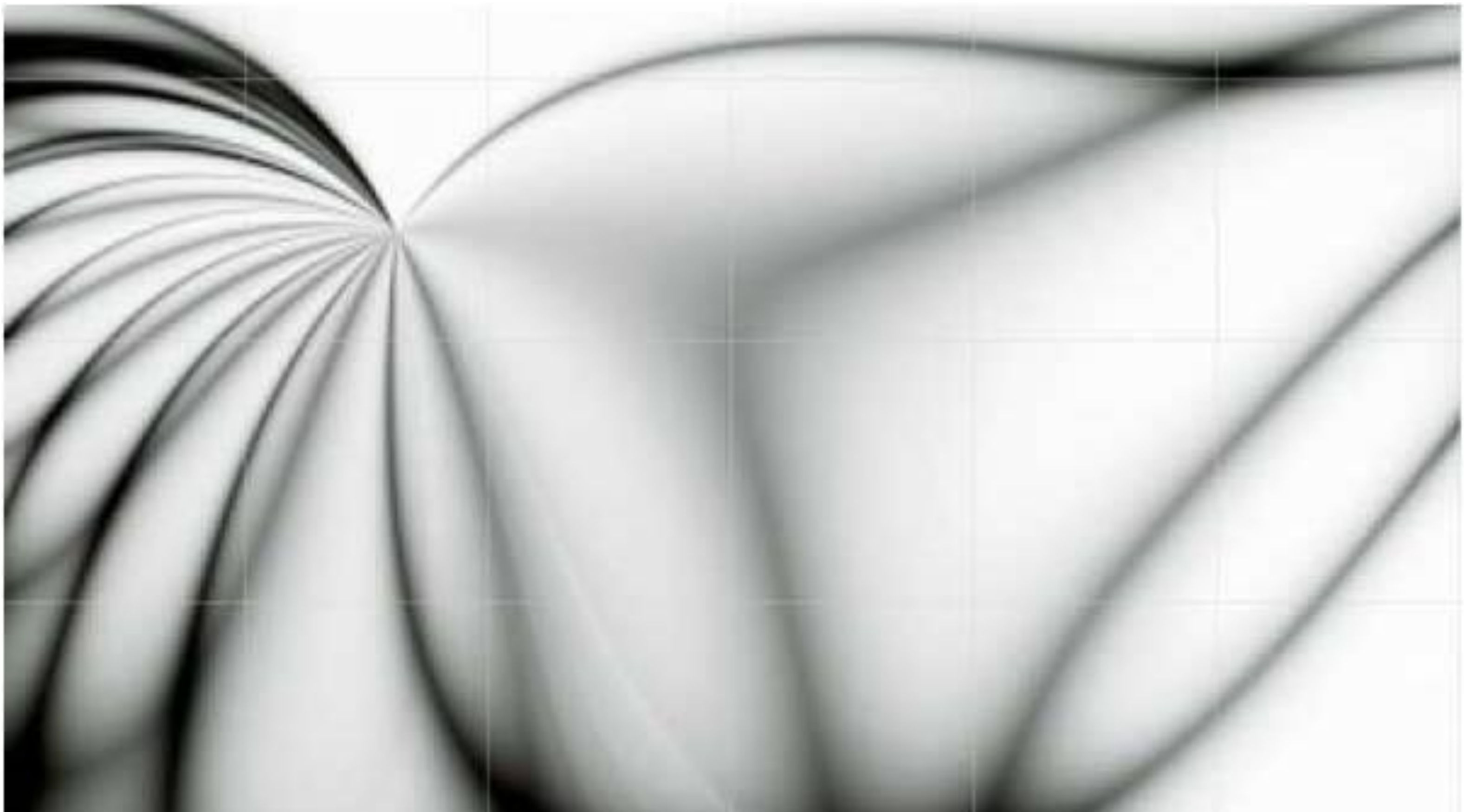
ISSN:2146-0957
eISSN:2146-5703

Volume:6 Number:1
January 2016

An International Journal of Optimization and Control: Theories & Applications



www.ijocta.com

ISSN: 2146-0957
eISSN: 2146-5703
www.ijocta.com
Email: info@ijocta.com

An International Journal of Optimization and Control: Theories & Applications

Volume: 6 Number: 1
January 2016

*"IJOCTA is an international, bi-annual, and peer-reviewed journal indexed/abstracted by **Cabell's Directories, DOAJ, EBSCO Databases, Genamics JournalSeek, Google Scholar, Index Copernicus, International Abstracts in Operations Research, JournalTOCs, Mathematical Reviews (MathSciNet), ProQuest, Ulrich's Periodical Directory,** and **Zentralblatt Math.**"*

An International Journal of Optimization and Control:
Theories & Applications
Volume: 6 Number: 1
January 2016

CONTENTS

An International Journal of Optimization
and Control: Theories & Applications
Vol.6, No.1, pp.1-7 (2016) © IJOCTA
ISSN: 2146-0957eISSN: 2146-5703
DOI: 10.11121/ijocta.01.2016.00247
http://www.ijocta.com

Generalized (Φ,ρ)-convexity in nonsmooth vector optimization over cones

S. K. Suneja[a] , Sunila Sharma[b] and Malti Kapoor[c]

[a,b] Department of Mathematics, Miranda House , University of Delhi, Delhi, India.
Email: surjeetsuneja@gmail.com, sunilaomhari@yahoo.co.in

[c] Department of Mathematics, Motilal Nehru College, University of Delhi, Delhi, India.
Email: maltikapoor1@gmail.com

(*Received February 6, 2015; in final form January 17, 2016*)

Abstract. In this paper, new classes of cone-generalized (Φ,ρ)-convex functions are introduced for a nonsmooth vector optimization problem over cones, which subsume several known studied classes. Using these generalized functions, various sufficient Karush-Kuhn-Tucker (KKT) type nonsmooth optimality conditions are established wherein Clarke's generalized gradient is used. Further, we prove duality results for both Wolfe and Mond-Weir type duals under various types of cone-generalized (Φ,ρ)-convexity assumptions.

Keywords: Nonsmooth vector optimization over cones; cone-generalized (Φ,ρ)-convexity; nonsmooth optimality conditions; duality.
AMS Classification: 90C26, 90C29, 90C46.

1. Introduction

Convexity plays an important role in many aspects of optimization theory including sufficient optimality conditions and duality theorems. In a quest to weaken the convexity hypothesis various generalized convexity notions have been introduced. Hanson and Mond [8] introduced *F*-convexity and Vial [10] defined ρ-convexity. Preda [9] unified the two concepts and gave the notion of an (*F*,ρ)-convex function.

Another generalization of convexity is invexity, introduced by Hanson [7]. The concept of (Φ,ρ)-invexity has been introduced by Caristi et al. [3]. Sufficient optimality conditions and duality results have been studied under (Φ,ρ)-invexity for differentiable single-objective and multiobjective programs [3,6]. (Φ,ρ)-invexity notion has been extended to the nonsmooth case by Antczak and Stasiak [2].

In this paper, we use the concept of cones to define new classes of nonsmooth functions that we call *K*-generalized (Φ,ρ)-convex, *K*-generalized (Φ,ρ)-pseudoconvex and *K*-generalized (Φ,ρ)-quasiconvex functions, where *K* is a closed convex pointed cone with nonempty interior. Sufficient optimality conditions are proved for a nonsmooth vector optimization problem over cones using the above defined functions. Further, both Wolfe and Mond-Weir type duals are formulated and weak and strong duality results are established.

2. Definitions and preliminaries

Let S be a nonempty open subset of $\mathbf{R}^n$.

Definition 2.1. A function $\theta : S \to \mathbf{R}$ is said to be locally Lipschitz at a point $u \in S$ if for some $l_u > 0$,

Corresponding Author. Email: maltikapoor1@gmail.com

$$|\theta(x)-\theta(\bar{x})|\le l_u\,\|x-\bar{x}\|$$

for all $x, \bar{x}$ in a neighborhood of u. We say that $\theta: S \to \mathbf{R}$ is locally Lipschitz on S if it is locally Lipschitz at each point of S.

Let $f=(f_1, f_2,...., f_m)^t : S \to \mathbf{R}^m$ be a vector-valued function. Then f is said to be locally Lipschitz on S if each f_i is locally Lipschitz on S.

Definition 2.2. [4] Let $\theta: S \to \mathbf{R}$ be a locally Lipschitz function on S. The Clarke's generalized directional derivative of θ at $u\in S$ in the direction v, denoted as $\theta^0(u;v)$, is defined by

$$\theta^0(u;v)=\limsup_{\substack{y\to u\\ t\to 0^+}}\frac{\theta(y+tv)-\theta(y)}{t}$$

Definition 2.3. [4] The Clarke's generalized gradient of θ at $u\in S$, denoted as $\partial\theta(u)$, is given by

$$\partial\theta(u)=\{\xi\in\mathbf{R}^n : \theta^0(u;v)\ge\langle\xi,v\rangle,\ \forall\, v\in\mathbf{R}^n\}.$$

The generalized directional derivative of a locally Lipschitz function $f=(f_1,..., f_m)^t : S\to\mathbf{R}^m$ at $u\in S$ in the direction v is given by

$$f^0(u;v)=(f_1^0(u;v),\ f_2^0(u;v),....,f_m^0(u;v))^t.$$

The generalized gradient of f at u is the set $\partial f(u)=\partial f_1(u)\times\partial f_2(u)\times...\times\partial f_m(u)$, where $\partial f_i(u)$ is the generalized gradient of f_i at u for i = 1, 2,...,m. An element $A=(A_1,...,A_m)^t\in\partial f(u)$ is a continuous linear operator from $\mathbf{R}^n$ to $\mathbf{R}^m$ and

$Au=(A_1^t u,...,A_m^t u)^t\in\mathbf{R}^m$ for all $u\in\mathbf{R}^n$.

Let $K\subseteq\mathbf{R}^m$ be a closed convex pointed cone with nonempty interior and let int$K$ denote the interior of K. The positive dual cone K^* and the strict positive dual cone K^{s^*} of K, are respectively defined as

$$K^*=\{y^*\in\mathbf{R}^m : \langle y,y^*\rangle\ge 0 \text{ for all } y\in K\}, \text{and}$$

$$K^{s*}=\{y^*\in\mathbf{R}^m : \langle y,y^*\rangle> 0 \text{ for all } y\in K\setminus\{0\}\}.$$

Throughout the paper, we shall denote an element of $\mathbf{R}^{n+1}$ by the ordered pair (a, r), where $a\in\mathbf{R}^n$ and $r\in\mathbf{R}$. Consider a function $\varphi : S\times S\times\mathbf{R}^{n+1}\to\mathbf{R}$ such that $\varphi(x, u; \cdot)$ is convex on $\mathbf{R}^{n+1}$ and $\varphi(x, u; (0, r))\ge 0$ for every x, $u\in S$ and any real number $r\in\mathbf{R}_+$. Let $f : S\to\mathbf{R}^m$ be a locally Lipschitz function, $u\in S$, $A=(A_1,...,A_m)^t\in\partial f(u)$, $\rho=(\rho_1,...,\rho_m)^t\in\mathbf{R}^m$ and $\Phi(x,u;(A,\rho))$ denote the vector $(\varphi(x,u;(A_1,\rho_1)),...,\varphi(x,u;(A_m,\rho_m)))^t$.

We introduce the following definitions:

Definition 2.4. The function f is said to be K-generalized (Φ,ρ)-convex at u on S if for every $x\in S$

$$f(x)-f(u)-\Phi(x,u;(A,\rho))\in K,\ \ \forall A\in\partial f(u).$$

Definition 2.5. The function f is said to be K-generalized (Φ,ρ)-pseudoconvex at u on S if for every $x\in S$, $A\in\partial f(u)$

$$-\Phi(x,u;(A,\rho))\notin\operatorname{int}K \Rightarrow -(f(x)-f(u))\notin\operatorname{int}K.$$

Equivalently, if for every $x\in S$

$$f(x)-f(u)\in-\operatorname{int}K\Rightarrow\Phi(x,u;(A,\rho))\in-\operatorname{int}K,\quad \forall A\in\partial f(u).$$

Definition 2.6. The function f is said to be K-generalized (Φ,ρ)-quasiconvex at u on S if for every $x\in S$

$$f(x)-f(u)\notin\operatorname{int}K\Rightarrow-\Phi(x,u;(A,\rho))\in K,\quad \forall A\in\partial f(u).$$

If f is K-generalized (Φ,ρ)-convex (K-generalized (Φ,ρ)-pseudoconvex, K-generalized (Φ,ρ)-quasiconvex) at every $u\in S$ then f is said to be K-generalized (Φ,ρ)-convex (K-generalized (Φ,ρ)-pseudoconvex, K-generalized (Φ,ρ)-quasiconvex) on S.

Remark 2.7: 1) If $K=\mathbf{R}_+^m$ and $\varphi: S\times S\times\mathbf{R}^{n+1}\to\mathbf{R}$ is of the form

$$\varphi(x, u; (A, \rho)) = F(x, u, A) + \rho d(x, u)$$

where $F(x, u, \cdot\,)$ is sublinear, ρ is a constant and $d : S\times S\to\mathbf{R}_+$, then K-generalized (Φ,ρ)-convexity reduces to (F,ρ)-convexity introduced by Preda [9].

2) If f is a scalar valued function and $K=\mathbf{R}_+$, then Definition 2.4 becomes the definition of (Φ,ρ)-invexity given by Antczak and Stasiak [2].

3) If f is a differentiable function and $K=\mathbf{R}_+^m$, then the above definitions reduce to the corresponding definitions introduced in [6].

4) If $K=\mathbf{R}_+^m$ then Definition 2.4 becomes the definition of (Φ,ρ)-invexity introduced by Antczak [1].

Now we give an example of a *K*-generalized (Φ,*ρ*)-convex function.

Example 2.8. Let $S=\mathbf{R}^2$ and $K=\{(x,y): x\le 0,\ y\ge x\}$. Consider the following nonsmooth function $f:S\to\mathbf{R}^2$, $f(x)=(f_1(x),\, f_2(x))$.

$$f_1(x_1,x_2)=\begin{cases}-x_1, & x_1\ge 0\\ 2x_1x_2, & x_1<0\end{cases}$$

$$f_2(x_1,x_2)=\begin{cases}\frac{1}{2}x_1+\frac{1}{3}x_2^4, & x_1\ge 0\\ x_1^2+x_2^2, & x_1<0\end{cases}$$

Here,

$\partial f_1(0,0)=(A_{11},A_{12}),\ A_{11}\in[-1,0],\ A_{12}\in\{0\}$

and $\partial f_2(0,0)=(A_{21},A_{22}),\ A_{21}\in[0,\frac{1}{2}],\ A_{22}\in\{0\}$.

Define $\varphi: S\times S\times\mathbf{R}^3\to\mathbf{R}$ as

$$\varphi(x,u;(a,\rho))=\begin{cases}(x_1+x_2^4)\rho, & x_1\ge 0\\ (x_1^2+x_2^2)e^{-(a_1+a_2)}, & x_1<0\end{cases}.$$

Note that $\varphi(x,u;(.,.))$ is convex on $\mathbf{R}^3$, $\varphi(x,u;(0,\ r))\ge 0$, for every $(x,\ u)\in S\times S$ and any $r\in\mathbf{R}_+$.

Set $\rho=(0,\frac{1}{3})$. Then, at $u=(0,0)$ we have

$$f(x)-f(u)-\Phi(x,u;(A,\rho))$$
$$=\begin{cases}(-x_1,\frac{1}{6}x_1), & x_1\ge 0\\ (2x_1x_2-(x_1^2+x_2^2)e^{-(A_{11}+A_{12})}, & \\ (x_1^2+x_2^2)(1-e^{-(A_{21}+A_{22})})), & x_1<0\end{cases}$$

which gives that,

$f(x)-f(u)-\Phi(x,u;(A,\rho))\in K$, for every $x\in S$ and $A\in\partial f(0,0)$.

Hence, *f* is *K*-generalized (Φ,*ρ*)-convex at *u* on *S*.

It is clear that every *K*-generalized (Φ,*ρ*)-convex function is *K*-generalized (Φ,*ρ*)-pseudoconvex. Converse of this statement may not be true as shown by the following example.

Example 2.9. Let $S=\mathbf{R}^2$ and $K=\{(x,y): x\ge 0,\ y\ge x\}$. Consider the following nonsmooth function $f:S\to\mathbf{R}^2$, $f(x)=(f_1(x),\, f_2(x))$.

$$f_1(x_1,x_2)=\begin{cases}-x_1, & x_1\ge 0\\ 0, & x_1<0\end{cases}$$

$$f_2(x_1,x_2)=\begin{cases}-x_1-2x_2, & x_1\ge 0\\ x_1^2+x_2^2, & x_1<0\end{cases}$$

Here,

$\partial f_1(0,0)=(A_{11},A_{12}),\ A_{11}\in[-1,0],\ A_{12}\in\{0\}$

and $\partial f_2(0,0)=(A_{21},A_{22}),\ A_{21}\in[-1,0],\ A_{22}\in[-2,0]$.

Define $\varphi: S\times S\times\mathbf{R}^3\to\mathbf{R}$ as

$$\varphi(x,u;(a,\rho))=\begin{cases}(x_1+x_2^2)\rho, & x_1\ge 0\\ (x_1^2+x_2^2)e^{a_1+a_2}, & x_1<0\end{cases}.$$

Note that, $\varphi(x,u;(.,.))$ is convex on $\mathbf{R}^3$, $\varphi(x,u;(0,\ r))\ge 0$, for every $(x,\ u)\in S\times S$ and any $r\in\mathbf{R}_+$.

Set $\rho=(-\frac{1}{2},-1)$. Then, at $u=(0,0)$ we have

$$f(x)-f(u)\in-\operatorname{int}K\Rightarrow x_1>0,x_2>0$$
$$\Rightarrow\Phi(x,u;(A,\rho))\in-\operatorname{int}K,$$

for every $x\in S$ and $A\in\partial f(0,0)$.

Thus *f* is *K*-generalized (Φ,*ρ*)-pseudoconvex at *u* on *S*. But *f* fails to be *K*-generalized (Φ,*ρ*)-convex at *u* on *S* because for *x* = (4,1),

$$f(x)-f(u)-\Phi(x,u;(A,\rho))=\left(-\frac{3}{2},-1\right)\notin K.$$

3. Optimality conditions

Consider the following nonsmooth vector optimization problem over cones.

(NVOP) *K*-minimize *f*(*x*)

subject to $-g(x)\in Q$,

where $f: S\to\mathbf{R}^m$, $g: S\to\mathbf{R}^p$ are locally Lipschitz vector-valued functions and *S* is a nonempty open subset of $\mathbf{R}^n$. *K* and *Q* are closed convex pointed cones with nonempty interiors in $\mathbf{R}^m$ and $\mathbf{R}^p$ respectively.

Let $S_0=\{x\in S: -g(x)\in Q\}$ denote the set of feasible solutions of (NVOP).

Definition 3.1. A point $\bar{x}\in S_0$ is said to be

(i) a weak minimum of (NVOP) if for every $x \in S_0$

$$f(x) - f(\bar{x}) \notin -\operatorname{int} K.$$

(ii) a minimum of (NVOP) if for every $x \in S_0$

$$f(x) - f(\bar{x}) \notin -K \setminus \{0\}.$$

The following constraint qualification and Karush-Kuhn-Tucker type necessary optimality conditions are a direct precipitation from Craven [5].

Definition 3.2. (Slater-type cone constraint qualification). The problem (NVOP) is said to satisfy Slater-type cone constraint qualification at $\bar{x}$ if, for all $B \in \partial g(\bar{x})$, there exists a vector $\Omega \in \mathbf{R}^n$ such that $B\Omega \in -\operatorname{int} Q$.

Theorem 3.3. If a vector $\bar{x} \in S_0$ is a weak minimum for (NVOP) with $S = \mathbf{R}^n$ at which Slater-type cone constraint qualification holds, then there exist Lagrange multipliers $\bar{\lambda} \in K^* \setminus \{0\}$ and $\bar{\mu} \in Q^*$, such that

$$0 \in \partial(\bar{\lambda}^t f + \bar{\mu}^t g)(\bar{x})$$

$$\bar{\mu}^t g(\bar{x}) = 0.$$

Note that, for $\bar{\lambda} = (\bar{\lambda}_1, ..., \bar{\lambda}_m)^t \in \mathbf{R}^m$ and $\bar{\mu} = (\bar{\mu}_1, ..., \bar{\mu}_p)^t \in \mathbf{R}^p$,

$$\partial(\bar{\lambda}^t f + \bar{\mu}^t g)(\bar{x}) \subseteq (\partial f(\bar{x})^t \bar{\lambda} + \partial g(\bar{x})^t \bar{\mu}).$$

Now we give the generalized form of nonsmooth KKT sufficient optimality conditions for (NVOP).

Theorem 3.4. Let f be K-generalized (Φ,ρ)-convex and g be Q-generalized (Φ,σ)-convex at $\bar{x} \in S_0$ on S_0. If there exist $\bar{\lambda} \in K^* \setminus \{0\}$ and $\bar{\mu} \in Q^*$, such that

$$0 \in (\partial f(\bar{x})^t \bar{\lambda} + \partial g(\bar{x})^t \bar{\mu}), \quad (1)$$

$$\bar{\mu}^t g(\bar{x}) = 0, \quad (2)$$

$$\sum_{i=1}^{m} \bar{\lambda}_i + \sum_{j=1}^{p} \bar{\mu}_j > 0, \quad (3)$$

$$\bar{\lambda}^t \rho + \bar{\mu}^t \sigma \geq 0, \quad (4)$$

then $\bar{x}$ is a weak minimum for (NVOP).

Proof: Suppose to the contrary that $\bar{x}$ is not a weak minimum for (NVOP). Then there exists $\hat{x} \in S_0$ such that

$$f(\hat{x}) - f(\bar{x}) \in -\operatorname{int} K. \quad (5)$$

By virtue of (1), there exist

$$\bar{A} = (\bar{A}_1, ..., \bar{A}_m)^t \in \partial f(\bar{x})$$

and $\bar{B} = (\bar{B}_1, ..., \bar{B}_p)^t \in \partial g(\bar{x})$

such that,

$$\bar{A}^t \bar{\lambda} + \bar{B}^t \bar{\mu} = 0. \quad (6)$$

Since f is K-generalized (Φ,ρ)-convex at $\bar{x}$ on S_0, we have

$$f(\hat{x}) - f(\bar{x}) - \Phi(\hat{x}, \bar{x}; (\bar{A}, \rho)) \in K. \quad (7)$$

Adding (5) and (7) we get,

$$\Phi(\hat{x}, \bar{x}; (\bar{A}, \rho)) \in -\operatorname{int} K. \quad (8)$$

Since $\bar{\lambda} \in K^* \setminus \{0\}$, we have

$$\bar{\lambda}^t \Phi(\hat{x}, \bar{x}; (\bar{A}, \rho)) < 0. \quad (9)$$

Also, since g is Q-generalized (Φ,σ)-convex at $\bar{x}$ on S_0 and $\bar{\mu} \in Q^*$, therefore

$$\bar{\mu}^t \{g(\hat{x}) - g(\bar{x}) - \Phi(\hat{x}, \bar{x}; (\bar{B}, \sigma))\} \geq 0.$$

However, $\hat{x} \in S_0, \bar{\mu} \in Q^*$ and (2) together imply

$$\bar{\mu}^t \Phi(\hat{x}, \bar{x}; (\bar{B}, \sigma)) \leq 0. \quad (10)$$

From (9) and (10), we have

$$\bar{\lambda}^t \Phi(\hat{x}, \bar{x}; (\bar{A}, \rho)) + \bar{\mu}^t \Phi(\hat{x}, \bar{x}; (\bar{B}, \sigma)) < 0. \quad (11)$$

Define $\tau = \dfrac{1}{\sum_{i=1}^{m} \bar{\lambda}_i + \sum_{j=1}^{p} \bar{\mu}_j}$,

$$\xi_i = \tau \bar{\lambda}_i, \; i = 1, 2, ..., m,$$

$$\zeta_j = \tau \bar{\mu}_j, \; j = 1, 2, ..., p.$$

Let $\bar{\xi} = (\bar{\xi}_1, ..., \bar{\xi}_m)^t$ and $\bar{\zeta} = (\bar{\zeta}_1, ..., \bar{\zeta}_p)^t$.

(3), (4) and (6) respectively imply $\tau > 0$, $\bar{\xi}^t \rho + \bar{\zeta}^t \sigma \geq 0$ and $\bar{\xi}^t \bar{A} + \bar{\zeta}^t \bar{B} = 0$.

Also, by definition $\sum_{i=1}^{m} \bar{\xi}_i + \sum_{j=1}^{p} \bar{\zeta}_j = 1$.

Thus, using the properties of φ, we have

$$0 \le \varphi(\hat{x}, \bar{x}; (\bar{\xi}^t \bar{A} + \bar{\zeta}^t \bar{B},\ \bar{\xi}^t \rho + \bar{\zeta}^t \sigma))$$
$$= \varphi(\hat{x}, \bar{x}; (\sum_{i=1}^{m} \bar{\xi}_i \bar{A}_i + \sum_{j=1}^{p} \bar{\zeta}_j \bar{B}_j, \sum_{i=1}^{m} \bar{\xi}_i \rho_i + \sum_{j=1}^{p} \bar{\zeta}_j \sigma_j))$$
$$\le \sum_{i=1}^{m} \bar{\xi}_i \varphi(\hat{x}, \bar{x}; (\bar{A}_i, \rho_i)) + \sum_{j=1}^{p} \bar{\zeta}_j \varphi(\hat{x}, \bar{x}; (\bar{B}_j, \sigma_j))$$
$$= \tau(\bar{\lambda}^t \Phi(\hat{x}, \bar{x}; (\bar{A}, \rho)) + \bar{\mu}^t \Phi(\hat{x}, \bar{x}; (\bar{B}, \sigma))) < 0 \quad \text{(by (11))},$$

which is a contradiction.

Hence, $\bar{x}$ is a weak minimum for (NVOP).

Theorem 3.5. Let f be K-generalized (Φ,ρ)-pseudoconvex and g be Q-generalized (Φ,σ)-quasiconvex at $\bar{x} \in S_0$ on S_0 and suppose there exist $\bar{\lambda} \in K^* \setminus \{0\}$ and $\bar{\mu} \in Q^*$ such that (1), (2), (3) and (4) hold, then $\bar{x}$ is a weak minimum for (NVOP).

Proof. Let, if possible, $\bar{x}$ be not a weak minimum for (NVOP). Then there exists $\hat{x} \in S_0$ such that (5) holds.

In view of (1) there exist $\bar{A} \in \partial f(\bar{x})$ and $\bar{B} \in \partial g(\bar{x})$ such that (6) is satisfied.

Since f is K-generalized (Φ,ρ)-pseudoconvex at $\bar{x}$ on S_0, therefore from (5), we have

$$-\Phi(\hat{x}, \bar{x}; (\bar{A}, \rho)) \in \operatorname{int} K.$$

Now $\bar{\lambda} \in K^* \setminus \{0\}$ gives $\bar{\lambda}^t \Phi(\hat{x}, \bar{x}; (\bar{A}, \rho)) < 0$.

As $\hat{x} \in S_0$ and $\bar{\mu} \in Q^*$, we have $\bar{\mu}^t g(\hat{x}) \le 0$. On using (2), we get

$$\bar{\mu}^t \{g(\hat{x}) - g(\bar{x})\} \le 0. \tag{12}$$

If $\bar{\mu} \ne 0$, then (12) implies $g(\hat{x}) - g(\bar{x}) \notin \operatorname{int} Q$.

Since g is Q-generalized (Φ,σ)-quasiconvex at $\bar{x}$ on S_0, therefore

$$-\Phi(\hat{x}, \bar{x}; (\bar{B}, \sigma)) \in Q,$$

so that, $\bar{\mu}^t \Phi(\hat{x}, \bar{x}; (\bar{B}, \sigma)) \le 0.$ (13)

If $\bar{\mu} = 0$, then also (13) holds.

Now proceeding as in the last part of Theorem 3.4, we get a contradiction. Hence $\bar{x}$ is a weak minimum for (NVOP).

Theorem 3.6. Let f be K-generalized (Φ,ρ)-convex and g be Q-generalized (Φ,σ)-convex at $\bar{x} \in S_0$ on S_0. Suppose there exist $\bar{\lambda} \in K^{s*}$ and $\bar{\mu} \in Q^*$ such that (1), (2), (3) and (4) hold, then $\bar{x}$ is a minimum for (NVOP).

Proof. Let if possible $\bar{x}$ be not a minimum for (NVOP), then there exists $\hat{x} \in S_0$ such that

$$f(\bar{x}) - f(\hat{x}) \in K \setminus \{0\}. \tag{14}$$

As (1) holds, there exist $\bar{A} \in \partial f(\bar{x})$ and $\bar{B} \in \partial g(\bar{x})$ such that (6) holds.

Since f is K-generalized (Φ, ρ)-convex at $\bar{x}$ on S_0, therefore proceeding on the similar lines as in proof of Theorem 3.4 and using (14) we have

$$-\Phi(\hat{x}, \bar{x}; (\bar{A}, \rho)) \in K \setminus \{0\}.$$

As $\bar{\lambda} \in K^{s*}$, we have $\bar{\lambda}^t \Phi(\hat{x}, \bar{x}; (\bar{A}, \rho)) < 0.$

This leads to a contradiction as in Theorem 3.4. Hence $\bar{x}$ is a minimum for (NVOP).

4. Duality

We associate with the primal problem (NVOP), the following Wolfe-type dual problem (NWOD):

(NWOD)K-maximize $f(y) + \mu^t g(y) l$

$$\text{subject to } 0 \in (\partial f(y)^t \lambda + \partial g(y)^t \mu), \tag{15}$$

$$y \in S,\ l \in \operatorname{int} K,\ \lambda \in K^* \setminus \{0\},\ \mu \in Q^* \text{ and } \lambda^t l = 1.$$

We now establish duality results between (NVOP) and (NWOD).

Let W denote the set of feasible solutions of (NWOD) and Y_W be the subset of S given by $Y_W = \{y \in S : (y, \lambda, \mu) \in W\}$.

Theorem 4.1.(Weak Duality). Let x be feasible for (NVOP) and (y, λ, μ) be feasible for (NWOD). If f is K-generalized (Φ,ρ)-convex at y on $S_0 \cup Y_W$, g is Q-generalized (Φ,σ)-convex at y on $S_0 \cup Y_W$, $\sum_{i=1}^{m} \lambda_i + \sum_{j=1}^{p} \mu_j > 0$ and $\lambda^t \rho + \mu^t \sigma \ge 0$, then

$$f(y) + \mu^t g(y) l - f(x) \notin \operatorname{int} K. \tag{16}$$

Proof. Let if possible,

$$f(y) + \mu^t g(y) l - f(x) \in \operatorname{int} K. \tag{17}$$

Since (y, λ, μ) is feasible for (NWOD), therefore by (15), there exist

$\bar{A} \in \partial f(y)$ and $\bar{B} \in \partial g(y)$ such that

$$\bar{A}^t \lambda + \bar{B}^t \mu = 0. \quad (18)$$

Since f is K-generalized (Φ,ρ)-convex at y on $S_0 \bigcup Y_W$, therefore

$$f(x) - f(y) - \Phi(x, y; (\bar{A}, \rho)) \in K. \quad (19)$$

Adding (17) and (19), we get

$$\mu^t g(y) l - \Phi(x, y; (\bar{A}, \rho)) \in \operatorname{int} K.$$

As $\lambda \in K^* \setminus \{0\}$ and $\lambda^t l = 1$, we have

$$\mu^t g(y) - \lambda^t \Phi(x, y; (\bar{A}, \rho)) > 0 . \quad (20)$$

Again, since $x \in S_0$, g is Q-generalized (Φ,σ)-convex at y on $S_0 \bigcup Y_W$ and $\mu \in Q^*$, therefore

$$\mu^t [g(x) - g(y) - \Phi(x, y; (\bar{B}, \sigma))] \geq 0. \quad (21)$$

From (20) and (21), we have $\lambda^t \Phi(x, y; (\bar{A}, \rho)) + \mu^t \Phi(x, y; (\bar{B}, \sigma)) < \mu^t g(x)$.

Since x is feasible for (NVOP) and $\mu \in Q^*$, $\mu^t g(x) \leq 0$, so that we have

$$\lambda^t \Phi(x, y; (\bar{A}, \rho)) + \mu^t \Phi(x, y; (\bar{B}, \sigma)) < 0 .$$

Now proceeding as in proof of Theorem 3.4, we obtain a contradiction. Hence (16) holds.

This weak duality result allows us to obtain strong duality result as follows.

Theorem 4.2. (Strong Duality). Let $\bar{x}$ be a weak minimum for (NVOP) at which Slater-type cone constraint qualification is satisfied. Then there exist $\bar{\lambda} \in K^* \setminus \{0\}$ and $\bar{\mu} \in Q^*$ such that $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is feasible for (NWOD). Moreover, if the conditions of Theorem 4.1, are satisfied for each feasible solution of (NWOD), then $\bar{x}$ is a weak maximum for (NWOD).

Proof. Since $\bar{x}$ is a weak minimum of (NVOP), therefore by Theorem 3.3, there exist $\bar{\lambda} \in K^* \setminus \{0\}$, $\bar{\mu} \in Q^*$ such that (1) and (2) hold.

Thus $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is feasible for (NWOD). Now assume on the contrary that $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is not a weak maximum for (NWOD), then there exists a feasible solution (y, λ, μ) for (NWOD) such that

$$\{f(y) + \mu^t g(y) l\} - \{f(\bar{x}) + \bar{\mu}^t g(\bar{x}) l\} \in \operatorname{int} K ,$$

which on using (2) gives

$$f(y) + \mu^t g(y) l - f(\bar{x}) \in \operatorname{int} K .$$

This contradicts Weak Duality Theorem 4.1. Hence $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a weak maximum for (NWOD).

Now we consider the following Mond-Weir type dual (NMOD) related to problem (NVOP):

(NMOD)K-maximize $f(y)$

$$\text{subject to } 0 \in \partial f(y)^t \lambda + \partial g(y)^t \mu \quad (22)$$

$$\mu^t g(y) \geq 0 , \quad (23)$$

$$y \in S,\ \lambda \in K^* \setminus \{0\} \text{ and } \mu \in Q^* .$$

Let M denote the set of feasible solutions of (NMOD) and Y_M be the subset of S defined by $Y_M = \{y \in S : (y,\ \lambda, \mu) \in M\}$.

Theorem 4.3. (Weak Duality). Let x be feasible for (NVOP) and $(y,\ \lambda, \mu)$ be feasible for (NMOD). Suppose f is K-generalized (Φ,ρ)-pseudoconvex and g is Q-generalized (Φ,σ)-quasiconvex at y on $S_0 \bigcup Y_M$ such that $\sum_{i=1}^{m} \lambda_i + \sum_{j=1}^{p} \mu_j > 0$ and $\lambda^t \rho + \mu^t \sigma \geq 0$, then

$$f(y) - f(x) \notin \operatorname{int} K . \quad (24)$$

Proof. Assume on the contrary,

$$f(y) - f(x) \in \operatorname{int} K . \quad (25)$$

Since $(y,\ \lambda, \mu)$ is feasible for (NMOD), there exist $\bar{A} \in \partial f(y)$ and $\bar{B} \in \partial g(y)$ such that (18) holds.

As f is K-generalized (Φ,ρ)-pseudoconvex at y on $S_0 \bigcup Y_M$, therefore from (25), we have

$$-\Phi(x, y; (\bar{A}, \rho)) \in \operatorname{int} K.$$

Since $\lambda \in K^* \setminus \{0\}$, we get $\lambda^t \Phi(x, y; (\bar{A}, \rho)) < 0$.

Also, $x \in S_0$ and $\mu \in Q^*$ so that $\mu^t g(x) \leq 0$. This together with (23) gives $\mu^t \{g(x) - g(y)\} \leq 0$.

Now proceeding on similar lines as in proof of Theorem 3.5 we get a contradiction. Hence (24) holds.

Theorem 4.4. (Strong Duality). Let $\bar{x}$ be a weak minimum of (NVOP) at which Slater-type cone constraint qualification is satisfied. Then there exist $\bar{\lambda} \in K^* \setminus \{0\}$ and $\bar{\mu} \in Q^*$ such that

$(\bar{x}, \bar{\lambda}, \bar{\mu})$ is feasible for (NMOD). Moreover, if the conditions of Weak Duality Theorem 4.3 are satisfied for each feasible solution (y, λ, μ) of (NMOD), then $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a weak maximum of (NMOD).

Proof. The proof is similar to that of Theorem 4.2 except that we invoke Theorem 4.3 instead of Theorem 4.1.

Acknowledgments

The authors would like to thank the reviewers for their valuable comments that helped to improve the manuscript.

References

[1] Antczak, T., On nonsmooth (Φ,ρ)-invexmultiobjective programming in finite-dimensional Euclidean spaces, Journal of Advanced Mathematical Studies, 7, 127-145 (2014).

[2] Antczak, T., Stasiak, A., (Φ,ρ)-invexity in nonsmooth optimization, Numerical Functional Analysis and Optimization, 32(1), 1-25 (2010).

[3] Caristi, G., Ferrara, M., Stefanescu, A., Mathematical programming with (Φ,ρ)-invexity. In: Konnov, I.V., Luc, D.T., Rubinov, A.M. (eds.) Generalized Convexity and Related Topics. Lecture Notes in Economics and Mathematical Systems, 583, 167-176. Springer, Berlin-Heidelberg-New York (2006).

[4] Clarke, F. H., Optimization and Nonsmooth Analysis, Wiley, New York (1983).

[5] Craven, B. D., Nonsmooth multiobjective programming, Numerical Functional Analysis and Optimization, 10(1-2), 49-64 (1989).

[6] Ferrara, M., Stefanescu, M.V., Optimality conditions and duality in multiobjective programming with (Φ,ρ)-invexity. Yugoslav Journal of Operations Research, 18, 153-165 (2008).

[7] Hanson, M.A., On sufficiency of the Kuhn-Tucker conditions, Journal of Mathematical Analysis and Applications, 80, 545-550 (1981).

[8] Hanson, M.A., Mond, B., Further generalization of convexity in mathematical programming. Journal of Information and Optimization Sciences, 3, 25-32 (1982).

[9] Preda, V., On efficiency and duality for multiobjective programs, Journal of Mathematical Analysis and Applications, 166, 365-377 (1992).

[10] Vial, J.P., Strong and weak convexity of sets and functions, Mathematics of Operations Research, 8, 231-259 (1983).

Surjeet Kaur Suneja *has recently retired as Associate Professor in Department of Mathematics, Miranda House, University of Delhi (India). She completed her Ph.D. in 1984 from the Department of Mathematics, University of Delhi and has more than 75 publications in journals of international repute to her credit. Ten students have completed their Ph.D. while over 21 have done their M.Phil. under her exemplary supervision. Her areas of interest include vector optimization, nonsmooth optimization, generalized convexity and variational inequality problems.*

Sunila Sharma *is an Associate Professor in Department of Mathematics, Miranda House, University of Delhi. She completed her Ph.D. in 1999 from the Department of Mathematics, University of Delhi. Her areas of interest include vector optimization, nonsmooth optimization and generalized convexity.*

Malti Kapoor *is an Assistant Professor in Department of Mathematics, Motilal Nehru College, University of Delhi. She completed her Masters in 2004, M.Phil. in 2006 and is currently pursuing Ph.D. from the Department of Mathematics, University of Delhi. Her areas of interest include generalized convexity, vector optimization and nonsmooth optimization.*

An International Journal of Optimization and Control:
Theories & Applications

Volume: 6 Number: 1
January 2016

This page intentionally left blank.

An International Journal of Optimization
and Control: Theories & Applications
Vol.6, No.1, pp.9-22 (2016) © IJOCTA
ISSN: 2146-0957 eISSN: 2146-5703
DOI: 10.11121/ijocta.01.2016.00258
http://www.ijocta.com

Transit network design and scheduling using genetic algorithm – a review

Amita Johar, S. S. Jain and P. K. Garg

Department of Civil Engineering, Indian Institute of Technology Roorkee, India
Email: avamitamit@gmail.com, profssjain@gmail.com, pkgiitr@gmail.com

(Received April 1, 2015; in final form July 15, 2015)

Abstract. The aim of this paper is to summarize the findings of research concerning the application of genetic algorithm in transit network design and scheduling. Due to the involvement of several parameters the design and scheduling of transit network by means of traditional optimization technique is very difficult. To overcome these problems, most of the researchers have applied genetic algorithm for designing and scheduling of transit network. After the review of various studies involved in design and scheduling of transit network using genetic algorithm, it was concluded that genetic algorithm is an efficient optimization technique.

Keywords: Transit network, genetic algorithm, optimization technique
AMS Classification: 90B10, 90B15.

1. Introduction

In developing countries like India traffic congestion, slow speed of vehicle and poor level of service are the major problems encountered in our daily life. These problems are due to huge growth of vehicular population specially the private and intermediate transport service [1-3]. In this view of rapid development it necessary to plan and design the public transport system in an efficient manner so that the use of private and intermediate transport service is reduced.

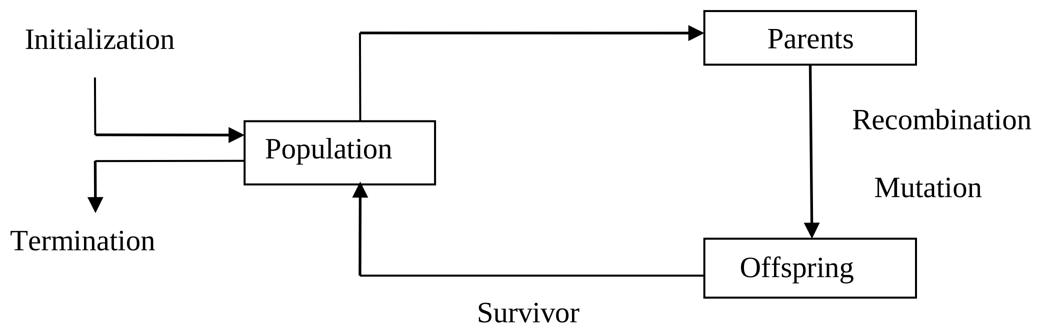
The transport system is efficient if design and schedule of transit network is efficient. From the user point of view, the system is efficient if it meets the demand by providing cheap and direct service to the passenger, and from the operator point of view the system is efficient if it makes as much profit as possible. This is the main challenge in the transit planning to find balance between these conflicting objectives, various optimization techniques come in to the game [4].

Among various optimization techniques the genetic algorithm offers a new strategy with enormous potential for many tasks in planning and designing of transit network. It is an area of interest indicating how genetic algorithm addresses the shortcoming of conventional optimization techniques. In the present study an attempt has been made to explore the application of genetic algorithm in routing, scheduling, combined routing and scheduling and integration of mass transit planning.

2. Genetic algorithm

Genetic algorithms, search optimization techniques are based on the mechanics of natural selection. It is an evolutionary algorithm. The basic mechanics of genetic algorithm are simple involving copying strings and swapping partial strings. The major steps involve in the GA implementation algorithm are generation of population, finding the fitness function, and application of genetic operator and evaluation of population [5] as shown in Figure 1.

Corresponding Author. Email: avamitamita@gmail.com



Pseudo –code

Begin

INITIALISE population with random candidate solution

EVALUATE each candidate.

REPEAT UNTILL (termination condition) is satisfied DO

1. SELECT Parents
2. RECOMBINE Pairs of parents
3. MUTATE the resulting offspring.
4. SELECT individuals or the next generation.

END

Figure 1. General scheme of evolutionary process [6]

GA starts with the population of randomly created string structure representing the decision variable. The size of the population depends upon the string length and problem being optimized. These strings are called chromosome in biological system. The associate value with the chromosome is called the "fitness function". These strings consist of coding and binary coding which is most common coding method and GA performs best if adopted [7] i.e. ones and zeros or "bits". Every position in chromosome consists of "genes" having value as "allele" (e.g., 0 or 1). Initially the allele of chromosome is generated as simple tossing of an unbiased coin and consecutive flips (head=1 and tail=0) can be used to decide genes in coding of a population strings. Thus population having individuals is generated by pseudorandom generator whose individuals represent a feasible solution in solution space called initial solution. After deciding the encoding method as binary strings, its length is determined according to desired precision. While in the process of coding, the corresponding point can be found using fixed mapping rule [7]. Suppose the function of n variables, $f(x_1, x_2,\dots,x_n)$: $R^n \in R$ to be minimized for each decision variable s_i then linear mapping rule is:

$$X_i = X_{i\min} + \frac{X_{i\max} - X_{i\min}}{2^{m_i} - 1} \times decoded\_value \ (S_i) \qquad (1)$$

Where,

X_{imin} = lower bound on decision variable X_i

X_{imax} = upper bound on decision variable X_i

The variable X_i is coded into substring S_i of length m_i then,

Decoded value (s_i) is $\sum_{i=0}^{i=1} 2^i S_i$, where S_i € (0, 1) and string is represented as (S_{m-1}, S_{m-2}, S_2, S_1, S_0). After decoding all decision variables using above mapping rule, the function value can be calculated by substituting the variables in the given objective function F(x). The objective function value is used as a measure of "goodness" of the string and called as "fitness" in GA terminology. The obtained accuracy of a variable for a m_i-bit coding is

$$\frac{X_{i\max} - X_{i\min}}{2^{m_i} - 1} \qquad (2)$$

In the next step, fitness function f(x) is derived from the objective function and used in successive genetic operators. GAs is naturally suitable for solving maximization problems while for minimization problems are transformed into maximization problems using suitable transformation. For maximization problems, fitness function is the same as objective function i.e. f(x) = F(x). For minimization problems, the fitness function is an equivalent maximization problem chosen such that the optimum point remains unchanged. The fitness function used is [8]:

$$F(x) = \frac{1}{[1 + f(x)]} \quad (3)$$

With fitness function value of each string in particular generator, maximum, minimum and average fitness values of all strings in a population are calculated and checked for termination criteria. If the termination criterion is not satisfied then new population is created by applying three main genetic operators – reproduction, crossover and mutation.

Reproduction / Selection: [7, 9] is a process in which individual is copied considering their fitness function values to make more copies of better string in a population. This represents a measure of the utility or goodness related to what we want to maximize. Copying strings according to their fitness function values means that string with a high value have higher probability of contribution to one or more off-spring in the next generation. In all selection schemes the essential idea is to pick strings with more than average fitness value from current population and their multiple copies are inserted in the mating pool in a probabilistic manner as shown in Figure 2. The most commonly selection operator are uniform random selection, roulette selection and tournament selection. The former selects member of pool at random, ignoring fitness or other factors. Thus the chromosome is likely to be selected. The simplest way to implement the reproduction operator is to create a biased roulette wheel where each string in the current population has a slot sized proportionally to its fitness function value. The formula used to calculate the slot size of roulette wheel, corresponding to the reproduction probability $p_r(i)$ of the string is as follows:

Reproduction probability $P_r(i) = f_i / \sum_{i=1}^{n} f_i$ (4)

Where, n = Population size

f_i = fitness value of i string.

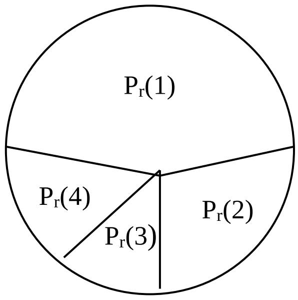


Figure 2. Reproduction operator

Crossover: [7, 9] after reproduction, crossover is applied to the string of mating pool. A crossover is used to combine two strings with the hope of creating better string. It can performed with the probability (P_c) to restrict some of the good string found previously. Two strings are chosen at random for crossover. The most commonly used crossover operators are single point crossover, double point crossover. A crossing site (represented by vertical line) is chosen at random. The contents in the right side of the crossing side are swapped between the strings. The essential idea of crossover is to exchange bits between two good strings to obtain a string that is generally better than the parent. For example, a single point crossover on five bit string is shown in Figure 3.

Mutation: [7, 9] adds new information in a random way to genetic search process and prevents an irrecoverable loss of potentially useful information which reproduction and crossover can cause. It operates at bit level, when bits are copied from current string to new string. Mutation operates with a very small mutation probability (p_m). It introduces the diversity in the population whenever the population trends to become homogeneous due to iterative use of reproduction and crossover. A coin toss mechanism is used; if a random number between 0 and 1 is less than the mutation probability, then bit is inverted *i.e.* 0 become 1 and 1 become 0. There are different type of mutation operator flip-bit, boundary, uniform, non-uniform and Gaussian. Flip-bit operator is used for binary gene; boundary, uniform, non-uniform and Gaussian operator is used for integer and float gene. In this paper the example using binary gene has been considered. To understand how to use integer gene refer [10].

Selected Strings	Single point crossover	New Strings
1 0 0 10	1 0 | 0 1 0	1 0 0 0 1
1 1 0 0 1	1 1 | 0 0 1	1 1 0 1 0

Figure 3. Crossover operator

The newly created strings are evaluated by decoding and calculating their objective function values (fitness). This whole process completes one cycle of GAs, normally called as generation. Such iterative process continues until the termination criterion is satisfied.

Termination Criteria: the population is said to be converged, when the average fitness of all the string in a population is equal to best fitness. When the population is converged, the GAs is terminated.

3. Review of GAs in design and scheduling of transit network

Various attempts have been made in designing and scheduling of transit network by different researchers. The design and scheduling of transit network using GAs has been classified into four broad categories: Routing, scheduling, combined routing and scheduling and integration of mass transit planning. Table 1 provide an overview of the approaches reported in the literature.

3.1 Routing approaches

Design of a route is an important step in planning the transit system. Bus/rail takes a major share of public transport demand. However, in most of the service areas the distribution of passenger travel is not homogenous; and therefore such location may not be cost effective in terms of operator or user point of view. For better passenger accessibility and saving of cost, reconstruction of bus route and its associated frequency must be done to suit the travel demand results in better passenger accessibility and saving of operating cost. Transit operator and commuter both give preference to short routes so that the operator cost and the travel time can be decreased, respectively.

Generally, the passengers also prefer those routes that can be easily accessed from their origin or destination trip. The route set is efficient if it satisfies the following rules:

i. The transit demand of all commuters.
ii. Transit demand of all commuters with zero transfer.
iii. Less time to travel.

Pattnaik *et al*. [11] focused on route network design and calculated associated frequencies for a given set of routes using genetic algorithm (GAs). Design consisted of two phase; first of all candidate route set was generated and then using GAs optimum route set was selected**.** The GAs was solved by fixed string length coding scheme assuming a solution set route size, and tried to find many best routes from the candidate route set. Newly proposed varaible string length method was used to found the solution route set size and set of solution routes.

Bielli *et al.* [17] focused on a new method of computing fitness function (ff) values in genetic algorithm for bus optimization. Each set was evaluated by computing a number of performance indicators obtained by analysis and aimed to achieve best bus network satisfying both demand and offer of transport. The algorithm was used to generate iteratively new set of bus networks.

Ngamchai and Lovell [23] proposed a new model to optimize bus route design which incorporates unique frequency setting for each route using GAs. The model design the bus route in three phases; firstly an initial set of route is constructed which are feasible and good. Secondly the service frequency on each route was assigned and headway coordination techniques were applied by ranking of transfer demand at transfer terminal. Lastly the existing route was modified to identify the shortest paths between origin and destination. The efficiency of the model was tested by applying it on the benchmark network. The performance result shows that proposed model is better than binary-coded genetic algorithm.

Chien *et al.* [21] develop a model using GAs to optimize bus transit system. The total cost function consisting of supplier and user costs was minimized subject to realistic demand distribution and irregular street pattern. The quality and quantity of the data can be improved by incorporating boarding demand data of census block and information of GIS (Geographical Information Systems) street network.

Table 1. Classification of papers dealing with design and scheduling of transit network

Year	Author	Objectives	Decision variables	Network structure
1998	Pattnaik *et al.*[11]	Total cost (user + operator)	Route, frequencies	Irregular grid
1998	Deb & Chakroborty [12]	Passenger waiting time	Arrival time, departure time	Not specified
2000	Gundaliya *et al.* [13]	Total cost (user + operator)	Route, frequencies	Irregular grid
2001	Chien *et al.* [14]	Total cost (user + operator)	Route, headway	Rectangular grid
2001	Kalaga *et al.*[15]	Crowding level	Route, frequencies	Irregular grid
2001	Chakroborty *et al.* [16]	Total waiting time (passenger)	Arrival time, departure time	Not specified
2002	Bielli *et al.* [17]	Multi-objective	Route, frequencies	Irregular grid
2002/ 2006	Shrivastava & Dhingra [18,19]	Total cost (user + operator)	Route, frequencies	Not specified
2002	Chakroborty & Diwedi [20]	Multi-objective	Route	Irregular grid
2003	Chien *et al.* [21]	Total cost (user + operator)	Route spacing, headway, stop spacing	Rectangular grid
2003	Tom and Mohan [22]	Total cost (user + operator)	Route, frequencies	Irregular grid
2003	Ngamchai & Lovell [23]	Total cost (user + operator)	Route, frequencies	Irregular grid
2004	Agrawal & Mathew[24]	Total cost (user + operator)	Route, frequencies	Not Specified
2005	Kidwai *et al.* [25]	Passenger wait time and crowding level	Route, frequencies	Irregular grid
2006	Zhao & Zeng [26]	Coverage, transfer, user cost	Route, frequencies	Irregular grid
2006	Kaun *et al.* [27]	Total cost (user + operator)	Route	Not Specified
2006	Verma & Dhingra [28]	Total cost (user + operator)	Route, headway	Irregular grid
2006/ 2007	Shrivastava & O'Mahony [29,30]	Total cost (user + operator)	Route, frequencies	Irregular grid
2010	Wang and Lin [31]	Operator cost, passenger travelling cost	Route, frequencies, capacity, headway	Not specified
2012	Chew and Lee [32]	Passenger Cost	Route	Irregular grid

The developed model determines the optimal solution and generates relationship among variables. The service area was divided into various sub-regions to deal with multiple bus routes situation. Agrawal and Mathew [24] proposed two parallel computational models for urban transit network using parallel genetic algorithm (PGA). The first model i.e., global parallel virtual machine (PVM) calculates the fitness function. The second model i.e., global message passing interface (MPI) substitutes MPI environment for PVM libraries. The large transit network consisting of 1332 nodes, 4076 links and 6000 demand pairs was used as a case study. Both the models were tested for better performance using various factors, like efficiency, computation time and speedup. Based on the performance measure, it is identified that global PVM model is efficient than the other model.

Wang and Lin [31] developed a bi-level programming model using genetic algorithm for mass transit route network design (MTRND). The TRTC (Taipei Rapid Transit Corporation) mass transit network was used as a case study. To attain better search space for initial feasible solution, the algorithm was formulated with smart generating methodology. The computation time of an efficient network model can be minimized using redundancy checking mechanism and gene repairing strategy. The solution quality is improved by embedding the passenger assignment model and improved fitness function. Based on the comparison between performance measures, such as two initial solution generating methods (1-car and Minimum –car) and three operators (1-point, 1-point mutation and 2-point crossover), it was found that for MTRND problem proper combination of minimum-car method and the 2-point crossover operator was suitable. The result indentifies that development model and algorithm is effective for solving MTRND problem.

Chew and Lee [32] developed a framework using GAs to solve urban transit routing problem (UTRP). In this study, the infeasible solution was converted into feasible solution using adding-node procedure. Minimizing the passenger cost was the main objective of the study. To perform genetic operation, route crossover and identical-point mutation were proposed. The Mandl's benchmark data set was used to carry out the computational experiment. The result shows that the proposed algorithm performs more efficiently when compared to other researchers as shown in Table 2.

3.2 Scheduling approaches

Careful and detail scheduling computation and precise presentation of schedulers are extremely important aspects of transit system operation. They affect efficiency and economy of operation, regularity and reliability of service and facility with which public can use system. Good scheduling means spacing transit vehicle at appropriate intervals throughout the day and evening to maintain an adequate level of service. Therefore, it minimizes both waiting time for passengers as well as transfer time from one route to another. Total waiting time of passengers is the sum of the total initial waiting time (IWT) and total transfer time (TT) of the passenger. As an effect, it will provide a better level of service to passenger at no extra cost. The resource and service related constraints are as follows:

i. Minimum fleet size: number of buses available should be finite for running on different routes.
ii. Minimum Bus capacity: capacity of the bus should be fixed.
iii. Stopping time bounds: buses cannot stop for a very little or a very long time at a stop.
iv. Policy headway: minimum frequency level should be maintained on a given route.
v. Maximum transfer time: transfer time for the passenger should not be too long.

Deb and Chakroborty [12] formulated the time scheduling problem of transit system into mixed –integer nonlinear programming problem (MINLP) while considering large number of resources and service related constraint like size of fleet, stopping time and headway. Minimize the total waiting time (initial waiting time + transfer time) of the passenger is the main objective of MINLP. Genetic algorithm based approach was selected to solve transit scheduling problem as difficulties rises while using conventional optimization techniques. This research shows that the GA_S based approach with least modification can handle various transit scheduling problems, such as minimum versus maximum capacity of bus, static versus dynamic arrival time, and single versus multiple transfer stations. Result shows that GAs based approach is capable of finding optimal results.

Kalaga *et al.* [15] presented a two-step based heuristic technique for the distribution of buses on urban bus route. In the first step, bus frequencies required to manage the peak demand on each route was worked out. To compute base frequencies, buses overcrowding at certain location was also included.

Table 2. Comparison of GAs model with models developed by other researchers [18]

Authors		**Case**	**Number of routes**	**Parameters**				
				% of Demand satisfy through transfer				
				Zero	**One**	**Two**	**Unsat**	**ATT**
Mandll [33]		I	4	69.94	29.93	0.13	0.00	12.90
Baaj and Mahmassani [34]				N/A	N/A	N/A	N/A	N/A
Kidwai [25]				72.95	26.92	0.13	0.00	12.72
Chakroborty and Dwivedi [20]				86.86	12.00	1.14	0.00	11.90
Fan and Mumford [35]				93.26	6.74	0.00	0.00	11.37
Proposed GAs	Avg Results			92.88	6.91	0.20	0.00	11.16
	Best Results			93.71	6.29	0.00	0.00	10.82
Mandll [33]		II	6	N/A	N/A	N/A	N/A	N/A
Baaj and Mahmassani [34]				78.61	21.39	0.00	0.00	11.86
Kidwai [25]				77.92	19.68	2.40	0.00	11.87
Chakroborty and Dwivedi [20]				86.04	13.96	0.00	0.00	10.30
Fan and Mumford [35]				91.52	8.48	0.00	0.00	10.48
Proposed GAs	Avg Results			93.85	5.88	0.24	0.03	10.51
	Best Results			95.57	4.43	0.00	0.00	10.28
Mandll [33]		III	7	N/A	N/A	N/A	N/A	N/A
Baaj and Mahmassani [34]				80.99	19.01	0.00	0.001	12.50
Kidwai [25]				93.91	6.09	0.00	0.00	10.69
Chakroborty and Dwivedi [20]				89.15	10.85	0.00	0.00	10.15
Fan and Mumford [35]				93.32	6.36	0.32	0.00	10.42
Proposed GAs	Avg Results			96.47	3.53	0.00	0.00	10.31
	Best Results			95.57	4.43	0.00	0.00	10.27
Mandll [33]		IV	8	N/A	N/A	N/A	N/A	N/A
Baaj and Mahmassani [34]				79.96	20.04	0.00	0.00	11.86
Kidwai [25]				84.73	15.27	0.00	0.00	11.22
Chakroborty and Dwivedi [20]				90.38	9.62	0.00	0.00	10.46
Fan and Mumford [35]				94.54	5.46	0.00	0.00	10.36
Proposed GAs	Avg Results			96.16	3.84	0.00	0.00	10.31
	Best Results			97.82	2.18	0.00	0.00	10.19

Avg: Average, Unsat: Unsatisfied, ATT: Average Travel Time

In the second step, additional frequencies were allocated in order to minimize the level of overcrowding in the network. The problem of commuters' discomfort because of overcrowding was formulated as non-linear objective function. The problem, such as allocation of superfluous was solved using GAs. Route overlapping has been considered and a frequency of buses according to one transfer was set. The model concentrates on overcrowding as measure of user cost; other factors, like waiting time and vehicle operating cost are not taken into account, which play a major role in allocation of buses. At intermediate node maximum of one transfer was considered that is generally not an actual case.

They generated frequency-setting (FRESET) model and minimized the objective function subject to constraints of feasibility loading, assignment of commuter flow and size of fleet. Loading feasibility constraint ensured that the passengers demand across each route was fulfilled by allocation of frequency of buses across each route. Two stages of FRESET model were: Base frequency allocation and surplus allocation.

Chakroborty *et al*. [16] presented a genetic algorithm based approach for optimizing the problem related to allocation of fleet size and development of schedule with transfer consideration as well as minimizing the passenger waiting time. From the past experiences, it has been identified that it is impossible to get optimal result for simple problem using conventional optimization method but it is possible to get optimal result with minimum computation effort using GAs. Limitation of the developed method is that string length in case of a larger network is generally large. Two points that needs attention are; (i) developed a real- coded GAs based approach (ii) proposed procedure must be included with transfer stops.

Shrivastava and Dhingra [18] developed a Schedule Optimization Model (SOM) for coordinating schedule of BEST buses determined on existing feeder routes. Minimizing transfer time between buses and train and operator cost are the main aims of the proposed model. The problem becomes nonlinear and non-convex due to large number of variables and constraint in the objective function, making it difficult to solve due to traditional approaches. Therefore, for coordination of sub-urban train and buses a genetic algorithm was used which is a robust optimization technique. The proposed model provides a better level of service to the commuters because they consider load factor and transfer time from train to bus rather than fleet size. It was found that less number of buses is required on existing feeder routes and it is a specific contribution towards integration of public transport mode.

Kidwai *et al*. [25] presented a bi-level optimization model for bus scheduling problem. In first level, load feasibility was determined for each route individually and by adding the number of buses across each route, the fleet size was determined. In second level, using GA_S the fleet size obtained from the first level is again minimized. Model is applied to real life network and based on the result, it is concluded that proposed algorithm yields significant saving for transit network with overlapping of routes.

3.3 Routing and scheduling approaches

Problems related to vehicle routing and scheduling (VRS) involve four decisions; (a) vehicle fleet size, (b) customer are assigned to a vehicle, (c) assigning a sequence to the vehicle which travel to the assigned customer, and (d) completing its route the actual time that vehicle travel. To solve the problem, various techniques have been used but no technique has included all the practical options or restriction confronting to a company. Unfortunately, the analysts had to be satisfied with the existing method or with slight modification they can develop their custom solution techniques. Routing and scheduling also mean an approach which deals with the route configuration and respective frequency simultaneously. This combined process of routing and scheduling involves two decisions parameters: number of routes and associated frequency.

Gundaliya *et al.* [13] proposed a GAs to develop a model for routing and scheduling**.** Objective function is minimization of total cost that is user and operator costs and the related constraints are load factor, fleet size and overloading of links. User cost is a combination of in-vehicle time, waiting time and transfer time and operator cost is vehicle operating cost of buses. Mandl's Swiss network of fifteen nodes was used to test the model. Model gives the better optimized results found by other researcher on same network and demand matrix.

Chakroborty and Dwivedi [20] proposed an algorithm using GAs that provides an efficient transit route networks. In this paper, four cases with different number of routes in the route set were used. A comparison of the proposed algorithm with algorithm developed by other author is done using various measures of effectiveness, such as (i) percentage of demand satisfied directly (ii) proportion of demand satisfied with one transfer (iii) proportion of demand satisfied with two transfers (iv) proportion of demand unsatisfied (4) average travel time per user in minutes (5) total man-hours saved per day. They also state that the developed procedure uses non-heuristic techniques in optimization process. Further they mentioned that the proposed procedure is useful for transit planners and designers.

Tom and Mohan [22] designed a route network for transit system involving selection of route set and its related frequency. The problem was formulated as an optimization function that minimized the overall cost of the system (operating cost of bus + travel time of passenger). Design of the route network was done in two stages: In the first stage, a large set of candidate route set was generated. In the second step, a solution route set with associated frequencies was chosen using GAs, from the large set of candidate route set generated during first step. The proposed model considered route frequency as the variable, thus making it different from the existing model in terms of adopted coding method. The model was studied on small size network and found that the performance of the model can be evaluated using adopted coding method for design of transit network. The SRFC model provides a solution with minimum operational cost, minimum fleet size and maximum allocation of demand which is directly satisfied. Using asymmetric demand matrix and demand sensitiveness to the service quality, this study can be extended.

Chakroborty [36] discussed the optimal routing and optimal scheduling problems and described that the problem of routing can be classified as vehicle routing (TSP (travelling salesman problem), SVPDP (single vehicle pick-up and delivery problem) and SVPDPTW (single vehicle pick-up and delivery problem with time windows) and the transit routing problem. To develop schedules for bus arrival and departure at all the stops of network for a given set of route was the optimal scheduling problem. The genetic algorithm was used to solve optimization problem which was difficult to solve using conventional optimization tools. Results for various routing and scheduling problem were obtained by applying GAs technique.

Zhao and Zeng [26] demonstrated a mathematical based stochastic methodology for optimizing transit route network using integrated Simulated Annealing (SA) and GAs search method. A computer program was developed to implement the methodology, and previously available results were used to test the feasibility of the proposed methodology. By developing time-dependent transit network optimization methods, the present study can be enhanced further to optimize a transit network for both periods that is peak and off-peak which also takes into account the waiting time and transfer penalties. It is also necessary to analyze the objective functions defined in terms of commuters and operator costs. To correctly identify the difference between two lines (i.e., is bus and rapid transit line), it is necessary to use the travel time instead of travel distance.

3.4 Integration of mass transit planning

Integration of mass transit planning means development of feeder routes and schedule coordination simultaneously. In integrated system, all the trips involve more than one mode and since passengers are subjected to transfer. The transfer is one of negative aspect of any trip but cannot be neglected. Transfer is essential because it make the integrated system quick and convenient.

Dhingra and Shrivastava [37] described the methodology for co-ordination of suburban railway and BEST buses at Mumbai. The aim of this study was to achieve optimal coordinated schedules for optimally designed feeder route network with due consideration to user and operator costs and better level of service. To meet these objectives, network optimisation and transfer optimisation models were proposed. The problem was of multi-objective nature therefore a strong optimization technique GAs was proposed for optimisation. The objective function contains minimisation of in-vehicle travel time, standing passengers and vehicle operating cost.

Kaun *et al.* [27] presented a methodology for solving the problem related to feeder bus network design using meta-heuristic (combination of GAs and Ant Colony Optimization (ACO)). To compare the performance of meta-heuristic in terms solution quality and computational efficiency, a comparison was done between randomly generated 20 test solutions. In this study, each route was sequentially developed as follows: firstly the station was randomly selected, secondly the selected stop was added to the route linking to selected station, and lastly the route length was checked. The current route is terminated if it exceeds the maximum route length, and a similar procedure is used to develop new route. The procedure continues until all the stops have been included in the routes. To test the base problem, the results are compared with those published in the literature. Computational experiments have shown that both heuristics (simulated annealing and tab search) are comparable to the state-of-the-art algorithms.

Results also indicate that procedure can be further improved by using a 3-opt procedure (used to optimize each route) formed by applying GAs and ACO, so that the performance could be as good as that of tabu search with intensification.

Shrivastava and O'Mahony [29, 30] build a model using GAs for generating the optimized feeder routes and indentifying associated frequencies for coordinating schedule of feeder buses. Thus instead of decomposing the problem in two steps (i) feeder route development (ii) schedule coordination of feeder buses, the two steps were optimized together that were complementary to each other. In this study, the authors determined the schedules coordination of feeder buses for the existing given schedules of main transit. The model produced better results in terms of improved load factor as compared to previous technique accepted by author for the same study area. The proposed model involves real life objective and constraints therefore it is specific involvement towards realistic modeling of integrated public transport system.

Shrivastava and Dhingra [19] developed a methodology that determines the feeder routes and coordinated schedules for coordinating feeder buses with suburban trains. Feeder routes were developed using heuristic feeder route generation algorithm, and GAs was used for optimizing coordinated schedules. Based on the load factor and bus waiting time, the schedules were decided. To maintain better level of service and waiting time within satisfactory limit, the load factor lies between minimum and a maximum values. From the results, it is found that optimal values can be obtained in lesser time using genetic algorithm.

Verma and Dhingra [28] described a model for building optimal coordinated schedules for urban rail and feeder bus operation. An optimization technique GAs was used for developing a model. In this study, the optimal coordinated of urban train and feeder buses was done in two parts; (i) sub-model was developed for train scheduling, and (ii) sub-model was developed for schedule coordination. The train scheduling objective function is taken as minimization of train operating cost and passenger waiting time cost (boarding the train) subject to constraint load factor and waiting time. The schedule coordination objective function is taken as minimization of sum of bus operating cost, passenger transfer time cost (transferring from train to feeder buses), and passenger waiting time cost (boarding along the feeder route) subject to constraint load factor and transfer time. Two cases were considered for coordination, in first case, buses of different types were considered, and in second case single-decker fleet buses were considered. A comparison between both the cases was done to choose the strategy that is best. On the basis of comparison, it was found that mixed fleet size buses produce optimum result for coordinated feeder bus schedules.

After examining and interpreting various literatures it is concluded that studies up until recent times are limited to operational integration of mass transit planning. The efficiency of different mass transit modes and main transit facility can be enhanced by overall system integration. The overall system integration includes operational integration, institutional integration and physical integration as shown in Figure 4.

4. Conclusion

In this review paper, we have presented the classification and analysis of studies on design and scheduling of transit network using GAs. It was found that problem related to design and scheduling of transit network are highly complex and non-linear in terms of decision variable and difficult to achieve using classical programming. GAs an optimization technique is computationally more efficient to solve the problem requiring large number of resources and services related constraints such as design and scheduling of transit network. Based on the review, it is concluded that GAs have advantage over traditional optimization techniques, as they work with clusters of points rather than a single point. Due to simulataneous prosess of more than one string it increases the posibility of global optimum solution. But still there exists some limitations: the solution for the complex problems is efficient if the evaluation of the fitness function is good. But to evaluate a good fitness function is mostly the hardest part.

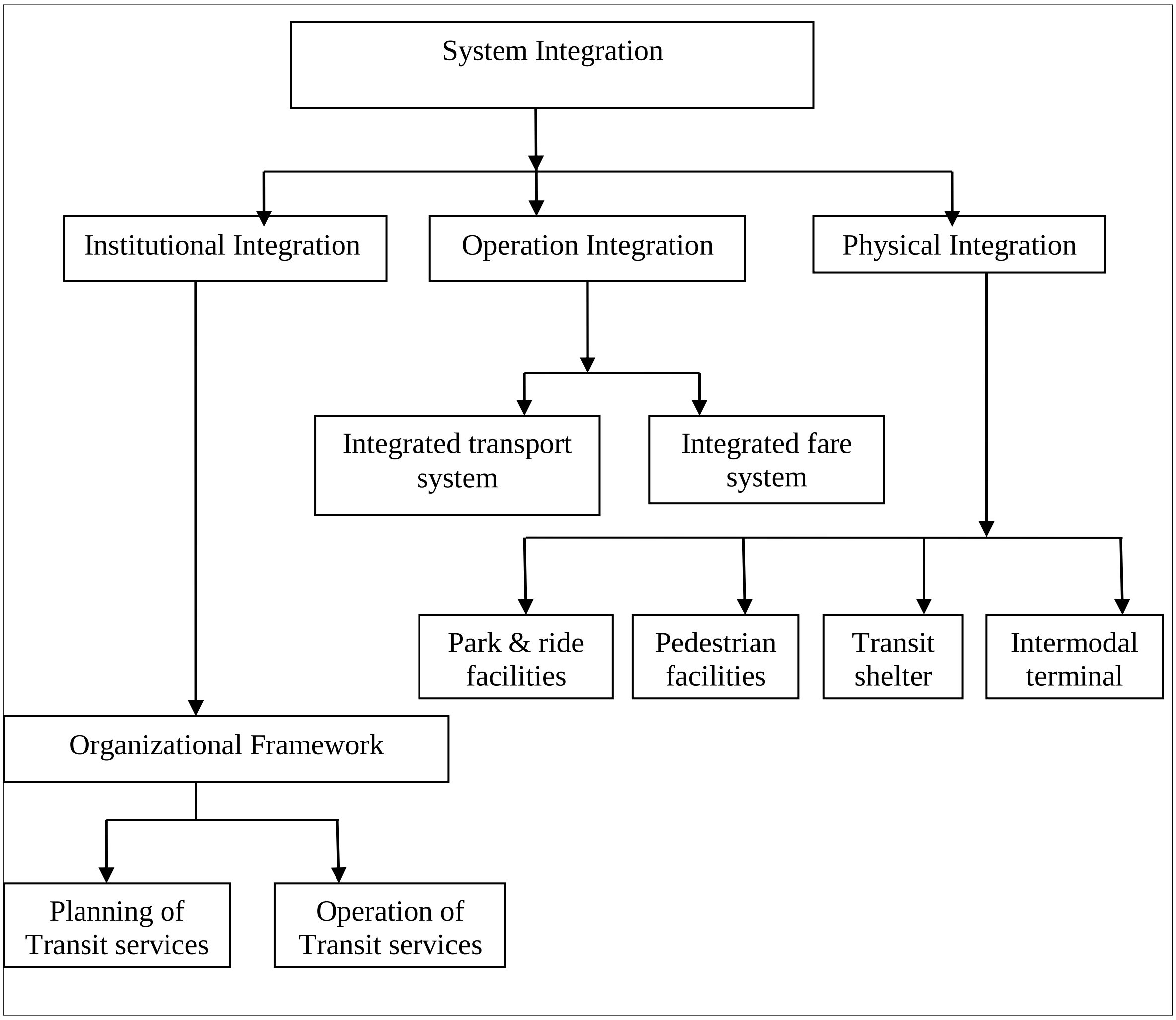


Figure 4. Overview of system integration

5. Future scope

For future scope the following points are recommended:

1. The effeciency of the integrated transport sytem can be enhanced by overall system integration.
2. The developed integrated transport system can be extended by developing integrated fare system between integrated modes which is a part of operation integration.
3. Up till now, the developed integrated systems consider the train as the main mode and the bus and the intermediated transport system (auto-rickshaw and taxi) as feeder modes. In order to develop a fully integrated transport system, personalized modes (car and two-wheeler) and non-motorized modes (cycle-rickshaw & cycle) should also be considered.

List of Abbreviation

Gas : Genetic Algorithm
SA : Simulated Annealing
ACO : Ant Colony Optimization
ff : Fitness function
GIS : Geographical Information System
PGA : Parallel Genetic Algorithm
PVM : Parallel Virtual Machine
MPI : Message Passing Interface
MTRND: Mass Transit Route Network Design
TRTC : Taipei Rapid Transit Corporation
UTRP : Urban Transit Routing Problem `
Avg : Average
Unsat : Unsatisfied
ATT : Average Travel Time
IWT : Initial Waiting Time
TT : Transfer Time
MINLP: Mixed Integer-Non-Linear Programming Problem
FRESET: Frequency-Setting
SOM : Schedule Optimization Model

VRS : Vehicle Routing and Scheduling
TSP : Travelling Salesman Problem
SVPDP : single vehicle pick-up and delivery problem
SVPDPTW : Single Vehicle Pick-up and Delivery Problem with Time Windows

Acknowledgments

The financial support in the form of MHRD fellowship from Indian Institute of Technology Roorkee, India, to Ms Amita Johar is gratefully acknowledged.

References

[1] Sinha, K. C., Sustainability and Urban Public Transportation. Journal of Transportation Engineering, 1294, 331–341 (2003).
[2] Transport Research Board (TRB), Making Transit Work. Special Report, 257, National Academy, Washington, D.C (2001).
[3] European Conference on Ministers of Transport (ECMT), Implementing Sustainable Urban Transport Policies. Final Report, ECMT/OECD Publication Service, Paris, France (2002).
[4] Guihaire, V. and Hao, J.-K., Transit Network Design and Scheduling: A Global Review. Transpotation Research Part A, 42, 1251-1273 (2008).
[5] Johar, A., Jain, S.S. and Garg, P.K., Genetic Algorithm for Transit Route Network Design, National Conference on “Twenty-Eight National Convention of Civil Engineers and National Seminar on Role of Infrastructure for Sustainable Development (RinforSD 2012), Institute of Engineers (India), Roorkee Local Centre, Roorkee, Oct 12-14, (2012).
[6] Chakroborty, R. C (2010), Fundamental of Genetic Algorithm. See. www.myreaders.info/html/aritifical_intelligence.html. Access on 14/04/12
[7] Goldberg, E. D., Genetic Algorithm in Search Optimization and Machine Learning. Addison-Wesley, Reading, MA (1989).
[8] Deb, K., Optimization for Engineering Design—Algorithms and Examples. Prentice Hall of India Pvt. Ltd, New Delhi, India, 290–320 (1995).
[9] Michalewicz, Z., Genetic Algorithm + Data structure = Evolution program. Springer, Berlin/Heidelberg (1992).
[10] Chhavi, K., Genetic Algorithm. See http://www.slideworld.com/slideshow.aspx/Genetic-Algorithms-By-Chhavi-Kashyap-ppt-732521. Access on 15/04/12.
[11] Pattanaik, S.B., Mohan, S. and Tom, V.M., Urban Bus Transit Route Network Design Using Genetic Algorithm. Journal of Transportation Engineering, 124 (4), 368-375 (1998).
[12] Deb, K. and Chakroborty, P., Time Scheduling of Transit System with Transfer Consideration Using Genetic Algorithm. Journal of Evolutionary Computation, 1, 1-24 (1998).
[13] Gundaliya, P.J., Shrivastava, P. and Dhingra, S.L., Model For Simultaneous Routing and Scheduling Using Genetic Algorithm. Transpoti Europei, Quaterly Jouranl of Transport law, Economics and Engineering, 16, 10-19 (2000).
[14] Chien, S., Yang, Z., and Hou, E., Genetic Algorithm Approach for Transit Route Planning and Design. Journal of Transportation Engineering, 127 (3), 200-207 (2001).
[15] Kalaga, R. R., Datta, R. N. and Reddy, K. S., Allocation of Buses on Interdependent Regional Bus Transit Routes. Journal of Transportation Engineering, 127 (3), 208-214 (2001).
[16] Chakroborty, P., Deb, K. and Sharma, R. k., Optimal Fleet Size Distribution and Scheduling of Transit System Using Genetic Algorithms. Transportation Planning and Technology, 24 (3), 209-225 (2001).
[17] Bielli, M., Caramia, M. and Carotenuto, P., Genetic Algorithm in Bus Network Optimization. Transportation Research Part C, 10, 19-34 (2002).
[18] Shrivastava, P. and Dhingra, S. L, Development of Coordinated Schedules Using Genetic Algorithm. Journal of Transportation Engineering, 128 (1), 89-96 (2002).
[19] Shrivastava, P. and Dhingra, S. L., Operational Integration of Suburban Railway and Public Buses—Case Study of Mumbai. Journal of Transportation Engineering, 132 (6), pp. 518-522 (2006).
[20] Chakroborty, P. and Dwivedi, T., Optimal Route Network Design for Transit Systems Using Genetic Algorithms. Engineering optimization, 34 (1), 83-100, (2002).
[21] Chien, S. I., Tsai, F. and Edwin, H., Optimization of Multiple-Route Feeder Bus Service- An Apllication of GIS. TRB Annual Meeting (2003).

[22] Tom, V.M and Mohan, S., Transit Route Network Design Using Frequency Coded Genetic Algorithm. Journal of Transportation Engineering, 129 (2), 186-195 (2003).

[23] Ngamchai, S. and Lovell, D.J., Optimal Time Transfer in Bus Transit Route Network Design Using a Genetic Algorithm. Journal of Transportation Engineering, ASCE, 129 (5), 510-521 (2003).

[24] Agrawal, J. and M, T. V., Transit Route Network Design Using Parallel Genetic Algorithm. Journal of Transportation Engineering, 18(3), 248-256 (2004).

[25] Kidwai, F. Ahmad, M. R., Deb, K. and Karim, M. R., A Genetic Algorithm Based Bus Scheduling Model for Transit Network. Proceedings of the Eastern Asia Society for Transportation Studies, 5, 477 – 489 (2005).

[26] Zhao, F. and Zeng, X., Simulated Annealing-Genetic Algorithm for Transit Network Optimization. Journal of Computing In Civil Engineering, 20 (1), 57-68 (2006).

[27] Kaun, S.N, Ong, H.L. and Ng, K.M., Solving the Feeder Bus Network Design Problem by Genetic Algorithm and Ant Colony Optimization. Advances in Engineering Software, 37, 351-359 (2006).

[28] Verma, A. and Dhingra, S. L., Developing Integrated Schedules for Urban Rail and Feeder Bus Operation. Journal of Urban Planning and Development, 132 (3), 138-146 (2006).

[29] Shrivastava, P. and O'Mahony, M., A Model for Development of Optimized Feeder Routes and Coordinated Schedules- A Genetic Algorithm Approach. Transport Policy, 13, 413-425 (2006).

[30] Shrivastava, P. and O'Mahony, M., Design of Feeder Route Network Using Combined Genetic Algorithm and Specialized Repair Heuristic. Journal of Public Transportation, Vol. 10, No. 2, pp. 99-123 (2007).

[31] Chew, J. S. C. and Lee, L. S., A Genetic Algorithm for Urban Transit Routing Problem. International Journal of Modern Physics: Conference Series, 9, 411-421 (2012).

[32] Wang, J.-Y. and Lin, C.-M., Mass Transit Route Network Design Using Genetic Algorithm. Journal of Chinese Institute of Engineers, 33 (2), 301-315 (2010).

[33] Mandl, C.E., Applied Network Optimization. Academic. Academic Press, London (1979).

[34] Baaj, M. H. and Mahmassani, H. S., An AI-based Approach for Transit Route System Planning and Design. Journal of Advanced Transportation, 25(2), 287-210 (1990).

[35] Fan, L. and Mumford, C. L., A Metaheuristic Approach to the Urban Transit Routing Problem. Journal of Heuristic. 16, 353-372 (2010).

[36] Chakroborty, P., Optimal Routing and Scheduling Transportation: Using Genetic Algorithm to Solve Difficult Optimization Problems. Direction IITK Newsletter, 6 (3), 29-40 (2004).

[37] Dhingra, S.L. and, Shrivastava P., Modelling for Co-Ordinate Bus Train Network. 6th International Conference on Computer for Urban Planning and Urban Management (CUPUM), Venue Italy (1999).

Johar Amita received his master's degree in computer science from Kurukshetra University of Haryana, India. Presently she is doing Ph.D. in the Department of Centre for Transportation System (CTRANS), Indian Institute of Technology, Roorkee. Title of his Ph.D. thesis is "Public Transport System Planning and Operation using Geospatial Techniques" under the supervision of Prof S.S Jain, Associate Faculty, CTRANS and Prof. P.K Garg, Associate Faculty, CTRANS.

Jain S.S is a Professor, Transportation Engineering Group in Department of Civil Engineering and Associated faculty at Centre for Transportation Systems (CTRANS) at Indian Institute of Technology, Roorkee. He has contributed over 430 research papers in Indian and foreign Journals and Conference Proceedings. He has completed 32 sponsored research projects including three nationally coordinated projects on Urban Transport Environment Interaction, Road Traffic Safety and Integrated Development of Public Transport System. Dr. Jain is the Fellow of Institution of Engineers (India) and Fellow of Indian Institution of Bridge Engineers. He is the recipient of 42 Medals/Prizes/Awards which includes University Gold Medal, Khosla Research Prizes and Medals, First Recipient of Jawahar Lal Nehru Award from Indian Roads Congress, Present of India Prize from Institution of Engineers, U.P. Government National Award from Indian Society for Technical Education, Prof. S.R. Mehra National Award. Prof. Jain had been the Member, Editorial Board of Institution of Civil Engineers (UK) Journal. Prof. Jain is the Member of various Technical Committees of Indian Roads Congress

and Member, World Roads Congress, Technical Committee, PIARC, France.

Garg P.K received his master degree from University of Roorkee, India in 1982 and his Ph.D degree in remote sensing from University of Bristol, U.K in 1991and his post-doctoral in Remote sensing and GIS from University of Reading, U.K in 1999. He was awarded a gold medal in M.E from University of Roorkee. He is working as a Professor at IIT Roorkee. His main areas of specializations are Satellite Image Analysis, Landuse Mapping, Land Surveying, Digital Image Processing, GPS Survey and GIS. He has contributed over 251 research papers in national and international journals and conference proceedings

An International Journal of Optimization
and Control: Theories & Applications
Vol.6, No.1, pp.23-31 (2016) © IJOCTA
ISSN: 2146-0957 eISSN: 2146-5703
DOI: 10.11121/ijocta.01.2016.00262
http://www.ijocta.com

The analysis of the interaction of the migration, diversity and permeability in parallel genetic algorithms

Gültekin Kuvat [a] and Nihat Adar [b]

[a] Department of Computer Engineering, Balıkesir University, Turkey
Email: gkuvat@balikesir.edu.tr

[b] Department of Computer Engineering, Eskişehir Osmangazi University, Turkey
Email: nadar@ogu.edu.tr

(*Received June 11, 2015; in final form December 20, 2015*)

Abstract: Diversity is an important factor for genetic algorithms to do a successful search. High diversity obtained owing to the migration is one of the most important reasons of generation of successful results by Parallel Genetic Algorithms. While permeability is a new term that analyses the effect of the migration parameters in Parallel Genetic Algorithms. In this study, the effect of changing migration intervals on the permeability and the performance of algorithm has been examined and it has been showed that the migration done in exploration phase has made much more contribution to the performance of the algorithm. In addition to this, for the different migration parameters of the individuals that will migrate the values of diversity and permeability have been calculated and the obtained results have been analysed.

Keywords: Parallel genetic algorithms; migration; diversity; permeability
AMS Classification: 65Y05, 68W10, 80M50.

1. Introduction

Genetic Algorithms (GAs) do search by using individuals that represent different regions of search space. As for Parallel Genetic Algorithms (PGAs), they have been formed by practicing GA on different subpopulations. The most important property that separates PGAs from GAs is the migration that enables the individuals to change places between populations. The migration in PGAs enables the transmission of the required amount of individuals in subpopulations to different subpopulations. Owing to the individuals shared between subpopulations by migration, the risk of being caught at the local bests in PGAs decreases [1]. The migration process by sending good individuals to other subpopulations reduces the possibility of deterioration or loss of these individuals in the big population as a result of GA operations. This result enables PGAs to generate better results than GAs.

The migration method extensively used in PGAs is ring migration method (R_PGA) [2-8]. In addition to this, the random ring method [9, 10], where migration process is carried out to a randomly chosen subpopulation, is another method that is used. The other method to be used is fully-connected migration method (FC_PGA) [3, 4, 11], where all subpopulations are connected to each other. According to this method, all subpopulations exchange individuals among themselves during the migration process. However, since this structure includes too many connections its communication cost is high. Elitist Migration Method (E_PGA) is a migration method that aims at generating good results earlier by transmitting good individuals to the right subpopulations faster. E_PGA by doing goodness evaluation among subpopulations, enables good individual to go to the good subpopulation [12-16].

Corresponding Author. Email: gkuvat@balikesir.edu.tr

In PGAs, migration parameters are used in addition to GA parameters. Migration parameters are topology, migration method, number of the migrated individuals, migration interval, the selection method of migration individuals, the method of changing places after migration, subpopulation number, subpopulation size and communication model. The selection of these parameters affects the search success in PGAs. Therefore, the studies and settings done on the migration are important for the algorithm to give faster and more efficient results.

In GAs, diversity is used for the measure of the affinity between the individuals in the population. High diversity values are necessary to obtain more successful search results [17]. However, only the examination of the individuals in the population is not enough to exhibit the relationship between diversity and PGA performance. In this study, the term of permeability is introduced as the diversity measure between populations. Thus, the effects of migration parameters to PGA have been examined with the measures of diversity in subpopulation and diversity between subpopulations.

In the following part of the study, the terms of diversity and permeability have been explained, in part 3 the effect of the migration interval changing linearly and parabolically to the permeability has been examined, in part 4 and 5 the permeability and diversity of the individuals that will migrate have been examined. As for part 6 the obtained findings have been interpreted.

2. Diversity and permeability in parallel genetic algorithms

2.1. Diversity and permeability

GAs are formed of individuals representing different regions of the search space. Owing to this, they can reach the solution in different points at the same time. However, as the iteration steps proceed if the reached solutions are out of the region in which there is the best solution, it becomes difficult to obtain a good result. Therefore, there is a need for a balanced selection pressure which will gather the individuals of the population around the best solution. Besides this, the gathering of the individuals only in a specific region is not a desired situation as well. If the region in which the individuals gathered is not the right region to do the search, a good solution cannot be caught and the possibility of getting caught to the local bests increases. For this reason, sufficient diversity and selection pressure are needed in order to reach good solutions in GAs [17]. As for PGAs the migrated individuals in them, since they are not same with the individuals in the target subpopulation, they contribute to the formation of the diversity. High diversity in all subpopulations (big population) is shown as the main reason of the PGAs' generation of good results [18].

The diversity in GAs is calculated by two methods considering fitness values or gene structure. The diversity is calculated in the first method by looking at the closeness of fitness values of population individuals with each other and in the second method according to the differences in the gene structures of the individuals. According to the second method the affinity rate between the genes are determined by using Hamming Distance (HD) and the diversity is calculated [19].

Another used gene-based diversity analysis is entropy calculation carried out on bits forming the chromosomes. In this approach, 0 and 1s in bit positions belonging to all chromosomes are counted, the rate being used in entropy expression, is evaluated as the diversity belonging to the result population obtained as chromosome length [2]. Since this approach is completely independent from fitness value and fitness function, it is easy to adapt to structures with different parameters.

As for permeability, it is calculated with the diversity analysis between the individuals by examining the subpopulations as a whole population. Since permeability is obtained by the evaluation of the individuals in all subpopulations together, it is an approach used to determine whether the genes spreaded in the big population correctly, and thus used to interpret the quality of the gene change. When permeability is provided, low permeability values; when it is not provided sufficiently high permeability values are obtained. Since the correct spread of the genes provides a faster convergence to a good solution, it is a desired situation that the permeability takes minimum values. If the migration is not carried out successfully, the permeability value doesn't come close to the minimum. In this case, subpopulations by being caught to the local best points, do not generate good results.

2.2. The calculation of diversity and permeability by entropy

The used diversity analysis method is the entropy approach calculating the diversity by counting the 0 and 1s on the bit points of chromosomes in subpopulation. In Figure 1 subpopulations in PGA and the gene distribution of the individuals belonging to them are given representatively. In this way, i expresses bit position, l chromosome length, and n expresses the number of individuals in subpopulation. 0 and 1s are counted by using Equation (1) on subpopulations formed by using binary coding and the rate of 0 and 1s in ith bit position is determined. In Equation (2) [2] where l expression represents the chromosome length, diversity value belonging to kth subpopulation is determined by using 0 and 1 rates in all bit positions. A general diversity value is calculated by averaging the diversity values belonging to subpopulations as in Equation (3) [16, 20].

Equation (1) is the expression used to calculate the rate of 0 and 1s in the ith bit position. The summand term $c_i^k(r,t)$ takes its values depending on the value at the ith bit position which belongs to the rth individual of the kth subpopulation. It takes 1 if the corresponding bit value is equal to t and 0 if not.

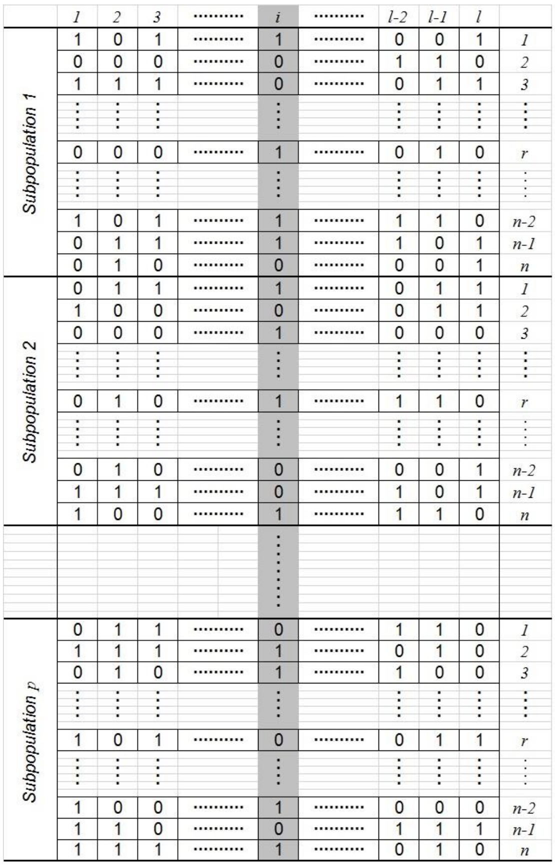


Figure 1. Subpopulation structure

Hence, $P_i^k(t)$ gives the rate of the t values at the ith bit position, which is obtained by summing the $c_i^k(r,t)$ values over all the individuals of the kth subpopulation and normalizing the summation by their total number n [16, 20].

$$P_i^k(t) = \frac{1}{n}\sum_{r=1}^{n} c_i^k(r,t)$$

$$c_i^k(r,t) = \begin{cases} 1, & \text{if the value at the } i\text{th bit position belonging to the } r\text{th individual of the } k\text{th subpopulation is equal to } t,\ t \in (0,1) \\ 0, & \text{if not,} \end{cases} \quad (1)$$

$$H^k\left[P(t)\right] = -\frac{1}{l}\sum_{i=1}^{l}\left(P_i^k(0)\times\log_2 P_i^k(0) + P_i^k(1)\times\log_2 P_i^k(1)\right) \quad (2)$$

$H^k[P(t)]$ [2] given in Equation (2) is the expression where the entropy value has been calculated by using the rate of 0 and 1s in ith bit position belonging to kth subpopulation.

The expression given in Equation (3) enables to calculate the diversity value (DIV) among all individuals by averaging the diversity values calculated independently in subpopulations. p in Equation (3) and (4) expresses the subpopulation number [16, 20].

$$DIV = \frac{1}{p}\sum_{k=1}^{p} H^k \quad (3)$$

As for the calculation of permeability in PGAs, when it is wanted, diversity analysis is done by examining subpopulations as a whole population. First of all, the rates of 0 and 1s in ith bit position within the subpopulations are determined by applying Equation (1) exactly. As for the next step, the rates of 0 and 1s in ith bit position when all populations are considered, are calculated by applying the following Equation (4) [16, 20].

$$WP_i(t) = \frac{1}{p}\sum_{k=1}^{p} P_i^k(t) \qquad t \in (0,1) \quad (4)$$

In Equation (5) given below, permeability (PRM) has been obtained by using the rates of 0 and 1 in ith bit position obtained by using Equation (4).

$$PRM = -\frac{1}{l}\sum_{i=1}^{l}\left(WP_i(0)\times\log_2 WP_i(0) + WP_i(1)\times\log_2 WP_i(1)\right) \quad (5)$$

3. The effect of changing migration interval to permeability and algorithm performance

The optimization process in GAs is carried out in two phases as exploration and exploitation. The exploration phase is the phase where the different regions of solution space are searched, as for the exploitation phase it is a phase in which a better solution has been obtained by using the genes gathered in the region where there are good solutions [21]. The number of the migration and the iteration step in which it has been done, is important in order to obtain a successful PGA. It is obvious that a great number of migration process will increase the cost, but few migrations will not make enough contribution to the performance. Therefore, it is important to carry out the migration process in the correct steps. In this stage of the study, it is put forward that if the migration process in exploration phase is done densely in PGAs, better results are produced. The reason of this, is that there are more genes in the first iteration steps which will enable variety among subpopulations. As for the following iteration steps, since the genes in subpopulations resemble each other, the variety will decrease, the exchange of the individuals among subpopulations will not make enough contribution to the performance.

In this part of the study, the values of performance and permeability of PGAs have been obtained by taking stable migration interval 80 with increasing and decreasing migration intervals as linear and parabolic and the results have been exhibited. E_PGA [12-16] has been used in this analysis. The results obtained from Rastrigin (f_{Ras}) function [10, 22, 23] given in Equation (6), have been given by using subpopulation size 640, migration rate 7.2%. The trials have been carried out by using 8 subpopulations. In these trials, equal number of migrations have been done at the iteration length. In all of the real environment tests done in this study “send the best, replace with the worst in target subpopulation” method which is used widely as the migration policy [24], has been used.

$$f_{Ras}(x_i|_{i=1\ldots n}) = a\cdot n + \sum_{i=1}^{n}\left[x_i^2 - a\cdot Cos(\omega\cdot x_i)\right]$$
$$a = 10,\ \omega = 2\pi \quad (6)$$
$$-5.12 \le x_i \le 5.12$$

3.1. Linear changing migration interval

In this part of the study, the maximum fitness value and permeability results obtained by migration interval increasing and decreasing linearly are presented. As is seen in Figure 2(a), the approach whose migration interval increases linearly, namely which carries out the majority of the migration at the initial steps, has shown a faster climb and has reached the flexion point earlier. As for the migration interval decreasing linearly, compared with increasing and stable migration interval approach, has converged more slowly and has not produced a good result. When the permeability results in Figure 2(b) are examined, it is seen that increasing migration interval approach has reached lower values faster compared with other models however, at the end of the iteration it has finished at a close value with the decreasing migration interval. The reason of this is that at the increasing migration interval approach migration cannot be carried out towards the end of the iteration and due to this, the result does not change and the obtained permeability value advances horizontally. As for the decreasing migration interval, the migration exhibits a reverse behaviour to increasing approach showing its effect on permeability at the end of the iteration steps.

The real environment tests presented in this part, show that by doing the migration process at initial iterations better results can be reached at early steps. It has been put forward that migration done frequently at initial iterations decrease the permeability quickly. When Figure 2(a) and Figure 2(b) are examined together, it is seen that as the permeability decreases, the performance of the algorithm increases. This result exhibits the effect of the migration process to performance obviously. The reason of increasing migration interval approach to reach good result at early iteration steps is the low possibility of migration individuals to exist in the target subpopulation initially. The sent individual will contribute to the diversity in the target subpopulation, and good results will be able to be obtained with the gathering of good genes. As for the following iterations, since all subpopulations will start to search in the correct regions of the search space, the sent individuals will be similar to the present individuals, and thus a low migration effect will be formed [16].

3.2. Parabolic changing migration interval

As is shown in the previous part, it provides to obtain better results to do the migration process frequently when the iteration has started. In this part of the study, it has been searched how it will perform when the migration process is more frequent initially, compared with the linear change. Thus, trials have been carried out by using migration intervals changing parabolically.

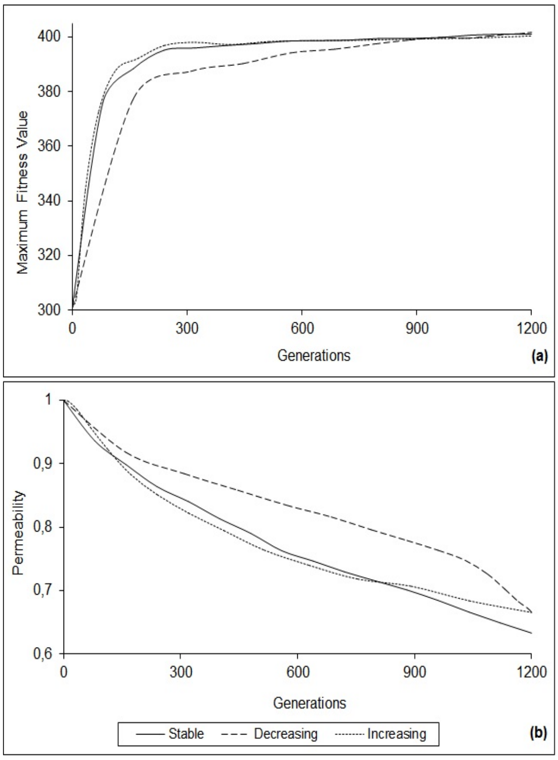


Figure 2. For linear changing intervals (a) maximum fitness value and (b) permeability results

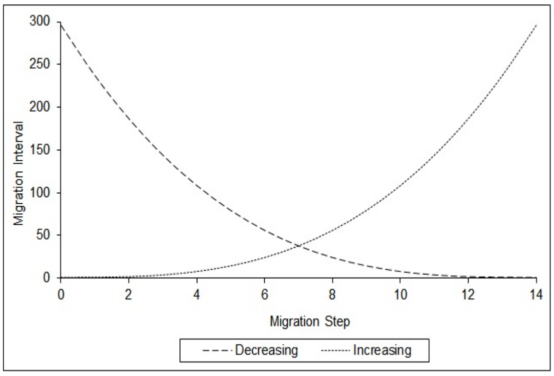


Figure 3. Parabolically changing migration intervals

In Figure 3, migration point-migration interval graphic has been given. By selecting 1200 as total iteration number, the obtained values have been used normalizing with a determined coefficient. As it is understood from Figure 3, in each migration step the process has been done by defining after how many iterations the migration process will be carried out.

The test results for parabolically changing migration interval and the change in permeability values have been given in Figure 4. In parabolically increasing migration interval, initial iterations carry out more migrations. As a result of this, a very fast convergence to flexion point has been obtained. As for parabolically decreasing migration interval, since few migrations are carried out at initial iterations, in all iteration steps it is seen to produce worse results than other methods as is seen in Figure 4(a). As is given in Figure 4(b) the permeability belonging to parabolically increasing migration interval has decreased very quickly at first, but in the latter iteration steps it has not reached low values since it could not carry out migration. As for parabolically decreasing migration interval, the permeability has shown very little change at initial steps in it, but since the migration process increases approaching the last steps, it enters into a fast decrease period [16].

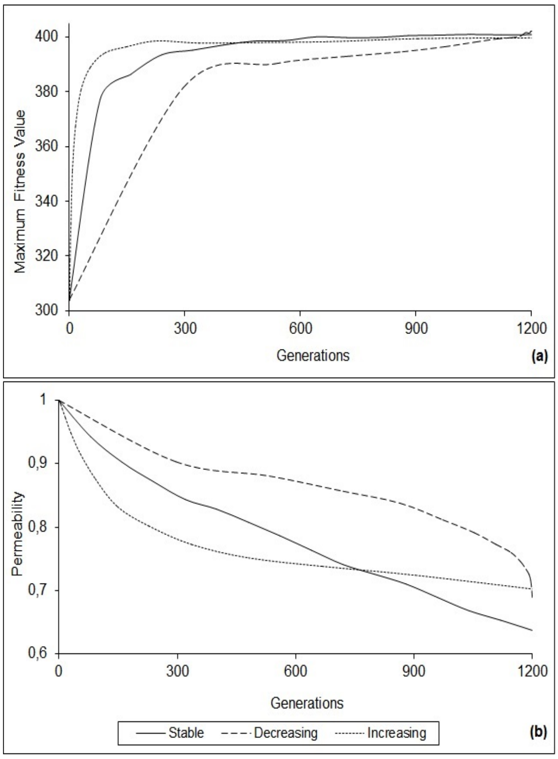


Figure 4. For parabolically changing migration intervals (a) maximum fitness value and (b) permeability results

4. Migration individuals and permeability

New individuals are added to the subpopulations as a result of migration process. The new individuals since they are most probably not the individuals existing in the subpopulation, reform the space where the search is done. For this reason, the analysis of the individuals that are selected to the migration process is important. The permeability results for migration interval 5, 20, 80, 160 and migration method E_PGA, obtained from the whole of all populations and from the individuals that will migrate are given in Figure 5. In the given graphic permeability has been shown as PRM, permeability belonging to the migration individuals as $mPRM$. As for the migration interval, it is the value presented with this illustration. According to the results, the permeability obtained from the whole of the subpopulations takes bigger values than the permeability obtained only for the individuals that will migrate. The reason of this is the selection of the best individuals for migration and thus sending of the individuals resembling each other. According to the results obtained from the real environment tests, the permeability values obtained for migration interval 160 are very high. Since the migration interval is too much, very few migrations have been done, many GA steps have been carried out between two migration points, and the populations have acted completely independent from each other. For these reasons, the individuals joining the subpopulation by migration have not been effective, the permeability obtained by the evaluation of the whole of the populations and the permeability values obtained from the migration individuals have shown a little difference. When the migration interval is 80, the permeability difference between the population and the migration individuals increases. Since the migration interval decreases, migration individuals take place in the target subpopulations more efficiently. When the migration interval is taken as 5 and 20, the difference between the permeability values belonging to the population and the migration individuals increases quickly. The permeability values belonging to the population and the migration individuals similarly get close to each other in the following iteration steps for migration interval 5 and 160. When the migration interval is taken as 5, because of the migration done very frequently, the majority of the target subpopulation is formed of migration individuals and the whole of the population consists of similar individuals. In the following steps, since both population and the migration individuals are formed with the effect of the migration, the permeability difference between them decreases very much. The other point that must be considered in this study is that as the migration interval decreases, the permeability value belonging both to the population and to the migration individuals decreases quickly. This result shows that the permeability term that has been put forward, has analysed the migration effect correctly [16].

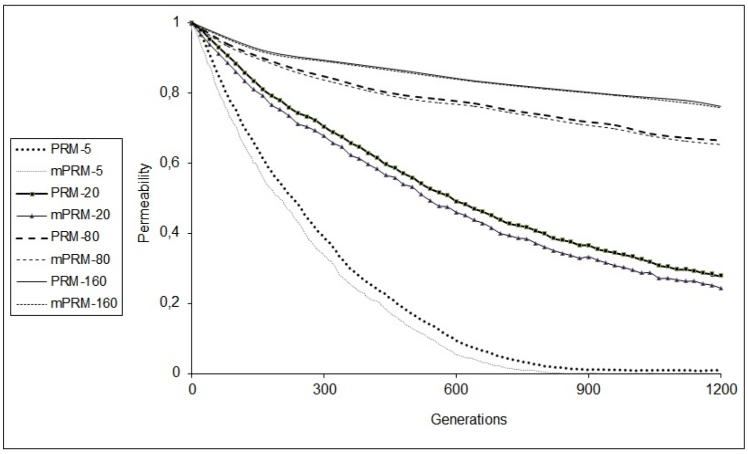


Figure 5. The permeability values for migration interval 5, 20, 80, 160 obtained from the population individuals and the individuals that will migrate

5. Migration individuals and diversity

Migration in PGAs is an important process for increasing the diversity and the performance since it adds new individuals to the subpopulation. Therefore, the analysis of the individuals that will migrate becomes important. In this study, diversity values of migration individuals for different migration methods and migration intervals have been given. In the diversity analysis, the differences of the individuals of all subpopulations that will migrate, have been calculated and by averaging it a final diversity value has been produced. In Figure 6, the diversity values obtained from the individuals that E_PGA, R_PGA and FC_PGA used in the migration respectively for the migration interval 20 and 160 are presented. In Figure 6(a) it is seen that for migration interval 20, the migration individuals that E_PGA has used, have higher diversity values. Especially, in the first half of the iteration steps the diversity formed by E_PGA is much more successful than the other methods. For this reason, it provides a faster convergence to good results. In Figure 6(b), in the trials done for the migration interval 160 while the migration individuals belonging to R_PGA produce more successful diversity values, in the following steps they take close values.

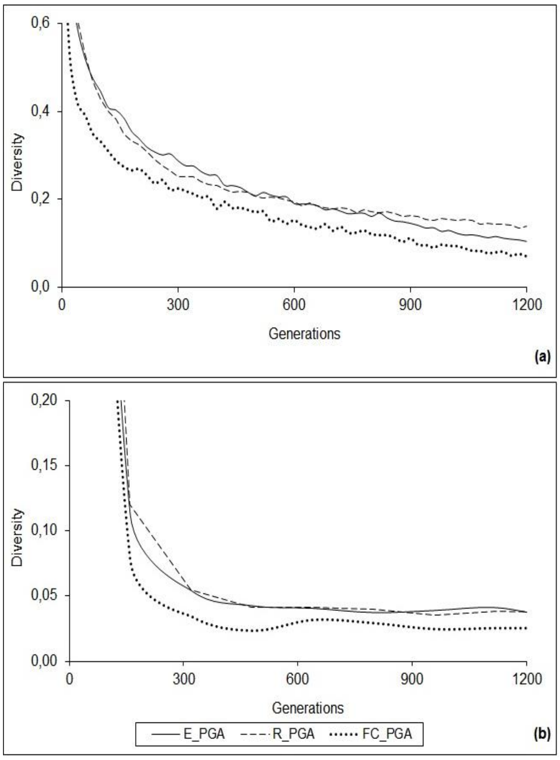


Figure 6. Diversity values of individuals belonging to different migration methods for migration interval (a) 20 and (b) 160

In both graphics it is seen that FC_PGA has produced low diversity results. The reason of this is that while the diversity analysis is being done, 7.2% of the individuals in E_PGA ordered according to the fitness value in the subpopulations is exposed to evaluation and 10% of it in R_PGA, but this analysis is done for 1.25% in FC_PGA. Although E_PGA is exposed to evaluation with lower rate than R_PGA, it produces close diversity values in migration individuals. This result is one of the reasons of E_PGA's production of successful results [16].

6. Results

In order PGAs to produce more successful results, it is important in which phase the migration process will be done. It has been shown with the trials done for linear and parabolic changing migration intervals that more frequently done migration in exploration phase has been effective to obtain good results in early steps. With this study, the contribution of the migration process to the permeability has been exhibited. In order to increase the contribution of the migration to the performance the analysis of the migration individual is also important as well as the phase in which the migration has been done. For the purpose of doing this evaluation, the permeability belonging to the migration individuals and the whole of the subpopulations for different migration intervals has been examined. According to the obtained results, the permeability obtained from the whole of the subpopulations takes bigger values than the permeability obtained only for the individuals that will migrate. It has been shown that as the migration interval increases, the permeability values of migration individuals and the whole populations get close to each other. Besides this, it has been put forward that as the migration interval decreases, both permeability values decrease, namely better permeability has been provided. In addition to this, diversity results belonging to the migration individuals for different migration methods and migration intervals have been given. The differences of the migration individuals have been examined for E_PGA, R_PGA and FC_PGA. Although E_PGA is evaluated for lower migration rate than R_PGA, it has produced close diversity values to R_PGA both at 20 and 160 migration intervals.

References

[1] Rebaudengo, M. and Reorda, M.S., An experimental analysis of effects of migration in parallel genetic algorithms, EWPDP93: IEEE/Euromicro Workshop on Parallel and Distributed Processing, Gran Canaria (E), Gennaio, 232-238 (1992).

[2] Alba E. and Troya, J.M., Analyzing synchronous and asynchronous parallel distributed genetic algorithms, Future Generation Computer Systems 17, 451-465 (2001).

[3] Cantú-Paz, E., Markov chain models of parallel genetic algorithms, IEEE Transactions of Evolutionary Computation, 4(3), 216-226 (2000).

[4] Cantú-Paz, E., Migration policies, selection pressure, and parallel evolutionary algorithms, In Brave, S., Wu, A. (Eds.) Late Breaking Papers at the Genetic and Evolutionary Computation Conference. Orlando, FL (1999).

[5] Cantú-Paz, E., Topologies, migration rates, and multi-population parallel genetic algorithms, GECCO-99:Proceedings of the Genetic and Evolutionary Computation Conference, San Francisco, CA: Morgan Kaufmann, 91-98 (1999).

[6] Nowostawski, M. and Poli, R., Parallel genetic algorithm taxonomy, Proceedings of Third International Conference on Knowledge-based Intelligent Information Engineering Systems KES'99 (1999).

[7] Surry, P.D. and Radcliffe, N.J., RPL2: A language and parallel framework for evolutionary computing, Springer-Verlag LNCS 866, 628-637 (1994).

[8] Alba, E. and Troya, J.M., A survey of parallel distributed genetic algorithms, Complexity 4, 31-52 (1999).

[9] Maeda, Y., Ishita, M.and Li, O., Fuzzy adaptive search method for parallel genetic algorithm with island combination process, International Journal of Approximate Reasoning 41, 59-73 (2006).

[10] Hiroyasu, T., Miki, M. and Negami, M., Distributed genetic algorithms with randomized migration rate, IEEE Proc. of Systems, Man and Cybernetics Conference (SMC'99), 1, 689-694 (1999).

[11] Berntsson, J. and Tang, M., A convergence model for asynchronous parallel genetic algorithms, The 2003 Congress on Evolutionary Computation, CEC'03, 4, 2627-2634 (2003).

[12] Kuvat, G., Adar, N., Canbek, S., Seke, E., Fast converging migration method in Parallel Genetic Algorithms, UMES 2007–National Young Researchers Symposium on Technical Training, Engineering and Education Sciences (in Turkish), 71-74, Kocaeli, Turkey, June (2007).

[13] Kuvat, G., Adar, N., Seke, E., Canbek, S.,

New migration schemes for parallel genetic algorithms on cluster systems, INISTA 2007- International Symposium on Innovations in Intelligent Systems and Applications, 268-272, Istanbul, Turkey (2007).

[14] Kuvat, G., Adar, N., Canbek, S., Seke, E., Analysis the fast converging migration method for various test functions, 12th National Congress and Exhibition on Electrical, Electronics, Computer, Biomedical Engineering (in Turkish), 313-316, Eskişehir, Turkey (2007).

[15] Kuvat, G., Adar, N., Canbek, S., Seke, E., Migration methods, migration rate and parallel genetic algorithms, ASYU 2008 - International Symposium on Innovations in Intelligent Systems and Applications (in Turkish), 138-142, Isparta, Turkey (2008).

[16] Kuvat, G., Migration Methods and Dynamic Determination of Migration Parameters in Parallel Genetic Algorithms, Doctoral Dissertation (in Turkish), Department of Electrical and Electronics Engineering, Eskişehir Osmangazi University (2009).

[17] Lazano, M., Herrera F. and Cano, J.R., Replacement strategies to preserve useful diversity in steady-state genetic algorithms, Information Sciences 178, 4421-4433 (2008).

[18] Denzinger, J. and Kidney, J., Improving migration by diversity, The 2003 Congress on Evolutionary Computation, CEC'03, 1, 700-707 (2003).

[19] Li, Q. and Maeda, Y., Distributed adaptive search method for genetic algorithm controlled by fuzzy reasoning, IEEE International Conference on Fuzzy Systems, 2022-2027 (2008).

[20] Adar, N. and Kuvat, G., Diversity and permeability in parallel genetic algorithms, Journal of the Institute of Science and Technology of Dumlupınar University (in Turkish), Num: 27, 55-66, April (2012)

[21] Goldberg, D.E., Genetic algorithms in search, optimization & machine learning, Addison Wesley Publishing Company, USA (1989).

[22] Alba, E., Luna, F., Nebro A.J. and Troya, J.M., Parallel heterogeneous genetic algorithms for continuous optimization, Parallel Computing 30, 699-719 (2004).

[23] Oh, S., Kim, C.T. and Lee, J., Balancing the selection pressures and migration schemes in parallel genetic algorithms for planning multiple paths, Proceedings of the 2001 IEEE International Conference on Robotics & Automation, Seoul, Korea, May 21-26, 3314-3319 (2001).

[24] Cantú-Paz., E., On the effects of migration on the fitness distribution of parallel evolutionary algorithms, In Workshop on Evolutionary Computation and Parallel Processing at GECCO-2000, 3-6 (2000).

Gültekin Kuvat received the B.Sc. and M.Sc. degrees in 2000 and 2003 from the Department of Electrical-Electronics Engineering from Dumlupınar University, Kütahya, Turkey respectively. He received Ph.D. degree from Eskişehir Osmangazi University in 2009. He is working as Assistant Professor at Computer Engineering department of Balıkesir University. His research interests include genetic algorithms, parallel programming and optimization techniques.

Nihat Adar was born in Konya, Turkey. He received the B.Sc. and M.S.E. degrees in Electrical-Electronics Engineering from Anadolu University, Turkey in 1986 and 1988 and the Ph.D. degree in electrical engineering from Lehigh University, USA in 1994. Currently, he is a full asst. professor at Computer Engineering department of Eskisehir Osmangazi University, Turkey. His research interests are in the application of image processing, parallel programming and genetic algorithms.

An International Journal of Optimization and Control:
Theories & Applications

Volume: 6 Number: 1
January 2016

This page intentionally left blank.

An International Journal of Optimization
and Control: Theories & Applications
Vol.6, No.1, pp.33-51 (2016) © IJOCTA
ISSN:2146-0957 eISSN:2146-5703
DOI:10.11121/ijocta.01.2016.00270
http://www.ijocta.com

Free terminal time optimal control problem for the treatment of HIV infection

Amine Hamdache[a], Smahane Saadi[b], and Ilias Elmouki[c]

Laboratory of Analysis, Modeling and Simulation, Department of Mathematics and Computer Science, Faculty of Sciences Ben M'sik, Hassan II University, Casablanca, Morocco

Email: [a] hamdacheamine@gmail.com, [b] smahanesaadi@gmail.com, [c] i.elmouki@gmail.com

(Received September 1, 2015; in final form January 27, 2016)

Abstract. In this work, an optimal control approach is presented in order to propose an optimal therapy for the treatment HIV infection using a combination of two appropriate treatment strategies. The optimal treatment duration and the optimal medications amount are considered. The main objective of this study is to be able to maximize the benefit based on number of healthy $CD4^+$ T-cells and CTL immune cells and to minimize the infection level and the overall treatment cost while optimizing the duration of therapy. The free terminal time optimal control problem is formulated and the Pontryagin's maximum principle is employed to provide the explicit formulations of the optimal controls. The corresponding optimality system with the additional transversality condition for the terminal time is derived and solved numerically using an adapted iterative method with a Runge-Kutta fourth order scheme and a gradient method routine.

Keywords: Interleukin-2 immunotherapy; Highly active antiretroviral therapy; Pontryagin's maximum principle; Free terminal time optimal tracking control problem, Forward backward sweep method.
AMS Classification: 34H05, 49J15.

1. Introduction

Recent data from the World Health Organization [19] show that approximately 34 million people worldwide are infected with HIV, more than 30 million people died of AIDS-related causes since twenty years. HIV/AIDS is the sixth leading cause of death overall, and the third leading cause of death in poor countries, where an estimated 3.4 million children are infected with HIV/AIDS. Mathematical modeling allows public health officials to compare, plan, implement, evaluate and optimize various programs for the detection, prevention, treatment and control of this disease. Mathematical modeling of infectious diseases at the molecular level is a relatively new science. If epidemiology has a long history, it is only recently that mathematicians and immunologists have begun to work together to create models to predict the evolution of a disease. Since the discovery of human immunodeficiency virus (HIV) and the assertion that it is the cause of the acquired immune deficiency syndrome (AIDS), many scientific studies have focused on the HIV infection [8, 9, 11, 12, 23, 31, 39] and various mathematical models have been developed in order to suggest possible optimal treatment strategies for HIV infection [6, 7, 13, 29, 30, 33, 49, 51].

The HIV infection [19, 36] affects the immune system and particularly the body's natural defenses against disease. If the infection is not treated, serious illnesses can occur. Normally, harmless infections like flu or bronchitis can get worse and become very difficult to

Corresponding Author. Email: hamdacheamine@gmail.com.

treat sometimes involving even death of the infected patients. The human immunodeficiency virus (HIV) approaches the Antigen-presenting cells (APCs) [54], once entered by phagocytosis; it joined the molecular recognition system of the cell. The HIV virus is a retrovirus, the RNA of this virus is converted into DNA inside the $CD4^+$ T-cell. Thus, when the infected $CD4^+$ T-cells begin to multiply for fighting this pathogen, eventually more viruses are produced in parallel.

The scientific research continues for the development of an effective drug therapy hence the interest of optimal control theory [33] which is presented as an indispensable tool for a better understanding of the dynamics of immune system and the evolution of HIV infection in order to propose an appropriate treatment strategy [6, 13, 20, 29, 30].

The HIV infection is usually treated with highly active antiretroviral therapy (HAART) [1, 18] which commonly refers to the combination of antiretroviral treatments struggling against the HIV. The different classes of antiretroviral agents act by disrupting different stages of the HIV replication cycle. This has the effect of reducing the number of virions in the body. The HAART has proven to be very effective limiting significantly the progression of HIV in order to minimize the viral load and to reduce both morbidity and mortality. There are several classes of antiretroviral drugs including: Reverse transcriptase inhibitors [24], HIV fusion inhibitor [48], CCR5 receptor antagonist class [37] and Protease inhibitors [40].

The Interleukin-2 [32, 34] is one of the chemical signals used by immune cells to communicate. This cytokine plays a role in the activation and the proliferation of healthy $CD4^+$ T-cells that are the target cells for HIV virus. The Interleukin-2 is currently used in addition to the antiretroviral therapy (HAART) for increasing the natural immunity of HIV patients. Indeed, the HAART controls the replication of the virus in the blood and IL-2 helps to regenerate more healthy $CD4^+$ T-cells causing effectively the maturation and the proliferation of target immune cells.

In this work, an optimal control approach with free terminal time is proposed for the treatment of HIV infection during an optimal therapeutic period. This approach is based on the introduction of two optimal controls characterizing a combination treatment using both HAART and IL-2 immunotherapy. A free terminal time optimal tracking control problem [3, 27, 28, 41, 46, 47] is formulated by defining a suitable objective function that summarizes the main objectives of the adopted treatment strategy. The corresponding optimality system is expanded to include the necessary condition on free terminal time. However, the Pontryagin maximum principle [17, 44, 45] is used to characterize the formulation of optimal controls. Finally, for the numerical resolution of the optimality system with the additional transversality condition for the terminal time, an adapted iterative method known as the Forward backward sweep method (FBSM) [33, 38]is implemented using a Runge-Kutta [33] fourth order scheme and a gradient method routine [3].

This paper is organized as follows: Section 2 describes the mathematical control model of HIV treatment using a combination treatment of both HAART and IL-2 immunotherapy. The analysis of the free terminal time optimal tracking control problem is also presented in the same section. In section 3, the iterative method is introduced and the numerical simulations are discussed. Finally, the results of this therapeutic approach are explored in the conclusion in section 4.

2. Mathematical model

2.1. Presentation of the treatment model

In this section, a system of ordinary differential equations modeling the treatment of HIV infection is presented. The adopted therapeutic approach is based on the introduction of a treatment strategy using combination of both Highly active antiretroviral therapy (HAART) and IL-2 immunotherapy with tolerated doses. The basic HIV dynamics model was originally discussed by Roy et al. in [50] and the control model providing optimal treatment strategies has been studied in [20].

The HIV dynamics model [50] explores the possible interactions between immune cells and HIV-producing cells in the presence of appropriate therapeutic agents. The obtained biological results have provided a better understanding of dynamics and behavior of the immune system, especially after stimulation of CTL cells that are produced after a maximum proliferation of $CD4^+$ T-cells, which ultimately enables to design the biological reasons that led to such a reaction of the immune system [42, 43, 56, 57].

Note with interest that it has been proven that results from mathematical analysis of the model is fully compatible with clinical and experimental observations. In addition, it was verified analytically that this system is globally asymptotically stable under specific conditions [50].

The main purpose of this work [50] is the development of an adequate mathematical framework which must be consistent with medical experiments and biological observations in order to provide thereafter a set of optimal therapeutic strategies for the treatment of HIV infection. Clinical findings from biological results of treatment strategies that exploit antiretroviral therapy using Lamivudine and Zidovudine show that these treatment strategies enable reducing the viral load (10 *to* 100 %) and allow increasing the concentration of healthy $CD4^+$ T-cells by almost 25 %, provided that the treatment duration must exceed one year [42, 43, 58].

Since this study is interested primarily in the possible biological changes resulting from the introduction of an appropriate treatment in the equilibrium state [50], the mathematical analysis shows that any state variable relating to the dynamics of HIV particles can be omitted [50], which explains the absence of any specific compartment that characterizes the evolution of HIV concentration in the studied model.

However, it should be noted that it was necessary to introduce in this same mathematical model, a new state variable $z(t)$ that describes the behavior and models the dynamics of CTL cells during HIV infection [50]. Three compartments characterizing the different biological populations are defined as follows: $x(t)$ the uninfected $CD4^+$ T-cells, $y(t)$ the infected $CD4^+$ T-cells and $z(t)$ the immune response measured by the rate of the cytotoxic T-cells (CTL). Therefore, the mathematical control model representing the immune system dynamics in presence of appropriate treatments is governed by the following equations:

$$\begin{aligned}
\frac{dx}{dt} &= \lambda + px(1-\frac{x}{T_m}) - dx - (1-u_1)\beta xy \\
&\quad +u_2x, \\
\frac{dy}{dt} &= (1-u_1)\beta xy - ay - lyz, \\
\frac{dz}{dt} &= sy - bz.
\end{aligned} \tag{1}$$

where $X(t) = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix}$ is the state vector and $u(t)=(u_1(t),u_2(t))$ is the control function which describes the medication used for the treatment of HIV infection. For biological specificities characterizing the HIV infection at AIDS stage, the initial values estimations assigned to state variables of the system (1) are measured in units of cells $mm^{-3}day^{-1}$ [50] and verify [16, 26] at $t = 0$:

$$x_0 = 50, \quad y_0 = 50, \quad z_0 = 2. \tag{2}$$

Note that $u_1(t)$ represents the HAART control function which inhibits the viral production in order to reduce the number of infected $CD4^+$ T-cells. It is important to observe that the parameter β represents both rates of infection and viral replication, which explains the choice of introduction of control u_1. The values of $u_1(t)$ vary between 0 if no treatment is used and 1 if totally effective HAART therapy is exploited.

However, $u_2(t)$ represents the IL-2 immunotherapy control function that stimulates immune cells and restores the immune response. The Interleukin-2 is administered to patients with HIV by daily injections following a continuous process for an optimal immunotherapy period where $u_2(t) = \alpha = 0.003$ is the maximum tolerated dose (MTD) [25, 30] producing the desired effect without unacceptable toxicity.

The descriptions of parameters used in the state system (1) are ranged in the table (1). Notice that the experimental observation period is fixed $T = 600$ days [50] and the main objective of this study is to find the optimal duration of treatment T^* which allows to reach all goals set in the optimal control problem.

Note with interest that the scientific works [15, 21] present results of an optimal control approach which aims to introduce a notion of isoperimetric constraint representing the exact total amount of immunotherapy that could be administered to the patient during the treatment period reducing subsequently the total cost of therapy. Furthermore, the biological results observed during the discontinuous administration of immunotherapeutic agents to patients, following a pulse vaccination process, are the subject of a recent study [52] presenting an optimal control problem with a view to suggesting optimal treatment strategies.

Finally, in the presence of an additional initial pathogen concentration, the enhancement of immune response via immunotherapy was adopted using a neighboring optimal control approach in order to restore the optimality conditions of control system [22].

Table 1. The parameters descriptions [50].

Parameters	Descriptions
λ	Production rate of healthy $CD4^+$ T cells
β	Infection rate and viral replication rate
d	Natural mortality rate of healthy $CD4^+$ T cells
p	Maximum proliferation rate of healthy $CD4^+$ T cell
a	Natural mortality rate of infected $CD4^+$ T cells
l	Mortality rate of virus-producing cells by CTL cells
s	Production rate of CTL cells
b	Natural mortality rate of CTL cells
T_m	Number of $CD4^+$ T cells after a maximum proliferation

2.2. The optimal control problem

A free terminal time optimal tracking control problem is formulated in order to propose an optimal therapeutic schedule for an optimal treatment duration. For that purpose, an objective function is defined as follows:

$$\begin{aligned} J(u_1, u_2, T) &= \frac{1}{2}\int_0^T x^2(t) + z^2(t) - y^2(t) \\ &\quad -A_1 u_1^2(t) - A_2 u_2^2(t)dt, \end{aligned} \tag{3}$$

where the positive parameters $A_1 \geq 0$ and $A_2 \geq 0$ balance the terms size and characterize weight factors which are based on benefits and costs of the treatment.

The principal aim of this therapeutic strategy suggested for the treatment of HIV infection is to maximize the benefit based on the count of healthy CD4$^+$ T-cells and CTL immune cells while minimizing the number of infected CD4$^+$ T-cells and the concentration of infectious HIV population allowing thereafter to minimize the harmful side effects and costs based on the percentage effect of HAART and IL-2 immunotherapy given (i.e. u_1^* and u_2^*).

All elements constituting the objective function (3) are quadratic to ensure a better homogeneity of optimal control problem. Note with interest that the optimal duration T^* of the treatment program is also considered. Mathematically, the optimal controls $(u_1^*, u_2^*) \in U$ are sought such that:

$$J(u_1^*, u_2^*, T^*) = \max J(u_1, u_2, T), \tag{4}$$

Over the control set U defined as follows: $U = U_1 \times U_2$ where

$$U_1 = \{u_1 \ continuous, 0 \leq u_1(t) \leq 1, t \in [0, T]\},$$

and

$$U_2 = \{u_2 \ continuous, 0 \leq u_2(t) \leq \alpha, t \in [0, T]\}.$$

Notice that the scientific work [14] dealing with an optimal control problem has outlined the study results of a same objective function $J(u)$, presenting initially a quadratic cost and subsequently a linear cost [14].

The control system (1) is rewritten implicitly as follows:

$$\begin{aligned} X'(t) &= f(t, X(t), u_1(t), u_2(t)), \\ X(0) &= X_0 \quad given. \end{aligned} \tag{5}$$

where $X(t) = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix}$ is the state vector and $u(t) = (u_1(t), u_2(t))$ is the control pair. Thus, the objective function (3) is implicitly defined at control $u(t) = (u_1, u_2)$ as follows:

$$\begin{aligned} J(u_1, u_2, T) &= \int_0^T g(t, X(t), u_1(t), u_2(t))dt \\ &\quad +\theta(T, X(T)), \end{aligned} \tag{6}$$

Consider the optimal control problem:

$$\begin{aligned} &\max \int_0^T g(t, X(t), u_1(t), u_2(t))dt + \theta(T, X(T)), \\ &subject\ to\ X'(t) = f(t, X(t), u_1(t), u_2(t)), \\ &where\ X(0) = X_0\ given, \end{aligned} \tag{7}$$

The corresponding adjoint system is expressed as follows:

$$\begin{aligned} \psi'(t) &= -g_X(t, X(t), u_1(t), u_2(t)) \\ &\quad -\psi f_X(t, X(t), u_1(t), u_2(t)), \\ &\quad where\ \psi(T^*) = \theta_X(T^*, X(T^*)), \\ &\quad and\ 0 \leq u_1(t) \leq 1\ and\ 0 \leq u_2(t) \leq \alpha. \end{aligned} \tag{8}$$

The Pontryagin's Maximum Principle [17, 44, 45] is used to determine the precise formulation of the optimal control pair $u^*(t) = (u_1^*(t), u_2^*(t))$.

In order to characterize the optimal control u^*, the Hamiltonian is defined from the formulation of cost function (3) as follows:

$$H(X,u,\psi) = g(t,X,u_1,u_2) + \psi f(t,X,u_1,u_2)$$

where $\psi(t) = \begin{pmatrix} \psi_1(t) \\ \psi_2(t) \\ \psi_3(t) \end{pmatrix}$ is the adjoint variable vector.

Explicitly:

$$\begin{aligned} H(X,u,\psi) &= \frac{1}{2} \times (x^2 + z^2 - y^2 \\ &-A_1 u_1^2 - A_2 u_2^2) \\ &+\psi_1[\lambda + px(1 - \frac{x}{T_m}) \\ &-dx - (1-u_1)\beta xy + u_2 x] \\ &+\psi_2[(1-u_1)\beta xy - ay - lyz] \\ &+\psi_3[sy - bz]. \end{aligned} \tag{9}$$

The existence of an optimal control solution is satisfied using a classical result of existence that was developed by Fleming in [17]. Indeed, the following properties have to be checked:

(1) The class of all initial conditions with controls u_1 and u_2 in the admissible control set $U = U_1 \times U_2$ along with state system equations (1) is not empty;

(2) The admissible control set U is convex and closed;

(3) The right-hand side of the state system is continuous, is bounded above by a sum of the bounded control and the state, and can be expressed as a linear function of controls u_1 and u_2 with coefficients depending on time and state.

(4) The integrand $g(t,X,u_1,u_2)$ of the objective functional $J(u,T)$ is concave on U;

(5) There exist positive constants $b_1,\ b_2 > 0$ and $\beta > 1$ such that the integrand of the objective functional $J(u,T)$ is bounded below by $g(t,X,u_1,u_2) \leq b_2 - b_1(|u_1|^2 + |u_2|^2)^{\frac{\beta}{2}}$;

(6) The payoff function $\theta(T,X(T))$ in the objective functional $J(u)$ is continuous at the terminal time T.

Proof. Since the system has bounded coefficients and any state system solution is bounded on a finite interval $[0,T]$ [5], a classical result established by Lukes [35] is used to prove the existence of solutions for the state system (1). The admissible control set U is convex and closed by definition.

The system (1) is bilinear in controls u_1 and u_2 and each right-hand side of this state system (1) is continuous since each term has a nonzero denominator and can be written as a linear function of controls u_1 and u_2 with coefficients depending on time and state.

Moreover, the fact that state variables x, y, z and controls u_1 and u_2 are bounded on time interval $[0,T]$ involves the rest of the third property. In order to verify that the integrand $g(t,X,u_1,u_2)$ in the objective functional (3) is concave on U, the following condition should be verified:

$$\begin{aligned} h(t,X,(1-\lambda)u_i + \lambda v_i) &\leq (1-\lambda)h(t,X,u_i) \\ &+\lambda h(t,X,v_i) \end{aligned} \tag{10}$$

where

$$\begin{aligned} h(t,X(t),u_i(t)) &= -g(t,X(t),u_i(t)) \\ &= \frac{1}{2} \times (-x^2(t) - z^2(t) + y^2(t) \\ &+A_i u_i^2(t)), \end{aligned} \tag{11}$$

This inequality (10) is rewritten in the following form:

$$\begin{aligned} \mathcal{A} &= h(t,X,(1-\lambda)u_i + \lambda v_i) - (1-\lambda)h(t,X,u_i) \\ &-\lambda h(t,X,v_i) \leq 0 \end{aligned} \tag{12}$$

where $\lambda \in [0,1]$, $u_i, v_i \geq 0$ and with $i = 1,2$.

$$\mathcal{A} = \frac{1}{2}A_i((1-\lambda)^2u_i^2 - (1-\lambda)u_i^2 + \lambda^2v_i^2 - \lambda v_i^2 + 2\lambda(1-\lambda)u_iv_i)$$
$$= \frac{1}{2}A_i(u_i^2((1-\lambda)^2 - (1-\lambda)) + v_i^2(\lambda^2-\lambda) + 2\lambda(1-\lambda)u_iv_i)$$
$$= \frac{1}{2}A_i(u_i^2(\lambda^2-\lambda) + v_i^2(\lambda^2-\lambda) - 2(\lambda^2-\lambda)u_iv_i)$$
$$= \frac{1}{2}A_i(\lambda^2-\lambda)(u_i^2+v_i^2-2u_iv_i)$$
$$= \frac{1}{2}A_i(\lambda^2-\lambda)(u_i-v_i)^2 \leq 0 \quad (13)$$

Since $\lambda \in [0,1]$, this implies that $\lambda \geq \lambda^2$. Thus, the inequality (10) is verified which proves that the integrand $g(t, X, u_1, u_2)$ is concave. Thus, since h is convex on $U \Longrightarrow g$ is concave on U. In addition, notice that there exists positive constants $b_1,\ b_2 > 0$ and $\beta > 1$ satisfying:

$$g(t, X(t), u_1(t), u_2(t)) \leq x^2(t) + z^2(t) - A_1u_1^2(t) - A_2u_2^2(t) \leq b_2 - b_1(|u_1|^2 + |u_2|^2)^{\frac{\beta}{2}} \quad (14)$$

where the positive constant b_2 depends on the upper bounds on x and z and by analogy it would be appropriate to set $b_1 = inf(A_1, A_2)$ and $\beta = 2$.

If the control pair $u(t) = (u_1(t), u_2(t))$ and the corresponding state $X(t)$ are optimal, there exists an adjoint vector $\psi(t)$ such that the Hamiltonian $H(t, X, u_1, u_2, \psi)$ reaches its maximum on the set U at u^*,T^*. It ensues the following theorem:

Theorem 1. *Given an optimal control vector $u^* = (u_1^*, u_2^*)$, an optimal terminal time T^*, and solutions of corresponding state system (1), there exists an adjoint vector $\psi = [\psi_1,\ \psi_2,\ \psi_3]$ satisfying the following equations:*

$$\begin{aligned} \psi_1'(t) &= -x + \psi_1(\frac{2px^*}{T_m} + d - p - u_2^*) + \beta y^*(1-u_1^*)(\psi_1-\psi_2), \\ \psi_2'(t) &= y + \beta x^*(1-u_1^*)(\psi_1-\psi_2) + \psi_2(a + lz^*) - \psi_3 s, \\ \psi_3'(t) &= -z + \psi_2 l y^* + \psi_3 b. \end{aligned} \quad (15)$$

with final conditions

$$\psi_j(T) = 0,\ j = 1, 2, 3.$$

The transversality condition for the terminal time is defined as follows:

$$\frac{1}{2} \times (x^2(t) + z^2(t) - y^2(t) - A_1u_1^2(t) - A_2u_2^2(t)) = 0 \ at \ t = T^* \quad (16)$$

Further, u_1^* and u_2^* are represented by:

$$u_1^*(t) = \min(1, \max(0, \frac{\beta x^*(t)y^*(t)(\psi_1(t)-\psi_2(t))}{A_1})), \quad (17)$$

and

$$u_2^*(t) = \min(\alpha, \max(0, \frac{x^*(t)\psi_1(t)}{A_2})). \quad (18)$$

Proof. Due to the existence of an optimal couple (X^*, u^*) which maximizes the objective function J subject to the state system (1), the adjoint equations can be obtained using Pontryagin's maximum principle [17, 44, 45] such that:

$$\psi_1' = -\frac{\partial H}{\partial x}, \quad \psi_2' = -\frac{\partial H}{\partial y}, \quad \psi_3' = -\frac{\partial H}{\partial z}. \quad (19)$$

The terminal time T variable of the objective function J (3) should be exploited to provide all necessary information concerning the optimal final time T^* [33]. For this, consider a real number $\sigma \geq -T^*$ in order that $T^* + \sigma$ is an admissible final time and $T^* + \sigma \in \mathbb{R}^+$.

Note that the corresponding state X^* and the control function u^* are considered on an interval larger than $[0, T^*]$ [33]. Suppose that u^* is left-continuous at T^*, then set $u^*(t) = u^*(T^*)$ for all $t > T^*$ in order that u^* is continuous at T^*. Now, x^* and z^* are also defined for $t > T^*$. As $J(u_1, u_2, T)$ reaches its maximum at $u^* = (u_1^*, u_2^*)$,T^*, the following equality is established [33]:

$$0 = \lim_{\sigma \to 0} \frac{J(u^*, T^* + \sigma) - J(u^*, T^*)}{\sigma}, \tag{20}$$

Hence,

$$\begin{aligned} 0 &= \lim_{\sigma \to 0} [\int_0^{T^*+\sigma} g(t, X^*(t), u^*(t))dt \\ &+\theta(T^* + \sigma, X^*(T^* + \sigma)) \\ &- \int_0^{T^*} g(t, X^*(t), u^*(t))dt - \theta(T^*, X^*(T^*))]. \end{aligned} \tag{21}$$

$$\begin{aligned} 0 &= \lim_{\sigma \to 0} [\int_0^{T^*+\sigma} g(t, X^*(t), u^*(t))dt \\ &+\theta(T^* + \sigma, X^*(T^* + \sigma)) \\ &- \int_0^{T^*} g(t, X^*(t), u^*(t))dt - \theta(T^*, X^*(T^*))] \\ &= \lim_{\sigma \to 0} \int_{T^*}^{T^*+\sigma} g(t, X^*(t), u^*(t))dt \\ &+ \frac{\theta(T^* + \sigma, X^*(T^* + \sigma)) - \theta(T^*, X^*(T^*))}{\sigma} \\ &= g(T^*, X^*(T^*), u^*(T^*)) + \theta_t(T^*, X^*(T^*)) \\ &+\theta_X(T^*, X^*(T^*))\frac{X^*}{dt}(T^*) \\ &= g(T^*, X^*(T^*), u^*(T^*)) \\ &+\psi(T^*)f(T^*, X^*(T^*), u^*(T^*)) \\ &+\theta_t(T^*, X^*(T^*)) \\ &= H(T^*, X^*(T^*), u^*(T^*), \psi(T^*)) \\ &+\theta_t(T^*, X^*(T^*)). \end{aligned} \tag{22}$$

Taking into account that $\theta_t(T^*, X^*(T^*)) = 0$ and $\psi_j(T) = 0 \; for \; j = 1, 2, 3$. Thus, the transversality condition (16) for the terminal time is obtained.

Since controls $u_1(t)$ and $u_2(t)$ are bounded, the optimal controls u_1^* and u_2^* can be solve from the following optimality conditions:

$$\frac{\partial L}{\partial u_1} = 0 \; and \; \frac{\partial L}{\partial u_2} = 0.$$

In order to find the characterization of optimal controls (17) and (18), the Lagrangian L is used and defined as follows:

$$L = H + \omega_{11}(1 - u_1) + \omega_{12}u_1 + \omega_{21}(\alpha - u_2) + \omega_{22}u_2 \tag{23}$$

where $\omega_{11}, \; \omega_{12}, \; \omega_{21}, \; \omega_{22} \geqslant 0$ are the penalty multipliers which ensure the boundedness of controls $u_1(t)$ and $u_2(t)$ and satisfy the two following conditions [5, 27]:

$$\begin{aligned} \omega_{11}(1 - u_1^*) &= \omega_{12}u_1^* = 0 \; at \; u_1 = u_1^*, \\ \omega_{21}(\alpha - u_2^*) &= \omega_{22}u_2^* = 0 \; at \; u_2 = u_2^*. \end{aligned} \tag{24}$$

The maximization problem (4) is redefined as follows:

$$\begin{aligned} &L(T^*, X^*, u_1^*, u_2^*, \psi, \omega_{ij}) \\ &= \max L(T, X^*, u_1, u_2, \psi, \omega_{ij}) \end{aligned} \tag{25}$$

Differentiating the Lagrangian L with respect to u_1 on the set $U_1 : \{t \mid 0 \leq u_1(t) \leq 1\}$ allows to obtain the following optimality equation:

$$\frac{dL}{du_1} = -A_1u_1 + \beta xy(\psi_1 - \psi_2) - \omega_{11} + \omega_{12} = 0$$
at $u_1 = u_1^*$.

Thus, the control is expressed:

$$u_1^*(t) = \frac{\beta x^*(t)y^*(t)(\psi_1(t) - \psi_2(t)) - \omega_{11} + \omega_{12}}{A_1}$$

According to the conditions (24), three cases are distinguished:

⋆ if $0 < u_1^*(t) < 1$ then $w_{11} = w_{12} = 0$. Therefore, the control is expressed as follows:

$$u_1^*(t) = \frac{\beta x^*(t)y^*(t)(\psi_1(t) - \psi_2(t))}{A_1}$$

⋆ if $u_1^*(t) = 0$ then $w_{11} = 0$. Therefore, the control is expressed as follows:

$$u_1^* = \frac{\beta xy(\psi_1 - \psi_2) + \omega_{12}}{A_1} = 0$$
$$\Longrightarrow \omega_{12} = A_1u_1 - \beta xy(\psi_1 - \psi_2).$$

Knowing that $\omega_{12}(t) \geq 0$ and $A_1 > 0$, the control is expressed as follows:

$$u_1^* = 0 \leq \beta xy(\psi_1 - \psi_2) \leq \frac{\beta xy(\psi_1 - \psi_2)}{A_1}$$

⋆ if $u_1^*(t) = 1$ and $w_{12}(t) = 0$ then the control is expressed as follows:

$$u_1^* = \frac{\beta xy(\psi_1 - \psi_2) - \omega_{11}}{A_1} = 1$$

$\Longrightarrow \omega_{11}(t) = -A_1 u_1 + \beta xy(\psi_1 - \psi_2)$.

Given that $w_{11}(t) \geq 0$ and $A_1 > 0$, the control is expressed as follows:

$$u_1^* = 1 \geq \frac{\beta xy(\psi_1 - \psi_2)}{A_1}$$

Combining these three results, the optimal control $u_1^*(t)$ is characterized as follows:

$$\begin{aligned} u_1^*(t) &= \frac{\beta xy(\psi_1 - \psi_2)}{A_1} \quad if\ 0 < \frac{\beta xy(\psi_1 - \psi_2)}{A_1} < 1, \\ u_1^*(t) &= 0 \quad\quad\quad\quad if\ \frac{\beta xy(\psi_1 - \psi_2)}{A_1} \leq 0, \\ u_1^*(t) &= 1 \quad\quad\quad\quad if\ \frac{\beta xy(\psi_1 - \psi_2)}{A_1} \geq 1. \end{aligned} \tag{26}$$

Thus, the optimal control $u_1^*(t)$ is formulated as follows:

$$u_1^*(t) = min(max(0, \frac{\beta x^*(t) y^*(t)(\psi_1(t) - \psi_2(t))}{A_1}, 1))$$

Differentiating the Lagrangian L with respect to u_2 on the set $U_2 : \{t \mid 0 \leq u_2(t) \leq \alpha\}$ allows to obtain the following optimality equation:

$$\frac{dL}{du_2} = -A_2 u_2 + \psi_1 x - \omega_{21} + \omega_{22} = 0 \text{ at } u_2 = u_2^*.$$

Thus, the control is expressed as follows:

$$u_2^*(t) = \frac{\psi_1 x - \omega_{21} + \omega_{22}}{A_2}$$

According to the conditions (24), three cases are distinguished:

⋆ if $0 < u_2^*(t) < \alpha$ then $w_{11} = w_{12} = 0$. Therefore, the control is expressed as follows:

$$u_2^*(t) = \frac{\psi_1(t) x^*(t)}{A_2}$$

⋆ if $u_2^*(t) = 0$ then $w_{11} = 0$. Therefore, the control is expressed as follows:

$$u_2^* = 0 = \frac{\psi_1 x + \omega_{22}}{A_2}$$

$\Longrightarrow \omega_{22} = A_2 u_2 - \psi_1 x$.

Knowing that $\omega_{22}(t) \geq 0$ and $A_2 > 0$, the control is expressed as follows:

$$u_2^* = 0 \leq \frac{\psi_1(t) x}{A_2}$$

⋆ if $u_2^*(t) = \alpha$ and $w_{12}(t) = 0$, then the control is expressed as follows:

$$u_2^* = \alpha = \frac{\psi_1 x - \omega_{21}}{A_2}$$

$\Longrightarrow \omega_{21}(t) = \psi_1 x - A_2 u_2$.

Given that $w_{21}(t) \geq 0$ and $A_2 > 0$, the control is expressed as follows:

$$u_2^* = \alpha \geq \frac{\psi_1(t) x}{A_2}$$

Combining these three results, the optimal control $u_2^*(t)$ is characterized as follows:

$$\begin{aligned} u_2^*(t) &= \frac{\psi_1 x}{A_2} \quad if\ 0 < \frac{\psi_1 x}{A_2} < \alpha, \\ u_2^*(t) &= 0 \quad\quad if\ \frac{\psi_1 x}{A_2} \leq 0, \\ u_2^*(t) &= \alpha \quad\quad if\ \frac{\psi_1 x}{A_2} \geq \alpha. \end{aligned} \tag{27}$$

Thus, the optimal control $u_2^*(t)$ is formulated as follows:

$$u_2^*(t) = min(max(0, \frac{\psi_1 x^*(t)}{A_2}, \alpha))$$

3. Numerical simulations

3.1. Model parameters

The main purpose of the theoretical analysis developed by Roy et al. [50] was intended to explore the equilibrium of dynamical system and to study the various aspects of the stability of solutions in order to determine the threshold values of studied model parameters for which the disease can be controlled.

In this sense, any optimal control approach elaborated for the studied dynamical model and which aims to provide treatment strategies for the HIV infection, should absolutely explore the equilibria and consider the theoretical results of stability analysis [50], which implies respecting the established conditions and constraints that characterize the different model parameters, thus allowing to define specific parametric regions where the equilibrium is locally or globally stable.

Indeed, the standard values of parameters [50] have been chosen in the context of this theoretical analysis in order to observe the particular dynamical behavior of state variables x, y and z with the threshold values that enable controlling the disease from a well-defined initial state ($x(0) = 50$, $y(0) = 50$, and $z(0) = 2$). Note with interest that analytical study and numerical resolution of the system have been developed entirely on the basis of these model parameters set to their standard values [50].

Subsequently, studying the influence of model parameters, allowed to observe the impact of these parameters on the dynamical behavior of state variables. The stability analysis shows that for the positive equilibrium of the dynamical system, the disease can only be controlled if the parameter p (Proliferation constant rate of $CD4^+$ T-cells) is greater than the parameter d (Death rate of Uninfected $CD4^+$ T-cells) [50].

Moreover, it is observed that if the parameter p increases, we note a considerable growth in the concentration of immune cells (Healthy $CD4^+$ T-cells and CTL immune cells) and we notice a significant decrease of oscillations characterizing the evolution of the state variables that manage to converge more quickly to their respective equilibrium states [50].

The study also allows to note that increasing of the parameter value β ($0,0008$ to $0,01$) denoting the rate of infection, causes a development of the HIV infection followed by a rise in the number of CTLs and a substantial decline in the concentration of healthy $CD4^+$ T-cells [50].

However, we note that any increase of parameters k ($0,001$ to $0,005$) and s ($0,01$ to $0,05$) implies a significant growth in the count of active immune cells (Healthy $CD4^+$ T-cells and CTL immune cells) and an important decrease in the level of virus producing cells [50]. In addition, the theoretical analysis enables to determine a specific stability criteria of the equilibrium in the parametric space of β, p and k [50].

Finally, it is clear that the possibility of proposing a control approach for the treatment of HIV infection requires the exploitation of numerical results obtained in this analytical study.

Therefore, since the main purpose of this study is to use optimal control theory in the context of a free terminal time optimal tracking control problem which should be coherent and compatible with the parametric conditions obtained analytically in [50], in order to suggest an optimal strategy for the treatment of HIV infection during an optimal therapeutic period, the basic parameters set to their standard values and found in [4, 12, 42, 50, 56] are kept and it is stated that the stability properties [50] of the state system (1) are stored for these parameters which are rearranged in the table (2).

Table 2. The standard parameter values [50].

Parameters	Values
λ	$10\ mm^{-3}day^{-1}$ [12, 42]
β	$0.002\ mm^{-3}day^{-1}$ [4]
d	$0.01\ day^{-1}$ [42]
p	$0.03\ day^{-1}$ [42, 56]
a	$0.024\ day^{-1}$ [12]
l	$0.001\ mm^{-3}day^{-1}$ [4]
s	$0.2\ day^{-1}$ [4]
b	$0.02\ day^{-1}$ [4]
T_m	$1500\ mm^{-3}$ [42, 56]

3.2. Numerical method

Various numerical methods are used to solve the optimality system and find an optimal solution for controls u_1 and u_2 [10, 55]. In this work, an iterative method known as the Forward-Backward sweep method (FBSM) [33, 38] is developed using a Runge-Kutta [33] fourth order scheme in order to characterize numerical solutions for the optimality system resulting from the studied free terminal time optimal tracking control problem (4).

The general principle of this numerical method is that from an initial guess for the control variables u_1 and u_2 and terminal time T, the state system (1) with initial conditions is solved forward in time and subsequently the adjoint system (15) with terminal conditions is solved backward in time. Taking into account the nature of the optimal control problem with free terminal time (4), a specific numerical technique is considered for the numerical resolution of the optimality system.

Indeed, an adapted iterative Forward backward sweep method is extended using a gradient algorithm with the Formulae of sensitivity to change of end-time [3] view to finding the optimal solutions u_1^*,u_2^* and T^* while considering the transversality condition for the terminal time (16).

This numerical resolution process comprises a number of numerical computation techniques summarized in the algorithm given below. Here the vector approximations for state variable $\vec{X}$=(X^1,...,X^{N+1}) and adjoint variable $\vec{\psi}$=(ψ^1,...,ψ^{N+1}).

Algorithm

Step 0:
. Make an initial guess for the terminal time T;
. Make an initial guess for the controls $\vec{u_1}$ and $\vec{u_2}$ over the time interval;

Step 1:
. Solve the state system (1) with initial conditions X^1=$X(0)$ forward in time using the stored values for the controls $\vec{u_1}$ and $\vec{u_2}$;

Step 2:
. Solve the adjoint system (15) with terminal conditions ψ^{N+1}=$\psi(T)$ backward in time using the stored values for the controls $\vec{u_1}$ and $\vec{u_2}$ and the state variable $\vec{X}$;

Step 3:
. Update the controls $\vec{u_1}$ and $\vec{u_2}$ using by the Forward backward sweep method;
. Update the terminal time by the gradient method defined as follows:

$$\begin{aligned} &T_{i+1} \\ &= T_i - h[H(T_i, X_i(T_i), \psi_i(T_i), u_{1_i}(T_i), u_{2_i}(T_i)) \\ &-\nabla J(T_i, X_i(T_i))], \end{aligned} \tag{28}$$

for $i = 1, .., n$ with h is a small positive constant,

. Test the convergence: If the difference of values of these variables in this iteration and the last iteration is sufficiently small, output the obtained current values as solutions. If the difference is not considerably small, go to Step 1.

3.3. Numerical results

The estimates of initial values assigned to the state variables at time $t = 0$ (2) and specifically the number of healthy CD4$^+$ T-cells which is far below than 200 cell units, indicate that the disease has reached the AIDS stage [16].

This biological phase of HIV infection is generally characterized by the progressive weakening of the immune system and the occurrence of various anomalies and opportunistic diseases [26]. Without therapeutic intervention, the state variables converge logically to their respective equilibrium points [20, 50]. The concentration of healthy CD4$^+$ T-cells after an observation period which lasts 600 days shows that the immune system is weak and defective and the general condition of the HIV patient is clearly deteriorated [20].

However, introducing a treatment strategy using both highly active antiretroviral therapy and IL-2 immunotherapy provides biological results which are satisfactory and especially promising (Figures 1, 2 and 3). Indeed, at the end of an observation therapeutic period of 600 days, the treatment effectively helps to maximize the number of healthy CD4$^+$ T-cells which reached 1400 cell units (Figure 1).

Similarly, the infection level has gradually decreased and the number of infected CD4$^+$ T-cells has achieved values lower than 5 units towards the end of the therapeutic period (Figure 3). Henceforth, the immune system makes full use of its defensive function and the immune response reacts actively to the evolution of the HIV infection: Any increase in the concentration of infected cells is followed immediately by a considerable proliferation of CTL immune cells (Figure 3).

However, it was observed that the count of immune cells which are stimulated for the immune response has naturally decreased after the minimization of the viral load thereby reducing the side effects resulting from a prolonged maximization of the immune cells level (Figure 3).

Considering the shape and the behavior of the optimal controls u_1^* and u_2^* (Figures 4 and 5) during the optimal duration of treatment, it is noted with interest that the therapeutic process has adopted an appropriate treatment approach which takes into consideration the progression of HIV infection and the development of infected cells in order to achieve the objectives of the optimal control problem (4).

Compared to the initial observation period lasting 600 days, this free terminal time optimal tracking control problem (4) situated within the framework of a treatment strategy of HIV infection, has allowed to find an optimal terminal time T^* (Figure 6) satisfying the transversality condition (16) and has enabled to define an optimal treatment duration of 512 days, ensuring therefore a more consequent reduction of the overall cost of treatment and a minimization of the side effects resulting from the adopted therapy.

In fact, from the optimal terminal time T^*, it is observed that even after stopping the treatment process, the optimality conditions remain satisfied (Figures 1, 2 and 3), allowing subsequently to generate an important increase in the level of healthy $CD4^+$ T-cells (Figure 1) and a large reduction of infection level (Figure 2). Finally, for testing the effectiveness of the treatment approach which is adopted in this study, a new terminal time is fixed $T = 500$ days with the aim of finding a new optimal terminal time T^* able to further reduce the treatment duration and thereby allowing to further minimize the overall cost of treatment.

However, it is noticed that the obtained biological results (Figures 7, 8 and 9) show that this new therapeutic approach has not achieved the key objectives defined in the optimal control problem (4). Although the number of healthy $CD4^+$ T-cells is significantly important during the treatment period (Figure 7), the gradual reduction of the controls concentration in the last 30 days of the treatment period generates a substantial increase in the level of infected $CD4^+$ T-cells (Figure 8).

Moreover, despite a maximum stimulation of CTL immune cells (Figure 9), the HIV infection remains unstable and the concentration of infected $CD4^+$ T-cells increases abruptly just before the end of treatment, which explains the inability of the immune system that fails to limit the HIV infection progression and to restrict the action of the HIV particles. These recent observations prove the efficiency of the initial optimal control approach with free terminal time T^* for the treatment of HIV infection in an optimal duration which lasts 512 days.

Using standard parameter values given in table (2) [50], the behavior of the state variables has been observed in the presence of the natural immune response and without the intervention of any specific therapy. Indeed, the state variables converge respectively towards their equilibrium states [20, 50].

However, from a biological point of view, despite the weak growth in the level of immune cells and the limited reduction in viral load, this equilibrium state fails to reach the expected biological objectives since the concentration of healthy $CD4^+$ T-cells is still low and the general condition of the HIV patient remains critical [26]. By exploiting the different results of the study conducted by Roy et al. [50], the interest of adopting an appropriate therapeutic strategy for treatment of the HIV infection is well confirmed, thereby justifying the introduction of the control u_1 that limits the growth of the parameter β in order to reduce the level of infection and viral load and the control u_2 that stimulates the proliferation of active immune cells. Finally it is important to note that the effectiveness of drug used in the treatment process is assumed to be fully controlled by drug dose level.

The continuous character [6, 13, 29, 30, 33] of optimal solutions u_1^* and u_2^* (Figures 4 and 5) is essentially acquired from the definition of the admissible control set U. This continuity aspect characterizing the controls u_1 and u_2 permits theoretically to find optimal solutions that achieve the objectives set in the optimal control problem, thus enabling to provide a general profile of therapeutic strategies to be adopted with a view to treating the HIV infection (Figures 4 and 5).

For clinical tests and trials, the treatment strategies relating to the optimal controls u_1^* and u_2^* that are represented by continuous functions would be difficult to implement from a practical point of view. As part of an optimal control problem presenting an objective function with linear control, the optimal control may just take the extreme constant values (The solution is of the bang-bang type) [33] provided that it is possible to prove the absence of singular arcs [33].

However, the problem studied in this work defines a quadratic objective function in order to ensure more consistency to the optimal control problem by minimizing the contributions of small variations [53]. Hence the interest to provide functions approaching the optimal solutions and which are much easier to prescribe practically in the context of the adopted treatment strategy.

At first, the curves illustrating the evolution of optimal controls u_1^* and u_2^* were fitted [2] with the aim of reducing the irregularities and the singularities characterizing these curves (Figures 10,11,20 and 21) and thus enabling to mitigate the observed disturbances. Then, on the basis of obtained results, piecewise constant functions are defined to characterize the control functions u_1 and u_2 (Figures 10,11,20 and 21).

The impact of applying these new treatment regimens $u_1^{'}$ and $u_2^{'}$ (Figures 10 and 11) on the behavior of the state variables is observed by illustrating graphically the evolution of the variables x, y and z using the two types of control functions (Figures 12,13 and 14). Over an observation period of 600 days, although the results of the optimal treatment strategy are better than those obtained with piecewise constant control functions, the treatments $u_1^{'}$ and $u_2^{'}$ (Figures 10 and 11) allow to obtain satisfactory biological results and manage to reach all objectives of optimal control problem.

Indeed, using constant control functions modeling the adopted treatment, the healthy CD4$^+$ T-cells follow an increasing evolution. The proliferation peak occurs around the 600th day by reaching the count of 1320 cell units (Figure 12). Similarly, the immune response is active allowing the stimulation of immune cells thus generating a significant proliferation of the CTL immune cells (Figure 14) when the viral load is growing. The infection is reduced considerably and the concentration of virus producing cells reaches very low levels towards the end of the observation period (Figure 13).

A number of scientific works [29, 30, 33] show that the administration of a treatment strategy during early stages of the HIV infection is more beneficial for the therapeutic process. For example, the immunotherapy adopted in an earlier stage increases the levels of healthy CD4$^+$ T-cells [29, 30].

In this respect, we use numerical data suggested in the scientific work developed by Butler et al. [6] and which characterizes a new initial state corresponding to a clinical case presenting an infection appeared since only 74 days ($x(0) = 494.3$ and $y(0) = 0.04$) [6]. During this stage of the disease which is known as the Acute HIV syndrome that precedes the stage of clinical latency, we note a wide spread of the virus particles in the body and a replication of HIV in lymphoid organs.

Indeed, towards the 20th day of treatment, a severe increase in the concentration of infected CD4$^+$ T-cells is observed due to the biological resistance of these virus producing cells to the introduction of therapeutic agents involved in the treatment process (Figure 16). The stimulation of cells involved in immune response (Figure 17) and the action of optimal controls (Figures 18 and 19), allow to reduce the viral load in the short term from the 23th day (Figure 16). The level of infection is stabilizing from the 200th day.

Furthermore, the count of infected CD4$^+$ T-cells reached values below 10 cell units from the 420th day of treatment (optimal final time T^* (Figure 22)) and it eventually reached values below 2 cell units towards the end of the observation period (Figure 16). In addition, we note a gradual growth in the number of healthy CD4$^+$ T-cells from the 30th day, thus enabling to reach a count of 1492 cell units by the end of the clinical observation period (Figure 15).

Finally, note with interest that the introduction of an appropriate treatment strategy at an early stage of HIV infection has achieved all the objectives set in the optimal control problem thereby allowing to further stimulating the immune cell proliferation (Figures 15 and 17) and reducing the viral load (Figure 16) while minimizing the optimal treatment duration (Figure 22).

Compared to the first studied case ($x(0) = 50$, $y(0) = 50$, $z(0) = 2$), the optimal treatment duration was considerably minimized (T^*=420 days) (Figure 22) and the concentration of controls used in the therapeutic process has decreased significantly (Figures 18,19,20 and 21).

The results obtained have helped to reduce side effects and overall costs of the adopted treatment leading to a marked improvement in the quality of life of HIV patients.

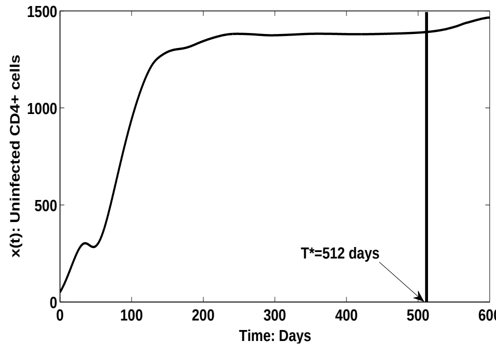


Figure 1. The state variable x with $x(0)$=50 units $mm^{-3}day^{-1}$, initial terminal time $T = 600$ days and optimal terminal time $T^* = 512$ days.

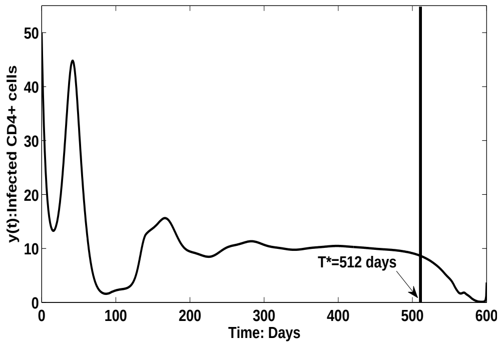


Figure 2. The state variable y with $y(0)$=50 units $mm^{-3}day^{-1}$, initial terminal time $T = 600$ days and optimal terminal time $T^* = 512$ days.

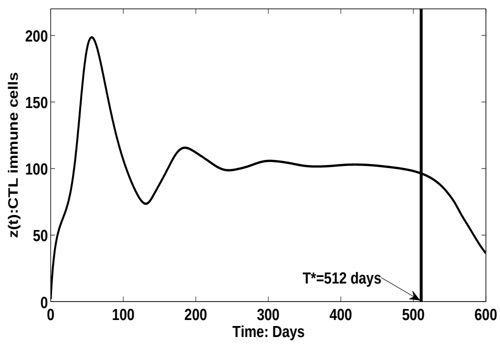


Figure 3. The state variable z with $z(0)$=2 units $mm^{-3}day^{-1}$, initial terminal time $T = 600$ days and optimal terminal time $T^* = 512$ days.

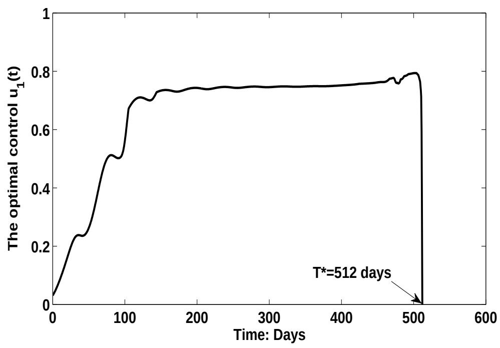


Figure 4. The optimal control $u_1^*(t)$ with $x(0)$=50 units $mm^{-3}day^{-1}$, $y(0)$=50 units $mm^{-3}day^{-1}$, $z(0)$=2 units $mm^{-3}day^{-1}$ and $T^* = 512$ days.

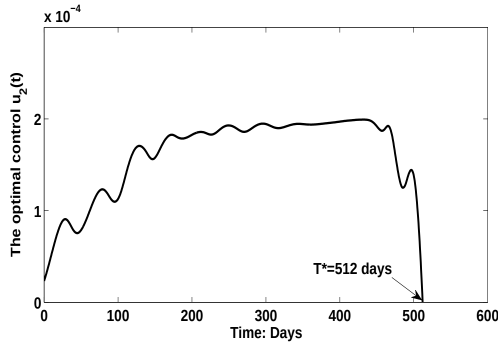


Figure 5. The optimal control $u_2^*(t)$ with $x(0)$=50 units $mm^{-3}day^{-1}$, $y(0)$=50 units $mm^{-3}day^{-1}$, $z(0)$=2 units $mm^{-3}day^{-1}$ and $T^* = 512$ days.

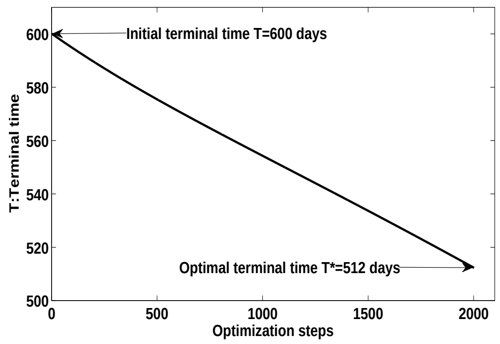


Figure 6. Estimation of optimal terminal time T^*, zero of ∇J with initial terminal time $T = 600$ days, $x(0)$=50 units $mm^{-3}day^{-1}$, $y(0)$=50 units $mm^{-3}day^{-1}$ and $z(0)$=2 units $mm^{-3}day^{-1}$.

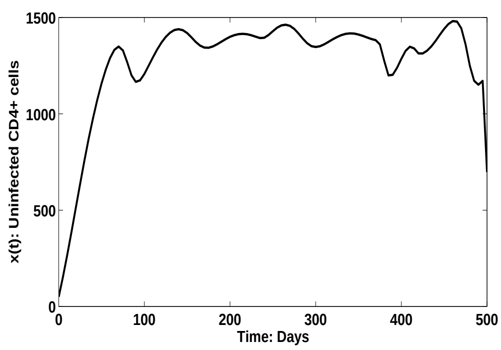


Figure 7. The state variable x with $x(0)$=50 units $mm^{-3}day^{-1}$ and terminal time $T = 500$ days.

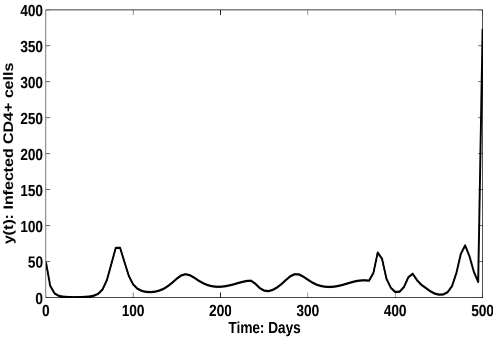


Figure 8. The state variable y with $y(0)$=50 units $mm^{-3}day^{-1}$ and terminal time $T = 500$ days.

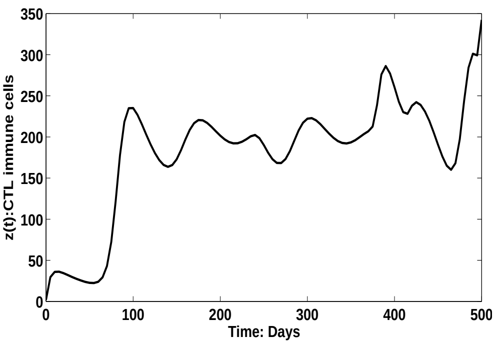


Figure 9. The state variable z with $z(0)$=2 units $mm^{-3}day^{-1}$ and terminal time $T = 500$ days.

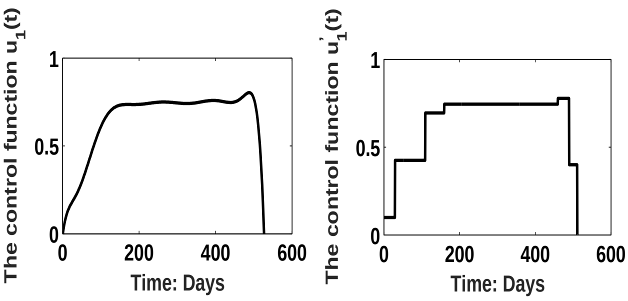


Figure 10. The profile fitting of the optimal control function $u_1^*(t)$ (Left) and the piecewise constant control function $u_1^{'}(t)$ (Right) with $x(0)$=50 units $mm^{-3}day^{-1}$, $y(0)$=50 units $mm^{-3}day^{-1}$, $z(0)$=2 units $mm^{-3}day^{-1}$ and $T^* = 512$ days.

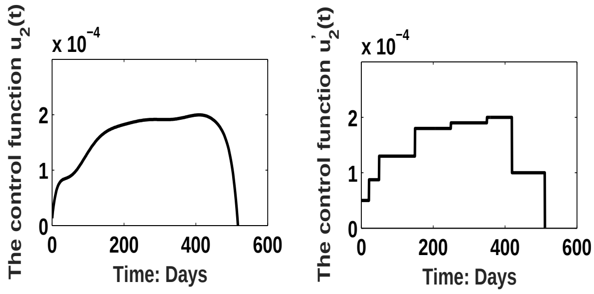


Figure 11. The profile fitting of the optimal control function $u_2^*(t)$ (Left) and the piecewise constant control function $u_2^{'}(t)$ (Right) with $x(0)$=50 units $mm^{-3}day^{-1}$, $y(0)$=50 units $mm^{-3}day^{-1}$, $z(0)$=2 units $mm^{-3}day^{-1}$ and $T^* = 512$ days.

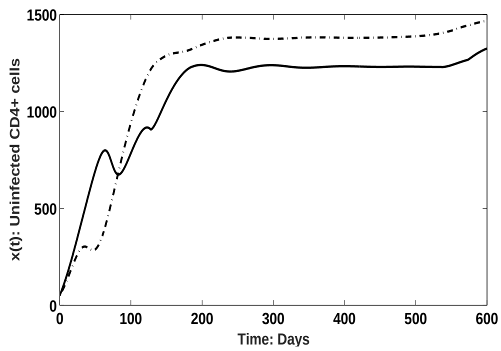


Figure 12. The state variable x using optimal controls $u_1^*(t)$ and $u_2^*(t)$ (DashDot) and using piecewise constant control functions $u_1^{'}(t)$ and $u_2^{'}(t)$ (Solid) with $x(0)$=50 units $mm^{-3}day^{-1}$, initial terminal time $T = 600$ days and optimal terminal time $T^* = 512$ days.

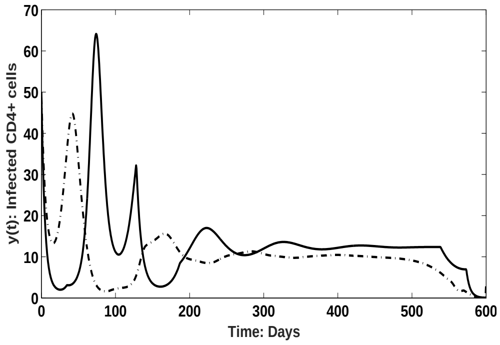


Figure 13. The state variable y using optimal controls $u_1^*(t)$ and $u_2^*(t)$ (DashDot) and using piecewise constant control functions $u_1^{'}(t)$ and $u_2^{'}(t)$ (Solid) with $y(0)$=50 units $mm^{-3}day^{-1}$, initial terminal time $T = 600$ days and optimal terminal time $T^* = 512$ days.

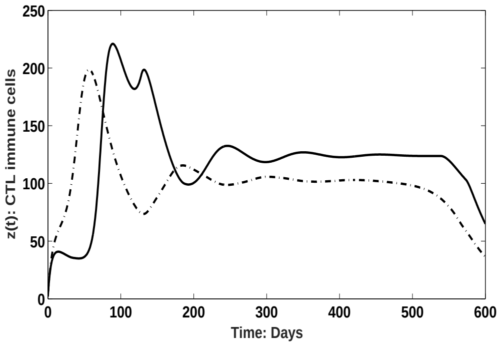


Figure 14. The state variable z using optimal controls $u_1^*(t)$ and $u_2^*(t)$ (DashDot) and using piecewise constant control functions $u_1^{'}(t)$ and $u_2^{'}(t)$ (Solid) with $z(0)$=2 units $mm^{-3}day^{-1}$, initial terminal time $T = 600$ days and optimal terminal time $T^* = 512$ days.

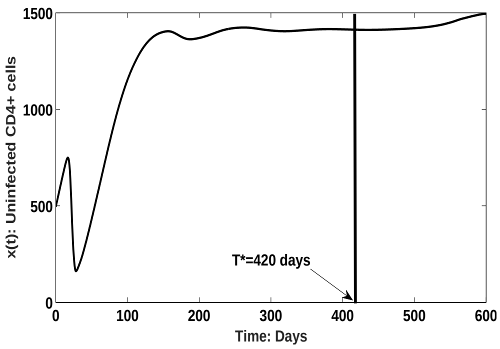


Figure 15. The state variable x with $x(0)$=494.3 units $mm^{-3}day^{-1}$, initial terminal time $T = 600$ days and optimal terminal time $T^* = 420$ days.

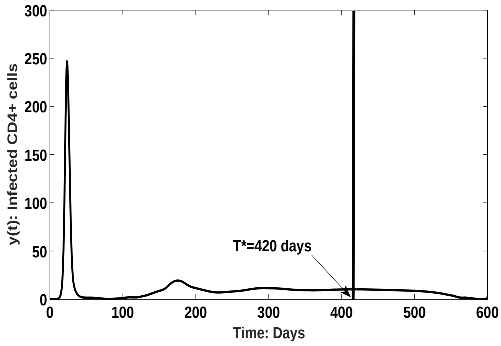


Figure 16. The state variable y with $y(0)$=0.04 units $mm^{-3}day^{-1}$, initial terminal time $T = 600$ days and optimal terminal time $T^* = 420$ days.

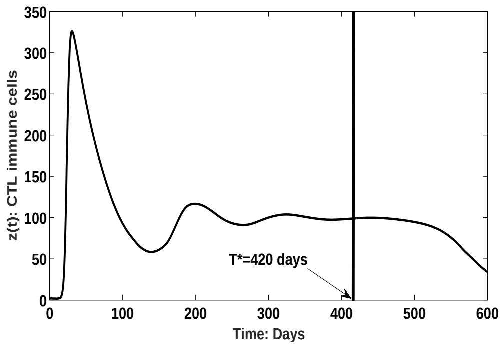


Figure 17. The state variable z with $z(0)$=2 units $mm^{-3}day^{-1}$, initial terminal time $T = 600$ days and optimal terminal time $T^* = 420$ days.

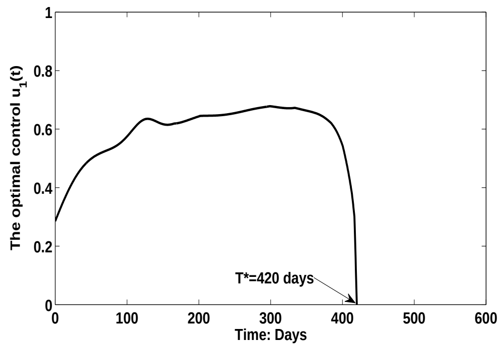


Figure 18. The optimal control $u_1^*(t)$ with $x(0)$=494.3 units $mm^{-3}day^{-1}$, $y(0)$=0.04 units $mm^{-3}day^{-1}$, $z(0)$=2 units $mm^{-3}day^{-1}$ and $T^* = 420$ days.

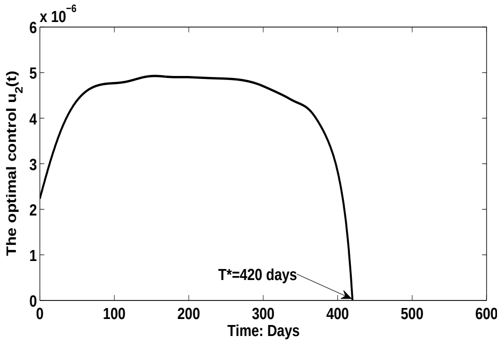


Figure 19. The optimal control $u_2^*(t)$ with $x(0)$=494.3 units $mm^{-3}day^{-1}$, $y(0)$=0.04 units $mm^{-3}day^{-1}$, $z(0)$=2 units $mm^{-3}day^{-1}$ and $T^* = 420$ days.

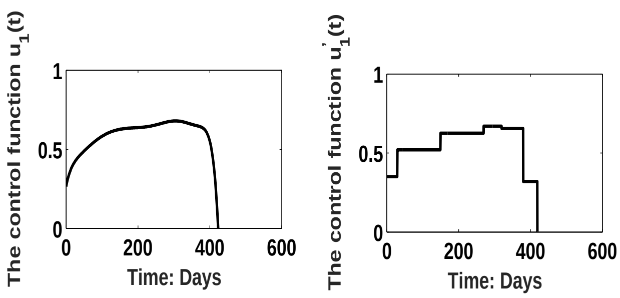


Figure 20. The profile fitting of the optimal control function $u_1^*(t)$ (Left) and the piecewise constant control function $u_1^{'}(t)$ (Right) with $x(0)$=494.3 units $mm^{-3}day^{-1}$, $y(0)$=0.04 units $mm^{-3}day^{-1}$, $z(0)$=2 units $mm^{-3}day^{-1}$ and $T^* = 420$ days.

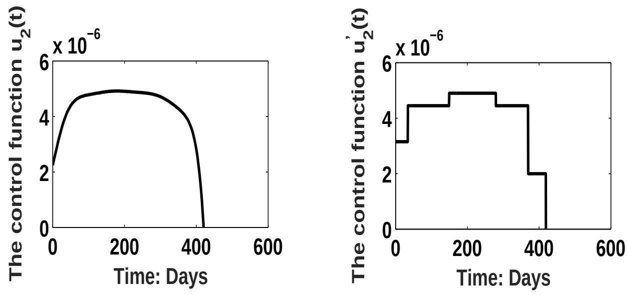


Figure 21. The profile fitting of the optimal control function $u_2^*(t)$ (Left) and the piecewise constant control function $u_2^{'}(t)$ (Right) with $x(0)$=494.3 units $mm^{-3}day^{-1}$, $y(0)$=0.04 units $mm^{-3}day^{-1}$, $z(0)$=2 units $mm^{-3}day^{-1}$ and $T^* = 420$ days.

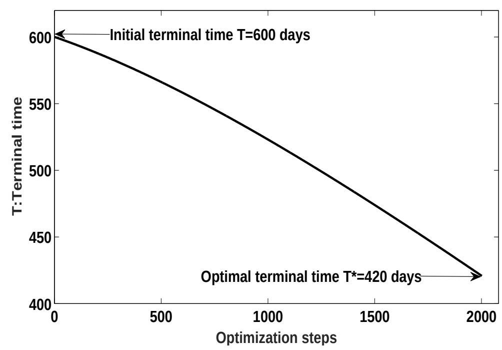


Figure 22. Estimation of optimal terminal time T^*, zero of ∇J with initial terminal time $T = 600$ days, $x(0)$=494.3 units $mm^{-3}day^{-1}$, $y(0)$=0.04 units $mm^{-3}day^{-1}$ and $z(0)$=2 units $mm^{-3}day^{-1}$.

4. Conclusion

In this work, a therapeutic approach has been suggested with the aim of treating the HIV infection by adopting a treatment strategy that uses both highly active antiretroviral therapy (HAART) to limit the virus evolution and an IL-2 immunotherapy to stimulate the active immune response.

In this sense, techniques of the optimal control theory have been used to develop an appropriate mathematical framework relating to this treatment approach. Indeed, a free terminal time optimal control problem was formulated by identifying a specific objective function that includes all the main objectives of the adopted therapeutic strategy.

The Pontryagin's maximum principle is used to characterize the optimal controls related to the used treatments. An adapted forward backward sweep method is implemented using a Runge-Kutta fourth order scheme and a gradient method routine for numerical resolution of the optimality system with the additional transversality condition for the terminal time.

Taking into account all the theoretical and numerical techniques used in the context of this research work, the treatment strategy suggested for the treatment of HIV infection has achieved all the objectives defined in the optimal control problem. Indeed, the adopted treatments have led to maximize the healthy $CD4^+$ T-cells and to establish an active immune response while reducing both the infection concentration and the treatment duration.

Finally, this optimal control approach has enabled the minimization of side effects and therefore the overall cost of the medication treatment allowing a significant improvement of the quality of life of HIV patients.

Acknowledgments

The authors would like to express their gratitude to the honorable reviewers who suggested many worthwhile changes to improve the quality of the manuscript.

References

[1] Agosto, L. M., Zhong, P., Munro, J. and Mothes, W. Highly active antiretroviral therapies are effective against HIV-1 cell-to-cell transmission. PLoS Pathog, 10, e1003982 (2014).

[2] Arlinghaus, S. Practical handbook of curve fitting. CRC press (1994).

[3] Boccia, A., Falugi, P., Maurer, H. and Vinter, R. B. Free time optimal control problems with time delays. In Decision and Control (CDC), 2013 IEEE 52nd Annual Conference, 520-525 (2013).

[4] Bonhoeffer, S., Coffin, J. M. and Nowak, M. A. Human immunodeficiency virus drug therapy and virus load. Journal of Virology, 71(4),3275-3278 (1997).

[5] Burden, T. N., Ernstberger, J. and Fister, K. R. Optimal control applied to immunotherapy. Discrete and Continuous Dynamical Systems Series B, 4(1), 135-146 (2004).

[6] Butler, S., Kirschner, D. and Lenhart, S. Optimal control of chemotherapy affecting the infectivity of HIV. Ann Arbor, 1001, 48109-0620 (1997).

[7] Cai, L., Guo, S. and Wang, S. Analysis of an extended HIV/AIDS epidemic model with treatment. Applied Mathematics and Computation, 236, 621-627 (2014).

[8] Callaway, D. S. and Perelson, A. S. HIV-1 infection and low steady state viral loads. Bulletin of mathematical biology, 64(1), 29-64 (2002).

[9] Cassels, S., Jenness, S. M. and Khanna, A. S. Conceptual Framework and Research Methods for Migration and HIV Transmission Dynamics. AIDS and Behavior, 18(12), 2302-2313 (2014).

[10] Cheney, E. and Kincaid, D. Numerical mathematics and computing. Cengage Learning, (2012).

[11] Coffin, J. M. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science, 267(5197), 483-489 (1995).

[12] Culshaw, R. V. and Ruan, S. A delay-differential equation model of HIV infection of $CD4^+$ T-cells. Mathematical biosciences, 165(1), 27-39 (2000).

[13] Culshaw, R. V., Ruan, S. and Spiteri, R. J. Optimal HIV treatment by maximising immune response. Journal of Mathematical Biology, 48(5), 545-562 (2004).

[14] Elmouki, I., Saadi, S. Quadratic and linear controls developing an optimal treatment for the use of BCG immunotherapy in superficial bladder cancer. Optimal Control Applications and Methods. (2015).

[15] Elmouki, I., Saadi, S. BCG immunotherapy optimization on an isoperimetric optimal control problem for the treatment of superficial bladder cancer. International Journal of Dynamics and Control. 1-7, (2014).

[16] Fauci, A.S., Desrosiers, R.C. Pathogenesis of HIV and SIV, 587-636. Cold Spring Harbor Laboratory Press, New York (1997).

[17] Fleming, W.H., Rishel, R.W. Deterministic and stochastic optimal control. Springer Verlag, New York (1975).

[18] Gray, C.M., Lawrence, J., Schapiro, J.M., Altman, J.D., Winters, M.A., Crompton, M., Loi, M., Kundu, S.K., Davis, M.M. and Merigan, T.C. Frequency of Class I HLA-Restricted anti-HIV $CD8^+$ T-cells in individuals receiving Highly Active Antiretroviral Therapy (HAART). The journal of immunology. 162, 1780-1788 (1999).

[19] Gumel, A.B. Spread and control of HIV: a mathematical model. Accromath. 26(8), (2013).

[20] Hamdache, A., Saadi, S., Elmouki, I., Zouhri, S. Two Therapeutic Approaches for the Treatment of HIV Infection in AIDS Stage. Journal of Applied Mathematical sciences. 7(105), 5243-5257 (2013).

[21] Hamdache, A., Elmouki, I., Saadi, S. Optimal Control with an Isoperimetric Constraint Applied to Cancer Immunotherapy. International Journal of Computer Applications. 94(15), 31-37 (2014).

[22] Hamdache, A., Saadi, S. and Elmouki, I. Nominal and neighboring-optimal control approaches to the adoptive immunotherapy for cancer. International Journal of Dynamics and Control, 1-16 (2015).

[23] Hlavacek, W.S., Wofsy, C. and Perelson, A.S. Dissociation of HIV-1 from follicular dendritic cells during HAART: mathematical analysis. Proceedings of the National Academy of Sciences. 96(26), 14681-14686 (1999).

[24] Iversen, A.K., Shafer, R.W., Wehrly, K., Winters, M.A., Mullins, J.I., Chesebro, B. and Merigan, T.C. Multidrug-resistant human immunodeficiency virus type 1 strains resulting from combination antiretroviral therapy. Journal of Virology. 70(2), 1086-1090 (1996).

[25] Jacobson, E.L., Pilaro, F. and Smith, A.K. Rationnal IL-2 therapy for HIV positifs individuals: daily low doses enhance immune function without toxicity. Proc. Natl. Acad. Sci USA. 93, 10405-10410 (1996).

[26] Janeway, C., Murphy, K. P., Travers, P. and Walport, M. Janeway's immunobiology, 530-535. Garland Science, London (2008).

[27] Jang, T., Kwon, H. D. and Lee, J. Free terminal time optimal control problem of an HIV model based on a conjugate gradient method. Bulletin of mathematical biology, 73(10), 2408-2429 (2011).

[28] Jiang, C., Lin, Q., Yu, C., Teo, K. L. and Duan, G. R. An exact penalty method for free terminal time optimal control problem with continuous inequality constraints. Journal of Optimization Theory and Applications, 154(1), 30-53 (2012).

[29] Joshi, H. R. Optimal control of an HIV immunology model. Optimal control applications and methods, 23(4), 199-213 (2002).

[30] Kirschner, D.E., Webb, G.F. Immunotherapy of HIV-1 infection. Journal of Biological Systems. 6(1), 71-83 (1998).

[31] Khanna, A. S., Dimitrov, D. T. and Goodreau, S. M. What can mathematical models tell us about the relationship between circular migrations and HIV transmission dynamics?. Mathematical biosciences and engineering: MBE, 11(5), 1065-1090 (2014).

[32] Klatzmann, D. and Abbas, A. K. The promise of low-dose interleukin-2 therapy for autoimmune and inflammatory diseases. Nature Reviews Immunology, (2015).

[33] Lenhart, S., Workman, T. Optimal control applied to biological models, 49-55. Chapman and Hall/CRC Mathematical and Computational Biology Series, New York (2007).

[34] Lévy, Y. Immunothérapie de l'infection par le VIH par l'utilisation de cytokines: un état des lieux. M/S: médecine sciences. 22(8-9), 751-754 (2006).

[35] Lukes, D. L. Differential Equations: Classical to Controlled, Mathematics in Science and Engineering, Academic Press, New York, (1982).

[36] Maartens, G., Celum, C. and Lewin, S. R. HIV infection: epidemiology, pathogenesis, treatment, and prevention. The Lancet, 384(9939), 258-271 (2014).

[37] MacArthur, R.D., Novak, R.M. Maraviroc: The First of a New Class of Antiretroviral Agents. Oxford journals. 47(2), 236-241 (2008).

[38] McAsey, M., Mou, L., Han, W. Convergence of the Forward-Backward Sweep Method in optimal control. Comput Optim Appl. **3**, (2012).

[39] Mastroberardino, A., Cheng, Y., Abdelrazec, A. and Liu, H. Mathematical modeling of the HIV/AIDS epidemic in Cuba. International Journal of Biomathematics, 1550047 (2015).

[40] Merry, C., Barry, M.G., Mulcahy, F., Ryan, M., Heavey, J., Tjia, J.F., Gibbons, S.E., Breckenridge, A.M. and Back, D.J. Saquinavir pharmacokinetics alone and in combination with ritonavir in HIV-infected patients. AIDS. 11(4), (1997).

[41] Palanki, S., Kravaris, C. and Wang, H. Y. Optimal feedback control of batch reactors with a state inequality constraint and free terminal time. Chemical engineering science, 49(1), 85-97 (1994).

[42] Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. and Ho, D. D. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science, 271(5255), 1582-1586 (1996).

[43] Perelson, A. S., Kirschner, D. E. and De Boer, R. Dynamics of HIV infection of $CD4^+$ T cells. Mathematical biosciences, 114(1), 81-125 (1993).

[44] Pontryagin, L. S. Mathematical theory of optimal processes. CRC Press, (1987).

[45] Pontryagin, L. S., Boltyanskii, V. G. and Gamkrelidze, R. V. EF Mishchenko The Mathematical Theory of Optimal Processes. New York: Interscience (1962).

[46] Pooseh, S., Almeida, R. and Torres, D. F. Fractional order optimal control problems with free terminal time. arXiv preprint arXiv: 1302.1717 (2013).

[47] Qun, L., Loxton, R., Teo, K. L. and Wu, Y. H. A new computational method for a class of free terminal time optimal control problems. Pacific Journal of Optimization. 7(1), 63-81 (2011).

[48] Raffi, F.: Enfuvirtide, premier inhibiteur de fusion dans le traitement de l'infection par le virus de l'immunodéficience humaine: mécanisme d'action et pharmacocintique. Médecine et maladies infectieuses. 34, 3-7 (2004).

[49] Roshanfekr, M., Farahi, M. H. and Rahbarian, R. A different approach of optimal control on an HIV immunology model. Ain Shams Engineering Journal, 5(1), 213-219 (2014).

[50] Roy, P.K., Chatterjee, A.N. T-cell proliferation in a mathematical model of CTL activity through HIV-1 infection. Proceedings of the World Congress on Engineering. 1, (2010).

[51] Roy, P. K., Saha, S. and Al Basir, F. Effect of awareness programs in controlling the disease HIV/AIDS: an optimal control theoretic approach. Advances in Difference Equations, 2015(1), 1-18 (2015).

[52] Saadi, S., Elmouki, I. and Hamdache, A. Impulsive control dosing BCG immunotherapy for non-muscle invasive bladder cancer. International Journal of Dynamics and Control, 1-11 (2015).

[53] Stengel, R. F., Ghigliazza, R. M. and Kulkarni, N. V. Optimal enhancement of immune response Bioinformatics,18(9), 1227-1235 (2002).

[54] Su, B., Lederle, A., Laumond, G., Schmidt, S., Decoville, T., Ducloy, C. and Moog, C. Antibody Inhibition of HIV-1 Transmission from Antigen-presenting Cells to CD4 T Lymphocytes Involves Immune Cell Activation. AIDS research and human retroviruses, 30(S1), A154-A154 (2014).

[55] Trélat, E. Controle optimal: théorie et applications. Paris: Vuibert (2005).

[56] Wang, L. and Li, M. Y. Mathematical analysis of the global dynamics of a model for HIV infection of $CD4^+$ T-cells. Mathematical Biosciences, 200(1), 44-57 (2006).

[57] Zhou, X., Song, X. and Shi, X. A differential equation model of HIV infection of CD4+ T-cells with cure rate. Journal of Mathematical Analysis and Applications, 342(2), 1342-1355 (2008).

[58] Zurakowski, R. and Teel, A. R. A model predictive control based scheduling method for HIV therapy. Journal of Theoretical Biology, 238(2), 368-382 (2006).

Amine Hamdache *is a researcher at the Laboratory of Analysis, Modeling and Simulation, Department of Mathematics and Computer Science, Faculty of Sciences Ben M'sik, Hassan II University, Casablanca, Morocco. His current area of research includes the application of optimal control theory to dynamical models for providing (deterministic, hybrid, neighboring and stochastic) optimal control approaches that allow to suggest optimal treatment strategies for cancer and HIV infection.*

Smahane Saadi *is a Professor of Analysis at the Faculty of Sciences Ben M'sik and a researcher at the Laboratory of Analysis, Modeling and Simulation, Department of Mathematics and Computer Science, Faculty of Sciences Ben M'sik, Hassan II University, Casablanca, Morocco. Her current area of research includes the application of optimal control theory to dynamical models for providing (deterministic, hybrid, neighboring and stochastic) optimal control approaches that allow to suggest optimal treatment strategies for cancer and HIV infection.*

Ilias Elmouki *is a researcher at the Laboratory of Analysis, Modeling and Simulation, Department of Mathematics and Computer Science, Faculty of Sciences Ben M'sik, Hassan II University, Casablanca, Morocco. His current area of research includes the application of optimal control theory to dynamical models for providing (deterministic, hybrid, neighboring and stochastic) optimal control approaches that allow to suggest optimal treatment strategies for cancer and HIV infection.*

An International Journal of Optimization and Control:
Theories & Applications

Volume: 6 Number: 1
January 2016

This page intentionally left blank.

An International Journal of Optimization
and Control: Theories & Applications
Vol.6, No.1, pp.53-61 (2016) © IJOCTA
ISSN:2146-0957 eISSN:2146-5703
DOI:10.11121/ijocta.01.2016.00254
http://www.ijocta.com

Optimization of cereal output in presence of locusts

Nacima Moussouni[a] and Mohamed Aidene[b]

University Mouloud Mammeri of Tizi-Ouzou, Faculty of Science,
Department of Mathematics, L2CSP Laboratory

Email: [a]nmoussouni@yahoo.fr, [b]aidene@mail.ummto.dz

(Received March 22, 2015; in final form January 27, 2016)

Abstract. In this paper, we study a modelization of the evolution of cereal output production, controlled by adding fertilizers and in presence of locusts, then by adding insecticides. The aim is to maximize the cereal output and meanwhile minimize pollution caused by adding fertilizers and insecticides. The optimal control problem obtained is solved theoretically by using the Pontryagin Maximum Principle, and then numerically with shooting method.

Keywords: Optimal control; optimization; Pontryagin maximum principal.
AMS Classification: 49J15.

1. Statement of the problem

Consider $x(t)$, $t \in [0,T]$ the rate of pollution at time t. In cereal field, If the farmer does not put fertilizers and insecticides then the evolution of pollution satisfies

$$\dot{x}(t) = -\alpha x(t), \quad t \in [0,T];$$

where α is the natural decreasing rate. Note that the rate $x(t)$, $t \in [0,T]$ is decreasing.
In order to increase the cereal output, we add fertilizers and insecticides to protect the crop harm locusts. Denoting by $u(t)$ and $v(t)$, $t \in [0,T]$ the quantities of fertilizers and insecticides respectively, in this case $x(t)$ is evolving as

$$\dot{x}(t) = -\alpha x(t) + u(t) + v(t), \qquad (1)$$
$$x(0) = x_0 > 0, t \in [0,T]$$

Our goal is to minimize the pollution generated by fertilizers and insecticides, and optimize the cereal output from the seed to the harvest.

In practice, we choose typically $T = 10\, months$, which corresponds to a cycle of cereal production from September to July.
Let be $y(t)$, $t \in [0,T]$ the quantity of the cereal production. Adding fertilizers, the production increases, this production decreases with the presence of locusts and by adding large quantities of fertilizers.

Denoting by $z(t), t \in [0,T]$, the quantity of locusts presents in cereal field. In this case the evolution of cereal output is given for $t \in [0,T]$ by:

$$\dot{y}(t) = -by(t)z(t) + \sqrt{(M - u(t))(m + u(t))} \qquad (2)$$
$$y(0) = 0.$$

and $z(t)$ verify the following equation

$$\dot{z}(t) = z(t)(c(t)y(t) - d(t)) - v(t), \qquad (3)$$
$$z(0) = z_0 > 0, \quad t \in [0,T]$$

where $m > 0, M > 0$ are real numbers, b is the rate of reproduction of cereal, $c(t), t \in [0,T]$ is the rate of reproduction of locusts and $d(t), t \in [0,T]$, is the rate of extinction of locusts. All those parameters will be identified subsequently.

Note that if we add a too large quantity of fertilizers and insecticides, this causes the death of locusts but also the death of cereals.

The functions $u(.)$ and $v(.)$ are considered as controls. Those controls $u(.)$ and $v(.)$ considered are submitted to constraints. They are such that

$$0 \leq v(.) \leq V, \ (*)$$

and

$$0 \leq u(.) \leq M. \ (**)$$

Note that $V > 0$ will be by identified in the following section.

In a cereal field, the aim is to maximize the production of cereals and to minimize the bad effects of pollution given by insecticides and fertilizers. For this our criterion is:

$$J(u) = \beta x(T) - y(T) \rightarrow \min_{u,\ v},$$

where $\beta > 0$ is a real number to be chosen, $x(.)$ is solution of (1) and $y(.)$ is solution of (2).
Minimizing J corresponds to realizing a compromise between maximizing the cereal output and minimizing the bad effects of pollution given by insecticides and fertilizers.

Finally, our problem is as follows

$$\begin{cases} \dot{x}(t) = -\alpha x(t) + u(t) + v(t), \\ x(0) = x_0 > 0 \\ \\ \dot{y}(t) = -by(t)z(t) \\ +\sqrt{(M - u(t))(m + u(t))}, \\ y(0) = 0 \\ \\ \dot{z}(t) = z(t)(c(t)y(t) - d(t)) - v(t), \\ z(0) = z_0 > 0 \\ \\ 0 \leq v(t) \leq V \\ \\ 0 \leq u(t) \leq M. \quad t \in [0, T]. \end{cases}$$

Here we consider that the final time T is fixed.

This problem is inspired by a model used in [9], where the authors formulated a model without presence of locusts. They calculated the quantities of fertilizers to put in cereal field to get a better production. The reader can refer also to [11].

This article is structured as follows. In Section 2, we provide an identification of the parameters considered in the model with real life measures used in Algeria see [4,7]. In this section, we calculated the rate of reproduction of cereals using a simple dechotomy method. we calculated also the durations of maturity of locusts, then the reproduction rate and the rate of extinction of locusts, in hot and cold seasons.

Section 3 is devoted to the study of necessary condition of optimality based on the Pontryagin Maximum Principle see[10,12]. We make a rigorous mathematical analysis of the extremal equations leading to a precise expression of the optimal control. In Section 4, we provide numerical simulations based on the rigorous mathematical analysis, using the shooting method and we comment these results. Note that these numerical results describe the best possible way for a farmer to realize a good compromise between maximizing the cereal output and minimizing pollution effects consequences of fertilizers and insecticides.

2. Identification of the parameters of the model

In what follows, the time t is given in months, and T corresponds to a cycle of cereal production, $T = 10$ months.

The quantities of fertilizers used in Algeria are given by [7]:

$$u(t) = \begin{cases} 100kg/ha \ if \\ t \in [0, 1] = [September, October] \\ \\ \frac{100}{3}kg/ha \ if \\ t \in [2, 3] = [November, December] \\ \\ \frac{200}{3}kg/ha \ if \\ t \in [6, 7] = [March, April] \\ \\ 0 \\ otherwise([1, 2] \cup [3, 6] \cup [7, 10]) \end{cases} \tag{4}$$

According to [7], the quantity of cereal output without fertilizers is equal to 500 *kilograms* per *hectare*.

From this in our model corresponds to $u(t) = 0$, and hence, using equation (2);
$\dot{y}(t) = \sqrt{Mm}$, then we obtain

$$10\sqrt{Mm} = 500. \tag{5}$$

From [7], the cereal output with the addition of fertilizers and in absence of locusts as described

by (2) is equal to 4500 *kilograms* per *hectare*.
In our model, this leads to

$$y(T) = 4500 = \int_0^1 \sqrt{(M-100)(m+100)}dt + \int_2^3 \sqrt{(M-100/3)(m+100/3)}dt + \int_6^7 \sqrt{(M-200/3)(m+200/3)}dt + 7\sqrt{Mm}.$$

solving the system

$$\begin{cases} 10\sqrt{Mm} = 500 \\ \sqrt{(M-100)(m+100)} \\ +\sqrt{(M-100/3)(m+100/3)} \\ +\sqrt{(M-200/3)(m+200/3)} \\ +7\sqrt{Mm} = 4500 \end{cases}.$$

and leads to

$$M = 300,83, \quad m = 0.00083.$$

Let us now compute the value of the decreasing rate α of pollution, according to real-life data. In the absence of fertilizers and insecticides, $u(t) = 0$, we have $x(0) = x_0 = 119mg/l$ at $t = 0$, and $x(T) = 28mg/l$ at $T = 10$ *months*. From this by using formula (1), we obtain:

$$x(T) = x_0 e^{-\alpha T}$$
$$\Leftrightarrow 28 = 119 e^{-10\alpha},$$

then

$$\alpha = 0.12.$$

Note that the locust attack held in May. From equation (2), and in absence of fertilizers ($u(t) = 0$),

$$\dot{y}(t) = -b\ y(t)\ z(t) + \sqrt{mM},\ t \in [0, T].$$

the larval density causing damage is 5000 locusts per hectare [4], they consume 80 % of cereal a day.
For $t_1 = \frac{1}{30}$ *month* = 1 day, we will have:

$$y(\frac{1}{30}) = 0.2y_0.$$

Note that $y_0 = y(8)$ is cereal production at the time of the attack of locusts. the value of b is determined by solving the following differential equation:

$$\dot{y}(t) = -5000\ b\ y(t) + \sqrt{Mm}$$

under the initial conditions:

$$y(8) = y_0, y(\frac{1}{30}) = 0.2y_0.$$

Using these data, we will have

$$\frac{d}{dt}(y(t) - \frac{\sqrt{Mm}}{bz(t)}) = -bz(t)(y(t) - \frac{\sqrt{Mm}}{bz(t)})$$

$$\Rightarrow \quad y(t) - \frac{\sqrt{Mm}}{bz(t)} = cste \times e^{-bz(t)t};$$

then

$$y(t) = \frac{\sqrt{Mm}}{bz(t)} + (y_0 - \frac{\sqrt{Mm}}{bz(t)})e^{-bz(t)t},\ t \in [0, T].$$

For $t = t_1$, we will have:

$$\frac{\sqrt{Mm}}{bz(t)} + (y_0 - \frac{\sqrt{Mm}}{bz(t)})e^{-bz(t)t_1} = 0.2y_0.$$

To determine the value of y_0, we set the following assumptions:

- The locust come in May .
- Insects attack a fraught field of cereal.
- The field has not been attacked before May.

To calculate y_0, we solve the following equation:

$$y(8) = \int_0^1 \sqrt{(M-100)(m+100)}dt + \int_2^3 \sqrt{(M-100/3)(m+100/3)}dt + \int_6^7 \sqrt{(M-200/3)(m+200/3)}dt + 5\sqrt{Mm}.$$

Such that $M = 300$ and $m = 0.00083$, then $y(8) = 4379kg/ha$. To determine the value of b, plot the graph of the following function:

$$b \mapsto \frac{\sqrt{Mm}}{bz} + (y_0 - \frac{\sqrt{Mm}}{bz})e^{-bzt_1} - 0.2y_0$$

where $t_1 = \frac{1}{30}$ month, $y_0 = 4379kg/ha$ and z = $5000 locusts/ha$.

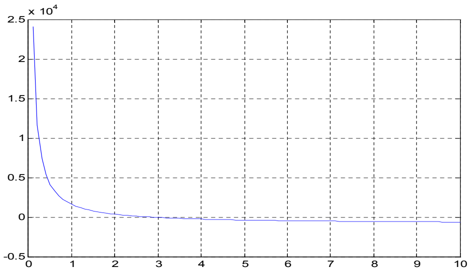


Figure 1. $t \mapsto b(t)$.

By Dichotomy, we obtain $b = 2.85$ (see Figure 1).

Population of locusts on which is based our work is 55000 larvae and 375 adults. A female lays 140 eggs in two generations [4]. Average losses of eggs are about 33% [4]. The Table 1 shows the duration of maturity of locusts. For method for calculating incubation periods see [4].

Table 1. Durations of maturity of locusts [4].

Locusts temperatures	high	bass
Eggs	IT: 11 days	IT: 41 days
Larvae	DT: 80 days	DT: 21 days
Adults	AM : 20 days	AM: 6 months

Indications: IT: Incubation time, DT: Development time and AM: Adults maturity.

After hatching of eggs, the larvae pass from five larval stages L_1, L_2, L_3,L_4, L_5. The percentages of mortality in different stages of larvae are given in Table 2. For more informations see[4]

Table 2. Larval mortality [4].

Stages	L_1	L_2	L_3	L_4	L_5
percentages	70%	20%	10%	10%	10%

There, and using these data, we calculate the number of locusts that can produce a viable female in hot and cold seasons.

Hot seasons:

$$N_1 = 140 \times 0.66 \times 0.3 \times 0.8 \times (0.9)^3 = 16.16$$
$$\simeq 17 locusts$$

Cold seasons:

$$N_2 = 140 \times 0.35 \times 0.3 \times 0.8 \times (0.9)^3 = 8.57$$
$$\simeq 9 \ locusts$$

In other words: On 55375 Locust (larvae, immature adults , mature adults) we assume that 100 females lay their eggs in two generations. They will generate 17 locusts viable in the hot season and 9 locusts viable in the cold season. The rate of reproduction of locusts $c(t)$, $t \in [0, T]$ is represented in Figure 2 and calculated as follows:

$$c(t) = \begin{cases} \frac{200}{55375} * 17, & in\ hot\ season \\ \frac{200}{55375} * 9, & in\ cold\ season \end{cases} .$$

Then

$$c(t) = \begin{cases} 0.0613, & in\ hot\ season \\ 0.0288, & in\ cold\ season \end{cases} .$$

Analytic expression of $c(t), t \in [0, T]$ is

$$c(t) = 0.0288 + (0.0613 - 0.0288)\frac{(t-5)^2}{25}$$

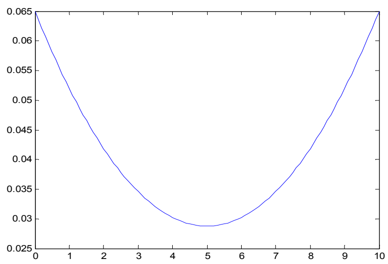


Figure 2. $t \mapsto c(t)$.

The average lifespan of a locust is 3 months in hot periods, it is 8 months in cold periods (see [4]). we assume that we are at 0 when 90 % of the population locusts disappeared (this assumption is possible because after elimination of locusts, the solitary locusts do not disappear).

From Constraint (3), in the absence of insecticides and food,

$$\dot{z}(t) = -d\ z(t).$$

This differential equation has the solution:

$$z(t) = z_0 e^{-dt}, \quad t \in [0, T],$$

where $z_0 = z(0)$.

In hot season, $t = 3\ months$,

$$z_0\ e^{-3d} = 0.1\ z_0$$

$$\Rightarrow e^{-3d} = 0.1$$

$$d = -\frac{1}{3}ln(0.1) = 0.767$$

In cold season $t = 8\ months$:

$$z_0\ e^{-8d} = 0.1\ z_0$$

$$\Rightarrow e^{-8d} = 0.1$$

$$d = -\frac{1}{8}ln(0.1) = 0.287.$$

In other words:

$$d(t) = \begin{cases} 0.767 & \text{in hot season} \\ 0.287 & \text{in cold season} \end{cases}.$$

The analytical expression for the rate of extinction of locusts $d(t)$ represented in Figure 3 and it is given by:

$$d(t) = 0.287 + (0.767 - 0.287)\frac{(t-5)^2}{25}, \quad t \in [0, T].$$

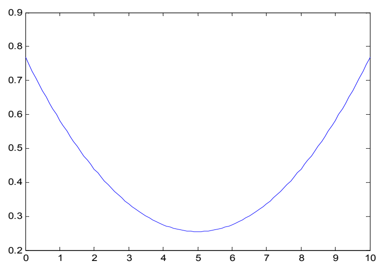


Figure 3. $t \mapsto d(t)$.

For the value of $V = 5l/ha$ see [4].

3. Theoretical solving of the optimal control problem

In this section, we solve the problem (4) theoretically by employing the Pontryagin Maximum Principle.

Let us first recall a version of the Pontryagin Maximum Principle (see [10,12]).

Theorem 1. *We consider the control system on* $\mathbb{R}^n$

$$\dot{x}(t) = f(t, x(t), u(t)), \tag{6}$$

where $f : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ *in* C^1*. Where the controls are the measurable and bounded functions in* $[0, t_e(u)]$ *of* $\mathbb{R}^+$ *in values in* $\Omega \subset \mathbb{R}^m$*. Let* M_0 *and* M_1 *two subsets of* $\mathbb{R}^m$*. Note by* U *the set of an admissible controls* u *whose corresponding trajectories connect one point of* M_0 *to a final point in* M_1 *in time* $t(u) < t_e(u)$*. Note the quality criterion by*

$$C(t, u) = \int_0^t f^0(s, x(s), u(s))ds + g(t, x(t)),$$

where $f^0 : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ *and* $g : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ *in* C^1*, and* $x(.)$ *is solution of (7) associated to the control* u*.*

We consider the following optimal control problem: determine one trajectory connecting M_0 *to* M_1 *and minimize the cost.*

If a control u *is optimal in* $[0, T]$*, then there exist an application* $p(.)$ *absolutely continuous on [0,T], with values in* $\mathbb{R}^n$*, called adjoint vector, and a real nonpositive number* p^0 *such that* $(p(.), p^0)$ *is nontrivial, and for almost all* $t \in [0, T]$

$$\dot{x}(t) = \frac{\partial H}{\partial p}(t, x(t), p(t), p^0, u(t)), \tag{7}$$

$$\dot{p}(t) = \frac{-\partial H}{\partial x}(t, x(t), p(t), p^0, u(t)). \tag{8}$$

where

$$H(t, x, p, p^0, u) = p'(t)f(t, x, u) + \ p^0 f^0(t, x, u)$$

is the Hamiltonian of the system (7).

Moreover, we have a condition of maximization

$$\begin{aligned} &H(t, x(t), p(t), p^0, u(t)) \\ &= \max_{v \in U} H(t, x(t), p(t), p^0, v), \end{aligned} \tag{9}$$

for all $t \in [0, T]$*.*

Moreover, if M_0 *and* M_1 *are two submanifolds of* $\mathbb{R}^n$ *having tangent spaces in* $x(0) \in M_0$ *and* $x(T) \in M_1$*, then the adjoint vector satisfies the transversality conditions*

$$p(0) \perp T_{x(0)} M_0, \tag{10}$$

and

$$p(T) - p^0 \frac{\partial g}{\partial x}(T, x(T)) \perp T_{x(T)} M_1. \tag{11}$$

We apply the Pontryagin Maximum Principle to our specific optimal control problem.
The Hamiltonian of System (4) is

$$H(x,p,p^0,u,v) = p_x(-\alpha\ x+u+v)+p_y(-b\ y\ z+ \sqrt{(M-m)(m+u)}) + p_z(z(cy-d)-v)$$

where $p(t) = \begin{pmatrix} p_x(t) \\ p_y(t) \\ p_z(t) \end{pmatrix}$, $t \in [0,T]$ is adjoint vector solution of the following system:

$$\begin{cases} \dot{p}_x = -\frac{\partial H}{\partial x} = \alpha p_x, \\ \dot{p}_y = -\frac{\partial H}{\partial y} = bp_y z - cp_z z = \\ bzp_y - czp_z = z(bp_y - cp_z), \\ \dot{p}_z = -\frac{\partial H}{\partial z} = byp_y - p_z(cy-d) = \\ y(bp_y - cp_z) + p_z d. \end{cases} \quad (12)$$

The final transversality condition
We know that $x(T)$ is free, the final transversality condition leads to

$$p(T) = p^0 \nabla g(x(T)).$$

In other words:

$$p_x(T) = p^0 \frac{\partial}{\partial x}(\beta x(T) - y(T)),$$

$$p_y(T) = p^0 \frac{\partial}{\partial y}(\beta x(T) - y(T)),$$

$$p_z(T) = p^0 \frac{\partial}{\partial z}(\beta x(T) - y(T)).$$

Then $p_x(T) = -\beta$, $p_y(T) = 1$, $p_z(T) = 0$.
It is easy to see that

$$p_x(t) = -\beta e^{\alpha(t-T)}, \quad t \in [0,T]$$

Remark 1. *Since $\beta > 0$, it follows that, for every $t \in [0,T]$, $p_x(t) < 0$.*

Lemma 1. $p^0 \neq 0$.

Proof. We argue by contradiction. If $p^0 = 0$ then $p(T) = 0$ so $(p(T),p^0) = (0,0)$; this is in contradiction with the Pontryagin Maximum Principle. □

Remark 2. *If $p^0 = 0$ then the extremal $(x(.),p(.),u(.))$ is said normal and in this case it is usual to normalize the adjoint vector so that $p^0 = -1$. If $p^0 = 0$, then the extremal $(x(.),p(.),u(.))$ is said abnormal. Note that several works have been devoted to the investigation of abnormal extremals in a generic context (see [1],[2],[3]). In our example, the abnormal case does not occur.*

Lemma 2. *$p_y(.)$ is not canceled identically on a sub interval.*

Proof. Assume that $p_y \equiv 0$ is canceled identically on a sub interval of $[0,T]$,

$$p_y \equiv 0 \Rightarrow \dot{p_y} \equiv 0$$
$$\Rightarrow z(t)(bp_y(t) - c(t)p_z(t)) = 0$$
$$\Rightarrow p_z \equiv 0$$

So by unicity of Cauchy,

$$p_y \equiv p_z \equiv 0 \ sur\ [0,T],$$

this is in contradiction with $p_y(T) = 1$. □

For the proof of the following lemma see [11].

Lemma 3. *$p_x(.) - p_z(.)$ does not vanish identically on a subset interval.*

To study the maximization condition, we search the maximum on v and u of the following function

$$p_x u(t) + p_y(t)\sqrt{(M-u(t))(m+u(t))} + (p_x(t) - p_z(t))v(t)$$

It is clear that

$$v(t) = \begin{cases} 0 & if\ p_x(t) - p_z(t) < 0 \quad , \\ V & if\ p_x(t) - p_z(t) > 0 \quad . \end{cases} \quad (13)$$

To determine the optimal control $u(.)$, we search in $[-m,M]$, the maximum of the following function

$$\phi(u) = p_x u + p_y\sqrt{(M-u)(m+u)}.$$

Function ϕ is defined on $[-m,M]$. To found its absolute maximum proceed as follows

$$\phi'(u) = p_x + p_y \frac{-u + \frac{M-m}{2}}{\sqrt{(M-u)(m+u)}}$$

$$\phi'(u) = 0 \Leftrightarrow p_y \frac{-u + \frac{M-m}{2}}{\sqrt{(M-u)(m+u)}} = -p_x$$

Note that $p_x(t) < 0$, then

$$p_y(-u + \frac{M-m}{2}) > 0$$

$$(p_x^2 + p_y^2)u^2 - (M-m)(p_x^2+p_y^2)u + \frac{(M-m)^2}{2}p_y^2$$

$$-p_x^2 Mm = 0.$$

The absolute maximum of ϕ on $[-m, M]$ is

$$u_\phi = \frac{M-m}{2} + \frac{M+m}{2\sqrt{p_x^2(t)+p_y^2(t)}} p_x sign(p_y(t))$$

We deduce two cases:

$p_y(t) > 0$, in this case the maximum of ϕ on $[0, M]$ is

$$0 \;\; if \;\; u_\phi < 0,$$

$$u_\phi \;\; if \;\; u_\phi \geq 0,$$

$p_y(t) < 0$

$$\phi(0) = p_y(t)\sqrt{Mm} < 0$$

and

$$\phi(M) = p_x(t)M < 0.$$

The maximum of ϕ on $[0, M]$ is

$$0 \;\; if \;\; p_y(t)\sqrt{Mm} > p_x(t)M$$

$$M \;\; if \;\; p_x(t)M > p_y(t)\sqrt{Mm}, \; t \in [0, T]$$

Conclusion

The optimal control of system (4) is

$$u(t) = \begin{cases} 0 \; if \; p_y(t) > 0 \\ and \; \frac{M-m}{2} + \frac{M+m}{2\sqrt{p_x^2(t)+p_y^2(t)}} p_x(t) \leq 0, \\ \\ \frac{M-m}{2} + \frac{(M+m)p_x(t)}{2\sqrt{p_x^2(t)+p_y^2(t)}} \; if \;\; p_y(t) > 0 \\ and \; \frac{M-m}{2} + \frac{M+m}{2\sqrt{p_x^2(t)+p_y^2(t)}} p_x(t) > 0, \\ \\ 0 \;\; if \;\; p_y(t) < 0 \\ and \;\; p_y(t)\sqrt{m} > p_x(t)\sqrt{M} \\ \\ M \;\; if \;\; p_y(t) < 0 \\ and \;\; p_y(t)\sqrt{m} < p_x(t)\sqrt{M}. \end{cases} \tag{14}$$

We proved the following theorem.

Theorem 2. *If* $p_y(t) > 0$ *and* $\frac{M-m}{2} + \frac{M+m}{2\sqrt{p_x^2(t)+p_y^2(t)}} p_x(t) < 0$, *then*

$$u(t) = 0, \quad t \in [0, T].$$

If $p_y(t) > 0$ *et* $\frac{M-m}{2} + \frac{M+m}{2\sqrt{p_x^2(t)+p_y^2(t)}} p_x(t) > 0$, *then*

$$u(t) = \frac{M-m}{2} + \frac{(M+m)p_x(t)}{2\sqrt{p_x^2(t)+p_y^2(t)}}, t \in [0, T].$$

If $p_y(t) < 0$ *and* $p_y(t)\sqrt{m} > p_x(t)\sqrt{M}$, *then*

$$u(t) = 0, \quad t \in [0, T].$$

If $p_y(t) < 0$ *and* $p_y(t)\sqrt{m} < p_x(t)\sqrt{M}$, *then*

$$u(t) = M, \; t \in [0, T].$$

4. Numerical simulations

We give here a brief overview of the indirect method, this method is based on the Pontryagin Maximum Principle, which gives necessary condition for optimality, and states that every optimal trajectory is the projection of an extremal. If one is able from the condition of maximization to express the extremal control in function of $(x(t), p(t))$, then the extremal system is a differential system of the form $\dot{z}(t) = F(t, z(t))$, where $z(t) = (x(t), p(t))$, and the values of initial, final and transversality conditions are put in the form $R(z(0), z(T)) = 0$.

Finally we obtain a problem of the form

$$\begin{cases} \dot{z}(t) = F(t, z(t)), \\ R(z(0), z(T)) = 0 \end{cases}. \tag{15}$$

Let $z(t, z_0)$ the solution of Cauchy's problem

$$\dot{z}(t) = F(t, z(t)), \quad z(0) = z_0.$$

Put $G(z_0) = R(z_0, z(T, z_0))$. The problem (16) it equivalent to

$$G(z_0) = 0$$

which is solved using the Newton's method. For more details on the shooting method, the reader can refer to [12].

Let us consider now that the ground is fertilized from September to July, and the insecticides is put in a continuous way, from May to July. To solve the problem, the indirect method based on the Pontryagin Maximum Principle is used. We provide in Table 3 numerical results of $x(T)$, $y(T)$ and $z(T)$ for several values of the weight parameter β. The numerical simulations were led using *Matlab* on a desktop computer.

Table 3. Pollution, cereal output and locusts final as time function of β.

0	909	5.94	248
10	36.40	16.90	13.39
50	35.97	20.76	2.50
70	35.91	18.47	1.92
100	35.90	15.65	1.39
150	35.87	12.18	1
200	35.84	9.85	1

We note that pollution decreases slowly, yield is decreased significantly, this is due to the fact that the wheat been ravaged by locusts.

Variations of controls $u(.)$ and $v(.)$ depending on t for $\beta = 0$ and $\beta = 50$ respectively are shown in the Figure 4 and Figure 5 :

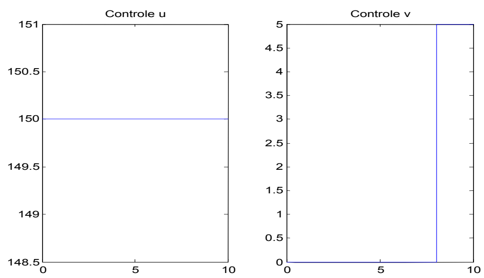


Figure 4. Optimal controls for $\beta = 0$.

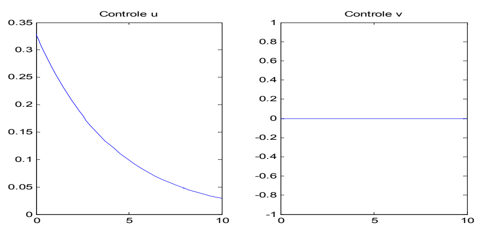


Figure 5. Optimal controls for $\beta = 50$.

It is visible from Figure 4 that when the farmer is only interested to increase the output, without taking into account the pollution, the optimal weight to be considered is of course "$\beta = 0$". Then $u(t) = 150qx/ha$, for all $t \in [0, 10]$ and quantities of insecticides are zero before the arrival of locusts i.e before May, $v(t) = 0$, but from May $v(t) = 5l/ha$.

Whereas, if he does not want to pollute the ground, he should use smaller quantities of fertilizers, for example, $\beta = 50$, the optimal control $u(.)$ decrease from $u = 0.34\ qx/ha$ at time $t = 0$ to $u = 0.05\ qx/ha$ at time $t = 10$. The quantities of insecticides in this case are zero (Figure 5).

Figure 6 shows the variations of pollution, cereal output and the number of locusts function of time t, for $\beta = 0$.

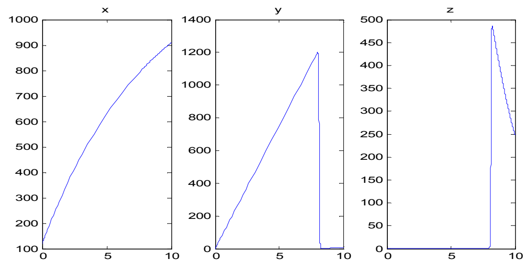


Figure 6. Optimal trajectories versus to t for $\beta = 0$.

We note that when $\beta = 0$, pollution increases, cereal output also increases until time $t = 8$ i.e until the coming of the locusts. At the same time, the number of locusts is maximum, then the curve decreases because of the maximum amount of insecticides $v(t) = 5l/ha$ which is applied.

If $\beta = 50$, in other words, taking into account the pollution, this last decreases according to t, cereal output increases until time $t = 8$ months, in May, then curve y decreases. But the curve representing the number of locusts does not decrease because no amount of insecticides is applied (see Figure 7).

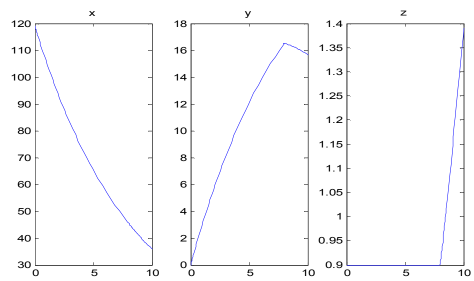


Figure 7. Optimal trajectories versus to t for $\beta = 50$.

5. Conclusion

In this work, we have modeled a practical problem in agriculture which is the optimization problem of a cereal production by introducing the constraint of the presence of locusts that are a real nuisance in Algeria. Controls resulting from the model are nonlinear. Different parameters of the model are identified using real-life data from the

National Institute of Plant Protection (INPV) located in the capital of Algeria Algiers. The theoretical resolution is done using the Pontryagin maximum principle. For the numerical resolution, we used the shooting method based on the Pontryagin maximum principle.

Our simulations show that the strategy of spreading fertilizers and insecticides can be improved in Algeria compared to what is done at present, so as to increase the rate of production and however minimizing the pollution effect.

Acknowledgements

The authors are very indebted to prof. Emmanuel Trélat (Univ. Pierre et Marie Curie, France) for his help and many fruitful discussions. Prof. E. Trélat and the second author are the PhD codirectors of the first author.

References

[1] Chitour, Y. Jean, F. and Trélat, E., Singular trajectories of driftless and control-affine systems. In Proceedings of 44th IEEE CDC-ECC'05, Séville, Spain, (2005).

[2] Chitour, Y. Jean, F. and Trélat, E., *Propriétées génériques des trajectoires singulières*, C. R. Acad. Sci. Paris, Ser. I, vol. 337, (2003).

[3] Chitour, Y., Jean, F., Trélat, E., Genericity results for singular curves. *J. Differential Geom.*, 73 (1), 45-73 (2006).

[4] Cressman, K., Dobson, H.M., *Diréctives sur le criquet pèlerin: Annexes*, Organisation des nations unies pour l'alimentation et l'agriculture (2001).

[5] Frankowska, H., Regularity of minimizers and of adjoint states for optimal control problems under state constraints. *J. Convex Analysis*, 13, 299-328 (2006).

[6] Stoer, J., Bulirsh, R., *Introduction to numerical analysis*. Springer-Verlag, (2002).

[7] *Ministry of Agriculture, Statistics* (2006).

[8] Murray, J.D., *Mathematical Biology: II. Spatial Models and Biomedical Applications*, (2003).

[9] Moussouni, N., Aidene, M., An Algorithm for optimization of cereal output. *Acta Applicandae Mathematicae*, 119, 113-127 (2012).

[10] Pontryaguin, L. et al., *Mathematical theory of optimal processes*. Moscou (1974).

[11] Sethi, S.P and Thompson, G.L. *Optimal Control Theory: Applications to Management Science and Economics*, Second Edition, Springer (2000).

[12] Trélat, E., *Contrôle optimal: théorie et applications*, Vuibert, Collection "Mathématiques Concrètes" (2005).

Nacima MOUSSOUNI *is currently a professor at the University Mouloud Mammeri of Tizi Ouzou in Algeria, which she joined in November 2002. She began her research in the laboratory, L2CSP at the University of Tizi Ouzou with Professor Mohamed AIDENE in the same laboratory, subsequently, this allowed us to publish an article entitled An Algorithm for Optimization of Cereal Output. Our works are based on real data of the Ministry of Agriculture in order to improve cereal production in Algeria.*

Aidene MOHAMED *is currently a professor at the University Mouloud Mammeri of Tizi Ouzou in Algeria. He is also director of laboratory L2CSP and Graduate School of Operations Research and Optimization.*

INSTRUCTIONS FOR AUTHORS

Aims and Scope

This journal shares the research carried out through different disciplines in regards to optimization, control and their applications.

The basic fields of this journal are linear, nonlinear, stochastic, parametric, discrete and dynamic programming; heuristic algorithms in optimization, control theory, game theory and their applications. Problems such as managerial decisions, time minimization, profit maximizations and other related topics are also shared in this journal.

Besides the research articles expository papers, which are hard to express or model, conference proceedings, book reviews and announcements are also welcome.

Journal Topics

- Applied Mathematics,
- Financial Mathematics,
- Control Theory,
- Game Theory,
- Fractional Calculus,
- Fractional Control,
- Modeling of Bio-systems for Optimization and Control,
- Linear Programming,
- Nonlinear Programming,
- Stochastic Programming,
- Parametric Programming,
- Conic Programming,
- Discrete Programming,
- Dynamic Programming,
- Optimization with Artificial Intelligence,
- Operational Research in Life and Human Sciences,
- Heuristic Algorithms in Optimization,
- Applications Related to Optimization on Engineering.

Submission of Manuscripts

All manuscripts should be submitted to An International Journal of Optimization and Control: Theories & Applications (IJOCTA) via the journal's online web-based submission system at the following link: http://ijocta.balikesir.edu.tr. Authors are responsible for verifying all files have uploaded correctly.

IMPORTANT: Please check whether you already have an account in the system before trying to create a new one. If you have reviewed or authored for the Journal in the past year it is likely that you will have created an account.

File types

Preferred first submission format for your manuscript is **Portable Document File (.pdf)** while preferred final submission formats are **Word (.docx)** and **LaTeX (.tex)**.

IMPORTANT: Manuscripts with heavy mathematical content will only be accepted as LaTeX files.

After you have logged in, click the "Submit a Manuscript" link in the menu bar. Enter data and answer questions as prompted. Click on the "Next" button on each screen to save your work and advance to the next screen. You will be prompted to upload your files: Click on the "Browse" button and locate the file on your computer. Select the description of the file in the drop down next to the Browse button. When you have selected all files you wish to upload, click the "Upload" button. Review your submission before sending to the Editors. Click the "Submit" button when you are done reviewing.

Each manuscript must be accompanied by a statement that it has not been published elsewhere and that it has not been submitted simultaneously for publication elsewhere.

English Editing

Papers must be in English. Oxford English Dictionary or American spelling is acceptable, provided usage is consistent within the manuscript. Manuscripts that are written in English that is ambiguous or incomprehensible, in the opinion of the Editor, will be returned to the authors with a request to resubmit once the language issues have been improved. This policy does not imply that all papers must be written in "perfect" English, whatever that may mean. Rather, the criteria require that the intended meaning of the authors must be clearly understandable, i.e., not obscured by language problems, by referees who have agreed to review the paper.

Presentation of Papers

Manuscript style

Use a standard font of the **11-point type: Times New Roman** is preferred. It is necessary to single line space your manuscript. Normally manuscripts are expected not to exceed 25 single-spaced pages including text, tables, figures and bibliography. All illustrations, figures, and tables are placed within the text at the appropriate points, rather than at the end.

During the submission process you must enter: (1) the full title, (2) names and affiliations of all authors and (3) the full address, including email, telephone and fax of the author who is to check the proofs. Supply a brief **biography** of each author at the end of the manuscript after references.

- Include the name(s) of any **sponsor(s)** of the research contained in the paper, along with **grant number(s)**.
- Enter an **abstract** of no more than 250 words for all articles.

Keywords. Authors should prepare no more than 5 keywords for their manuscript.

Maximum five **AMS Classification number** (http://www.ams.org/mathscinet/msc/msc2010.html) of the study should be specified after keywords.

Writing Abstracts

An abstract is a concise summary of the whole paper, not just the conclusions. The abstract should be no more than 250 words and convey the following:

1. An introduction to the work. This should be accessible by scientists in any field and express the necessity of the experiments executed.
2. Some scientific detail regarding the background to the problem.
3. A summary of the main result.
4. The implications of the result.
5. A broader perspective of the results, once again understandable across scientific disciplines.

It is crucial that the abstract conveys the importance of the work and be understandable without reference to the rest of the manuscript to a multidisciplinary audience. Abstracts should not contain any citation to other published works.

Reference Style

Reference citations in the text should be identified by numbers in square brackets. All references must be complete and accurate. If necessary, cite unpublished or personal work in the text but do not include it in the reference list. Online citations should include date of access. References should be listed in the following style:

Journal article

Author, A.A. and Author, B., Title of article. *Title of Journal*, Vol (issue), pages (Year).

Evans, W.A., Approaches to intelligent information retrieval. *Information Processing and Management*, 7 (2), 147–168 (1994).

Book

Author, A., *Title of book*. Publisher, Place of Publication (Year).

Mercer, P.A. and Smith, G., *Private Viewdata in the UK*. 2nd ed. Longman, London (1993).

Chapter

Author, A., Title of chapter. *In*: A. Editor and B. Editor, eds. *Title of book*. Publisher, Place of publication, pages (Year).

Bantz, C.R., Social dimensions of software development. *In*: J.A. Anderson, ed. *Annual review of software management and development*. CA: Sage, Newbury Park, 502–510 (1995).

Internet document

Author, A., Year. *Title of document* [online]. Source. Available from: URL [Accessed date Month Year].

Holland, M., 2004. *Guide to citing Internet sources* [online]. Poole, Bournemouth University. Available from: http://www.bournemouth.ac.uk/library/using/guide_to_citing_internet_sourc.html [Accessed 4 November 2004].

Newspaper article

Author, A. (or Title of Newspaper), Title of article. *Title of Newspaper*, day Month, page, column (Year).

Independent, Picking up the bills. *Independent*, 4 June, p. 28a (1992).

Thesis

Author, A., *Title of thesis*. Type of thesis (degree). Name of University (Year).

Agutter, A.J., *The linguistic significance of current British slang*. Thesis (PhD). Edinburgh University (1995).

Illustrations

Illustrations submitted (line drawings, halftones, photos, photomicrographs, etc.) should be clean originals or digital files. Digital files are recommended for highest quality reproduction and should follow these guidelines:

- 300 dpi or higher
- Sized to fit on journal page
- TIFF or JPEG format only
- Embedded in text files and submitted as separate files (if required)

Tables and Figures

Tables and figures (illustrations) should be embedded in the text at the appropriate points, rather than at the end. A short descriptive title should appear above each table with a clear legend and any footnotes suitably identified below.

Proofs

Page proofs are sent to the designated author using IJOCTA EProof system. They must be carefully checked and returned within 48 hours of receipt.

Offprints/Reprints

Each corresponding author of an article will receive a PDF file of the article via email. This file is for personal use only and may not be copied and disseminated in any form without prior written permission from IJOCTA.

Submission Preparation Checklist

As part of the submission process, authors are required to check off their submission's compliance with all of the following items, and submissions may be returned to authors that do not adhere to these guidelines.

1. The submission has not been previously published, nor is it before another journal for consideration (or an explanation has been provided in Comments to the Editor).
2. The submission file is in Portable Document Format (.pdf).
3. Where available, URLs for the references have been provided.
4. The text is single line spaced; uses a 11-point font; employs italics, rather than underlining (except with URL addresses); and all illustrations, figures, and tables are placed within the text at the appropriate points, rather than at the end.
5. The text adheres to the stylistic and bibliographic requirements outlined in the Author Guidelines, which is found in About the Journal.
6. Maximum five **AMS Classification number** (http://www.ams.org/mathscinet/msc/msc2010.html) of the study should be specified after keywords.
7. After the acceptance of manuscript (before copy editing), **Word (.docx)** or **LaTeX (.tex)** version of the paper should be submitted.

An International Journal of Optimization and Control: Theories & Applications

Volume: 6 Number: 1
January 2016

CONTENTS

INDEXES

IJOCTA is indexed/abstracted by:

- Cabell's Directories
- DOAJ
- EBSCO Databases
- Genamics JournalSeek
- Google Scholar
- Index Copernicus
- International Abstracts in Operations Research
- JournalTOCs
- Mathematical Reviews (MathSciNet)
- ProQuest
- Tubitak ULAKBIM
- Ulrich's Periodical Directory
- Zentralblatt Math

ISSN 2146-0957

9 772146 095004

www.ijocta.com
info@ijocta.com