INFORMS Optimization Society
2020 Student Prize

Expository Article:
Duality Gap Estimation via a Refined Shapley-Folkman Lemma

Yingjie Bi
Department of Industrial Engineering and Operations Research
University of California, Berkeley

Ao Tang
School of Electrical and Computer Engineering
Cornell University

Published online by the INFORMS Optimization Society on March 2, 2021.
http://connect.informs.org/optimizationsociety/home

I am deeply honored to have been awarded the 2020 INFORMS Optimization Society Student Paper Prize for the paper *Duality Gap Estimation via A Refined Shapley–Folkman Lemma* (joint with my advisor Ao Tang). I would like to thank the committee members Fatma Kilinc-Karzan, Maryam Fazel, Daniel Kuhn and Andrea Lodi for their effort and appreciation of our work. This article surveys the major results and applications in our winning work [1].

1 Introduction

We consider nonconvex optimization problems with separable objective and separable constraints, which provide the framework for many important problems in fields such as communication [2] and machine learning [3, 4]. It is of particular interest to estimate the duality gap for such problems, mainly because of its relation to approximation algorithms based on dual methods.

In this article, for simplicity we focus on separable problems with linear constraints, whose general formulation is given as follows:

$$\begin{aligned} \min \quad & \sum_{i=1}^{n} f_i(x^i) \\ \text{s.t.} \quad & \sum_{i=1}^{n} A_i x^i \leq b. \end{aligned} \tag{1}$$

Here $x^i \in \mathbb{R}^{n_i}$ are the decision variables. The function $f_i : \mathbb{R}^{n_i} \to (-\infty, +\infty]$ is lower semicontinuous, and its domain is bounded. A_i is a matrix of size $m \times n_i$, so there are m linear constraints in total. The reader could refer to Section 6 of our paper [1] for the generalization of our result to the cases with separable nonlinear constraints (not necessarily convex).

The Lagrange dual problem of (1) is

$$\begin{aligned} \max \quad & -\sum_{i=1}^{n} f_i^*(-A_i^T y) - b^T y \\ \text{s.t.} \quad & y \geq 0, \end{aligned} \tag{2}$$

where f_i^* is the conjugate function of f_i. We always assume the feasibility on the primal problem (1). Furthermore, under our assumptions on the functions f_i, the dual problem (2) is guaranteed to be feasible, and we denote the optimal value of the primal problem (1) and dual problem (2) as p and d, respectively. In general, there will be a positive duality gap $p - d > 0$ if some function f_i is not convex.

The authors of [5] presented the following upper bound for the duality gap of (1):

$$p - d \leq \min\{m+1, n\} \max_{i=1,\ldots,n} \rho(f_i). \tag{3}$$

Here $\rho(f)$ is the nonconvexity of a function f defined by

$$\rho(f) = \sup \left\{ f\left(\sum_j \alpha_j x^j\right) - \sum_j \alpha_j f(x^j) \right\} \tag{4}$$

over all finite convex combinations of points $x^j \in \operatorname{dom} f$, i.e., $f(x^j) < +\infty$, $\alpha_j \geq 0$ with $\sum_j \alpha_j = 1$.

In [6], an improved bound for the duality gap was given by

$$p - d \leq \sum_{i=1}^{\min\{m,n\}} \rho(f_i), \tag{5}$$

where we assume that $\rho(f_1) \geq \cdots \geq \rho(f_n)$. Although the bound (5) is only a slight improvement over the original bound (3) by a factor of $m/(m+1)$, it nevertheless shows that (3) can never be tight except for some trivial situations. But as will be demonstrated by the examples in Section 3, the bound (5) can still be very conservative.

In our paper [1], we presented a tighter upper bound for the duality gap of (1). Compared with the previous bounds, the improvements of our result are attributed to two sources. First, instead of using a single number measurement, a series of numbers are introduced to characterize the nonconvexity of a function in a finer manner. Second, for a separable nonconvex problem, instead of approximating each subproblem individually, we consider all of them jointly. To combine the above two ideas, a refined version of the Shapley–Folkman lemma has been proposed (see Section 2 of our paper [1]). This new lemma is the key tool used in our proof of the bound and may have its own independent interest.

In Section 2 below, we summarize the major results in our paper [1]. Next, in Section 3, we apply our result to two examples, a network utility maximization problem in networking and the dynamic spectrum management problem in communication, to demonstrate the effectiveness of our new bound in practical problems.

2 Main Results

To improve the bound (5), some finer characterization of the nonconvexity of a function has to be introduced. Define the *k-th nonconvexity* $\rho^k(f)$ of a function f to be the supremum in (4) taken over the convex combinations of k points instead of an arbitrary number of points. These numbers $\rho^k(f)$ satisfy the obvious inequality

$$0 = \rho^1(f) \leq \rho^2(f) \leq \cdots \leq \rho(f),$$

and it is not hard to prove that $\rho^{n+1}(f) = \rho(f)$ if the function f has n variables (see Proposition 3.1 of our paper [1]).

Based on the above definition, our new bound for the duality gap is stated in the following theorem:

Theorem 1. *Assume that the primal problem* (1) *is feasible, i.e., $p < +\infty$. Then there exist integers $1 \leq k_i \leq m+1$ such that $\sum_{i=1}^{n} k_i \leq m+n$ and the duality gap*

$$p - d \leq \sum_{i=1}^{n} \rho_i^{k_i}.$$

Here $\rho_i^k = \rho^k(f_i)$ is the k-th nonconvexity of function f_i.

From a computational viewpoint, since we do not know the k_i that appear in Theorem 1, in order to find a number for the bound, we have to find the worst case k_i by solving the following optimization problem:

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} \rho_i^{k_i} \\ \text{s.t.} \quad & 1 \leq k_i \leq m+1,\, k_i \in \mathbb{Z}, \quad \forall i = 1, \ldots, n, \\ & \sum_{i=1}^{n} k_i \leq m+n. \end{aligned} \tag{6}$$

Let B be the optimal value of (6); then

$$B \leq \sum_{i=1}^{n} \rho(f_i).$$

On the other hand, since for any feasible solution of (6), the number of k_i with $k_i \geq 2$ is bounded by m,

$$B = \sum_{i:k_i \geq 2} \rho_i^{k_i} \leq \sum_{i=1}^{m} \rho(f_i)$$

if $\rho(f_1) \geq \cdots \geq \rho(f_n)$. The above argument shows that the bound B given by the optimization problem (6) is at least as tight as the bound (5) in [6].

Generally, solving the optimization problem (6) directly to obtain B is not an easy task. However, as demonstrated in Section 3, in many applications, (6) can be further relaxed to obtain an upper bound for the duality gap that is still much tighter than the previous bound (5).

3 Applications

In this section, we provide two examples to illustrate how our new bound in Theorem 1 can be applied in engineering problems and demonstrate the superiority of our bound over the previous results.

3.1 Network Utility Maximization

In this part, we will first apply our result to the network utility maximization problem. Consider a network with N users and L links. Let a strictly positive vector $c \in \mathbb{R}^L$ contain the capacity of each link. Each user i has K^i available paths to send its commodity. We assume that the users are sorted such that $K^1 \geq \cdots \geq K^N$. The routing matrix of user i, denoted by R^i, is an $L \times K^i$ matrix with $R^i_{lk} = 1$ if the k-th path of user i passes through link l and $R^i_{lk} = 0$ otherwise.

Let $x^i \in \mathbb{R}^{K^i}$ be the vector in which x^i_k is the amount of commodity sent by user i on its k-th path. Assume that each user i has a utility function $U_i(\cdot)$ depending on the vector x^i, then the network utility maximization problem can be written as

$$\begin{aligned} \max \quad & \sum_{i=1}^{N} U_i(x^i) \\ \text{s. t.} \quad & \sum_{i=1}^{N} R^i x^i \leq c, \\ & x^i \geq 0, \quad \forall i = 1, \ldots, N. \end{aligned} \tag{7}$$

If all the utility functions $U^i(\cdot)$ are concave, then the above problem (7) can be solved by standard convex optimization techniques. Difficulty arises when $U^i(\cdot)$ is not concave. For example, if we restrict each user to choose only one path (single-path routing) and want to maximize the total throughput of the network, then the corresponding utility function is

$$U_i(x^i) = \max_{s=1,\ldots,K^i} x^i_s.$$

Define

$$f_i(x^i) = \begin{cases} \min_{s=1,\ldots,K^i}(-x^i_s), & \text{if } 0 \le x^i \le \|c\|_\infty, \\ +\infty, & \text{otherwise.} \end{cases} \tag{8}$$

Here $\|c\|_\infty$ is the maximum link capacity in the network. Now the original network utility maximization problem (7) is equivalent to the following problem:

$$\begin{aligned} \min \quad & \sum_{i=1}^{N} f_i(x^i) \\ \text{s. t.} \quad & \sum_{i=1}^{N} R^i x^i \le c. \end{aligned} \tag{9}$$

The above problem is a particular case of the general optimization problem with separable objectives (1) studied in this paper. The k-th nonconvexity of the functions f_i can be computed as shown in Example 3.4 of our paper [1]:

$$\rho^k(f_i) = \frac{k-1}{k}\|c\|_\infty, \quad k = 1, \ldots, K^i,$$

$$\rho^{K^i+1}(f_i) = \rho(f_i) = \frac{K^i - 1}{K^i}\|c\|_\infty.$$

In the following, suppose each user has a large number of paths to select. More explicitly, $K^i \ge L+1$ is assumed for user i. Based on the bound (5), the duality gap is bounded by

$$\sum_{i=1}^{\min\{N,L\}} \frac{K^i - 1}{K^i}\|c\|_\infty,$$

which is at least

$$\min\{N, L\}\frac{L}{L+1}\|c\|_\infty.$$

In contrast, by Theorem 1, the duality gap is bounded by the optimal value of the following optimization problem:

$$\begin{aligned} \max \quad & \sum_{i=1}^{N} \frac{k_i - 1}{k_i}\|c\|_\infty \\ \text{s. t.} \quad & 1 \le k_i \le L+1,\, k_i \in \mathbb{Z}, \quad \forall i = 1, \ldots, N, \\ & \sum_{i=1}^{N} k_i \le N + L. \end{aligned} \tag{10}$$

Let N' be the number of users whose $k_i \ge 2$; then $0 \le N' \le \min\{N, L\}$. If $N' > 0$, using the inequality between arithmetic mean and harmonic mean,

$$\begin{aligned} \sum_{i=1}^{N} \frac{k_i - 1}{k_i} &= \sum_{i:k_i \ge 2} \frac{k_i - 1}{k_i} = N' - \sum_{i:k_i \ge 2} \frac{1}{k_i} \\ &\le N' - \frac{N'^2}{\sum_{i:k_i\ge 2} k_i} \le N' - \frac{N'^2}{N' + L} \\ &= \frac{L}{1 + L/N'} \le \min\{N, L\}\frac{L}{L + \min\{N, L\}}. \end{aligned}$$

The above analysis provides an upper bound for the problem (10), which in turn is an upper bound for the duality gap. Taking the $N \geq L$ case as an example, by the above inequality, we can bound the duality gap by $L\|c\|_\infty/2$, essentially half of the bound given by (5). The same result was obtained by a specialized technique in [7].

Next, we consider another case in which each user has logarithmic utility but still must choose only one path. The utility function of user i can be written as

$$U_i(x^i) = \log \max_{s=1,\dots,K^i} x^i_s.$$

Define functions g_i similar to the f_i defined in (8) above, whose k-th nonconvexity is given by

$$\begin{aligned}\rho^k(g_i) &= \log k, \quad k = 1, \dots, K^i,\\ \rho^{K^i+1}(g_i) &= \rho(g_i) = \log K^i.\end{aligned}$$

Then the network utility maximization problem (7) is equivalent to the problem obtained by replacing f_i with g_i in (9).

Applying the bound (5) to this case, we can bound the duality gap by

$$\sum_{i=1}^{\min\{N,L\}} \log K^i, \tag{11}$$

which is at least $\min\{N, L\} \log(L + 1)$. On the other hand, by Theorem 1, the duality gap is bounded by the optimal value of the following optimization problem

$$\begin{aligned}\max \quad & \sum_{i=1}^{N} \log k_i\\ \text{s.t.} \quad & 1 \leq k_i \leq L + 1,\, k_i \in \mathbb{Z}, \quad \forall i = 1, \dots, N,\\ & \sum_{i=1}^{N} k_i \leq N + L.\end{aligned} \tag{12}$$

If we still let N' be the number of users whose $k_i \geq 2$, then $0 \leq N' \leq \min\{N, L\}$ and the above bound

$$\begin{aligned}\sum_{i=1}^{N} \log k_i &= \sum_{i:k_i\geq 2} \log k_i\\ &= \log \prod_{i:k_i\geq 2} k_i \leq \log \left(\frac{\sum_{i:k_i\geq 2} k_i}{N'}\right)^{N'}\\ &\leq \log \left(\frac{N' + L}{N'}\right)^{N'}\\ &\leq \min\{N, L\} \log \left(1 + \frac{L}{\min\{N, L\}}\right),\end{aligned}$$

where in the last step the monotonicity of the function $(1 + 1/x)^x$ is used. Note that the new bound is qualitatively tighter than the bound (11) provided by (5) by removing a logarithm factor of $O(\log L)$ when $N \geq L$.

3.2 Dynamic Spectrum Management

Consider a communication system consisting of L users sharing a common band. The band is divided equally into N tones. Each user l has a power budget p_l, which can be allocated across all the tones. Let x_l^i be the power of user l allocated on tone i. Due to the crosstalk interference between users, the total noise for a user on tone i is the sum of a background noise σ_i and the power of all other users on the same tone. Therefore, the achievable transmission rate of user l on tone i is given by

$$u_l^i = \frac{1}{N} \log \left(1 + \frac{x_l^i}{\|x^i\|_1 - x_l^i + \sigma_i} \right).$$

The dynamic spectrum management problem is to maximize the total throughput of all users under the power budget constraints, which can be formulated as the following nonconvex optimization problem:

$$\begin{aligned} \max \quad & \sum_{l=1}^{L} \sum_{i=1}^{N} u_l^i \\ \text{s. t.} \quad & \sum_{i=1}^{N} x_l^i \le p_l, \quad \forall l = 1, \ldots, L, \\ & x_l^i \ge 0, \quad \forall i = 1, \ldots, N, \, \forall l = 1, \ldots, L. \end{aligned} \tag{13}$$

For simplicity, we assume that the noises $\sigma_i \le 1$ and the power budgets $p_l \le 1$ (if not, then scale all the σ_i and p_l simultaneously). The latter requires all the variables $x_l^i \le 1$. Define

$$h_\sigma(x) = h_\sigma(x_1, \ldots, x_n) = \sum_{s=1}^{n} \log \frac{\|x\|_1 - x_s + \sigma}{\|x\|_1 + \sigma}.$$

The objective function of (13) can be rewritten as a sum of separable objectives:

$$\sum_{l=1}^{L} \sum_{i=1}^{N} u_l^i = -\frac{1}{N} \sum_{i=1}^{N} h_{\sigma_i}(x^i).$$

For the purpose of designing dual algorithms, it is of great interest to estimate the duality gap for the problem (13). In [8], the authors showed that the duality gap will tend to zero if the number of users L is fixed and the number of tones N goes to infinity. [2] further determined the convergence rate of the duality gap to be $O(1/\sqrt{N})$. Using the bound (5), we now demonstrate how to improve the convergence rate estimation to $O(1/N)$, which can be only achieved by the method in [2] in the special case where all the noises σ_i are the same.

As proved in Example 3.6 of our paper [1],

$$\rho^k(h_{\sigma_i}) \le \log \frac{k}{\sigma_i} \le \log \frac{k}{\sigma}, \quad k = 1, \ldots, L+1,$$

where σ is the minimum among all the noises σ_i, so (5) implies that the duality gap is upper bounded by

$$\frac{\min\{N, L\}}{N} \log \frac{L+1}{\sigma}, \tag{14}$$

which is in the order of $O(1/N)$ if L is fixed and N increases.

In order to further improve the estimation (14) for the duality gap, we can resort to Theorem 1 and follow the exact same steps for solving (12), which shows that the duality gap is upper bounded by

$$\frac{\min\{N, L\}}{N} \log \frac{1 + L/\min\{N, L\}}{\sigma}.$$

Like the previous example, our bound is still tighter than the one (14) from (5) by removing a logarithm factor.

4 Conclusion

In this article, we focus on estimating the duality gap without consideration of actually solving the primal problem (1) and present a novel bound for the duality gap estimation. A natural future direction is to design approximate algorithms for the primal problem and analyze the quality of the obtained solution based on our deeper understanding of the nonconvexity achieved by our new tools.

References

[1] Y. Bi and A. Tang, "Duality gap estimation via a refined Shapley–Folkman lemma," *SIAM J. Optim.*, vol. 30, no. 2, pp. 1094–1118, 2020.

[2] Z.-Q. Luo and S. Zhang, "Duality gap estimation and polynomial time approximation for optimal spectrum management," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2675–2689, Jul. 2009.

[3] H. Zhang, J. Shao, and R. Salakhutdinov, "Deep neural networks with multi-branch architectures are intrinsically less non-convex," in *Proc. Mach. Learn. Res.*, vol. 89, Apr. 2019, pp. 1099–1109.

[4] A. Askari, A. d'Aspremont, and L. El Ghaoui, "Naive feature selection: Sparsity in naive Bayes," in *Proc. Mach. Learn. Res.*, vol. 108, Aug. 2020, pp. 1813–1822.

[5] J. P. Aubin and I. Ekeland, "Estimates of the duality gap in nonconvex optimization," *Math. Oper. Res.*, vol. 1, no. 3, pp. 225–245, Aug. 1976.

[6] M. Udell and S. Boyd, "Bounding duality gap for separable problems with linear constraints," *Comput. Optim. Appl.*, vol. 64, no. 2, pp. 355–378, Jun. 2016.

[7] Y. Bi, C. W. Tan, and A. Tang, "Network utility maximization with path cardinality constraints," in *Proc. INFOCOM*, Apr. 2016.

[8] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, Jul. 2006.