Solving Convex Programs via Lagrangian Decomposition

M. Knobloch[1]

Abstract

We consider general convex large-scale optimization problems in finite dimensions. Under usual assumptions concerning the structure of the constraint functions, the considered problems are suitable for decomposition approaches. Lagrangian-dual problems are formulated and solved by applying a well-known cutting-plane method of level-type. The proposed method is capable to handle infinite function values. Therefore it is no longer necessary to demand the feasible set with respect to the non-dualized constraints to be bounded.

The paper primarily deals with the description of an appropriate oracle. We first discuss the realization of the oracle under requisite assumptions for generic convex problems. Afterwards we show that for convex quadratic programs the algorithm of the oracle is universally applicable. Moreover, a method to construct approximate feasible and approximate optimal, primal solutions is given.

MSC: 90C25, 90C30, 90C06, 65K05

Keywords: level method, cutting-plane methods, decomposition methods, convex programming, nonsmooth programming.

1 Problem Formulation and Assumptions

We consider optimization problems of the following form

$$(P)\begin{cases} f(x) \to \inf\limits_x \\ g(x) \leq \mathbb{O} \\ h(x) \leq \mathbb{O} \\ x \in \mathbb{R}^n, \end{cases}$$

where $g(x) = (g_1(x), g_2(x), \ldots, g_m(x))^\top$ and $h(x) = (h_1(x), h_2(x), \ldots, h_p(x))^\top$. The functions f, g_i and h_j are supposed to be convex on $\mathbb{R}^n$ for all i and for all j. Moreover, it is supposed that all functions f, g_i and h_j are differentiable on the entire space $\mathbb{R}^n$.
We define the optimal value of (P) as

$$f^* \;:=\; \inf_{x\in\mathbb{R}^n}\{f(x) : g(x) \leq \mathbb{O}, h(x) \leq \mathbb{O}\}$$

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

and we denote the optimal set of (P) by

$$X^* = \{x \in \mathbb{R}^n : g(x) \leq \mathbb{O}, h(x) \leq \mathbb{O}, f(x) = f^*\}.$$

As usual, we assume the feasible set of (P) to be nonempty, which means $f^* < +\infty$. Of course, the case $X^* = \emptyset$ is not excluded. We only have to suppose problem (P) to be solvable in the sense that the optimal value of (P) is bounded below, i.e. $f^* > -\infty$.
Finally, we demand that a regularity condition holds for the feasible set of (P) with respect to the constraints connected with function g. For instance, it is possible to assume the existence of a vector $\bar{x} \in \mathbb{R}^n$ with $h(\bar{x}) \leq \mathbb{O}$ such that

$$g_i(\bar{x}) < 0 \;\; \forall i = 1, 2, \ldots, m.$$

This assumption is usually refered as the Slater-condition.
We remark that in the special cases, where (P) is a linear or a convex quadratic program, the theory developed in this article is applicable as well. In these cases an additional regularity condition will be superfluous. Section 5 deals with convex, quadratic programs.

2 Dual Decomposition Approach

To solve problem (P) we choose an approach involving a dual problem corresponding to (P). With the help of the Lagrangian function $L(x, \lambda)$ defined by

$$L(x, \lambda) := f(x) + \langle \lambda, g(x) \rangle, \;\; \lambda \geq \mathbb{O}$$

we construct the so-called dual function for (P) with respect to the constraints connected with g as

$$\varphi(\lambda) := \inf_{h(x) \leq \mathbb{O}} L(x, \lambda).$$

We remark that $\varphi(\lambda)$ is defined as the optimal value of an optimization problem. In what follows let this program be denoted by $(\varphi(\lambda))$. In the usual way the dual problem

$$(D) \begin{cases} \varphi(\lambda) \to \max\limits_{\lambda} \\ \lambda \geq \mathbb{O} \end{cases}$$

can be assigned to the primal problem (P). Let φ^* denote the optimal value of (D).
First we discuss some known facts from duality theory. Between the primal and the dual objective function the relation

$$\varphi(\lambda) \leq f(x)$$

holds for all dual feasible λ and all primal feasible x. This relation is known as “weak duality”. It immediately follows that $\varphi^* \leq f^*$ and since (P) is supposed to be feasible one gets $\varphi^* < +\infty$.

Since we have additionally assumed that a regularity condition holds for (P), we can conclude "strong duality", i.e. $f^* = \varphi^*$ and moreover, there exists $\lambda^* \geq \mathbb{O}$ such that

$$f^* \;=\; \varphi(\lambda^*).$$

This means that (D) has at least one optimal solution.
We remind the reader that the set $\{x \in \mathbb{R}^n : h(x) \leq \mathbb{O}\}$ was not supposed to be bounded in general. This means that the optimal value of $(\varphi(\lambda))$ may equal to $-\infty$ for certain $\lambda \geq \mathbb{O}$ although its objective function is differentiable on the entire space. Let therefore denote $\operatorname{dom}\varphi$ the effective domain of $\varphi(\lambda)$, i.e.

$$\operatorname{dom}\varphi \;=\; \{\lambda \geq \mathbb{O} : \varphi(\lambda) > -\infty\}.$$

To avoid confusion, we remark that independent of $\operatorname{dom}\varphi$ we consider the set $\mathbb{R}^m_+$ as the feasible set of problem (D).
The set $\operatorname{dom}\varphi$ is always nonempty under the previous assumptions since at least the aforementioned dual optimal solution λ^* is an element of $\operatorname{dom}\varphi$.
Unfortunately, the existence of a primal minimizer can not be concluded without additional assumptions. Consider for instance the case $f(x) = e^x$ and $g(x) = x$ with nonexisting function $h(x)$. It can be easily seen that $f^* = 0 = \varphi^*$. Moreover, $\lambda^* = 0$ is dual optimal and $\operatorname{dom}\varphi = \{0\}$ but there is no x^* with $e^{x^*} = 0$.
It is a well known fact from duality theory that the function $\varphi(\lambda)$ is concave on the convex set $\operatorname{dom}\varphi$. We set $\hat{\varphi}(\lambda) := -\varphi(\lambda)$ and consider the problem

$$\begin{cases} \hat{\varphi}(\lambda) \to \min\limits_{\lambda} \\ \lambda \geq \mathbb{O} \end{cases} \tag{1}$$

instead of considering (D). Then (1) is a convex problem of minimization.
Obviously, vectors from $\mathbb{R}^m_+ \setminus \operatorname{dom}\varphi$ can not be optimal in (1). But simply changing the feasible set of (1) to "$\operatorname{dom}\varphi$" is actually impossible since $\operatorname{dom}\varphi$ can not be described explicitely in general.
It can be seen that the chosen approach is not of practical use, if problem (P) has no special structure. In general it is supposed that (P) is difficult to solve, whereas the problems $(\varphi(\lambda))$ have nice properties and are easier to solve. The typical situation is (P) having a block-angular structure and f having compatible structure. The difficulty is the presence of the constraints connected with g. These constraints couple all variables. They are therefore refered as the so-called "coupling constraints". Since our dual approach makes it possible to get rid of these constraints, the problems $(\varphi(\lambda))$ decompose into smaller subproblems which can be solved separately or one can solve them simultaneously on parallel computer architectures. Another possible application is the computation of lower bounds in integer programming via Lagrangian relaxation.

3 Algorithm of the Level Method

It is convenient to solve problems of the type, which problem (1) is of, with the help of cutting plane methods. The algorithm we use to solve (1) is a so-called level method, a special cutting plane method, and was first mentioned in [7]. Essentially we use a variant described in [2]. This algorithm generates a sequence $\{\lambda^i\}_{i=1,2,\dots}$ to find the optimal solution of a convex minimization problem with compact, polyhedral feasible set. Obviously, the feasible set of (1) is polyhedral but unbounded. For $U > 0$ we define

$$\Lambda \;:=\; \{\lambda \in \mathbb{R}^m_+ : \lambda \le U \cdot \mathbb{1}\},$$

where $\mathbb{1}$ is the vector consisting of ones. Since (1) has minimizers, it is possible to find a suitable U such that Λ contains at least one minimizer of (1). Therefore, we henceforth consider the following problem

$$(\hat{D}) \begin{cases} \hat{\varphi}(\lambda) \to \inf\limits_{\lambda} \\ \lambda \in \Lambda \end{cases} \tag{2}$$

instead of (D). It is explicitely allowed that $\hat{\varphi}(\lambda) = +\infty$ for certain $\lambda \in \Lambda$.
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, separating 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* $\lambda^1 \in \Lambda$, *the level parameter* $\Theta \in (0,1)$ *and* $C > 0$. *We set* $\hat{\varphi}^*_0 = \infty$ *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* $\lambda^k \in dom\,\hat{\varphi}$ *then* b_k *is a subgradient of* $\hat{\varphi}$ *at* λ^k *and* $\beta_k = \hat{\varphi}(\lambda^k)$,
- *if* $\lambda^k \notin dom\,\hat{\varphi}$ *then for* b_k *and* $\beta_k > 0$ *holds*

$$\langle b_k, \lambda - \lambda^k\rangle + \beta_k \quad \le 0 \;\; \forall \lambda \in dom\,\hat{\varphi}. \tag{3}$$

(2.) If $\lambda^k \in dom\,\hat{\varphi}$ *then set* $\hat{\varphi}^*_k = \min\{\hat{\varphi}^*_{k-1}, \hat{\varphi}(\lambda^k)\}$ *and update the values* β^k_i *using the following rule*

$$\beta^k_i \;=\; \begin{cases} \hat{\varphi}(\lambda^i) - \hat{\varphi}^*_k & \text{if } \lambda^i \in dom\,\hat{\varphi}, \;\; i = 1, 2, \dots, k \\ \beta^{k-1}_i & \text{if } \lambda^i \notin dom\,\hat{\varphi}, \;\; i = 1, 2, \dots, k-1. \end{cases} \tag{4}$$

Otherwise $\hat{\varphi}^*_k := \hat{\varphi}^*_{k-1}$ *and* $\beta^k_k := \beta_k$.

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

$$\begin{cases} t \to \max \\ \langle b_i, \lambda - \lambda^i\rangle + \beta^k_i + \|b_i\| t \le 0 \quad i = 1, 2, \dots, k \\ \lambda \in \Lambda. \end{cases} \tag{5}$$

If $\hat{\varphi}_k^* < \infty$ and $\Delta_k < C \cdot \varepsilon$ additionally compute Δ_k' as the optimal value of

$$\begin{cases} t \to \max \\ \langle b_i, \lambda - \lambda^i \rangle + \beta_i^k + t \leq 0 & \forall i : \lambda^i \in dom\,\hat{\varphi} \\ \langle b_i, \lambda - \lambda^i \rangle + \beta_i^k \leq 0 & \forall i : \lambda^i \notin dom\,\hat{\varphi} \\ \lambda \in \Lambda. \end{cases} \tag{6}$$

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

$$\begin{cases} \|\lambda - \lambda^k\|^2 \to \min \\ \langle b_i, \lambda - \lambda^i \rangle + \beta_i^k + \Theta \cdot \|b_i\| \cdot \Delta_k \leq 0 \quad i = 1, 2, \dots, k \\ \lambda \in \Lambda \end{cases} \tag{7}$$

as new iterate λ^{k+1}.

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

Remarks

- To prove convergence of the algorithm, several assumptions must hold: $\mathrm{dom}\,\hat{\varphi}$ has to be a proper, convex function, all vectors b^i generated by the algorithm must have a norm bounded by some constant $L > 0$, Λ is supposed to be a polyhedron and, furthermore, the set $\mathrm{int}(\mathrm{dom}\,\varphi \cap \Lambda)$ must be nonempty. The last assumption especially means that we have to demand the interior of $\mathrm{dom}\,\varphi$ to be nonempty. This property can not always be ensured. In [4] we give an example with $\mathrm{int}(\mathrm{dom}\,\varphi) = \emptyset$. Ways to overcome this problem have already be found. For the sake of brevity we will skip these results here. They shall be published later.

- The following fact can be proven: If $\varepsilon > 0$ then the method stops after a finite number k_0 of iterations and we have $0 \leq \hat{\varphi}_{k_0}^* - \hat{\varphi}^* \leq \varepsilon$. The proof for linear problems of primal decomposition can be found in [3] and can easily be generalized.

- The described algorithm already contains certain modifications.
  - In problems (5) and (7) 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.
  - The described method exploits the strategy of so-called “Deeper Cuts”. If we set $\beta_i^k = 0\ \forall i : \lambda^i \in \mathrm{dom}\,\hat{\varphi}$ we have the standard method.

4 Realization of the Oracle

After having described the proposed solution method, our next aim is to apply Algorithm 1 to problem (1). This algorithm demands an oracle to be given. The core problem of an implementation of the method will be to make the oracle available. A brief look at the steps of the algorithm reveals that the used oracle is supposed to be able to:

- decide, whether a current iterate λ^k is in $\operatorname{dom}\varphi$ or not, i.e. to compute $\hat{\varphi}(\lambda^k)$
- compute subgradients of the objective function $\hat{\varphi}$,
- construct separating hyperplanes.

The aim of the forthcoming sections will therefore be to show, how the oracle can be taught to meet all three requirements.

4.1 Determinig Infinity

Algorithm 1 consists of two different types of iterations. If $\lambda^k \in \operatorname{dom}\varphi$ then a standard cutting-plane step is done. If $\lambda^k \notin \operatorname{dom}\varphi$ then the oracle must deliver data to construct a hyperplane seperating λ^k from $\operatorname{dom}\varphi$. As already mentioned, there is no way to describe the set $\operatorname{dom}\varphi$ explicitely in general. Therefore, when the oracle is provided with a current iterate λ^k, its first task is to find out, if $\lambda^k \in \operatorname{dom}\varphi$. Henceforth points from $\mathbb{R}^m_+ \setminus \operatorname{dom}\varphi$ shall be called INFINITY POINTS.
We remember the reader that the value $\varphi(\lambda^k)$ is the optimal value of a convex optimization problem. Therefore, it is possible to use duality theory to obtain results concerning the finiteness of $\varphi(\lambda^k)$. The dual problem corresponding to problem $(\varphi(\lambda))$ is

$$(\varphi(\lambda)_D) \begin{cases} \psi_\lambda(\mu) \to \sup\limits_{\mu} \\ \mu \geq \mathbb{O}, \end{cases}$$

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

$$\psi_\lambda(\mu) \quad := \quad \inf_{x \in \mathbb{R}^n} \{f(x) + \langle \lambda, g(x) \rangle + \langle \mu, h(x) \rangle\} \tag{8}$$

Let ψ^*_λ denote the optimal value of $(\varphi(\lambda)_D)$. Problem $(\varphi(\lambda)_D)$ and the function $\psi_\lambda(\mu)$ both depend on the parameter λ.
The value $\psi_\lambda(\mu)$ is defined as the optimal value of an unconstrained optimization problem. Evaluating first order necessary and sufficient conditions for this program and regarding duality results leads to the construction of an INFINITY POINT INDICATOR FUNCTION.
Let $\eta(\lambda)$ be the optimal value function of the problem

$$(\eta(\lambda)) \begin{cases} \left\| \nabla f(x) + G(x)^\top \lambda + H(x)^\top \mu \right\|^2 \to \inf\limits_{x,\mu} \\ \mu \geq \mathbb{O}, x \in \mathbb{R}^n, \end{cases} \tag{9}$$

where $G(x)$ respectively $H(x)$ is the Jacobian of $g(x)$ respectively $h(x)$. Obviously, $\eta(\lambda) \geq 0$ holds for all $\lambda \in \mathbb{R}^m_+$ and moreover, $\eta(\lambda) < \infty \quad \forall \lambda \in \mathbb{R}^m_+$ since $(\eta(\lambda))$ has feasible solutions for arbitrary λ.
To prove statements concerning $\varphi(\lambda)$ we will sometimes need an additional assumption.

Assumption 2 *An arbitrary regularity condition holds for the feasible set of problem* $(\varphi(\lambda))$.

We remark that the feasible set of $(\varphi(\lambda))$ does not depend on λ.
Solving problem $(\eta(\lambda))$ is closely related to determining the infinity of $\varphi(\lambda)$ and to the construction of separating hyperplanes, too. The following theorem shows the possible use of $(\eta(\lambda))$.

Theorem 3 *Let Assumption 2 hold. If $\lambda^0 \in dom\,\varphi$ then $\eta(\lambda^0) = 0$.*

<u>*Proof.*</u>
Let $\lambda^0 \in \operatorname{dom}\varphi$, i.e. $\varphi(\lambda^0) > -\infty$. Since owing to Assumption 2 problem $(\varphi(\lambda^0))$ is regular, it follows that strong duality holds between $(\varphi(\lambda^0))$ and its dual $(\varphi(\lambda^0)_D)$ and the latter possesses at least one optimal solution $\mu^0 \geq \mathbb{O}$, i.e. it holds

$$\varphi(\lambda^0) \quad = \quad \psi^*_{\lambda^0} \quad = \quad \psi_{\lambda^0}(\mu^0).$$

Since $\lambda^0 \in \operatorname{dom}\varphi$ we have

$$\begin{aligned} -\infty &< \varphi(\lambda^0) \\ &= \psi^*_{\lambda^0} \\ &= \inf_{x\in\mathbb{R}^n} \left\{ f(x) + \langle \lambda^0, g(x)\rangle + \langle \mu^0, h(x)\rangle \right\}. \end{aligned}$$

Therefore, the function $f(x) + \langle \lambda^0, g(x)\rangle + \langle \mu^0, h(x)\rangle$ is bounded below and it is lower semicontinuous because of the differentiability of the functions f, g_i and h_j. We can therefore apply Theorem 6.3 from [6]. It follows that there is a sequence $\{x^k\}_{k=1}^{\infty}$ such that

$$\begin{aligned} \lim_{k\to\infty} \left(\nabla f(x^k) + G(x^k)^\top \lambda^0 + H(x^k)^\top \mu^0\right) &= 0 \\ \Rightarrow \lim_{k\to\infty} \left\|\nabla f(x^k) + G(x^k)^\top \lambda^0 + H(x^k)^\top \mu^0\right\|^2 &= 0. \end{aligned}$$

The sequence $\{(x^k, \mu^0)^\top\}_{k=1}^{\infty}$ is feasible in $(\eta(\lambda^0))$. Moreover, this sequence realizes the optimal value of $(\eta(\lambda^0))$ since $\eta(\lambda^0) \geq 0$. It follows $\eta(\lambda^0) = 0$ and the theorem is proven.
■

The hereafter stated corollary is a direct consequence of the previous theorem.

Corollary 4 *Let Assumption 2 hold. If $\eta(\lambda^0) > 0$ then $\lambda^0 \notin dom\,\varphi$.*

It can be seen that Theorem 3 is not sufficient to decide, whether a current iterate λ^k is in $\operatorname{dom}\varphi$ or not. To make $\eta(\lambda)$ a true infinity point indicator function the converse statement to Theorem 3 is needed. Unfortunately, couterexamples show that this statement is not valid in general. We will therefore have to make additional assumptions.

Assumption 5 *For arbitrary $\lambda \in \mathbb{R}^m_+$ there exists an optimal solution of problem $(\eta(\lambda))$.*

Theorem 6 *Let Assumption 5 hold. If $\eta(\lambda^0) = 0$ then $\lambda^0 \in dom\,\varphi$.*

Proof.
Let $\lambda^0 \in \mathbb{R}^m_+$ such that $\eta(\lambda^0) = 0$. Since Assumption 5 holds, there exists an optimal solution $(x^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^p_+$ of $(\eta(\lambda^0))$. For this solution

$$\nabla f(x^*) + G(x^*)^\top \lambda^0 + H(x^*)^\top \mu^* \;=\; \mathbb{O}$$

must hold. This is the necessary and sufficient condition for x^* to be a global minimizer of problem $(\psi_{\lambda^0}(\mu^*))$. From weak duality relation we get

$$\begin{aligned}
\varphi(\lambda^0) &\geq \psi^*_{\lambda^0} \\
&= \sup_{\mu \geq \mathbb{O}} \{\psi_{\lambda^0}(\mu)\} \\
&\geq \psi_{\lambda^0}(\mu^*) \\
&= f(x^*) + \langle \lambda^0, g(x^*) \rangle + \langle \mu^*, h(x^*) \rangle \\
&> -\infty
\end{aligned}$$

which proves the theorem. ■

Consequently, if Assumptions 2 and 5 hold simultaneously, then $\lambda^0 \in \operatorname{dom} \varphi$ if and only if $\eta(\lambda^0) = 0$. In this case the oracle is able to make a definite decision regarding the property to be an infinity point for a current iterate.

4.2 Supergradients of the Dual Function

For iterates $\lambda^k \in \operatorname{dom} \varphi$ the level method from section 3 has to solve problem $(\varphi(\lambda^k))$ to compute the objective function value and it has to compute a subgradient of $\hat{\varphi}$ at λ^k, which is of course directly connected with a supergradient of φ.
The following result is well-known. Nevertheless, it shall be stated and proven here, since its message is of crucial interest for realizing Algorithm 1 .

Theorem 7 (ε-Supergradients of the dual function)
Let $\varepsilon \geq 0$, $\lambda^0 \in \operatorname{dom} \varphi$ and let x^0 be an ε-optimal solution of problem $(\varphi(\lambda^0))$. Then the vector $g(x^0)$ is an ε-supergradient of the function $\varphi(\cdot)$ at λ^0.

Proof.
Since $\lambda^0 \in \operatorname{dom} \varphi$ it follows that $\varphi(\lambda)$ is finite. The vector x^0 is an ε-optimal solution of $(\varphi(\lambda^0))$. That means

$$\varphi(\lambda^0) \;\geq\; f(x^0) + \langle \lambda^0, g(x^0) \rangle - \varepsilon. \tag{10}$$

Moreover, x^0 is feasible in $(\varphi(\lambda))$ for arbitrary $\lambda \in \mathbb{R}^m$. Therefore, it holds

$$\varphi(\lambda) \;\leq\; f(x^0) + \langle \lambda, g(x^0) \rangle \;\; \forall \lambda \in \mathbb{R}^m. \tag{11}$$

Multiplying (10) by (-1) and adding it to (11) yields

$$\varphi(\lambda) - \varphi(\lambda^0) \;\leq\; \langle g(x^0), \lambda - \lambda^0 \rangle + \varepsilon \;\; \forall \lambda \in \mathbb{R}^m \tag{12}$$

which means that $g(x^0)$ is in the ε-superdifferential of $\varphi(\cdot)$ at λ^0. ■

One easily checks that under the assumptions of Theorem 7 $-g(x^0)$ is an ε-subgradient of $\hat{\varphi}$ at λ^0. The above stated theorem is therefore one of the needed keys for the realization of the oracle.

4.3 Construction of Separating Hyperplanes

The current section is devoted to the construction of so-called "domain-cuts". We remember the reader that a domain-cut at $\lambda^0 \notin \operatorname{dom} \varphi$ consists of a vector a and a number α such that the following two conditions hold:

$$\begin{aligned} \langle a, \lambda - \lambda^0 \rangle + \alpha &\leq 0 \ \ \forall \lambda \in \operatorname{dom} \varphi \\ \alpha &> 0, \end{aligned}$$

i.e. the hyperplane connected with a and α separates λ^0 from $\operatorname{dom} \varphi$ and λ^0 has positive distance from the hyperplane. The basic result for the realization of this aim is the following.

Theorem 8 *Let Assumptions 2 and 5 hold and let $\lambda^0 \notin dom\, \varphi$. Let the vector s 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 constructed, i.e.*

$$\langle s, \lambda - \lambda^0 \rangle + \eta(\lambda^0) \leq 0 \ \ \forall \lambda \in dom\, \varphi \tag{13}$$

$$\eta(\lambda^0) > 0. \tag{14}$$

Proof.
Let $\lambda^0 \in \mathbb{R}^m_+ \setminus \operatorname{dom} \varphi$. From Theorem 6 we immediately get (14).
Since $s \in \partial \eta(\lambda^0)$ we know

$$\eta(\lambda) - \eta(\lambda^0) \geq \langle s, \lambda - \lambda^0 \rangle \ \ \forall \lambda \in \mathbb{R}^m. \tag{15}$$

Considering Theorem 3 we have moreover that $\eta(\lambda) = 0 \ \ \forall \lambda \in \operatorname{dom} \varphi$. Writing down (15) only for $\lambda \in \operatorname{dom} \varphi$ yields the validity of (13). ■

Obviously, Theorem 8 does not contain statements concerning the existence of subgradients of η. But we know that the subdifferential of a convex function is nonempty for points in the relative interior of the function's domain and, moreover, it is obvious that the domain of η coincides with $\mathbb{R}^m$. Therefore, convexity of $\eta(\lambda)$ should be ensured. It is for example possible to use the hereafter stated lemma.

Lemma 9 *Let the function*

$$\|\nabla f(x) + H(x)^\top \mu + G(x)^\top \lambda\|^2 \tag{16}$$

be jointly convex with respect to (x, μ, λ) on $\mathbb{R}^n \times \mathbb{R}^p_+ \times \mathbb{R}^m$. Then the optimal value function $\eta(\lambda)$ is convex on $\mathbb{R}^m$.

Proof.
See [10] Proposition 2.6. ■

Unfortunately, the assumption of Lemma 9 is not always valid. An analogue of Theorem 8 is also possible to be proven, if we assume that η is quasi-convex and the needed vector s is from a generalized, quasi-convex subdifferential. To prove quasi-convexity of η it is sufficient that (16) is jointly quasi-convex. The next example will show that even that property is not possible to be proven.

Example 10 *We consider problem (P) with*

$$f(x_1, x_2) \;=\; \sqrt{1 + x_1^2 + x_2^2} + x_1^2$$

and with arbitrary convex differentiable functions g and h. For (16) to be quasi-convex it is necessary that (16) is quasi-convex with respect to x for fixed μ and λ. Let $\mu := \lambda := \mathbb{O}$. It is then necessary that $\|\nabla f(x_1, x_2)\|^2$ is quasi-convex. Considering

$$\|\nabla f(x_1, x_2)\|^2 \;=\; \left(\frac{x_1}{\sqrt{1 + x_1^2 + x_2^2}} + 2x_1 \right)^2 + \frac{x_2^2}{1 + x_1^2 + x_2^2} \tag{17}$$

one easily checks that (17) is not quasi-convex.

From Lemma 9 one can immediately conclude that for linear programs and for convex, quadratic programs function $\eta(\lambda)$ is convex. See Section 5 for more details.
Besides Lemma 9 there are more cases where η is a convex function. They shall not be stated and proven here since they are beyond the scope of this paper. We will work with a new assumption in the forthcoming parts of this article instead.

Assumption 11 *The function $\eta(\cdot)$ is convex on $\mathbb{R}^m$.*

The above assumption ensures that for every iterate $\lambda^k \notin \operatorname{dom} \varphi$ it is possible to build up a separating hyperplane with the help of a subgradient of η.
Finally, we have to discuss how the needed subgradient can be computed. We will show that this subgradient is connected with an optimal solution of $(\eta(\lambda))$. Since (16) is not supposed to be jointly convex it is not possible to cite standard references.

Theorem 12 *Let Assumptions 5 and 11 hold. For $\lambda \in \mathbb{R}^m$ let $(x^*(\lambda), \mu^*(\lambda))$ be a globally optimal solution of $(\eta(\lambda))$. Then*

$$\nabla \eta(\lambda) \;=\; 2 \cdot G(x^*(\lambda)) \left(\nabla f(x^*(\lambda)) + H(x^*(\lambda))^\top \mu^*(\lambda) + G(x^*(\lambda))^\top \lambda \right)$$

holds.

Proof.
Let $\lambda^0 \in \mathbb{R}^m$ arbitrary but fixed. Since η is convex by assumption there is a subgradient of η for each $\lambda \in \mathbb{R}^m$. Let therefore be $a \in \partial\eta(\lambda^0)$. It holds

$$\begin{aligned} \eta(\lambda) - \eta(\lambda^0) &\geq \langle a, \lambda - \lambda^0 \rangle \ \ \forall \lambda \in \mathbb{R}^m \\ \Leftrightarrow \eta(\lambda) - \langle a, \lambda \rangle &\geq \eta(\lambda^0) - \langle a, \lambda^0 \rangle \ \ \forall \lambda \in \mathbb{R}^m, \end{aligned}$$

which means that λ^0 is a globally optimal solution of the program

$$\inf_{\lambda} \left\{ \eta(\lambda) - \langle a, \lambda \rangle \right\}.$$

Applying the definition of η we get that λ^0 is a globally optimal solution of problem

$$\inf_{\lambda} \left\{ \inf_{\mu \geq \mathbb{O}, x} \left\{ \left\| \nabla f(x) + H(x)^\top \mu + G(x)^\top \lambda \right\|^2 \right\} - \langle a, \lambda \rangle \right\}.$$

Considering that $(x^*(\lambda^0), \mu^*(\lambda^0))$ is a globally optimal solution of problem $(\eta(\lambda^0))$ it follows that $(x^*(\lambda^0), \mu^*(\lambda^0), \lambda^0)$ is a globally optimal solution of the following optimization problem:

$$\inf_{\mu \geq \mathbb{O}, x, \lambda} \left\{ \left\| \nabla f(x) + H(x)^\top \mu + G(x)^\top \lambda \right\|^2 - \langle a, \lambda \rangle \right\}.$$

Since $(x^*(\lambda^0), \mu^*(\lambda^0), \lambda^0)$ is an optimal solution, the first order necessary conditions must hold. This means especially that

$$2 \cdot G(x^*(\lambda^0)) \cdot \left(\nabla f(x^*(\lambda^0)) + G(x^*(\lambda^0))^\top \lambda^0 + H(x^*(\lambda^0))^\top \mu^*(\lambda^0) \right) = a. \qquad (18)$$

Since a was an arbitrary element of $\partial\eta(\lambda^0)$ and since we have proven that a can be represented by the above formula, we have actually proven that $\partial\eta(\lambda^0)$ only consists of one single element. Considering that the domain of function $\eta(\lambda)$ coincides with $\mathbb{R}^m$ and together with the convexity of $\eta(\lambda)$ we have proven that the right-hand-side of (18) is the gradient of η at λ^0. ■

Therefore, the above theorem yields the last property, which is necessary to realize the oracle.

4.4 Algorithm of the Oracle

The theorems stated in the previous subsections make it clear how the oracle has to be realized. Nevertheless, we will give the steps of the detailed algorithm of the oracle for the sake of completeness. We suppose Assumptions that 2, 5 and 11 hold. Moreover, we assume that for every $\lambda \in \operatorname{dom} \varphi$ there is an optimal solution of problem $(\varphi(\lambda))$.
Assume that the algorithm of the oracle has been started with λ^k being the input.

Algorithm 13 (Oracle)

(1.) Compute the optimal value $\eta(\lambda^k)$ *and an optimal solution* $(x^*(\lambda^k), \mu^*(\lambda^k))$ *of problem*

$$\begin{cases} \left\|\nabla f(x) + G(x)^\top \lambda^k + H(x)^\top \mu\right\|^2 \to \inf\limits_{x,\mu} \\ \mu \geq \mathbb{O}, x \in \mathbb{R}^n. \end{cases}$$

(2.) If $\eta(\lambda^k) > 0$ *then*

$$\begin{aligned} b^k &:= G(x^*(\lambda^k)) \cdot \left(\nabla f(x^*(\lambda^k)) + G(x^*(\lambda^k))^\top \lambda^k + H(x^*(\lambda^k))^\top \mu^*(\lambda^k)\right) \\ \beta_k^k &:= \eta(\lambda^k) \end{aligned}$$

and **STOP**. *Otherwise continue with step (3.).*

(3.) Compute the optimal value $\varphi(\lambda^k)$ *and an optimal solution* x^k *of problem*

$$\begin{cases} f(x) + \langle \lambda^k, g(x) \rangle \to \inf\limits_{x} \\ h(x) \leq \mathbb{O} \end{cases}$$

and set

$$b^k = -g(x^k).$$

5 Linear and Convex Quadratic Programs

Of course, the theory of the previous sections covers linear and convex quadratic programming problems. Nevertheless, we will devote the current section to a brief overview over this class of programs because of its importance in practical applications. See [4] for a more detailed survey of dual decomposition in convex quadratic programming.
In the sequel we consider the special case of problem (P) of the following form:

$$(P_Q) \begin{cases} \frac{1}{2}\langle x, Qx \rangle + \langle c, x \rangle \to \inf\limits_{x} \\ Gx \leq g \\ Hx \leq h. \end{cases}$$

The matrix Q is supposed to be symmetric and positive semidefinite. The case $Q = \mathbb{O}$ is allowed. Since the constraint functions are affine, the needed regularity condition for our dual approach holds.
The following statement is the key to the realization of the oracle for the considered class of problems.

Proposition 14 *Assumptions 2, 5 and 11 are valid for* (P_Q).

Proof.
For the considered class of programs, the problems ($\varphi(\lambda)$) look as follows

$$\varphi(\lambda) \quad = \quad \inf_{Hx \leq h} \{ \tfrac{1}{2}\langle x, Qx\rangle + \langle c, x\rangle + \langle \lambda, Gx - g\rangle \}$$

and have a polyhedral feasible set. Therefore these problems are regular and Assumption 2 holds.
For $\lambda \in \mathbb{R}^m_+$ the function $\eta(\lambda)$ is defined by

$$\eta(\lambda) \quad = \quad \inf_{x, \mu \geq \mathbb{O}} \|Qx + c + G^\top \lambda + H^\top \mu\|^2$$

The function $\|\cdot\|$ is convex. Precomposition with an affine function a convex function, too. The function $(\cdot)^2$ is convex and nondecreasing for nonnegative arguments. It can be concluded that the function $\|Qx + c + G^\top \lambda + H^\top \mu\|^2$ is jointly convex with respect to (x, μ, λ) on $\mathbb{R}^n \times \mathbb{R}^p_+ \times \mathbb{R}^m$. Using Lemma 9 it follows that $\eta(\lambda)$ is convex on $\mathbb{R}^m$ and Assumption 11 holds.
Moreover, the objective function of ($\eta(\lambda)$) is convex, quadratic function which is bounded below by zero and the constraints are affine. Therefore, for all $\lambda \in \mathbb{R}^m_+$ problem ($\eta(\lambda)$) has minimizers which means that Assumption 5 holds. ■

Instead of using the squared Euclidean norm in problem ($\eta(\lambda)$) it is possible to use $\|\cdot\|_1$, where $\|x\|_1 = \sum_{i=1}^{n} |x_i|$. The advantage of using this norm is the fact that ($\eta(\lambda)$) can be written as a linear program. The dual problem corresponding to this program is

$$\begin{cases} \langle c + G^\top \lambda, \kappa\rangle \to \min\limits_{\kappa} \\ Q\kappa = \mathbb{O} \\ H\kappa \geq \mathbb{O} \\ -\mathbb{1} \leq \kappa \leq \mathbb{1}. \end{cases} \tag{19}$$

To indicate the difference to function $\eta(\lambda)$ we denote the optimal value function of (19) by $\hat{\eta}(\lambda)$. Since strong duality holds, we only solve (19) to compute the value of $\hat{\eta}(\lambda)$ in each iteration. Obviously $\hat{\eta}(\lambda)$ has essentially the same properties as function $\eta(\lambda)$ does, i.e. it is an infinity point indicator function.
Moreover, problem (19) will help us to find subgradients of $\hat{\eta}(\lambda)$.

Theorem 15 *Consider problem (P_Q). Let $\lambda^0 \notin dom\, \varphi$ and let κ^* be an optimal solution of problem (19) for $\lambda = \lambda^0$. Then*

$$G\kappa^* \quad \in \quad \partial\hat{\eta}(\lambda^0).$$

Proof.
See [4] Theorem 4. ■

For the sake of completeness we summarize the oracles responses for the considered class of convex, quadratic programs. By x^i we denote an optimal solution of problem ($\varphi(\lambda^i)$)

and by κ^i we denote an optimal solution of problem (19) for $\lambda = \lambda^i$. After k iterations the following holds:

$$b^i = \begin{cases} g - Gx^i & : \lambda^i \in \operatorname{dom}\varphi \\ G\kappa^i & : \lambda^i \notin \operatorname{dom}\varphi \end{cases} \tag{20}$$

$$\beta_i^k = \begin{cases} -\left(\frac{1}{2}\langle x^i, Qx^i\rangle + \langle c, x^i\rangle + \langle \lambda^i, Gx^i - g\rangle\right) - \hat{\varphi}_k^* & : \lambda^i \in \operatorname{dom}\varphi \\ \langle c + G^\top \lambda^i, \kappa^i\rangle & : \lambda^i \notin \operatorname{dom}\varphi. \end{cases} \tag{21}$$

5.1 Obtaining Primal Solutions

After having performed the iteration process for problem $(\hat{D})$ and having obtained an approximate optimal dual solution, it is of interest to construct primal optimal solutions. Primal information can be obtained from the method itself. Because of the fact that we solve $(\hat{D})$ only ε-optimal, we can only expect to find approximate feasible and approximate optimal, primal solutions. The next theorem shows, how to use the information of the iteration process to construct such approximate solutions.
For the sake of brevity for $k \geq 1$ we define two sets of indices by

$$\begin{aligned} I'_k &:= \{i \leq k : \lambda^i \in \operatorname{dom}\varphi\} \\ I''_k &:= \{i \leq k : \lambda^i \notin \operatorname{dom}\varphi\}. \end{aligned}$$

Obviously, $I'_k \cap I''_k = \emptyset$ and $I'_k \cup I''_k = \{1, 2, \ldots, k\}$.

Theorem 16 *After k iterations the stopping criterion of Algorithm 1 may have been fulfilled. The method may have generated the sequence of trial points $\{\lambda^i\}_{i=1,2,\ldots,k}$. Let λ^* be an optimal solution of (D) and define $U := \max_{j=1}^{m}\{\lambda_j^*\} + 1$. Let the first k components of a dual optimal solution of problem (6) be denoted by $\{s_i^k\}_{i=1,2,\ldots,k}$. Define the vectors*

$$\bar{x} := \sum_{i \in I'_k} s_i^k \cdot x^i, \quad \bar{\kappa} := \sum_{i \in I''_k} s_i^k \cdot \kappa^i, \quad x^\varepsilon := \bar{x} - \bar{\kappa}.$$

Then the following properties are true:

$$Hx^\varepsilon - h \leq \mathbb{O} \tag{22}$$

$$Gx^\varepsilon - g \leq \varepsilon \cdot \mathbb{1} \tag{23}$$

$$\tfrac{1}{2}\langle x^\varepsilon, Qx^\varepsilon\rangle + \langle c, x^\varepsilon\rangle \in [f^* - U \cdot \varepsilon \cdot m, f^* + \varepsilon]. \tag{24}$$

<u>*Proof.*</u>
First we construct the dual problem corresponding to (6). This problem can be written

down as follows:

$$\begin{cases} \sum\limits_{i \in I'_k \cup I''_k} s_i \cdot (\langle b^i, \lambda^i \rangle - \beta_i^k) + U\langle \mathbb{1}, t\rangle \to \inf\limits_{s,t} \\ \sum\limits_{i \in I'_k \cup I''_k} s_i \cdot b^i + t \geq \mathbb{O} \\ \sum\limits_{i \in I'_k} s_i = 1 \\ s \geq \mathbb{O}, t \geq \mathbb{O}. \end{cases} \tag{25}$$

Since the points x^i are optimal solutions of $(\varphi(\lambda^i))$, we get that $Hx^i - h \leq \mathbb{O}$ holds for $i \in I'_k$. From the third line of (25) it follows that $\bar{x}$ is a convex combination of the points x^i and, therefore, $H\bar{x} - h \leq \mathbb{O}$ holds, too. For $i \in I''_k$ the vectors κ^i are optimal in (19) for $\lambda = \lambda^i$, they are especially feasible with respect to the constraints $H\kappa \geq \mathbb{O}$. Since $s \geq \mathbb{O}$ we get $H\bar{\kappa} \geq \mathbb{O}$ and finally it follows

$$Hx^{\varepsilon} - h \;=\; \underbrace{H\bar{x} - h}_{\leq \mathbb{O}} - \underbrace{H\bar{\kappa}}_{\geq \mathbb{O}} \;\leq\; \mathbb{O}.$$

Between (6) and its dual problem (25) strong duality holds. Moreover, the vector $s^k = (s_1^k, s_2^k, \ldots, s_k^k)^\top$ is optimal in (25). It follows

$$\begin{aligned} \Delta'_k &= \inf_{s,t} \left\{ \sum_{i \in I'_k \cup I''_k} s_i \cdot (\langle b^i, \lambda^i \rangle - \beta_i^k) + U\langle \mathbb{1}, t\rangle : \right. \\ &\qquad \left. \sum_{i \in I'_k \cup I''_k} s_i \cdot b^i + t \geq \mathbb{O}, \sum_{i \in I'_k} s_i = 1, s \geq \mathbb{O}, t \geq \mathbb{O} \right\} \\ &= \sum_{i \in I'_k \cup I''_k} s_i^k \cdot (\langle b^i, \lambda^i \rangle - \beta_i^k) + U \cdot \sum_{j=1}^m \max\left\{ 0, - \sum_{i \in I'_k \cup I''_k} s_i^k \cdot b^i \right\}. \end{aligned} \tag{26}$$

Let G_j be the jth row of G and let g_j be the jth component of g. Substituting the oracles responses (20) and (21) in equation (26) one gets

$$\begin{aligned} \Delta'_k &= \sum_{i \in I'_k} s_i^k \cdot (\tfrac{1}{2}\langle x^i, Qx^i \rangle + \langle c, x^i \rangle + \hat{\varphi}_k^*) + \sum_{i \in I''_k} s_i^k \cdot (-\langle c, \kappa^i \rangle) + \\ &\qquad + U \cdot \sum_{j=1}^m \max\left\{ 0, - \left(\sum_{i \in I'_k} s_i^k \cdot (\langle -G_j, x^i \rangle + g_j) + \sum_{i \in I''_k} s_i^k \cdot \langle G_j, \kappa^i \rangle \right) \right\} \\ &= \hat{\varphi}_k^* + \sum_{i \in I'_k} s_i^k \cdot (\tfrac{1}{2}\langle x^i, Qx^i \rangle + \langle c, x^i \rangle) - \langle c, \bar{\kappa} \rangle \\ &\qquad + U \cdot \sum_{j=1}^m \max\left\{ 0, \langle G_j, \bar{x} \rangle - g_j - \langle G_j, \bar{\kappa} \rangle \right\}. \end{aligned} \tag{27}$$

Since the function $\frac{1}{2}\langle x, Qx\rangle + \langle c, x\rangle$ is convex, it holds:

$$\tfrac{1}{2}\langle \bar{x}, Q\bar{x}\rangle + \langle c, \bar{x}\rangle \quad \leq \quad \sum_{i\in I'_k} s_i^k \cdot \left(\tfrac{1}{2}\langle x^i, Qx^i\rangle + \langle c, x^i\rangle\right).$$

Together with (27) this means

$$\tfrac{1}{2}\langle \bar{x}, Q\bar{x}\rangle + \langle c, \bar{x}\rangle - \langle c, \bar{\kappa}\rangle + U \cdot \sum_{j=1}^{m} \max\left\{0, \langle G_j, \bar{x}\rangle - g_j - \langle G_j, \bar{\kappa}\rangle\right\} \quad \leq \quad \Delta'_k - \hat{\varphi}_k^*. \tag{28}$$

For $i \in I''_k$ the vectors κ^i are feasible solutions of (19) for $\lambda = \lambda^i$, which means especially that $Q\kappa^i = \mathbb{O}$. It follows that $\frac{1}{2}\langle \bar{\kappa}, Q\bar{\kappa}\rangle - \langle \bar{x}, Q\bar{\kappa}\rangle = 0$. Considering the definition of x^ε we get:

$$\begin{aligned}
& \tfrac{1}{2}\langle \bar{x}, Q\bar{x}\rangle + \langle c, \bar{x}\rangle - \langle c, \bar{\kappa}\rangle + U \cdot \sum_{j=1}^{m} \max\left\{0, \langle G_j, \bar{x}\rangle - g_j - \langle G_j, \bar{\kappa}\rangle\right\} \\
= \quad & \tfrac{1}{2}\langle \bar{x}, Q\bar{x}\rangle + \tfrac{1}{2}\langle \bar{\kappa}, Q\bar{\kappa}\rangle - \langle \bar{x}, Q\bar{\kappa}\rangle + \langle c, \bar{x}\rangle - \langle c, \bar{\kappa}\rangle \\
& \quad + U \cdot \sum_{j=1}^{m} \max\left\{0, \langle G_j, \bar{x}\rangle - g_j - \langle G_j, \bar{\kappa}\rangle\right\} \\
= \quad & \tfrac{1}{2}\langle x^\varepsilon, Qx^\varepsilon\rangle + \langle c, x^\varepsilon\rangle + U \cdot \sum_{j=1}^{m} \max\left\{0, \langle G_j, x^\varepsilon\rangle - g_j\right\}.
\end{aligned} \tag{29}$$

Taking into account that the stopping criterion of Algorithm 1 is fulfilled, from (28) and (29) we get:

$$\begin{aligned}
\tfrac{1}{2}\langle x^\varepsilon, Qx^\varepsilon\rangle + \langle c, x^\varepsilon\rangle + U \cdot \sum_{j=1}^{m} \max\left\{0, \langle G_j, x^\varepsilon\rangle - g_j\right\} \quad & \leq \quad \Delta'_k - \hat{\varphi}_k^* \\
& \leq \quad \varepsilon - \hat{\varphi}_k^* \\
& \leq \quad \varepsilon - \hat{\varphi}^* \\
& = \quad \varepsilon + f^*.
\end{aligned} \tag{30}$$

Since λ^* is an optimal solution of (D) and since strong duality holds between (P_Q) and (D) it follows:

$$\begin{aligned}
& \tfrac{1}{2}\langle x^\varepsilon, Qx^\varepsilon\rangle + \langle c, x^\varepsilon\rangle + U \cdot \sum_{j=1}^{m} \max\left\{0, \langle G_j, x^\varepsilon\rangle - g_j\rangle\right\} \quad \leq \quad f^* + \varepsilon \\
= \quad & \inf_{Hx\leq h}\left\{\tfrac{1}{2}\langle x, Qx\rangle + \langle c, x\rangle + \langle \lambda^*, Gx - g\rangle\right\} + \varepsilon \\
\leq \quad & \tfrac{1}{2}\langle x^\varepsilon, Qx^\varepsilon\rangle + \langle c, x^\varepsilon\rangle + \langle \lambda^*, Gx^\varepsilon - g\rangle + \varepsilon.
\end{aligned} \tag{31}$$

Subtracting $\frac{1}{2}\langle x^\varepsilon, Qx^\varepsilon\rangle + \langle c, x^\varepsilon\rangle$ on both sides of this inequality one gets:

$$U \cdot \sum_{j=1}^{m} \max\left\{0, \langle G_j, x^\varepsilon\rangle - g_j\right\} \quad \leq \quad \langle \lambda^*, Gx^\varepsilon - g\rangle + \varepsilon. \tag{32}$$

Performing a distinction of cases for each j, it is possible to obtain

$$0 \quad \leq \quad U \cdot \max\left\{0, \langle G_j, x^{\varepsilon}\rangle - g_j\right\} - \lambda_j^* \left(\langle G_j, x^{\varepsilon}\rangle - g_j\right),$$

and from that we conclude in connection with (32)

$$U \cdot \max\left\{0, \langle G_j, x^{\varepsilon}\rangle - g_j\right\} - \lambda_j^* \left(\langle G_j, x^{\varepsilon}\rangle - g_j\right) \quad \leq \quad \varepsilon.$$

Since $0 \leq \lambda_j^* < U$ we immediately get

$$\langle G_j, x^{\varepsilon}\rangle - g_j \quad \leq \quad \frac{\varepsilon}{U - \lambda_j^*}. \tag{33}$$

Using the definition of U it follows $U - \lambda_j^* > 1 \;\; \forall j$ and we can conclude

$$\langle G_j, x^{\varepsilon}\rangle - g_j \quad \leq \quad \varepsilon \;\; \forall j \tag{34}$$

which proves (23).

Obviously, the value $U \cdot \sum_{j=1}^{m} \max\left\{0, \langle G_j, x^{\varepsilon}\rangle - g_j\right\}$ is nonnegative. This means together with (30)

$$\tfrac{1}{2}\langle x^{\varepsilon}, Qx^{\varepsilon}\rangle + \langle c, x^{\varepsilon}\rangle \quad \leq \quad f^* + \varepsilon. \tag{35}$$

We have assumed that the vector λ^* is an optimal solution of (D). Moreover, x^{ε} is a feasible solution of problem $(\varphi(\lambda^*))$ since (22) is already proven. Therefore, the following holds:

$$\begin{aligned}
f^* \quad &= \quad \inf_{Hx - h \leq \mathbb{O}} \left\{\tfrac{1}{2}\langle x, Qx\rangle + \langle c, x\rangle + \langle \lambda^*, Gx - g\rangle\right\} \\
&\leq \quad \tfrac{1}{2}\langle x^{\varepsilon}, Qx^{\varepsilon}\rangle + \langle c, x^{\varepsilon}\rangle + \langle \lambda^*, Gx^{\varepsilon} - g\rangle \\
&\overset{(34)}{\leq} \quad \tfrac{1}{2}\langle x^{\varepsilon}, Qx^{\varepsilon}\rangle + \langle c, x^{\varepsilon}\rangle + \varepsilon \cdot \langle \lambda^*, \mathbb{1}\rangle \\
&\overset{\lambda^* \in \Lambda}{\leq} \quad \tfrac{1}{2}\langle x^{\varepsilon}, Qx^{\varepsilon}\rangle + \langle c, x^{\varepsilon}\rangle + U \cdot \varepsilon \cdot \langle \mathbb{1}, \mathbb{1}\rangle \\
&= \quad \tfrac{1}{2}\langle x^{\varepsilon}, Qx^{\varepsilon}\rangle + \langle c, x^{\varepsilon}\rangle + U \cdot \varepsilon \cdot m.
\end{aligned}$$

If we additionally take (35) into account, we get the following pair of estimates for the function value of x^{ε}

$$f^* - U \cdot \varepsilon \cdot m \quad \leq \quad \tfrac{1}{2}\langle x^{\varepsilon}, Qx^{\varepsilon}\rangle + \langle c, x^{\varepsilon}\rangle \quad \leq \quad f^* + \varepsilon \tag{36}$$

and the proof is finished. ■

Remark

- The above stated result is still true if we use our original function $\eta(\lambda)$ for the construction of separating hyperplanes but some additional results are used for that proof. We have used problem (19) for the sake of simplicity.

- Let $\lambda^{k,*}$ be an optimal solution of (6). If $\lambda^{k,*} < U \cdot \mathbb{1}$ then we can prove that $Gx^{\varepsilon} - g \leq \mathbb{O}$. Moreover, the lower bound for the function value of x^{ε} changes to f^*.

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", Optimization, Vol. 51(6), 2002, pp. 819 - 840

[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] K. Beer, M. Knobloch: "Utilization of the Level Method for Dual Decomposition in Convex Quadratic Programming", Preprint 2002-4, Faculty of Mathematics, University of Technology, Chemnitz

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

[6] I. Ekeland, R. Témam: "Convex analysis and variational problems", Unabridged, corrected republication of the 1976 English original, Society for Industrial and Applied Mathematics, Philadelphia, 1999

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

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

[9] K. Richter: "Lösungsverfahren für konvexe Optimierungsaufgaben mit Umrandungsstruktur auf der Grundlage gleichzeitiger primal-dualer Dekomposition", Dissertation, TU Chemnitz, 2000

[10] H. Tuy: "Convex Analysis and Global Optimization", Kluwer Academic Publishers, 1998

[11] T. Unger: "Erweiterungen der Levelmethode zur Lösung konvexer Optimierungsaufgaben" , Dissertation, TU Chemnitz, 2003