DOI: 10.1016/j.cam.2013.11.010

A new exact penalty method for semi-infinite programming problems

Qun Lin Ryan Loxton Kok Lay Teo Yong Hong Wu Changjun Yu

Department of Mathematics and Statistics, Curtin University, Perth, Western Australia

Abstract

In this paper, we consider a class of nonlinear semi-infinite optimization problems. These problems involve continuous inequality constraints that need to be satisfied at every point in an infinite index set, as well as conventional equality and inequality constraints. By introducing a novel penalty function to penalize constraint violations, we form an approximate optimization problem in which the penalty function is minimized subject to only bound constraints. We then show that this penalty function is exact—that is, when the penalty parameter is sufficiently large, any local solution of the approximate problem can be used to generate a corresponding local solution of the original problem. On this basis, the original problem can be solved as a sequence of approximate nonlinear programming problems. We conclude the paper with some numerical results demonstrating the applicability of our approach to PID control and filter design.

Keywords:
Exact penalty function, Semi-infinite programming, Constrained optimization, Nonlinear programming

1. Introduction

In this paper, we consider semi-infinite programming problems of the following form:

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) && \text{(1a)} \\ \text{subject to} \quad & \varphi_i(\boldsymbol{x}, \omega) \leq 0, \quad \omega \in \Omega_i, \quad i \in \mathcal{C}, && \text{(1b)} \\ & g_i(\boldsymbol{x}) \leq 0, \quad i \in \mathcal{I}, && \text{(1c)} \\ & h_i(\boldsymbol{x}) = 0, \quad i \in \mathcal{E}, && \text{(1d)} \\ & a_j \leq x_j \leq b_j, \quad j = 1, \ldots, n, && \text{(1e)} \end{aligned}$$

where $\boldsymbol{x} = [x_1, \ldots, x_n]^\top \in \mathbb{R}^n$ is the decision vector; f, g_i, $h_i : \mathbb{R}^n \to \mathbb{R}$ and $\varphi_i : \mathbb{R}^n \times \Omega_i \to \mathbb{R}$ are continuously differentiable functions; a_j and b_j are given constants satisfying $a_j < b_j$; and $\Omega_i \subset \mathbb{R}$ are compact intervals of positive measure. We refer to this problem as Problem (P).

If $\mathcal{C} = \emptyset$, then Problem (P) is a standard nonlinear programming problem that can be solved efficiently using well-known methods such as sequential quadratic programming (see [9, 10]). Thus, the main difficulty with Problem (P) is the continuous inequality constraints (1b), which arise in a wide range of important applications such as signal processing [11], circuit design [1, 12], and optimal control [6, 16]. Each continuous inequality constraint in (1b) actually defines an infinite number of constraints—one for each point in Ω_i.

Teo and Goh in [14] have proposed a simple approach for tackling Problem (P). This approach involves transforming the continuous inequality constraints (1b) into the following set of equivalent equality constraints:

$$c_i \int_{\Omega_i} \left[\max\{\varphi_i(\boldsymbol{x}, \omega), 0\} \right]^2 d\omega = 0, \qquad i \in \mathcal{C}, \tag{2}$$

where $c_i > 0$, $i \in \mathcal{C}$ are given weights. Thus, the continuous inequality constraints are replaced by a finite set of equality constraints, and the resulting optimization problem can, in principle, be solved using conventional techniques. The downside of this approach, however, is that the equality constraints (2) do not satisfy the standard linear independence constraint qualification, and thus numerical convergence cannot be guaranteed.

To overcome this limitation, Jennings and Teo in [7] proposed an alternative method in which the continuous inequality constraints (1b) are approximated as follows:

$$\int_{\Omega_i} \mathcal{L}_\epsilon(\varphi_i(\boldsymbol{x},\omega))d\omega \leq \tau, \qquad i \in \mathcal{C}, \tag{3}$$

where $\epsilon > 0$ and $\tau > 0$ are adjustable parameters and $\mathcal{L}_\epsilon : \mathbb{R} \to \mathbb{R}$ is a smooth approximation of $\max\{\cdot, 0\}$. The approximating function $\mathcal{L}_\epsilon$ is specially designed so that $\mathcal{L}_\epsilon(\eta) \geq 0$ and $\mathcal{L}_\epsilon(\eta) \to \max\{\eta, 0\}$ as $\epsilon \to 0$. Replacing (1b) with (3) yields an approximate nonlinear programming problem. It can be shown that for each $\epsilon > 0$, if τ is chosen sufficiently small, then any solution of the approximate problem is feasible for Problem (P). Furthermore, the optimal cost of the approximate problem converges to the optimal cost of Problem (P) as $\epsilon \to 0$. On this basis, a solution of Problem (P) can be obtained by solving a sequence of approximate problems, where the parameters ϵ and τ are adjusted appropriately according to certain rules.

This idea was further developed in [15] with the introduction of the following penalty function, which is based on the constraint approximation (3):

$$f(\boldsymbol{x}) + \sigma \sum_{i\in\mathcal{C}} \int_{\Omega_i} \mathcal{L}_\epsilon(\varphi_i(\boldsymbol{x},\omega))d\omega, \tag{4}$$

where $\sigma > 0$ is the penalty parameter. Note that violations of the continuous inequality constraints (1b) are penalized by the integral term in (4). It can be shown that for each $\epsilon > 0$, if σ is made sufficiently large, then any minimizer of (4) on the region defined by (1c)-(1e) is feasible for Problem (P). Thus, a solution of Problem (P) can be obtained by minimizing (4) for appropriate choices of the parameters ϵ and σ.

Although the constraint approximation methods in [7, 15] generally perform well, numerical convergence is only guaranteed when the approximate problems are solved in a global sense. However, in practice, the approximate problems (and the original problem) are usually non-convex, and thus we can only expect to solve them locally. Unfortunately, conditions under which a local solution of the approximate problem converges to a local solution of the original problem are not known.

Motivated by this drawback, Yu et al. in [17, 18] recently introduced a new penalty function defined as follows:

$$F(\boldsymbol{x},\epsilon) \triangleq \begin{cases} f(\boldsymbol{x}), & \text{if } \epsilon = 0,\ \Delta(\boldsymbol{x},\epsilon) = 0, \\ f(\boldsymbol{x}) + \epsilon^{-\alpha}\Delta(\boldsymbol{x},\epsilon) + \sigma\epsilon^{\beta}, & \text{if } \epsilon > 0, \\ \infty, & \text{otherwise,} \end{cases} \tag{5}$$

where

$$\Delta(\boldsymbol{x},\epsilon) \triangleq \sum_{i\in\mathcal{C}} \int_{\Omega_i} \left[\max\{\varphi_i(\boldsymbol{x},\omega) - \epsilon^{\gamma}W_i, 0\}\right]^2 d\omega. \tag{6}$$

Here, ϵ is a new decision variable, $\sigma > 0$ is the penalty parameter, and $\alpha > 0$, $\beta > 2$, $\gamma > 0$, and $W_i \in (0, 1)$, $i \in \mathcal{C}$ are fixed constants.

Unlike (4), the penalty function (5) only involves one adjustable parameter (the penalty parameter σ). Furthermore, when σ is sufficiently large (and certain technical conditions are satisfied), any local minimizer of (5) can be used to generate a corresponding local minimizer of Problem (P) (with $\mathcal{E} = \mathcal{I} = \emptyset$). This result is more practical than the convergence results given in [7, 15], which are only applicable when the approximate problems are solved globally.

The penalty function (5) is a clear improvement over (4). However, it still has two disadvantages:

(i) Equations (5) and (6) involve $|\mathcal{C}| + 3$ fixed parameters, each of which needs to be selected judiciously.

(ii) Convergence is only guaranteed when there are no standard equality or inequality constraints (i.e. $\mathcal{I} = \mathcal{E} = \emptyset$) and none of the bound constraints are active at an optimal solution.

The aim of this paper is to address these issues by proving new convergence results under less stringent conditions. In particular, we will show that the parameters W_i in (6) are actually unnecessary, and (5) is still an effective penalty function when $W_i = 0$. Accordingly, the number of fixed parameters in the penalty function can be significantly reduced from $|\mathcal{C}| + 3$ to just 2. This simplified penalty function is still exact in the sense that the penalty parameter is not required to reach infinity for the constraints in Problem (P) to be satisfied. Furthermore, the numerical results in Section 5 show that our simplified penalty function is just as effective as the original one proposed in [17, 18].

2. An Exact Penalty Function for Problem (P)

2.1. Definition of the Penalty Function

Let

$$\mathcal{X} \triangleq \{ \boldsymbol{x} \in \mathbb{R}^n : a_j \leq x_j \leq b_j, \, j = 1, \ldots, n \}.$$

Define a *constraint violation function* on $\mathcal{X}$ as follows:

$$G(\boldsymbol{x}) \triangleq \sum_{i \in \mathcal{E}} \big[h_i(\boldsymbol{x})\big]^2 + \sum_{i \in \mathcal{I}} \big[\max\{g_i(\boldsymbol{x}), 0\}\big]^2 + \sum_{i \in \mathcal{C}} \int_{\Omega_i} \big[\max\{\varphi_i(\boldsymbol{x}, \omega), 0\}\big]^2 d\omega.$$

Clearly, $G(\boldsymbol{x}) = 0$ if and only if constraints (1b)-(1d) are satisfied. Furthermore, by Leibniz's rule,

$$\frac{\partial G(\boldsymbol{x})}{\partial \boldsymbol{x}} = 2\sum_{i \in \mathcal{E}} h_i(\boldsymbol{x}) \frac{\partial h_i(\boldsymbol{x})}{\partial \boldsymbol{x}} + 2\sum_{i \in \mathcal{I}} \max\{g_i(\boldsymbol{x}), 0\} \frac{\partial g_i(\boldsymbol{x})}{\partial \boldsymbol{x}} + 2\sum_{i \in \mathcal{C}} \int_{\Omega_i} \max\{\varphi_i(\boldsymbol{x}, \omega), 0\} \frac{\partial \varphi_i(\boldsymbol{x}, \omega)}{\partial \boldsymbol{x}} d\omega. \quad (7)$$

Thus, the constraint violation function is continuously differentiable.

Let $\epsilon \in [0, \bar{\epsilon}]$ be a new decision variable, where $\bar{\epsilon} > 0$ is a given upper bound. We define the following *penalty function* on $\mathcal{X} \times [0, \bar{\epsilon}]$:

$$F_\sigma(\boldsymbol{x}, \epsilon) \triangleq \begin{cases} f(\boldsymbol{x}), & \text{if } \epsilon = 0, \; G(\boldsymbol{x}) = 0, \\ f(\boldsymbol{x}) + \epsilon^{-\alpha} G(\boldsymbol{x}) + \sigma \epsilon^{\beta}, & \text{if } \epsilon \in (0, \bar{\epsilon}], \\ \infty, & \text{if } \epsilon = 0, \; G(\boldsymbol{x}) \neq 0, \end{cases}$$

where $\sigma > 0$ is the penalty parameter and α and β are positive constants satisfying $1 \leq \beta \leq \alpha$. Note that this condition on α and β is different to the condition in [17, 18], which requires $\alpha > 0$ and $\beta > 2$.

The penalty function F_σ works as follows: when σ is large, the term $\sigma \epsilon^\beta$ forces ϵ to be small, which in turn causes the term $\epsilon^{-\alpha} G(\boldsymbol{x})$ to severely penalize constraint violations. Hence, minimizing the penalty function for large values of σ will likely yield a feasible point of Problem (P). Accordingly, we consider the following *penalty problem* in which the penalty function F_σ is minimized on $\mathcal{X} \times [0, \bar{\epsilon}]$:

$$\min_{(\boldsymbol{x}, \epsilon) \in \mathcal{X} \times [0, \bar{\epsilon}]} F_\sigma(\boldsymbol{x}, \epsilon).$$

We refer to this problem as Problem (P_σ).

2.2. Three Preliminary Lemmas

For each fixed $\sigma > 0$ and $\boldsymbol{x} \in \mathcal{X}$, consider the following one-dimensional subproblem of Problem (P_σ):

$$\min_{\epsilon \in [0, \bar{\epsilon}]} F_\sigma(\boldsymbol{x}, \epsilon).$$

We refer to this subproblem as Problem ($\mathrm{P}_{\sigma, \boldsymbol{x}}$). Our first result characterizes the solution of Problem ($\mathrm{P}_{\sigma, \boldsymbol{x}}$).

Lemma 1. *Let $\sigma > 0$ and $\boldsymbol{x} \in \mathcal{X}$ be arbitrary but fixed. Define*

$$\delta = \frac{\sigma\beta\bar{\epsilon}^{\alpha+\beta}}{\alpha},$$
$$\tau = \left(\frac{\alpha G(\boldsymbol{x})}{\sigma\beta}\right)^{\frac{1}{\alpha+\beta}},$$
$$\epsilon^* = \begin{cases} \tau, & \text{if } G(\boldsymbol{x}) < \delta, \\ \bar{\epsilon}, & \text{if } G(\boldsymbol{x}) \geq \delta. \end{cases}$$

Then Problem ($P_{\sigma,\boldsymbol{x}}$) has a unique global minimizer $\epsilon = \epsilon^*$. Furthermore, Problem ($P_{\sigma,\boldsymbol{x}}$) does not have any non-global local minimizers.

Proof. We consider two cases: (i) $G(\boldsymbol{x}) = 0$ and (ii) $G(\boldsymbol{x}) \neq 0$. For case (i), $\epsilon^* = \tau = 0$ because $\delta > 0$. Also,

$$F_\sigma(\boldsymbol{x}, \epsilon) = f(\boldsymbol{x}) + \sigma\epsilon^\beta, \qquad \epsilon \in [0, \bar{\epsilon}]. \tag{8}$$

Since $\sigma > 0$ and $\beta \geq 1$, it is clear from (8) that $\epsilon^* = 0$ is the unique global minimizer of $F_\sigma(\boldsymbol{x}, \cdot)$ on $[0, \bar{\epsilon}]$. Furthermore,

$$\frac{\partial F_\sigma(\boldsymbol{x}, \epsilon)}{\partial \epsilon} = \sigma\beta\epsilon^{\beta-1} > 0, \qquad \epsilon \in (0, \bar{\epsilon}],$$

which shows that $F_\sigma(\boldsymbol{x}, \cdot)$ is strictly increasing on $(0, \bar{\epsilon}]$. Hence, there are no local minimizers of $F_\sigma(\boldsymbol{x}, \cdot)$ on the interval $(0, \bar{\epsilon}]$.

Now, consider case (ii). We have

$$F_\sigma(\boldsymbol{x}, \epsilon) = f(\boldsymbol{x}) + \epsilon^{-\alpha}G(\boldsymbol{x}) + \sigma\epsilon^\beta, \qquad \epsilon \in (0, \bar{\epsilon}].$$

Differentiating F_σ with respect to ϵ yields

$$\frac{\partial F_\sigma(\boldsymbol{x}, \epsilon)}{\partial \epsilon} = -\alpha\epsilon^{-\alpha-1}G(\boldsymbol{x}) + \sigma\beta\epsilon^{\beta-1}, \qquad \epsilon \in (0, \bar{\epsilon}].$$

Thus, for each $\epsilon \in (0, \bar{\epsilon}]$,

$$\frac{\partial F_\sigma(\boldsymbol{x}, \epsilon)}{\partial \epsilon} \begin{cases} < 0, & \text{if } \epsilon < \tau, \\ = 0, & \text{if } \epsilon = \tau, \\ > 0, & \text{if } \epsilon > \tau. \end{cases} \tag{9}$$

This shows that $F_\sigma(\boldsymbol{x}, \cdot)$ is strictly decreasing from $\epsilon = 0$ to $\epsilon = \tau$, and then strictly increasing from $\epsilon = \tau$ onwards. Thus, there is only one possible minimizing point for Problem ($P_{\sigma,\boldsymbol{x}}$): $\epsilon = \tau$ if $\tau < \bar{\epsilon}$, and $\epsilon = \bar{\epsilon}$ if $\tau \geq \bar{\epsilon}$.

If $G(\boldsymbol{x}) \geq \delta$, then

$$\tau = \left(\frac{\alpha G(\boldsymbol{x})}{\sigma\beta}\right)^{\frac{1}{\alpha+\beta}} \geq \left(\frac{\alpha\delta}{\sigma\beta}\right)^{\frac{1}{\alpha+\beta}} = \bar{\epsilon},$$

which implies that $\epsilon^* = \bar{\epsilon}$ is the unique global minimizer of Problem ($P_{\sigma,\boldsymbol{x}}$).

On the other hand, if $G(\boldsymbol{x}) < \delta$, then

$$\tau = \left(\frac{\alpha G(\boldsymbol{x})}{\sigma\beta}\right)^{\frac{1}{\alpha+\beta}} < \left(\frac{\alpha\delta}{\sigma\beta}\right)^{\frac{1}{\alpha+\beta}} = \bar{\epsilon}.$$

Hence, it follows that $\epsilon^* = \tau$ is the unique global minimizer of Problem ($P_{\sigma,\boldsymbol{x}}$). □

Recall that the constraint violation G is a continuous function defined on the compact set $\mathcal{X}$. Thus, G is bounded on $\mathcal{X}$; that is, there exists a positive real number $C > 0$ such that

$$G(\boldsymbol{x}) \leq C, \qquad \boldsymbol{x} \in \mathcal{X}.$$

Let

$$\bar{\sigma} = \frac{C\alpha}{\beta\bar{\epsilon}^{\alpha+\beta}}.$$

Then for all $\sigma > \bar{\sigma}$,

$$G(\boldsymbol{x}) \leq C = \frac{\bar{\sigma}\beta\bar{\epsilon}^{\alpha+\beta}}{\alpha} < \frac{\sigma\beta\bar{\epsilon}^{\alpha+\beta}}{\alpha} = \delta, \qquad \boldsymbol{x} \in \mathcal{X}.$$

Thus, the following result follows immediately from Lemma 1.

Lemma 2. *For each $\sigma > 0$ and $\boldsymbol{x} \in \mathcal{X}$, let*

$$\tau(\sigma, \boldsymbol{x}) = \left(\frac{\alpha G(\boldsymbol{x})}{\sigma\beta}\right)^{\frac{1}{\alpha+\beta}}.$$

Then for all sufficiently large $\sigma > 0$, the unique global minimizer of Problem $(P_{\sigma,\boldsymbol{x}})$ is $\epsilon = \tau(\sigma, \boldsymbol{x})$.

On the basis of Lemma 1, we now derive the following result.

Lemma 3. *Let $(\boldsymbol{x}^*, \epsilon^*)$ be a local solution of Problem (P_σ). Then $\boldsymbol{x}^*$ is a local solution of Problem (P) if and only if $\epsilon^* = 0$.*

Proof. Since $(\boldsymbol{x}^*, \epsilon^*)$ is a local solution of Problem (P_σ), ϵ^* is a local solution of Problem $(\mathrm{P}_{\sigma,\boldsymbol{x}^*})$. But we know from Lemma 1 that Problem $(\mathrm{P}_{\sigma,\boldsymbol{x}^*})$ does not have any non-global local solutions, and hence ϵ^* must be the unique global solution of Problem $(\mathrm{P}_{\sigma,\boldsymbol{x}^*})$.

Now, suppose that $\boldsymbol{x}^*$ is a local solution of Problem (P). Then clearly $G(\boldsymbol{x}^*) = 0$. Hence, Lemma 1 implies that the unique global solution of Problem $(\mathrm{P}_{\sigma,\boldsymbol{x}^*})$ is

$$\epsilon^* = \tau = \left(\frac{\alpha G(\boldsymbol{x}^*)}{\sigma\beta}\right)^{\frac{1}{\alpha+\beta}} = 0,$$

as required.

Conversely, suppose that $\epsilon^* = 0$. Then $\epsilon^* = 0$ is the unique global solution of Problem $(\mathrm{P}_{\sigma,\boldsymbol{x}^*})$. Thus, since $\bar{\epsilon} > 0$, it follows from Lemma 1 that $\tau = \epsilon^* = 0$. Therefore,

$$\tau = \left(\frac{\alpha G(\boldsymbol{x}^*)}{\sigma\beta}\right)^{\frac{1}{\alpha+\beta}} = 0.$$

This implies that $G(\boldsymbol{x}^*) = 0$ and hence $\boldsymbol{x}^*$ is feasible for Problem (P).

Next we will prove that $\boldsymbol{x}^*$ is locally optimal for Problem (P). Since $(\boldsymbol{x}^*, \epsilon^*) = (\boldsymbol{x}^*, 0)$ is a local solution of Problem (P_σ), there exists a real number $\gamma > 0$ such that

$$f(\boldsymbol{x}^*) = F_\sigma(\boldsymbol{x}^*, 0) \leq F_\sigma(\boldsymbol{x}, 0), \qquad \boldsymbol{x} \in \mathcal{N}_\gamma(\boldsymbol{x}^*),$$

where

$$\mathcal{N}_\gamma(\boldsymbol{x}^*) = \{\, \boldsymbol{y} \in \mathcal{X} : \|\boldsymbol{y} - \boldsymbol{x}^*\| < \gamma \,\}.$$

In particular, for all feasible points $\boldsymbol{x} \in \mathcal{N}_\gamma(\boldsymbol{x}^*)$ satisfying constraints (1b)-(1d),

$$f(\boldsymbol{x}^*) \leq F_\sigma(\boldsymbol{x}, 0) = f(\boldsymbol{x}).$$

This shows that $\boldsymbol{x}^*$ is a local solution of Problem (P), as required. □

3. Convergence Results and Algorithm

Recall that the penalty function F_σ is designed to penalize constraint violations more and more severely as σ is increased. In this section, we will show that F_σ is a so-called *exact penalty function*. This means that when the penalty parameter σ is sufficiently large, any local solution of Problem (P_σ) can be used to generate a corresponding local solution of Problem (P). As we will see, this result forms the basis of a computational algorithm for solving Problem (P).

Our first result shows that Problem (P_σ), the penalty problem, is well-posed.

Theorem 1. *For each $\sigma > 0$, Problem (P_σ) has an optimal solution.*

Proof. Let $H(\boldsymbol{x})$ denote the optimal cost of Problem ($\mathrm{P}_{\sigma,\boldsymbol{x}}$). That is,

$$H(\boldsymbol{x}) = \min_{\epsilon\in[0,\bar{\epsilon}]} F_\sigma(\boldsymbol{x},\epsilon).$$

Furthermore, let

$$\mathcal{X}_1 = \{\, \boldsymbol{x}\in\mathcal{X} : G(\boldsymbol{x}) \le \delta \,\}$$

and

$$\mathcal{X}_2 = \{\, \boldsymbol{x}\in\mathcal{X} : G(\boldsymbol{x}) \ge \delta \,\},$$

where δ is as defined in Lemma 1. Then the sets $\mathcal{X}_1$ and $\mathcal{X}_2$ are clearly compact because G is continuous.

Let

$$\tau(\sigma,\boldsymbol{x}) = \left(\frac{\alpha G(\boldsymbol{x})}{\sigma\beta}\right)^{\frac{1}{\alpha+\beta}}$$

It is easy to see that if $G(\boldsymbol{x}) = \delta$, then $\tau(\sigma,\boldsymbol{x}) = \bar{\epsilon}$. Hence, by Lemma 1, for each $\boldsymbol{x}\in\mathcal{X}_1$,

$$H(\boldsymbol{x}) = F_\sigma(\boldsymbol{x},\tau(\sigma,\boldsymbol{x})) = f(\boldsymbol{x}) + M\big[G(\boldsymbol{x})\big]^{\frac{\beta}{\alpha+\beta}}, \tag{10}$$

where

$$M = \sigma^{\frac{\alpha}{\alpha+\beta}}\left[\left(\frac{\beta}{\alpha}\right)^{\frac{\alpha}{\alpha+\beta}} + \left(\frac{\alpha}{\beta}\right)^{\frac{\beta}{\alpha+\beta}}\right].$$

Since f and G are continuous, H defined by (10) is also continuous. Thus, H achieves its minimum value on the compact set $\mathcal{X}_1$. In other words, there exists a vector $\boldsymbol{x}_1^*\in\mathcal{X}_1$ such that

$$F_\sigma(\boldsymbol{x}_1^*,\tau(\sigma,\boldsymbol{x}_1^*)) = H(\boldsymbol{x}_1^*) = \min_{\boldsymbol{x}\in\mathcal{X}_1} H(\boldsymbol{x}) = \min_{\boldsymbol{x}\in\mathcal{X}_1}\left\{\min_{\epsilon\in[0,\bar{\epsilon}]} F_\sigma(\boldsymbol{x},\epsilon)\right\} = \min_{(\boldsymbol{x},\epsilon)\in\mathcal{X}_1\times[0,\bar{\epsilon}]} F_\sigma(\boldsymbol{x},\epsilon). \tag{11}$$

Similarly, it follows from Lemma 1 that for each $\boldsymbol{x}\in\mathcal{X}_2$,

$$H(\boldsymbol{x}) = F_\sigma(\boldsymbol{x},\bar{\epsilon}) = f(\boldsymbol{x}) + \frac{G(\boldsymbol{x})}{\bar{\epsilon}^\alpha} + \sigma\bar{\epsilon}^\beta.$$

Thus, there exists a vector $\boldsymbol{x}_2^*\in\mathcal{X}_2$ such that

$$F_\sigma(\boldsymbol{x}_2^*,\bar{\epsilon}) = H(\boldsymbol{x}_2^*) = \min_{\boldsymbol{x}\in\mathcal{X}_2} H(\boldsymbol{x}) = \min_{\boldsymbol{x}\in\mathcal{X}_2}\left\{\min_{\epsilon\in[0,\bar{\epsilon}]} F_\sigma(\boldsymbol{x},\epsilon)\right\} = \min_{(\boldsymbol{x},\epsilon)\in\mathcal{X}_2\times[0,\bar{\epsilon}]} F_\sigma(\boldsymbol{x},\epsilon). \tag{12}$$

Since $\mathcal{X}_1\cup\mathcal{X}_2 = \mathcal{X}$, equations (11) and (12) give

$$\begin{aligned}\min_{(\boldsymbol{x},\epsilon)\in\mathcal{X}\times[0,\bar{\epsilon}]} F_\sigma(\boldsymbol{x},\epsilon) &= \min\left\{\min_{(\boldsymbol{x},\epsilon)\in\mathcal{X}_1\times[0,\bar{\epsilon}]} F_\sigma(\boldsymbol{x},\epsilon),\ \min_{(\boldsymbol{x},\epsilon)\in\mathcal{X}_2\times[0,\bar{\epsilon}]} F_\sigma(\boldsymbol{x},\epsilon)\right\}\\ &= \min\big\{F_\sigma(\boldsymbol{x}_1^*,\tau(\sigma,\boldsymbol{x}_1^*)),\ F_\sigma(\boldsymbol{x}_2^*,\bar{\epsilon})\big\}.\end{aligned}$$

Hence, either $(\boldsymbol{x}_1^*,\tau(\sigma,\boldsymbol{x}_1^*))$ or $(\boldsymbol{x}_2^*,\bar{\epsilon})$ is a global solution of Problem (P_σ). □

We now show that if a sequence of local solutions of Problem (P_σ) converges to a limit $(\boldsymbol{x}^*, \epsilon^*)$ as $\sigma \to \infty$, then $\boldsymbol{x}^*$ must be feasible for Problem (P).

Theorem 2. *Let $\{\sigma_k\}_{k=1}^\infty$ be an increasing sequence of penalty parameters such that $\sigma_k \to \infty$ as $k \to \infty$. Furthermore, let $(\boldsymbol{x}^{k,*}, \epsilon^{k,*})$ be a local solution of Problem (P_{σ_k}). Suppose that $\{F_{\sigma_k}(\boldsymbol{x}^{k,*}, \epsilon^{k,*})\}_{k=1}^\infty$ is bounded and $\boldsymbol{x}^{k,*} \to \boldsymbol{x}^*$ as $k \to \infty$. Then $\boldsymbol{x}^*$ is feasible for Problem (P).*

Proof. Since $\{\boldsymbol{x}^{k,*}\} \subset \mathcal{X}$ and $\mathcal{X}$ is compact, it is clear that $\boldsymbol{x}^* \in \mathcal{X}$. We complete the proof by showing that $G(\boldsymbol{x}^*) = 0$.

Note that $G(\boldsymbol{x}^{k,*}) \to G(\boldsymbol{x}^*)$ as $k \to \infty$ because G is a continuous function. Thus, if $G(\boldsymbol{x}^{k,*}) = 0$ for all sufficiently large k, then clearly $G(\boldsymbol{x}^*) = 0$ and the proof is complete. We may therefore assume that there exists a subsequence $\{\boldsymbol{x}^{k_l,*}\}_{l=1}^\infty$ such that $G(\boldsymbol{x}^{k_l,*}) > 0$ for each integer l.

Since $(\boldsymbol{x}^{k_l,*}, \epsilon^{k_l,*})$ is a local solution of the penalty problem with $\sigma = \sigma_{k_l}$, Lemma 1 implies that $\epsilon^{k_l,*}$ is the unique global solution of Problem ($\mathrm{P}_{\sigma,\boldsymbol{x}}$) with $\sigma = \sigma_{k_l}$ and $\boldsymbol{x} = \boldsymbol{x}^{k_l,*}$. Hence, since $\sigma_{k_l} \to \infty$ as $l \to \infty$, it follows from Lemma 2 that there exists an integer $l_1 > 0$ such that

$$\epsilon^{k_l,*} = \tau(\sigma_{k_l}, \boldsymbol{x}^{k_l,*}) = \left(\frac{\alpha G(\boldsymbol{x}^{k_l,*})}{\sigma_{k_l} \beta} \right)^{\frac{1}{\alpha+\beta}} > 0, \qquad l \geq l_1. \tag{13}$$

Now, assume that $G(\boldsymbol{x}^*) > 0$. Since f and G are continuous, there exists a real number $\gamma > 0$ and an integer $l_2 \geq l_1$ such that for all $l \geq l_2$,

$$f(\boldsymbol{x}^{k_l,*}) > f(\boldsymbol{x}^*) - \gamma,$$

and

$$G(\boldsymbol{x}^{k_l,*}) > G(\boldsymbol{x}^*) - \gamma > 0. \tag{14}$$

Without loss of generality, we may assume that for all $l \geq l_2$,

$$\sigma_{k_l} > \frac{\alpha}{\beta} \left(\frac{M - f(\boldsymbol{x}^*) + \gamma}{\big(G(\boldsymbol{x}^*) - \gamma\big)^{\frac{\beta}{\alpha+\beta}}} \right)^{\frac{\alpha+\beta}{\alpha}}, \tag{15}$$

where M is an upper bound for $F_{\sigma_{k_l}}(\boldsymbol{x}^{k_l,*}, \epsilon^{k_l,*})$ (recall that the sequence of penalty function values is assumed to be bounded). Then for each integer $l \geq l_2$,

$$F_{\sigma_{k_l}}(\boldsymbol{x}^{k_l,*}, \epsilon^{k_l,*}) = f(\boldsymbol{x}^{k_l,*}) + \frac{G(\boldsymbol{x}^{k_l,*})}{(\epsilon^{k_l,*})^\alpha} + \sigma_{k_l}(\epsilon^{k_l,*})^\beta \geq f(\boldsymbol{x}^{k_l,*}) + \frac{G(\boldsymbol{x}^{k_l,*})}{(\epsilon^{k_l,*})^\alpha} > f(\boldsymbol{x}^*) - \gamma + \frac{G(\boldsymbol{x}^{k_l,*})}{(\epsilon^{k_l,*})^\alpha}.$$

It thus follows from (13)-(15) that

$$\begin{aligned}
F_{\sigma_{k_l}}(\boldsymbol{x}^{k_l,*}, \epsilon^{k_l,*}) &> f(\boldsymbol{x}^*) - \gamma + \frac{G(\boldsymbol{x}^{k_l,*})}{(\epsilon^{k_l,*})^\alpha} \\
&= f(\boldsymbol{x}^*) - \gamma + G(\boldsymbol{x}^{k_l,*})^{\frac{\beta}{\alpha+\beta}} \left(\frac{\sigma_{k_l}\beta}{\alpha} \right)^{\frac{\alpha}{\alpha+\beta}} \\
&> f(\boldsymbol{x}^*) - \gamma + \big(G(\boldsymbol{x}^*) - \gamma\big)^{\frac{\beta}{\alpha+\beta}} \left(\frac{\sigma_{k_l}\beta}{\alpha} \right)^{\frac{\alpha}{\alpha+\beta}} \\
&> f(\boldsymbol{x}^*) - \gamma + \big(G(\boldsymbol{x}^*) - \gamma\big)^{\frac{\beta}{\alpha+\beta}} \frac{M - f(\boldsymbol{x}^*) + \gamma}{\big(G(\boldsymbol{x}^*) - \gamma\big)^{\frac{\beta}{\alpha+\beta}}} \\
&= M,
\end{aligned}$$

which contradicts the definition of M. Therefore, our initial assumption that $G(\boldsymbol{x}^*) > 0$ is false. This completes the proof. □

Our next result is concerned with the behavior of global solutions of Problem (P_σ) as $\sigma \to \infty$.

Theorem 3. *Let $\{\sigma_k\}_{k=1}^{\infty}$ be an increasing sequence of penalty parameters such that $\sigma_k \to \infty$ as $k \to \infty$. Furthermore, let $(\boldsymbol{x}^{k,*}, \epsilon^{k,*})$ be a global solution of Problem (P_{σ_k}). Then the sequence $\{\boldsymbol{x}^{k,*}\}_{k=1}^{\infty}$ has at least one limit point, and any limit point is a global solution of Problem (P).*

Proof. It is clear that $\{\boldsymbol{x}^{k,*}\}_{k=1}^{\infty} \subset \mathcal{X}$ has at least one limit point because $\mathcal{X}$ is compact. Let $\bar{\boldsymbol{x}}$ denote one such limit point. Then by passing to a subsequence if necessary, we may assume that $\boldsymbol{x}^{k,*} \to \bar{\boldsymbol{x}}$ as $k \to \infty$.

Now, let $\boldsymbol{x}^*$ be a global solution of Problem (P). Then

$$F_{\sigma_k}(\boldsymbol{x}^{k,*}, \epsilon^{k,*}) \leq F_{\sigma_k}(\boldsymbol{x}^*, 0) = f(\boldsymbol{x}^*), \qquad k \geq 1. \tag{16}$$

Thus, the sequence $\{F_{\sigma_k}(\boldsymbol{x}^{k,*}, \epsilon^{k,*})\}_{k=1}^{\infty}$ is bounded. Theorem 2 then implies that the limit point $\bar{\boldsymbol{x}}$ is feasible for Problem (P).

Since $\sigma_k < \sigma_{k+1}$,

$$F_{\sigma_k}(\boldsymbol{x}, \epsilon) \leq F_{\sigma_{k+1}}(\boldsymbol{x}, \epsilon), \qquad (\boldsymbol{x}, \epsilon) \in \mathcal{X} \times [0, \bar{\epsilon}],$$

which implies

$$F_{\sigma_k}(\boldsymbol{x}^{k,*}, \epsilon^{k,*}) \leq F_{\sigma_k}(\boldsymbol{x}^{k+1,*}, \epsilon^{k+1,*}) \leq F_{\sigma_{k+1}}(\boldsymbol{x}^{k+1,*}, \epsilon^{k+1,*}).$$

Hence, the sequence $\{F_{\sigma_k}(\boldsymbol{x}^{k,*}, \epsilon^{k,*})\}_{k=1}^{\infty}$ is non-decreasing. Since this sequence is also bounded, it must be convergent. From (16), we obtain

$$\lim_{k\to\infty} F_{\sigma_k}(\boldsymbol{x}^{k,*}, \epsilon^{k,*}) \leq f(\boldsymbol{x}^*). \tag{17}$$

On the other hand, since $\bar{\boldsymbol{x}}$ is feasible for Problem (P),

$$\lim_{k\to\infty} F_{\sigma_k}(\boldsymbol{x}^{k,*}, \epsilon^{k,*}) \geq \lim_{k\to\infty} f(\boldsymbol{x}^{k,*}) = f(\bar{\boldsymbol{x}}) \geq f(\boldsymbol{x}^*). \tag{18}$$

Combining (17) and (18) gives

$$f(\bar{\boldsymbol{x}}) = f(\boldsymbol{x}^*).$$

This shows that the limit point $\bar{\boldsymbol{x}}$ is a global solution of Problem (P). □

We now focus on the behavior of local solutions of Problem (P_σ) when $\sigma \to \infty$. Let $\{\sigma_k\}_{k=1}^{\infty}$ be an increasing sequence of penalty parameters such that $\sigma_k \to \infty$ as $k \to \infty$ and suppose that each $(\boldsymbol{x}^{k,*}, \epsilon^{k,*})$ is a local solution of Problem (P_{σ_k}). We assume that any limit point $\boldsymbol{x}^*$ of $\{\boldsymbol{x}^{k,*}\}_{k=1}^{\infty}$ satisfies the following regularity conditions.

(A1) The vectors $\frac{\partial h_i(\boldsymbol{x}^*)}{\partial \boldsymbol{x}}$, $i \in \mathcal{E}$ are linearly independent (assuming that $\mathcal{E} \neq \emptyset$).

(A2) There exists a vector $\boldsymbol{p} = [p_1, \ldots, p_n]^\top \in \mathbb{R}^n$ and real numbers $\vartheta_1 < 0$ and $\vartheta_2 < 0$ such that

$$\begin{aligned}
&\boldsymbol{p}^\top \frac{\partial h_i(\boldsymbol{x}^*)}{\partial \boldsymbol{x}} = 0, \qquad i \in \mathcal{E}, \\
&\boldsymbol{p}^\top \frac{\partial g_i(\boldsymbol{x}^*)}{\partial \boldsymbol{x}} < 0, \qquad i \in \bar{\mathcal{I}}(\boldsymbol{x}^*) \triangleq \{\, j \in \mathcal{I} : \, g_j(\boldsymbol{x}^*) = 0 \,\}, \\
&\boldsymbol{p}^\top \frac{\partial \varphi_i(\boldsymbol{x}^*, \omega)}{\partial \boldsymbol{x}} < \vartheta_1, \qquad \omega \in \Omega_i^{\vartheta_2}(\boldsymbol{x}^*) \triangleq \{\, \eta \in \Omega_i : \, \varphi_i(\boldsymbol{x}^*, \eta) \geq \vartheta_2 \,\}, \qquad i \in \mathcal{C}, \\
&p_j \begin{cases} > 0, & \text{if } x_j^* = a_j, \\ < 0, & \text{if } x_j^* = b_j. \end{cases}
\end{aligned}$$

(A3) There exists real numbers $L > 0$ and $\gamma > 0$ such that for each $i \in \mathcal{C}$,

$$\big[\max\{\varphi_i(\boldsymbol{x}, \omega), 0\}\big]^2 \leq L \int_{\Omega_i} \big[\max\{\varphi_i(\boldsymbol{x}, \eta), 0\}\big]^2 d\eta, \qquad (\boldsymbol{x}, \omega) \in \mathcal{N}_\gamma(\boldsymbol{x}^*) \times \Omega_i,$$

where $\mathcal{N}_\gamma(\boldsymbol{x}^*) \triangleq \{\, \boldsymbol{y} \in \mathcal{X} : \, \|\boldsymbol{y} - \boldsymbol{x}^*\| < \gamma \,\}$.

Note that Assumption (A2) is similar to the standard *Mangasarian-Fromovitz constraint qualification* in mathematical programming (see [4, 5, 8]). Note also that in (A1) and (A2), the partial derivatives $\frac{\partial h_i}{\partial \boldsymbol{x}}$, $\frac{\partial g_i}{\partial \boldsymbol{x}}$, and $\frac{\partial \varphi_i}{\partial \boldsymbol{x}}$ are assumed to be column vectors. We will continue this convention throughout the remainder of the paper.

To prove our main convergence result, we will need the following lemma.

Lemma 4. *Let $\{\sigma_k\}_{k=1}^{\infty}$ be an increasing sequence of penalty parameters such that $\sigma_k \to \infty$ as $k \to \infty$. Furthermore, let $(\boldsymbol{x}^{k,*}, \epsilon^{k,*})$ be a local solution of Problem (P_{σ_k}). Suppose that $\boldsymbol{x}^{k,*} \to \boldsymbol{x}^*$ as $k \to \infty$. Then for all sufficiently large k, the following inequalities hold:*

$$\varphi_i(\boldsymbol{x}^{k,*}, \omega) < 0, \qquad \omega \in \Omega_i \setminus \Omega_i^{\vartheta_2}(\boldsymbol{x}^*), \qquad i \in \mathcal{C}, \tag{19}$$

$$\boldsymbol{p}^\top \frac{\partial \varphi_i(\boldsymbol{x}^{k,*}, \omega)}{\partial \boldsymbol{x}} < \vartheta_1, \qquad \omega \in \Omega_i^{\vartheta_2}(\boldsymbol{x}^*), \qquad i \in \mathcal{C}, \tag{20}$$

where $\boldsymbol{p}$, ϑ_1, and ϑ_2 are as defined in Assumption (A2).

Proof. We first consider the ith inequality in (19). Suppose, to the contrary of the lemma, that there exists a subsequence $\{\boldsymbol{x}^{k_l,*}\}_{l=1}^{\infty}$ and another sequence $\{\omega^l\}_{l=1}^{\infty} \subset \Omega_i \setminus \Omega_i^{\vartheta_2}(\boldsymbol{x}^*)$ such that

$$\varphi_i(\boldsymbol{x}^{k_l,*}, \omega^l) \geq 0, \qquad l \geq 1. \tag{21}$$

Recall that Ω_i is a compact interval. Hence, the sequence $\{\omega^l\}_{l=1}^{\infty} \subset \Omega_i$ is bounded. By invoking the Bolzano-Weierstrass theorem and passing to a subsequence if necessary, we may assume that there exists a $\bar{\omega} \in \Omega_i$ such that $\omega^l \to \bar{\omega}$ as $l \to \infty$. Since φ_i is continuous, taking the limit in (21) gives

$$\varphi_i(\boldsymbol{x}^*, \bar{\omega}) = \lim_{l \to \infty} \varphi_i(\boldsymbol{x}^{k_l,*}, \omega^l) \geq 0. \tag{22}$$

Now, since $\{\omega^l\}_{l=1}^{\infty} \subset \Omega_i \setminus \Omega_i^{\vartheta_2}(\boldsymbol{x}^*)$,

$$\varphi_i(\boldsymbol{x}^*, \omega^l) < \vartheta_2 < 0, \qquad l \geq 1.$$

Thus,

$$\varphi_i(\boldsymbol{x}^*, \bar{\omega}) = \lim_{l \to \infty} \varphi_i(\boldsymbol{x}^*, \omega^l) \leq \vartheta_2 < 0.$$

But this contradicts (22). Hence, there must exist an integer $k_i^1 > 0$ such that the ith inequality in (19) holds for each $k \geq k_i^1$. Since $\boldsymbol{x}^*$ satisfies Assumption (A2), in a similar manner, we can show that there exists an integer $k_i^2 > 0$ such that the ith inequality in (20) holds for each $k \geq k_i^2$. It follows that both (19) and (20) hold for all integers $k \geq \max\{\, k_i^1, k_i^2 : i \in \mathcal{C} \,\}$. This completes the proof. ☐

We are now ready to state our main convergence result showing that the penalty function F_σ is exact. The proof is given in the next section.

Theorem 4. *Let $\{\sigma_k\}_{k=1}^{\infty}$ be an increasing sequence of penalty parameters such that $\sigma_k \to \infty$ as $k \to \infty$. Furthermore, let $(\boldsymbol{x}^{k,*}, \epsilon^{k,*})$ be a local solution of Problem (P_{σ_k}). Suppose that $\{F_{\sigma_k}(\boldsymbol{x}^{k,*}, \epsilon^{k,*})\}_{k=1}^{\infty}$ is bounded. Then for all sufficiently large k, $\boldsymbol{x}^{k,*}$ is a local solution of Problem (P).*

Let $(\boldsymbol{x}^*, \epsilon^*)$ be a local solution of Problem (P_σ). Then according to Theorem 4, when the penalty parameter σ is sufficiently large, $\boldsymbol{x}^*$ is a local solution of Problem (P). Recall also from Lemma 3 that $\boldsymbol{x}^*$ is a local solution of Problem (P) if and only if $\epsilon^* = 0$. On this basis, we can solve Problem (P) by solving Problem (P_σ) sequentially for increasing values of σ, stopping once $\epsilon^* = 0$. The following algorithm is based on this idea.

Algorithm 1. Input $\boldsymbol{x}^0 \in \mathcal{X}$ (initial guess), $\sigma^0 > 0$ (initial penalty parameter), ρ (tolerance), and $\sigma_{\max}$ (maximum penalty parameter).

Step 1. Set $\bar{\epsilon} \to \epsilon^0$ and $\sigma^0 \to \sigma$.

Step 2. Starting with $(\boldsymbol{x}^0, \epsilon^0)$ as the initial guess, use a standard nonlinear programming algorithm to solve Problem (P_σ). Let $(\boldsymbol{x}^*, \epsilon^*)$ denote the local minimizer obtained.

Step 3. If $\epsilon^* < \rho$, then stop: $\boldsymbol{x}^*$ is a (local) solution of Problem (P). Otherwise, set $10\sigma \to \sigma$ and go to Step 4.

Step 4. If $\sigma \leq \sigma_{\max}$, then set $(\boldsymbol{x}^*, \epsilon^*) \to (\boldsymbol{x}^0, \epsilon^0)$ and go to Step 2. Otherwise, stop: the algorithm cannot find a solution of Problem (P).

If Algorithm 1 terminates without finding a solution of Problem (P), then there are two possible remedies that we can try: (i) Choose a different initial guess; or (ii) Adjust the parameters α and β in the exact penalty function.

4. Proof of Theorem 4

We follow the approach suggested in [5]. Suppose that the theorem is false. Then there exists a subsequence $\{\boldsymbol{x}^{k_l,*}\}_{l=1}^{\infty}$ such that for each $l \geq 1$, $\boldsymbol{x}^{k_l,*}$ is *not* a local solution of Problem (P). We will now proceed to derive a contradiction in five steps.

4.1. Preliminaries

Note that $\{\boldsymbol{x}^{k_l,*}\} \subset \mathcal{X}$ and $\mathcal{X}$ is compact. Therefore, by invoking the Bolzano-Weierstrass theorem and passing to a subsequence if necessary, we may assume that there exists a limit point $\boldsymbol{x}^* \in \mathcal{X}$ such that $\boldsymbol{x}^{k_l,*} \to \boldsymbol{x}^*$ as $l \to \infty$. It then follows from Theorem 2 that $G(\boldsymbol{x}^*) = 0$. Also, since $(\boldsymbol{x}^{k_l,*}, \epsilon^{k_l,*})$ is a local solution of the penalty problem, but $\boldsymbol{x}^{k_l,*}$ is not a local solution of Problem (P), Lemma 3 implies that $\epsilon^{k_l,*} > 0$ for each $l \geq 1$.

Now, recall that $\sigma_{k_l} \to \infty$ as $l \to \infty$ and $\epsilon^{k_l,*}$ is the unique global solution of Problem ($\mathrm{P}_{\sigma,\boldsymbol{x}}$) with $\sigma = \sigma_{k_l}$ and $\boldsymbol{x} = \boldsymbol{x}^{k_l,*}$. Thus, it follows from Lemma 2 that there exists an integer l_1 such that

$$\epsilon^{k_l,*} = \left(\frac{\alpha G(\boldsymbol{x}^{k_l,*})}{\sigma_{k_l}\beta}\right)^{\frac{1}{\alpha+\beta}} > 0, \qquad l \geq l_1. \tag{23}$$

Consequently,

$$G(\boldsymbol{x}^{k_l,*}) > 0, \qquad l \geq l_1. \tag{24}$$

Furthermore, by (23) and our requirement that $\alpha \geq \beta$,

$$\lim_{l\to\infty} \frac{(\epsilon^{k_l,*})^{\alpha}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} = \lim_{l\to\infty} G(\boldsymbol{x}^{k_l,*})^{\frac{\alpha-\beta}{2(\alpha+\beta)}} \left(\frac{\alpha}{\sigma_{k_l}\beta}\right)^{\frac{\alpha}{\alpha+\beta}} = 0. \tag{25}$$

4.2. Definition of $\{\boldsymbol{z}^l\}$

By virtue of (24), the following vectors are well-defined for each $l \geq l_1$:

$$\begin{aligned}
\boldsymbol{z}_1^l &= \left[z_{1,i}^l\right]_{i\in\mathcal{E}} \triangleq \frac{1}{\sqrt{G(\boldsymbol{x}^{k_l,*})}}\left[h_i(\boldsymbol{x}^{k_l,*})\right]_{i\in\mathcal{E}}, \\
\boldsymbol{z}_2^l &= \left[z_{2,i}^l\right]_{i\in\mathcal{I}} \triangleq \frac{1}{\sqrt{G(\boldsymbol{x}^{k_l,*})}}\left[\max\{g_i(\boldsymbol{x}^{k_l,*}), 0\}\right]_{i\in\mathcal{I}}, \\
\boldsymbol{z}_3^l &= \left[z_{3,i}^l\right]_{i\in\mathcal{C}} \triangleq \frac{1}{\sqrt{G(\boldsymbol{x}^{k_l,*})}}\left[\left(\int_{\Omega_i}\left[\max\{\varphi_i(\boldsymbol{x}^{k_l,*}, \omega), 0\}\right]^2 d\omega\right)^{\frac{1}{2}}\right]_{i\in\mathcal{C}}, \\
\boldsymbol{z}^l &\triangleq \left[(\boldsymbol{z}_1^l)^\top, (\boldsymbol{z}_2^l)^\top, (\boldsymbol{z}_3^l)^\top\right]^\top.
\end{aligned}$$

It is clear that $\|\boldsymbol{z}^l\| = 1$ and hence the sequence $\{\boldsymbol{z}^l\}_{l=l_1}^{\infty}$ is bounded. Thus, by invoking the Bolzano-Weierstrass theorem and passing to a subsequence if necessary, we may assume that there exists a $\boldsymbol{z}^* \neq \boldsymbol{0}$ such that $\boldsymbol{z}^l \to \boldsymbol{z}^*$ as $l \to \infty$, where

$$\boldsymbol{z}^* \triangleq \left[(\boldsymbol{z}_1^*)^\top, (\boldsymbol{z}_2^*)^\top, (\boldsymbol{z}_3^*)^\top\right]^\top.$$

Clearly, $\|\boldsymbol{z}^*\| = 1$. Furthermore,

$$\begin{aligned} z_{1,i}^* &= \lim_{l\to\infty} \frac{1}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} h_i(\boldsymbol{x}^{k_l,*}), \qquad i \in \mathcal{E}, \\ z_{2,i}^* &= \lim_{l\to\infty} \frac{1}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \max\{g_i(\boldsymbol{x}^{k_l,*}), 0\}, \qquad i \in \mathcal{I}, \\ z_{3,i}^* &= \lim_{l\to\infty} \frac{1}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \left(\int_{\Omega_i} \left[\max\{\varphi_i(\boldsymbol{x}^{k_l,*}, \omega), 0\}\right]^2 d\omega \right)^{\frac{1}{2}}, \qquad i \in \mathcal{C}. \end{aligned}$$

Recall from Section 4.1 that $G(\boldsymbol{x}^*) = 0$. Thus, $g_i(\boldsymbol{x}^*) \leq 0$ for each $i \in \mathcal{I}$. If $g_i(\boldsymbol{x}^*) < 0$, then $g_i(\boldsymbol{x}^{k_l,*}) < 0$ and $\max\{g_i(\boldsymbol{x}^{k_l,*}), 0\} = 0$ whenever l is sufficiently large. Hence,

$$z_{2,i}^* \begin{cases} = 0, & i \in \mathcal{I} \setminus \bar{\mathcal{I}}(\boldsymbol{x}^*), \\ \geq 0, & i \in \bar{\mathcal{I}}(\boldsymbol{x}^*), \end{cases} \tag{26}$$

where $\bar{\mathcal{I}}(\boldsymbol{x}^*)$ is as defined in Assumption (A2).

4.3. Definition of $\{\boldsymbol{u}^l\}$

By (24), we know that the following vectors are well-defined for each $l \geq l_1$:

$$\boldsymbol{u}^l \triangleq \frac{1}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \frac{\partial G(\boldsymbol{x}^{k_l,*})}{\partial \boldsymbol{x}}.$$

It follows from (7) that for each $l \geq l_1$,

$$\begin{aligned} \boldsymbol{u}^l &= \frac{2}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \Bigg[\sum_{i\in\mathcal{E}} h_i(\boldsymbol{x}^{k_l,*}) \frac{\partial h_i(\boldsymbol{x}^{k_l,*})}{\partial \boldsymbol{x}} + \sum_{i\in\mathcal{I}} \max\{g_i(\boldsymbol{x}^{k_l,*}), 0\} \frac{\partial g_i(\boldsymbol{x}^{k_l,*})}{\partial \boldsymbol{x}} \\ &\qquad + \sum_{i\in\mathcal{C}} \int_{\Omega_i} \max\{\varphi_i(\boldsymbol{x}^{k_l,*}, \omega), 0\} \frac{\partial \varphi_i(\boldsymbol{x}^{k_l,*}, \omega)}{\partial \boldsymbol{x}} d\omega \Bigg] \\ &= 2 \Bigg[\sum_{i\in\mathcal{E}} z_{1,i}^l \frac{\partial h_i(\boldsymbol{x}^{k_l,*})}{\partial \boldsymbol{x}} + \sum_{i\in\mathcal{I}} z_{2,i}^l \frac{\partial g_i(\boldsymbol{x}^{k_l,*})}{\partial \boldsymbol{x}} \\ &\qquad + \sum_{i\in\mathcal{C}} \int_{\Omega_i} \frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*}, \omega), 0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \frac{\partial \varphi_i(\boldsymbol{x}^{k_l,*}, \omega)}{\partial \boldsymbol{x}} d\omega \Bigg]. \end{aligned} \tag{27}$$

Next we will show that $\{\boldsymbol{u}^l\}_{l=l_1}^{\infty}$ (or an appropriate subsequence) is convergent.

Since the partial derivatives of φ_i are continuous on the compact set $\mathcal{X} \times \Omega_i$, there exists a real number $M_1 > 0$ such that

$$\left| \frac{\partial \varphi_i(\boldsymbol{x}, \omega)}{\partial x_j} \right| \leq M_1, \qquad (\boldsymbol{x}, \omega) \in \mathcal{X} \times \Omega_i, \qquad j = 1, \ldots, n, \qquad i \in \mathcal{C}.$$

Thus, for each $l \geq l_1$,

$$\left| \int_{\Omega_i} \frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\omega),0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \frac{\partial \varphi_i(\boldsymbol{x}^{k_l,*},\omega)}{\partial x_j} d\omega \right| \leq \int_{\Omega_i} \frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\omega),0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \left| \frac{\partial \varphi_i(\boldsymbol{x}^{k_l,*},\omega)}{\partial x_j} \right| d\omega$$
$$\leq M_1 \int_{\Omega_i} \frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\omega),0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} d\omega. \tag{28}$$

Applying the Hölder inequality (with $p = q = 2$) gives

$$\begin{aligned}\left| \int_{\Omega_i} \frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\omega),0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \frac{\partial \varphi_i(\boldsymbol{x}^{k_l,*},\omega)}{\partial x_j} d\omega \right| &\leq M_1 \int_{\Omega_i} \frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\omega),0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} d\omega \\ &\leq M_1 \sqrt{|\Omega_i|} \left(\int_{\Omega_i} \left[\frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\omega),0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \right]^2 d\omega \right)^{\frac{1}{2}} \\ &= z_{3,i}^l M_1 \sqrt{|\Omega_i|}.\end{aligned}$$

Recall that $z_{3,i}^l \to z_{3,i}^*$ as $l \to \infty$. Thus, the inequalities above show that there exists a real number $M_2 > 0$ such that

$$\left| \int_{\Omega_i} \frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\omega),0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \frac{\partial \varphi_i(\boldsymbol{x}^{k_l,*},\omega)}{\partial x_j} d\omega \right| \leq M_1 \int_{\Omega_i} \frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\omega),0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} d\omega \leq M_2, \qquad l \geq l_1.$$

Therefore, by invoking the Bolzano-Weierstrass theorem and passing to a subsequence if necessary, we may assume that there exist vectors $\boldsymbol{q}^i \in \mathbb{R}^n$ and constants $r^i \in \mathbb{R}$ such that

$$\lim_{l\to\infty} \int_{\Omega_i} \frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\omega),0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \frac{\partial \varphi_i(\boldsymbol{x}^{k_l,*},\omega)}{\partial \boldsymbol{x}} d\omega = \boldsymbol{q}^i, \qquad i \in \mathcal{C}, \tag{29}$$

and

$$\lim_{l\to\infty} \int_{\Omega_i} \frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\omega),0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} d\omega = r^i, \qquad i \in \mathcal{C}. \tag{30}$$

From (26), (27), and (29), we obtain that

$$\lim_{l\to\infty} \boldsymbol{u}^l = \lim_{l\to\infty} \frac{1}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \frac{\partial G(\boldsymbol{x}^{k_l,*})}{\partial \boldsymbol{x}} = 2\left[\sum_{i\in\mathcal{E}} z_{1,i}^* \frac{\partial h_i(\boldsymbol{x}^*)}{\partial \boldsymbol{x}} + \sum_{i\in\bar{\mathcal{I}}(\boldsymbol{x}^*)} z_{2,i}^* \frac{\partial g_i(\boldsymbol{x}^*)}{\partial \boldsymbol{x}} + \sum_{i\in\mathcal{C}} \boldsymbol{q}^i \right]. \tag{31}$$

Let $\boldsymbol{u}^*$ denote the limit of $\boldsymbol{u}^l$. In the next part of the proof, we will investigate the signs of the components of $\boldsymbol{u}^*$.

4.4. Signs of the Components of $\boldsymbol{u}^*$

Recall that $\boldsymbol{x}^{k_l,*} \to \boldsymbol{x}^*$ as $l \to \infty$. Let

$$\begin{aligned} I_1 &= \{\, j \in \{1,\ldots,n\} : x_j^* = a_j \,\}, \\ I_2 &= \{\, j \in \{1,\ldots,n\} : a_j < x_j^* < b_j \,\}, \\ I_3 &= \{\, j \in \{1,\ldots,n\} : x_j^* = b_j \,\}. \end{aligned}$$

Then I_1, I_2, and I_3 form a partition of $\{1,\ldots,n\}$. It is clear that there exists an integer $l_2 \geq l_1$ such that for each $l \geq l_2$,

$$\begin{aligned} a_j &\leq x_j^{k_l,*} < b_j, \qquad j \in I_1, \\ a_j &< x_j^{k_l,*} < b_j, \qquad j \in I_2, \\ a_j &< x_j^{k_l,*} \leq b_j, \qquad j \in I_3. \end{aligned}$$

Recall that $\epsilon^{k_l,*} > 0$ and $(\boldsymbol{x}^{k_l,*}, \epsilon^{k_l,*})$ is a local solution of the penalty problem with $\sigma = \sigma_{k_l}$. Then,

$$\frac{\partial F_{\sigma_{k_l}}(\boldsymbol{x}^{k_l,*}, \epsilon^{k_l,*})}{\partial x_j} = \frac{\partial f(\boldsymbol{x}^{k_l,*})}{\partial x_j} + (\epsilon^{k_l,*})^{-\alpha} \frac{\partial G(\boldsymbol{x}^{k_l,*})}{\partial x_j} \begin{cases} \geq 0, & \text{if } x_j^{k_l,*} = a_j, \\ = 0, & \text{if } a_j < x_j^{k_l,*} < b_j, \\ \leq 0, & \text{if } x_j^{k_l,*} = b_j. \end{cases}$$

Thus, for each $l \geq l_2$,

$$\frac{\partial f(\boldsymbol{x}^{k_l,*})}{\partial x_j} + (\epsilon^{k_l,*})^{-\alpha} \frac{\partial G(\boldsymbol{x}^{k_l,*})}{\partial x_j} \begin{cases} \geq 0, & \text{if } j \in I_1, \\ = 0, & \text{if } j \in I_2, \\ \leq 0, & \text{if } j \in I_3, \end{cases}$$

and

$$\frac{(\epsilon^{k_l,*})^{\alpha}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \frac{\partial f(\boldsymbol{x}^{k_l,*})}{\partial x_j} + \frac{1}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \frac{\partial G(\boldsymbol{x}^{k_l,*})}{\partial x_j} = \frac{(\epsilon^{k_l,*})^{\alpha}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \frac{\partial f(\boldsymbol{x}^{k_l,*})}{\partial x_j} + u_j^l \begin{cases} \geq 0, & \text{if } j \in I_1, \\ = 0, & \text{if } j \in I_2, \\ \leq 0, & \text{if } j \in I_3. \end{cases}$$

Consequently, by virtue of (25) and (31),

$$u_j^* \begin{cases} \geq 0, & \text{if } j \in I_1, \\ = 0, & \text{if } j \in I_2, \\ \leq 0, & \text{if } j \in I_3. \end{cases} \tag{32}$$

4.5. A Contradiction

We now complete the proof by showing that $\boldsymbol{z}^* = \boldsymbol{0}$. This is a contradiction because we have already seen in Section 4.2 that $\|\boldsymbol{z}^*\| = 1$.

Note that Assumptions (A1)-(A3) hold for $\boldsymbol{x} = \boldsymbol{x}^*$. By Assumption (A2), there exists a vector $\boldsymbol{p} \in \mathbb{R}^n$ and real numbers $\vartheta_1 < 0$ and $\vartheta_2 < 0$ such that

$$\boldsymbol{p}^\top \frac{\partial h_i(\boldsymbol{x}^*)}{\partial \boldsymbol{x}} = 0, \qquad i \in \mathcal{E}, \tag{33}$$

$$\boldsymbol{p}^\top \frac{\partial g_i(\boldsymbol{x}^*)}{\partial \boldsymbol{x}} < 0, \qquad i \in \bar{\mathcal{I}}(\boldsymbol{x}^*), \tag{34}$$

$$\boldsymbol{p}^\top \frac{\partial \varphi_i(\boldsymbol{x}^*, \omega)}{\partial \boldsymbol{x}} < \vartheta_1, \qquad \omega \in \Omega_i^{\vartheta_2}(\boldsymbol{x}^*), \qquad i \in \mathcal{C},$$

$$p_j \begin{cases} > 0, & \text{if } j \in I_1, \\ < 0, & \text{if } j \in I_3. \end{cases} \tag{35}$$

By (32) and (35),

$$\boldsymbol{p}^\top \boldsymbol{u}^* \geq 0. \tag{36}$$

Furthermore, (26), (31), (33), and (34) give

$$\begin{aligned} \boldsymbol{p}^\top \boldsymbol{u}^* &= 2\left[\sum_{i \in \mathcal{E}} z_{1,i}^* \boldsymbol{p}^\top \frac{\partial h_i(\boldsymbol{x}^*)}{\partial \boldsymbol{x}} + \sum_{i \in \bar{\mathcal{I}}(\boldsymbol{x}^*)} z_{2,i}^* \boldsymbol{p}^\top \frac{\partial g_i(\boldsymbol{x}^*)}{\partial \boldsymbol{x}} + \sum_{i \in \mathcal{C}} \boldsymbol{p}^\top \boldsymbol{q}^i\right] \\ &= 2\left[\sum_{i \in \bar{\mathcal{I}}(\boldsymbol{x}^*)} z_{2,i}^* \boldsymbol{p}^\top \frac{\partial g_i(\boldsymbol{x}^*)}{\partial \boldsymbol{x}} + \sum_{i \in \mathcal{C}} \boldsymbol{p}^\top \boldsymbol{q}^i\right] \\ &\leq 2 \sum_{i \in \mathcal{C}} \boldsymbol{p}^\top \boldsymbol{q}^i. \end{aligned} \tag{37}$$

Note that equality holds in (37) if and only if $\boldsymbol{z}_2^* = \mathbf{0}$ (according to (26), we automatically have $\boldsymbol{z}_2^* = \mathbf{0}$ when $\bar{\mathcal{I}}(\boldsymbol{x}^*) = \emptyset$).

Since $\boldsymbol{x}^{k_l,*} \to \boldsymbol{x}^*$ as $l \to \infty$, Lemma 4 implies that there exists an integer $l_3 \geq l_2$ such that for each $l \geq l_3$,

$$\varphi_i(\boldsymbol{x}^{k_l,*}, \omega) < 0, \qquad \omega \in \Omega_i \setminus \Omega_i^{\vartheta_2}(\boldsymbol{x}^*), \qquad i \in \mathcal{C},$$

and

$$\boldsymbol{p}^\top \frac{\partial \varphi_i(\boldsymbol{x}^{k_l,*}, \omega)}{\partial \boldsymbol{x}} < \vartheta_1, \qquad \omega \in \Omega_i^{\vartheta_2}(\boldsymbol{x}^*), \qquad i \in \mathcal{C}.$$

It thus follows from (29) and (30) that

$$\begin{aligned}
\sum_{i\in\mathcal{C}} \boldsymbol{p}^\top \boldsymbol{q}^i &= \sum_{i\in\mathcal{C}} \lim_{l\to\infty} \int_{\Omega_i} \frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*}, \omega), 0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \boldsymbol{p}^\top \frac{\partial \varphi_i(\boldsymbol{x}^{k_l,*}, \omega)}{\partial \boldsymbol{x}} d\omega \\
&= \sum_{i\in\mathcal{C}} \lim_{l\to\infty} \int_{\Omega_i^{\vartheta_2}(\boldsymbol{x}^*)} \frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*}, \omega), 0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \boldsymbol{p}^\top \frac{\partial \varphi_i(\boldsymbol{x}^{k_l,*}, \omega)}{\partial \boldsymbol{x}} d\omega \\
&\leq \vartheta_1 \sum_{i\in\mathcal{C}} \lim_{l\to\infty} \int_{\Omega_i^{\vartheta_2}(\boldsymbol{x}^*)} \frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*}, \omega), 0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} d\omega \\
&= \vartheta_1 \sum_{i\in\mathcal{C}} \lim_{l\to\infty} \int_{\Omega_i} \frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*}, \omega), 0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} d\omega \\
&= \vartheta_1 \sum_{i\in\mathcal{C}} r^i. \qquad (38)
\end{aligned}$$

Recall that $\vartheta_1 < 0$. Note also that $r^i \geq 0$ for each $i \in \mathcal{C}$. Hence, by (37) and (38),

$$\boldsymbol{p}^\top \boldsymbol{u}^* \leq 2 \sum_{i\in\mathcal{C}} \boldsymbol{p}^\top \boldsymbol{q}^i \leq 2\vartheta_1 \sum_{i\in\mathcal{C}} r^i \leq 0. \qquad (39)$$

Combining (36) and (39) gives

$$\boldsymbol{p}^\top \boldsymbol{u}^* = 0. \qquad (40)$$

Hence, the equality in (37) holds, so we must have $\boldsymbol{z}_2^* = \mathbf{0}$.

Now, inequalities (39) and (40) give

$$r^i = \lim_{l\to\infty} \int_{\Omega_i} \frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*}, \omega), 0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} d\omega = 0, \qquad i \in \mathcal{C}. \qquad (41)$$

Thus, by (28) and (30),

$$\lim_{l\to\infty} \int_{\Omega_i} \frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*}, \omega), 0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \frac{\partial \varphi_i(\boldsymbol{x}^{k_l,*}, \omega)}{\partial x_j} d\omega = 0, \qquad j = 1, \ldots, n, \qquad i \in \mathcal{C}.$$

That is, $\boldsymbol{q}^i = \mathbf{0}$ for each $i \in \mathcal{C}$. Substituting $\boldsymbol{z}_2^* = \mathbf{0}$ and $\boldsymbol{q}^i = \mathbf{0}$ into (31) gives

$$\boldsymbol{u}^* = \lim_{l\to\infty} \boldsymbol{u}^l = 2 \sum_{i\in\mathcal{E}} z_{1,i}^* \frac{\partial h_i(\boldsymbol{x}^*)}{\partial \boldsymbol{x}}.$$

But we know that $\boldsymbol{p}^\top \boldsymbol{u}^* = 0$, and so it follows from (32) and (35) that $\boldsymbol{u}^* = \mathbf{0}$. That is,

$$\boldsymbol{u}^* = 2 \sum_{i\in\mathcal{E}} z_{1,i}^* \frac{\partial h_i(\boldsymbol{x}^*)}{\partial \boldsymbol{x}} = \mathbf{0}.$$

Recall from Assumption (A1) that $\frac{\partial h_i(\boldsymbol{x}^*)}{\partial \boldsymbol{x}}$, $i \in \mathcal{E}$ are linearly independent vectors. Thus, we must have $\boldsymbol{z}_1^* = \mathbf{0}$.

To conclude the proof, we will show that $\boldsymbol{z}_3^* = \mathbf{0}$. From equation (41), there exists a subsequence, which we denote by the original sequence, such that

$$\lim_{l\to\infty} \frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\omega),0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} = 0, \qquad \text{a.e. } \omega \in \Omega_i, \qquad i \in \mathcal{C}. \tag{42}$$

Now, by Assumption (A3), there exists an integer $l_4 \geq l_3$ such that for each $l \geq l_4$,

$$\big[\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\omega),0\}\big]^2 \leq L \int_{\Omega_i} \big[\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\eta),0\}\big]^2 d\eta, \qquad \omega \in \Omega_i, \qquad i \in \mathcal{C}. \tag{43}$$

For each fixed $l \geq l_4$, we consider two cases:

(i) $\int_{\Omega_i} \big[\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\eta),0\}\big]^2 d\eta = 0$.

(ii) $\int_{\Omega_i} \big[\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\eta),0\}\big]^2 d\eta > 0$.

In case (i), the continuity of φ_i implies that $\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\eta),0\} = 0$ for all $\eta \in \Omega_i$. In case (ii), inequality (43) implies that

$$\left[\frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\omega),0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}}\right]^2 \leq \frac{\big[\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\omega),0\}\big]^2}{\int_{\Omega_i} \big[\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\eta),0\}\big]^2 d\eta} \leq L, \qquad \omega \in \Omega_i.$$

Hence, in both cases, the quotient of $\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\omega),0\}$ and $\sqrt{G(\boldsymbol{x}^{k_l,*})}$ is uniformly bounded with respect to ω and l.

This means that we can use (42) in conjunction with Lebesgue's dominated convergence theorem to obtain

$$\begin{aligned} z_{3,i}^* &= \lim_{l\to\infty} \frac{1}{\sqrt{G(\boldsymbol{x}^{k_l,*})}} \left(\int_{\Omega_i} \big[\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\omega),0\}\big]^2 d\omega \right)^{\frac{1}{2}} \\ &= \left(\lim_{l\to\infty} \int_{\Omega_i} \left[\frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\omega),0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}}\right]^2 d\omega \right)^{\frac{1}{2}} \\ &= \left(\int_{\Omega_i} \lim_{l\to\infty} \left[\frac{\max\{\varphi_i(\boldsymbol{x}^{k_l,*},\omega),0\}}{\sqrt{G(\boldsymbol{x}^{k_l,*})}}\right]^2 d\omega \right)^{\frac{1}{2}} = 0, \qquad i \in \mathcal{C}. \end{aligned}$$

Hence $\boldsymbol{z}_3^* = \mathbf{0}$, as required. This completes the proof.

5. Numerical Examples

To demonstrate the effectiveness and power of Algorithm 1, we consider three example problems, each of which involves difficult continuous inequality constraints. To solve Problem (P_σ), we use the Fortran subroutine NLPQLP [13]. The integrals in the constraint violation function G are evaluated using Simpson's rule. The fixed parameters in Algorithm 1 are $\bar{\epsilon} = 10$, $\sigma^0 = 1$, $\sigma_{\max} = 10^6$, and $\rho = 10^{-6}$.

Example 1. Consider the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) = \frac{x_2(122 + 17x_1 + 6x_3 - 5x_2 + x_1x_3) + 180x_3 - 36x_1 + 1244}{x_2(408 + 56x_1 - 50x_2 + 60x_3 + 10x_1x_3 - 2x_1^2)} \\ \text{subject to} \quad & \varphi(\boldsymbol{x},\omega) = \mathrm{Im}(T(\boldsymbol{x},\omega)) - 3.33\big[\mathrm{Re}(T(\boldsymbol{x},\omega))\big]^2 + 1 \leq 0, \qquad \omega \in [10^{-6}, 30], \\ & 0 \leq x_1,\ x_3 \leq 100, \qquad 0.1 \leq x_2 \leq 100, \end{aligned}$$

where

$$T(\boldsymbol{x},\omega) = 1 + \frac{\left(x_1 + \frac{x_2}{\mathrm{i}\omega} + \mathrm{i}x_3\omega\right)}{(3+\mathrm{i}\omega)(2-\omega^2+\mathrm{i}2\omega)}, \qquad \mathrm{i}^2 = -1.$$

This problem is a PID control problem in which the aim is to choose the PID compensator gains x_1, x_2, and x_3 to minimize the mean-square error [3, 7, 15, 17]. We solve the problem using Algorithm 1 with the following inputs:

$$\boldsymbol{x}^0 = [50, 50, 50]^\top, \qquad \alpha = 2, \qquad \beta = 2.$$

Note that this initial guess of $\boldsymbol{x}^0 = [50, 50, 50]^\top$ was also used in [17]. Here, we choose 10^{-8} as the error tolerance for NLPQLP and partition the integration interval into 30,000 subintervals.

Algorithm 1 gives the following optimal solution:

$$x_1^* = 16.9559238246, \qquad x_2^* = 45.4397319897, \qquad x_3^* = 34.6736696052.$$

The corresponding optimal cost is $f(\boldsymbol{x}^*) = 0.1746273739$. The value of $\varphi(\boldsymbol{x}^*, \omega)$ on $[10^{-6}, 30]$ is always less than -1.5×10^{-9}, and thus the solution produced by Algorithm 1 is feasible. Note that this solution is slightly better than the solution reported in [17], which has an optimal cost of 0.174778004.

Example 2. Consider the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) = (x_1 + x_2 - 2)^2 + (x_1 - x_2)^2 + 30\big[\min\{0, x_1 - x_2\}\big]^2 \\ \text{subject to} \quad & \varphi(\boldsymbol{x}, \omega) = x_1 \cos(\omega) + x_2 \sin(\omega) - 1 \le 0, \qquad \omega \in [0, \pi]. \end{aligned}$$

This problem was considered in [17]. Since there are no bound constraints, we choose sufficiently large finite numbers for the bounds on x_1 and x_2.

As in Example 1, we choose 10^{-8} as the error tolerance for NLPQLP and partition the integration interval into 30,000 subintervals. We run Algorithm 1 with the initial guess $\boldsymbol{x}^0 = [0.5, 0.5]^\top$ and fixed parameters

$$\alpha = 2, \qquad \beta = 2.$$

Algorithm 1 gives the following optimal solution:

$$x_1^* = 0.7071240599, \qquad x_2^* = 0.7070895002.$$

The corresponding optimal cost is $f(\boldsymbol{x}^*) = 0.3431457543$ and the solution satisfies the given continuous inequality constraint (the value of $\varphi(\boldsymbol{x}^*, \omega)$ on $[0, \pi]$ is always less than -1.7×10^{-9}). As with Example 1, Algorithm 1 here outperforms the algorithm in [17], which gives an optimal cost of 0.3432592109 for this example.

Example 3. Consider the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) = \boldsymbol{x}^\top H \boldsymbol{x} - 2\boldsymbol{c}^\top \boldsymbol{x} \\ \text{subject to} \quad & g_1(\boldsymbol{x}, \omega) = \left|\boldsymbol{\phi}(\omega)^\top \boldsymbol{x} - 1\right| - 0.05 \le 0, \qquad \omega \in [0, 0.05], \\ & g_2(\boldsymbol{x}, \omega) = \left|\boldsymbol{\phi}(\omega)^\top \boldsymbol{x}\right| - 0.01 \le 0, \qquad \omega \in [0.1, 0.5], \end{aligned}$$

where $\boldsymbol{x} \in \mathbb{R}^{18}$ and

$$\begin{aligned} \boldsymbol{\phi}(\omega) &= [2\cos(34\pi\omega),\, 2\cos(32\pi\omega),\, \ldots\, , 2\cos(2\pi\omega),\, 1]^\top, \\ H &= \int_0^{0.05} \boldsymbol{\phi}(\omega)\boldsymbol{\phi}(\omega)^\top d\omega + 1000 \int_{0.1}^{0.5} \boldsymbol{\phi}(\omega)\boldsymbol{\phi}(\omega)^\top d\omega, \\ \boldsymbol{c} &= \int_0^{0.05} \boldsymbol{\phi}(\omega) d\omega. \end{aligned}$$

x_1^*	x_2^*	x_3^*	x_4^*	x_5^*	x_6^*
0.0056459702	0.0030387544	0.0006012952	-0.0032460689	-0.0079542542	-0.0143351706
x_7^*	x_8^*	x_9^*	x_{10}^*	x_{11}^*	x_{12}^*
-0.0203065017	-0.0224669279	-0.0224005907	-0.0143936721	-0.0008034870	0.0195827365
x_{13}^*	x_{14}^*	x_{15}^*	x_{16}^*	x_{17}^*	x_{18}^*
0.0453308083	0.0733901025	0.1001082981	0.1234568179	0.1379288243	0.1436543078

Table 1: Optimal solution for Example 3

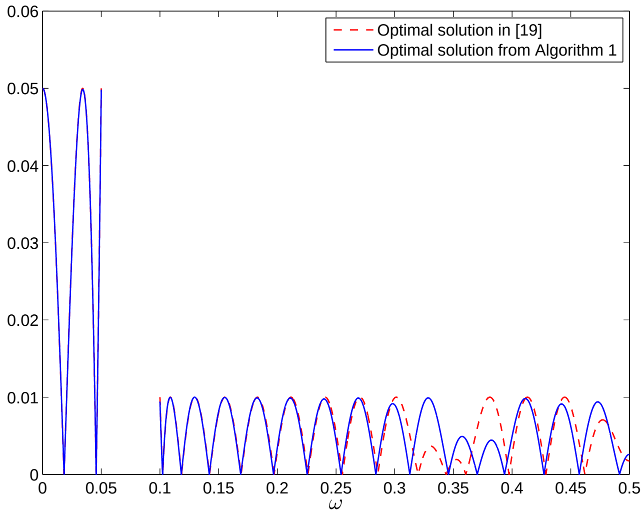

Figure 1: Plots of $|\boldsymbol{\phi}(\omega)^\top \boldsymbol{x} - 1|$ (left) and $|\boldsymbol{\phi}(\omega)^\top \boldsymbol{x}|$ (right) for the optimal solutions from [19] and Algorithm 1.

This problem is a FIR filter design problem [2, 19].

Here, we choose 10^{-5} as the error tolerance for NLPQLP and partition the integration interval into 2,000 subintervals. The inputs to Algorithm 1 are

$$\boldsymbol{x}^0 = [1, \dots, 1]^\top, \qquad \alpha = 2, \qquad \beta = 1.$$

The optimal solution obtained from Algorithm 1 is given in Table 1 and the corresponding cost is -0.0386933156. Note that this optimal solution is feasible:

$$\max_{\omega \in [0, 0.05]} g_1(\boldsymbol{x}^*, \omega) < -8.1 \times 10^{-6}, \qquad \max_{\omega \in [0.1, 0.5]} g_2(\boldsymbol{x}^*, \omega) < -1.7 \times 10^{-6}.$$

The cost of the optimal solution given in [19] is -0.0383876371, which is similar to our result. However, the solution in [19] slightly violates the continuous inequality constraints, with the maximum values of g_1 and g_2 being greater than or equal to 8.9×10^{-8} and 2.29×10^{-7}, respectively. The continuous inequality constraints g_1 and g_2 are plotted in Figure 1 for the optimal solution in [19] and the optimal solution produced by Algorithm 1.

6. Conclusion

In this paper, we have introduced a new exact penalty method for solving nonlinear semi-infinite programming problems. Our new method is based on the exact penalty function in [17, 18]. Our work in this paper has shown that: (i) the convergence results in [17, 18] can be considerably strengthened; and (ii) the number of fixed parameters in the exact penalty function in [17, 18] can be reduced. Future research will involve developing rules for selecting optimal values of the fixed parameters to yield fast convergence.

References

[1] R. Brayton, G. D. Hachtel and A. L. Sangiovanni-Vincentelli, A survey of optimization techniques for integrated-circuit design, Proceedings of the IEEE 69 (1981) 1334-1362.

[2] H. H. Dam, K. L. Teo, S. Nordebo and A. Cantoni, The dual parameterization approach to optimal least square FIR filter design subject to maximum error constraints, IEEE Transactions on Signal Processing 48 (2000) 2314-2320.

[3] C. Gonzaga, E. Polak and R. Trahan, An improved algorithm for optimization problems with functional inequality constraints, IEEE Transactions on Automatic Control AC-25 (1980) 49-54.

[4] R. Hettich and K. O. Kortanek, Semi-infinite programming: Theory, methods, and applications, SIAM Review 35 (1993) 380-429.

[5] W. Huyer and A. Neumaier, A new exact penalty function, SIAM Journal on Optimization 13 (2003) 1141-1158.

[6] C. Jiang, Q. Lin, C. Yu, K. L. Teo and G. Duan, An exact penalty method for free terminal time optimal control problem with continuous inequality constraints, Journal of Optimization Theory and Applications 154 (2012) 30-53.

[7] L. S. Jennings and K. L. Teo, A computational algorithm for functional inequality constrained optimization problems, Automatica 26 (1990) 371-375.

[8] M. López and G. Still, Semi-infinite programming, European Journal of Operational Research 180 (2007) 491-518.

[9] D. G. Luenberger and Y. Ye, Linear and Nonlinear Programming, 3rd edition. New York: Springer, 2008.

[10] J. Nocedal and S. J. Wright, Numerical Optimization, 2nd edition. New York: Springer, 2006.

[11] S. Nordebo, Z. Zang and I. Claesson, A semi-infinite quadratic programming algorithm with applications to array pattern synthesis, IEEE Transactions on Circuits and Systems – Analog and Digital Signal Processing 48 (2001) 225-232.

[12] E. R. Panier and A. L.Tits, A globally convergent algorithm with adaptively refined discretization for semi-infinite optimization problems arising in engineering design, IEEE Transactions on Automatic Control 34 (1989) 903-908.

[13] K. Schittkowski, NLPQLP: A Fortran implementation of a sequential quadratic programming algorithm with distributed and non-monotone line search, Version 2.24, User's Guide, Bayreuth: University of Bayreuth, 2007.

[14] K. L. Teo and C. J. Goh, A simple computational procedure for optimization problems with functional inequality constraints, IEEE Transactions on Automatic Control AC-32, (1987) 940-941.

[15] K. L. Teo, V. Rehbock and L. S. Jennings, A new computational algorithm for functional inequality constrained optimization problems, Automatica 29 (1993) 789-792.

[16] L. Y. Wang, W. H. Gui, K. L. Teo, R. Loxton and C. H. Yang, Time delayed optimal control problems with multiple characteristic time points: Computation and industrial applications, Journal of Industrial and Management Optimization 5 (2009) 705-718.

[17] C. Yu, K. L. Teo, L. Zhang and Y. Bai, A new exact penalty function method for continuous inequality constrained optimization problems, Journal of Industrial and Management Optimization 6 (2010) 895-910.

[18] C. Yu, K. L. Teo, L. Zhang and Y. Bai, On a refinement of the convergence analysis for the new exact penalty function method for continuous inequality constrained optimization problem, Journal of Industrial and Management Optimization 8 (2012) 485-491.

[19] L. Zhang, S. Y. Wu and M. A. López, A new exchange method for convex semi-infinite programming, SIAM Journal on Optimization 20 (2010) 2959-2977.