

# Zhiliang's Blog and Notes

## Intuition behind Lagrange Duality

📅 2016-09-10

---

We all knew what to do, but don't ask me why. Or you insist?

---

### Mysterious “dual”

Suppose we are going to solve the optimization problem

$$\begin{aligned} & \min_x f(x) \\ & s.t. \ h_i(x) \leq 0, \ i = 1 \dots m \\ & \mathcal{D} = \text{dom } f \cap \bigcap_{i=1}^m \text{dom } h_i \end{aligned}$$

*note, the equation constrains can be represented by two inequation constrains. To simplify our problem, here we use only inequation constrains.*

*Besides, all the functions we discussed in this blog are nicely functions, they are always continuous and derivable, as whatever you need.*

Due to the standard procedure, we firstly construct a Lagrangian function, which can be considered as adding constrains as perturbation terms.

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^T h(x), \text{ with } \lambda \geq 0$$

Then we define a “mysterious” dual function as

$$\begin{aligned} g(\lambda) &= \inf_{x \in \mathcal{D}} (\mathcal{L}(x, \lambda)) \\ &= \inf_{x \in \mathcal{D}} (f(x) + \lambda^T h(x)) \end{aligned}$$

*note, to get a infimum, x is still in the domain  $\mathcal{D}$ , not feasible regions from constrains.*

Finally, by solving the dual problem

$$\max_{\lambda} g(\lambda) \quad s.t. \ \lambda \geq 0$$

We can get a satisfying lower bound of the primal problem.

Here is a short proof about  $g(\lambda)$  is the lower bound of primal problem.

$$\begin{aligned} g(\lambda) &= \inf_{x \in D} (\mathcal{L}(x, \lambda)) \\ &\leq f(x) + \lambda^T h(x) \\ &\leq f(x) \quad \text{if } x \text{ in feasible region} \end{aligned}$$

These are what we learned from text book, surely there are also deduction and proofing.

However it's difficult for me to explain these to other people, when they keep asking why, because I didn't have a clear intuition about this standard procedure. For myself, the most agonizing questions were all about this dual problem. The algorithm is clear, but why we choose to solve the dual problem to estimate the lower bound of primal problem instead of solving primal problem directly. What is the intuition behind this dual problem?

In the following sections, I will try to explain my intuitive understanding. At the same time I will try to avoid the mathematic proofs, which is beautiful and logically perfect but hard for mortal to speak.

Be careful, all the examples in this blog are special cases, which are helpful to understand the intuition, but since they are special cases, their properties can not be simply generalized. If I made any mistakes, please notify me to correct, thank you.

## Why Lagrangian function

Why we need to construct this Lagrangian function ?

The answer is about perturbation

We can not solve the primal problem by just calculating  $\frac{\partial f(x)}{\partial x} = 0$ , because of the feasible region of constraints. One possible solution is adding the constraints as perturbation terms, and then estimate the best lower bound as Eq.2 shows.

Here is an example. The optimization problem is

$$\begin{aligned} \min_x & x^2 \\ \text{s.t. } & x - 1 \leq 0 \\ & D = \mathbb{R} \end{aligned}$$

The optimal solution is easy to guess, which is  $p^* = 0$ , when  $x = 0$ .

The Lagrangian function is

$$\mathcal{L}(x, \lambda) = x^2 + \lambda(x - 1), \text{ with } \lambda \geq 0$$

The dual function is

$$g(\lambda) = \inf_{x \in D} (x^2 + \lambda(x - 1)), \text{ with } \lambda \geq 0$$

consider  $\inf_{x \in D} (x^2 + \lambda(x - 1))$  is a function about  $x$ , the infimum exists only when  $\frac{\partial \mathcal{L}(x, \lambda)}{\partial x} = 0$ .

Therefor we have  $x = -\frac{\lambda}{2}$ . plug this back to the dual function, we have

$$g(\lambda) = -\frac{1}{4}\lambda^2 - \lambda, \text{ with } \lambda \geq 0$$

And the dual problem is

$$\max_{\lambda} g(\lambda) \quad \text{s.t. } \lambda \geq 0$$

And the dual solution is  $d^* = 0$ , with  $\lambda = 0$

Now let's plot the solution with intuition. In Fig.1, the black curve is our original problem, blue and green curve is original problem with perturbations as Lagrange function. We can see the values of  $\inf_{x \in D} \mathcal{L}(x, \lambda)$ , e.g. saddle points on blue and green curve, are always equals or smaller than the optimal solution  $p^*$ , which illustrates  $g(\lambda)$  is a lower bound of  $f(x)$ .

The red curve represents the dual function  $g(\lambda)$ . It is easily to find out, when  $\lambda = 0$ ,  $g(\lambda)$  achieve its maximum, and the  $\max_{\lambda} g(\lambda)$  exactly equals to the  $\min_x f(x)$ .

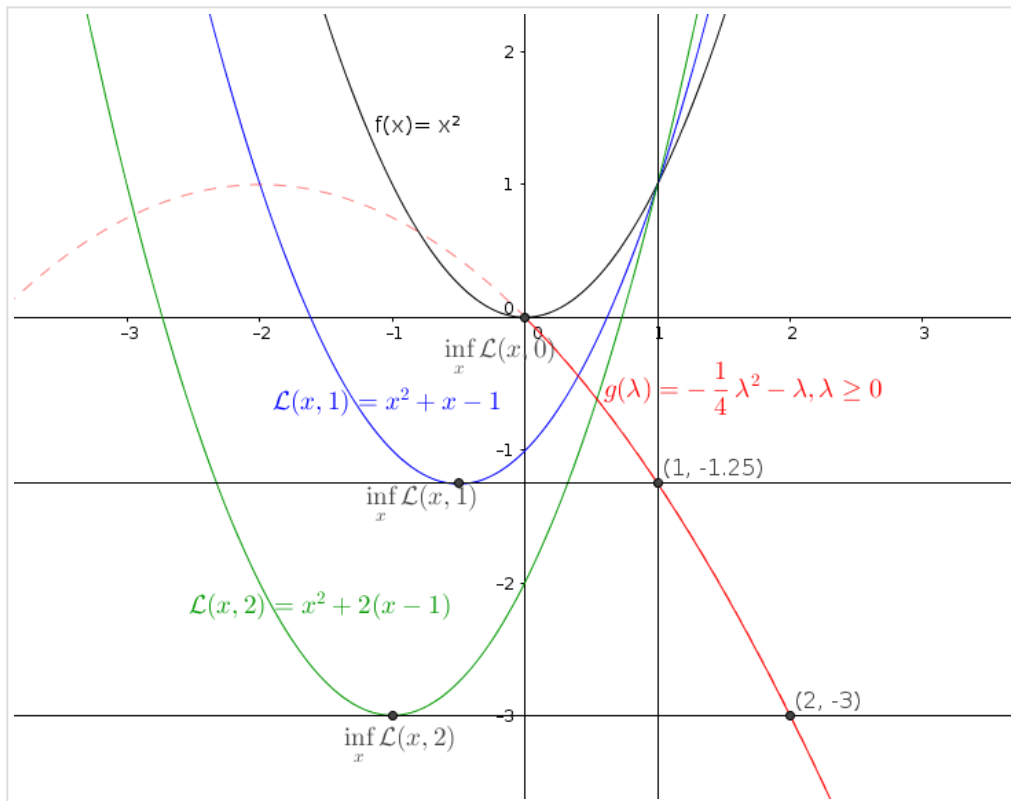


Fig.1 Lagrangian function and dual function. note,  $f(x)$  and  $g(\lambda)$  have independent variables, for a more concise expression, the red curve should be plotted on an additional figure.

This example and figure give us a simple intuition about how the optimization works. But are there any other reasons, that we prefer to find the lower bound by solving its dual problem?

## Why dual problem

Why we choose to solve the dual problem to estimate the lower bound? is it easier to solve?

The answer is YES. It is easier to solve.

In fact dual function  $g(\lambda)$  is concave, or we say  $-g(\lambda)$  is convex.

From Eq.(3), we know  $-g(\lambda)$  is pointwise supremum of a set of affine functions about  $\lambda$ , which means  $-g(\lambda)$  is convex. About this fact, here we skip the mathematic proof, but give an intuitive example in Fig.2.

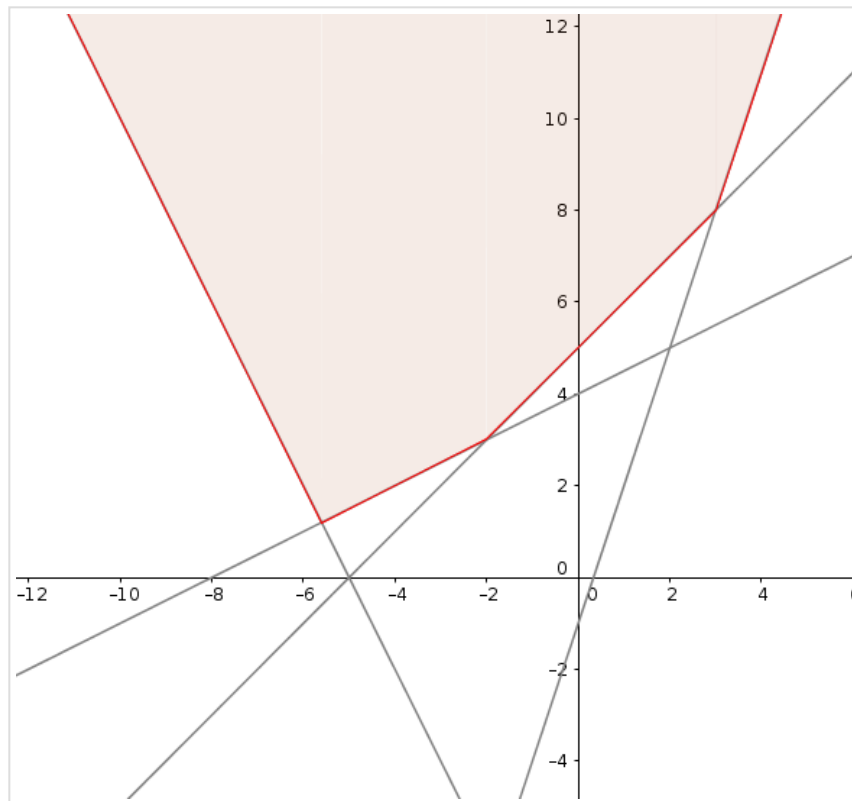


Fig.2 pointwise supremum of the set of affine functions is convex function

Fig.2 shows a set of affine functions, the red lines stands for the pointwise supremum. It is obvious that in this figure the red line function is a convex function. We know for convex function and set, it's easy to solve the global extreme. No matter  $f(x)$  is convex or not, its dual function is concave, which is a convenient property.

## Connection with conjugate function

When the constrains in Eq.(1) are linear, the dual function looks often similar to the conjugate function. In the following parts, I will try explain the connection between conjugate function and lagrange dual functions with my intuitive understanding.

## Conjugate function

The conjugate function of  $f(x)$  is defined as

$$f^*(\lambda) = \sup_x (\lambda^T x - f(x))$$

Here is an example about conjugate function.

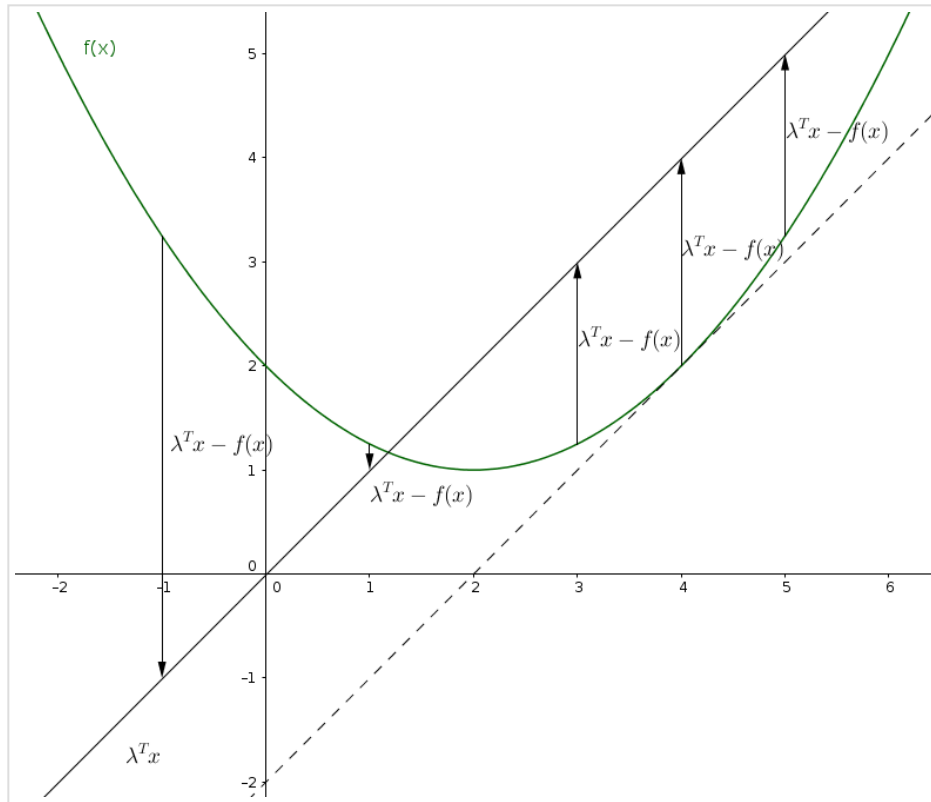


Fig.3 Conjugate function.

As Fig.3 illustrated, the green curve is  $f(x)$ , the black solid line is  $\lambda^T x$ , the vectors (with sign) between  $\lambda^T x$  and  $f(x)$  are  $\lambda^T x - f(x)$ . So the conjugate function  $f^*(\lambda)$  is about the biggest gap between line  $\lambda^T x$  and  $f(x)$ . From Fig.3 we can see, intuitively conjugate function describes lines with slope and intercept, which tangent to  $f(x)$ .

In this example, the maximum of  $\lambda^T x - f(x)$  is the intercept of tangent line of  $f(x)$  with slope  $\lambda$ . More generally, for convex function, the meaning of conjugate function is, given a slope  $\lambda$ , if  $\sup_x (\lambda^T x - f(x))$  exists, then  $g(\lambda)$  is the intercept of the tangent line of  $f(x)$  with slope  $\lambda$ .

Another intuitive thinking is, if  $f(x)$  is “nicely” convex, when conjugate function  $f^*(\lambda)$  is given, which means all the tangent lines of  $f(x)$  are given, the intersection of their epigraph is exactly the epigraph of  $f(x)$ , which means  $f(x)$  is known too.

So conjugate function  $f^*(\lambda)$  is a function about tangent lines, which approximate of  $f(x)$  with tangent lines. Conjugate function is another view of  $f(x)$ .

## When constraints are linear

When constraints are linear, our optimization problem looks like

$$\min_x f(x) \quad s.t. \quad Ax \leq b$$

The Lagrange dual function is

$$\begin{aligned} g(\lambda) &= \inf_x (f(x) + \lambda^T (Ax - b)) && \text{with } \lambda \geq 0 \\ &= -\sup_x (-\lambda^T Ax - f(x)) - \lambda^T b \\ &= -f^*(-\lambda^T A) - \lambda^T b && \text{due to Eq.9} \end{aligned}$$

Intuitively, since  $-\lambda^T b$  is a linear term, the extreme values only are effected by the domain, so if we focus on the  $-f^*(-\lambda^T A)$  term, to solve  $\max_{\lambda} g(\lambda)$  means to solve  $\min f^*(-\lambda^T A)$ .

What is the minimum of a conjugate function? again, let's see an example.

if the optimization problem is

$$\min x^2 \quad s.t. \quad x \leq 1$$

we have the dual function as

$$\begin{aligned} g(\lambda) &= -f^*(-\lambda^T A) - \lambda^T b \\ &= -f^*(-\lambda) - \lambda \end{aligned}$$

here the conjugate function is

$$\begin{aligned} f^*(\lambda) &= \sup_x (\lambda x - f(x)) \\ &= \frac{1}{4} \lambda^2 \end{aligned}$$

The dual problem is defined as

$$\begin{aligned} &\max_{\lambda} g(\lambda) \\ \text{dom } g &= \{-\lambda \in \text{dom } f^*, \lambda \geq 0\} \end{aligned}$$

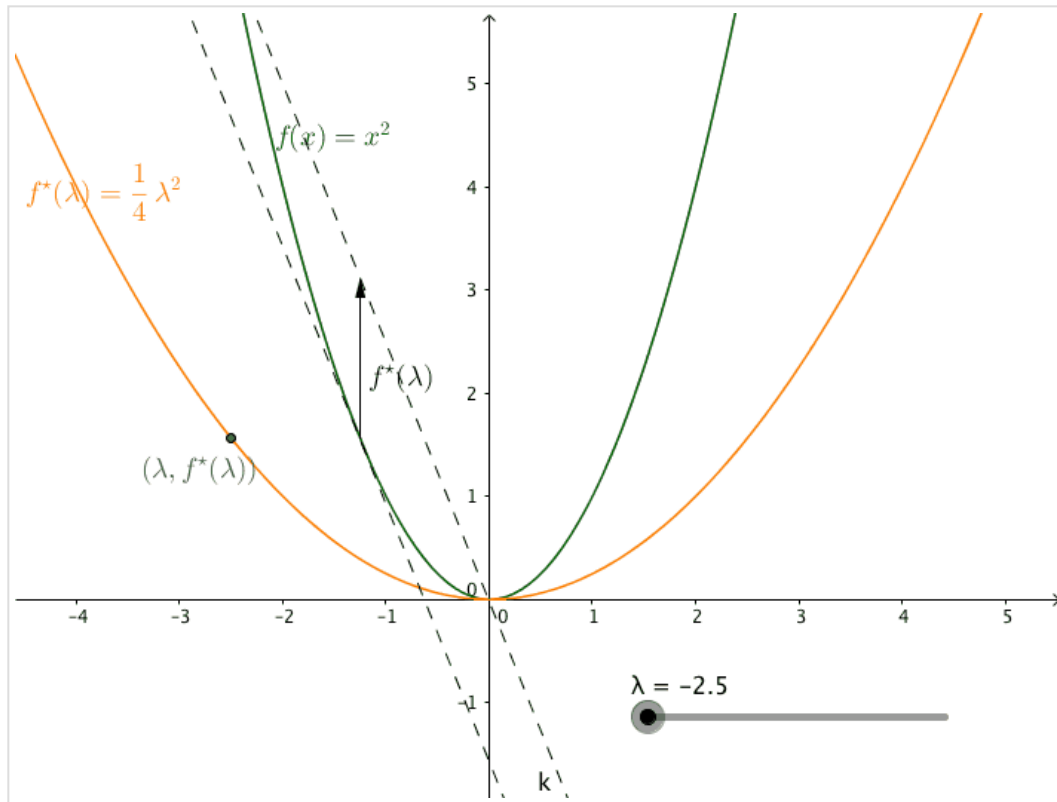
if we plug Eq.11 into Eq.12, we have dual problem like

$$\max_{\lambda} -f^*(-\lambda) - \lambda$$

As we discussed before,  $-\lambda$  is the linear term, the extreme value is only effected by its domain. Now let's consider the extreme of  $-f^*(-\lambda)$ . The problem is equivalent to solve

$$\min_{\lambda} f^*(-\lambda)$$

In Fig.4, we can see, it is obvious when  $\lambda = 0$ , the conjugate function  $f^*(-\lambda)$  is minimized.



Fi.4 minimum of conjugate function. The same as Fig.1,  $f(x)$  and  $f^*(\lambda)$  have independent variables, for a more concise expression, the orange curve should be plotted on an additional figure.

Now lets go back to our questions, conjugate function is another way to describe the original function with support tangent lines. For convex function, the minimum of original function often means the tangent line in conjugate function, which has minimum intercept. *note the intercept has sign here.*

When our optimization problem has linear constrains, the dual function is similar to the form of conjugate function, intuitively solve the dual problem is more likely to search the support tangent line on the extreme point.

## References

1. "Convex Optimization", The Book, you can find every thing you need
2. A helpful discuss thread
3. Proof about conjugate function's properties, in chinese

## Copyright

© Zhiliang Zhou and <https://masszhou.github.io>, 2016.

You are free repost them, but please repost them with source and author's information, i.e. me.

Thanks :)

If you find any mistakes in my blog, please notify me or correct me.

Thanks a lot!:)

# fundamental math

---

[Bayes Inference by Hand](#) ➤

© 2018  Zhiliang Zhou

Powered by [Hexo](#) | Theme — [NexT.Mist](#) v5.1.4