Chapter 6

Augmented Lagrangian Cone Method for MOPs with an Application to an Optimal Control Problem

6.1 Introduction

The augmented Lagrangian Method (ALM) is a well-established method specifically developed for solving nonlinear constrained scalar optimization problems. This method has gained popularity in various applications due to its effectiveness and efficiency. Furthermore, it has been extensively analyzed and discussed in the literature from a theoretical perspective. For those interested in a comprehensive overview of the ALM, we recommend consulting reference [14] .

The ALM stands out as a powerful technique for addressing constrained optimization challenges. It adeptly manages constraints through an iterative approach, which enhances the precision of the solutions generated. This iterative process not only improves solution accuracy but also helps in navigating the complexities often associated with constrained optimization.

6.2 Motivation

The motivation for developing new methods to solve MOPs arises from the inherent limitations and specific requirements of existing approaches. First, there is a clear need to adapt conventional numerical optimization methods to handle multiple objectives effectively, as traditional methods are often designed for single objective problems. Additionally, it is crucial to develop methods that perform well for both convex and nonconvex problems, as exemplified by the cone method, which provides a versatile framework for addressing these challenges [60]. Furthermore, existing methods often struggle with problems that involve numerous constraints, particularly when both equality and inequality constraints are present, making it essential to devise approaches that can manage these complexities more effectively [14]. Another key consideration is improving computational efficiency, as MOPs are typically computationally intensive; hence, there is a need for methods that can efficiently solve subproblems for each direction in the objective space. Robustness is also a critical factor, with a demand for algorithms that incorporate global convergence results to ensure reliability across various problem instances [14]. Finally, the practical applicability of these methods is of paramount importance, as they should be versatile enough to be applied across a wide range of realistic and diverse applications, thereby demonstrating their value in real-world scenarios [2, 7, 51].

In this chapter, we propose an ALM with exact penalization to solve MOPs without needing prior information about Pareto surface locations or function convexity. The cone method is implemented to convert an MOP into parametric single objective problems, which are then transformed into unconstrained subproblems. These are solved using steepest descent with a max-type nonmonotone line search. We prove the global convergence of the algorithm and demonstrate its effectiveness through standard test problems and a real-world unemployment control model.

The decision to employ the ALM with exact penalization and a nonmonotone approach is driven by the need to effectively handle the complexities of MOPs, particularly when dealing with nonconvex objective functions and constraints. The ALM, with its exact penalization, transforms constrained problems into unconstrained ones, ensuring that the penalized problems closely approximate the original MOP, thus preserving the integrity of the solution. The incorporation of a nonmonotone line search enhances the algorithm's convergence behavior, allowing it to escape local minima and accelerate convergence by permitting occasional increases in the objective function. This combination not only guarantees convergence to globally optimal and feasible solutions but also enhances the algorithm's robustness and efficiency, making it particularly suitable for complex real-world applications, such as the unemployment optimal control model discussed in your research. This approach provides a strong theoretical foundation while offering practical advantages in solving large-scale and challenging MOPs.

6.3 Contribution

The main contributions of this research are

(i) we integrate the cone method with the ALM to effectively solve MOPs. The cone method, initially proposed by Ghosh and Chakraborty [60], converts an MOP into parametric singleobjective subproblems, but lacked a detailed solution method. Our research addresses this gap by applying ALM to these subproblems.

(ii) Unlike traditional optimization methods that may struggle with numerous equality and inequality constraints, our use of ALM is flexible and efficient, providing a robust solution framework that is not constrained by these limitations [14].

(iii) We prove the global convergence of the proposed algorithm, ensuring that any subsequential limit of the generated sequence is a global minimizer of the infeasibility measure, even in complex scenarios with nonempty feasible regions.

(iv) A notable feature of the proposed method is its robustness to variable scaling. The solution quality remains unaffected by changes in the scale of variables, enhancing the method's reliability and applicability across diverse problem settings.

(v) The proposed method is not only theoretically sound but also applicable to real-world problems [20,21,23]. We demonstrate its effectiveness through standard test problems and a practical application to a deterministic unemployment optimal control model, showcasing its utility in addressing realistic challenges [2, 7, 51].

(vi) By providing a detailed method to solve the direction-based parametric singleobjective optimization problems obtained from the cone method, this research addresses a gap in the original cone method proposal [60].

In this chapter we consider the constrained MOP:

$$\left.\begin{array}{lll} \min & F(x) = \ (f_1(x), f_2(x), \ldots, f_r(x))^\top, & r \geq 2, \\ \text{subject to} & g_i(x) \leq 0, & i = 1, 2, \ldots, p, \\ & h_j(x) = 0, & j = 1, 2, \ldots, q, \end{array}\right\} \tag{6.1}$$

where $x = (x_1, x_2, \ldots, x_n)^\top \in \mathbb{R}^n$ and f_m, $m = 1, 2, \ldots, r$, g_i and h_j are real-valued functions. In this chapter, we assume that each f_m, g_i and h_j are continuously differentiable.

Set of all efficient points and nondominated points are denoted by $\mathcal{X}_E$ and $\mathcal{Y}_N$, respectively. The collection of all weakly efficient points and weakly non-dominated points are represented by $\mathcal{X}_{wE}$ and $\mathcal{Y}_{wN}$, respectively. For further detail, we refer to [5, 40, 87].

The primary requirement to apply the cone method is that $\mathcal{Y}$ must lie in the nonnegative hyperoctant of $\mathbb{R}^r$ [60]. In order to make $\mathcal{Y}$ a subset of the nonnegative hyperoctant of $\mathbb{R}^r$, we take the following assumption.

A(0) We assume that for each $m = 1, 2, \ldots, r$, $\min_{x \in \mathcal{X}} f_m(x)$ exists.

Then, we redefine $f_m(x)$ by $f_m(x) - f_m^*$, where $f_m^* = \min_{x \in \mathcal{X}} f_m(x)$, $m = 1, 2, \ldots, r$.

To capture a (weakly) efficient point of the MOP (6.1), the cone method [60,61,119] solves the following singleobjective optimization problem:

$$\left.\begin{array}{lll} \min & t & \\ \text{subject to} & t\hat{\beta} \geqq F(x) & \\ & g_i(x) \leq 0, & i = 1, 2, \ldots, p, \\ & h_j(x) = 0, & j = 1, 2, \ldots, q, \\ & t \geq 0, & \end{array}\right\} \tag{6.2}$$

where $\hat{\beta} \in \mathbb{S}_+^{r-1}$, and $\mathbb{S}_+^{r-1}$ is the intersection of the first octant $\mathbb{R}_+^r$ and the unit sphere S^{r-1} in $\mathbb{R}^r$.

To find the complete Pareto set (see [60]) of the MOP (6.1), we need to solve (6.2) for each

$$\hat{\beta} = (\beta_1, \beta_2, \ldots, \beta_r) = \left(\cos\theta_1,\ \cos\theta_2 \sin\theta_1,\ \cos\theta_3 \sin\theta_2 \sin\theta_1, \ldots,\ \cos\theta_{r-1} \prod_{i=1}^{r-2} \sin\theta_i,\ \prod_{i-1}^{r-1} \sin\theta_i\right) \in \mathbb{S}_+^{r-1}, \tag{6.3}$$

where $\theta_i \in [0, \frac{\pi}{2}]$, $i = 1, 2, \ldots, (r-1)$. In the proposed algorithm, we follow the suggested way in [60] on number of grid points for θ_i's to obtain uniformly spreaded Pareto points.

In the next section, we construct an augmented Lagrangian function corresponding to (6.1) using the auxiliary problems (6.2) for different $\hat{\beta}$'s.

6.4 Construction of Augmented Lagrangian Function for MOPs

Consider the MOP (6.1) with the assumptions that the objective functions and the constraints are continuously differentiable in $\mathcal{X}$. To obtain a (weakly) Pareto optimal solution of (6.1) we attempt to solve (6.2). Notice that (6.2) is equivalent to

$$\left.\begin{array}{llll} \min & t & & \\ \text{subject to} & -t\hat{\beta} + F(x) + z = 0, & z \geqq 0, & \\ & \quad g_i(x) + s_i = 0, & s_i \geq 0, & i = 1, 2, \ldots, p, \\ & \quad h_j(x) = 0, & & j = 1, 2, \ldots, q, \\ & -t + a = 0, & a \geq 0, & \end{array}\right\} \tag{6.4}$$

where $z = (z_1, z_2, \ldots, z_r)^\top$, $s = (s_1, s_2, \ldots, s_p)^\top$ and a are slack variables. The ALM for the problem (6.4) involves solving the following problem (see [12]):

$$\min_{\substack{(x,t)\in\mathbb{R}^n\times\mathbb{R}_+\\ z\geq 0, s\geq 0, a\geq 0}} \overline{L_\eta}(x, t, z, s, a, \mu, v, \lambda, w), \tag{6.5}$$

where

$$\begin{aligned}
&\overline{L_\eta}(x, t, z, s, a, \mu, v, \lambda, w)\\
&= t + \sum_{m=1}^{r} \left\{ \mu_m \left(-t\beta_m + f_m(x) + z_m\right) + \tfrac{\eta}{2}\left(-t\beta_m + f_m(x) + z_m\right)^2 \right\}\\
&\quad + \sum_{i=1}^{p} \left\{ v_i \left(g_i(x) + s_i\right) + \tfrac{\eta}{2}\left(g_i(x) + s_i\right)^2 \right\}\\
&\quad + \sum_{j=1}^{q} \left\{ \lambda_j h_j(x) + \tfrac{\eta}{2}\left(h_j(x)\right)^2 \right\} + \left\{ w\left(-t + a\right) + \tfrac{\eta}{2}\left(-t + a\right)^2 \right\},
\end{aligned}$$

$\mu = (\mu_1, \mu_2, \ldots, \mu_r) \in \mathbb{R}^r_+$, $v = (v_1, v_2, \ldots, v_p) \in \mathbb{R}^p_+$, $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_q) \in \mathbb{R}^q$, $w \in \mathbb{R}_+$ and $\eta > 0$. The minimization in (6.5) can be done [12] by first minimizing $\overline{L_\eta}(x, t, z, s, a, \mu, v, \lambda, w)$ over the slack variables $z \geqq 0$, $s \geqq 0$ and $a \geq 0$, obtaining

$$L_\eta(x, t, \mu, v, \lambda, w) = \min_{z\geqq 0, s\geqq 0, a\geq 0} \overline{L_\eta}(x, t, z, s, a, \mu, v, \lambda, w), \tag{6.6}$$

and then by minimizing $L_\eta(x, t, \mu, v, \lambda, w)$ over $(x, t) \in \mathbb{R}^n \times \mathbb{R}_+$.

Note that the functions in first, second, and fourth braces of the expression of $\overline{L_\eta}$ are quadratic in z_m, s_i and a, respectively. Thus, it is easy to get a closed form expression of the function in the left of (6.6) for each fixed (x, t). Certainly, for a given value of

(x, t), we have

$$\begin{aligned}
&\min_{z \geqq 0, s \geqq 0, a \geq 0} \overline{L_\eta}(x, t, z, s, a, \mu, v, \lambda, w) \\
&= t + \sum_{m=1}^{r} \min_{z_m \geq 0} \left\{ \mu_m \left(-t\beta_m + f_m(x) + z_m \right) + \tfrac{\eta}{2} \left(-t\beta_m + f_m(x) + z_m \right)^2 \right\} \\
&\quad + \sum_{i=1}^{p} \min_{s_i \geq 0} \left\{ v_i \left(g_i(x) + s_i \right) + \tfrac{\eta}{2} \left(g_i(x) + s_i \right)^2 \right\} \\
&\quad + \sum_{j=1}^{q} \left\{ \lambda_j h_j(x) + \tfrac{\eta}{2} \left(h_j(x) \right)^2 \right\} + \min_{a \geq 0} \left\{ w(-t + a) + \tfrac{\eta}{2} (-t + a)^2 \right\}.
\end{aligned} \tag{6.7}$$

Minimum points of the expressions in the first, second and third braces on the right of (6.7) are

$$\begin{aligned}
&\hat{z}_m = \max \left\{ 0, -\left(\tfrac{\mu_m}{\eta} - t\beta_m + f_m(x) \right) \right\} \text{for all } m = 1, 2, \ldots, r, \\
&\hat{s}_i = \max \left\{ 0, -\left(\tfrac{v_i}{\eta} + g_i(x) \right) \right\} \text{ and } \hat{a} = \max \left\{ 0, -\left(\tfrac{w}{\eta} - t \right) \right\}, \text{respectively.}
\end{aligned}$$

Denoting

$$\begin{aligned}
f_{m+}(x, t, \mu_m, \eta) &= \max \left\{ -t\beta_m + f_m(x), -\tfrac{\mu_m}{\eta} \right\}, \\
g_{i+}(x, v_i, \eta) &= \max \left\{ g_i(x), -\tfrac{v_i}{\eta} \right\} \\
\text{and } t_+(t, w, \eta) &= \max \left\{ -t, -\tfrac{w}{\eta} \right\},
\end{aligned}$$

we have $-t\beta_m + f_m(x) + \hat{z}_m = f_{m+}(x, t, \beta_m, \mu_m, \eta)$, $g_i(x) + \hat{s}_i = g_{i+}(x, v_i, \eta)$, and $-t + \hat{a} =$

$t_+(t, w, \eta)$. Substituting these expressions in (6.7), we get a closed form expression for

$$\begin{aligned}
&L_\eta(x, t, \mu, v, \lambda, w) \\
&= t + \sum_{j=1}^{q} \left\{ \lambda_j h_j(x) + \tfrac{\eta}{2} \left(h_j(x) \right)^2 \right\} \\
&\quad + \sum_{m=1}^{r} \left\{ \mu_m f_{m+}(x, t, \mu_m, \eta) + \tfrac{\eta}{2} \left(f_{m+}(x, t, \mu_m, \eta) \right)^2 \right\} \\
&\quad + \sum_{i=1}^{p} \left\{ v_i g_{i+}(x, v_i, \eta) + \tfrac{\eta}{2} \left(g_{i+}(x, v_i, \eta) \right)^2 \right\} + \left\{ w t_+(t, w, \eta) + \tfrac{\eta}{2} \left(t_+(t, w, \eta) \right)^2 \right\}, \qquad (6.8)
\end{aligned}$$

and we regard it as the augmented Lagrangian function for the problem (6.4).

The ALM [13] for the problem (6.4) can be described by sequences of minimization in the following form:

$$(x^{k+1}, t^{k+1}) \in \underset{(x,t) \in \mathbb{R}^n \times \mathbb{R}_+}{\arg\min}\ L_{\eta^k}(x, t, \mu^k, v^k, \lambda^k, w^k) \qquad (6.9)$$

followed by the multiplier-iterations

$$\left.\begin{aligned}
\mu_m^{k+1} &= \mu_m^k + \eta^k f_{m+}(x^k, t^k, \mu^k, \eta^k), && m = 1, 2, \ldots, r, \\
v_i^{k+1} &= v_i^k + \eta^k g_{i+}(x^k, v^k, \eta^k), && i = 1, 2, \ldots, p, \\
\lambda_j^{k+1} &= \lambda_j^k + \eta^k h_j(x^k), && j = 1, 2, \ldots, q, \\
w^{k+1} &= w^k + \eta^k t_+(t^k, w^k, \eta^k).
\end{aligned}\right\} \qquad (6.10)$$

We denote

$$\left.\begin{aligned}
\mathcal{F}_m^k &= \min\left\{ t^k \beta_m - f_m(x^k), \tfrac{\mu_m^k}{\eta^k} \right\}, && m = 1, 2, \ldots, r, \\
\mathcal{G}_i^k &= \min\left\{ -g_i(x^k), \tfrac{v_i^k}{\eta^k} \right\}, && i = 1, 2, \ldots, p, \\
\mathcal{T}^k &= \min\left\{ t^k, \tfrac{w^k}{\eta^k} \right\}.
\end{aligned}\right\}$$

The update of the sequence $\{\eta^k\}$ is given by

$$\eta^{k+1} = \begin{cases} \eta^k & \text{if } k = 1 \text{ or } \xi_k \leq \gamma \xi_{k-1} \text{ for some } \gamma \in (0,1), \\ \alpha \eta^k \text{ for some } \alpha > 1 & \text{otherwise,} \end{cases} \tag{6.11}$$

where

$\xi_k = \max\left\{\left\|h(x^k)\right\|, \left\|\mathcal{F}^k\right\|, \left\|\mathcal{G}^k\right\|, \left\|\mathcal{T}^k\right\|\right\}$ and $\xi_{k-1} = \max\left\{\left\|h(x^{k-1})\right\|, \left\|\mathcal{F}^{k-1}\right\|, \left\|\mathcal{G}^{k-1}\right\|, \left\|\mathcal{T}^{k-1}\right\|\right\}$.

The quantities η^k's are termed as "penalty parameters." Mainly, at each iteration, we intend to solve (6.9). If adequate progress has been made in terms of improvement of complementarity and feasibility, then we may use the same penalty parameter at the next iteration. Otherwise, the penalty parameter must be increased.

The quantities $\mu^k/\eta^k, \upsilon^k/\eta^k$, $w^k/\eta^k \in \mathbb{R}_+$ and $\lambda^k/\eta^k \in \mathbb{R}$ are called "shifts." Careful choice of shift makes a desirable progress to the solution of (6.9).

Shifts are updated according to the equation (6.13) in Algorithm 8 below. The working mechanism of shifts are as follows. Assume that for a fixed direction $\hat{\beta}$ and for some $k \in \mathbb{N}$, we get the approximate minimizer (x^k, t^k) by solving (6.12) with the shifts $\mu^k/\eta^k, \upsilon^k/\eta^k$ and λ^k/η^k. If $-t^k\hat{\beta} + f(x^k) > 0$, i.e., (x^k, t^k) violates the constraint $-t^k\hat{\beta} + F(x^k) \leq 0$, then the shift has to be increased by the quantity by which (x^k, t^k) has violated, i.e., $\mu^{k+1}/\eta^k = \mu^k/\eta^k + (-t^k\hat{\beta} + F(x^k))$, or equivalently $\mu^{k+1} = \mu^k + \eta^k(-t^k\hat{\beta} + F(x^k))$. Furthermore, if the feasibility has not been violated and $-\mu^k/\eta^k < -t^k\hat{\beta} + F(x^k) \leq 0$, the shift has become extremely big and reduction is required, i.e., $\mu^{k+1}/\eta^k = \mu^k/\eta^k + (-t^k\hat{\beta} + F(x^k))$, to get $-t\hat{\beta} + F(x) = 0$ at the next iteration along with a desired improvement of the objective function (see p.34, [14]). At last, if $-t^k\hat{\beta} + F(x^k) < -\mu^k/\eta^k$, we may predict that the shift was not required and should be null for the next iterate. Similar interpretation are applied for the update of the shifts for the rest of the inequality and equality constraints of MOP (6.2). In the context of augmented Lagrangian, some of the authors [14] suggest to update either penalty parameters or Lagrange multipliers but not both. Although our formulation in Algorithm 8 permits updating of either of the parameters, still we choose to update both the parameters simultaneously.

The following Algorithm 8 provides a complete sequential procedure to obtain the Pareto set of a triobjective optimization problem. First, we evaluate the minimum f_m^* for each of the objectives f_m. Then, we replace the objective function f_m by $f_m - f_m^*$ that translate the objective feasible region to nonnegative orthant $\mathbb{R}^3_+$. Then, we convert the MOP into a parametric singleobjective optimization problem as given in [60]. We chose $E_{\text{sup}} > 0$, λ_{inf} and λ_{sup} arbitrarily to bound the range of multipliers corresponding to inequality and equality constraints. The first element of the sequence of penalty parameter, $\eta^1 > 0$, is chosen arbitrarily, and for each iteration, it is updated according to (6.11). For a triobjective optimization problem, it is necessary to run two 'for' loops (corresponding to θ_1 and θ_2), and in the case of r objectives, the number of 'for' loops will be $r - 1$, for each θ_m, $m = 1, 2, \ldots, r - 1$ (see [60]). With the given setting of θ_1 and θ_2, once we get a direction $\hat{\beta} \in \mathbb{S}^2_+$, we then formulate a direction-based parametric unconstrained optimization problem using ALM. Thus, the solution of the augmented Lagrangian subproblem results a Pareto optimal point on the Pareto surface. We remark here that Algorithm 8 does not require convexity assumption of the objective and constraint functions.

Algorithm 8 Augmented Lagrangian cone method to generate the complete Pareto set of (6.1) with three objectives

Aim: To generate a discrete approximation of the complete Pareto set of the problem (6.1)

1: Provide n, and the functions g_i's and h_j's of the feasible set $\mathcal{X} = \{x \in \mathbb{R}^n : g_i(x) \leq 0, h_j(x) = 0, i = 1, 2, \ldots, p, j = 1, 2, \ldots, q\}$
2: Provide $F(x) = (f_1(x), f_2(x), f_3(x))^\top$, where each of f_1, f_2 and f_3 is a continuously differentiable function on $\mathbb{R}^n$
3: Evaluate $f_m^* = \min\{f_m(x) : x \in \mathcal{X}\}$ for each $m = 1, 2, 3$
4: Replace the objective $f_m(x)$ by $f_m(x) - f_m^*$ for all $m = 1, 2, 3$
5: Provide the tolerance level $\varepsilon > 0$ for the optimum solution to the problem (6.9)
6: Choose $E_{\sup}, \lambda_{\inf}, \lambda_{\sup}$ arbitrarily such that $E_{\sup} > 0$ and $\lambda_{\inf} < \lambda_{\sup}$
7: Choose $\overline{\mu}^1 \in [0, E_{\sup}]^3$, $\overline{v}^1 \in [0, E_{\sup}]^p$, $\overline{w}^1 \in [0, E_{\sup}]$, and $\overline{\lambda}^1 \in [\lambda_{\inf}, \lambda_{\sup}]^q$ arbitrarily
8: Set the Pareto set $\mathcal{S} \leftarrow \emptyset$
9: Choose m_1, the number of grid points for θ_1
10: **for** $\theta_1 = 0 : \frac{\pi}{2m_1} : \frac{\pi}{2}$ **do**
11: Set $m_2 \leftarrow round(m_1 \sin\theta_1)$
12: **for** $\theta_2 = 0 : \frac{\pi}{2m_2} : \frac{\pi}{2}$ **do**
13: (By the equation (6.3)) Set $\hat{\beta} \leftarrow (\cos\theta_1, \cos\theta_2 \sin\theta_1, \sin\theta_2 \sin\theta_1)$
14: Choose $\eta^1 > 0$ (the penalty parameter) arbitrarily
15: **for** $k = 1 : 1 : 2$ **do**
16: Find the expression of $L_{\eta^k}(x, t, \overline{\mu}^k, \overline{\lambda}^k, \overline{v}^k, \overline{w}^k)$ using (6.8)
17: Find $(x^k, t^k) \in \mathbb{R}^n \times \mathbb{R}_+$ such that

$$(x^k, t^k) = \underset{t \geq 0}{\operatorname{argmin}}\ L_{\eta^k}(x, t, \overline{\mu}^k, \overline{\lambda}^k, \overline{v}^k, \overline{w}^k) \tag{6.12}$$

18: Compute a new approximation of augmented Lagrange multipliers by

$$\left.\begin{aligned} \lambda^{k+1} &\leftarrow \overline{\lambda}^k + \eta^k h(x^k) \\ \mu^{k+1} &\leftarrow \max\{0, \overline{\mu}^k + \eta^k(-t\hat{\beta} + f(x^k))\} \\ v^{k+1} &\leftarrow \max\{0, \overline{v}^k + \eta^k g(x^k)\} \\ w^{k+1} &\leftarrow \max\{0, \overline{w}^k + \eta^k(-t^k)\} \end{aligned}\right\} \tag{6.13}$$

19: **Update** η^k by η^{k+1} according to (6.11)
20: Choose $\overline{\mu}^{k+1} \in [0, E_{\sup}]^3, \overline{v}^{k+1} \in [0, E_{\sup}]^p, \overline{w}^{k+1} \in [0, E_{\sup}]$, and $\overline{\lambda}^{k+1} \in [\lambda_{\inf}, \lambda_{\sup}]^q$
21: **end for**
22: **while** $\left|t^k - t^{k-1}\right| > \varepsilon$ **do**
23: Set $k \leftarrow k + 1$
24: Execute the steps 16–20
25: **end while**
26: **return** $\bar{x} = x^k$
27: Update the set $\mathcal{S} \leftarrow \mathcal{S} \bigcup \{F(\bar{x})\}$
28: **end for**
29: **end for**
30: **return** $\mathcal{S}$ as a discrete approximation of the complete Pareto set of the problem (6.1)

Remark 1 We remark here that the well-definedness of Algorithm 8 depends upon

Steps 3 and 17. Step 3 is well-defined due to Assumption A(0). The well-definedness of Step 17 is shown in Section 6.6.

Remark 2 In Step 20, the multipliers $(\overline{\lambda}^{k+1}, \overline{\mu}^{k+1}, \overline{v}^{k+1}, \overline{w}^{k+1})$ can be chosen as the projection of multipliers $(\lambda^k, \mu^k, v^k, w^k)$ onto the given safeguarded intervals in Step 20 of Algorithm 8. For instance, the projection rule for λ^k can be taken as follows:

(i) If $\lambda^{k+1} > \lambda_{\sup}$, then $\overline{\lambda}^{k+1} = \lambda_{\sup}$.

(ii) If $\lambda^{k+1} < \lambda_{\inf}$, then $\overline{\lambda}^{k+1} = \lambda_{\inf}$.

(iii) If $\lambda^{k+1} \in [\lambda_{\inf}, \lambda_{\sup}]$ then $\overline{\lambda}^{k+1} = \lambda^{k+1}$.

6.5 Results on Global Minimization

In this section, we look into the subproblem (6.12) of Algorithm 8 in terms of global optimization. We assume that at each inner iteration, corresponding to each direction $\hat{\beta} = (\beta_1, \beta_2, \ldots, \beta_r)^\top \in \mathbb{S}_+^{r-1}$, (x^k, t^k) is an approximate global minimizer of L_{η_k}.

Throughout this section, we take the following assumption.

A(1) For each given $\hat{\beta} = (\beta_1, \beta_2 \ldots, \beta_r)^\top \in \mathbb{S}_+^{r-1}$, the point (x^k, t^k), obtained from Algorithm 8, satisfies

$$L_{\eta^k}(x^k, t^k, \hat{\beta}, \overline{\mu}^k, \overline{\lambda}^k, \overline{v}^k, \overline{w}^k) \leq L_{\eta^k}(x, t, \hat{\beta}, \overline{\mu}^k, \overline{\lambda}^k, \overline{v}^k, \overline{w}^k) + \tau_k \tag{6.14}$$

for all $(x, t) \in \mathbb{R}^n \times \mathbb{R}_+$, where $\{\tau_k\}$ is a given bounded sequence of tolerance.

This assumption states that after each inner iteration for each direction, we obtain a point which is an approximate global minimizer of L_{η^k} with tolerance τ_k. Assuming A(1) holds, we prove in the next theorem that a subsequence of $\{(x^k, t^k)\}$ converges to a global minimizer of the infeasibility measure. We note that, for any $(x, t) \in \mathbb{R}^n \times \mathbb{R}_+$, the infeasibility measure of the problem (6.12) is given by

$$\left\|h(x)\right\|_2^2 + \left\|g(x)_+\right\|_2^2 + \left\|(-t\hat{\beta} + F(x))_+\right\|_2^2 + \left\|(-t)_+\right\|_2^2,$$

where $g(x)_+ = \max\{0, g(x)\}$, $(-t\hat{\beta} + F(x))_+ = \max\{0, -t\hat{\beta} + F(x)\}$ and $(-t)_+ = \max\{0, -t\}$.

Theorem 6.1 (Feasibility result). *Assume that A(0) and A(1) holds. Let $\{(x^k, t^k)\}$ be the sequence, for a $\hat{\beta} \in \mathbb{S}_+^{r-1}$, generated by Algorithm 8. Let $K \underset{\infty}{\subset} \mathbb{N}$ be such that $(x^k, t^k) \xrightarrow{k\in K} (x^*, t^*)$. Then, for all $(x, t) \in \mathbb{R}^n \times \mathbb{R}_+$, we have*

$$\begin{aligned} &\left\|h(x^*)\right\|_2^2 + \left\|g(x^*)_+\right\|_2^2 + \left\|(-t^*\hat{\beta} + F(x^*))_+\right\|_2^2 + \left\|(-t^*)_+\right\|_2^2 \\ \leq\ &\left\|h(x)\right\|_2^2 + \left\|g(x)_+\right\|_2^2 + \left\|(-t\hat{\beta} + F(x))_+\right\|_2^2 + \left\|(-t)_+\right\|_2^2. \end{aligned} \tag{6.15}$$

Proof: Since $\mathbb{R}^n \times \mathbb{R}_+$ is a closed set and $(x^k, t^k) \in \mathbb{R}^n \times \mathbb{R}_\geq$, we have $(x^*, t^*) \in \mathbb{R}^n \times \mathbb{R}_+$. We consider two cases on the sequence of penalty parameters:

(i) The sequence $\{\eta^k\}$ is bounded.

(ii) The sequence $\{\eta^k\}$ tends to positive infinity.

Case (i). From the definition of sequence of penalty parameter (6.11), the terms of sequence η^k either satisfies $\eta^{k+1} = \eta^k$ or $\eta^{k+1} = \alpha\eta^k$ with $\alpha > 1$, which indicates that η^k is a monotonic increasing sequence. Thus, for $\{\eta^k\}$ to be bounded, the number of times the equality $\eta^{k+1} = \alpha\eta^k$ satisfies should be finitely many. So, there exists k' such that $\eta^k = \eta^{k'}$ for all $k \geq k'$. Therefore, (6.11) holds for all $k \geq k'$. This implies $\left\|h(x^k)\right\|_2 \to 0$, $\left\|\mathcal{F}^k\right\|_2 \to 0$, $\left\|\mathcal{G}^k\right\|_2 \to 0$, and $\left\|\mathcal{T}^k\right\|_2 \to 0$. Hence, $(-t^k\beta_m + F(x^k))_+ \to 0$ for all $m = 1, 2, \ldots, r$, $g_i(x^k)_+ \to 0$ for all $i = 1, 2, \ldots, p$ and $(-t^k)_+ \to 0$. Thus, (6.15) holds trivially.

Case (ii). Let $\eta^k \to \infty$. Let $K \underset{\infty}{\subset} \mathbb{N}$ be such that $(x^k, t^k) \xrightarrow{k\in K} (x^*, t^*)$. Assume that there exists $(x,t) \in \mathbb{R}^n \times \mathbb{R}_+$, such that

$$\begin{aligned}&\left\|h(x^*)\right\|_2^2 + \left\|g(x^*)_+\right\|_2^2 + \left\|(-t^*\hat{\beta} + F(x^*))_+\right\|_2^2 + \left\|(-t^*)_+\right\|_2^2 \\ &> \left\|h(x)\right\|_2^2 + \left\|g(x)_+\right\|_2^2 + \left\|(-t\hat{\beta} + F(x))_+\right\|_2^2 + \left\|(-t)_+\right\|_2^2.\end{aligned}$$

By the continuity of h, g and F, the boundedness of $\{\overline{\mu}^k\}, \{\overline{v}^k\}, \{\overline{\lambda}^k\}$ and $\{\overline{w}^k\}$, and the fact that $\eta^k \to \infty$, there exist $c > 0$ and $k' \in \mathbb{N}$ such that for all $k \geq k'$,

$$\begin{aligned}&\left\|h(x^k) + \tfrac{\overline{\lambda}^k}{\eta^k}\right\|_2^2 - \left\|\tfrac{\overline{\lambda}^k}{\eta^k}\right\|_2^2 + \left\|\left(g(x^k) + \tfrac{\overline{v}^k}{\eta^k}\right)_+\right\|_2^2 - \left\|\tfrac{\overline{v}^k}{\eta^k}\right\|_2^2 + \left\|\left(-t^k\hat{\beta} + F(x^k) + \tfrac{\overline{\mu}^k}{\eta^k}\right)_+\right\|_2^2 - \\ &\left\|\tfrac{\overline{\mu}^k}{\eta^k}\right\|_2^2 + \left\|\left(-t^k + \tfrac{\overline{w}^k}{\eta^k}\right)_+\right\|_2^2 - \left\|\tfrac{\overline{w}^k}{\eta^k}\right\|_2^2 \\ > &\left\|h(x) + \tfrac{\overline{\lambda}^k}{\eta^k}\right\|_2^2 - \left\|\tfrac{\overline{\lambda}^k}{\eta^k}\right\|_2^2 + \left\|\left(g(x) + \tfrac{\overline{v}^k}{\eta^k}\right)_+\right\|_2^2 - \left\|\tfrac{\overline{v}^k}{\eta^k}\right\|_2^2 + \left\|\left(-t\hat{\beta} + F(x) + \tfrac{\overline{\mu}^k}{\eta^k}\right)_+\right\|_2^2 - \\ &\left\|\tfrac{\overline{\mu}^k}{\eta^k}\right\|_2^2 + \left\|\left(-t + \tfrac{\overline{w}^k}{\eta^k}\right)_+\right\|_2^2 - \left\|\tfrac{\overline{w}^k}{\eta^k}\right\|_2^2 + c.\end{aligned}$$

Therefore, for all $k \in K$, $k \geq k'$, we have

$$L_{\eta^k}(x^k, t^k, \overline{\mu}^k, \overline{\lambda}^k, \overline{v}^k, \overline{w}^k) > L_{\eta^k}(x, t, \overline{\mu}^k, \overline{\lambda}^k, \overline{v}^k, \overline{w}^k) + \tfrac{\eta^k c}{2} + t^k - t. \tag{6.16}$$

Since $t^k \xrightarrow{k \in K} t^*$, and $\{\tau_k\}$ is a given bounded sequence of tolerance, there exists $k'' \geq k'$ such that for all $k \geq k''$

$$\tfrac{\eta^k c}{2} + t^k - t > \tau_k \;\; \text{(see Theorem 5.1 of [14])}. \tag{6.17}$$

Therefore, for $k > k''$, form (6.16) and (6.17), we can write

$$L_{\eta^k}(x^k, t^k, \overline{\mu}^k, \overline{\lambda}^k, \overline{v}^k, \overline{w}^k) > L_{\eta^k}(x, t, \overline{\mu}^k, \overline{\lambda}^k, \overline{v}^k, \overline{w}^k) + \tau_k,$$

which contradicts A(1). Hence, (6.15) holds. □

Theorem 6.1 states that a subsequential limit of $\{(x^k, t^k)\}$ obtained from Algorithm 8 corresponding to each direction is feasible if the original problem (6.4) is feasible. Note that for the proof of Theorem 6.1, the only assumption we made on the sequence $\{\tau_k\}$ is that it is bounded. In the next theorem, we show that if we assume $\tau_k \downarrow 0$ and the optimization problem (6.4) is feasible, then corresponding to each direction $\hat{\beta}$, we can achieve the global minimizer of (6.4).

Theorem 6.2 (Optimality result). *Assume that for a fixed direction $\hat{\beta} \in \mathbb{S}_+^{r-1}$, $\{(x^k, t^k)\}$ is the sequence generated by Algorithm 8 under assumptions A(0), A(1) and $\tau_k \downarrow 0$. Let $K \underset{\infty}{\subset} \mathbb{N}$ be such that $(x^k, t^k) \xrightarrow{k \in K} (x^*, t^*)$. Suppose that the problem* (6.4) *is feasible. Then, (x^*, t^*) is a global minimizer of* (6.4).

Proof: Let $\hat{\beta} \in \mathbb{S}_+^{r-1}$ and $K \underset{\infty}{\subset} \mathbb{N}$ be such that $(x^k, t^k) \xrightarrow{k \in K} (x^*, t^*)$. Since the problem (6.4) is feasible, by Theorem 6.1, the point (x^*, t^*) is feasible. Let $(x, t) \in \mathbb{R}^n \times \mathbb{R}_+$ be such that $-t\hat{\beta} + F(x) \leqq 0$, $g(x) \leqq 0$, $-t \leq 0$ and $h(x) = 0$. We consider two cases on the sequences of penalty parameter:

(i) $\{\eta^k\}$ tends to positive infinity.

(ii) $\{\eta^k\}$ is bounded.

Case (i). By assumption A(1), for all $k \in \mathbb{N}$, we have

$$\begin{aligned} & t^k + \tfrac{\eta^k}{2}\left[\left\|h(x^k) + \tfrac{\overline{\lambda}^k}{\eta^k}\right\|_2^2 + \left\|\left(g(x^k) + \tfrac{\overline{v}^k}{\eta^k}\right)_+\right\|_2^2 + \left\|\left(-t^k\hat{\beta} + F(x^k) + \tfrac{\overline{\mu}^k}{\eta^k}\right)_+\right\|_2^2 + \left\|\left(-t^k + \tfrac{\overline{w}^k}{\eta^k}\right)_+\right\|_2^2\right] \\ & \leq t + \tfrac{\eta^k}{2}\left[\left\|h(x) + \tfrac{\overline{\lambda}^k}{\eta^k}\right\|_2^2 + \left\|\left(g(x) + \tfrac{\overline{v}^k}{\eta^k}\right)_+\right\|_2^2 + \left\|\left(-t\hat{\beta} + F(x) + \tfrac{\overline{\mu}^k}{\eta^k}\right)_+\right\|_2^2 + \left\|\left(-t + \tfrac{\overline{w}^k}{\eta^k}\right)_+\right\|_2^2\right] + \tau_k. \end{aligned} \tag{6.18}$$

Since $-t\hat{\beta} + F(x) \leqq 0$, $h(x) = 0$, $g(x) \leqq 0$ and $-t \leq 0$, we obtain

$$\begin{aligned} & \left\|\left(-t\hat{\beta} + F(x) + \tfrac{\overline{\mu}^k}{\eta^k}\right)_+\right\|_2^2 \leq \left\|\tfrac{\overline{\mu}^k}{\eta^k}\right\|_2^2, \ \left\|h(x) + \tfrac{\overline{\lambda}^k}{\eta^k}\right\|_2^2 = \left\|\tfrac{\overline{\lambda}^k}{\eta^k}\right\|_2^2, \\ & \left\|\left(g(x) + \tfrac{\overline{v}^k}{\eta^k}\right)_+\right\|_2^2 \leq \left\|\tfrac{\overline{v}^k}{\eta^k}\right\|_2^2 \text{and} \left\|\left(-t + \tfrac{\overline{w}^k}{\eta^k}\right)_+\right\|_2^2 \leq \left\|\tfrac{\overline{w}^k}{\eta^k}\right\|_2^2. \end{aligned} \tag{6.19}$$

Therefore, by (6.18) and (6.19), we can write the following inequality

$$\begin{aligned} t^k &\le t^k + \frac{\eta^k}{2}\left[\left\|h(x^k) + \frac{\overline{\lambda}^k}{\eta^k}\right\|_2^2 + \left\|\left(g(x^k) + \frac{\overline{v}^k}{\eta^k}\right)_+\right\|_2^2 + \left\|\left(-t^k\hat{\beta} + F(x^k) + \frac{\overline{\mu}^k}{\eta^k}\right)_+\right\|_2^2 + \left\|\left(-t^k + \frac{\overline{w}^k}{\eta^k}\right)_+\right\|_2^2\right] \\ &\le t + \frac{\left\|\overline{\lambda}^k\right\|_2^2}{2\eta^k} + \frac{\left\|\overline{v}^k\right\|_2^2}{2\eta^k} + \frac{\left\|\overline{\mu}^k\right\|_2^2}{2\eta^k} + \frac{\left\|\overline{w}^k\right\|_2^2}{2\eta^k} + \tau_k. \end{aligned} \tag{6.20}$$

Since $(x^k, t^k) \xrightarrow{k\in K} (x^*, t^*)$, $\left(\frac{\left\|\overline{\lambda}^k\right\|_2^2}{\eta^k}, \frac{\left\|\overline{v}^k\right\|_2^2}{\eta^k}, \frac{\left\|\overline{\mu}^k\right\|_2^2}{\eta^k}, \frac{\left\|\overline{w}^k\right\|_2^2}{\eta^k}\right) \xrightarrow{k\in K} (0,0,0,0)$ and $\tau_k \downarrow 0$, so by taking limits on the inequality in (6.20) for $k \in K$, we get $t^* \le t$.

Since t is arbitrarily chosen, (x^*, t^*) is a global minimizer of (6.4) .

Case (ii). In this case, there exists $k' \in \mathbb{N}$ such that $\eta^k = \eta^{k'}$ for all $k \ge k'$. Therefore, by the assumption A(1), for all $k \ge k'$, we have

$$\begin{aligned} & t^k + \frac{\eta^{k'}}{2}\left[\left\|h(x^k) + \frac{\overline{\lambda}^k}{\eta^{k'}}\right\|_2^2 + \left\|\left(g(x^k) + \frac{\overline{v}^k}{\eta^{k'}}\right)_+\right\|_2^2 + \left\|\left(-t^k\hat{\beta} + F(x^k) + \frac{\overline{\mu}^k}{\eta^{k'}}\right)_+\right\|_2^2 + \left\|\left(-t^k + \frac{\overline{w}^k}{\eta^{k'}}\right)_+\right\|_2^2\right] \\ \le\ & t + \frac{\eta^{k'}}{2}\left[\left\|h(x) + \frac{\overline{\lambda}^k}{\eta^{k'}}\right\|_2^2 + \left\|\left(g(x) + \frac{\overline{v}^k}{\eta^{k'}}\right)_+\right\|_2^2 + \left\|\left(-t\hat{\beta} + F(x) + \frac{\overline{\mu}^k}{\eta^{k'}}\right)_+\right\|_2^2 + \left\|\left(-t + \frac{\overline{w}^k}{\eta^{k'}}\right)_+\right\|_2^2\right] + \tau_k. \end{aligned} \tag{6.21}$$

Therefore, by (6.19), for all $k \ge k''$, we can write (6.21) as

$$\begin{aligned} & t^k + \frac{\eta^{k'}}{2}\left[\left\|h(x^k) + \frac{\overline{\lambda}^k}{\eta^{k'}}\right\|_2^2 + \left\|\left(g(x^k) + \frac{\overline{v}^k}{\eta^{k'}}\right)_+\right\|_2^2 + \left\|\left(-t^k\hat{\beta} + F(x^k) + \frac{\overline{\mu}^k}{\eta^{k'}}\right)_+\right\|_2^2 + \left\|\left(-t^k + \frac{\overline{w}^k}{\eta^{k'}}\right)_+\right\|_2^2\right] \\ \le\ & t + \frac{\eta^{k'}}{2}\left[\left\|\frac{\overline{\lambda}^k}{\eta^{k'}}\right\|_2^2 + \left\|\frac{\overline{v}^k}{\eta^{k'}}\right\|_2^2 + \left\|\frac{\overline{\mu}^k}{\eta^{k'}}\right\|_2^2 + \left\|\frac{\overline{w}^k}{\eta^{k'}}\right\|_2^2\right] + \tau_k. \end{aligned} \tag{6.22}$$

Let $K_0 \underset{\infty}{\subset} K, \lambda^* \in \mathbb{R}^q, v^* \in \mathbb{R}_+^p, \mu^* \in \mathbb{R}_+^r$ and $w^* \in \mathbb{R}_+$ such that $(\overline{\lambda}^k, \overline{v}^k, \overline{\mu}^k, \overline{w}^k) \xrightarrow{k\in K_0} (\overline{\lambda}^*, \overline{v}^*, \overline{\mu}^*, \overline{w}^*)$.

By the feasibility of (x^*, t^*), taking the limits in the inequality (6.22) for $k \in K_0$, we get

$$\begin{aligned} & t^* + \frac{\eta^{k'}}{2}\left[\left\|\frac{\overline{\lambda}^*}{\eta^{k'}}\right\|_2^2 + \left\|\left(g(x^*) + \frac{\overline{v}^*}{\eta^{k'}}\right)_+\right\|_2^2 + \left\|\left(-t^*\hat{\beta} + F(x^*) + \frac{\overline{\mu}^*}{\eta^{k'}}\right)_+\right\|_2^2 + \left\|\left(-t^* + \frac{\overline{w}^*}{\eta^{k'}}\right)_+\right\|_2^2\right] \\ \le\ & t + \frac{\eta^{k'}}{2}\left[\left\|\frac{\overline{\lambda}^*}{\eta^{k'}}\right\|_2^2 + \left\|\frac{\overline{v}^*}{\eta^{k'}}\right\|_2^2 + \left\|\frac{\overline{\mu}^*}{\eta^{k'}}\right\|_2^2 + \left\|\frac{\overline{w}^*}{\eta^{k'}}\right\|_2^2\right]. \end{aligned}$$

Therefore,

$$t^* + \frac{\eta^{k'}}{2}\left[\left\|\left(g(x^*) + \frac{\overline{v}^*}{\eta^{k'}}\right)_+\right\|_2^2 + \left\|\left(-t^*\hat{\beta} + F(x^*) + \frac{\overline{\mu}^*}{\eta^{k'}}\right)_+\right\|_2^2 + \left\|\left(-t^* + \frac{\overline{w}^*}{\eta^{k'}}\right)_+\right\|_2^2\right]$$
$$\le t + \frac{\eta^{k'}}{2}\left[\left\|\frac{\overline{v}^*}{\eta^{k'}}\right\|_2^2 + \left\|\frac{\overline{\mu}^*}{\eta^{k'}}\right\|_2^2 + \left\|\frac{\overline{w}^*}{\eta^{k'}}\right\|_2^2\right].$$

Thus,

$$t^* + \frac{\eta^{k'}}{2}\left[\sum_{i=1}^{p}\left(g_i(x^*) + \frac{\overline{v}_i^*}{\eta^{k'}}\right)_+^2 + \sum_{m=1}^{r}\left(-t^*\beta_m + f_m(x^*) + \frac{\overline{\mu}_m^*}{\eta^{k'}}\right)_+^2 + \left(-t^* + \frac{\overline{w}^*}{\eta^{k'}}\right)_+^2\right]$$
$$\le t + \frac{\eta^{k'}}{2}\left[\sum_{i=1}^{p}\left(\frac{\overline{v}_i^*}{\eta^{k'}}\right)^2 + \sum_{m=1}^{r}\left(\frac{\overline{\mu}_m^*}{\eta^{k'}}\right)^2 + \left(\frac{\overline{w}^*}{\eta^{k'}}\right)^2\right]. \tag{6.23}$$

Now, if $g_i(x^*) = 0$, $-t^*\beta_m + f_m(x^*) = 0$, and $-t^* = 0$, since $\overline{v}_i^*/\eta^{k'} \ge 0$, $\overline{\mu}_m^*/\eta^{k'} \ge 0$, and $\overline{w}^*/\eta^{k'} \ge 0$, we obtain

$$\left(g_i(x^*) + \frac{\overline{v}_i^*}{\eta^{k'}}\right)_+ = \frac{\overline{v}_i^*}{\eta^{k'}},\ \left(-t^*\beta_m + f_m(x^*) + \frac{\overline{\mu}_m^*}{\eta^{k'}}\right)_+ = \frac{\overline{\mu}_m^*}{\eta^{k'}},\ \text{and}\ \left(-t^* + \frac{\overline{w}^*}{\eta^{k'}}\right)_+ = \frac{\overline{w}^*}{\eta^{k'}}.$$

Therefore, by (6.23),

$$t^* + \frac{\eta^{k'}}{2}\left[\sum_{g_i(x^*)<0}\left(g_i(x^*) + \frac{\overline{v}_i^*}{\eta^{k'}}\right)_+^2 + \sum_{-t^*\beta_m + f_m(x^*)<0}\left(-t^*\beta_m + f_m(x^*) + \frac{\overline{\mu}_m^*}{\eta^{k'}}\right)_+^2 + {}_{-t^*<0}\left(-t^* + \frac{\overline{w}^*}{\eta^{k'}}\right)_+^2\right]$$
$$\le t + \frac{\eta^{k'}}{2}\left[\sum_{g_i(x^*)<0}\left(\frac{\overline{v}_i^*}{\eta^{k'}}\right)^2 + \sum_{-t^*\beta_m + f_m(x^*)<0}\left(\frac{\overline{\mu}_m^*}{\eta^{k'}}\right)^2 + {}_{-t^*<0}\left(\frac{\overline{w}^*}{\eta^{k'}}\right)^2\right], \tag{6.24}$$

where ${}_{-t^*<0}\left(-t^* + \frac{\overline{w}^*}{\eta^{k'}}\right)_+ = \max_{-t^*<0}\left\{0, -t^* + \frac{\overline{w}^*}{\eta^{k'}}\right\}$.

However, we know that

$\lim_{k\to\infty}\max\{g_i(x^k),\ -\overline{v}_i^k/\eta^{k'}\} = 0$, $\lim_{k\to\infty}\max\{-t^k\beta_m + f_m(x^k), -\overline{\mu}_m^k/\eta^{k'}\} = 0$ and $\lim_{k\to\infty}\max\{-t^k, -\overline{w}^k/\eta^{k'}\} = 0$ (see Theorem 4.1 of [14]).

Therefore, if $g_i(x^*) < 0, -t^*\beta_m + f_m(x^*) < 0$ and $-t^* < 0$, then $v_i^* = 0, \mu_m^* = 0$ and $w^* = 0$. Thus, (6.24) implies that $t^* \le t$. Since t is an arbitrary feasible point, the result follows. $\square$

In the next theorem, we discuss the effect of variable scaling on the optimal solution of (6.12).

Let the independent variables x and t of the problem (6.12) are scaled by the following rule:

$$t \leftarrow bt \text{ and } x_s \leftarrow d_s x_s, \quad s = 1, 2, \ldots, n,$$

where b and d_s are positive scalars. If d is the vector $(d_1, d_2, \ldots, d_n)^\top$ and $D = \text{diag}(d)$, then $x \leftarrow Dx$. Therefore, the scaled form of the problem (6.12) can be written as

$$(\overline{x}^k, \overline{t}^k) = \underset{t \geq 0}{\text{argmin}}\ L_{\eta^k}(Dx, bt, \mu^k, v^k, \lambda^k, w^k), \tag{6.25}$$

where $(\overline{x}, \overline{t}) \in \mathbb{R}^n \times \mathbb{R}_+$.

In order to analyze the effect of the scaling of variables on Algorithm 8, we take the following assumptions:

A(2) For each given $\hat{\beta} \in \mathbb{S}_+^{r-1}$, there exists a unique optimal solution of the problems (6.12) and (6.25).

A(3) For each given $\hat{\beta} \in \mathbb{S}_+^{r-1}$, for each $k \in \mathbb{N}$, the points $(\overline{x}^k, \overline{t}^k)$ and (x^k, t^k) satisfy

$$L_{\eta^k}(D\overline{x}^k, b\overline{t}^k, \mu^k, v^k, \lambda^k, w^k) \leq L_{\eta^k}(x^k, t^k, \mu^k, v^k, \lambda^k, w^k) + \zeta_k,$$

where the sequence of tolerances $\{\zeta_k\}$ is bounded.

Theorem 6.3 *Let corresponding to each direction $\hat{\beta} \in \mathbb{S}_+^{r-1}$, $\{(\overline{x}^k, \overline{t}^k)\}$ be the sequence generated by Algorithm 8 for the problem* (6.25)*, under the assumptions A(0)–A(3),* $\lim\limits_{k\to\infty} \tau_k = 0$ *and* $\lim\limits_{k\to\infty} \zeta_k = 0$. *Moreover, we choose $\eta^{k+1} = \eta^k$ if* (6.11) *holds. Let $K \underset{\infty}{\subset} \mathbb{N}$ be such that $(\overline{x}^k, \overline{t}^k) \xrightarrow{k \in K} (\overline{x}^*, \overline{t}^*)$, and $(\overline{x}^*, \overline{t}^*)$ be the global minimizer of the problem* (6.25)*. Then $(D\overline{x}^*, b\overline{t}^*) = (x^*, t^*)$.*

Proof: We consider two cases on the sequences of penalty parameter:

(i)$\{\eta^k\}$ tends to positive infinity.

(ii)$\{\eta^k\}$ is bounded.

Case (i). For all $k \in \mathbb{N}$, we have

$$b\bar{t}^k \leq b\bar{t}^k + \frac{\eta^k}{2}\left[\left\|h(D\overline{x}^k) + \frac{\overline{\lambda}^k}{\eta^k}\right\|_2^2 + \left\|\left(g(D\overline{x}^k) + \frac{\overline{v}^k}{\eta^k}\right)_+\right\|_2^2 + \left\|\left(-b\bar{t}^k\hat{\beta} + F(D\overline{x}^k) + \frac{\overline{\mu}^k}{\eta^k}\right)_+\right\|_2^2 + \right.$$
$$\left.\left\|\left(-b\bar{t}^k + \frac{\overline{w}^k}{\eta^k}\right)_+\right\|_2^2\right]. \tag{6.26}$$

From the assumptions A(1) and A(3), we can write the inequality (6.26) as

$$b\bar{t}^k \leq t + \frac{\eta^k}{2}\left[\left\|h(x) + \frac{\overline{\lambda}^k}{\eta^k}\right\|_2^2 + \left\|\left(g(x) + \frac{\overline{v}^k}{\eta^k}\right)_+\right\|_2^2 + \left\|\left(-t\hat{\beta} + F(x) + \frac{\overline{\mu}^k}{\eta^k}\right)_+\right\|_2^2 + \right.$$
$$\left.\left\|\left(-t + \frac{\overline{w}^k}{\eta^k}\right)_+\right\|_2^2\right] + \tau_k + \zeta_k. \tag{6.27}$$

From (6.20), we can write the inequality (6.27) as

$$b\bar{t}^k \leq t + \frac{\left\|\overline{\lambda}^k\right\|_2^2}{2\eta^k} + \frac{\left\|\overline{v}^k\right\|_2^2}{2\eta^k} + \frac{\left\|\overline{\mu}^k\right\|_2^2}{2\eta^k} + \frac{\left\|\overline{w}^k\right\|_2^2}{2\eta^k} + \tau_k + \zeta_k. \tag{6.28}$$

Since $(\overline{x}^k, \bar{t}^k) \xrightarrow{k\in K} (\overline{x}^*, \bar{t}^*)$, $\left(\frac{\left\|\overline{\lambda}^k\right\|_2^2}{\eta^k}, \frac{\left\|\overline{v}^k\right\|_2^2}{\eta^k}, \frac{\left\|\overline{\mu}^k\right\|_2^2}{\eta^k}, \frac{\left\|\overline{w}^k\right\|_2^2}{\eta^k}\right) \xrightarrow{k\in K} (0,0,0,0)$, $\tau_k \xrightarrow{k\in K} 0$ and $\zeta_k \xrightarrow{k\in K} 0$, by taking limits on the inequality in (6.28) for $k \in K$, we get $b\bar{t}^* \leq t$. Since t is arbitrarily chosen, $(D\overline{x}^*, b\bar{t}^*)$ is a global minimizer of the problem (6.12).

From the Assumption A(2), we get $(D\overline{x}^*, b\bar{t}^*) = (x^*, t^*)$.

Case (ii). In this case, there exists $k' \in \mathbb{N}$ such that $\eta^k = \eta^{k'}$ for all $k \geq k'$. Therefore, by A(1), A(3) and using (6.20), we get the following for all $k \geq k'$:

$$b\bar{t}^k + \frac{\eta^{k'}}{2}\left[\left\|h(D\overline{x}^k) + \frac{\overline{\lambda}^k}{\eta^{k'}}\right\|_2^2 + \left\|\left(g(D\overline{x}^k) + \frac{\overline{v}^k}{\eta^{k'}}\right)_+\right\|_2^2 + \left\|\left(-b\bar{t}^k\hat{\beta} + f(D\overline{x}^k) + \frac{\overline{\mu}^k}{\eta^{k'}}\right)_+\right\|_2^2 + \right.$$
$$\left.\left\|\left(-b\bar{t}^k + \frac{\overline{w}^k}{\eta^{k'}}\right)_+\right\|_2^2\right] \leq t + \frac{\eta^{k'}}{2}\left[\left\|\frac{\overline{\lambda}^k}{\eta^{k'}}\right\|_2^2 + \left\|\frac{\overline{v}^k}{\eta^{k'}}\right\|_2^2 + \left\|\frac{\overline{\mu}^k}{\eta^{k'}}\right\|_2^2 + \left\|\frac{\overline{w}^k}{\eta^{k'}}\right\|_2^2\right] + \tau_k + \zeta_k \tag{6.29}$$

Let $K_0 \underset{\infty}{\subset} K, \lambda^* \in \mathbb{R}^q, v^* \in \mathbb{R}^p_+, \mu^* \in \mathbb{R}^r_+$ and $w^* \in \mathbb{R}_+$ such that $(\overline{\lambda}^k, \overline{v}^k, \overline{\mu}^k, \overline{w}^k) \xrightarrow{k \in K_0}$ $(\overline{\lambda}^*, \overline{v}^*, \overline{\mu}^*, \overline{w}^*)$.

By the feasibility of $(\overline{x}^*, \overline{t}^*)$, taking the limit on the inequality (6.29) for $k \in K_0$, we get

$$b\overline{t}^* + \frac{\eta^{k'}}{2}\left[\left\|\frac{\overline{\lambda}^*}{\eta^{k'}}\right\|_2^2 + \left\|\left(g(D\overline{x}^*) + \frac{\overline{v}^*}{\eta^{k'}}\right)_+\right\|_2^2 + \left\|\left(-b\overline{t}^*\hat{\beta} + F(D\overline{x}^*) + \frac{\overline{\mu}^*}{\eta^{k'}}\right)_+\right\|_2^2 + \left\|\left(-b\overline{t}^* + \frac{\overline{w}^*}{\eta^{k'}}\right)_+\right\|_2^2\right] \leq t + \frac{\eta^{k'}}{2}\left[\left\|\frac{\overline{\lambda}^*}{\eta^{k'}}\right\|_2^2 + \left\|\frac{\overline{v}^*}{\eta^{k'}}\right\|_2^2 + \left\|\frac{\overline{\mu}^*}{\eta^{k'}}\right\|_2^2 + \left\|\frac{\overline{w}^*}{\eta^{k'}}\right\|_2^2\right].$$

Therefore,

$$b\overline{t}^* + \frac{\eta^{k'}}{2}\left[\left\|\left(g(D\overline{x}^*) + \frac{\overline{v}^*}{\eta^{k'}}\right)_+\right\|_2^2 + \left\|\left(-b\overline{t}^*\hat{\beta} + F(D\overline{x}^*) + \frac{\overline{\mu}^*}{\eta^{k'}}\right)_+\right\|_2^2 + \left\|\left(-b\overline{t}^* + \frac{\overline{w}^*}{\eta^{k'}}\right)_+\right\|_2^2\right]$$
$$\leq t + \frac{\eta^{k'}}{2}\left[\left\|\frac{\overline{v}^*}{\eta^{k'}}\right\|_2^2 + \left\|\frac{\overline{\mu}^*}{\eta^{k'}}\right\|_2^2 + \left\|\frac{\overline{w}^*}{\eta^{k'}}\right\|_2^2\right].$$

Now, using the same procedure for the scaled problem (6.25), as used in equations (6.23) and (6.24), we get $b\overline{t}^* \leq t$.

Since t is arbitrarily chosen, $(D\overline{x}^*, b\overline{t}^*)$ is a global minimizer of the problem (6.12).

By A(2), we get $(D\overline{x}^*, b\overline{t}^*) = (x^*, t^*)$. Hence, the result follows. □

For a numerical evidence of Theorem 6.3 we have evaluated the optimal solutions of the scaled as well as unscaled comet problem [85] for 15 values of $\hat{\beta}$, which is listed below in Table 6.1, the parameter values used for the experiments are $d_s = 3$ for all $s = 1, 2, 3$ and $b = 2$.

6.6 An Approach to Solve the Unconstrained Subproblem (6.12) in Algorithm 8

Algorithm 8 requires that we minimize the problem in (6.12) at each iteration, i.e., for a given $\hat{\beta}$, η^k and $(\overline{\mu}^k, \overline{\lambda}^k, \overline{v}^k, \overline{w}^k)$ for each $k = 1, 2, \ldots,$ we need to solve the problem

Direction ($\hat{\beta}$)	Unscaled optimal solution (x_1^*, x_2^*, x_3^*, t^*)	Scaled optimal solution ($\overline{x_1}^*, \overline{x_2}^*, \overline{x_3}^*, \overline{t}^*$)	($D\overline{x}^*, b\overline{t}^*$)
(0.8766, 0.4699, 0.0952)	(1.5078, −0.5834, 0, 40.1136)	(0.5026, −0.1945, 0, 20.0568)	(1.5078, −0.5835, 0, 40.1136)
(0.9950, 0.0954, 0.0295)	(1.6745, −0.4259, 0, 183.4462)	(0.5582, −0.1420, 0, 91.7231)	(1.6746, −0.4260, 0, 183.4462)
(0.4536, 0.8209, 0.3471)	(2.1112, 0.2125, 0, 29.8855)	(0.7037, 0.0709, 0, 14.9428)	(2.1111, 0.2127, 0, 29.8856)
(0.4536, 0.8912, 0.0036)	(1, 1.9996, 0, 46.5024)	(0.3333, 0.6663, 0, 23.2512)	(0.9999, 1.9989, 0, 46.5024)
(0.8253, 0.5646, 0.0056)	(1.0352, −1.2402, 0, 38.0557)	(0.3448, −0.4135, 0, 19.0315)	(1.0344, −1.2405, 0, 38.0630)
(0.9960, 0.0326, 0.0838)	(2.7344, −0.0978, 0, 231.9457)	(0.9115, −0.0326, 0, 115.9728)	(2.7345, −0.0978, 0, 231.9456)
(0.5403, 0.2251, 0.8108)	(2.8835, −0.0834, 0, 27.0634)	(0.9612, −0.0278, 0, 13.5317)	(2.8836, −0.0834, 0, 27.0742)
(0.8473, 0.5286, 0.0525)	(1.2932, −0.9247, 0, 38.4262)	(0.4311, −0.3083, 0, 19.2131)	(1.2933, −0.9249, 0, 38.4262)
(0.5403, 0.8415, $8.4147e^{-04}$)	(1.0054, 1.4868, 0, 39.4941)	(0.3351, 0.4956, 0, 19.7471)	(1.0053, 1.4868, 0, 39.4942)
(1, 0.0090, $1.8000e^{-05}$)	(1, −1.9986, 0, $2.3437e^{+03}$)	(0.3356, −0.6532, 0, $1.1726e^{+03}$)	(1.0068, −1.9596, 0, $2.3452e^{+03}$)
(0.5403, 0.4546, 0.7081)	(2.4857, −0.1302, 0, 21.9416)	(0.8286, −0.0434, 0, 10.9708)	(2.4858, −0.1302, 0, 21.9416)
(0.9090, 0.3007, 0.2887)	(2.2307, −0.1802, 0, 41.3188)	(0.7436, −0.0601, 0, 20.6594)	(2.2308, −0.1803, 0, 41.3188)
(1, 0.0020, $1.0000e^{-05}$)	(1, −2, 0, $1.0547e^{+04}$)	(0.3390, −0.6338, 0, $5.2802e^{+03}$)	(1.0170, −1.9014, 0, $1.0560e^{+04}$)
(0.9801, 0.1982, 0.0133)	(1.2099, −1.1293, 0, 104.6089)	(0.4033, −0.3764, 0, 52.3045)	(1.2099, −1.1292, 0, 104.6090)
(0.8376, 0.3257, 0.4386)	(2.4098, −0.1429, 0, 32.8777)	(0.8033, −0.0477, 0, 16.4389)	(2.4099, −0.1431, 0, 32.8778)

Table 6.1: Optimal solution of scaled and unscaled comet problem in [85]

(6.12). For simplicity, we denote $\eta = \eta^k$, $\mu = \overline{\mu}^k$, $\lambda = \overline{\lambda}^k$, $v = \overline{v}^k$ and $w = \overline{w}^k$, and

$$H(x, t) = L_\eta(x, t, \mu, \lambda, v, w).$$

In this section, we denote $\{(x^k, t^k)\}$ as the sequence of iterates generated by Algorithm 9. The readers must not be confused with the iterates $\{(x^k, t^k)\}$ of Algorithm 8.

Algorithm 9 shows a step-wise procedure to solve the subproblem (6.12) at each iteration of Algorithm 8. In Algorithm 9 the direction d^k is the steepest descent direction. In the line search, to find the step size φ we use a nonmonotone Armijo-like condition $H((x^k, t^k) + \varphi d^k) \leq c_k + \rho\varphi\nabla H(x^k, t^k)^\top d^k$. The line search used in Algorithm 9 is commonly known as max-type nonmonotone line search propsed by Grippo [66]. The numerical results reported in [156] indicate that the max-monotone line search converge faster than the monotone one, thus adopting a max-monotone stretegy with any method may improve the performance of the method. The choice of c_k is such that it takes the maximum of the recent objective function values [111].

Algorithm 9 Algorithm to solve the subproblem (6.12)

Aim: To solve the subproblem (6.12) for a given $\hat{\beta} \in \mathbb{S}_+^{r-1}$

1: Choose a starting point $(x^0, t^0) \in \mathbb{R}^n \times \mathbb{R}_+$

2: Choose $\rho \in (0,1)$, $\varsigma \in (0,1)$ and a nonnegative integer M arbitrarily

3: Evaluate the function $H(x,t) = L_\eta(x,t,\mu,\lambda,v,w)$ using (6.8)

4: Initialize $k \leftarrow 0$

5: Provide $\varrho > 0$, the tolerance level for the optimum solution to the problem (6.12)

6: Set $c_0 \leftarrow H(x^0,t^0)$, $m_0 \leftarrow 0$

7: Compute $d^k \leftarrow -\nabla H(x^k,t^k) \in \mathbb{R}^{n+1}$

8: **if** $\|\nabla H(x^k,t^k)\| < \varrho$, **then** stop and **return** (x^k,t^k) as a minimizer of the subproblem (6.12)

9: **else**

10: Set $\varphi = 1$

11: **while** $H((x^k,t^k) + \varphi d^k) > c_k + \rho\varphi\nabla H(x^k,t^k)^\top d^k$ **do**

12: $\varphi \leftarrow \varsigma\varphi$

13: **end while**

14: Set $(x^{k+1},t^{k+1}) \leftarrow (x^k,t^k) + \varphi d^k$

15: $k \leftarrow k+1$

16: Choose an integer m_k such that $0 \le m_k \le \min\{m_{k-1}+1, M\}$, and set

$$c_k = \max_{0\le j\le m_k} H(x^{k-j},t^{k-j}) \tag{6.30}$$

17: **end if**

18: **go to** Step 7

Lemma 6.1 *For each iteration k of Algorithm 9, we have $0 \le H(x^k,t^k) \le c_k$.*

Proof: Note that

$$\begin{aligned}0 \le t \le t + \tfrac{\eta}{2}\left[\left\|h(x) + \tfrac{\lambda}{\eta}\right\|_2^2 + \left\|\left(g(x) + \tfrac{v}{\eta}\right)_+\right\|_2^2 + \left\|\left(-t\hat{\beta} + F(x) + \tfrac{\mu}{\eta^k}\right)_+\right\|_2^2 + \left\|\left(-t + \tfrac{w}{\eta}\right)_+\right\|_2^2\right] \\ = L_\eta(x,t,\hat{\beta},\mu,\lambda,v,w) = H(x,t).\end{aligned}$$

Thus, $H(x,t)$ is bounded from below. Hence, the result easily follows from the definition of c_k in (6.30). ∎

We assume that H has continuous first derivatives for all $(x^k,t^k) \in \mathbb{R}^{n+1}$ without mentioning second derivatives at all. This omission is convenient since the augmented

Lagrangian function has second derivative discontinuities when the original optimization problem has inequality constraints [13].

The following theorem shows that the line search used in Algorithm 9 is well-defined.

Theorem 6.4 *Let (x^k, t^k) be an iterate of Algorithm 9. If d^k is a descent direction, i.e., $\nabla H(x^k, t^k)^\top d^k < 0$, then for any $\rho \in (0, 1)$, there exists a $\varphi > 0$ satisfying*

$$H((x^k, t^k) + \varphi d^k) \leq c_k + \rho\varphi\nabla H(x^k, t^k)^\top d^k.$$

Proof: Assume that $\nabla H(x^k, t^k) \neq 0$. By the differentiablity of $H(x^k, t^k)$, we have

$$\lim_{\xi\to 0}\frac{H((x^k, t^k) + \xi d^k) - H(x^k, t^k)}{\xi} = \nabla H(x^k, t^k)^\top d^k < 0.$$

Thus,

$$\lim_{\xi\to 0}\frac{H((x^k, t^k) + \xi d^k) - H(x^k, t^k)}{\xi\nabla H(x^k, t^k)^\top d^k} = 1 > \rho.$$

So, there exists $\bar{\delta} > 0$ such that for some $\varphi \in [0, \bar{\delta}]$, we have

$$\frac{H((x^k, t^k) + \varphi d^k) - H(x^k, t^k)}{\varphi\nabla H(x^k, t^k)^\top d^k} \geq \rho$$

i.e., $H((x^k, t^k) + \varphi d^k) \leq H(x^k, t^k) + \rho\varphi\nabla H(x^k, t^k)^\top d^k$ since $\nabla H(x^k, d^k)^\top d < 0$

i.e., $H((x^k, t^k) + \varphi d^k) \leq c_k + \rho\varphi\nabla H(x^k, t^k)^\top d^k$ since $\nabla H(x^k, d^k)^\top d \leq c_k$ by Lemma 6.1.

This completes the proof. □

In order to establish the global convergence of nonmonotone line searches, the following assumption concerning the search directions d^k is considered [111].

A(4) There exists a positive constants Γ_1 and Γ_2 such that

$$\nabla H(x^k, t^k)^\top d^k \leq -\Gamma_1\|\nabla H(x^k, t^k)\|^2 \quad \text{and} \quad \|d^k\| \leq \Gamma_2\|\nabla H(x^k, t^k)\|.$$

The first inequality of the above assumption means that the search direction d^k is a sufficient descent direction, and the second one means that d^k is not too large in magnitude [111].

Theorem 6.5 *Assume that the assumption A(4) holds. If H is bounded from below, then the sequence* $\{(x^k, t^k)\}$ *generated by Algorithm 9 has the property that*

$$\lim_{k\to\infty} \nabla H(x^k, t^k) = 0.$$

Proof: Since the steepest descent direction $d^k = -\nabla H(x^k, t^k)$ clearly satisfies Assumption A(4) for $\Gamma_1 = \Gamma_2 = 1$, the result holds for Algorithm 9 (see [111], Theorem 3 on p.70). □

6.7 Discussion

- *Is Algorithms 8 always well-defined?*

  Algorithm 8 gets started by choosing a direction $\hat{\beta}$ in $\mathbb{S}_+^{r-1}$ as defined in (6.3). Corresponding to each direction $\hat{\beta}$, we solve the subproblem (6.12) to obtain (x^{k+1}, t^{k+1}) by Algorithm 9. Thus, the well-definedness of Algorithm 8 depends upon the well definedness of Algorithm 9. As Algorithm 9 is well-defined by Theorem 6.4, Algorithm 8 is always well-defined.

- *Why do we choose new Lagrange multipliers* $\overline{\lambda}^{k+1}$, $\overline{\mu}^{k+1}$, $\overline{v}^{k+1}$, *and* $\overline{w}^{k+1}$ *in place of* λ^{k+1}, μ^{k+1}, v^{k+1}, *and* w^{k+1}, *respectively, in every inner iteration of Algorithm 8?*

  The idea behind choosing the multipliers $\overline{\lambda}^{k+1}$, $\overline{\mu}^{k+1}$, $\overline{v}^{k+1}$, and $\overline{w}^{k+1}$ in place of λ^{k+1}, μ^{k+1}, v^{k+1}, and w^{k+1} is just to assure the boundedness of the sequences $\{\overline{\lambda}^k\}, \{\overline{\mu}^k\}, \{\overline{v}^k\}$ and $\{\overline{w}^{k+1}\}$. This would guarantee an important property—the shift $(\overline{\lambda}_j^k/\eta^k, \overline{\mu}_m^k/\eta^k, \overline{v}_i^k/\eta^k, \overline{w}^k/\eta^k) \xrightarrow{\eta^k\to\infty} (0,0,0,0)$, which is due to the fact that it may be possible that the Lagrange multipliers sequences may be unbounded for ALM without this strategy. Thus, the choice of the bounded sequences $\{\overline{\lambda}^k\}, \{\overline{\mu}^k\}, \{\overline{v}^k\}$, and $\{\overline{w}^{k+1}\}$ ensures all the ratios $\overline{\lambda}_j^k/\eta^k, \overline{\mu}_m^k/\eta^k, \overline{v}_i^k/\eta^k$, and $\overline{w}^k/\eta^k$ goes to 0 as $\eta^k \to \infty$, for details see [14, 84]. In practice, the multipliers $(\overline{\lambda}^{k+1}, \overline{\mu}^{k+1}, \overline{v}^{k+1}, \overline{w}^{k+1})$ can be chosen as the projection of multipliers $(\lambda^k, \mu^k, v^k, w^k)$ onto the safeguarded intervals (see Remark 2).

- *How does the inequality*

$$\max\left\{\left\|h(x^k)\right\|, \left\|\mathcal{F}^k\right\|, \left\|\mathcal{G}^k\right\|, \left\|\mathcal{T}^k\right\|\right\} \leq \gamma \max\left\{\left\|h(x^{k-1})\right\|, \left\|\mathcal{F}^{k-1}\right\|, \left\|\mathcal{G}^{k-1}\right\|, \left\|\mathcal{T}^{k-1}\right\|\right\}$$

ensure progress in terms of feasibility and complementarity measures?

For a fixed value of $\gamma \in (0, 1)$, the above inequality discusses the progress of two different quantities. First, the improvement in the feasibility of equality constraints which is measured by $\left\|h(x^k)\right\|$. Secondly, one needs to reduce the quantities $\left\|\mathcal{F}^k\right\|, \left\|\mathcal{G}^k\right\|$, and $\left\|\mathcal{T}^k\right\|$ that are related to inequality constraints. To get a clear picture, for each $m = 1, 2, \ldots, r$, let us consider $F_m^k = \min\left\{t^k\beta_m - f_m(x^k), \frac{\mu_m^k}{\eta^k}\right\}$. Since the shift $\frac{\mu_m^k}{\eta^k}$ is nonnegative, $\min\left\{t^k\beta_m - f_m(x^k), \frac{\mu_m^k}{\eta^k}\right\}$ gives the improvement in terms of attainment of inequality constraint $-t^k\beta_m + f_m(x^k) \leq 0$. In fact, if $-t^k\beta_m + f_m(x^k)$ goes to zero, then there is a high chance of the progress of $\min\left\{t^k\beta_m - f_m(x^k), \frac{\mu_m^k}{\eta^k}\right\}$ which is independent of the shifts $\frac{\mu_m^k}{\eta^k}$. Similar observation is applied to G_i^k and T^k.

- *What is the advantage of the proposed method over the existing ones?*

Many of the existing methods either cannot generate the whole Pareto set or require some prior information of the Pareto set, such as convexity (see [32, 40, 87, 104, 108, 154, 155]), but the proposed method does not require any piece of such information, and it efficiently solves convex and nonconvex MOPs.

6.8 Numerical Results

In this section, we have shown the performance of the proposed method through several well known test problems which are required to see the efficacy of the Pareto set generating method. These collection of test problems include convex as well as nonconvex test problems. All tests were carried out within MATLAB software (version R2018b).

In our work, we use the condition $\left|t^k - t^{k+1}\right| \leq \varepsilon$ as a stopping criterion to solve (6.12) corresponding to every direction $\hat{\beta}$. The choice of this stopping criterion is motivated by the cone method, which states that optimal solution x^* corresponding to each $\hat{\beta}$ can only be obtained when t is the minimum of the problem (6.2) (see [60]).

6.8.1 Numerical Parameters for the Experiments

For the computations, we use the following parameter values:

(i) $E_{\text{sup}} \in \{10, 10^2, 10^3, 10^4, 10^5\}$, $m_1 \in \{40, 45, 120, 200, 250, 300\}$, and $\alpha \in \{1.1, 1.3, 1.4, 1.6, 1.8, 2,\ 2.1, 2.4, 2.5, 2.9\}$.

(ii) For a chosen $\alpha > 0$, the penalty parameter is updated by $\eta^{k+1} = \alpha\eta^k$, $k = 1, 2, 3, \ldots$.

(iii) The tolerance level $\varepsilon \in \{10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}\}$.

Table 6.2 shows the chosen combination of parameters for each of the specific problems.

Problem name and reference	No. of objectives and decision variables	Problem type	m_1	α	E_{sup}	ε (Tolerance)	Pareto Font Type		Figure No.
							Convex	Connected	
SCH [164]	2,1	Convex	200	2.5	10	10^{-4}	Y	Y	6.2b
FONa [164]	2,3	Non-convex	200	2	10	10^{-3}	N	Y	6.2a
FONb [164]	2,50	Non-convex	200	2	10^2	10^{-3}	N	Y	6.2a
FONc [164]	2,100	Non-convex	200	2.4	10^2	10^{-3}	N	Y	6.2a
JOSa [75]	2,50	Convex	300	1.4	10	10^{-4}	Y	Y	6.2c
JOSb [75]	2,100	Convex	300	2.5	10	10^{-4}	Y	Y	6.2c
JOSc [75]	2,500	Convex	300	1.4	10	10^{-4}	Y	Y	6.2c
FDSa [50]	3,50	Convex	120	2.5	10^2	10^{-3}	Y	Y	6.1d
FDSb [50]	3,100	Convex	120	2.5	10^2	10^{-3}	Y	Y	6.1d
COMET [85]	3,3	Non-convex	40	2.1	10^5	10^{-4}	N	Y	6.1a
DLTZ1 [164]	3,3	Non-convex	45	2.9	10	10^{-3}	Y	Y	6.1b
DLTZ2 [164]	3,3	Non-convex	45	1.6	10^5	10^{-5}	N	Y	6.2e
DLTZ5 [61]	3,3	Non-convex	200	1.1	10^3	10^{-7}	N	Y	6.1c
TNK [60]	2,2	Non-convex	200	1.3	10	10^{-4}	N	N	6.2i
ZDT2 [60]	2,2	Non-convex	200	1.8	10^4	10^{-6}	N	Y	6.2f
PNR [60]	2,3	Convex	300	2.4	10	10^{-5}	Y	Y	6.2d
DGO1 [75]	2,1	Non-convex	200	2.5	10	10^{-3}	Y	Y	6.2h
FF1 [75]	2,2	Non-convex	250	2.4	10	10^{-3}	N	Y	6.2j
IM1 [75]	2,2	Non-convex	200	1.6	10	10^{-3}	N	Y	6.2g

Table 6.2: Result obtained by Algorithm 8 on test problems

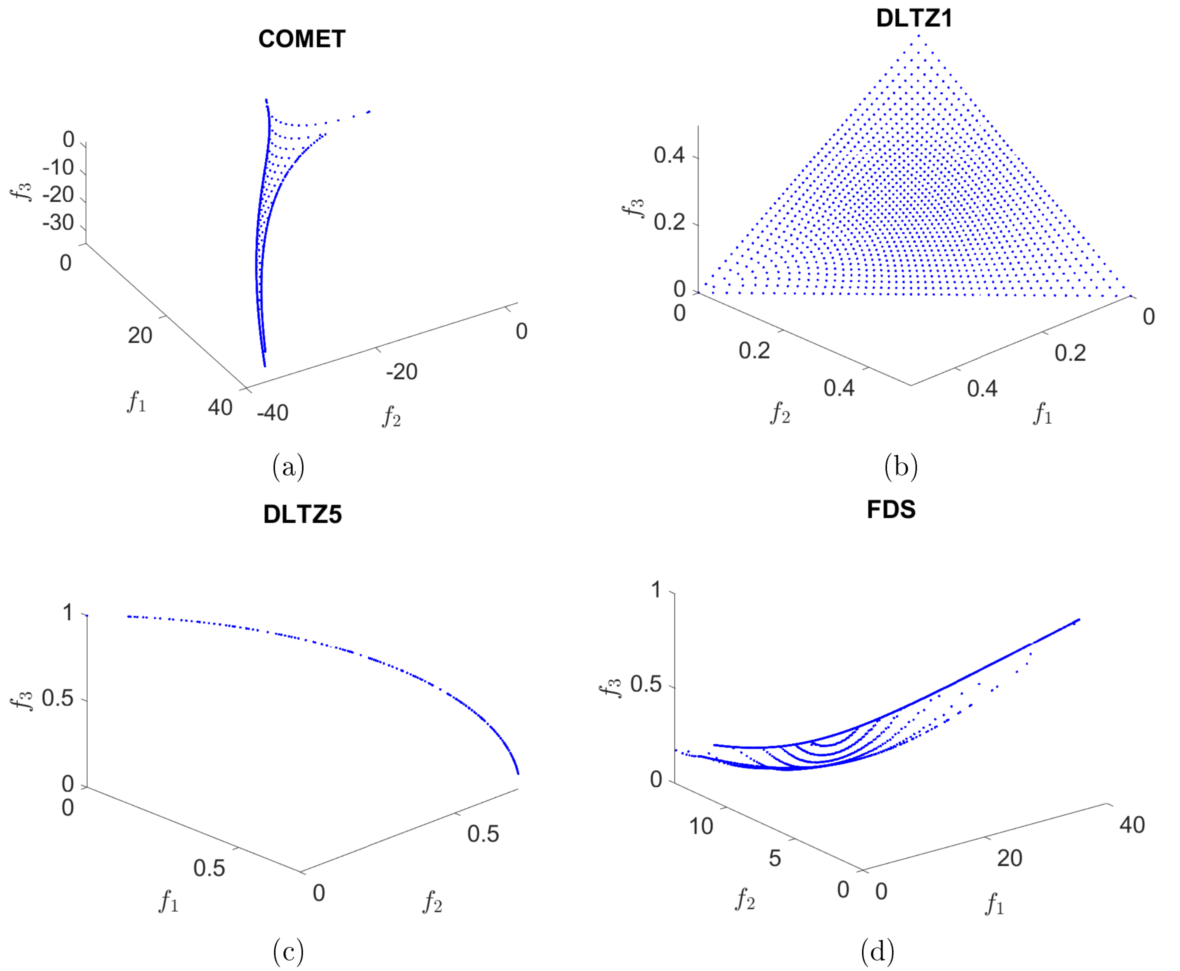


Figure 6.1: Performance of the proposed method (Algorithm 8) on comet, DLTZ1, DLTZ5, FDS problems

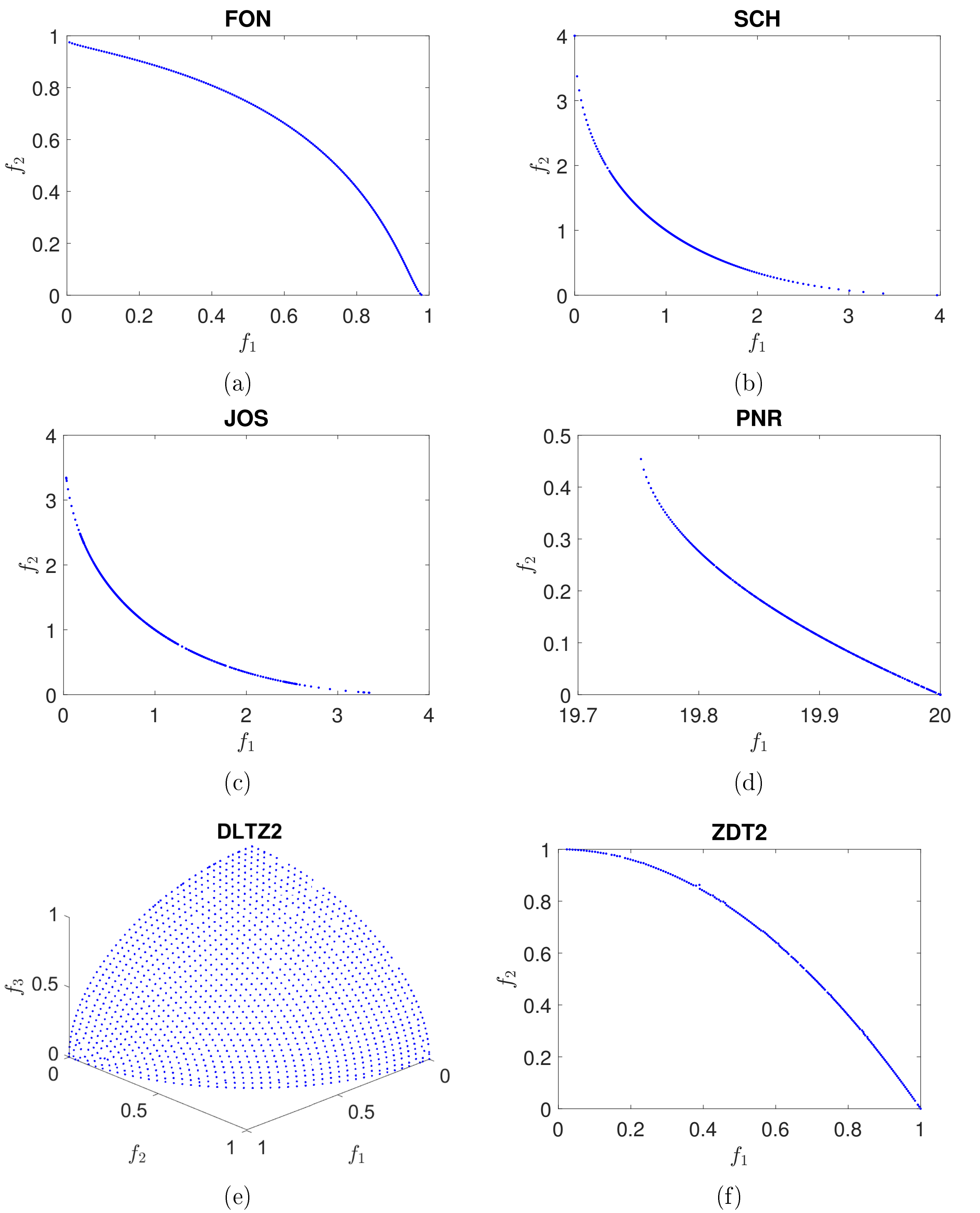


Figure 6.2: Performance of the proposed method (Algorithm 8) on FON, SCH, JOS, PNR, DLTZ2, ZDT2, IM1, DGO1, TNK, FF1 problems

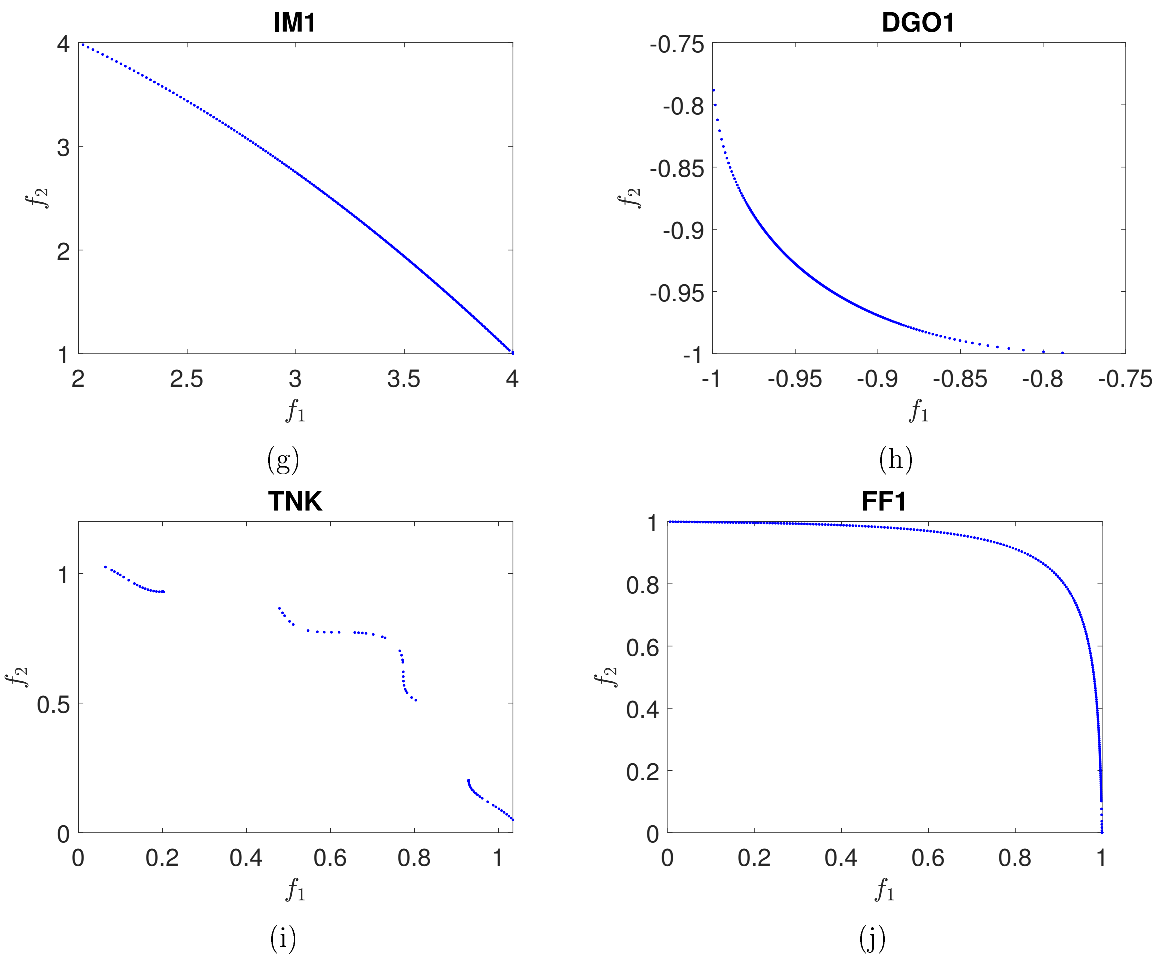


Figure 6.2: continued from Figure 6.2

6.9 An application

Worldwide, unemployment is increasing due to the rapid growth of the population. This current section proposes an optimal control policy for a deterministic unemployment model [115]. An outbreak of unemployment in a society can be prevented if early preparations are made to curb it. A common strategy is to create jobs for unemployed people and creation of new vacancies created by the government. Hence, we consider the rate at which employment is created for the unemployed and the rate at which new vacancies are created, as their control variable.

6.9.1 Model Formulation

For the application, we consider a three dimensional deterministic unemployment model with unemployment, employment, and newly created vacancies as their state variable [115].

Based on the given model system (6.31), the population is divided into the following three categories:

- U indicates the number of unemployed individuals;
- E indicates the number of employed individuals;
- V indicates the number of vacancies.

The control variables of the model system are as follows:

- $u_1(t)$ represents the rate at which the government is implementing its policies to create employment for unemployed people.
- $u_2(t)$ represents the the rate at which the government is implementing its policies to create new vacancies.

According to Munoli and Gani [115], the deterministic unemployment model with $u_1(t)$ and $u_2(t)$ as their control variables is modeled by the following system of nonlinear time-varying state equations:

$$\left.\begin{aligned} \frac{dU(t)}{dt} &= \Lambda - (1-u_1(t))kU(t)V(t) - \alpha_1 U(t) + \gamma E(t), \\ \frac{dE(t)}{dt} &= (1-u_1(t))kU(t)V(t) - \alpha_2 E(t) - \gamma E(t), \\ \frac{dV(t)}{dt} &= \alpha_2 E(t) + \gamma E(t) - \delta V(t) + \phi u_2(t)U(t), \end{aligned}\right\} \tag{6.31}$$

with the initial conditions

$$U(0) = 10000,\ \ E(0) = 1000 \text{ and } V(0) = 100, \tag{6.32}$$

where parameters in the system (6.31) are provided in the following table:

Symbols	Description	Value
Λ	The rate at which the number of unemployed persons is increasing continuously	5×10^{3}
k	Denotes the rate at which the unemployed person are being employed	9×10^{-6}
α_1	Represents the rate of migration as well as death of unemployed persons	4×10^{-2}
α_2	Represents the rate of retirement as well as death of employed persons	5×10^{-2}
γ	Represents the rate of persons who are fired from their jobs	1×10^{-3}
ϕ	Represents the rate of creating new vacancies	7×10^{-3}
δ	Denotes the diminution rate of vacancies due to lack of funds	5×10^{-2}
t_f	Total simulation duration	5 yr

Table 6.3: Description and values of the parameters in the model problem (6.31)

We have employed the values of rates $\Lambda, k, \alpha_1, \alpha_2, \gamma, \phi,$ and δ from [115].

6.9.2 Optimal Control Problem Formulation

Our aim is to determine the optimal values u_1^* and u_2^* of the controls u_1 and u_2 such that (i) the corresponding state trajectories $U^*, E^*,$ and V^* constitute a solution to the system (6.31) and (6.32) in the time interval $[0, t_f]$ and (ii) (u_1^*, u_2^*) minimizes the following objective functional:

$$J(u_1, u_2) = \int_0^{t_f} \left[U(t) + P_1 u_1^2(t) + P_2 u_2^2(t)\right] \, dt, \tag{6.33}$$

where t_f is the final time the constants P_1 and P_2 give relative cost of the intervention associated with the controls u_1 and u_2, respectively. Thus, the terms $P_1 u_1^2$ and $P_1 u_2^2$ denote the costs associated with the implemented policies of government to provide employment to unemployed persons and to create new vacancies, respectively [115]. In our study, we have considred $P_1 = P_2 = 1$.

Consider the model system (6.31). Suppose the set of permissible control functions is given by the following Lebesgue measurable set

$$\eta = \left\{(u_1, u_2) : 0 \leq u_1(t), u_2(t) \leq 1, \ \forall t \in [0, t_f]\right\}.$$

Hence, the proposed problem is to minimize the objective functional (6.33), The following problem

$$J(u_1^*, u_2^*) = \min_{(u_1,u_2)\in\eta} J(u_1, u_2) \tag{6.34}$$

has two main aspects: (i) it aims to minimizing the number of unemployed individuals and (ii) minimizing the cost to implement government policies to provide employment and create vacancies.

6.9.3 Multiobjective Approach to Optimal Control Problem

The available approaches that use the idea of optimal control theory have a significant disadvantage that it can only yield a single optimal solution for the problem (6.34). The determination of a single optimal solution requires decision-maker's preferences involving a pair of weight constants P_1 and P_2, for detail see [36], and the references therein. However, the choice of the parameters P_1 and P_2 is not straightforward. It requires some prior knowledge about the problem and decision maker's preferences, which may not always be available. Nevertheless a single optimal solution for problem (6.34) does not provide all helpful insights for the optimal strategies and for the corresponding dynamics. Thus, many optimal alternatives remains unexplored by using the approaches based on optimal control theory.

The current work proposes an approach that uses the idea of multiobjective optimization, which may provide the Pareto solutions instead of a single optimal solution. In the proposed approach, we decompose the cost functional (6.33) into two components

f_1 and f_2, where

$$f_1(u_1, u_2) = \int_0^{t_f} U(t)\, dt \text{ and} \tag{6.35}$$

$$f_2(u_1, u_2) = \int_0^{t_f} \left(u_1^2(t) + u_2^2(t)\right)\, dt. \tag{6.36}$$

Thus a biobjective formulation corresponding to (6.34) can be defined by the following:

$$\left.\begin{array}{ll} \text{minimize} & \left(f_1, f_2\right) \\ \text{subject to} & \dfrac{dU(t)}{dt} = \Lambda - (1 - u_1(t))kU(t)V(t) - \alpha_1 U(t) + \gamma E(t), \\ & \dfrac{dE(t)}{dt} = (1 - u_1(t))kU(t)V(t) - \alpha_2 E(t) - \gamma E(t), \\ & \dfrac{dV(t)}{dt} = \alpha_2 E(t) + \gamma E(t) - \delta V(t) + \phi u_2(t)U(t), \\ & U(0) = 10000,\ E(0) = 1000,\ V(0) = 100, \\ & 0 \leq u_1 \leq 1, 0 \leq u_2 \leq 1. \end{array}\right\} \tag{6.37}$$

Note that objective f_1 represents the total number of individual who become unemployed during the period $[0, t_f]$, and f_2 represents the cost associated to the implementation of control policy during the period $[0, t_f]$.

6.9.3.1 Algorithm

The following Algorithm 10 describes a complete procedure to obtain the Pareto set corresponding to the biobjective optimal control problem (6.37). We first numerically integrate the model system (6.31) by using the fourth-order Runge-Kutta method on the interval $[0, t_f]$, where $t_f = 5$ years. We discretize the range of the controls u_1 and u_2 on 80 equally spaced time intervals. Then, the integrals in (6.35) and (6.36) are calculated by applying the trapezoidal rule. To solve the biobjective problem (6.37) we use Algorithm 8.

Algorithm 10 Algorithm to obtain the Pareto set of the optimal control problem (6.37)

1: Numerically solve the system (6.31) to get a discrete approximation of U by using the fourth-order Runge-Kutta method on the interval of time $[0, t_f]$, where $t_f = 5$ years.
2: Discretize the controls u_1 and u_2 based on 80 equally spaced time intervals.
3: Compute the integrals in (6.35) and (6.36) by applying the trapezoidal rule in order to obtain discrete approximations of f_1 and f_2.
4: Compute the quadratic approximation of f_1 and f_2 with the help of curve fitting.
5: Apply Algorithm 8 to the biobjective optimization problem (f_1, f_2) subject to $0 \leq u_1 \leq 1$, $0 \leq u_2 \leq 1$ to find a discrete approximation of the complete Pareto set.

6.9.3.2 Experimental results

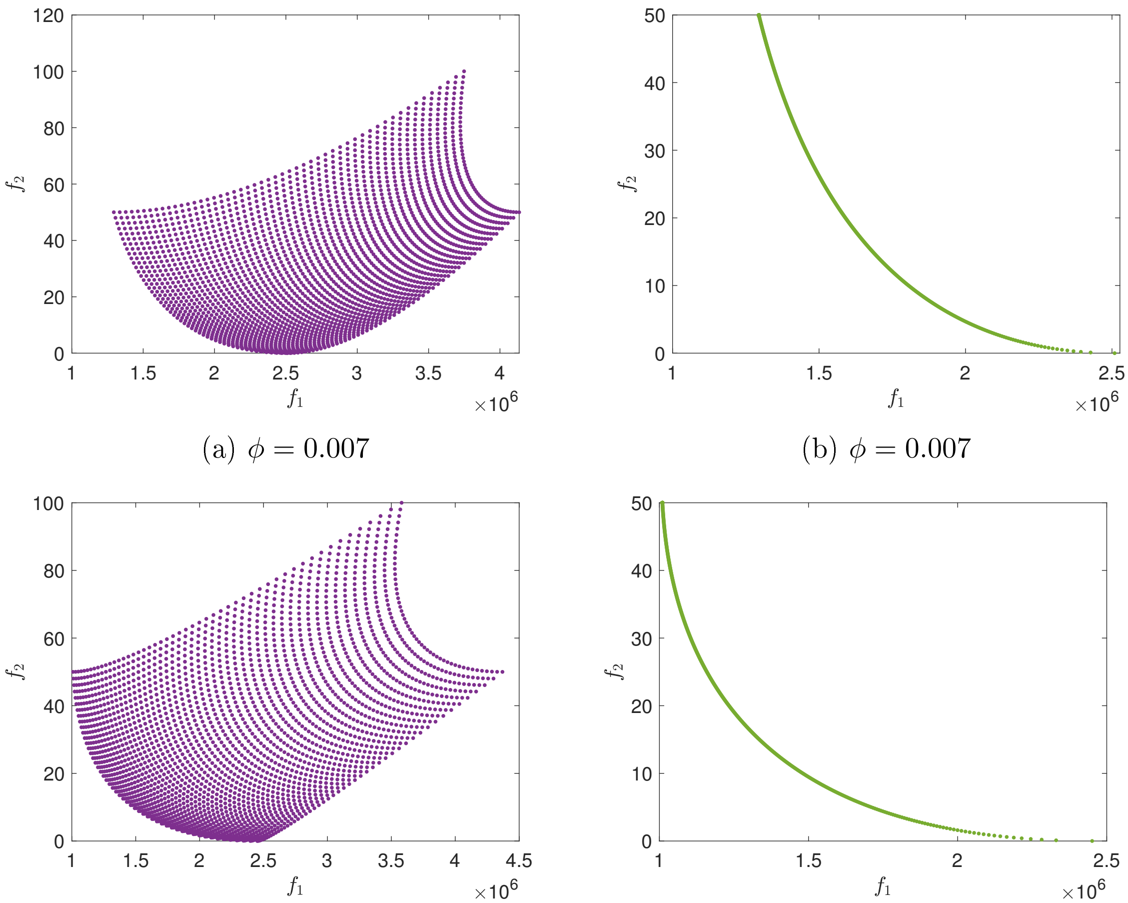
In the following part, we describe the obtained optimal solutions to the problem (6.37), taking into account the variations of the parameter ϕ.

Figures 6.3a, 6.3b and 6.3c, 6.3d plot the objective feasible region and Pareto front of problem (6.37) for $\phi = 0.007$ and 0.02, respectively.

Figure 6.4 plots the trade-off curve obtained for two different values of transmission coefficient ϕ. From the figures, we observe that the minimum value of of f_1 decrease when ϕ increases.

6.10 Conclusions and Future Work

In this chapter, an augmented Lagrangian cone method to solve MOPs with continuous objectives and constraints has been proposed. In the proposed approach, firstly, we have converted the given MOP into a set of parametric singleobjective optimization problems by using the cone method. Then, we have subsequently applied ALM to convert the singleobjective optimization problem into a sequence of parametric un-



(a) $\phi = 0.007$ (b) $\phi = 0.007$

(c) $\phi = 0.02$ (d) $\phi = 0.02$

Figure 6.3: (a)–(d) objective feasible region (left) and Pareto fronts (right column) of the problem (6.37) for $\phi = 0.007$ and 0.02

constrained optimization problems. The augmented Lagrangian subproblem is solved by steepest descent method under an max-type nonmonotone line search method. We have observed that corresponding to each parameter $\hat{\beta} \in \mathbb{S}_+^{r-1}$, by solving the sequence of unconstrained optimization problem obtained from ALM, one can capture a Pareto optimal point on the Pareto surface. Under common assumptions (see Theorems 6.1 and 6.2), the numerical Algorithm 8 globally converges to each Pareto optimal point on the Pareto surface. We have also shown that the solutions obtained by the proposed method is not affected by variable scaling. Algorithm 8 has been tested on various MOPs with diverse properties which can be found in [60, 61, 85, 164].

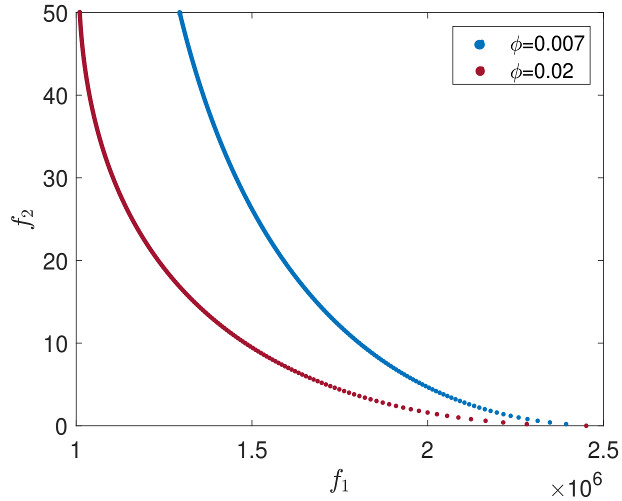


Figure 6.4: Pareto fronts of the problem (6.37) for two different values of ϕ

To show the diversity of the proposed method, we have shown a realistic application. We have applied the proposed method to a deterministic unemployment optimal control model with implementation of government policies to create employment for unemployed individuals and to create vacancies as their controls. We have proposed a multiobjective optimization approach to find the optimal strategies to minimize both of the number of unemployed individuals and of the implementation cost of the control strategies. Numerical results are analyzed for two different values of rate of creating new vacancies, i.e., $\phi = 0.007, 0.02$. We have obtained the trade-off curves for two values of ϕ which shows that the minimum value of f_1 decreases when ϕ increases which also indicates that as the rate of creating vacancies increases, minimum unemployed individuals in the population decreases.

As a future research, we are interested in extending our result to a vector optimization in which the objective functions are defined from Hilbert space X to a Banach space Y with a partial order induced by a closed, convex, and pointed cone $C \subset Y$ with a non-empty interior. More theoretical properties and experimental performance of augmented Lagrangian cone method can also be studied in future.

****