Journal of Global Optimization **22:** 49–57, 2002.
© 2002 *Kluwer Academic Publishers. Printed in the Netherlands.*

A class of problems where dual bounds beat underestimation bounds

MIRJAM DÜR
Department of Statistics, Vienna University of Economics and Business Administration, A–1090 Vienna, Austria (E-mail: mirjam.duer@wu-wien.ac.at)

Abstract. We investigate the problem of minimizing a nonconvex function with respect to convex constraints, and we study different techniques to compute a lower bound on the optimal value: The method of using convex envelope functions on one hand, and the method of exploiting nonconvex duality on the other hand. We investigate which technique gives the better bound and develop conditions under which the dual bound is strictly better than the convex envelope bound. As a byproduct, we derive some interesting results on nonconvex duality.

Key words: Nonconvex duality, Dual bounds, Convex underestimation

Dedicated to Reiner Horst on the occasion of his 60th birthday.

1. Introduction

In this paper, we consider the global optimization problem of minimizing a nonconvex function subject to convex constraints:

$$(P) \quad \begin{array}{l} \min f(x) \\ \text{s.t. } h_i(x) \leqslant 0, \quad i = 1, \ldots, m, \\ x \in X, \end{array}$$

where $f : X \to \mathbb{R}$ is a lower semicontinuous function, $h_i : X \to \mathbb{R}$ $(i = 1, \ldots, m)$ are convex functions, and $X \subset \mathbb{R}^m$ is a convex compact set. Our aim is to study different methods to obtain lower bounds for the optimal value of (P).

The first technique which has been used since many years is to replace the objective function with some easier (i.e., convex or linear) subfunctional and solve the resulting problem. Obviously, the quality of a bound obtained by this means depends on the quality of the underestimating function. The best possible result is achieved when the so called convex envelope function is used.

DEFINITION 1. *Let $X \subset \mathbb{R}^n$ be convex and compact, and let $f : X \to \mathbb{R}$ be lower semicontinuous on X. A function $\varphi_f : X \to \mathbb{R}$ is called the convex envelope of f on X if it satisfies*

(a) $\varphi_f(x)$ is convex on X,

(b) $\varphi_f(x) \leqslant f(x)$ *for all* $x \in X$,

(c) there is no function $\psi : X \to \mathbb{R}$ *satisfying (a), (b) and* $\varphi_f(\bar{x}) < \psi(\bar{x})$ *for some point* $\bar{x} \in X$.

A detailed discussion of convex envelopes, their properties and their use for computing lower bounds can be found in Horst and Tuy (1996).

To calculate a lower bound using the convex envelope, we must solve the following convexified version of (P):

$$(\bar{P}) \quad \begin{aligned} &\min \varphi_f(x) \\ &\text{s.t. } h_i(x) \leqslant 0, \quad i = 1, \dots, m, \\ &x \in X. \end{aligned}$$

The second possibility to obtain a bound is to use nonconvex duality, see Dür (2001), Dür and Horst (1997), Nowak (2000), and Thoai (2001). Recall that the Lagrange–dual problem of (P) is

$$(D) \quad \sup_{\lambda \in \mathbb{R}^m_+} \inf_{x \in X} \left\{ f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) \right\}.$$

Using the notation $\min(P)$ and $\sup(D)$ to denote the optimal values of (P) and (D), respectively, the weak duality theorem tells us that always

$$\sup(D) \leqslant \min(P),$$

and therefore the dual optimal value always gives a lower bound for the primal one.

Thus equipped with two possible bounding techniques, the question arises which of the two is more powerful. Comparing (D) with the dual problem $(\bar{D})$ of $(\bar{P})$,

$$(\bar{D}) \quad \sup_{\lambda \in \mathbb{R}^m_+} \inf_{x \in X} \left\{ \varphi_f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) \right\},$$

and assuming strong duality for the pair of convex problems $(\bar{P})$ and $(\bar{D})$, it is easy to see that

$$\sup(D) \geqslant \sup(\bar{D}) = \min(\bar{P}). \tag{1}$$

So the dual bound is always at least as good as the convex envelope bound. Falk (1969) showed that in the case of linear constraints the two bounds $\sup(D)$ and $\min(\bar{P})$ coincide, see also Dür and Horst (1997). But little seems to be known for the nonlinear case.

The following example raises the expectation that in some cases the dual bound may be strictly better than the convex envelope bound.

EXAMPLE 2. Consider the one dimensional problem

$$\min_{x\in[-2,3]}\{-x^2 : x^2 - x - 2 \leqslant 0\}.$$

The optimal value is $\min(P) = -4$, attained at $x = 2$. The convexified problem $(\bar{P})$ takes the form

$$\min_{x\in[-2,3]}\{-x - 6 : x^2 - x - 2 \leqslant 0\}.$$

Its optimal value is $\min(\bar{P}) = -8$, also attained at $x = 2$. The dual (D) of (P) is

$$\sup_{\lambda\in\mathbb{R}_+}\min_{x\in[-2,3]}\{(\lambda - 1)x^2 - \lambda x - 2\lambda\},$$

which takes the optimal value $\sup(D) = -4.2$ at $\lambda = 6/5$ and $x = 3$. The poor lower bound provided by $(\bar{P})$ is therefore improved considerably.

In the remainder of the paper we develop conditions which guarantee that the dual bound is strictly better than the convex envelope bound.

2. Some results on nonconvex duality

It is well known that in convex programming, Slater's constraint qualification ensures strong duality for (P) and (D), see, e.g., Geoffrion (1971). Since in nonconvex programming this condition turns out to be very useful as well, recall that problem (P) is said to fulfill Slater's condition if there exists a point $\hat{x} \in X$ such that $h_i(\hat{x}) < 0$ for all $i = 1, \dots, m$.

In the sequel, we will use the notation

$$\mathcal{S} := \{\hat{x} \in X : h_i(\hat{x}) < 0 \text{ for all } i = 1, \dots, m\}$$

to denote the set of all Slater points,

$$\bar{h}(x) := \max_{i=1,\dots,m} h_i(x)$$

to denote the pointwise maximum of the constraint functions,

$$L(x, \lambda) := f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$$

to denote the Lagrangean function of (P), and

$$\Theta(\lambda) := \min_{x\in X} L(x, \lambda)$$

to denote the dual objective function.

The following result shows that the supremum of the dual problem is attained at some finite point, provided that $\mathcal{S} \neq \emptyset$.

THEOREM 3. *Assume the functions f and h_i $(i = 1, \ldots, m)$ in (P) are lower semicontinuous and the set X is compact. Assume further Slater's condition to be fulfilled for (P). Then*

$$\sup(D) = \max(D),$$

and the maximum of the dual objective function $\Theta(\lambda)$ is attained at some $\bar{\lambda} \in \mathbb{R}^m_+$ with

$$\|\bar{\lambda}\|_1 \leqslant \inf_{\hat{x} \in \mathcal{S}} \frac{\min_{x \in X} f(x) - f(\hat{x})}{\bar{h}(\hat{x})}.$$

Proof. For any $\hat{x} \in \mathcal{S}$ and $\lambda \in \mathbb{R}^m_+$, we have

$$\begin{aligned}\Theta(\lambda) &= \min_{x \in X} \left\{ f(x) + \sum_{i=1}^m \lambda_i h_i(x) \right\} \\ &\leqslant f(\hat{x}) + \sum_{i=1}^m \lambda_i \bar{h}(\hat{x}) \\ &= f(\hat{x}) + \bar{h}(\hat{x}) \sum_{i=1}^m \lambda_i \\ &= f(\hat{x}) + \bar{h}(\hat{x}) \|\lambda\|_1.\end{aligned}$$

Now define a number $\rho(\hat{x})$, which is easily seen to be nonnegative (note that $\bar{h}(\hat{x}) < 0$):

$$\rho(\hat{x}) := \frac{\sup(D) - f(\hat{x})}{\bar{h}(\hat{x})}.$$

Then for every $\lambda \in \mathbb{R}^m_+$ fulfilling $\|\lambda\|_1 > \rho(\hat{x})$, we get a dual objective value which is strictly smaller than the optimal one:

$$\begin{aligned}\Theta(\lambda) &\leqslant f(\hat{x}) + \bar{h}(\hat{x}) \|\lambda\|_1 \\ &< f(\hat{x}) + \bar{h}(\hat{x}) \rho(\hat{x}) \\ &= \sup(D).\end{aligned}$$

As Θ is concave and hence continuous, it follows that

$$\begin{aligned}\sup(D) &= \sup \left\{ \Theta(\lambda) : \lambda \in \mathbb{R}^m_+, \|\lambda\|_1 \leqslant \rho(\hat{x}) \right\} \\ &= \max \left\{ \Theta(\lambda) : \lambda \in \mathbb{R}^m_+, \|\lambda\|_1 \leqslant \rho(\hat{x}) \right\} \\ &= \Theta(\bar{\lambda}).\end{aligned}$$

Finally, using

$$\sup(D) = \sup_{\lambda \in \mathbb{R}^m_+} \Theta(\lambda) \geqslant \Theta(0) = \min_{x \in X} f(x),$$

we obtain

$$\rho(\hat{x}) = \frac{\sup(D) - f(\hat{x})}{\bar{h}(\hat{x})} \leqslant \frac{\min_{x \in X} f(x) - f(\hat{x})}{\bar{h}(\hat{x})}.$$

Obviously, $\|\bar{\lambda}\|_1 \leqslant \rho(\hat{x})$ for any Slater point $\hat{x} \in \mathcal{S}$. Hence we get the desired upper bound for $\|\bar{\lambda}\|_1$:

$$\|\bar{\lambda}\|_1 \leqslant \inf_{\hat{x} \in \mathcal{S}} \frac{\min_{x \in X} f(x) - f(\hat{x})}{\bar{h}(\hat{x})}.$$

□

This result seems interesting in its own right, but it may also prove useful in a numerical context: Note that any Slater point $\hat{x}$ gives the a priori bound

$$\|\bar{\lambda}\|_1 \leqslant \frac{\min_{x \in X} f(x) - f(\hat{x})}{\bar{h}(\hat{x})},$$

which may be helpful when solving dual problems with bundle–type methods. Of course, finding a Slater point is a difficult task in general, but may be easy when the constraints are simple, e.g. box constraints. This reasoning also applies to the so called standard quadratic problem of maximizing an indefinite quadratic form on the standard simplex, see Bomze (1998).

3. When are dual bounds better?

In this section we return to the question which of the bounds $\min(\bar{P})$ and $\sup(D)$ is better. Theorem 4 states under which assumptions on objective and constraint functions the dual bound beats the convex envelope bound.

But first observe that it may happen that $\min(P) = \min(\bar{P})$. In this case, it follows from (1) and weak duality that

$$\sup(D) = \min(\bar{P}) = \min(P),$$

in other words, the duality gap is zero and both bounds are equal. For this reason, the mentioned case is excluded in the theorem.

THEOREM 4. *In problem (P), let* $f : \mathbb{R}^n \to \mathbb{R}$ *be strictly concave, let* $h_i(x) : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \ldots, m)$ *be strictly convex and continuously differentiable, let* $X \subset \mathbb{R}^n$ *be convex and compact. Assume that Slater's condition is fulfilled for (P),* *and that* $\min(P) > \min(\bar{P})$*. Assume further that the convex envelope* φ_f *of* f *on* X *is not constant on any interval contained in* X*. Then the dual bound is strictly better than the convex envelope bound, i.e.*

$$\sup(D) > \min(\bar{P}).$$

Proof. Let $\bar{x}$ be a minimizer of problem $(\bar{P})$, and denote by $I(\bar{x})$ the set of indices of the constraints which are active at $\bar{x}$, i.e.

$$I(\bar{x}) = \{i \in \{1, \ldots, m\} : h_i(\bar{x}) = 0\}.$$

It follows from the assumption $\min(P) > \min(\bar{P})$ that there exists a solution $\bar{x}$ with $I(\bar{x}) \neq \emptyset$: If $I(\bar{x}) = \emptyset$ for all solutions $\bar{x}$, then all minimizers of $(\bar{P})$ would fulfill Slater's condition. As φ_f is convex, all those minimizers would also solve $\min_{x \in X} \varphi_f(x)$. Since

$$\arg\min_{x \in X} \varphi_f(x) \supset \arg\min_{x \in X} f(x), \tag{2}$$

all minimizers of $\min_{x \in X} f(x)$ would also fulfill Slater's condition and would therefore minimize (P). It would follow that $\min(P) = \min(\bar{P})$.

So choose an $\bar{x}$ with $I(\bar{x}) \neq \emptyset$ and define a partly linearized version $(\tilde{P})$ of $(\bar{P})$ as follows:

$$(\tilde{P}) \quad \begin{array}{lll} \min \varphi_f(x) & & \\ \text{s.t. } \langle x - \bar{x}, \nabla h_i(\bar{x}) \rangle \leqslant 0 & & i \in I(\bar{x}), \\ \qquad\qquad\quad h_i(x) \leqslant 0 & & i \notin I(\bar{x}), \\ x \in X. & & \end{array}$$

From Theorem 5 in the appendix we get $\min(\bar{P}) = \min(\tilde{P})$, from strong duality (problem $(\tilde{P})$ also fulfills Slater's qualification) we get $\min(\tilde{P}) = \sup(\tilde{D})$. What remains to show is that $\sup(\tilde{D}) < \sup(D)$.

For abbreviation, let $\Lambda(\bar{x})$ denote the set of all $\lambda \in \mathbb{R}^n_+$ which fulfill $\lambda_i \neq 0$ for at least one index $i \in I(\bar{x})$. The Lagrangean $\tilde{L}(x, \lambda)$ of $(\tilde{P})$ then fulfills for $\lambda \in \Lambda(\bar{x})$ and $x \neq \bar{x}$:

$$\begin{aligned} \tilde{L}(x, \lambda) &= \varphi_f(x) + \sum_{i \in I(\bar{x})} \lambda_i \langle x - \bar{x}, \nabla h_i(\bar{x}) \rangle + \sum_{i \notin I(\bar{x})} \lambda_i h_i(x) \\ &\leqslant f(x) + \sum_{i \in I(\bar{x})} \lambda_i \langle x - \bar{x}, \nabla h_i(\bar{x}) \rangle + \sum_{i \notin I(\bar{x})} \lambda_i h_i(x) \\ &< f(x) + \sum_{i \in I(\bar{x})} \lambda_i h_i(x) + \sum_{i \notin I(\bar{x})} \lambda_i h_i(x) \qquad (3) \\ &= L(x, \lambda). \end{aligned}$$

Inequality (3) holds because for $x \neq \bar{x}$ we have $h_i(x) > \langle x - \bar{x}, \nabla h_i(\bar{x}) \rangle$, as all h_i are strictly convex.

Since $\varphi_f(\bar{x}) < f(\bar{x})$ (recall that $\min(\bar{P}) < \min(P)$ by assumption), we obviously have $\tilde{L}(\bar{x}, \lambda) < L(\bar{x}, \lambda)$ for every $\lambda \in \mathbb{R}^m_+$, and hence

$$\tilde{L}(x, \lambda) < L(x, \lambda) \qquad \forall\, x \in X, \quad \forall \lambda \in \Lambda(\bar{x}).$$

Therefore, we get for the dual objective function $\tilde{\Theta}(\lambda)$ of $(\tilde{P})$

$$\tilde{\Theta}(\lambda) = \min_{x \in X} \tilde{L}(x, \lambda) < \min_{x \in X} L(x, \lambda) = \Theta(\lambda) \qquad \forall \lambda \in \Lambda(\bar{x}).$$

Next we show that $\max\limits_{\lambda \in \mathbb{R}^m_+} \tilde{\Theta}(\lambda)$ is attained at some $\tilde{\lambda} \in \Lambda(\bar{x})$:

Let $\tilde{\lambda}$ denote a solution of $(\tilde{D})$, i.e., $\sup(\tilde{D}) = \tilde{\Theta}(\tilde{\lambda})$, and recall that $\bar{x}$ is the optimal solution of $(\tilde{P})$. Since $(\tilde{P})$ is a convex problem, the optimal primal dual pair $(\bar{x}, \tilde{\lambda})$ fulfills the complementary slackness condition (see Geoffrion (1971))

$$\sum_{i \in I(\bar{x})} \tilde{\lambda}_i \langle \bar{x} - \bar{x}, \nabla h_i(\bar{x}) \rangle + \sum_{i \notin I(\bar{x})} \tilde{\lambda}_i h_i(\bar{x}) = 0.$$

It follows that $\tilde{\lambda}_i = 0$ for all $i \notin I(\bar{x})$.

But from the assumptions of the theorem it follows that $\tilde{\lambda} \neq 0$: Assume that $\tilde{\lambda} = 0$. Then, since $(\bar{x}, \tilde{\lambda})$ is a saddle point of $\tilde{L}(x, \lambda)$, it follows that $\tilde{L}(\bar{x}, 0) \leqslant \tilde{L}(x, 0)$ for all $x \in X$, in other words, $\varphi_f(\bar{x}) \leqslant \varphi_f(x)$ for all $x \in X$, and therefore

$$\varphi_f(\bar{x}) = \min_{x \in X} \varphi_f(x).$$

As $\min_{x \in X} \varphi_f(x) = \min_{x \in X} f(x)$, we have $\varphi_f(\bar{x}) = \min_{x \in X} f(x)$. Since f is a concave function, the minimum of f over X is attained at some extremal point of X. Therefore, either $\bar{x}$ is an extreme point. Then $\varphi_f(\bar{x}) = f(\bar{x})$ and $\bar{x}$ would solve both $(\bar{P})$ and (P), a contradiction.

Or there exist $k \leqslant n+1$ extremal points $v^1, \ldots, v^k$ of X, such that $\bar{x}$ is a convex combination of these extremal points and $f(v^j) = \varphi_f(\bar{x})$. But then φ_f would be constant on the convex hull of $\{v^1, \ldots, v^k\}$, which contradicts the assumptions as well.

Therefore, we conclude that $\tilde{\lambda} \in \Lambda(\bar{x})$.

To sum up, let $\bar{\rho}$ denote the maximum of the $\|\lambda\|_1$–bounds obtained via Theorem 3 for problems (P) and $(\tilde{P})$, respectively, and get

$$\begin{aligned} \sup(\tilde{D}) &= \max \left\{ \tilde{\Theta}(\lambda) : \lambda \in \mathbb{R}^m_+, \|\lambda\|_1 \leqslant \bar{\rho}, \lambda \in \Lambda(\bar{x}) \right\} \\ &< \max \left\{ \Theta(\lambda) : \lambda \in \mathbb{R}^m_+, \|\lambda\|_1 \leqslant \bar{\rho}, \lambda \in \Lambda(\bar{x}) \right\} \\ &\leqslant \max \left\{ \Theta(\lambda) : \lambda \in \mathbb{R}^m_+, \|\lambda\|_1 \leqslant \bar{\rho} \right\} \\ &= \sup(D). \end{aligned}$$

This completes the proof. □

In Theorem 4, instead of requiring that φ_f is not constant on any interval contained in X, it is also possible to impose any other assumption which guarantees that there exists a dual optimal solution $\tilde{\lambda} \neq 0$, e.g., the assumption that $\varphi_f(\bar{x}) \neq \min_{x \in X} \varphi_f(x)$.

Theorem 4 may be generalized to the setting of nondifferentiable strictly convex constraint functions. The gradients $\nabla h_i(\bar{x})$ must then be replaced with subgradients of the functions h_i at the point $\bar{x}$ and the proof becomes more technical.

The Theorem tells us that, in the special setting described there, dual bounds should be preferred to convex envelope bounds. It would, of course, be nice to have a measure of the difference of the two bounds, in order to know how much can be gained. Maybe such an estimate is possible for special instances, e.g. for concave quadratic functions. If this difference turns out to be very high (as it was in Example 2) this could be numerically interesting, because it must be admitted that in many cases the dual bound is more difficult to compute than the convex envelope bound or some other bound using subfunctionals of f. However, this quantification of the bound improvement is left to future research.

Appendix

The following technical result is used in the proof of Theorem 4. In can be generalized to the case of nondifferentiable constraint functions in the same way as pointed out above by substituting gradients with subgradients.

THEOREM 5. *Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be convex, let $h_i(x) : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \ldots, m)$ be strictly convex and continuously differentiable, let $X \subset \mathbb{R}^n$ be convex and compact. Let $\bar{x} \in \mathbb{R}^n$ be a solution to the problem*

$$\min\{\varphi(x) : x \in X, h_i(x) \leqslant 0 \ (i = 1, \ldots, m)\} \tag{4}$$

and assume Slater's condition holds for this problem. Denote by $I(\bar{x}) \subseteq \{1, \ldots, m\}$ the set of indices of the constraints active at $\bar{x}$. Then $\bar{x}$ is also a solution to the problem

$$\begin{aligned} \min\{\varphi(x) : x \in X, \ & \langle x - \bar{x}, \nabla h_i(\bar{x})\rangle \leqslant 0 \ \textit{for } i \in I(\bar{x}), \\ & h_i(x) \leqslant 0 \ \textit{for } i \notin I(\bar{x})\}. \end{aligned} \tag{5}$$

Proof. Let $\tilde{x}$ denote the minimizer of problem (5). Clearly, the feasible set of problem (4) is contained in that of problem (5), hence

$$\varphi(\bar{x}) \geqslant \varphi(\tilde{x}).$$

Now assume that $\varphi(\bar{x}) > \varphi(\tilde{x})$. Since $\bar{x}$ is optimal for problem (4), there does not exist a feasible descent direction of φ at $\bar{x}$, i.e. there does not exist a direction d with

$$\langle d, \nabla h_i(\bar{x})\rangle < 0 \quad \text{and} \quad \varphi'_d(\bar{x}) < 0, \tag{6}$$

where φ'_d denotes the directional derivative of φ in direction d.

Because of the strict convexity of all constraint functions and because of Slater's condition, we can assume that there exists a point $\check{x}$ feasible for (5) such that

$$\langle \check{x} - \bar{x}, \nabla h_i(\bar{x}) \rangle < 0 \quad \text{for all} \quad i \in I(\bar{x}),$$

and $\varphi(\check{x}) < \varphi(\bar{x})$. But then $d := \check{x} - \bar{x}$ is a feasible descent direction of φ at $\bar{x}$, since it fulfills conditions (6). This contradicts the optimality assumption on $\bar{x}$. □

References

Bomze, I.M. (1998), On Standard Quadratic Optimization Problems. *Journal of Global Optimization*, 13: 369–387.

Dür, M., Dual Bounding Procedures Lead to Convergent Branch-and-Bound Algorithms. Forthcoming in Mathematical Programming, 2001.

Dür, M. and Horst, R. (1997), Lagrange–Duality and Partitioning Techniques in Nonconvex Global Optimization. *Journal of Optimization Theory and Applications*, 95: 347–369.

Falk, J.E. (1969), Lagrange Multipliers and Nonconvex Programs. *SIAM Journal on Control*, 7: 534–545, 1969.

Geoffrion, A.M. (1971), Duality in Nonlinear Programming: A Simplified Application–Oriented Development. *SIAM Review*, 13: 1–37.

Horst, R. and Tuy, H. (1996), *Global Optimization*. Springer, Berlin.

Nowak, I. (2000), Dual Bounds and Optimality Cuts for All–Quadratic Programs with Convex Constraints. Journal of Global Optimization 18: 337–356.

Thoai, N.V. On Convergence and application of a Decomposition Method Using Duality Bounds for Nonconvex Global Optimization. Forthcoming in *Journal of Optimization Theory and Applications*, 2001.