(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년11월16일
(11) 등록번호 10-2326733
(24) 등록일자 2021년11월10일

(51) 국제특허분류(Int. Cl.)
G05B 13/02 (2006.01) *G05B 13/04* (2006.01)
G06F 17/11 (2006.01)
(52) CPC특허분류
G05B 13/021 (2013.01)
G05B 13/041 (2013.01)
(21) 출원번호 10-2020-7014310
(22) 출원일자(국제) 2018년08월10일
심사청구일자 2021년05월24일
(85) 번역문제출일자 2020년05월19일
(65) 공개번호 10-2020-0081407
(43) 공개일자 2020년07월07일
(86) 국제출원번호 PCT/EP2018/071753
(87) 국제공개번호 WO 2019/076512
국제공개일자 2019년04월25일
(30) 우선권주장
102017218811.1 2017년10월20일 독일(DE)
(56) 선행기술조사문헌
US20090271340 A1

(73) 특허권자
로베르트 보쉬 게엠베하
독일 데-70442 슈투트가르트 포스트파호 30 02 20
(72) 발명자
비쇼프, 바슈티안
독일 73734 에쓰링엔 뭇첸라이스슈트라쎄 14
비노그라드스카, 율리아
독일 70469 슈트트가르트 키프해우저슈트라쎄 67
페터스, 얀
독일 64342 제하임-유겐하임 (오테 말헨) 임 슈타인빙에르트 7
(74) 대리인
조영현

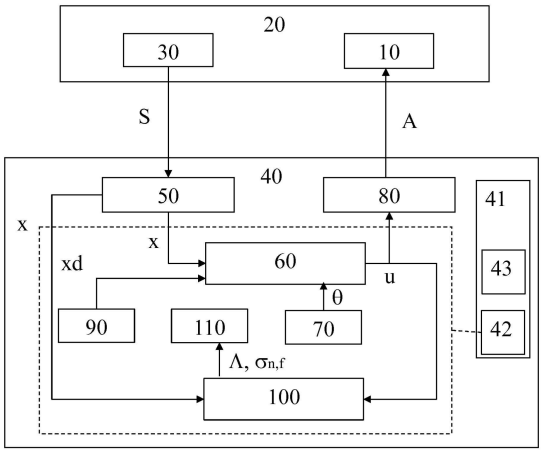
전체 청구항 수 : 총 13 항 심사관 : 권보람

(54) 발명의 명칭 **엑츄에이터 조절 시스템을 작동시키기 위한 방법 및 장치, 컴퓨터 프로그램 및 기계 판독가능한 저장 매체**

(57) 요 약

본 발명은 엑츄에이터(20)의 조절 변수(x)를 미리 정의가능한 목표 변수(x)로 조절하도록 설계되는 엑츄에이터 조절 시스템(45)을 작동시키기 위한 방법에 있어서, 상기 엑츄에이터 조절 시스템(45)은 조절 전략(π)을 특성화하는 변수(Θ)의 함수로서 교정 변수(u)를 생성하고 또한 이 교정 변수(u)의 함수로서 상기 엑츄에이터(20)를 제어하도록 설계되고, 상기 조절 전략(π)을 특성화하는 변수(Θ)는 가치 함수(V*)의 함수로서 결정되는, 방법에 관한 것이다.

대 표 도 - 도1



(52) CPC특허분류
G06F 17/11 (2013.01)

명 세 서

청구범위

청구항 1

엑츄에이터(20)의 조절 변수(x)를 미리 정의가능한 목표 변수(xd)로 조절하도록 설정되는 엑츄에이터 조절 시스템(45)을 작동시키기 위한 방법에 있어서,

상기 엑츄에이터 조절 시스템(45)은 조절 전략(π)을 특성화하는 변수(Θ)의 함수로서 교정 변수(u)를 생성하고 또한 이 교정 변수(u)의 함수로서 상기 엑츄에이터(20)를 제어하도록 설정되고,

이때 상기 조절 전략(π)을 특성화하는 변수(Θ)는 가치 함수(V*)의 함수로서 결정되고, 또한 상기 가치 함수(V*)는 반복되는 가치 함수($\hat{V}^t$)의 연속적인 반복들에 의해, 벨만 방정식을 이용해 상기 가치 함수(V*)를 점진적으로 근사하는 것에 의해 반복적으로 결정되고, 이때 후속하는 반복의 반복되는 가치 함수($\hat{V}^{t+1}$)는 이전 반복의 반복되는 가치 함수로부터의 벨만 방정식을 이용해 결정되고,

이때 상기 벨만 방정식의 해에 대하여, 상기 이전 반복의 상기 반복되는 가치 함수($\hat{V}^t$) 대신, 제1 기본 함수들(ϕ_i^t)의 제1 집합(B1)에 의해 포괄되는 함수 공간으로의 그 투사만이 이용되는, 방법.

청구항 2

제 1 항에 있어서, 또한 상기 후속하는 반복의 반복되는 가치 함수($\hat{V}^{t+1}$) 대신 제2 기본 함수들(ϕ_i^{t+1})의 제2 집합(B2)에 의해 포괄되는 함수 공간으로의 그 투사만이 결정되는, 방법.

청구항 3

제 2 항에 있어서, 가우시안 함수들이 제1 기본 함수들(ϕ_i^t) 및 제2 기본 함수들(ϕ_i^{t+1}) 중 적어도 어느 하나로서 사용되는, 방법.

청구항 4

제 3 항에 있어서, 상기 벨만 방정식의 적분 값은 수치구적법에 의해 결정되는, 방법.

청구항 5

제 4 항에 있어서, 제2 기본 함수들(ϕ_i^{t+1})의 후속하는 제2 집합(B2)은 상기 반복되는 가치 함수($\hat{V}^t$)와 상기 제1 집합(B1)에 의해 포괄되는 함수 공간 상으로의 그 투사 사이의 최대 잔사($R^{k,l}$))가 얼마나 큰지에 따라서 적어도 하나의 추가적인 기본 함수(ϕ_{l+1}^t)를 상기 제1 집합(B1)에 부가하는 것에 의해 반복적으로 결정되는, 방법.

청구항 6

제 5 항에 있어서, 상기 적어도 하나의 추가적인 기본 함수(ϕ_{l+1}^t)는 상기 잔사($R^{k,l}$)가 최대가 되는 상기 조절 변수(x)의 최대점(x_*)에 따라서 선택되는, 방법.

청구항 7

제 6 항에 있어서, 상기 적어도 하나의 추가적인 기본 함수(ϕ_{l+1}^t)는 최대점(x_*)에서 그 최대값을 가정하는, 방법.

청구항 8

제 7 항에 있어서, 상기 적어도 하나의 추가적인 기본 함수(ϕ_{l+1}^t)는 상기 최대점(x_*)에서 상기 잔사($R^{k,l}$)의 곡

률을 특성화하는 변수, 또는 상기 최대점(x_s)에서 상기 잔사($R^{k,l}$)의 헤세 행렬($H^{t,l}$)에 따라 선택되는, 방법.

청구항 9

제 8 항에 있어서, 상기 적어도 하나의 추가적인 기본 함수(ϕ_{l+1}^{t})는 상기 최대점(x_s)에서의 헤세 행렬이 상기 잔사($R^{k,l}$)의 헤세 행렬($H^{t,l}$)과 같아지는 이러한 방식으로 선택되는, 방법.

청구항 10

제 1 항 내지 제 9 항 중 어느 한 항에 있어서, 상기 벨만 방정식이 따르는 조건부 확률(p)은 상기 엑츄에이터(20)의 모델을 이용해 결정되는, 방법.

청구항 11

제 10 항에 있어서, 상기 모델은 가우시안 프로세스(g)인, 방법.

청구항 12

제 11 항에 있어서, 상기 조절 전략(π)을 특성화하는 상기 변수(Θ)의 결정 후, 상기 모델(g)은 상기 교정 변수(u)의 함수로서 조정되고, 상기 교정 변수(u)는 상기 엑츄에이터(20)의 조절 동안 상기 엑츄에이터(20)로 공급되고, 상기 엑츄에이터 조절 시스템(45)는 상기 조절 전략(π), 및 그때 최종 조절 변수(x)를 고려하고, 이때 상기 모델(g)의 조정 후 상기 조절 전략(π)을 특성화하는 상기 변수(Θ)는 제 11 항에 따른 방법에 의해 다시 결정되고, 이때 상기 조건부 확률(p)은 그후 이제 조정된 모델(g)을 이용해 결정되는, 방법.

청구항 13

제 12 항에 있어서, 상기 교정 변수(u)는 상기 조절 전략(π)을 특성화하는 상기 변수(Θ)의 함수로서 생성되고 또한 상기 엑츄에이터(20)는 이 교정 변수(u)의 함수로서 제어되는, 방법.

청구항 14

삭제

청구항 15

삭제

청구항 16

삭제

청구항 17

삭제

청구항 18

삭제

발명의 설명

기 술 분 야

[0001] 본 발명은 엑츄에이터 조절 시스템을 작동시키기 위한 방법, 학습 시스템, 엑츄에이터 조절 시스템, 이 방법을 실행하기 위한 컴퓨터 프로그램 및 이 컴퓨터 프로그램이 저장되는 기계 판독가능한 저장 매체에 관한 것이다.

배 경 기 술

[0002] 선공개되지 않은 DE 10 2017 211 209로부터, 엑츄에이터 조절 시스템의 적어도 하나의 파라미터(parameter)의

자동 설정을 위한 방법이 알려져 있는데, 이것은 엑츄에이터의 조절 변수(regulation variable)를 미리 정의가능한 목표 변수(target variable)로 조절하도록 설계되고, 이때 엑츄에이터 조절 시스템은, 적어도 하나의 파라미터, 목표 변수 및 조절 변수에 따라서, 교정 변수(correcting variable)를 생성하고 또한 이 교정 변수의 함수로서 엑츄에이터를 제어하도록 설계되고, 이때 적어도 하나의 파라미터의 새로운 값은 장기 비용 함수(long-term cost function)의 함수로서 선택되고, 이때 이 장기 비용 함수는 엑츄에이터의 조절 변수의 확률 분포의 예측되는 시간 진행(predicted time evolution)의 함수로서 결정되고 그후 파라미터는 이 새로운 값으로 설정된다.

발명의 내용

해결하려는 과제

[0003] 대조적으로, 독립항 1 항의 특징들을 갖는 방법은 특히 엑츄에이터 조절 시스템의 최적 조절이 보장될 수 있다는 장점을 가진다. 유리한 추가적인 개선들은 종속항들의 요지이다.

과제의 해결 수단

[0004] 제1 측면에 있어서, 본 발명은 엑츄에이터의 조절 변수를 미리 정의가능한 목표 변수로 조절하도록 설정되는 엑츄에이터 조절 시스템을 작동시키기 위한 방법에 있어서, 상기 엑츄에이터 조절 시스템은 조절 전략(regulation strategy)을 특성화하는 변수의 함수로서, 특히 또한 목표 변수 및/또는 조절 변수의 함수로서, 교정 변수를 생성하고 또한 이 교정 변수의 함수로서 상기 엑츄에이터를 구동하도록 설정되고,

[0005] 이때 상기 조절 전략을 특성화하는 변수는 가치 함수(value function)의 함수로서 결정되는, 방법에 관한 것이다.

[0006] 이 가치 함수를 결정함으로써, 심지어 상태 변수들(state variables) 및/또는 행위들(actions)이 불연속적인 값들에 한정되지 않고 연속적인 값들을 획득할 수 있는 경우들에 있어서조차도, 엑츄에이터 조절 시스템의 최적 조절을 보장하는 것이 가능하다.

[0007] 특히, 조절 전략은 각각의 조절 변수에 대하여 교정 변수가 유도되는 행위가 결정되어, 이것이 가치 함수를 최대화하는 이러한 방식으로 결정될 수 있다.

[0008] 추가적인 개선에 있어서, 상기 가치 함수는 반복되는 가치 함수의 연속적인 반복들에 의해, 벨만 방정식(Bellmann equation)을 이용해 상기 가치 함수를 점진적으로 근사하는 것에 의해 반복적으로 결정되고, 이때 후속하는 반복의 반복되는 가치 함수는 벨만 방정식을 이용해 이전 반복의 반복되는 가치 함수로부터 결정되고, 이때 상기 이전 반복의 상기 반복되는 가치 함수 대신, 단지 기본 함수들(basic functions)의 집합에 의해 포괄되는, 선형 함수 공간으로의 그 투사만이 상기 벨만 방정식을 풀기 위해 이용되는 것이 제안된다.

[0009] 특히, 이것은 반복적으로 결정되는 가치 함수가, 특히 장기로 또한 시스템 역학을 고려한, 선-정의된 보상을 최대화하는 것을 보장한다. 투사들을 이용하는 것에 의해, 벨만 방정식을 푸는 것이 가능하고, 이것은 그 안에 포함된 최대 값 형성 때문에, 특히 근사에 의해 용이하게, 분석적으로 하나씩 풀릴 수 있다.

[0010] 만약 상기 후속하는 반복의 반복되는 가치 함수 대신 기본 함수들의 제2 집합에 의해 포괄되는, 함수 공간으로의 그 투사만이 결정된다면 특히 유리하다.

[0011] 따라서 후속하는 반복의 반복되는 가치 함수 그 자체를 완전히 계산할 필요 없이 이 투사를 결정하는 것이 가능하다.

[0012] 벨만 방정식의 적분은, 특히 분석적으로 푸는 것이 용이한데, 가우시안 함수들이 기본 함수들로서 사용될 때 획득된다. 이것은 이 방법이 수치적으로 특히 효율적이게 해준다.

[0013] 벨만 방정식의 최대 값 형성 때문에, 이것은 일반적으로 개별 점들에서만 평가될 수 있다. 그럼에도 불구하고 완전한 해는 벨만 방정식에 있어서의 적분이 수치구적법(numerical quadrature)을 이용해 계산된다면 가능하다. 따라서 수치구적법의 이용은 수치적으로 특히 효율적이다.

[0014] 본 발명의 다른 일 측면에 있어서, 기본 함수들의 후속하는 집합이 적어도 하나의 추가적인 기본 함수를 이에 따른 상기 집합에 부가하는 것에 의해 반복적으로 결정된다면, 상기 반복되는 가치 함수와 이 집합에 의해 포괄되는 함수 공간 상으로의 그 투사 사이의 최대 잔사(maximum residuum)가 얼마나 큰지가 제공된다.

[0015] 이 반복적인 절차에 의해, 이 방법의 수치적인 오류는 선-정의가능한 최대 값에 특히 효율적으로 제한될 수 있고 이로써 엑츄에이터 조절 시스템은 특히 용이하게 작동될 수 있다.

[0016] 추가적인 개선에 있어서, 상기 적어도 하나의 추가적인 기본 함수는 상기 잔사가 최대가 되는 상기 조절 변수의 최대점에 따라서 선택되는 것이 제공된다.

[0017] 이것은 이 방법을 특히 효율적으로 만들어 주는데, 이는 수치적 오류가 기본 함수들의 집합에 의해 포괄되는 함수 공간으로의 투사에 의해 특히 빠르게 감소될 수 있기 때문이다.

[0018] 만약 최대점에서 적어도 하나의 추가적인 기본 함수가 그 최대 값을 취한다면, 효율성은 특히 높다.

[0019] 대안적으로 또는 추가적으로, 만약 적어도 하나의 추가적인 기본 함수가 상기 최대점에서 상기 잔사의 곡률을 특성화하는 정도, 특히 상기 최대점에서 상기 잔사의 헤세 행렬(Hesse matrix)에 따라 선택된다면 이 방법의 효율성이 더 증가된다.

[0020] 만약 적어도 하나의 추가적인 기본 함수가 상기 최대점에서 그 헤세 행렬이 상기 잔사의 헤세 행렬과 같아지는 이러한 방식으로 선택된다면, 특히 다차원 조절 변수들의 경우에 있어서는, 특히 용이하다.

[0021] 본 발명의 다른 일 측면에 있어서, 상기 벨만 방정식이 따르는 조건부 확률이 상기 엑츄에이터의 모델을 이용해 결정되는 것이 제공될 수 있다. 이것은 또한 이 방법을 특히 효율적으로 만들어 주는데, 이는 엑츄에이터의 실제 행위를 다시 결정할 필요가 없기 때문이다.

[0022] 여기서 상기 모델이 가우시안 프로세스라면 특히 유리하다. 상기 기본 함수들이 가우시안 함수들에 의해 주어진다면 특히 유리한데, 이는 발생하는 적분들(occurring integrals)이 그후 가우시안 함수들의 곱을 이용한 적분으로서 분석적으로 풀릴 수 있기 때문이고, 이것은 특히 효율적인 구현을 가능하게 해준다.

[0023] 엑츄에이터 조절 시스템의 특히 좋은 조절 행동을 획득하기 위해서, 본 발명의 다른 일 측면에 따라 엑츄에이터 조절 시스템의 교습(teaching) 및 모델의 교습은 사건적 절차(episodic procedure)로 결정되는 것이 제공될 수 있는데, 이것은 상기 조절 전략을 특성화하는 상기 변수의 결정 후, 상기 모델이 상기 교정 변수에 따라서 생성되고, 상기 교정 변수는 엑츄에이터 조절 시스템을 갖는 엑츄에이터의 조절의 경우에 있어서 상기 엑츄에이터로 공급되고, 조절 전략을 고려하고, 또한 최종 조절 변수로 조정되고, 이때 상기 모델의 조정 후, 상기 조절 전략을 특성화하는 상기 변수는 상기에서 설명된 방법으로 다시 결정되고, 이때 상기 조건부 확률은 그후 이제 조정된 모델(now adapted model)을 이용해 결정되는 것을 의미한다.

[0024] 다른 일 측면에 있어서, 본 발명은 엑츄에이터 조절 시스템의 조절 전략을 특성화하는 변수를 자동으로 설정하기 위한 학습 시스템에 관한 것으로서, 이것은 엑츄에이터의 조절 변수를 선-정의가능한 목표 변수로 조절하도록 배치되고, 이 학습 시스템은 상기에서 언급된 방법들 중 하나를 수행하도록 배치된다.

[0025] 다른 일 측면에 있어서, 본 발명은 조절 전략을 특성화하는 변수가 상기에서 언급된 방법들 중 하나에 따라 결정되고 그후, 조절 전략을 특성화하는 변수에 따라서, 조작된 변수가 생성되고 또한 엑츄에이터는 이 교정 변수에 따라서 제어되는, 방법에 관한 것이다.

[0026] 다른 일 측면에 있어서, 본 발명은 이 방법을 이용하는 엑츄에이터를 제어하도록 설정하는 엑츄에이터 조절 시스템에 관한 것이다.

[0027] 또 다른 일 측면에 있어서, 본 발명은 상기에서 언급된 방법들 중 하나를 수행하도록 설정되는 컴퓨터 프로그램에 관한 것이다. 다시 말하면, 컴퓨터 프로그램은, 컴퓨터 상에서 실행될 때, 컴퓨터가 이 방법을 수행하도록 야기시키는 지시들(instructions)을 포함한다.

[0028] 본 발명은 이 컴퓨터 프로그램이 저장되는 기계 판독가능한 저장 매체에 관한 것이다.

도면의 간단한 설명

[0029] 이어서, 본 발명의 실시예들이 첨부된 도면들을 참조하여 더 상세하게 설명된다.

도 1은 학습 시스템과 엑츄에이터 사이의 상호작용을 나타내는 대략도이다.

도 2는 엑츄에이터 조절 시스템과 엑츄에이터 사이의 상호작용을 나타내는 대략도이다.

도 3은 흐름도로서, 엑츄에이터 조절 시스템을 훈련시키기 위한 방법의 일 실시예이다.

도 4는 흐름도로서, 반복되는 가치 함수를 결정하기 위한 방법의 일 실시예이다.

도 5는 흐름도로서, 기본 함수들의 집합을 결정하기 위한 방법의 일 실시예이다.

도 6은 흐름도로서, 교정 변수를 결정하기 위한 방법들의 일 실시예이다.

발명을 실시하기 위한 구체적인 내용

[0030] 도 1은 학습 시스템(40)과 상호작용하는 그 환경(20)에 있어서의 엑츄에이터(10)를 보여준다. 엑츄에이터(10)와 환경(20)은 집합적으로 이하에서 엑츄에이터 시스템으로 지칭된다. 엑츄에이터 시스템의 상태는 센서(30)에 의해 검출되는데, 이것은 또한 복수의 센서들에 의해 제공될 수 있다. 센서(30)의 출력 신호(S)는 학습 시스템(40)으로 전송된다. 학습 시스템(40)은 이로부터 엑츄에이터(10)가 수신하는, 구동 신호(A)를 결정한다.

[0031] 엑츄에이터(10)는, 예를 들어 (부분적으로) 자율 로봇, 예를 들어 (부분적으로) 자율 차량, (부분적으로) 자율 잔디깍기기계일 수 있다. 이것은 또한 차량의 엑츄에이터의 작동(actuation), 예를 들어 쓰로틀 밸브 또는 유휴 제어를 위한 바이패스 엑츄에이터일 수 있다. 이것은 또한 난방 장치 또는 밸브 엑츄에이터와 같이, 난방 장치의 일부일 수 있다. 엑츄에이터(10)는 특히 또한 내연 기관 또는 차량의 구동 트레인(가능하다면 하이브리드된) 또는 브레이크 시스템과 같이, 더 큰 시스템들일 수 있다.

[0032] 센서(30)는, 예를 들어 하나 또는 복수의 비디오 센서들 및/또는 하나 또는 복수의 라이더 센서들 및/또는 하나 또는 복수의 초음파 센서들 및/또는 하나 또는 복수의 위치 센서들(예를 들어 GPS)일 수 있다. 다른 센서들, 예를 들어 온도 센서가 고려될 수 있다.

[0033] 다른 일 실시예에 있어서, 엑츄에이터(10)는 제조 로봇일 수 있고 센서(30)는 이때, 예를 들어 제조 로봇의 제조 제품들의 특성을 검출하는 광학 센서일 수 있다.

[0034] 학습 시스템(40)은 출력 신호(S)를 조절 변수(x)로 변환하는, 선택적 수신 유닛(50)에서 센서(30)의 출력 신호(S)를 수신한다(또는, 출력 신호(S)는 또한 조절 변수(x)로서 직접 인계될 수 있다). 조절 변수(x)는, 예를 들어 출력 신호(S)의 일부 또는 추가적 처리일 수 있다. 조절 변수(x)는 조절기(60)로 공급된다. 조절기에, 조절 전략(π) 또는 가치 함수(V*)가 구현될 수 있다.

[0035] 파라미터 메모리(70)에, 파라미터들(Θ)이 저장되는데, 이것은 조절기(60)로 공급된다. 파라미터들(Θ)은 조절 전략(π) 또는 가치 함수(V*)를 파라미터로 나타낸다. 파라미터들(Θ)은 하나 또는 복수의 파라미터들일 수 있다.

[0036] 블록(90)은 조절기(60)로 선-정의가능한 목표 변수(xd)를 공급한다. 블록(90)이 예를 들어 블록(90)에 대하여 미리 정의된 센서 신호의 함수로서, 선-정의가능한 목표 변수(xd)를 생성하는 것이 제공될 수 있다. 블록(90)이 목표 변수(xd)가 존재하는 전용 메모리 영역으로부터 이를 판독하는 것 또한 가능하다.

[0037] 조절 전략(π) 또는 가치 함수(V*)에 따라서, 목표 변수(xd) 및 조절 변수(x)에 따라서, 조절기(60)는 교정 변수(u)를 생성한다. 이것은 예를 들어 조절 변수(x)와 목표 변수(xd) 사이의 차(x-xd)에 따라서, 결정될 수 있다.

[0038] 조절기(60)는 교정 변수(u)를 이로부터 구동 신호(A)를 결정하는, 출력 유닛(80)으로 전송한다. 출력 유닛이 먼저 교정 변수(u)가 선-정의가능한 변수 범위 내에 있는지 여부를 점검하는 것이 가능하다. 이 경우라면, 제어 신호(A)는 예를 들어 교정 변수(u)의 함수로서 특성 필드로부터 판독되는 연관된 구동 신호(A)에 의해, 교정 변수(u)의 함수로서 결정된다. 이것은 정상적인 경우이다. 한편, 교정 변수(u)가 선-정의가능한 값 범위 내에 있지 않다면, 제어 신호(A)는 엑츄에이터(A)가 안전 모드로 돌입하도록 야기시키는 이러한 방식으로 설계되는 것이 제공될 수 있다.

[0039] 수신 유닛(50)은 조절 변수(x)를 블록(100)으로 전송한다. 유사하게, 조절기(60)는 대응하는 교정 변수(u)를 블록(100)으로 전송한다. 블록(100)은 일련의 시간들에서 수신되는 조절 변수(x)의 시계열들 및 각각의 대응하는 교정 변수(u)를 저장한다. 블록(100)은 이때 이 시계열들에 기초하여 모델(g)의 모델 파라미터들(Λ, σ_n, σ_f)을 조정할 수 있다. 이 모델 파라미터들(Λ, σ_n, σ_f)은 블록(110)으로 공급되는데, 이것은 예를 들어 전용 저장 위치에, 이들을 저장한다. 이것은 이하의 도 4의, 단계(1010)에서 더 상세하게 설명될 것이다.

[0040] 학습 시스템(40)은, 일 실시예에 있어서, 컴퓨터(41)에 의해 실행될 때, 학습 시스템(40)의 설명된 기능을 수행하도록 야기시키는 컴퓨터 프로그램이 저장되는, 기계 판독가능한 저장 매체(42)를 갖는 컴퓨터(41)를

포함한다. 이 실시예에 있어서, 컴퓨터(41)는 GPU(43)를 포함한다.

[0041] 이 모델(g)은 가치 함수(V*)의 결정을 위해 이용될 수 있다. 이것은 이하에서 설명된다.

[0042] 도 2는 엑츄에이터(10)와 엑츄에이터 조절 시스템(45)의 상호작용을 보여준다. 엑츄에이터 조절 시스템(45)의 구조 및 엑츄에이터(10) 및 센서(30)와의 상호작용은 학습 시스템(40)의 구조와 많은 부분들에서 유사하고 여기서는 단지 차이점들만 설명한다. 학습 시스템(40)과 달리, 엑츄에이터 조절 시스템(45)은 블록(100) 및 블록(110)을 가지지 않는다. 블록(100)으로의 변수들의 전송은 이로써 생략된다. 엑츄에이터 조절 시스템(45)의 파라미터 메모리(70)에는, 파라미터들(Θ)이 저장되는데, 이것은 예를 들어 도 4에 도시된 바와 같이, 본 발명에 따른 방법에 의해 결정되었다.

[0043] 도 3은 본 발명에 따른 방법의 일 실시예를 보여준다. 먼저(1000), 조절 변수(x)의 초기 값(x_0)이 선-정의가능한 초기 확률 분포($p(x_0)$)로부터 선택된다. 사건 인덱스(e)는 값 e=1으로 초기화되고, 이 사건 인덱스(e)에 할당된 가치 함수($\hat{V}_e$)는 값 $\hat{V}^e = 0$ 으로 초기화된다.

[0044] 이에 더하여, 교정 변수들(u_0, u_1, ..., u_{T-1})은 선-정의가능한 시계(time horizon, T)까지 랜덤하게 선택되고 이로써 엑츄에이터(10)는 도 1에서 설명된 바와 같이 제어된다. 엑츄에이터(10)는 환경(20)을 통해 센서(30)와 상호작용하고, 그 센서 신호(S)는 조절기(60)로부터 직접 또는 간접적으로 조절 변수(x_1, ..., x_{T-1}, x_T)로서 수신된다.

[0045] 이것들은 데이터 집합 $\mathcal{D} = \{(x_0, u_0, x_1), \ldots, (x_{T-1}, u_{T-1}, x_T\}$ 으로 결합된다.

[0046] 블록(100)은 조절 변수(x) 및 교정 변수(u)의 시계열들을 수신하고 집계하고(1030), 이는 조절 변수(x)와 교정 변수(u)의 쌍(z) $z_t = (x_t^1, \ldots, x_t^D, u_t^1 \ldots u_t^F)^T$ 으로 귀결된다.

[0047] D는 이로써 조절 변수(x)의 차원수(dimensionality)이고 F는 교정 변수(u)의 차원수이다. 즉 $x \in \mathbb{R}^D, u \in \mathbb{R}^F$.

[0048] 이 상태 궤적(state trajectory)에 따라서, 이때 가우시안 프로세스(g)는 연속하는 시간들(t, t+1) 사이에 이하가 적용되는 이러한 방식으로 조정된다.

[0049] $$x_{t+1} = x_t + g(x_t, u_t) \quad (1)$$

[0050] 여기서

[0051] $$u_t = \pi_\theta(x_t) \quad (1')$$

[0052] 가우시안 프로세스(g)의 공분산 함수(k)는, 예를 들어 이하와 같이 주어진다.

[0053] $$k(z, w) = \sigma_f^2 \exp\left(-\frac{1}{2}(z-w)^T \Lambda^{-1}(z-w)\right) \quad (2)$$

[0054] 파라미터(σ_f^2)는 신호 분산이고, $\Lambda = diag(l_1^2 \ldots l_{D+F}^2)$ 는 D+F 입력 차원들 각각에 대한 제곱 길이 축척들(squared length scales) $l_1^2 \ldots l_{D+F}^2$ 의 집합이다.

[0055] 공분산 행렬(K)은 이하에 의해 정의된다.

[0056] $$K(Z, Z)_{i,j} = k\left(z^i, z^j\right) \quad (3)$$

[0057] 가우시안 프로세스(g)는 이때 2 개의 함수들에 의해 특성화된다: 평균(μ) 및 분산(Var). 이것들은 이하와 같이 주어진다.

[0058] $$\mu(z_*) = k(z_*, Z)(K(Z, Z) + \sigma_n^2 I)^{-1} y, \quad (4)$$

[0059] $$Var(z_*) = k(z_*, z_*) - k(z_*, Z)(K(Z, Z) + \sigma_n^2 I)^{-1} k(Z, z_*) \quad (5)$$

[0060] 여기서 y는 백색소음(ϵ^i)를 가지고, 보통 $y^i = f\left(z^i\right) + \epsilon^i$ 에 의해 주어진다.

[0061] 파라미터들($\Lambda, \sigma_n, \sigma_f$)은 이때 로그주변우도 함수(logarithmic *marginal likelihood* function)를 최대화하는 것에 의해, 알려진 방식으로 쌍들(z^i, y^i)로 매칭된다.

[0062] 그후 (1020) 사건 인덱스(e)와 연관된 반복되는 가치 함수들($\hat{V}_e^1, \hat{V}_e^2, \ldots \hat{V}_e^*$)이 결정되고, 이 반복되는 가치 함수들의 마지막은 사건 인덱스(e)와 연관된 수렴하는 반복되는 가치 함수($\hat{V}_e^*$)이다. 사건 인덱스(e)에 할당된 반복되는 가치 함수들($\hat{V}_e^1, \hat{V}_e^2, \ldots \hat{V}_e^*$)을 결정하기 위한 본 발명의 일 실시예는 도 5에 도시되어 있다.

[0063] 그후 (1030) 예를 들어 현재 사건 인덱스(e)에 할당된 수렴하는 반복되는 가치 함수들 및 이전의 사건 인덱스(e-1)에 할당된 반복되는 가치 함수들($\hat{V}_e^*$, $\hat{V}_{e-1}^*$)이 함수의 제1 선-정의가능한 한계(Δ_1)보다 더 작게 차이가 있는지 여부를 점검하는 것에 의해, 사건 인덱스(e)와 연관된 수렴하는 반복되는 가치 함수($\hat{V}_e^*$)가 수렴되는지 알기 위해 점검된다. 즉 $\|\hat{V}_e^* - \hat{V}_{e-1}^*\| < \Delta_1$. 이 경우라면, 단계(1080)가 뒤따른다.

[0064] 하지만, 수렴이 아직 달성되지 않았다면(1040), 사건 인덱스(e)에 연관된 최적 조절 전략(π_e)은 이하에 의해 정의된다.

[0065] $$\pi_e(x) = \mathrm{argmax}_u \int p(x'|x,u)\hat{V}_e^*(x')dx' \qquad (6)$$

[0066] 그후 (1050) 조절 변수(x)의 초기 값(x_0)은 다시 초기 확률 분포($p(x_0)$)로부터 선택된다.

[0067] 식 (6)에 정의된 최적 조절 전략(π_e)을 이용해, 일련의 조절 변수들($\pi_e(x_0), \ldots, \pi_e(x_{T-1})$)은 이제 (1060) 반복적으로 결정되고 이로써 엑츄에이터(10)는 제어된다. 이때 수신된 센서(30)의 출력 신호들(S)로부터, 최종 상태 변수들($x_1, \ldots, x_T$)이 그후 결정된다.

[0068] 이제 (1070) 사건 인덱스(e)는 1 더 증가되고, 단계(1030)으로 다시 분기된다.

[0069] 단계(1030)에서 사건들에 대한 반복이 사건 인덱스(e)에 할당된 반복되는 가치 함수들($\hat{V}_e^*$)의 수렴으로 안내하는 것으로 결정되었다면, 가치 함수(V^*)는 사건 인덱스(e)에 할당된 반복되는 가치 함수들($\hat{V}_e^*$)과 같도록 설정된다. 이것으로 이 방법의 측면이 끝난다.

[0070] 도 4는 사건 인덱스(e)에 할당된 반복되는 가치 함수들($\hat{V}_e^1, \hat{V}_e^2, \ldots \hat{V}_e^*$)을 결정하기 위한 방법의 일 실시예를 보여준다. 명확함을 위해, 사건 인덱스(e)는 이하에서 생략된다. 윗첨자는 이하에서 문자 t로 지칭된다. 이 방법은 항상 이전의 가치 함수($\hat{V}^t$)에 기초하여, 항상 후속하는 반복되는 가치 함수($\hat{V}^{t+1}$)를 계산한다. 이 이전의 반복되는 가치 함수($\hat{V}^t$)는 기본 함수들($\{\phi_i^t\}_{i \le N_t}$) 및 계수들($\{\alpha_i^t\}_{i \le N_t}$)의 선형 결합 $\hat{V}^t = \sum_{i=1}^{N_t} \alpha_i^t \cdot \phi_i^t$ 으로서 주어진다. 이 계수들($\{\alpha_i^t\}_{i \le N_t}$) 또한 계수 벡터($\alpha^t$)로 간략히 요약된다. 이 방법은 인덱스 $t = 0$로 시작한다(1500).

[0071] 먼저, 기본 함수들($\{\phi_i^{t+1}\}_{i \le N_{t+1}}$)의 집합(B)이 결정된다(1510). 이것들은 미리 정의될 수 있거나, 또는 도 6에 도시된 알고리즘을 이용해 결정될 수 있다.

[0072] 그후 (1520) $i, j = 1 \ldots N_{t+1}$ 에 대하여 스칼라 곱들 $M_{ij} = \langle \phi_i^{t+1} | \phi_j^{t+1} \rangle_{L^2}$ 이 결정된다.

[0073] 이어서 (1530), 노드들($\xi_1, \ldots, \xi_K$) 및 연관된 가중치들($w_1, \ldots, w_K$)이 수치구적법을 이용해 정의된다.

[0074] 이 노드들($\xi_1, \ldots, \xi_K$) 및 가중치들($w_1, \ldots, w_K$)의 도움으로 그후 (1540) 모든 인덱스들 $i = 1 \ldots N_{t+1}$에 대하여 벡터(b^{t+1})의 계수들(b_i^{t+1})이 이하로 결정된다.

[0075] $$b_i^{t+1} = \sum_{k=1}^{K} w_k \phi_i^{t+1}(\xi_k) A\hat{V}^t(\xi_k) \quad (7)$$

[0076] 계수 벡터(α^{t+1})는 이제 (1550) $\alpha^{t+1} = M^{-1}b^{t+1}$로 결정되고, 이때 질량 행렬(mass matrix, M)은 $M = (M_{ij})_{i,j \le N_{t+1}}$에 의해 주어진다.

[0077] 연산자(A)는 이하로 정의된다.

[0078] $$A\hat{V}^t(x) = \max_u \int \left(p(x'|x,u) \cdot (\mathrm{r}(\mathrm{x}') + \gamma \hat{\mathrm{V}}^t(x'))\right) \mathrm{dx}' \quad (8)$$

[0079] 여기서, $0<\gamma<1$ 는 특정할 수 있는 가중치 인자이고, 예를 들어 $\gamma = 0.85$이다. r는 보상 값을 조절 변수(x)의 값에 할당하는 보상 함수(reward function)이다. 유리하게도, 보상 함수(r)는 목표 변수(xd)로부터 조절 변수(x)의 편차(deviation)가 작을수록, 그 가정되는 값이 더 커지는 이러한 방식으로 선택된다.

[0080] 이전의 조절 변수(x) 및 조작된 변수(u)가 주어질 때 조절 변수(x')의 조건부 확률($p(x'|x,u)$)은 가우시안 프로세스(g)를 이용해 식 (8)에서 결정될 수 있다.

[0081] 식 (8)의 max 연산자는 분석적 해에 접근가능하지 않음에 유의해야 한다. 하지만, 주어진 조절 변수(x)에 대하여, 최대화는 경사상승법(gradient ascent method)을 이용해 각각의 경우에 있어서 발생할 수 있다.

[0082] 이 정의들은 이러한 방식으로 정의된, 후속하는 반복되는 가치 함수($\hat{V}^{t+1} = \sum_{i=1}^{N_{t+1}} \alpha_i^{t+1} \cdot \phi_i^{t+1}$)가 기본 함수들(B)에 의해 포괄되는 공간으로의 실제 반복되는 가치 함수(V^{t+1})의 투사에 대응하는 것을 보장하고, 이때 실제 반복되는 가치 함수들은 이하의 벨만 방정식을 만족한다.

[0083] $$V^{t+1}(\mathrm{s}) = \max_u \int \left(p(x'|x,u) \cdot (\mathrm{r}(\mathrm{x}') + \gamma \mathrm{V}^t(x'))\right) \mathrm{dx}' \quad (9)$$

[0084] 벡터(b^{t+1})는 이로써 방정식 $b_i^{t+1} = \langle \phi_i^{t+1} | V^{t+1} \rangle_{L^2}$을 적절하게 만족시키고, 이때 이 방정식은, 단지 예외적인 경우들에 있어서 풀릴 수 있지만, 실제 가치 함수(V^{t+1})가 기본 함수들(B)에 의해 포괄되는 공간으로의 그 투사에 의해, 즉 반복되는 가치 함수($\hat{V}^{t+1}$)에 의해 대체되고, 또한 수치구적법으로 최종 적분 방정식이 적절하게 풀린다면, 풀릴 수 있다.

[0085] 이제 (1560) 종료 기준이 만족되었는지 점검된다. 이 종료 기준(termination criteria)은, 예를 들어 반복되는 가치 함수($\hat{V}^{t+1}$)가 수렴된다면, 예를 들어 이전의 반복되는 가치 함수($\hat{V}^t$)와의 차가 함수의 제2 한계(Δ_2)보다 더 작게 된다면, 즉 $\|\hat{V}^{t+1} - \hat{V}^t\| < \Delta_2$라면, 만족될 수 있다. 종료 기준은 또한 인덱스(ˡ)가 선-정의가능한 시계(T)에 도달한다면 만족된 것으로 간주될 수 있다.

[0086] 이 종료 기준이 만족되지 않는다면, 인덱스(ˡ)는 1 더 증가된다(1570). 한편, 종료 기준이 만족된다면, 가치 함수(V*)는 마지막 반복의 반복되는 가치 함수($\hat{V}^{t+1}$)와 같도록 설정된다.

[0087] 이것으로 이 방법의 이 부분은 끝난다.

[0088] 도 5는 벨만 방정식의 실제 반복되는 가치 함수(V^t)에 대한 기본 함수들의 집합(B)을 결정하기 위한 방법의 일 실시예를 보여준다. 이를 위해 먼저 (1600) 기본 함수들의 집합(B)은 빈 집합으로 초기화되고, 인덱스(l)는 값 l=0으로 초기화된다. 기본 함수들의 집합(B)으로 투사된 반복되는 가치 함수($\hat{V}^{t,l}$) 또한 값 0으로 초기화된다.

[0089] 그후 (1610) 잔사 $R^{t,l}(x) = |\hat{V}^t(x) - \hat{V}^{t,l}(x)|$는 반복되는 가치 함수($\hat{V}^t$)와 대응하는 투사된 반복되는 가치 함수($\hat{V}^{t,l}$) 사이의 편차로서 정의된다.

[0090] 그후 (1620) 잔사의 최대점 $x_* = arg\max_s R^{t,l}(x)$이 예를 들어 경사상승법으로, 결정되고 또한 잔사($R^{t,l}$)의 헤세 행렬($H^{t,l}$)이 최대 자리수(maximum digit, x_*)에서 결정된다.

[0091] 이제 (1630) 기본 함수들의 집합(B)에 추가될 새로운 기본 함수(ϕ_{l+1}^t)가 결정된다. 추가될 새로운 기본 함수(ϕ_{l+1}^t)는 바람직하게 중간 값(s_*) 및 공분산 행렬(Σ^*)을 갖는 가우시안 함수로서 선택된다. 공분산 행렬(Σ^*)은 이 식을 충족하는 이러한 방식으로 계산된다.

[0092] $$\Sigma_*^{-1} = -R^{t,l}(x_*)^{(-2)}\nabla^T R^{t,l}(\mathrm{x})|_{\mathrm{x}=\mathrm{x}_*}\nabla R^{t,l}(\mathrm{x})|_{\mathrm{x}=\mathrm{x}_*} + R(x_*)^{-1}H^{t,l} \quad (10)$$

[0093] 그후 (1640) 이 기본 함수(ϕ_{l+1}^t)는 기본 함수들의 집합(B)에 추가된다.

[0094] 이제 (1650) 투사된 반복되는 가치 함수($\hat{V}^{t,l+1}$)는 기본 함수들의 이제 확장된 집합(B)에 의해 포괄되는 함수 공간으로 반복되는 가치 함수($\hat{V}^t$)의 투사에 의해 결정된다.

[0095] 이어서 (1660) 투사된 반복되는 가치 함수($\hat{V}^{t,l+1}$)의 결정이, 예를 들어 연관된 편차의 표준(norm)(예를 들어 L_∞ 표준)이 함수의 제3 선-정의가능한 한계 이하로 떨어지는지 점검하는 것에 의해, 즉 $\|\hat{V}^{t,l+1} - \hat{V}^t\|_{L_\infty} < \Delta_3$, 충분히 수렴되는지 점검한다.

[0096] 이 경우가 아니라면, 인덱스(l)은 1 더 증가되고 이 방법은 단계(1610)으로 다시 분기된다.

[0097] 그렇지 않다면, 결정된 집합($B = \{\phi_i^t\}_{i\le l+1}$)은 검색된 기본 함수들의 집합으로 복귀되고 이 방법의 이 부분은 끝난다.

[0098] 도 6은 교정 변수를 결정하기 위한 방법의 실시예들을 보여주고 또한 도 7a는 파라미터 저장부(70)에 저장된 파라미터들(Θ)이 조절 전략(π)을 파라미터로 나타내는 경우에 대한 일 실시예를 보여준다. 이를 위해, 먼저 (1700) 테스트 점들(x_i)의 집합이, 예를 들어 소볼 설계 계획(Sobol design plan)으로서 정의된다.

[0099] 그후 (1710) 테스트 점들(u_i)에 할당된 최적 교정 변수들(x_i)이 이하의 식을 이용해 계산된다.

[0100] $$u_i = \mathrm{argmax}_{u\in U}\int p(x'|x_i,u)V^*(x')dx' \quad (11)$$

[0101] 예를 들어, 경사상승법을 이용해 결정되고, 훈련 집합($M = \{(x_1,u_1),(x_2,u_2),\dots\}$)은 각각에 할당된 최적 조작된 변수들($u_i$)을 갖는 테스트 점들($x_i$)의 쌍들로부터 생성된다.

[0102] 이 훈련 집합(M)으로, 데이터에 기반한 모델, 예를 들어 가우시안 프로세스(g_θ)가 그후 (1720) 교습되어, 데이터에 기반한 모델은 조절 변수(x)에 대하여 할당되는 최적 교정 변수(u)를 효율적으로 결정하게 된다. 가우시안 프로세스(θ)를 특성화하는 파라미터들(g_θ)은 파라미터 저장부(70)에 저장된다.

[0103] 단계들 (1700) 내지 (1720)은 바람직하게 학습 시스템(40) 내에서 실행된다.

[0104] 엑츄에이터 조절 시스템(45)의 작동 동안 (1730), 이 시스템은 이때 가우시안 프로세스(g_θ)를 이용해 주어진 조절 변수(x)에 대하여 연관된 교정 변수(u)를 결정한다.

[0105] 이것으로 이 방법은 끝난다.

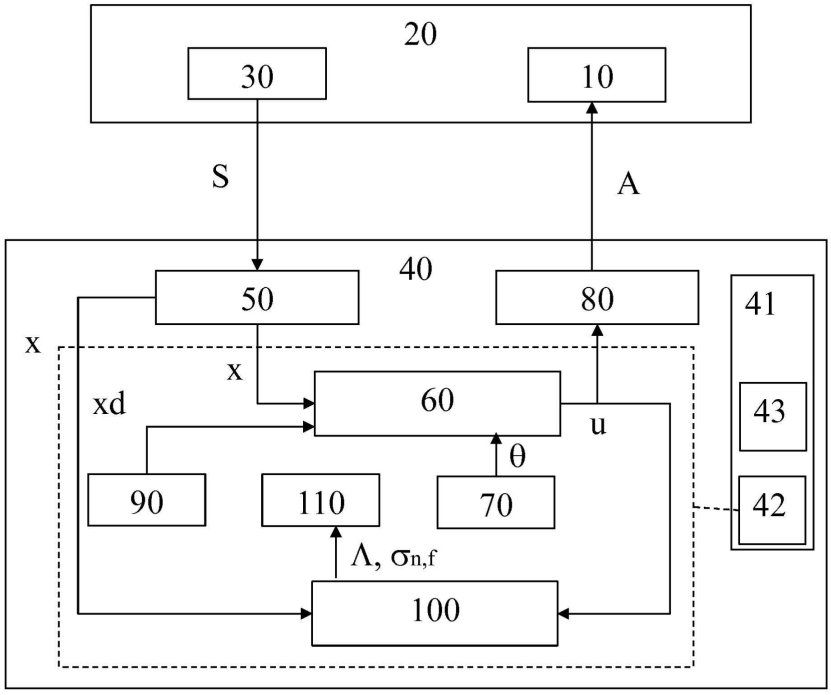
[0106] 도 7b는 파라미터 저장부(70)에 저장된 파라미터들(Θ)이 가치 함수(V*)를 파라미터로 나타내는 경우에 대한 일 실시예를 보여준다. 이를 위해, 주어진 조절 변수(x)에 대하여 단계(1800)에서, 단계(1710)과 유사하게, 이하의 식에 의해 정의되는 연관된 교정 변수(u)는 경사상승법으로 결정된다.

[0107] $$u = \mathrm{argmax}_u\int p(x'|x,u)V^*(x')dx'$$

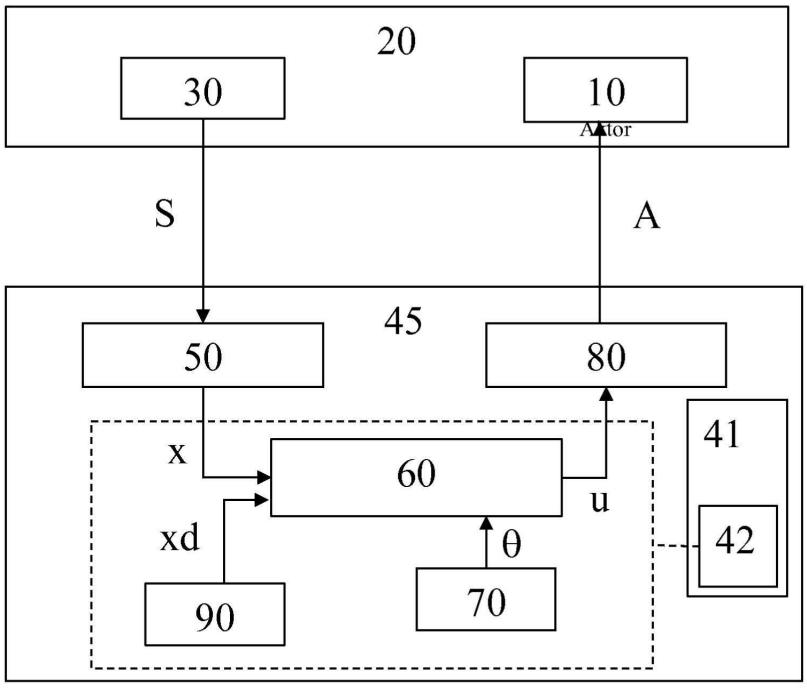
[0108] 이것으로 이 방법은 끝난다.

도면

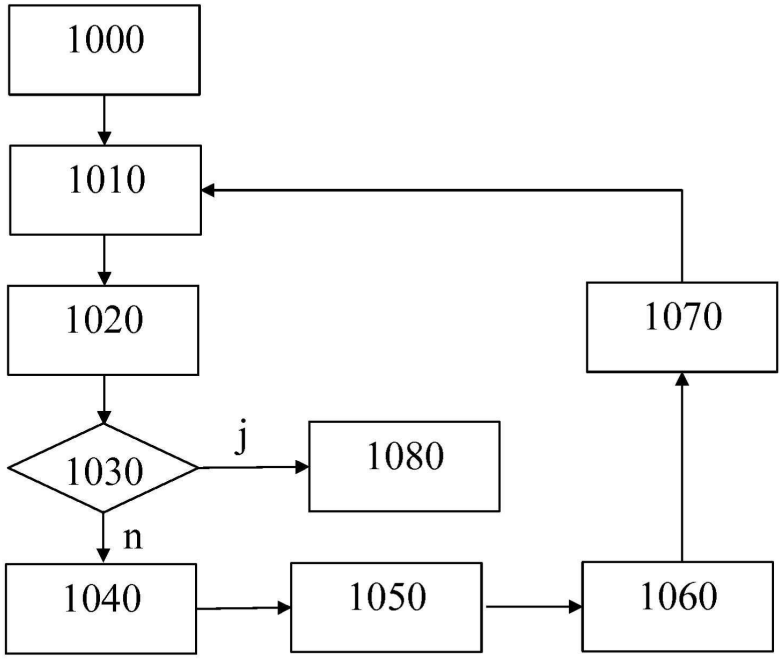
도면1



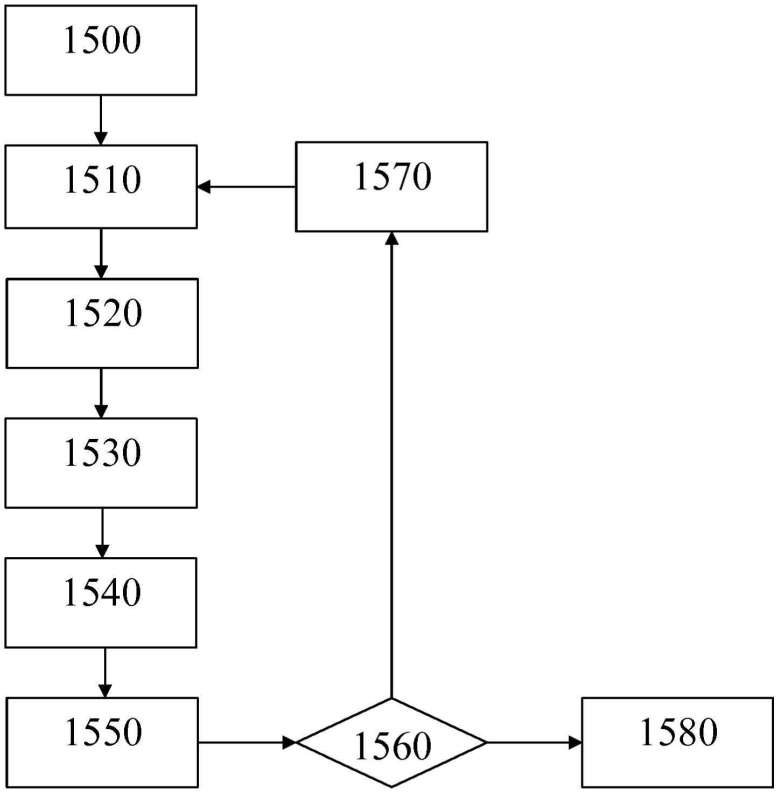
도면2



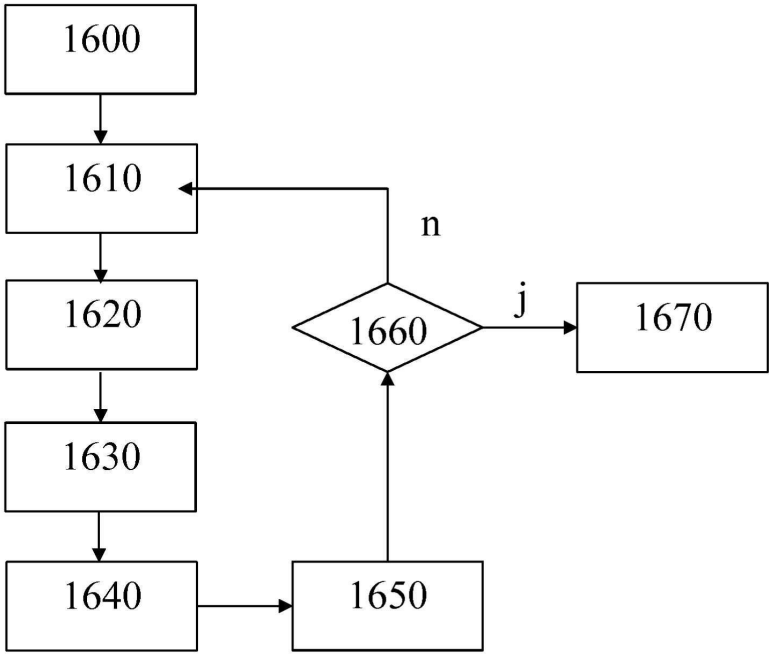
도면3


1000
1010
1020
1030
j
1080
n
1040
1050
1060
1070

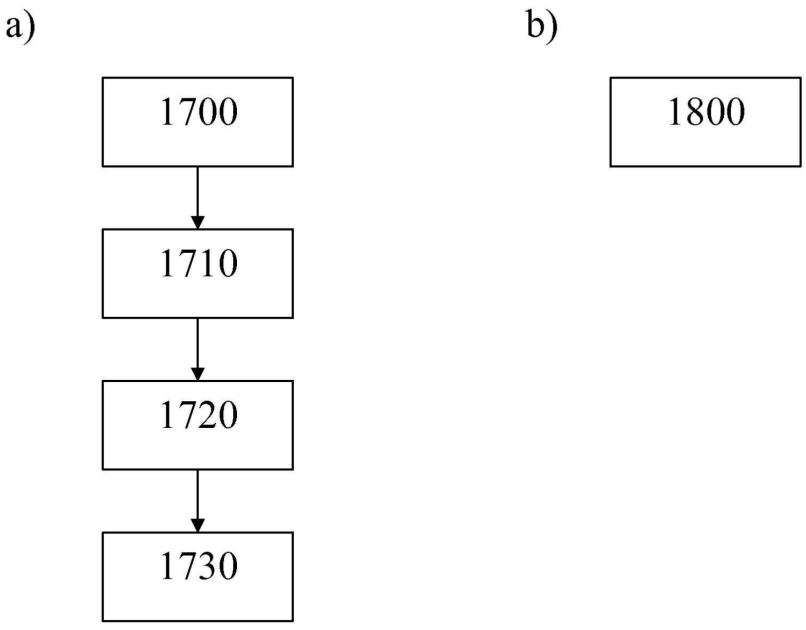
도면4


1500
1510
1570
1520
1530
1540
1550
1560
1580

도면5


1600
1610
1620
1630
1640
1650
1660
1670
n
j

도면6


a)
b)
1700
1710
1720
1730
1800