Remove Duality Gap Using Improved ℓ_1 Penalty Method

Cindy Zhang

Washington University, St. Louis, 63130

Abstract

Removing duality gap is an important problem in optimization area. Using penalty method to penalize the violated constraints is a simple but powerful way to minimize the objective function. However, in practice, the penalty parameters are often large for most penalty methods, which will cause the problem ill-conditioned and algorithms hard to converge. This paper proposed to use ℓ_1^m method, which uses multiple penalty parameters for ℓ_1 penalty, and evaluate the method using well-known test problems G1 to G10. The result suggests that the ℓ_1^m method has better performance in finding the constrained global minimum.

1 Introduction

Constrained nonlinear optimization plays a crucial role across diverse fields such as machine learning, engineering, and economics, among others. It involves optimizing a cost function while adhering to specific constraints that represent real-world limitations or requirements. The objective is often to minimize the cost or maximize a desired outcome. The formal expression is below.

Goal:

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad h_1(x) = 0, \ldots, h_i(x) = 0, \\ &\qquad\qquad g_1(x) \leq 0, \ldots, g_j(x) \leq 0, \end{aligned}$$

where $x \in X$, X is a compact subset of $\mathbb{R}^n$. We assume that the objective function f is lower bounded and is continuous and differentiable with respect to x, whereas the constraint functions $g = (g_1, \ldots, g_j)^T$ and $h = (h_1, \ldots, h_i)^T$ are continuous.

Duality is a fundamental concept in optimization that provides an alternative perspective on solving optimization problems by associating a dual problem with the original primal problem. The dual problem is formulated by relaxing the constraints of the primal problem, typically through techniques such as Lagrangian multipliers, resulting in a related optimization problem with its own objective function.In optimization, we often prefer to solve dual problems because they can be computationally easier to solve than the original ”primal” problem, particularly when the primal has a large number of constraints, as the dual may have fewer constraints and can provide valuable insights into the problem through its variables.

A key aspect of duality is the duality gap, which is the difference between the optimal values of the primal and dual problems. In many cases, especially for convex optimization problems, strong duality holds, meaning that the duality gap is zero, and the optimal values of the primal and dual problems are equal. By the principle of weak duality, the value of the dual objective function at any feasible point provides a lower bound (for minimization problems) or an upper bound (for maximization problems) to the primal objective value. Thus, removing the duality gap becomes a key problem.

2 Background

2.1 Primal Problem

Goal:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h_1(x) = 0, \dots, h_i(x) = 0, \\ & g_1(x) \leq 0, \dots, g_j(x) \leq 0, \end{aligned}$$

where $x \in X$, X is a compact subset of $\mathbb{R}^n$. We assume that the objective function f is lower bounded and is continuous and differentiable with respect to x, whereas the constraint functions $g = (g_1, \dots, g_j)^T$ and $h = (h_1, \dots, h_i)^T$ are continuous.

Since the function f is a constrained function, which is challenging to solve directly, we want to transform it into an unconstrained problem while ensuring that the solution satisfies the original constraints. A natural approach is to impose a significantly large penalty for constraint violations, effectively making the cost of violating the constraints approach infinity. To achieve this, we combine the objective function and the constraint functions into a single formulation, creating a unified representation of the optimization problem.

Let

$$J(x) = f(x) + \sigma \cdot |h(x)| + \theta \cdot g(x)$$

$$\sigma_i = \begin{cases} 0 & \text{if } h_i(x) = 0, \\ \infty & \text{if } h_i(x) \neq 0, \end{cases} \quad \theta_j = \begin{cases} 0 & \text{if } g_j(x) \leq 0, \\ \infty & \text{if } g_j(x) > 0. \end{cases}$$

The value of J becomes infinite when the constraints are violated, while it equals the value of the original objective function f when all constraints are satisfied.

However, function J is not continuous, which is difficult to solve. This, we approximate σ, θ with constant functions λ, μ, where $\mu \in [0, \infty)$.

Let function

$$L(x, \lambda, \mu) = f(x) + \lambda h(x) + \mu g(x)$$

$$L(x, \lambda, \mu) = \begin{cases} f(x) & \text{if constraints satisfied,} \\ f(x) + \lambda h(x) + \mu g(x) & \text{otherwise.} \end{cases}$$

$$\forall x, \quad \max_{\lambda \geq 0} L(x, \lambda, \mu) = J(x) \quad \text{by letting } \lambda \to \pm\infty \text{ or } \mu \to \infty.$$

Thus, minimizing f under constraints is the same as minimizing J, which is similar to get

$$\min_{x} \max_{\lambda \geq 0} L(x, \lambda, \mu) \tag{1}$$

2.2 Dual Problem

While the primal problem directly optimizes the original objective function subject to constraints, the dual problem focuses on maximizing or minimizing a related function (the dual function), which depends on the Lagrange multipliers and indirectly reflects the primal constraints.

Dual function:

$$q(\lambda, \mu) := \min_{x} L(x, \lambda, \mu).$$

Dual problem:

$$\max_{\lambda \geq 0} q(\lambda, \mu),$$

which is

$$\max_{\lambda \geq 0} \min_{x} L(x, \lambda, \mu) \tag{2}$$

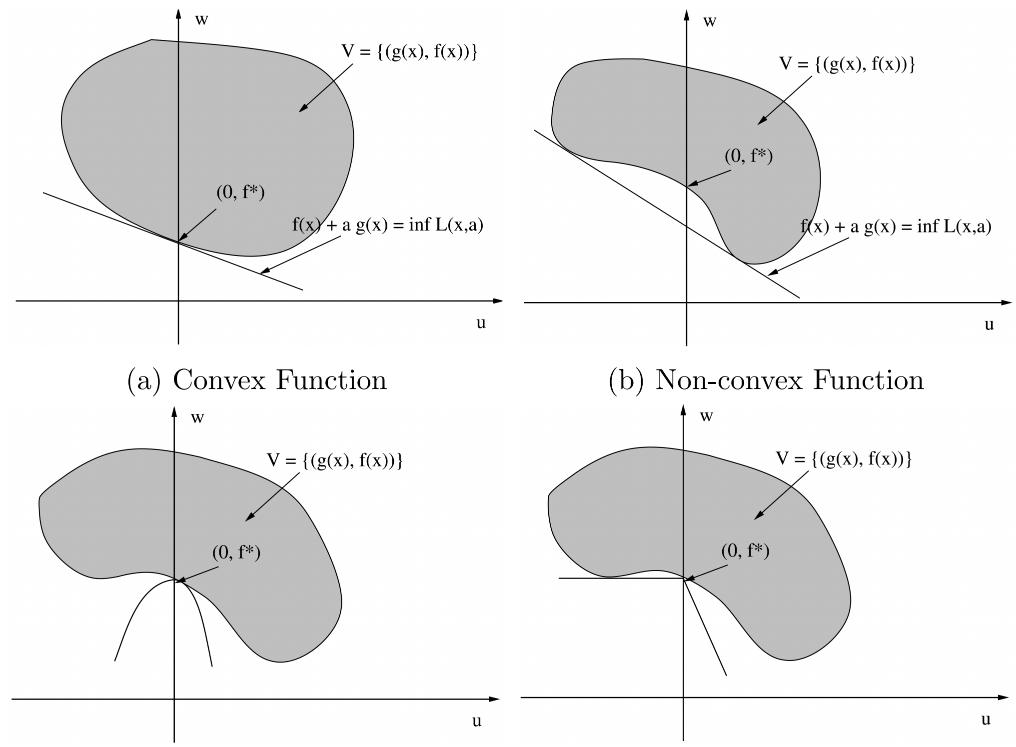
We can see that the difference between 1 and 2 is the exchange of *min* and *max*. The exchange of the *min* and *max* operations in the calculation is not always valid, and 1 and 2 are equivalent only under specific conditions, namely when strong duality holds.

Duality Gap:

$$\text{Duality gap} = \max q(\lambda, \mu) - f(x^*).$$

where x^* is the optimal solution of $f(x)$, x^* subject to constraints.

There are many efforts in regards to remove the duality gap. It is well known that the existence of duality gap is closely related to the geometric problem of finding the hyperplane supporting the set V of constraint-objective pairs. [4]



(a) Convex Function (b) Non-convex Function

(c) Augmented Lagrange Method (d) L1 Penalty

Figure 1: Geometric Interpretation of Duality

Let $u = g(x)$, $v = f(x)$, $V = \{(g(x), f(x))\}$. The Figure 1 provides a geometric interpretation of duality for an objective function with inequality constraints. Since $L(x, a) = f(x) + ag(x)$, which can be transformed to $f(x) = -ag(x) + L(x, a)$, it can be visualized as a straight line with the slope of $-a$ and w-intercept of $L(x, a)$. Since the constraint is $g(x) \leq 0$, the part of V which is in the left of the w-axis is the feasible set. The lowest point of the feasible part of V is the lowest value of f(x) under constraints, which is the optimal solution. The f* represents the optimal solution. To find the global minimum, we need to change the Lagrange function, so the shape of it can fit the feasible set better. The detail interpretation of Figure 1 is in Appendix A

3 Methodology

Definition 1 *The ℓ_1^m-penalty function for optimization problem in is defined as follows:*

$$L_m(x, \alpha, \beta) = f(x) + \alpha^T |h(x)| + \beta^T g^+(x),$$

where $|h(x)| = (|h_1(x)|, \ldots, |h_m(x)|)^T$ and $g^+(x) = \left(g_1^+(x), \ldots, g_r^+(x)\right)^T$, where we define $\phi^+(x) = \max(0, \phi(x))$ for a function ϕ, and $\alpha \in \mathbb{R}^m$ and $\beta \in \mathbb{R}^r$ are penalty multipliers.

In the term ℓ_1^m-*penalty*, the subscript 1 denotes the fact that L_m uses an ℓ_1 transformation of the constraints, while the superscript m denotes the fact that L_m has multiple penalty multipliers as opposed to the single penalty multiplier used by the conventional ℓ_1-penalty.

We consider the *dual function* defined for $\alpha \in \mathbb{R}^m$ and $\beta \in \mathbb{R}^r$ as:

$$q(\alpha, \beta) = \min_{x \in X} L_m(x, \alpha, \beta).$$

It is straightforward to show that the dual function $q(\alpha, \beta)$ is concave over $\alpha \geq 0$ and $\beta \geq 0$.

We define the *dual problem* as:

$$\begin{aligned} &\text{maximize} \quad q(\alpha, \beta) \\ &\text{subject to} \quad \alpha \geq 0, \quad \text{and} \quad \beta \geq 0, \end{aligned}$$

and the *optimal dual value* as:

$$q^* = \max_{\alpha \geq 0,\, \beta \geq 0} q(\alpha, \beta).$$

For continuous problems, we need the following constraint-qualification condition in order to rule out the special case in which all continuous constraints have zero derivative along a direction.

Definition 2 *A point $x \in X$ meets the constraint qualification if there exists no direction $p \in \mathbb{R}^n$ along which the directional derivatives of the objective is non-zero but the directional derivatives of continuous equality and continuous active inequality constraints are all zero. That is,*

$$\nexists\, p \in \mathbb{R}^n \quad \textit{such that}$$

$$f'(x; p) \neq 0, \quad h_i'(x; p) = 0 \quad \textit{and} \quad g_j'(x; p) = 0, \quad \forall i \in C_h \textit{ and } j \in C_g,$$

where C_h and C_g are, respectively, the sets of indices of continuous equality and continuous active inequality constraints. The constraint qualification is satisfied if both C_h and C_g are empty.

Theorem 1 *Suppose $x^* \in X$ is a constrained global minimum and x^* satisfies the constraint qualification, then there exist finite $\alpha^* \geq 0$ and $\beta^* \geq 0$ such that*

$$f(x^*) = \min_{x \in X} L_m(x, \alpha^{**}, \beta^{**}), \quad \textit{for any } \alpha^{**} \geq \alpha^*,\ \beta^{**} \geq \beta^*.$$

Proof: See Appendix A

The idea of this theorem is that when we find the $x*$, increasing the penalty parameter will not change the result of the optimal solution. This is intuitive as the increasing penalty parameters will not influence the value of function for feasible $x*$. With this theorem, it's easy to prove that ℓ_1^m method removes the duality gap A.

4 Technical details

procedure (l_1^m_Duality_Search($P, x, \alpha^{\max}, \beta^{\max}$);

1. $\alpha \leftarrow 0$; $\beta \leftarrow 0$;
2. **repeat**
3. **for** $(i = 1, \ldots, m)$ **if** $(h_i(x) \neq 0$ **and** $\alpha_i < \alpha_i^{\max})$ **then** increase α_i;
4. **for** $(j = 1, \ldots, r)$ **if** $(g_j(x) \nleq 0$ **and** $\beta_j < \beta_j^{\max})$ **then** increase β_j;
5. solve the dual problem and set $z \leftarrow \arg\min_{z \in X \times Y} L_m(z, \alpha, \beta)$;
6. **if** (penalty decrease condition satisfied) scale down α and β;
7. **until** $(\alpha_i > \alpha_i^{\max}$ **for all** $h_i(x) \neq 0$ **and** $\beta_j > \beta_j^{\max}$ **for all** $g_j(x) \nleq 0)$ **or a** $x*$ **of** P **is found;**
8. **return** $x*$ **if found;**
9. **end_procedure**

Pseudo Code: Iterative procedures to look for optimal value of problem P. The bounds on α and β, $\alpha^{\max}$ and $\beta^{\max}$, are user-provided.

Figure 4 is the pseudo code of the algorithm. The goal is to use the algorithm to find the global minimum by looking for x^*, y^*, α^{**}, and β^{**} by theorem1. Given an *alpha* and *beta*, the inner loop looks for a minimum of $L_m(x, y, \alpha, \beta)$ in order to find x^* and y^*. If a feasible solution to P_m is not found at the point minimizing $L_m(x, y, \alpha, \beta)$, the penalties corresponding to the violated constraints are increased. The process is repeated until a constrained global minimum is found or when α^{**} (resp. β^{**}) is larger than the user-provided maximum bound $\alpha^{\max}$ (resp. $\beta^{\max}$), where $\alpha^{\max}$ (resp. $\beta^{\max}$) is chosen to be so large that it exceeds α^* (resp. β^*).

The implementation details of the algorithm is described below. A key component is the strategy for updating the ℓ_1^m-penalties. In our experiments, after each unconstrained optimization in Line 5 of Figure, we use the following strategy to update the penalty values α and β corresponding to violated constraints (Lines 3 and 4):

$$\alpha \leftarrow \alpha + \rho^T |h(z)|, \qquad \beta \leftarrow \beta + \varrho^T g^+(z),$$

where ρ and ϱ are vectors for controlling the rate of updating α and β. Thus, after initialization, the update of each element of α and β is proportional to the violation of the corresponding constraint and stops when it is satisfied.

We update each element of ρ and ϱ dynamically until the corresponding constraint is satisfied. For each constraint h_i, $i = 1, \ldots, m$, we use c_i to count the number of consecutive iterations in which h_i is violated since the last update of ρ_i. After an iteration, we increase c_i by 1 if h_i is violated; if c_i reaches threshold K, which means that h_i has not been satisfied in K consecutive iterations, we increase ρ_i by:

$$\rho_i \leftarrow \rho_i \cdot \delta, \qquad \text{where } \delta > 1,$$

and reset c_i to 0. If h_i is satisfied, we reset ρ_i to ρ_0 and c_i to 0. In our implementation, we choose $\rho_0 = 0.01$, $K = 2$ and $\delta = 2.25$. We update ϱ in the same manner.

5 Experimental results

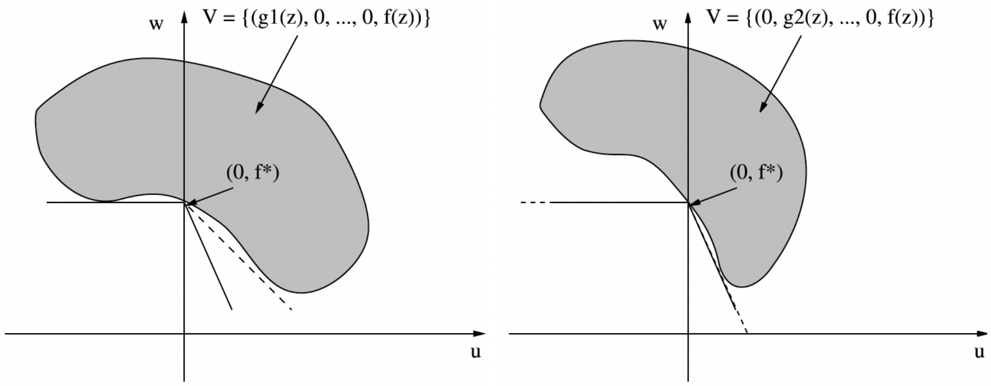
In this section, I use test sets from engineering application to test the algorithm. The G1 to G13 problems are a well-known suite of nonlinear constrained optimization problems used to evaluate the performance of optimization algorithms. Each problem has unique characteristics, including different types of nonlinear objectives, equality constraints, and inequality constraints.

	ℓ_1^M			ℓ_1		
	Q_{avg}	Q_{best}	P_{succ}	Q_{avg}	Q_{best}	P_{succ}
G1	-15.00	-15.00	1.0	-13.84	**-15.00**	0.05
G2	-0.77	-0.80	1.0	-	-	0
G3	-1.00	-1.00	1.0	-0.03	-0.96	0.57
G4	-30665.54	-30665.54	1.0	-30637.92	-30665.54	0.13
G5	-4221.96	-4221.96	1.0	-	-	0
G6	-6961.81	-6961.81	1.0	-3573.06	-6961.81	1
G7	24.31	24.31	1.0	-	-	0
G8	-0.096	-0.096	1.0	-0.014	-0.095	0.15
G9	680.63	680.63	1.0	-	-	0
G10	1.00	1.00	1.0	-	-	0

Table 1: Test Result for ℓ_1^M and ℓ_1 running 100 times

Table 1 shows the result of running both methods for 100 times. We can see that the ℓ_1^M method performs better than ℓ_1 method especially for G2, G5, G7, G9, and G10, which all require especially large penalty parameters. Taking G5 as an example, "the method seems to provide too strong penalties: often the factor grows too fast to be useful". By using ℓ_1^M method, we make the algorithm easier to converge and find the solution.

Figure 2 gives a geometric interpretation of it. Consider the shadowed area in graph a and b as 2 different section planes for the same set, and the line as the section planes for the same supporting hyperplane. The slope of the supporting vector for graph 2b should be very large to support the set from below, which makes the penalty parameter for the corresponding constraint extremely large. For other dimensions, the penalty parameter can be smaller. If we only have one penalty parameter, it will be the max of the penalty parameters needed for all constraints so the hyperplane can support the set from below. If we have multiple of them, each constraint can have their own penalty parameter, which don't have to be very large.



(a) Parameter dif in some dimensions (b) Same parameter for some dimension

The solid line represents $\ell_1$method, while the dotted line represents ℓ_1^m method

Figure 2: Comparison of 2 methods.

6 Conclusion

This paper introduces ℓ_1^m method, which requires less penalty values to achieve zero duality gap comparing to traditional penalty methods like ℓ_1 method. A search algorithm is proposed that can optimally solved many classic optimization problems for non-convex functions. The future work may include ways to further refine the algorithm and build a systematic way to determine the suitable hyperparameters based on the geometric shape of the set.

A Appendix: Additional Mathematical Proof

A.1 Detail explanation of Figure 1

In figure (a), let the objective function be convex and the constraint function be affine. V is convex by the definition of convexity and affinity. This is the case that satisfied the Slater's condition, which is a sufficient condition for strong duality to hold for a convex optimization problem.

Figure (b) shows the circumstance where the duality gap is non-zero because the w-intercept of the line is lower than the optimal value.

Figure (c) shows how the Augmented Lagrange Method removes duality gap. By adding a quadratic penalty term, the shape of the line becomes a parabola. The penalty parameter controls how narrow the parabola is, and theoretically remove the duality gap when the penalty parameter approaches to infinity.

Figure(d) shows how the L1 penalty method removes the duality gap. By adding a linear penalty term that becomes 0 when the constraints are satisfied, it creates a sharp-corner at the w-axis. The penalty parameter controls the slope of the line in infeasible part so that the line can support the set from below.

A.2 Proof of Theorem 1

Proof Since we have:

$$L_m(x^*, \alpha^{**}, \beta^{**}) = f(x^*) + \sum_{i=1}^{m} \alpha_i^{**}|h_i(x)| + \sum_{j=1}^{r} \beta_j^{**} g_j^+(x) = f(x^*), \tag{3}$$

$$\forall \alpha^{**} \geq 0, \beta^{**} \geq 0,$$

it is equivalent to show that there exist finite $\alpha^* \geq 0$ and $\beta^* \geq 0$ such that

$$f(x^*) \leq L_m(x, \alpha^{**}, \beta^{**}), \quad \text{for any } x \in X, \alpha^{**} > \alpha^*, \beta^{**} > \beta^*. \tag{4}$$

We prove is divided in three parts. First, we prove that the theorem is valid for any point x in the feasible set $\mathscr{F}$. Then, we show that the theorem is valid for any point x within $\mathscr{F}^+_{\epsilon_{\min}}$ for a small $\epsilon_{\min} > 0$. Last, we prove the theorem for the points in $X - \mathscr{F}^+_{\epsilon_{\min}}$. For simplicity, we assume that x^* is the only constrained global minimum in X. The case of more than one constrained global minimum can be proved similarly.
Before we start, we will first specify the constraint qualification. *Part (a)* For every feasible point $x' \in \mathscr{F}$, theorem holds for any $\alpha^{**} \geq 0$ and $\beta^{**} \geq 0$ since

$$L_m(x', \alpha^{**}, \beta^{**}) = f(x') \geq f(x^*), \tag{5}$$

noting that $h(x') = 0$, $g(x') \leq 0$, and $f(x') \geq f(x^*)$ by the definition of constrained global minimum.

Part (b) We show that theorem holds in $\mathscr{F}^+_\epsilon$ when ϵ is small enough. To this end, we show that for each feasible point $x' \in \mathscr{F}$, any point x in the close neighborhood of x' satisfies the condition.

For any feasible $x' \in \mathscr{F}$ that is not in the neighborhood of x^*, we have $f(x') - f(x^*) \geq \xi > 0$ for a finite positive ξ since x^* is the only constrained global minimum. Let $x = x' + \epsilon p$, $p \in \mathbb{R}^n$ is a unit-length direction vector with $\|p\| = 1$ and $\epsilon = \|x - x'\|$. When ϵ is small enough, we have:

$$\begin{aligned} L_m(x, \alpha^{**}, \beta^{**}) &= f(x) + \sum_{i=1}^{m} \alpha_i^{**}|h_i(x)| + \sum_{j=1}^{r} \beta_j^{**} g_j^+(x) \\ &\geq f(x) = f(x') + \epsilon f'(x'; p) + o(\epsilon^2) \\ &\geq f(x^*) + \xi + \epsilon f'(x'; p) + o(\epsilon^2) \geq f(x^*). \end{aligned} \tag{6}$$

For any point x in the neighborhood of x^*, let $x = x^* + \epsilon p$, where $p \in \mathbb{R}^n$, $\|p\| = 1$ is a unit-length direction vector and $\epsilon = \|x - x^*\|$. We show that when ϵ is small enough, there always exist finite α^* and β^* such that the condition holds. We consider the following two cases:

(1) If at x^* all the constraints are inactive inequality constraints, then when ϵ is small enough, x is also a feasible point. Hence, x^* being a constrained global minimum implies that $f(x) \geq f(x^*)$ and, regardless of the choice of the penalties,

$$L_m(x, \alpha^{**}, \beta^{**}) = f(x) + \sum_{i=1}^{m} \alpha_i^{**}|h_i(x)| + \sum_{j=1}^{r} \beta_j^{**} g_j^+(x) = f(x) \geq f(x^*). \tag{7}$$

(2) In this case, other than inactive inequality constraints, if there are equality or active inequality constraints at x^*. According to the constraint-qualification condition, unless $f'(x^*; p) = 0$, in which case x^* minimizes $L_m(x^*, \alpha^{**}, \beta^{**})$ in $\mathscr{F}_\epsilon^+$ for small enough ϵ, there must exist an equality constraint or an active inequality constraint that has non-zero derivative along p. Suppose there exists an equality constraint h_k that has non-zero derivative along p, which means $|h'_k(x^*; p)| > 0$. If we set $\alpha_k^{**} > \frac{|f'(x^*;p)|}{|h'_k(x^*;p)|}$ and ϵ small enough, then:

$$\begin{aligned} L_m(x, \alpha^{**}, \beta^{**}) &= f(x) + \sum_{i=1}^{m} \alpha_i^{**}|h_i(x)| + \sum_{j=1}^{r} \beta_j^{**} g_j^+(x) \\ &\geq f(x) + \alpha_k^{**}|h_k(x)| \\ &\geq f(x^*) + \epsilon f'(x^*; p) + o(\epsilon^2) + \alpha_k^{**}\epsilon|h'_k(x^*; p)| \\ &\geq f(x^*) + \epsilon\left(\alpha_k^{**}|h'_k(x^*; p)| - |f'(x^*; p)|\right) + o(\epsilon^2) \\ &\geq f(x^*). \end{aligned} \tag{8}$$

Part (c) Parts (a) and (b) have proved that condition holds for any point $x \in \mathscr{F}_{\epsilon_{\min}}^+$. We now prove that the theorem is true for any point $x \in X - \mathscr{F}_{\epsilon_{\min}}^+$.

For a point $x \in X - \mathscr{F}_{\epsilon_{\min}}^+$, there exists finite $\xi > 0$ such that:

$$\|h(x)\|^2 + \|g^+(x)\|^2 \geq \xi. \tag{9}$$

Let $f_{\min} = \min_{x \in X} f(x)$. Since $f(x)$ is lower bounded, $f_{\min}$ is finite. We set:

$$\alpha_i^* = \frac{f(x^*) - f_{\min}}{\xi|h_i(x)|}, \quad i = 1, \ldots, m, \tag{10}$$

$$\beta_j^* = \frac{f(x^*) - f_{\min}}{\xi g_j^+(x)}, \quad j = 1, \ldots, r. \tag{11}$$

Note that $\alpha^* \geq 0$ and $\beta^* \geq 0$ since $f(x^*) \geq f_{\min}$. We have, for any $\alpha^{**} \geq \alpha^*, \beta^{**} \geq \beta^*$:

$$\begin{aligned} L_m(x, \alpha^{**}, \beta^{**}) &= f(x) + \sum_{i=1}^{m} \alpha_i^{**}|h_i(x)| + \sum_{j=1}^{r} \beta_j^{**} g_j^+(x) \\ &\geq f(x) + \frac{f(x^*) - f_{\min}}{\xi}\left(\|h(x)\|^2 + \|g^+(x)\|^2\right) \\ &\geq f(x) + f(x^*) - f_{\min} \\ &\geq f(x^*). \end{aligned} \tag{12}$$

□

A.3 Proof of zero Duality Gap

Proof First, we have $q^* \leq f(x^*)$ since

$$q^* = \max_{\alpha \geq 0, \beta \geq 0} q(\alpha, \beta) = \max_{\alpha \geq 0, \beta \geq 0}\left(\min_{x \in X} L_m(x, \alpha, \beta)\right)$$

$$\leq \max_{\alpha \geq 0, \beta \geq 0} L_m(x^*, \alpha, \beta) = \max_{\alpha \geq 0, \beta \geq 0} f(x^*) = f(x^*).$$

Also, according to Theorem 1, there are $\alpha^{**} \geq 0$ and $\beta^{**} \geq 0$ such that $q(\alpha^{**}, \beta^{**}) = f(x^*)$, we have:

$$q^* = \max_{\alpha \geq 0, \beta \geq 0} q(\alpha, \beta) \geq q(\alpha^{**}, \beta^{**}) = f(x^*).$$

Since $q^* \leq f(x^*)$ and $q^* \geq f(x^*)$, we have $q^* = f(x^*)$ $\square$.

References

[1] Burke, J.V. (1991). Calmness and exact penalization. *SIAM Journal on Control and Optimization*, 29, 493–497.

[2] Burke, J.V. (1991). An exact penalization viewpoint of constrained optimization. *SIAM Journal on Control and Optimization*, 29, 968–998.

[3] Michalewicz, Z., & Schoenauer, M. (1996). Evolutionary Algorithms for Constrained Parameter Optimization Problems. *Evolutionary Computation*, 4(1), 1–32. MIT Press.

[4] Nedić, A., & Ozdaglar, A. (2008). A geometric framework for nonconvex optimization duality using augmented Lagrangian functions. *Journal of Global Optimization*, 40(4), 545–573.