Utilization of the Level Method for Dual Decomposition in Convex Quadratic Programming

K. Beer[1], M. Knobloch[2]

Abstract

In this paper we describe a method, which enables us to solve optimization problems resulting from dual decomposition approaches. We make use of a well-known cutting plane method of level-type. This method allows us, to get rid of the compactness-condition for the resulting inner problems because the described level method can handle infinite function values.

The principal aim of this paper is the description of an appropriate oracle. The case of optimization problems with convex quadratic objective functions and affine-linear constraints is fully exploited and a detailled algorithm is given for the mentioned class of problems. Some numerical tests for problems with exclusively finite function values close the paper. We compare the standard method with a method using normalized subgradients.

MSC: 90C25, 65K05, 90C20

Keywords: level method, decomposition methods, quadratic programming, cutting plane methods, nonsmooth programming

1 Problem Formulation

The goal we aim at throughout the entire paper is to find the optimal value of the quadratic programming problem

$$\begin{cases} \frac{1}{2}\langle x, Qx\rangle + \langle c, x\rangle \to \min\limits_{x} \\ Gx \quad \leq \quad g \\ Hx \quad \leq \quad h \\ x \in \mathbb{R}^n \end{cases} \tag{1}$$

via a dual problem. We suppose the matrix $Q \in \mathbb{R}^{n\times n}$ to be symmetric and positive semidefinite. The matrices G respectively H may be real matrices of type $[m \times n]$ respectively $[p \times n]$. Further on we suppose vectors $c \in \mathbb{R}^n, g \in \mathbb{R}^m$ and $h \in \mathbb{R}^p$ to be given.

[1]Faculty of Mathematics, TU Chemnitz, 09107 Chemnitz, Germany; k.beer@mathematik.tu-chemnitz.de

[2]Faculty of Mathematics, TU Chemnitz, 09107 Chemnitz, Germany; m.knobloch@mathematik.tu-chemnitz.de

Let the value f^*, defined by

$$f^* \;\; := \;\; \inf_{x\in\mathbb{R}^n} \left\{ \tfrac{1}{2}\langle x, Qx\rangle + \langle c, x\rangle : Gx \leq g, Hx \leq h \right\}, \tag{2}$$

denote the optimal value of problem (1). We suppose problem (1) to be solvable in the sense that the value f^* is finite. That means particularly that the feasible set of problem (1) is nonempty and that the objective function is bounded below on this set. We remark, that we do not suppose the set $\{x \in \mathbb{R}^n : Hx \leq h\}$ to be bounded. Since problem (1) is a convex quadratic optimization problem, we can conclude the existence of global optimizers. This can for instance be read in [1] and [4].
Obviously, if we choose Q to be the matrix consisting of zeros, problem (1) is a pure linear programming problem. This case is fully covered by the theory developed in this article and shall not be excluded. But we will make use of some special properties of the solution method for these problems, the simplex method, in Section 4.5 where we treat the case of linear programming seperately.

2 Dual Decomposition Approach

In general we suppose problem (1) to be a large-scale problem. That justifies the decomposition approach. Instead of problem (1) we process its (Lagrange-)dual problem. We call the function $\varphi(\lambda)$ defined by

$$\varphi(\lambda) \;\; := \;\; \inf_{x\in\mathbb{R}^n} \left\{ \tfrac{1}{2}\langle x, Qx\rangle + \langle c, x\rangle + \langle \lambda, Gx - g\rangle : Hx \leq h \right\}, \quad \lambda \in \mathbb{R}^m_+ \tag{3}$$

the dual function with respect to the constraints $Gx \leq g$. The objective function of problem (3) is convex due to the positive semidefiniteness of Q. Moreover, its feasible set is closed but not necessarily bounded. For that reason it is possible that $\varphi(\lambda) = -\infty$ for some $\lambda \in \mathbb{R}^m_+$.
Let $\operatorname{dom}\varphi$ be the effective domain of the function $\varphi(\lambda)$, i.e.,

$$\operatorname{dom}\varphi \;\; := \;\; \{\lambda \in \mathbb{R}^m_+ : \varphi(\lambda) > -\infty\}. \tag{4}$$

Unfortunately, $\operatorname{dom}\varphi$ is not explicitely known in general. However, it is well-known that $\varphi(\lambda)$ is concave and that $\operatorname{dom}\varphi$ is a convex set. The theory of parametric optimization yields furthermore that $\operatorname{dom}\varphi$ is a polyhedral set for the considered class of functions. This fact will be proven later.
Since we assumed f^* to be finite and Q to be positive semidefinite it follows that $\operatorname{dom}\varphi$ is nonempty.
It is now possible to define the dual problem corresponding to (1) as

$$\begin{cases} \varphi(\lambda) \to \max\limits_{\lambda} \\ \lambda \in \mathbb{R}^m_+. \end{cases} \tag{5}$$

If we reformulate (5) as a problem of minimization, we get a problem with convex objective function and convex feasible set. In this formulation it is possible to use the level method for solving (5). We remark that we have to demand that the used method has the ability to handle infinite function values.
Because of the fact that the constraints of (1) are affine-linear functions, the positive semidefiniteness of Q and the solvability of problem (1) itself we can conclude the equality of f^* and φ^* from strong duality where φ^* is the optimal value of problem (5).
It should be remarked that the inner problems for the computation of $\varphi(\lambda)$ have in general to be easily solvable. Especially these problems may be supposed to decompose such that the solution process is parallelizable. Another application is the computation of a lower bound for the primal optimal value in discrete optimization.

3 Algorithm of the Level Method

The algorithm we use to solve (5) is a so called level method, a special cutting plane method, and was first mentioned in [5]. Essentially we use a variant described in [2].
This algorithm generates a sequence $\{\lambda^i, i = 1, 2, \ldots\}$ to find the optimal solution of problems of the form

$$\begin{cases} f(\lambda) \to \inf\limits_{\lambda} \\ \lambda \in \Lambda \end{cases}$$

with proper, convex, Lipschitzean f and convex, bounded G. It is explicitely allowed that $f(\lambda) = +\infty$ for certain $\lambda \in G$. Additionally we have to suppose $\operatorname{int}(\Lambda \cap \operatorname{dom} f)$ where $\operatorname{dom} f$ is the effective domain of function f.
The level method demands a so-called oracle to be given. The oracle can be imagined as a black box which is capable to produce desired output (subgradients, function values, seperating hyperplanes) if it is provided with special input data (current iterate).
Let us now describe the steps of the level method.

Algorithm 1 (Level Method)

(0.) Choose precision $\varepsilon > 0$, starting point $\lambda \in \Lambda$ and level paramater $\Theta \in (0, 1)$. We set $f_0^* = \infty$ and start with $k = 1$.

(1.) Provide the oracle with λ^k and let the oracle compute a vector b_k and a number β_k such that the following conditions hold:

- *if $\lambda^k \in dom f$ then b_k is a subgradient of f at λ^k and $\beta_k = f(\lambda^k)$,*
- *if $\lambda^k \notin dom f$ then for b_k and β_k holds*

$$\langle b_k, \lambda - \lambda^k \rangle + \beta_k \quad \leq 0 \;\; \forall \lambda \in dom f. \tag{6}$$

(2.) If $\lambda^k \in dom\, f$ then set $f_k^* = \min\{f_{k-1}^*, f(\lambda^k)\}$ and update the values β_i^k using the following rule

$$\beta_i^k \quad = \quad \begin{cases} f(\lambda^i) - f_k^* & \text{if } \lambda^i \in dom\, f, \quad i = 1, 2, \ldots, k \\ \beta_i^{k-1} & \text{if } \lambda^i \notin dom\, f, \quad i = 1, 2, \ldots, k-1 \end{cases} . \tag{7}$$

Otherwise $f_k^* = f_{k-1}^*$ and $\beta_k^k = \beta_k$.

(3.) Compute Δ_k as the optimal value of the problem

$$\begin{cases} t \to \max \\ \langle b_i, \lambda - \lambda^i \rangle + \beta_i^k + \|b_i\| t \leq 0 \quad i = 1, 2, \ldots, k \\ \lambda \in \Lambda. \end{cases} \tag{8}$$

If $f_k^* < \infty$ and $\Delta_k < 10 \cdot \varepsilon$ additionally compute Δ_k' as the optimal value of

$$\begin{cases} t \to \max \\ \langle b_i, \lambda - \lambda^i \rangle + \beta_i^k + t \leq 0 & \forall i : \lambda^i \in dom\, f \\ \langle b_i, \lambda - \lambda^i \rangle + \beta_i^k \leq 0 & \forall i : \lambda^i \notin dom\, f \\ \lambda \in \Lambda. \end{cases} \tag{9}$$

(4.) If $\Delta_k' < \varepsilon$ then **STOP**. *Otherwise use the minimizer of problem*

$$\begin{cases} \|\lambda - \lambda^k\|^2 \to \min \\ \langle b_i, \lambda - \lambda^i \rangle + \beta_i^k + \Theta \cdot \|b_i\| \cdot \Delta_k \leq 0 \quad i = 1, 2, \ldots, k \\ \lambda \in \Lambda \end{cases} \tag{10}$$

as new iterate λ^{k+1}.

(5.) Set $k = k + 1$ and return to step (1.).

Remarks

- The following fact can be easily proven: If $\varepsilon > 0$ then the method stops after a finite number k_0 of iterations and we have $0 \leq f_{k_0}^* - f^* \leq \varepsilon$. The proof for linear problems of primal decomposition can be found in [3] and it can be generalized.
- The described algorithm already contains certain modifications.
  - In problems (8) and (10) we use normalized subgradients. If we substitute the coefficient $\|b_i\|$ by 1 the subgradients are unnormalized. It is not a priori clear which one of the methods yields the better performance. Comparisons can be found in Section 6.
  - The described method exploits the strategy of so-called "Deeper Cuts". If we set $\beta_i^k = 0$ $\forall i : \lambda^i \in \mathrm{dom}\, f$ we have the standard method. Both strategies will be compared in Section 6, too.

4 Realization of the Oracle

To use Algorithm 1 we need an oracle which must be able

- to calculate function values, especially to recognize whether $\lambda^k \in \operatorname{dom}\varphi$ or not,
- to compute subgradients of the objective function
- to construct domain-cuts.

In the next subsections we will describe how these three tasks may be realized.

4.1 Recognizing Infinite Function Values

Obviously, it is numerically difficult, to determine whether a current iterate λ^k has finite function value or not and the algorithmic definition of $\varphi(\lambda)$ makes it even more difficult. The aim of this section is therefore the description of a way that makes it possible to decide the mentioned question.

We remember that for computing $\varphi(\lambda)$ we have to solve problem (3). Now it is possible once again to consider a dual problem. The dual problem corresponding to problem (3) is

$$\begin{cases} \psi_\lambda(\mu) \to \sup\limits_{\mu} \\ \mu \in \mathbb{R}^p_+ \end{cases} \tag{11}$$

where the function $\psi_\lambda(\mu)$ is defined by

$$\psi_\lambda(\mu) \quad := \quad \inf_{x\in\mathbb{R}^n} \{\tfrac{1}{2}\langle x, Qx\rangle + \langle c, x\rangle + \langle \lambda, Gx - g\rangle + \langle \mu, Hx - h\rangle\}. \tag{12}$$

Let ψ^*_λ denote the optimal value of (11). We remark that problem (11) as well as the function ψ_λ depend on the parameter λ. By evaluating the first order necessary and sufficient conditions for (12) we get the following auxiliary problem

$$\begin{cases} \langle \mathbb{1}, z + z'\rangle \to \min\limits_{x,\mu,z,z'} \\ Qx + H^\top\mu + z - z' = -(c + G^\top\lambda) \\ x \in \mathbb{R}^n,\ \mu \in \mathbb{R}^p_+,\ z, z' \in \mathbb{R}^n_+ \end{cases} \tag{13}$$

where $\mathbb{1}$ denotes the vector $(1, 1, \dots, 1)^\top$. Once again it should be remarked that (13) is a parametric linear programming problem depending on the parameter λ. Obviously, for each $\lambda \in \mathbb{R}^m_+$ the optimal value of (13) is bounded below by 0. We define $\eta(\lambda)$ as the optimal value function of problem (13).

Determining whether $\lambda \in \operatorname{dom}\varphi$ or not as well as the construction of domain-cuts is closely related to computing $\eta(\lambda)$. With the help of the hereafter stated theorem it will become possible to decide if an actual iterate λ^k is in $\operatorname{dom}\varphi$.

Theorem 1 *Considering problem (5) for arbitrary $\lambda^0 \in \mathbb{R}^m_+$ it holds*

$$\lambda^0 \in dom\,\varphi \iff \eta(\lambda^0) = 0.$$

<u>*Proof.*</u>
Let us first prove sufficiency. Let $\lambda^0 \in \mathrm{dom}\,\varphi$. Immediately, we get $\varphi(\lambda^0) > -\infty$. Because of the linearity of the feasible set of problem (3) and the positive semidefiniteness of Q strong duality holds and moreover we can conclude the existence of $\mu^0 \in \mathbb{R}^p_+$ such that

$$\varphi(\lambda^0) \;=\; \psi_{\lambda^0}(\mu^0).$$

Hence, there exists a global optimizer $x^0 \in \mathbb{R}^n$ such that

$$\varphi(\lambda^0) \;=\; \tfrac{1}{2}\langle x^0, Qx^0\rangle + \langle c, x^0\rangle + \langle \lambda^0, Gx^0 - g\rangle + \langle \mu^0, Hx^0 - h\rangle.$$

Since (12) is convex the first-order necessarry and sufficient condition holds at x^0, i.e.

$$Qx^0 + c + H^\top\mu^0 + G^\top\lambda^0 \;=\; \mathbb{O}.$$

Therefore, the vector $[x^0, \mu^0, \mathbb{O}, \mathbb{O}]$ is feasible in problem (13) and yields function value 0 and hence it is a global optimal solution of (13).
Let us now prove necessity. Let therefore be $\eta(\lambda^0) = 0$. Since (13) is a linear program we know there exists a global optimizer $[x^1, \mu^1, z^1, z^{1'}]$. Moreover, since the optimal value is equal to zero we know that $z^1 = z^{1'} = \mathbb{O}$. Hence

$$Qx^1 + c + H^\top\mu^1 + G^\top\lambda^0 \;=\; \mathbb{O}$$

holds. This condition is the necessary and sufficient condition for a global minimum in problem (12). Hence

$$\psi_{\lambda^0}(\mu^1) \;=\; \tfrac{1}{2}\langle x^1, Qx^1\rangle + \langle c, x^1\rangle + \langle \lambda^0, Gx^1 - g\rangle + \langle \mu^1, Hx^1 - h\rangle \;>\; -\infty$$

holds and by weak duality we can conclude that $\varphi(\lambda^0) \geq \psi_{\lambda^0}(\mu^1) > -\infty$ which means $\lambda^0 \in \mathrm{dom}\,\varphi$. ■

Having obtained the above stated result we can easily conclude the following statement.

Corollary 1 *Considering problem (5) for $\lambda^0 \in \mathbb{R}^m_+$ it holds*

$$\lambda^0 \notin dom\,\varphi \iff \eta(\lambda^0) > 0.$$

If the oracle is told an actual iterate λ^k then the first task is to solve problem (13). The resulting optimal value gives a hint how to continue.

4.2 Supergradients of the Dual Function

The actual section deals with a well-known result from duality theory. For the sake of completeness it shall be noted here.

Theorem 2 (Supergradients of the dual function)
Let $\lambda^0 \in dom\,\varphi$ and let $x^0 \in \mathbb{R}^n$ be a global optimal solution of (3) for $\lambda = \lambda^0$, i.e.

$$\begin{cases} \varphi(\lambda^0) = \frac{1}{2}\langle x^0, Qx^0\rangle + \langle c, x^0\rangle + \langle \lambda^0, Gx^0 - g\rangle \\ Hx^0 \leq h. \end{cases}$$

Then the vector $Gx^0 - g$ is a supergradient of $\varphi(\lambda)$ at λ^0.

Obviously, if z is a supergradient of $\varphi(\lambda)$ at λ^0 then $-a$ is a subgradient of the function $-\varphi(\lambda)$ at λ^0. In the linear-quadratic problem (3) the optimum is always attained if finite. Therefore in the case when $\lambda^k \in \operatorname{dom}\varphi$ we can compute a subgradient of $-\varphi$.

4.3 Properties of the Set dom φ

Although not explicitely needed to describe the algorithm of the oracle it could be helpful to explore something more about the nature of $\operatorname{dom}\varphi$. With the help of linear parametric programming we can prove the following statement.

Proposition 1 *The set $dom\,\varphi$ is a convex, polyhedral subset of $\mathbb{R}^m_+$.*

Proof.
From Theorem 1 we know that $\operatorname{dom}\varphi$ is the set of all parameters $\lambda \in \mathbb{R}^m_+$ such that the optimal value of the problem

$$\begin{cases} \langle \mathbb{1}, z + z'\rangle \to \min\limits_{x,\mu,z,z'} \\ Qx + H^\top\mu + z - z' = -(c + G^\top\lambda) \\ x \in \mathbb{R}^n,\ \mu \in \mathbb{R}^p_+,\ z, z' \in \mathbb{R}^n_+. \end{cases}$$

is equal to zero. This means nothing else than that the linear parametric system

$$\begin{cases} Qx + H^\top\mu = -(c + G^\top\lambda) \\ x \in \mathbb{R}^n,\ \mu \in \mathbb{R}^p_+ \end{cases} \tag{14}$$

is solvable if and only if $\lambda \in \operatorname{dom}\varphi$. We now consider the multifunction mapping each parameter $\lambda \in \mathbb{R}^m_+$ to the solution set of (14). Obviously, $\operatorname{dom}\varphi$ is the feasible parameter set of this multifunction, i.e. the set of all parameters with nonempty images. Applying Proposition 7.1 from [6] yields the desired statement. ■

The variant of the level method we use demands $\Lambda \cap \operatorname{dom}\varphi$ to be nonempty. To ensure convergence of the method the set $\operatorname{dom}\varphi$ must have a nonempty interior. Therefore it would be advantageous if it could be proven that the interior of $\operatorname{dom}\varphi$ is always nonempty. But the next example will show that it is possible that $\operatorname{int}(\operatorname{dom}\varphi) = \emptyset$.

Example 1 *Let us choose matrices Q, G, H and a vector c as follows*

$$Q = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad G = \begin{pmatrix} -1 & 0 & -1 \\ 0 & -2 & -1 \\ -3 & 2 & -1 \end{pmatrix} \quad H = \begin{pmatrix} -4 & 7 & 0 \\ -2 & 3 & 0 \\ 1 & -1 & 0 \end{pmatrix} \quad c = \begin{pmatrix} -2 \\ -1 \\ 1 \end{pmatrix} \tag{15}$$

and set $g := h := \mathbb{O}$. The optimal solution of problem (5) corresponding to this problem data is $\lambda^* = (1, 0, 0)^\top$ which yields the optimal value $\varphi^* = 0$.
We now remember that $\operatorname{dom}\varphi$ is the feasible parameter set of the linear parametric system of equations (14). If we evaluate (14) for the given example we get the following system of equations and inequalities.

$$\begin{array}{rcrcrcrcrcrcrcr} 2x_1 & & & - & 4\mu_1 & - & 2\mu_2 & + & \mu_3 & = & 2 & + & \lambda_1 & & & + & 3\lambda_3 \\ & & 2x_2 & + & 7\mu_1 & + & 3\mu_2 & - & \mu_3 & = & 1 & & & + & 2\lambda_2 & - & 2\lambda_3 \\ & & & & & & & & 0 & = & -1 & + & \lambda_1 & + & \lambda_2 & + & \lambda_3 \\ & & & & & & & & \mu_1, \mu_2, \mu_3 & \geq & 0 & & & & & & \end{array} \tag{16}$$

Since $x \in \mathbb{R}^3$ can be chosen without any restrictions the first two equations in (16) are solvable for arbitrary right-hand-sides and therefore the set $\operatorname{dom}\varphi$ is only determined by the third equation. It follows that $\operatorname{dom}\varphi$ is the set of all $\lambda \in \mathbb{R}^3_+$ which fulfil the equation

$$1 \;=\; \lambda_1 \;+\; \lambda_2 \;+\; \lambda_3. \tag{17}$$

Obviously $\operatorname{dom}\varphi$ has no interior points since it is the intersection of $\mathbb{R}^3_+$ with an affine hyperplane.
Moreover, we see that $\operatorname{dom}\varphi$ is compact in our example because all components are bounded above by 1. Setting the third component c_3 of the vector c to 0 would cause further degeneration. In this case $\operatorname{dom}\varphi$ would only consist of the origin. Setting c_3 to negative values would cause $\operatorname{dom}\varphi$ to be empty.

Example 1 shows that it is possible that $\operatorname{int}(\operatorname{dom}\varphi) = \emptyset$. The question arises if there are examples where $\operatorname{dom}\varphi$ has interior points but does not coincide with the entire $\mathbb{R}^m_+$.

Example 2 *We now choose matrices Q and G and the vectors c, h, g exactly as in (15) and define a slightly modified matrix H as*

$$H \;=\; \begin{pmatrix} 4 & 7 & 0 \\ -2 & 3 & 0 \\ 1 & -1 & -1 \end{pmatrix}. \tag{18}$$

The optimal solution of problem (5) corresponding to this problem data is now $\lambda^* = \mathbb{O}$ which yields the new optimal value $\varphi^* \approx -0.8654$.
Once again we use the fact that $\operatorname{dom}\varphi$ is the feasible parameter set of system (14). In our example that system is similar to system (16) with the difference that the third equation changes to

$$-\mu_3 \;=\; -1 \;+\; \lambda_1 \;+\; \lambda_2 \;+\; \lambda_3. \tag{19}$$

Considering that μ_3 has to be nonnegative it follows that λ belongs to $\operatorname{dom}\varphi$ if and only if it fulfils the inequality

$$1 \;\geq\; \lambda_1 \;+\; \lambda_2 \;+\; \lambda_3. \tag{20}$$

For this example $\operatorname{dom}\varphi$ obviously has interior points and on the other hand $\operatorname{dom}\varphi$ does not coincide with the entire $\mathbb{R}^3_+$.

Obviously, slight changes in the problem data can cause extreme changes of the set $\operatorname{dom}\varphi$. It is therefore complicated to formulate useful criteria for the interior of $\operatorname{dom}\varphi$ to be nonempty. Later publications may contain such criteria and a method that does not explicitely demand $\operatorname{int}(\operatorname{dom}\varphi)$ to be nonempty.

4.4 How to construct Domain-Cuts

In this section we will describe a possible way to construct domain-cuts. The most interesting question arising in connection with the previous consideration is how to make use of the already known, i.e. how is it possible to utilize the solution of problem (13) for constructing domain-cuts. Let us therefore return to the exploration of $\eta(\lambda)$. The most important property of $\eta(\lambda)$ needed for the construction of a domain-cut is described in the following theorem.

Theorem 3 (Convexity of $\boldsymbol{\eta(\lambda)}$)
The optimal value function $\eta(\lambda)$ of problem (13) is convex on $\mathbb{R}^m_+$.

<u>*Proof.*</u>
For fixed $\lambda^0 \in \mathbb{R}^m_+$ consider problem (13). The dual problem corresponding to (13) is

$$\begin{cases} \langle c + G^\top \lambda^0, \kappa\rangle \to \max\limits_{\kappa\in\mathbb{R}^n} \\ Q\kappa = \mathbb{O} \\ H\kappa \geq \mathbb{O} \\ -\mathbb{1} \leq \kappa \leq \mathbb{1}. \end{cases} \tag{21}$$

Let Ω denote the feasible set of problem (21). Obviously, Ω is bounded and nonempty since $\mathbb{O} \in \Omega$ and (13) is a linear program. We can therefore conclude strong duality, i.e.

$$\eta(\lambda^0) \;=\; \max_{\kappa\in\Omega} \big\{ \langle c + G^\top \lambda^0, \kappa\rangle \big\}. \tag{22}$$

In this formulation it is easy to check the convexity of $\eta(\lambda)$. Let $\lambda^1, \lambda^2 \in \mathbb{R}^m_+$ and $t \in [0,1]$.

$$\begin{aligned} \eta(t\lambda^1 + (1-t)\lambda^2) &= \max_{\kappa\in\Omega} \big\{ \langle c + G^\top (t\lambda^1 + (1-t)\lambda^2), \kappa\rangle \big\} \\ &= \max_{\kappa\in\Omega} \big\{ t\langle c + G^\top \lambda^1, \kappa\rangle + (1-t)\langle c + G^\top \lambda^2, \kappa\rangle \big\} \\ &\leq t \cdot \max_{\kappa\in\Omega} \big\{ \langle c + G^\top \lambda^1, \kappa\rangle \big\} + (1-t) \cdot \max_{\kappa\in\Omega} \big\{ \langle c + G^\top \lambda^2, \kappa\rangle \big\} \\ &= t \cdot \eta(\lambda^1) + (1-t) \cdot \eta(\lambda^2) \end{aligned}$$

$\blacksquare$

Besides convexity of $\eta(\lambda)$ Theorems 1 and 3 yield another result.

Corollary 2 *If the matrix Q is positive definite then $dom\, \varphi = \mathbb{R}^m_+$.*

Proof.
Consider problem (21) and its feasible set Ω. A vector κ^0 can only be feasible if

$$Q\kappa^0 \quad = \quad \mathbb{O} \tag{23}$$

holds. Since Q is positive definite $\kappa^0 = \mathbb{O}$ is the only vector satisfying (23). Therefore, $\eta(\lambda) = \langle \mathbb{O}, c + G^\top \lambda \rangle = 0$ on $\mathbb{R}^m_+$. $\blacksquare$

Since we have proven the convexity of $\eta(\lambda)$ it is in general possible to calculate subgradients of $\eta(\lambda)$.

Proposition 2 *Consider problem (13) and its optimal value function $\eta(\lambda)$ and let $\lambda^0 \notin dom\, \varphi$. Let the vector $s \in \mathbb{R}^n$ be a subgradient of $\eta(\lambda)$ at λ^0. Then with the vector s and the number $\eta(\lambda^0)$ a domain-cut can be formed, i.e.*

$$\langle s, \lambda - \lambda^0 \rangle + \eta(\lambda^0) \quad \leq \quad 0 \;\; \forall \lambda \in dom\, \varphi \tag{24}$$

$$\eta(\lambda^0) \quad > \quad 0. \tag{25}$$

Proof.
Let $\lambda^0 \notin \mathrm{dom}\, \varphi$. Hence (25) holds by Theorem 1. Moreover, we know that s is a subgradient of $\eta(\lambda)$ at λ^0. Therefore, the following holds:

$$\eta(\lambda) - \eta(\lambda^0) \quad \geq \quad \langle s, \lambda - \lambda^0 \rangle \;\; \forall \lambda \in \mathbb{R}^m_+. \tag{26}$$

In particular this inequality holds for all $\lambda \in \mathrm{dom}\, \varphi$. Again from Theorem 1 we know that $\eta(\lambda) = 0 \; \forall \lambda \in \mathrm{dom}\, \varphi$ which yields the validity of (25). $\blacksquare$

Proposition 2 describes a possibility to construct a domain-cut. With the knowledge from Theorem 3 we know that in general computing a subgradient of $\eta(\lambda)$ is possible. The next theorem shows how to obtain the searched subgradient.

Theorem 4 (Subgradient of $\boldsymbol{\eta(\lambda)}$)
Let $\lambda^0 \notin dom\, \varphi$ and let κ^* be an optimal solution of problem (21) for λ^0. Then the vector $G\kappa^*$ is a subgradient of the function $\eta(\lambda)$ at λ^0.

Proof.
Since κ^* is an optimal solution of (21) it holds

$$\eta(\lambda^0) \quad = \quad \langle \kappa^*, c + G^\top \lambda^0 \rangle. \tag{27}$$

Moreover, $\eta(\lambda^0)$ is positive by Theorem 1. Since Ω does not depend on the parameter λ the vector κ^* is feasible for all λ. For arbitrary $\lambda \in \mathbb{R}^m_+$ we get

$$\begin{aligned} \eta(\lambda) - \eta(\lambda^0) &= \max_{\kappa \in \Omega} \left\{ \langle c + G^\top \lambda, \kappa \rangle \right\} - \langle \kappa^*, c + G^\top \lambda^0 \rangle \\ &\geq \langle \kappa^*, c + G^\top \lambda \rangle - \langle \kappa^*, c + G^\top \lambda^0 \rangle \ \ \forall \lambda \in \mathbb{R}^m_+ \\ &= \langle \kappa^*, G^\top (\lambda - \lambda^0) \rangle \ \ \forall \lambda \in \mathbb{R}^m_+ \\ &= \langle G\kappa^*, \lambda - \lambda^0 \rangle \ \ \forall \lambda \in \mathbb{R}^m_+ \end{aligned}$$

and this means $G\kappa^*$ is a subgradient of $\eta(\lambda)$ at λ^0. ■

Remarks

- If a current iterate λ^k is not in $\operatorname{dom} \varphi$ we have to solve (21) which is the dual corresponding to (13) to construct a domain-cut. Since strong duality holds with Theorem 1 and (22) we are able to decide if $\lambda^k \in \operatorname{dom} \varphi$ with the help of the optimal value of (21).
- We demand that the vector defining the domain-cut is normed. Therefore the oracle has to deliver vector $\frac{G\kappa^*}{\|G\kappa^*\|}$ and number $\frac{\eta(\lambda^k)}{\|G\kappa^*\|}$.

4.5 Linear Programming Problems

The theory developed in the previous sections is, of course, applicable for pure linear programming problems, too. But for these problems it is possible to obtain the data needed for the domain-cut with much less expense.
The oracle is provided with a current iterate λ^k in the k-th iteration. The oracles first task is to solve the linear programm

$$\begin{cases} \langle c, x \rangle + \langle \lambda^k, Gx - g \rangle \to \min\limits_x \\ Hx \leq h. \end{cases} \tag{28}$$

Suppose $\lambda^k \notin \operatorname{dom} \varphi$. We solve (28) using a simplex method. It is well known that after a finite number of iterations the simplex method recognizes that the optimal value of (28) is $-\infty$. Moreover, it provides a direction $\hat{x}$ which has the following two properties.

- The direction $\hat{x}$ is in the recession cone of the feasible set of (28), i.e. for arbitrary feasible $\bar{x}$ and for all $\alpha > 0$ it holds

$$H(\bar{x} + \alpha \hat{x}) \quad \leq \quad h. \tag{29}$$

- The direction $\hat{x}$ is a direction of recession of the objective function of (28), i.e. for arbitrary feasible $\bar{x}$ it holds

$$\lim_{\alpha \to \infty} \left\{ \langle c, \bar{x} + \alpha \hat{x} \rangle + \langle \lambda^k, G(\bar{x} + \alpha \hat{x}) - g \rangle \right\} \quad = \quad -\infty. \tag{30}$$

The direction $\hat{x}$ can be used to construct a domain-cut without having to solve (21).

Proposition 3 *Let $\lambda^0 \notin dom\,\varphi$ and let $\hat{x}$ be a vector fulfilling (29) and (30) for $\lambda^k = \lambda^0$. Then the following two inequalities hold:*

$$\begin{aligned} \langle -G\hat{x}, \lambda - \lambda^0 \rangle - (\langle c, \hat{x} \rangle + \langle \lambda^0, G\hat{x} \rangle) &\leq 0 \;\; \forall \lambda \in dom\,\varphi & (31) \\ -(\langle c, \hat{x} \rangle + \langle \lambda^0, G\hat{x} \rangle) &> 0. & (32) \end{aligned}$$

<u>*Proof.*</u>
Let $\bar{x}$ be feasible in problem (28). From (30) for sufficiently large α we immediately get:

$$\begin{aligned} \langle c, \bar{x} + \alpha\hat{x} \rangle + \langle \lambda^0, G(\bar{x} + \alpha\hat{x}) - g \rangle &= \alpha(\langle c, \hat{x} \rangle + \langle \lambda^0, G\hat{x} \rangle) + \langle c, \bar{x} \rangle + \langle \lambda^0, G\bar{x} - g \rangle \\ &< \langle c, \bar{x} \rangle + \langle \lambda^0, G\bar{x} - g \rangle \end{aligned}$$

which proves (32).
Let $\lambda^1 \in \mathrm{dom}\,\varphi$ which means $\varphi(\lambda^1) > -\infty$. From (29) we know that the entire ray $\bar{x} + \alpha\hat{x}$ is feasible in (28). For arbitrary $\alpha > 0$ we get:

$$\begin{aligned} &\Rightarrow \langle c, \bar{x} + \alpha\hat{x} \rangle + \langle \lambda^1, G(\bar{x} + \alpha\hat{x}) - g \rangle \geq \varphi(\lambda^1) > -\infty \\ &\Rightarrow \alpha(\langle c, \hat{x} \rangle + \langle \lambda^1, G\hat{x} \rangle) + \langle c, \bar{x} \rangle + \langle \lambda^1, G\bar{x} - g \rangle > -\infty \end{aligned}$$

Since α can be made arbitrary large we get

$$\langle c, \hat{x} \rangle + \langle \lambda^1, G\hat{x} \rangle \geq 0. \tag{33}$$

The substitution $\lambda = \lambda^1$ in the left-hand-side of (31) yields

$$\langle -G\hat{x}, \lambda^1 - \lambda^0 \rangle - (\langle c, \hat{x} \rangle + \langle G\hat{x}, \lambda^0 \rangle) = -(\langle G\hat{x}, \lambda^1 \rangle + \langle c, \hat{x} \rangle) \overset{(33)}{\leq} 0.$$

Since this inequality holds for all $\lambda^1 \in \mathrm{dom}\,\varphi$ the proof is finished. ■

5 Description of the Algorithm

With the knowledge of the previous sections it is possible now to decsribe the detailled algorithm appropriate for solving (5). The level method itself demands that Λ is compact and convex[3]. Therefore we have to choose a convex, compact subset of $\mathbb{R}^m_+$ to apply the level method for solving problem (5). We can do this by choosing sufficiently large upper bounds U and defining the set Λ as $\{\lambda \in \mathbb{R}^m_+ : \lambda \leq U \cdot \mathbb{1}\}$. Problem (1) is a convex quadratic optimization problem with affine-linear constraints and we supposed it to be solvable. Therefore the existence of optimal solutions in the dual problem (5) is guaranteed and this means that it exists a finite number U such that all components of at least one optimal solution of problem (5) are smaller than U.

[3]see Section 3

Remember that we have to define a new objective function

$$f(\lambda) \quad := \quad -\varphi(\lambda) \ \ \forall \lambda \in \mathbb{R}^m_+$$

since our algorithm is of minimization type. Furthermore, the interior of $\Lambda \cap \operatorname{dom} \varphi$ must be supposed to be nonempty.

Algorithm 2 (Level Method for Dual Decomposition)
We choose starting point $\lambda^1 \in \mathbb{R}^m_+$, upper bound $U > 0$, level parameter $\Theta \in (0,1)$ and precision $\varepsilon > 0$. We start with $f_0^* = \infty$ and $k = 1$.

A) Compute $\eta(\lambda^k)$ as the optimal value and κ_k^* as the optimal solution of the problem

$$\begin{cases} \langle c + G^\top \lambda^k, \kappa \rangle \to \max \\ Q\kappa = \mathbb{O} \\ H\kappa \geq \mathbb{O} \\ -\mathbb{1} \leq \kappa \leq \mathbb{1}. \end{cases} \tag{34}$$

If $\eta(\lambda^k) > 0$ then set $\beta_k^k = \frac{\eta(\lambda^k)}{\|G\kappa_k^*\|}$ and $b_k = \frac{G\kappa_k^*}{\|G\kappa_k^*\|}$ and go to step C). Otherwise continue with step B).

B) Compute $\varphi(\lambda^k)$ as the optimal value and x_k^* as a global minimizer of

$$\begin{cases} \frac{1}{2}\langle x, Qx \rangle + \langle c, x \rangle + \langle \lambda^k, Gx - g \rangle \to \min\limits_{x \in \mathbb{R}^n} \\ Hx \leq h. \end{cases} \tag{35}$$

Set $f(\lambda^k) = -\varphi(\lambda^k)$ and update the function value record by $f_k^* = \min\{f_{k-1}^*, f(\lambda^k)\}$. Set $b_k = -Gx_k^* + g$.

C) Update the values $\beta_i^k, i = 1, 2, \ldots, k$ using formula (7)

D) Compute the optimal value Δ_k of problem (8). If $f_k^* < \infty$ and $\Delta_k < 10 \cdot \varepsilon$ additionally compute the optimal value Δ_k' of problem (9). If $\Delta_k' < \varepsilon$ then **STOP**. *Otherwise select a new iterate as the global minimizer of problem (10).*

E) Set $k = k + 1$ and return to step A).

Remark

- In step B) the function value $\varphi(\lambda^k) = -f(\lambda^k)$ has to be computed. Some algorithms for quadratic programming do not reliably recognize the case $\lambda^k \notin \operatorname{dom} \varphi$. To avoid doubts regarding the finiteness of $\varphi(\lambda^k)$ we solve (34) in each iteration.

- For practical use of Algorithm 2 we have to demand that the method we use to solve problem (34) is capable to compute the exact optimal value. If $\eta(\lambda^k)$ can only be computed with a precision $\delta > 0$ then no decision will be possible if $0 < \eta(\lambda^k) < \delta$.

6 Numerical Experiments

We tested Algorithm 2 with several modifications. Although we have described a method which is capable to handle infinite function values we have so far generated only problems where the set $\operatorname{dom} \varphi$ coincides with the entire $\mathbb{R}^m_+$. The problem arising together with the presence of infinite function values is the necessity to solve (34) exactly. At the current stage of implementating Algorithm 2 this is impossible since the used LP-solver is a primal-dual type interior-point algorithm. Further work has to be done and numerical results for problems with possible infinite function values will be published later.
We have therefore focused our experiments on a comparison of the performance of the method with normalized subgradients and the method with unnormalized subgradients.

6.1 Problem Generation

For testing the algorithm we solved several problems of type (1) using the dual decomposition approach described in this paper. The problems were generated using the following rules:

- We chose the dimension m of the decision variables in the dual problems from the set $\{10, 20, 30, 40, 50\}$. We set $n = 2m$ and $p = m$.
- The matrices G and H as well as the vectors c, g and h were chosen by random completely. The matrix Q was chosen as a diagonal matrix with randomly generated positive diagonal elements such that the positive definiteness of Q was ensured.
- The upper bound U was set to $1 + \tilde{U}$ where $\tilde{U}$ was the largest component of the (already known) optimal solution of problem (5).
- All generated random numbers were chosen from the interval $(-10, 10)$ except the diagonal elements of Q which were chosen from $(0, 10)$.
- The level parameter Θ was set to $\frac{1}{4}$. We chose precision $\varepsilon = 10^{-3}$.
- For each combination of m and the selected modification we solved 200 test problems. Each problem was solved using normalized subgradients in problems (8) and (10) and then the same problem was solved once again without normalizing the subgradients. The numbers in the tables are the numbers of iterations, needed to achieve the desired accuracy, averaged over the 200 problems.
- The algorithms were implemented in MATLAB version 6.1. All the resulting linear and quadratic programming problems were solved using algorithms from the MOSEK OPTIMIZATION TOOLBOX version 2.0.

6.2 Comparison of the Algorithms

We first implemented the standard method without using deeper cuts. The solution process was terminated when the absolute accuracy of ε was achieved. That was tested by solving the problems (9) at the last stage of the solution process exactly as described in Algorithm 2. Table 1 shows the average numbers of iterations.

m	unnormalized	normalized
10	51.10	51.37
20	66.29	65.40
30	75.05	74.69
40	85.56	85.34
50	93.77	95.36

Table 1: Performance without Deeper Cuts

Obviously, using normalized subgradients does not yield significantly improved performance as espected. The dependence between the average number of iterations and the problem dimension m seems to be weak.
The next set of problems were solved using deeper cuts. The results for average numbers of iteration are noted in table 2.

m	unnormalized	normalized
10	28.15	28.19
20	38.05	38.36
30	46.01	46.46
40	53.93	54.04
50	58.98	60.26

Table 2: Performance with Deeper Cuts

It can be easily seen that the level method performs significantly better when we use deeper cuts. But once again there is no significant difference between the algorithm with normalized subgradients and the algorithm with unnormalized subgradients.
To check if at least in the first stage of the iteration process the variant with normalized subgradients shows improved performance and to make our results comparable to earlier results we solved another set of test problems with one difference. Instead of terminating the solution process using absolute accuracy $\varepsilon > 0$ we now stopped the algorithm when the relative accuracy goal

$$\frac{|\varphi(\lambda^k) - \varphi^*|}{1 + |\varphi^*|} \quad \leq \quad \varepsilon \tag{36}$$

was reached where φ^* is the optimal value of problem (5). Results can be seen in table 3.

m	unnormalized	normalized
10	17.43	17.59
20	19.76	19.63
30	21.44	21.57
40	23.16	22.88
50	24.30	24.23
200	32.70	32.08

Table 3: Performance using stopping criterion (36)

Obviously, the relative accuracy stopping criterion can be fulfilled in average after less iterations compared to the average number of iterations needed to fulfil the absolute accuracy stopping criterion. Dependence between the needed number of iterations and problem dimension is very weak again. The set of problems of dimension 200 has been solved to emphasize this fact. But once again using normalized subgradients did not improve performance.

References

[1] A. Auslender: "Existence of optimal solutions and duality results under weak conditions", Mathematical Programming 88 (2001), pp. 45-59

[2] K. Beer, E.G. Gol'stejn, N.A. Sokolov: " Minimization of a nondifferentiable, convex function, defined not everywhere", Preprint 98–9, Faculty of Mathematics, University of Technology, Chemnitz (to appear in "Optimization")

[3] K. Beer, E.G. Gol'stejn, N.A. Sokolov: "Utilization of the level-method for primal decomposition in linear programming problems", Preprint 2000–13, Faculty of Mathematics, University of Technology, Chemnitz

[4] E.G. Belousov, "Introduction to Convex Analysis and Integer Programming" (Russian), Izdatel'stvo Moskovskogo Universiteta, Moskau, 1977

[5] C. Lémarechal, A. Nemirovskii, Y. Nesterov: "New Variants of Bundle Methods", Mathematical Programming 69 (1995), pp. 111-147

[6] F. Nožička, J. Guddat, H. Hollatz, B. Bank: "Theorie der linearen parametrischen Optimierung ", Akademie-Verlag Berlin, 1974