Gaining traction: on the convergence of an inner approximation scheme for probability maximization

Csaba I. Fábián¹

Published online: 17 July 2020
© The Author(s) 2020

Abstract

We analyze an inner approximation scheme for probability maximization. The approach was proposed in Fábián et al. (Acta Polytech Hung 15:105–125, 2018), as an analogue of a classic dual approach in the handling of probabilistic constraints. Even a basic implementation of the maximization scheme proved usable and endured noise in gradient computations without any special effort. Moreover, the speed of convergence was not affected by approximate computation of test points. This robustness was then explained in an idealized setting. Here we work out convergence proofs and efficiency arguments for a nondegenerate normal distribution. The main message of the present paper is that the procedure gains traction as an optimal solution is approached.

Keywords Convex optimization · Stochastic optimization · Probabilistic problems · Cutting-plane method

1 Introduction

Let $F(\mathbf{z})$ denote an n -dimensional nondegenerate standard normal distribution function. Due to logconcavity of the normal distribution, the probabilistic function $\phi(\mathbf{z}) = -\log F(\mathbf{z})$ is convex. We consider a probability maximization problem in the form

$$\min \phi(T\mathbf{x}) \quad \text{subject to} \quad A\mathbf{x} \leq \mathbf{b}, \quad (1)$$

This research is supported by EFOP-3.6.1-16-2016-00006 project titled 'The development and enhancement of the research potential at John von Neumann University'. The Project is supported by the Hungarian Government and co-financed by the European Social Fund.

✉ Csaba I. Fábián
fabian.csaba@gamf.uni-neumann.hu

¹ Department of Informatics, GAMF, Faculty of Engineering and Computer Science, John von Neumann University, Izsáki út 10, Kecskemet 6000, Hungary

where vectors are $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{b} \in \mathbb{R}^r$, and the matrices T and A are of sizes $n \times m$ and $r \times m$, respectively.

We are going to examine a simple and straightforward solution method, the one proposed in Fábíán et al. (2018). It builds an inner approximation of the epigraph of $\phi(\mathbf{z})$, based on function evaluations in certain test points. A master problem is formulated using this approximation. Further test points are selected in the course of the procedure, with a view of gradually improving the optimum of the master problem.

In the computational study of Fábíán et al. (2018), the speed of the convergence was not affected by approximate computation of test points. This robustness was then explained in an idealized setting, considering a globally well-conditioned objective function. In this paper we work out convergence proofs for the probabilistic function $\phi(\mathbf{z})$ specified above.

1.1 The broader context: inner approximation in probabilistic problems

Our scheme is analogous to the well-known dual approach in the handling of probabilistic constraints of the form

$$T\mathbf{x} \in \mathcal{L}_p = \{\mathbf{z} \mid F(\mathbf{z}) \geq p\}, \quad (2)$$

where $p \gg 0$ is a prescribed probability. The classic dual approach applies an inner approximation of the level set $\mathcal{L}_p$. The approach has been initiated by Prékopa (1990), and the inner approximation was first applied to a probabilistic constraint by Prékopa et al. (1998). In Dentcheva et al. (2000), a cone generation scheme was developed for the continual improvement of the approximation. In this scheme, new test points are found by minimizing a tractable function over the level set $\mathcal{L}_p$. Since this minimization entails a substantial computational effort, the master part of the decomposition framework should succeed with as few test points as possible. Efficient solution methods were developed by Dentcheva et al. (2004) and Dentcheva and Martinez (2013), approximating the original distribution by a discrete one and applying regularization to the master problem. More recently, van Ackooij et al. (2017) employed a special bundle-type method for the solution of the master problem, based on the on-demand accuracy approach of de Oliveira and Sagastizábal (2014). Each of these methods are quite complex, and, looking at them in a chronological order, an increasing level of complexity is noticeable. An effective solver based on this approach demands a sophisticated implementation.

In the classic scheme for probabilistic constraints, finding a new test point amounts to minimization over the level set $\mathcal{L}_p$. In the scheme we are going to examine, a new approximation point is found by unconstrained minimization, with considerably less computational effort. The unconstrained problems can be approximately solved by a simple gradient descent method. Of course the probability maximization problem (1) is easier than the handling of a probabilistic constraint (2). A Newton-type scheme for the handling of the latter was proposed in Fábíán et al. (2019). It requires the approximate solution of a short sequence of problems of the former type. (Initial

problems in this sequence are solved with a large stopping tolerance, and the accuracy is gradually increased.)

Typically, gradient computations of probabilistic functions require a far greater effort than function evaluations. Computing a single non-zero component of a gradient vector will involve an effort comparable to that of computing a function value. (An alternative means of alleviating the difficulty of gradient computation in case of multivariate normal distribution has recently been proposed by Hantoute et al. 2018.)

In comparison with the outer approximation approach widely used in probabilistic programming, we mention that it requires a sophisticated implementation to deal with noise in gradient computation. Even a fairly accurate gradient may result in a cut cutting into the level set. The analogous difficulty does not occur in case of inner approximation. We can easily keep the inner approximating model from bulging out of the epigraph. A procedure that employs an inner approximation of the epigraph will endure noise in gradient computation without any special effort, provided function values are evaluated with appropriate accuracy. Inherent stability of the model enables the application of randomized methods of simple structure. Such methods have been worked out in Fábíán et al. (2019).

1.2 Overview of the paper

Section 2 contains problem and model formulation, under mild assumptions. We formulate the probability maximization problem and its Lagrangian dual. Moreover, we construct respective polyhedral models of this primal-dual pair of convex problems.

Section 3 is a brief overview of the solution method proposed in Fábíán et al. (2018). This is a simple procedure, formulated here as a column generation scheme with a linear programming master problem. The column generating subproblems are unconstrained convex problems, conveniently solved by the steepest descent method. In the computational study reported in Fábíán et al. (2018), we applied approximate solutions of the subproblems, performing just a few line searches in each subproblem. This heuristic procedure never resulted in any substantial increase in the number of the iterations needed to solve a test problem. We actually found that a single line search performs well in practice.

This paper presents a convergence analysis of the simple procedure of Sect. 3, and an explanation of the experimental efficiency of a single line search, in case the objective function is derived from a nondegenerate normal distribution. The objects defined in the remaining sections are not employed in the algorithm itself. Their sole role is to enable convergence proofs and to support efficiency arguments. (The only exception is the box $\mathcal{Z}^\lambda$ in Sect. 8.2, which has some minor role in a modified algorithm.)

Spadework for the analysis is done in Sect. 4, where we show that, with proper initialization of the test points, the solutions of the successive model problems will remain within well-defined bounds. These bounds, however, may be huge and have no practical relevance.

The bounds are applied in Sect. 5, where we present a theoretical estimate concerning the efficiency of a single line search in the column generating subproblem of Sect. 3. Though this estimate is not suitable for practical considerations.

In Sect. 6, we look at the whole column generation scheme from a dual viewpoint, and interpret it as an inexact cutting-plane scheme. The theoretical efficiency estimate of Sect. 5 is interpreted here as a rule that controls cut tolerance. The advantage of this viewpoint is that we can apply classic results concerning the uniform convergence of convex functions and the convergence of exact cutting-plane schemes. We prove that the resulting inexact cutting-plane scheme is convergent. Though this proof does not directly yield practical efficiency results.

In Sect. 7, we return to the primal viewpoint of the column generation scheme. We assume that the probabilistic function $\phi(z)$ is well-conditioned in the optimal solution of the original probability maximization problem. The convergence results of Sect. 6 imply that, starting with a certain iteration in the column generation scheme, we can focus on a certain neighborhood of the optimal solution. We show that, starting with that certain iteration, a practicable version of the theoretical efficiency estimate of Sect. 5 will hold.

In Sect. 8, we consider the master problem in a linear programming form, to be solved by the simplex method. In the simplex context, we look on the estimate of Sect. 7 as a guarantee of the workability of the column-selection rule of our simplex variant. Our arguments will then be based on the practical efficiency of the simplex method.

In Sect. 9, we illustrate the locally well-conditioned character of the probabilistic objective function, and a closing discussion is included in Sect. 10.

2 Problem and model formulation

We assume that the feasible domain of the probability maximization problem (1) is not empty, and is contained in the box $\mathcal{X} = \{x \in \mathbb{R}^m \mid -1 \leq x_j \leq 1 \ (j = 1, \dots, m)\}$. Exploiting the monotonicity of the objective function, problem (1) can be written as

$$\min \phi(z) \quad \text{subject to} \quad z - Tx \leq 0, \quad Ax - b \leq 0, \quad x \in \mathcal{X}. \quad (3)$$

Assumption 1 A significantly high probability can be achieved in the probability maximization problem. Specifically, a feasible point $\check{z}$ is known such that $F(\check{z}) \geq 0.5$.

The purpose of this assumption is not just to provide a starting solution. It will also help keeping the solution process in a safe area where the slope of the $\log(\cdot)$ function is moderate.

A further speciality of the normal distribution function is the existence of a bounded box $\mathcal{Z}$ outside which the probability weight can be ignored. For the sake simplicity, we assume that this box has the form $\mathcal{Z} = \{z \in \mathbb{R}^n \mid -1 \leq z_j \leq 1 \ (j = 1, \dots, n)\}$. Including the constraint $z \in \mathcal{Z}$ in (3) results in the approximating problem

$$\min \phi(z) \quad \text{subject to} \quad z - Tx \leq 0, \quad Ax - b \leq 0, \quad z \in \mathcal{Z}, \quad x \in \mathcal{X}. \quad (4)$$

As observed in Fábían et al. (2019), the difference between the respective optima of problems (3) and (4) is insignificant. (In the proof, we need an upper bound on the

gradients of the objective function $\phi(z) = -\log F(z)$ at the optimal solutions. Such a bound can be derived from Assumption 1.)

The bound $z \in \mathcal{Z}$ in (4) allows regularization of the objective function, in the form of

$$\phi(z) = -\log F(z) + \frac{\rho}{2} \|z\|^2 \quad (5)$$

with some $\rho > 0$. Substituting this regularized objective in (4) makes no significant variation in the objective value of $z \in \mathcal{Z}$, provided ρ is small enough. On the other hand, the regularizing term improves the condition of the objective and ensures that the conjugate function $\phi^*(.)$ is finite valued. In this paper, we are going to work with this regularized objective, with an appropriately small ρ . The aim is just to improve numerical characteristics of $\phi(z)$. We do not wish to omit $\mathcal{Z}$ in the problem formulation, because it will have a role in the algorithm.

Splitting the variables z , we transform (4) into the equivalent form

$$\min \phi(z) \quad \text{subject to} \quad z - z' = 0, \quad z' - Tx \leq 0, \quad Ax - b \leq 0, \quad z' \in \mathcal{Z}, \quad x \in \mathcal{X}. \quad (6)$$

Problem (6) has an optimal solution because the feasible domain is nonempty and bounded.

We relax the constraints $z - z' = 0$; $z' - Tx \leq 0$ and $Ax - b \leq 0$ by introducing respective multiplier vectors $u \in \mathbb{R}^n$; $-v \in \mathbb{R}^n$, $-v \geq 0$ and $-y \in \mathbb{R}^r$, $-y \geq 0$. The Lagrangian is

$$\begin{aligned} L(z, z', x, -u, -v, -y) = & \phi(z) - u^T z \\ & + u^T z' - v^T z' + v^T Tx - y^T Ax + y^T b. \end{aligned} \quad (7)$$

(In a strict sense, we actually applied the multiplier $-u$ to the first constraint.)

The relaxed problem falls apart into three separate minimization problems:

$$\min_z \left\{ \phi(z) - u^T z \right\} + \min_{z' \in \mathcal{Z}} (u - v)^T z' + \min_{x \in \mathcal{X}} (v^T T - y^T A) x + y^T b. \quad (8)$$

The first minimum is by definition the negative of the conjugate function value $\phi^*(u)$. Due to the special form of the box $\mathcal{Z}$, the second minimum is $-\|u - v\|_1$. The third minimum can be computed in a similar manner, and the optimum of the relaxed problem is

$$-\phi^*(u) - \|u - v\|_1 - \|T^T v - A^T y\|_1 + b^T y. \quad (9)$$

Introducing the function

$$\nu(u) := - \sup_{v, y \leq 0} \left\{ -\|u - v\|_1 - \|T^T v - A^T y\|_1 + b^T y \right\} \quad (u \in \mathbb{R}^n), \quad (10)$$

the Lagrangian dual of (6) can be written as

$$\max_{\mathbf{u} \in \mathbb{R}^n} \{ -\phi^*(\mathbf{u}) - v(\mathbf{u}) \} = - \min_{\mathbf{u} \in \mathbb{R}^n} \{ \phi^*(\mathbf{u}) + v(\mathbf{u}) \}. \quad (11)$$

According to the theory of convex duality, this problem has an optimal solution. Even strong duality relationship holds between (6) and (11), because we assumed linear constraints in the primal problem. We could even allow nonlinear convex constraints, assumed that Slater's condition is satisfied. For a recent treatise on Lagrangian duality, see, e.g., Chapter 4 in the book Ruszczyński (2006).

An optimal solution of the primal problem (6) has the form of $(\mathbf{z}^*, \mathbf{z}'^*, \mathbf{x}^*)$. With some laxity, I will call a vector $\mathbf{z}^* \in \mathbb{R}^n$ an optimal solution of (6), if it can be extended into a strict-sense optimal solution $(\mathbf{z}^*, \mathbf{z}'^*, \mathbf{x}^*)$. This $\mathbf{z}^*$ is unique due to the strict convexity of the regularized objective function $\phi(\mathbf{z})$. The objective function is also smooth, hence its conjugate is strictly convex (see Theorem 26.1 in Rockafellar (1970)). Hence the dual problem (11) has a unique optimal solution $\mathbf{u}^*$.

Observation 2 Let $\mathbf{z}^*$ and $\mathbf{u}^*$ denote the respective optimal solutions of the primal problem (6) and the dual problem (11). We have

$$\mathbf{u}^* = \nabla \phi(\mathbf{z}^*). \quad (12)$$

Proof Let us extend $\mathbf{z}^*$ to a strict-sense optimal solution $(\mathbf{z}^*, \mathbf{z}'^*, \mathbf{x}^*)$ of the primal problem (6). Given $\mathbf{u}^*$, let $\mathbf{v}^*, \mathbf{y}^*$ denote an optimal solution of the supremum problem in (10). Then $((\mathbf{z}^*, \mathbf{z}'^*, \mathbf{x}^*), (-\mathbf{u}^*, -\mathbf{v}^*, -\mathbf{y}^*))$ is a saddle point of the Lagrangian (7). Hence the Karush–Kuhn–Tucker conditions for the optimality of $(\mathbf{z}^*, \mathbf{z}'^*, \mathbf{x}^*)$ in (6) hold with Lagrange multipliers $(-\mathbf{u}^*, -\mathbf{v}^*, -\mathbf{y}^*)$, and (12) is part of the Karush–Kuhn–Tucker conditions. Proofs of the cited statements can be found, e.g. in Ruszczyński (2006), Theorems 4.9, 4.7 and 3.34. $\square$

2.1 Polyhedral models

Suppose we have evaluated the function $\phi(\mathbf{z})$ at points $\mathbf{z}_i$ ($i = 0, 1, \dots, k$); we introduce the notation $\phi_i = \phi(\mathbf{z}_i)$ for respective objective values. An inner approximation of $\phi(\cdot)$ is

$$\phi_k(\mathbf{z}) = \min \sum_{i=0}^k \lambda_i \phi_i$$

such that (13)

$$\lambda_i \geq 0 \ (i = 0, \dots, k), \quad \sum_{i=0}^k \lambda_i = 1, \quad \sum_{i=0}^k \lambda_i \mathbf{z}_i = \mathbf{z}.$$

If $\mathbf{z} \notin \text{Conv}(\mathbf{z}_0, \dots, \mathbf{z}_k)$, then let $\phi_k(\mathbf{z}) := +\infty$. A polyhedral model of problem (6) is

$$\begin{aligned} \min \phi_k(\mathbf{z}) \quad \text{subject to} \quad & \mathbf{z} - \mathbf{z}' = \mathbf{0}, \quad \mathbf{z}' - T\mathbf{x} \leq \mathbf{0}, \\ & A\mathbf{x} - \mathbf{b} \leq \mathbf{0}, \quad \mathbf{z}' \in \mathcal{Z}, \quad \mathbf{x} \in \mathcal{X}. \end{aligned} \quad (14)$$

We assume that (14) is feasible, i.e., its optimum is finite. This can be ensured by proper selection of the initial $\mathbf{z}_0, \dots, \mathbf{z}_k$ points. With some laxity, I will call a vector $\bar{\mathbf{z}} \in \mathbb{R}^n$ an optimal solution of (14), if it can be extended into a strict-sense optimal solution $(\bar{\mathbf{z}}, \mathbf{z}', \bar{\mathbf{x}})$.

The convex conjugate of $\phi_k(\mathbf{z})$ is

$$\phi_k^*(\mathbf{u}) = \max_{0 \leq i \leq k} \{\mathbf{u}^T \mathbf{z}_i - \phi_i\}. \quad (15)$$

As $\phi_k^*(\cdot)$ is a cutting-plane model of $\phi^*(\cdot)$, the following problem is a polyhedral model of problem (11):

$$\max_{\mathbf{u} \in \mathbb{R}^n} \{-\phi_k^*(\mathbf{u}) - v(\mathbf{u})\} = -\min_{\mathbf{u} \in \mathbb{R}^n} \{\phi_k^*(\mathbf{u}) + v(\mathbf{u})\}. \quad (16)$$

At the same time, (16) is the linear programming dual of (14).

Let $\bar{\mathbf{z}}$ denote an optimal solution of the primal model problem (14), and let $\bar{\mathbf{u}}$ denote an optimal solution of the dual model problem (16)—existing due to our assumption concerning the feasibility of (14). The following observation was proved in Fábíán et al. (2018).

Observation 3 *We have $\phi_k(\bar{\mathbf{z}}) + \phi_k^*(\bar{\mathbf{u}}) = \bar{\mathbf{u}}^T \bar{\mathbf{z}}$ and hence $\bar{\mathbf{u}}$ is a subgradient of $\phi_k(\mathbf{z})$ at $\mathbf{z} = \bar{\mathbf{z}}$.*

3 A simple solution method

The convex problem (6) can be solved through approximation. The approximating problem (14) can be formulated as a linear programming problem. The idea is to progressively improve the approximation by adding further approximation points, a traditional and widely used approach.

In Fábíán et al. (2018), we applied a straightforward and unsophisticated method for the construction of approximation points. That is an iterative method. The initialization will be presented in Sect. 3.1. An iteration consist in building and solving the model problem as described in Sect. 3.2, and finding an improving approximation point as described in Sect. 3.3.

3.1 Initialization of the model

We initialized the model function (13) with $n + 1$ points (n being the dimension of the distribution). The 2-dimensional case is illustrated in Fig. 1. The shaded area represents the bounded box $\mathcal{Z}$ outside which the probability weight can be ignored.

Fig. 1 Initial test points in a two-dimensional case

The ‘most positive’ vertex of $\mathcal{Z}$ was selected as z_0 , and $z_1, \dots, z_n$ were respective points from the edges adjoining z_0 . In this paper we follow this way of initialization, further adding the vector of Assumption 1, as $z_{n+1} = \check{z}$. This of course always results in a feasible primal model problem (14).

3.2 A linear programming formulation of the model problem

Taking into account the model function definition (13), the primal model problem (14) can obviously be formulated as follows.

$$\begin{aligned}
 \min \quad & \sum_{i=0}^k \lambda_i \phi_i \\
 \text{such that} \quad & \lambda_i \geq 0 (i = 0, \dots, k), \\
 & \sum_{i=0}^k \lambda_i = 1, \quad \perp \quad \vartheta \in \mathbb{R} \\
 & \sum_{i=0}^k \lambda_i z_i - z' = \mathbf{0}, \quad \perp \quad \mathbf{u} \in \mathbb{R}^n \\
 & z' - T\mathbf{x} \leq \mathbf{0}, \\
 & A\mathbf{x} \leq \mathbf{b}, \\
 & z' \in \mathcal{Z}, \\
 & \mathbf{x} \in \mathcal{X}.
 \end{aligned} \tag{17}$$

The variable z of (14) is represented here as the convex combination $\sum \lambda_i z_i$. Hence the splitting constraint $z - z' = \mathbf{0}$ of (14) takes the form $\sum \lambda_i z_i - z' = \mathbf{0}$.

The columns corresponding to the approximation points $z_0, \dots, z_k$ are involved only in the convexity constraint $\sum \lambda_i = 1$ and in the splitting constraint. The dual variable (shadow price) belonging to the convexity constraint will be denoted by ϑ , and the dual vector belonging to the splitting constraint will be denoted by u . (These are shown above, attached to the respective constraints with $\perp$ signs.)

The linear programming problem (17) is easy to solve (and an optimal solution always exists, due to the initialization described above.) Let $(\bar{\lambda}_0, \dots, \bar{\lambda}_k)$ denote a part of an optimal solution of (17). Then an optimal solution $\bar{z}$ of (14) can be constructed in the form

$$\bar{z} = \sum_{i=0}^k \bar{\lambda}_i z_i. \quad (18)$$

3.3 Column generation

We aim to find a new approximation point z_{k+1} whose addition to the model function (13) will result in an improvement in the optimal objective value of (17).

Let $\bar{\vartheta}$ denote the optimal dual variable belonging to the convexity constraint in (17). Similarly, let $\bar{u}$ denote the optimal dual vector belonging to the splitting constraint.

Given a vector $z \in \mathbb{R}^n$, we can add the corresponding column $(1, z, \mathbf{0})$ to the linear programming problem (17), with objective component $\phi(z)$. This is an improving column if its reduced cost is positive; formally, if $\bar{q}(z) > 0$ holds for

$$\bar{q}(z) := (\bar{\vartheta}, \bar{u})^T (1, z) - \phi(z) = \bar{\vartheta} + \bar{u}^T z - \phi(z). \quad (19)$$

A vector yielding a good reduced cost can be found by approximately solving the column generating problem

$$\max_{z \in \mathbb{R}^n} \left\{ \bar{u}^T z - \phi(z) \right\}. \quad (20)$$

In the context of the simplex method, the optimal solution of this problem corresponds to the Markowitz column selection rule. This is a classic column selection rule that is still considered fairly efficient and is widely used.

3.4 Our experience with the iterative method

As we have seen, an improving column for (17) can be found by maximizing the reduced cost function $\bar{q}(z)$. In the computational study reported in Fábíán et al. (2018), we applied a steepest descent method to $-\bar{q}(z)$, a natural starting point being $\bar{z}$. In the first round of test runs, we solved the column generating problems to near-optimality. However, we found the computational effort excessive. (Most of the computational effort was spent in the computation of probabilistic function gradients.) Hence we

decided to perform just a few line searches in each column generating problem (20). This heuristic procedure never resulted in any substantial increase in the number of the iterations needed to solve a test problem. We actually found that a single line search performs well in practice.

3.5 Convergence arguments for an idealized setting

We thought that the experimental efficiency of the above mentioned simplified scheme needs explanation. In Fábián et al. (2018) we explained it in an idealized setting, considering a globally well-conditioned objective function. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ such that the following assumption holds.

Assumption 4 (Bounded Hessians) The function $f(\mathbf{z})$ is twice continuously differentiable, and real numbers α, ω ($0 < \alpha \leq \omega$) exist such that

$$\alpha I \preceq \nabla^2 f(\mathbf{z}) \preceq \omega I \quad (\mathbf{z} \in \mathbb{R}^n).$$

Here $\nabla^2 f(\mathbf{z})$ is the Hessian matrix, I is the identity matrix, and the relation $U \preceq V$ between matrices means that $V - U$ is positive semidefinite.

Let me note that lower-bounded Hessians imply strong convexity – this is an easy consequence of Taylor's theorem. The following well-known theorem can be found e.g., in Chapter 8.6 of Luenberger and Ye (2008). (Ruszczynski 2006 in Chapter 5.3.5, Theorem 5.7 presents a slightly different form.)

Theorem 5 *Let Assumption 4 hold. We minimize $f(\mathbf{z})$ over $\mathbb{R}^n$ using a steepest descent method, starting from a point $\mathbf{z}^0$. Let $\mathbf{z}^1, \dots, \mathbf{z}^j, \dots$ denote the iterates obtained by applying exact line search at each step. Then we have*

$$f(\mathbf{z}^j) - \mathcal{F} \leq \left(1 - \frac{\alpha}{\omega}\right)^j \left[f(\mathbf{z}^0) - \mathcal{F}\right], \quad (21)$$

where $\mathcal{F} = \min_{\mathbf{z}} f(\mathbf{z})$.

If our objective function $\phi(\mathbf{z})$ should have bounded Hessians, then the function $-\bar{q}(\mathbf{z})$ of (19) would inherit this property. (The column generating subproblem can be formulated as minimization of $-\bar{q}(\mathbf{z})$.)

In Fábián et al. (2018), we applied Theorem 5 to $f(\mathbf{z}) = -\bar{q}(\mathbf{z})$ with $\mathbf{z}^0 = \bar{\mathbf{z}}$. We showed that

$$\bar{q}(\mathbf{z}^j) \geq \left[1 - \left(1 - \frac{\alpha}{\omega}\right)^j\right] \max_{\mathbf{z}} \bar{q}(\mathbf{z}) \quad (22)$$

is a straight consequence of the theorem. In view of the Markowitz rule mentioned above, we find a fairly good improving vector in the column generation scheme in a very few iterations, provided the condition number α/ω is not extremely bad.

The efficiency of the whole iterative procedure can then be explained by the practical efficiency of the simplex method.

4 Bounds on optimal solutions

Our first aim is to construct a compact set that contains every optimal solution $\bar{z}$ of any model problem (14). An obvious choice would be the box $\mathcal{Z}$. But, for reasons that will become apparent shortly, we need a set well inside $\mathcal{Z}$.

The constraints $z = z' \leq T\mathbf{x}$ and $\mathbf{x} \in \mathcal{X}$ appear in the model problem (14). Hence the set $\{T\mathbf{x} \mid \mathbf{x} \in \mathcal{X}\} + \mathcal{N}$, where $\mathcal{N}$ denotes the negative (closed) orthant in $\mathbb{R}^n$, contains every optimal solution $\bar{z}$ of (14).

As we include $\tilde{z}$ of Assumption 1 among the initial test points, every optimal solution satisfies also $\phi(\bar{z}) \leq \phi(\tilde{z})$. We have set the regularization term in the objective (5) in such a way that the variation of $\frac{\rho}{2}\|z\|^2$ on $\mathcal{Z}$ is small. This results a reasonably large lower bound for the probability belonging to the optimal solution, like $F(\bar{z}) \geq 0.49$. It follows that the set

$$\mathcal{O}_z = (\{T\mathbf{x} \mid \mathbf{x} \in \mathcal{X}\} + \mathcal{N}) \cap \{z \in \mathbb{R}^n \mid F(z) \geq 0.49\} \quad (23)$$

contains every optimal solution $\bar{z}$ of any model problem (14). In a similar manner, it can be shown that $z^* \in \mathcal{O}_z$ with the optimal solution z^* of the primal problem (6).

The set $\mathcal{O}_z$ is compact. Indeed, let us consider the two intersecting sets on the right-hand side of (23): the former has the negative orthant as its recession cone, while the recession cone of the latter is the positive orthant.

Now we are going to construct an upper bound for the subgradients of the model function on $\mathcal{O}_z$. The idea is to have a wide border around $\mathcal{O}_z$, on which border the model function is finite valued.

Initialization of the approximation points was described in Sect. 3.1, and illustrated in Fig. 1. (The 'most positive' vertex of $\mathcal{Z}$ is selected as z_0 , and $z_1, \dots, z_n$ are respective points from the edges adjoining z_0 .) Let $\mathcal{S} = \text{Conv}(z_0, \dots, z_n)$ denote the convex hull of these $n + 1$ test points. As the box $\mathcal{Z}$ can be extended arbitrarily, we may assume not only that $\mathcal{O}_z \subset \mathcal{S}$ holds, but even that

$$\mathcal{O}_z \subset \mathcal{S}_\eta \quad \text{holds with some } \eta \in (0, 1), \quad (24)$$

where $\mathcal{S}_\eta$ denotes the simplex obtained by shrinking $\mathcal{S}$ towards its barycenter, the ratio of decrease being 1 to η . In what follows, we work with a fixed value η satisfying (24).

Observation 6 *Assume that the model functions have been initialized in such a way that (24) holds.*

Then there exists a finite upper bound Γ_u such that $\|u\| \leq \Gamma_u$ holds for every subgradient u of any model function $\phi_k(z)$ on $\mathcal{O}_z$.

Moreover, with the true objective function, we have $\|\nabla\phi(z)\| \leq \Gamma_u$ ($z \in \mathcal{O}_z$) as well.

The proof is straightforward but somewhat technical, hence I include it in the "Appendix, section A.1".

5 On the efficiency of a single line search

Let $\bar{z}$ denote an optimal solution of the model problem, constructed in the form (18). Let moreover $\bar{u}$ be the optimal dual vector belonging to the splitting constraint in the linear programming form (17) of the model problem. According to the column generating procedure of Sect. 3.3, the next test point $z_{k+1} = z^\lambda$ is obtained by approximate maximization of $\bar{u}^T z - \phi(z)$.

The column generating subproblem is conveniently reformulated as approximate minimization of the function $f(z) = \phi(z) - \bar{u}^T z$, and a steepest descent method can be applied from the convenient starting point $\bar{z}$. In accordance with our experience described in Sect. 3.4, we are actually going to work out proofs for a single line search performed in the opposite direction of the gradient $\bar{g} = \nabla f(\bar{z}) = \nabla \phi(\bar{z}) - \bar{u}$.

We have $\bar{z} \in \mathcal{O}_z$ and $\|\bar{g}\| \leq 2\Gamma_u$ as shown in Sect. 4. Hence any line search is performed on a ray of the form

$$\{ z - t\bar{g} \mid t \geq 0 \} \quad \text{where } z \in \mathcal{O}_z, \quad \|\bar{g}\| \leq 2\Gamma_u. \quad (25)$$

Due to the regularizing term in the objective (5), all eigenvalues of the Hessians are increased by ρ , hence $\nabla^2 \phi(z) \succeq \rho I$ ($z \in \mathbb{R}^n$) holds. Though there exists no finite upper bound on the eigenvalues, it turns out that a local version of Assumption 4 is sufficient.

Given $z \in \mathbb{R}^n$, let $\omega(z)$ denote the largest eigenvalue of $\nabla^2 f(z)$. Continuity of this function is a straight consequence of a theorem of Ostrowski that I cite as Theorem 17 in “Appendix” B. Since the set $\mathcal{O}_z$ is compact, it follows that

$$\Omega = \sup_{z \in \mathcal{O}_z} \omega(z)$$

is finite. With this, let

$$\bar{\mathcal{O}} = \{ z - t\bar{g} \mid z \in \mathcal{O}_z, \quad \|\bar{g}\| \leq 2\Gamma_u, \quad t \in [0, 1/\Omega] \}.$$

This again is a compact set, hence

$$\bar{\Omega} = \sup_{z \in \bar{\mathcal{O}}} \omega(z) \quad (26)$$

is again finite. As $\bar{\mathcal{O}} \supseteq \mathcal{O}_z$, it follows that $\bar{\Omega} \geq \Omega$.

Proposition 7 *With z^λ found by performing a single line search, we have*

$$f(z^\lambda) - \mathcal{F} \leq (1 - \rho/\bar{\Omega}) [f(\bar{z}) - \mathcal{F}], \quad (27)$$

where $\mathcal{F} = \min_z f(z)$.

The proof will be an adaptation of that of Theorem 5, as related in Chapter 8.6 of Luenberger and Ye (2008). We restrict our investigation to $\bar{\mathcal{O}}$.

Proof of Proposition 7 Due to the above construction, we have $\bar{z} - t\bar{g} \in \bar{\mathcal{O}}$ for t such that $0 \leq t \leq 1/\Omega$. Hence $\nabla^2\phi(z) \preceq \bar{\Omega}I$ holds for $z = \bar{z} - t\bar{g}$ with these t values. It follows that

$$f(\bar{z} - t\bar{g}) \leq f(\bar{z}) - t\bar{g}^T\bar{g} + \frac{\bar{\Omega}}{2}t^2\bar{g}^T\bar{g} \quad (0 \leq t \leq 1/\Omega)$$

holds (a consequence of Taylor's theorem). We consider the respective minima in $t \in \mathbb{R}$ separately of the two sides. The right-hand side is a quadratic expression, yielding minimum at $t = 1/\bar{\Omega}$. (Note that $1/\bar{\Omega} \leq 1/\Omega$.) It follows that

$$\min_{t \in \mathbb{R}} f(\bar{z} - t\bar{g}) \leq f(\bar{z}) - \frac{1}{2\bar{\Omega}} \|\bar{g}\|^2. \quad (28)$$

From $\nabla^2\phi(z) \succeq \rho I$ ($z \in \mathbb{R}^n$), it follows in a similar manner that

$$f(z) \geq f(\bar{z}) + \bar{g}^T(z - \bar{z}) + \frac{\rho}{2} \|z - \bar{z}\|^2 \quad (z \in \mathbb{R}^n). \quad (29)$$

The right-hand side expression is a quadratic function of z , which yields its minimum at $z = \bar{z} - \frac{1}{\rho}\bar{g}$. Hence

$$\mathcal{F} \geq f(\bar{z}) - \frac{1}{2\rho} \|\bar{g}\|^2. \quad (30)$$

The proof is concluded by simple transformations of (28) and (30). The improving column z^λ is found by solving the minimization problem in the left-hand side of (28). Subtracting $\mathcal{F}$ from both sides of (28), and applying (30) to underestimate $\|\bar{g}\|^2$, we get (27). $\square$

6 Dual viewpoint: an approximate cutting-plane method

The method to be described in this section is not new. It is just the simple solution method of Sect. 3, described from a new viewpoint that will enable us to apply some classic tools in the convergence proof.

The column generation procedure, as described in Sects. 3.2 and 3.3, yields an inner approximation of the epigraph of the objective function $\phi(z)$. Looking at the same procedure from a dual viewpoint we can see a cutting-plane method. This relationship between the primal and dual approaches is well known, see, e.g., Frangioni (2002, 2018), but it is immediately apparent in the present case: (16) is clearly a cutting-plane model of (11). From the dual point of view, the procedure yields an outer approximation of the epigraph of the conjugate function $\phi^*(u)$.

As the linear programming problem (17) is just a re-formulation of the primal model problem (14), it follows that the dual of (17) is equivalent to the dual model problem (16). Specifically, let $\bar{u}$ denote an optimal dual vector belonging to the splitting

constraint in (17). Then, at the same time, $\bar{\mathbf{u}}$ is an optimal solution of the dual model problem (16).

If we wish to apply a cutting-plane method for the minimization of a convex function $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$, we must be able to construct an approximate support function at any given point $\hat{\mathbf{u}} \in \mathbb{R}^n$. This is a linear function $\ell(\mathbf{u})$ such that $\ell(\mathbf{u}) \leq \varphi(\mathbf{u})$ ($\mathbf{u} \in \mathbb{R}^n$) holds, and the difference $\varphi(\hat{\mathbf{u}}) - \ell(\hat{\mathbf{u}})$ is under control. If the difference is 0 then $\ell(\mathbf{u})$ is an exact support function to $\varphi(\mathbf{u})$ at $\hat{\mathbf{u}}$.

6.1 Problem specification and dual interpretation of former results

Introducing the function $d(\mathbf{u}) := \phi^*(\mathbf{u}) + v(\mathbf{u})$, the dual problem (11) can be solved in the form

$$\min_{\mathbf{u} \in \mathbb{R}^n} d(\mathbf{u}). \quad (31)$$

The k th model function is $d_k(\mathbf{u}) := \phi_k^*(\mathbf{u}) + v(\mathbf{u})$, with $\phi_k^*(\mathbf{u})$ defined in (15). The dual model problem (16) can be solved in the form

$$\min_{\mathbf{u} \in \mathbb{R}^n} d_k(\mathbf{u}). \quad (32)$$

Let $\bar{\mathbf{u}}$ denote an optimal solution of the model problem (32). Moreover let

$$\mathcal{O}_{\mathbf{u}} = \{ \mathbf{u} \in \mathbb{R}^n \mid \|\mathbf{u}\| \leq \Gamma_{\mathbf{u}} \}. \quad (33)$$

with the bound $\Gamma_{\mathbf{u}}$ of Observation 6. This is a compact ball that contains the optimal solution $\bar{\mathbf{u}}$ of the model problem (32), as well as the optimal solution $\mathbf{u}^*$ of the convex problem (31).

An approximate support function to $d(\mathbf{u})$ at $\bar{\mathbf{u}}$ is constructed in the form

$$\ell(\mathbf{u}) = \ell'(\mathbf{u}) + \ell''(\mathbf{u}), \quad (34)$$

where the right-hand-side functions are separate support functions to $\phi^*(\mathbf{u})$ and $v(\mathbf{u})$, respectively.

As for $v(\mathbf{u})$, we can construct an exact support function $\ell''(\mathbf{u})$ by solving the problem (10: $\mathbf{u} = \bar{\mathbf{u}}$). This can be formulated as a linear programming problem whose optimal dual variables will give a subgradient at $\bar{\mathbf{u}}$.

The approximate support function to $\phi^*(\mathbf{u})$ is constructed in the form

$$\ell'(\mathbf{u}) := \mathbf{u}^T \mathbf{z}^\lambda - \phi(\mathbf{z}^\lambda), \quad (35)$$

with an appropriate vector $\mathbf{z}^\lambda$. (Note that $\ell'(\cdot) \leq \phi^*(\cdot)$ holds by the definition of the conjugate function.) An exact support function at $\bar{\mathbf{u}}$ could be obtained by setting $\mathbf{z}^\lambda$ to be the exact maximizer of $\bar{\mathbf{u}}^T \mathbf{z} - \phi(\mathbf{z})$. This is just the column generating problem (20). An approximate solution of the column generating problem yields an approximate support function. Approximate solution of the column generation problem was described in

Sect. 3. Let $\bar{z}$ denote a primal optimal solution, constructed in the form (18). Let moreover z^λ be found by a single line search starting from $\bar{z}$. We analyzed this procedure in Sect. 5. Translating Proposition 7 to the present setting, we get

Corollary 8 *The support function (35) satisfies*

$$\ell'(\bar{u}) \geq \phi^*(\bar{u}) - (1 - \theta) \left[\phi^*(\bar{u}) - \phi_k^*(\bar{u}) \right] = \theta \phi^*(\bar{u}) + (1 - \theta) \phi_k^*(\bar{u}), \quad (36)$$

with the constant $\theta = \rho/\bar{\Omega}$. Here ρ is the regularization constant of (5), and $\bar{\Omega}$ defined in (26). Of course we have $0 < \theta < 1$, and it is independent of the iteration count k .

Proof of Corollary 8 We start with formulating (27) in terms of $\phi(z)$ and $\phi^*(u)$. By the definition $f(z) = \phi(z) - \bar{u}^T z$, we have $\mathcal{F} = -\phi^*(\bar{u})$. Substituting these, we get

$$\phi^*(\bar{u}) + \phi(z^\lambda) - \bar{u}^T z^\lambda \leq (1 - \theta) \left[\phi^*(\bar{u}) + \phi(\bar{z}) - \bar{u}^T \bar{z} \right]. \quad (37)$$

A simple transformation results in

$$\ell'(\bar{u}) \geq \phi^*(\bar{u}) - (1 - \theta) \left[\phi^*(\bar{u}) + \phi(\bar{z}) - \bar{u}^T \bar{z} \right]. \quad (38)$$

By Observation 3, we have $\bar{u}^T \bar{z} = \phi_k^*(\bar{u}) + \phi_k(\bar{z})$. Substituting this in the bracketed expression, and taking into account $\phi(z) \leq \phi_k(z)$ ($z \in \mathbb{R}^n$), we get $\phi^*(\bar{u}) + \phi(\bar{z}) - \bar{u}^T \bar{z} \leq \phi^*(\bar{u}) - \phi_k^*(\bar{u})$, directly yielding (36). $\square$

The support function (34) obviously inherits the property of Corollary 8. That is,

$$\begin{aligned} \ell(\bar{u}) &= \ell'(\bar{u}) + \ell''(\bar{u}) \\ &\geq \left[\theta \phi^*(\bar{u}) + (1 - \theta) \phi_k^*(\bar{u}) \right] + \left[\theta v(\bar{u}) + (1 - \theta) v(\bar{u}) \right] \\ &= \theta d(\bar{u}) + (1 - \theta) d_k(\bar{u}) \end{aligned} \quad (39)$$

holds with the same constant $\theta = \rho/\bar{\Omega}$.

6.2 Convergence

A more detailed notation will be needed, including indexing of the iterates and support functions in the cutting-plane scheme. Let $\bar{u}_1, \dots, \bar{u}_k$ denote the known iterates. Let $\ell_i(u)$ ($i = 1, \dots, k$) denote approximate support functions at the respective iterates. These take the form $\ell_i(u) = \ell'_i(u) + \ell''_i(u)$, where $\ell'_i(u)$ is an approximate support function to $\phi^*(u)$, and $\ell''_i(u)$ is an exact support function to $v(u)$, at $\bar{u}_i$.

The model function $d_k(u)$ is the upper cover of the support functions $\ell_i(u)$ ($i = 1, \dots, k$), and the next iterate $\bar{u}_{k+1}$ is a minimizer of $d_k(u)$. At $\bar{u}_{k+1}$, again, we construct a support function to $d(u)$. By (39),

$$\ell_{k+1}(\bar{u}_{k+1}) \geq \theta d(\bar{u}_{k+1}) + (1 - \theta) d_k(\bar{u}_{k+1}) \quad (40)$$

holds with the constant $\theta = \rho/\sqrt{\Omega}$.

Theorem 9 *We have seen that the compact set $\mathcal{O}_u$ of (33) contains all the iterates $\bar{u}_k$. Moreover, all the approximate supporting functions satisfy (40) with the constant θ independent of k .*

The approximate cutting plane method generates a sequence of models and points satisfying

$$\lim_{k \rightarrow \infty} D_k = D \quad \text{and} \quad \lim_{k \rightarrow \infty} d(\bar{u}_k) = D, \quad (41)$$

where D_k and D are the minima of $d_k(u)$ and $d(u)$, respectively.

A nice elementary proof of the convergence of the exact cutting plane method can be found in Ruszczyński (2006), Theorem 7.7. The following proof uses some of the ideas presented there. To handle inexactness, I'll need a well-known though strong theorem from convex analysis. Using that, the proof becomes surprisingly simple.

Proof of Theorem 9 We have $d_1(u) \leq d_2(u) \leq \dots \leq d(u)$ ($u \in \mathbb{R}^n$), hence

$$d_\infty(u) = \lim_{k \rightarrow \infty} d_k(u) \quad (u \in \mathbb{R}^n)$$

exists and is finite. $d_\infty(u)$ is a convex function and the sequence of the model functions converges uniformly on the compact set $\mathcal{O}_u$ —see, e.g., Theorem 10.8 in Rockafellar's book Rockafellar (1970).

We have $D_k = \min_{u \in \mathbb{R}^n} d_k(u) = d_k(\bar{u}_{k+1})$. Moreover, let $D_\infty = \min_{u \in \mathbb{R}^n} d_\infty(u)$. These are finite and $D_k \leq D_\infty \leq D$. With any index k , we have

$$d_k(\bar{u}_{k+1}) = D_k \leq D_\infty \leq d_\infty(\bar{u}_{k+1}). \quad (42)$$

Let us now take into account the uniform convergence of the sequence of the model functions. Given $\epsilon > 0$, there exist a finite number K_ϵ such that $|d_k(u) - d_\infty(u)| \leq \epsilon$ holds for $k \geq K_\epsilon$, $u \in \mathcal{O}_u$. Specifically, this holds with $u = \bar{u}_{k+1}$, showing that the difference between the left-hand side and the right-hand side of (42) is small for large enough k . It follows that

$$\lim_{k \rightarrow \infty} D_k = D_\infty = \lim_{k \rightarrow \infty} d_\infty(\bar{u}_{k+1}). \quad (43)$$

Now let $(\bar{u}_{k_i+1})$ be a convergent subsequence of $(\bar{u}_{k+1})$, and let $\tilde{u} := \lim_{i \rightarrow \infty} \bar{u}_{k_i+1}$. We have

$$d(\bar{u}_{k_i+1}) \rightarrow d(\tilde{u}) \quad \text{as} \quad i \rightarrow \infty, \quad (44)$$

due to the continuity of $d(u)$.

With any index k , we have $d_\infty(\bar{u}_{k+1}) \geq \ell_{k+1}(\bar{u}_{k+1})$. Taking into account (40), we get

$$d_\infty(\bar{u}_{k+1}) \geq \theta d(\bar{u}_{k+1}) + (1 - \theta) D_k. \quad (45)$$

Here we have $d_\infty(\bar{\mathbf{u}}_{k+1}) \rightarrow D_\infty$ and $D_k \rightarrow D_\infty$ as $k \rightarrow \infty$, according to (43). As for the remaining expression $d(\bar{\mathbf{u}}_{k+1})$, we have (44). Taking limits in (45), we get $D_\infty \geq d(\tilde{\mathbf{u}})$. But $d(\tilde{\mathbf{u}}) \geq D \geq D_\infty$ by definition. It follows that $d(\tilde{\mathbf{u}}) = D = D_\infty$ holds.

The proof can be completed in an indirect fashion. Assume that (41) does not hold, i.e., there exists some $\epsilon > 0$ and a subsequence $(\bar{\mathbf{u}}_{k_j})$ such that $|d(\bar{\mathbf{u}}_{k_j}) - D| > \epsilon$ for every j . But due to the compactness of the domain $\mathcal{O}_u$, there exists a convergent subsequence of $(\bar{\mathbf{u}}_{k_j})$, which is a contradiction with the argument above. $\square$

Using the notation of Sect. 2, the $\mathbf{z}$ -part of the optimal solution of the true problem (6) will be denoted by $\mathbf{z}^*$, and the optimal solution of the true dual problem (11) will be denoted by $\mathbf{u}^*$.

Corollary 10 *We have $\bar{\mathbf{u}}_k \rightarrow \mathbf{u}^*$ and $\bar{\mathbf{z}}_k \rightarrow \mathbf{z}^*$.*

Proof Strict convexity of $\phi^*(\mathbf{u})$ results in strict convexity of $d(\mathbf{u})$. The latter, combined with the second statement of Theorem 9, yields $\bar{\mathbf{u}}_k \rightarrow \mathbf{u}^*$.

Using the notation of Theorem 9, $-D_k$ is the maximum of the dual model problem (16) which in turn equals the minimum of the primal model problem (14), the latter minimum being $\phi_k(\bar{\mathbf{z}}_{k+1})$. Similarly, $-D$ is the maximum of the true dual problem (11) which in turn equals the minimum of the true primal problem (6), the latter minimum being $\phi(\mathbf{z}^*)$. As we mentioned in Sect. 2, a strong duality relationship holds between the true primal and the true dual problem.

Hence the statement $D_k \rightarrow D$ of Theorem 9 translates to primal setting in the form

$$\phi_k(\bar{\mathbf{z}}_{k+1}) \rightarrow \phi(\mathbf{z}^*). \quad (46)$$

As $\phi_k(\mathbf{z})$ is an upper approximation of $\phi(\mathbf{z})$, we have $\phi_k(\bar{\mathbf{z}}_{k+1}) \geq \phi(\bar{\mathbf{z}}_{k+1}) \geq \phi(\mathbf{z}^*)$ for every k . Taking into account (46), we get $\phi(\bar{\mathbf{z}}_{k+1}) \rightarrow \phi(\mathbf{z}^*)$. Strict convexity of $\phi(\mathbf{z})$ then yields $\bar{\mathbf{z}}_k \rightarrow \mathbf{z}^*$. $\square$

7 Line search revisited

In this section we review the efficiency of applying a single line search in the solution of the column generating problem. We assume that the probabilistic objective function is locally well-conditioned. (I am going to make a case for this assumption in Sect. 9.) We are going to consider the convergence results of Sect. 6 in the primal setting. They will enable us to focus on a locality where the objective function is well-conditioned. Our aim is to develop a practicable version of Proposition 7.

Assumption 11 (*Locally well-conditioned objective*) The function $\phi(\mathbf{z})$ is well-conditioned in the optimal solution $\mathbf{z}^*$ of the true problem (6).

Formally, $\alpha^*/\omega^* \gg 0$ holds, where ω^* and α^* denotes the smallest and the largest eigenvalue, respectively, of the Hessian matrix $\nabla^2\phi(\mathbf{z}^*)$.

Let $\alpha(\mathbf{z})$ and $\omega(\mathbf{z})$ denote the smallest and the largest eigenvalue, respectively, of the Hessian matrix $\nabla^2\phi(\mathbf{z})$ as a function of $\mathbf{z} \in \mathbb{R}^n$. Continuity of both functions follows from a theorem of Ostrowski that I cite in “Appendix B”.

Let the real numbers α^- , ω^+ be such that $0 < \alpha^- < \alpha^*$ and $\omega^* < \omega^+$. (The idea is to have α^- near α^* , and ω^+ near ω^* .) From the continuity of the functions $\alpha(z)$ and $\omega(z)$, it follows that there exists a ball $\mathcal{B}_r$ around z^* with radius $r > 0$, such that $\alpha(z) \geq \alpha^-$ and $\omega(z) \leq \omega^+$ holds for every $z \in \mathcal{B}_r$, and hence

$$\alpha^- I \leq \nabla^2 \phi(z) \leq \omega^+ I \quad (z \in \mathcal{B}_r). \quad (47)$$

Let us consider the convergence results of Sect. 6 in the primal setting. I start by reviewing the notation. $\bar{z}_k$ denotes an optimal solution of the $(k-1)$ th primal model problem, and $\bar{u}_k$ denotes an optimal solution of the $(k-1)$ th dual model problem, for $k = 1, 2, \dots$. In the column generating subproblem, an improving column is found by approximate minimization of the function $f_k(z) = \phi(z) - \bar{u}_k^T z$. Specifically, z_k° denotes the point found by performing a single line search from $\bar{z}_k$, in the opposite direction of $\bar{g}_k = \nabla f_k(\bar{z}_k)$.

Observation 12 *We have $\bar{g}_k \rightarrow 0$.*

Proof We have

$$\bar{g}_k = \nabla \phi(\bar{z}_k) - \bar{u}_k = [\nabla \phi(\bar{z}_k) - \nabla \phi(z^*)] + [u^* - \bar{u}_k],$$

the latter equality being a consequence of Observation 2. In the right-hand side, we have $\bar{u}_k \rightarrow u^*$ by Corollary 10, and $\nabla \phi(\bar{z}_k) \rightarrow \nabla \phi(z^*)$ by Corollary 10 and $\phi(z)$ being continuously differentiable. $\square$

For technical reasons, we'll need some further notation. Let z_k° be the exact minimizer of the function $f_k(z)$. (This is never computed in the course of the column generation scheme.)

Observation 13 *We have $z_k^\circ \rightarrow z^*$.*

The proof is simple but rather technical, hence I include it in the ‘‘Appendix, section A.2’’.

7.1 Focusing on a locality

Given the radius $r > 0$ of the ball $\mathcal{B}_r$ of (47), there exists a natural number K_r such that

$$\|\bar{z}_k - z^*\| \leq \frac{1}{2} r, \quad \|\bar{g}_k\| \leq \frac{1}{2} r \omega^+, \quad \text{and} \quad \|z_k^\circ - z^*\| \leq r \text{ holds with each } k \geq K_r. \quad (48)$$

The above statement holds because we have $\bar{z}_k \rightarrow z^*$, $\bar{g}_k \rightarrow 0$, $z_k^\circ \rightarrow z^*$ by Corollary 10 and Observations 12 and 13.

In the remainder of this section, we consider a fixed $k \geq K_r$. For the sake of simplicity, we are going to omit the lower index k . Let me sum up this simplified notation. An improving column is found by approximate minimization of the function

$f(z) = \phi(z) - \bar{u}^T z$. Specifically, z^λ denotes the point found by performing a single line search from $\bar{z}$, in the opposite direction of $\bar{g} = \nabla f(\bar{z})$. The exact minimizer of $f(z)$ is denoted by z° .

Theorem 14 *We have*

$$f(z^\lambda) - \mathcal{F} \leq (1 - \alpha^-/\omega^+) [f(\bar{z}) - \mathcal{F}] \quad (49)$$

where $\mathcal{F} = \min_z f(z)$.

The proof will be an adaptation of the proof of Proposition 7. It turns out that we only need $\nabla^2 \phi(z) \preceq \omega^+ I$ for $z = \bar{z} - t\bar{g}$ such that $0 \leq t \leq \frac{1}{\omega^+}$. Moreover, $\nabla^2 \phi(z) \succeq \alpha^- I$ is only needed for $z \in [\bar{z}, z^\circ]$.

Proof of Theorem 14 Due to the first and the second assumption in (48), we have $\bar{z} - t\bar{g} \in \mathcal{B}_r$ for $t \in [0, 1/\omega^+]$. Let $T = 1/\omega^+$. According to (47), $\nabla^2 \phi(z) \preceq \omega^+ I$ holds for $z = \bar{z} - t\bar{g}$ with $t \in [0, T]$. It follows that

$$f(\bar{z} - t\bar{g}) \leq f(\bar{z}) - t\bar{g}^T \bar{g} + \frac{\omega^+}{2} t^2 \bar{g}^T \bar{g} \quad (0 \leq t \leq T)$$

holds (a consequence of Taylor's theorem). We consider the respective minima in $t \in \mathbb{R}$ separately of the two sides. The right-hand side is a quadratic expression, yielding minimum at $t = 1/\omega^+ \in [0, T]$. It follows that

$$\min_{t \in \mathbb{R}} f(\bar{z} - t\bar{g}) \leq f(\bar{z}) - \frac{1}{2\omega^+} \|\bar{g}\|^2. \quad (50)$$

Coming to lower bounds, we have $[\bar{z}, z^\circ] \subset \mathcal{B}_r$ according to the first and the third assumption in (48).

By (47), we have $\nabla^2 \phi(z) \succeq \alpha^- I$ ($z \in [\bar{z}, z^\circ]$). From this, it follows by Taylor's theorem that

$$f(z^\circ) \geq f(\bar{z}) + \bar{g}^T (z^\circ - \bar{z}) + \frac{\alpha^-}{2} \|z^\circ - \bar{z}\|^2. \quad (51)$$

The left-hand side is $\mathcal{F}$ by definition. A lower estimate of the right-hand side is obtained by taking its minimum in z° . The minimum is easily computed as the right-hand side expression is a quadratic function of z° . We get

$$\mathcal{F} \geq f(\bar{z}) - \frac{1}{2\alpha^-} \|\bar{g}\|^2. \quad (52)$$

The proof is concluded by simple transformations of (50) and (52). The improving column z^λ is found by solving the minimization problem in the left-hand side of (50). Subtracting $\mathcal{F}$ from both sides of (50), and applying (52) to underestimate $\|\bar{g}\|^2$, we get (49). $\square$

8 Arguments based on the efficiency of the simplex method

We are going to interpret the column generation scheme of Sect. 3 as the solution process, by the simplex method, of a single linear programming problem. Specifically, the procedure for finding improving columns, described in Sect. 3.3, will be interpreted as a column selection procedure in the context of the simplex method.

We assume a locally well-conditioned $\phi(\mathbf{z})$, and are going to apply the results of Sect. 7. As the column generation scheme progresses, we are bound to arrive at a neighborhood on which the objective function is well-conditioned. For a more precise exposition, let me include a summary of the first half of Sect. 7. Let the real numbers α^- , ω^+ be such that $0 < \alpha^- < \alpha^*$ and $\omega^* < \omega^+$. There exists a ball $\mathcal{B}_r$ around $\mathbf{z}^*$ with radius $r > 0$, such that (47) holds. For the radius r , in turn, there exists a natural number K_r such that (48) holds.

From that point on (i.e., in any iteration $k > K_r$), our column selection procedure will conform to a practical simplex rule, as I am going to show in the remainder of this section. The practical efficiency of the simplex method will then explain the behavior of our method.

In accordance with former notation, let $\bar{\mathbf{z}}$ denote the vector constructed in the form of (18), using optimal weights from the linear programming problem (17). Let moreover $\bar{\vartheta}$ and $\bar{\mathbf{u}}$ denote the optimal dual variable and the optimal dual vector belonging to the convexity constraint and the splitting constraint, respectively, in (17).

In our column generation scheme, an improving column is found by approximate minimization of the function $f(\mathbf{z}) = \phi(\mathbf{z}) - \bar{\mathbf{u}}^T \mathbf{z}$. This is of course the same problem as the approximate minimization of $-\bar{\varrho}(\mathbf{z})$, the negative of the reduced cost function whose definition I copy here for convenience: $-\bar{\varrho}(\mathbf{z}) = \phi(\mathbf{z}) - \bar{\mathbf{u}}^T \mathbf{z} - \bar{\vartheta}$.

Let $\mathbf{z}^\backslash$ denote the point found by performing a single line search from $\bar{\mathbf{z}}$, in the opposite direction of $\bar{\mathbf{g}} = \nabla \phi(\bar{\mathbf{z}}) - \bar{\mathbf{u}}$. Theorem 14 is applicable to $f(\mathbf{z}) = -\bar{\varrho}(\mathbf{z})$, and (49) transforms to

$$-\bar{\varrho}(\mathbf{z}^\backslash) + \mathcal{R} \leq (1 - \alpha^-/\omega^+) [-\bar{\varrho}(\bar{\mathbf{z}}) + \mathcal{R}] \quad (53)$$

where $\mathcal{R} = \max_{\mathbf{z}} \bar{\varrho}(\mathbf{z})$.

In Fábían et al. (2018), we have shown that $\bar{\varrho}(\bar{\mathbf{z}}) \geq 0$ holds (as an obvious consequence of the convexity of $\phi(\mathbf{z})$.) Hence the right-hand side of (53) is majorated by $(1 - \alpha^-/\omega^+) \mathcal{R}$. A simple transformation then results in

$$\bar{\varrho}(\mathbf{z}^\backslash) \geq \frac{\alpha^-}{\omega^+} \mathcal{R}. \quad (54)$$

In the simplex context, (54) means that the vector $\mathbf{z}^\backslash$ has a substantial reduced cost in the linear programming master problem, hence we may expect a substantial improvement from including it as a new column. Equation (54) is a realistic counterpart of (22) that was obtained in an idealized setting. More importantly, (54) is an approximate version of the Markowitz column selection rule, well-known in the context of the simplex method.

8.1 Arguments based on experimental results

A long-standing rule of thumb states that, in solving a linear programming problem having M rows and N columns (with $N \gg M$), the number of pivot steps needed is proportional with M . As confirmation, let me cite from the computational study of Cutler and Wolfe (1963): 'It would appear that the rule of "2M iterations" from folklore is fairly good ...' After a more detailed examination, they observe: '... an estimate of between M and $3M$ iterations will almost always be correct.'

In our case, the linear programming model problem (17) was built using a finite set of test points $z_0, z_1, \dots, z_k$. Let us include every vector $z \in \mathbb{R}^n$ among the initial test points; the resulting infinite problem will then be equivalent to the convex problem (6). Though the number of the columns is infinite, the construction (18) of the optimal solution $\bar{z}$ is still usable, because the number of the rows and hence the cardinality of the bases remains finite.

Our column generation scheme is equivalent to a simplex method applied to this infinite linear programming problem. As we have seen in the previous section, the column selection rule is such that the incoming column x^i satisfies (54) from iteration K_r . From this point on, we can expect the simplex method to be fairly efficient.

8.2 Arguments based on theoretical results

Most of the theoretical results concerning the behavior of the simplex method are incapable of explaining its practical efficiency. Specifically, results based on worst-case analysis are generally very far from experimental results. The results of Karl Heinz Borgwardt are fundamental to bridging this gap. He studies average behavior. A further important contribution is Spielman and Teng (2004). The results of Dempster and Merkovsky (1995) are related, from a dual viewpoint. They presented a geometrically convergent version of the cutting-plane method, for a continuously differentiable, strictly convex case.

The review paper Borgwardt (2009) deals with rudimentary problem types of maximizing a linear function subject to $Ax \leq b$ or $Ax \leq b, x \geq 0$ or $Ax = b, x \geq 0$. The matrix has M rows and N columns. The matrix and the objective vectors are populated with random variables of special joint distributions. Borgwardt focuses on the rotation symmetry model: $b = 1$ and the rows of A and the objective vectors are distributed on $\mathbb{R}^n \setminus \{0\}$ independently, identically and symmetrically under rotations. Taking into account several rotation-symmetric distributions and several special simplex-type algorithms, he examines the expected number of the pivot steps required. In Borgwardt (2009), he summarizes several related bounds for problems with $M \geq N$, and also asymptotic results for N fixed and $M \rightarrow \infty$. These are typically of the form

$$M^{\frac{1}{N}} N^{\nu} C \quad (55)$$

with a moderate ν and a constant C . Let us play, for a moment, with the idea that bounds of the type (55) grasp the average behavior of the simplex method in a general sense.

Let us make a minor extension to the column generation scheme described in Sect. 3. There exists a bounded box $\mathcal{Z}^\lambda$ and a natural number K^λ such that $\mathcal{Z}^\lambda$ contains every possible improving vector z^λ that we may obtain, in iterations after K^λ , by approximately solving the column generating problem. In the following arguments, we need to know $\mathcal{Z}^\lambda$ explicitly, but need not know K^λ . The construction of $\mathcal{Z}^\lambda$ is easy but somewhat technical, hence I include it in the “Appendix, section A.3”.

We could include every $z \in \mathcal{Z}^\lambda$ among the initial test points of the linear programming model problem (17), like we did in the previous section. But here we wish to avoid an infinite problem. Though we can settle for a closely approximating problem that has a finite (huge) number of columns. This is easily constructed. Given a prescribed precision $\varepsilon > 0$, let us divide the box $\mathcal{Z}^\lambda$ into small boxes whose edges are not longer than ε . Let us include the center of each cell among the initial test points. The number of the initial test points, considered as a function of the prescribed accuracy, is on the order of $(1/\varepsilon)^n$.

We can now make a slight modification in the column generation process. In a certain iteration, let z^λ denote the vector obtained by performing a line search. If this falls into the box $\mathcal{Z}^\lambda$, then we add no new test point, because there is a cell center close enough among the initial test points. In case $z^\lambda \notin \mathcal{Z}^\lambda$, we add z^λ as a new test point, like in the original scheme. We know that there exists a natural number K^λ such that no new columns are added after iteration K^λ . I.e., we have an unchanging master problem from that point.

For iterations $k > K^\lambda$, this modified column generation scheme can be modelled as the solution, by the simplex method, of the unchanging master problem. Let z^λ denote the vector obtained by a line search in the column generation scheme. In the simplex method, we can then select an existing column (i.e., cell center) closest to z^λ . For iterations $k > \max\{K^\lambda, K_r\}$, this is a practicable selection rule in the simplex context. Let us consider the number of the iterations needed to finish from here, as a function of the prescribed accuracy ε .

Let us assume that a bound of the form (55) applies in the present setting. In the dual of our huge linear programming problem, the number M of the rows is on the order of $(1/\varepsilon)^n$. As for the number N of columns, we have $N > n$ (and N is of a moderate magnitude.) Hence we may expect the iteration count to be on the order of $1/\varepsilon$, that is by no means discouraging.

9 Illustration of the locally well-conditioned character of the probabilistic objective function

In Sect. 7, we assumed that the objective function is well-conditioned in the optimal solution, i.e., $\alpha^*/\omega^* \gg 0$ holds with the smallest and largest eigenvalue, respectively, of the Hessian matrix $\nabla^2\phi(z^*)$ (Assumption 11). Let me make a case for this assumption by a simple illustration.

Figure 2 shows contour lines of a two-dimensional standard normal distribution function $F(z)$, where the covariance between the marginals is 0.5.

Fig. 2 Contour lines of a two-dimensional normal distribution function $F(z)$, and optimal solutions z^* , z^{**} . From top right, the contour lines belong to the probabilities 0.99, 0.95, and 0.90, respectively. The contour lines tend to be straight as we move away from the diagonal $z_1 = z_2$

Let us consider a probability maximization problem of the form (3), with the unregularized objective function $\phi(z) = -\log F(z)$. The feasible region for the z vector has the negative orthant for regression cone.

Let z^* in Fig. 2 denote an optimal solution, with $F(z^*) = 0.9$. Let z^{**} denote another point on the same contour line, i.e., we have $F(z^{**}) = 0.9$ as well. As the section of the contour line between z^* and z^{**} is practically straight and horizontal, we have $z^{**} \leq z^*$, implying that z^{**} is a feasible solution. Hence z^{**} is also an optimal solution. This shows that an optimal solution can be located 'near' the curved part of a contour line, i.e., 'near' the diagonal $z_1 = z_2$.

In case the optimal probability falls between 0.9 and 0.99, we may locate an optimal z^{**} in the neighborhood of (2, 2), where the objective function is well-conditioned. Eigenvalues of the Hessian matrix $\nabla^2[-\log F(2, 2)]$ are $\alpha(2, 2) = 0.0929$ and $\omega(2, 2) = 0.1140$, their ratio beeing above 0.8.

Contour lines of the eigenvalue functions $\alpha(z)$ and $\omega(z)$ for $z \in [-6, +6]^2$ are shown in Fig. 3, taken from Fábíán et al. (2018). $\alpha(z)$ and $\omega(z)$ are the smaller and the larger eigenvalue, respectively, of the Hessian matrix $\nabla^2\phi(z) = \nabla^2[-\log F(z)]$.

10 Discussion

We worked out a convergence proof and presented efficiency arguments, for a simple column generation scheme. The rigorous convergence statements of Sect. 6 do not

Fig. 3 Contour lines of the smaller and the larger eigenvalue, respectively, of $\nabla^2[-\log F(z)]$ as a function of z , for a two-dimensional normal distribution function $F(z)$. In the left-hand figure (smaller eigenvalue), contour lines from top right are $1e - 5$, $1e - 4$, $1e - 3$, $1e - 2$. In the area not shaded gray, the smaller eigenvalue is above $1e - 5$. In the right-hand figure (larger eigenvalue), contour lines from top right are 1, 1.2, 1.4, 1.6. In the area not shaded gray, the larger eigenvalue is below 1.6

directly yield a practicable efficiency estimate. However, they enable us to focus on a neighborhood of the optimal solution (where we may depend on the locally well-conditioned character of the objective function.) In Sect. 7, we argued that, starting with iteration K_r , the process remains in a well-defined neighborhood. The threshold K_r may, in theory, be large. But in our experiments, a rough approximate solution was always found relatively quickly, and the main computational effort was devoted to refine the approximation. The diagrams included in Fábíán et al. (2019) show similar behavior. In that paper, we developed a randomized version of the method, and performed an experimental study. We always found a rough approximate solution with a relatively small effort.

In (5), we regularized the probabilistic objective function by adding the term $\frac{\rho}{2} \|z\|^2$ with an appropriate $\rho > 0$. In the present paper, we worked with a fixed ρ . However, from a practical point of view, it makes sense to start with a large ρ , and decrease it gradually as the optimal solution is approached. The resulting procedure, when observed from a dual viewpoint, shows an interesting relationship with bundle methods. In the context of Sect. 6, the regularization (5) can be interpreted as a smoothing operation applied to the conjugate function $\phi^*(u)$ in (31). This smoothing operation improves the conditioning of $\phi^*(u)$. (On the inverse relationship between Hessian matrices of $\phi(z)$ and $\phi^*(u)$, see, e.g., Gorni 1991.)

Acknowledgements Open access funding provided by John von Neumann University. The author gives his thanks to the editors of the special issue 'VOCAL and Hungarian OR Conference' and to the anonymous reviewers for many helpful suggestions and constructive remarks.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If

material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Technical proofs

A.1 Bounding the subgradients of the model functions

Before proving Observation 6, we need some preliminary observations.

It is easily seen that the shrunk simplex $\mathcal{S}_\eta$ of (24) can be obtained in the form

$$\mathcal{S}_\eta = \left\{ \sum_{i=0}^n \lambda_i \mathbf{z}_i \mid \sum_{i=0}^n \lambda_i = 1, \lambda_i \geq \frac{1}{n+1} (1-\eta) \right\}. \quad (56)$$

Due to $\eta < 1$, the common lower bound of the weights is positive.

Let $\mathcal{R} := \{\boldsymbol{\varrho} \in \mathbb{R}^n \mid \|\boldsymbol{\varrho}\| = 1\}$ denote the unit sphere.

Lemma 15 *There exists a positive constant Δ such that*

$$\max_{0 \leq i \leq n} \boldsymbol{\varrho}^T (\mathbf{z}_i - \mathbf{z}) \geq \Delta$$

holds for any $\mathbf{z} \in \mathcal{S}_\eta$ and $\boldsymbol{\varrho} \in \mathcal{R}$.

Proof Let $\mathbf{z} = \sum_{i=0}^n \lambda_i \mathbf{z}_i$ be the representation according to (56). We have

$$\sum_{i=0}^n \lambda_i \boldsymbol{\varrho}^T (\mathbf{z}_i - \mathbf{z}) = 0, \quad (57)$$

because $\sum \lambda_i (\mathbf{z}_i - \mathbf{z}) = \mathbf{0}$ holds by the definition of the weights λ_i . As the simplex $\mathcal{S}$ is non-degenerate, we cannot have $\boldsymbol{\varrho}^T (\mathbf{z}_i - \mathbf{z}) = 0$ for all i . Moreover, as the weights λ_i are all positive, a nonzero $\boldsymbol{\varrho}^T (\mathbf{z}_i - \mathbf{z})$ results in a nonzero addend in (57). Hence there exists i such that $\boldsymbol{\varrho}^T (\mathbf{z}_i - \mathbf{z}) > 0$. It means that the function

$$(\boldsymbol{\varrho}, \mathbf{z}) \mapsto \max_{0 \leq i \leq n} \boldsymbol{\varrho}^T (\mathbf{z}_i - \mathbf{z})$$

is positive-valued on the compact set $\mathcal{R} \times \mathcal{S}_\eta$. Since the function is also continuous, it obtains its infimum by Weierstrass' extreme value theorem. Let $\Delta > 0$ denote the minimum. $\square$

Proof of Observation 6 Given $\mathbf{z} \in \mathcal{O}_z$, we have $\mathbf{z} \in \mathcal{S}_\eta$ according to (24). Let ϕ denote the current model function value at $\mathbf{z}$. Moreover, let $\mathbf{u}$ be a subgradient of the current model function at $\mathbf{z}$. We put it in the form $\mathbf{u} = t\boldsymbol{\varrho}$ with some $t \geq 0$ and $\boldsymbol{\varrho} \in \mathcal{R}$, and will find a bound on $t = \|\mathbf{u}\|$. By the definition of the subgradient, we get

$$t\boldsymbol{\varrho}^T (\mathbf{z}_i - \mathbf{z}) \leq \phi_i - \phi \quad (i = 0, \dots, n), \quad (58)$$

where ϕ_i denotes the objective value at z_i . As the model function is an upper approximation of the true objective, the model function value satisfies

$$\phi \geq \phi(z) \geq \min_{z \in \mathcal{S}_\eta} \phi(z). \quad (59)$$

The latter minimum exists and is finite as the objective function is continuous and $\mathcal{S}_\eta$ is compact. Hence all the right-hand sides in (58) are bounded from above by $\max_{0 \leq i \leq n} \phi_i - \min_{z \in \mathcal{S}_\eta} \phi(z)$.

According to Lemma 15, we have $q^T(z_i - z) > \Delta$ for some i . It follows that

$$t\Delta \leq \max_{0 \leq i \leq n} \phi_i - \min_{z \in \mathcal{S}_\eta} \phi(z).$$

Recall that $t = \|u\|$. This gives an upper bound Γ_u as claimed in the first part of Observation 6. The statement with the true objective function, $\|\nabla \phi(z)\| \leq \Gamma_u$ ($z \in \mathcal{S}_\eta$), can be proved in a similar manner. $\square$

A.2 Convergence of improving columns

Proof of Observation 13 In accordance with former notation, $\bar{z}_k$ denotes an optimal solution of the current primal model problem, and $\bar{u}_k$ denotes an optimal solution of the current dual model problem. Moreover, z_k° denotes the exact maximizer of the function $\bar{u}_k^T z - \phi(z)$.

By the definition of z_k° , we have

$$\bar{u}_k^T \bar{z}_k - \phi(\bar{z}_k) \leq \bar{u}_k^T z_k^\circ - \phi(z_k^\circ). \quad (60)$$

Due to the regularized objective (5), we have $\nabla^2 \phi(z) \succeq \rho I$ ($z \in \mathbb{R}^n$), which, in turn, implies strong convexity. Specifically,

$$\phi(z_k^\circ) \geq \phi(z^*) + [\nabla \phi(z^*)]^T (z_k^\circ - z^*) + \frac{\rho}{2} \|z_k^\circ - z^*\|^2 \quad (61)$$

holds with the optimal solution z^* . – In the right-hand side, we can substitute u^* for $\nabla \phi(z^*)$, according to Observation 2.

Coming back to (60), let us subtract $\bar{u}_k^T z^*$ from both sides, and let us substitute the lower bound of (61) for $\phi(z_k^\circ)$ in the right-hand side. We get

$$\bar{u}_k^T (\bar{z}_k - z^*) - \phi(\bar{z}_k) \leq \bar{u}_k^T (z_k^\circ - z^*) - \phi(z^*) - u^{*T} (z_k^\circ - z^*) - \frac{\rho}{2} \|z_k^\circ - z^*\|^2.$$

Simple transformations yield

$$\frac{\rho}{2} \|z_k^\circ - z^*\|^2 - (\bar{u}_k - u^*)^T (z_k^\circ - z^*) \leq [\phi(\bar{z}_k) - \phi(z^*)] - \bar{u}_k^T (\bar{z}_k - z^*).$$

Applying the Cauchy-Bunyakovsky-Schwarz inequality on both sides, and multiplying by $\frac{2}{\rho}$, we get

$$\|z_k^\circ - z^*\|^2 - \frac{2}{\rho} \|\bar{u}_k - u^*\| \|z_k^\circ - z^*\| \leq \frac{2}{\rho} \left[\phi(\bar{z}_k) - \phi(z^*) \right] + \frac{2}{\rho} \|\bar{u}_k\| \|\bar{z}_k - z^*\|. \quad (62)$$

An elementary transformation will then yield a square on the right-hand side:

$$\begin{aligned} & \left(\|z_k^\circ - z^*\| - \frac{1}{\rho} \|\bar{u}_k - u^*\| \right)^2 \\ & \leq \frac{2}{\rho} \left[\phi(\bar{z}_k) - \phi(z^*) \right] + \frac{2}{\rho} \|\bar{u}_k\| \|\bar{z}_k - z^*\| + \frac{1}{\rho^2} \|\bar{u}_k - u^*\|^2. \end{aligned} \quad (63)$$

We have $\bar{u}_k \rightarrow u^*$ and $\bar{z}_k \rightarrow z^*$ according to Corollary 10. Taking into account the continuity of $\phi(z)$, the right-hand side of (63) converges to zero. Hence $\|z_k^\circ - z^*\| \rightarrow 0$ must hold. $\square$

Remark 16 In the above proof, we applied the lower bound $\nabla^2 \phi(z) \succeq \rho I$ ($z \in \mathbb{R}^n$). However, if we can focus on the locality $\mathcal{B}_r$ of (47), we'll have the stronger bound $\nabla^2 \phi(z) \succeq \alpha^- I$ ($z \in \mathcal{B}_r$).

This yields faster convergence, because in (63), we can substitute α^- for ρ .

A.3 Computing a moderate bound for improving columns

In Sect. 8.1, we need a bounded box $\mathcal{Z}^\lambda$ that contains every possible improving vector z_k^λ that we may obtain by approximately solving the column generating problem in iterations $k \geq K^\lambda$, with some natural number K^λ . (We do not need to know K^λ in Sect. 8.1, only its existence is important.) I'm going to sketch the construction of a workable box. We can apply the inequalities of section A.2.

Since z_k^λ is obtained by a line search starting from $\bar{z}_k$, the analogue of (60) holds:

$$\bar{u}_k^T \bar{z}_k - \phi(\bar{z}_k) \leq \bar{u}_k^T z_k^\lambda - \phi(z_k^\lambda).$$

Applying the transformations of section A.2, we arrive at the following analogue of (63):

$$\begin{aligned} & \left(\|z_k^\lambda - z^*\| - \frac{1}{\rho} \|\bar{u}_k - u^*\| \right)^2 \\ & \leq \frac{2}{\rho} \left[\phi(\bar{z}_k) - \phi(z^*) \right] + \frac{2}{\rho} \|\bar{u}_k\| \|\bar{z}_k - z^*\| + \frac{1}{\rho^2} \|\bar{u}_k - u^*\|^2. \end{aligned} \quad (64)$$

Like section A.2, it follows that $\|z_k^\lambda - z^*\| \rightarrow 0$.

Hence, in theory, any box containing z^* in its interior is a feasible candidate for $\mathcal{Z}^\lambda$.

For a practicable construction, let us consider $k \geq K_r$ of (48). Then we can focus on the locality $\mathcal{B}_r$ of (47), where $\nabla^2\phi(z) \succeq \alpha^- I$ ($z \in \mathcal{B}_r$) holds. In accordance with Remark 16, we can substitute α^- for ρ in (64).

B On the convergence of eigenvalues

The following theorem is due to Ostrowski. (I learned it from the textbook Szidarovszky 1974 that cited it in Chapter 7.4.)

Theorem 17 (from Appendix K in Ostrowski 1960) *Given matrices $A, B \in \mathbb{R}^{n \times n}$ having components a_{ij} and b_{ij} ($1 \leq i, j \leq n$), respectively, let*

$$M = \max_{i,j} \{|a_{ij}|, |b_{ij}|\} \quad \text{and} \quad \delta = \frac{1}{nM} \sum_{i,j} |a_{ij} - b_{ij}|.$$

For any eigenvalue λ of A , there exists an eigenvalue μ of B such that

$$|\lambda - \mu| \leq (n + 2)M\delta^{1/n}.$$

References

- Borgwardt KH (2009) Probabilistic analysis of simplex algorithms. In: Floudas CA, Pardalos PM (eds) Encyclopedia of optimization. Springer
- Cutler L, Wolfe P (1963) Experiments in linear programming. In: Graves R, Wolfe P (eds) Recent advances in mathematical programming. McGraw-Hill, first published as a technical report of the RAND Corporation
- de Oliveira W, Sagastizábal C (2014) Level bundle methods for oracles with on-demand accuracy. Optim Methods Softw 29:1180–1209
- Dempster MAH, Merkovsky RR (1995) A practical geometrically convergent cutting plane algorithm. SIAM J Numer Anal 32:631–644
- Dentcheva D, Martinez G (2013) Regularization methods for optimization problems with probabilistic constraints. Math Program 138:223–251
- Dentcheva D, Prékopa A, Ruszczyński A (2000) Concavity and efficient points of discrete distributions in probabilistic programming. Math Program 89:55–77
- Dentcheva D, Lai B, Ruszczyński A (2004) Dual methods for probabilistic optimization problems. Math Methods Oper Res 60:331–346
- Fábián CI, Csizmás E, Drenyovszki R, van Ackooij W, Vajnai T, Kovács L, Szántai T (2018) Probability maximization by inner approximation. Acta Polytech Hung 15:105–125, special issue dedicated to the memory of András Prékopa (editors: A. Bakó, I. Maros and T. Szántai)
- Fábián CI, Csizmás E, Drenyovszki R, Vajnai T, Kovács L, Szántai T (2019) A randomized method for handling a difficult function in a convex optimization problem, motivated by probabilistic programming. Ann Oper Res. <https://doi.org/10.1007/s10479-019-03143-z>. To appear in S.I.: Stochastic Modeling and Optimization, in memory of András Prékopa (editors: E. Boros, M. Katehakis, A. Ruszczyński). Open access
- Frangioni A (2002) Generalized bundle methods. SIAM J Optim 13:117–156
- Frangioni A (2018) Standard bundle methods: untrusted models and duality. In: Bagirov A, Gaudioso M, Karmitsa N, Mäkelä MM (eds) Numerical nonsmooth optimization. Springer, Berlin, pp 61–116 (Former version: Technical report, Department of Informatics, University of Pisa, Italy)
- Gorn G (1991) Conjugation and second-order properties of convex functions. J Math Anal Appl 158:293–315

- Hantoute A, Henrion R, Pérez-Aros P (2018) Subdifferential characterization of probability functions under Gaussian distribution. *Math Program* 174:167–194
- Luenberger DG, Ye Y (2008) Linear and nonlinear programming. International Series in Operations Research and Management Science. Springer, Berlin
- Ostrowski A (1960) Solution of equations and systems of equations. Academic Press, New York
- Prékopa A (1990) Dual method for a one-stage stochastic programming problem with random RHS obeying a discrete probability distribution. *ZOR Methods Models Oper Res* 34:441–461
- Prékopa A, Vizvári B, Badics T (1998) Programming under probabilistic constraint with discrete random variable. In: Giannesi F, Rapcsák T, Komlósi S (eds) *New trends in mathematical programming*. Kluwer, Dordrecht, pp 235–255
- Rockafellar RT (1970) *Convex analysis*. Princeton University Press, Princeton
- Ruszczynski A (2006) *Nonlinear optimization*. Princeton University Press, Princeton
- Spielman D, Teng SH (2004) Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time. *J ACM* 51:385–463
- Szidarovszky F (1974) *Introduction to numerical methods*. Közgazdasági és Jogi Könyvkiadó, Budapest (in Hungarian)
- van Ackooij W, Berge V, de Oliveira W, Sagastizábal C (2017) Probabilistic optimization via approximate p -efficient points and bundle methods. *Comput Oper Res* 77:177–193

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.