Eliminating Duality Gap by α Penalty

Can KIZILKALE

15 May 2010

Abstract

In this short and self contained paper we are proposing a simple penalty, which is a modification on the constraint set. We show that duality gap is eliminated under certain conditions.

Introduction

We are considering the primal and dual solutions of the following problem.

$$\begin{aligned} &\text{minimize} \quad && f(x) \\ &\text{subject to} \quad && x \in X,\ g_j(x) \leq 0, j = 1, \ldots, r, \end{aligned} \tag{1}$$

where X is a subset of R^n ,and $f : R^n \to R$, $g_j : R^n \to R$ are given functions. We denote $g(x)$ the vector of constraint functions.

$$g(x) = (g_1(x), \ldots, g_r(x)).$$

Since this is a standard formulation I will skip further explanation on it. We will also omit any well known background information such as the definition of duality gap, mc/mc framework etc. A very good treatment of this subject can be found in [1],[2].

The paper is organized as follows: First we will define the α penalty function, assumptions, give the main result and the intuition behind it. The third part will be the proof of our main result. You will see an application of the idea on a toy problem in the fourth part and the final part is the conclusion.

Figure 1: Intution of the penalty

Figure 2: The case where no non-horizontal hyperplane supports M

The α Penalty

We are going to scale the constraint function $g(x)$'s nonnegative part(where it is infeasible) by α and form the new function $g_\alpha(x)$ which we will refer to the α penalty.

$$g_\alpha(x) = \begin{cases} g(x) & \text{for } g(x) \leq 0 \\ \alpha g(x) & \text{for } g(x) > 0 \end{cases} \tag{2}$$

Under the assumptions (1,2) , when we use $g_\alpha(x)$ instead of $g(x)$ the set M gets flatter around point 0 and it can be supported by hyperplanes having smaller(element wise) normal vectors (μ) which causes the value of the dual function to get closer to the optimal value. An intuitive graph showing this idea is given in figure (1). The grey curve belongs to $(f(x), g(x))$ and the dashed red curve to $(f(x), g_\alpha(x))$, while $q_\alpha*$ is closer to $f*$ then $q*$.

Throughout the paper f^* and g^* denote the optimal values for the primal $f(x)$ and dual $g(x)$ functions and (we will assume that we have a duality gap, $f^* > q^*$) we will have the following assumptions.

Assumption 1. *$\exists \mu \in R^r$ such that $\mu > 0$ and $q(\mu)$ attains a finite value.*

Assumption 2. *The set $M = \{(w, u) \in R^{r+1} \| \exists x \text{ such that } w \geq f(x), u \geq g(x)\}$ is closed.*

Assumption (1) guarantees the existence of a non vertical and non horizontal hyperplane. The existence of a non vertical hyperplane guarantees the dual function $q(\mu)$ having a finite value. The assumption of the existence of a non horizontal hyperplane is just as crucial, otherwise the supporting hyperplane is already flat for some $g_i(x)$ and the scaling will have no effect on the duality gap, this problem can be seen in figure (2).
Assumption (2) eliminates the possibility of a gap on the boundary of set M which will result in the supporting hyperplane stucking at a non optimal point, an example can be seen on figure (3). Now that we have given the idea behind the main result we can move on to the actual theorem and its proof.

Figure 3: M is open at $g(x) = 0$, the dual function can not got higher then $q*$

Main Theorem and Proof

We will show that by increasing α we can make the dual optimal value as close to the primal optimal as possible. We will start with showing that the feasible portion of set M can be supported by any flat enough hyperplane(trivial but the result is crucial for the remaining proof)then for such flat hyperplane (small enough μ) we will show that there is a large enough α such set M_α stays on one half-space separated by the same hyperplane.

Proposition 1. *For any q where $f^* > q > q^*$ there exists a $\mu_l > 0$ in R^{r+1} where $\forall x$ such that $g(x) \leq 0$ we have $f(x) + \mu_l^t g(x) \geq q$.*

By definition $\forall x \in X$ such that $g(x) \leq 0$ we have $f(x) \geq f^*$. From assumption 1 we know that $\exists \mu > 0$ where $f(x) + \mu^t g(x) \geq c$ for $\forall x \in X, g(x) \leq 0$ and $c > -\infty$. Now take some $\gamma > 0$ in R. $\forall x \in X$ such that $g(x) \leq 0$ we have

$$\begin{aligned} f(x) + \gamma\mu^t g(x) &= f(x)(1-\gamma) + \gamma(f(x) + \mu^t g(x)) \\ &\geq f(x)(1-\gamma) + \gamma c \\ &\geq f^*(1-\gamma) + \gamma c \end{aligned}$$

where as $\gamma \to 0$, $f^*(1-\gamma) + \gamma c$ gets arbitrarily close to f^* which is greater then q. Then for small enough γ we have

$$f^*(1-\gamma) + \gamma c > q$$

and we conclude that $f(x)+\mu_l^t g(x) > q$ for $\mu_l = \gamma\mu$. Since $\gamma > 0$ we have $\mu_l = \gamma\mu > 0$ this proves proposition 1.

The function $g_\alpha(x)$ can also be written as :

$$g_\alpha(x) = \alpha g^+(x) + g^-(x)$$

, where $\forall j \; g_j^+(x) = max\{g_j(x), 0\}$ and $g_j^-(x) = min\{g_j(x), 0\}$. Let f_α^* denote the optimal primal for the case where exchange $g(x)$ with $g_\alpha x$ in the original problem 1. Likewise let $M_\alpha = \{(w, u) \in R^{r+1} \| \exists x \text{ such that } w \geq f(x), u \geq g_\alpha(x)\}$

Proposition 2. *For $\alpha \geq 1$ we have $f_\alpha^* = f^*$.*

This is trivial, since $g_\alpha(x) = g(x)$ for $g(x) \leq 0$ and $\forall x \in X, g(x) > 0 \Rightarrow g_\alpha(x) = \alpha g(x) > 0$ then the feasible region is the same for the modified problem so the optimal value for the primal problem is the same.

Proposition 3. *For every $\alpha \geq 1$ we have $M_\alpha \supseteq M$.*

The proof is trivial.

Lemma 1. *For α large enough and $\forall q$ such that $f^* > q > q^*$ there exists a μ_q where $f(x) + \mu_q^t g_\alpha(x) \geq q$.*

We will prove this lemma by contradiction. From proposition 1 we know that for q_l such that $q < q_l < f^*$, $\exists \mu_l > 0$ in R^{r+1} where $\forall x$ satisfying $g(x) \leq 0$ we have $f(x) + \mu_l^t g(x) \geq q_l > q$. Assume that for every α there exists an $x \in X$ such that $\exists j, g_j(x) > 0$ and $f(x) + \mu_l^t g_\alpha(x) < q$. Now for $n \in Z^+$ let

$$x_n \in \{x \in X \| f(x) + \mu_l^t g_{\alpha=n}(x) < q\}$$

. Note that such x_n can not be feasible (result of proposition 1).
Now since $g^+(x_n) \geq 0$ and $\mu_l > 0$ then will have $\liminf_{n\to\infty}(\mu_l^t g^+(x_n)) \geq 0$. We will look into two cases for this inequality.

Case 1: Let $\liminf_{n\to\infty}(\mu_l^t g^+_{\alpha=n}(x_n)) > 0$:
Then for large enough N there exists a $c_g > 0$ in R such that $\forall n \geq N$ $\mu_l^t g^+(x_n) \geq c_g$. From assumption 1 we have a finite c such that $\exists \mu \in R^r$ such that $\mu > 0$ and $q(\mu) = c$. Then

$$\begin{aligned}
f(x_n) + \mu^t g(x_n) &\geq c \\
f(x_n) &\geq c - \mu^t g(x_n) \\
f(x_n) + \mu_l^t g_{\alpha=n}(x_n) &\geq c - \mu^t g(x_n) + \mu_l^t g_{\alpha=n}(x_n) \\
&\geq c - \mu^t(g^+(x_n) + g^-(x_n)) + \mu_l^t(ng^+(x_n) + g^-(x_n)) \\
&= c + (n\mu_l - \mu)^t g^+(x_n) + (\mu_l - \mu)^t g^-(x_n)
\end{aligned}$$

Let's remember the proof of proposition 1 for a moment. We chose μ_l as a scaled version (there exist a small enough $1 > \gamma > 0$ such that $\mu_l = \gamma\mu$) of μ. Then for large enough n.

$$\begin{aligned}
c + (n\mu_l - \mu)^t g^+(x_n) + (\mu_l - \mu)^t g^-(x_n) &= c + (n - \tfrac{1}{\gamma})c_g + (\mu_l - \mu)^t g^-(x_n) \\
&= c + (n - \tfrac{1}{\gamma})c_g + (\gamma - 1)\mu^t g^-(x_n)
\end{aligned} \tag{3}$$

Here the term $(\gamma - 1)\mu^t g^-(x_n) \geq 0$ since $\mu > 0 \Rightarrow \mu^t g^-(x_n) \leq 0$ and $(\gamma - 1) < 0$. Then by increasing n we can make $(n - \frac{1}{\gamma})c_g$ hence $c + (n - \frac{1}{\gamma})c_g + (\gamma - 1)\mu^t g^-(x_n)$ arbitrarily large such that it exceeds q, but then for large enough n we get

$$f(x_n) + \mu_l^t g_{\alpha=n}(x_n) \geq c + (n - \frac{1}{\gamma})c_g + (\gamma - 1)\mu^t g^-(x_n) > q \tag{4}$$

which is a contradiction to the initial claim.

Case 2: Let $\liminf_{n\to\infty}(\mu_l^t g^+_{\alpha=n}(x_n)) = 0$:

Since $\mu_l > 0$ and $g^+_{\alpha=n}(x_n) \geq 0$ then we have to have $\liminf g^+_{\alpha=n}(x_n) = 0$ moreover there must be a subsequence $x_{p(n)}$ of x_n such that $\lim g^+_{\alpha=n}(x_{p(n)}) = 0$.
Now let's focus on $\mu_l^t g^-_{\alpha=n}(x_{p(n)})$.
i) Let $\liminf \mu_l^t g^-_{\alpha=n}(x_{p(n)}) = -\infty$ then there exists a subsequence $x_{g(n)}$ of $x_{p(n)}$ such that $(\gamma - 1)\mu^t g^-(x_{g(n)})$ gets arbitrarily large as n gets larger so does the expression in equation 3. Hence as n gets large enough we get the inequality in 4 which is a contradiction to the initial assumption.
ii) Let $\liminf \mu_l^t g^-_{\alpha=n}(x_{p(n)}) > -\infty$ (in other words there exists a finite $c_p < 0$ in R such that there exists an $N \in Z^+$ such that $\forall n \geq N$ we have $\mu_l^t g^-_{\alpha=n}(x_{p(n)}) \geq c_p$) since $0 \geq g^-_{\alpha=n}(x_{p(n)})$ and $\mu > 0$ then the set $\{u \| 0 \geq u, \mu^t u \geq c_p\}$ is compact(trivial) and $g^-(x_{p(n)}) \in \{u \| 0 \geq u, \mu^t u \geq c_p\}$. Then there exists a subsequence $x_{g(n)}$ of $x_{p(n)}$ such that $\lim g_{\alpha=n}(x_{g(n)}) \to u_g \in R^r$ where $u_g^+ = 0$.

Now let (w_n, u_n) be a sequence in $M_{\alpha=n}$ such that $w_n = f(x_{g(n)})$ and $u_n = g_{\alpha=n}(x_{g(n)})$ from proposition 3 we know that $(w_n, u_n) \in M$.
Since $u_n \to u_g$, for every $\epsilon > 0$ there exists an N_ϵ such that $\forall n > N_\epsilon$, $\|u_n - u_g\| < \epsilon$. From $w_n + \mu_l^t u_n < q$ (initial claim) and $w_n + \mu^t u_n \geq c$ (assumption 1), for $n > N_\epsilon$ we have,

$$q + \epsilon\|\mu_l\| + \mu_l^t u_g \geq w_n \geq c - \epsilon\|\mu\| + \mu^t u_g \tag{5}$$

which means w_n is in a compact set. Then there exists a convergent subsequence $w_{k(n)}$ which converges to a point w_g in $[q + \epsilon\|\mu_l\| + \mu_l^t u_g, c - \epsilon\|\mu\| + \mu^t u_g]$ hence $\lim_{n\to\infty}(w_{k(n)}, u_{k(n)}) = (w_g, u_g)$.
Since M is closed (assumption 2) then $(w_g, u_g) \in M$. We have $w_g + \mu_l^t u_g < q$ (definition of x_n) and since $u_g \leq 0$ from proposition 1 we have $w_g + \mu_l^t u_g \geq q_l$. But then

$$q > w_g + \mu_l^t u_g \geq q_l > q \tag{6}$$

which is a contradiction. Case1 and case 2 together conclude the proof of lemma 1.

Now we can write our main theorem.

Theorem 1. *Under assumptions (1) and (2), as $\alpha \to \infty$ the dual optimal value q^*_α approaches the primal optimal value f^*.*

The proof of the theorem follows directly from (proposition 2) and lemma 1. For every α the optimal value stays the same. For every $q_l < f^*$ there exists a μ_l such that for large enough α we have $q_\alpha(\mu_l) \geq q_l$ (lemma 1) so $q^*_{alpha} \geq q_\alpha(\mu_l) \geq q_l$. From weak duality we have $f^*_\alpha \geq q^*_\alpha \geq q_l$ and from proposition 2 we get

$$f^* \geq q^*_\alpha \geq q_l$$

. A fairly obvious observation is that for $\alpha_1 \leq \alpha_2$ we have $q^*_{\alpha_1} \leq q^*_{\alpha_2}$. Then for every $n \geq 2$, $\alpha_1 = 0$ there exists an $\alpha_n \geq \alpha_{n-1}$ such that $f^* \geq q^*_{\alpha_n} \geq f^* - \frac{1}{2}^n$ which proves that asl $\alpha \to \infty$ the dual optimal q^*_α converges to f^*.

A Toy Example

In this part we are going to look at a simple nonconvex polynomial with a linear constraint to put the result in use. We will try a nonconvex polynomial.

$$\begin{aligned} &\text{minimize} \quad && x^4 - 18x^3 + 48x^2 - 32x + 16 \\ &\text{subject to} \quad && x \leq 0, \end{aligned} \tag{7}$$

The optimum point of the primal problem is 16. We are going to solve the dual of the problem with the α penalty function. For $\alpha \geq 1$ we have $g_\alpha(x) = max\{g(x), \alpha g(x)\}$. Then the dual function is

$$q_\alpha(\mu) = inf_x(x^4 - 18x^3 + 48x^2 - 32x + 16 + \mu(max\{x, \alpha x\}))$$

. To find this minimum a sun-gradient based steepest descent method is used. After finding $q(\mu)$ the optimum is found by generalized lp like convexification of the set M_α. On the table below you will find the optimal values for the dual problem versus α.

q^*_α	α
1	-7.00E+03
2	-6.63E+03
4	-5.88E+03
8	-4.41E+03
16	-2.11E+03
20	-168.6933
32	16

At $\alpha = 32$ the optimum value is reached.

Conclusion

We have proposed a penalty method to eliminate duality gap under 2 simple assumptions. The example is easy however shows that the method can work in practice.

The futurework will include some real world examples.

Assumption (1) can be relaxed by rotating the set M small enough so that the primal optimal does not change a lot and a non-vertical hyperplane can be found. To relax the non-horizontal hyperplane condition we can use large enough bounds on the problem. Relaxing the assumption (2) is a tedious task which may involve checking for lowersemicontinuity on the boundaries.

References

[1] Dimitri P. Bertsekas, *Nonlinear Probramming*, Athena Scientific; 2nd edition (September 1, 1999)

[2] Dimitri P. Bertsekas, Angelia Nedic, *Convex Analysis and Optimization*, Athena Scientific (April 1, 2003)