システム制御理論と統計的機械学習

第 10 章：マルコフ連鎖と定常状態

加嶋 健司

October 10, 2025

京都大学 情報学研究科

© 2025 Kenji Kashima

本章の流れ

10.1 有限状態確率システム
状態変数と入力変数が有限個の値しかとらない確率システムを対象とした制御問題

10.2 確率システムの定常状態
確率システムの状態変数の分布の収束性，収束先の設計と工学応用

有限状態確率システム

ベルマン方程式と線形計画法 (1) 確率による表現

状態変数 $x_k \in \lceil n \rceil := \{1, 2, \ldots, n\}$，入力変数 $u_k \in \lceil m \rceil := \{1, 2, \ldots, m\}$

遷移確率

$$\mathbb{P}(x_{k+1} = \mathrm{x}' | x_k = \mathrm{x}, u_k = \mathrm{u}) = \Psi_k(\mathrm{x}' | \mathrm{x}, \mathrm{u}) \tag{10.1}$$

ベルマン方程式などここまでの結果はそのまま成り立つ

- 有限時間

$$V(k, \mathrm{x}) = \min_{\mathrm{u} \in U_k(\mathrm{x})} \left(\ell(k, \mathrm{x}, \mathrm{u}) + \sum_{\mathrm{x}' \in \lceil n \rceil} V(k+1, \mathrm{x}') \Psi_k(\mathrm{x}' | \mathrm{x}, \mathrm{u}) \right) \tag{10.2}$$

- 無限時間

$$V(\mathrm{x}) = \min_{\mathrm{u} \in U_k(\mathrm{x})} \left(\ell(\mathrm{x}, \mathrm{u}) + \beta \sum_{\mathrm{x}' \in \lceil n \rceil} V(\mathrm{x}') \Psi(\mathrm{x}' | \mathrm{x}, \mathrm{u}) \right) \tag{10.3}$$

- $\min$ は網羅的な有限回の評価で計算可能

ベルマン方程式と線形計画法 (2) 下界

連続最適化により状態価値関数の下界を評価

下界

$$\breve{\mathrm{V}}(k,\mathrm{x}) \leq \ell(k,\mathrm{x},\mathrm{u}) + \sum_{\mathrm{x}'\in\lceil n\rceil} \Psi_k(\mathrm{x}'|\mathrm{x},\mathrm{u})\breve{\mathrm{V}}(k+1,\mathrm{x}') \ , \forall k \in \lfloor \bar{k} \rfloor,\ x,\ u \tag{10.4a}$$

$$\breve{\mathrm{V}}(\bar{k},\mathrm{x}) \leq \ell_f(\mathrm{x}) \ , \forall x \tag{10.4b}$$

のもとでの $\{\breve{\mathrm{V}}(k,\mathrm{x})\}$ を決定変数とする

$$\mathbb{E}[\breve{\mathrm{V}}(0,x_0)] = \sum_{\mathrm{x}\in\lceil n\rceil} \mathbb{P}(\mathrm{x}_0=\mathrm{x})\breve{\mathrm{V}}(0,x) \tag{10.5}$$

の最大値が最適制御性能に一致する（V^* が最適解のひとつ）

- 制約を満たす任意の $\breve{\mathrm{V}}$ に対して $V^* \geq \breve{\mathrm{V}}$
- 無限時間の場合，$\breve{\mathrm{V}}(\mathrm{x}) \leq \ell(\mathrm{x},\mathrm{u}) + \beta \sum_{\mathrm{x}\in\lceil n\rceil} \Psi(\mathrm{x}'|\mathrm{x},\mathrm{u})\breve{\mathrm{V}}(\mathrm{x})$

ベルマン方程式と線形計画法 (3) 上界

連続最適化により状態価値関数の上界を評価

任意の確率的状態フィードバック π に対して $\zeta_k(\mathrm{x},\mathrm{u}) := \mathbb{P}^{\pi}(x_k = \mathrm{x}, u_k = \mathrm{u})$

上界

$$\begin{aligned} &\zeta_k(\mathrm{x},\mathrm{u}) \geq 0, \quad \sum_{\mathrm{x},\mathrm{u}} \zeta_k(\mathrm{x},\mathrm{u}) = 1, \quad \sum_{\mathrm{u}} \zeta_0(\mathrm{x},\mathrm{u}) = \mathbb{P}(x_0 = \mathrm{x}) \\ &\sum_{\mathrm{u}'} \zeta_{k+1}(\mathrm{x}',\mathrm{u}') = \sum_{\mathrm{x},\mathrm{u}} \Psi_k(\mathrm{x}'|\mathrm{x},\mathrm{u})\zeta_k(\mathrm{x},\mathrm{u}), \end{aligned} \tag{10.10}$$

のもとでの $\{\zeta_k(\mathrm{x},\mathrm{u})\}$ を決定変数とする

$$\sum_{k,\mathrm{x},\mathrm{u}} \ell(k,\mathrm{x},\mathrm{u})\zeta_k(\mathrm{x},\mathrm{u}) + \sum_{\mathrm{x}'} \ell_f(\mathrm{x}')\zeta_{\bar{k}}(\mathrm{x}') \tag{10.11}$$

の最大値が最適制御性能に一致する（$\mathbb{P}^{\pi^*}(x_k = \mathrm{x}, u_k = \mathrm{u})$ が最適解のひとつ）

ベルマン方程式と線形計画法 (4) マルコフ連鎖

ベクトル p の i 番目の要素を $(p)_i$，行列 P の (i,j) 成分を $(P)_{ij}$，j 列を $(P)_{:j}$ と表記

$P^\top \mathbb{1} = \mathbb{1}$ を満たす非負行列 $P \in \mathbb{R}^{n\times n}$ を**確率行列**（stochastic matrix）

定義 10.1.1 – マルコフ連鎖

遷移行列 P をもつ**マルコフ連鎖**（Markov chain）$x_:$

$$\mathbb{P}(x_{k+1} = i | x_k = j) = (P)_{ij},\ \forall i,j \in \lceil n \rceil \tag{10.16}$$

定理 10.1.2 – マルコフ連鎖の分布の時間発展

$(p_k)_i := \mathbb{P}(x_k = i)$ により定義される確率ベクトル $p_k \in \mathbb{R}^n$（和が 1 の非負ベクトル）

$$p_{k+1} = Pp_k,\ \forall k \in \mathbb{Z}_+ \tag{10.17}$$

カルバック・ライブラー制御 (1) 定式化

ベルマン方程式が線形化されるなど様々な興味深い性質をもつ制御問題

問題 10.1.3 – KL 制御

初期分布 p_0，遷移行列 P^π をもつマルコフ過程 p_k^π

状態ステージコスト $\ell \in \mathbb{R}^n$，割引率 $\beta \in (0,1)$, 遷移行列 P^0

$$\sum_{k=0}^{\infty} \beta^k \left\{ \ell^\top p_k^\pi + D_{\mathrm{KL}}(P^\pi p_k^\pi \parallel P^0 p_k^\pi) \right\} \tag{10.18}$$

を最小化する遷移行列 P^π を求めよ．

- 確率ベクトル p, q に対するカルバック・ライブラー距離 $D_{\mathrm{KL}}(p \parallel q) := \sum_{i=1}^n (p)_i \log \frac{(p)_i}{(q)_i}$
- 制御しない場合の遷移確率が P^0 であり，遷移行列を P^0 から P^π に整形することが制御
- KL 距離の項が一種の入力コスト

カルバック・ライブラー制御 (2) ベルマン方程式

定理 10.1.4 – KL 制御問題の解

KL 制御問題に対して，状態価値関数 V^* は

$$L_j^{\mathrm{KL}}(V) := (V)_j - (\ell)_j + \log\left(\sum_{i=1}^{n} \mathrm{e}^{-\beta(V)_i}(P^0)_{ij}\right) = 0,\ \forall j \in \lceil n \rceil \tag{10.20}$$

の一意解である．また，**コール・ホップ変換**（Cole-Hopf transformation）

$$(\mathsf{Z}^*)_j := \mathrm{e}^{-\beta(\mathsf{V}^*)_j},\ \forall j \tag{10.21}$$

により**適合関数**（desirability function）$\mathsf{Z}^* \in \mathbb{R}^n$ を定義すると，最適な制御則 P^π は

$$(P^*)_{ij} := \frac{(\mathsf{Z}^*)_i (P^0)_{ij}}{(\mathsf{Z}^*)^\top (P^0)_{:j}} \tag{10.22}$$

カルバック・ライブラー制御 (3) 線形可解性

- 適合関数 Z^* の値が大きいほど望ましい状態
- 最適制御則の式 (10.22) では，適合関数の値に比例して P^0 が $(P^*)_{ij} \propto (\mathsf{Z}^*)_i (P^0)_{ij},\ \forall j$ と調整
- 式(10.20)を Z^* に関して書き下すと，方程式

$$Z^{1/\beta} = \mathrm{diag}(\mathrm{e}^{-(\ell)_1}, \ldots, \mathrm{e}^{-(\ell)_n})(P^0)^\top Z \qquad (10.28)$$

（左辺の $1/\beta$ 乗は要素ごと）が得られ，とくに $\beta = 1$ のとき線形方程式に帰着する．こうした性質をもつ制御問題は**線形可解マルコフ決定過程**（linearly solvable MDP）とよばれ，容易に適合関数が求められるだけでなく，有限時間制御問題における終端コストの重ね合わせが可能といった有用な性質をもつ．

- 式 (10.22) は Z^* の定数倍に依存しない．これは，ステージコスト ℓ の全状態に一様な定数を加える自由度に対応する．

カルバック・ライブラー制御 (4) 凸性

log-sum-exp

$\mathbb{R}^n$ 上の実数値関数 **log-sum-exp** 関数

$$\mathrm{LSE}(\mathrm{x}) := \log\left(\sum_{j=1}^{n} \mathrm{e}^{(\mathrm{x})_j}\right) \in \mathbb{R} \tag{10.29}$$

は凸関数である．また，$\mathrm{LSE}(\mathrm{x}) = c + \mathrm{LSE}(\mathrm{x} - c\mathbb{1}),\ \forall c \in \mathbb{R}$ は指数関数を含む確率計算に伴う桁落ちを防ぐために用いられる．

- 式(10.28)を状態価値関数に関する方程式として書き下すと，

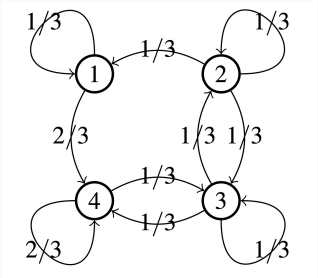
$$(V)_j = l_j - \log\left(\sum_{k=1}^{n} (P^0)_{kj} \exp(-\beta (V)_j)\right),\ \forall j \tag{10.27}$$

右辺は V に関して凸関数であり，最適化手法により方程式を解く場合に有用

カルバック・ライブラー制御 (5) 例

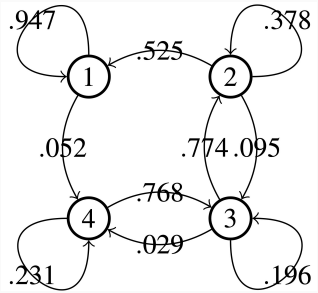
例 10.1.6 – KL 制御

ステージコスト $\ell = \sigma \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}$



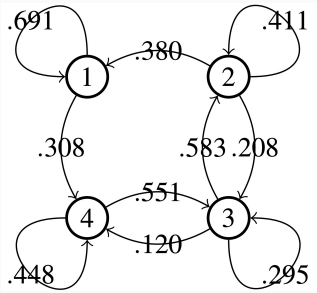
P_0

$\begin{bmatrix} .076 & .153 & .307 & .461 \end{bmatrix}^\top$



$\sigma = 1$

$\begin{bmatrix} .800 & .079 & .063 & .056 \end{bmatrix}^\top$



$\sigma = 0.5$

$\begin{bmatrix} .296 & .240 & .242 & .219 \end{bmatrix}^\top$

逆強化学習 (1) 定式化

最適制御問題：与えられた制御対象およびコスト関数に対して最適制御則を求める

制御対象および最適制御則が与えられたときに，対応するコスト関数を求める

例

LQR 問題において，フィードバックゲイン行列が与えられたときに，そのゲインが最適となるコスト関数の係数行列は?

逆強化学習（inverse reinforcement learning）：

制御対象および最適制御則により制御された状態の軌道データを用いて，対応するコスト関数を求める問題

逆強化学習 (2) 非決定性の要因

最適制御則 $u = \pi^*(\mathrm{x})$，状態価値関数 $\mathsf{V}^*(\mathrm{x})$，遷移確率密度関数 Ψ が既知の場合にステージコスト $\ell(\mathrm{x}, \mathrm{u})$ を求める．
無限区間のベルマン方程式を考えると，

$$\pi^*(\mathrm{x}) = \underset{\mathrm{u} \in U(\mathrm{x})}{\arg\min} \left\{ \ell(\mathrm{x}, \mathrm{u}) + \beta \sum_{\mathrm{x}' \in \lceil n \rceil} \mathsf{V}^*(\mathrm{x}') \Psi(\mathrm{x}' | \mathrm{x}, \mathrm{u}) \right\}$$

が成り立つような関数 ℓ を見つければ良いが，
これだけでは一意に定まらない．

最適制御問題

遷移確率密度関数 Ψ と初期状態 $x_0 \in \mathrm{rv}(\mathbb{X})$ をもつ確率システム，ステージコスト $\ell : \mathbb{X} \times \mathbb{U} \to \mathbb{R}_+$，入力許容集合 $U(\mathrm{x}) \subset \mathbb{U}$ および $\beta \in (0, 1)$ が与えられたとき，

$u_k \in U(x_k)$ を満たし，期待値

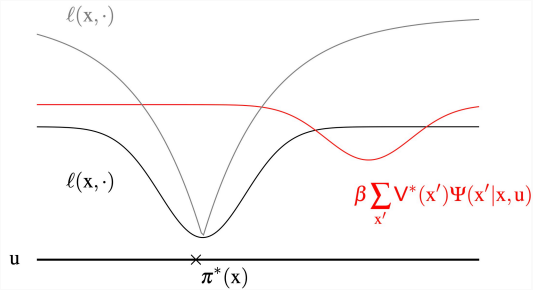
$$J(u_:) := \mathbb{E}^u \left[\sum_{k=0}^{\infty} \beta^k \ell(x_k, u_k) \right] \quad (10.28)$$

を最小化する因果的な制御入力 $u_:$ を求めよ

逆強化学習 (3) 確定方策の場合

最適制御入力値 $\pi^*(\mathrm{x})$ は $\ell(\mathrm{x},\cdot)$ に関して一点のみ（$\mathrm{u}=\pi^*(\mathrm{x})$）における値の情報しか与えない
$\Longrightarrow$ 最適制御則が確率的状態フィードバックで与えられる KL 制御問題を考える

- 状態コスト ℓ が未知であり，制御をしない場合の状態 j から i への遷移確率 $\Psi^0(i|j) := (P^0)_{ij}$ は既知，最適制御則のもとでの軌道データ $\mathrm{x}^*_{0:\bar{k}}$ が利用可能
- ステージコスト ℓ，適合関数 Z^*，状態価値関数 V^*，最適制御則 P^* のいずれかを求めればよい



同一の $\pi(\mathrm{x})$ となる評価関数

逆強化学習 (4) 最尤推定による解法

- 最適軌道 $x^*_{0:\bar{k}}$ の負の対数尤度は

$$
\begin{aligned}
-\log\varphi(\mathrm{x}^*_{0:\bar{k}}) &= -\log\left(\varphi(\mathrm{x}^*_0)\prod_{k=1}^{\bar{k}}(P^*)_{\mathrm{x}^*_k,\mathrm{x}^*_{k-1}}\right)\\
&= -\sum_{k=1}^{\bar{k}}\log(\mathsf{Z}^*)_{\mathrm{x}^*_k}+\sum_{k=1}^{\bar{k}}\log\sum_{i=1}^{n}(P^0)_{i,\mathrm{x}^*_{k-1}}(\mathsf{Z}^*)_i + \text{定数}
\end{aligned}
$$

- $a(i) := \#\{k \in \lceil\bar{k}\rceil : \mathrm{x}^*_k = i\}$, $b(j) := \#\{k \in \lfloor\bar{k}\rfloor : \mathrm{x}^*_k = j\}$ として書き換えると

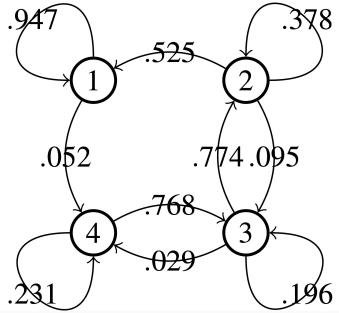
$$
L[\mathsf{V}^*] := \beta\sum_{i=1}^{n}a(i)(\mathsf{V}^*)_i + \sum_{j=1}^{n}b(j)\log\left(\sum_{i=1}^{n}(P^0)_{ij}\exp(-\beta(\mathsf{V}^*)_i)\right) \qquad (10.31)
$$

は V^* の要素について凸になり，効率よく最尤推定（$L[\mathsf{V}^*]$ の最小化）が計算できる．

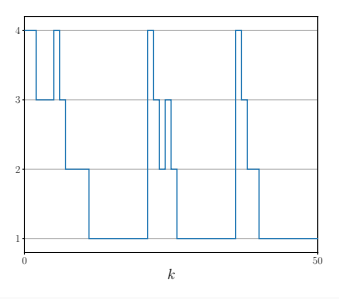
逆強化学習 (5) 例

例 10.1.6 – KL 制御と逆強化学習

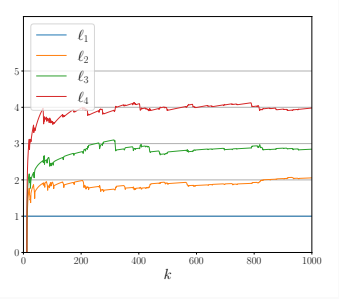
ステージコスト $\ell = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}$



遷移確率



P^* による標本経路（$\sigma = 1$）



ステージコストの推定値

確率システムの定常状態

マルコフ連鎖の収束 (1) 大域的な遷移に関する性質

対象を取りうる状態の数が有限なマルコフ連鎖に限定して分布の収束性を示す.

定義 10.2.1 – 既約性

非負行列 $A \in \mathbb{R}^{n \times n}$ が**既約**（irreducible）であるとは，任意の $i, j \in \lceil n \rceil$ に対して，ある $k_1(i,j) \in \mathbb{N}$ が存在し，$(A^{(k_1(i,j))})_{ij} > 0$ が成り立つことをいう.

定義 10.2.2 – 非周期性

遷移確率行列 $P \in \mathbb{R}^{n \times n}$ をもつマルコフ連鎖の状態 $i \in \lceil n \rceil$ について，ある $k_2(i) \in \mathbb{N}$ が存在して，任意の $k \geq k_2$ に対して $(P^k)_{ii} > 0$ となるとき，状態 i は**非周期的**（aperiodic）であるという.

マルコフ連鎖の収束 (2) 固有値に関する定理

定理 10.2.3 – ペロン・フロベニウス（Perron-Frobenius）の定理

非負正方行列 A に対して，

1. A は非負の実固有値をもち，そのうち最大のものを $\bar{\lambda}$ （A の**フロベニウス根**とよぶ）とおけば，$\bar{\lambda}$ に対応する非負固有ベクトルが存在し，
2. A の任意の固有値 λ は $|\lambda| \leq \bar{\lambda}$ を満たす．

また，A が既約な非負行列ならば，$\bar{\lambda}$ は単純固有値（重複度が 1）である．さらに A が正行列ならば，$\bar{\lambda}$ を除いた A の全ての固有値 λ は $|\lambda| < \bar{\lambda}$ を満たす．

定理 10.2.4 – 定常分布

確率行列 $P \in \mathbb{R}^{n \times n}$ のフロベニウス根は 1 である．

マルコフ連鎖の収束 (3) 定常分布への収束

定理 10.2.5 – マルコフ連鎖の収束

既約な遷移確率行列 $P \in \mathbb{R}^{n \times n}$ をもつマルコフ連鎖のある状態 $i \in [n]$ が非周期的

- $Pp_{\mathrm{st}} = p_{\mathrm{st}}$ をみたす正値の確率ベクトル p_{st} が存在
- 任意の初期分布 p_0 に対して，

$$\lim_{k \to \infty} P^k p_0 = P_{\mathrm{st}} \tag{10.34}$$

- p_{st} をマルコフ連鎖の**定常分布**（stationary distribution）もしくは**不変分布**（invariant distribution）とよぶ
- 固有値 1 に対応する P の非負固有ベクトル $v^\star \in \mathbb{R}^n$ をとり，

$$p_{\mathrm{st}} := \frac{v^\star}{\sum_{i=1}^n (v^\star)_i} \tag{10.33}$$

定常分布の設計 (1) 連続状態の場合

入力をもたない確定システム $x(k+1) = f(x(k))$ に対して $f(x_{\mathrm{eq}}) = x_{\mathrm{eq}}$ を満たす x_{eq} は平衡点とよばれ，$x(k) = x_{\mathrm{eq}}$ ならば $x(l) = x_{\mathrm{eq}},\ \forall l \geq k$

定義 10.2.6 – 定常分布

密度関数 Ψ をもつ $\mathbb{X}$-値マルコフ過程 $x_{:}$ に対して，$\mathbb{X}$ 上の確率密度関数 φ_{st} が

$$\varphi_{\mathrm{st}}(\mathrm{x}') = \int_{\mathbb{X}} \Psi(\mathrm{x}'|\mathrm{x})\varphi_{\mathrm{st}}(\mathrm{x})\mathrm{d}\mathrm{x} \tag{10.35}$$

を満たすとき，φ_{st} を x_k の**定常分布**（stationary –）とよぶ．

$\varphi_{x_k} = \varphi_{\mathrm{st}}$ ならば，$\varphi_{x_l} = \varphi_{\mathrm{st}},\ \forall l \geq k$ である（$x_l = x_k$ ではない）．

定常分布の設計 (2) 平衡分布

定義 10.2.7 – 詳細つりあい条件

遷移確率密度関数 Ψ を持つ $\mathbb{X}$ 値マルコフ過程 $x_:$ 対して，$\mathbb{X}$ 上の確率密度関数 φ_{eq} が

$$\Psi(\mathrm{x}'|\mathrm{x})\varphi_{\mathrm{eq}}(\mathrm{x}) = \Psi(\mathrm{x}|\mathrm{x}')\varphi_{\mathrm{eq}}(\mathrm{x}'),\ \mathrm{x} \neq \mathrm{x}' \tag{10.36}$$

を満たすとき，$x_:$ は**詳細つりあい条件**（detailed balance condition）を満たすといい，φ_{eq} を**平衡分布**（equilibrium distribution）とよぶ．

- 平衡分布は定常分布である．
- x から x' への移動量はその逆方向の移動量に等しいことを要求しており，見方を変えると時間的に可逆であることを意味している．

定常分布の設計 (3) 標本生成への応用

与えられた非負値関数 φ_{d} の規格化定数が未知の場合の標本生成方法

メトロポリス・ヘイスティングス（Metropolis-Hastings）法

非負値関数 $\psi_{\mathrm{d}} : \mathbb{X} \to \mathbb{R}_+$ および $q : \mathbb{X} \times \mathbb{X} \to \mathbb{R}_+$ が

$$c_{\mathrm{d}} := \int_{\mathbb{X}} \psi_{\mathrm{d}}(\mathrm{x})\mathrm{d}\mathrm{x} < \infty, \tag{10.37}$$

$$q(\mathrm{x}|\mathrm{x}') = q(\mathrm{x}'|\mathrm{x}), \ \int_{\mathbb{X}} q(\mathrm{x}'|\mathrm{x})\mathrm{d}\mathrm{x}' = 1 \tag{10.38}$$

を満たすとする．このとき，マルコフ過程 x_k を

1. x_k^+ は $\varphi_{x_k^+}(\mathrm{x}'|x_k = \mathrm{x}) = q(\mathrm{x}'|\mathrm{x})$，$r_k \sim \mathrm{Uni}([0,1])$ の独立な確率変数，
2. $r_k \leq \alpha_k := \min\{1, \psi_{\mathrm{d}}(x_k^+)/\psi_{\mathrm{d}}(x_k)\}$ ならば $x_{k+1} = x_k^+$，それ以外の場合は $x_{k+1} = x_k$

により定義する．このとき，$\varphi_{\mathrm{d}}(\mathrm{x}) := \psi_{\mathrm{d}}(\mathrm{x})/c_{\mathrm{d}}$ は x_k の平衡分布である．

定常分布の設計 (4) ポテンシャルゲーム

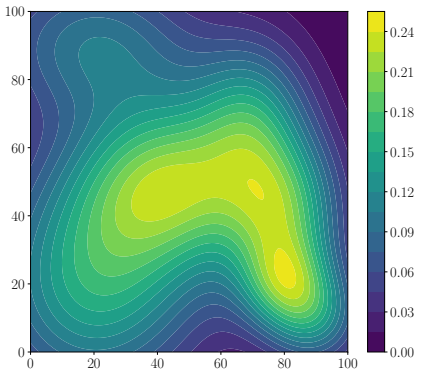
例 10.2.9 – センサーネットワーク配置問題

- $\mathsf{z} \in \mathsf{P} := \lceil 100 \rceil \times \lceil 100 \rceil$ において重要度 $\mathsf{R}(\mathsf{z})$
- 12 台のセンサー $j \in \lceil 12 \rceil$ を配置
- 各センサーは，自身を中心とした半径 10 の円内しか観測できない
- 配置 $\mathrm{x} = (\mathrm{x}^1, \ldots, \mathrm{x}^{12})$ に対してセンサー群全体で観測可能な領域

$$\mathsf{P}(\mathrm{x}) := \{\mathsf{z} \in \mathsf{P} : \exists j \text{ s.t. } \|\mathsf{z} - \mathrm{x}^j\| \leq 10\}$$

- 配置の評価値

$$\ell(\mathrm{x}) := \sum_{\mathsf{z} \in \mathsf{P}(\mathrm{x})} \mathsf{R}(\mathsf{z}) \qquad \textbf{(10.41)}$$

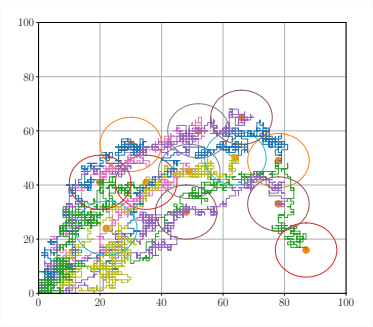


重要度 $\mathsf{R}(\mathsf{z})$

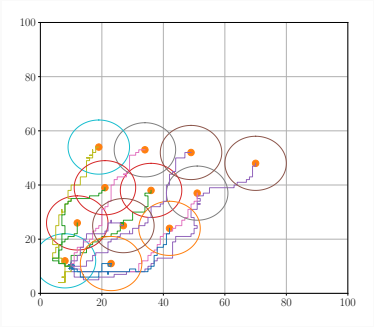
定常分布の設計 (5) ポテンシャルゲーム（つづき）

例 10.2.9 – センサーネットワーク配置問題（つづき）

- 位置を更新するセンサー $\iota \in [12]$，移動方向を上下左右から等確率で選択：$\hat{\mathrm{x}}^{\iota}$
- 確率 $\mathrm{e}^{\beta \ell(\mathrm{x})} / (\mathrm{e}^{\beta \ell(\mathrm{x})} + \mathrm{e}^{\beta \ell(\hat{\mathrm{x}}^{\iota})})$ で現在の位置 x にとどまるとすると，$\mathbb{P}_{\infty}(\mathrm{x}) \propto \mathrm{e}^{\beta \ell(\mathrm{x})}$



$\beta = 1$



確定的（評価値が上がる場合のみ更新）

第 10 章

1. 有限状態確率システム
    - ベルマン方程式と線形計画法
    - カルバック・ライブラー制御
    - 逆強化学習
2. 確率システムの定常状態
    - マルコフ連鎖の収束
    - 定常分布の設計