산발적인 데이터를 이용한 HIV 변이모델의 파라미터 추정

Parameter Estimation of an HIV Model with Mutants using Sporadically Sampled Data

김 석 균, 김 정 수, 윤 태 웅*

(Seok-Kyoon Kim¹, Jung-Su Kim², and Tae-Woong Yoon¹)

¹Korea University

²Seoul National University of Science and Technology

Abstract: The HIV (Human Immunodeficiency Virus) causes AIDS (Acquired Immune Deficiency Syndrome). The process of infection and mutation by HIV can be described by a 3rd order state equation. For this HIV model that includes the dynamics of the mutant virus, we present a parameter estimation scheme using two state variables sporadically measured, out of the three, by employing a genetic algorithm. It is assumed that these non-uniformly sampled measurements are subject to random noises. The effectiveness of the proposed parameter estimation is demonstrated by simulations. In addition, the estimated parameters are used to analyze the equilibrium points of the HIV model, and the results are shown to be consistent with those previously obtained.

Keywords: HIV model, mutant virus, parameter estimation, non-uniformly sampled output data, genetic algorithm

I. 서론

1981년 처음 알려진 후천성 면역 결핍 증후군(AIDS)은 RNA 바이러스인 HIV로 말미암아 발병하는 심각한 질병이다. 따라서 HIV의 증식과정을 표현하는 수학적 모델을 얻고 그 동특성을 분석하는 일은 AIDS의 진행 과정을 이해하고 그 치료전략을 고안하는 데 중요한 단서를 제공한다 [1,2]. HIV 모델의 파라미터를 정확하게 추정하는 것은 그래서 중요하다. [3,4]에서는 변이 바이러스의 동특성을 고려하지 않은 모델을 이용하여 파라미터 추정 알고리즘을 제안했다. 하지만, 환자에게 약물을 투여할 때 발생하는 변이 바이러스의 영향이 심각하다는 점을 헤아리면, 이 연구 결과는 현실적으로 한계가 있다. 이 문제를 보완하기 위해 변이 바이러스 동특성을 고려하고 더불어 측정 가능한 상태 변수인 건강한 CD4+T 세포와 감염된 CD4+T 세포 농도만을 사용한 파라미터 추정 알고리즘이 개발되었다 [5]. 만약, 측정 가능한 상태 변수의 데이터가 연속적이거나 연속적이라 가정할 수 있을 만큼 충분하다면, [5]에서 제시한 파라미터 추정 알고리즘을 이용하여 파라미터를 추정할 수 있을 것이다. 그러나 환자들에게 직접 얻는 데이터는 연속적이라 가정할 수 있을 만큼 충분하지 않고 각 데이터를 얻은 시간 간격도 일정하지 않다. 또한, 데이터 측정 과정에서 생기는 측정 오차도 무시할 수 없다. 따라서 [5]에서 제

안된 방법으로 파라미터를 추정하는 것에도 비현실적인 면이 있다. 실험으로 얻은 데이터를 이용하여 모델의 파라미터를 추정하는 방법으로는 멀티플 슈팅(multiple shooting) [10], 비선형 최소제곱법 등이 있다. 그러나 이 방법들도 모델과 문제 크기에 따라 국소 최소점으로 수렴하는 현상 등의 문제가 있어 성공적으로 파라미터를 추정하기에는 한계가 있다.

이 논문에서는 균일하지 않은 시차를 두고 산발적으로 측정된 출력 데이터를 이용해 HIV 모델의 파라미터를 추정하는 기법을 제안한다. 이를 위해 파라미터 추정 문제를, 출력 변수에 대한 추정 오차를 최소화하는 최적화 문제로 변환한다. 그리고 유전 알고리즘(genetic algorithm) [7,8,11]을 사용하여 변이를 고려한 HIV 모델의 파라미터를 추정한다. 마지막으로 추정된 파라미터를 이용하여 모델의 평형점을 분석하여 추정한 모델이 실제 모델의 동특성을 잘 반영함을 보인다.

II-1에서는 HIV 변이 모델을 소개한다. II-2에서는 파라미터 추정을 위한 최적화 문제를 정의하고, 이를 푸는 알고리즘을 제안한다. 그리고 이 알고리즘에 대한 모의실험 결과를 III 장에서 보인다. IV 장에서는 추정한 파라미터를 이용하여 모델의 평형점을 분석함으로써 추정한 모델도 원래 모델과 같은 성질을 가지고 있음을 보인다. 마지막으로 V 장에서 결론을 맺는다.

II. 산발적 출력 데이터를 이용한 HIV 모델의 파라미터 추정

여기서는 파라미터 추정을 위한 최적화 문제를 정의하고 이를 푸는 알고리즘을 제안한다. 이를 위해 II-1에서 변이

* 책임저자(Corresponding Author)

논문접수: 2011. 5. 20., 수정: 2011. 6. 5., 채택확정: 2011. 6. 20.

김석균: 고려대학교 전기전자전파공학과(lotus45@korea.ac.kr)

김정수: 서울과학기술대학교 제어측정공학과(jungsu@seoultech.ac.kr)

윤태웅: 고려대학교 전기전자전파공학부(twy@korea.ac.kr)

※ 본 논문은 연구재단 기초연구사업-일반연구지원사업 (과제번호: 2010-0016589, 2010-0006131)에서 지원하여 연구하였음.

동특성을 포함한 HIV 모델을 고려한다. 그리고 II-2에서 파라미터 추정문제를 최적화 문제로 표현하고 그 해법으로 혼합유전 알고리즘(hybrid genetic algorithm)을 도입한다.

1. 모델

HIV 변이 모델은, 건강한 CD4+T세포와 HIV에 감염된 CD4+T 세포의 농도, 그리고 변이 HIV에 감염된 세포의 농도를 상태변수로 하며, 다음과 같이 표현된다[2].

$$\begin{aligned}\dot{x} &= \lambda - dx - (1-e)\beta'xy - \beta_m'xy_m \\ \dot{y} &= (1-\mu)(1-e)\beta'xy - ay \\ \dot{y}_m &= \mu(1-e)\beta'xy + \beta_m'xy_m - ay_m\end{aligned}\quad (1)$$

여기서 x 는 정상 CD4+T 세포의 농도, y 는 HIV에 감염된 CD4+T 세포의 농도, y_m 은 변이 바이러스에 감염된 CD4+T 세포의 농도다. 그리고 λ 는 건강한 CD4+T 세포가 생성되어 x 가 커지는 속도, βxy 는 HIV가 건강한 CD4+T 세포를 감염시켜 x 는 작아지고 y 가 커지는 속도, e 는 약물의 효과이며 $0 \leq e < 1$ 이다. μ 는 변이 때문에 변이 바이러스에 감염된 CD4+T세포가 생성될 확률을 나타낸다. $\beta_m xy_m$ 은 변이 바이러스가 건강한 CD4+T 세포를 감염시켜 x 가 작아지고 y_m 이 커지는 속도, dx, ay, ay_m 은 각 세포가 자멸하여 x, y, y_m 이 감소하는 속도를 나타낸다.

모델의 동특성 해석을 위해서는 측정할 수 있는 상태변수들을 이용해 모델 (1)의 파라미터를 추정해야 한다. 이 논문에서는 상태 변수 x 와 y 가 측정 가능하고[6] 이 데이터들이 산발적으로 존재한다는 현실적인 가정 아래, 모델의 파라미터를 추정하고자 한다. 이를 위해 다음과 같이 유전 알고리즘을 이용한다.

2. 최적화 문제 설정과 혼합 유전 알고리즘

일반적으로 환자에게서 건강한 세포 농도 x 와 감염된 세포 농도 y 를 일정한 주기로 측정하지는 않는다. 또한, 이 데이터들을 측정하는 빈도도 높지 않다. 이 절에서는 이렇게 균일하지 않고 산발적으로 측정된 출력 데이터를 이용해 모델 (1)의 파라미터를 추정하는 문제를 최적화 문제로 표현한다. 이를 위해 다음과 같은 비용함수를 고려한다.

$$J(\theta) := \sum_{i=1}^N [(x_{ex}(t_i) - x(t_i, \theta))^2 + (y_{ex}(t_i) - y(t_i, \theta))^2] \quad (2)$$

여기서 $x_{ex}(t_i)$ 와 $y_{ex}(t_i)$ 는 시간 t_i , $i = 1, 2, \dots, N$, 에서 실제로 측정된 데이터이고 $x(t_i, \theta)$ 와 $y(t_i, \theta)$ 는 t_i 에서 모델 (1)과 파라미터 $\theta := [a \ \beta' \ \lambda \ d \ \beta_m' \ \mu]^T$ 을 이용하여 얻은 모델의 출력값이다. 출력 데이터는 일정한 주기로 측정하지 않으므로 측정 시간을 나타내는 t_i 의 간격은 일정하지 않다고 가정한다. 또 환자에게는 아직 약물을 투여하지 않았다고(즉, $e = 0$) 가정한다.

미지의 파라미터를 찾는 한 가지 방법으로 비용 함수 (2)의 값을 최소로 만드는 최적 파라미터 θ^* 를 찾는 것을 생각할 수 있다. 즉, 모델의 출력값이 실제 측정값과 최대한 비슷하도록 하는 파라미터를 찾는 것이다. 따라서 파라미터 추정 문제를 미지의 파라미터 θ 에 대해 비용함수 (2)를 최소화하는 문제로 볼 수 있다. 이러한 파라미터 추정 문제를

정의하면 다음과 같다.

$$\begin{aligned}\min_{\theta} J(\theta) \\ \text{subject to (1)}.\end{aligned}\quad (3)$$

이 최적화 문제는 볼록 문제(convex problem)가 아니다. 또한, 일반적으로 많은 국소 최소점이 존재하기 때문에 볼록 성질을 이용한 최적화 알고리즘은 최적의 파라미터 θ 를 찾는 데 어려움이 있다. 이 논문에서는 이러한 조건에서 최적화 문제 (3)의 해를 효율적으로 찾기 위해 유전 알고리즘을 이용한다.

진화 연산에서는 세대를 형성하는 개체들 가운데 환경에 대한 적합도가 높은 개체들이 더 높은 확률로 선택(selection)된다. 그리고 다음 세대의 개체 집합은 교차(crossover)와 변이(mutation) 그리고 대체(replacement)를 통해 형성된다[8]. 이러한 개체들이 최적해의 후보들이며, 개체 집합의 초기 원소 수를 개체 집합의 크기(population size)라 한다. 유전 알고리즘은 최적해를 찾는 과정에서 다음과 같은 연산자를 이용한다[8].

① 선택(selection)

현재 세대에서 다음 세대 개체가 될 수 있는 후보를 선택하는 연산이다. 여러 가지 선택 방법이 있으나, 본 논문에서는 룰렛휠(roulette wheel) 방법을 사용했다.

② 교차(crossover)

다음 세대 개체는 전 세대의 두 개체로부터 서로 다른 위치의 유전자에 의해 생성된다. 이때 전 세대 두 개체의 일부분이 특정 위치를 기준으로 서로 바뀌어 결합되는 경우가 있다. 이 현상을 구현한 연산자를 교차라고 한다.

③ 변이(mutation)

다음 세대 개체를 생성할 때, 전 세대의 두 개체를 이용한 교배 뿐 아니라 낮은 확률로 한 개체가 변이를 일으켜 다음 세대의 개체가 되는 경우가 있다. 이 현상을 구현한 연산자를 변이라고 한다.

④ 대체(replacement)

교차와 변이 연산을 이용하여 다음 세대의 개체를 생성하기도 하지만, 현재 세대의 열등한 개체를 우수한 개체로 대체하여 다음 세대의 개체를 생성하기도 한다. 이 현상을 반영한 것이 대체 연산이다.

유전 알고리즘은 주어진 최적화 문제의 전역적 최소점을 찾을 확률이 높지만, 항상 전역해를 찾는다고 보장할 수는 없다. 그리고 알고리즘의 초기 조건에 따라서 결과가 달라질 수도 있고 해를 찾는 데 많은 시간이 걸리기도 한다. 반면, 국소 최적화 방법들(local optimization methods)은 확률에 의존하지 않고, 유전 알고리즘보다 더 빨리 해를 찾을 수 있다. 그러나 전역 최소점에 수렴하지 못하고 국소 최소점에 머무는 경우가 발생할 수 있다. 따라서 이 논문에서는 유전 알고리즘의 장점과 국소 최적화 알고리즘의 장점을 함께 지니는 혼합 유전알고리즘(hybrid genetic algorithm)을 이용한다[7,8]. 이 논문에서는 국소 최적화 알고리즘으로 SQP(Sequential Quadratic Programming)를 사용하였다[9]. 혼합 유전 알고리즘의 흐름도는 그림 1과 같다.

그림 1. 혼합 유전 알고리즘의 흐름도[8].

Fig. 1. The flowchart of the hybrid genetic algorithm [8].

위 알고리즘에서 W_{GA}^k 와 W_{Local}^k 는 국소 최적화 알고리즘의 선택 확률 P_{GA}^k 를 결정하기 위한 초기 파라미터다. 유전 알고리즘이 전역적 최소점을 찾을 확률이 높기에 첫 단계에서는 유전 알고리즘을 선택하는 것이 유리하다. 따라서 초기 W_{GA}^k 값은 초기 W_{Local}^k 값보다 훨씬 더 크다. 그리고 개체 집합의 크기를 정하고, 초기 개체 집합 P 를 생성한다. 알고리즘 각 단계(k)마다 유전 알고리즘 혹은 국소 최적화 알고리즘은 최적해 θ_{best}^k 와 비용함수 값 J_{best}^k 를 계산하고 이를 이용하여 두 알고리즘이 비용함수 최소화해 공헌한 정도(W_{GA}^k , W_{Local}^k)를 구한다. 이들을 이용하여 두 알고리즘의 선택 확률을 정하는 P_{GA}^k 또한 갱신하여 기여도에 따라 두 알고리즘 중 어느 것을 선택할지를 결정한다. 이와 같이 두 알고리즘의 단점을 보완해 혼합 유전 알고리즘의 성능을 높이하고자 하였다. 다음 절에서는 이 알고리즘을 이용해 HIV 변이 모델의 파라미터를 추정한다.

III. 유전 알고리즘을 이용한 파라미터 추정

여기서는 위에서 설명한 유전 알고리즘을 이용해 최적화 문제 (3)을 풀고 모델 (1)의 파라미터를 추정한다. 모의실험에 사용한 모델의 실제 파라미터는 표 1과 같다. 그리고 파라미터 추정을 위한 출력 데이터는 다음과 같이 샘플링되었다고 가정한다.

$x(t)$ 의 측정 데이터:

$$\begin{aligned} x_{ex}(0) &= 1000, x_{ex}(0.2) = 932, x_{ex}(0.5) = 1.3, \\ x_{ex}(0.7) &= 0.4, x_{ex}(1) = 0.4, x_{ex}(2) = 0.7, \\ x_{ex}(2.3) &= 0.8, x_{ex}(2.9) = 1 \end{aligned}$$

$y(t)$ 의 측정 데이터:

$$\begin{aligned} y_{ex}(0) &= 0.1, y_{ex}(0.2) = 66.3, y_{ex}(0.5) = 897.1, \\ y_{ex}(0.7) &= 814.6, y_{ex}(1) = 703.8, y_{ex}(2) = 434.6, \\ y_{ex}(2.3) &= 376.7, y_{ex}(2.9) = 284 \end{aligned}$$

표 1. 파라미터 참값.

Table 1. The true parameter values.

a	0.5	β'	0.0333	λ	10
d	0.01	β_m'	0.0167	μ	0.0001

이 데이터는 건강한 상태 근방의 초기조건($x(0) = 1000$, $y(0) = 0.1$, $y_m(0) = 0$)에서 모델 (1)과 표 1을 이용해 계산한 궤적 $x(t), y(t), y_m(t)$ 을 0, 0.2, 0.5, 0.7, 1, 2, 2.3, 2.9 초에 샘플링한 값들이다.

이 데이터들을 실제 실험에서 얻은 측정치라 가정하고 II에서 설명한 유전 알고리즘을 이용해 모델 (1)의 파라미터를 추정한다. 이 논문에서 사용한 유전 알고리즘의 파라미터는 다음과 같다.

- 초기 개체수(Initial population size): 150
- 선택(Selection) 확률: 0.6
- 교배(Crossover) 확률: 0.9
- 변이(Mutation) 확률: 0.005

그리고 유전 알고리즘에서 사용한 모델의 초기치는 다음과 같다.

$$x(0) = x_{ex}(0) = 1000, y(0) = y_{ex}(0) = 0.1, y_m(0) = 0$$

여기서 $x(0)$ 와 $y(0)$ 는 측정해서 아는 값인 $x_{ex}(0)$ 과 $y_{ex}(0)$ 과 같게 정했고, $y_m(0)$ 은 아직 바이러스 변이가 일어나지 않았다는 가정 아래 0으로 놓았다.

출력 데이터에 측정 오차가 없는 경우, 최적화를 통해 얻은 추정 파라미터 $\theta^* = [a^* \ \beta'^* \ \lambda^* \ d^* \ \beta_m'^* \ \mu^*]^T$ 는 다음과 같다.

$$a^* = 0.5, \beta'^* = 0.0333, \lambda^* = 10, d^* = 0.01, \\ \beta_m'^* = 0.0157, \mu^* = 0.103 \times 10^{-3}$$

추정 파라미터들이 참값과 거의 일치함을 알 수 있다.

그림 2. 출력 데이터에 측정 오차가 없는 경우 추정 파라미터를 이용하여 추정된 상태변수.

Fig. 2. The estimated states using the estimated parameters in the absence of the measurement error at the output.

그림 3. 비용함수 J 의 변화 추이.

Fig. 3. The cost versus iterations.

그리고 이 파라미터 추정값을 이용하여 추정된 상태변수는 모델(1)을 이용하여 얻을 수 있으며 그림 2와 같다. 또한, 비용함수 (2)의 변화 추이는 그림 3과 같다.

그림 2에서 작은 원들은 출력의 측정치 $x_{ex}(t_i)$ 와 $y_{ex}(t_i)$ 들을 나타내며 실선은 추정 파라미터와 모델 (1)을 이용하여 계산한 $x(t_i, \theta^*)$ 와 $y(t_i, \theta^*)$ 의 궤적이다. 그림에서 알 수 있듯이 추정 파라미터를 이용한 모델의 출력이 실제 출력값과 거의 같음을 알 수 있다. 그림 3에서 수평축은 알고리즘 반복 횟수를 의미하고 작은 원들은 각 반복 횟수에서의 비용함수 값을 나타낸다. 점선은 비용 함수값의 변화 추이다. 예상대로 반복 횟수가 증가함에 따라 비용함수 값이 점차 감소함을 알 수 있다.

이로써 측정 오차가 없는 경우 유전 알고리즘 기반의 최적화 기법을 통해 HIV 모델 (1)의 파라미터를 성공적으로 추정하였음을 알 수 있다. 그런데 환자 데이터를 얻는 실제 과정에서는 측정 오차가 생길 수밖에 없다. 즉, 출력의 측정치는 실제 출력값과 측정 오차의 합이 된다. 이러한 상황을 고려하기 위하여 출력 데이터에 측정 오차가 포함된 경우를 고려한다. 측정 오차를 반영하기 위해 실험 데이터를 다음과 같이 생성한다.

$x(t)$ 와 $y(t)$ 의 실험 데이터:

$$\overline{x_{ex}}(t_i) = (1 + \alpha d) x_{ex}(t_i)$$

$$\overline{y_{ex}}(t_i) = (1 + \alpha d) y_{ex}(t_i), \quad t_i = 1, 2, \dots, 8$$

여기서 $x_{ex}(t_i)$ 와 $y_{ex}(t_i)$ 는 이전 모의실험에서 사용한 측정 오차가 없는 출력 측정값이고 $d \in [-1, 1]$ 는 확률적으로 결정되는 임의의 값이다. 그리고 $\alpha \in [0, 1]$ 이며 측정 오차의 강도 또는 신호 대 잡음비(signal-to-noise ratio)를 의미한다. 예를 들어 $\alpha = 0$ 이면 측정 데이터는 오차가 없는 경우이고 $\alpha = 0.1$ 이면 실험 데이터에 10%의 측정 오차가 있음을 의미한다. 이후의 파라미터 추정에서는 실제 출력값 대비 10%, 20%, 30%의 측정 오차가 있는 경우를 고려한다. 이 경우 알고리즘에 사용되는 측정치만 $x_{ex}(t_i)$ 와 $y_{ex}(t_i)$ 에서 $\overline{x_{ex}}(t_i)$ 와 $\overline{y_{ex}}(t_i)$ 으로 바뀌고, 그 밖의 모든 조건은

표 2. 측정 오차에 따른 파라미터 추정 결과.

Table 2. The estimated parameters versus different measurement errors strengths.

파라미터 추정 결과	
10% 측정오차	$a^* = 0.47, \beta'^* = 0.0325, \lambda^* = 15.2,$ $d^* = 0.0008, \beta_m'^* = 0.0254,$ $\mu^* = 0.155 \times 10^{-3}$
20% 측정오차	$a^* = 0.66, \beta'^* = 0.034, \lambda^* = 15.2,$ $d^* = 0.0152, \beta_m'^* = 0.0253,$ $\mu^* = 0.148 \times 10^{-3}$
30% 측정오차	$a^* = 0.35, \beta'^* = 0.035, \lambda^* = 0.8,$ $d^* = 0.0008, \beta_m'^* = 0.0013,$ $\mu^* = 0.76 \times 10^{-5}$

그림 4. 측정오차에 따른 추정 오차와 비용함수의 변화.

Fig. 4. The estimation errors and the cost versus different measurement errors strengths.

이전의 모의실험에서와 같다고 가정한다. 모의실험 결과는 표 2와 같다. 그리고 측정 오차의 양에 따른 추정 오차와 비용함수의 변화를 그림 4에 나타낸다.

그림 4에서 $\theta_1 = a, \theta_2 = \beta', \theta_3 = \lambda, \theta_4 = d, \theta_5 = \beta_m', \theta_6 = \mu$ 이고 측정 오차가 커질수록 각 파라미터 추정 오차와 최적 비용함수 값이 커지는 경향을 볼 수 있다. 그리고 이 파라미터 추정치들과 모델을 이용하여 추정된 상태변수의 궤적은 그림 5와 같다.

그림 5. 출력 데이터에 측정 오차가 있는 경우 추정 파라미터를 이용하여 추정된 상태변수.

Fig. 5. The estimated states using the estimated parameters in the presence of the different measurement errors at the output.

위 그림 5에서 작은 원들은 오차가 포함된 출력의 측정치인 $x_{ex}(t_i)$ 와 $y_{ex}(t_i)$ 를 나타내며 선들은 추정된 파라미터를 이용하여 계산한 $x(t_i, \theta^*)$ 와 $y(t_i, \theta^*)$ 를 나타낸다. 파라미터 추정 오차가 있는데도 모델의 출력이 실제 출력과 비슷함을 알 수 있다. 이로부터 출력에 측정 오차가 어느 정도 있더라도 추정된 모델이 실제 HIV 모델과 동특성이 유사함을 알 수 있다.

추정된 파라미터의 값이 실제 값과 비슷하다고 해도 실제 비선형 HIV 모델과 추정된 모델 사이에 질적인 차이가 생길 가능성도 존재한다. 다음 절에서는 그러한 현상이 발생하는지 알아보기 위해 추정된 파라미터로 모델의 평형점을 분석한다.

IV. 추정된 모델의 평형점 분석

이번 절에서는 III에서 추정된 파라미터들을 이용하여 약물 효과 e 의 변화에 따른 HIV 모델의 평형점을 분석한다. HIV 변이 모델의 평형점은 모두 세 개며 다음과 같다.

$$1) x_{10} = \frac{\lambda}{d}, y_{10} = 0, y_{m10} = 0$$

$$2) x_{20} = \frac{a}{(1-e)(1-\mu)\beta'}, y_{20} = \frac{\lambda - dx_{20} - \beta_m' x_{20} y_{m20}}{(1-e)\beta' x_{20}},$$

$$y_{m20} = \frac{\mu(\lambda - dx_{20})}{(1-\mu)\beta_m' x_{20} + a}$$

$$3) x_{30} = \frac{a}{\beta_m'}, y_{30} = 0, y_{m30} = \frac{\lambda - dx_{30}}{\beta_m' x_{30}}$$

위 평형점들은 모두 파라미터의 함수임을 알 수 있다. 실험 데이터가 측정 오차를 포함하지 않는 경우와 20%의 측정 오차를 포함하는 경우로 나누어 파라미터 추정치를 이용한 모델 (1)의 평형점을 분석한다. 각각의 경우 약물 효과 e 의 변화에 따라 평형점을 분석한 결과는 그림 6과 그림 7에 나타낸다.

그림 6. 측정 오차가 없는 경우 추정 파라미터를 이용하여 평형점을 분석한 결과.

Fig. 6. The equilibrium analysis using the estimated parameters in the absence of the measurement error.

그림 7. 20%의 측정 오차가 있는 경우 추정 파라미터를 이용하여 평형점을 분석한 결과.

Fig. 7. The equilibrium analysis using the estimated parameters in the presence of the 20% measurement error.

그림 6과 그림 7에서 수직축은 건강한 세포 농도 값이고 수평축은 e 값을 나타낸다. 그리고 괄호 안의 수치는 파라미터 참값을 이용하여 계산한 결과를 나타낸다. 위 그림 6처럼 평형점 1 ($x_{10} = \frac{\lambda^*}{d^*}, y_{10} = 0, y_{m10} = 0$)은 모든 e 에 대해서 불안정하다. 이 평형점에 대해 선형화한 시스템이 불안정하기 때문이다. 그리고 건강한 세포 농도 수치 x_{10} 가 큰 값이고, 감염된 세포 농도 y_{10} 와 변이 세포 농도 y_{m10} 가 0이므로 생물학적으로 건강한 상태를 의미한다. 따라서 HIV가 인체에 발생하면(즉, y_{10} 이 약간이라도 증가하면) 현재 상태를 유지하지 못하고 건강한 세포 농도는 급격하게 줄고 감염된 세포 농도와 변이 세포 농도는 함께 증가한다.

평형점 2 ($x_{20} = \frac{a^*}{(1-e)(1-\mu^*)\beta_m'^*}, y_{20} = \frac{\lambda^* - dx_{20} - \beta_m'^* x_{20} y_{m20}}{(1-e)\beta_m'^* x_{20}}$, $y_{m20} = \frac{\mu^*(\lambda^* - dx_{20})}{(1-\mu^*)\beta_m'^* x_{20} + a^*}$)는 $e < e' (= 1 - \frac{\beta_m'^*}{(1-\mu)\beta^*})$ 이면 안정하고 $e > e'$ 이면 사라진다. 이 평형점이 안정하면 건강한 세포 농도 수치 x_{20} 가 x_{10} 의 10% 이하이므로 생물학적

으로 AIDS 상태를 의미한다. ($e < e'$ 이면 초기치가 평형점 1 근처에 있더라도 상태 궤적은 곧 평형점 2를 향해 움직이고 건강한 세포 농도 $x(t)$ 는 급격히 줄어든다.)

평형점 3 ($x_{30} = \frac{a^*}{\beta_m'^*}, y_{30} = 0, y_{m30} = \frac{\lambda^* - d^* x_{30}}{\beta_m'^* x_{30}}$)은 낮

은 약물 효과 $e < e'$ 에 대해서는 불안정하다. 하지만 $e > e'$ 이면 안정하다. 그리고 건강한 세포 농도 수치 x_{30} 가 x_{10} 의 10% 이하이고 감염된 세포 농도 y_{30} 은 0이지만 변이 세포 농도 y_{m30} 은 y_{m20} 보다 상당히 크다. 따라서 이 평형점이 안정하면 생물학적으로 AIDS 상태를 의미한다. ($e > e'$ 이면 초기치가 평형점 1 근처에 있더라도 상태 궤적은 곧 평형점 3을 향해 움직이고 건강한 세포 농도 $x(t)$ 는 줄어들고 $y(t)$ 는 0을 향하며 변이세포 농도 $y_m(t)$ 는 급격히 증가한다.)

위 분석을 종합하면 약물 효과가 작을 때는($e < e'$), 평형점 2가 안정하여 AIDS 상태가 존재하게 된다. 여기서 벗어나려고 약물 효과를 높이면($e > e'$), 평형점 2는 사라지지만 대신 평형점 3이 안정하게 돼 여전히 AIDS 상태가 존재하게 된다. 이는 파라미터의 참값을 이용해 평형점을 분석한 결과와 일치한다.

그림 7은 그림 6에 비해 분기(bifurcation)가 발생하는 지점 e' 이 다르다. 하지만, 이 경우에도 평형점들의 성질은 질적으로 같음을 알 수 있다. 이 결과는 이 논문에서 제안한 유전 알고리즘 기반 최적화 기법이 측정 오차가 있을 때에도 효과적임을 나타낸다.

V. 결론

이 논문에서는 균일하지 않은 시차를 두고 산발적으로 측정된 출력 데이터를 이용해 HIV 모델의 파라미터를 추정하는 알고리즘을 제안했다. 이를 위해 우선 파라미터 추정 문제를 실제 출력과 모델 출력의 거리를 최소화하는 최적화 문제로 정식화하였다. 그리고 블록 문제가 아닌 이 문제를 효율적으로 풀기 위해 유전 알고리즘을 사용하였다. 모의실험을 통해 측정 오차가 있는 경우에도 모델의 파라미터를 잘 추정함을 알 수 있었다. 추정된 파라미터가 실제 값과 양적으로 비슷할 뿐만 아니라, 평형점 분석을 통해 추정된 파라미터를 이용한 모델이 실제 모델과 질적으로 특성이 비슷함을 확인하였다. 이로써, 데이터가 산발적으로 샘플링 된 데다 측정오차까지 포함된, 아주 실제적인 상황에서도 제안된 추정 알고리즘이 효과적일 수 있음을 추론할 수 있다.

참고문헌

- [1] M. A. Nowak and C. R. Bangham, "Population dynamics of immune responses to persistent viruses," *Science*, vol. 272, no. 5258, pp. 74-79, 1996.
- [2] M. A. Nowak, S. Bonhoeffer, G. M. Shaw, and R. M. May, "Anti-viral drug treatment: dynamics of resistance in free virus and infected cell populations," *Journal of Theoretical Biology*, vol. 184, pp. 203-217, 1997.

- [3] X. Xia, "Estimation of HIV/AIDS parameters" *Automatica*, vol. 39, pp. 1983-1988, 2003.
- [4] X. Xia and C. H. Moog, "Identifiability of nonlinear systems with applications to HIV/AIDS models," *IEEE Transaction on Automatic Control*, vol. 48, no. 2, pp. 330-336, 2003.
- [5] S.-K. Kim, C.-N. Kim, and T.-W. Yoon, "Output feedback parameter estimation for a HIV model with mutants," *Proc. of Modelling, Identification, and Control 2011*, pp. 399-405, 2011.
- [6] R. A. Filter, X. Xia, and C. M. Gray "Dynamic HIV/AIDS parameter estimation with application to a vaccine readiness study in southern africa," *IEEE Transaction on Biomedical Engineering*, vol. 52, no. 5, pp. 784-791, 2005.
- [7] C. Houck, J. Joins, and M. Kay, "A genetic algorithm for function optimization: a matlab implementation," NCSU-IE, TR95-09, 1995.
- [8] P. P. Menon, J. Kim, D. G. Bates, and I. Postlethwaite, "Clearance of nonlinear flight control laws using hybrid evolutionary optimization," *IEEE Transaction on Evolutionary Computation*, vol. 10, no. 6, pp. 689-699, 2006.
- [9] "Optimization Toolbox User's Guide," ver.2, The MathWorks, 2000.
- [10] H. Bock, "Recent advances in parameter identification for ordinary differential equations," *Progress in Scientific Computing*, vol. 2, pp. 95-121, 1983.
- [11] K.-R. Cho, S.-W. Baek, and D.-W. Lee, "Fitness change of mission scheduling algorithm using genetic theory ac-

cording to the control constants," *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 16, no. 6, June 2010.

김 석 군

2004년 서울산업대학교 매체공학과 졸업. 2004년~현재 고려대학교 전기공학과 석·박사 통합과정 재학중. 관심분야는 비선형 제어, 적응 제어, 전력계통 안정화, 분산제어, 모폴로지 기반 최단 경로 생성.

김 정 수

1998년 고려대학교 전기공학과 학사. 2000년, 2005년 동 대학원 석사, 박사. 2005년~2008년, 서울대학교 자동화기술공동 연구소, 독일 Stuttgart 대학, 영국 Leicester대학 박사후 연구원. 현재 서울과학기술대학교 제어계측공학과 조교수. 관심분야는 비선형 제어, 모델예측제어, 최적화 응용, 멀티에이전트 시스템, 시스템 생물학 등.

윤 태 응

1984년 서울대학교 제어계측공학과 학사. 1986년 동 대학원 석사. 1994년 옥스퍼드대학교 박사. 1995년~현재 고려대학교 전기전자전파공학부 교수. 연구분야는 시스템·제어 이론과 응용. 관심분야는 과학적 사고와 소통.