitry Golembiovsky
Lomonosov Moscow State University
GSP-1, Leninskie Gory,
119991 Moscow, Russia
golemb@cs.msu.su

Dmitry Denisov
Lomonosov Moscow State University
GSP-1, Leninskie Gory,
119991 Moscow, Russia
dvden@bk.ru

Evgeny Antonov
Lomonosov Moscow State University
GSP-1, Leninskie Gory,
119991 Moscow, Russia
eugeneP.antonov@gmail.com

Abstract

The program of optimal quantization of a continuous distribution suggested by Heitsch H. and W. Romisch in 2003 is generalized for arbitrage exclusion in financial models. It is a non-convex problem, which belongs to the class of NP-hard problems. In the paper, three different approximate algorithms are developed for finding the global extremum. Two of them are based on the separation of variables according to their power in the objective function while the other is a SQP algorithm. The numerical results of using the algorithms are provided. The effectiveness and speed of the problem solving are compared.

1 Introduction

Currently one of the most rapidly developing optimization approaches is the stochastic dual dynamic programming (SDDP) algorithm (see [Pereira & Pinto, 1991]). The algorithm belongs to the class of dynamic optimization algorithms and allows to solve problems where one or more parameters are represented by random variables. SDDP is a procedure of consecutive solving of linear problems, thus obtaining new constraints of objective function in a form of cutting hyperplanes and, as a result, obtaining upper and lower bounds of an optimal solution of the problem. One of the essential parts of SDDP algorithm is constructing a set of possible scenarios of random variables. Scenario models for stochastic optimization in finance must exclude arbitrage possibilities (look, for example [Consiglio et al., 2014], [Geyer et al., 2010]).

The paper analyzes the problem of constructing a scenario lattice for the joint evolution of a set of variables representing the values of economic indicators such as interest rates, exchange rates, prices of goods, securities or other financial instruments. It is assumed that the distribution of such variables is known, and a number of its realizations can be simulated. Also it is assumed that forward prices of such variables can be calculated at each

In: Yu. G. Evtushenko, M. Yu. Khachay, O. V. Khamisov, Yu. A. Kochetov, V.U. Malkova, M.A. Posypkin (eds.): Proceedings of the OPTIMA-2017 Conference, Petrovac, Montenegro, 02-Oct-2017, published at http://ceur-ws.org

stage of the simulated trajectories. The lattice considered in the paper satisfies several restrictions: positivity of each variable in the nodes, equality to one of the sum of the transition probabilities from a node and the distinctive feature of the lattice is an absence of arbitrage opportunities. It means that for each node of the current stage, an expected value of variables at the next stage is equal to its forward prices in the current node. The optimal lattice is obtained through solving an optimization problem which is not convex.

The second section of the paper presents formal definition of the problem, while 2.1.1, 2.1.2 and 2.2.1 subsections contain algorithms suggested for its solution. The first two are based on a separation of the variables of the optimization problem according to their power in the objective function while the latter uses a gradient of the objective function to iteratively approach the optimal value. The third section of the paper contains the comparison of the algorithms in particular cases.

2 Problem Definition and Methodology

Consider n_x variables which represent particular economic indicators. Let the number of the nodes at the stage t is l_t (equals 1 for the first stage) and the nodes till the stage $t-1$ have been already formed. Denote by $F^k_{t-1}, k = 1,\ldots,l_{t-1}$ the n_x-elements vector of prices in the nodes of the previous stage and by $F^{k(frw)}_{t-1}, k = 1,\ldots,l_{t-1}$ the vector of corresponding forward prices. The stage t of the scenario lattice is constructed by solving the next optimization problem.

Denote by $f^j_{tk}, j = 1,\ldots,L_t$ the realizations of the vector of considered variables at the stage t generated by the Monte-Carlo method from the node k of the previous stage $t-1$. (Like l_t parameter L_t equal to 1 for $t = 1$, $L_t >> l_t$). The vectors of the values of the random variables which are assigned to the nodes of the stage t are denoted by $F^i_t, i = 1,\ldots,l_t$. They are variables of the optimization problem.

We also introduce variables $g^{ij}_{tk} \geq 0$, where $i = 1,\ldots,l_t$ is a number of the according node at the stage t; $k = 1,\ldots,l_{t-1}$ is a number of the node at the stage $t-1$; $j = 1,\ldots,L_t$ is a number of realization of the random variables generated from the node k of the previous stage. Each variable g^{ij}_{tk} shows to what extent the according random pattern f^j_{tk} belongs to the corresponding node. So, the condition

$$\sum_{i=1}^{l_t} g^{ij}_{tk} = 1 \tag{1}$$

must be fulfilled for all j and k. The transition probability from node k of the stage $t-1$ to node i of stage t is obtained as

$$\rho^i_{tk} = \frac{\sum_{j=1}^{L_t} g^{ij}_{tk}}{L_t}, k = 1,\ldots,l_{t-1}, i = 1,\ldots,l_t \tag{2}$$

It follows from (1) and (2) that $\sum_{i=1}^{l_t} \rho^i_{tk} = 1$ for any $k = 1,\ldots,l_{t-1}$.

The objective function of the optimization problem, similar to [Heitsch & Romisch, 2003], is a sum of the distances between the values of random variables relevant to the nodes of the stage t and realizations of the random patterns that were produced by Monte-Carlo:

$$\min_{\substack{F^i_t, i=1,\ldots,l_t \\ g^{ij}_{tk}, k=1,\ldots,l_{t-1}, i=1,\ldots,l_t, j=1,\ldots,L_t}} \sum_{k=1}^{l_{t-1}} \sum_{i=1}^{l_t} \sum_{j=1}^{L_t} g^{ij}_{tk} (f^j_{tk} - F^i_t)^T (f^j_{tk} - F^i_t) \tag{3}$$

The problem includes the following constraints:

$$F^i_t \geq 0, g^{ij}_{tk} \geq 0, k = 1,\ldots,l_{t-1}, i = 1,\ldots,l_t, j = 1,\ldots,L_t \tag{4}$$

$$\sum_{i=1}^{l_t} g^{ij}_{tk} = 1, k = 1,\ldots,l_{t-1}, j = 1,\ldots,L_t \tag{5}$$

$$\rho^i_{tk} = \frac{\sum_{j=1}^{L_t} g^{ij}_{tk}}{L_t}, k = 1,\ldots,l_{t-1}, i = 1,\ldots,l_t \tag{6}$$

$$\sum_{i=1}^{l_t} F^i_t \rho^i_{tk} = F^{k(frw)}_{t-1}, k = 1,\ldots,l_{t-1} \tag{7}$$

Constraints (7) ensure a non-arbitrage condition on the scenario lattice: for each node, the expected values of random variables at the next stage must be equal to their forward values in the node.

The objective function of the considered problem is non-convex. There are well-known methods of the global extremum search in box-constraint optimization problems developed in [Evtushenko & Posypkin, 2013], [Khamisov, 1999], [Strongin & Sergeyev, 2000]. In such methods, the objective function is approximated from below by linear and quadratic functions on the elements of the feasible set partition. The feasible set of the problem considered in the paper is determined by bilinear equality constraints and contains non-regular points. The goal of this paper is to compare effectiveness of methods based on a consecutive solving of less complex problems. The constraints structure of the problem allows us to explicitly define a starting feasible point.

Denote by X the set of feasible points (g, F) for constraints of the problem (3) – (7). The problem is obviously has a solution if the set $\{F : (g, F) \in X\}$ is bounded. This property is fulfilled in a number of cases, namely the following lemma takes place.

Lemma 1: If for each point $(g, F) \in X$ the sum $\sum_{j=1}^{L_t} g_{tk}^{ij} \geq \delta$ for some $\delta > 0$, then X is bounded.

Proof: Assume that there exists the sequence $\{(g(n), F(n))\} \subset X, n = 1, 2, \ldots$, for which $\| F(n) \| \to \infty, n \to \infty$. It follows from here that there exists an index $m \in \{1, 2, \ldots, n_x\}$, for which an element $F_{t,m}^i(n) \to \infty$ for some i. From here and from (7) it follows that $F_{t-1,m}^k = \infty$, that contradicts the meaning of the vector F_{t-1}^k.

Obviously, the problem (3) – (7) is not convex. To solve it, two classes of algorithms are suggested. The first one is based on separation of variables according to their power in the objective function. The second one is the sequential quadratic programming (SQP) algorithm.

2.1 Variables Separation Algorithms

2.1.1 Algorithm 1.1

The problem is solved by a consecutive minimization of the objective function over the group of variables g_{tk}^{ij} that have a power of 1 in the objective function and F_t^i that have a power of 2 afterwards.

A. Put $Y = \infty$

B. Get starting values for F_t^i and g_{tk}^{ij}

C. Solve the the problem (3) – (7) for the variables $F_t^i, i = 1, \ldots, l_t$. (It is a quadratic problem with linear constraints).

D. If the absolute value of the difference between Y and the objective function value is less than ε, then End.

E. Fix the obtained values of the variables $F_t^i, i = 1, \ldots, l_t$.

F. Save the value of the objective function to the variable Y.

G. Solve the problem (3) – (7) for the variables $g_{tk}^{ij}, k = 1, \ldots, l_{t-1}, i = 1, \ldots, l_t, j = 1, \ldots, L_t$. (It is a linear problem).

H. Fix the obtained values of the variables $g_{tk}^{ij}, k = 1, \ldots, l_{t-1}, i = 1, \ldots, l_t, j = 1, \ldots, L_t$.

I. Go to the item C.

2.1.2 Algorithm 1.2

The following algorithm is based on the representation of the problem (3) – (7) in a form of a parametric linear problem

$$\max_g (-c(F), g) \tag{8}$$

$$g \geq 0, A(F)g = b \tag{9}$$

The idea of the algorithm is to consecutively update vector of parameters F based on the problem dependency on this parameter. Along with this it is assumed that linear problems, presented below are feasible in a neighborhood of the sought-for solution.

A. Set the tolerance parameters $\varepsilon, \beta_{min} > 0$ for a stopping criteria. Put $n = 0$, get starting values $(g(n), F(n)) \in X$. Set the initial shift value $\beta(n)$.

B. For fixed $F(n)$ solve the problem (3) – (7) as a linear problem over variables g. Fix $f_0^*(F(n))$ as an optimal value of the objective function, and its corresponding optimal solution $g(n+1)$, defined by the set $F(n)$.

C. Define $G(n)$ from the following rule:

$$G^i_{t,m}(1) = max(0, F^i_{t,m}(n) - \beta(n)\sum_{k=1}^{l_{t-1}}\sum_{j=1}^{L_t} g^{ij}_{tk}(n+1)(F^i_{t,m}(n) - f^j_{tk,m})), i = 1, 2, \ldots, l_t, m = 1, 2, \ldots, n_x.$$

D. Solve the problem (3) – (7) over g for the fixed set of variables $G(n)$ instead of $F(n)$. Fix $f_0^*(G(n))$ as an optimal value of the objective function.

E. If $f_0^*(G(n)) - f_0^*(F(n)) < -\varepsilon$, then put $F(n+1) = G(n)$, $n = n+1$. Go to the item B.

F. If $\beta(n) > \beta_{min}$, then set $\beta(n+1) = \beta(n) \setminus 2$, $F(n+1) = F(n)$, $n = n+1$. Go to the item B.

G. $(g(n), F(n))$ is the problem solution.

The issue of the feasibility of linear problems specified in Algorithm 1.2 can be reduced to the issue of the stability of the linear program solution g^*, where (g^*, F^*) is an optimal solution of the problem (3)–(7). The following lemma takes place

Lemma 2: The problem (8) – (9) is stable in its solution g^*, if $g^* > 0$, (g^*, F^*) is the solution of the problem (3) – (7), in which vectors F^i_t are linearly independent.

Proof: The stability of the linear program in the solution g^* is equivalent to the existence of such solutions in a feasible region of primal and dual linear programs, in which all inequality constraints are strict and fulfilled, provided that constraints matrix has a full rank. For a primal program such solution is g^*. The existence of such solution for a dual program is obvious. From a linear independency of vectors F^i_t if follows that the constraints matrix has a full rank.

2.2 SQP Algorithm

The Algorithm described hereinafter belongs to the class of SQP algorithms. It implies solving an auxiliary quadratic program at each stage when choosing a descent direction. This class of algorithms is implemented for Optimization and Variational Problems (look, for example [Izmailov & Solodov, 2014], [Gould & Robinson, 2010 (1)] and [Gould & Robinson, 2010 (2)].

2.2.1 Algorithm 2

Denote by x a pair of (g, F), by $q(x)$ – expression $\sum_{i=1}^{l_t} g^{ij}_{tk} - 1$, by $r(x)$ – expression $\frac{1}{L_t}\sum_{i=1}^{l_t} F^i_t \sum_{j=1}^{L_t} g^{ij}_{tk} - F^k_{t-1}$. The problem (3) – (7) can be equivalently rewritten

$$\min_x f_0(x) = 0.5\sum_{k=1}^{l_{t-1}}\sum_{i=1}^{l_t}\sum_{j=1}^{L_t} g^{ij}_{tk}||F^i_t - f^j_{tk}||^2 \tag{10}$$

$$x \geq 0, q(x) = 0, r(x) = 0 \tag{11}$$

Here $x \in R^{l_{t-1}\cdot l_t \cdot L_t + n_x \cdot l_t}$, $q(x) \in R^{l_{t-1}\cdot L_t}$, $r(x) \in R^{n_x \cdot l_{t-1}}$

The problem (10) – (11) has a large dimensionality, the objective function and its constraints are not convex. Besides that, constraints gradients are not linearly independent. The problem contains equation constraints as well as equality constraints. SQP-algorithms for problems with equality constraints in which constraints gradients are not linearly independent are presented in [Izmailov & Uskov, 2017].

The problem (10) – (11) can be solved iteratively by choosing the vector of descent $s_n = (s^g_n, s^F_n)$ on the step n, $s^g_n \in R^{l_{t-1}\cdot l_t \cdot L_t}$, $s^F_n \in R^{n_x \cdot l_t}$ which optimizes the following quadratic problem:

$$\min_{s_n}\left[(f_0'(x_n, s_n) + \frac{1}{2}(Ds_n, s_n)\right] \tag{12}$$

$$q(x_n) + (q'(x_n), s_n) = 0, r(x_n) + (r'(x_n), s_n) = 0, x_n + s_n \geq 0 \tag{13}$$

where D is a positive-definite matrix.

The problem (12) – (13) is feasible under some assumptions. In particular, constraints gradients should be linearly independent.

For the particular case the problem (12) – (13) can be rewritten as follows

$$\min_{s_n} \left[0.5 \sum_{k=1}^{lt_1} \sum_{i=1}^{l_t} \sum_{j=1}^{L_t} s_{nk}^{gij} ||F_{tn}^i - f_{tk}^j||^2 + \sum_{i=1}^{lt} \sum_{m=1}^{n_x} s_{nm}^{Fi} \sum_{k=1}^{lt_1} \sum_{i=1}^{l_t} (F_{tn,m}^i - f_{tk,m}^j) g_{tnk}^{ij} + \frac{1}{2}(Ds_n, s_n) \right] \tag{14}$$

$$g_{tnk}^{ij} + s_{nk}^{gij} \geq 0, k = 1, \ldots, l_{t-1}, i = 1, \ldots, l_t, j = 1, \ldots, L_t \tag{15}$$

$$F_{tn,m}^i + s_{nm}^{Fi} \geq 0, i = 1, \ldots, l_t, m = 1, \ldots, n_x \tag{16}$$

$$\sum_{i=1}^{l_t} s_{nk}^{gij} = 1, k = 1, \ldots, l_{t-1}, j = 1, \ldots, L_t \tag{17}$$

$$\frac{1}{L_t} \sum_{j=1}^{L_t} g_{tnk}^{ij} \sum_{i=1}^{l_t} F_{tn,m}^i - F_{t-1n,m}^k + \frac{1}{L_t} \sum_{j=1}^{L_t} \sum_{i=1}^{l_t} s_{nk}^{gij} F_{tn,m}^i + \frac{1}{L_t} \sum_{i=1}^{l_t} s_{nm}^{Fi} \sum_{j=1}^{L_t} g_{tnk}^{ij} = 0, k = 1, \ldots, l_{t-1}, m = 1, \ldots, n_x \tag{18}$$

As mentioned above,the matrix D should be positive-definite. For example, D can be an identity matrix of a corresponding dimension $(l_{t-1} \cdot l_t \cdot L_t + n_x \cdot l_t \times l_{t-1} \cdot l_t \cdot L_t + n_x \cdot l_t)$ or, alternatively, this can be a matrix of second-order derivatives on each step:

$$D = \begin{pmatrix} \frac{\partial^2 f_0}{\partial g^2}(x_n) & \frac{\partial^2 f_0}{\partial g \partial F}(x_n) \\ \frac{\partial^2 f_0}{\partial F \partial g}(x_n) & \frac{\partial^2 f_0}{\partial F^2}(x_n) \end{pmatrix}$$

where $\frac{\partial^2 f_0}{\partial g^2}(x_n) = 0$, $\frac{\partial^2 f_0}{\partial F^2}(x_n) = 1$, $\frac{\partial^2 f_0}{\partial g \partial F}(x_n) = \frac{\partial^2 f_0}{\partial F \partial g}(x_n) = F_{t,m}^i - f_{tk,m}^j$.

Another way to set D is to evaluate second-order derivatives only in the starting vector $x_0 = (g_0, F_0)$.

The Algorithm is the following.

A. Get a starting vector $x_0 = (g_0, F_0)$.

B. Solve the problem (12) – (13) and obtain the vectors $s_n = (s_n^g, s_n^F)$ and $y_n = (y_{qn}, y_{rn})$ of optimal and dual solutions correspondingly, where n is the index of the current step

C. Until $\varphi(x_n + \beta s_{n+1}) \leq \varphi(x_n) + \frac{1}{2}\beta\left((f_0'(x_n), s_{n+1}) - G\psi(x_n)\right)$
do $\beta := \frac{1}{2}\beta$,
where $\varphi(x) = f_0(x) + G\psi(x)$,
$\psi(x) = \sum_{i=1}^{n_x \cdot l_{t-1}} |r_i(x)| + \sum \min(0, x)$,
$G = \max_i |y_{ni}| + a, a > 0$ is a penalty parameter,
$\min(0, x)$ is a component-wise minimum of vector x.

D. If β is less than ε, then End.

E. Obtain new values of vector x as $x_{n+1} := x_n + \beta s_{n+1}$

F. Go to the item B

2.3 Scenario Lattice Construction Algorithm

The algorithm of the scenario lattice constructing is the following. We suppose that the principal component analysis has been produced and all selected principal components are presented by the according ARIMA-GARCH models. Denote by v_{tk}^j the vector of the parameters of the ARIMA-GARCH model with the vector of forward prices f_{tk}^j. Further we will address the vectors v_{tk}^j and f_{tk}^j as "a k, j - pattern at the stage t". For the construction of this algorithm put $l_0 = 1$ also.

A. Assign the current values of the forward prices to the vector F_1^1 and the current parameters of ARIMA-GARCH models to the vector $v_{1,1}^1$ which form 1,1-pattern at the stage 1 of the scenario lattice.

B. Construct the degenerated probability distribution of the patterns at the first stage $p_{t-1,m}^n = 1, m = 1, n = 1$.

C. $t := 2$

D. $k := 1$

E. For $j = 1, \dots, L_t$ Do:
Select one from the vectors $v_{t-1,m}^n, m = 1, \dots, l_{t-2}, n = 1, \dots, L_{t-1}$ per the probability distribution $p_{t-1,m}^n, m = 1, \dots, l_{t-2}, n = 1, \dots, L_{t-1}$. Using this vector generate one k, j-pattern for the stage t by Monte-Carlo method.

F. If $k < l_{t-1}$ then $k := k + 1$ and go to the item E.

G. Solve problem (3) – (7) using Algorithms presented above. Assign forward and futures prices $F_t^i, i = 1, , l_t$ to the according nodes of the stage t.

H. Calculate transition probabilities $\rho_{tk}^i, k = 1, \dots, l_{t-1}, i = 1, \dots, l_t$ using formula (2).

I. If $t < T$ then $t := t + 1$, else End.

J. Construct the distribution $p_{t-1,m}^n = \frac{\sum_{k=1}^{l_{t-2}} g_{t-1,k}^{mn}}{l_{t-1} \times L_{t-1}}, m = 1, \dots, l_{t-1}, n = 1, \dots, L_{t-1}$. Go to the item D.

3 Empirical Results

Table 1: Comparison of the algorithms

Parameter	Set 1	Set 2
Number of stages, T	1	3
Vector dimensionality, n_x	5	5
Number of nodes at each stage, l_t, l_{t-1}	6	6
Number of simulations at each stage, L_t	50	50
Value of objective function in the last node, Algorithm 1.1	3.84	6324138.50
Value of objective function in the last node, Algorithm 1.2	2711.96	44598540.30
Value of objective function in the last node, Algorithm 2	1042820.69	2088855976.00
Time consumed (seconds), Algorithm 1.1	45.05	2067.00
Time consumed (seconds), Algorithm 1.2	47.70	896.83
Time consumed (seconds), Algorithm 2	33.90	117.44
Total number of iterations, Algorithm 1.1	422	1211
Total number of iterations, Algorithm 1.2	275	699
Total number of iterations, Algorithm 2	23	32

As an empirical example, the joint dynamic of 5 variables was examined. The first variable represents a spot exchange rate for the currencies USD and RUB, while others are interest rates of different maturities in Russia and the United States. Principal component analysis (PCA) was applied to these variables to construct 5 independent time series. After that ARIMA-GARCH model parameters were estimated for each time series and all principal components were simulated. This allowed to produce, after using the factor loadings matrix, the set of Monte-Carlo realizations of the original variables. Then the scenario lattice was constructed using Algorithms presented in Section 2. The results of the study are presented in the table 1.

As it can be seen from the presented table, the SQP algorithm despite being the most time-efficient in all cases gives the worst approximation of optimal solution. Among algorithms that use variables separation the one based on consecutive optimization over each group of variables is preferable. It should be also mentioned, that after completing of both Algorithm 1.1 and Algorithm 1.2 its solutions were utilized as a starting value for SQP algorithm. The latter one has not improved the value of objective function which means that obtained solutions represent stationary points of the objective function.

References

[Pereira & Pinto, 1991] Pereira, M.V.F. & Pinto, L.M.V.G. (1991). Multi-stage stochastic optimization applied to energy planning. *Mathematical programming, 52(1-3)*, 359-375.

[Heitsch & Romisch, 2003] Heitsch, H. & Romisch, W. (2003). Scenario Reduction Algorithms in Stochastic Programming. *Computational Optimization and Applications, 24(2-3)*, 187-206.

[Consiglio et al., 2014] Consiglio, Andrea & Carollo, Angelo & Zenios, Stavros A. (2014). Generating Multi-factor Arbitrage-Free Scenario Trees with Global Optimization. *Working Papers of The Wharton Financial Institutions Center,13-35.*

[Geyer et al., 2010] Geyer, Alois, Hanke, Michael & Weissensteiner, Alex, (2010). No-arbitrage conditions, scenario trees, and multi-asset financial optimization. *European Journal of Operational Research, 206(3)*, 609-613.

[Izmailov & Solodov, 2014] Izmailov A.F., Solodov M.V. (2014). *Newton-Type Methods for Optimization and Variational Problems.* Switzerland: Springer.

[Evtushenko & Posypkin, 2013] Evtushenko, Y., Posypkin, M. (2013). A deterministic approach to global box-constrained optimization. *Optimization Letters, 7(4)*, 819-829.

[Khamisov, 1999] Khamisov, O. (1999). On Optimization Properties of Functions, with a Concave Minorant. *Journal of Global Optimization, 14(1)*, 79-101.

[Strongin & Sergeyev, 2000] Strongin, R.G., Sergeyev, Y.D. (2000). *Global Optimization with Non-Convex Constraints. Sequential and Parallel Algorithms.*. New York: Springer.

[Gould & Robinson, 2010 (1)] Gould, N.I.M., Robinson, D.P. (2010). A Second Derivative SQP Method: Global Convergence. *SIAM Journal on Optimization, 20(4)*, 2023-2048.

[Gould & Robinson, 2010 (2)] Gould, N.I.M., Robinson, D.P. (2010). A Second Derivative SQP Method: Local Convergence and Practical Issues. *SIAM Journal on Optimization, 20(4)*, 2049-2079.

[Izmailov & Uskov, 2017] Izmailov A.F. & Uskov E.I. (2017). Subspace-stabilized sequential quadratic programming. *Computational Optimization and Applications, Springer, 67(1)*, 129-154.