Kybernetika

Matthias Knobloch

Solving convex program via Lagrangian decomposition

Kybernetika, Vol. 40 (2004), No. 5, [595]--610

Persistent URL: http://dml.cz/dmlcz/135619

KYBERNETIKA — VOLUME 40 (2004), NUMBER 5, PAGES 595 – 610

SOLVING CONVEX PROGRAMS VIA LAGRANGIAN DECOMPOSITION

MATTHIAS KNOBLOCH

We consider general convex large-scale optimization problems with finite dimensional decision variables. 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 also capable to handle infinite function values. Therefore it is no longer necessary to assume that the feasible set with respect to the non-dualized constraints is bounded.

The paper primarily deals with the description of an appropriate oracle. We first discuss the realization of the oracle under appropriate assumptions for generic convex problems. Afterwards we show that for convex quadratic programs the algorithm of the oracle is universally applicable.

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

AMS Subject Classification: 90C25, 90C30, 90C06, 65K05

1. INTRODUCTION

Large-scale optimization problems has been attracting the attention of specialists already for many years. The reason for such an enduring interest comes from applications. As an example, one can consider portfolio optimization, where the dimension depend on the number of possible assets and can thus be very large. Another example comes from models for power plant optimization. These models are another class of large-scale problems.

Usually these problems have a structure which enables to use decomposition approaches. In the last years decomposition has become even more important since the fast development of parallel computers has revealed new areas where decomposition can be used to handle problems of very large scale with the help of computers. Primal decomposition approaches are for instance described in [3] and [16]. A method using simultaneous primal-dual decomposition can be found in [13].

Methods for dual decomposition often need the assumption that the feasible set of the considered program is bounded. Our method enables to get rid of this additional assumption. In [4] we have already discussed the basic theory of our method for convex quadratic problems. The aim of the paper is to show some application aspects associated with the solution of general convex programs.

2. PROBLEM FORMULATION AND ASSUMPTIONS

We consider optimization problems of the following form

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

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

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

and we denote the optimal set of (P) by

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

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 the problem (P) to be solvable in the sense that the optimal value of (P) is bounded from below, i.e. $f^* > -\infty$.

Finally, we demand that a constraint qualification 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

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

This assumption is usually referred as the *Slater*-condition.

We note 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 6 deals with convex, quadratic programs.

3. DUAL DECOMPOSITION APPROACH

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

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

we construct the so-called dual function for (P) with respect to the constraints $g(x) \leq \mathbb{O}$ as

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

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

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

can be associated with the primal problem (P). Let φ^* denote the optimal value of (D).

First we discuss some known facts from the 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 follows immediately that $\varphi^* \leq f^*$ and since (P) is supposed to be feasible we get $\varphi^* < +\infty$.

Since we have additionally assumed that a constraint qualification holds for (P), we can establish "strong duality", i. e. $f^* = \varphi^*$ and moreover, the existence of some $\lambda^* \geq \mathbb{O}$ such that

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

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 be equal to $-\infty$ for certain $\lambda \geq \mathbb{O}$ although its objective function is differentiable on the entire space. Let us therefore denote by $\operatorname{dom}\varphi$ the effective domain of $\varphi(\lambda)$, i. e.

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

To avoid confusion, we remark that independently 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 cannot be ensured without additional assumptions. Consider for instance the case $f(x) = e^x$ and $g(x) = x$ without any constraint 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^* such that $e^{x^*} = 0$.

It is well known from the 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

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

instead of considering (D). Then (1) is a convex problem of minimization.

Obviously, the vectors from $\mathbb{R}^m_+ \setminus \operatorname{dom} \varphi$ cannot be optimal in (1). Also simply changing the feasible set of (1) to "$\operatorname{dom} \varphi$" is not possible since $\operatorname{dom} \varphi$ cannot be described explicitly in general.

It can be seen that the chosen approach is not of practical use, unless the problem (P) has a specific 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 arises when (P) has block-angular constraints and f has a compatible structure. The difficulty is the presence of the constraints modeled by g. These constraints couple all variables. They are therefore referred as the so-called "coupling constraints". Since our dual approach makes it possible to get rid of these constraints, the problems $(\varphi(\lambda))$ can be decomposed 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 using the methods of integer programming via Lagrangian relaxation.

Problem (1) is a problem of minimizing a convex objective function subject to nonnegativity conditions. Unfortunately, the objective function is not given in an explicit form. Therefore, we will have to choose an appropriate method to solve (1). In the next section the proposed method will be described.

4. ALGORITHM OF THE LEVEL METHOD

It is convenient to solve problems of the type, including also the problem (1), 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, was first mentioned in [10]. 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

$$\Lambda \quad := \quad \{\lambda \in \mathbb{R}^m_+ : \lambda \leq U \cdot \mathbf{1}\},$$

where $\mathbf{1}$ 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

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

instead of (D). It is explicitly 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)

Step 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 $k = 1$.

Step 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 $\lambda^k \in \operatorname{dom}\hat{\varphi}$ then b_k is a subgradient of $\hat{\varphi}$ at λ^k and $\beta_k = \hat{\varphi}(\lambda^k)$,
- if $\lambda^k \notin \operatorname{dom}\hat{\varphi}$ then for b_k and $\beta_k > 0$ holds

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

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

$$\beta_i^k = \begin{cases} \hat{\varphi}(\lambda^i) - \hat{\varphi}_k^* & \text{if } \lambda^i \in \operatorname{dom}\hat{\varphi}, \ i = 1,2,\ldots,k \\ \beta_i^{k-1} & \text{if } \lambda^i \notin \operatorname{dom}\hat{\varphi}, \ i = 1,2,\ldots,k-1. \end{cases} \tag{4}$$

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

Step 3. Compute Δ_k as the optimal value of the problem

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

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

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

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

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

as new iterate λ^{k+1}.

Step 5. Set $k = k + 1$ and return to Step 1.

Remarks.

- To prove convergence of the algorithm, several assumptions must be fulfilled: $\hat{\varphi}$ 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 $\operatorname{int}(\operatorname{dom}\varphi \cap \Lambda)$ must be nonempty. The last assumption especially means that we have to demand the interior of $\operatorname{dom}\varphi$ to be nonempty. This property cannot always be ensured. In [4] we give an example with $\operatorname{int}(\operatorname{dom}\varphi) = \emptyset$. Methods how to overcome this problem have not been published yet.

- 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 of this fact can be found in [3]. Note that in [3] only linear problems and primal decomposition methods are discussed. Nevertheless, the proof can be easily adapted.

- The described algorithm already contains certain modifications.
  - In problems (5) and (7) we use normalized subgradients. If we replace 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 the so-called "Deeper Cuts". Instead of using the cuts

    $$\langle b^i, \lambda - \lambda^i \rangle \leq 0 \quad \forall i : \lambda^i \in \operatorname{dom}\hat{\varphi}$$

    we use cuts of the following form

    $$\langle b^i, \lambda - \lambda^i \rangle \leq \hat{\varphi}^*_k - \hat{\varphi}(\lambda^i) \quad \forall i : \lambda^i \in \operatorname{dom}\hat{\varphi}$$

    in the kth iteration. Since $\hat{\varphi}^*_k - \hat{\varphi}(\lambda^i) \leq 0$ the feasible sets of (5) and (7) become smaller in general and therefore these cuts are called "deeper". Practical experience (see [16]) has shown that the level method performs a lot better when deeper cuts are used.

- Further modifications for fixing the problem that our method demands unbounded storage are known. Techniques for subgradient selection are for instance described in [16] and techniques for subgradient aggregation are contained in [9].

Since the algorithm demands an oracle to be given we have to construct such an oracle which is capable to compute the data needed for the outer iterations. In the next section we will describe how the oracle can be constructed.

5. 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.

Therefore the aim of the forthcoming sections is to show, how the oracle can be constructed to meet all three requirements.

5.1. Determining 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 compute data to construct a hyperplane separating λ^k from $\operatorname{dom}\varphi$. As already mentioned, there is no way to describe the set $\operatorname{dom}\varphi$ explicitly 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 remind 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

$$\begin{cases} \psi_\lambda(\mu) \to \sup_\mu \\ \mu \geq \mathbb{O}, \end{cases}$$

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

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

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, $\eta(\lambda) < \infty \ \forall \lambda \in \mathbb{R}^m_+$ since $(\eta(\lambda))$ has feasible solutions for arbitrary λ.

To prove statements concerning $\varphi(\lambda)$ we will sometimes need an additional assumption.

Assumption 2. A constraint qualification holds for the feasible set of problem $(\varphi(\lambda))$.

The feasible set of $(\varphi(\lambda))$ is $\{x \in \mathbb{R}^n : h(x) \leq \mathbb{O}\}$. We note that this set obviously does not depend on the parameter 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 $\lambda^0 \in \operatorname{dom}\varphi$ then $\eta(\lambda^0) = 0$.

Proof. Let $\lambda^0 \in \operatorname{dom}\varphi$, i.e. $\varphi(\lambda^0) > -\infty$. Since in virtue of 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

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

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 from 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 next corollary is a direct consequence of the previous theorem.

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

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, counterexamples show that this statement is not valid in general. Therefore we 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 $\lambda^0 \in \operatorname{dom}\varphi$.

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 the 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.

5.2. Supergradients of the dual function

For iterates $\lambda^k \in \operatorname{dom}\varphi$ the level method from Section 4 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 will be stated and proven here, since its message is of crucial interest for implementing 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

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

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

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

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 necessary keys for the implementation of the oracle.

5.3. Construction of separating hyperplanes

This section is devoted to the construction of the 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:

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

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 $\lambda^0 \notin \operatorname{dom}\varphi$. 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.

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

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

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

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

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 3 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. To this purpose it is possible to use the next 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 [15] Proposition 2.6. □

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

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 6 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 make use of known results.

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

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

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

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

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

$$\inf_\lambda \left\{ \inf_{\mu \geq 0, x} \left\{ \left\| \nabla f(x) + H(x)^\top \mu + G(x)^\top \lambda \right\|^2 \right\} - \langle a, \lambda \rangle \right\}.$$

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

$$\inf_{\mu \geq 0, x, \lambda} \left\{ \left\| \nabla f(x) + H(x)^\top \mu + G(x)^\top \lambda \right\|^2 - \langle a, \lambda \rangle \right\}.$$

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

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

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 left-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.

5.4. Algorithm of the oracle

The theorems stated in the previous subsections clarify how the oracle has to be implemented. Nevertheless, we will give the steps of the detailed algorithm of the oracle for the sake of completeness. We suppose that Assumptions 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)

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

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

Step 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.

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

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

and set

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

After the algorithm of the oracle has been described, we are able use the algorithm to solve (1). Therefore, we can devote our interest to further topics which we intend to discuss, namely, to a very important subclass of convex programs.

6. 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:

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

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 necessary constraint qualification 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

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

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

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

The function $\|\cdot\|$ is convex. Composition with an affine function yields 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 from 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

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

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 $\lambda^0 \notin \operatorname{dom}\varphi$ and let κ^* be an optimal solution of problem (19) for $\lambda = \lambda^0$. Then

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

Proof. See [4] Theorem 4. □

Remarks.

- The advantage of using $\|\cdot\|_1$ instead of $\|\cdot\|_2^2$ is the fact that for linear and quadratic programs the oracle has to solve a linear program instead of a convex quadratic program.

- For the considered class of programs it is possible to describe a way to construct approximate feasible and approximate optimal primal solutions using the iteration data of the level method. For the sake of brevity we skip this theorem.

(Received September 30, 2003.)

REFERENCES

[1] A. Auslender: Existence of optimal solutions and duality results under weak conditions. Math. Programming *88* (2001), 45–59.

[2] K. Beer, E. G. Gol'stejn, and N. A. Sokolov: Minimization of a nondifferentiable, convex function, defined not everywhere. Optimization *51* (2002), 6, 819–840.

[3] K. Beer, E. G. Gol'stejn, and 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 2000.

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

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

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

[7] J. L. Goffin and J. P. Vial: Convex nondifferentiable optimization: a survey focused on the analytic center cutting plane method. Optimization Methods and Software *17* (2002), 5, 805–867.

[8] K. C. Kiwiel: Methods of Descent for Nondifferentiable Optimization. (Lecture Notes in Mathematics 1133), Springer Verlag, Heidelberg 1985.

[9] K. C. Kiwiel: Proximal level bundle methods for convex nondifferentiable optimization, saddle-point problems and variational inequalities. Math. Programming *69B* (1995), 89–105.

[10] C. Lémarechal, A. Nemirovskii, and Y. Nesterov: New variants of bundle methods. Math. Programming *69* (1995), 111–147.

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

[12] R. Mifflin: An algorithm for constrained optimization with semi-smooth functions. Math. Oper. Res. *2* (1997), 2, 191–207.

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

[14] N. Z. Shor: Minimization Methods for Non-differentiable Functions. Springer Verlag, Berlin 1985.

[15] H. Tuy: Convex Analysis and Global Optimization. Kluwer, Dordrecht 1998.

[16] T. Unger: Erweiterungen der Levelmethode zur Lösung konvexer Optimierungsaufgaben. Dissertation, Shaker Verlag, Aachen 2003.

Matthias Knobloch, Faculty of Mathematics, TU Chemnitz, 09107 Chemnitz. Germany.
e-mail: m.knobloch@mathematik.tu-chemnitz.de